Maria Elisa Manetti

Caracterização de retrotransposons similares a Tnt1 em espécies silvestres e cultivadas do gênero *Solanum*

São Paulo

2007

Maria Elisa Manetti

Caracterização de retrotransposons similares a Tnt1 em espécies silvestres e cultivadas do gênero *Solanum*

Tese apresentada ao Instituto de Biociências da Universidade de São Paulo, para a obtenção de Título de Doutor em Ciências, na Área de Botânica.

Orientadora: Prof. Dra. Marie-Anne Van Sluys

São Paulo
2007

Manetti, Maria Elisa

Caracterização de retrotransposons similares a Tnt1 em espécies silvestres e cultivadas do gênero *Solanum*

111 páginas

Tese Doutorado - Instituto de Biociências da Universidade de São Paulo. Departamento de Botânica.

1. retrotransposon 2. *Solanum* 3. Tnt1

I. Universidade de São Paulo. Instituto de Biociências. Departamento de Botânica.

Comissão Julgadora:

_____ _____ _____
Prof(a). Dr(a). Prof(a). Dr(a).

_____ _____
Prof(a). Dr(a). Prof(a). Dr(a).

_____
Prof(a). Dr(a).
Orientador(a)

A mi madre Matilde

A mi familia, Jorge Francisco y Alicia

En memoria de mi querido padre, Adolfo

AGRADECIMENTOS

Agradeço a hospitalidade e a oportunidade que o Brasil me ofereceu. Senti-me em casa desde o começo e a bondade com que recebeu minha família foi muito profunda.

Agradeço á Prof. Dra. Marie-Anne Van Sluys pela oportunidade, orientação, amizade e apoio ao longo de todo o projeto.

Quero agradecer especialmente á Prof.Dra. Maria Magadalena Rossi, pelo apoio e ajuda imprescindível ao longo de todo o doutorado. Sua amizade e a de Gerson são muito importantes para minha família.

Quero agradecer aos colegas e amigos do Laboratório de Biologia Molecular de Plantas (GaTE lab): Alessandro, Ana Paula, Andrés, Breno, Bruno, Carlos, Claudia, Dani K, Dani Milstein, Douglas, Edgar, Érika, Fabiana, Guilherme, Hana, Jonas, Juliane, Leonor, Luísa, Marcelo, Marisa, Myna, Nilo, Regina, Silvia, Úrsula, Vanessa, Vivian, Wanessa. Eu vou ter muitas saudades de vocês...

Agradeço especialmente á Érika e Marisa pela sensibilidade, conversa, e amizade ao longo dos anos.

Ficou muito feliz de ter conhecido Luiza e compartilhado “muchos cafecitos”.

Um obrigado especial a Douglas, Marcelo e Vivi, disfrutei muito das conversas e obrigada pelo “proofreading” da tese.

A Silvia pela amizade e a ajuda no laboratório.

A todo o pessoal do Laboratório de Fisiologia Vegetal pela amizade e pelo apoio.

Este trabalho foi realizado com o apoio financiero da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior.

LISTA DE ABREVIAÇÔES

α: “alfa” (letra grega)

π: “pi” (letra grega)

BLAST: do inglês, *Basic Local Alignment Search TooL*

cDNA: DNA complementar

cult.: *cultivar*

CV: coeficiente de variação

DNA: ácido desoxirribonucléico

DNAse: desoxirribonuclease

dNTP: desoxinucleotídeo trifosfato

E: eficiência da amplificação por PCR em tempo real

EDTA: ácido etileno diamônio tetracético

EN: endonuclease

kb: 1000 bases

kbp: mil pares de bases

LINEs: do inglês *long interspersed nuclear elements*

LTRs: longas terminações repetidas (do inglês *long terminal repeats)*

MITE: do inglês *miniature inverted-repeat transposable elements*

Mb: milhões de bases

RNAm: RNA mensageiro

msnm: metros sobre nível do mar

ORF: quadro aberto de leitura (do inglês, *open reading frame*)

pb: pares de base

PCR: reação em cadeia da polimerase (do inglês, *polymerase chain reaction*)

PBS: do inglês, *primer binding site*

PPT: do inglês, *polypurine tract*

PROT: protease

R: do inglês, *repeated RNA*

RNA: ácido ribonucléico

RNAseH: ribonuclease H

rpm: rotações por minuto

RT: transcriptase reversa (do inlgês, reverse transcriptase)

S: desvio padrão

SINEs: do inglês, *short interspersed nuclear elements*

SNPs: do inglês, single nucleotide polymorphisms

subsp.: subespécie

TDS: do inglês, *target site duplication*

TIRs: repetições invertidas terminais (do inglês, *terminal inverted repeat*)

tRNA: RNA transportador

U3: do inglês, *unique 3' RNA*

U5: do inglês, *unique 5' RNA*

UV: luz ultravioleta

VLP: do inglês, *virus-like particle*

X-gal: 5-bromo-4-cloro-3-indolil-β-D-galactosídeo

XP: xeroderma pigmentosum

INDICE

I. -Introdução.....1
I.1- Descoberta dos elementos de transposição.....1
I.2- Elementos de transposição.....2
I.3- Ciclo de vida de um retrotransposon.....6
I.4- O impacto dos TEs na evolução dos genomas.....8
I.5- O retrotransposon Tnt1.....13
I.6- O gênero *Solanum*.....16
II.-Objetivos.....18
III.-Materiais e Métodos.....19
III. 1-Material botânico.....19
III. 2-Plasmídeos, iniciadores e bactérias utilizados.....21
III. 3-Soluções estoque e meios de cultura.....23
III. 3.1-Soluções estoque.....23
III. 3.2-Meios de cultura.....25
III. 3.2.1-Meios de cultura para plantas.....25
III. 3.2.1-Meios de cultura para bactérias.....26
III. 4-Cultivo das espécies de *Solanum* e extração de DNA.....26
III. 4.1-Cultivos das espécies de *Solanum*.....26
III. 4.2-Extração de DNA.....27
III. 4.3-Quantificação de DNA.....28
III. 5- Manipulações gerais com bactérias *Escherichia coli*.....28
III. 5.1-Obtenção de bactérias competentes.....28
III. 5.2-Transformação de bactérias competentes.....29
III. 5.3-Extração de DNA plasmidial em pequena escala (Miniprep)….....30
III. 5.4-Extração de DNA plasmidial em pequena escala em placa de 96 poços (Miniprep de placa).....31
III. 5.5-Extração de DNA plasmidial em média escala (Midiprep).....33
III. 6- Procedimentos gerais de manipulação de DNA.....33
III. 6.1-Amplificação de fragmento por PCR.....33

III. 6.2-Purificação de fragmento de DNA.....34
III. 6.3-Ligação de fragmento de DNA.....34
III. 6.4-Determinação da seqüência de nucleotídeos de uma molécula de DNA.....35
III. 6.5-Hidrólise de DNA com endonuclease de restrição.....36
III. 6.6-Transferência alcalina de DNA para membrana e Southern blot.....37
III. 6.7-Marcação radiativa de fragmento de DNA empregado como sonda.....37
III. 6.8-Hibridação molecular, lavagem e exposição de membrana.....38
III. 7-Análise da diversidade das seqüências.....39
III. 7.1-Divergência das seqüências.....39
III. 7.2-Análises das relações entre as seqüências usando métodos filogenéticos.....39
III. 7.3-Análises das das relações entre as seqüências usando o programa NETWORK.....40
III. 8-Determinação do número de cópias usando Quantificação Absoluta por qPCR.....41
III. 8.1-Normalização utilizando um gene de cópia única.....45
III. 9- Isolamento do elemento Retrosol.....46
IV –Resultados.....48
IV. 1- Obtenção e análise de seqüências similares a Tnt1.....48
IV.1.1-Amplificação e seqüenciamento de fragmentos similares a Tnt1 em espécies de S*olanum*.....48
IV.1.2- Análise da diversidade das seqüências.....49
IV. 1.2.1- Relações filogenéticas das seqüências similares a Tnt1.....53
IV. 1.2.2- Análise das regiões U3 de Retrosol.....57
IV. 1.2.3- Análise da região RNAseH com NETWORK.....59
IV. 1.2.4- Distribuição do retrotransposon Retrosol.....63
IV. 2- Determinação do número de cópias de retrotransposons similares a Tnt1 em diversas espécies de *Solanaceae*.....65
IV.2.1-Clonagem de XPB para normalizar a determinação do número de cópias por qPCR.....74
IV. 2.2- Distribuição do número de cópias por hibridação molecular.....78
IV. 3- Isolamento e caracterização do elemento completo similar

a Tnt1 de *Solanum oplocence*, Retrosol1 80
V. – Discussão 87
VI – Conclusões 98
VII – Resumo 99
VIII – *Abstract* 101
IX – Referências Bibliográficas 102
X – Anexos 111

I - INTRODUÇÃO

I.1 – Descoberta dos elementos de transposição

Bárbara McClintock descreveu a existência dos elementos de transposição (TEs) na década de 40, estudando grãos de milho variegados. Identificou dois loci dominantes que interagiam entre si denominados *Dissociator* (Ds) e *Activator* (Ac). Em 1948, fez a descoberta surpreendente de que Ds e Ac podiam se transpor, ou seja, mudar de posição no cromossomo. Os efeitos da transposição foram constatados através da mudança na coloração do pericarpo dos grãos de milho, e a quebra do braço curto do cromossomo 9. Analisando cruzamentos entre linhagems controladas, McClintock concluiu que Ac controla a transposição de Ds e que o movimento do último promovia a ruptura do cromossomo. Quando Ds se excisava daquela posição, o gene responsável pela coloração na camada de aleurona era liberado do efeito supressor, e voltava ao estado nativo permitindo a síntese do pigmento. O efeito de mosaico de alguns grãos deve-se à coexistência de células contendo o gene da coloração inativado pela inserção de Ds e células com o gene ativo (figura 1). Entre 1948 e 1950, McClintock desenvolveu uma teoria postulando que os elementos móveis (transponíveis) regulam os genes, inibindo ou modulando sua ação. Denominou Ds e Ac como "elementos controladores" para distingui-los dos genes. A descoberta de McClintock desafiou o conceito de que o genoma é um conjunto estático de informações transmitidas entre gerações. Em decorrência destas descobertas, Bárbara McClintock foi agraciada com o prêmio Nobel de Fisiologia e Medicina em 1983.

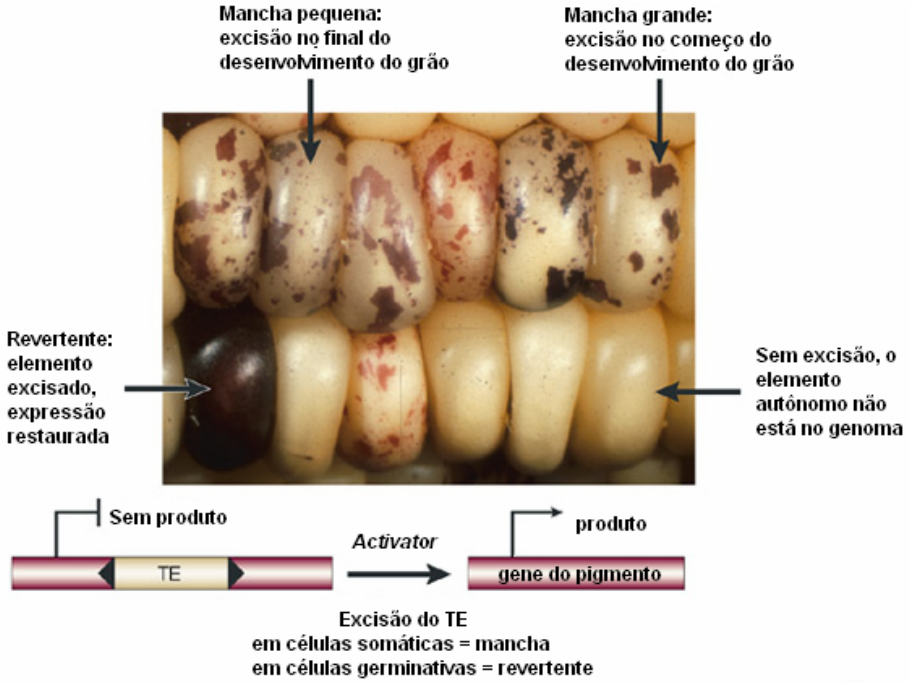

Figura 1. Diferentes fenótipos observados em grãos de milho devido à ação de TEs.
A instabilidade dos fenótipos se deve à inserção ou excisão do elemento em um gene que codifica uma enzima que participa da rota de biossíntese do pigmento antocianina. Setores revertentes (com pigmento) resultam da excisão do elemento em células da camada de aleurona. O tamanho do setor reflete o momento durante o desenvolvimento onde a excisão aconteceu. Adaptado de Feschotte *et al.* 2002.

I.2- Elementos de transposição

Atualmente, caracterizam-se os TEs como segmentos de DNA que possuem a capacidade de catalisar seu próprio movimento e inserir-se em novas regiões do genoma. Como conseqüência da sua atividade os TEs podem gerar alterações funcionais nos genes aos quais se associam. Quando inseridos dentro da região codificante de um determinado gene podem inativá-lo ou alterar seu produto. Quando inseridos dentro ou próximo a regiões promotoras podem alterar os padrões de expressão do gene (Kashkush *et al.* 2003). A atividade dos TEs pode ainda ser responsável pela mobilização de outras porções do DNA durante a transposição, ou de mecanismos de recombinação ilegítima. Já os

elementos replicativos apresentam como conseqüência do seu modo de transposição o aumento do tamanho do genoma hospedeiro (Kidwell, 1997; Fedoroff, 2000). O seqüenciamento de vários genomas evidenciou que os TEs são o componente majoritário dos genomas de eucariotos, compondo pelo menos 45% do genoma humano (Lander *et al*. 2001) e de 50 a 80% dos genomas em algumas gramíneas (San Miguel *et al*. 1998, Vicient *et al*. 1999).

Os TEs podem ser classificados em duas classes dependendo do intermediário de transposição, que pode ser uma molécula de RNA (classe I) ou de DNA (classe II) (Voytas *et al*. 1992, Feschotte *et al*. 2002). Os elementos da Classe I ou retroelementos possuem um mecanismo replicativo de transposição, que produz como resultado de sua mobilização um aumento do número de cópias e, concomitantemente, um aumento no tamanho do genoma hospedeiro (San Miguel and Bennetzen, 1998).

Os retroelementos podem ou não possuir longas terminações repetidas (LTRs). Nos retrotransposons com LTRs as mesmas localizam-se em suas extremidades na orientação direta. A LTR é subdividida nas regiões U3 (3' RNA), R (RNA repetido) e U5. (5' RNA) (figura 2). Os sinais regulatórios tais como o região promotora de transcrição, regiões de término de transcrição e os locais da poliadenilação estão localizados nas LTRs (Kumar and Bennetzen 1999). Além da importância funcional destas seqüências, diversos estudos relataram que as LTRs são as regiões que evoluem mais rapidamente nos retrotransposons (Kalmykova and Gvozdev 2004; Vicient *et al.* 2005). Os elementos com LTRs codificam uma proteína (*gag*) similar a um capsídeo viral e uma poliproteína (*pol*). Esta, por sua vez, quando processada resulta em várias proteínas maduras: a protease (PROT), que está envolvida no processo de maturação da poliproteína; a integrase (INT), necessária para a integração da cópia de DNA no genoma hospedeiro; e a proteína transcriptase

reversa/RNAseH (RT/RNAseH), responsável pela síntese do cDNA a partir do mRNA (figura 2). De acordo com a posição da integrase na poliproteína e o grau de semelhança entre as seqüências, os retrotransposons são divididos em duas famílias: Ty1-cópia e Ty3-gypsy (Boeke and Corces, 1989; Zachar *et al*. 1985) (figura 2).

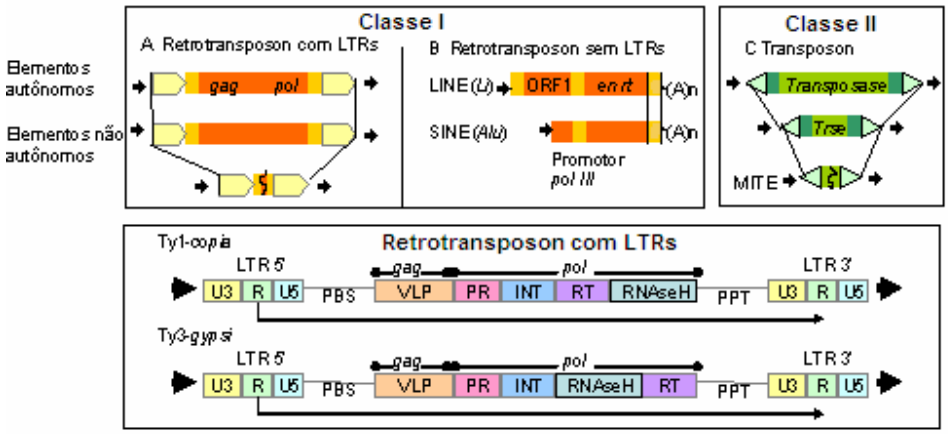

Figura 2. Estrutura geral dos Elementos de Transposição.

A Classe I ou retroelementos é composta por: retrotransposons com LTRs; sem LTRs; e retrovírus (não representado). Nos retrotransposons com LTRs (B), a orientação das mesmas é de maneira direta (triangulo amarelo claro). Elementos autônomos possuem dois genes: *gag* e *pol* (caixas laranja). A maioria dos elementos não autônomos não possui toda a região codificante. Os retrotransposons sem LTRs (B) são divididos em LINEs e SINEs. A região codificante inclui: ORF1 (similar a *gag*); *en* (endonuclease); e *rt* (reverse transcriptase) genes (caixas laranjas). A Classe II está composta por transposons de DNA com TIRs (terminal inverted repeats) (triângulo verde claro) e um gene (*transposase*) que codifica para uma transposase (caixa verde). Elementos não-autonomos são geralmente derivados de elementos autônomos por uma deleção. Caixa inferior: estrutura mais detalhada de um retrotransposon com LTR. Cada LTRs é dividida em três regiões: U3, R, e U5, que apresentam as sinais para a iniciação e terminação da transcrição. O transcrito (seta fina) é mostrado abaixo de cada bloco, começa no final 5' da região R da LTR'5 e termina na posição '3 da região R pertencente a LTR'3. Os genes do retrotransposon codificam varias proteínas: proteína do capsídeo viral (VLP), integrase (INT), protease (PR), transcriptase reversa (RT) e ribonuclease H (RNAase-H). Outras regiões características das seqüências são: PBS (*primer binding sites*), PPT (*polypurine tracts*). Pol III B sitio de reconhecimento da RNA polymerase III. Setas: direção do elemento. Adaptado de Kumar *et al*.1999 e Feschotte *et al.* 2002.

Os retrovírus têm estrutura similar a um retrotransposon com LTRs, com a presença adicional de uma ORF contendo o domínio *env*, que lhes confere o caráter infeccioso (Xiong and Eickbush, 1990). Já os retroelementos sem LTRs compreendem os LINEs (long interspersed nuclear elements) e SINEs (short interspersed nuclear elements). Os LINEs codificam duas ORFs, que são transcritas como um mRNA bicistrônico composto de uma ORF1 (proteína que se une ao RNA) e uma ORF2 (endonuclease e atividade de transcriptase reversa). Os LINEs e os SINEs terminam em uma seqüência simples repetitiva, geralmente uma cauda poli-A (figura 2). Os transcritos dos LINEs iniciam-se a partir de um promotor dentro da extremidade 5' do elemento e terminam freqüentemente junto à seqüência simples repetitiva. Os SINEs são caracterizados por um promotor interno de pol III de RNA próximo a extremidade 5′. Os SINEs são um grupo heterogêneo de elementos, que variam no comprimento de 90 a 300 pb e são derivados de uma variedade de genes de tRNA ou do RNA 7SL. Os SINEs são considerados elementos defectivos que parasitam a maquinaria de transposição dos LINEs (Ogiwara *et al*. 1999; Wessler 2006).

Muito pouco se conhece sobre a origem dos elementos da Classe I nas plantas. A ubiqüidade dos retrotransposons nas plantas e a sua heterogeneidade indicam que estiveram presentes nos primeiros vegetais, mas evoluíram variando seu número de cópias, localização, e estrutura das seqüências (Kumar and Bennetzen, 1999). Os TEs de Classe II, ou transposons clássicos, codificam geralmente para uma única proteína, a transposase, responsável pela excisão e integração do elemento. Possuem em suas extremidades repetições invertidas terminais (TIRs)(Feschotte *et al*. 2002) (figura 2).

Cada família de TEs contém elementos autônomos e não autônomos. Os elementos autônomos têm ORFs que codificam os produtos requeridos para a transposição. Já os

elementos não autônomos não codificam proteínas da transposição, mas podem se transpor porque retêm as seqüências *cis* necessárias para a transposição. A integração da maioria dos TEs resulta na duplicação de uma seqüência genômica curta (TSD: do inglês, *target site duplication*) no sitio de inserção (Wessler 2006) (figura 2).

I.3 – Ciclo de vida de um retrotransposon

O ciclo de vida do retrotransposon resulta na amplificação do genoma e por esta razão sua atividade é controlada finamente pelo hospedeiro. Além dos ciclos de amplificação, o tamanho do genoma do hospedeiro pode diminuir devido a eventos de recombinação homóloga dentro de um elemento ou entre elementos homólogos.

O processo de replicação de um retrotransposon é chamado retrotransposição porque o RNA é reverso-transcrito em DNA (oposto ao especificado pelo Dogma central da Biologia). Acredita-se que o ciclo de vida de um retrotransposon compartilha muitas características com o ciclo de um retrovírus tal como o da AIDS (Lentivirus). Ambos os ciclos de vida são mutagênicos e propensos a erros, já que a transcriptase reversa tem baixa atividade de processamento quanto à incorporação de bases errôneas não complementares à fita molde de RNA. Há também a possibilidade de recombinação durante a transcrição reversa. Uma vez que uma cópia replicada é introduzida no genoma as mutações neutras no decorrer do tempo irão gerar polimorfismos nas seqüências das LTRs, sinais de processamento e regiões codificantes, incluindo o surgimento de novos codons de parada de tradução (*stop codons*). Portanto, os retrovírus e os retrotransposons geram populações de seqüências similares denominadas *quasispecies* (Casacuberta *et al.* 1995).

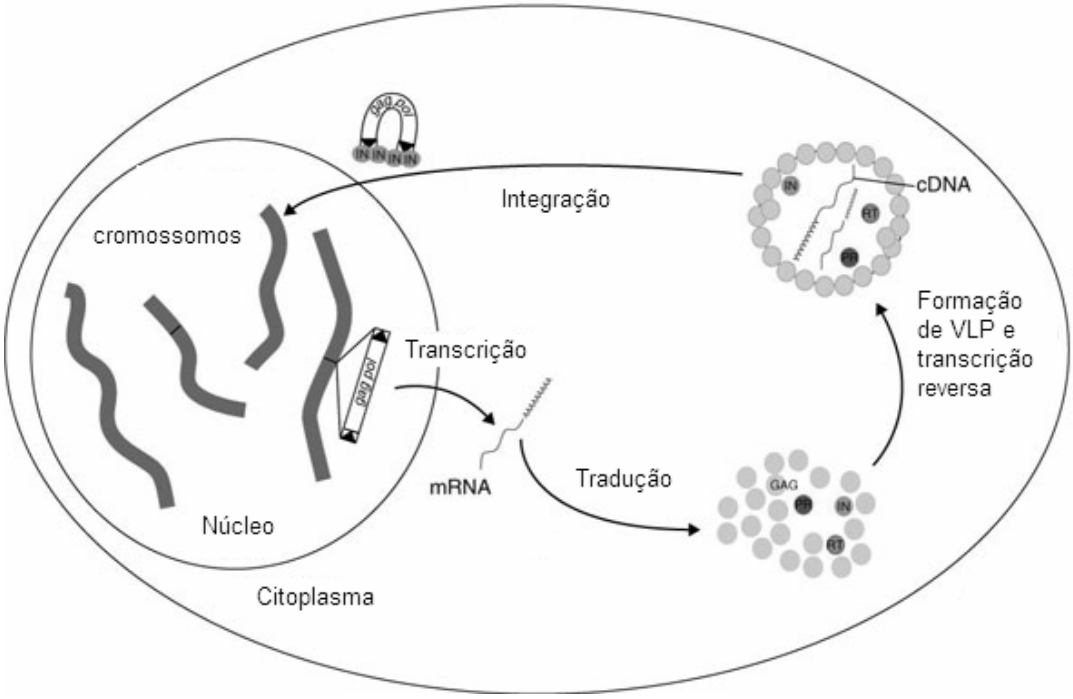

Figura 3. Ciclo de vida de um retrotransposon com LTRs.

IN: integrase; PR: protease; RT: transcriptase reversa; VLP: partícula tipo viral (do inglês *virus-like particle*). Triângulos pretos representam LTRs. Adaptado de Havecker *et al*. 2004.

O ciclo de vida de um elemento envolve primeiramente a transcrição do mesmo pela RNA polimerase II celular que reconhece o promotor localizado na região 5'LTR. O RNA é traduzido no citoplasma para formar as proteínas que formam as VLP (*virus-like particles*) e as que levam a cabo a transcrição reversa e integração da cópia de cDNA (Havecker *et al*. 2004) (figura 3). A transcrição reversa é um processo complexo e com muitas etapas que acontece dentro das VLPs no citoplasma (Kazazian 2004) (figura 4).

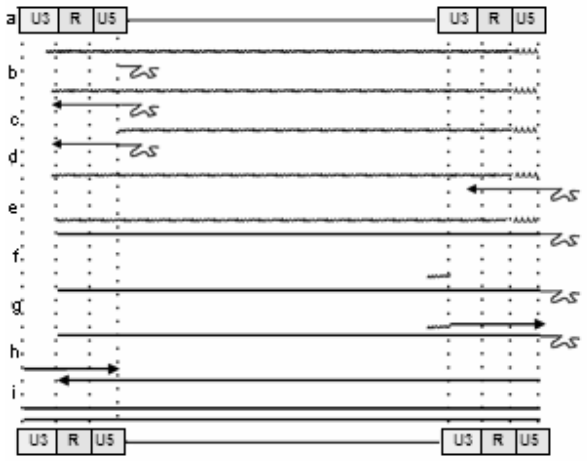

Figura 4. Transcrição reversa de transposons com LTRs.

A transcrição reversa começa com a cópia do RNA a DNA próximo do final 5' do RNA usando um tRNA como iniciador (a e b), seguido da degradação do RNA 5'(c). O novo DNA muda para a posição final 3' do RNA (d) e se completa a síntese da primeira fita de DNA (e). A RNAseH codificada pelo elemento degrada a maior parte do RNA (f). Uma pequena porção do RNA é utilizada como iniciador para a síntese da porção a direita do elemento (g). Uma segunda mudança de posição da segunda fita de DNA permite completar a síntese da segunda fita de DNA (i). Durante o processo as LTRs são formadas. Caixa colorida: LTRs e linhas verticais pontilhadas indicam regiões das mesmas (U3, R e U5); línea seguida de A: mRNA; línea irregular: tRNA; seta preta: direção da síntese. Adaptado de Kazazian 2004.

I.4 – O impacto dos TEs na evolução dos genomas

Por que os organismos mantêm uma alta proporção de TEs nos seus genomas? Eles cumprem alguma função? Como teriam influenciado a evolução da vida?

Vários processos participam da evolução dos genomas eucarióticos, tais como quebras cromossômicas e sua posterior translocação a outro cromossomo, duplicações de genes e segmentos, reordenação de domínios funcionais tais como os exons, criação de novos genes e conversão gênica. Durante as últimas décadas muitos destes processos foram

relatados como estando associados a TEs (Babushok *et al*. 2007; Cordaux *et al.* 2006; Deininger *et al*. 2003).

Desde o descobrimento dos TEs por Bárbara McClintock, muito tem se especulado e questionado sobre os TEs e seu possível papel na evolução dos genomas. McClintock chamou os TEs de "elementos controladores" ao observar mudanças nos fenótipos de grãos de milho devido à mobilização de TEs, propondo que controlariam o desenvolvimento normal do milho. Esta idéia foi rejeitada e anos depois ela propôs que os TEs seriam parte de uma resposta global do organismo ao estresse. Com a caracterização do primeiro TE em bactéria (Jordan *et al*. 1967) e o posterior reconhecimento dos mesmos como sendo componentes antigos dos genomas eucarióticos e procarióticos, especulações surgiram em relação ao papel dos TEs na evolução dos genomas.

Duas publicações (Doolittle and Sapienza 1980; Orgel and Criel 1980) postularam que a seleção natural, que opera dentro dos genomas, resultará inevitavelmente no surgimento de fragmentos de DNA sem nenhuma expressão fenotípica cuja única função é sua sobrevivência dentro dos genomas. Os TEs podem ser vistos como tais DNAs sem função fenotípica ou evolutiva. A partir destas informações surgiu a teoria dos TEs como parasitas (*selfish* DNA) ou DNA lixo (*junk* DNA) e a aceitação de que os TEs são componentes dos genomas sem um papel importante na evolução.

A teoria do DNA parasita foi amplamente aceita até a constatação de que os TEs geralmente perfazem a maior fração dos genomas dos organismos eucariontes. Estas revelações conduziram a novas questões: como os organismos e seus TEs coexistem? (Bowen and Jordan 2002).

O conteúdo de TEs varia de espécie para espécie tanto em relação às classes de TEs presente quanto quanto em relação à proporção no genoma. A levedura (*Saccharomyces*

cerevisiae) tem somente retrotransposon com LTRs (chamados elementos *Ty*), já nos genomas dos mamíferos predominam retrotransposons sem LTRs e os transposons da Classe II compõem menos de 5% da fração total de TEs. Os genomas das plantas com flores possuem uma grande variedade de ambas as classes de elementos, sendo os retrotransposons com LTRs a maior fração dos genomas caracterizados até agora. O potencial de amplificação dos retrotransposons com e sem LTRs foi primeiramente ilustrado em pesquisas com o gênero *Lilium.* As 14 espécies analisadas possuem genomas que variam entre 30.000-45.000 MB, cujo tamanho pode ser explicado por amplificações massivas de retrotransposons (Leeton *et al*. 1993). Diversos estudos mostraram que a amplificação de retrotransposon com LTRs esclarece a maior parte do paradoxo do número C entre importantes espécies de gramíneas. O paradoxo do valor C é a falta de correlação entre o tamanho do genoma e a complexidade alcançada pelo organismo. Na família das gramíneas a fração do genoma compreendida por retrotransposons com LTRs parece estar correlacionada com o tamanho de genoma: no arroz que apresenta o menor genoma (430 Mb), os retrotransposons com LTRs somam 15% do genoma (Jiang *et al.* 2001), por sua vez no genoma do milho (2.800 Mb) os retrotransposons representam 50-80 % (San Miguel and Bennetzen, 1998), e na cevada (4.800 Mb) estes constituem mais de 70 % do genoma (Vicient *et al*. 1999). A variação do conteúdo de TEs entre espécies pode ter influenciado tanto a qualidade quanto a quantidade da variação genômica, e como conseqüência a evolução de seus genomas. Dois estudos muito interessantes permitiram aprofundar o conhecimento na dinâmica dos genomas, particularmente nas gramíneas. Bennetzen e colaboradores (Science,1996) analisaram uma região de 280 kb em torno do gene de milho *adh1* e encontraram retrotransposons com LTRs uns dentro dos outros (*nested*). Posteriormente, conseguiram demonstrar que expansões dos retrotransposons com LTRs

dobraram o tamanho do genoma de milho dentro dos 6 milhões de anos depois da separação entre este e o sorgo. O componente temporal da sua análise foi possível já que as LTRs de um mesmo elemento são idênticas no momento da inserção e, quanto mais divergentes elas são, mais tempo decorrido desde a inserção do elemento. Outro estudo, realizado por Kalendar e colaboradores (2000), apresenta um exemplo da reestruturação do genoma mediada por TEs numa população natural de cevada (*Hordeum spontaneum)*. Neste caso, a reestruturação toma a forma de uma variação no tamanho do genoma intra-específico devido a amplificação do retrotransposon com LTRs BARE-1. O número de cópias do elemento BARE-1 varia entre populações geograficamente próximas sujeitas a diferentes níveis de estresse hídrico. A correlação entre o aumento do número de cópias do retrotransposon BARE-1, o tamanho do genoma e as condições climáticas, sugere a ativação do elemento como resposta adaptativa ao ambiente. Os resultados de ambos os estudos mostraram pela primeira vez *in vivo* ou em uma população natural que os TEs podiam reestruturar um genoma.

Outro exemplo do impacto dos TEs na evolução dos genomas são os elementos de transposição em miniatura (MITEs) em duas subespécies de arroz: *japonica* e *indica*. Os MITEs são uma classe especial de elementos da Classe II não-autônomos, encontrados em um número muito elevado de cópias nos genomas, preferencialmente localizados em regiões ricas em genes. Um elemento MITE denominado mPing é ativo no genoma das subespécies de arroz. A diferença no número estimado de cópias do elemento em *japonica* e *indica* é de 70 versus 11 cópias, respectivamente. Esta diferença no número sugere uma amplificação recente desta família de MITEs, talvez durante a domesticação e possivelmente associada a diferentes condições ambientais em que cada subespécie de arroz se submeteu (Jiang *et al*. 2003, Wessler 2006).

Apesar da constatação da grande quantidade de TEs nos genomas e os ciclos de expansão observados, pouca atividade foi experimentalmente demonstrada. Alguns retrotransposons com LTRs podem ser ativados transcricionalmente sob condições de estresse biótico e abiótico. O primeiro retrotransposon ativo encontrado em plantas, Tnt1, foi isolado de fumo através da seleção de inserções no gene da nitrato redutase (NR), préviamente clonado. A família Tnt1 é o objeto de estudo deste trabalho.

A transcrição de retrotransposons com LTRs não está relacionada necessariamente a inserções novas no genoma (transposição). O ciclo de vida de um retrotranspon com LTRs inclui quatro etapas: transcrição, tradução, transcrição reversa e integração do elemento. A regulação pode acontecer em qualquer uma das etapas e como resultado limitar a taxa de transposição. A atividade dos TEs é altamente regulada no nível transcricional de maneira a estabelecer um equilíbrio que garanta a presença do elemento sem inviabilizar o genoma hospedeiro (Kumar and Bennetzen, 1999).

A variação do conteúdo de TEs entre espécies pode ter influenciado tanto a qualidade quanto a quantidade da variação genômica, e portanto impactado a evolução dos genomas. Por exemplo, quase 50 % do genoma humano é derivado de elementos L1 (LINEs) e elementos *Alu* (SINEs). Os elementos *Alu* são os elementos mais prolíficos nos genomas dos primatas, acumulando aproximadamente 1,2 milhões de cópias durante 65 milhões de anos de evolução. Os elementos *Alu* são uma fonte de instabilidade genômica (Sen *et al*. 2006). Devido a sua abundância e semelhança de seqüências, eles podem servir de sítios de recombinação resultando em deleções, duplicações e translocações (Callinan *et al.*. 2005). Os elementos *Alu* podem interromper a edição (*splicing*) normal de um gene integrando-se dentro de introns; mudar o padrão de expressão gênica, inserindo-se em regiões promotoras ou silenciando um gene através da inserção dentro do mesmo

(Deininger and Batzer, 1999). Alguns elementos *Alu* se encontram em introns de genes gerando sítios para que ocorra um *splicing* alternativo. Especificamente uns 5% dos introns de *splicing* alternativo são derivados das seqüências de *Alu,* e até tornaram-se exons; um processo que foi chamado de exonização de *Alu* (Lev-Maor *et al*. 2003).

Existem estudos que demonstram a coevolução dos elementos com os genomas hospedeiros. Casacuberta E. (2005) demonstrou que os elementos *HeT-A* e *TART* de *Drosophila melanogaster* se integram exclusivamente nas extremidades dos cromossomos, e a transposição deles é essencial para manter a integridade dos telômeros e em conseqüência a sobrevivência da espécie. A domesticação de um TE é outro exemplo de coevolução. SETMAR, um gene específico dos primatas, originou-se há 40-58 milhões de anos atrás pela fusão do domínio *histone methyltransferase* SET, com uma transposase similar ao elemento *Mariner* (Cordaux *et al*. 2006).

Os exemplos anteriores mostram as variadas formas como os TEs influenciaram a evolução dos organismos, indicando que atuaram ativamente no desenvolvimento e complexidade dos genomas.

I.5 – O retrotransposon Tnt1

Tnt1 foi o primeiro elemento de transposição caracterizado na família *Solanaceae*, sendo um dos mais bem caracterizados retrotransposons em plantas. O retroelemento Tnt1 foi isolado do genoma de fumo (*Nicotiana tabacum*) após sua inserção no gene estrutural *nitrato redutase* (NR). Para selecionar um evento tão raro Grandbastien *et al*. (1989), utilizaram protoplastos a partir dos quais podem ser regeneradas plantas. Normalmente a transcrição de Tnt1 é baixa mas fortuitamente quando a planta é exposta ao extrato de

fungo (necessário para degradar a parede celular e gerar protoplastos) é ativada. Também foi demonstrado que o retrotransposon Tnt1 pode ser ativado por vários estresses bióticos e abióticos, incluindo estresse por injúria, agentes oxidantes, infecção de patógenos e elicitores microbianos (Pouteau *et al.* 1991).

Tnt1 é um membro típico da família *Ty1/copia* de *Saccharomyces cereviseae* e *D melanogaster*, mostrando grande similaridade tanto na organização quanto em extensiva similaridade em nível de aminoácidos com esse grupo. É encontrado em elevado número de cópias no genoma de fumo (>100), e possui 5,3 kpb, sendo flanqueado por pequenas duplicações de 5 pb em seu sítio de inserção. Contém duas LTRs terminadas pelas seqüências palindrômicas 5'- TG...CA -3', comuns à grande maioria dos elementos descritos, assim como um único quadro aberto de leitura (ORF) de 1.328 aminoácidos homólogos a dois domínios codificantes de retrovírus (gag-pol) (Grandbastien *et al.* 1989). Tnt1 também contém regiões não codificantes, adjacente às LTRs: o PBS (sítio de ligação para iniciador) na extremidade 5'; e o PPT (trato de polipurinas) na extremidade 3'. Tanto o PBS quanto o PPT correspondem a sítios importantes no processo de transposição dos retroelementos com LTRs.

A expressão do elemento Tnt1 é rigidamente regulada e não é detectada na maioria dos tecidos de fumo, excetuando as raízes (Pouteau *et al.* 1991). Porém sua expressão é altamente induzida em preparos frescos de protoplastos, derivados de tecidos de mesófilo foliar de fumo; e também pode ser induzida por fatores de origem microbianos e estresses abióticos relacionados com injúria celular. Análises de expressão transiente em protoplastos de fumo mostraram que na região U3 da LTR estão à maioria dos componentes em *cis* que atuam na ativação transcricional do elemento.

O retrotransposon Retrolyc1-1, isolado de *L. peruvianum*, possui uma organização similar aos retrotransposons do grupo *Ty1/copia* e contém regiões de homologia com o elemento Tnt1 sendo considerado um membro da família Tnt1. Em *L. esculentum*, foi detectada a presença de aproximadamente 20 cópias do elemento, sendo que em *L. peruvianum* o sinal de hibridação apresentou-se bem mais forte, indicando um maior número de cópias. Esse resultado sugere que a domesticação poderia ter levado a uma perda de número de cópias homólogas a Tnt1/Retrolyc1 em *L. esculentum*. Verificou-se também que os domínios internos do elemento Tnt1 (gag, prot, endo e RT) são homólogos ao elemento Retrolyc1 de *L. peruvianum*, com exceção da região U3 presente nas LTRs do elemento, região associada com a ativação transcricional do elemento.

O retrotransposon Tnt1 é um de poucos retrotransposons de plantas com atividade transposicional demonstrada (Grandbastien *et al.* 1989). As seqüências similares a Tnt1 foram detectadas em diversas espécies em *Solanaceae*. Pelo menos três grupos principais podem ser diferenciados baseando-se na divergência da região regulatória U3, originando os grupos Tnt1A, B e C (Vernhettes *et al.* 1998; Grandbastien *et al.* 2005) (figura 5). Os elementos Retrolyc1 detectados no tomate compartilham extensas similaridades de nucleotídeo com o elemento Tnt1, exceto na região U3 (Costa *et al*. 1999; Araújo *et al*. 2001) (figura 5). A subfamília Retrolyc1, por sua vez compreende dois grupos, Retrolyc1A e B, distinguindo-se também por sua região regulatória U3 (Araújo *et al.* 2001) (figura 5).

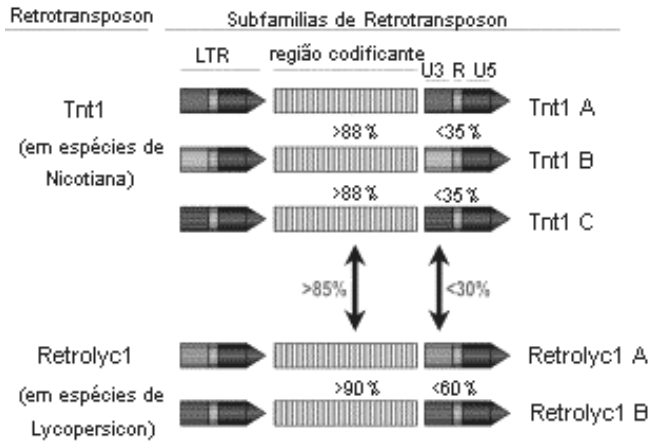

Figura 5. Comparação das regiões de Tnt1 e Retrolyc1.
Porcentagens de identidade entre as regiões U3 e regiões codificantes (baseadas no domínio da RT) entre os elementos das diferentes subfamílias Tnt1 e Retrolyc1. Adaptado de Casacuberta e Santiago 2003

I.6 - O gênero *Solanum*

A família *Solanaceae* compreende aproximadamente 90 gêneros distribuídos no mundo todo, sendo especialmente abundante nas Américas, incluindo entre 1400-1700 espécies (Mc Lean and Ivimey-Cook, 1956). Muitas espécies têm sido exploradas pelo homem desde a Antiguidade, muitas das quais foram domesticadas. Dentre as espécies com maior importância econômica, encontram-se a batata, o tomate e o fumo, que pertencem a gêneros diferentes: *Solanum*, *Lycopersicon* e *Nicotiana*, respectivamente. Com respeito ao gênero *Lycopersicon*, ainda existe controvérsia se as espécies que ele abrange pertenceriam ao gênero *Solanum,* especialmente na literatura referente ao melhoramento genético. A clasificação das espécies silvestres de tomate baseadas na morfologia as colocam dentro do um gênero deferente: *Lycopersicon* (Muller, 1940; Luckwill, 1943 e Child, 1990). Embora quando concetos biológicos das espécies são considerados *Lycopersicon* é uma subseção de *Solanum* (Rick, 1979) (para uma revisão ver Peralta e Spooner, 2000).

Dentro de *Solanum*, as batatas encontram-se no subgênero *Solanum*, seção *Petota* Dumort, subseção *Potatoe*. Todas as espécies desta seção são consideradas batatas por produzir tubérculos. Estão amplamente distribuídas nas Américas desde a região sudoeste

dos Estados Unidos (Nebraska e Colorado) até o extremo sul dos Andes. Segundo Hawkes (1990) existem 232 espécies dentro desta subseção, das quais 7 são cultivadas e 225 silvestres. As espécies cultivadas surgiram depois da domesticação, há 7000 anos atrás quando os índios começaram a coletar, armazenar e plantar os tubérculos (baseado nas análises de espécies da batata coletadas numa caverna no Perú).

