

Universidade de Brasília
Pós-Graduação em Biologia Molecular

História evolutiva de um Retrotransposon-LTR nos dois genomas componentes
do amendoim

Fernando Campos de Assis Fonseca

Brasília
2007

Universidade de Brasília
Pós-Graduação em Biologia Molecular

História evolutiva de um Retrotransposon-LTR nos dois genomas componentes
do amendoim

Fernando Campos de Assis Fonseca

Orientador: Renato de Oliveira Resende

Co-orientador: David John Bertioli

Dissertação apresentada ao Departamento de
Biologia Celular do Instituto de Ciências
Biológicas da Universidade de Brasília, como
requisito parcial à obtenção do grau de mestre em
Biologia Molecular.

Brasília

2007

Universidade de Brasília
Pós-Graduação em Biologia Molecular

História evolutiva de um Retrotransposon-LTR nos dois genomas componentes
do amendoim

Banca Examinadora

Dr. Márcio de Carvalho Moretzsohn

Prof. Dr. Rinaldo Wellerson Pereira

Dr^a. Patrícia Messenberg Guimarães (suplente)

Brasília
2007

“Num dia excessivamente nítido,
Dia em que dava a vontade de ter trabalhado muito
Para nele não trabalhar nada,
Entrevi, como uma estrada por entre as árvores,
O que talvez seja o Grande Segredo,
Aquele Grande Mistério de que os poetas falsos falam.

Vi que não há Natureza,
Que a Natureza não existe,
Que há montes, vales, planícies,
Que há árvores, flores, ervas,
Que há rios e pedras,
Mas que não há um todo a que isso pertença,
Que um conjunto real e verdadeiro
É uma doença das nossas idéias.”

(Fernando Pessoa/ Alberto Caeiro)

Dedico este trabalho aos meus pais, Vanessa e Sérgio, por todas as suas contribuições na minha formação pessoal, social e cultural que me permitiram, mesmo nas horas mais difíceis discernir entre certo e errado e escolher o melhor caminho a ser seguido. Por serem tão pacientes e presentes em momentos de extrema dificuldade.

Aos meus irmãos, Isabel e Leonardo, por todos os momentos vividos, desde as brigas mais intensas às brincadeiras mais ternas, que sempre ajudaram a “sair” um pouco do mundo e relaxar, mesmo que por um breve período.

A todos os meus amigos: os de “antigamente” e aqueles que estão sempre surgindo. Com carinho e bom humor, eles souberam dar conselhos valiosos, me ajudando a enfrentar vários desafios. Por todas as horas sofridas que passamos e pelo sucesso conquistado juntos.

À minha namorada Bruna, por ser sempre compreensível, amorosa e acreditar na minha capacidade mais do que eu mesmo.

Àqueles que acreditavam que ao final deste processo eu teria desenvolvido um super amendoim que resolveria todos os problemas, gastronômicos, amorosos e etc.

Agradecimentos

À Embrapa Recursos Genéticos e Biotecnologia, pela oportunidade de estágio que propiciou a minha entrada no universo da pesquisa, pelos recursos oferecidos e pela experiência adquirida.

Ao Departamento de Biologia Celular da Universidade de Brasília, pela oportunidade de fazer a pós-graduação. Aos professores do programa, pelos ensinamentos e à secretária Ana Tiberti, por se fazer tão prestativa em todos os momentos.

Ao professor Dr. Renato de Oliveira Resende, pela orientação, participação da banca examinadora e toda a ajuda na resolução de problemas burocráticos.

Ao professor Dr. David John Bertoli, pela co-orientação, pelo inestimável conhecimento transmitido e confiança no meu trabalho.

Ao Dr. Stephan Nielen, pelos ensinamentos, treinamento e companheirismo no laboratório, além da incomensurável ajuda na análise de vários dados e revisão do trabalho.

À Dra. Soraya Leal-Bertoli, pela ajuda na realização deste trabalho, sempre fazendo críticas e dando sugestões construtivas.

Aos Drs. Rinaldo Pereira, Márcio Moretzsohn e Patrícia Guimarães, por aceitarem participar, com grande disposição da banca examinadora e por ajudarem na elaboração de um trabalho de qualidade.

A todos do laboratório LPPIII, pela camaradagem, trocas de experiências e apoio.

Ao Leandro Mesquita, pelo cuidado com as plantas nas casas de vegetação.

Aos que, diretamente ou indiretamente, contribuíram para a realização deste projeto.

À CAPES e Comunidade Européia, pelo apoio financeiro muito bem-vindo.

A todos os supracitados,

Obrigado!

Sumário

Lista de Figuras	x
Lista de Tabelas	xiii
Abreviaturas	xiv
Resumo	1
Abstract	3
1. Introdução	4
1.01 O gênero <i>Arachis</i>	4
1.02 Anfidiplóides	5
1.03 Sintenia	6
1.04 Seqüências repetitivas	7
1.05 DNA móvel	8
1.06 Controle de atividade	12
1.07 Diversidade e utilização	14
1.08 Caracterização do retroelemento	15
2. Hipótese	17
3. Objetivos	18
3.01 Objetivos gerais	18
3.02 Objetivos específicos	18
4. Material e Métodos	19
4.01 Material Vegetal	19
4.02 Desenho de iniciadores	20
4.03 Reação em cadeia da polimerase	20
4.04 Ligação e Transformação	21
4.05 Preparação de DNA plasmidial	21
4.06 Seqüenciamento e análise de seqüências	22
4.07 Alinhamento e análise filogenética	22
4.08 Análise do número de cópias	23
4.09 Análise de Metilação	27
4.10 Extensão da seqüência	28
4.11 Identificação de clones em biblioteca BAC	28
5. Resultados	30
5.01 Extração de DNA vegetal	30
5.02 Reação em cadeia da polimerase	30
5.03 Análise Filogenética	33
5.04 Número de cópias	43
5.05 Análise de Metilação	47
5.06 Extensão da seqüência	49

6. Discussão	52
6.01 Distribuição do elemento nos genomas	53
6.02 Análise Filogenética	55
6.03 Número de cópias	59
6.04 Análise de metilação.....	60
6.05 Extensão da seqüência	61
7. Conclusões	63
8. Perspectivas	65
9. Referências Bibliográficas	66
Apêndice	74
ANEXO I	76
ANEXO II	78

Lista de Figuras

Figura 1. Mapa demonstrativo dos diversos tipos de sintenia utilizando-se diferentes marcadores moleculares (A-P). À esquerda, sintenia completa na ordem e posição dos cromossomos de espécies diferentes. Ao centro, comparação do mesmo mapa com cromossomos de várias outras espécies. À direita, comparação entre dois cromossomos de espécies tetraplóides (1 e 2) com o cromossomo da espécie diplóide, (Schmidt, 2000)..... 7

Figura 2. Ciclo de vida hipotético do retrotransposon LTR. (a) transcrição da fita de mRNA. (b) tradução e síntese de proteínas das regiões do GAG e POL; POL inclui as proteínas AP (protease), RT (transcriptase reversa), RNaseH e INT (integrase). (c) formação do dímero de RNA. (d) empacotamento do RNA e das proteínas RT, RNaseH e INT e início da transcrição reversa. (e) degradação da matrix de RNA e síntese da segunda fita de cDNA. (f) término da fita dupla de cDNA e ligação de INT junto à seqüência LTR. (g) integração da dupla fita em um novo sítio do genoma do hospedeiro. A linha cinza mais espessa ilustra a membrana nuclear. Adaptado: Sabot e Schulman, 2006. 10

Figura 3. Diferentes grupos de retrotransposons LTR e Não-LTR, mostrando suas seqüências terminais e posições das seqüências codificadoras. Adaptado: Kumar e Bennetzen, 1999. 11

Figura 4. Padrões de distribuição de seqüências repetitivas e de genes ao longo de um cromossomo vegetal. Dependendo da espécie e do tipo de retrotransposon a distribuição muda consideravelmente. Adaptado: Schmidt e Heslop-Harrison, 1998 e Kumar e Bennetzen, 1999. 12

Figura 5. Ilustração dos dois modelos hipotéticos a serem testados neste trabalho. 17

Figura 6. Posição e orientação dos iniciadores desenhados para amplificação do gene da enzima transcriptase reversa. Obs: os tamanhos dos iniciadores não estão em escala..... 31

Figura 7. A) Amplificação da seqüência gênica da enzima transcriptase reversa em *A. hypogaea*. 1 (iniciadores F1 x R1), 2 (F1 x R2), 3 (F1 x R3), 4 (F2 x R1), 5 (F2 x R2), 6 (F2 x R3), 7 (F3 x R1), 8 (F2 x R2), 9 (F3 x R3). B) Digestão dos plasmídios mostrando os diferentes tamanhos de insertos obtidos. Tatu (*A. Hypogaea*), D6 (*A. duranensis*), Ipa (*A. ipaënsis*), M (1Kb DNA ladder, Invitrogen), LM (Low DNA Mass ladder, Invitrogen)..... 32

Figura 8. A) Mapa de todos os 87 contigs construídos mostrando as posições de cada par de iniciadores. Cada linha entre duas barras verticais corresponde a um contig. B) Exemplo de alinhamento de seqüências de um contig. Marcada em azul está a seqüência do iniciador Rep-RT-F3. 32

Figura 9. Relações filogenéticas obtidas a partir das seqüências de nucleotídeos do gene de RT. Os resultados de bootstrap são indicados para valores acima de 50 (1000 réplicas). *M. truncatula* foi utilizado como grupo externo..... 35

Figura 10. Relações filogenéticas obtidas a partir das seqüências de aminoácidos do gene de RT. Os resultados de bootstrap são indicados para valores acima de 50 (1000 réplicas). *M. truncatula* foi utilizado como grupo externo..... 37

Figura 11. Distribuição dos valores de dS apresentados pelas 80 seqüências dos contigs em relação à seqüência consenso. A) Valores obtidos para <i>A. duranensis</i> , B) Valores de <i>A. hypogaea</i> e C) Valores de <i>A. ipaënsis</i> . A linha de tendência foi calculada utilizando-se a função polinomial.	39
Figura 12. Distribuição dos valores de dS + dN apresentados pelas seqüências dos contigs em relação à seqüência consenso. A) Valores obtidos para <i>A. duranensis</i> , B) Valores de <i>A. hypogaea</i> e C) Valores de <i>A. ipaënsis</i> . A linha de tendência foi calculada utilizando-se a função polinomial.	40
Figura 13. Relações filogenéticas entre diversos grupos de elementos do tipo Athila. Resultados de bootstrap estão indicados para valores acima de 50 (1000 replicações).	42
Figura 14. Resultados das hibridizações de DNA genômico utilizando o clone Ah-Athila-9 contendo o gene de RT em <i>A. hypogaea</i> . A) Dot blot (Linhas 1 e 3: DNA genômico, linhas 2 e 4: Controle), B) Curva de calibragem.	44
Figura 15. Resultados das hibridizações de DNA genômico utilizando o clone Ah-Athila-9 contendo o gene de RT em <i>A. duranensis</i> . A) Dot blot (azul=DNA genômico, verde=Controle), B) Curva de calibragem e C) Curva de DNA genômico.	45
Figura 16. Resultados das hibridizações de DNA genômico utilizando o clone Ah-Athila-9 contendo o gene de RT em <i>A. ipaënsis</i> . A) Dot blot (azul=DNA genômico, verde=Controle), B) Curva de calibragem e C) Curva de DNA genômico.	46
Figura 17. Gel de agarose 1% mostrando o padrão de digestão dos DNAs genômicos. M: 1Kb Plus ladder, 1: <i>A. hypogaea</i> (<i>Mbo</i> I), 1: <i>A. hypogaea</i> (<i>Sau</i> 3AI), 3: <i>A. duranensis</i> (<i>Mbo</i> I), 4: <i>A. duranensis</i> (<i>Sau</i> 3AI), 5: <i>A. ipaënsis</i> (<i>Mbo</i> I), 6: <i>A. ipaënsis</i> (<i>Sau</i> 3AI).	48
Figura 18. Análise de metilação através de Southern blot. Tatu (<i>A. hypogaea</i>), D6 (<i>A. duranensis</i>) e Ipa (<i>A. ipaënsis</i>). Foi utilizado como sonda o clone Ah-Athila-9.	49
Figura 19. A) Resultado da amplificação para o genome walker, 1: <i>Pvu</i> II + Rep-B1-OUT-L, 2: <i>Pvu</i> II + Rep-B1-OUT-R, 3: <i>Dra</i> I +Rep-B1-OUT-L, 4: <i>Dra</i> I + Rep-B1-OUT-R. B)Exemplo de análise da presença e tamanho do inserto de clones obtidos de <i>A. hypogaea</i> para a extensão da seqüência do retroelemento. M (1kb ladder), P (Low mass ladder).	50
Figura 20. A) Seqüência traduzida do retroelemento, após extensão utilizando o Kit Genome Walker, em negrito estão os motivos conservados de cada gene. B) Tamanho e composição de um retrotransposon do tipo Athila de <i>A. thaliana</i> , adaptado: Hansen e Heslop-Harrison, 2004.	51
Figura 21. À esquerda, hibridização com DAPI de uma célula contendo 40 cromossomos de <i>A. hypogaea</i> . Notar a presença de 20 cromossomos com banda centromérica (genoma AA) e 20 com ausência de qualquer banda (genoma BB). À direita, hibridização da mesma célula com o elemento repetitivo em verde. Cortesia: Seijo <i>et al.</i> , 2005 (dados não publicados).	53
Figura 22. Padrão de distribuição do elemento Athila, em vermelho, nos cromossomos de <i>M. truncatula</i>	54

- Figura 23.** Colinearidade, em vermelho, entre os cromossomos de *M. truncatula* e *L. japonicus*. Adaptado: Cannon *et al.*, 2006. 55
- Figura 24.** Dot blot dos clones obtidos após subclonagem de fragmentos oriundos de digestão de DNA genômico de *A. hypogaea* com *HindIII* (vermelho). Como sonda foi utilizado DNA genômico do próprio amendoim. Os clones com maior sinal de hibridização foram escolhidos para seqüenciamento automático. Em azul estão os clones E12 e F05 obtidos em trabalhos anteriores. 74
- Figura 25.** Fragmentos polimórficos de DNA obtidos por AFLP utilizando-se iniciadores desenhados exclusivamente para amplificação da seqüência do clone E12. As amostras foram amplificadas no genoma de *A. duranensis*, *A. stenosperma* (genoma AA) e da planta híbrida entre elas. As bandas marcadas com as setas foram escolhidas para seqüenciamento automático. Após o seqüenciamento dos clones e das bandas polimórficas, as seqüências alinharam-se para comporem um contig (consenso) apenas. 74
- Figura 26.** Southern blot para análise de metilação em *A. ipaënsis* (esquerda) e no Anfidiplóide sintético (direita), utilizando as enzimas *MboI* e *Sau3AI* e como sonda a seqüência de E12 que agora acredita-se fazer parte do retrotransposon utilizado nesta dissertação. Nenhum sinal de metilação foi encontrado..... 75
- Figura 27.** Distribuição da seqüência do clone E12 no genoma de *A. hypogaea*. Em azul coloração com DAPI. Em verde a hibridização da mesma lâmina utilizando o clone repetitivo com sonda. O padrão obtido é muito semelhante ao do retrotransposon caracterizado neste trabalho. Foto: Stephan Nielen, 2005 (dados não publicados). 75

Lista de Tabelas

- Tabela 1.** Denominação e seqüência dos iniciadores específicos desenhados para amplificação do gene codificador da enzima transcriptase reversa em *Arachis* sp..... 20
- Tabela 2.** Quantidade utilizada de DNA e número de cópias esperado do plasmídio contendo o clone Ah-Athila-9 para o genoma das três espécies de *Arachis*..... 26
- Tabela 3.** Combinação de iniciadores para extensão do retroelemento de amostras de DNA genômico de *A. hypogaea*, digeridas com as enzimas *DraI* e *PvuII* e os iniciadores desenhados para a seqüência de RT. 28
- Tabela 4.** Distância genética entre as seqüências de DNA e de aminoácidos. Ah (*A. hypogaea*), Ad (*A. duranensis*) e Ai (*A. ipaënsis*). O valor médio corresponde a todas as 80 seqüências de DNA e/ou aminoácidos. 33
- Tabela 5.** Distância genética entre as seqüências de nucleotídeos de cada grupo obtido na árvore filogenética e o genoma relacionado. O valor médio foi obtido para todas as seqüências dentro de cada grupo. N/A indica a ausência de seqüências ou a presença de apenas uma seqüência. 34
- Tabela 6.** Distância genética entre as seqüências de aminoácidos de cada grupo obtido na árvore filogenética e o genoma relacionado. O valor médio foi obtido para todas as seqüências dentro de cada grupo. N/A indica a ausência de seqüências ou a presença de apenas uma seqüência. 36
- Tabela 7.** Taxas de substituições sinônimas (dS) e não-sinônimas (dN) dentro e entre os grupos. 38

Abreviaturas

AFLP - Amplified Fragment Length Polymorphism

BAC - Bacterial Artificial Chromosome

BSA - Bovine Serum Albumin

1C – Genoma haplóide

2C – Genoma diplóide

°C – Graus Celsius

cDNA – DNA complementar

CTAB - Cetyl-trimethylammoniumbromide

DAPI - 4',6-diamidino-2-phenylindole

d-CTP - 2'-deoxycytidine 5'-triphosphate

dN – Substituições não sinônimas

DNA – Ácido desoxirribonucleico

dNTP - Deoxyribonucleoside Triphosphates

DraI - *Deinococcus radiophilus* I

dS – Substituições sinônimas

EcoRI - *Escherichia coli* RI

EDTA – Ethylene DiamineTetrAcetic acid

EST - Expressed Sequence Tag

GAG - Group Specific Antigen

HindIII - *Haemophilus influenzae* Rd

IPTG - Isopropyl β-D-1-thiogalactopyranoside

IRAP - Inter-Retrotransposon Amplified Polymorphism

Kb – Kilobase

LB – Meio de cultura Luria-Bertani

LINE - Long Interspersed Repetitive Elements

LTR – Long Terminal Repeat

M – Molar

MboI - *Moraxella bovis* I

mg – Miligrama

MgCl₂ – Cloreto de Magnésio

mL – Mililitro

mM – Milimolar
Mpb – Mega pares de bases
mRNA – RNA Mensageiro
 μ FD – Microfaraday
 μ g – Micrograma
 μ L – Microlitro
 μ M - Micromolar
NaCl - Cloreto de Sódio
NaOH – Hidróxido de Sódio
ng – Nanograma
 32 P – Fósforo 32
pb – Pares de bases
PCR - Polymerase Chain Reaction
pH – Potencial Hidrogeniônico
PvuII - *Proteus vulgaris* II
R² - Coeficiente de correlação
RAPD - Random Amplified Polymorphic DNA
RBIP - Retrotransposon-based Insertion Polymorphism
REMAP - Retrotransposon-Microsatellite Amplified Polymorphism
RFLP - Restriction Fragment Length Polymorphism
RNA - Ácido Ribonucleico
RNase – Ribonuclease
rpm – Rotações por minuto
RT – Transcriptase reversa
Sau3AI - *Staphylococcus aureus* 3A
SDS - Sodium Dodecyl Sulfate
SINE - Short Interspersed Repetitive Elements
SSAP - Sequence Specific Amplified Polymorphism
SSC - Sodium Salt Citrate
SSR - Simple Sequence Repeats
Taq - *Thermus aquaticus*
U – Unidade
V – Volts
X-GAL - 5-bromo-4-chloro-3-indolyl-b-D-galactopyranoside

Resumo

O amendoim cultivado (*Arachis hypogaea*) é uma cultura de grande importância econômica, sendo cultivado em vários países do mundo. As safras na agricultura são reduzidas por estresses bióticos e abióticos aos quais as espécies silvestres são resistentes. Visando a obtenção de plantas mais resistentes, foi desenvolvida uma estratégia onde as espécies silvestres seriam as doadoras de genes de resistência que seriam transferidos através de cruzamentos interespecíficos. O primeiro passo seria a construção de mapas genéticos que identifiquem as distâncias e as posições dos genes de interesse no genoma de cada espécie. O segundo passo seria a correção da diferença de ploidia entre as espécies, já que a cultivada é uma alotetraplóide ($2n=4x=40$) com genoma AABB e as silvestres diplóides ($2n=2x=20$) com genomas AA ou BB. Porém, problemas de incompatibilidade entre os cruzamentos são comuns. Isso se deve muitas vezes por causa de diferenças nas composições das regiões não-codantes e de repetições, o que faz necessário um estudo detalhado dessas regiões. Dentre as seqüências repetitivas de maior importância estão os retrotransposons, elementos que se multiplicam no genoma e inserem uma nova fita de DNA em diferentes sítios. Neste trabalho, foram isoladas e caracterizadas seqüências de um retrotransposon LTR do amendoim cultivado e de seus parentais silvestres (*A. duranensis* e *A. ipaënsis*). A seqüência codificadora da enzima transcriptase reversa foi amplificada através de PCR utilizando combinações de iniciadores específicos. O Southern blot não mostrou nenhum sinal de metilação na seqüência desse elemento, mas através de sondagem de 20 mil seqüências de ESTs nenhuma similaridade foi encontrada, sugerindo que esse elemento não é expresso. Análises filogenéticas e estatísticas foram realizadas utilizando-se 87 contigs, onde seqüências nucleotídicas e de aminoácidos formaram grupos específicos nas árvores de similaridade, indicando que o retrotransposon evoluiu diferentemente nos genomas das espécies parentais. De acordo com o cálculo do número de substituições sinônimas na seqüência desse elemento, observou-se que ele é mais variável no genoma BB. Já no genoma AA de *A. duranensis* ele parece ter sofrido modificações recentes, provavelmente decorridas da transposição desse elemento há cerca de 2-6 milhões de anos, fato que não pode ser observado no genoma de *A. ipaënsis*. O número de cópias calculado por genoma diplóide nas três espécies foi, em média, 800 em *A. ipaënsis*, 3 mil em *A. duranensis* e 5 mil em *A. hypogaea*. Esses resultados mostram que o retrotransposon é muito antigo e provavelmente fazia parte do genoma do ancestral comum à diferentes grupos vegetais. Após a diferenciação nas espécies parentais o

retrotransposon provavelmente multiplicou seu número de cópias no genoma AA e permaneceu praticamente inalterado no genoma BB. Apesar de ser encontrado em regiões eucromáticas e de não estar metilado ele se mantém inativo.

Palavras-chave: *Arachis*, Retrotransposon, Número de cópias, Substituições sinônimas.

Abstract

The cultivated peanut (*Arachis hypogaea*) is a crop of economic importance and it is cultivated in several countries around the world. Yields in agriculture are reduced by biotic and abiotic stresses to which most the wild species are resistant. With the aim of developing more resistant plants, a strategy of crossing where wild species are the donors of resistance genes to the cultivated peanut is used. The first step is to develop a linkage map to identify the distances and positions of genes of interest for each species. The second step is the correction of the number of chromosomes since the cultivated peanut is an allotetraploid ($2n=4x=40$) with an AABB genome and the wild species are diploid ($2n=2x=20$) with AA or BB genomes. However, problems of incompatibility during crosses are common. One possible cause for incompatibility is the difference in the intergenic noncoding regions of different species. For this and other reasons, the study of repetitive DNA is interesting. Among the repetitive sequences, retrotransposable elements are found to be of major importance because they can copy of themselves into new chromosome sites. In this work we reported the isolation and characterization of an LTR retrotransposon from the cultivated peanut and its wild parental genomes (*A. duranensis* and *A. ipaënsis*). The coding region of the reverse transcriptase gene was amplified by PCR using specific primer combinations derived from genomic sequences of the *Arachis* data base. No methylation was found using Southern blot but also no significant homology to 20.000 ESTs sequences on the GenBank, suggesting that this element is not expressed. Phylogenetic and statistical analyses were done with 87 contigs. Both the nucleotide and aminoacid sequences formed specific groups indicating that the retroelement evolved differently in the wild species genomes. According to the synonymous substitutions content it was observed that the retrotransposon has accumulated more mutations in BB genome than in the *A. duranensis* genome, probably because of a recent amplification that occurred about 2-6 myr ago. The copy number calculated per diploid genome for the three species was approximately: 800 for *A. ipaënsis*, 3.000 for *A. duranensis* and 5.000 for *A. hypogaea*. These results show that the element is ancient and it was a part of the genome of a species ancestral to many different groups of plants. After the differentiation between the two parental species, this element probably multiplied its copy number in the AA genome and remained almost silent in the BB genome. Although the element is dispersed in euchromatic regions and does not present any methylation no sign of activity was found.

Keywords: *Arachis*, Retrotransposon, Copy number, Synonymous substitutions.

1. Introdução

1.01 O gênero *Arachis*

O amendoim (*Arachis hypogaea* L.) é uma cultura de grande importância agrônômica em todo o mundo, pois pode ser utilizado tanto como forragem para a alimentação do gado quanto para o consumo humano, como fonte de proteínas e óleo (Godoy *et al.*, 1999). Desde a época dos povos pré-colombianos esta leguminosa já era cultivada, tendo se disseminado pela América do Sul por meio das antigas rotas comerciais e, pelo restante do mundo, por meio das grandes viagens de portugueses e espanhóis (Stalker, 1997; Simpson *et al.*, 2001). Hoje, o amendoim é cultivado em mais de 80 países, na Ásia, na África e nas Américas (Singh U e Singh B, 1991 citado por: Moretzsohn *et al.*, 2004), tornando-se a terceira leguminosa mais cultivada no mundo (Duke, 1981 citado por: Seijo *et al.*, 2004).

Acredita-se que o gênero *Arachis* tenha surgido primeiramente na região correspondente ao centro oeste brasileiro e nordeste paraguaio (Simpson *et al.*, 2001). Evidências geográficas, estudos de dispersão, testes de cruzabilidade e análises morfológicas levam à divisão do gênero em nove seções taxonômicas, dentre elas a seção *Arachis*, que compreende o amendoim e outras 30 espécies silvestres que cruzam com *A. hypogaea* (Krapovickas e Gregory, 1994; Valls e Simpson, 2005). Estudos similares evidenciaram que a espécie *A. hypogaea* surgiu entre o norte da Argentina e o sul da Bolívia (Krapovickas e Gregory, 1994).

A produção nacional de amendoim está muito aquém da sua capacidade. Segundo o Instituto Brasileiro de Geografia e Estatística (IBGE), em 2005 a produção de amendoim em casca no Brasil foi de aproximadamente 314906,00 toneladas, o que gerou uma renda de quase 290 milhões de reais. O estado com maior produção foi São Paulo, com cerca de 70% de toda a produção nacional. Em comparação, a produção de soja em grão durante o mesmo período superou 50 milhões de toneladas (Fonte: www.ibge.gov.br – ver referências).