Através de cruzamentos e hibridizações interespecíficas naturais entre as quatro espécies silvestres *S. acaule* (4x) *S. sparsipilum* (2x), *S. leptophyes* (2x) e *S. megistacrolubum* (2x) se originaram todas as espécies cultivadas. Elas variam em sua ploidia, resistência a fatores bióticos ou abióticos, morfologia e qualidade para consumo. Durante o cultivo da batata em tempos pré-históricos a espécie *Solanum tuberosum* (4x) se dispersou pelo mundo e transformou-se na quarta lavoura mundial, diferenciou-se em duas subespécies geograficamente muito distantes; *tuberosum* e *andígena*. A subespécie *andígena* se originou nos Andes peruanos e a subespécie *tuberosum* na ilha de Chiloé, Chile. A subespécie *tuberosum* teria originado os cultivares atualmente usados, após sua introdução na Europa (Ochoa, 1989).

As espécies silvestres possuem características de ervas daninhas adaptadas a condições ecológicas pouco favoráveis para o crescimento. Apresentam estolões, alto conteúdo de alcalóides, sementes e tubérculos adaptados a longos períodos de secas e geadas (Hawkes, 1999).

Outra subseção da seção *Petota*, *Etuberosum*, contém uma espécie, *Solanum brevidens*, que não produz tubérculos, mas que é utilizada em programas de melhoramento genético por possuir resistência a nemátóides e ser de fácil cruzamento com as batatas.

II – OBJETIVOS

Este trabalho tem por objetivo geral identificar e caracterizar retrotransposons similares a Tnt1 em espécies silvestres e cultivadas do gênero *Solanum.*

Os objetivos especificos são:

1. Obtenção e análise de seqüências similares a Tnt1.
2. Determinação do número de cópias de retrotransposons similares a Tnt1 em diversas espécies de *Solanaceae*.
3. Isolamento e caracterização do elemento completo similar a Tnt1 de *Solanum oploscence*, Retrosol1.

III - MATERIAIS E MÉTODOS

III. 1- Material botânico

A coleção de espécies vegetais do Banco de Germoplasma do INTA-Balcarce (Argentina), pertencentes ao gênero *Solanum,* foi o principal objeto de estudo. Todas as espécies exceto uma, *Solanum brevidens*, são consideradas batatas por produzir tubèrculos. As espécies silvestres foram crescidas a partir de sementes e os cultivares Chacarera 109, Moradita 64 e Tuni 207 de minitubérculos. Detalhes das espécies e cultivares coletados na Argentina e na Bolívia estão apresentados na tabela 1, e suas respectivas localizações geográficas no tabela 2. Os acessos usados representam amostras da diversidade e localização geográfica de cada espécie nos correspondentes centro de dispersão.

Além do material vegetal do Banco de Germoplasma, material vegetal fresco dos cultivares Kennebec e Huinkul foi obtido de clones *in vitro* gentilmente cedidos pelo INTA-Castelar (Argentina). Outras espécies foram escolhidas para diversos estudos: *Lycopersicon peruvianum* Mill., *Nicotiana tabacum* L. (fumo) e *Oryza sativa* (arroz).

Todas as espécies e cultivares foram crescidos e mantidos no *Gate lab* (Genomics and Transposable Elements), Instituto de Biociências, Departamento de Botânica, Universidade de São Paulo.

Tabela 1. Lista de espécies de *Solanum* utilizadas.

Acesso	Subgênero	Seção	Serie	Espécies	Cód.	Pl.
Bal20002	Potatoe	Etuberosum		*Solanum brevidens*	bre	2x
Bal80009		Petote	Commersoniana	*S.commersonii* subp.*malmeanum*	cmm	2x
Bal01280			Yungasensia	*S.chacoense*	chc	2x
Bal03017				*S.tarijense*	tar	2x
Bal01113			Megistracroloba	*S.megistacrolobum*	mga	2x
Bal74326				*S.sanctae-rosae*	sat	2x
Bal01156			Cuneolate	*S.infundibuliforme*	ifd	2x
Bal86017			Tuberosa	*S.xdodsssii*	dds	2x
Bal87066				*S.leptophyes*	lph	2x
Bal90081				*S.microdontum*	mcd	2x
Bal01121				*S.oplocense*	opl	4x
Bal86053				*S.xsucrense*	scr	4x
Bal01007				*S.gourlayi*	gou	2x
BAl83042				*S.hannemanii*	ha	2x
BAl73042				*S.incamayoense*	inc	2x
Bal90044				*S.kurtzianum*	ktz	2x
Bal74282				*S.okadae*	oka	2x
Bal03013				*S.spegazzinii*	spg	2x
Bal74330				*S.vernei*	vrn	2x
				S.tuberosum		
				cult. Kennebec	Ke	4x
				cult. Huinkul	Hu	4x
				*S.t.*subp. *andigenum*		4x
				cult. Chacarera 109	Cha	4x
				cult. Moradita 64	Mo	4x
				cult. Tuni 207	Tuni	4x
Bal01109			Acaulia	*S.acaule*	acl	4x

Cód: código. P.: ploidia. cult.: cultivar.

Tabela 2. Lista das localidades geográficas de cada acesso.

Bal N°	N°donante	Pais	Provincia	Latitude	Longitude	Alt. msnm	Código
Bal 01109	CCS 1263	Argentina	Jujuy	22°26'S	66°09'O	3818	acl
Bal 20002	CI 1096	Argentina	Neuquen			1150	bre
Bal 01280	CS 1435	Argentina	Jujuy				chc
Bal 8009	OCI 7263	Argentina	Entre Rios	31°53'S	58°14'O	40	cmm
Bal 86017	HAO 27	Bolívia		18°10'S	65°10'O	2150	dds
Bal 01007	CCS 1162	Argentina	Jujuy	23°25'S	65°22'O	2845	gou
Bal 83042	Oka 7531	Argentina	Salta	24°37'S	65°35'O	3350	ha
Bal 73042	OL 4852	Argentina	Salta	24°45'S	65°44'O	2280	inc
Bal 01156	CCS 1311	Argentina	Jujuy	22°05'S	66°15'O	3700	ifd
Bal 90044	SCI 4568	Argentina	La Rioja	29°15'S	67°39'O	1800	ktz
Bal 87066	HOHL 242	Argentina		18°01'S	66°22'O	3700	lph
Bal 01113	CSC 1268	Argentina	Jujuy	22°16'S	66°01'O	3722	mga
Bal 90081	SCI 4630	Argentina	Catamarca	27°19'S	66°02'O	2310	mcd
Bal 74282	Oka 5645	Argentina	Salta	25°09'S	65°50'O	2740	oka
Bal 01121	CCS 1276	Argentina	Jujuy	21°52'S	66°11'O	3604	opl
Bal 74326	Oka 5697	Argentina	Tucumàn	26°46'S	65°45'O	2700	sat
Bal 03013	CIE 1581	Argentina	Salta	25°10'S	65°50'O	3005	spg
Bal 86053	HAO 64	Bolívia		19°11'S	65°10'O	2820	scr
Bal 03017	CIE 1585	Argentina	Salta	25°10'S	65°46'O	2139	tar
Bal 74330	Oka 5702	Argentina	Tucumàn	26°46'S	65°46'O	2900	vrn

Código corresponde às abreviaturas utilizadas para indicar cada genótipo. Alt.: altitude. msnm: metros sobre o nível do mar.

III. 2- Plasmídeos, iniciadores e bactérias utilizados

Plasmídeos

PGEM: pGEM®-T Easy Vector System (Promega Cat#A1360), vetor para clonagem de produtos amplificados em reação de PCR. Possui 3,0 kpb e confere resistência a ampicilina.

Iniciadores

Iniciadores foram amplamente utilizados para amplificação por PCR e/ou amplificação por qPCR. Os niciadores específicos encontram-se listados a seguir (tabela 3) e sua localização respeito a Tnt1-94 e Retrosol1 (anexo 2) na figura 6. Para seqüenciamento de insertos clonados no vetor pGEM®-T, foram utilizados os iniciadores T7 e SP6, e para insertor clonados no vetor pMOS*Blue* T7 e U19 (de acordo com as recomendações dos fabricantes). Todos os iniciadores foram deshenados no presente trabalho, com exceção de Avi/Ol16 utilizados por Araujo e colaboradores (2001).

Tabela 3. Iniciadores específicos utilizados e localização respeito a Tnt1 e Retrosol.

Nome	Seqüência (5'→ 3')	Proc.	PosiçãoTnt1 pb	Posição Retrosol pb
Avi	GCTGACCAAGGTGGTAC	P	4609-4625	4527-4543
OL16	TTCCCACCTCACTACAATATCGC	P	5041-5064	4945
PPT	CAATCTCCCCCTCCAG	P/Q	4710-4726	4630- 4645
GI-F	CCATCCCCATGTGCC	P	-	4651- 4665
GI-R	GAGAACAAATGCAGTAGC	P	-	Não se encontra
GII-F	CCATCCCCATTTGCC	P	-	4651- 4665
GII-R	GATAACAAGTGCAG	P	-	4795- 4809
XPB-F	GTGAGAGGGAAGGGAAGAAGA	P/Q	-	-
XPB-R	TCTACCATCTGCGCAAGCCC	P/Q	-	-
LTR-F	GGGTCTGGAGGGGAAGATTGT	Q	-	4626- 4646
LTR-R	CTGCCAACCTTTTGACATTGTGA	Q	-	4669- 4691

Proc.: procedimentos, P: PCR, Q: qPCR, S: sequenciamento.

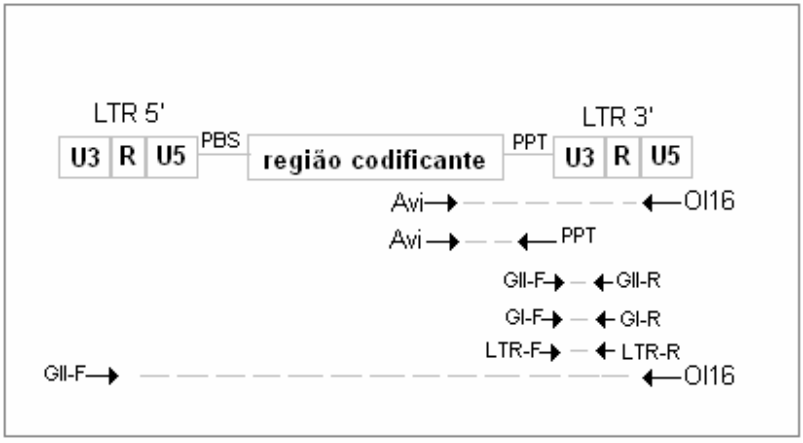

Figura 6. Esquema dos iniciadores utilizados e regiões amplificadas comparadas com um retrotransposon com LTR.

Os iniciadores estão indicados com setas e as regiões amplificadas, por lnhass descontínuais. LTR: longas repetições terminais. PBS: primer binding site. PPT: *poly purine track*. U3: *unique 3'RNA*, R: *repeated RNA*, U5: *unique 5'RNA*.

Bactéria

Escherichia coli cepa: DH10B – genótipo F- mcrA Δ(mrr-hsdRMS-mcrBC) Φ80d lacZ ΔMIS ΔlacX74 endA1recA1 deoRΔ (ara,leu) 7697 araD139 galU galK nupG rpsL.

A cepa foi mantida em placas com meio sólido LB e repicadas mensalmente. Para conservar a cepa por perídos de tempo longo (até 1 ano), foram feitos estoques de glicerol obtidos adicionando-se 1mL de meio LB com 40% de glicerol a 1mL de cultura bacteriana em LB.

III. 3- Soluções estoque e meios de cultura

III. 3.1- Soluções estoque

Dodecil sulfato de sódio – SDS 20%: O SDS (200g/L) deve ser dissolvido em água destilada ou ultra filtrada previamente esterilizada em autoclave (120ºC, 30 minutos).

Tampão TBE 10X : Tris (0,9M), Ácido bórico (0,9M) e EDTA (20mM), dissolvidos em água destilada ou ultra filtrada. Esterilizado em autoclave (120ºC, 15 minutos).

Brometo de etídeo (10mg/mL): Preparado em água destilada previamente esterilizada em autoclave (120ºC, 30 minutos).

RNAseA(10mg/mL) : A ribonuclease A é dissolvida em uma solução de Tris pH7,5 (10mM) e Cloreto de sódio (15mM), utilizando as soluções estoque citadas. Em seguida, é incubada a 100ºC por 15 minutos e mantida à temperatura ambiente até esfriar. Deve ser estocada a -20ºC.

SSC 20X: Cloreto de sódio (3M), dissolvidos em água destilada ou ultrafiltrada. Esterilizado em autoclave (120ºC, 15 minutos).

5-bromo-4-cloro-3-indolil- β-D-galactosídeo - X-gal (50mg/mL): Solução preparada em dimetilformamida armazenada a -20ºC e protegida da luz.

Hidroxido de sódio – NaOH 10N: As pastilhas de hidróxido de sódio (400g/L) são dissolvidos em água destilada previamente esterilizada em autoclave (120ºC, 30 minutos).

Àcido etileno diamono tetracético – EDTA 0,5M (pH8,0): O EDTA (186,1g/L) é dissolvido em água ultrafiltrada e utilizam-se pastilhas de hidróxido de sódio para acertar o pH. Esterilizado em autoclave (120ºC, 15 minutos).

Tris 1M (pH7,5 ou 8,0): O Tris (121,1g/L) é dissolvido em água ultrafiltrada e utiliza-se ácido clorídrico fumante para acertar o pH. Esterilizado em autoclave (120ºC, 15 minutos).

Cloreto de sódio – NaCl 5M: O cloreto e sódio (292,2g/L) é dissolvido em água ultra filtrada e esterilizado em autoclave (120ºC, 15 minutos).

Cloreto de magnésio – $MgCl_2$ 1M: O cloreto de magnésio 6 H_2O (203,3g/L) é dissolvido em água ultra filtrada e esterilizado em autoclave (120ºC, 15 minutos).

Tampão TE 1X: Tris pH8,0 (10mM) e EDTA (1mM). Preparado a partir das soluções estoque de Tris e EDTA citadas, utilizando-se água ultra filtrada previamente esterilizada em autoclave (120ºC, 30 minutos).

Acetato de potássio 3M: O acetato de potássio (294,2g/L) e Ácido acético glacial (2M), são dissolvidos em água ultra filtrada. Esterilizado em autoclave (120ºC, 15 minutos).

Antibióticos:

Ampicilina (100mg/mL): Solução preparada em água ultra filtrada, esterilizada em filtro Millipore (malha de 0,45µm) e armazenada a -20°C. A concentração empregada no meio de cultura para a seleção de bactérias contendo plasmídeo de interesse foi de 100µg/mL.

III. 3.2- Meios de cultura

III. 3.2.1- Meios de cultura para plantas

MS20 (pH5,7): Sais MS(*), Sigma M-0153 (2,18g/L) Sacarose (20g/L), Vitaminas de Nitch (2mL/L) e Phytagar, Gibco BRL (6,5g/L). Preparado com água destilada ou ultra filtrada e esterilizado em autoclave (120ºC, 15 minutos).

Vitaminas de Nitch: Glicina (1mg/mL), Mioinositol (50mg/mL), Ácido nicotínico (2,5mg/mL0, Piridoxina (0,25mg/mL), Tiamina (0,25mg/mL), Biotina(0,025mg/mL), e Ácido fólico (0,25mg/mL). A solução é preparada em água destilada ou ultra filtrada previamente esterilizada em autoclave (120ºC, 30 minutos) e armazenada a -20ºC.

(*) Trata-se dos sais do meio MS descrito por Murashige e Skoog (1962), com redução na concentração dos macronutrientes (metade da concentração original).

III. 3.2.2- Meios de cultura para bactérias

LB (pH7,5): Triptona (1 %), Extrato de levedura (0,5%) e Cloreto de sódio (1 %). Para preparar meio sólido são adicionados 20g/L de agar.

SOB (pH6,8): Triptona (2%), Extrato de levedura (0,5%), Cloreto de sódio (10mM), Cloreto de potássio (2,5mM), Cloreto de magnésio (10mM) e Sulfato de magnésio (10mM).

SOC: Trata-se do meio SOB acrescido de 20mM de glicose.

Os meios de cultura são preparados com água destilada ou ultra filtrada e esterilizados em autoclave (120ºC, 15 minutos).

III. 4- Cultivo das espécies de *Solanum* e extração de DNA

III. 4.1- Cultivo das espécies de *Solanum*

As sementes foram primeiramente desinfetadas com hipoclorito de sódio al 2% durante 15 minutos. Posteriormente foram enxaguadas com água estéril e logo colocadas em acido giberélico 0,1% durante 24 horas no escuro. Concluído o tratamento com ácido giberélico foram semeadas em placas de petri com algodão previamente esterilizado em autoclave (120ºC, 30 minutos) e molhadas com água destilada previamente esterilizada em autoclave (120ºC, 30 minutos). Após a germinação as plântulas foram transferidas para recipientes Magentas® com orifícios, em meio de cultivo MS20. O cultivo foi mantido em uma câmara nas condições de temperatura de 22,5 ºC ao dia e 18,5 ºC à noite, com um fotoperíodo de 16 h. Foram estabelecidos clones para os posteriores experimentos.

Os mini tubérculos foram plantados em terra com vermiculita e areia, numa proporção 1:1:1, mescla que foi esterilizada em autoclave (120ºC, 60 minutos). O cultivo

foi mantido em uma sala de crescimento nas condições de temperatura mínima 21ºC e máxima de 24ºC, sob fotoperíodo de 16 h. Depois de crescidas as plantas, pequenos brotos foram cortados é lavados em água ultra filtrada previamente esterilizada em autoclave (120ºC, 30 minutos). Os brotos foram cultivados em Magentas® com meio MS20 nas mesmas condições das plantas provenientes de sementes.

A manipulação foi realizada em câmara de fluxo de ar laminar, com uso de material estéril.

III. 4.2- Extração de DNA

A extração do DNA genômico foi realizada segundo a metodologia descrita por Fulton *et al.* (1995). O material vegetal foi pesado e 100mg de folhas foram maceradas com 200µL de Tampão Microprep em um tubo de 1,5mL, posteriormente foram adicionados mais 550µL do Tampão. Após homogeneização delicada do material, este foi mantido em banho-maria a 65°C por um período de 120 min.

Completou-se o tubo com clorofórmio:isoamílico (24:1) misturando bem. A seguir se centrifugaram a 10.000rpm por 5minutos (centrífuga refrigerada Hettich Micro 22R). A fase aquosa foi transferida para um novo tubo de 1,5mL e adicionaram-se ¾ a 1 o volume de isopropanol. Os tubos foram invertidos repetidamente até o DNA precipitar. Imediatamente os tubos foram centrifugados a 10.000rpm por 5minutos. O isopropanol foi descartado e o precipitado de DNA lavado com etanol 70%. Depois de secar o precipitado foi ressuspendido em 50µL de tampão TE.

Com o intuito de degradar o RNA extraído junto com o DNA, as amostras foram incubadas com 2µL de RNAse (10mg/mL) por 60 minutos à 37°C.

Tampão de Preparação: Tampão de extração (41,6%v/v), Tampão de lise (41,6%v/v), Sarcosil 5% (16,6%v/v) e Bissulfito de sódio (3,4mg/mL). O preparo é realizado no momento da extração, a partir das soluções citadas, que são previamente confeccionadas.

-Tampão de Extração (pH7,5): Sorbitol (0,35M), Tris pH8,0 (0,1M) e EDTA (5mM).

-Tampão de Lise: Tris pH8,0 (0,2M), EDTA (0,05M), NaCl (2M) e Brometo de cetiltrimetilamônio – CTAB (2%).

A solução de Sarcosil e os tampões de extração e lise são feitos com água ultrafiltrada previamente esterilizada em autoclave (120ºC, 30 minutos).

III. 4.3- Quantificação de DNA

A quantificação do DNA foi feita no fluorômetro HoeferDyNA-Quant de acordo com as especificações do fabricante, e com o *Spectrophotometer* Nanodrop® ND-100 segundo especificações do fabricante.

Solução de ensaio para o uso do fluorômetro: 10µL de H33258 (10mg/µL), TNE10X 10mL, água destilada até um volume de 100mL.

TNE 10X: Tris 100mM, EDTA $Na_2.2H_2O$ 10mM, NaCl 2M.

III. 5- Manipulações gerais com bactérias *Escherichia coli*

III. 5.1- Obtenção de bactérias competentes

Para a obtenção de bactérias competentes, um inóculo de colônia isolada de DH10B foi colocado para crescer em 2mL de meio SOB líquido e a cultura foi mantida sob agitação

durante 18 horas a 37ºC. Após este período, acrescentaram-se mais 50mL de SOB e realizou-se nova incubação a 37 ºC por uma hora até que medida de absorbância em 600nm tivesse alcançado valores em torno de 0,6.

Em seguida, foram adicionados 0,5mL de $MgCl_2$ (1M) mantendo-se no gelo por 15 minutos. Centrifugo-se 3.000rpm por 20 minutos a 4 ºC (Centrifuga refrigerada Hettich Universal 30RF, rotor angular 1414). O precipitado foi ressuspendido em 15mL de solução RFI. Passados 10 minutos no gelo, nova centrifugação foi realizada e as bactérias foram ressuspendidas em 2mL de solução RFII. Este volume foi dividido em alíquotas de 50µL, armazenadas -70ºC.

Toda a manipulação foi realizada em câmara de fluxo de ar laminar, com uso de material estéril.

RFI: Acetato de potássio (30mM), cloreto de manganês (50mM), Cloreto de potássio (100mM), Cloreto de cálcio (10mM) e Glicerol (15%).

RFII: MOPs (3-(N-morpholino)propanesulfonic acid) pH7,0 (10mM), Cloreto de cálcio (75mM), Cloreto de potássio (10mM) e Glicerol (15%).

As soluções de $MgCl_2$, RFI e RFII são preparadas com água ultrafiltrada e esterilizadas em autoclave (120ºC, 15 minutos).

III. 5.2 - Transformação de bactérias competentes

Para a transformação de bactérias competentes foi utilizado o método de choque térmico, baseado em Hanahan (1983). Bactérias competentes foram retiradas de estoque a -70ºC e mantidas em gelo por 15 minutos. Adicionou-se o plasmídeo desejado ou 5µL da ligação (item III.6.3) a um volume de 50µL de bactérias. Os tubos permaneceram no gelo

durante 30 minutos e, então, foram submetidos a um banho a 42ºC por exatamente 45 segundos e imediatamente retornaram ao gelo por mais 3 minutos.

Adicionaram-se 900µL de SOC em cada tubo, seguindo-se uma incubação de 40 minutos sob agitação a 37ºC. A cultura foi inoculada em meio LB sólido com 40µg/mL de X-Gal (5-bromo-4-chloro-3-indolyl-*β*-D-galactoside), na presença do antibiótico adequado, permanecendo cerca de 16 horas a 37ºC para multiplicação das bactérias transformadas.

Toda a manipulação foi realizada em câmara de fluxo de ar laminar, com uso de material estéril.

III. 5.3- Extração de DNA plasmidial em pequena escala (Miniprep)

Após a seleção das bactérias transformadas prosseguiu-se com a extração do DNA plasmidial. Em fluxo laminar, uma colônia de bactérias foi coletada com palito de madeira esterilizado em autoclave, colocado em tubo de vidro para cultura de bactérias contendo 3mL de meio Circle Grow® com o antibiótico de seleção na concentração desejada, e levada ao agitador de tubos a 37°C por um período de 18 horas.. Após incubação, centrifugou-se a 14000rpm durante 30-60 segundos (centrífuga refrigerada Hettich Universal 30RF, rotor angular 1412) e o sobrenadante foi descartado.

O sobrenadante foi retirado e as bactérias foram ressuspendidas em 200µL de P1 com auxílio de agitador de tubos (Vortex “AP 56” – Phoenix). Em seguida, foram adicionados 200µL da solução P2. O tubo foi agitado por inversão suave e incubado a temperatura ambiente durante 5 minutos..Adicionaram-se 200µL de P3 e o tubo foi agitado por inversão suave logo após a adição da solução até que se formou um precipitado branco. A suspensão foi centrifugada a 20.000rpm por 15 minutos a 4ºC. Decorrido o tempo, o

sobrenadante foi transferido para um tubo novo e adicionou-se 400µL de isopropanol 65% e homogeneizou-se o conteúdo dos tubos por inversão. Em seguida, este foi centrifugado a 20.000rpm por 15 minutos a temperatura ambiente. O sobrenadante foi descartado e o precipitado lavado com etanol 70% (500µL), misturando-se por inversão dos tubos. Centrifugou-se a 20.000g, a 4°C, por 5 minutos.. O precipitado obtido foi seco em Speed-vac (Centrivap Console Labconco), e ressuspendido em 30 µL de água deionizada. Estas minipreparações foram então tratadas com RNAse I-A a 50µg/µL, ficando incubadas por 1 hora a 37°C, e estocadas a -20°C..

P1: GET (glicose anidra 0,92%, EDTA pH8,0 10mM, Tris 26mM)

P2: NaOH 200mM e SDS1%.

P3 - KoAc 3M pH 5,5 acertado com ácido acético glacial.

III. 5.4- Extração de DNA plasmidial em pequena escala em placas de 96 poços (Miniprep de placa)

Em placas de 96 poços, própria para crescimento de bactérias, adicionou-se 1mL de meio Circle Grow com o antibiótico desejado. O inóculo das bactérias foi coletado das placas de estoque de glicerol ou de uma colonia de bácterias, e depositado em placas de cultivo (Mega Titer Plate) contendo o meio de cultura com o antibiótico,. Estas foram seladas com adesivos ("Adhesive Plate Sealer" – AB Gene e "Adhesive Sealing Films for Microplates" – Simport). Com agulha flambada, um furo foi feito em cima de cada poço, para aeração das bactérias. A placa foi levada a uma incubadora horizontal com agitação, a 37°C, por um período de aproximadamente 20 horas. Em seguida foi centrifugada por 6 minutos a 2.000g (Centrífuga "Universal 30 F", Rotor "1422" – Hettich Zentrifugen). O

adesivo foi removido e o sobrenadante descartado, a placa foi invertida sobre papel absorvente por 5 minutos. Adicionou-se 240µL de GTE em cada poço, utilizando-se micropipetador multicanal Finnpipette. A placa foi selada com adesivo e levada ao agitador de tubos por 2 minutos. Foi centrifugada por 6 minutos a 2.000g, e o sobrenadante foi descartado, a placa foi invertida sobre papel absorvente por 5 minutos. Adicionou-se 80µL de GTE em cada poço. A placa foi selada novamente e levada ao agitador de tubos, para ressuspender o precipitado. Foram adicionados 2,5µL de RNase a 10mg/mL (RNase A – USB – fonte: pâncreas bovino) em cada poço da placa de polipropileno de 96 poços com fundo em "U" (Costar). Transferiu-se 60µL da suspensão de células para cada poço contendo RNase. Em seguida adicionou-se a cada poço 60µL de uma solução de NaOH 0,2N 1% SDS em água deionizada. A placa foi selada com adesivo, invertida 10 vezes para homogeneização, e deixada à temperatura ambiente por 10 minutos. Adicionou-se então, a cada poço, 60µL de uma solução de KoAc 3M gelada. Novamente a placa foi selada com adesivos, invertida 10 vezes e deixada à temperatura ambiente por 10 minutos. O adesivo foi removido, e a placa levada à estufa a 90°C. Após 30 minutos, foi colocada no gelo por 10 minutos. Foi centrifugada por 4 minutos a 2.000g, temperatura ambiente. Todo o volume de cada poço (180µL), foi transferido para uma placa de 96 poços ("MultiScreen Filter Plate" (cód. MAGVN 22-5) – Millipore), previamente fixada no topo da placa de polipropileno de 96 poços com fundo em "V" (Costar), de maneira que os poços ficassem alinhados. Centrifugando-se este conjunto, o precipitado ficou retido nas placas Millipore, enquanto a fase aquosa contendo o DNA foi filtrada para as placas com fundo em "V". Adicionou-se, então, 110µL de isopropanol gelado a cada poço contendo o filtrado. A placa foi selada com adesivo, invertidas por 15 vezes, e centrifugada a 2.000g por 45 minutos à

temperatura ambiente. O sobrenadante foi descartado, e a placa invertida sobre papel absorvente. Adicionou-se 200µL de etanol 70% gelado a cada poço. A placa foi centrifugada a 2.000g por 10 minutos, à temperatura ambiente, e vertida sobre papel absorvente para remover-se o sobrenadante. Para secar, a placa foi colocada em estufa a 37°C por 30 minutos. O DNA precipitado no fundo das placas foi ressuspendido em 80µL de água deionizada. A manipulação de bactérias foi realizada em fluxo laminar.

GTE - 23mL de Glicose Anidra 20%; 10mL de EDTA 0,5M pH8,0; 13mL de Tris-HCl 1M pH 7,0; água deionizada para completar um volume de 500mL.

III. 5.5 - Extração de DNA plasmidial em escala média (Midiprep)

A extração de DNA plasmidial em escala media foi feita usando o *Qifilter Plasmid Midi and Maxi Kits* (Quiagen Cat#12243) seguindo as especificações do fabricante.

III. 6- Procedimentos gerais de manipulação de DNA

III. 6.1- Amplificação de fragmentos por PCR

Para as reações de PCR utilizou-se: 300ng de DNA genômico, água destilada, ultrafiltrada e autoclavada, tampão de PCR 1X, 1,5mM $MgCl_2$, 10mM de dNTPs (dATP, dCTP, dGTP e dTTP), 10µM de cada iniciador e *Taq Polymerase Recombinan"* (Invitrogen Brasil Cat#11615-010). Sempre foi feito um controle negativo contendo todos os reagentes menos a amostra de DNA, para garantir a ausência de qualquer tipo de contaminação. Todas as reações foram feitas em um termociclador "*Gene Amp PCR Sytem 2400* "(Applied Biosystem, Foster City, Califórnia), conforme com o ciclo da figura 7.

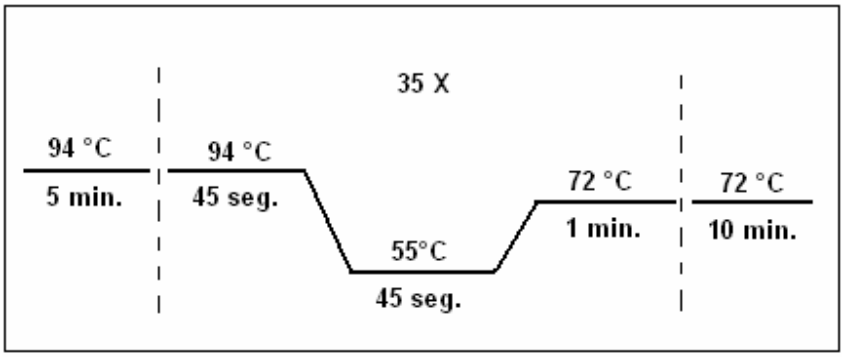

Figura 7- Ciclo de PCR utilizados nas amplificações.

III. 6.2- Purificação de fragmentos de DNA

A purificação de fragmentos de DNA foi feita através de colunas de purificação "GFX^{TM} *PCR and GEL BAND PURIFICATION* "(Amersham Bioscences Cat# 27-9602-01).

III. 6.3- Ligação de fragmentos de DNA

Os fragmentos originados em reações de amplificação por PCR com *Taq DNA Polymerase Recombinant* (Invitrogen Brazil Cat#11615-010) foram clonados no vetor pGEM, utilizando-se a enzima de ligação e demais reagentes fornecidos pela empresa Promega. Para a realização dos ensaios de ligação, foi utilizada a proporção inserto: vetor 3:1. Quando a eficiência de transformação não se apresentou adequada outras proporções foram testadas (1:1 e 1:3).

III. 6.4- Determinação da seqüência de nucleotídeos de uma molécula de DNA

O princípio do seqüenciamento automático de nucleotídeos é a incorporação de

nucleotídeos marcados com moléculas fluorescentes aos fragmentos de DNA de interesse. O seqüenciador automático identifica a seqüência de bases do fragmento de DNA através do reconhecimento das moléculas de fluorescência. Esta incorporação de nucleotídeos marcados é feita através de uma reação de PCR

As amostras de DNA utilizadas em reações de seqüenciamento de nucleotídeos foram provenientes de extrações de DNA plasmidial em pequena ou média escala (itens III.5.2 e III.5.3).

Cada reação inclui: 100-200ng de DNA, 5μM do respectivo iniciador, 2μL de *BigDye™ Terminator V3.1 Cycle Sequencing Ready Kit* (Applied Biosystem Part#4336919), e água deionizada para completar um volume final de 10μL. O "kit" é composto por uma mistura pronta de reagentes que contém: dNTPs, dideoxinucleosídeos (ddNTPs) com marcadores fluorescentes, tampão, cloreto de magnésio e a enzima *Taq* polimerase.

Os tubos ou placas de seqüenciamento foram levados para o termociclador "PTC 100™ Programmable Thermal Controller" (MJ Research, Inc.), passando pelo termociclo que consistiu de: 3 minutos a 96°C, 25 vezes a repetição da seguinte sequência: 96°C por 10 segundos, 48°C por 20 segundos, 60°C por 4 minutos. O DNA marcado com as moléculas fluorescentes foi então precipitado. Para a precipitação, adicionou-se a cada reação 100μL de isopropanol 65% e agitaram-se os tubos ou placas no "vortex". Os mesmos foram deixados por 20 minutos em temperatura ambiente, e em seguida centrifugados a 2.000g por 50 minutos, temperatura ambiente. O sobrenadante foi descartado vertendo-se os tubos ou placas sobre papel absorvente. Em seguida foi realizada uma lavagen do precipitado com 150μL de etanol 60%, e centrifugação a 2000g por 10 minutos em temperatura ambiente. O sobrenadante foi descartado vertendo-se os tubos ou

placas sobre papel absorvente. Os mesmos foram levados à estufa a 37°C por 15 minutos, para que o etanol secasse completamente. Os DNAs marcados foram ressuspendidos em 10µL de formamida ("Hi-Di Formamide" P/N 4311320 – Perkin Elmer Biosystems), desnaturados, ficando por 4 minutos a 96°C, e deixados no gelo por 10 minutos. Logo as amostras foram colocadas para sequenciar no aparelho "3700 Genetic Analyser" ou "3100 Genetic Analyser" - ABI PRISM. Os eletroferogramas das sequências foram visualizados no programa Sequencing Analysis 3.4 (ABI PRISM).

Os clones foram seqüenciados em ambas as direções. Todas as seqüências consensos foram geradas com uma qualidade Phred maior que 20. Os clones foram montados com o pacote Phred/Phrap/Consed (Gordon *et al*. 1998; Gordon *et al*. 2001; Ewing *et al*. 1998) com uma taxa de erro final de 1bp a cada 10kb.

III. 6.5- Hidrólises de DNA com endonucleases de restrição

Para os ensaios de hidrólise de DNA foi usada a proporção de 1 unidade de enzima de restrição para cada micrograma de DNA plasmidial e 10 unidades para cada micrograma de DNA genômico. As condições de reação foram determinadas de acordo com as indicações do fabricante. As endonucleases utilizadas foram: *Eco*RI e *Bam*HI (New England Biolabs).

III. 6.6- Transferência alcalina de DNA para membrana e Southern Blot

A passagem de DNA de um gel de agarose para uma membrana foi realizada através do processo de transferência alcalina.

Após uma eletroforese o gel foi mantido em uma solução de hidrólise por 10 minutos sob agitação. Depois foi lavado em água ultrafiltrada e imediatamente mantido em uma solução de desnaturação por 30 minutos. Logo, foi colocado no aparelho de transferência com uma membrana de nylon (*GeneScreenPlus® Cat# NEF988 Perkin Elmer*). Em cima da membrana foi montada uma coluna de cerca de 10cm, constituída de papel de filtro e toalha, permitindo a transferência por capilaridade da solução de desnaturação. A coluna de papel foi compactada usando um bloco de 500gr e o processo de transferência durou 16-18 horas. Ao final a membrana foi mantida na solução de neutralização por 15 minutos e posteriormente deixada secar a temperatura ambiente pelo tempo necessário.

Hidrólise: HCL (0,25N): preparado na hora ou utilização máxima de 10 dias.

Desnaturação: NaOH (0,4N): preparado na hora ou utilização máxima de 10 dias.

Neutralização: SSC 2X

As soluções são preparadas no momento de serem utilizadas a partir das soluções estoque descritas no item III.3.1.

III. 6.7- Marcação radiativa de fragmentos de DNA utilizadas como sondas

As sondas foram preparadas através de amplificação por PCR (item III.6.1) dos fragmentos escolhidos dos plasmídeos correspondentes. Posteriormente foram purificadas segundo o procedimento do item III.6.2.

A incorporação de (α-^{32}P) dCTP foi realizada com *Random Primers DNA Labelling Sytem* (Invitrogen Cat#18187-013), seguindo as instruções do fabricante. Após da marcação

foi adicionado DNA de esperma de salmão à sonda, para que ficasse a uma concentração de 100µg/mL na solução de hibridação. Antes de ser adicionada, a sonda foi desnaturada colocando-se o tubo da reação a 94°C por 4 minutos e posteriormente em gelo por 3 minutos.

III. 6.8- Hibridação molecular, lavagem e exposição da membrana

A membrana foi incubada em 10mL de solução de hibridação a 65°C por 4 horas. Após esse tempo à solução de hibridação foi trocada por uma nova e a sonda foi adicionada continuando com a incubação por mais 24 horas.

Solução de hibridação: SDS 1%, NaCL 1M, Dextran (Dextran Sulfate Sodium Salt from *Leuconostoc* spp.) 10%.

Foram efetuadas quatro lavagens de 30 minutos nas membranas. De acordo com as exigências e necessidades dos experimentos foram feitas de duas maneiras diferentes: utilizando uma mesma solução I (condições menos adstringentes) ou fazendo as primeiras duas lavagens com a solução I e as duas seguintes com a solução II (condições mais adstringentes).

Solução I: SSC 0,5X, SDS 0,1%.

Solução II: SSC 0,1X, SDS 0,1%.

Após as lavagens, as membranas foram acondicionadas em chassis contendo filmes de raios-X (Kodak), e ficaram mantidas a -70°C pelo tempo necessário para a impressão do filme. A revelação do filme foi feita em tanque de revelação em sala escura, utilizando revelador e fixador Kodak.

III. 7- Análises da diversidade das seqüências

A identidade das seqüências obtidas foi confirmada por BLAST (*National Center for Biotechnology Information, NCBI, Bethesda, Md*) (Altschul *et al*. 1990). As seqüências consensos foram alinhadas usando o programa CLUSTAL W *multiple-alignment* (version 1.5) (Thompson *et al*. 1994) e ajustes manuais foram feitos quando necessário no programa Bioedit version 7.0.5.3 (Hall, 1999).

III. 7.1 - Divergências das seqüências

A divergência das seqüências previamente alinhadas foi estimada usando o programa Dnasp version 4.10.9 (Rozas and Rozas, 1999). O programa permite determinar diversos parâmetros entre seqüências, entre os quais se encontra a diversidade de nucleotídeos o parâmetro Pi (π), que representa a media do número de nucleotídeos diferentes em cada posição do alinhamento. O modelo da substituição utilizado foi o Jukes-Cantor (Jukes and Cantor 1969) onde todas as substituições têm a mesma taxa.

III. 7.2 - Análises das relações entre as seqüências usando métodos filogenéticos.

O programa PAUP (Swofford, 1993) foi utilizado para analisar filogeneticamente as seqüências geradas, usando o método de distancia com o modelo Kimura 2-parâmetros. A sustentação para grupos foi avaliada com 1.000 réplicas de boostrap.

A identidade de nucleotídeos entre todos os pares de seqüências (*pair-wise)* foi calculada com Strecher da EMBOSS (Rice, 2000).

III. 7.3.- Análises das relações entre as seqüências usando o programa NETWORK

Para a análise em rede, as seqüências obtidas foram alinhadas usando DNA *Alignment* 1.1.21 *software* e analisadas usando o programa NETWORK 4.2.0.1 e *median-joining* (Bandelt, 1999) com os parâmetros *default*. Ambos os programas estão disponíveis no site: http:/www.fluxus-engineering.com/sharenet.htm.

O NETWORK é um programa desenvolvido para reconstruir árvores filogenéticas parcimoniosas calculando distâncias entre seqüências.

O NETWORK aceita seqüências alinhadas em um formato determinado, o mesmo pode ser gerado manualmente ou usando o programa "Fluxus commercial DNA alignment". Este programa, utilizada um algoritmo que alinha as seqüências completas sem penalizar os *gaps*, comparando as com uma seqüência referencia escolhida arbitrariamente. Normalmente a seqüência referência é escolhida do mesmo grupo de seqüências a serem alinhadas.

Foram analisadas 197 seqüências geradas neste estudo provenientes de espécies de *Solanum*, 116 seqüências provenientes de espécies de *Lycopersicon* e 140 seqüências, provenientes de espécies de *Nicotiana*.

Os números de acesso no "GenBank" são:, [GenBank: EF620567-EF620763] para as espécies de *Solanum*, [GenBank: AJ227998-AJ228137] para as espécies de *Nicotiana,* [GenBank: X13777] para Tnt1-94, e para as espécies de *Lycopersicon* [GenBank: EF602630-EF602746].

III. 8- Determinação do número de cópias usando Quantificação Absoluta por qPCR.

O número absoluto de cópias foi determinado comparando-se as amostras de DNA de concentração determinada com uma curva padrão utilizando a PCR quantitativo (qPCR). A curva padrão foi feita com diluições seriadas de um plasmídeo contendo a fragmento de interesse.

Os plasmídeos utilizados na curva padrão foram obtidos por extração de MidiPrep (ver item III.5.4), confirmando a presença do fragmento de interesse por seqüenciamento. As concentrações foram cuidadosamente determinadas no Fluorômetro segundo o item III.4.3 utilizando uma curva padrão de *Lamda DNA Hind III (*Biolabs) e posteriormente confirmadas em gel de agarose. Foram feitas diluições de 10ng/µL do plasmídeo a partir da solução estoque e confirmadas por quantificação. A partir destas diluições foram preparadas as diluições de trabalho 1,0ng/µL. Uma serie de diluições seriadas a partir da solução de trabalho foram realizadas para produzir: $5x10^{-1}$, $5x10^{-2}$, $5x10^{-3}$, $5x10^{-4}$, $5x10^{-5}$, $5x10^{-6}$ ng/µL do plasmídeo recombinante.

As espécies do estudo foram escolhidas por terem os tamanhos de seus genomas já determinados (*Plant DNA C-value database Royal Botanic Kew Gardens*): *Nicotiana tabacum* (4.459 Mpb), *Lycopersicon peruvianum* (1.127 Mpb), *Solanum tuberosum* (1.715 Mpb) cult. Huinkul e cult. Kennebec, *S. chacoense* (613 Mpb), *S. vernei* (784 Mpb) e *S. brevidens* (740 Mpb). O DNA foi extraído usando *DNeasy®Plant Mini kit* (Quiagen Cat #69104) seguindo as especificações do fabricante, com o intuito de obter DNA de alta qualidade e livre de possíveis contaminantes e inibidores.

As reações de qPCR para as diluições padrão e o DNA das amostras foram feitas no equipamento *7300 Real Time PCR Systems* (Applied Biosystems) utilizando 12,5µL de *SYBR®Green PCR Máster Mix* (Applied Biosystems P/N4344463), 6,5 µL de água ultra filtrada, 0,5µL de cada iniciador e 5µL do DNA (1ng/µL) da amostra. Os parâmetros e tempo da corrida, utilizados estão indicados na figura 8. No caso da determinação do número de cópias do retrotransposon similares a Tnt1, a temperatura de hibridação dos iniciadores foi mudada para 57ºC.

A concentração ideal dos iniciadores foi determinada testando diferentes concentrações. Foi utilizada aquela concentração que produziu o máximo sinal fluorescente com o DNA das amostras e o correspondente controle negativo sem emissão de sinal (não produz dímeros de iniciadores).

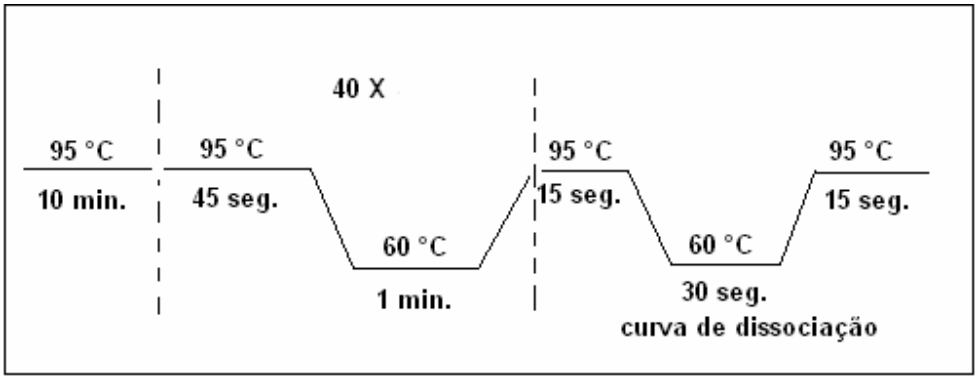

Figura 8. Ciclo de qPCR utilizado.

Os iniciadores utilizados para determinar o número de cópias dos retrotransposons foram Avi e RT, e para determinar o número de LTRs: LTR-F e LTR-R (ver item III.2.2). O iniciador Avi foi utilizado para caracterizar todas as famílias e grupos dos elementos similares a Tnt1 em diferentes espécies de *Solanaceae* (Vernhettes *et al*. 1998, Araujo *et al*.

2001), e por tanto se considerou adequado na determinação do número de cópias. Já o iniciador PPT foi desenhado tendo em conta seqüências de Retrosol determinadas neste trabalho, a seqüência de Tnt1-94 (GenBank: X13777] e as seqüências de Retrolyc1 em *Lycopersicon peruvianum* [GenBank: EF602710-EF602734]. Um retrotransposon em um genoma ao longo do tempo gerará uma população de seqüências e como a amplificação por PCR depende do reconhecimento dos iniciadores o resultado da mesma será uma amostra da população.

A curva padrão foi produzida determinando a linha de tendência da regressão entre os valores de C_T (valor da fluorescência acima do *threshold*) e o logaritmo do número de cópias. A curva padrão é uma curva linear descrita pelos parâmetros [$x = my + b$], onde m representa –[1/log] (eficiência da PCR), b é o logaritmo da quantidade do produto amplificado no *threshold* dividido pelo logaritmo da eficiência da PCR. A eficiência da amplificação por PCR (E) foi calculada segundo a fórmula (Kubista *et al*. 2006):

$$E = 10^{-1/\text{pendente}} - 1$$

O número inicial de cópias do plasmídeo foi calculado considerando-se a massa e o tamanho do plasmídeo, segundo a fórmula:

massa de uma cópia de plasmídeo = [3015 bp (PGEMEASY Vector) + pb do fragmento] [1,096.10-21g/pb]

O número de cópias do DNA das espécies foi calculado a partir do valor obtido de C_T e utilizando as quantidades de genomas da tabela 4.

Tabela 4. Número de genomas por massa de DNA.

	1C pg	1ng	5ng
Nt	4,55	220	1100
Lp	1,15	869	4348
Hu	1,75	571	2857
Ke	1,75	571	2857
chc	0,63	1587	793
vrn	0,80	1250	625
bre	0,76	1316	657

Nt: *Nicotiana tabacum*, Lp: *Lycopersicon peruvianum*, Hu: *Solanum tuberosum* cult. Huinkul, Ke: *S.tuberosum* cult. Kennebec, chc: *S. chacoense*, vrn: *S. vernei*, e bre: *S.brevidens*. 1C: conteúdo genômico haplóide. 1ng: número de genomas por 1ng de massa genômica. 5ng: números de genomas em 5ng de massa genómica.

Diversos parâmetros da qPCR são importantes para ter confiança nos resultados da quantificação. Um deles é a eficiência de amplificação (E). A eficiência está influenciada por vários fatores incluindo o tamanho do produto amplificado, estrutura secundaria do mesmo e qualidade dos iniciadores. Um valor de E = 1 (100%) corresponde a uma inclinação da curva padrão de -3,3 e uma amplificação (A) de 2 (duplica o número de moléculas em cada ciclo de amplificação). Consideraremos razoável utilizar curvas padrão com uma inclinação que varie entre -2,92 e -3,90 gerando uma E=120% e A= 2.2, e E= 80.5% A=1.8 respectivamente. Outro parâmetro a ser considerado é o número de pontos (diluições de plasmídeos) que vão definir a curva padrão, recomenda-se um mínimo de cinco pontos e as quantificações das amostras devem estar contidas entre eles. Por último considerar o valor do coeficiente de correlação (R^2).

III.8.1- Normalizações utilizando um gene de cópia única

Com o objetivo de ter um controle interno da quantidade de DNA utilizado em cada experimento, o número de cópias das amostras foi normalizado utilizando um gene de copia única: *XPB* (*Xeroderma Pigmentosum* B).

Fragmentos do gene *XPB* das espécies a quantificar foram amplificados e confirmados por seqüenciamento. O número absoluto de cópias de *XPB* por PCR em Tempo Real foi determinado através da curva padrão do plasmídeo com o incerto de *XPB*.

Foram quantificadas as mesmas diluições das amostras de DNA utilizadas na determinação do número de cópias de retrotransposons similares a Tnt1. Os valores absolutos do número de cópias de cada genótipo foram divididos pelos valores correspondentes do número de cópias de *XPB*.

Os valores normalizados podem ser utilizados para comparar as quantidades relativas em diferentes amostras ou experimentos. Uma manera de fazer esa comparação é designar uma amostra como calibrador e dividir os valores quantificados por ela para obter as quantidades relativas. Arbitrariamente o número de cópias normalizado de *S.brevidens* foi escolhido como calibrador por ter o menor de número de cópias (Livak e Schmittgen 2001).

Os coefficientes de variação (CV) dos valores normalizados foram determinados utilizando a seguinte fórmula:

$$CV_{\text{da normalização}} = \sqrt{CV_{amostra}^{2} + CV_{XPB}^{2}}$$

Onde o desvio padrão (S):

$$S = (CV_{\text{da normalização}}) \times (\text{Média}_{\text{normalizada}})$$

A metodologia da PCR em tempo real foi adaptada do manual: *Sequence Detection Systems, Applied Biosystem, Chemistry Guide.* 2003.

III.9- Isolamento do elemento Retrosol

O DNA da espécie *Solanum oplocense*, foi extraído usando *DNeasy®Plant Mini kit* (Quiagen Cat #69104) seguindo as especificações do fabricante, com o intuito de obter DNA de alta qualidade e livre de possíveis contaminantes e inibidores.

Para a reação de PCR utilizou-se: 250ng de DNA genômico; água destilada ultrafiltrada e autoclavada, tampão de PCR A 5X e B 5X, enzima *Elongase® Enzyme Mix* (Invitrogen Cat#10480-010), 10mM de dNTPs (dATP, dCTP, dGTP e dTTP) e 10µM de cada iniciador (Ol16 e GII-R). Sempre foi feito um controle negativo contendo todos os reagentes menos a amostra de DNA, para garantir a ausência de qualquer tipo de contaminação.

A reação foi feita em um termociclador "*Gene Amp PCR Sytem 2400* "(Applied Biosystem, Foster City, Califórnia), conforme o ciclo da figura 9.

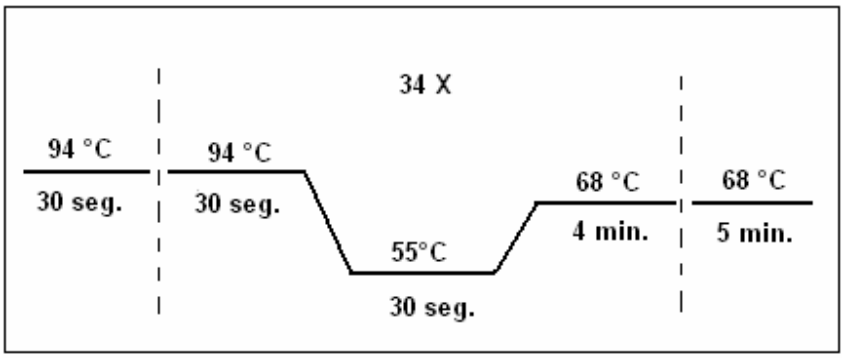

Figura 9. Ciclo PCR utilizado para amplificar Retrosol.

A purificação do fragmento de DNA foi feita através de colunas de purificação "GFX^{TM} *PCR and GEL BAND PURIFICATION* "(Amersham Bioscences Cat#27-9602-01). O fragmento foi clonado no vetor pMOS*Blue* Blunt Ended Cloning Kit (Amersham Biosciences Cat# RPN5110). O DNA plasmidial foi extraído em escala media usando o *Qifilter Plasmid Midi and Maxi Kits* (Quiagen Cat#12243) seguindo as especificações do fabricante. Duas hidrólises totais e duplas foram feitas com 1ug de DNA plasmidial. A primeira hidrólise utilizou 40U das endonucleases *Hind* III e *Pst* I e a segunda hidrólise utilizou 40U das enzimas *EcoR* I e *Kpn* I. O produto das hidrólises foi clonado no vetor *pMOSBlue Blunt Ended Cloning Kit* (Amersham Biosciences Cat#RPN5110).

As seqüências obtidas foram analisadas no programa Phred/Phrap/Consed (Gordon *et al*. 1998, Gordon *et al*. 2001, Ewin *et al.* 1998) e a seqüência consenso foi analisada em banco de dados (Genebank/NCBI). A qualidade do concenso foi melhorada quando necessario, clonando e sequenciando fragmentos do clone B9. As seqüências parciais das LTRs foram alinhadas usando o programa CLUSTAL W *multiple-alignment* (version 1.5) (Thompson *et al.* 1994). A análise evolutiva foi conducida usando o programa MEGA 3.1(Kumar *et al*. 2004) com o parámetro Kimura (Kimura, 1980).

IV - RESULTADOS

IV. 1- Obtenção e análise de seqüências similares a Tnt1.

IV. 1.1- Amplificação e seqüenciamento de fragmentos similares a Tnt1 em espécies de *Solanum*

A presença de seqüências similares a Tnt1 em vinte e cinco genótipos de *Solanum* (tabela 1) foi analisada através da amplificação por PCR (item III.6.1) utilizando os iniciadores Avi e OL16 (item III.2.). Os iniciadores delimitam a última porção da RNAseH, a região entre o final da poliproteína e o inicio da LTR na posição 3' chamado de Linker, toda a região U3 de tamanho variável, a região R e uma porção da região U5 (Figura 10). DNA extraído de *Lycopersicon peruvianum* (Lp) foi utilizado como controle positivo da amplificação já que os iniciadores foram desenhados sobre as seqüências de Retrolyc1-1.

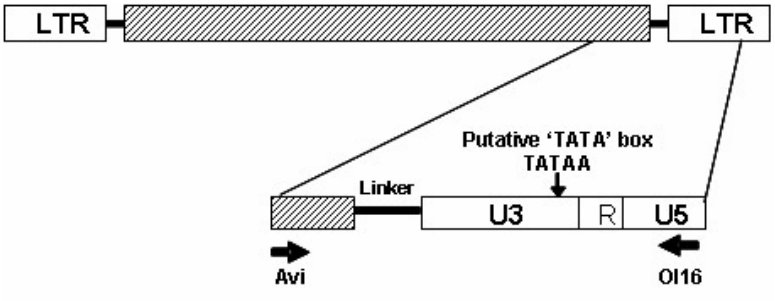

Figura 10. Representação esquemática do fragmento de retrotransposon amplificado.

LTR: *long terminal repeat*, caixa listrada grande: região codificadora, caixa listrada pequena: RNAseH: ribonuclease H, Linker: região no codificadora com PPT (*polypurine tract*), U3: *unique 3'RNA*, R: *repeated RNA*, U5: *unique 5'RNA*. Setas: posição dos iniciadores utilizados**.**

Fragmentos do tamanho esperado (500pb) foram obtidos em todos os genótipos como pode ser observado na figura 11. As bandas de tamanho esperado foram purificadas e clonadas em pGEM-T. Trezentos e dezesete clones foram seqüenciados (tabela 5). O tamanho do fragmento variou de 400-470 pb.

Figura 11. Amplificação de fragmentos similares a Tnt1 em 24 genótipos de *Solanum* usando os iniciadores Avi e OL16.

Os nomes dos genótipos correspondem aos códigos utilizados na tabela 1. M: marcador molecular.

IV. 1.2- Análise da diversidade das seqüências

A análise das seqüências contra o banco de dados do NCBI confirmou que o total dos clones apresentava similaridade com Tnt1. O novo retrotransposon isolado de *Solanum* foi denominado de Retrosol. O alinhamento de nove seqüências de Retrosol, escolhidas em forma randômica, junto com três seqüências representativas de Tnt1 (de *Nicotiana*) e duas de Retrolyc1 (*Lycopersicon peruvianum*) revelou que a similaridade em nucleotídeos se concentra nas regiões RNAseH, R e U5. A região U3 apresenta as maiores diferenças entre as seqüências (figura 12).

Tabela 5. Parâmetros de variabilidade genética das seqüências similares a Tnt1 dos genótipos de *Solanum*

code	Domest.	total	unique	size bp	π JC	I	II	III	premature	no stop/cod
bre		25	13 52%	456(±7)	0,057	13	-	-	1	5
cmm		3	3 100%	423(±5)	0,093	1	2	-	-	-
chc		1	1 100%	428	-	1	-	-	-	-
tar		15	9 60%	422(±14)	0,038	1	8	-	1	-
mga		17	10 58%	439(±19)	0,171	6	3	1	-	-
sat		7	6 85%	440(±4)	0,069	4	2	-	-	-
inf		23	10 43%	428(±9)	0,053	4	6	-	2	-
dds		1	1 100%	435	-	1	-	-	1	-
lph		1	1 100%	436	-	1	-	-	-	-
mcd	silvestre	2	2 100%	428(±11)	0,078	1	1	-	-	-
opl		2	2 100%	419(±2)	-	-	2	-	1	-
scr		12	12 100%	463(±11)	0,132	6	5	1	1	1
gou		3	3 100%	430(±8)	0,093	2	1	-	1	-
ha		19	12 63%	428(±9)	0,249	2	7	3	-	-
inc		21	11 52%	421(±4)	0,041	2	9	-	3	-
ktz		18	14 77%	443(±8)	0,287	7	-	7	-	-
oka		16	7 43%	434(±10)	0,273	2	3	2	-	-
spg		20	9 45%	433(±7)	0,055	5	4	-	3	-
vrn		12	4 27%	429(±9)	0,072	2	2	-	-	-
Ke		8	8 100%	426(±14)	0,078	2	6	-	-	-
Hu		4	4 100%	421	0,026	-	4	-	-	-
Cha	cultivada	23	17 74%	424(±7)	0,064	7	10	-	-	-
Mo		29	17 58%	423(±6)	0,057	6	11	-	-	-
Tu		17	14 82%	428(±10)	0,058	8	6	-	-	-
acl	silvestre	18	7 38%	419(±19)	0,191	3	3	1	-	4
total		317	197 62%			87	95	15	14	10

Código: espécies e cultivares (**code**). Domesticação (**Domest**.): silvestre ou cultivada. **Total**: número total de seqüências obtidas; **unique**: número total de seqüências diferentes em porcentagem, **size bp**: tamanho médio dos fragmentos em pb, **πJC**: diversidade de nucleotídeos com a correção de Jukes e Cantor's (1969); **I, II e III:** grupos filogenéticos da figura 13. **premature**: seqüências com *stop codon* prematuro e **no stop/cod:** são seqüências sem *stop codon* no domínio RNAseH.

```
           10        20        30        40        50        60        70        80        90       100
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
TntlA      CAAGGAACAA-TTCGAGCTATGCAAAGAACTTGTCGGCATGCACTCAAACTAGAAGAC--AGTGCTACCTCCTCTAGATGAATGAGACTGGAGGGGGAGA
TntlB      ..........G.............................................A..--....T..............A....G...............
TntlC      ....AG....G..T....G........T........C....A..............ACTGTG...T..T.C.T.TC..G............A.........
RetlyA     -..CAG....-..T.........................T......G......AC.-TG.AA.........TC..G..C...G.T.............
RetlyB     -.T..G....G..TA.............................G......AC.-C..A..........TC..G.T....G.TA....AA......
mcd28      ....AG....G.................................G......ACT-TC.A........T.TC..G.....AG.T.A...........
spg2       ....AG....G.................................G......ATT-T..A........T.TC..-......G.T.A...........
Tu19       ....AG....G.................................G......ACT-T..A........T.T...G......G.TT............
ha16       .G.AAG....G..T...T................G..T...AG...TCT...AGGA.GATG.A.A......T.TCCAG......G..............
ktz23      .G.AAG....G..T...T........A.......G..T...AG...TCT...A.GA.GATGAA.A........TCCAG......G..............
ktz11      .G.AAG....G..T...T................G..T...AG...TCT...A.GA.GATGAA.A........TCCAG......G..............
Tu2        T...AG....G.....A...............................G......ACT-T..A..........TC..G......G.T.............
ver3       ....AG....G..T....................A............G......ACT-T..A..........TC..G......G.T......A......
inc21      ..G.AG....G..T....................A............G......ACT-T..A..........TC..G......G.T......A......
```

RNAseH ★ Linker

```
           110       120       130       140       150       160       170       180       190       200
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
TntlA      TTGATGATGTCCAT-CTCATTGAAGAAGTATTAGGCATGTGCCTAATAAGAGTTTTCTTTGGTTTGGTAGCCAACCT-TGTTGAC--------TTGGTTT
TntlB      ..TG..GG......A.C......T------.....--------------A......TC...A---------....T.-....TT.--------AAA...-
TntlC      ..TG...G.....-G.C...AT-----..GAA..AA----------G..A.C..C.TC..TC.--------TCCTT.-...CA.G--------CC...CA
RetlyA     ..-G...G......C.C...-------...CC.AA.----------.TT..A.AACT..GT.C--------...TT.-CAAA.T---------.....AG
RetlyB     ..TG..GA......A.C....------...GATAAA----------C.TA..GAAAG..AA----------..TG.A.AAATGG--------...AA..
mcd28      ..-G..GG......C.C...-------T.GCC.AA---------------------TC.CA----------...TG.-CAAAAGG--------....CAG
spg2       ..-G..GG......C.C...-------T.GCC.AA---------------------TC.CA----------...TG.-CAAAAGG--------....CAG
Tu19       ..-G..GG......C.C...-------T.GCC.AA---------------------TC.CA----------...TG.-CAAAAGG--------....CAG
ha16       ..TG..GG.....GT.C..C--------....CCATT.T---TG.C.TCA.CAA..G..GCA.GCAA.T.T...TTGT..GCA.ATTGCAATAGCAACAA
ktz23      ..TG..GG.....GT.C..C--------....CCATT.T---T..C.TCA.CAA..GC.GCA.GCAA.T.T...TTGT..GCA.ATTGCAATAGCA.CAA
ktz11      ..TG..GG......C.C...--------....CAATTATAAGTGTT.TCA.CA...GC.GCA--CCA.T.A...TTAT..C.---------TT.CAACAA
Tu2        ..-G..GGA.....C.C...-------..G.C.AA.----------.TT..-.AAATC.CA----------...TG.-CAAA.GG--------....CAG
ver3       ..-G..GG......C.C...-------..G.C.AA.----------.TT..ACAA.TC.CT----------...TT.-CAAA.G---------....CAG
inc21      ..-G..GG.....-C.C...--------..G.C.AA.----------.TT..ACAA.TC.CT----------...TT.-CAAA.G---------....CAG
```

U3

```
           210       220       230       240       250       260       270       280       290       300
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
TntlA      GGTTGGTAGCCAACCTTGTTGAATTAGTTT--GGTTTGGTAGCC-AACCTTGTTGAATTTCTTTGG--------------------------------
TntlB      -.C..C..AAG..------------------------.C....-C.TG..T.C--....T...-----------------------------------
TntlC      AT.CA.A.AGATTTGG----------------------.CGA.T-G.TT.-------...GC.-----------------------------------
RetlyA     A.......AT.T.TAA----AG.AG.AAAACTT....T..G-..-..TT.CAAA--G...GG.A.AGTTGGTAAGCTACAAAGAAGAAAAACAAGTTTTG
RetlyB     TT..T..T..TT.TTG.TAAAGTC..CAAA----.GCT.C.A.A-..AT.CAAA--GC.CACC------------------------------ATTTCTT
mcd28      AT....-.AGAT.TTA----A.TA..CA---------T..G-.A-G.TT.-------...-----------------------------------
spg2       AT....-.AGAT.TTG----..TA..CA---------T..G-.A-G.TT.-------...-----------------------------------
Tu19       AT....-.AGAT.TTG----..TA..CA---------TA.G-.A-G.TT.-------...-----------------------------------
ha16       ATG.T.A.AAGC..AGC----.....AAAACGCT..C...G..AC...T...ACAT.C-.A.GCATGG--------------------------------
ktz23      ATG.T.A.AAGC..AAC----.....AAAACGCT..C...G..AC...T...ACAT.C-.A.G.ATGG--------------------------------
ktz11      TTGCT.C.T------GC----....GTCAATG.T.--..C...A-..TG...AGA..CA.A.CG.TGGA--------------------------------
Tu2        AT....-.AGAT.TTG----..TA..CA---------T..G-.T-..TT.-------...T..----------------------------CCAGCTTTC
ver3       AT....-.--------------TA..CA---------T..G-.T-G.TT.-------A..T..----------------------------C-AGCTTTC
inc21      AT....-.--------------TA..CA---------T..G-.T-G.TT.-------A..T..----------------------------C-AG----C
```

U3

```
           310       320       330       340       350       360       370       380       390       400
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
TntlA      -TTTGGTAGCCAACTTTGTTG---------AA----TTGTGAAAAGTGTGTGTAAATTG-------TCAAATATTGTAGGCTTTAGAGGGTGAAGCT---
TntlB      --------.T...AAA..ACA---------GG----C.ACAG..GT..-----.....-------.....-----..A.T.------..CA...T.---
TntlC      ---------GT.GTG...A..------------TCAC..CAT.C.TCA---AC.----------T----------------------A.AT.TT..---
RetlyA     TGCCAA.TT.A..G....G.AGAGTAGGTG.G----C.A....G.AGA---AA..G...-----TT.TT-C.TC.A...CA.A.TTG.A.G.-TAG.---
RetlyB     T...CAACTAA..TC..AGCA---------.GATTTC.ACAT.GGTGA---AA.GG...AAATCTT.TGTT.C....GATG.CACTG.T.AC.TCA.TAG
mcd28      ---------.A..GG...G.A---------.G----C..CA..G.AGA---AA...A..-----TT.TTTC.TC.....CA.AGTTG.AAAA.TA..---
spg2       ---------.A..GG...G.A---------.G----C..CA..G.AGA---AG...A..-----TT.TT-C.TC.....CA.AATTG.AAAA.TA..---
Tu19       ---------.A..GG...G.A---------.G----C..CA..G.AGA---AAC.....-----TT.TT-C.TC.....CACAGTTG.AAAA.TAT.---
ha16       ------.GTTG..A....CA.----------------C..CA..TTTCT.C.A.....AA---------GG.GCCCT.T.CTCA.TCCTAAAC.CA.CAAT
ktz23      ------.GTTG..A....CA.----------------C..CA..TTTCT.C.A.....A.---------GG.GCC.T.T.CTCA.TCCAAAAC.CA.CAAT
ktz11      ------.GTTG..AAA.....----------------C..---.TTTCT.C.A....-A.---------G.G.T.CT...CTCAATT-.AAAT.CA.CAAT
Tu2        C.CAAAAGT.A..TG...G.A---------.G----.......G.AGA---AA..T...-----T----------------------.A.AC-TA..---
ver3       C.CAAAAGT.A..TG...G.A---------.G----C.A....G.AGA---AA.....-----TT.TT-C.TC.....TACA.TTG.A.GC-TA..---
inc21      C.CAAAAGT.A..TG...G.A---------.G----C......G.AGA---AA.....-----TT.TT-C.TC.....TA.A.TTG.A.GC-TA..---
```

U3

```
           410       420       430       440       450       460       470       480       490
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....
TntlA      -------TTGGC-TATAAAAG--GAGAGCTTCA-----ACTCTCATTTCTACACACCAACA-------------AAGAGAGAAAGAAAGAGTGA
TntlB      -------.GCC.-.......T--.........G-----G.........TC...T-.----------------------.------.----....
TntlC      -------GCAA.-......T---A..T.TC..T-----..CT.....GT-.AGATA..C..A---------CAA...................
RetlyA     GTATCTG..CT.-.......---.G...GACA.-----TTCT.....CT--A.GTA..C..AAAAGAATTACAA.G...........-------
RetlyB     GAGTAGA..TT.C......---TAGAGAGC..-----TTGCTCAT.-..A.A.-TG..C..AAAAGAATTACAA.G..C........-------
mcd28      GTACTTG..AT.-.......---.G...AG..------TTCT.....CTC.A.ATG..C..AAAA------CAA.G........AG..------
spg2       GCACTTG..AT.-...G...---.G...AG..------TTCT.....CTC.A.ATG..C..AAAA------CAA.G........AG..------
Tu19       GCACTTG..AT.-.......---.G...AG..------TTCT.....CTC.A.ATG..C..AAAAAT----CAA.G...G....AG..------
ha16       TCACAGAGA.AG-..A...TATTT....AG.TGTGAGGTT.TC...----.G..TAT.TA.GAAAATAGT-CTGTGA...A...T.G..T....
ktz23      TCACAGAGA.AG-..A...TATTT....AGCTGTGAGGTT.TC...----.GG.TAT.TA.GAAAATAGT-CTGTGA...A...T.G..T....
ktz11      TCACAGAGA.AG-..A...TATTT....AGCTGTGAGGTT.TC...----.GG.TAT.TA.GAAAATAGT-CTGTGA...A...T.G..T....
Tu2        CCATTTG..AT.-.......---.G...GG...------TTCT.....CTC.A.-TG..C..AAAA------TAA.GA.......AG..------
ver3       GCATTTG..CT.-.......---.G...GG...------TTCT.....GTA.A.-TG..C..AAAA------CAA.G........AG..------
inc21      GCATTTG..CT.-.......---.G...GG...-----TTCT.....GTA.A.-TG..C..AAAA------CAA.T........AG..------
```

TATA U3 R U5

Página anterior.

Figura 12. Alinhamento das seqüências parciais de Retrosol clonadas.

Foram incluídas seqüências representativas de Tnt1A, B e C; Retrolyc1A e B (RetlyA e B); e 6 seqüências de Retrosol (negritas). Regiões sublinhadas indicam: RNAseH: ribonuclease H, Linker: região no codificadora com PPT (*polypurine tract*), U3: *unique 3'RNA*, R: *repeated RNA*, U5: *unique 5'RNA* e TATA.

As seqüências redundantes obtidas nos genótipos não foram consideradas nas análises posteriores porque não é possível saber se elas pertencem a cópias diferentes do elemento ou não (tabela 5). Poucas seqüências apresentaram um *stop codon* prematuro ou nenhum *stop codon* na região codificante. Como nenhum dos genótipos em estudo, tem seu genoma caracterizado, não é possível assegurar que a população de seqüências amplificadas representa a diversidade realmente existente. Entretanto, esta aproximação parece apropriada para estudar a evolução da sub-família deste elemento como foi previamente usada na caracterização das subfamílias Tnt1 e Retrolyc1 (Vernhettes *et al*. 1998, Araújo *et al*. 2001).

Foi calculada a diversidade das seqüências (πJC) do fragmento completo dentro de cada genótipo. Os índices variam de 0,026 até 0,287 e permite classificar os genótipos em três classes (tabela 5). Os genótipos com maior diversidade (πJC maior ou igual a 0,2) são *S. hannemanii, S. kurtzianum e S. okadae.* A segunda classe inclui *S. acaule, S. megistacrolobum* e *S. x sucrense*, com um índice de diversidade entre 0,1 e 0,2. A terceira classe agrupa o resto dos genótipos com πJC menor que 0,1. As batatas cultivadas (*S. tuberosum* cultivar Huinkul e Kennebec, e as subsp. *andigenum* cult. Moradita, Chacarera e Tuni) foram as que apresentaram seqüências menos diversas.

IV. 1.2.1- Relações filogenéticas das seqüências similares a Tnt1

A diversidade do elemento Retrosol em todos os genótipos foi estudado através de uma análise filogenética. As 197 seqüências únicas amplificadas junto com representantes de Tnt1 A, B, C e Retrolyc1 A e B foram alinhadas e comparadas utilizando o método de distância. Três grupos foram obtidos com um *bootstrap* de 98%. (figura 13). O grupo I possui 87 seqüências incluindo tanto espécies silvestres quanto cultivadas e as 15 seqüências pertencentes a *S. brevidens* (Tabela 5). A identidade de nucleotídeos entre as seqüências varia entre 80-90 %. O grupo II tem 95 seqüências pertencentes a espécies silvestres e cultivadas com uma identidade de nucleotídeos maior a 90 % entre as seqüências. O grupo III possui 15 seqüências pertencentes a seis espécies silvestres: *S. acaule*, *S. hannemanni*, *S. kurtzianum*, *S. okadae* e *S. x sucrense* (tabela 5) com uma variabilidade de nucleotídeo entre 60-95 %, entretanto o grupo III apresenta uma alta variabilidade e não pode ser considerado como um grupo homogêneo.

O índice de πJC calculado para o total de seqüências únicas de Retrosol foi de 0,087 ± 0,0000571. Considerando os grupos previamente descritos, a variabilidade é menor para as seqüências do grupo II (0,030 ± 0,0000016) e para o grupo I 0,063 ± 0,0000035. O grupo III apresentou o maior valor (0,220 ± 0,000948) provavelmente representando mais de um grupo de seqüências relacionadas.

Foi realizada uma análise da diversidade de nucleotídeos considerando cada domínio do fragmento amplificado separadamente. A análise foi realizada para todas as seqüências juntas e para cada grupo obtido da figura 13. A região U3 foi a mais variável na análise do total de seqüências (figura 14). Porém quando a diversidade foi analisada em

cada grupo, o grupo II apresentou a região U3 menos diversa comparada com o resto dos grupos.

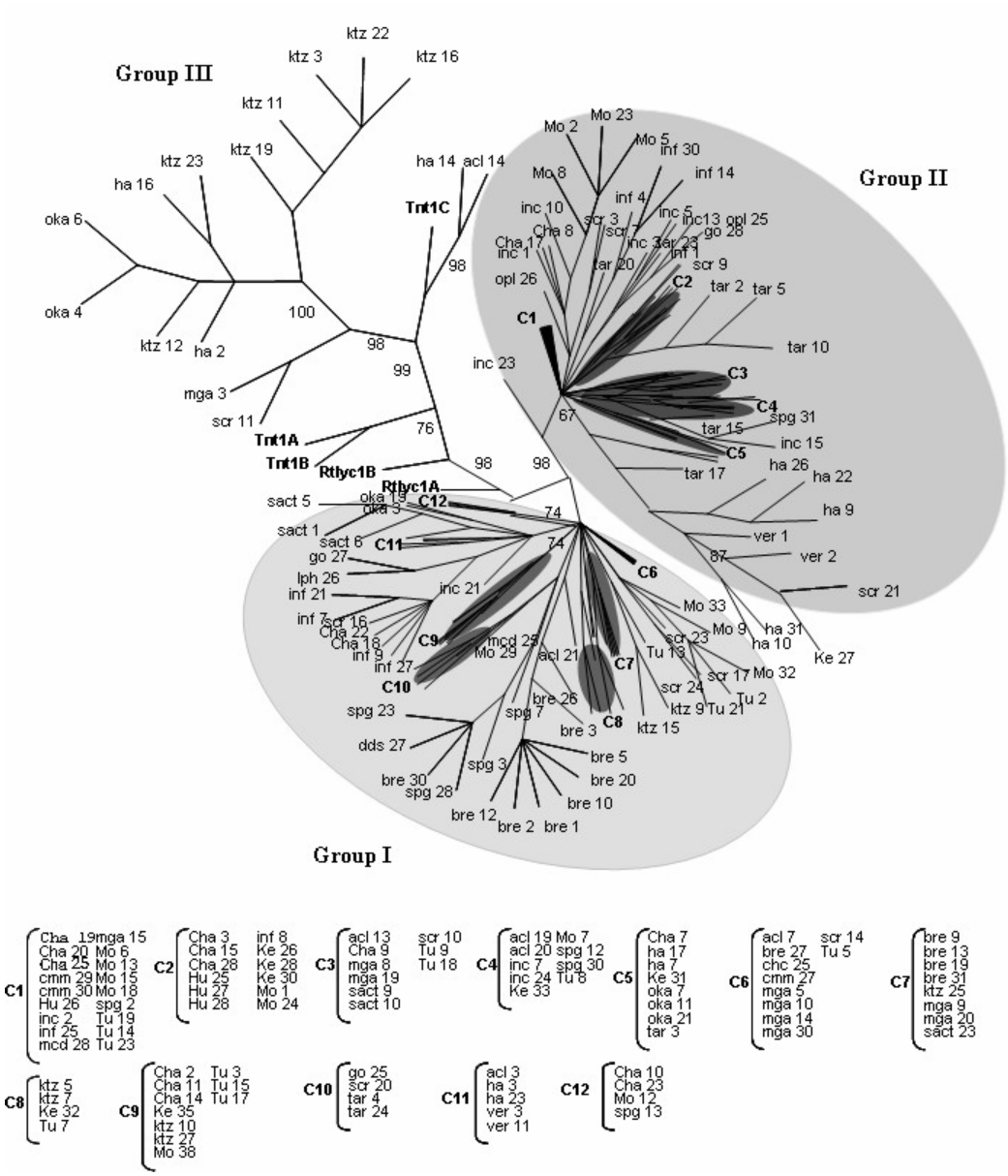

Figura 13. Análise filogenética da família Tnt1 em *Solanaceae*.

A análise filogenética foi feita com 197 seqüências amplificadas de diferentes espécies de *Solanum*. As espécies estão indicadas pelo código indicado na tabela 2 e diferentes clones estão indicados pelos números. Foram incluídos representantes de Tnt1 A, B e C, e Retrolyc1 A e B. As seqüências alinhadas abrangem o domínio da RNAseH, o Linker, a U3, R e parte da região U5. C seguido de um número indica o grupo de clones indicado entre chaves.

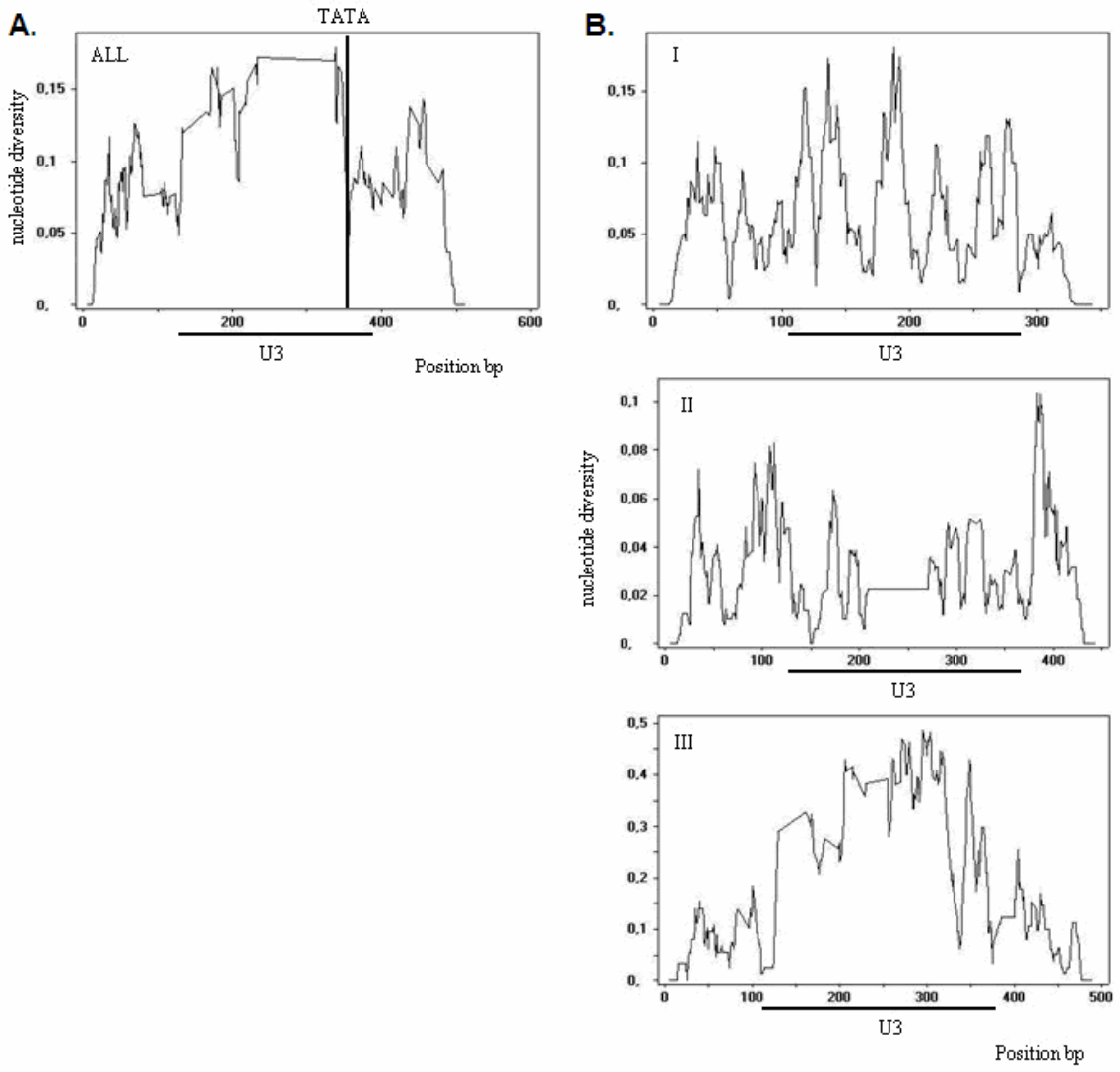

Figura 14. Divergência das seqüências ao longo do fragmento amplificado.

Os valores no eixo *x* correspondem à posição dos nucleotídeos no fragmento amplificado. Os valores no eixo *y* indicam a diversidade de nucleotídeos π, a porcentagem de divergência de nucleotídeos relativa ao número de bases informativas foi calculado usando uma janela de 10 pb e uma base como *step*, pelo programa DnaSp. A posição do *TATA box* e a região U3 estão indicados. A: análise para o total das seqüências amplificadas (All). B:análise para cada grupo da figura 13 separadamente (**I**: grupo I; **II**: grupo II; **III**: grupo III.) (os alinhamentos foram realizados sem considerar os *gaps*).

Os grupos derivados da análise filogenética foram relacionados com as espécies, o lugar de coleta e a altura sobre o nível do mar onde os acessos foram coletados. Nenhuma associação em particular foi observada (figura 15 e 16A e B), exceto o aparecimento de LTRs do grupo III acima de 1500m sobre nível do mar.

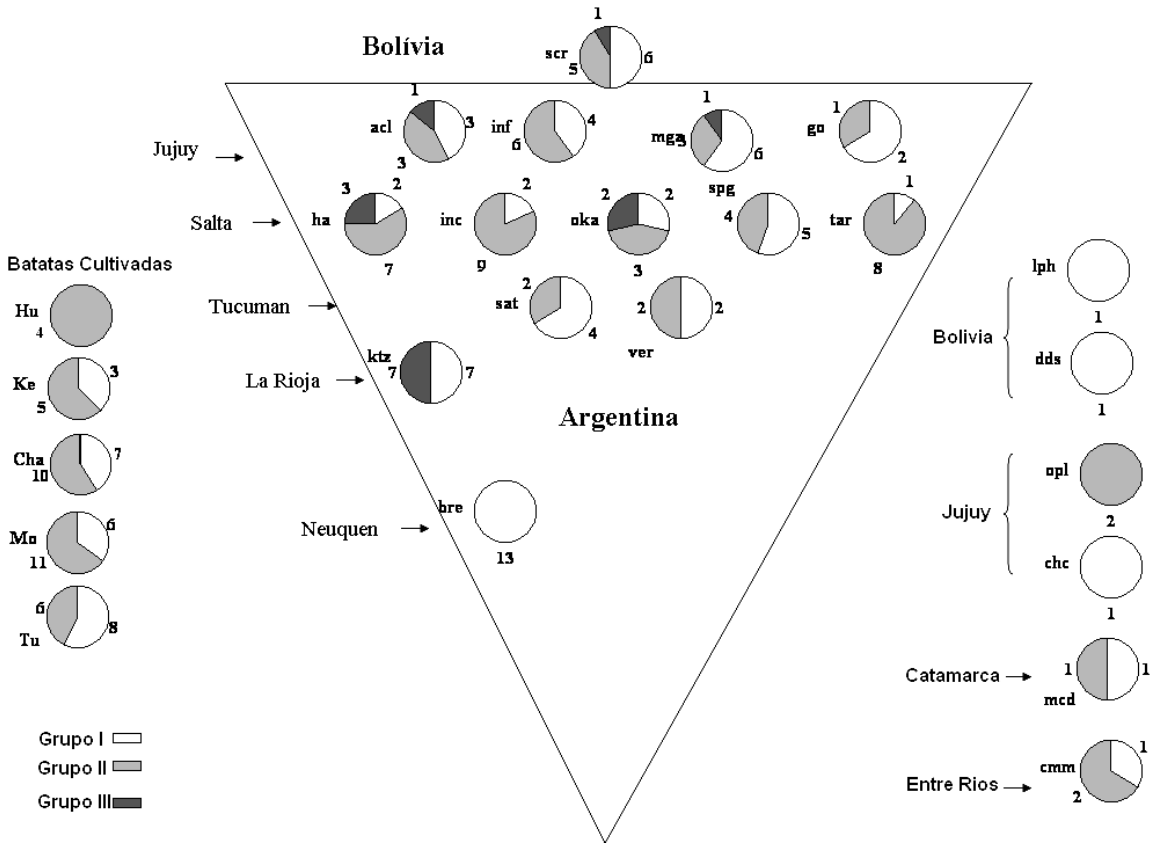

Figura 15. Esquema representando as localizações geográficas dos acessos das espécies e o número de seqüências dos grupos filogenéticos.

Os números indicam número de seqüências obtidas. Províncias da Argentina: Salta, Tucumán, La Rioja, Neuquén, Jujuy, Catamarca, Entre Rios. Códigos dos genótipos: Ke-*Solanum tuberosum* var Kennebec, Cha-*S.t.*ssp.*andigena* var.Chacarera, Mo-*S.t.*ssp.*andigena* var.Moradita, Tu-*S.t.*ssp.*andigena* var.Tuni, bre-*S.brevidens*, acl-*S.acaule*, chc-*S.chacoense*, cmm-*S.commersonii* ssp *malmeanum*, dds-*S.xdoddsii*, go-*S.gourlayi*, ha-*S.hannemanii*, ifd-S.*infundibuliforme,* inc-*S.incamayoense*, ktz-*S.kurtzianum*, lph-*S.leptophyes*, mga-*S.megistacrolubom,* mcd-

S.microdontum, oka-*S.okadae*, opl-*S.oplosence*, sat-*S.sanctae-rosae*, spg-*S.spegazzinii*, scr-*S.xsucrense*, tar-*S.tarijense*, vrn-*S.vernei*.