Análises citogenéticas mostram que a seção *Arachis* apresenta 29 espécies diplóides ($2n = 2x = 20$) e duas espécies tetraplóides ($2n = 4x = 40$), incluindo o próprio amendoim (Krapovickas e Gregory, 1994; Stalker, 1997; Valls e Simpson, 2005). Além disso, existem diferenças no tamanho, forma e arranjos dos cromossomos que permitem identificar três genomas distintos: A, B e D, sendo que o amendoim apresenta um cariótipo AABB (Stalker, 1991; Krapovickas e Gregory, 1994). Acredita-se que o genoma tetraplóide do amendoim

tenha surgido a partir do cruzamento de duas espécies de genomas distintos AA e BB (*A. duranensis* e *A. ipaënsis*, respectivamente) (Kochert *et al.*, 1991;1996; Seijo *et al.*, 2004) e que o híbrido estéril tenha sofrido duplicação espontânea dos seus cromossomos, recuperando assim sua fertilidade (Halward *et al.*, 1991; Young *et al.*, 1996).

Atualmente existem seis variedades descritas de amendoim, classificadas com base em características morfológicas e reprodutivas. Apesar dessa diversidade, a variabilidade genética entre elas é extremamente baixa (Kochert *et al.*, 1991; Moretzsohn, 2006; Moretzsohn *et al.*, 2004; Stalker *et al.*, 1995), tornando-a muito susceptível a doenças causadas por fungos, nematóides e vírus assim como a fatores abióticos. Técnicas como AFLP, RAPD, RFLP e SSR buscam constantemente encontrar marcadores polimórficos em *A. hypogaea*, mas encontram dificuldades (Kochert *et al.*, 1996; Moretzsohn *et al.*, 2004; Stalker, 1997).

Apesar dos empecilhos encontrados ao trabalhar-se com o amendoim cultivado, o mesmo não acontece com os representantes silvestres e diplóides da seção *Arachis* (Krapovickas e Gregory, 1994). Muitas dessas espécies apresentam algum tipo de resistência ou uma menor susceptibilidade a doenças em relação a *A. hypogaea*, tornando interessante o mapeamento dos genes responsáveis por essas características com o intuito de transferi-los para o amendoim cultivado. Um grande problema a ser enfrentado é a diferença de ploidia entre as espécies, o que leva à necessidade de realizar cruzamentos complexos.

1.02 Anfidiplóides

De acordo com Simpson (2001), existem três formas principais de introgressão de genes de interesse para o amendoim cultivado. A primeira consiste no cruzamento de uma espécie silvestre diplóide ($2n = 2x = 20$) com *A. hypogaea*, gerando um híbrido triplóide ($2n = 3x = 30$) e estéril. Esse híbrido sofre tratamento com colchicina para duplicar o seu número de cromossomos ($2n = 6x = 60$), tornando-o fértil. Então retrocruzamentos com *A. hypogaea* são realizados até que a progênie tenha o número de cromossomos reduzido para 40. Uma dificuldade característica desta técnica é que ela funciona apenas com algumas espécies de *Arachis*, enquanto que em outras tentativas de cruzamentos há um índice de esterilidade muito alto, com ausência de florescimento e frutificação (Company *et al.*, 1982 citado por: Fávero, 2004). A segunda maneira seria o tratamento de duas espécies silvestres (uma de genoma AA e outra de genoma BB) com colchicina gerando indivíduos tetraplóides. Estes são cruzados entre si e o híbrido é então utilizado em retrocruzamentos com *A. hypogaea*. Essa técnica não tem se mostrado muito eficaz, pois após o primeiro retrocruzamento obtém-se um índice de

esterilidade altíssimo. A terceira técnica consiste no cruzamento de duas espécies de genomas diferentes AA e BB, gerando um híbrido AB estéril. Trata-se essa planta com colchicina para duplicação de cromossomos e ela é então cruzada com *A. hypogaea* para seleção de características desejáveis. Essa é a técnica que tem fornecido melhores resultados, servindo também como apoio para a hipótese da origem do amendoim cultivado. Acredita-se que, com a presença dos genomas A e B, os eventos de pareamento de cromossomos durante a meiose ocorram de uma maneira que seja a mais regular e mais estável possível (Fávero, 2004), o que favorece a introgressão de genes de interesse para a cultivar. Essa metodologia foi utilizada para a produção de duas cultivares, COAN (Simpson e Starr, 2001) e NEMATAN (Starr *et al.*, 2002; Simpson *et al.*, 2003) resistentes a nematóides, sendo esses os únicos exemplos de introgressão de genes silvestres para o amendoim cultivado até agora conhecidos.

Anfidiplóides sintéticos têm sido produzidos na Embrapa Recursos Genéticos e Biotecnologia (em colaboração com Simpson) sendo que até agora dois cruzamentos, (*A. ipaënsis* KG30076 x *A. duranensis* V14167) e (*A. gregoryii* V6389 x *A. linearifolia* V9401), permitiram obter híbridos férteis com o amendoim (Fávero *et al.*, 2006). Esses anfidiplóides junto com o mapa genético oferecem a perspectiva de introgressão controlada de genes silvestres para o amendoim.

1.03 Sintenia

Para que esta técnica possa ser utilizada como uma das principais vias de transferência de genes de interesse para o amendoim, é necessário o conhecimento de quais genomas silvestres apresentam maior compatibilidade com o genoma AA e com o genoma BB de *A. hypogaea*. Quando os genomas de duas espécies são comparados entre si, é possível determinar seqüências extremamente conservadas dentro de um mesmo grupo de espécies e até entre famílias diferentes. Espécies que apresentam um alto grau de sintenia ou colinearidade podem apresentar não somente uma grande similaridade na seqüência gênica como também a mesma ordem de genes nos cromossomos (Bennetzen e Freeling, 2000), mudando apenas as distâncias entre eles (Figura 1). Como relatado em diversos trabalhos, a utilização de seqüências de genes de interesse agrônômico de espécies tais como *Arabidopsis thaliana* e *Lotus japonicus*, que possuem genomas relativamente pequenos e que apresentam alta similaridade com algumas leguminosas, tem suprido a dificuldade de seqüenciamento de genomas maiores e guiado a transferência através de seleção assistida por marcadores

moleculares (Grant *et al.*, 2000; Schmidt, 2000; Ku *et al.*, 2001; Armstead *et al.*, 2004; Stracke *et al.*, 2004; Young *et al.*, 2005; Young *et al.*, 1996). Porém, em alguns casos, apesar da compatibilidade gênica os cruzamentos entre as espécies podem ser dificultados devido ao pareamento cromossômico irregular, causado por diferenças nos seus tamanhos ou pelas suas composições diferenciadas nas regiões não-codantes indicando que, apesar de terem conservado as funções de alguns genes, o restante do genoma pode ter passado por diferentes pressões seletivas.

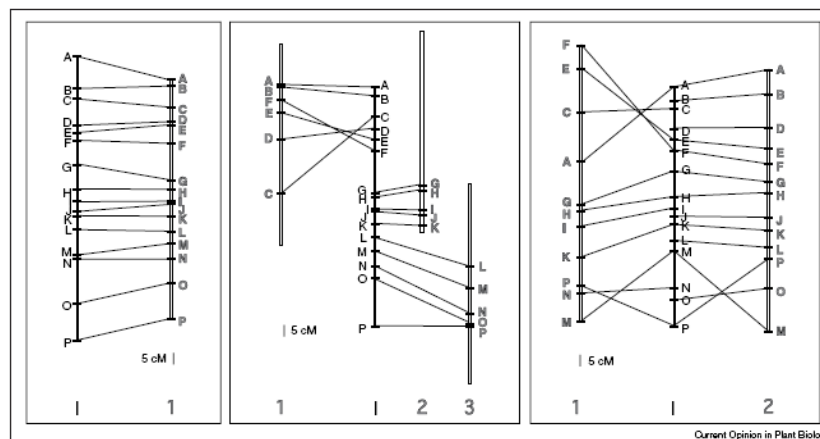


Figura 1. Mapa demonstrativo dos diversos tipos de sintenia utilizando-se diferentes marcadores moleculares (A-P). À esquerda, sintenia completa na ordem e posição dos cromossomos de espécies diferentes. Ao centro, comparação do mesmo mapa com cromossomos de várias outras espécies. À direita, comparação entre dois cromossomos de espécies tetraplóides (1 e 2) com o cromossomo da espécie diplóide, (Schmidt, 2000).

1.04 Seqüências repetitivas

O genoma das plantas, assim como dos demais eucariotos complexos, é organizado de uma forma em que se podem distinguir dois tipos padrões de seqüências de DNA: o DNA não-repetitivo e o repetitivo, isso pode ser mostrado a partir de ensaios de reassociação cinética das fitas de DNA (Flavell, 1974). É no DNA não repetitivo ou de baixo número de cópias que estará contida a mensagem para a síntese de determinada proteína, seja ela estrutural, metabólica, sinalizadora ou de defesa. O restante do genoma é denominado por

alguns de DNA “lixo” (Junk DNA), sendo formado quase que inteiramente por seqüências repetitivas geralmente sem função conhecida (Lewin, 2004) e, durante muito tempo, foi descartado dos estudos científicos.

Famílias de seqüências repetitivas possuem diferentes padrões de organização dentro do genoma e muitas vezes essa organização é específica para um determinado grupo de espécies. Cada padrão surge através dos resultados de diferenças nas respostas às pressões seletivas que podem levar a uma diferenciação na sua composição nucleotídica como também na sua freqüência dentro do genoma. Vegetais com genomas pequenos possuem poucas seqüências repetitivas, que são normalmente encontradas em blocos. Espécies com genomas maiores possuem uma ampla diversidade de seqüências repetitivas que se encontram espalhadas em diversos sítios dos cromossomos (Bennetzen, 2000).

Alguns tipos de repetições também podem apresentar seqüências conservadas entre espécies, muitas vezes essas seqüências possuem uma função conhecida e bem determinada. É o caso de repetições tais como DNA ribossômico (rDNA), que possui os genes 18S; 5,8S e 26S, responsáveis pela síntese das fitas de RNA, que participam na formação das subunidades dos ribossomos. Essas seqüências foram utilizadas por Seijo e colaboradores (2004) para determinar as possíveis espécies que deram origem ao amendoim cultivado. Telômeros são estruturas encontradas nos finais de cada fita de DNA e atuam na estabilização da replicação e proteção dos cromossomos. Centrômeros são repetições que mantiveram sua função extremamente conservada entre os mais variados grupos de plantas e animais, pois apresentam participação na segregação cromossômica durante a mitose e a meiose (Heslop-Harrison, 2000).

Existem também seqüências repetitivas em tandem, satélites e microssatélites que, apesar de não apresentarem uma função conhecida, podem exibir polimorfismos entre espécies, o que é muito importante para o desenvolvimento de marcadores moleculares.

1.05 DNA móvel

Dentre as seqüências repetitivas com maior capacidade de afetar a estrutura e função do genoma estão os elementos transponíveis ou elementos móveis. Esses elementos foram encontrados em todos os organismos estudados até o momento e são importantes constituintes do genoma vegetal. Apresentam diferentes composições, tamanhos e padrões de organização nas mais variadas espécies (Kumar e Bennetzen, 1999). A principal característica desse grupo

é a sua capacidade de se mover dentro do genoma, inserindo-se em novos sítios, próximos ou até mesmo dentro de seqüências gênicas, o que pode causar mutações de inserção, alteração da estrutura e função de genes, rearranjos cromossômicos, mudanças na regulação gênica e aumento exagerado do tamanho do genoma, podendo dessa forma servir como fonte de diversidade (Deragon e Capy, 2000 citado por: Alix e Heslop-Harrison, 2004). Todo o aparato utilizado para a sua transposição é sintetizado na maioria das vezes pelo próprio elemento (elementos autônomos), mas em alguns casos ele pode utilizar as enzimas sintetizadas por outros elementos para realizar a sua movimentação (elementos não-autônomos) (Wessler, 2006).

Existem duas classes principais de elementos transponíveis: Classe I composta pelos Retrotransposons, elementos genéticos que se movem no interior de uma mesma célula via um intermediário de mRNA, seguida de transcrição reversa e inserção da cópia de cDNA em um novo sítio no genoma (Figura 2); classe II composta pelos Transposons, que são elementos que se movem no genoma através da excisão de sua seqüência de um sítio e inserção em um outro. No caso dos retrotransposons, estes apresentam uma importância predominante na formação dos genomas, já que seu mecanismo de transposição permite a transcrição reversa de uma fita de mRNA em várias cópias de cDNA, multiplicando intensamente seu número de cópias e espalhando-se potencialmente, por todos os cromossomos.

Os retrotransposons podem ser divididos em dois grupos: LTR e Não-LTR. O grupo LTR é subdividido nos tipos Ty1-Copia e Ty3-Gypsy, que se distinguem um do outro de acordo com a sua similaridade na seqüência e ordem em que os genes aparecem, sendo ambos os tipos amplamente distribuídos no reino vegetal. Além disso, o tipo Ty3-Gypsy assemelha-se muito com os retrovirus, pois apresenta parte de um gene parecido com o que codifica a síntese das proteínas do envelope, que auxilia a sua transmissão para outras células (Wright e Voytas, 2001; Marco e Marín, 2005). O grupo Não-LTR divide-se nos tipos LINEs e SINEs (Figura 3), também distribuídos por todo o reino vegetal, mas possuindo uma freqüência maior nos genomas de mamíferos (Kumar e Bennetzen, 1999; Hansen e Heslop-Harrison, 2004). De acordo com os modelos de cromossomo vegetal propostos (Figura 4), percebe-se que a maior parte do genoma é constituída por seqüências repetitivas, sendo que os retrotransposons LTR podem constituir cerca de 50% (Schmidt e Heslop-Harrison, 1998; Kumar e Bennetzen, 1999), enquanto os genes perfazem cerca de 5% de todo o genoma, fato que torna importante o estudo desses elementos na composição e função em cada organismo.

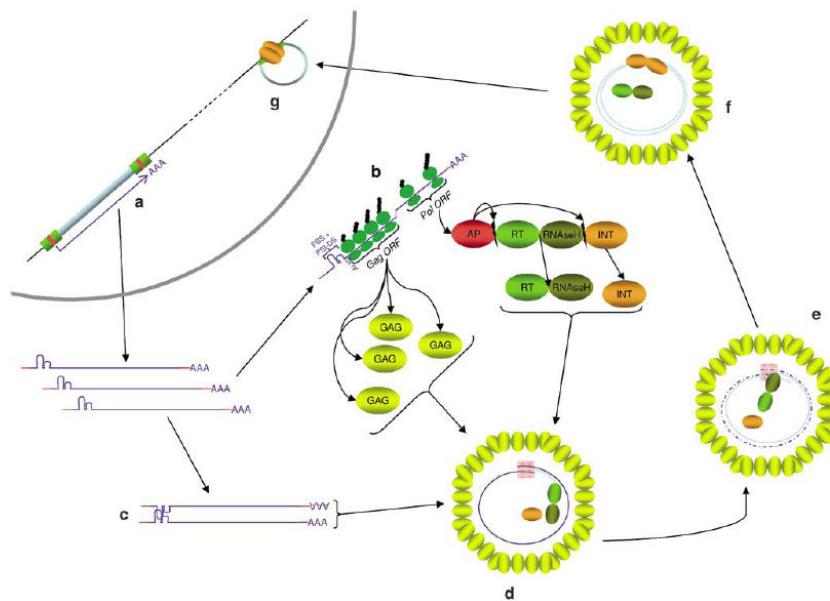
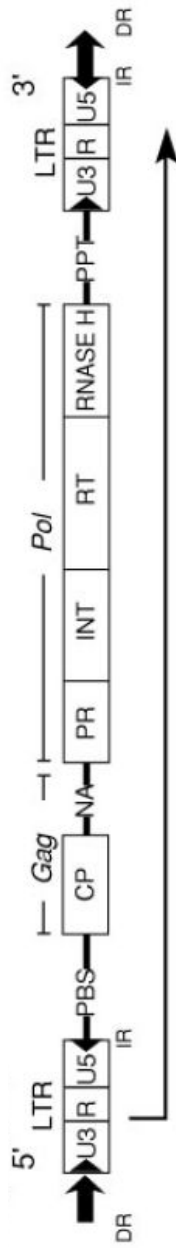


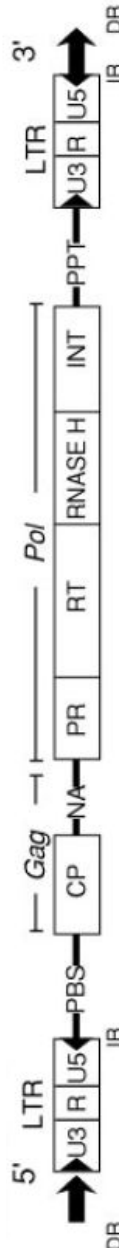
Figura 2. Ciclo de vida hipotético do retrotransposon LTR. (a) transcrição da fita de mRNA. (b) tradução e síntese de proteínas das regiões do GAG e POL; POL inclui as proteínas AP (protease), RT (transcriptase reversa), RNaseH e INT (integrase). (c) formação do dímero de RNA. (d) empacotamento do RNA e das proteínas RT, RNaseH e INT e início da transcrição reversa. (e) degradação da matrix de RNA e síntese da segunda fita de cDNA. (f) término da fita dupla de cDNA e ligação de INT junto à sequência LTR. (g) integração da dupla fita em um novo sítio do genoma do hospedeiro. A linha cinza mais espessa ilustra a membrana nuclear. Adaptado: Sabot e Schulman, 2006.

Retrotransposons LTR

Grupo Ty1-copia



Grupo Ty3-gypsy



Retrotransposons Não - LTR

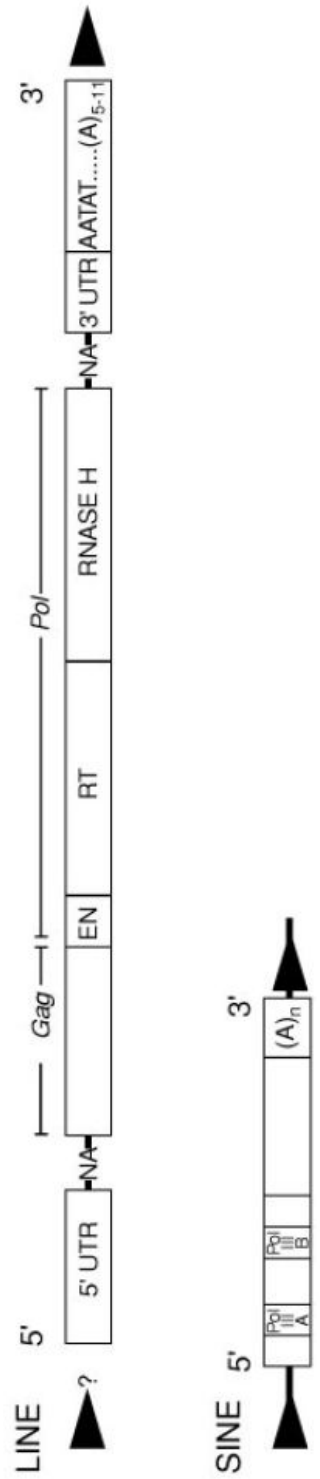


Figura 3. Diferentes grupos de retrotransposons LTR e Não-LTR, mostrando suas seqüências terminais e posições das seqüências codificadoras. Adaptado: Kumar e Bennetzen, 1999.

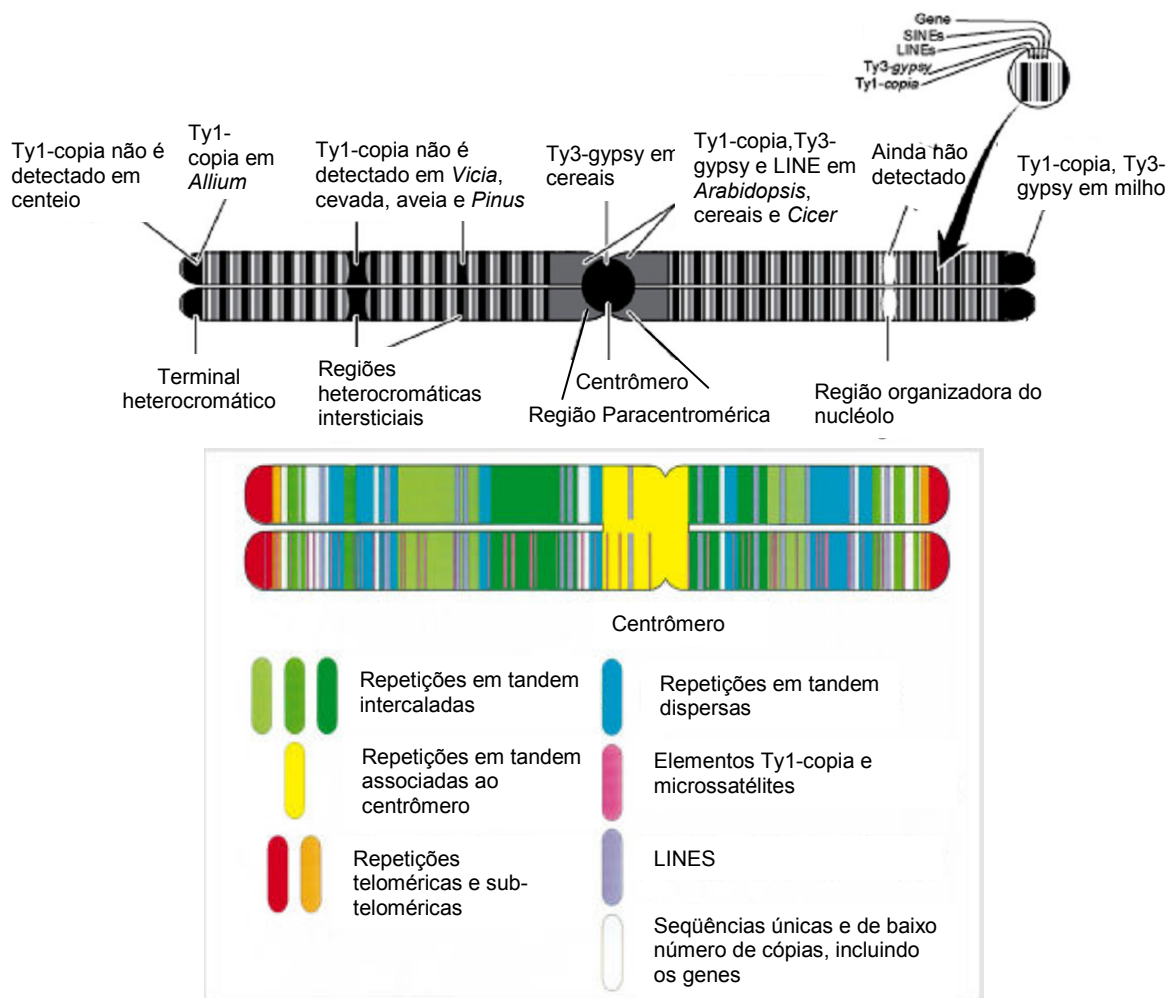


Figura 4. Padrões de distribuição de seqüências repetitivas e de genes ao longo de um cromossomo vegetal. Dependendo da espécie e do tipo de retrotransposon a distribuição muda consideravelmente. Adaptado: Schmidt e Heslop-Harrison, 1998 e Kumar e Bennetzen, 1999.

1.06 Controle de atividade

Apesar da intensa participação na formação dos genomas de diversas espécies, os retrotransposons, na maioria das vezes, encontram-se inativos. Isso se deve aos vários mecanismos de regulação existentes, alguns deles realizados pelo hospedeiro e outros realizados pelo próprio elemento (Grandbastien, 1998; Bennetzen, 2000). É de se imaginar que caso sua transposição não fosse controlada, a multiplicação do retrotransposon poderia acarretar na morte celular, o que não é interessante para a sobrevivência do hospedeiro nem do retroelemento. Dessa forma, os organismos desenvolveram técnicas que, na maioria das

vezes, permitem a presença do elemento repetitivo, mas inativando-o. Um dos mecanismos desenvolvidos pelo hospedeiro, principalmente em plantas, é a metilação de suas seqüências de DNA, que normalmente é utilizada para o controle da expressão dos genes em geral. Cerca de 80% das citosinas das regiões CG e CNG do genoma encontram-se no estado metilado, enquanto que pequenas regiões não-metiladas estão relacionadas com genes ativos (Hirochika *et al.*, 2000). Em muitos casos, a metilação da seqüência do retroelemento resultará na sua inativação, mas esse mecanismo não é necessariamente utilizado por todos os eucariotos, já que em *Drosophila sp.* e *Saccharomyces cerevisiae* ele não é observado (Dorer e Henikoff, 1994; Grunstein, 1998).

A preferência por sítios de inserção pode exercer um excelente controle de atividade. Apesar de alguns grupos de retrotransposons possuírem locais preferenciais para a sua inserção, como centrômeros e telômeros, muitos outros encontram-se em regiões não-gênicas ou de heterocromatina que, normalmente, não apresentam grande atividade (Kumar e Bennetzen, 1999). Em outros casos, eles podem se inserir uns nos outros ou em íntrons, evitando qualquer ativação do sistema de proteção do hospedeiro (Martienssen, 1998). Além disso, qualquer forma de modificação na seqüência dos elementos pode acarretar na sua inativação. Elementos que perdem uma das seqüências LTR ficam impedidos de se inserirem em um novo local. Há também aqueles que sofrem deleção total, sendo praticamente eliminados do genoma (Vitte e Panaud, 2005).

Entretanto, devido ao seu mecanismo de replicação utilizando um intermediário de RNA, o modo aparentemente mais simples de controlar sua atividade seria durante a transcrição. Estudos realizados com tabaco mostraram que a quantidade de mRNAs oriundos da transcrição dos elementos Tto1 e Tos17, ambos do tipo Ty1-copia, está diretamente relacionada com a sua taxa de transposição (Hirochika, 1993; Hirochika *et al.*, 1996). Por outro lado, em espécies de cevada, há relatos de que alguns tipos de retrotransposons são intensamente transcritos em diversos tecidos, mas sua transposição ou o surgimento de novos sítios de inserção não foram encontrados, sugerindo que além do controle transcricional existam outras formas de regulação da atividade desses elementos, tal como controle por RNA interferente (Kumar e Bennetzen, 1999; Navarro-Quezada e Schoen, 2002). Acredita-se que, durante o desenvolvimento embrionário, os diferentes tecidos passem por controles de expressão gênica distintos e, dessa forma, a atividade dos retroelementos também acabe sofrendo regulação durante a transcrição, na tradução e na inserção.

Em condições adversas de estresses bióticos e abióticos tais como: infecções, cruzamentos interespecíficos e poliploidização (Zhao *et al.*, 1998; Liu e Wendel, 2000),

salinidade, exposição à radiação, cultura de tecidos e fusão de protoplastos, os retrotransposons podem readquirir sua atividade, multiplicando-se no interior do genoma (Wessler, 1996; Grandbastien, 1998). Acredita-se que os elementos “percebam” uma situação de dificuldade que possa levar à morte do hospedeiro e dessa forma eles se multiplicam, na esperança de serem transmitidos para outros genomas, ou por cruzamentos ou por transmissão horizontal, apesar de esta última ter sido pouco observada (Flavell, 1999; Diao *et al.*, 2006). Além disso, a multiplicação dos retrotransposons no genoma pode causar o surgimento de novas características desejáveis através de mutações gênicas, o que pode auxiliar na sobrevivência do retroelemento.

1.07 Diversidade e utilização

De acordo com a vasta diversidade e heterogeneidade de retrotransposons, acredita-se que eles sejam muito antigos e que possivelmente estavam presentes em espécies ancestrais de vários grupos de organismos. Essa característica pode ser percebida através de análises filogenéticas que agrupam seqüências tais como a LTR, a da transcriptase reversa e da integrase de acordo com o grau de parentesco entre as espécies (Kumekawa *et al.*, 1999; Langdon *et al.*, 2000; Dixit *et al.*, 2006). Entretanto, em alguns casos fica difícil determinar a história evolutiva de uma espécie com base na distribuição de retroelementos, principalmente se ela apresentar uma origem complexa como é o caso da família *Brassicaceae* que possui espécies diplóides com genomas AA, BB e CC além de espécies tetraplóides com genomas AACC, AABB e BBCC. Nesse caso, a filogenia pode diferir consideravelmente das relações genéticas entre as espécies hospedeiras, obtidas por diferentes marcadores moleculares, como AFLP, RAPD, RFLP e SSR (Alix e Heslop-Harrison, 2004).

O sucesso de colonização do DNA vegetal por esses elementos deve-se principalmente ao fato de que as plantas, em geral, aceitam melhor os eventos de poliploidização, o que permite, além da multiplicação do genoma como um todo, a amplificação em larga escala dos retrotransposons (Heslop-Harrison, 2000). Características reprodutivas dos vegetais, tais como formação de órgãos reprodutivos a partir de tecidos meristemáticos também auxiliam a distribuição desses elementos, pois permitem a transmissão de qualquer característica genética adquirida para as próximas gerações (Kumar e Bennetzen, 1999). Além disso, as taxas de reprodução dos vegetais são muito altas, já que vários grãos de pólen podem entrar em contato com várias flores num mesmo ciclo reprodutivo. Nos vegetais, também há uma taxa de cruzamentos interespecíficos muito alta e essa mistura de genomas, apesar de em alguns

casos gerar um descendente estéril, também leva a um estresse causado pela incompatibilidade genômica, o que pode levar à ativação dos retrotransposons. Associada a esses fatores, a alta taxa de erro de correção da enzima RNA polimerase em regiões repetitivas e da enzima transcriptase reversa, facilitam o surgimento de mutações nas seqüências dos retroelementos, o que permite que eles se diferenciem rapidamente (Bennetzen, 2000).

Devido a essa diversidade, o padrão polimórfico gerado pelos retrotransposons pode ser utilizado para o desenvolvimento de novos marcadores moleculares. Técnicas recentemente desenvolvidas vêm sendo utilizadas para o desenvolvimento dos marcadores baseados em retrotransposons (Kumar e Bennetzen, 1999). Uma dessas técnicas, o SSAP é uma variação do AFLP, com a diferença de usar iniciadores específicos para a seqüência LTR, permitindo a análise de polimorfismo e o mapeamento genético (Rodriguez *et al.*, 2006). Uma outra técnica (RBIP) permite detectar novos sítios de inserção, a partir do uso de iniciadores específicos para esses elementos que amplificam a região flanqueadora do DNA genômico do hospedeiro (Kumar e Bennetzen, 1999). Técnicas como IRAP e REMAP surgiram visando o uso de uma metodologia diferente das técnicas anteriores. Ao invés de utilizar enzimas de restrição, que muitas vezes são sensíveis e não cortam o DNA, essas técnicas utilizam iniciadores que amplificam regiões entre retrotransposons (IRAP) e entre retrotransposons e microssatélites (REMAP), gerando um alto padrão polimórfico de fragmentos de DNA (Kalendar *et al.*, 1999).

Além do uso como marcadores moleculares, os retrotransposons também podem ser utilizados para o estudo de expressão gênica e de mutação, já que, quando se inserem próximos ou dentro de uma seqüência gênica, eles podem alterar por completo a função do gene (Watson, 1992; Kumar e Bennetzen, 1999). Acredita-se também que alguns grupos de retrotransposons possam ser infecciosos, já que apresentam seqüências similares às do gene do envelope. Caso isso seja confirmado, esses elementos podem ser utilizados como vetores para transformação de diferentes organismos (Wright & Voytas, 2001).

1.08 Caracterização do retroelemento

Com o advento de novas metodologias, ferramentas de bioinformática e com o crescente número de seqüências depositadas em bancos de dados de acesso público, está sendo possível reavaliar a composição genômica de diversas espécies e caracterizar novas

famílias de retrotransposons. Os elementos LTR melhor caracterizados são os do grupo Ty1-Copia. Até o presente momento esses elementos estão sendo estudados com maior rigor em *D. melanogaster*, *Saccharomyces* sp e em espécies vegetais que já possuem seu genoma totalmente sequenciado como *Arabidopsis thaliana* e *Oryza sativa*. Pouco se sabe a respeito desses elementos em outras espécies. Na família das leguminosas, as seqüências repetitivas melhor caracterizadas são as da soja (*Glycine max*) e da ervilha (*Pisum sativum*), ainda que em um índice muito inferior ao grupo das gramíneas e de algumas brassicáceas (Vitte e Panaud, 2005). No caso do amendoim, não há relatos conhecidos ou de grande importância a respeito de retrotransposons. As seqüências repetitivas encontradas no GenBank são oriundas, basicamente, de trabalhos de seqüenciamento de ESTs, RGAs ou de microssatélites. Os maiores esforços dos grupos que trabalham com biologia molecular em *Arachis* estão concentrados na caracterização de proteínas alergênicas e na construção de mapas genéticos para espécies silvestres e para o amendoim cultivado, visando possibilitar o direcionamento dos programas de melhoramento genético da cultura. Até agora as técnicas mais utilizadas para a construção desses mapas foram AFLP, RAPD, RFLP e SSR, buscando principalmente a identificação de marcadores que estejam ligados a genes de resistência a doenças e a estresses abióticos (José, 2006). Os retrotransposons que venham a ser obtidos por esses grupos normalmente serão considerados como restos de seqüenciamento e serão possivelmente descartados.

Recentemente foram construídas duas bibliotecas BAC para as espécies *A. duranensis* e *A. ipaënsis* (Proite *et al.*, 2006). Espera-se que com essas bibliotecas o seqüenciamento dos genomas seja facilitado, assim como a identificação de genes e de elementos repetitivos de interesse, além de possibilitar a ligação do mapa genético com o mapa físico. Dessa maneira, a saturação do mapa genético de espécies silvestres e do amendoim poderá facilitar a identificação de genomas mais compatíveis, indicando ao melhorista quais as melhores espécies a serem utilizadas para a síntese dos anfidiplóides e acelerar o processo de melhoramento da cultura através de cruzamentos interespecíficos, a partir da identificação de marcadores ligados a genes que controlem características importantes agronomicamente.

2. Hipótese

De acordo com as características citadas e com base em resultados previamente obtidos (discutidos no trabalho), imaginou-se duas situações hipotéticas as quais serão testadas neste trabalho. A primeira delas infere que a espécie ancestral de *A. duranensis* e *A. ipaënsis* apresentava um alto número de cópias do retrotransposon isolado neste trabalho e que este número foi mantido no genoma AA e reduzido no genoma BB, muito antes da origem do amendoim cultivado. A segunda hipótese é a de que esse ancestral possuía um baixo número de cópias que então foi multiplicado no genoma AA e praticamente inalterado no genoma BB, também em uma época muito anterior ao aparecimento de *A. hypogaea* (Figura 5).

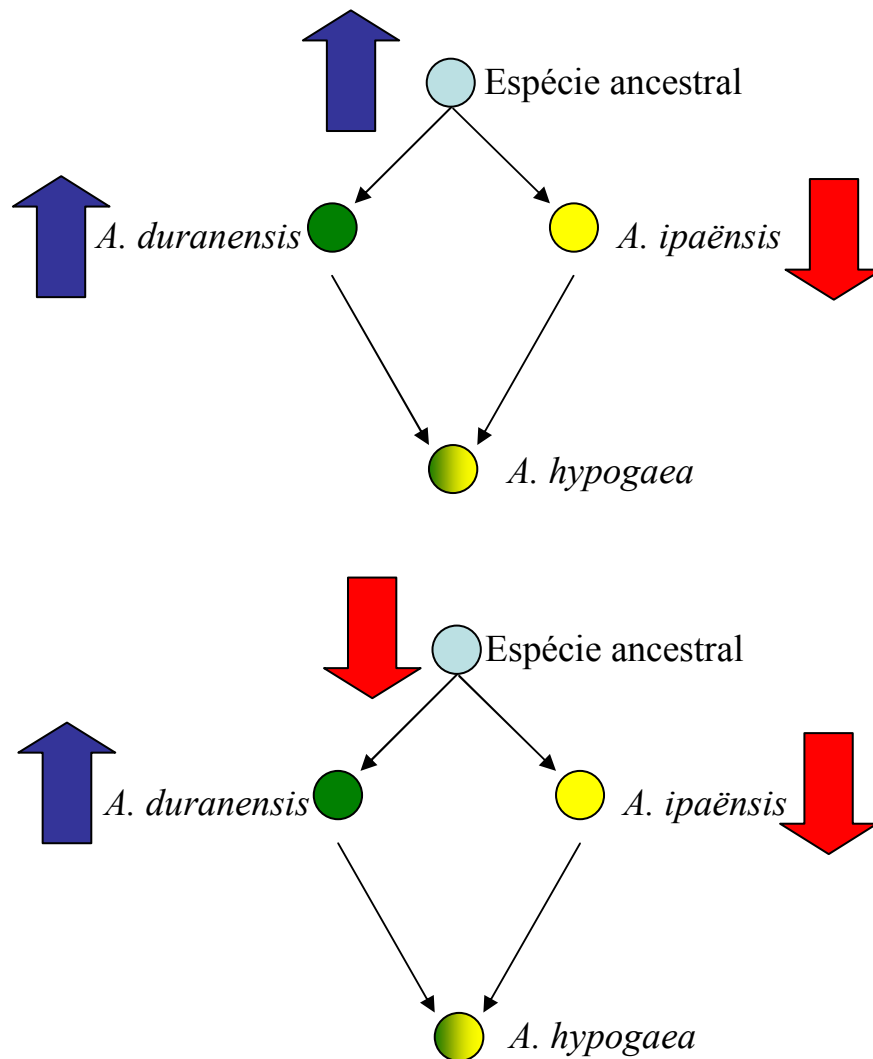


Figura 5. Ilustração dos dois modelos hipotéticos a serem testados neste trabalho.

3. Objetivos

3.01 Objetivos gerais

Tendo em vista a baixa informação a respeito da caracterização de elementos repetitivos e sua contribuição na formação dos genomas de diversas espécies de *Arachis*, o objetivo geral deste trabalho foi o isolamento e a caracterização de um retrotransposon-LTR, a partir de espécies de genoma AA e BB, visando o conhecimento da organização desse elemento e entender seu processo evolutivo nos genomas das espécies que melhor se relacionam com o amendoim cultivado.

3.02 Objetivos específicos

- Isolar elementos repetitivos através de reação em cadeia da polimerase (PCR) utilizando-se iniciadores específicos e ligação em plasmídeo.
- Caracterizar esses elementos quanto a sua seqüência nucleotídica e de aminoácidos.
- Determinar o número de cópias desses elementos nos genomas de *A. hypogaea* (AABB), *A. duranensis* (AA) e *A. ipaënsis* (BB).
- Estudar a metilação dos elementos isolados dos genomas das três espécies de *Arachis* através de Southern blot.
- Analisar as distâncias genéticas entre as seqüências assim como o número de substituições sinônimas e não-sinônimas.
- Agrupar filogeneticamente essas seqüências entre si, comparando com as seqüências de outros grupos vegetais, e entender a evolução desses elementos repetitivos nos três genomas.
- Verificar qual modelo hipotético (Ítem 2) explica melhor os resultados obtidos.

4. Material e Métodos

4.01 Material Vegetal

O DNA genômico utilizado foi adquirido de plantas mantidas em casas de vegetação na EMBRAPA Recursos Genéticos e Biotecnologia. Foram extraídos DNA das espécies: *A. hypogaea* (tetraplóide AABB – variedade Tatu), *A. ipaënsis*, acesso KG30076 (diplóide BB) e *A. duranensis*, acesso V14167 (diplóide AA). Folhas jovens recém-surgidas e antes da expansão foliar foram escolhidas, para reduzir a extração de compostos fenólicos, polissacarídeos e proteínas presentes nas folhas adultas. Um protocolo de extração de DNA baseado em CTAB, publicado por Grattapaglia e Sederoff (1994) foi modificado para o uso em *Arachis* sp, no qual um novo passo de precipitação com 1,2M de NaCl foi adicionado: 200 mg de tecido vegetal foram macerados em cadinho com o auxílio de Nitrogênio líquido e adição de 700µL de CTAB (2x) e 2-mercaptoetanol (2 µL/mL de CTAB) em cada tubo de microcentrifuga de 2 mL. As amostras foram incubadas a 65°C por 60 minutos e agitadas levemente a cada 10 minutos. Após esse período, foram adicionados 700µL de clorofórmio/álcool-isoamílico (24:1) para cada amostra e misturou-se até formar uma emulsão. As amostras foram centrifugadas a 14000 rpm por 5 minutos. A fase aquosa foi transferida para dois tubos novos de 1,5 mL. Foram adicionados 600µL de tampão CTAB (1x), agitou-se lentamente e centrifugou-se a 14000 rpm por 5 minutos. O sobrenadante foi descartado cuidadosamente e o DNA dos dois tubos foi ressuspensão em 300µL de NaCl (1,2M) até ser dissolvido completamente. Em seguida, 600µL de etanol absoluto gelado foram adicionados e o DNA precipitado em freezer por uma hora. As amostras foram centrifugadas a 14000 rpm por 5 minutos e descartou-se o sobrenadante. O precipitado de cada tubo foi lavado duas vezes com 500µL de etanol 70%. Os precipitados foram secos e ressuspensos em 100µL de água Milli-Q estéril com 0,01mg/mL de RNase-A. As amostras foram quantificadas em gel de agarose 1% usando um marcador de massa (High Mass Ladder – Invitrogen).

4.02 Desenho de iniciadores

Alguns clones obtidos através de seqüenciamento de clones de bibliotecas genômicas de trabalhos anteriores, foram utilizados em Dot blot e hibridizados com DNA genômico de *A. duranensis* e *A. ipaënsis*. Os clones que apresentaram um forte sinal de hibridização foram selecionados e comparados com outras seqüências do GenBank. Vários desses clones apresentaram certa similaridade com retrotransposons e as seqüências foram incluídas em um banco de dados de *Arachis*, no qual foi feito o alinhamento com todas elas para aumentar a seqüência conhecida desses elementos. A seqüência obtida foi submetida à comparação a enzima transcriptase reversa através de BLASTx (Altschul *et al.*, 1990). Utilizando o programa Staden Package (Staden *et al.*, 1998), três pares de iniciadores foram desenhados para a amplificação da seqüência completa da enzima transcriptase reversa (RT) dos genomas das espécies *A. hypogaea*, *A. ipaënsis* e *A. duranensis*. Quando se tornou necessário, novo iniciador foi desenhado para completar a seqüência interna do inserto (Tabela 1).

Tabela 1. Denominação e seqüência dos iniciadores específicos desenhados para amplificação do gene codificador da enzima transcriptase reversa em *Arachis* sp.

Nome	Seqüência
Rep-RT-F1	5'-AAGGACACACAAGACAGCTC-3'
Rep-RT-F2	5'-GTACGCACAAGATCCTATTG-3'
Rep-RT-F3	5'-CTAAATCCAGCCATGAAGG-3'
Rep-RT-R1	5'-GTCAGCTACAAGGAGATTGC-3'
Rep-RT-R2	5'-GGAGATGATAGGTGCAGAAG-3'
Rep-RT-R3	5'-TCACACATCAGTTCAAATGG-3'
Athila-internal-1	5'-GGATCATTTTCCTTTACCA-3'

4.03 Reação em cadeia da polimerase

De acordo com Wright e Voytas (2001), foi realizada uma reação em cadeia da polimerase para amplificação da seqüência da enzima transcriptase reversa, onde 100ng de DNA genômico de *A. hypogaea*, *A. ipaënsis* e *A. duranensis* foram adicionados à reação contendo: 3µM de cada iniciador; 8mM de dNTPs; 125mM de MgCl₂; 2,5U de *Taq* DNA polimerase (Invitrogen); 10% de tampão de PCR em um volume final de 20µL de reação. O programa utilizado continha: desnaturação inicial a 92°C por 5 minutos, 30 ciclos de (92°C por 20 segundos, 50°C por 30 segundos e 72°C por 90 segundos) e extensão final a 72°C por 5 minutos. Visando obter uma maior variabilidade de seqüências amplificadas e garantir que

pelo menos um par de iniciadores iria amplificar a seqüência desejada, os três pares de iniciadores foram utilizados ao mesmo tempo na mesma reação.

Para amplificação de insertos em plasmídios foi utilizado o seguinte protocolo: 0,2ng de DNA; 7,5µM de cada iniciador universal (T7 e SP6); 4mM de dNTPs; 37,5mM de MgCl₂; 2,5U de *Taq* DNA polimerase, 10% de tampão de PCR em volume final de 25 µL. O programa utilizado continha: Desnaturação inicial a 94°C por 5 minutos, extensão final a 72°C por 5 minutos, 34 ciclos de (94°C por 30 segundos, 55°C por 50 segundos e 72°C por 1 minuto e 30 segundos).

4.04 Ligação e Transformação

A ligação dos produtos de PCR foi realizada em pGEM-T Easy (Promega), de acordo com o protocolo da empresa: 50ng de vetor, 3U de enzima T4 DNA ligase, 10% tampão de ligação e a quantidade de inserto variando de acordo com o seu peso molecular num volume final de 10µL. A reação foi realizada a 4°C por aproximadamente 16 horas.

As amostras foram diluídas entre 10-20 vezes e 1µL foi utilizado para transformação de células de *Escherichia coli* por eletroporação da seguinte forma: adicionar aproximadamente 10ng de produto de ligação a 40µL de células competentes (XL1-Blue), incubar em gelo por 1 minuto, transferir para cuvetas, submeter ao eletroporador (resistência de 200 Ohms, capacitância de 25 µFD e 1,8 Kvolts), acrescentar 1mL de meio LB líquido, incubar por 1 hora a 37°C. Foram plaqueados 100µL em meio LB sólido contendo ampicilina (100mg/mL), IPTG (100nM) e X-GAL (50ng/mL) seguido de incubação a 37°C por 16 horas. As colônias recombinantes foram selecionadas e amplificadas em meio LB líquido contendo ampicilina (100mg/mL) e com agitação a 37°C.

4.05 Preparação de DNA plasmidial

Para extração de DNA plasmidial foi utilizado o protocolo de lise alcalina publicado por Ahn *et al.*, 2000. 1) Baixar as células por centrifugação a 11.000g por 1 minuto, descartar o sobrenadante e ressuspender o precipitado em 100µL de tampão de ressuspensão (50mM Tris-HCl; pH= 8; 10mM EDTA; 20µg de RNase A). 2) Adicionar 100µL de tampão de lise (200mM NaOH; 1% SDS). 3) Misturar delicadamente à temperatura ambiente. 4) Adicionar 120µL de tampão de neutralização (3M acetato de potássio, pH= 5,5). 5) Misturar por 3 minutos à temperatura ambiente. 6) Remover os resíduos de células por centrifugação a

11.000g por 1 minuto e transferir o sobrenadante para novo tubo. 7) Adicionar 200µL de isopropanol para precipitar o plasmídeo. 8) Misturar por 1 minuto à temperatura ambiente. Centrifugar a 11.000g por 1 minuto e descartar o sobrenadante. 9) Lavar o precipitado com etanol 70% gelado, secar o precipitado e ressuspender em 100µL de água Milli-Q estéril. Após a extração do DNA plasmidial, a presença do inserto foi detectada a partir de digestão com a enzima de restrição *EcoRI* e separação dos fragmentos por eletroforese em gel de agarose 1%, junto com marcador de peso molecular 1Kb DNA ladder (Invitrogen).

4.06 Seqüenciamento e análise de seqüências

Para o seqüenciamento das amostras foi utilizado o Kit DYEnamic ET Terminator (Amersham, Biosciences). Cerca de 500ng de DNA foram adicionados à solução contendo 3µM de iniciadores para ambos os sentidos, 2µL de premix para um volume final de 10µL. O seqüenciamento foi realizado utilizando um seqüenciador automático ABI377 (Applied Biosystems).

As seqüências obtidas foram processadas utilizando-se o programa Staden Package. A ferramenta pregap permitiu o tratamento inicial das seqüências, removendo as seqüências dos vetores, além de se fazer o teste de qualidade de cada base nucleotídica através de PHRED (Ewung *et al.*, 1998). Utilizando o programa gap4, foi possível alinhar as fitas que foram seqüenciadas em ambos os sentidos, permitindo construir vários contigs -um para cada clone obtido na ligação- e também corrigir a orientação de cada um para uso em novos ensaios.

Após o processamento, as seqüências foram submetidas ao programa BLASTn e BLASTx (Altschul *et al.*, 1997) contra os dados totais do GenBank, visando à análise de similaridade com seqüências de retrotransposons já publicadas. Além disso, foi feita a tradução das seqüências de DNA utilizando-se a ferramenta “*Translate – TrEMBL*” no site www.expasy.ch/tools/dna.html. A seqüência protéica foi então submetida à análise pelo programa BLASTp para a detecção de possíveis regiões conservadas.

4.07 Alinhamento e análise filogenética

As seqüências de DNA e de aminoácidos dos contigs obtidos foram utilizadas para alinhamento múltiplo utilizando-se o programa ClustalX (Chenna *et al.*, 2003). As correções manuais necessárias, tais como remoção de terminais de seqüências não-alinhados, foram feitas utilizando-se o programa Jalview (Clamp *et al.*, 2004 - www.jalview.org). Os dados

obtidos foram utilizados para a construção de árvores filogenéticas através do programa Mega, versão 3.1 (Kumar *et al.*, 2004). A distância evolutiva das seqüências de DNA foi calculada pelo método de correção de dois parâmetros de Kimura (Kimura, 1980). As relações filogenéticas foram determinadas pelo método de *neighbor joining* e a reprodutibilidade de cada ramo calculada através de bootstrap. A distância das seqüências protéicas foi calculada pelo método *p distance* (Nei e Kumar, 2000), sendo que as relações filogenéticas assim como a confiança de cada ramo foram determinadas da mesma forma. Também foram construídas árvores filogenéticas junto com as seqüências protéicas utilizadas por Wright e Voytas (2001), visando um estudo evolutivo da seqüência do retrotransposon em diferentes grupos vegetais. As distâncias, assim como as taxas de substituições sinônimas e não-sinônimas, foram calculadas utilizando-se o programa Mega 3.1.

4.08 Análise do número de cópias

O cálculo do número de cópias dos retrotransposons nos genomas das espécies *A. hypogaea*, *A. ipaënsis* e *A. duranensis* foi feito através de Dot blot, no qual diluições em série de DNA genômico e do plasmídio contendo um inserto com a seqüência completa da enzima transcriptase reversa foram plotadas em membrana Hybond-N⁺ (Amersham, Pharmacia Biotech) e fixadas sob luz ultravioleta. As diluições foram feitas calculando-se a quantidade de inserto em pg para cada 500ng de DNA genômico das três espécies estudadas (Pearce *et al.*, 1996; Nouzová *et al.*, 1999). Assim, sabendo-se que 1pg de DNA equivale a 965 Mpb, que o plasmídio pGEM-T Easy e o inserto possuem, respectivamente, 3015pb e 1100pb e a quantidade de DNA no genoma das três espécies: *A. hypogaea* (2C = 5,93pg), *A. duranensis* (2C = 2,61pg) e *A. ipaënsis* (2C = 2,8pg) (Temsch e Greilhuber, 2000 e 2001), foi possível determinar o peso molecular, em pg, de uma cópia do plasmídio contendo uma única cópia do inserto. Dessa maneira, foram escolhidas diluições variando entre 10 a 100.000 cópias para o plasmídio e de 0,32ng a 500ng de DNA genômico. A sonda não-radioativa foi construída utilizando-se o kit DIG DNA Labeling and Detection (Roche), de acordo com o protocolo do fabricante. Cerca de 3µg de DNA foram adicionados em água Milli-Q estéril num volume final de 15µL e, em seguida, desnaturados a 95°C por 10 minutos e rapidamente colocados em gelo. Adicionaram-se 2µL da mistura de hexanucleotídeos (10X), 2µL de dNTPs marcados e 1µL de enzima, incubando-se a 37°C por aproximadamente 16 horas. A reação foi terminada através de aquecimento a 65°C por 10 minutos. A eficiência da marcação foi determinada

aplicando-se diluições de DNA controle e da sonda desejada em membrana Hybond-N⁺ e comparando-se os sinais obtidos na detecção imunológica.

As amostras de DNA foram aplicadas na membrana manualmente, o volume máximo para cada aplicação foi de 1µL. Pela dificuldade de pipetar amostras muito concentradas decidiu-se utilizar DNAs com, no máximo, 200ng/µL e esperar que eles secassem na membrana, repetindo-se a aplicação de 1µL no mesmo ponto até alcançar a concentração desejada. As amostras de DNA genômico foram aplicadas na membrana junto com diluições do clone Ah-Athila-9, que contém o inserto com o gene de RT. Todas as amostras foram aplicadas em duplicata nas membranas, tanto o controle como o DNA genômico. As concentrações necessárias para o experimento foram calculadas de acordo com a quantidade de DNA para o genoma diplóide de cada espécie como pode ser visto na Tabela 2.

A membrana do Dot blot contendo as diluições para análise do número de cópias foi então colocada em saco plástico vedado contendo solução de pré-hibridização a 42°C com agitação constante por aproximadamente 5 horas. Removeu-se essa solução e adicionou-se a solução de hibridização contendo a sonda marcada e desnaturada. O saco plástico foi novamente vedado e incubado (submerso) em banho-maria a 42°C por aproximadamente 16 horas. Após a hibridização, a membrana foi tratada da seguinte maneira: duas lavagens de 15 minutos com agitação constante em solução de lavagem I a temperatura ambiente, duas lavagens de 15 minutos com agitação constante em solução de lavagem II a 68°C. A stringência utilizada em cada lavagem permitiu a remoção de fragmentos com menos de 90% de similaridade. Após as lavagens stringentes, os próximos passos são realizados a temperatura ambiente. Lavagem em solução de ácido maléico por 5 minutos. Incubação em solução de bloqueio com agitação constante por 1 hora. Incubação em solução com anticorpo por uma hora. Três lavagens de 15 minutos em solução de ácido maléico + tween 20. Incubação em solução de detecção por 5 minutos. Transferência da membrana para saco plástico e adicionar substrato, com incubação por 5 minutos. Transferência da membrana para filme de PVC e incubação por 15 minutos a 37°C. Sensibilização de filme radiográfico (Kodak-Bio). Os protocolos das soluções podem ser vistos no ANEXO I – soluções utilizadas no Dot blot.