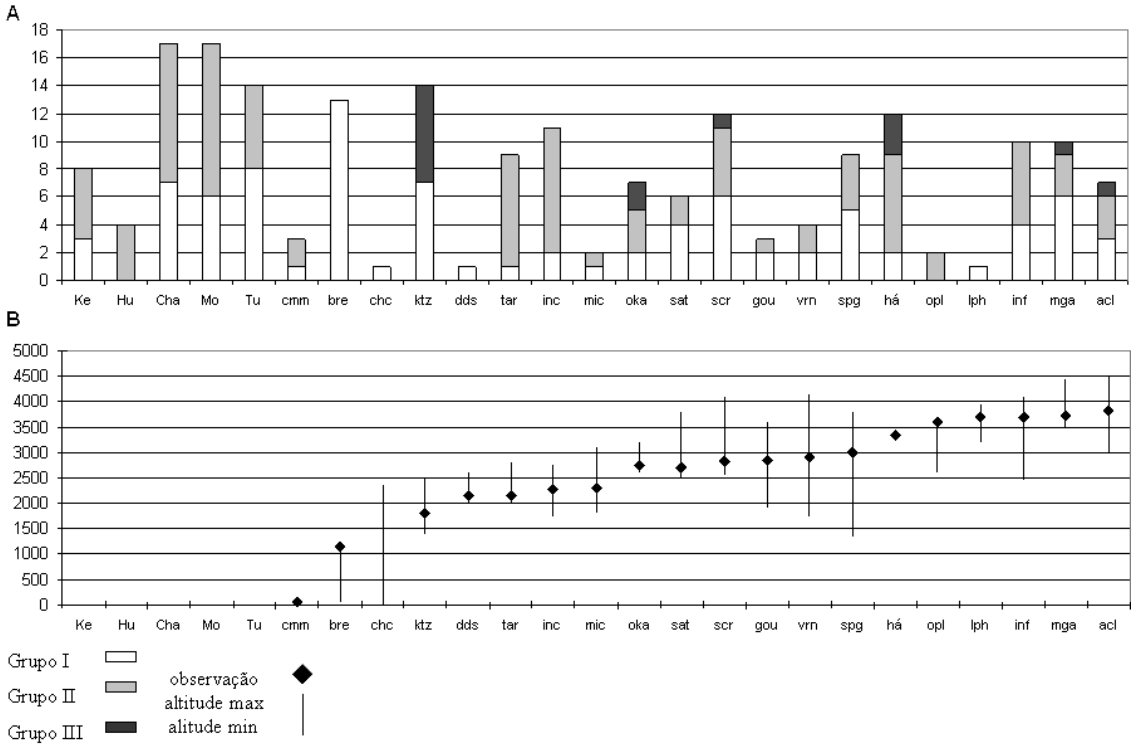

Figura 16. Distribuição dos grupos filogenéticos por genótipos e a altura sobre o nível do mar do acesso.

A. Distribuição do número de seqüências dos grupos filogenéticos por genótipo. B. Altura sobre o nível do mar do acesso utilizado (triangulo) e distribuição em altura da espécie na natureza. Códigos dos genótipos: estão detalhados na figura anterior.

IV. 1.2.2- Análises das regiões U3 de Retrosol

Analisando o alinhamento das regiões U3 dos grupos I e II de Retrosol, foi constatada a existência de lacunas sem seqüências chamadas *gaps* ou deleções. As deleções estavam na sua maioria flanqueadas por repetições de poucas bases na mesma direção

(tabela 6) (figura 17). O tamanho das deleções variou entre uma e 79 pb, mas a média foi de 21pb sem levar em consideração as deleções de uma base. O número de bases nas duplicações também é variável. Estas *deleções* flanqueadas por duplicações são um indício da perda de DNA dos elementos por recombinação ilegítima. O grupo III não foi considerado por não ser um grupo homogêneo.

Tabela 6- Associação de *deleções* e duplicações diretas

N°	N° pb da *deleção*	Grupos I/II	N° de bases repet.	N°de seq. com essa deleção	*Deleções* flanq.por DRs
1	4	I/II	2 (CT)	175	S
2	2	I	-	3	N
3	7	I	2 (CT)	1	S
4	8	I	3 (GGT)	3	S
5	11	II	3 (AAA)	95	S
6	10	I	4 (TTGG)	47	S
7	12	I/II	5 (GATTT)	95	S
8	20	II	2 (TC)	95	S
9	2	I/II	1 (A)	194	S
10	15	I/II	6(AAAGAA)	190	S
11	79	I	2 (TC)	1	S
12	2	I/II	2 (AA)	192	S
13	7	I	1 (A)	2	S
14	9	II	1 (A)	4	S
15	2	II	1 (G)	1	S
16	2	II	2 (AG)	13	S
17	4	II	2 (TC)	1	S
18	2	I	-	12	N
19	1	I	2 (AA)	78	S
20	3	I/II	3 (AAA)	178	S
21	16	I	3 (AAA)	1	S
22	1	I/II	1 (A)	75	S
23	3	I	1 (T)	1	S
24	2	I	2 (G)	3	S

N°:número da deleção. N° pb da deleção: número de pares de bases da deleção. N° de bases nas repet.: número de bases nas repetições. N° de seq. com essa deleção: número de seqüências com essa deleção. Deleções flanq. por DRs: deleções flanqueadas por duplicações repetidas. N: ausência. S: presença. grupos I/II: grupos filogenéticos da figura 13.

```
           10        20        30        40        50        60        70        80        90
   ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
1  TGGGGCCCATCCCCATGTGCCAAACTTTGTAAAATCTCACAATTTCAAAG-GTTGGCAAATTTGG----------GTATACATGTGCTGA
2  TGAGGTCCATCCCCATGTGCCAAACTTTGTTAAATCTCACAATTTCAAAG-GTTGACAGATTTGG----------ATATACACGTGCTGA
3  TGGGGTCCATCCCCATGTGCCAAACTTTGATAAATCCCACAATGTCAAAG-GTTGGCAGATTTGGAAGAAATTGGATATACATGTGCCAA
4  TGGGGTCCATCCC-ATGTGCCAAACTTTGATAAATCCCACAATGTCAAAG-GTTGGCAGATTTGGAAGAAATTGGATATACATGTGCCAA
5  TGGGGTCCATCCCCATGTGCCAAACTTTGATAAATCCCACAATGTCAAAG-GTTGGCAGATTTGGAAGAAATTGGATATACATGTGCCAA
6  TGGGGTCCATCCTCATGTCCCAAACTTTGATAAATCTCACAATGTTAAAGTGTTGGCAGATTTGGAAGATATTGGATATATATGTGCTGA
7  TGGGGTCCATCCCCATTTGCCAAA----------TCTCACAATGTCAAAAGGTTGGCAGATTTGGAAGATATTGGATATACATGTGCAGA
8  TGGGGTCCATCCCCATTTGCCAAA----------TCTCACAATGTCAAAAGGTTGGCAGATTTGGAAGATATTGGATATACATGTGCAGA
9  TGGGGTCCATCCCCATTTGCCAAA----------TCTCACAATGTCAAAAGGTTGGCAGATTTGGAAGATATTGGATATACATGTGCAGA
10 TGGGGCCCATCCCCATTTGCCAAA----------TCTCACAATGTCAAAAGGTTGGCAGATTTGGAAGATATTAGATATACATGTGCATA
11 TGGGGTCCATCCCCATTTGCCAAA----------TCTCACAATATCAAAAGGTTGGCAGATTTGGAAGATATTGGATATACATGTGCATA
12 TGGGGTCCATCCCCATTTGCCAAA----------TCTCACAATATCAAAAGGTTGGCAGATTTGGAAGATATTGGATATACATGTGCATA

           100       110       120       130       140       150       160       170       180
   ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
1  TTTATTT-------------TTCAGCTTTCCTCAAAAGTCAAATGTTGGTAAGCTGTGAAGAAGAAAAAA---------------TTGTT
2  TTTATTT-------------TTCAGCTTTCCTCAAAAGTCAAATGTTGGTAACGTGTGAAGAAGA---TT---------------TTTTT
3  TTTACTTGGCTGATTTTTCTTCCAACTTCCCTTAAA-GTCAAATGTTGGTAAGCTGTGAAGAAGAAAAAAA---------------TGTT
4  TTTACTTGGCTGATTTTTCTTCCAACTTCCCTTAAA-GTCAAATGTTGGTAAGCTGTGAAGAAGAAAAAAA---------------TGTT
5  TTTACTTGGCTGATTTTTCTTCCAACTTCCCTTAAA-GTCAAATGTTGGTAAGCTGTGAAGAAGAAAAAAA---------------TGTT
6  TTT--------GTTTT----TCCAGCTATCCTCAAAAGTCAAATGTTGGTAAGCTGTGAAAAAGAAAAAA--------------------
7  TTT---------TTTC------------------AAA-------GGTTGGTAAGCTGCAAAGAAGAAAAATAAAAAGAAGAAAAAATTATT
8  TTT---------TTTC------------------AAA-------GGTTGGTAAGCTGCAAAGAAGAAAAATAAAAAGAAGAAAAAATTATT
9  TTT---------TTTC------------------AAT-------GGTTGGTAAGCTGCAAAGAAGAAAAAT---------------TTGTT
10 TTT---------TTTC------------------AAA-------GGTTGGTAAGCTGTAAAGAAGAAAAA---------------ATAGTT
11 TTT---------TTTC------------------AAA-------GGTTGGTAAGCTGCAAAGAAGAAAAA---------------AATGTT
12 TTT---------TTT-----------------------------------------------------------------------------

           190       200       210       220       230       240       250       260       270
   ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
1  TTTCTTCGGTAGCATAGTTGGAGGC-TACTGCATTTGTTCTCTATAAAAGGGAGGGTCATTCTTCATTCTTAAA-TGCACCAAAAACAAA
2  TTTCTTCTTTAGCATATTTGGAGGT-TGCTGCATTTGTTATCTATAAAAGGGAGGGTCATTCTTCATTCTCAAA-TGCACCAAAAACAAA
3  TTTCTTCTGTAGCATATTTGGAGGC-TACTGCATTTGTTCTCTATAAAAGGGAGGATCATTCTTCATTCT-AAAATTCATCAAAAACAAA
4  TTTCTTCTGTAGCATATTTGGAGGC-TACTGCATTTGTTCTCTATAAAAGGGAGGATCATTCTTCATTCT-AAAATTCATCAAAAACAAA
5  TTTCTTCTGTAGCATATTTGGAGGC-TACTGCATTTGTTCTCTATAAAAGGGAGGATCATTCTTCATTCT-AAAATTCATCAAAAACAAA
6  ----------------TTGTGAGGT-TACTCCATTTGTTATCTATAAAAGGGAGGGCCATTCTTCATTCTCAAA-TGCACCAGAAACAAA
7  TTTCTTCTGTAGCGTAGTTGGAAAAATACTGCACTTGTTATCTATAAAAGGGAGAGTC-TTCTTCATTCTCAAAATGCACCAAAAACAAA
8  TTTCTTCTGTAGCGTAGTTGGAAAAATACTGCACTTGTTATCTATAAAAGGGAGAGTC-TTCTTCATTCTCAAAATGCACCAAAAACAAA
9  TTTCTTCTGTAGCATAGTTGGAAATATACTGCACTTGTTATCTATAAAAGGGAGAGTC-TTCTTCATTCTCAAATTGCACCAAAAACAAA
10 TTTCATCTGTAGCATAGTTGAAAAAATACTGCACTTGTTATCTATAAAAGGGAGAGTC-TTCTTTATTCTCAAA-TGCACCAAAAACAAA
11 TTTCTTCTGTAGCATAGTTGGAAAAATATTGCACTTGTTATCTATAAAAGGGAGAGTC-TTCTTCATTCTCAAA-TGCACGAAAAACAAA
12 ---CTTCTGTAGCATAGTTGGAAAAATACTGCACTTGTTATCTATAAAAGGGAG--TC-TTCTTCATTCTCAAA-TGCACCAAAAACAAA
```

Figura 17. Alinhamento da região U3 de seqüências de Retrosol.

As seqüências completas foram alinhadas usando Clustal W e a região da LTR foi extraída e editada manualmente com Bioedit. Os traços indicam deleções. Caixas cinza: duplicações repetidas nos limites das deleções. Bases sublinhadas:bases das duplicações repetidas remanescentes. (1-12): regiões U3 de 12 seqüências diferentes de Retrosol. 1-6: grupo I. 7-12: grupo II.

IV. 1.2.3 - Análise da região RNAseH com NETWORK

As análises do NETWORK foram projetadas para investigar as conexões entre seqüências relacionadas, permitindo a identificação de nós ancestrais persistentes, ligações múltiplas e redes, que não são resolvidas ao aplicar um programa filogenético

convencional. No NETWORK, as seqüências idênticas mais representadas, nomeadas α ou cópia mestre, ocupam uma posição central na rede com os números mais elevados de conexões sugerindo que a seqüência forma parte de um elemento ativo. Esta aproximação é baseada no princípio da parcimônia, que conecta séries de dados com o número mínimo de etapas evolutivas. As redes resultam de recombinações e/ou homoplasia.

Considerando a família Tnt1 nos gêneros *Solanum*, *Lycopersicon* e *Nicotiana* como uma população de seqüências com um antepassado comum, uma analise NETWORK foi realizada com o intuito de testar a hipótese da existência de uma seqüência comum α ou mestre na família Tnt1. Os 53 nucleotídeos que correspondem à última porção do domínio da RNAseH de um total de 453 seqüências similares a Tnt1 (197 seqüências de Retrosol relatadas neste estudo, 140 seqüências de Tnt1 caracterizadas em 7 espécies de *Nicotiana* e 116 seqüências de Retrolyc1 de 4 espécies de *Lycopersicon*) foram usadas na análise. Primeiramente as seqüências foram alinhadas utilizando o programa DNA *Alignment* 1.1.21 *software* (anexo 1) e logo analisadas com o programa NETWORK 4.2.0.1. A rede resultante (figura 18) revelou que Tnt1, Retrolyc1 e Retrosol não compartilharam nenhuma seqüência de RNAseH, o que foi revelado pela ausência de nós compartilhados. Espécies diferentes dentro de cada gênero podem ou não compartilhar uma determinada seqüência α.

Nas seqüências de Retrosol, o nó original α observado (α1) representa 25.5% de todas as seqüências de Retrosol e abrange as espécies: *S. hannemanii, S. incamayoense, S. infundibuliforme*, *S. microdontum, S. megistacrolobum*, *S. okadae*, *S. sanctarosae*, *S. spegazzinii*, *S. tarijense*, *S. x sucrense, S. x vernei* e todos os cultivares de *S. tuberosum*. Vale a pena mencionar que 46 das 50 seqüências α1 pertencem ao grupo II, que foi o grupo filogenético menos diverso. Além disso, dentro do grupo α1 foi possível identificar U3

idênticas que pertencem a espécies e cultivares diferentes. Um total de 13 seqüências não redundante de *S. brevidens* são conectadas com a seqüência α da rede através da mudança de um nucleotídeo. Nenhumas das seqüências que pertencem ao grupo III aparecem no conjunto α1. As seqüências do grupo III estão situadas em uma posição mais afastada da rede sugerindo um relacionamento mais distante com a seqüência α e provavelmente não foram originadas de α1.

As seqüências Tnt1 originaram-se de dois nós α: α2 com 24% (34 seqüências) e α3 com 12% (17 seqüências) das seqüências (figura 18 B). As seqüências α2 pertencem a: *Nicotiana tabacum*, *N. debneyi*, *N. benthamiana, N. plumbaginifolia*, *N. sylvestris*, abrangendo as três subfamílias Tnt1 (A, B e C). α3 reune seqüências de *N.tabacum*, *N. glauca*, *N. debneyi*, *N. benthamiana*, *N. plumbaginifolia*, *N. sylvestris*, agrupando as subfamílias Tnt1 A e C. É interessante ressaltar que Tnt1-94, para o qual foi demonstrada atividade transpositional, agrupa no nó α3.

Retrolyc1 mostrou três nós α diferentes: α4 com 18.9% (22 seqüências), α5 com 17% (20 seqüências) e α6 com 13% (15 seqüências) das seqüências Retrolyc1 (figura 18 B). α4 inclui seqüências de *Lycopersicon pimpinellifolium* e de *L. esculentum*, todas as seqüências da subfamília Retrolyc1 B. α5 tem seqüências pertencentes a *L. peruvianum* e a subfamília Retrolyc1 A, enquanto α6 possui 14 seqüências que pertencem a *L. peruvianum,* e somente uma a *L. hirsutum* . U3 idênticas de Retrolyc1 A foram compartilhados por espécies diferentes no nó α4

Outros nós com números diferentes de seqüências estão conectados com os nós α1, α2, α3, α4, α5 e α6 centrais, sugerindo que estes membros poderiam ser ativos e poderiam

contribuir à expansão de cada elemento dentro do seu gênero correspondente (figura 18 B). A rede não mostrou nenhum excesso de reticulações multidimensionais e por conseqüência existe uma homoplasia insignificante.

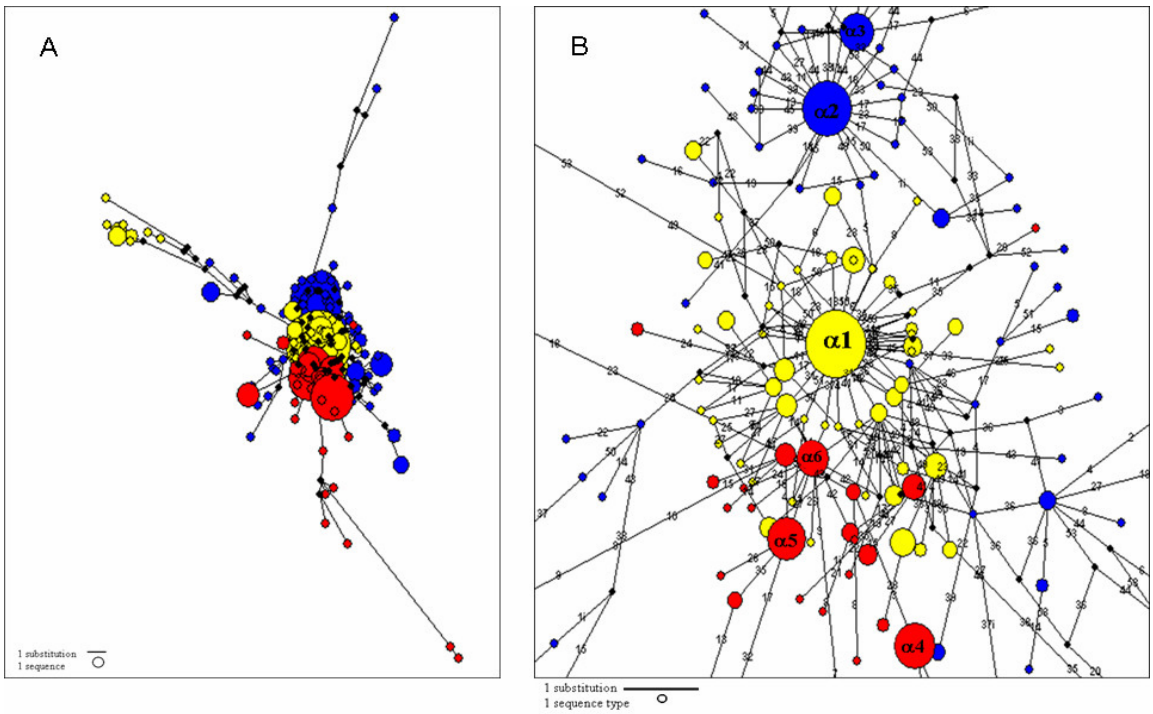

Figura 18. ***Median-joining Networks*** **da região codificante RNAseH.**

A rede foi construída usando 453 seqüências da família Tnt1. Os círculos denotam tipos de seqüências: círculos amarelos são seqüências de Retrosol, círculos azuis são seqüências de Tnt1, e círculos vermelhos denotam seqüências de Retrolyc1 (A). O tamanho do círculo é proporcional ao número de seqüências que abrange. As linhas denotam substituições. Os nós reconstruídos são identificados como círculos pretos; são ausentes na série de dados porque não foram amostrados ou estão extintos na população de seqüências. (B) Detalhe da rede de *median-joining* da região RNAseH. Os rótulos nos nós ilustram a nomenclatura usada na seqüência mestre original de cada gênero, chamada α: α1 é a cópia mestre de Retrosol, α2 e α3 são de Tnt1 e α4, α5 e α6 são de Retrolyc1.

IV. 1.2.4 - Distribuição do retrotransposon Retrosol

Foi estudado o padrão de distribuição do elemento Retrosol por Southern blot utilizando como sondas a região codificante (RNAseH) e as regiões U3 dos grupos I e II caracterizados na secção IV.1.2.1. O DNA genômico dos 24 genótipos de *Solanum*, mais *Lycopersicon peruvianum* e *Oryza sativa* (arroz), foram submetidas à hidrólise enzimática pela endonuclease de restrição *Eco*RI e transferidos a uma membrana. As sondas utilizadas na hibridação provieram de amplificações por PCR dos plasmídeos com os fragmentos de interesse, e sua posterior purificação. A sonda da região codificante RNAseH junto com o Linker foi obtida usando os iniciadores Avi e PTT com um tamanho de 119pb. O padrão de hibridação usando esta sonda foi correlacionado com o de um elemento inteiro, já que possui uma porção da região codificante. As sondas das regiões U3 dos grupos I e II, foram obtidas usando os iniciadores GI-F e GI-R com um tamanho de fragmento de 177pb e os iniciadores GII-F e GII-R com um tamanho de 158pb, respectivamente. Os iniciadores dos grupos I e II são específicos para cada grupo. O padrão de hibridação do grupo III não foi determinado devido à impossibilidade de desenhar iniciadores para o conjunto de seqüências correspondentes ao grupo como um todo dada sua diversidade. Como já foi mencionado o grupo III é heterogêneo.

Analisando as auto-radiografias do padrão de hibridação da sonda da região codificante (RNAseH) pode ser observado que o elemento Retrosol é especifico de *Solanaceae* já que não se observa sinal na gramínea arroz. *Lycopersicon peruvianum* mostra um sinal de hibridação mais intensa que as das espécies do gênero *Solanum* mostrando alta similaridade na região codificante entre os elementos Retrolyc1 e Retrosol. Comparando as espécies de *Solanum,* o padrão de hibridação é heterogêneo, e nehuma

relação foi possível de estabelecer a respeito do caracter cultivado ou silvestre das espécies, nem a respeito das diferentes ploídias ou relações filogenéticas (tabela 1) (figura 19). A espécie *S. brevidens* (não produz tubérculos) apresenta poucas copias a respeito da região codificante e dos grupos I (dados não mostrados) ou II (figura 19).

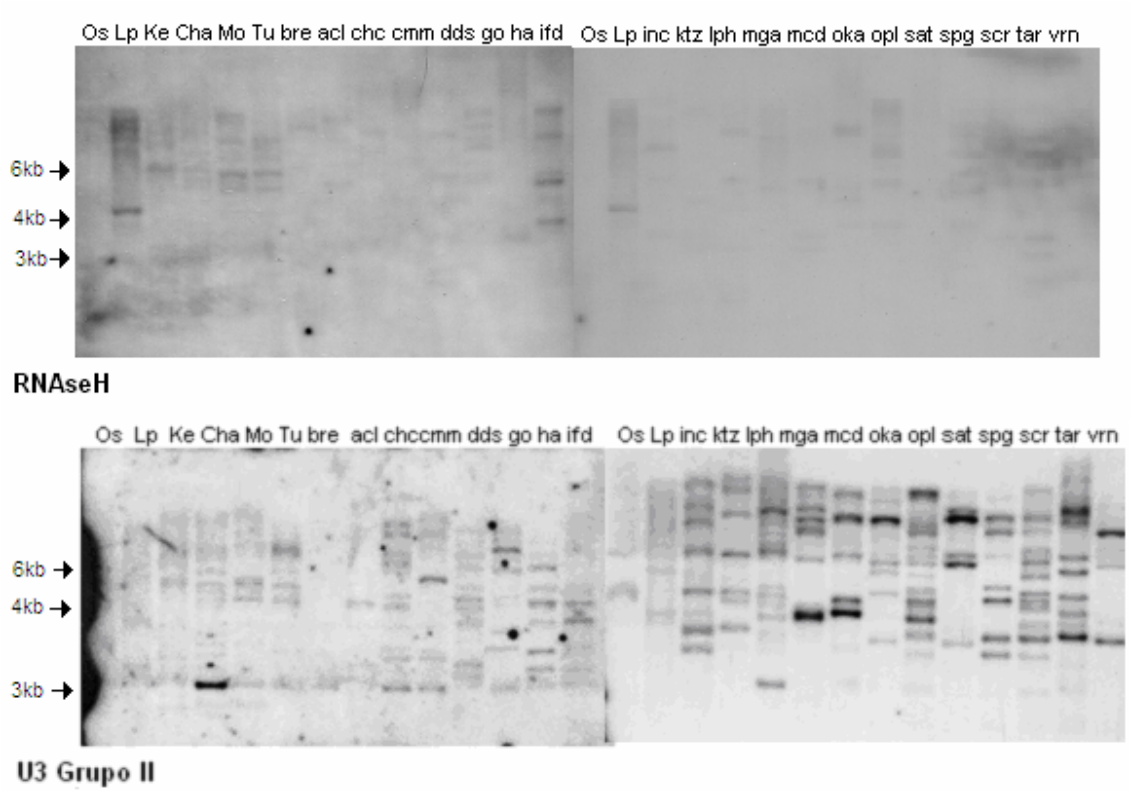

Figura 19. Análise por hibridação molecular do polimorfismo existente entre as espécies a partir das seqüências da região RNAseH e U3 dentro dos genomas das espécies de *Solanum*.

DNA genômico foi submetido a hidrólise com a endonuclease de restrição *Eco* RI. As sondas utilizadas foram: região U3 grupo II (158pb), e região RNAseH (119pb). O tamanho molecular está indicado pelas setas e números na lateral. Códigos dos genótipos: Os-*Oryza sativa* (arroz), Lp-*Lycopersicon peruvianum,* Ke-*Solanum tuberosum* var Kennebec, Cha-*S.t.*ssp.*andigena* var.Chacarera, Mo-*S.t.*ssp.*andigena* var.Moradita, Tu-*S.t.*ssp.*andigena* var.Tuni, bre-*S.brevidens*, acl-*S.acaule*, chc-*S.chacoense*, cmm-*S.commersonii* ssp *malmeanum*, dds-*S.xdoddsii*, go-*S.gourlayi*, ha-*S.hannemanii*, ifd-S.*infundibuliforme,* inc-*S.incamayoense*, ktz-*S.kurtzianum*, lph-*S.leptophyes*, mga-*S.megistacrolubom,* mcd-*S.microdontum*, oka-*S.okadae*, opl-*S.oplosence*, sat-*S.sanctae-rosae*, spg-*S.spegazzinii*, scr-*S.xsucrense*, tar-*S.tarijense*, vrn-*S.vernei*.

IV.2 – Determinaçõa do número de cópias de retrotransposons similares a Tnt1 em diversas espécies de *Solanaceae*.

Uma PCR quantitativa (qPCR) em tempo-real foi realizada com o objetivo de determinar o número absoluto de cópias dos retrotransposons da família Tnt1 em diferentes espécies pertencentes a *Solanaceae*. Do mesmo modo, foi determinado o número total de LTRs do retrotransposon Retrosol em espécies de *Solanum*. As espécies foram escolhidas por terem os tamanhos de seus genomas determinados (ver item III.8).

O número absoluto de cópias foi determinado comparando o valor das amostras de DNA das espécies com uma curva padrão. A curva padrão foi definida utilizando diluições seriadas de um plasmídeo contendo a fragmento de interesse. Depois da extração de DNA, tanto das espécies quanto dos plasmídeos, foram preparadas diluições de estoque de 10ng/µL a partir das quais as diluições de trabalho (1ng/µL) foram obtidas. Foram utilizados 5µL das diluições de trabalho para realizar as quantificações.

A curva padrão foi produzida determinando a linha de tendência da regressão entre os valores de C_T (valor da fluorescência acima do *threshold*) e o logaritmo do número de cópias. A curva padrão é uma curva linear descrita pelos parâmetros $x=my + b$, onde m representa –1/log (eficiência da PCR), b é o logaritmo da quantidade do produto amplificado no *threshold* dividido pelo logaritmo da eficiência da PCR. A eficiência da amplificação por PCR (E) foi calculada segundo a fórmula (Kubista *et al*. 2006):

$$E = 10^{-1/\text{pendente}} - 1$$

O número inicial de cópias do plasmídeo de cada diluição foi calculado tendo em conta a massa, concentração e tamanho do plasmídeo, segundo a fórmula:

Massa de uma copia de plasmídeo = [3015 bp (PGEMEASY Vector) + pb do fragmento] [1,096.10-21g/pb]

Para determinar o número de cópias do retrotransposon similar a Tnt1 foram utilizados os iniciadores Avi-PPT gerando um fragmento de 119pb contendo a porção final da RNAseH (53 nucleotídeos), e o valor obtido foi considerado cópia inteira do elemento por conter uma porção da região codificante. Por outro lado, para a quantificação das LTRs de Retrosol foram utilizados os iniciadores LTR-F e LTR-R que amplificam uma porção de 68 pb da região U3 tanto do grupo I quanto do grupo II gerados na análise filogenética (ver item IV.1.2.1). Ambas as combinações de iniciadores foram testadas em uma reação de PCR utilizando os DNA genômicos e plasmidial gerando bandas únicas do tamanho esperado visualizadas em gel de agarose com brometo de etídeo. Foi escolhido o plasmídeo contendo o clone vrn1 amplificado de *Solanum vernei* para quantificar a RNAseH e as LTRs.

As diluições da curva padrão (figura 20) e das amostras foram amplificadas usando as mesmas diluições de iniciadores e em idênticas condições de amplificação. Por tanto, foi considerada a eficiência da curva padrão como a mesma das amostras em cada experimento.

Foram realizados quatro experimentos. Os experimentos 1 (Exp. 1) e 2 (Exp. 2) determinaram o número de cópias dos elementos (iniciadorea Avi-Ol16) e os experimentos 4 (Exp. 4) e 5 (Exp. 5) determinaram o número de cópias das LTRs de Retrosol (iniciadores LTR-F-LTR-R).

A figura 21 apresenta a curva padrão do Exp.1, indicando os diferentes parâmetros. Os dados completos de todos os experimentos estão detalhados na tabela 7.

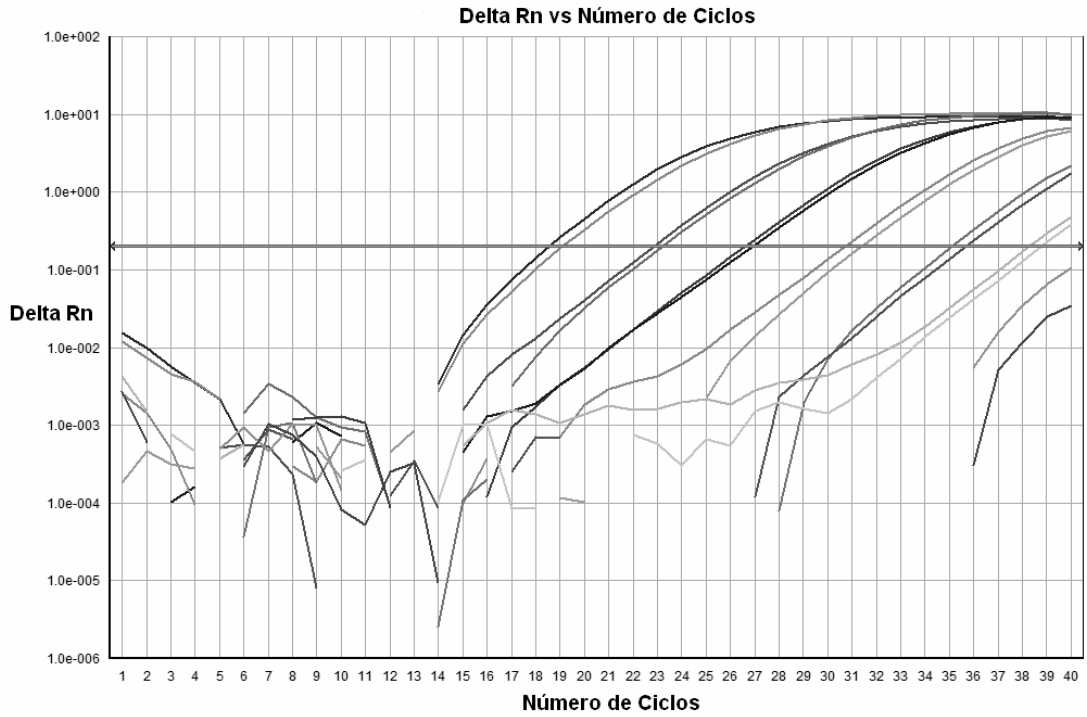

Figura 20. Detecção do produto amplificado por qPCR das diluições de plasmídeo contendo o domínio RNAseH (Exp1).

Foram usadas seis diluições (de 5 X 10^{-1} até 5 X 10^{-6} ng de DNA de plasmídeo). A emissão de fluorescência foi medida em cada ciclo no final da fase. Delta Rn: emissão de fluorescência com subtração da linha de base de PCR.

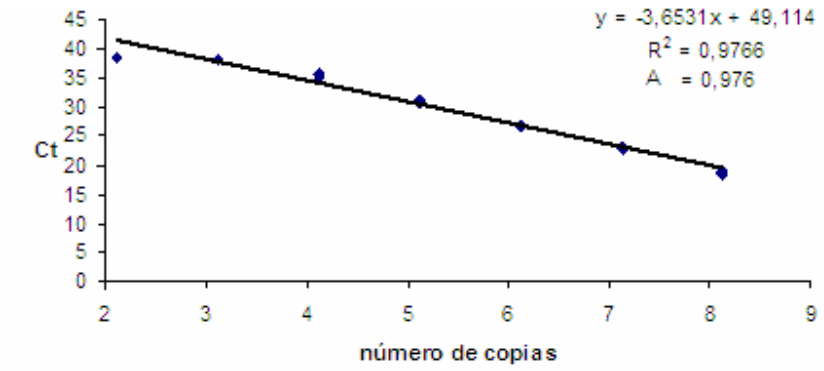

Figura 21. Curva padrão das diluições de plasmídeo contendo o domínio RNAseH.

O número de copias está expresso em escala logarítmica e o C_T representa a fluorescência acima do *threshold*. A inclinação e a interseção foram calculadas a partir da equação que descreve a curva padrão.

Tabela 7. Dados da curva padrão dos diferentes experimentos.

Parâmetros	Exp.1	Exp. 2	Exp. 3	Exp. 4
Inclinação	-3,65	-3,97	-3,13	-3,06
Interseção	49,1	49,4	38,5	38,12
Eficiência	87,82%	78,4%	108,3%	112%
Amplificação	0,976	1,78	2,08	2,12
R^2	0,98	0,99	0,98	0,98

As curvas de dissociação dos produtos foram geradas para todos os experimentos. A análise da curva de dissociaçaõ permite identificar a amplificação de um produto único ou produtos múltiplos, e/ou dímeros de iniciadores. As curvas de dissociação do Exp.1 das amostras sugere que fumo (Nt) e tomate (Lp) amplificaram produtos múltiplos por apresentar curvas largas; e as espécies restantes amplificaram produtos únicos por apresentar curvas estreitas (figura 22A). A curva de dissociação da curva padrão do Exp.1 mostrou curvas estreitas (figura 22B). O Exp. 2 mostrou resultados similares ao Exp.1. No caso dos Exp. 3 e 4 (LTRs) os resultados das curvas de dissociação sugerem produtos únicos (dados não apresentados).

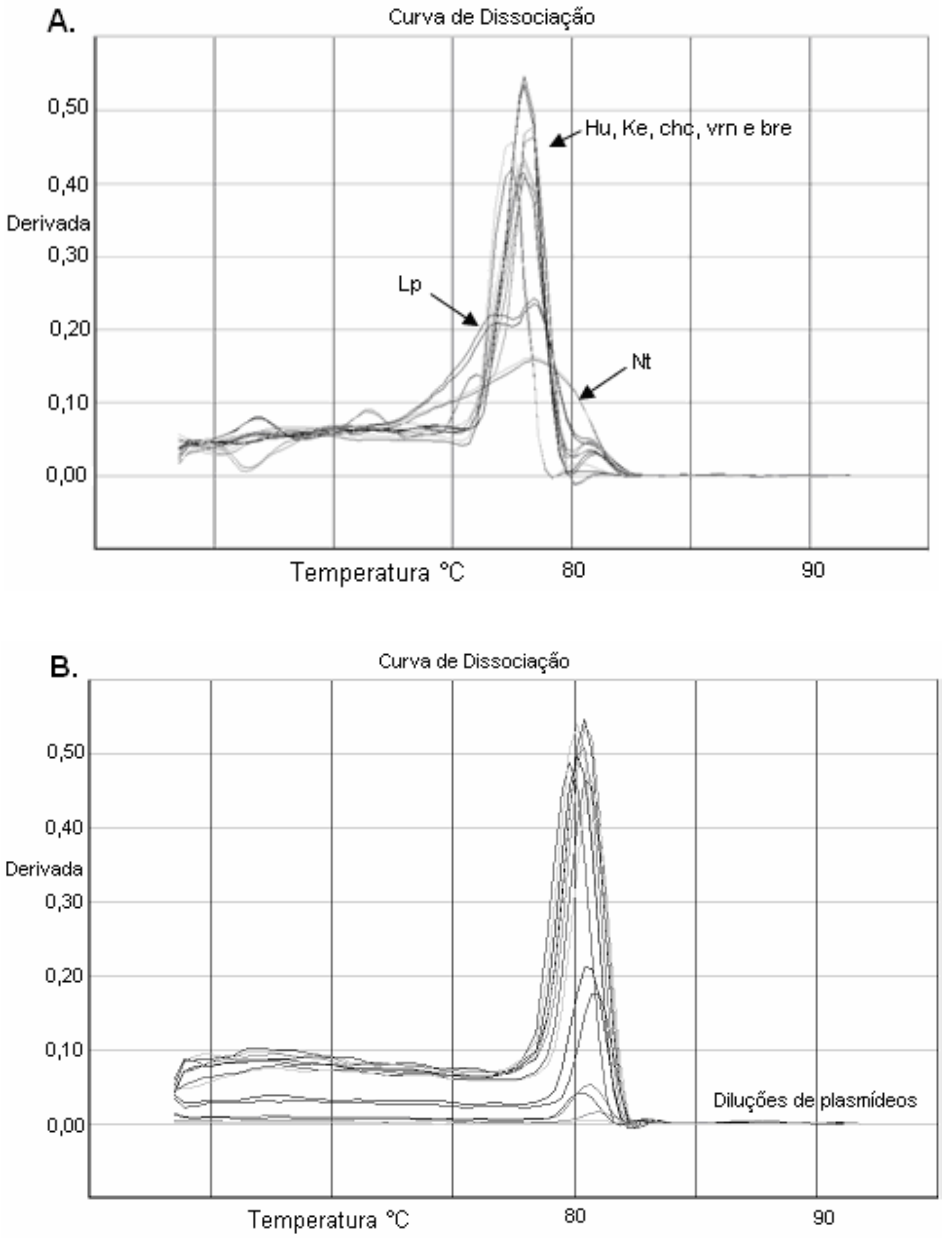

Figura 22 A e B. Curva de Dissociação das diluições das amostras (A) e (B) do plasmídeo.

Derivada: derivada negativa da emissão de fluorescência do Exp1. Nt: *Nicotiana tabacum*, Lp: *Lycopersicon peruvianum*, Hu: *Solanum tuberosum* cult. Huinkul, Ke: *S. tuberosum* cult. Kennebec, chc: *S. chacoense*, vrn: *S. vernei*, bre: *S. brevidens.*

Os produtos da qPCR da RNAseH foram migrados em gel de agarose e posteriormente visualizados com brometo de etídeo. Esse procedimento permitiu verificar a existência de uma banda única e do tamanho esperado (119pb) para todas as amostras (figura 23).

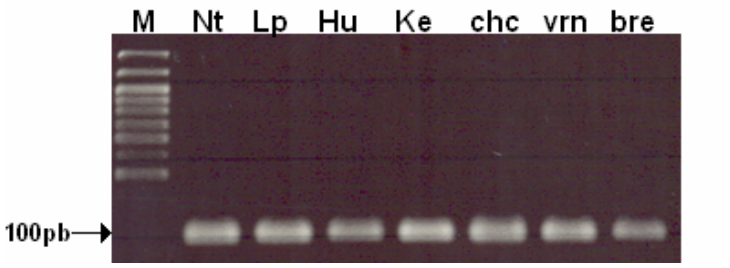

Figura 23. Gel de agarose dos produtos da PCR em tempo real amplificando o domínio RNAseH.

M: marcador molecular. Nt-*Nicotiana tabacum*, Lp-*Lycopersicon peruvianum*, Hu-*Solanum tuberosum* cult. Huinkul, Ke-*S.tuberosum* cult. Kennebec, chc-*S. chacoense*, vrn-*S. vernei*, bre-*S. brevidens.*

Com o objetivo de conhecer a especificidade dos iniciadores Avi-PTT na amplificação de retrotransposons similares a Tnt1, os produtos da PCR em tempo real foram clonados e posteriormente seqüenciados. Obteve-se um total de 62 clones distribuídos da seguinte forma: *Nicotiana tabacu,* 9; *Lycopersicon peruvianum,*10; *Solanum tuberosum* cult. Huinkul, 10; *S. tuberosum* cult. Kennebec, 8; *S. chacoense,* 8; *S. vernei,* 9; e *S. brevidens,* 8. A análise das seqüências contra o banco de dados do NCBI confirmou que o total de clones apresentava similaridade com Tnt1 confirmando a especifidade da amplificação.

As seqüências obtidas da qPCR foram analisadas com o intuito de explicar as curvas de dissociaçãos de fumo (Nt) e tomate (Lp). Segundo Rire e colaboradoes (1996) a forma da curva de dissociação é função do tamanho, conteúdo e seqüência do produto

amplificado. As seqüências não apresentaram diferenças significativas em tamanho ou conteúdo de GC (figura 24), portanto os múltiplos produtos poderiam ser resultado de fragmentos diferentes, com o mesmo conteúdo GC, mas com distribuição distinta deste conteúdo.

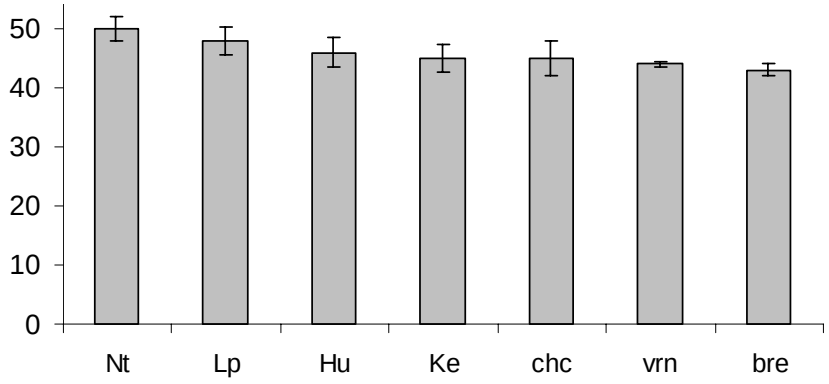

Figura 24. Porcentagem de GC das seqüências obtidas na qPCR.