O Filme radiográfico foi digitalizado e foi feita a análise da imagem a partir de comparação direta do tamanho e intensidade do sinal gerado para a diluição do DNA genômico com a diluição controle através do programa MultiGauge (FUJIFILM). Os sinais de hibridização geraram uma tabela com dados numéricos que foram utilizados para a construção de gráficos de regressão linear na planilha Excel, permitindo desta forma obter confiabilidade

nos resultados e, a partir da fórmula obtida, calcular o valor do número de cópias do elemento nos três genomas. Os valores que não se encaixaram ou que destoaram demais o cálculo do número de cópias foram descartados da análise. As curvas foram construídas buscando-se o melhor coeficiente de correlação (R^2) que indica quão significativos são os dados de dispersão.

Tabela 2. Quantidade utilizada de DNA e número de cópias esperado do plasmídeo contendo o clone Ah-Athila-9 para o genoma das três espécies de *Arachis*.

Clone Ah-Athila-9 - <i>A. hypogaea</i>		Clone Ah-Athila-9 - <i>A. duranensis</i>		Clone Ah-Athila-9 - <i>A. ipaënsis</i>	
Quantidade de DNA	nº de cópias	Quantidade de DNA	nº de cópias	Quantidade de DNA	nº de cópias
0,718pg	1	1,632pg	1	1,52pg	1
7,18pg	10	16,32pg	10	15,2pg	10
71,8pg	100	163,2pg	100	152pg	100
0,359ng	500	0,816ng	500	0,760ng	500
0,718ng	1000	1,632ng	1000	1,52ng	1000
3,59ng	5000	8,16ng	5000	7,6ng	5000
7,18ng	10000	16,32ng	10000	15,2ng	10000
35,9ng	50000	81,6ng	50000	76ng	50000
71,8ng	100000	163,2ng	100000	152ng	100000

4.09 Análise de Metilação

A análise de metilação foi realizada utilizando-se a técnica de Southern blot (Southern, 1975). Cerca de 5µg de DNA genômico de *A. duranensis*, *A. ipaënsis* e de *A. hypogaea* foram digeridos com 50 unidades da enzima de restrição *MboI* e de seu esteroisômero *Sau3AI* em um volume final de 150µL. Essas enzimas reconhecem o sítio GATC, mas *Sau3AI* é sensível à metilação na citosina dos sítios CG e CNG. As amostras foram precipitadas com 10% de acetato de sódio 3M e 2 volumes de etanol absoluto gelado em freezer por aproximadamente 16 horas. Foi feita a centrifugação a 14.000rpm por 5 minutos. Descartou-se o sobrenadante e foram feitas duas lavagens com etanol 70%, centrifugando-se a 14.000rpm por 5 minutos. O precipitado foi seco e ressuspensão em 20µL de água Milli-Q a 65°C por 15 minutos. As amostras foram separadas em gel de agarose 1% a 20V por aproximadamente 16 horas. Após a separação dos fragmentos, o gel foi tratado da seguinte maneira: 1) Adicionar solução de depurinação por 15 minutos com agitação leve constante. 2) Lavar com água destilada e adicionar solução de desnaturação por 30 minutos com agitação. 3) Lavar novamente em água destilada e adicionar solução de neutralização por 30 minutos com agitação. 4) Transferir os fragmentos para membrana Hybond-N⁺ por capilaridade em SSC 20X por aproximadamente 24 horas.

Para a pré-hibridização, a membrana foi colocada em garrafa de vidro contendo solução de pré-hibridização a 65°C com agitação lenta por aproximadamente 5 horas. A sonda foi amplificada por PCR do clone Ah-Athila-9 (um plasmídeo contendo o gene da RT) e cerca de 50ng utilizados para marcação utilizando-se o Kit *Ready-to-Go* d-CTP (Amersham, Pharmacia) juntamente com 5 µL de ³²P. A sonda foi incubada por 30 minutos a 37°C em banho-maria e então adicionada à garrafa contendo a membrana com a solução de pré-hibridização. A hibridização ocorreu a 65°C por 24 horas. Após a hibridização, as membranas foram lavadas da seguinte maneira: uma lavagem com solução I a 65°C por 15 minutos, duas lavagens com solução II por 15 minutos cada, a temperatura ambiente. A membrana foi seca e colocada para sensibilizar filme radiográfico (Kodak-Bio). O filme foi revelado da seguinte maneira: 2-5 minutos em revelador 1X (Dektol – Kodak), 1 minuto em água e 3 minutos em fixador 1X (Kodak). Ver soluções no ANEXO I – soluções utilizadas no Southern blot.

4.10 Extensão da seqüência

Para obter mais informação sobre a seqüência do retrotransposon, foram realizados experimentos utilizando o Kit Genome Walker (Clontech Laboratories, Inc.). Foram digeridos 1,5µg de DNA genômico de *A. hypogaea* com as enzimas de restrição *PvuII* e *DraI*. Para a ligação do adaptador aos fragmentos de DNA, foram utilizados: 4µL da digestão total, 25µM de adaptador, 10% de tampão e 0,5U de enzima T4 DNA ligase em um volume final de 8µL. A reação aconteceu a 16°C por 16 horas. Após ligação do adaptador aos terminais dos sítios de restrição, foi realizada uma reação de PCR utilizando-se iniciadores que amplificam-na para fora da seqüência da enzima transcriptase reversa e iniciadores complementares aos adaptadores da seguinte maneira: 3µM de iniciador específico; 2µM de iniciador adaptador; 10% de tampão; 0,6µL de reação de ligação; 2mM de dNTP; 75mM de MgCl₂, 1,5U de *Taq* polimerase em volume final de 15µL. O programa utilizado consistia de: 94°C por 15 segundos; 6 ciclos de 94°C por 2 segundos e 72°C por 3 minutos e 35 ciclos de 94°C por 2 segundos; 67°C por 3 minutos; e mais 4 minutos a 67°C ao final da reação. Os iniciadores utilizados e as reações feitas estão listados na Tabela 3.

Tabela 3. Combinação de iniciadores para extensão do retroelemento de amostras de DNA genômico de *A. hypogaea*, digeridas com as enzimas *DraI* e *PvuII* e os iniciadores desenhados para a seqüência de RT.

	Enzimas	
	<i>DraI</i>	<i>PvuII</i>
Iniciadores	Adapter I x Rep-B1-OUT-L	Adapter I x Rep-B1-OUT-L
	Adapter I x Rep-B1-OUT-R	Adapter I x Rep-B1-OUT-R

4.11 Identificação de clones em biblioteca BAC

De duas bibliotecas BAC, contendo cada uma cerca de 70 mil clones representativos de todo o genoma de *A. duranensis* e de *A. ipaënsis*, foram escolhidos 192 clones (96 de cada biblioteca) para uma varredura inicial para identificar a presença de retrotransposons contendo a mesma seqüência que codifica a enzima RT utilizada nos experimentos anteriores. O protocolo para o crescimento dos clones e extração do DNA foi de acordo com Heslop-Harrison (www.molcyt.com): 10µL de células em estoque foram incubadas em 5mL de meio

LB contendo cloranfenicol (12,5mg/mL) a 37°C com agitação rápida por até 24 horas. Centrifugou-se a 1500g por 10 minutos, descartando-se o sobrenadante em seguida. Adicionou-se 100µL de tampão de ressuspensão, misturou-se gentilmente, seguindo de incubação em gelo por 5 minutos. Foram adicionados 400µL de tampão de lise, misturando-se gentilmente e incubando-se em gelo por 5 minutos. Adicionaram-se 300µL de tampão de neutralização e incubou-se em gelo por 15 minutos. Centrifugou-se a 2800g por 15 minutos e transferiu-se o sobrenadante para tubo de microcentrífuga. Adicionaram-se 450µL de isopropanol e centrifugou-se a 12000g por 5 minutos. Descartou-se o sobrenadante e lavou-se o precipitado em etanol 70%, centrifugando-se em seguida a 12000g por 2 minutos. O precipitado foi seco ao ar e então ressuspensão em 40µL de água Milli-Q estéril. As amostras foram então tratadas com 20µg de RNaseA a 37°C por 30 minutos. As condições de reação da PCR foram as mesmas utilizadas para a amplificação do gene de RT do DNA genômico das três espécies (pg. 20).

5. Resultados

5.01 Extração de DNA vegetal

A extração do DNA genômico realizada de acordo com o protocolo modificado de Grattapaglia e Sederoff (1994) permitiu isolar grandes quantidades de DNA, em média 15µg por extração. Porém, mesmo com as modificações citadas, a qualidade dos DNAs extraídos algumas vezes não permitiam a sua utilização, pois eles apresentavam grandes quantidades de compostos fenólicos e polissacarídeos, principalmente para as extrações realizadas nos meses frios do ano. Existem protocolos específicos para a extração de DNA genômico de amendoim de ótima qualidade (Sharma *et al.*, 2000), mas os custos para cada extração seriam altos. Portanto, a melhor estratégia adotada foi a de extrair material genético de folhas jovens ainda não expandidas, principalmente durante os meses mais quentes e conservar folhas colhidas nesta época do ano em freezer -80°C para utilização ao longo do trabalho. Além disso, os cuidados durante o processo de extração são muito importantes, como misturar as soluções delicadamente, evitar o aquecimento das amostras de DNA desnecessariamente, pipetar o mínimo possível e não secar o precipitado em speed-vac, pois podem aumentar o risco de degradação do material genético.

5.02 Reação em cadeia da polimerase

A busca por seqüências homólogas à da enzima transcriptase reversa no banco de dados de *Arachis* permitiu encontrar uma seqüência completa do gene com alto índice de similaridade à de um retrotransposon LTR primeiramente caracterizado em *A. thaliana* (Pelissier *et al.*, 1995). Junto com os melhores resultados de índice de similaridade no GenBank, foi possível classificar o retrotransposon obtido como sendo um elemento tipo-Athila (*Athila-like*), pertencente à subordem *Retrotransposineae* e família *Metaviridae* (Ty3/Gypsy) (Hansen e Heslop-Harrison, 2004). Assim, foi possível desenhar os três pares de iniciadores para amplificação completa do gene (Figura 6).



Figura 6. Posição e orientação dos iniciadores desenhados para amplificação do gene da enzima transcriptase reversa. Obs: os tamanhos dos iniciadores não estão em escala.

A amplificação da seqüência de RT nas três espécies com os três pares de iniciadores gerou fragmentos de DNA com tamanhos variando entre 600pb e 1500pb (Figura 7). O produto da PCR foi então purificado em colunas S400 HR (Amersham Pharmacia Biotech) e utilizado para ligação em pGEM-T Easy. Após a transformação de células de *E. coli* por eletroporação foram selecionadas 32 colônias recombinantes para cada espécie, totalizando 96 colônias. A presença e tamanho do inserto foram conferidos após digestão do plasmídeo com a enzima de restrição *EcoRI* e eletroforese em gel de agarose 1%. Os clones que apresentaram insertos variando entre 800pb e 1200pb foram selecionados para seqüenciamento automático. Em alguns casos foi necessário repetir o seqüenciamento dos clones devido à baixa qualidade dos resultados. O iniciador Athila-internal-1 foi utilizado apenas para auxiliar o seqüenciamento dos insertos maiores. As 273 seqüências geradas foram manipuladas através do programa Pregap4 para remoção das partes de vetores, análise de qualidade das bases nucleotídicas através de PHRED e remoção de seqüências com baixo índice de confiança. Destas, 130 foram alinhadas pelo programa Gap4, possibilitando a identificação de 87 contigs - um para cada clone - com as leituras direta e reversa. Além disso, as seqüências dos iniciadores foram identificadas e utilizadas para correção da orientação dos contigs, quando necessária (Figura 8).

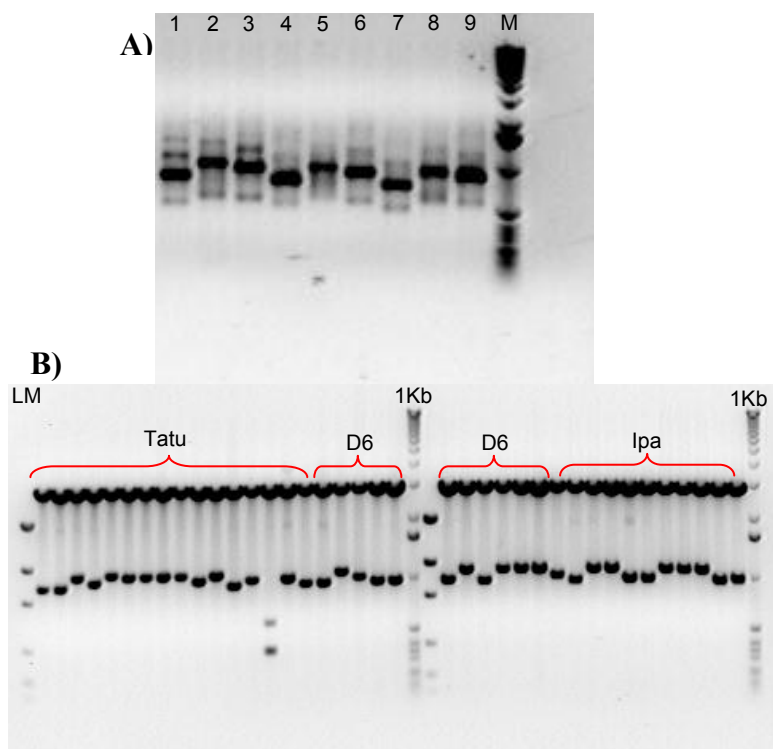


Figura 7. A) Amplificação da seqüência gênica da enzima transcriptase reversa em *A. hypogaea*. 1 (iniciadores F1 x R1), 2 (F1 x R2), 3 (F1 x R3), 4 (F2 x R1), 5 (F2 x R2), 6 (F2 x R3), 7 (F3 x R1), 8 (F2 x R2), 9 (F3 x R3). B) Digestão dos plasmídios mostrando os diferentes tamanhos de insertos obtidos. Tatu (*A. Hypogaea*), D6 (*A. duranensis*), Ipa (*A. ipaënsis*), M (1Kb DNA ladder, Invitrogen), LM (Low DNA Mass ladder, Invitrogen).

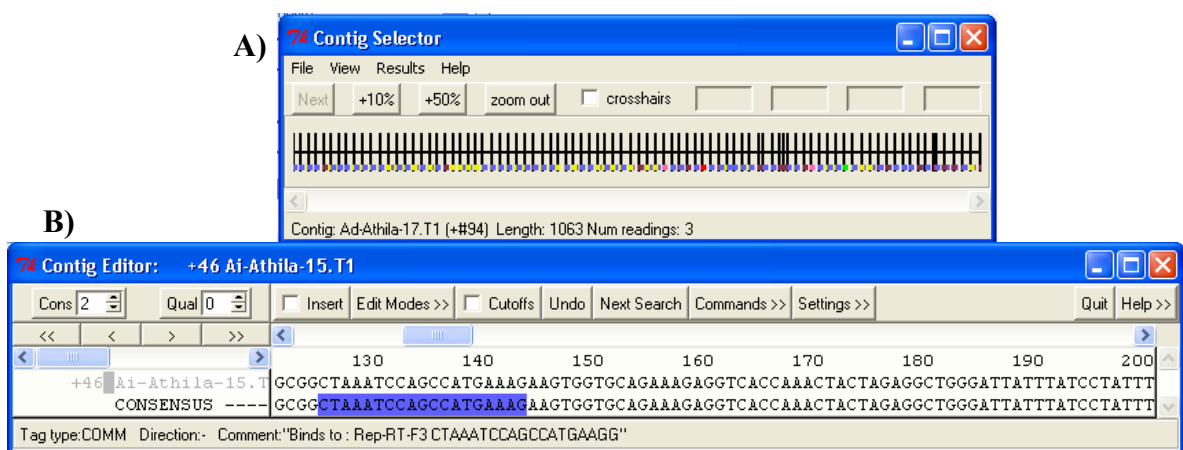


Figura 8. A) Mapa de todos os 87 contigs construídos mostrando as posições de cada par de iniciadores. Cada linha entre duas barras verticais corresponde a um contig. B) Exemplo de alinhamento de seqüências de um contig. Marcada em azul está a seqüência do iniciador Rep-RT-F3.

5.03 Análise Filogenética

Após a correção da orientação, as seqüências foram alinhadas em Clustal X e visualmente editadas no programa Jalview. O resultado do alinhamento foi então utilizado para análises estatísticas e construção de árvores filogenéticas utilizando-se o programa Mega, versão 3.1. Foram construídas árvores de similaridade utilizando-se as seqüências nucleotídicas e de aminoácidos. Dos 87 contigs obtidos, 80 foram utilizados no estudo de filogenia junto com o melhor resultado de BLASTx, a seqüência da transcriptase reversa de um retrotransposon-LTR de *Medicago truncatula* (GI:92891377), utilizada como grupo externo (Figura 9).

Ainda utilizando o programa Mega 3.1 foi possível calcular as distâncias genéticas entre as seqüências. A tabela 3 mostra os valores das distâncias genéticas para todas as seqüências utilizadas na análise filogenética. Foram montados três grupos, cada um contendo todas as seqüências de uma espécie. Foram analisadas as distâncias dentro de um mesmo grupo assim como entre os grupos para as seqüências de DNA e de aminoácidos. Para as análises de distância genética, foram descartadas as seqüências Ah-Athila-21 e Ai-Athila-7, que ficaram muito discrepantes de todas as outras na árvore de similaridade (Figura 9). Os dados mostram que os elementos do genoma de *A. duranensis* sofreram menos modificações, se comparados a *A. hypogaea* e *A. ipaënsis*, sendo que, esta última apresentou as maiores distâncias genéticas. As modificações sofridas nas seqüências de DNA refletiram-se com maior intensidade na cadeia polipeptídica, já que as distâncias praticamente dobraram de valor (Tabela 4), provavelmente por causa de mudanças no quadro de leitura que podem ter gerado uma seqüência de aminoácidos diferente.

Tabela 4. Distância genética entre as seqüências de DNA e de aminoácidos. Ah (*A. hypogaea*), Ad (*A. duranensis*) e Ai (*A. ipaënsis*). O valor médio corresponde a todas as 80 seqüências de DNA e/ou aminoácidos.

Grupos	Distância entre as seqüências (%)	
	Seq. DNA	Seq. Aminoácidos
Ah x Ah	8,68	18,83
Ad x Ad	8,20	17,44
Ai x Ai	8,95	19,87
Ah x Ad	8,44	18,13
Ah x Ai	9,09	19,88
Ad x Ai	9,30	20,14
Valor médio	8,86	19,16

A análise filogenética para as seqüências de nucleotídeos permitiu identificar cinco grupos principais. De acordo com o número de seqüências e do cálculo das distâncias entre elas foi possível determinar com qual genoma cada grupo apresenta melhor relacionamento (Tabela 5). O grupo 1 foi formado por 11 (57,9%) seqüências de *A. duranensis*, 7 (36,8%) de *A. hypogaea* e 1 (5,3%) de *A. ipaënsis*. O grupo 2 foi formado por 9 seqüências (64,2%) de *A. ipaënsis*, 4 (28,6%) de *A. hypogaea* e 1 (7,2%) de *A. duranensis*. O grupo 3 é composto de 5 seqüências (83,3%) de *A. ipaënsis* e 1 (16,7%) de *A. hypogaea*. O grupo 4 é composto de 12 seqüências (48%) de *A. duranensis*, 10 (40%) de *A. hypogaea* e 3 (12%) de *A. ipaënsis*. O grupo 5 foi formado por 8 seqüências (80%) de *A. ipaënsis* e 2 (20%) de *A. hypogaea*.

Tabela 5. Distância genética entre as seqüências de nucleotídeos de cada grupo obtido na árvore filogenética e o genoma relacionado. O valor médio foi obtido para todas as seqüências dentro de cada grupo. N/A indica a ausência de seqüências ou a presença de apenas uma seqüência.

Seqüências	Distância entre as seqüências (%)				
	Grupo1	Grupo2	Grupo3	Grupo4	Grupo5
Ah x Ah	4,88	6,05	n/a	6,45	4,76
Ad x Ad	5,82	n/a	n/a	5,91	n/a
Ai x Ai	n/a	5,71	8,69	5,83	6,82
Ah x Ad	5,07	6,34	n/a	6,15	n/a
Ah x Ai	5,37	5,79	8,20	6,18	6,18
Ad x Ai	5,39	6,56	n/a	5,93	n/a
Valor médio	5,45	5,87	8,52	6,12	6,55
Genoma	AA	BB	BB	AA	BB

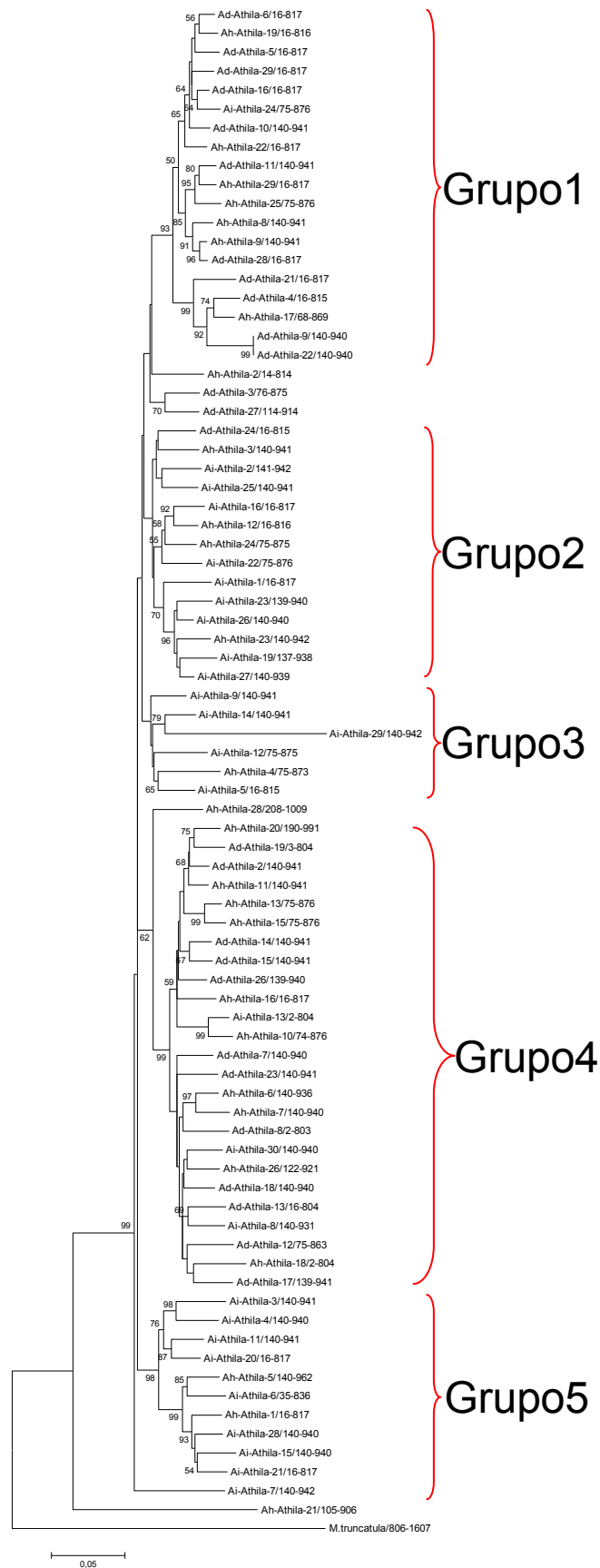


Figura 9. Relações filogenéticas obtidas a partir das seqüências de nucleotídeos do gene de RT. Os resultados de bootstrap são indicados para valores acima de 50 (1000 réplicas). *M. truncatula* foi utilizado como grupo externo.

A análise filogenética obtida a partir das seqüências deduzidas de aminoácidos permitiu identificar 5 grupos principais (Figura 10). Assim como na análise anterior, a distância entre as seqüências das espécies foi calculada para cada grupo (Tabela 6). O grupo 1 foi formado por 11 (57,9%) seqüências de *A. duranensis*, 7 (36,8%) de *A. hypogaea* e 1 (5,3%) de *A. ipaënsis*. O grupo 2 foi formado por 9 seqüências (64,2%) de *A. ipaënsis*, 4 (28,6%) de *A. hypogaea* e 1 (7,2%) de *A. duranensis*. O grupo 3 é composto de 5 seqüências (83,3%) de *A. ipaënsis* e 1 (16,7%) de *A. hypogaea*. O grupo 4 é composto de 12 seqüências (48%) de *A. duranensis*, 10 (40%) de *A. hypogaea* e 3 (12%) de *A. ipaënsis*. O grupo 5 foi formado por 8 seqüências (80%) de *A. ipaënsis* e 2 (20%) de *A. hypogaea*.

Tabela 6. Distância genética entre as seqüências de aminoácidos de cada grupo obtido na árvore filogenética e o genoma relacionado. O valor médio foi obtido para todas as seqüências dentro de cada grupo. N/A indica a ausência de seqüências ou a presença de apenas uma seqüência.