Código dos genótipos: Nt-*Nicotiana tabacum*, Lp-*Lycopersicon peruvianum*, Hu-*Solanum tuberosum* cult. Huinkul, Ke-*S.t.* cult. Kennebec, chc-*S. chacoense*, vrn-*S. vernei*, bre-*S. brevidens.*

O número absoluto de cópias das espécies da RNAseH e das LTRs, foi determinado utilizando 5µL da solução de trabalho (1ng/µL). O número de cópias por genoma surge a partir da relação do valor anterior com o número de genomas haplóides contidos nesse volume quantificado. O valor da solução de trabalho de Hu foi corregido em seguida a quantificação do DNA no fluorômetro.

Os resultados dos Exp. 1 e 2, do número de cópias absoluto dos elementos similares a Tnt1, estão representados na figura 25. Os valores de cada experimento são diferentes comparando as mesmas amostras, embora a tendência é mantida. Fumo (Nt) apresenta um número muito alto de cópias, em tomate o valor é menor e nas espécies de *Solanum* poucas cópias são determinadas. O desvio padrão é alto nas amostras de fumo (Nt) e tomate (Lp)

mais vale ressaltar que a determinação do número absoluto de cópias se basea na curva padrão onde um dos eixos (número de cópias) está em escala logarítmica o que geraria desvios padrão altos quando convertidos em cópias. Os dados do número de cópias por genoma haplóide estão descritos na tabela 8.

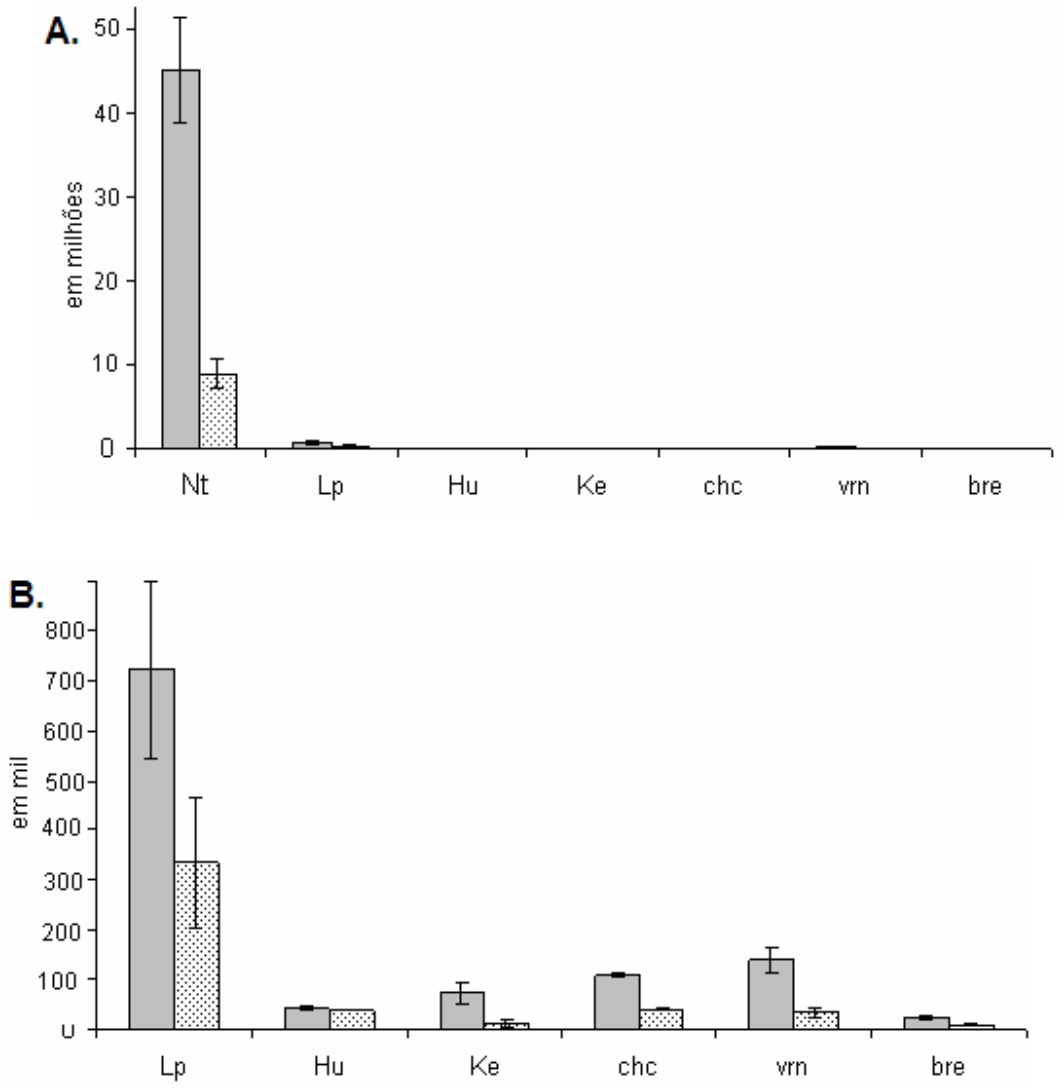

Figura 25. Número de copias de retrotransposon similares a Tnt1 nos diferentes genótipos.

Cinza: Experimento 1. Pontilhado: Experimento 2. A: com todos os genótipos. B: sem Nt. Nt: *Nicotiana tabacum*, Lp: *Lycopersicon peruvianum*, Hu: *Solanum tuberosum* cult. Huinkul, *tuberosum* cult. Kennebec, chc: *S. chacoense*, vrn: *S. vernei*, bre: *S. brevidens.*

Os resultados da determinação do número de cópias das LTRs dos Exp. 3 e 4 estão apresentados na figura 26 e na tabela 9.

Tabela 8. Número de cópias dos experimentos 1 e 2 por genoma haplóide (1C).

Genotipos	Exp. 1	Exp. 2	Tamanho (pg)	1C (5µl)
NT	41.000	8.129	4,55	1.100
Lp	166	77	1,15	4.347
Hu	26	23	1,75	1.657
Ke	28	4	1,75	2.571
chc	17	6	0,63	6.508
vrn	25	6	0,8	5.625
bre	4	2	0,76	5.460

Exp.: Experimento. Nt: *Nicotiana tabacum*, Lp: *Lycopersicon peruvianum*, Hu: *Solanum tuberosum* cult. Huinkul, *tuberosum* cult. Kennebec, chc: *S. chacoense*, vrn: *S. vernei*, bre: *S. brevidens.* Tamanho (pg): conteúdo genômico haplóide em picogramas. 1C (5µl): número de genomas haplóides em 5µl, quantificados na reação de PCR em tempo real.

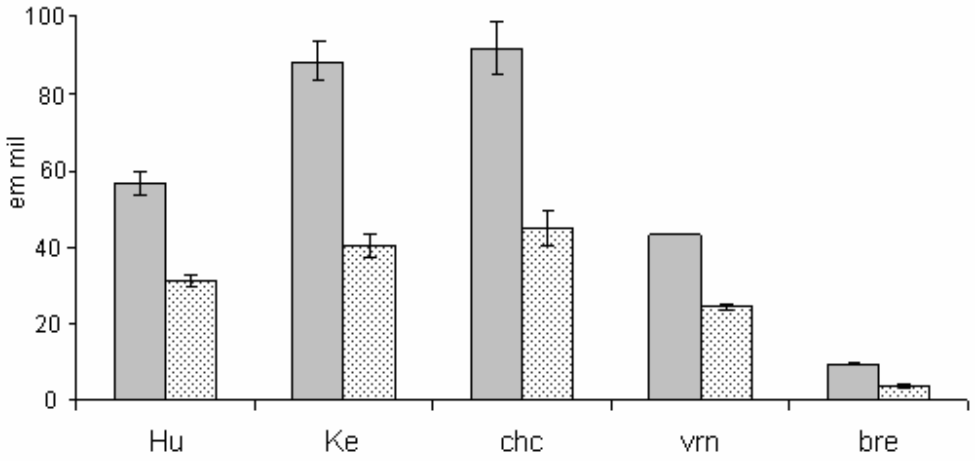

Figura 26. Número de copias das LTRs de Retrosol.

Experimento 3.:cinza. Experimento 4:pontilhado. Hu: *Solanum tuberosum* cult. Huinkul, *tuberosum* cult. Kennebec, chc: *S. chacoense*, vrn: *S. vernei*, bre: *S. brevidens.*

Tabela 9. Número de copias dos experimentos das LTRs de Retrosol.

Genótipos	Exp. 3	Exp. 4	Tamanho (pg)	1C (5µL)
Hu	34	19	1,75	1.657
Ke	34	16	1,75	2.571
chc	14	7	0,63	6.508
vrn	8	4	0,8	5.625
bre	2	1	0,76	5.460

Exp.: experimento. Hu: *Solanum tuberosum* cult. Huinkul, *tuberosum* cult. Kennebec, chc: *S. chacoense*, vrn: *S. vernei*, bre: *S. brevidens.* Tamanho (pg): conteúdo genômico haplóide em picogramas. 1C (5µl): número de genomas haplóides em 5µl, quantificados na reação de PCR em tempo real.

IV. 2.1– Clonagem de XPB para normalizar a determinação do número de cópias por qPCR.

Com o objetivo de ter um controle interno da quantidade de DNA utilizado em cada experimento, o número de cópias das amostras foi normalizado utilizando um gene de cópia única: *Xeroderma Pigmentosum* B (XPB). Para isso foram desenhados os iniciadores XPB-F e XPB-R a partir do alinhamento de duas seqüências de ESTs (do inglês *Expressed sequence tag*): uma de *Solanum tuberosum* (CN216239 DNA *excision repair protein* ERCC3-*like partial*) obtida do J. Craig Venter Institute, e outra de *Lycopersicon peruvianum* (SGN-U345011) obtida do *SOL Genomics Network*. Ambos os fragmentos alinham no segundo exon do gene *XPB* de *Arabidopsis thaliana*. Os iniciadores determinam um fragmento de 92 pb.

Com o objetivo de confirmar a identidade dos fragmentos amplificados com os iniciadores XPB-F e XPB-R, foi realizada uma PCR com as espécies: *Lycopersicon peruvianum*, *Solanum tuberosum* cult. Huinkul, *S. t.* cult. Kennebec, *S. chacoense*, *S. vernei* e *S. brevidens*. Os fragmentos amplificados foram clonados e dois clones de cada genótipo

foram seqüenciados. A identidade das seqüências obtidas foi confirmada por BLAST. Não foi possível obter a seqüência referente a XPB em *Nicotiana tabacum* utilizando estes iniciadores e como conseqüência, os dados do número de cópias de fumo não foram normalizados.

A confirmação do gene XPB como gene de cópia única nos genótipos analisados foi realizada através de uma hibridação molecular utilizando como sonda o fragmento de XPB amplificado (92pb) (figura 27). Todos os genótipos apresentaram uma banda exceto vrn (*Solanum vernei*) que apresenta duas bandas com menor intensidade, sugerindo alelos diferentes de um gene de cópia única ou sítios internos de restrição dada a menor intensidade. As condições de lavagens foram mudadas para obter sinal em *Lycopersicon peruvianum* e *Nicotiana tabacum.*

As mesmas diluições das amostras nos Exp. 1, 2, 3 e 4 foram quantificadas para XPB utilizando curvas padrão com o plasmídeo Ke1 contendo o fragmento de XPB de *Solanum tuberosum* cult. Kennebec. Os parâmetros de cada curva padrão de XPB de cada experimento estão expostos na tabela 10 e os valores absolutos quantificados na figura 28.

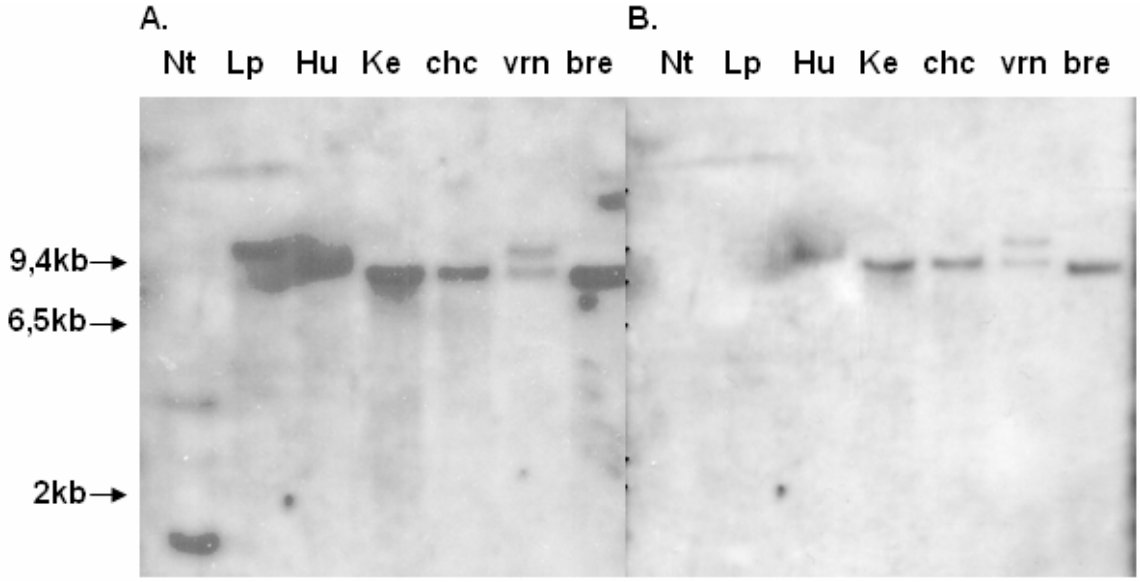

Figura 27. Análise por hibridação molecular do número de cópias de XPB (92pb) dentro dos genomas de espécies pertencentes à *Solanaceae*.

A: condições menos *estringentes* de lavagens que **B**. O DNA genômico das espécies foi digerido com a enzima *Eco*RI. Nt: *Nicotiana tabacum*, Lp: *Lycopersicon peruvianum*, Hu: *Solanum tuberosum* cult. Huinkul, *tuberosum* cult. Kennebec, chc: *S. chacoense*, vrn: *S. vernei*, bre: *S. brevidens*. Os números indicam tamanho molecular.

Tabela 10. Parâmetros da curva padrão dos diferentes experimentos com XPB.

Parâmetros	Exp.1	Exp. 2	Exp. 3	Exp. 4
Inclinação	-3,46	-3,32	-3,5	-3,18
Interseção	42,1	41,6	42,1	39,9
Eficiência	94,2%	99,7%	93,1%	106%
Amplificação	1,94	1,99	1,93	2,06
R^2	0,99	0,99	0,98	0,97

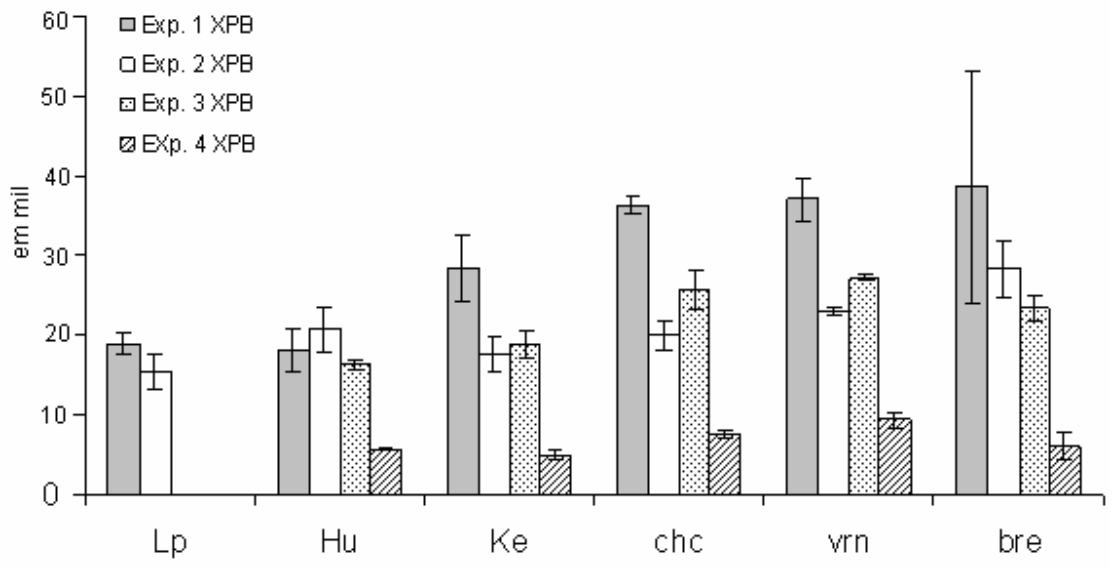

Figura 28. Número de copias de XPB.

Lp: *Lycopersicun peruvianum*. Hu: *Solanum tuberosum* cult. Huinkul, *tuberosum* cult. Kennebec, chc: *S. chacoense*, vrn: *S. vernei*, bre: *S. brevidens.*

O valor de número de cópias absoluto (RNAseH e LTRs) obtido em cada experimento foi dividido pelo correspondente valor de número de cópias de XPB para determinar o valor normalizado. A fim de comparar os valores obtidos na quantificação de ambos os experimentos com o fragmento da RNAseH (Exp.1 e 2) e também com os fragmentos das LTRs (Exp. 3 e 4), os dados foram relativizados. Foi utilizado o valor de *S. brevidens* como calibrador já que nos experimentos apresentou o menor valor. A média dos valores normalizados das amostras foi dividida pela média do calibrador para obter os valores relativos (Livak *et al*. 2001) (tabela 11).

Tabela 11. Valores normalizados e relativos do número de cópias (RNAseH) de elementos similares a Tnt1 e LTRs de Retrosol.

	N° cópias Exp. 1		N° cópia Exp. 2		N° LTR Exp. 3		N° LTR Exp. 4	
Genotipos	Norm.[a]	Rel.	Norm.[a]	Rel.[b]	Norm.[a]	Rel.[b]	Norm.[a]	Rel.[b]
Lp	38 ± 9,5	63 ± 0,3	22 ± 9	72 ± 0,4				
Hu	2 ± 0,4	3 ± 0,1	2 ± 0,3	7 ± 0,1	4 ± 0,3	10 ± 0,1	6 ± 0,3	10 ± 0,6
Ke	3 ± 1	5 ± 0,3	0,6 ± 0,5	2 ± 0,8	5 ± 0,5	13 ± 0,1	8 ± 0,8	13 ± 0,1
chc	3 ± 0,1	5 ± 0,1	2 ± 0,2	7 ± 0,1	4 ± 0,4	10 ± 0,1	6 ± 0,6	10 ± 0,1
vrn	4 ± 0,8	7 ± 0,2	1,5 ± 0,4	5 ± 0,3	2 ± 0,01	5 ± 0,01	2 ± 0,01	5 ± 0,01
bre	0,6 ± 0,7	1 ± 0,4	0,3 ± 0,1	1 ± 0,2	0,4± 0,02	1 ± 0,06	0,6 ± 0,2	1 ± 0,3

a. Os valores normalizados foram determinados dividindo a média do número de cópias pela média do valor de XPB (Norm.).
b. Os valores normalizados foram divididos pelo valor do calibrador (*Solanum brevidens*) (Rel.). Como é uma divisão por uma constante arbitraria o cv do resultado é o mesmo cv da normalização.

Os valores normalizados e relativos das batatas cultivadas, Hu e Ke, devem ser considerados como o dobro já que os genomas são tetraploides. Os valores relativos do número de cópias não são consistentes entre os experimentos embora sempre tomate apresente o número maior de cópias e *Solanum brevidens* o menor. Já os resultados relativos do número de LTRs são consistente nos dois experimentos. As batatas cultivadas apresentam o maior número de LTRs e *Solanum brevidens* o menor.

IV. 2.2 – Distribuição do número de cópias por hibridação molecular

A distribuição do número de cópias por hibridação molecular foi avaliada utilizando como sonda os iniciadores Avi-PTT (RNAseH) a fim de comparar o perfil de hibridação

com a determinação por qPCR. O resultado está representado na auto-radiografia da figura 30.

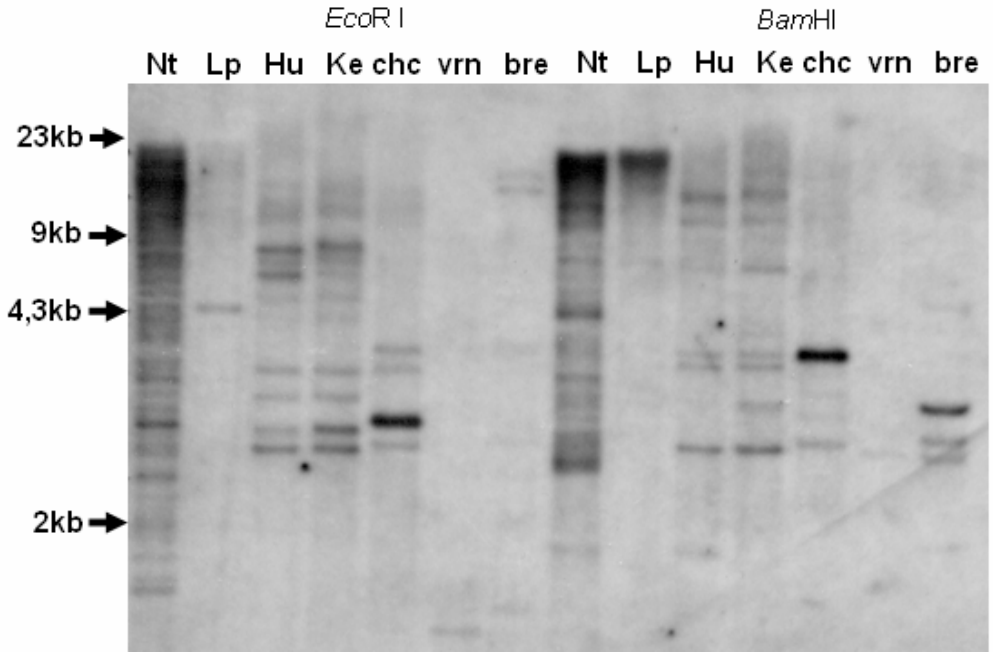

Figura 30. Análise por hibridação molecular do número de cópias da região RNAseH (119pb) dentro dos genomas de espécies pertencentes a *Solanaceae*.

O DNA genômico das espécies foi digerido com as enzimas *Eco*RI e *Bam*HI. Nt: *Nicotiana tabacum*, Lp: *Lycopersicon peruvianum*, Hu: *Solanum tuberosum* cult. Huinkul, *tuberosum* cult. Kennebec, chc: *S. chacoense*, vrn: *S. vernei*, bre: *S. brevidens*. Os números indicam tamanhos moleculares.

Fumo (Nt) apresenta maior sinal de hibridação. Tomate (Lp) apresenta maior sinal na digestão com *Bam*HII que com *Eco*R I possivelmente por conter menor quantidade de DNA genômico. As batatas cultivadas, Hu e Ke, apresentam varias bandas e *Solanum vernei* (vrn) apresenta poucas bandas e fracas por causa do pouco conteúdo de DNA no poço. O único resultado claramente coincidente com a determinação do fragmento Avi-PPT por qPCR, é que o maior número de cópias de fumo coincide com a maior sinal de hibridação.

IV. 3. Isolamento e caracterização do elemento completo similar a Tnt1 de *Solanum oplocence*, Retrosol1.

Com o intuito de isolar o elemento Retrosol, a espécie *Solanum oplocense* foi escolhida por apresentar o maior sinal de hibridação com a sonda contendo a região codificante, sugerindo ser uma das espécies com maior número de cópias (figura 19). Os iniciadores selecionados foram GIIF e Ol16 que permitiram a amplificação quase completa do elemento, desde a região U3 da LTR 5' até a região U5 da LTR 3' (figura 6). O iniciador Ol16 foi desenhado sobre o elemento Tnt1-94 e o iniciador GIIF foi desenhado tendo em vista as seqüências caracterizadas na análise filogenética pertencentes ao grupo II (ver item IV.1.2.1). Este grupo foi escolhido por apresentar a maior proporção de seqüências no nó $\alpha 1$ da análise do NETWORK.

A amplificação por PCR gerou bandas de diversos tamanhos. A banda que apresentou o tamanho esperado para um elemento inteiro (~6 kb) foi isolada, e um clone foi seqüenciado. A análise das seqüências das extremidades mostrou similaridade com seqüências de Retrosol determinadas no item IV-I. A extremidade 5' apresentou uma identidade de 95% com a porção da LTR (porção U3-R-U5) do clone scr_10 (Genbank: EF62079) e a extremidade 3' uma identidade de 98% com o clone total (RNAseH-Linker com PPT-U3-R-porção U5) inf_4 (Genbank: EF620640). Foi concluído que o clone é a versão completa do elemento Retrosol.

Através da restrição, subclonagem e posterior seqüenciamento dos fragmentos, foi possível obter a seqüência completa de Retrosol1.

Retrosol1 tem 4.945 pb e pertence ao grupo filogenético II mas sua RNAseH não faz parte do nó α1. A poliproteína de Retrosol1 apresentou uma identidade de 85% em 86% de cobertura com Tnt1-94 (GenBank: X13777), e uma identidade de 90% com 61% de cobertura com Retrolyc1-1 (Genbank: AF228701). A LTRs de Retrosol1 não apresentaram identidade significativa com Tnt1-94 e/ou Retrolyc1-1. A identidade das duas LTRs de Retrosol1 é de 94%.

Os retrotransposons com LTRs (longas terminações repetidas) apresentam LTRs em idêntica orientações, localizadas nas suas extremidades 5' e 3'. Uma das características intrínsecas das LTRs é que ambas são idênticas no momento da inserção do elemento. Com o tempo as LTR vão acumulando mutações e o número de substituições entre ambas pode servir para datar o momento da inserção (SanMiguel *et al.* 1998).

Com o objetivo de determinar o momento da inserção de Retrosol1 no genoma, as duas LTRs parciais (285 pb) foram alinhadas (figura 31) e a taxa de substituição calculada como distância. O valor da distância foi transformado em tempo de inserção (T) como descrito por San Miguel *et al*. (1996). O T se calcula através da formula: T= ***d***/2**S**, onde ***d*** é a taxa de substituição entre as duas LTRs e **S** é a média das substituições sinônimas, que neste caso foi utilizado o valor $1{,}95 \times 10^{-8}$ (Ammiraju J. *et al*. 2007). O valor ***d*** obtido entre as duas LTRs foi de 0,062 e por tanto a inserção de Retrosol1 aconteceu 1,6 Mya (do inglês, *millions of years*) no genoma de *Solanum oplocense*. Levando em consideração que a separação entre *Nicotiana* e *Lycopersicon* ocorreu a 14 Mya (Wikström *et al.* 2001) a inserção de Retrosol1 aconteceu apos a separação dos gêneros.

A estrutura geral da LTR é composta de três regiões: U3, R e U5. A região U3 compreende desde o sitio de inicio da LTR até o sitio de inicio da transcrição (SIT) que se

localiza 25-30 pb depois do TATA *box* (figura 31) (Perlman and Boeke, 2004). Ambas LTRs contem uma seqüência promotora para a RNA polimerase II, porem só atua como promotor forte na LTR5'. A seqüência da LTR3' poderia atuar como promotor débil para os genes adjacentes ao sitio de inserção (White *et al*. 1994). A região R está compreendida entre o SIT e o final da transcrição (Poly A) e se encontra em ambas as extremidades do transcrito. A região R permite a cadeia (-) curta de cDNA se deslocar da LTR5' para a LTR3' durante a transcrição reversa. A região U5 abrange desde o PolyA até o final da LTR.

Logo depois da LTR5' encontra-se o PBS (do inglês, *primer binding site)* que é o sitio de união ao iniciador que utilizam os retrotransposons para o inicio da síntese da primeira fita de cDNA (-). O PBS é uma seqüência complementar ao tRNA metionina. Retrosol1 apresenta a seqüência de PBS mais comum em plantas: TGGTATCAGAGC. A LTR3' está precedida por uma zona rica em purinas o PPT (do inglês, *poly purine track*) de 9-17 pb, que atua como sitio do inicio da síntese da segunda fita de cDNA (+).

As LTRs são importantes para a integração do elemento já que a enzima integrase reconhece seus extremos. Ambas LTR apresentam repetições curtas invertidas (IR, *inverted repeat*) nos extremos. O consenso descrito para os retrotransposon com LTR é de 5' – "TG..CA" -3', mais o número de nucleotídeos variam entre os retrotransposons. No caso de Retrosol1 não puderam determinar-se as IR porque a seqüência não esta completa.

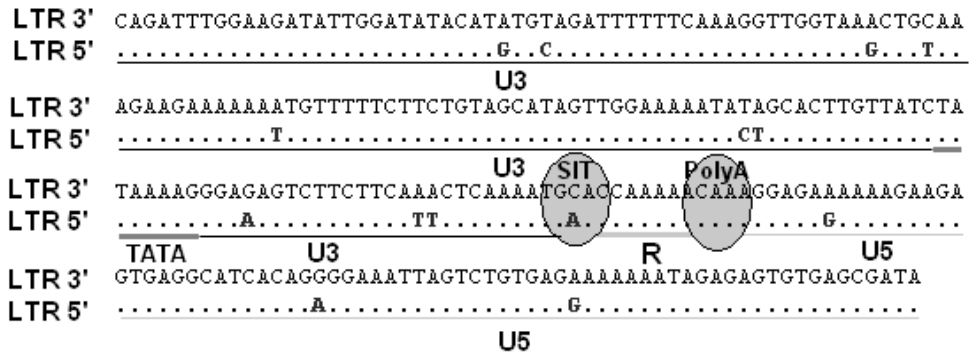

Figura 31. Alinhamento das LTRs de Retrosol1.

Pontos: nucleotídeos idênticos. Letras seguidas de pontos: nucleotídeos diferentes. As regiões da LTRs estão sublinhadas com o nome embaixo. U3: *unique 3'RNA*, R: *repeated RNA*, U5: *unique 5'RNA,* TATA: Tata box. SIT: sitio de inicio da transcrição na LTR5'. Poly A: sitio da poliadenilação na LTR3'.

A região codificante apresenta todas as atividades enzimáticas necessárias para o ciclo de vida autônomo do retrotransposon. A transcrição é policistronica com posterior síntese em forma de poli-proteína, que é posteriormente processada em diferentes peptídeos funcionais mediante atividade *protease* codificada pelo próprio elemento. Como já foi mencionado o domínio *gag* codifica a proteína estrutural do capsídeo viral (VLP do inglês, *virus like particle*). Esta região está pouco conservada em nível de seqüência nos distintos retrotransposons, embora diversos domínios ricos em Cisteína-Histidina são conservados em retrotransposons e retrovirus. Estes domínios denominados dedos de zinco (do inglês, *zinc finger*) estão relacionados com a união ao RNA ou DNA e a asitência no empacotamento das VLPs, e provavelmente em outros processos (Hansen and Heslop-Harrison, 2004). No caso de Retrosol1 o motivo GAG não está perfeitamente conservado (figura 32).

A região *pol* codifica diferentes atividades enzimáticas e está situado na continuação do gene *gag*, podendo ser codificado pelo mesmo quadro de leitura (ORF, *open reading frame*) do gene *gag* o por uma ORF própria. O gene *pol* codifica para as atividades enzimáticas como a protease (PR), integrase (INT), transcriptase reversa (RT) e RNAse H.

```
Copia      C H H C G R E G H I K K D C
Tnt1       C Y N C N Q P G H F K R D C
Retrosol1  C Y N C E Q P G Q F K R D C
```

Figura 32. Alinhamento do domínio Cisteína-Histidina referente ao domínio GAG em Tnt1, Cópia e Retrosol1.

O motivo inclui o dedo de zinco com os aminoacidos C-C-H-C. Tnt1-94: InterPro: IPR001878; Znf_CCHC. Cópia elemento de *Drosophila melanogaster*, GeneBank: X02599; InterPro: IPR001878; Znf_CCHC. Retrosol1: Anexo 2.

A protease codificada pelo próprio elemento provoca o auto-processamento da poliproteína dando lugar aos péptidos funcionais. As proteases dos retrotransposons e retrovirus são do tipo aspartato e contém um motivo altamente conservado que atua como centro ativo. No caso de Retrosol1 o A se localiza na posição 296.

A integrase é necessária para a inserção do cDNA do retrotransposon no genoma. Na região central da proteína encontra-se o domínio catalítico com a seqüência DD-35-E, perfeitamente conservado em INT de retrovirus e retrotransposon (Suoniemi *et al*. 1998) e também em Retrosol 1 (figura 33).

A RT transforma as moléculas de RNA em DNA de dupla cadeia. Para lograr este objetivo consta da atividade DNA-polimerase dependente de RNA [copia o RNA genómico a cDNA (-)], a atividade DNA-polimerasa dependente de DNA [sintetiza a segunda cadeia de cDNA(+)] e atividade Ribonucleasa H (elimina o molde de RNA do híbrido DNA-

RNA). O domínio RT é o mais conservado ao nível de aminoácidos dentro dos retrotransposons (Doolittle *et al.* 1989). Xiong e Eickbush (1990) alinharam 87 seqüências de RT pertencentes à retroelementos de plantas, animais, protozoarios e bacterias, e determinaram 7 domínios conservados. A figura 34 mostra o alinhamento desses 7 domínios na seqüência de Retrosol1, Tnt1-94 e o elemento Cópia.

```
Copia  DVCGPITPVTL .......DFQDFVAKSEAHFNLKVVYLYIDNGREYLSNEMRQFCVKKGISYHLTVPHTPQLNGVSERMIR
Tnt1   DASRKLWVYILKTKDQVFQVFQKFHALVERETGRKLKRLRSDNGGEYTSREFEEYCSSHGIRHEKTVPGTPQHNGVAERMNR
Retsl1 DASQKLRVYILKTKDQVFQVFQKFHALVERETGRKLKRLQTDNGGEYTSREFEEYYSNHGIKHEKIVPGIPQHNGVAERMNH
```

Figura 33. Alinhamento da região conservada da integrase (INT) de Tnt1, Cópia e Retrosol1. O motivo conservado DD-35-E está sinalizado em cinza. Tnt1-94: GenBank: X13777; InterPro: IPR001584; Integrase_cat-core. Cópia elemento de *Drosophila melanogaster*, GeneBank: X02599; IinterPro: IPR00158. Retsl1: Retrosol1, Anexo 2.

A atividade da RNAseH situa-se na região C-terminal da mesma proteína RT e é capaz de degradar o RNA de um hibrido de RNA-DNA. *In vitro* possui atividade endo e exonucleasa 3' e 5'. Durante o processo de transcrição a RNAseH elimina o molde de RNA (para permitir a síntese da segunda cadeia de DNA) e os iniciadores de RNA uma vez acabado o processo.

A poliproteína de Retrosol 1 é constituída por 1.328 aminoácidos e apresenta vários códons de parada ao longo da região codificante sugerindo que Retrosol1 seja uma cópia que não apresenta proficiente para a transposição. A seqüência completa de Retrosol1 e da poliproteina estão no anexo 2, junto com a localização das LTRs, PBS, PPT região codificadora e domínios catalíticos da poliproteína.

```
                10        20        30        40        50        60        70
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Retrosol1 YKLVKLPKGKRPLKCKWIFKLKKDGNGKLVRYKARLVVKGFEQKKGIDFDEIFSPVVKMTFIRTILSLAA
Tnt1-94   YKLVELPKGKRPLKCKWVFKLKKDGDCKLVRYKARLVVKGFEQKKGIDFDEIFSPVVKMTSIRTILSLAA
Copia     WTITKRPENKNIVDSRWVFSVKYNELGNPIRYKARLVARGFTQKYQIDYEETFAPVARISSFRFILSLVI
                  1                                   2
                80        90       100       110       120       130       140
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Retrosol1 SLDLEVEQLDVKTTFLHGDLEEKIYMEQLEGFEVFGKKHMVCKLNKSLYGLKQAPRQ*YKKFDSFMKNQT
Tnt1-94   SLDLEVEQLDVKTAFLHGDLEEEIYMEQPEGFEVAGKKHMVCKLNKSLYGLKQAPRQWYMKFDSFMKSQT
Copia     QYNLKVHQMDVKTAFLNGTLKEEIYMRLPQX-XISCNSDNVCKLNKAIYGLKQAARCWFEVFEQALKECE
                  3                                   4
               150       160       170       180       190       200
        ....|....|....|....|....|....|....|....|....|....|....|....|....|...
Retrosol1 YTKTYSDXHVYTSXRFSDNNFIILLLYVDDILIVGQDKGLIARLKGDLSKSFDIRDLGPAQQILGMKI
Tnt1-94   YLKTYSDXHVYTSXRFSENNFIILLLYVDDMLIVGKDKGLIAKLKGDLSKSFDMKDLGPAQQILGMKI
Copia     FVNSSVDRCIYILDKGNINENIYVLLYVDDVVIATGDMTKMNNFKRYLMEKFRMTDLNEIKHFIGIRI
                                     5            6            7
```

Figura 34. Alinhamento dos domínios conservados da transcriptase reversa (RT) de Tnt1-94, Retrosol1 e Copia.

As seqüências mostram os sete domínios (1-7) de Xiong e Eickbush (1990). Tnt1-94: GenBank: X13777. Copia: GeneBank: X02599. Retsl1: Retrosol1, Anexo 2.

V - DISCUSSÃO

Os elementos de transposição são os mais diversos e abundantes componentes do genoma dos organismos eucariontes. As primeiras análises extensivas foram realizadas nos genomas de importantes lavouras mundiais como milho, trigo e cevada, com o objetivo de compreender as modalidades de expansão dos TEs. Estes estudos revelaram que as inserções de retrotransposons com LTRs acontecem com maior freqüência em regiões entre genes do que dentro de exons ou introns. Estas inserções contribuíram significantemente para o aumento do tamanho dos genomas. Muito da expansão devido à atividade dos retrotransposons parece ter ocorrido recentemente quando considera-se o tempo evolutivo (isto é, nos últimos cinco milhões de anos). A redução de tamanho dos genomas devido à recombinação homóloga desigual entre as LTRs dos retrotransposons contrabalançou parcialmente a expansão em algumas espécies, conduzindo ao aparecimento de SOLO-LTRs (Morgante 2006).

O retrotransposon Tnt1 foi caracterizado inicialmente em *Nicotiana tabacum* e logo detectado por hibridação molecular em outros genomas de *Solanaceae* (Grandbastien *et al*. 1989). Dez anos mais tarde um novo elemento similar a Tnt1, Retrolyc1, foi caracterizado em *Lycopersicon peruvianum* (Costa *et al.* 1999). O presente estudo relata a caracterização de um membro novo da família, Retrosol, presente em diversos genótipos silvestres e cultivados de *Solanum.* Os conjuntos destes estudos sugerem uma associação precoce entre a família Tnt1 e a evolução de *Solanaceae*.

Além de sua importância funcional para garantir a transcrição do elemento, a LTR é uma das regiões de mais rápida diferenciação dentro de um retrotransposon. As LTRs contêm regiões funcionais importantes tais como os promotores de transcrição gênica,

enhancers, sítio de início da transcrição, sinais de poliadenilação do RNA mensageiro nascente, e elas ainda fazem parte do processo da integração da nova cópia. No retrotransposon BARE-1 de cevada, as LTRs apresentam variabilidade em todas suas regiões e por esta razão são consideradas heterogêneas (Vicient *et al*. 2005), embora na família Tnt1 a variabilidade está localizada específicamente na região U3 da LTR (Grandbastien *et al*. 2005). Os três grupos do retrotransposon Tnt1, assim como os grupos A e B de Retrolyc1, diferenciam-se principalmente pela região U3 (Araujo *et al*. 2001, Vernhettes *et al*. 1998). As seqüências de Retrosol aqui identificadas apresentaram o mesmo padrão, mostrando um acúmulo de INDELs (do inglês, *insertions or deletions*) e de SNPs (do inglês, *single nucleotide polymorphisms*) localizados específicamente na U3. Assim, a variabilidade da região U3 na família Tnt1 parece ser um fenômeno geral na evolução deste elemento dentro das espécies de *Solanaceae*.

A acentuada variação no tamanho do genoma de plantas têm sido explicada básicamente por dois processos: (a) eventos de poliploidização e/ou (b) amplificação dos TEs residentes. Em 1997, Bennetzen e Kellogg propuseram que na ausência de algum mecanismo para contrabalancear estes processos, os genomas das plantas tenderiam a *ficar obesos*. Dois processos têm sido propostos como responsáveis pela redução no conteúdo de DNA: recombinação homóloga desigual (do inglês, *unequal homologous recombination*) e recombinação ilegítima (do inglês, *illegitimate recombination*) (Devos *et al.* 2002, Ma *et al.* 2004, Bennetzen *et al*. 2005). Este último, recombinação ilegítima, foi sugerido como o responsável pela remoção da maior parte dos TEs em arroz e arabidopsis, embora este fenômeno requeira muitas gerações ao longo de milhões de anos para ser efetivo. O mecanismo molecular da recombinação ilegítima não é conhecido mas parece estar associado com o reparo de quebras do DNA (Gorbunova and Levy, 1999). No entanto,

ainda não está demonstrado que as diferenças entre os tamanhos de genomas sejam efetivamente o resultado de diferentes taxas entre o aumento do genoma e a eficiência na remoção por recombinação ilegítima (Bennetzen *et al*. 2005). No caso da análise das regiões U3 de Retrosol foi observado que 91% das deleções ou *gaps* estão associadas com repetições diretas, cicatrizes típicas da remoção de DNA por recombinação ilegítima. Estas observações poderiam sugerir que a variação na região U3 das cópias de Retrosol nos genomas de *Solanum* está associada com a remoção de DNA por recombinação ilegítima mais também, a geração de deleções poderia ser uma estratégia para gerar variabilidade nas regiões promotoras e *escapar* do silenciamento por parte do hospedeiro.

Três grupos de seqüências foram derivados ao ser aplicado o método filogenético nas seqüências amplificadas de Retrosol nas espécies de *Solanum*. Os grupos I e II são consistentes porque são representados por grupos de seqüências poucos variáveis na região U3. Ao contrário, o grupo III é composto por seqüências heterogêneas, provavelmente representando mais de um conjunto de elementos. Não é possível correlacionar uma versão particular de U3 às variáveis aqui tratadas como genótipo, espécie silvestre ou cultivada, ou ainda distribuição geográfica (tabelas 1, 2 e figura 15 e 16). No entanto, nota-se que o aparecimento do grupo III de seqüências ocorre apenas nas variedades coletadas acima e 1500 m. Seria necessário um conjunto amostral maior para que fosse testado este padrão. Independente deste comentário os resultados sugerem que a diferenciação de Retrosol em três grupos ocorreu antes da especiação de *Solanum.*

Com a intenção de avaliar as relações entre os elementos da família Tnt1 em *Solanum*, *Lycopersicon* e *Nicotiana* uma análise populacional foi aplicada. O objetivo foi testar a hipótese da existência de uma cópia mestre/fundador em cada subfamília. O resultado obtido pelo uso do NETWORK foi analisado assumindo os três modelos da

expansão propostos por Cordaux et *al*. (2004) para a subfamília de elementos *Alu* nos seres humanos e primatas: "modelo do gene mestre/fundador", modelo intermédiário e modelo transposon (figura 35). Segundo o "modelo do gene mestre/fundador", um único elemento pertencente ao grupo α gera todos os demais membros da subfamília, conduzindo a um relacionamento em forma de estrela (*star-like*) com todas as cópias inativas derivadas do elemento α. O modelo contrastante do transposon postula que todos os membros do elemento podem ser ativos e produzir cópias novas, tendo por resultado a ausência de uma estrutura radial a partir de um nó central α. O modelo intermédiário sugere que diversos membros sejam ativos e contribuam com a expansão do elemento. Os relacionamentos esperados neste último modelo são em parte na forma de estrela, mas igualmente com proporções diferentes de elementos potencialmente ativos que não são conectados diretamente com o elemento central descrito como α (figura 35).