Seqüências	Distância entre as seqüências (%)				
	Grupo1	Grupo2	Grupo3	Grupo4	Grupo5
Ah x Ah	11,33	11,55	n/a	13,1	9,36
Ad x Ad	11,98	n/a	n/a	10,87	n/a
Ai x Ai	n/a	13,58	14,96	10,94	14,33
Ah x Ad	11,06	12,06	n/a	11,91	n/a
Ah x Ai	11,84	12,45	14,65	11,95	12,87
Ad x Ai	10,67	13,69	n/a	10,96	n/a
Valor médio	11,71	12,95	14,85	11,74	13,7
Genoma	AA	BB	BB	AA	BB

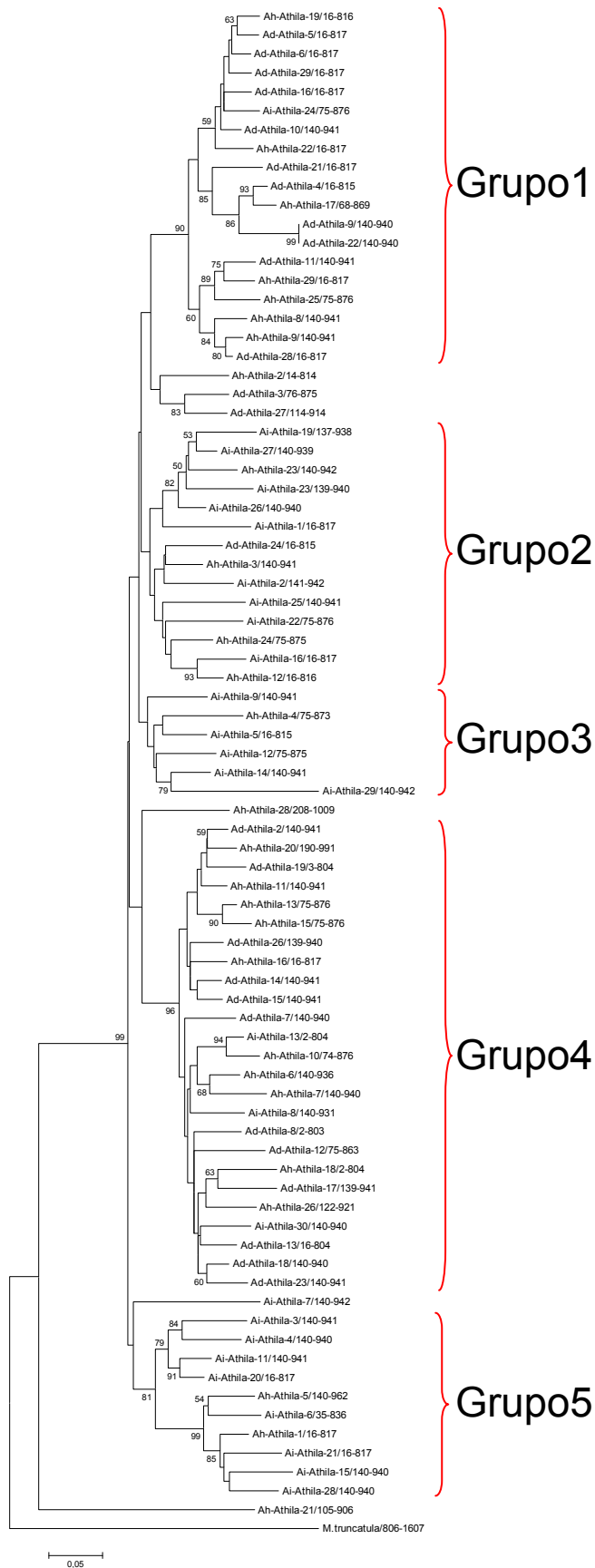


Figura 10. Relações filogenéticas obtidas a partir das seqüências de aminoácidos do gene de RT. Os resultados de bootstrap são indicados para valores acima de 50 (1000 réplicas). *M. truncatula* foi utilizado como grupo externo.

Também foram calculados os números de substituições sinônimas e não-sinônimas entre os grupos. Sabe-se que mutações ocorridas, principalmente na terceira base do códon, na maioria das vezes, não muda o padrão de leitura da fita de mRNA. Isso acontece devido ao código genético ser degenerado e, dessa forma, um gene de extrema importância para um organismo sofre menor pressão seletiva das taxas de mutações sinônimas porque elas evitam a mudança na síntese protéica (Nei e Kumar, 2000). Porém, em alguns casos e principalmente no caso dos retrotransposons, as mutações não-sinônimas ocorridas podem não ser eliminadas por seleção, gerando um espectro de diversidade. Os dados obtidos podem ser vistos na Tabela 7, da qual pode-se inferir que os genomas AA e BB das espécies silvestres evoluíram diferentemente.

Tabela 7. Taxas de substituições sinônimas (dS) e não-sinônimas (dN) dentro e entre os grupos.

Seqüências	dS (%)	dN (%)
Ah x Ah	7,34	8,51
Ad x Ad	6,60	7,81
Ai x Ai	7,73	9,17
Ah x Ad	6,99	8,17
Ah x Ai	7,80	9,08
Ad x Ai	7,75	9,16
Valor médio	7,42	8,70

Visando obter melhor conhecimento a respeito da evolução desses elementos, valores de dS e dN foram utilizados na construção de histogramas para determinar em quais espécies o retrotransposon sofreu modificações recentes ou em quais ele vem acumulando mais mutações ao longo do tempo. Para que não houvesse tendência de agrupar as seqüências apenas com o genoma AA ou o BB, optou-se pela construção de uma seqüência consenso para o grupo de *A. duranensis* e uma para *A. ipaënsis*, ao invés de um consenso de todos os 80 contigs. Novamente as taxas de dS e dN entre as seqüências das três espécies e a seqüência consenso foram calculadas. Os dados foram então distribuídos no histograma, de acordo com a sua freqüência (Figura 11). Os valores obtidos para os elementos isolados de *A. duranensis* agruparam-se entre 0,01 e 0,06; enquanto os de *A. ipaënsis* mostraram uma distribuição homogênea, mas também com valores muito altos. Os dados obtidos para *A. hypogaea* ficaram entre os valores para as espécies silvestres.

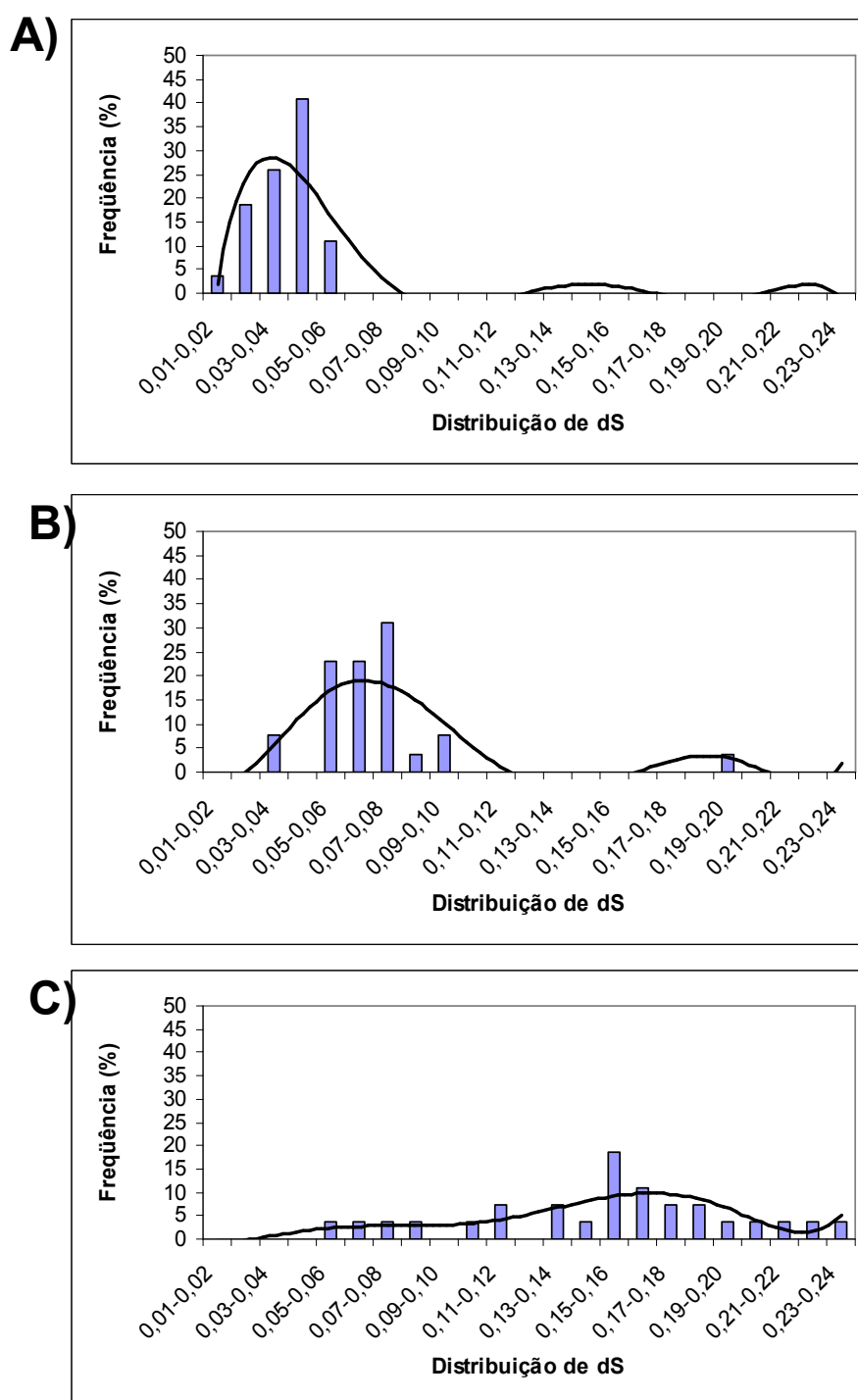


Figura 11. Distribuição dos valores de dS apresentados pelas 80 seqüências dos contigs em relação à seqüência consenso. A) Valores obtidos para *A. duranensis*, B) Valores de *A. hypogaea* e C) Valores de *A. ipaënsis*. A linha de tendência foi calculada utilizando-se a função polinomial.

Devido ao baixo número de seqüências para cada grupo, optou-se por utilizar os valores de dS em conjunto com os valores de dN para aumentar a quantidade de dados. Não foi possível observar grandes mudanças na topologia dos histogramas, apenas um leve deslocamento dos valores sobre o eixo das distribuições ou uma maior homogeneização das curvas da linha de tendência (Figura 12).

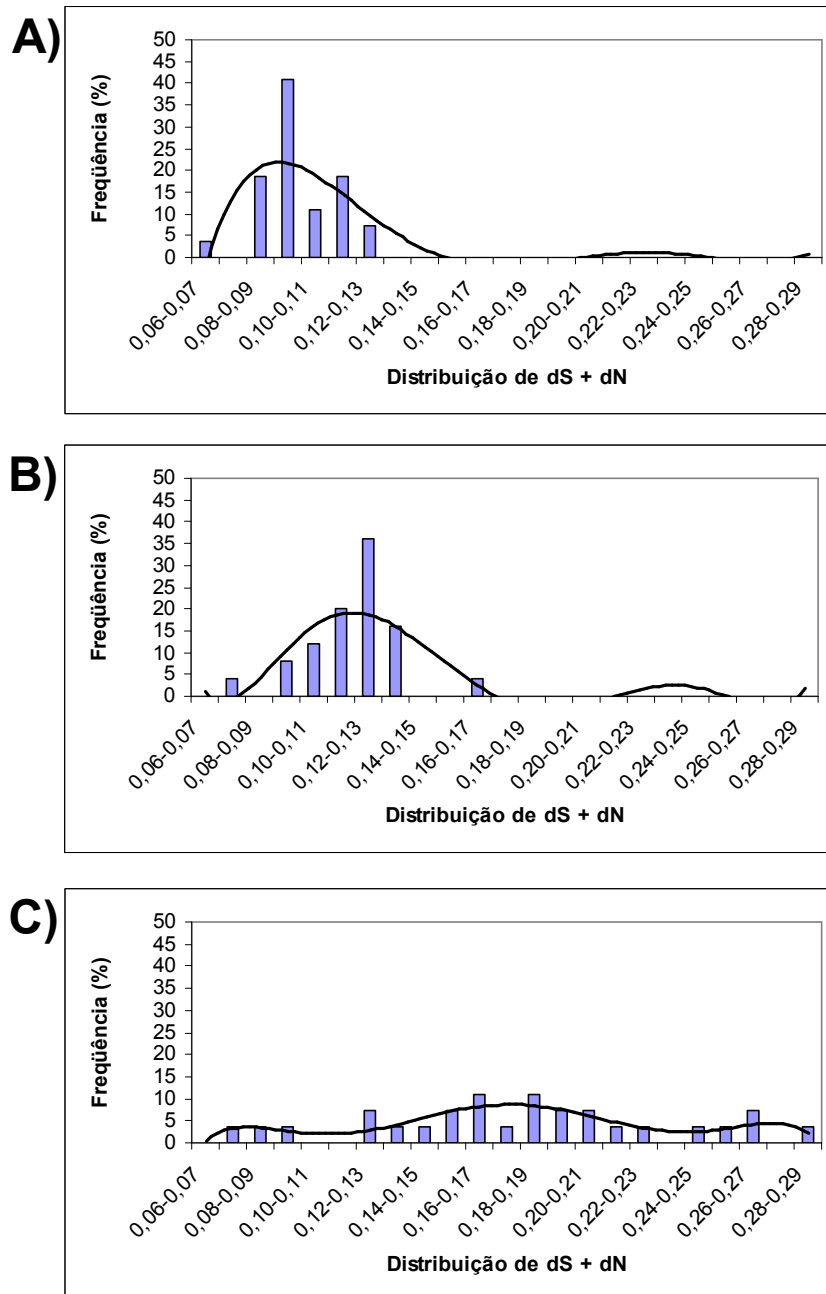


Figura 12. Distribuição dos valores de dS + dN apresentados pelas seqüências dos contigs em relação à seqüência consenso. A) Valores obtidos para *A. duranensis*, B) Valores de *A. hypogaea* e C) Valores de *A. ipaënsis*. A linha de tendência foi calculada utilizando-se a função polinomial.

Em um segundo momento, uma nova análise filogenética foi realizada, desta vez utilizando-se várias seqüências de elementos Athila (*Athila-like elements*), disponíveis no site <http://www.public.iastate.edu/~voytas/>. As seqüências pertencem a retrotransposons isolados de vários grupos vegetais, gramíneas, leguminosas, solanáceas e brassicáceas. Foi adicionada ao estudo, a seqüência consenso do elemento, chamada de Arachis, um representante de cada grupo obtido na árvore de similaridade para as seqüências de aminoácidos (Ad-Athila-16, Ah-Athila-10, Ai-Athila-25, Ah-Athila-1, Ai-Athla-12) e também a seqüência de *M. truncatula* utilizada nas análises anteriores. Essas seqüências foram traduzidas pela ferramenta EXPASy para a escolha do quadro de leitura (*frameshift*) correto e depois alinhadas. Como grupo externo, foram utilizadas as seqüências da linhagem TAT, como foi feito por Wright e Voytas (2001). As seqüências de *Arachis* agruparam-se com o grupo das leguminosas, mas formaram um subgrupo junto com seqüências de *M. truncatula*, ervilha e algumas de fava. As demais espécies agruparam-se de acordo com o grupo a que pertencem (Figura 13).

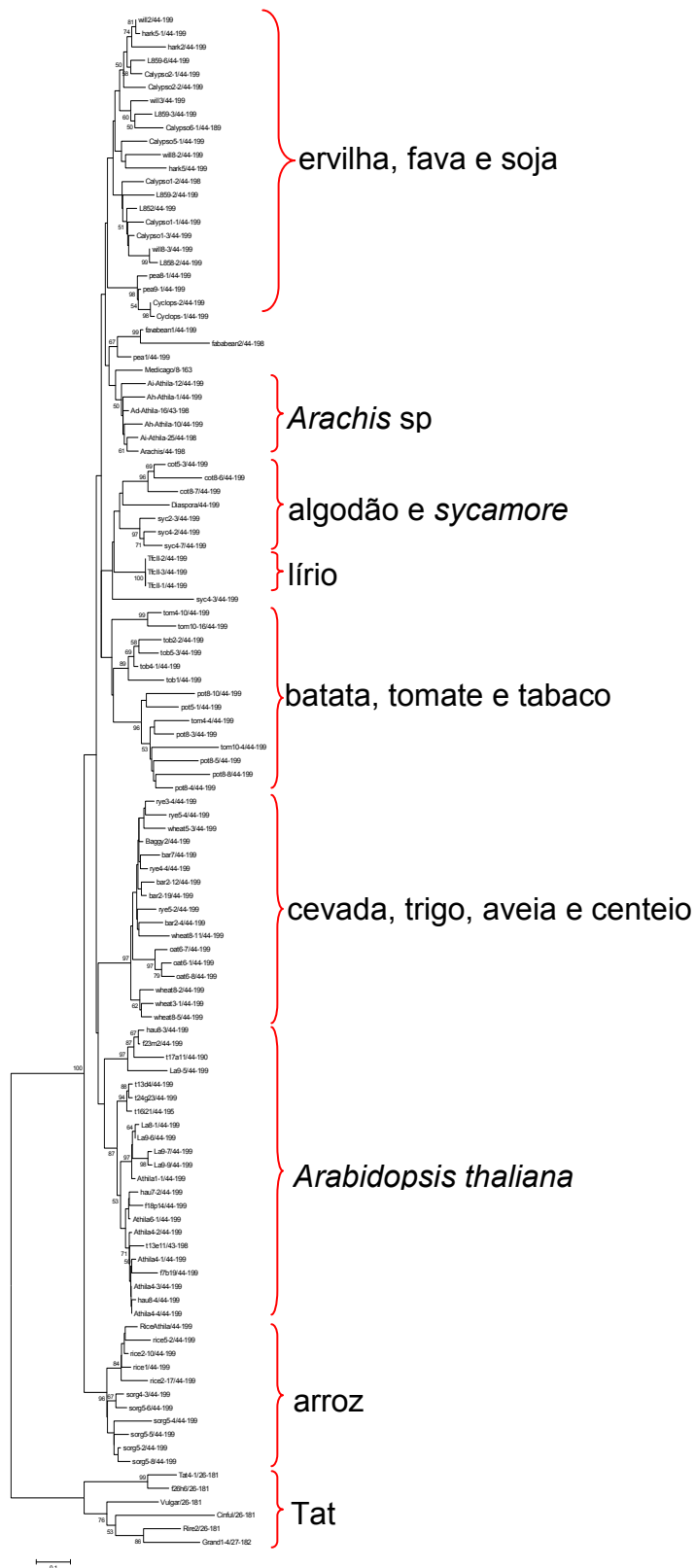
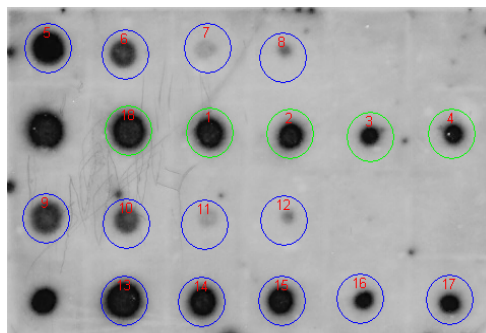


Figura 13. Relações filogenéticas entre diversos grupos de elementos do tipo Athila. Resultados de bootstrap estão indicados para valores acima de 50 (1000 replicações).

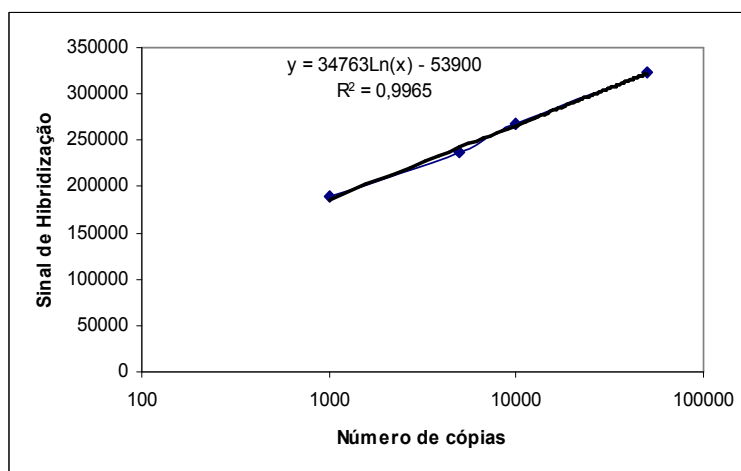
5.04 Número de cópias

Os sinais de hibridização obtidos, podem ser vistos nas Figuras (14 A, 15 A e 16 A) Em alguns casos, não foi possível a comparação direta entre os sinais gerados para o plasmídeo controle e as amostras, tendo em vista que pode ter havido uma saturação da emissão do sinal no filme radiográfico. Dessa forma, os dados numéricos obtidos pelo programa MultiGauge foram utilizados para a construção de gráficos de dispersão (Figura 14 B e C, figura 15 B e C, figura 16 B e C) e os dados que ficaram muito discrepantes foram eliminados da análise. Foram construídos gráficos para cada série de diluições e então desenhada uma linha de tendência onde a fórmula obtida foi utilizada para calcular o valor do número de cópias do elemento em cada genoma. Para o genoma de *A. hypogaea* foi calculado um número de cópias de $4893 \pm 1281,6$ por genoma diplóide. Para o genoma de *A. duranensis* calculou-se um número de $3122,8 \pm 2382,5$ por genoma diplóide. Em *A. ipaënsis* calculou-se um número de 829 ± 143 por genoma diplóide.

A)



B)



C)

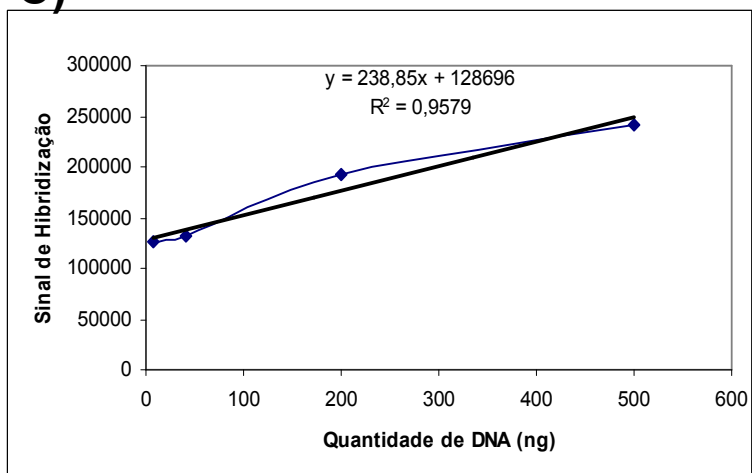
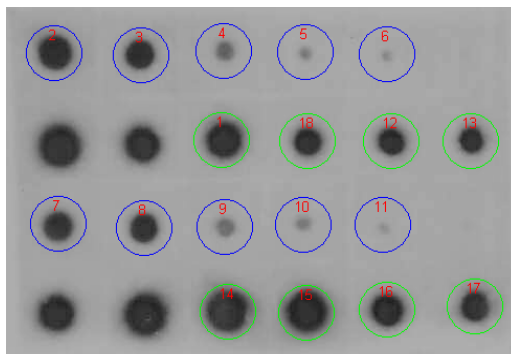
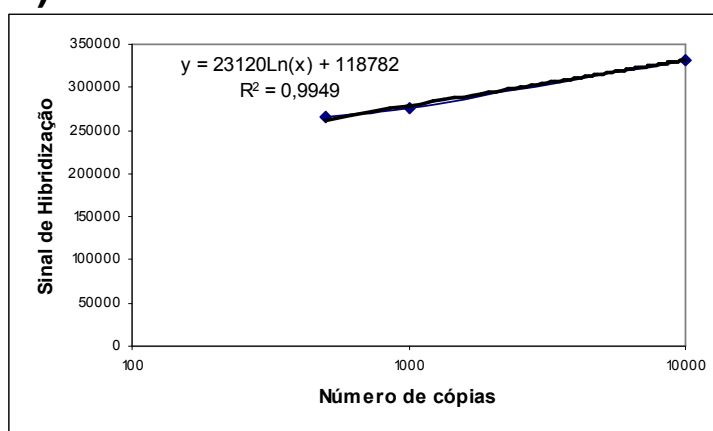


Figura 14. Resultados das hibridizações de DNA genômico utilizando o clone Ah-Athila-9 contendo o gene de RT em *A. hypogaea*. A) Dot blot (Linhas 1 e 3: DNA genômico, linhas 2 e 4: Controle), B) Curva de calibragem e C) Curva de DNA genômico.

A)



B)



C)

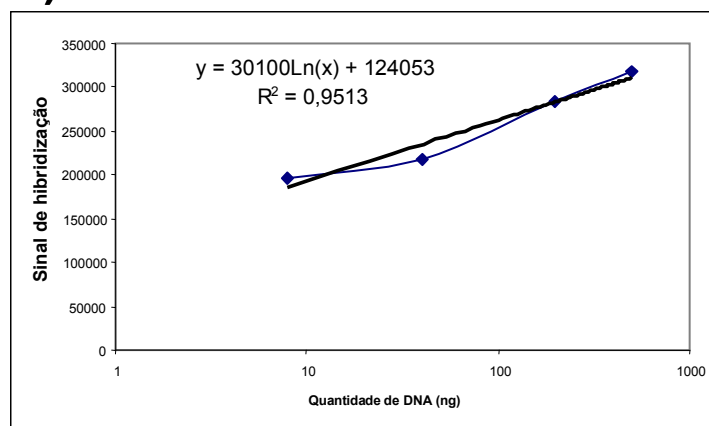
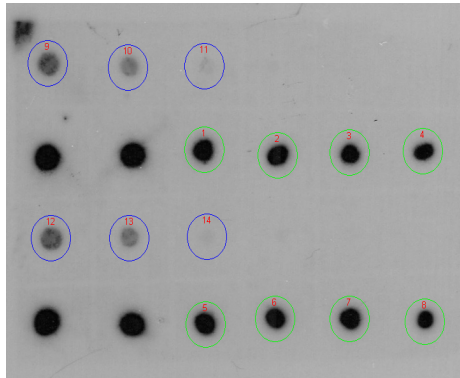
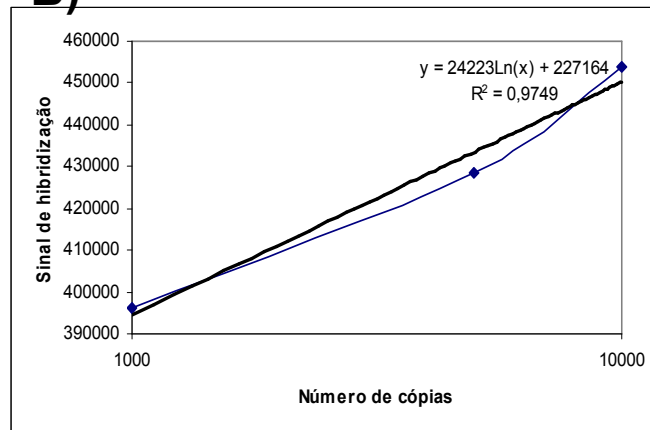


Figura 15. Resultados das hibridizações de DNA genômico utilizando o clone Ah-Athila-9 contendo o gene de RT em *A. duranensis*. A) Dot blot (azul=DNA genômico, verde=Controle), B) Curva de calibragem e C) Curva de DNA genômico.

A)



B)



C)

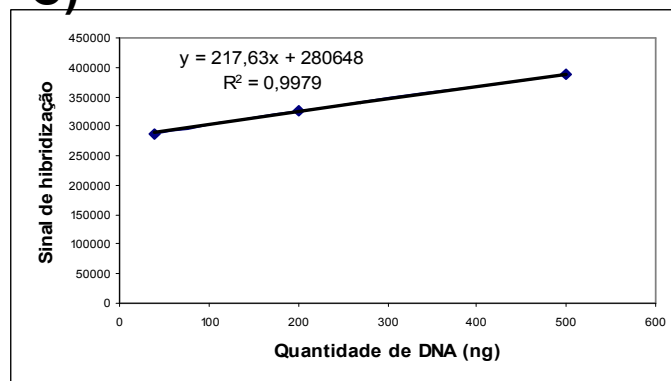


Figura 16. Resultados das hibridizações de DNA genômico utilizando o clone Ah-Athila-9 contendo o gene de RT em *A. ipaënsis*. A) Dot blot (azul=DNA genômico, verde=Controle), B) Curva de calibragem e C) Curva de DNA genômico.