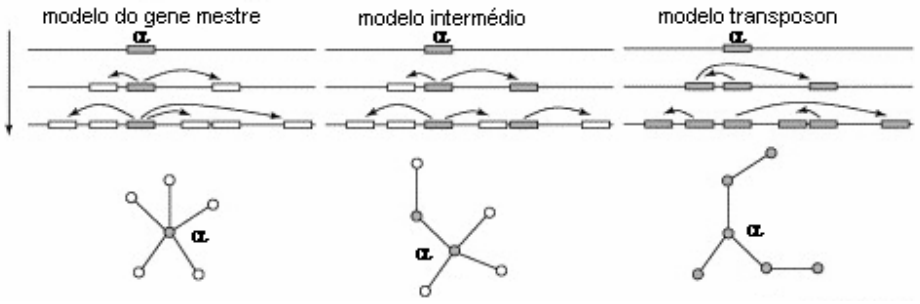

Figura 35. Esquema dos três modelos de expansão propostos para o elemento *Alu*.

Os elementos ativos estão representados com caixas e círculos em cinza e os inativos em branco. Seta vertical a esquerda da figura indica a passagem do tempo. O tamanho das setas e dos círculos é arbitrário. Adaptado de Cordaux *et al*. 2004.

A *rede* obtida com o alinhamento e análise da seqüência da RNAseH da família Tnt1 comporta modelos diferentes de expansão segundo o gênero. Esta observação da

existência de diferentes modelos de expansão sugere que cada elemento (Tnt1, Retrolyc1 e Retrosol) tem uma história particular em sua associação com o genoma vegetal. Em particular, temos que os tipos de seqüência de RNAseH, representados em cada nó na rede são distintos para cada gênero. Retrosol e Tnt1 mostram uma topologia em forma de estrela com um nó central α majoritário que corresponde ao tipo mais freqüente de seqüência, embora Tnt1 apresente mais de um nó α. Outros nós conectados não diretamente com o nó central α são igualmente encontrados, indicando que a seqüência imediatamente abaixo é igualmente capaz de se amplificar sugirindo mais de uma forma ativa do elemento no genoma. O nó original α de Retrosol encontra-se representado em 16 dos 25 genótipos analisados, levando em consideração que para a maioria dos outros 9 genótipos somente algumas seqüências foram clonadas. Em resumo, o modelo da expansão de Tnt1 nas espécies de *Nicotiana* e Retrosol em espécies de *Solanum* comporta o modelo intermédio. Ao contrário, o elemento Retrolyc1 apresenta um modelo com ausência de uma estrutura de irradiação com um nó central e com mais de dois nós α, sugerindo que muitos membros da família sejam capazes de produzirem novas cópias. Entretanto, estes resultados não podem ser considerados conclusivos a respeito da evolução de Retrolyc1 já que as seqüências usadas neste estudo foram obtidas somente para 4 das 9 espécies de *Lycopersicon*.

O nó α1 em Retrosol é constituído por 92% das seqüências do grupo II, representando 50% da amostra total do grupo II, e como já foi mencionado, o grupo II apresenta as regiões U3 menos variáveis (figura 14). Além disso, quando se compararam as regiões U3 das seqüências do grupo II pertencentes ao nó α1, foram encontradas seqüências com regiões U3 iguais em espécies diferentes. Por exemplo, o clone inc15 de *S. incamayoense* possui a região U3 idéntica ao o clone spg31 de *S. spegazzinii,* ambos os

clones com a porção da RNAseH do nó $\alpha 1$. Estas observaçõs sugeririam que as seqüências com RNAseH e regiões U3 idênticas são ativas desde antes da divergência das espécies que as contêm e que esta atividade foi recente o suficiente de modo que as cópias do elemento não tenham acumulado mutações na região promotora U3. Como hipótese adicional sugere-se que estas cópias tenham preservando a capacidade de transposição representando cópias ativas. De modo mais amplo e em função do relatado acima propomos que os elementos contendo a U3 do grupo II têm um papel preponderante na expansão de Retrosol como a linhagem transposicionalmente ativa.

O domínio codificante da RNAseH é conservado nas cópias de Retrosol nas diferentes espécies de hospedeiros do mesmo gênero, enquanto a região U3 diversificou rapidamente de acordo com um conceito de evolução modular (Lerat *et al*. 1999). A conservação do módulo de RNAseH em nível genérico está possivelmente relacionado com sua função biológica já que a enzima tem que interagir com outras moléculas. O domínio RNAseH está envolvido na degradação do molde original da molécula de RNA, na geração de uma trilha de poli-purinas, e na remoção final dos iniciadores de RNA das fitas de DNA recém sintetizadas (Malik and Eickbush, 2001). Estas funções são necessárias para a atividade transposicional do elemento. Por outro lado, o módulo U3 evoluiu rápidamente de acordo com sua função de região promotora. As seqüências novas dos promotores podem permitir a expressão do elemento em circunstâncias ambientais diversas, aumentando o potencial de sobrevivência do elemento no hospedeiro e/ou à probabilidade de superar os esforços de silenciamento transcricional. Por sua vez, o hospedeiro poderia ganhar uma vantagem em gerar diversidade no seu genoma pela inserção de cópias novas do retrotransposon, aumentando sua própria adaptação ao ambiente. Segundo Kidwell e Lish

(2000) a evolução e adaptação conjunta do retrotransposon e seu genoma hospedeiro, deve cumprir um papel particularmente importante na sobrevivência a longo prazo dos retrotransposons.

A determinação do número de cópias por PCR quantitativo usando o domínio RNAseH de retrotransposons similares a Tnt1 caracterizou que as cópias variam em ordens de grandeza entre as espécies nos diferentes gêneros. *Nicotiana tabacum* possui o maior número de cópias, *Lycopersicon peruvianum* apresentou um número intermédiario de cópias e as espécies de *Solanum* um baixo número de cópias. A determinação do valor absoluto do número de cópias deve ser considerada relativa em particular por duas causas, a primeira é que a amplificação depende de diversos parâmetros variáveis para sua determinação, como por exemplo a eficiência de amplificação que pode ser diferente entre experimentos e também entre genomas. E segundo, o número de cópias determinadas corresponde a uma amostra da população de retrotransposons. As poucas cópias de Retrosol sugerem uma baixa atividade mesmo dentro dos genótipos cultivados, em quanto Tnt1 possuindo um alto número de cópias aparenta ter uma alta expansão em fumo. Novamente Retrolyc1 aparece com sendo um elemento de atividade intermediaria em tomate.

Considerando a média do número de cópias, aproximadamente de 1 a 4% do genoma de *Nicotiana tabacum* (4.459 Mb) está ocupado por cópias de Tnt1, no caso do *Lycopersicon peruvianum* (1.127 Mb) 0,03% do genoma está ocupado por Retrolyc1 e nas espécies de *Solanum* perto de 0,003% dos genomas são cópias de Retrosol. Vale ressaltar que *S. brevidens* (740 Mb) possui a menor quantidade de cópias de Retrosol (0,0008% do genoma). Uma pergunta que surge da observação acima é se haveria um limite de tamanho de genoma para que ocorra a amplificação de retrotransposons ou, contariamente, a partir

de um determinado tamanho de genoma se faz necessária a amplificação destes elementos talvez para guardar a estrutura cromossômica organizada.

Os resultados da quantificação das LTRs de Retrosol revelaram que a relação LTR:RNAseH variou entre 1:1 a 2:1 entre as espécies de *Solanum,* considerando que cada elemento teria duas LTRs por região codificante (RNAseH). Estes resultados sugerem que a maioria das poucas cópias de Retrosol estariam completas. Soleimani e colaboradores (2005) utilizando a mesma técnica de PCR quantitativo determinaram ao contrario que o elemento BARE-1 em vários cultivares de cevada apresenta uma maior proporção de SOLO-LTRs que elementos inteiros e que em torno de 9% do genoma vegetal estaria formado por o retrotransposons BARE-1.

Quais são as causas da diferente abundância de Tnt1, Retrolyc1 e Retrosol em *Solanaceae*? Wright e Finnegan (2001) propuseram que a hipótese de DNA parasita seria o modelo dominante para explicar o controle do número de cópias. Esta hipotese postula que os elementos persistem nas populações de hospedeiros em função da sua habilidade de autoreplicar-se embora isto implicaria em potencial aumento na geração de inserções com efeitos deletéreos para o hospedeiro. Se considerarmos os TEs como parasitas, a taxa de multiplicação é um processo fundamental na sua sobrevivência e evolução. A reprodução sexual dos hospedeiros ofereceria aos TEs a capacidade de colonizar todos os organismos da população do hospedeiro. Em contraste, se o hospedeiro tem reprodução assexual, e na ausência de transmisão horizontal, a invasão de novos genomas pelos TEs na população do hospedeiro seria nula. A multiplicação do número de cópias dentro do genoma é a mesma em organismos com reprodução sexual ou assexual. Mas a propagação de TEs entre genomas através do cruzamentos só aconteceria em organismos com reprodução sexual. Vale ressaltar, que este modelo não leva em consideração a possibilidade que novas

inserções gerem benefícios Os autores citados concordam que ainda faltam estudos para entender os efeitos dos diferentes modos de reprodução dos hospedeiros e sua relação com a evolução dos TEs. As batatas, ao contario de fumo e tomate, possuem reprodução assexual por meio de tubérculos. Segundo a hipótese onde os TEs figuram como parasitas do genoma, este modo de reprodução poderia ter influenciado a taxa de dispersão do retrotransposon Retrosol dentro da população e gerar um número baixo de cópias. Mas duas questões merecem ser consideradas, primeiro que *Solanum brevidens* é a espécie com menor número de cópias e não produz tubérculos, e segundo que uma única família de retrotransposon foi considerada, a Tnt1.

A *Solanaceae* é uma das maiores famílias das plantas com flores, abrangendo 96 gêneros. Tem uma distribuição mundial, porém o centro de dispersão está situado na região da linha do equador e como conseqüência as espécies muito provavelmente não sofreram distúrbios durante as glaciações permitindo o acúmulo de variações genéticas adaptativas. Wikström e colaboradores sugerem que a *Solanaceae* se originou faz 41Myr no período Terciário da época Eoceno.

Com o objetivo de analisar a evolução da família Tnt1, as espécies de três gêneros foram consideradas: *Solanum*, *Lycopersicon* e *Nicotiana*. O *Solanum* é um dos maiores gêneros em número de espécies das Angiospermas e inclui todas as espécies silvestres e cultivadas que produzem tubérculos que se originaram numa extensão de área entre Bolívia e Peru. *Lycopersicon* inclui o tomate e seus parentes silvestres, cuja origem encontra-se confinada principalmente no Chile, Peru e Equador. O gênero *Nicotiana* abrange o fumo cultivado e as espécies silvestres encontrados principalmente na América do Sul e na Austrália. Estes três gêneros originaram-se na América do Sul, mas a distribuição atual da espécie reflete diversos fenômenos de dispersão (D'arcy 1991).

Os elementos similares a Tnt1 poderiam ser um determinante no processo de especiação em *Solanaceae*? Como proposto por Grandbastien (2005) a composição das populações atuais do elemento resultam de uma adaptação da população ancestral de Tnt1 aos diferentes estímulos aos que foram submetidos pelos hospedeiros. Esta adaptação esta relacionada com padrãos de expressão apropiados, mecanismos eficientes de replicação e integração, e inserção em locais genômicos que não resultem em efeitos deletérios. As cópias ancestrais de Tnt1 que exibiram essas respostas de adptação foram amplificadas seletivamente durante ou após a radiação da *Solanaceae*. Por outro lado, as cópias inativas não transcritas teriam sido perdidas rápidamente. As cópias de Retrosol, não apresentam uma correlação entre a espécie do hospedeiro, a região U3 e a versão da RNAseH, sugerindo que estes elementos não foram determinantes para a diversificação das espécies de *Solanum*. Resultados similares foram relatados para Tnt1 (Vernhettes *et al*. 1998) e Retrolyc1 (Araujo *et al*. 2001).

O uso de um modelo de população baseado na aproximação em rede (NETWORK) para avaliar a natureza da associação de seqüências similares a Tnt1 em *Solanaceae* suporta a existência de um elemento ancestral. Este resultado refuta a hipótese anterior da ocorrência de transferência horizontal recente do elemento na família. A natureza modular de tais retrotransposons é demonstrada claramente ao encontrar que a proteína funcional, representada aqui por RNAseH, está caracterizada como seqüências de diversificação lenta, enquanto a região regulatória U3 está sob uma força evolutiva menos restritiva. As populações atuais de Tnt1, Retrolyc1 e Retrosol são os resultados mais prováveis de sua transmissão vertical em *Solanaceae.* A base molecular que dirige a diferenciação da região U3 como uma seqüência em franca diferenciação pode ser o resultado de uma estratégia

para fugir do controle negativo do hospedeiro ou uma tentativa de se adequar rapidamente ao genoma, hipóteses estas que permanecem a ser demonstradas.

Como conclusão, os dados apresentados sugerem uma correlação entre o hospedeiro (gênero botânico) e a subfamílias Tnt1, Retrolyc1 e Retrosol. Esta correlação estaria dada principalmente na diversidade das régiões U3, porção conservada da RNAseH nos gêneros, e no número de cópias nos hospedeiros. Mas ainda falta uma maior compreensão da evolução conjunta entre Tnt1/*Solanaceae* (retrotransposon/hospedeiro) para entender como a família Tnt1 colonizou e se manteve nos genomas e as possíveis vantagens adaptativas geradas em *Solanaceae*.

VI - CONCLUSÕES

1. O gênero *Solanum* apresenta retrotransposons da família Tnt1 constituído por três grupos de seqüênciais que se diferenciam na região U3 da LTR.
2. Analisando a região codificante da RNAseH modelos de expansão foram propostos para cada gênero a saber, *Solanum*, *Lycopersicon* e *Nicotiana*.
3. A família de retrotransposons Tnt1 amplificou-se diferencialmente entre as espécies apresentando menor número cópias em *Solanum*.
4. Uma cópia completa denominada Retrosol1 foi isolada de *Solanum oplocense* e confirma similaridade estrutural e em nível de seqüência com os membros da família Tnt1. A família de retrotransposons Tnt1 se diferencia em nível genérico através da variabilidade de sua região U3 e da região RNAseH codificante.

VII – RESUMO

Tnt1 foi o primeiro retrotransposon ativo identificado em fumo após sua inserção no gene estrutural da *nitrato redutase*. Dez anos mais tarde, um novo membro foi caracterizado em *Lycopersicon*: Retrolyc1. Neste estudo analisamos seqüências similares a Tnt1 em 20 espécies silvestres de *Solanum* e cinco cultivares de *Solanum tuberosum*. As seqüências completas foram amplificadas a partir de DNA genômico utilizando uma estratégia baseada em PCR. Os fragmentos purificados foram clonados e seqüenciados, e a análise filogenética revelou três grupos que diferem na sua região U3 [Genbank: EF620567-EF620763]. Estes fragmentos clonados foram chamados de Retrosol.

Utilizando uma abordagem de "network" com um total de 453 seqüências não redundantes e isoladas de espécies de: *Solanum* (197), *Nicotiana* (140) e (116) *Lycopersicon*, foi demonstrado que a família Tnt1 pode ser tratada como uma população para tentar resolver ramificações filogenéticas multiforcadas. A rede resultante da RNAseH revelou que as seqüências agrupam de acordo com o gênero de *Solanaceae*, sustentando uma forte associação com o genoma hospedeiro, e que a variação das seqüências na região U3 associada, caracteriza o padrão evolutivo modular dentro da família Tnt1. Dentro de cada gênero, e independentemente das espécies, quase 20% das seqüências Tnt1 analisadas são idênticas, o que indica que foram provenientes de uma cópia ativa. As relações de "network" permitiram a identificação da "seqüência mestre ou fundadora" e forneceu provas de que dentro de cada gênero essas seqüências mestre são associadas com regiões U3 distintas.

O número de cópias de seqüências similares a Tnt1 foi determinado utilizando PCR quantitativo em tempo-real. Comparando-se o número de cópias, foi revelado que em fumo as cópias de Tnt1 são abundantes e representam cerca de 2% do genoma, enquanto em tomate as cópias de Retrolyc1 são menos abundantes. As espécies de *Solanum* mostraram um número baixo de cópias de Retrosol e uma relação LTR: domínio interno de aproximadamente 2:1, sugerindo que a maioria das cópias são completas.

Utilizando uma estratégia baseada em PCR, uma seqüência completa de Retrosol de quase 5 kb foi clonada. Esta cópia possui todas as características típicas de um retrotransposon tipo *Ty1-copia* e apresenta baixa identidade de nucleotídeos na região U3 quando comparada, aos outros membros da família.

Os resultados aqui apresentados corroboram a hipótese de que a família Tnt1 estava presente logo no início da evolução de *Solanaceae*. A evidência sugere também que a região da RNAseH ficou fixada ao nível de gênero no hospedeiro, e que dentro de cada gênero a propagação foi assegurado pela diversificação da região U3.

VIII – *Abstract*

Tnt1 was the first active plant retrotransposon identified in tobacco after nitrate reductase gene disruption. Ten years later a new member was characterized in *Lycopersicon:* Retrolyc1. In this study, we performed an analysis of Tnt1-like sequences of 20 wild species of *Solanum* and five cultivars of *Solanum tuberosum.* Sequences were amplified from total genomic DNA using a PCR-based approach. Purified fragments were cloned and sequenced, and clustering analysis revealed three groups that differ in their U3 region [GenBank: EF620567-EF620763]. These cloned fragments were named Retrosol.

Using a network approach with a total of 453 non-redundant sequences isolated from *Solanum* (197), *Nicotiana* (140) and *Lycopersicon* (116) species, it was demonstrated that the Tnt1 family can be treated as a population to resolve previous phylogenetic multifurcations. The resulting RNAseH network revealed that sequences group according to the *Solanaceae* genus, supporting a strong association with the host genome, whereas tracing the U3 region sequence association characterizes the modular evolutionary pattern within the Tnt1 family. Within each genus, and irrespective of species, nearly 20% of Tnt1 sequences analyzed are identical, indicative of being part of an active copy. The network approach enabled the identification of putative "master" sequences and provided evidence that within a genus these master sequences are associated with distinct U3 regions.

Copy number of Tnt1-like sequences was determined using quantitative real-time PCR. Comparing copy number among the species revealed that Tnt1 elements are abundant in tobacco, representing around 2% of the genome while Retrolyc1 elements in tomato are less abundant. *Solanum* species show low copy number of Retrosol elements and an LTR:internal domain ratio of about 2:1, indicating that most copies are complete.

A complete sequence of Retrosol, which is nearly 5 kb long was cloned using a PCR approach. It has all the features typical of a *Ty1-copia* like retrotransposon and show low nucleotide identity in the U3 region when compared with the other members of the family.

The results presented here support the hypothesis that the Tnt1 family was present early in the evolution of *Solanaceae*. The evidence also suggests that the RNAseH region of Tnt1 became fixed at the host genus level whereas, within each genus, propagation was ensured by the diversification of the U3 region.

IX - REFERÊNCIAS BIBLIOGRÁFICAS

AMMIRAJU, J. S. S.; ZUCCOLO, A.; YU, Y.; SONG, X.; PIEGU, B.; CHEVALIER, F.; WALING, J. G.; MA, J.; TALAG, J.; BRAR, D. S.; SAN MIGUEL, P. J.; JIANG, N.; JACKSON, S. A.; PANAUD, O.; WING, R. A. **Evolutionary dynamics of an ancient retrotransposon family provides insigths into evolution of the genome size in the genus *Oryza***. The Plant Journal. doi: 10.1111/j.1365-313x.2007.03242.x. 2007.

ARAUJO, P. G.; CASACUBERTA, J. M.; COSTA, A. P. P.; HASHIMOTO, R. Y.; GRANDBASTIEN, M. A.; VAN SLUYS, M. A. **Retrolyc1 subfamilies defined by different U3 LTR regulatory regions in the *Lycopersicon* genus.** Mol Genet Genomics**.** 266:35-41. 2001.

ALTSCHUL, S. F.; GISH, W.; MILLER, W.; MYERS, E. W.; LIPMAN, D. J. **Basic local alignment search tool.** J Mol Biol. 215:403-410. 1990.

BANDELT, H. J. **Median-Joining networks for inferring intraespecific phylogenies**. Mol Biol Evol. 16:37-48. 1999.

BABUSHOK, D. V.; OSTERTAG, E.M.; KAZAZIAN, H. H. **Current topics in genome evolution: Molecular mechanism of new gene formation.** Cell.MolLife Sci.64:542-554.2007.

BENNETZEN, J. L.; KELLOG, E. A. **Do plants have a one way ticket to genomic obesity?** The Plant Cell. 9:1509-1514. 1997.

BENNETZEN, J. L.; MA, J.; DEVOS, K. M. **Mechanisms of Recent Genome Size Variation in Flowering Plants.** Annals of Botany. 95:127-132. 2005.

BOEKE, J. D.; CORCES, V. G. **Transcription and reverse transcription of retrotransposon**. Annu Rev Microbiol. 43: 403-34. 1989.

BOWEN, N. J.; JORDAN, I. K. **Transposable elements and the evolution of eukaryotic complexity**. Curr. Issues Mol. Biol. 4: 65–76. 2002.

CALLINAN, P. A.;WANG, J.; HERKE, S. W.; GARBER, R. K.; LIANG, P., BATZER, M. A. ***Alu* Retrotransposition-mediated Deletion.** J. Mol. Biol. 348:791-800. 2005.

CASACUBERTA, E.; PARDUE, M. L. **HeT-A and TART, two Drosophila retrotransposon with a bona fide role in chromosome structure for more than 60 millions years**. Cytogenet. Genome Res. 110 (1-4): 152-9. 2005.

CASACUBERTA, J. M.; VERNHETTES, S.; GRANDBASTIEN, M. A. **Sequence variability within the tobacco retrotranspsoson Tnt1 population.** EMBO J. 14:2670-2678. 1995.

CASACUBERTA, J. M.; SANTIAGO, N. **Plant LTR-retrotransposons and MITEs: control of transposition and impact on the evolution of plant genes and genomes**. Gene. 311: 1-11 2003.

CASACUBERTA, J. M.; GRANDBASTIEN, M. A. **Characterisation of LTR sequences involved in the protoplast specific expression of the tobacco Tnt1 retrotransposon**. Nucleic Acids Res. 11: 21(9):2087-93. 1993.

CHILD, A. **A synopsis of *Solanum* subgenus *Potatoe* (G. Don) D'Arcy [*Tuberarium* (Dun.) Bitter** (s.l.)] *Feddes Repert*. 101:209-235. 1990.

CORDAUX, R.; UDIT, S.; BATZER, M. A.; FESCHOTTE, C. **Birth of a chimeric primate gene by capture of the transposase gene from a mobile element**. Proc Natl Acad Sci U S A. 23: 103(21):8101-6. 2006.

CORDAUX, R.; HEDGES, D. J.; BATZER, M. A. **Retrotransposition of Alu elements: how many sources?** Trends in Genetics**.** 20(10):464-467. 2004.

COSTA, A. P. P.; SCORTECCI, K.; HASHIMOTO, R. Y.; ARAUJO, P. G.; GRANDBASTIEN, M. A.; VAN SLUYS, M. A. **Retrolyc-1, a member of the Tnt1 retrotransposon super-family in the Lycopersicon peruvianum genome.** Genetica. 197:65-72. 1999.

D'ARCY, W. **The Solanaceae since 1976, with a Review of its Biogreography**. Solanaceae III:Taxonomy, Chemestry, Evolution. Edited by: Hawkes, Lester, Nee, Estrada. Royal Botanic Gardens Kew and Linnean Society of London; 1991.

DEININGER, P. L.; BATZER, M. A. ***Alu* repeats and human disease**. Mol. Genet. Metab. 67:183-193. 1999.

DEININGER, P. L.; MORAN, J. V.; BATZER, M. A.; KAZAZIAN, H. H. **Mobile elements and mammalian genome evolution.** Current Opinion in Genetics & Development. 13(6):651-658.2003.

DEVOS, K. M.; BROWN, J. K. M.; BENNETZEN, J. L. **Genome Size Reduction through Illegitimate Recombination Counteracts Genome Expansion in *Arabidopsis***. Genome Research. 12:1075-1079. 2002.

DOOLITTLE, W. F.; SAPIENZA, C. Selfish **genes, the phenotype paradigm and genome evolution.** Nature 284: 601–3. 1980.

DOOLITTLE, R. F.; FENG, D. F.; JOHNSON, M. S.; MCCLURE, M. A. **Origins and evolutionary relationships of retroviruses. Q. Rev. Biol. 64**(1): 1-30. 1989.

EWING, B.; HILLIER, L.; WEDL, M.; GREEN, P: **Basecalling of automated sequencer traces using phred. I. Accuracy assessment**. Genome Research. 8:175-185. 1998.

FLAVEL, A. J.; SMITH, D. B.; KUMAR, A**: Extreme heterogeneity of Ty1-copia group retrotransposons in plants.** Mol Gen Genet. 231:233-242. 1992.

FEDOROFF, N. **Transposons and genome evolution in plants**.Proc Natl Acad Sci U S A. 20;97(13):7002-7. Review. 2000.

FESCHOTTE, C.; JIANG, N.; WESSLER, R. **Plant transposable elements: where genetics meets genomics**. Nature Reviews. 3:329-341. 2002.

FULTON, T. M.; CHUNWONGSE, J.; TANKSLEY, S. D. **Microprep protocol for extraction of DNA from tomato and other herbaceous plants.** Plant Molecular Biology Reporter. 13:207-209. 1995.

GORBUNA, V.; LEVY, A. A. **How plants make ende meet: DNA double-strand break repair**. Trends in Plant Science. 4:263-269. 1997.

GORDON, D. P.; ABAJIAN, C.; GREEN, P**. Consed: a graphical tool for sequence finishing**. Genome Research. 8:195-202. 1998.

GORDON, D.; DESMARAIS, C.; GREEN, P**. Automated finishing with Autofinish.** Genome Res . 11:614-625. 2001.

GRANDBASTIEN, M. A.; SPIELMAN, A.; CABOCHE, M. **Tnt1, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics.** Nature. 26;337(6205)**:**376-80. 1989.

GRANDBASTIEN, M. A.; AUDEON, C.; BONNIVARD, E.; CASACUBERTA, J. M.; CHALHOUB, B.; COSTA, A. P. P.; LE, Q. H.; MELAYAH, D.; PETIT, M.; PONCET, C.; TAM, S. M.; VAN SLUYS, M. A.; MHIRI, C. **Stress activation and**

genomic impact of Tnt1 retrotransposons in Solanaceae. Cytogenet Genome Res. 110(1–4):229-41. 2005.

HALL, T. A. **BioEdit: a user-friendly biological sequence alignment editor and analysis program for Window 95/98/NT.** Nucl. Acids. Symp. Ser. 41:95-98. 1999.

HANAHAN, D. **Studies on transformation of E. coli with plasmids.** J. Mol. Biol. 156:557-580. 1983.

HANSON, C. N.; HESLOP-HARRISON, J. S. **Sequences and Phylogenies of Plant Pararetroviruses, Virus and Transposable Elements.** Advanced in Botanical Research. 41:165-193. 2004.

HAVECKER, E. R.; GAO, X.; VOYTAS, D. F. **The diversity of LTR retrotransposons**. Genome Biol. 5(6):225. Epub 2004 May 18. 2004.

HAWKES, J. G. **The potato: evolution, biodiversity and genetic resources**. Belhaven Press, London. 1990.

HAWKES, J. D. **The economic importance of the family *Solanaceae***. In Solanaceae IV. Edited by: Nee M, Symon DE, Lester RN, Jessop JP. Royal Botanic Gardens, Kew; 1999:1-8.

JIANG, N.; BAO, Z.; ZHANG, X.; MCCOUCH, S. R.; EDDY, S. R.; WESSLER, S. R. **An active DNA transposon in rice.** Nature 421: 163–7. 2003.

JIANG, N.; WESSLER, S. R. **Insertion preference of maize and rice miniature inverted repeat transposable elements as revealed by the analysis of nested elements**. Plant Cell. 13: 2553-2564. 2001.

JORDAN, M. A.; SAEDLER, H.; STARLINGER, P. **Stron-polar mutations in the trnasferase gene of the galactose operon E. Coli.** Mol.Gen.Genet.100:296-306. 1967.

JUKES, T. H.; CANTOR, C. R. Evolution of protein molecules. In Munro, H.N. (ed), **Mammalian Protein Metabolism.** Academic Press, New York, N.Y. pp. 21-132. 1969.

KALENDAR, R.; TANSKANEN, J.; IMMONEN, S.; NEVO, E.; SCHULMAN, A. H. **Genome evolution of wild barley (Hordeum spontaneum) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence**. Proc Natl Acad Sci U S A . 97:6603-6607. 2000.

KALMYKOVA, A. I.; GVOZDEV, V. A. **Selective expansion of the newly evolved genomic variants of retrotransposon 1731 in the Drosophila genomes**. Mol Biol Evo. 21(12):2281-9. 2004.

KASHKUSH, K.; FELDMAN, M.; LEVY, A. A. **Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat**. Nat. Genet. 33(1): 102-6. 2003.

KAZAZIAN, H. H. JR. **Mobile elements: drivers of genome evolution**. Science. 12;303(5664):1626-32. 2004.

KIDWELL, M. G.; LISCH, D. R. **Transposable elements as sources of variation in animals and plants**. Proc Natl Acad Sci U S A. 22;94(15):7704-11. 1997.

KIDWELL, M. G.; LISCH, D. R. **Transposable elements and host genome evolution**. Trends Ecol. Evol. 15:95-99. 2000.

KIMURA, M. **A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences.** J. Mol. Evol. 16: 111-120. 1980.

KUBISTA, M.; ANDRADE, J. M.; BENGTSSON, M.; FOROOTAN, A.; JONÁK, J.; LIND, K.; SINDELKA, R.; SJÖBACK, R.; SJÖGREEN, B.; STRÖMBOM, L.; STAHLBERG, A.; ZORIC, N. **The real-time polymerase chain reaction.** Mol Aspects Med. Apr-Jun;27(2-3):95-125. 2006.

KUMAR, A; BENNETZEN, J. L: **Plant Retrotransposons.** Annu Rev Genet. 33**:**479-532. 1999.

KUMAR, S.; TAMURA, K.; NEI, M. **MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment.** Briefings in Bioinformatics. 5:150-163. 2004.

LANDER, E. S, *et al*. **Initial sequencing and analysis of the human genome**. Nature. 409(6822):860-921. 2001.

LEETON, P. R. J.; SMYTH, D. R. **An abundant LINE-like element. amplified in the genome of Lilium speciosum.** Mol. Gen. Genet. 237(1-2):97-104. 1993.

LERAT, E.; BRUNET, F.; BAZIN, C.; CAPY, P**. Is the evolution of transposable elements modular?** Genetica. 107: 15-25. 1999.

LEV-MAOR, G.; SOREK, R.; SHOMRON, N.; AST, G. **The birth of an alternatively spliced exon: 3′ splice-site selection in *Alu* exons.** Science 300: 1288–91. 2003.

LIVAK, K. J.; SCHMITTGEN, T. D. **Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the $2^{-\Delta\Delta Ct}$Method**. Methods. 25:402-408. 2001.

LUCKWILL, L. C. **The genus *Lycopersicon*: an historical, biological, and taxonomical survey of the wild and cultivated tomatoes**. Aberdeen Univ. Stud. 120:1-44. 1943.

MA, J.; DEVOS, K. M.; BENNETZEN, J. L. **Analyses of LTR-Retrotransposon Structures Reveal Recent and Rapid Genomic DNA Loss in Rice.** Genome Research. 14:860-869. 2004.

MALIK, H. S.; EICKBUSH, T. H. **Phylogenetic Analysis of Ribonuclease H Domains Suggest a Late, Chimeric Origin of LTR Retrotransposable Elements and Retroviruses**. Genome Research**.** 11:1187-1197. 2001.

MC LEAN, R. C.; IVIMEY-COOK, W. R. **Textbook of Theorical Botany**. Vol II Longman. 1956.

MORGANTE, M. **Plant genome organisation and diversity: the year of the junk!** Current Opinion in Biotechnology. 17 (2):168-173. 2006.

MULLER, C. H. A revision of genus *Lycopersicon*. U.S.D.A. **Misc.Publ**. 382:1-28 + 10pl. 1940.

OCHOA, C. M**. Las papas de Sudamérica, Perú.** Centro Internacional da la Papa (CIP). 1989.

OGIWARA, I.; MIYA, M.; OHSHIMA, K.; OKADA, N. **Retropositional parasitism of SINEs on LINEs: identification of SINEs and LINEs in elasmobranchs.** Mol Biol Evol. 16(9):1238-50. 1999.

ORGEL, L. E.; CRICK, F. H. C. **Selfish DNA: the ultimate parasite**. Nature 284: 604–7. 1980.

PERLMAN, P.S.; BOEKE, J. D. **Ring Around the Retroelement**. Science. 303(5655): 182-184. 2004.

POUTEAU, S.; HUTTNER, E.; GRANDBASTIEN, M. A.; CABOCHE, M. **Specific expression of the tobacco Tnt1 retrotransposon in protoplasts.** EMBO J. 10(7):1911-8. 1991.

RIRIE, K. M.; RASMUSSEN, R. P.; WITTWER, C. T. **Product Differentiation by Analysis of DNA Melting Curves during the Polymerase Chain Reaction**. Analytical Biochemistry. 245:154-160. 1997.

RICE, P.; LONGDEN, I.; BLEASBY, A. **EMBOSS: The European Molecular Biology Open Software Suite.** Trends in Genetics. 16(6):276-277. 2000.

RICK, C. M. **Hybridization between *Lycopersicon esculentum* and *Solanum pennellii.* Phylogenetic and cytogenetic significance**. Proc Natl Acad Sci U S A. 46:78-82. 1960.

ROZAS, J.; ROZAS, R. **DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis**. Bioinformatics. 15:174-175. 1999.

SAN MIGUEL, P.; TIKHONOV, A.; JIN, Y. K. **Nested retrotransposons in the intergenic regions of the maize genome.** Science. 274: 765-768. 1996.

SAN MIGUEL, P.; BRANDON, S.; GAUT; TIKHONOV, A.; NAKAJIMA, Y.; BENNETZEN, J. L. **The paleontology of intergene retrotransposons of maize**. Nature Genetics .20, 43 – 45. 1998.

SAN MIGUEL, P.; BENNETZEN, J. L. **Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons**. Annals of Botany *82* (Supplement A)**:**37-44. 1998.

SEN, S. K.; HAN, K.; WANG, J.; LEE, J.; WANG, H.; CALLINAN, P. A.; DYER, M.; CORDAUX, R.; LIANG, P.; BATZER, M. A: **Human genomic deletions mediated by recombination between Alu elements**. Am J Hum Genet. 79(1):41-53. 2006.

SOLEIMANI, V. D.; BAUM, B. R.; JOHNSON, D. A. **Quantification of the retrotransposon BARE-1 reveals the dynamic nature of the barley genome.** Genome. 49:389-396. 2006.

SPOONER, D. M.; PERALTA, I. E. **Classification of wild tomatoes: a review**. Kurtziana. Tomo 28 (1): 45-54. 2000.

SUONIEMI, A.; TANKSKANEN, K.; SCHULMAN, A. H. **Gypsy-like retrotransposons are widespread in the plant kingdom.** Plant J. 13(5): 699-705. 1998.

SWOFFORD, D. L. **PAUP: Phylogenetic Analysis Using Parsimony, Version 3.0 s** (Computer program and manual distributed by the Center for Biodiversity, Illinois Natural History Survey, Champaign, IL 61820) 1992.

THOMPSON, J. D.; HIGGINS, D. G.; GIBSON, T. J. **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice.** Nucleic Acids Research, 22:4673-4680. 1994.

VERNHETTES, S.; GRANDBASTIEN, M. A.; CASACUBERTA, J. M. **The Evolutionary Analysis of the Tnt1 Retrotransposon in Nicotiana Species Reveals the High Variability of its Regulatory Sequences.** Mol Biol Evol. 15(7):827-836. 1998.

VICIENT, C. M.; SUONIEMI, A.; ANAMTHAWAT-JÓNSSON, K.; TANSKANEN, J.; BEHARAV, A.; NEVO, E.; SCHULMAN, A. H.. **Retrotransposon BARE-1 and its role in genome expansion in the genus Hordeum.** Plant Cell.11:1769-1784. 1999.

VICIENT, C. M.; KALENDAR, R.; SCHULMAN, A. H: **Variability, recombination, and mosaic evolution of the barley BARE-1 retrotransposo**n. J Mol Evol. 61(3):275-91. 2005.

VOYTAS, D. F.; CUMMINGS, M. P.; KONICZNY, A.; AUSUBEL, F. M.; RODERMEL, S. R. **Copia-like retrotransposons are ubiquitous among plants.** Proc Natl Acad Sci U S A. 1;89(15):7124-8. 1992.

WESSLER, S. R. **Eukariotic Transposable Elements: Teching Old Genomes New Tricks**. Chapter 8. The Implicit Genome (Caporale L, ed.). Oxford University Press. USA. 2006.

WHITE, S. E.; HABERA, L. F.; WESSLER, S. R**. Retrotransposons in the flanking regions of normal plant genes: A role for copia-like elements in the evolution of gene structure and expression.** Proc Natl Acad Sci U S A . 91: 11792-11796. 1994.

WIKSTRÖM, N.; SAVOLAINEN, V.; CHASE, M. W. **Evolution of the angiosperm calibrating he family tree.** Proc. R. Soc. Lond. 268:2211-2220. 2001.

WRIGHT, S.; FINNEGAN, D. **Genome evolution: Sex and the transposable element**. Current Biology**.** 11:R296-R299. 2001.

XIONG, Y.; EICKBUSH, T. H. **Origin and evolution of retroelements based upon their reverse transcriptase sequences.** EMBO J. 9(10):3353-62. 1990.

ZACHAR, Z.; DAVIDSON, D.; GARZA, D.; BINGHAM, P. M. **A detailed developmental and structural study of the transcriptional effects of insertion of the Copia transposon into the white locus of Drosophila melanogaster.** Genetics. 111(3):495-515. 1985.

http://www.tigr.org/index.shtml.	J.Craig Venter Institute
http://www.sgn.cornell.edu/index.pl.	Sol genomics Network.
http://www.ncbi.nlm.nih.gov/BLAST/	NCBI BLAST
http://www.ebi.ac.uk/clustalw/	CLUSTAW
http://www.ub.es/dnasp/	Dnasp
http://www.fluxus-technology.com/index.htm	NETWORK
http://www.megasoftware.net/	MEGA3

ANEXOS

ANEXO 1

Alinhamento das porções da RNAseH analisadas com o programa NETWORK.