A tragem inicial realizada nas bibliotecas BAC através de PCR mostrou uma evidente diferença no número de cópias desse elemento nos genomas AA e BB. De uma placa escolhida aleatoriamente contendo 96 clones BAC de *A. duranensis*, 36 seqüências amplificaram o gene da enzima transcriptase reversa. Para a placa contendo 96 clones BAC de *A. ipaënsis*, apenas 10 amplificaram a mesma seqüência. Além desses experimentos, para o cálculo do número de cópias foram feitas buscas contra todas as seqüências de *Arachis* publicadas no GenBank (Genome Survey Sequence database) utilizando a seqüência do clone Ah-Athila-9. Foram encontradas similaridades com 13 seqüências (0,45%) de *A. duranensis* de um total de 2905; 7 (0,15%) de *A. hypogaea* de um total de 4590 e 1 (0,03%) de *A. batizocoi* (genoma BB) de um total de 3276, todas elas com índices de similaridades (E-value) $\leq 1e^{-20}$. Não foram realizadas buscas contra *A. ipaënsis* por não haver seqüências depositadas no GenBank.

5.05 Análise de Metilação

Foi realizada uma análise da freqüência de digestão das enzimas de restrição *MboI* e *Sau3AI* sobre o clone Ah-Athila-9 e observou-se que esse inserto era cortado aproximadamente 3 vezes. Em seguida cerca de 5µg de DNA genômico das três espécies em questão foram digeridos com as mesmas enzimas, sendo necessária a utilização de 10U por µg de amostra. A Figura 17 representa um gel de agarose 1% contendo amostras de *A. hypogaea* (cultivar Tatu), *A. duranensis* e *A. ipaënsis*. Não foi possível identificar nenhuma banda específica sob luz ultravioleta. Os fragmentos foram transferidos para membrana e então hibridizados com a sonda contendo o clone Ah-Athila-9. As bandas que surgiram apresentaram tamanhos entre aproximadamente 40pb e 800pb (Figura 18). O Southern blot mostrou que, para esse par de enzimas, o retroelemento não está metilado em nenhuma das três espécies. No caso de *A. ipaënsis* um dos fragmentos apareceu com menor intensidade, mas apresentou o mesmo tamanho visto em outras digestões.

Comparando-se o resultado desse Southern-blot com outro obtido por nós em 2005 (dados não publicados), em trabalho utilizando um diferente elemento repetitivo, foi possível observar que não há metilação nos genomas de *A. ipaënsis* e do anfidiplóide sintético obtido pelo cruzamento de *A. ipaënsis* e *A. duranensis* (Apêndice) (Figura 26). Esse elemento possui um padrão de distribuição semelhante ao do retrotransposon discutido neste trabalho (Apêndice). Além de possuir uma seqüência muito parecida com a de uma região LTR,

acredita-se que este elemento seja um retrotransposon muito similar ao Athila, se não for o mesmo (Stephan Nielen, comunicado pessoal), o que reforça o resultado do Southern blot.

Em uma nova análise, foi realizada uma busca por homologia do clone Ah-Athila-9 contra todas as seqüências de ESTs de *Arachis* do nosso banco de dados contendo cerca de 8 mil clones e também no GenBank, contendo cerca de 12 mil seqüências, mas não foram encontradas quaisquer similaridades significativas. Esse resultado sugere a ausência ou uma quantidade muito baixa de cópias de mRNA desse elemento.

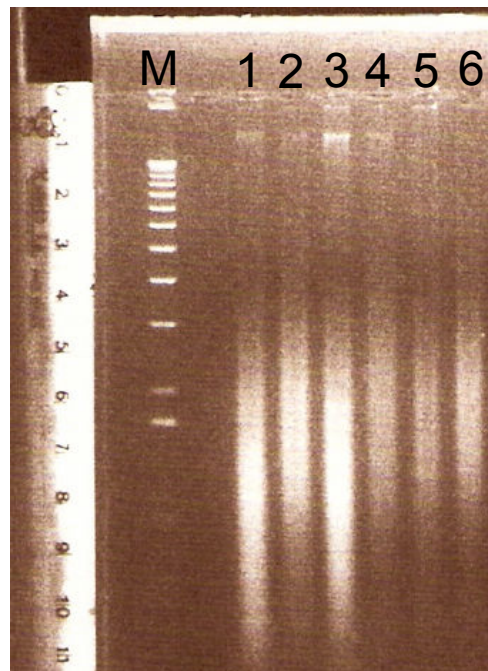


Figura 17. Gel de agarose 1% mostrando o padrão de digestão dos DNAs genômicos. M: 1Kb Plus ladder, 1: *A. hypogaea* (*Mbo*I), 2: *A. hypogaea* (*Sau*3AI), 3: *A. duranensis* (*Mbo*I), 4: *A. duranensis* (*Sau*3AI), 5: *A. ipaënsis* (*Mbo*I), 6: *A. ipaënsis* (*Sau*3AI).

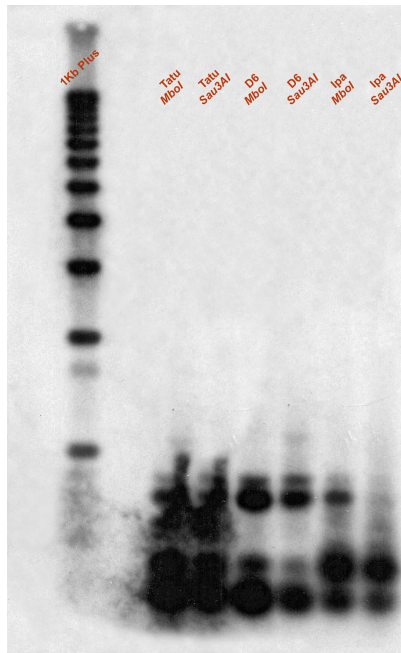


Figura 18. Análise de metilação através de Southern blot. Tatu (*A. hypogaea*), D6 (*A. duranensis*) e Ipa (*A. ipaënsis*). Foi utilizado como sonda o clone Ah-Athila-9.

5.06 Extensão da seqüência

Para a extensão da seqüência do retroelemento, o resultado da amplificação utilizando o Kit Genome Walker (Figura 19 A) foi então utilizado para ligação em pGEM-TEasy. Após transformação de células competentes por eletroporação, a presença e tamanho dos insertos foram conferidos em gel de agarose 1% após digestão dos plasmídios com *EcoRI* (Figura 19 B). Os clones que apresentaram insertos com tamanhos maiores do que 1Kb foram utilizados para seqüenciamento automático utilizando-se os iniciadores universais T7 e SP6. De um total de 24 clones obtidos, 12 apresentavam insertos acima de 1Kb e foram utilizados para seqüenciamento em ambos os sentidos. Das 24 seqüências geradas, 15 mostraram similaridade com o gene de RT ou com a região GAG (group specific antigen) no GenBank. Dessas amostras, 8 alinharam-se para compor a seqüência do retrotransposon, ampliando-a para aproximadamente 3,6Kb, cerca de 30% do tamanho médio do elemento Athila de *A. thaliana*.

Aplicando a seqüência total obtida para o retrotransposon no GenBank foi possível determinar a região completa do gene de RT com aproximadamente 200 aminoácidos e também uma seqüência de aproximadamente 250 aminoácidos muito similar à de *M.*

truncatula, mas que não codifica nenhuma proteína. Além disso, foi possível identificar o gene da enzima RnaseH e o gene da enzima Integrase. O restante da seqüência do retrotransposon não apresentou nenhuma similaridade significativa (Figura 20).

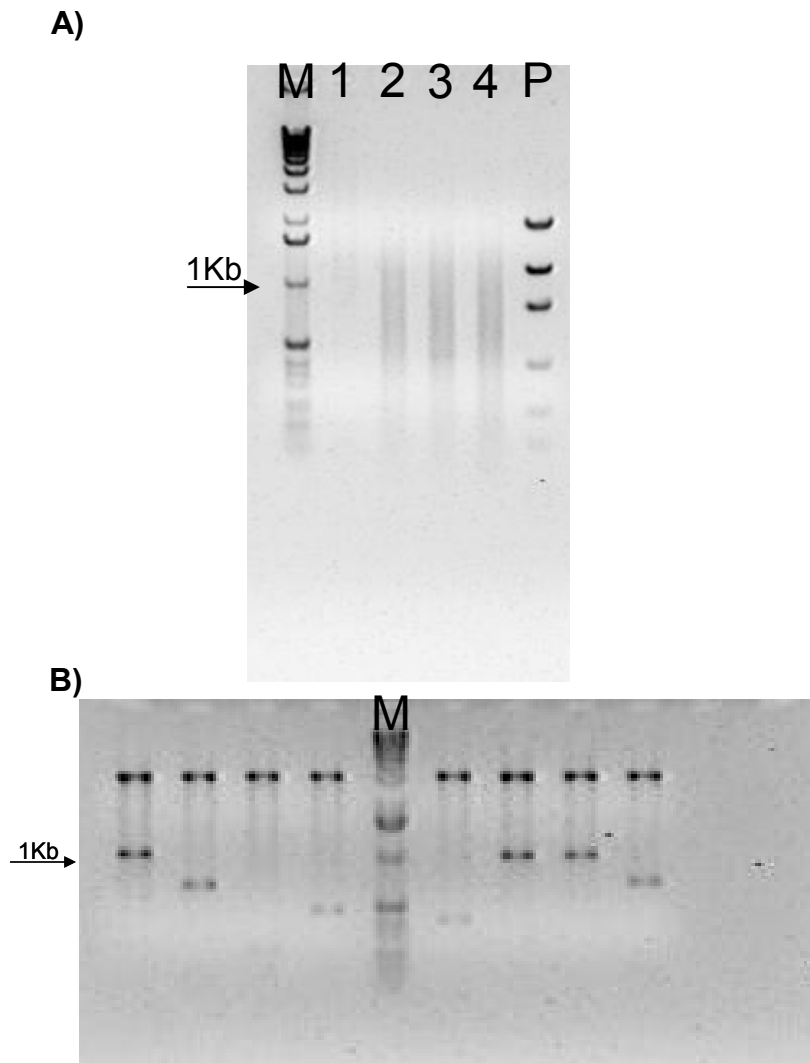


Figura 19. A) Resultado da amplificação para o genome walker, 1: *PvuII* + Rep-B1-OUT-L, 2: *PvuII* + Rep-B1-OUT-R, 3: *DraI* +Rep-B1-OUT-L, 4: *DraI* + Rep-B1-OUT-R. B)Exemplo de análise da presença e tamanho do inserto de clones obtidos de *A. hypogaea* para a extensão da seqüência do retroelemento. M (1kb ladder), P (Low mass ladder).

A)

XXXXXGAQKLIDTVARNQHMYLSSSESSMKEEAKVVSTELNPPEQVVELNQQLFITKLAEFKE
 MLQDTKKANKNMEAQLNQTRQQLSKQITEECQAVQIKEWEDIECLNSKQKAKKGTDRGPT
 HYPKSLGQEPREEFWHSNARKGKIGIKRPTHIQLWRSNARICRRHSNPGVQQGIRHLQVLITPLS
 RLLQPLLEVNQLQLRLKNIKPRCLILRNSAKRIGTICPLCRLSHDSNKDSVCRGTANTLLFVHERD
 LKPEGLERNKFFPHRMQCSHSQKLTKEAGSWKLYDTMHIRGLYQDSSMSWGKYQPNTCIHYQ
 KAWFDRSTNPDMPPTCSIKIPSGVIEDMIVKVGPFAPPTNFVLEMEEHKSATLILGRPFLATGRT
 LINVQKGEVTLRVNEDEFKLNKVDKMQHPDTSKDCMSIDIHDSLVEEVNMTESLKSELEDIFKDV
 QPDLEEPEKTEEPLKTPQEEENPPKPELKPLPPSLKYAFLGEDDIFSVIISSALNPLEEEALIQLVRT
 HKTALGWSISDLKGISPAKCTHKILLEDDAKPVVQPQRRLNPAMKEVVQKEVTKLLEAGHIYPIS
 DSPWVSPVQVVPKGGMTVVHNEKNELVPTRTVTGWRCIDYRRLNTATRKHDFPLPFIDQM
 LERLVGHEYYCFLDGYSYGNQIAVDPDQEKIAFTCPGVSFAFRRMPFGLCNAPVTFQRCLMSIF
 SDIVEKFLEVFMDDFSUYGDSFNLSCLNHLALVLKRCQETNLVLNWEKCHFMVTEGIVLGHKIS
 NKGIEVDQAKVEVIEKLPFANVKAIRSFLGHAFYRRFIKDFSKIARPLSNLLAADTPFMFDTEFL
 QAFETLKAKLVTAPVISSPDWTLPFELMCDASDHAIGAVLQGRHNKLLHVIIYASRVLNDAAQK
 NYTTTEKELLAVVYAIKFRSYLVGSKVIVYTDHAALKYLLTKQDSKPRLIRWVLLLQEFDIEIR
 DRKGTENQVADHLSQIEPVAGASLPSTEISETFPDEQLFAIQEAPWFADIANYSKDVRFIPQEYSRV
 QRKKLISDAKYLLWDEPYLFKRCADGMIRRCVPREEAQRILWHCHGSQYGGHFGSERTATKVL
 QCGFYWPTLYRDAREFVRNCDSCQRAGNLPHGYAMPQQGILEIELFDVWVIDFMGPFPRPSYSN
 TYILVAIDYVSKWVEAIATHNNDTKTVNSSRNTSSAGLVFPEYS

- Seqüência similar a de *M. truncatula*
- Transcriptase reversa
- RNaseH
- Integrase

B)

Elemento Athila de *A. thaliana*

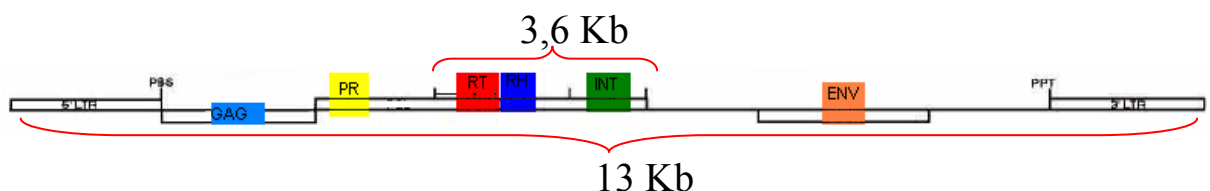


Figura 20. A) Seqüência traduzida do retroelemento, após extensão utilizando o Kit Genome Walker, em negrito estão os motivos conservados de cada gene. B) Tamanho e composição de um retrotransposon do tipo Athila de *A. thaliana*, adaptado: Hansen e Heslop-Harrison, 2004.

6. Discussão

Os retrotransposons compõem grande parte dos genomas vegetais, em alguns casos podem constituir cerca de 70% ou mais (Kumar e Bennetzen, 1999), dependendo da espécie. Sua contribuição principal está relacionada com o aumento do tamanho do genoma, rearranjos cromossômicos, silenciamento e duplicação gênica, podendo em alguns casos gerar especiação (Heslop-Harrison, 2000; Bennetzen, 2000; Kazazian Jr., 2004; Melayah *et al.*, 2004; Raina *et al.*, 2005). Os estudos realizados com esses elementos visam principalmente determinar sua evolução dentro dos genomas de diferentes espécies e sua contribuição para a evolução das mesmas. Pouco se sabe a respeito das suas taxas de amplificação ou preferência por sítios de inserção, mas acredita-se que em breve esse conhecimento fornecerá uma oportunidade de utilização mais intensa desses elementos como ferramentas de estudos genéticos. A utilização da enzima transcriptase reversa para os estudos deve-se ao fato de que sua função é altamente conservada e de grande importância para esses elementos, embora a sequência nucleotídica nem sempre permaneça a mesma (Dixit *et al.*, 2006). Além disso, vários trabalhos publicados utilizaram essa mesma metodologia, o que permite uma melhor comparação entre diferentes grupos de espécies.

O amendoim cultivado é uma planta que apresenta grandes oportunidades para o estudo de retrotransposons principalmente pela sua origem alotetraplóide complexa, com dois genomas diferentes, o que pode estimular a ativação desses elementos, como observado em outras espécies (Liu e Wendel, 2000; Zhao *et al.*, 1998). Este é o primeiro estudo da organização de um retrotransposon e sua contribuição na formação dos genomas de espécies silvestres e cultivadas de *Arachis*. Acredita-se que a utilização das prováveis espécies parentais *A. duranensis* e *A. ipaënsis*, junto com *A. hypogaea*, forneça um bom modelo para investigação da evolução dos genomas antes e após a diferenciação das espécies silvestres a partir do ancestral comum, além de permitir identificar em que momento esses genomas se encontravam antes do surgimento do amendoim cultivado.

Neste trabalho foram isoladas e caracterizadas seqüências de um retrotransposon contendo o gene da enzima transcriptase reversa, obtido em trabalhos anteriores de seqüenciamento de bibliotecas genômicas. Com esses dados foram feitas análises filogenéticas e estatísticas que permitiram contar a história desse elemento no genoma das espécies parentais e do amendoim cultivado.

6.01 Distribuição do elemento nos genomas

Após a obtenção do elemento repetitivo nos resultados do seqüenciamento de bibliotecas genômicas, foi feita uma busca por similaridades no GenBank e descobriu-se que a seqüência consistia no gene da enzima transcriptase reversa de um retrotransposon LTR do tipo Athila, primeiramente caracterizado em *A. thaliana*. Ensaio de hibridização *in situ* com fluorescência, realizados por Seijo e colaboradores (2005 – dados não publicados), permitiram distinguir um padrão diferenciado de distribuição desse elemento nos genomas AA e BB de *A. hypogaea* (Figura 21). Observou-se que essa seqüência apresentava uma maior hibridização no genoma AA e uma baixa hibridização no genoma BB. Além disso, observa-se que esse elemento localiza-se preferencialmente em regiões de eucromatina. De posse desses dados, foram desenhados iniciadores para amplificação do gene de RT. Das 273 seqüências geradas, 130 (48%) apresentaram qualidade suficiente e alinharam-se nos sentidos direto e reverso para comporem 87 contigs. Desses 87, 80 (91%) foram alinhados entre si e sua seqüência consenso foi utilizada para construção das árvores de similaridade.

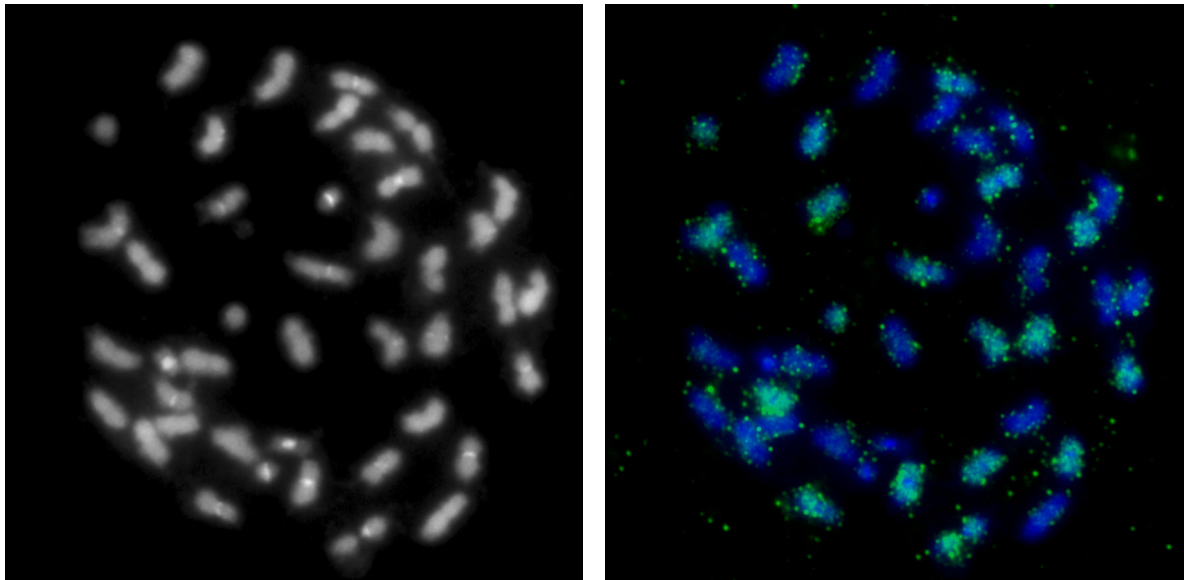


Figura 21. À esquerda, hibridização com DAPI de uma célula contendo 40 cromossomos de *A. hypogaea*. Notar a presença de 20 cromossomos com banda centromérica (genoma AA) e 20 com ausência de qualquer banda (genoma BB). À direita, hibridização da mesma célula com o elemento repetitivo em verde. Cortesia: Seijo *et al.*, 2005 (dados não publicados).

Aplicando a seqüência deduzida de aminoácidos do clone Ah-Athila-9 contra o banco de dados de *M. truncatula* (www.medicago.org/genome/cvit_blast.php) foi possível observar o padrão de dispersão do retrotransposon no genoma dessa espécie (Figura 22). Comparando esses resultados com um gráfico comparativo que mostra a colinearidade entre os cromossomos de *M. truncatula* e *L. japonicus* (Figura 23), foi possível observar um possível evento de rearranjo cromossômico, onde os locais de maior presença do elemento nos cromossomos de *M. truncatula* correspondem às regiões em que não há homologia com os cromossomos de *L. japonicus*. Esses resultados permitem inferir que esses elementos tiveram participação importante na formação do genoma de *Medicago* e possivelmente o mesmo pode ter ocorrido para os genomas das espécies de *Arachis*.

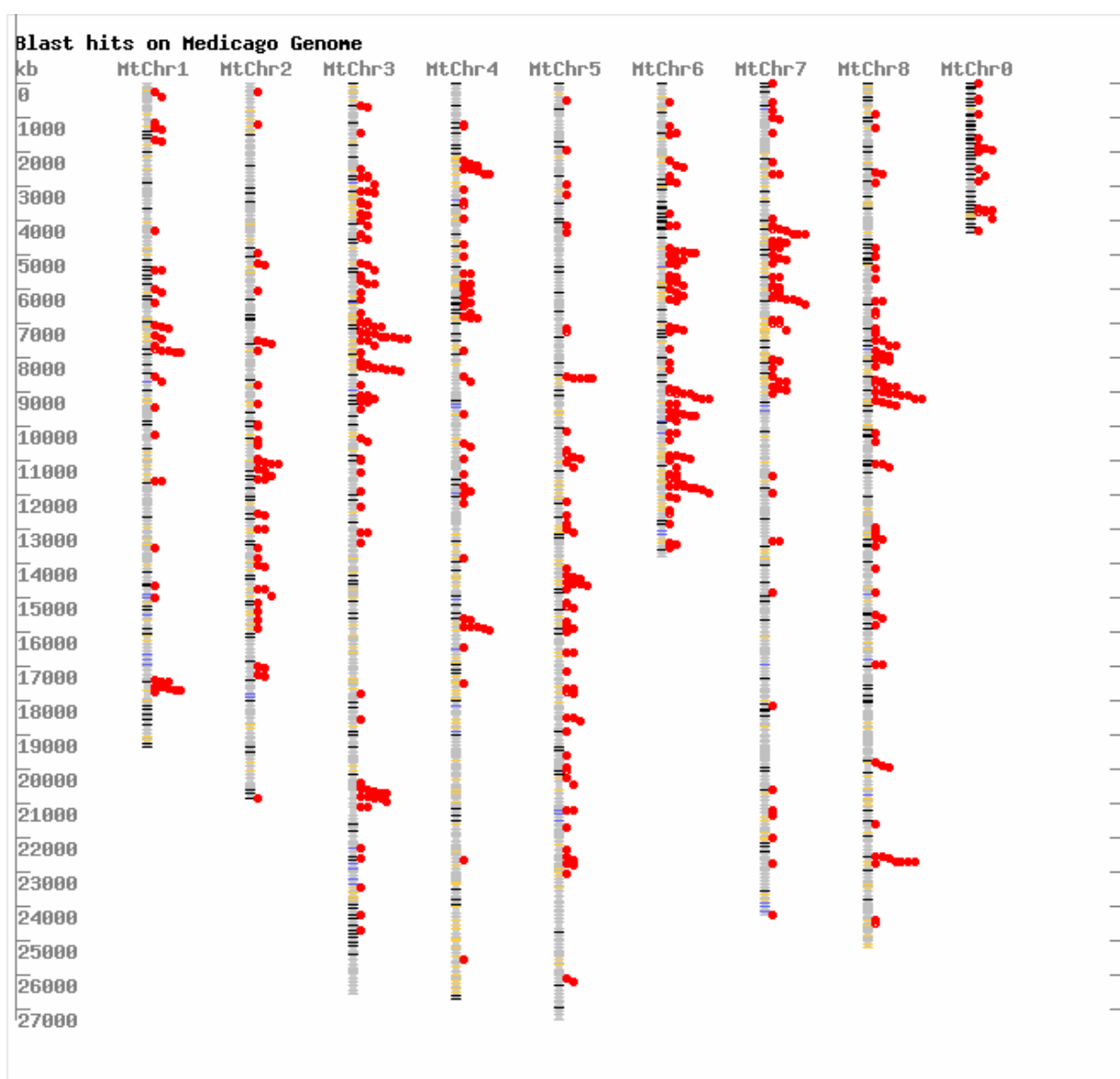


Figura 22. Padrão de distribuição do elemento Athila, em vermelho, nos cromossomos de *M. truncatula*.

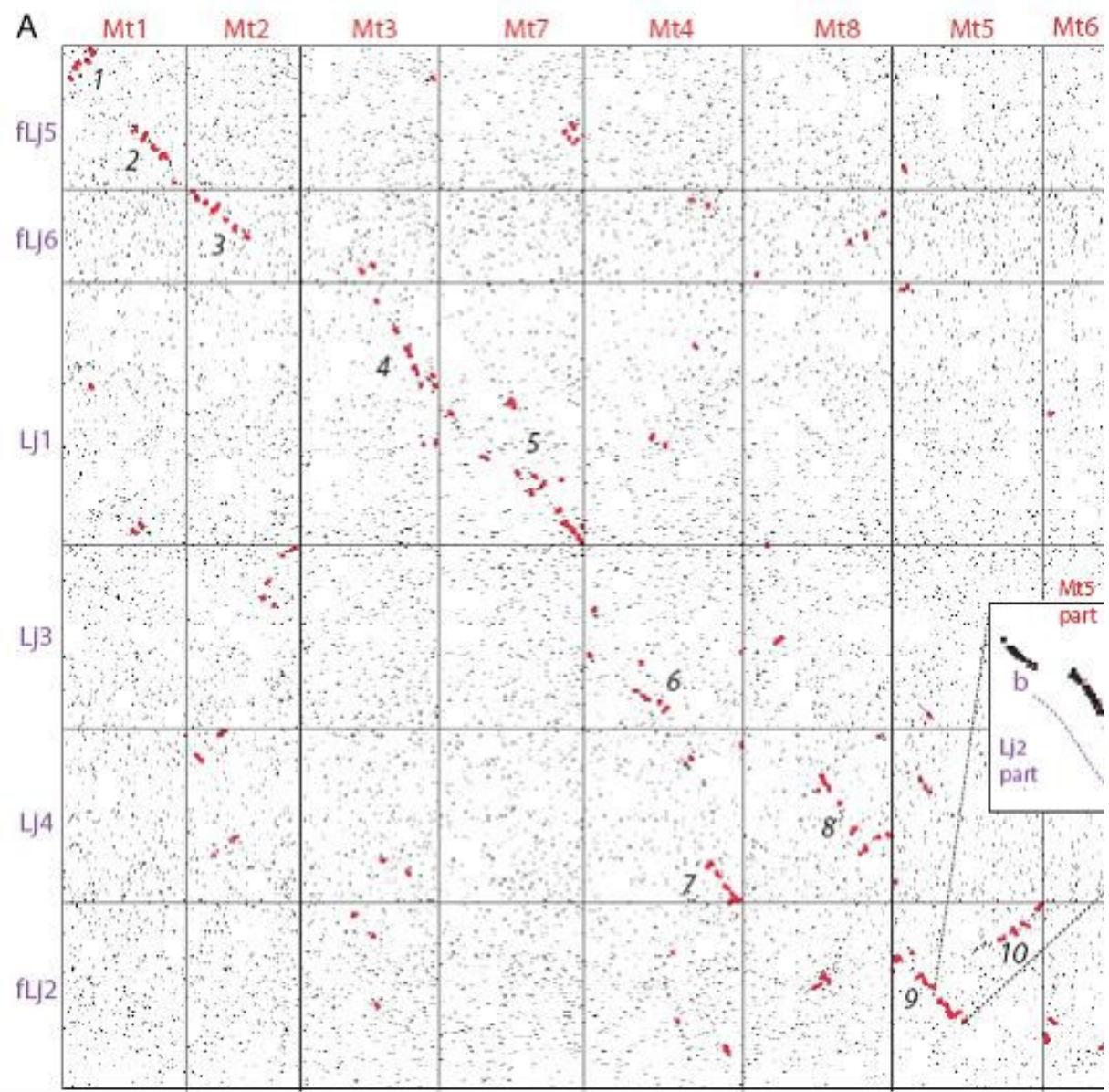


Figura 23. Colinearidade, em vermelho, entre os cromossomos de *M. truncatula* e *L. japonicus*. Adaptado: Cannon *et al.*, 2006.

6.02 Análise Filogenética

A topologia obtida na árvore contendo as seqüências nucleotídicas permitiu demonstrar que, os retrotransposons isolados de *A. duranensis* relacionam-se melhor com alguns que foram isolados de *A. hypogaea*, enquanto os de *A. ipaënsis* relacionam-se melhor com outros isolados do genoma do amendoim. Em alguns poucos casos observou-se a presença de seqüências das espécies silvestres em um mesmo grupo. Acredita-se que esse padrão de distribuição tenha ocorrido porque os elementos foram isolados de ambos os

genomas do amendoim cultivado, dessa forma os que se agruparam com *A. duranensis* provavelmente foram isolados do genoma AA de *A. hypogaea*, enquanto que os que se agruparam com *A. ipaënsis* provavelmente são oriundos do genoma BB.

A formação dos grupos permitiu inferir que, muito provavelmente, esse elemento já se encontrava no genoma do ancestral de *A. duranensis* e *A. ipaënsis*. Junto com os dados das distâncias genéticas obtidas para as seqüências é possível sugerir que esse elemento encontrava-se em um baixo número de cópias no genoma desse ancestral e que após a diferenciação nas duas espécies silvestres ele multiplicou-se no genoma de *A. duranensis*. Por outro lado, pode ter sofrido uma amplificação muito baixa, permaneceu com o mesmo número de cópias ou até mesmo começou a ser eliminado do genoma de *A. ipaënsis*. Caso esse elemento já tivesse um alto número de cópias no ancestral comum, esperar-se-ia uma distribuição mais homogênea das seqüências das três espécies na árvore de similaridade sem a formação de grupos isolados e que as distâncias genéticas entre elas fossem praticamente constantes entre os grupos.

O cálculo do número de substituições sinônimas e de substituições não-sinônimas mostrou claramente que, o genoma BB sofreu maior número de mutações, tanto silenciosas como não silenciosas, enquanto que o genoma AA de *A. duranensis* mostrou os menores índices de mutação. Esses dados sugerem que a seqüência gênica da enzima transcriptase reversa pode estar sofrendo maior modificação da função ou perda da função da proteína no genoma de *A. ipaënsis*. De um modo geral, os valores de dS e dN mostram que esses elementos não estão sofrendo nenhum tipo de pressão seletiva, sugerindo uma baixa ou nenhuma atividade sobre o genoma do hospedeiro.

A árvore obtida a partir das seqüências traduzidas também formou grupos isolados, apresentando as mesmas seqüências, porém, algumas em posições diferentes. Isso indica que nesses grupos as variações sofridas na sua composição nucleotídica refletiram-se diferentemente na tradução da seqüência protéica. Esses resultados vêm ao encontro com os obtidos para a distribuição de dS e dN para as três espécies. De acordo com Lynch e Conery (2000), é possível determinar a idade de eventos de duplicação gênica a partir da avaliação da freqüência da distribuição do número de substituições sinônimas para cada par de genes duplicados. Segundo Cui *et al* (2006), após a duplicação, alguns genes serão eliminados do genoma, enquanto outros irão acumular mutações, de preferência na terceira base do códon. Essas mutações normalmente escapam às pressões seletivas e podem mostrar o tempo aproximado desde a duplicação. Esses eventos serão observados como picos na distribuição

dos valores de dS e acredita-se que, da mesma maneira, esses picos possam mostrar o tempo decorrido desde a poliploidização do genoma.

No caso dos retrotransposons, imaginou-se que essa metodologia também pudesse prever eventos de multiplicação (aumento do número de cópias), já que esses elementos na maioria das vezes não estão sofrendo pressão seletiva e dessa forma acumulam mutações em sua seqüência ao longo do tempo. Devido aos mecanismos de controle da atividade desses elementos, a sua multiplicação no genoma, na maioria das vezes, acaba acontecendo em momentos distintos e únicos na história do genoma como, por exemplo, num possível evento de poliploidização e logo cessam sua atividade. Esses eventos acabariam sendo observados como os mesmos picos de distribuição dos valores de dS para a duplicação gênica e foi exatamente o que aconteceu após calcular o número de substituições sinônimas para a seqüência de cada espécie em relação à seqüência consenso, obtida a partir do alinhamento das seqüências de *A. duranensis* e *A. ipaënsis*.

A distribuição dos valores de dS para as seqüências de *A. duranensis* variou entre os primeiros intervalos 0,01-0,06; com um pico de freqüência no intervalo 0,04-0,05. Isso sugere que esses elementos sofreram transposições mais recentes. Para as seqüências de *A. ipaënsis*, a distribuição de dS foi mais homogênea, com um leve pico entre 0,15-0,16; mostrando que, nesse genoma, os retrotransposons acumularam maior quantidade de mutações ao longo do tempo e sugerindo que, além de serem mais antigas, poucas cópias desses elementos surgiram recentemente.

Para o genoma do amendoim esperava-se uma distribuição que apresentasse dois picos, um para cada genoma. Porém, o que foi observado é que a distribuição dos valores de dS para *A. hypogaea* ficou em um intervalo entre as das espécies silvestres, parecendo mais antigas do que *A. duranensis* e mais recentes que *A. ipaënsis*. O fato de a análise ter sido realizada com poucos dados e também devido à maior presença de seqüências isoladas do genoma AA, provavelmente está restringindo o gráfico de distribuição para *A. hypogaea*, impedindo a sobreposição das curvas e o aparecimento do pico indicativo da duplicação do genoma BB do amendoim. Numa tentativa de aumentar os dados estatísticos também foram construídos gráficos de distribuição contendo os valores de dS + dN, mas nenhuma mudança significativa foi observada.

De acordo com o valor médio de dS calculado para a distribuição em *A. duranensis*, foi possível determinar a data aproximada em que o retrotransposon se multiplicou no genoma (Cui *et al.*, 2006). Assim como feito por Bennetzen *et al* (2005), foi utilizada a taxa de mutação para a seqüência de genes dos elementos *adh1* e *adh2* da família *Poaceae* de

$6,5 \times 10^{-9}$ modificações por ano (Gaut *et al.*, 1996). O valor médio de dS (0,04) forneceu uma data de aproximadamente 6 milhões de anos. Utilizando a taxa de mutação específica para a seqüência gênica da enzima transcriptase reversa de gramíneas, de $2,08 \times 10^{-8}$, calculada por Langdon *et al.* (2000), uma data de aproximadamente 2 milhões de anos foi estimada.

Sabendo que a data de diversificação entre *M. truncatula* e *L. japonicus* é de aproximadamente 50 milhões de anos (Cannon *et al.*, 2006) e que a da soja, que possui uma data mais aproximada às de *Arachis* sp, de aproximadamente 60 milhões de anos, utilizou-se o dS indicativo da duplicação do genoma desta, de 0,5 (Cui *et al.*, 2006) para deduzir a data da multiplicação do retrotransposon no genoma de *A. duranensis*. Chegou-se ao resultado de aproximadamente 4,8 milhões de anos. A partir desses resultados definiu-se a data da explosão desses elementos entre 2-6 milhões de anos. Esses dados mostram que o genoma de *A. duranensis* sofreu essa modificação muito antes do surgimento do amendoim cultivado, que aconteceu provavelmente a apenas alguns milhares de anos.

Em um segundo momento, foi feita uma análise para determinar se esse elemento já poderia estar presente no genoma da espécie ancestral ao grupo das leguminosas. Acredita-se que esse ancestral possuía uma gama de retroelementos de diferentes tipos e que muito provavelmente eles foram transmitidos para as espécies surgidas. De acordo com os resultados da árvore filogenética pudemos observar que o elemento Athila já estava presente na espécie ancestral ao grupo. Porém ele foi transmitido para apenas algumas das espécies que surgiram, pois as seqüências claramente formaram subgrupos isolados dentro do grupo das leguminosas. De acordo com o cladograma obtido por Choi *et al* (2004), a subfamília *Papilionoideae*, à qual pertencem as espécies de *Arachis*, está entre as mais basais dentro da família das leguminosas. O agrupamento das seqüências utilizadas neste trabalho com as de *M. truncatula* da subfamília *Trifolieae*, ervilha (*Pisum sativum*) e fava (*Vicia faba*) da subfamília *Viceae* pode sugerir que os elementos repetitivos do ancestral foram transferidos diferentemente para algumas espécies. Mesmo formando um subgrupo, as seqüências dessas espécies são mais basais do que as de *Arachis*.

6.03 Número de cópias

O cálculo do número de cópias do retrotransposon nos três genomas foi realizado através de Dot blot. Essa técnica é aquela que permite obter uma melhor compreensão de uma seqüência com elevado número de cópias no genoma, quando comparada ao Southern blot, Colony blot e PCR (Navarro-Quezada e Schoen, 2002). Entretanto, características muito importantes dessa técnica são: a qualidade da amostra de DNA e também a sua quantificação. Nesse experimento, optamos por quantificar várias diluições das amostras em gel de agarose 1%, além de utilizar sempre as mesmas micropipetas e marca de ponteiras. As amostras contendo o plasmídeo com o inserto de RT, além de serem quantificadas em gel de agarose, também foram analisadas em espectrofotômetro. Não foi possível quantificar as amostras de DNA genômico em espectrofotômetro, pois esta era sempre superestimada, muito provavelmente devido a presença de compostos fenólicos, polissacarídeos e de proteínas, isolados durante a extração do material genético.

O sinal de hibridização obtido no filme radiográfico nem sempre forneceu uma escala linear. Normalmente as amostras mais concentradas ficaram muito parecidas umas com as outras. Acredita-se que, nesses casos, a emissão do sinal tenha ficado saturada no filme e, dessa forma, tiveram que ser eliminadas da análise. A utilização dos dados numéricos obtidos pelo programa MultiGauge, no programa Excel, visou obter as melhores curvas de diluição de DNA e também eliminar possíveis erros de manipulação das amostras. Em alguns casos tornou-se necessária a construção de gráficos em escala logarítmica, como no caso das curvas de calibração e no caso da curva de diluição do DNA genômico de *A. hypogaea*.

O número de cópias obtido para o genoma diplóide (2C) de cada espécie foi de $4893 \pm 1281,6$ para *A. hypogaea*, $3122,8 \pm 2382,5$ para *A. duranensis* e $829 \pm 143,02$ cópias para *A. ipaënsis*. Vale salientar que, em alguns casos, as seqüências dos clones diferiram entre si na sua composição nucleotídica em média de 8,5% entre *A. hypogaea* e *A. duranensis* e de 9% entre *A. hypogaea* e *A. ipaënsis*. Dessa forma, é muito provável que algumas das seqüências tenham sido removidas da membrana após as lavagens estridentes e, assim, o cálculo do número de cópias consistiu no número de moléculas do clone Ah-Athila-9 e daqueles com similaridade acima de 90% a ela. Portanto, os resultados obtidos servem apenas como uma estimativa da incidência desses elementos nos genomas.

Para auxiliar essa estimativa foram feitas sondagens iniciais na biblioteca BAC construída recentemente para as prováveis espécies parentais. De um total de 96 reações para

o genoma AA houve amplificação em 36 (37,5%) amostras. Para o genoma BB apenas 10 (10%) foram amplificadas, cerca de 3,5 vezes menos que no genoma AA, assim como obtido no Dot blot. Esses 96 clones correspondem a menos de 1% do genoma total para as duas espécies, mas espera-se que nos próximos resultados obtenhamos a mesma proporção de outrora, pois todas as vezes em que alguma sondagem era feita na biblioteca um número de clones muito maior era isolado do genoma AA do que no BB. Não foram realizadas novas reações devido à grande dificuldade de se trabalhar com os clones BAC, a extração do DNA nem sempre funciona e deve ser realizada com muito cuidado. A realização de extração em placas de 96 poços além de não permitir a extração de DNA em quantidade suficiente também termina na extração de amostras muito viscosas, péssimas para uma reação de PCR. A melhor maneira parece ser o crescimento bacteriano em tubos de 50mL com agitação muito forte (250rpm) e extração em tubos de 1,5mL, o que torna o trabalho muito dispendioso.

Em contribuição a essa análise, também foram feitas buscas eletrônicas por similaridades ao clone Ah-Athila-9 através de BLASTn e utilizando o banco de dados *Genome Survey Sequence*. Foram encontradas similaridades com seqüências de *A. duranensis*, *A. hypogaea* e *A. batizocoi* (genoma BB), mas apenas as que apresentaram índices de similaridades (E-value) $\leq 1e^{-20}$ foram consideradas significantes. Não foram realizadas buscas contra *A. ipaënsis* por não haver seqüências depositadas no GenBank. O que se pode discutir desses resultados é a presença desse elemento em um número muito maior no genoma AA do que no BB, cerca de 13 vezes mais o que indica que no genoma de *A. batizocoi* esse elemento possa conter um número de cópias ainda mais baixo do que em *A. ipaënsis*.

6.04 Análise de metilação

Sabe-se que os retrotransposons passam por diversas formas de controle da sua atividade, uma delas é a metilação do DNA genômico, que impede fisicamente que proteínas ou fatores de transcrição encontrem o seu sítio de reconhecimento (Hirochika *et al.*, 2000). Esse controle pode ser utilizado pela célula simplesmente para o controle de sua expressão gênica e também como proteção em situações que podem causar demasiados danos ao DNA (Kumar e Bennetzen, 1999). Entretanto, qualquer evento extremamente estressante para o genoma pode mudar o estado de metilação do material genético e causar a reativação de retrotransposons, dois desses eventos são a poliploidização e o cruzamento interespecífico, como é o caso do amendoim. Para avaliar o estado de metilação dos genomas das três

espécies, foi analisada a frequência de corte das enzimas *MboI* e *Sau3AI* sobre eles. O Southern blot mostrou que em nenhuma espécie o elemento repetitivo pareceu estar metilado. Além disso, em trabalhos anteriores utilizando uma seqüência repetitiva diferente também não foi observada a metilação desse elemento em *A. ipaënsis* e no anfidiplóide sintético.

Apesar de não estar metilado, esse elemento também parece não estar ativo, já que seu número de cópias não pode ser considerado muito alto, por não apresentar quaisquer similaridades significantes entre a seqüência de RT do retrotransposon e cerca de 20 mil seqüências de ESTs de *Arachis* sp, depositadas no GenBank e no banco de dados de *Arachis* e também devido às alterações sofridas no gene de RT, identificadas a partir do cálculo das distâncias genéticas entre as seqüências de aminoácidos. Alguns casos de transposons e retrotransposons não-metilados são conhecidos (Simmen *et al.*, 1999), inclusive em espécies conhecidas por possuírem o seu genoma altamente metilado como milho (*Zea mays*) (Martienssen, 1998). Isso mostra que nem sempre a metilação é o principal mecanismo de inativação desses elementos e que no caso de *Arachis*, as modificações na seqüência devem ser a causa principal para a inativação do retrotransposon.

De acordo com resultados anteriores obtidos com uma seqüência repetitiva diferente e com observações de Nielen *et al.* (2005 – dados não publicados), esse elemento parece ser o mais abundante no genoma do amendoim já que foi isolado diversas vezes através de diferentes técnicas tais como digestão e subclonagem de DNA genômico, seqüenciamento de fragmentos obtidos através de AFLP e também através de procura por seqüências depositadas no GenBank oriundas de seqüenciamento genômico (Apêndice).

6.05 Extensão da seqüência

Visando determinar a seqüência completa do retrotransposon, foram realizadas reações de genome walker utilizando iniciadores específicos para a seqüência desse elemento. O kit permitiu a ampliação da seqüência do elemento para aproximadamente 3,6Kb, cerca de 30% da média esperada para os elementos tipo Athila (Wright e Voytas, 2001). A seqüência obtida possui o gene completo da enzima transcriptase reversa de aproximadamente 200 aminoácidos além de uma região contendo cerca de 250 aminoácidos próxima ao gene, muito semelhante à do elemento de *M. truncatula*. O restante não apresentou nenhuma homologia significativa, o que mostra a grande diversidade desses elementos.

Uma grande dificuldade encontrada ao utilizar esse kit foi que os clones obtidos após a ligação continham vários tamanhos de insertos e somente aqueles maiores que 1Kb foram

úteis para a ampliação da seqüência conhecida do retrotransposon. Além disso, é necessário o desenho de novos iniciadores para cada nova reação de PCR e também a utilização de enzimas de restrição, o que acaba tornando o processo caro e demorado. Nessas circunstâncias e também com o desenvolvimento da biblioteca BAC, a melhor metodologia será identificar quais clones dessa biblioteca contêm o elemento e fazer o seqüenciamento a partir de produtos de PCR. Sequenciar o clone por completo sem a necessidade de ligação e transformação de plasmídios seria outra alternativa mas, extremamente cara. O seqüenciamento de produtos de PCR já vem sendo realizado com sucesso no laboratório e espera-se que em pouco tempo a seqüência do elemento esteja completa.

7. Conclusões

- I) O retrotransposon isolado pertence ao grupo LTR, Ty3-Gypsy, tipo Athila e parece ser o mais abundante no genoma do amendoim.
- II) Esse elemento é localizado em regiões de eucromatina no genoma do amendoim e apresenta distribuição diferenciada no genoma AA e BB de *A. hypogaea*, sendo mais abundante no primeiro. O número de cópias estimado para cada uma das espécies foi de: $4893 \pm 1281,6$ para *A. hypogaea*; $3122,8 \pm 2382,5$ para *A. duranensis* e de $823 \pm 143,02$ cópias em *A. ipaënsis*.
- III) Estudos filogenéticos e estatísticos permitiram demonstrar que esse elemento é muito antigo, estando presente no genoma de espécies de diferentes grupos vegetais.
- IV) Esse elemento provavelmente encontrava-se inativo no genoma do ancestral comum das espécies parentais do amendoim. Após a diferenciação em *A. duranensis* e *A. ipaënsis* o retroelemento aumentou seu número de cópias no genoma AA a aproximadamente 2-6 milhões de anos e permaneceu praticamente inativo no genoma BB.
- V) Ao mesmo tempo em que apresenta um moderado número de cópias, o elemento não sofreu metilação em nenhuma das espécies de *Arachis*. Além disso, a sondagem de 20 mil seqüências de ESTs não revelou nenhuma homologia com a do gene da transcriptase reversa, mostrando não haver a presença de cópias de mRNA transcritas.
- VI) A distribuição das seqüências nos cladogramas mostrou que este retrotransposon provavelmente cessou sua atividade antes do surgimento de *A. hypogaea*.
- VII) O acúmulo de mutações sinônimas e não-sinônimas nas seqüências desse elemento, isoladas das três espécies, permitiu observar que, apesar de ser abundante no genoma, a seqüência do gene da enzima transcriptase reversa vem

sofrendo modificações que indicam perda da função dessa proteína e que esta pode ser a causa principal da inativação do elemento.

VIII) A sondagem da biblioteca BAC vem mostrando que esse elemento pode vir a constituir um marcador cromossômico preferencial para o genoma AA.

8. Perspectivas

Dentre os possíveis trabalhos com retrotransposons a serem realizados, espera-se primeiramente terminar o seqüenciamento do elemento estudado nesta dissertação e nomeá-lo. Em seguida esperamos induzir uma possível ativação desse elemento no genoma de diversas espécies, a partir do uso de luz ultravioleta ou radiação e estudar os efeitos genéticos gerados, tais como mudanças morfológicas e surgimento de polimorfismos, visando o mapeamento dos genes envolvidos com essas características.

Como esse foi um trabalho pioneiro de caracterização de retrotransposons no genoma de *Arachis* sp, o isolamento e caracterização de novos elementos de diferentes tipos e classes pode ajudar a melhorar a compreensão da organização dos genomas dessas espécies.

9. Referências Bibliográficas

- Ahn, S.C., Baek, B.S., Oh, T., Song, C.S. e Chatterjee, B. 2000. Rapid Mini-Scale Plasmid Isolation for DNA Sequencing and Restriction Mapping. *BioTechniques*, 29 (3): 466-468.
- Alix, K.; Heslop-Harrison, J.S. 2004. The diversity of retroelements in diploid and allotetraploid *Brassica* species. *Plant Molecular Biology*, 54: 895-909.
- Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. 1990. "Basic local alignment search tool". *Journal of Molecular Biology*, 215: 403-410.
- Altschul, S.F.; Madden, T.L.; Schäffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.J. 1997. "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." *Nucleic Acids Research*, 25: 3389-3402.
- Armstead, I.P.; Turner, L.B.; Farrell, M.; Skøt, L.; Gomez, P.; Montoya, T.; Donnison, I.S.; King, I.P.; Humphreys, M.O. 2004. Synteny between a major heading-date QTL in perennial ryegrass (*Lolium perenne* L.) and the Hd3 heading-date locus in rice. *Theoretical and Applied Genetics*, 108: 822-828.
- Bennetzen, J.L.; Freeling, M. 2000. The Unified Grass Genome: Synergy in Synteny. *Genome Research*, 7: 301-306.
- Bennetzen, J.L. 2000. Transposable element contributions to plant gene and genome evolution. *Plant Molecular Biology*, 42: 251-269.
- Bennetzen, J.L.; Ma, J.; Devos, M. 2005. Mechanisms of Recent Genome Size Variation in Flowering Plants. *Annals of Botany*, 95: 127-132.
- Cannon, S.B.; Sterck, L.; Rombauts, S.; Sato, S.; Cheung, F.; Gouzy, J.; Wang, X.; Mudge, J.; Vasdewani, J.; Schiex, T.; Spannagl, M.; Monaghan, E.; Nicholson, C.; Humphray, S.J.; Schoof, H.; Mayer, K.F.X.; Rogers, J.; Quétier, F.; Oldroyd, G.E.; Debelle, F.; Cook, D.R.; Retzel, E.F.; Roe, B.A.; Town, C.D.; Tabata, S.; Van de Peer, Y.; Young, N.D. 2006. Legume genome evolution viewed through the *Medicago truncatula* and *Lotus japonicus* genomes. *Proceedings of the National Academy of Sciences of the United States of America*, 103: 14959-14964.
- Chenna, R.; Sugawara, H.; Koike, T.; Lopez, R.; Gibson, T.J.; Higgins, D.G.; Thompson, J.D. 2003. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Research*, 31 (13): 3497-3500.
- Choi, H.; Mun, J.; Kim, D.; Zhu, H.; Baek, J.; Mudge, J.; Roe, B.; Ellis, N.; Doyle, J.; Kiss, G.B.; Young, N.D.; Cook, D.R. 2004. Estimating genome conservation between crop and model legume species. *Proceedings of the National Academy of Sciences of the United States of America*, 101: 15289-15294.
- Clamp, M.; Cuff, J.; Searle, S.M.; Barton, G.J. 2004. The Jalview Java alignment editor. *Bioinformatics*, 20 (3): 426-427.