Os números representam a matriz gerada no alinhamento pelo programa DNA *Alignment* 1.1.21 *software.* Ref.: seqüência referente; foi utilizada uma seqüência pertencente ao nó central α1 de Retrosol. Pontos indicam nucleotídeos iguais e líneas falta de nucleotídeos ou *deleções*. N.tom: *Nicotiana* N.tab: *N. tabacum*, N. gla: *N. glauca*; N.deb: *N. debneyi*, N. bem: *N. benthamiana,* N. plu: *N. plumbaginifolia*, N. syl: *N. sylvestris*, Lp: *Lycopersicon peruvianum*; Lpp: *L. pimpinellifolium;* Lh: *L. hirsutum*; Le: *L. esculentum*. , Ke-*Solanum tuberosum* var Kennebec, Cha-*S.t.*ssp.*andigena* var.Chacarera, Mo-*S.t.*ssp.*andigena* var.Moradita, Tu-*S.t.*ssp.*andigena* var.Tuni, bre-*S.brevidens*, acl-*S.acaule*, chc-*S.chacoense*, cmm-*S.commersonii* ssp *malmeanum*, dds-*S.xdoddsii*, go-*S.gourlayi*, ha-*S.hannemanii*, ifd-S.*infundibuliforme,* inc-*S.incamayoense*, ktz-*S.kurtzianum*, lph-*S.leptophyes*, mga-*S.megistacrolubom,* mcd-*S.microdontum*, oka-*S.okadae*, opl-*S.oplosence*, sat-*S.sanctae-rosae*, spg-*S.spegazzinii*, scr-*S.xsucrense*, tar-*S.tarijense*, vrn-*S.vernei*. As chaves agrupam as seqüências de cada elemento: Tnt1, Retrosol (grupos I, II e III), e Retrolyc1.

```
        123456789111111111122222222233333333344444444445555
                 012345678901234567801235678901234567890123


Ref    CAAGAGACAAGTTCGAGCTATGCAAAGACTTGCGGCATGCACTCAAGCTAG
Ntom20 ................AT.G...................A....T.TA.GA 1
Ntom19 ....T...........T................A............A.... 1
Ntom18 ................AT.G...................A....T.TA.GA 1
Ntom17 ...A..................AA.............G.........A.... 1
Ntom16 ................T................A............A.C.. 1
Ntom15 ................AT.G............................T... 1
Ntom14 ..............A.T................A............A.... 1
Ntom13 ................T................A............A.... 1
Ntom12 .............T..AT.G.............................. 1
Ntom11 ............................T.....A...AA....T.TG.GA 1
Ntom10 ......................A...........CATC.CATTCA..AGCCA 1
Ntom9  ..............................................T.TA.GA 1
Ntom8  ..............T...................T....A....T.TA.GA 1
Ntom7  ...A..................AA.............G.........A.... 1
Ntom6  ..............................................A.... 1
Ntom5  ..............A.T................A............A.... 1
Ntom4  ................AT.G.T............................ 1
Ntom3  ...A.............T......................G.....A.... 1
Ntom2  ................AT.G...................A....T.TA.GA 1
```

Tnt1

```
Ntom1  ................AT.G...................A....T.TA.GA 1
Ntab20 ....GA..................................T....A.... 1
Ntab19 ....GA.......................................A.... 1
Ntab18 ....GA....-..................................A.... 1
Ntab17 ....GA..................................T....A.... 1
Ntab16 .............T............C...........A.......A...- 1
Ntab15 ....GA.......................................A.... 1
Ntab14 ....GA......-................................A.... 1
Ntab13 ....GA.......................-..............A.... 1
Ntab12 .......T.....T...................C....A.......A.... 1
Ntab11 ....GA.......................................A.... 1
Ntab10 .............T...................C....A.......A.... 1
Ntab9  ....GA.......................................A.... 1
Ntab8  ....GA.......................................A.... 1
Ntab7  ....GA.......................................A.... 1
Ntab6  ....GA..................................T....A.... 1
Ntab5  ....GA..................................T....A.... 1
Ntab4  ................AT.G............................... 1
Ntab3  ....GA.......................................A.... 1
Ntab2  .............T...................C....A.......A.... 1
Ntab1  ....GA.......................................A.... 1
Ngla20 ....GA........................-..............AT... 1
Ngla19 ....GA..................................T....A.... 1
Ngla18 .............T........................A.......A.... 1
Ngla17 ....GA.......................................AT..- 1
Ngla16 ....GA.......................................AT..- 1
Ngla15 ...A......A...A................T..T..A....G..A...- 1
Ngla14 ................AT.G...............T....G.........- 1
Ngla13 ....GA..............................T.........A.... 1
Ngla12 ...A...T..A...A................T..T.....T.G..A.... 1
Ngla11 ...A......A...A................T..T..A....G..A...- 1
Ngla10 ...A...T..A...A................T..T.....T.G..A.... 1
Ngla9  ....GA.......T...............................AT..- 1
Ngla8  .............T....G.......T......C....A.......A.... 1
Ngla7  ...A...T..A...A................T..T.....T.G..A.... 1
Ngla6  .............T...................C....A.......A.... 1
Ngla5  .............T.....T.............T....A.......A...A 1
Ngla4  ................AT.G..............CT....G.........- 1
Ngla3  .G.A.........T...................C....A.......A.... 1
Ngla2  .............T...................C....A..T....A...A 1
Ngla1  ....GA.......................................AT..- 1
Ndeb20 ....GA..........A............................A.... 1
Ndeb19 ..............A.AT.G..T...A........T....G.........A 1
Ndeb18 ....GA...........AG..........................A.... 1
Ndeb17 ....GA..................................T....A.... 1
Ndeb16 ....GA........................A..............A...A 1
Ndeb15 ....GA........................T..T.....T....A.... 1
Ndeb14 ....G...........................C....A.......A.... 1
Ndeb13 ....GA..................................T....A.... 1
Ndeb12 ................T.G..............T............T... 1
Ndeb11 ....GA..........T............................A.... 1
Ndeb10 ....GA...............................G........GA.... 1
Ndeb9  ....GA..................................T....A.... 1
Ndeb8  ....GA...........T......................T....A.... 1
Ndeb7  ....GA................T......................A.... 1
Ndeb6  ....GA.......................................A.... 1
Ndeb5  ....G........T...................C....A.......A.... 1
```

Tnt1

```
Ndeb4  ....GA.........G.AG...........................A.... 1
Ndeb3  ....GA........................A.........T....A.... 1
Ndeb2  ....GA..................................T....A.... 1
Ndeb1  ....GA..............................G.........A.... 1
Nben20 ....GA...........A................A...A.......A.... 1
Nben19 ....GA.......................................A.... 1
Nben18 ....GA..................................T....A.... 1
Nben17 ....GA.......................................A.... 1
Nben16 ....GA......................C..........G.....A.... 1
Nben15 ....GA..................................T....A.... 1
Nben14 ....GA..................................T....A.... 1
Nben13 ....GA.......................................A.... 1
Nben12 ....GA.......................................A.... 1
Nben11 ....G...........T........................T....A.... 1
Nben10 ....GA........................A.........T....A.... 1
Nben9  ....GA.......................................A.... 1
Nben8  ....GA..........................................
Nben7  ....GA...................................C...A.... 1
Nben6  ....GA.......................................A.... 1
Nben5  ....TA....................A..................A.... 1
Nben4  ....GA...........T...........................A.... 1
Nben3  ....GA..................................A....A.... 1
Nben2  ....GA................T........C..............A..GA 1
Nben1  ....G........T...................C....A.......A.... 1
Nplu20 .....T.......T...................C....A..T....A.... 1
Nplu19 ....GA.......................................A.... 1
Nplu18 ....GA........A..............................A.... 1
Nplu17 ........T....T...................C...........A.... 1
Nplu16 .............T....G.......T......C....A.......A.... 1
Nplu15 ....GA..................................T....A.... 1
Nplu14 .............T....G.......T......C....A.......A.... 1
Nplu13 ....G........T.....................CATACACTCA..CTAGA 1
Nplu12 ....GA.......................................A.... 1
Nplu11 .............T....G.......T......C....A.......A.... 1
Nplu10 .............T....G.......T......C....A.......A.... 1
Nplu9  ....GA.......................................A.... 1
Nplu8  .............T....G.......T......C....A.......A.... 1
Nplu7  ....GA.......................................A.... 1
Nplu6  ....GA.......................................A.... 1
Nplu5  ....GA.......................................A.... 1
Nplu4  ....GA.......................................A.... 1
Nplu3  ...TGA....T..................................A.... 1
Nplu2  ....GA.......................................A.... 1
Nplu1  ....GA.......................................A.... 1
Nsyl20 ....GA.......................................A.... 1
Nsyl19 ....GA........-..............................A.... 1
Nsyl18 ....GA.......................................A.... 1
Nsyl17 ....GA..................................T....A.... 1
Nsyl16 .............T...................C....A.......A.... 1
Nsyl15 ....GA..................................T....A.... 1
Nsyl14 ....GA..................................T....A.... 1
Nsyl13 ....GA..................................T....A.... 1
Nsyl12 ......................A......................A.... 1
Nsyl11 ....GA.......................................AT..- 1
Nsyl10 ...A.................AA.............G.........A.... 1
Nsyl9  ....GA.......................................A.... 1
Nsyl8  ....GA..................................TCA..CTAGA 1
```

Tnt1

```
Nsyl7  ....GA.......................................A.... 1
Nsyl6  ....GA.......................................A.... 1
Nsyl5  ....GA.......................................A.... 1
Nsyl4  ....GA.......................................A.... 1
Nsyl3  ....GA.......................................A.... 1
Nsyl2  ....GA.......................................A.... 1
Nsyl1  ....GA.......................................A.... 1
acl3   ............T...................A................. 1
acl7   ...A.........................................A... 1
acl21  ................................A................. 1
bre1   ............T...............................T.... 1
bre2   ............T...............................T.... 1
bre3   ...............................................T 1
bre5   ............T...............................T.... 1
bre9   ...............................................T 1
bre10  ............T...............................T.... 1
bre13  ...............................................T 1
bre19  ...............................................T 1
bre20  ............T...............................T.... 1
bre26  ............................A..................... 1
bre27  ............T..A........C......................... 1
bre30  ............T............T..A..................... 1
bre31  ...............................................T 1
Cha2   .........T...................................... 1
Cha10  ................................................... 1
Cha11  ................A....A........................... 1
Cha14  .........T...................................... 1
Cha18  ............T..................T.A................. 1
Cha22  ............T..................T.A................. 1
Cha23  ..........A.....................C................... 1
chc25  ................A.................................. 1
cmm27  ........C........T...........C...................... 1
dds27  ............T............T..A..................... 1
go25   ............................A..................... 1
go27   ............T..................T.A................. 1
ha3    ............T..................TA................. 1
ha23   ............T..................TA................. 1
inc21  ..G.........T...................A................. 1
inc23  ..G.........T...................A................. 1
inf7   ............T..................T.A................. 1
inf9   ............T........T........T.A................. 1
inf21  ............T..................T.A................. 1
inf27  ............T........T........T.A................. 1
Ke32   ...A...................G.......A...............A... 1
Ke35   .....A................T............................ 1
ktz5   .....................AT........................A... 1
ktz7   .....................AT........................A... 1
ktz9   ................................A................. 1
ktz10  .....................AT........................A... 1
ktz15  ................................A................. 1
ktz25  .....................AT........................A... 1
ktz27  .....A.A........................................... 1
lph26  ............T..................T.A................. 1
mcd25  .......A.....................A...................... 1
mga5   ...........................................A... 1
mga9   ...A.......................................A... 1
mga10  ...A.......................................A... 1
```

Tnt1

GI
Retrosol

```
mga14   ...A.........................................A... 1
mga20   ...A.........................................A... 1
mga30   ...A.........................................A... 1
Mo9     ................A.................................. 1
Mo12    ..........................................G..... 1
Mo29    ............................A...........T......... 1
Mo32    ..........A.....A................................. 1
Mo33    ................................................. 1
Mo38    .....A...............T............................ 1
oka3    ............T....................A................. 1
oka19   ............T....................A................. 1
sat1    ............T....................A................. 1
sat5    ............T....................A................. 1
sact6   ............T....................A................. 1
sat23   ...A.................CT........................A... 1
scr14   ................................................. 1
scr16   ............T..................T.A................. 1
scr17   ................A....A........................... 1
scr20   ............................A..................... 1
scr23   ................A....A........................... 1
scr24   ................A....A........................... 1
spg3    ............T............T..A..................... 1
spg7    .......................G.....T..........C........ 1
spg13   ................................................. 1
spg23   ............T............T..A..................... 1
spg28   ............T............T..A..................... 1
tar4    ............................A..................... 1
Tu2    T...............A................................. 1
Tu3     ............................A..................... 1
Tu5     ................A................................. 1
Tu7     ...A...................G.......A...............A... 1
Tu13    ............................A..................... 1
Tu15    .....A................T........................... 1
Tu17    .....A................T........................... 1
Tu21    ................A................................. 1
ver3    ............T....................A................. 1
ver11   ............T....................A................. 1
acl13   ..............................................C.. 1
acl19   ................A.............................C.. 1
acl20   ...................................A............C.. 1
Cha3    ............T...................................... 1
Cha7    ...............................A................. 1
Cha8    ...............................A................. 1
Cha9    ...............................A.....T........... 1
Cha15   ................................................. 1
Cha17   ................................................. 1
Cha19   ................................................. 1
Cha20   ............T...................................... 1
Cha25   .C.............................................. 1
Cha28   ...................................G............... 1
cmm29   ...A............................................ 1
cmm30   ........................A..A..................... 1
go28    ......................A......................... 1
ha7     ................................................. 1
ha17    ................................................. 1
ha9     ................A................................. 1
ha10    ................................................. 1
```

GI Retrosol

GII Retrosol

```
ha22   .................A.................................. 1
ha26   .............T..................TA................. 1
ha31   .................................................... 1
Hu25   ..............A..T.......G......................... 1
Hu26   .................................................... 1
Hu27   .................................................... 1
Hu28   .................................................... 1
inc1   .................................................... 1
inc2   .................................................... 1
inc3   ......................A............................ 1
inc5   ...A..................A............................ 1
inc7   ....................................A............... 1
inf8   .............T..................T.A................. 1
inc10  .................................................... 1
inc13  ...A..................A............................ 1
inc15  .................................................... 1
inc24  .................................................... 1
inf1   ......................A...........T................ 1
inf4   .................................................... 1
inf14  .................................................... 1
inf25  ......................A...........T................ 1
inf30  .................................................... 1
Ke26   ..............A..T.......G......................... 1
ke27   .................................................... 1
Ke28   .................................................... 1
ke30   ..............A..T.......G......................... 1
ke31   .............T.....T............................... 1
mcd28  .................................................... 1
mga8   .................................................... 1
mga15  .................................................... 1
mga19  .................................................... 1
Mo1    .................................................... 1
Mo2    ..............A..................................... 1
Mo5    ..............A..................................... 1
Mo6    ...............................................G... 1
Mo7    .................................................... 1
Mo8    ..............A..................................... 1
Mo13   .....A.............................................. 1
Mo15   ..............A..................................... 1
Mo18   .................................................... 1
Mo23   ..............A..................................... 1
Mo24   .................................................... 1
oka7   .................................................... 1
oka11  .................................................... 1
oka21  .................................................... 1
opl25  ......................A............................ 1
opl26  .............T..................T.A................. 1
sact9  .................................................... 1
sact10 .................................................... 1
scr3   ................................T................... 1
scr7   ................................T................... 1
scr9   ......................A...........T................ 1
scr10  .................................................... 1
scr21  .................................................... 1
spg2   .................................................... 1
spg12  .................................................... 1
spg30  .................................................... 1
```

GII
Retrosol

```
spg31  ..................................................... 1
tar2   .............A...................................... 1
tar3   ............T....................................... 1
tar5   .............A...................................... 1
tar10  .............A...................................... 1
tar15  ..................................................... 1
tar20  ............T....................................... 1
tar23  ....................A........A...........T....... 1
tar17  ................A................................... 1
Ke33   ..................................................... 1
Tu8    ..................................................... 1
Tu9    ..................................................... 1
Tu14   T.................................................... 1
Tu18   ..................................................... 1
Tu19   ..................................................... 1
Tu23   ....................T............................ 1
ver1   ..................................................... 1
ver2   ......................................T........... 1
ha2    .G.A........T...T.............GA.T...AG...CCT...A 1
ha16   .G.A........T...T.............G..T...AG...TCT...A 1
ktz12  .G.A........T...T.............G..T...AG...CCT...A 1
ktz23  .G.A........T...T.......A.....G..T...AG...TCT...A 1
oka4   .G.A........T...T.............G..T...AG...CCT...A 1
oka6   .G.A........T...T.............G..T...AG...CCT...A 1
acl14  ...A.........A.C.....T.GG.......A.....AG...TCT.CGA 1
ha14   .G.A............T.............G..T...AGT..TCT...A 1
mga3   ...A............T...CT........T......AG...TCTAAG. 1
scr11  ...A............T...CT........T......AG...TCTAAGA 1
ktz3   .G.A........T...T.......A.....G..T...AG...TCT...A 1
ktz11  .G.A........T...T.............G..T...AG...TCT...A 1
ktz16  .G.A........T...T.......A.....G..T...AG...TCT...A 1
ktz19  .G.A........T...T.....C.......G..T...AG...TCT...A 1
ktz22  .G.A........T...T.......A.....G..T...AG...TCT...A 1
Lp1    ...AG.......T..C............C......C.T........... 1
Lp2    ...AG.......T................T......T........... 1
Lp3    ...AG.......T......................T........... 1
Lp4    ...AG.......T......................T........... 1
Lp5    ...AG.......T......................T........... 1
Lp6    ...AG.......T................T......T........... 1
Lp7    ...AG......CT................T......T........... 1
Lp8    ...A........T......................T........... 1
Lp9    ...AG.......T................T......T........... 1
Lp10   ...A........T......................T........... 1
Lp11   ....G.......T...................A....T........... 1
Lp12   ...A........T...T...................T........... 1
Lp13   ...AG.......T......................T........... 1
Lp14   ...A........T......................T........... 1
Lp15   ...A........T......................T........... 1
Lp16   ...AG.......T......................T........... 1
Lp17   ...AG.......T......................T........... 1
Lp18   ....G.......T..................................... 1
Lp19   ....G.......T..................................... 1
Lp20   ....G.......T......................T........... 1
Lp21   ...AG.......T......................T........... 1
Lp22   -..A........T......................T........... 1
Lp23   ...A........T......................T........... 1
Lp24   ...AG.......T......................T........... 1
```

G II Retrosol

GIII Retrosol

Retrolvc1

```
Lp25   ....G........T.........................T........... 1
Lp26   ...A.........T.........................T........... 1
Lp27   --.AGAGAG.A..T..................................... 1
Lp28   -..AG........T.............T...........T........... 1
Lp29   -..A.........T.........................T........... 1
Lp30   -..A.........T.........................T........... 1
Lp31   -..A.........T.........................T........... 1
Lp32   -...G........T.........................T........... 1
Lp33   ...A.........T.........................T........... 1
Lp34   -..A.........T.........-...............T........... 1
Lp35   --.AG.GAC...AT..................A.......T........... 1
Lp36   --.ACAGAC.A..T.........................T........... 1
Lp37   -..AG........T.........................T........... 1
Lp38   ---A.A.G.CAG.TT.AGCTATGC...............T........... 1
Lp39   -..AG........T.........................T........... 1
Lp40   --.A.AGAC.A..T..............A..........T........... 1
Lp41   -..AG........T.........................T........... 1
Lp42   -...G........T..................................... 1
Lp43   ---A.A.G.CAG.TTGAGCTATGC...............T........... 1
Lp44   -..A.........T.........................T........... 1
Lp45   --.A.AGAC....T.........................T........... 1
Lh1    .............T.........................G........... 1
Lh2    .............T..A......................G........... 1
Lh3    .............T............A...........G........... 1
Lh4    .............T..A......................G........... 1
Lh5    .............T.........................G........... 1
Lh6    ...A.........T..................A.................. 1
Lh7    .............T.........................G........... 1
Lh8    .............T.........................G........... 1
Lh9    .............T.........................G........... 1
Lh10   .............T.........................G........... 1
Lh11   ...A.........T.........................T........... 1
Lh12   .............T............A...........G........... 1
Lh13   .............T.........................G........... 1
Lh14   .............T............A...........G........... 1
Lh15   .............T............A...........G........... 1
Lh16   .............T.........................G........... 1
Lh17   .............T..A......................G........... 1
Lh18   .............T.........................G.....G..... 1
Lh19   .............T....C...................G........... 1
Lh20   -............T..A...G..................G........... 1
Lh21   .............T..A......................G........... 1
Lh22   .............T............A...........G........... 1
Lpp1   -.T.G........TA.................................... 1
Lpp2   ..T.G........TA.................................... 1
Lpp3   -.T.G........TA.................................... 1
Lpp4   -.T.G........TA.................................... 1
Lpp5   -.........C........C............G.................. 1
Lpp6   -T..G..AGG...TT.........................T........... 1
Lpp7   -T..G..AGG...TT.........................T........... 1
Lpp8   -T..G..AGG...TT.........................T........... 1
Lpp9   .T..G..AGG...TT.........................T........... 1
Lpp10  -T..G..AGG...TT.........................T........... 1
Lpp11  .T..G..AGG...TT.........................T........... 1
Lpp20  ....G........T.........................T........... 1
Lpp12  ....G........T.........................T........... 1
Lpp13  ..T.G........TA.................................... 1
```

Retrolyc1

```
Lpp14  ..T.G........TA..................................... 1
Lpp15  ..T.G........TA..................................... 1
Lpp16  ..T.G........TA..................................... 1
Lpp17  ..T.G........TA..................................... 1
Lpp18  ....G........T...............A.........T........... 1
Lpp19  ....G........T...............A.........T........... 1
Le1    ..T.G........TA..................................... 1
Le2    ....G........T...A....TC-..............T........... 1
Le3    -...G........T.........................T........... 1
Le4    -.T.G....-...TA..........-..-......................... 1
Le5    ..T.G........TA..................................... 1
Le6    -.T.G........TA..................................... 1
Le7    -.T.G........TA.C................................... 1
Le8    -.T.G........TA..................................... 1
Le9    ..T.G........TA..................................... 1
Le10   ..T.G........TA..................................... 1
Le11   ..T.G........TA..................................... 1
Le12   ..T.G........TA..................................... 1
Le13   ..T.G........TA..................................... 1
Le14   ..T.G........TA..................................... 1
Le15   ..T.G........TA..................................... 1
Le16   ..T.G........TA..................................... 1
Le17   ..T.G........TA..................................... 1
lp1ME  ...AG........T.........................T........... 1
lp2ME  ...AG........T.........................T........... 1
lp3ME  ...A.........T.........................T........... 1
lp4ME  ...AG........T.........................T........... 1
lp5ME  ...AG........T.........................T........... 1
lp6ME  ...A.........T.........................T........... 1
lp7ME  ...AG........T.........................T........... 1
lp8ME  ...AG........T.........................T........... 1
lp9ME  ...AG........T.........................T........... 1
lp10ME ....G..A.....T.......C.......................... 1
lp11ME ...AG........T.........................T........... 1
lp12ME ...AG........T.........................T........... 1
```

Retrolyc1

ANEXO 2

Seqüência de Retrosol1.

Localização (bp)	Descrição	Cor
1-540	LTR5′	---
546-558	PBS	---
621-4594	Poliproteina	---
4632-4641	PPT	---
4645-4945	LTR3′	---

>Retrosol1
caaatctcacaatgtcaaaaggttggcagatttggaagatattggatatacatatgtagat
tttttcaaaggttggtaaactgcaaagaagaaaaaaatgtttttcttctgtagcatagttg
gaaaaatatagcacttgttatctataaaagggagagtcttcttcaaactcaaaatgcacca
aaaacaaaggagaaaaaagaagagtgaggcatcacaggggaaattagtctgtgagaaaaaa
atagagagtgtgagcgatattgtagtgaggtggaaaaataaaaaaaaggttatttcttttg
aagtgtgtagtggtcttttgagttttactcagacctacaagtgtaaaaatccttactatag
tgatatcagtttctcctctcgggccgtggtttttcccttattcaagagggtttccacgcaa
aatcttggcgtcattattgctctttttattcttgttgattaaccacattgttgtgttccgc
ttttattattttattaccgcaaatattattttgttgcgggtttattcctaataat**TGGT**
ATCAGAGCAcatgttctgctcatttacaaaaatatttttgcagcttatcgtactatactca
gtgaaaaaaaatgtctggagtaatgtatgaggtaacaaaattcaaatagcggtttctgaac
gtggcagagaaggatgagagacttgctcatccaacagggtttgcacaaggcactaggcggc
aaatccaaaagcccgaaaccatgaaacttggggattgggaagaaatggttgaaaaggcttc
tagtgcaatcaggttgcacttaacaaatgatgtggttaataacatcatcgatgaagagagt
gcatgtggcatttggacaaagttagaaaatctatacatgtccaaaacgctgacaaataaat
tgtacctaaaaaattagttatatgccctacatatgggtgaaggtacacatttttgtccca
tttaaatgtgcttaatagactaatcacacagctagcaaacctcggagtaaagatcgaggag
gaagataaagacatagtgctcctaaactcgttgccatcttcctacgatgagttggcaacaa
ccatcttacatggtaaggacactattgagttgaaggatgtcacatcagcccttctgcttaa
tgaaaagatgagaaaaaagcctgaaaatcacggacaggctctcatcacagaaagtagaggc
agaagttaccaaaggagttcaagcaactatggtagatccggagctcgtggaaagtctaagg
tccgatcaaaatcaaaggccaaaaattgctacaattgcgaacaaccaggtcagttcaaaag
agattgtccgaatctaaaaagaagcaaaggagaaagcagtggccagaagaatgttgacaac
acagctgccatggtgcaaaataatgacagtgttgttctcctcataaacgaggaagaggaat
gcatgcacttggaaggtgtagagttggaatgggtggtcgacacaacaacatcttaccatgc
cacaccagtaagagatgtcttttgcagatatgtagcaggtgattttggcaatgtgaaaatg
ggtaacacaagttactcaaagattgtggggatcagagatgtttgcataaagtcaaatgtcg
gatgcatattagttttgaaaaatgtgcgccacgtacctgatttgcggatgaacttgatctt
gggagttgctttggaccgagatggataagagaactactttgcaaataaaaaatggagactc
accaaaggggcattgatgatttcaaaaggagttgctcgtggcacgttgtacaggacaaatg
cagaaatatgccaaggtgaattcaatgcggctcacgaagagatttctacagatttgtggca
tagaagaatgggtcatatgagcgagaaaggattgcaaattctttccaaaaaatcactcatc
tcttttgccaaaggtacaacgataaaaccttgtaactatgtttatttggtaagcaacatag
aatctcatttctgacatcatctgaaagaaagttgaatatacttgatttggtatattctgat
gtctttggtccaatggagatagaatcgatgggaggtaataaatactttgttacctttattg

atgatgcttcacaaaaattgcgggtttatatcttgaaaaccaaagatcaggtgtttcaagt
atttcaaaagtttcatgctctggtagaaagagagacgggtcggaagctaaaacgtctccaa
actgacaatggaggtgagtacacttcaagagaatttgaagagtactattcaaatcatggaa
tcaaacatgaaaagatagttcctggaatcccacaacacaatggtgtagctgagagaatgaa
tcacaccattgttgagaatgtgagaagcatgctcagaatggctaaattacctaagtcattc
tggggtgaagcagttcggacagcatgttatcttatcaatcatagtccatcagttccgttgg
agtttgacattccggagagagtttggatcaacaaggagatatcttactcgcacctaaaggt
gttcggttgcaaagcttttgcacatgtaccgaaggagcagagaaccaagctagatgataat
tcagttccctgcatattcatcggatatggagatgaagaattcaggtacagattgtgggatc
ctgtaaagaagaaggtcatcaaaagcagagatgtaatcttccgagaaagtgaagttggaac
tgccgatgctctatccgagaaggccaaggaaaagaatggtatagtccctaatcttattaca
attaattctaaccatcccataagtgcagaatgtacgatcgatgaagttgctgaacaggaag
agcaacctaatgagattgttgagcaggaggagcaacttggtgataataccgagcaagtgga
gtacccggaagaagaagaacaatctcaacctatgaggagattagaaagacagggtagagtc
atccaagtacccttcctcaaagtatgtcctcatcaatgatgaaggggagccagaaagtctt
aaagaaatgttgtctcatccagaaaagaaccaatggatgaaagccatacaagaagagatgg
aatctctacaaaaaaatggcacatacaagctagttaaacttccaaaaggcaaaagaccgct
aaagtgcaagtggatctttaaactcaagaaagatggaaatggcaagttggtcaggtacaaa
gctcgattggttgtaaaaggcttcgaacaaaagaaaggtattgattttgatgaaatttttt
cacctgttgtcaaaatgacttttattcgaactattttgagcctagcagctagcctagatct
tgaagtggagcaattggacgtgaaaactacatttcttcacggagatttggaagaaaagatt
tatatggagcaactagagggatttgaagtatttggaaagaaacacatggtgtgcaaattga
ataagagtctttatgggttgaaacaggcaccaaggcaatgatacaaaaagtttgactcttt
catgaaaaatcaaacgtacacaaagacctattctgatccatgtgtatacttcaaaagattt
tctgacaacaattttattattttattgttgtatgtggatgacattttaattgtaggacaag
acaaaggtttgattgcaaggttgaagggggatttgtccaagtcatttgacataagggactt
gggaccggcacaacaaattctagggatgaaaattgtttgagaacgaacaaaaataaagttg
tggctatctcaagagaagtacattgaacgtgtactagaacgcttcaacatgaagagtatta
aacctgttagcacgcctcttgcgagtcatctgaagttgagcaaacaaatatgtcctacaac
aaaaaaggagaaagagggaatgaccaaagttccttattcttcagcagtggggagtttgatg
tatacaatggtgtgtacaagacctgatattgctcatgcagttggtgttgttagcagattcc
ttgaatatcctggaaaagagcactgggagtcaattaagtggatacttaggtacctaagagg
taccacaagagattgtttgtgttttgaaggatctgatccaatcttgaaaggctatacagat
gctgatatggtagctgacctcgataacagaaaattcactactggatatttgtttacatttt
taggggagctatatcatggtagtcaaagttgcagaagtgtgttgccctctctacaacgga
ggcagagtatattgcagctactaaagctggcaaagagatagtatggctgtaacgatttctt
caggaacttggattgcatcagaaagagtatgtcgtctattgtgacagtcaaagtgcaatag
tcttgagcaagaacacgatgtatcatgcaaggacgaagcatatcgacgtgagatatcactg
gattcgtgaaaagatggaggatggaatctatgcatgtcatgaagatccctactagtgaaaa
cccttcggatatgctgaccaaggtggtatcaagagacaagttcgagctatgtaaagaactt
gtcgacatgcactaaagctagaaacttagagctacctctttcaagtgaatgggtctGGAGG
GGAGAtttttggggtccatccacatttgccaaatctcacaatgtcaaaaggttggcagattt
ggaagatattggatatacatgtgcagatttttcaaaggttggtaagctgtaaagaagaaa
aaattgtttttcttctgtagcatagttggaaaaatactgcacttgttatctataaaaggga
aagtcttcttcattctcaaaatgaaccaaaaacaaaggagagaaaagaagagtgaggcatc
acagaggaaattagtctgtgaggaaaaaatagagagtgtgagcgatattgtagtgaggtgg
gaaa

Poliproteina

Localização (aa)	Descrição	Cor
229-246	Dedo de zinco	---
296	Domínio catalítico Protease	-
472-641	Domínio catalítico Integrase	---

MSGVMYEVTKFXXSGF*TWQRRMRDLLIQQGLHKALGGKSKXPETMKLGDWEEMVEKASSA
IRLHLTNDVVNNIIDEESACGIWTKLENLYMSKTLTNKLYLKN*LYALHMGEGTHFLSHLN
VLNRLITQLANLGVKIEEEDKDIVLLNSLPSSYDELATTILHGKDTIELKDVTSALLLNEK
MRKKPENHGQALITESRGRSYQRSSSNYGRSGARGKSKVRSKSKAKNCYNCEQPGQFKRDC
PNLKRSKGESSGQKNVDNTAAMVQNNDSVVLLINEEEECMHLEGVELEWVVDTTTSYHATP
VRDVFCRYVAGDFGNVKMGNTSYSKIVGIRDVCIKSNVGCILVLKNVRHVPDLRMNLILGV
ALDRDG*ENYFANKKWRLTKGALMISKGVARGTLYRTNAEICQGEFNAAHEEISTDLWHRR
MGHMSEKGLQILSKKSLISFAKGTTIKPCNXCLFGKQHRISFLTSSERKLNILDLVYSDVF
GPMEIESMGGNKYFVTFIDDASQKLRVYILKTKDQVFQVFQKFHALVERETGRKLKRLQTD
NGGEYTSREFEEYYSNHGIKHEKIVPGIPQHNGVAERMNHTIVENVRSMLRMAKLPKSFWG
EAVRTACYLINHSPSVPLEFDIPERVWINKEISYSHLKVFVSQAFAHVPKEQRTKLDDNSV
PCIFIGYGDEEFRYRLWDPVKKKVIKSRDVIFRESEVGTADALSEKAKEKNGIVPNLITIN
SNHPISAECTIDEVAEQEEQPNEIVEQEEQLGDNTEQVEYPXXEEEQSQPMRRLERXRVES
SKYPSSKYVLINDEGEPESLKEMLSHPEKNQWMKAIQEEMESLQKNGTYKLVKLPKGKRPL
KCKWIFKLKKDGNGKLVRYKARLVVKGFEQKKGIDFDEIFSPVVKMTFIRTILSLAASLDL
EVEQLDVKTTFLHGDLEEKIYMEQLEGFEVFGKKHMVCKLNKSLYGLKQAPRQ*YKKFDSF
MKNQTYTKTYSDPCVYFKRFSDNNFIILLLYVDDILIVGQDKGLIARLKGDLSKSFDIRDL
GPAQQILGMKIV*ERTKIKLWLSQEKYIERVLERFNMKSIKPVSTPLASHLKLSKQICPTT
KKEKEGMTKVPYSSAVGSLMYTMVCTRPDIAHAVGVVSRFLEYPGKEHWESIKWILRYLRG
TTRDCLCFEGSDPILKGYTDADMVADLDNRKFTTGYLFTFLGGAISW*SKLQKCVALSTTE
AEYIAATKAGKEIVWL*RFLQELGLHQKEYVVYCDSQSAIVLSKNTMYHARTKHIDVRYHW
IREKMEDESMHVMKIPTSENPSDMLTKVVSRDKFELCKELVDMH*S*

*: códon de parada (do inglês, *stop códon*)

Código de letras usadas para designar resíduos de amino ácidos: D=aspartato; E=acido glutâmico; H=histidina; S=serina; C=cisteina; K=lisina; G=glicina; A=alanina; L=leucina; V=valina; I=isoleucina; P=prolina; F=fenilalanina; T=treonina; Y=tirosina; N=asparagina; Q=glutamina; R=arginina; W=triptofano; M=metionina.

ANEXO 3

BMC Evolutionary Biology

BioMed Central

Research article

Open Access

Radiation of the Tnt1 retrotransposon superfamily in three Solanaceae genera

Maria E Manetti[1], Magdalena Rossi[1], Ana PP Costa[1,2], Andrea M Clausen[3] and Marie-Anne Van Sluys*[1]

Address: [1]GaTe lab, Departamento de Botânica, Instituto de Biociências – Universidade de São Paulo Departamento de Botânica – IBUSP, Rua do Matão, 277; 05508-090; São Paulo, S.P. Brasil, [2]Faculdade de Ciências Biológicas, Exatas e Experimentais, Universidade Presbiteriana Mackenzie, Brazil and [3]Estación Experimental Agropecuaria Balcarce, Instituto Nacional de Tecnología Agropecuaria (INTA), Argentina

Email: Maria E Manetti - maria_manetti@hotmail.com; Magdalena Rossi - magda1708@yahoo.com; Ana PP Costa - appcosta@yahoo.com; Andrea M Clausen - aclausen@balcarce.inta.gov.ar; Marie-Anne Van Sluys* - mavsluys@usp.br

* Corresponding author

Published: 7 March 2007

BMC Evolutionary Biology 2007, **7**:34 doi:10.1186/1471-2148-7-34

Received: 6 September 2006
Accepted: 7 March 2007

This article is available from: http://www.biomedcentral.com/1471-2148/7/34

© 2007 Manetti et al; licensee BioMed Central Ltd.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Tnt1 was the first active plant retrotransposon identified in tobacco after nitrate reductase gene disruption. The Tnt1 superfamily comprises elements from *Nicotiana* (Tnt1 and Tto1) and *Lycopersicon* (Retrolyc1 and Tlc1) species. The study presented here was conducted to characterise Tnt1-related sequences in 20 wild species of *Solanum* and five cultivars of *Solanum tuberosum*.

Results: Tnt1-related sequences were amplified from total genomic DNA using a PCR-based approach. Purified fragments were cloned and sequenced, and clustering analysis revealed three groups that differ in their U3 region. Using a network approach with a total of 453 non-redundant sequences isolated from *Solanum* (197), *Nicotiana* (140) and *Lycopersicon* (116) species, it is demonstrated that the Tnt1 superfamily can be treated as a population to resolve previous phylogenetic multifurcations. The resulting RNAseH network revealed that sequences group according to the Solanaceae genus, supporting a strong association with the host genome, whereas tracing the U3 region sequence association characterises the modular evolutionary pattern within the Tnt1 superfamily. Within each genus, and irrespective of species, nearly 20% of Tnt1 sequences analysed are identical, indicative of being part of an active copy. The network approach enabled the identification of putative "master" sequences and provided evidence that within a genus these master sequences are associated with distinct U3 regions.

Conclusion: The results presented here support the hypothesis that the Tnt1 superfamily was present early in the evolution of Solanaceae. The evidence also suggests that the RNAseH region of Tnt1 became fixed at the host genus level whereas, within each genus, propagation was ensured by the diversification of the U3 region. Different selection pressures seemed to have acted on the U3 and RNAseH modules of ancestral Tnt1 elements, probably due to the distinct functions of these regions in the retrotransposon life cycle, resulting in both co evolution and adaptation of the element population with its host.

Background

Retrotransposons are mobile genetic elements that transpose via an RNA intermediate. They are abundant and widespread components of eukaryotic genomes [1]. There are two major types of retrotransposons: long terminal repeat (LTR) retrotransposons and non-LTR retrotransposons such as LINEs (long interspersed nuclear elements) and SINEs (short interspersed nuclear elements) [2,3]. LTR retrotransposons are abundant in plant genomes and can constitute a very large fraction of the host genome [4,5].

LTR retrotransposons are flanked by LTRs in direct orientation. LTRs are subdivided into U3 (unique 3'RNA), R (repeated RNA) and U5 (unique 5'RNA) regions. Regulatory signals such as promoter, terminator regions and polyadenylation sites are found in the LTRs [2]. Besides the functional importance of these sequences, several studies have reported that the LTRs are the most-rapidly evolving regions of retrotransposons [6,7].

LTR retrotransposons encode a number of proteins derived from the *gag* and *pol* genes, which are usually transcribed as two main mRNAs. The proteins in *pol* are synthesised as a polyprotein that is cleaved by an internal protease activity releasing the internal protein activities endonuclease/integrase, reverse transcriptase and ribonuclease H (RNAseH). The *gag* gene encodes a protein involved in maturation and packaging of the RNA form of the retrotransposon. The life cycle of an element involves transcription by cellular RNA polymerase II, reverse transcription, packaging into virus-like particles, and integration of the cDNA copy back into the genome [2].

The retrotransposon Tnt1, first characterised in tobacco, is one of the few plant retrotransposons for which transpositional activity has been demonstrated [8]. Tnt1-like sequences have been detected in several *Solanaceae* and in *Nicotiana* at least three major groups can be differentiated based on U3 regulatory region sequence divergence giving rise to Tnt1A, B and C subfamilies [9-12]. Retrolyc1 elements detected in tomato share extensive nucleotide similarities to Tnt1 elements except in the U3 region [13,14]. Distinct from Tnt1, Retrolyc1 sequences comprise two subfamilies, Retrolyc1A and B, also distinguished at the regulatory U3 region [14].

The retrotransposon life cycle results in genome amplification and for this reason their activity is tightly controlled by the host. Besides rounds of amplification, host genome size can be scaled down by homologous recombination events within an element or between elements, usually leading to an increase in the copy number of SOLO-LTR sequences [15]. The work presented here aims to address the question of 'how' Tnt1-related sequences differentiated in three Solanaceae genera (*Nicotiana*, *Lycopersicon* and *Solanum*) and if there is any evidence for the occurrence of lateral gene transfer. These questions were addressed in a total of 453 non-redundant sequences using a population model approach based on recent studies of the amplification of *Alu* sequences among human and chimpanzee lineages [16,17].

The results presented here demonstrate the existence of Tnt1 superfamily related lineages in all species studied. Sequences analyses corroborate the hypothesis that the Tnt1-superfamily has evolved through differentiation of the U3 region as previously suggested [12]. RNAseH clustering studies revealed the existence of a major representative RNAseH sequence (α sequence) suggesting the presence of active copies. However, the highly diverse U3 region presents inter- and intra-specific diversification, which supports the hypothesis of a rapidly evolving sequence that can be interpreted as a strategy either to evade negative control by the host or to quickly adapt to the new evolving host genome thus contributing to host fitness.

Results

Tnt1-like sequences within different Solanum species

The presence of Tnt1-like sequences within 25 wild and cultivated *Solanum* genotypes (Table 1) was assayed by PCR using primers designed on the Tnt1 sequence anchored in the terminal part of the ribonuclease H (RNAseH) domain (Avi), and on the U5 region (Ol16) spanning part of RNAseH, Linker, U3, R and a portion of U5 region (Figure 1). Fragments of the expected size (500 bp) were detected in all genotypes analysed (Table 2). Sequence analysis using the NCBI Blast tool [35] confirmed that all 317 cloned fragments exhibited similarity to the Tnt1 superfamily and were therefore named "Retrosol" to designate them as retrotransposons from *Solanum*. Alignment of nine Retrosol sequences, chosen randomly from seven distinct species, to Tnt1 and Retrolyc1 retrotransposon sequences revealed that the similarity spans the RNAseH, R and U5 regions but not U3 (Figure 2).

Intra-genotype redundancy was eliminated and only unique sequences within species were considered for further sequence studies (Table 2). As these species do not have well-characterised genomes, it cannot be determined whether the amplified Retrosol fragments represent an unbiased random sample of sequence diversity. However, for a survey of the evolution of this element family, this approach seems appropriate as it was also used previously in the characterisation of the Tnt1 and Retrolyc1 families. A few sequences had a premature stop codon and others no stop codon at all in the RNAseH domain (Table 2).

Table 1: List of *Solanum* species used in this study

accession	section	subsection	Seria	Species	Code	Ploidy
Bal20002	Petota	Estolonifera		*S. brevidens*	Bre	2x
Bal80009		Potatoe	Commersoniana	*S. commersonii* subp.*malmeanum*	Cmm	2x
Bal01280			Yungasensia	*S. chacoense*	Chc	2x
Bal03017				*S. tarijense*	Tar	2x
Bal01113			Megistracroloba	*S. megistacrolobum*	Mga	2x
Bal74326				*S. sanctae-rosae*	Sat	2x
Bal01156			Cuneolata	*S. infundibuliforme*	Ifd	2x
Bal86017			Tuberosa	*S. x doddsii*	Dds	2x
Bal87066				*S. leptophyes*	Lph	2x
Bal90081				*S. microdontum*	Mcd	2x
Bal01121				*S. oplocense*	Opl	4x
Bal86053				*S. x sucrense*	Scr	4x
Bal01007				*S. gourlayi*	Gou	2x
Bal83042				*S. hannemanii*	Ha	2x
Bal73042				*S. incamayoense*	Inc	2x
Bal90044				*S. kurtzianum*	Ktz	2x
Bal74282				*S. okadae*	Oka	2x
Bal03013				*S. spegazzinii*	Spg	2x
Bal74330				*S. vernei*	Vrn	2x
				S. tuberosum		
				cultivar *Kennebec*	Ke	4x
				cul.*Huinkul*	Hu	4x
				S. tuberosum subp. *andigenum*		
				cul.*Chacarera* 109	Cha	4x
				cul . *Moradita* 64	Mo	4x
				cul.*Tuni* 207	Tu	4x
Bal01109			Acaulia	*S. acaule*	Acl	4x

Accession means collection, year and locality. x in species name, means hybrid.

The nucleotide divergence index (πJC), with Jukes and Castor's correction, was calculated for the full-length fragment sequence within each genotype. Numbers range from 0.026 to 0.287, and allow the genotypes to be classified within three sets (Table 2). The most diverse displays πJC values ≥ 0.2 and encompassed *S. hannemanii*, *S. kurtzianum* and *S. okadae* species. The second, involving *S. acaule*, *S. megistacrolubom* and *S. xsucrense*, presented values between 0.1 and 0.2; while the third, involving all remaining species had an πJC index of < 0.1. It is worth pointing out that cultivated potatoes (*S.tuberosum* cultivar Huinkul, Kennebec, and subsp. *andigenum* cul. Moradita, Chacarera and Tuni) were in the least diverse group.

Phylogenetic Relationships of Solanaceae Tnt1-related sequences

To visualise the relationship between all Solanaceae Tnt1-related sequences, a phylogenetic analysis was performed using all full-length fragments amplified. Retrosol sequences were aligned, together with representatives of Tnt1A B and C, and Retrolyc1A and B subfamily sequences. Three groups were obtained supported by a bootstrap of 98% or higher (Figure 3). Group I contains 87 sequences, including both wild and cultivated species and all 15 sequences amplified from *S. brevidens* (Table 2). Nucleotide identity between sequences varies from 80–90%. Group II has 95 sequences from wild and cultivated species (Table 2) and nucleotide identity is higher than 90% between sequences. Group III has 15 sequences belonging to 6 wild species: *S. acaule*, *S. hannemanii*, *S. kurtzianum*, *S. okadae* and *S. xsucrense* (Table 2) with variable nucleotide identity ranging from 60–95%. The Tnt1 C sequence clusters to Group III while Tnt1 A, B and Retrolyc1 A and B form a bridge between Groups I and II and the other sequences.

The πJC index calculated for all unique Retrosol sequences was 0.087 ± 0.0000571. Considering Retrosol phylogenetic groups separately, the nucleotide variability is lower for sequences of Group II (0.030 ± 0.0000016), and for Group I (0.063 ± 0.0000035). Group III has the highest value (0.220 ± 0.0009489), probably representing more than one group of related sequences. Nucleotide diversity is not uniformly distributed amongst the amplified fragments. When a comparative sliding window for nucleotide diversity is applied on all aligned sequences or within each group, as shown in Figure 4, higher substitution rates are seen to be in the U3 region. However, when nucleotide diversity distribution was analysed within each group, Group II showed a less diverse U3 region com-

Table 2: Genetic variability parameters of Tnt1-like sequences of *Solanum* wild and cultivated genotypes

code	Domest.	total	unique		size bp	π JC	I	II	III	premature	no stop/cod
bre	wild	25	13	52%	456(± 7)	0,057	13	-	-	1	5
cmm		3	3	100%	423(± 5)	0,093	1	2	-	-	-
chc		1	1	100%	428	-	1	-	-	-	-
tar		15	9	60%	422(± 14)	0,038	1	8	-	1	-
mga		17	10	58%	439(± 19)	0,171	6	3	1	-	-
sat		7	6	85%	440(± 4)	0,069	4	2	-	-	-
inf		23	10	43%	428(± 9)	0,053	4	6	-	2	-
dds		1	1	100%	435	-	1	-	-	1	-
lph		1	1	100%	436	-	1	-	-	-	-
mcd		2	2	100%	428(± 11)	0,078	1	1	-	-	-
opl		2	2	100%	419(± 2)	-	-	2	-	1	-
scr		12	12	100%	463(± 11)	0,132	6	5	1	1	1
gou		3	3	100%	430(± 8)	0,093	2	1	-	1	-
ha		19	12	63%	428(± 9)	0,249	2	7	3	-	-
inc		21	11	52%	421(± 4)	0,041	2	9	-	3	-
ktz		18	14	77%	443(± 8)	0,287	7	-	7	-	-
oka		16	7	43%	434(± 10)	0,273	2	3	2	-	-
spg		20	9	45%	433(± 7)	0,055	5	4	-	3	-
vrn		12	4	27%	429(± 9)	0,072	2	2	-	-	-
Ke	cult.	8	8	100%	426(± 14)	0,078	2	6	-	-	-
Hu		4	4	100%	421	0,026	-	4	-	-	-
Cha		23	17	74%	424(± 7)	0,064	7	10	-	-	-
Mo		29	17	58%	423(± 6)	0,057	6	11	-	-	-
Tu		17	14	82%	428(± 10)	0,058	8	6	-	-	-
acl	wild	18	7	38%	419(± 19)	0,191	3	3	1	-	4
total		317	197	62%			87	95	15	14	10

Note. Code refers to the codes use in table 1 for each species and cultivars. Domestication (Domest.) status: wild or cultivated (cult.) genotypes. Total: indicates total number of sequences obtained; unique genotypes: total number of different sequences and percentage; size: average fragment size in bp (size variation); πJC indicates nucleotide diversity with Jukes and Castor's (1969) correction; I, II and III refers to the sequence identity group obtained from the tree presented in figure 3 and numbers in these columns indicates numbers of sequences; premature: refers to number of sequences with premature stop codon and no stop/cod: refers to number of sequences with no stop codon in the RNAseH domain.

pared to the other groups, while Group III sequences exhibited the highest diversity in the U3 region. The data support the hypothesis that *Solanum* species harbour at least three versions of U3 regulatory regions, although based on previous Tnt1 superfamily U3 divergence studies [12], Group III could represent more than one version.

NETWORK analysis of Tnt1 superfamily ribonuclease H

Network approaches have been designed to investigate relationships between closely related sequences, allowing identification of persistent ancestral nodes, multifurcations and reticulations, which are not resolved when applying a conventional phylogenetic package [19]. In the network approach, the most-represented identical sequences – named α or master copy – occupy a central position in the net with high numbers of branches suggesting that the sequence represents part of an active element. This approach is based on the parsimony principle, which connects data sets with the minimum number of evolutionary steps. Reticulations on the net could result from recombination and/or homoplasy [18,19].

Considering the Tnt1 superfamily of retrotransposons in the genera *Solanum*, *Lycopersicon* and *Nicotiana* as a population of sequences with a common ancestor, a Network study was carried out with the aim of testing the hypothesis that a common α master sequence exists in Tnt1-superfamily. The 53 nucleotides corresponding to the last portion of the RNAseH coding domain from a total of 453 Tnt1 like-sequences (197 Retrosol sequences reported in this study, 140 Tnt1 sequences characterised from 7 *Nicotiana* species and 116 Retrolyc1 sequences from 4 species of *Lycopersicon*) were used for network analysis. The resulting network (Figure 5a) revealed that Tnt1, Retrolyc1 and Retrosol did not share any RNAseH sequence as revealed by the absence of shared nodes, although different species within each genus did share a master sequence.

In Retrosol sequences, the unique α node observed (α1) represents 25.5% of all Retrosol sequences and encompasses *S. hannemanii*, *S. incamayoense*, *S. infundibuliforme*, *S. microdontum*, *S. megistacrolobum*, *S. okadae*, *S. sanctarosae*, *S. spegazzinii*, *S. tarijense*, *S. x sucrense*, *S. vernei* and all

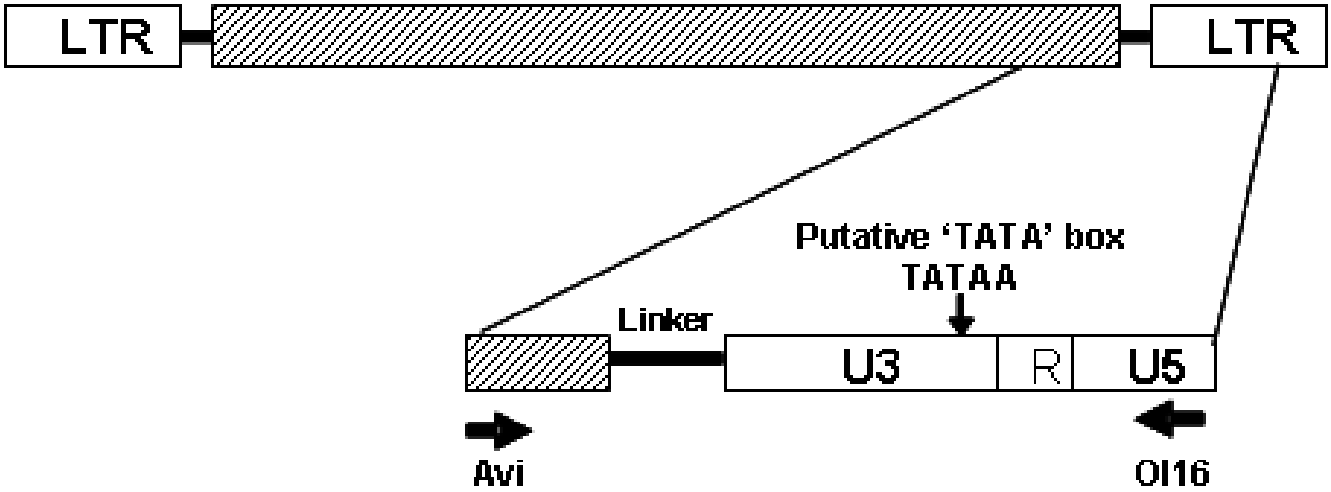

Figure 1
Schematic representation of the amplified retrotransposon fragment. LTR: long terminal repeat. RNAseH: ribonuclease H; Linker: noncoding region with PPT (polypurine track); U3: unique 3' RNA region; R: repeat RNA; and U5: unique 5'RNA region. Thick arrows indicate the position of the primers used in the amplifications.

S. tuberosum cultivars. It is worth mentioning that 46 of the 50 α1 sequences belong to Group II, which is the least diverse phylogenetic group. Moreover, within α1 group it was possible to identify identical U3 regions belonging to different species and cultivars. A total of 13 non-redundant *S. brevidens* sequences are connected with the α sequence in the net through a nucleotide change. None of the sequences belonging to group III appear in the α1 cluster; group III sequences are located at a more distant position in the net, suggesting a more distant relationship with the core sequence and probably do not originated from the α1 node.

Tnt1 sequences originated from two α nodes: α2 with 24% (34 sequences) and α3 with 12 % (17 sequences) of Tnt1 sequences (Figure 5b). α2 sequences belonged to: *Nicotiana tabacum*, *N. debneyi*, *N. benthamiana*, *N. plumbaginifolia* and *N. sylvestris*, embracing the three Tnt1 subfamilies (A, B and C). α3 clustered sequences from *Nicotiana tabacum*, *N. glauca*, *N. debneyi*, *N. benthamiana*, *N. plumbaginifolia* and *N. sylvestris*, grouping Tnt1 subfamilies A and C. It is interesting to note that Tnt1-94, for which transpositional activity has been demonstrated [8], grouped in the α3 node.

Retrolyc1 showed three different α nodes: α4 with 18.9% (22 sequences), α5 with 17% (20 sequences) and α6 with 13% (15 sequences) of Retrolyc1 sequences (Figure 5b). α4 include sequences from *Lycopersicon pimpinellifolium* and *L. esculentum* – all from the Retrolyc1 B subfamily. α5 has sequences that belong to *L. peruvianum* and to the Retrolyc1 A subfamily, while α6 has 14 sequences that belonged to *L. peruvianum* and only one from *L. hirsutum* – all from Retrolyc1 A. Identical U3 sequences were shared by different species in the α4 node.

Other data nodes with different numbers of sequences are connected with the central α1, α2, α3, α4, α5 and α6 nodes, suggesting that these members could be active and could contribute to the expansion of each element inside its corresponding genus (Figure 5b). The network did not show any excess of multidimensional reticulations so the influence of homoplasy can be considered as negligible [19].

Discussion

Genome expansion and contraction resulting from rounds of amplification/deletion of transposable elements is becoming accepted as a major component influencing the diversification of eukaryote genomes, as suggested by studies in particular plant and animal species [15,19]. Genome sequencing project also shed light on the ancient associations of known transposable element families with plant genomes. For example, in the work described by Rossi *et al.* (2004) [20], a phylogenetic study on Mutator-like elements (MULE) identified in sugarcane revealed the existence of 4 MULE lineages in Angiosperms prior to the divergence of Monocots and Eudicots.

The retrotransposon Tnt1 was initially characterised in *Nicotiana tabacum* and then detected by probe hybridisation in other Solanaceae genomes [8]. Ten years later, a new Tnt1-like element, Retrolyc1, was characterised in *Lycopersicon peruvianum* [13]. In this study we report a new

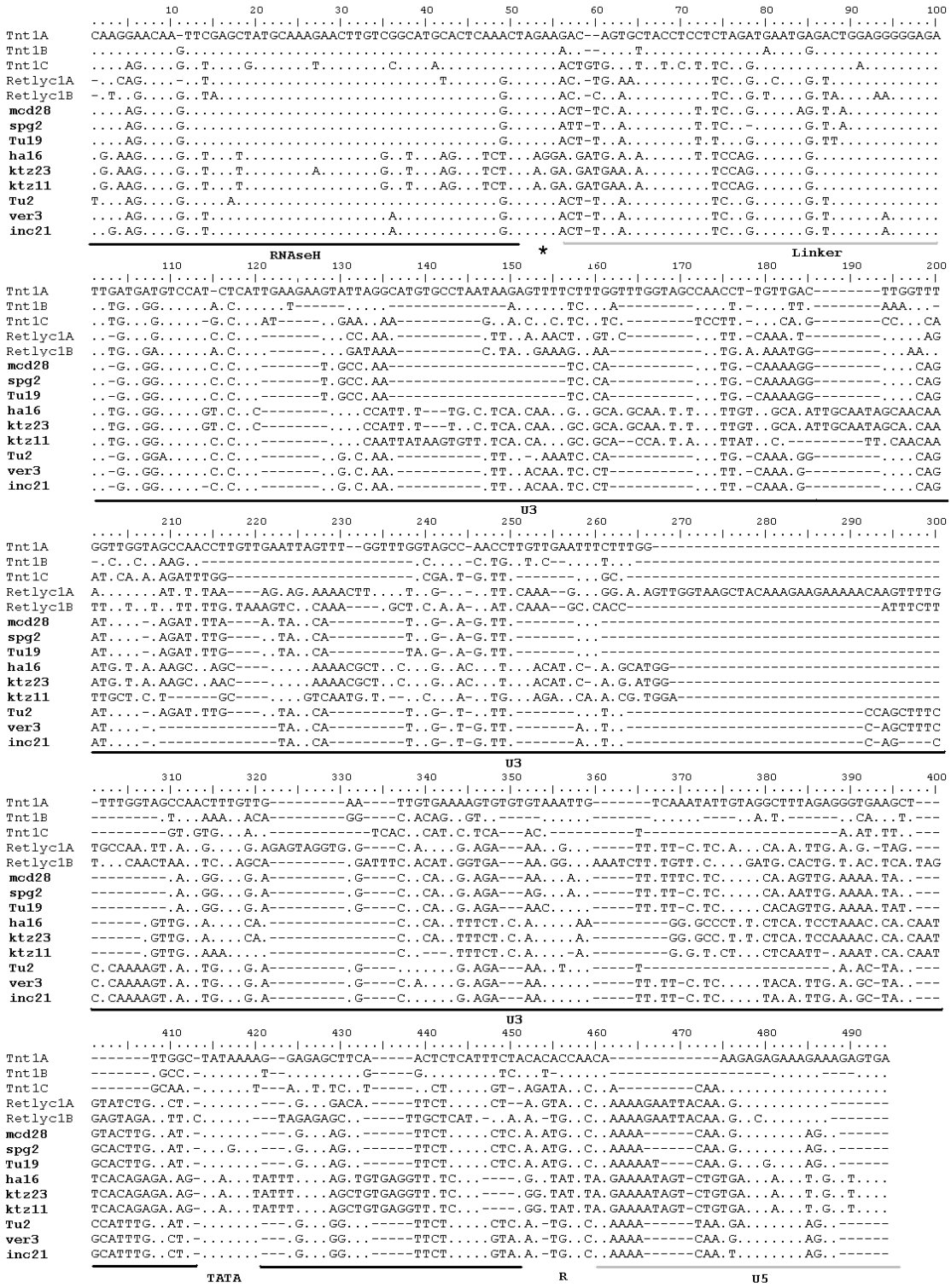

Figure 2

Sequence alignment of the partial amplified fragments. Representatives from Tnt1A, B and C; Retrolyc1A and B; and six sequences from Retrosol were included (names in bold). Underlined regions denote: RNAseH (ribonuclease), Linker, U3 (unique 3'RNA region), TATA, R (repeat RNA) and U5 (unique 5' RNA).

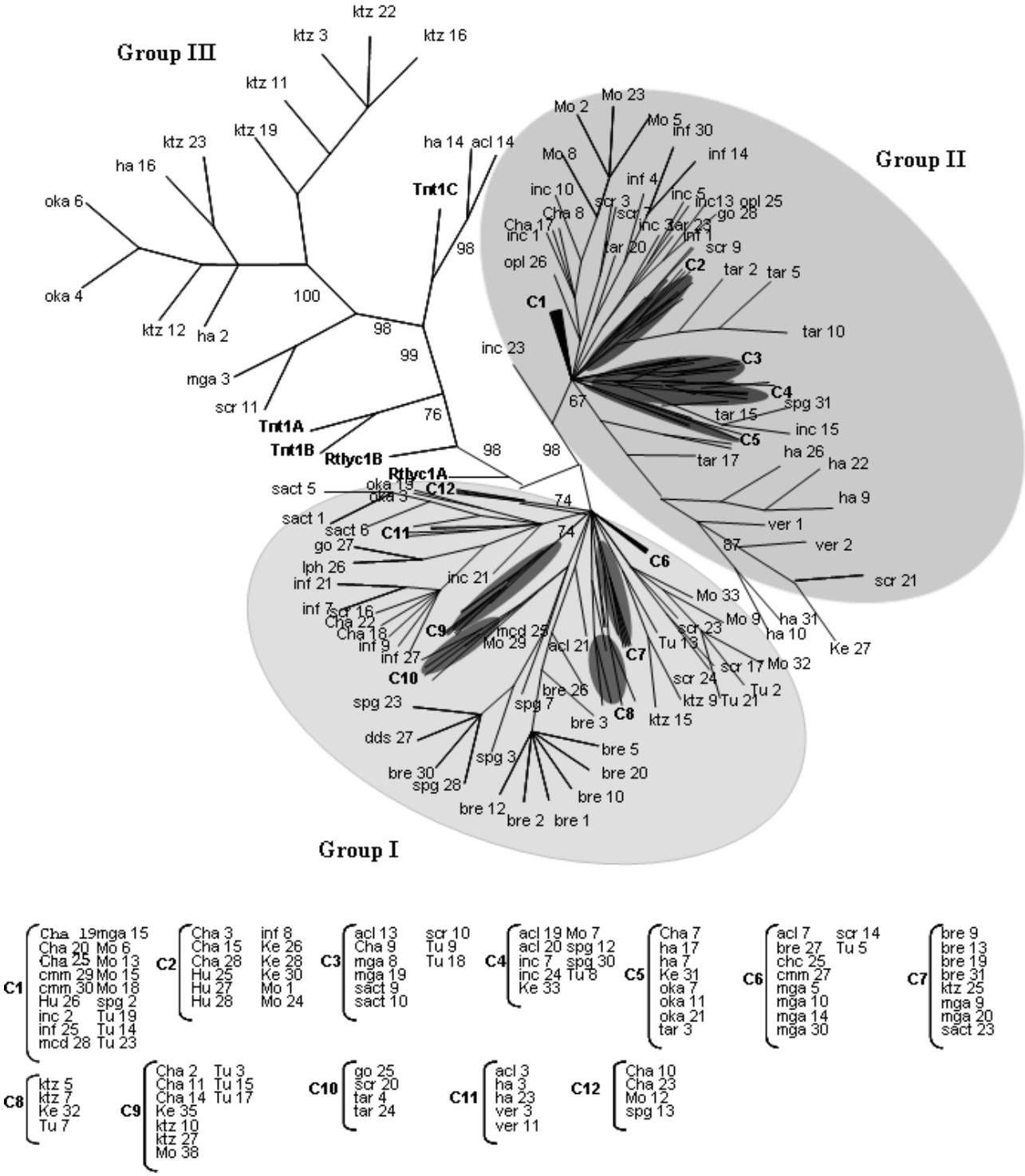

Figure 3

Phylogenetic analysis of Tnt1 Solanaceae superfamily. Phylogenetic analysis was performed with 197 sequences amplified from different species of *Solanum* (species are indicated by a code cited in Table 1 and different clones are indicated by numbers), and representative sequences from Tnt1 A, B and C, and Retrolyc1 A and B. The aligned nucleotide sequences span the last fragment of the RNAseH domain, the linker, U3, R and part of the U5 region.

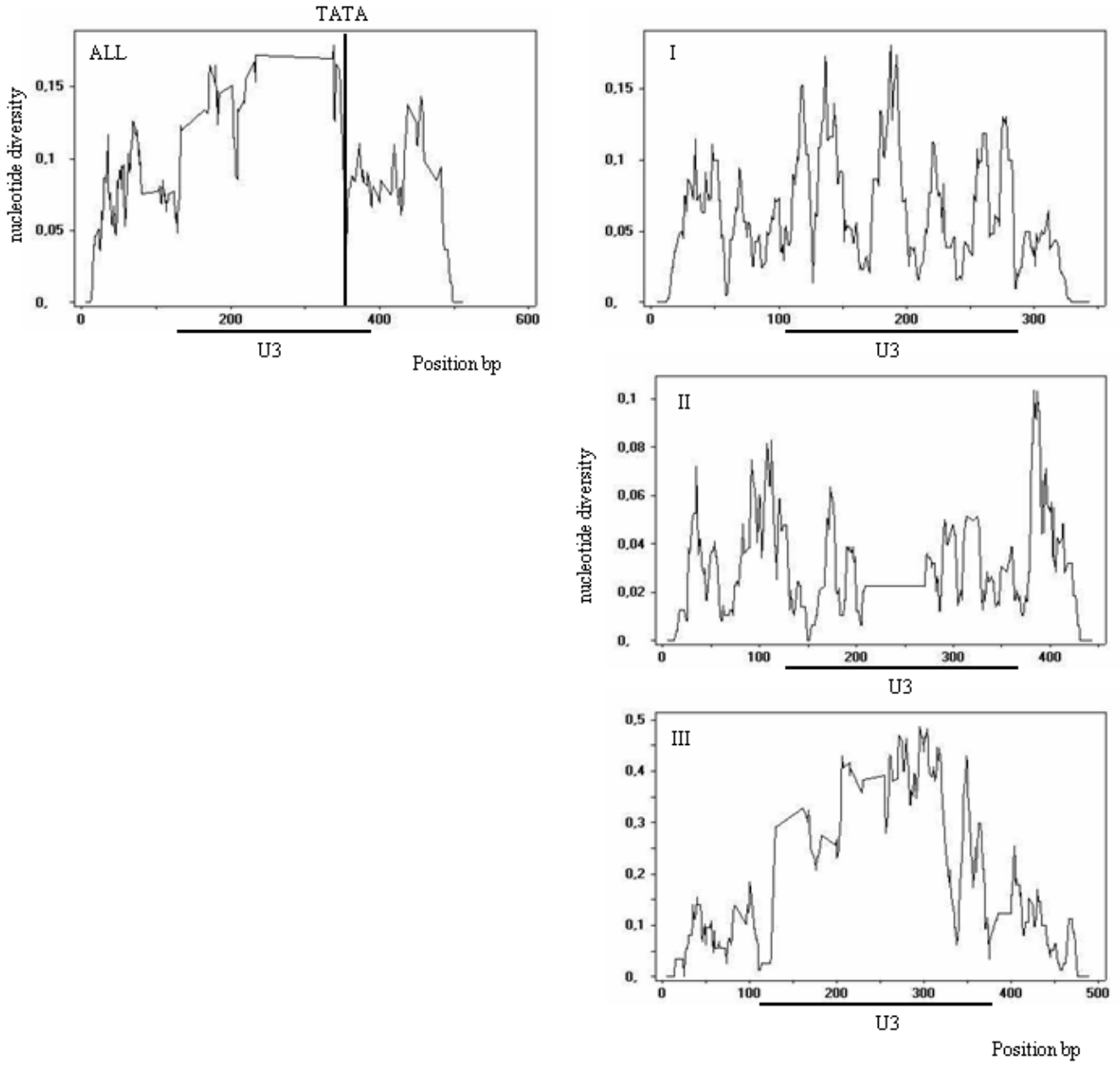

Figure 4

Sequence divergence distribution along the amplified fragments. The values on the x-axis correspond to the nucleotide position of the fragment amplified. The values on the y-axis are the nucleotide diversity (π) measure, the percentage of divergent nucleotides relative to the number of informative bases, calculated using a sliding window of 10 bp and a step of 1 bp, by the DnaSp program [31]. The position of the TATA box and U3 are indicated. Data are shown for all sequences amplified and for pair-wise comparisons within groups without gaps.

family member, Retrosol, present in several wild and cultivated *Solanum* genotypes from South America. These results suggest an early association of the Tnt1 superfamily in the evolution of this plant family (Solanaceae).

In addition to its functional importance, the LTR is one of the most rapidly evolving retrotransposon regions. LTRs contain important functional regions such as terminal segments, promoter and enhancer elements, RNA process-

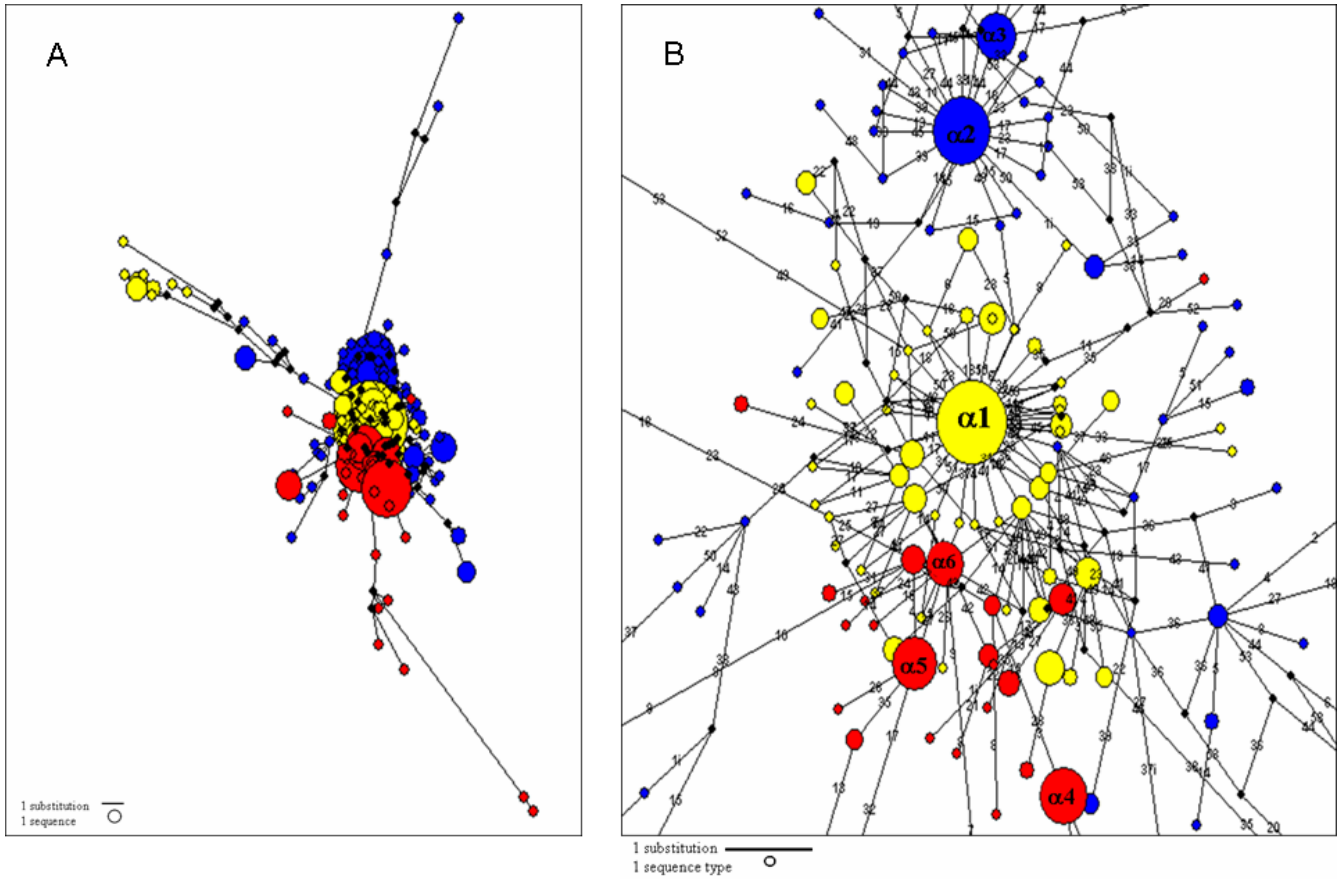

Figure 5

Median-joining networks of the RNAseH coding region. The network was constructed using 443 Tnt1 superfamily sequences. Circles denote sequence types: yellow circles denote sequences types from Retrosol, blue circles denote sequences from Tnt1, and red circles denote sequences from Retrolyc1. (A) The size of the circle is proportional to the number of sequences, with one sequence type indicated in the bottom-left corner. Lines denote substitutions; with one-step distance indicated in the bottom-left corner. Reconstructed nodes are identified as black circles; they are absent in the dataset because they were not sampled or lost. (B) Zoom of the median-joining networks of the RNAseH coding region. The size of the circles is proportional to the number of sequences, with one sequence type indicated in the bottom-left corner. Lines denote substitutions and numbers on the lines indicate the substitution position on the sequence. One-step distance is indicated in the bottom-left corner. The labels in the nodes illustrate the nomenclature used to refer to the original master sequence of each genus, called α type: α1 is the master copy from Retrosol, α2 and α3 are from Tnt1 and α4, α5 and α6 are from Retrolyc1.

ing signals, and it is recognised for the integration process. In the *BARE*-1 retrotransposon of barley, the whole LTR is heterogeneous [7], while in Tnt1 superfamily the variability is targeted specifically to the U3 region of the LTR [12]. The three subfamilies of Tnt1 retrotransposons, as well as Retrolyc1 subfamilies, are also differentiated mainly by the U3 region. Retrosol sequences present the same structure, showing an accumulation of INDELs and SNPs targeted specifically to the U3. Thus, the hyper variability of the U3 region in the Tnt1-superfamily seems to be a general phenomenon of this element's evolution within Solanaceae species.

Three clades emerged when applying phylogenetic methods to analyse amplified Retrosol sequences from *Solanum* sequences. Groups I and II are consistent because they represent cohesive versions of U3. In contrast, Group III encompasses the most diverse sequences, probably due more than one U3 version being represented. No correlation is observed between a particular U3 version and genotype, wild or cultivated species, or geographical distribution (Tables 1, 2). These results suggest that the amplification and differentiation of Retrosol occurred prior to *Solanum* speciation.

A population model was applied in this work to evaluate the relationships among Tnt1-derived elements in *Sola-*

num, *Lycopersicon* and *Nicotiana* with the aim of testing the existence of a unique master copy. To analyse the Network results, the three models of expansion proposed by Cordaux *et al* (2004) [19] for *Alu* subfamily in humans and primates were considered: the single 'master gene' model, the intermediate model and the transposon model. According to the 'master gene' model, a single α-type element generates all other subfamily members, leading to a star-like relationship with all inactive copies derived from the α element. The contrasting transposon model postulates that all element members can be active in producing new copies, resulting in the lack of a radiation structure from the α central node. The intermediate model suggests that several members are active and contribute to expansion. Relationships in this latter model are expected to be partly star-like, but also with different proportions of elements that are not connected directly with the central element depicted as the "α "-sequence.

The RNAseH network supports different models of expansion depending on the element. In addition, RNAseH sequence types, represented by nodes in the net, are distinctive for each genus. Retrosol and Tnt1 show a star-like topology, with the central α node corresponding to the most frequent sequence type, although Tnt1 presented more than one α node. Other nodes not directly connected with the central α node are also found, indicating that the sequence type immediately downstream is also capable of amplification. The unique Retrosol α node is represented in 16 out of the 25 genotypes analysed, taking into account that for most of the other 9 genotypes only a few sequences were cloned. In summary, the model of expansion of Tnt1 in *Nicotiana* sp. and Retrosol in *Solanum* sp. best suits the intermediate model. In contrast, Retrolyc1 present a model with the absence of a radiating structure from a central node and with more than two α nodes, suggesting that many family members are capable of producing new ones. However, these results cannot be used to make conclusions regarding Retrolyc1 evolution as the sequences used in this study were obtained from only a few species of *Lycopersicon*.

The α1 in Retrosol encompasses 92% of Group II sequences, representing 50% of the total Group II sample. As mentioned above, Group II presents the least variable U3 region (Figure 4), and several Group II sequences amplified from different genotypes have identical U3 regions. Thus, sequences with identical RNAseH and U3 regions transposed prior to the divergence of the host species. This activity is recent enough that element copies should not have accumulated mutations or deletions in the U3 variable region. All these lines of evidence point out the protagonist role of Group II in Retrosol expansion as the transpositionally active lineage.

RNAseH is conserved in Retrosol elements across host species from the same genus, while the U3 region has evolved rapidly, in agreement with the concept of modular evolution. The conservation of the RNAseH module at the genus level is related to its biological function, in which the enzyme has to interact with other molecules. RNAseH is involved in degradation of the original RNA template, generation of a polypurine track, and final removal of RNA primers from the newly synthesised minus and plus DNA strands [21]. These functions are necessary for the transpositional activity of the element. On the other hand, the U3 module evolved rapidly in accordance with its promoter function. New promoter sequences may permit the expression of the element in diverse environmental conditions, increasing the survival potential of the element in the host and/or the probability of overcoming host transcriptional silencing efforts. In addition, the host may gain an advantage in generating genome diversity by insertion of new retrotransposon copies, increasing its own environmental fitness. Co-evolution and co-adaptation of retrotransposon with its host genome is expected to play a particularly important role in the long-term survival of these genetic elements [22].

The Solanaceae is one of the largest flowering plant families, embracing 96 genera. It has a world-wide distribution but the greatest concentration of genera and species is found in South and Central America. About three-quarters of these genera and around half of the total number of species are found there; strongly suggesting that the family itself originated in the part of the ancient land mass that later became South America [23]. In order to analyse the evolution of the Tnt1 superfamily, species from three genera were considered: *Solanum*, *Lycopersicon* and *Nicotiana*. *Solanum* is one of the largest genera in Angiosperms and accounts for all wild and cultivated tuber-bearing species originating in the high Andes of southern Peru and northern Bolivia. *Lycopersicon* includes the tomato and its wild relatives, which are confined mainly to Chile, Peru and Ecuador. *Nicotiana* encompasses cultivated tobacco and wild species found mainly in South America and Australia. These three genera originated in South America but the present species distribution reflects several dispersal phenomena [24]. Could Tnt1-related elements be a determinant in the speciation of the Solanaceae?

As proposed by Grandbastien et al [12], the composition of present element populations results from the adaptive response of the ancestral Tnt1 population to different hosts. This adaptive element response refers to appropriate expression patterns, efficient mechanisms of replication and integration, and insertion into non-deleterious genomic sites. Ancestral copies exhibiting such adaptive responses were selectively amplified during or after the radiation of Solanaceae. On the other hand, non-tran-

scribed copies would have been rapidly inactivated and lost. In Retrosol elements, no correlation is found between host species, U3 region or RNAseH sequence type, suggesting that diversification of *Solanum* species was not a determinant of this retrotransposon population. Similar results have been reported for Tnt1 [10] and Retrolyc1 [14].

Conclusion

The use of a population model based on the network approach to evaluate the nature of the association of Tnt1-related sequences in Solanaceae supports the existence of an ancestral element rather than the occurrence of recent lateral gene transfer. The modular nature of such retrotransposons is clearly demonstrated by the finding that protein functions, represented here by RNAseH, are characterised as slowly evolving sequences, while the U3 regulatory region is under a less restrictive evolutionary force. Current populations of Tnt1, Retrolyc1 and Retrosol most likely result from their vertical transmission in Solanaceae. The molecular basis that drives U3 differentiation, and whether the rapidly evolving sequence is the result of a strategy to evade host negative control or an attempt to quickly adapt to the new evolving genome and contribute to host fitness, remains to be demonstrated.

Methods

Plant Material and DNA extraction

Solanum seeds and mini-tubercles were kindly provided by the Germplasm collection of INTA-Balcarce (Argentina). Details of species and cultivars collected from locations in Argentina and Bolivia are presented in Table 1. These localities are included in the geographical distribution of the species. *Solanum brevidens* was included in the study as the only species not producing tubers.

Fresh leaf material of commercial cultivars Kennebec and Huinkul were obtained from *in vitro* clones (INTA-Castelar-Argentina), while local Argentinian cultivars Chacarera 109, Moradita 64 and Tuni 207 were obtained from minitubers. The other *Solanum* wild species were germinated from seeds. All material was maintained and processed at Departamento de Botânica-IBUSP (Brazil).

Seeds were surface-sterilised with 10% (v/v) commercial bleach for 15 min, followed by three rinses with sterile water. Treatment with giberellic acid was performed for 24 hrs under dark conditions to break dormancy. Plants were grown in vitro in Murashige and Skoog (MS) basal medium supplemented with 20% sucrose under an 18 h photoperiod, and 22°C day/18°C night temperature regime, in a Percival growth chamber. After three months, leaves were harvested for DNA extraction. Minitubers were grown in soil under the same conditions. Total DNA was extracted as previously described [25] with an average yield of 600ng/mg of fresh tissue.

PCR amplification, cloning and sequencing

PCR amplifications were performed under standardised conditions using 300 ng of genomic DNA in a final volume of 100 μL. The reaction mix contained 0.23 μM of each primer (Avi; 5'GCTGACCAAGGTGGTAC3'; Ol16; 5'TTCCCACCTCACTACAATATCGC3'), 10 mM dNTPs, 1.5 mM MgCl2, and 5 U of *Taq* DNA polymerase (Invitrogen Brazil). Samples were amplified for 35 cycles with the following cycling conditions: denaturing at 94°C for 45s, annealing at 50°C for 45s, and extension at 72°C for 1 min. An initial denaturing step at 94°C for 5 min and a final extension step for 10 min were also performed.

PCR-amplified fragments were resolved by electrophoresis on agarose gels. Following purification (QIAquick Gel Extraction Kit-QIAGEN), fragments were cloned into the vector pGEM-T-easy (Promega) following the manufacturer's recommendations.

A total of 317 clones were sequenced in both directions using Big Dye Terminator Kit (Applied Biosystems) on an ABI 3700 sequencer (Applied Biosystems). All consensus sequences were generated with a quality Phred ≥ 20. Clone sequences were assembled with the Phred/Phrap/Consed package with a final error rate of < 1 bp/10 kb [26-28].

Sequence alignment and analysis

The identity of the sequences obtained was confirmed by a BLAST search [29] [available at the National Center for Biotechnology Information (NCBI) Bethesda, Md.]. Consensus sequences were aligned using the CLUSTAL W multiple-alignment program (version 1.5) [30] and manual adjustments of the alignments were performed when necessary.

Sequence divergence was analysed using the DnaSp program [31]. The PAUP program [32] was used to perform phylogenetic analysis based on the distance method with a nucleotide Kimura 2-parameters model. Support for groups was evaluated with 1000 bootstrap replicates. Nucleotide identity between all pairwise sequences was calculated with Strecher from EMBOSS [33].

For Network clustering, sequences were aligned using DNA Alignment 1.1.21 software and relationships were analysed using NETWORK 4.1 [34] both software available at fluxus-engineering.com [36].

The Tnt1 sequences were deposited with the GenBank nucleotide sequence database under the following accession numbers: [GenBank: AJ227998–AJ228017] for *Nico-*

tiana tabacum tab1-tab20, [GenBank: AJ228018–AJ228037] for *N. sylvestris* syl1-syl20, [GenBank: AJ228038–AJ228057] for *N. plumbaginifolia* plumb1-plumb20, [GenBank: AJ228058–AJ228077] for *N. benthamiana* ben1-ben20, [GenBank: AJ228078–AJ228097] for *N. debneyi* deb1-deb20, [GenBank: AJ228098–AJ228117] for *N. glauca* glau1-glau20, [GenBank: AJ228118–AJ228137] for *N. tomentosiformis* tom1-tom20, and for Tnt1-94, [GenBank: X13777]. Retrolyc1 sequences from *Lycopersicon peruvianum*, *L. hirsutum*, *L. pimpinellifolium* and *L. esculentum* were obtained from Araujo *et al*. 2001 [14], and from this study.

Authors' contributions

MEM carried out the molecular genetic studies, analysed and interpreted the data and drafted the manuscript. MR analysed and interpreted the data and drafted the manuscript. APPC and AMC revised the manuscript. MAVS conceived the study, participated in its design and coordination, analysed and interpreted the data, and contributed to writing the manuscript. All authors read and approved the final manuscript.

Acknowledgements

M.E. Manetti is a recipient of a CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) Ph.D. fellowship. We thank Silvia Regina Blanco Ribeiro for sequencing and Myna Nakabashi for technical assistance. Financial support for the work presented here was provided to MAVS from FAPESP (Brazil), CNPq (Brazil) and USP (Brazil).

References

1. Feschotte C, Jiang N, Wessler R: **Plant transposable elements: where genetics meets genomics.** *Nature Reviews* 2002, **3:**329-341.
2. Kumar A, Bennetzen JL: **Plant Retrotransposons.** *Annu Rev Genet* 1999, **33:**479-532.
3. Bennetzen JL: **Transposable element contributions to plant genome organization and genome evolution.** *Plant Molecular Biology* 2000, **42(1):**251-256.
4. Flavel AJ, Smith DB, Kumar A: **Extreme heterogeneity of Ty1-copia group retrotransposons in plants.** *Mol Gen Genet* 1992, **231:**233-242.
5. San Miguel P, Bennetzen JL: **Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons.** *Annals of Botany 82* 1998:37-44.
6. Kalmykova AI, Gvozdev VA: **Selective expansion of the newly evolved genomic variants of retrotransposon 1731 in the Drosophila genomes.** *Mol Biol Evo* 2004, **21(12):**2281-9.
7. Vicient CM, Kalendar R, Schulman AH: **Variability, recombination, and mosaic evolution of the barley BARE-1 retrotransposon.** *J Mol Evol* 2005, **61(3):**275-91.
8. Grandbastien MA, Spielman A, Caboche M: **Tnt1, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics.** *Nature* 1989, **26;337(6205):**376-80.
9. Casacuberta JM, Vernhettes S, Grandbastien MA: **Sequence variability within the tobacco retrotranspsoson Tnt1 population.** *EMBO J* 1995, **14:**2670-2678.
10. Vernhettes S, Grandbastien MA, Casacuberta JM: **The Evolutionary Analysis of the Tnt1 Retrotransposon in Nicotiana Species Reveals the High Variability of its Regulatory Sequences.** *Mol Biol Evol* 1998, **15(7):**827-836.
11. Hirochika H: **Activation of tobacco retrotransposons during tissue culture.** *EMBO J* 1993, **12:**2521-2528.
12. Grandbastien MA, Audeon C, Bonnivard E, Casacuberta JM, Chalhoub B, Costa APP, Le QH, Melayah D, Petit M, Poncet C, Tam SM, Van Sluys MA, Mhiri C: **Stress activation and genomic impact of Tnt1 retrotransposons in Solanaceae.** *Cytogenet Genome Res* 2005, **110(1–4):**229-41.
13. Costa APP, Scortecci K, Hashimoto RY, Araujo PG, Grandbastien MA, Van Sluys MA: **Retrolyc-1, a member of the Tnt1 retrotransposon super-family in the Lycopersicon peruvianum genome.** *Genetica* 1999, **197:**65-72.
14. Araujo PG, Casacuberts JM, Costa APP, Hashimoto RY, Grandbastien MA, Van Sluys MA: **Retrolyc1 subfamilies defined by different U3 LTR regulatory regions in the Lycopersicon genus.** *Mol Genet Genomics* 2001, **266:**35-41.
15. Kalendar R, Tanskanen J, Immonen S, Nevo E, Schulman AH: **Genome evolution of wild barley (Hordeum spontaneum) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence.** *Proc Natl Acad Sci USA* 2000, **97:**6603-6607.
16. Hedges DJ, Callinan PA, Cordaux R, Xing J, Barnes E, Batzer MA: **Differential Alu mobilization and polymorphism among the human and chimpanzee lineages.** *Genome Res* 2004, **14:**1068-1075.
17. Sen SK, Han K, Wang J, Lee J, Wang H, Callinan PA, Dyer M, Cordaux R, Liang P, Batzer MA: **Human genomic deletions mediated by recombination between Alu elements.** *Am J Hum Genet* 2006, **79(1):**41-53.
18. Posada D, Crandall KA: **Intraspecific gene genealogies: trees grafting into networks.** *Trends in Ecology & Evolution* 2001, **16(1):**37-45.
19. Cordaux R, Hedges DJ, Batzer MA: **Retrotransposition of Alu elements: how many sources?** *TRENDS in genetics* 2004, **20(10):**464-467.
20. Rossi M, Araujo PG, de Jesus EM, Varani AM, Van Sluys MA: **Comparative analysis of Mutator-like transposases in sugarcane.** *Mol Genet Genomics* 2004, **272:**194-203.
21. Malik HS, Eickbush TH: **Phylogenetic Analysis of Ribonuclease H Domains Suggest a Late, Chimeric Origin of LTR Retrotransposable Elements and Retroviruses.** *Genome Research* 2001, **11:**1187-1197.
22. Kidwell MG, Lisch DR: **Transposable elements and host genome evolution.** *Tree* 2000, **15:**95-99.
23. Hawkes JD: **The economic importance of the family Solanacea.** In *Solanaceae IV* Edited by: Nee M, Symon DE, Lester RN, Jessop JP. Royal Botanic Gardens, Kew; 1999:1-8.
24. D'arcy W: **The Solanaceae since 1976, with a Review of its Biogreography. Solanaceae III:Taxonomy, Chemestry, Evolution.** Edited by: Hawkes, Lester, Nee, estrada. Royal Botanic Gardens Kew and Linnean Society of London; 1991.
25. Doyle JJ, Doyle JL: **A rapid isolation procedure for small quantities of fresh leaf tissue.** *Phytochemical Bulletin* 1987, **19:**11-15.
26. Gordon DP, Abajian C, Green P: **Consed: a graphical tool for sequence finishing.** *Genome Research* 1998, **8:**195-202.
27. Gordon D, Desmarais C, Green P: **Automated finishing with Autofinish.** *Genome Res* 2001, **11:**614-625.
28. Ewing B, Hillier L, Wedl M, Green P: **Basecalling of automated sequencer traces using phred. I. Accuracy assessment.** *Genome Research* 1998, **8:**175-185.
29. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215:**403-410.
30. Higgins D, Thompson J, Gibson T, Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting position-specific gap penalties and weight matrix choice.** *Gene* 1988, **73:**237-244.
31. Rozas J, Sanchez del Barrio JC, Messegyer X, Rozas R: **DnaSP, DNA polymorphism analyses by the coalescent and other methods.** *Bioinformatics* 2003, **19:**2496-2497.
32. Swofford DL: **PAUP: phylogenetic analysis using parsimony (Version 3.0).** *Illinios Nat. History Survey, Champaign, Ill* 1993.
33. Rice P, Longden I, Bleasby A: **EMBOSS: The European Molecular Biology Open Software Suite.** *Trends in Genetics* 2000, **16(6):**276-277.
34. Bandelt HJ: **Median-Joining networks for inferring intraespecific phylogenies.** *Mol Biol Evol* 1999, **16:**37-48.
35. **NCBI Blast** [http://www.ncbi.nlm.nih.gov/BLAST/]
36. **Fluxus-engineering.com** [http://www.fluxus-technology.com/index.htm]