- Company, M.; Stalker, H.T.; Wynne, J.C. Cytology and leafspot resistance in *Arachis hypogaea* x wild species hybrids. 1982, citado por: Fávero, A.P. 2004. *Cruzabilidade entre espécies silvestres de Arachis visando à introgressão de genes de resistências a doenças no amendoim cultivado*. Tese (Doutorado) - Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 165pp.
- Cui, L.; Wall, P.K.; Leebens-Marck, J.H.; Lindsay, B.; Soltis, D.E.; Doyle, J.J.; Soltis, P.S.; Carlson, J.E.; Arumuganathan, K.; Barakat, A.; Albert, V.A.; Ma, H.; dePamphilis, C.W. 2006. Widespread genome duplications throughout the history of flowering plants. *Genome Research*, 16: 738-749.
- Deragon, J.M.; Capy, P. Impact of transposable elements on the human genome. 2000, citado por: Alix, K.; Heslop-Harrison, J.S. 2004. The diversity of retroelements in diploid and allotetraploid *Brassica* species. *Plant Molecular Biology*, 54: 895-909.
- Diao, X.; Freeling, M.; Lisch, D. 2006. Horizontal Transfer of a Plant Transposon. *Plos Biology*, 4 (1): 0119-0128.
- Dixit, A.; Ma, K.; Yu, J.; Cho, E.; Park, Y. 2006. Reverse transcriptase domain sequences from Mungbean (*Vigna radiata*) LTR retrotransposons: Sequence characterization and phylogenetic analysis. *Plant Cell Report*, 25: 100-111.
- Dorer, D.; Henikoff, S. 1994. Expansions of transgene repeats cause heterochromatin formation and gene silencing in *Drosophila*. *Cell*, 77: 993-1002.
- Duke, J.A. Handbook of legumes of world economic importance. 1981, citado por: Seijo, J.G.; Lavia, G.I.; Fernández, A.; Krapovickas, A.; Ducasse, D.; Moscone, E.A. 2004. Physical Mapping of the 5S and 18S-25S rRNA Genes by FISH as Evidence that *Arachis duranensis* and *A. ipaensis* are the Wild Diploid Progenitors of *A. hypogaea* (Leguminosae). *American Journal of Botany*. 91 (9): 1294-1303.
- Ewung, B.; Hillier, L.; Wendl, M.C.; Green, P. 1998. Base-calling of automated sequencer traces using PHRED. I. Accuracy assessment. *Genome Research*, 8:175-185.
- Fávero, A.P. 2004. *Cruzabilidade entre espécies silvestres de Arachis visando à introgressão de genes de resistências a doenças no amendoim cultivado*. Tese (Doutorado) - Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 165pp.
- Fávero, A.P.; Simpson, C.E.; Valls, J.F.M.; Vello, N.A. 2006. Study of the evolution of cultivated peanut through crossability studies among *A. ipaënsis*, *A. duranensis*, and *A. hypogaea*. *Crop Science*, 46: 1546-1552.
- Flavell, R.B.; Bennet, M.D.; Smith, J.B.; Smith, D.B. 1974. Genome size and proportion of repeated nucleotide sequence DNA in plants. *Biochemical Genetics*, 12: 257-269.
- Flavell, A.J. 1999. Long terminal repeat retrotransposons jump between species. *Proceedings of the National Academy of Sciences of the United States of America*, 96: 12211-12212.

- Gaut, B.S.; Morton, B.R.; McCaig, B.C.; Clegg, M.T. 1996. Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcL*. *Proceedings of the National Academy of Sciences of the United States of America*, 93: 10274-10279.
- Godoy, I.J.; Moraes, S.A.; Zanotto, M.D.; Santos, R.C. Melhoramento do amendoim. 1999. In: *Melhoramento de espécies cultivadas*. Borém, A. (Ed.). Universidade Federal de Viçosa. Viçosa, p. 51-94.
- Grandbastien, M.A. 1998. Activation of plant retrotransposons under stress conditions. *Trends in Plant Science*, 3 (5): 181-187.
- Grant, D.; Cregan, P.; Shoemaker, R.C. 2000. Genome organization in dicots: Genome duplication in *Arabidopsis* and synteny between soybean and *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America*, 97 (8): 4168-4173.
- Grattapaglia D.; Sederoff R. 1994. Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: Mapping strategy and RAPD markers. *Genetics*, 137: 1121-1137.
- Grunstein, M. 1998. Yeast heterochromatin: regulation of its assembly and inheritance by histones. *Cell*, 93: 325-328.
- Halward, T.M.; Stalker, H.T.; Larue, E.A.; Kochert, G. 1991. Genetic variation detectable with molecular markers among unadapted germ-plasm resources of cultivated peanut and related wild species. *Genome*, 34: 1013-1020.
- Hansen, C.N.; Heslop-Harrison, J.S. 2004. Sequences and Phylogenies of Plant Pararetroviruses, Viruses and Transposable Elements. *Advances in Botanical Research*, 41: 165-193.
- Heslop-Harrison, J.S. 2000. Comparative Genome Organization in Plants: From Sequence and Markers to Chromatin and Chromosomes. *The Plant Cell*, 12: 617-635.
- Hirochika, H. 1993. Activation of tobacco retrotransposons during tissue-culture. *EMBO J*, 12: 2521-2528.
- Hirochika, H.; Sugimoto, K.; Otsuki, Y.; Tsugawa, H.; Kanda, M. 1996. Retrotransposons of rice involved in mutations induced by tissue culture. *Proceedings of the National Academy of Sciences of the United States of America*, 93: 7783-7788.
- Hirochika, H.; Okamoto H.; Kakutani, T. 2000. Silencing of Retrotransposons in *Arabidopsis* and Reactivation by the *ddm1* Mutation. *Plant Cell*, 12: 357-368.
- Instituto Brasileiro de Geografia e Estatística. 2006. *Produção Agrícola Municipal: Cereais, Leguminosas e Oleaginosas*. Rio de Janeiro. www.ibge.gov.br/home/estatistica/economia/pamclo/2005/pamclo2005.pdf.

- José, A.C.V.F. 2006. *Desenvolvimento e Mapeamento de Marcadores Moleculares RGA em Espécies Silvestres de Arachis*. Dissertação (Mestrado), Ciências Genômicas e Biotecnologia, Universidade Católica de Brasília, Brasília, 154pp.
- Kazazian Jr, H.H. 2004. Mobile Elements: Drivers of Genome Evolution. *Science*, 303: 1626-1632.
- Kalendar, R.; Gorb, T.; Regina, M.; Suoniemi, A.; Schulman, A. 1999. IRAP and REMAP: two new retrotransposon-based DNA fingerprinting techniques. *Theoretical and Applied Genetics*, 98: 704-711.
- Kochert, G.; Halward, T.; Branch, W.D.; Simpson, C.E. 1991. RFLP Variability in peanut (*Arachis hypogaea* L.) Cultivars and Wild species. *Theoretical and Applied Genetics*, 81: 563-570.
- Kochert G, Stalker HT, Gimenes M, Galaro L, Lopes CR and Moore K. 1996. RFLP and cytogenetic evidence on the origin and evolution of allotetraploid domesticated peanut, *Arachis hypogaea* (Leguminosae). *American Journal of Botany*, 83: 1282-1291.
- Krapovickas, A.; Gregory, W.C. 1994. Taxonomía del género *Arachis* (Leguminosae). *Bonplandia*, 8: 1-186.
- Kimura, M. 1980. A simple method for estimating evolutionary rates of bases substitutions through comparative studies of nucleotide sequence. *Journal of Molecular Evolution*, 16: 11-129.
- Ku, H.; Liu, J.; Doganlar, S.; Tanksley, S.D. 2001. Exploitation of *Arabidopsis*-tomato synteny to construct a high-resolution map of the *ovate*-containing region in tomato chromosome 2. *Genome*, 44: 470-475.
- Kumar, A.; Bennetzen, J.L. 1999. Plant Retrotransposons. *Annual Review of Genetics*, 33: 479-532.
- Kumar, S.; Tamura, K.; Nei, M. 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Briefings in Bioinformatics*, 5: 150-163.
- Kumekawa, N.; Ohtsubo, E.; Ohtsubo, H. 1999. Identification and phylogenetic analysis of gypsy-type retrotransposon in the plant kingdom. *Genes & Genetics Systems*, 74: 299-307.
- Langdon, T.; Seago, C.; Mende, M.; Legget, M.; Thomas, H.; Forster, J.W.; Thomas, H.; Jones R.N.; Jenkins, G. 2000. Retrotransposon Evolution in Diverse Plant Genomes. *Genetics*, 156: 313-325.
- Lewin, B. 2004. Retroviruses and retroposons. In: *Genes VIII*. Carlson, G. (Ed.). Pearson Prentice Hall. NJ. pp.493-509.
- Liu, B.; Wendel, J.F. 2000. Retroelement activation followed by rapid repression in interspecific hybrid plants. *Genome*, 43: 1-7.

- Lynch, M.; Conery, J.S. 2000. The evolutionary fate and consequences of duplicated genes. *Science*, 290: 1151-1155.
- Marco, A.; Marín, I. 2005. Retrovirus-like elements in plants. *Recent Research Development in Plant Science*, 3: 1-10.
- Martienssen, R. 1998. Transposons, DNA methylation and gene control. *Trends in Genetics*, 14 (7): 263-264.
- Melayah, D.; Yoong Lim, K.; Bonnivard, E.; Chalhoub, B.; De Borne, F.D.; Mhiri, C.; Leitch, A.R.; Grandbastien, M. 2004. Distribution of the Tnt1 Retrotransposon family in the amphidiploid tobacco (*Nicotiana tabacum*) and its wild *Nicotiana* relatives. *Biological Journal of the Linnean Society*. 82: 639-649.
- Moretzsohn, M.C.; Hopkins, M.; Mitchell, S.E.; Kresovich, S.; Valls, J.F.M.; Ferreira, M.E. 2004. Genetic diversity of peanut (*Arachis hypogaea* L.) and its wild relatives based on the analysis of hypervariable regions of the genome. *BMC Plant Biology*. [<http://www.biomedcentral.com/1471-2229/4/11>].
- Moretzsohn, M.C. 2006. *Desenvolvimento e Mapeamento de Marcadores Microsatélites e Identificação de QTLs Ligados à Produtividade e à Resistência à Mancha Preta em Arachis spp.* Tese (Doutorado) - Departamento de Biologia Celular, Pós-Graduação em Biologia Molecular, Universidade de Brasília, Brasília, 142pp.
- Navarro-Quezada, A.; Schoen, D.J. 2002. Sequence evolution and copy number of Ty1-copia retrotransposons in diverse plant genomes. *Proceedings of the National Academy of Sciences of the United States of America*, 99: 268-273.
- Nei, M.; Kumar, S. 2000. *Molecular Evolution and Phylogenetics*. Oxford University Press, Inc, New York, 333p.
- Nouzová, M.; Kubaláková, M.; Zelová, M.D.; Koblízková, A.; Neumann, P.; Dolezel, J.; Macas, J. 1999. Cloning and Characterization of New Repetitive Sequences in Field Bean (*Vicia faba* L.). *Annals of Botany*. 83: 535-541.
- Pearce, S.R.; Harrison, G.; Li, D.; Heslop-Harrison, J.S.; Kumar, A.; Flavell, A.J. 1996. The Ty1-copia group retrotransposons in *Vicia* species: copy number, sequence heterogeneity and chromosomal localization. *Molecular and General Genetics*, 250: 305-315.
- Pelissier, T.; Tutois, S.; Deragon, J.M.; Tourment, S.; Genestier, S.; Picard, G. 1995. Athila, a new retroelement from *Arabidopsis thaliana*. *Plant Molecular Biology*, 29: 441-452.
- Proite, K.; Leal-Bertioli, S.C.M.; Bertioli, D.J.; Guimarães, P.M. 2006. *Construção de bibliotecas BAC de espécies silvestres de Arachis resistentes a fungos e nematóides.* Boletim de Pesquisa & Desenvolvimento, n.130. Embrapa Recursos Genéticos e Biotecnologia. Brasília.16pp.

- Raina, S.N.; Sharma, S.; Sasakuma, T.; Kishii, M.; Vaishnavi, S. 2005. Novel Repeated DNA Sequences in Safflower (*Carthamus tinctorius* L.) (*Asteraceae*): Cloning, Sequencing, and Physical Mapping by Fluorescence in situ Hybridization. *The American Genetic Association*, 96(4): 424-429.
- Rodriguez, M.; O'Sullivan, D.; Donini, P.; Papa, R.; Chiapparino, E.; Leigh, F.; Attene, G. 2006. Integration of retrotransposons-based markers in a linkage map of barley. *Molecular Breeding*, 17: 173-184.
- Sabot, F.; Schulman, A.H. 2006. Parasitism and the retrotransposon life cycle in plants: a hitchhiker's guide to the genome. *Heredity*, 97: 381-388.
- Schmidt, T.; Heslop-Harrison, J.S. 1998. Genomes, genes and junk: the large-scale organization of plant chromosomes. *Trends in Plant Science*, 3: 195-199.
- Schmidt, R. 2000. Synteny: recent advances and future prospects. *Current Opinion in Plant Biology*, 3: 97-102.
- Seijo, J.G.; Lavia, G.I.; Fernández, A.; Krapovickas, A.; Ducasse, D.; Moscone, E.A. 2004. Physical Mapping of the 5S and 18S-25S rRNA Genes by FISH as Evidence that *Arachis duranensis* and *A. ipaensis* are the Wild Diploid Progenitors of *A. hypogaea* (*Leguminosae*). *American Journal of Botany*, 91(9): 1294-1303.
- Sharma, K.K.; Lavanya, M.; Anjaiah, V. 2000. A Method for Isolation and Purification of Peanut Genomic DNA Suitable for Analytical Applications. *Plant Molecular Biology Reporter*, 18: 393a-393h.
- Simmen, M.W.; Leitgeb, S.; Charlton, J.; Jones, S.J.M.; Harris, B.R.; Clark, V.H.; Bird, A. 1999. Nonmethylated Transposable Elements and Methylated Genes in a Chordate Genome. *Science*, 283: 1164-1167.
- Simpson, C.E.; Krapovickas, A.; Valls, J.F.M. 2001. History of *Arachis* Including Evidence of *A. hypogaea* L. Progenitors. *Peanut Science*, 28: 78-80.
- Simpson, C.E. 2001. Use of Wild *Arachis* Species/Introgression of Genes into *A. hypogaea* L. *Peanut Science*, 28: 114-116, 2001.
- Simpson, C.E.; Starr, J.L. 2001. Registration of 'Coan' peanut. *Crop Science*, 41: 918.
- Simpson, C.E.; Starr, J.L.; Church, G.T.; Burow, M.D.; Paterson, A.H. 2003. Registration of 'NemaTAM' Peanut. *Crop Science*, 43: 1561.
- Singh, U.; Singh, B. Tropical grain legumes as important human foods. 1991. citado por: Moretzsohn, M.C.; Hopkins, M.; Mitchell, S.E.; Kresovich, S.; Valls, J.F.M.; Ferreira, M.E. 2004. Genetic diversity of peanut (*Arachis hypogaea* L.) and its wild relatives based on the analysis of hypervariable regions of the genome. *BMC Plant Biology*. [<http://www.biomedcentral.com/1471-2229/4/11>].
- Southern, E.M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology*, 98: 503-517.

- Staden, R.; Beal, K.F.; Bonfield, J.K. 1998. The Staden Package. In: Misener S., Krawetz. Computer Methods in Molecular Biology. *Bioinformatics Methods and Protocols*, 132: 115-130.
- Stalker, H.T. 1991. A new species in section *Arachis* of peanuts with D genome. *American Journal of Botany*, 78: 630-637.
- Stalker, H.T.; Simpson, C.E. Germplasm resources in *Arachis*. In: PATTEE, H.E.; Stalker, H.T. (Eds). *Advances in Peanut Science*, APRES, Stillwater, OK, 1995. Cap. 2, pp 14-53.
- Stalker, H.T. 1997. Peanut (*Arachis hypogaea* L.). *Field Crops Research*, 53: 205-217.
- Starr, J.L.; Morgan, E.R.; Simpson, C.E. 2002. Management of the peanut root-knot nematode, *Meloidogyne arenaria*, with host resistance. Online. Plant Health Progress, doi:10.1094/PHP-2002-1121-01-HM.
- Stracke, S.; Sato, S.; Sandal, N.; Kyoma, M.; Kaneko, T.; Tabata, S.; Parniske, M. 2004. Exploitation of colinear relationships between the genomes of *Lotus japonicus*, *Pisum sativum* and *Arabidopsis thaliana*, for positional cloning of a legume symbiosis gene. *Theoretical and Applied Genetics*, 108: 442-449.
- Temsch, E.M.; Greilhuber, J. 2000. Genome size variation in *Arachis hypogaea* and *A. monticola* re-evaluated. *Genome*, 43: 449-451.
- Temsch, E.M.; Greilhuber, J. 2001. Genome size in *Arachis duranensis*: a critical study. *Genome*, 44: 826-830.
- Valls, J.F.M; Simpson, C.E. 2005. New species of *Arachis* L. (Leguminosae) from Brazil, Paraguay and Bolivia. *Bonplandia*, 14: 35-64.
- Vitte, C.; Panaud, O. 2005. LTR retrotransposons and flowering plant genome size: emergence of the increase/decrease model. *Cytogenetic and Genome Research*, 110: 91-107.
- Young, N.D.; Cannon, S.B.; Sato, S.; Kim, D.; Cook, D.R.; Town, C.D.; Roe, B.A.; Tabata, S. 2005. Sequencing the Genespaces of *Medicago truncatula* and *Lotus japonicus*. *Plant Physiology*, 137: 1174-1181.
- Young, N.D.; Weeden, N.F.; Kochert, G. Genome mapping in legumes (Family Fabacea). In: Paterson, A.H. (Ed.). 1996. *Genome mapping in plants*. R.G. Landes, Austin, TX. p. 212-227.
- Watson, J. D. 1992. *Recombinant DNA*. 2nd.ed. New York: Scientific American Books: Distributed by W.H. Freeman.
- Wessler, S.R. 1996. Plant retrotransposons: Turned on by stress. *Current Biology*, 6 (8): 959-961.

- Wessler, S.R. 2006. Transposable elements and the evolution of eukaryotic genomes. *Proceedings of the National Academy of Sciences of the United States of America*, 103 (47): 17600-17601.
- Wright, D.A.; Voytas, D.F. 2001. *Athila4* of *Arabidopsis* and *Calypso* of Soybean Define a Lineage of Endogenous Plant Retroviruses. *Genome Research*, 12: 122-131.
- Zhao, X.; Si, Y.; Hanson, R.E.; Crane, C.F.; Price, H.J.; Stelly, D.M.; Wendel, J.F.; Paterson, A.H. 1998. Dispersed repetitive DNA has spread to new genomes since polyploidy formation in cotton. *Genome Research*, 8: 479-492.

Apêndice

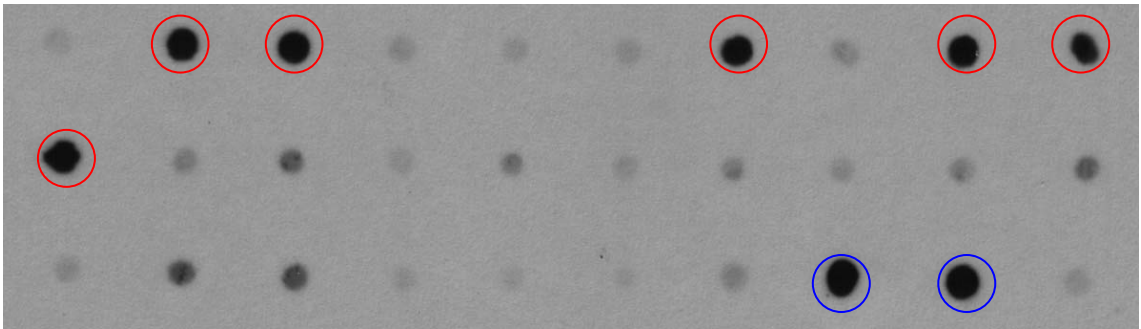


Figura 24. Dot blot dos clones obtidos após subclonagem de fragmentos oriundos de digestão de DNA genômico de *A. hypogaea* com *Hind*III (vermelho). Como sonda foi utilizado DNA genômico do próprio amendoim. Os clones com maior sinal de hibridização foram escolhidos para seqüenciamento automático. Em azul estão os clones E12 e F05 obtidos em trabalhos anteriores.

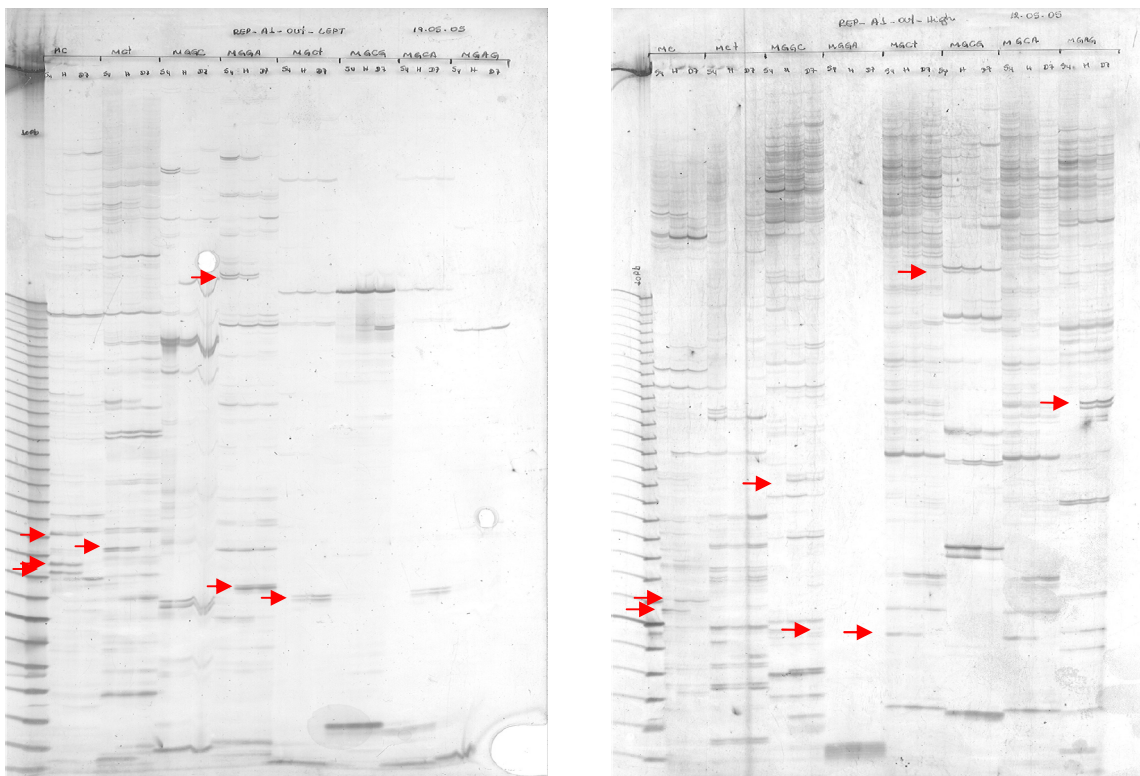


Figura 25. Fragmentos polimórficos de DNA obtidos por AFLP utilizando-se iniciadores desenhados exclusivamente para amplificação da seqüência do clone E12. As amostras foram amplificadas no genoma de *A. duranensis*, *A. stenosperma* (genoma AA) e da planta híbrida entre elas. As bandas marcadas com as setas foram escolhidas para seqüenciamento automático. Após o seqüenciamento dos clones e das bandas polimórficas, as seqüências alinharam-se para comporem um contig (consenso) apenas.

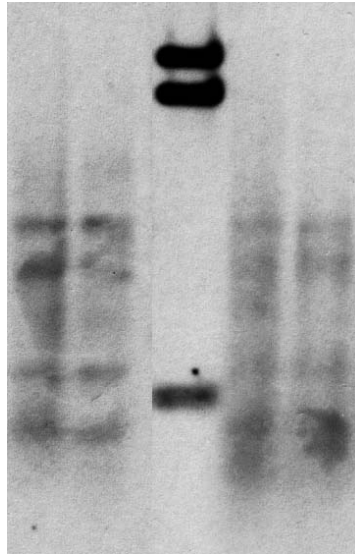


Figura 26. Southern blot para análise de metilação em *A. ipaënsis* (esquerda) e no Anfidiplóide sintético (direita), utilizando as enzimas *MboI* e *Sau3AI* e como sonda a seqüência de E12 que agora acredita-se fazer parte do retrotransposon utilizado nesta dissertação. Nenhum sinal de metilação foi encontrado.

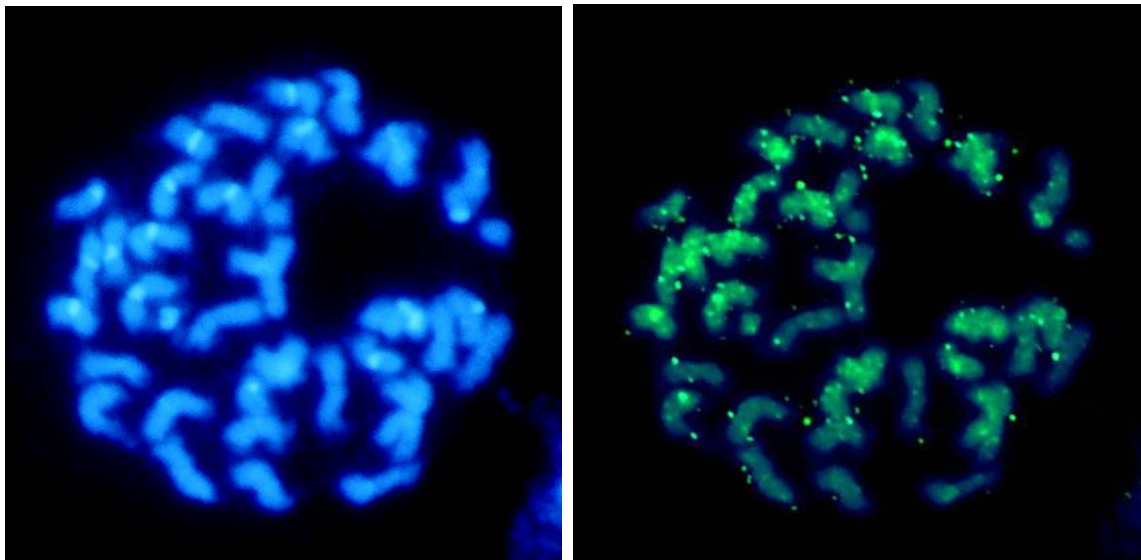


Figura 27. Distribuição da seqüência do clone E12 no genoma de *A. hypogaea*. Em azul coloração com DAPI. Em verde a hibridização da mesma lâmina utilizando o clone repetitivo com sonda. O padrão obtido é muito semelhante ao do retrotransposon caracterizado neste trabalho. Foto: Stephan Nielen, 2005 (dados não publicados).

ANEXO I

Extração de DNA

Tampão CTAB 2x

2% de CTAB
100mM Tris-HCl (ph=8)
20mM EDTA (ph=8)
1,4M NaCl
1% PVP 40

Tampão CTAB 1x

1% de CTAB
50mM Tris-HCl (ph=8)
20mM EDTA (ph=8)

Soluções utilizadas no Dot blot

Solução de pré-hibridização e hibridização:

SSC 5x
0,1% N-laurilsarcosine
0,02% SDS
50% formamida
5% blocking reagent

Solução de lavagem I

SSC 2x
0,1% SDS

Solução de lavagem II

SSC 0,25x
0,1% SDS

Solução Ac. Máleico

Ac máleico ph= 7,5
Tween (20) 0,3%

Solução de detecção (ph= 9,5)

Tris-Hcl 0,1M
NaCl 0,1M

Soluções utilizadas no Southern blot

Solução de Depurinação

HCl 250 mM

Solução de Desnaturação

NaCl 1,5M

NaOH 0,5M

Solução de Neutralização

NaCl 1,5M

Tris-HCl 0,5M

Solução Denhardtts 50x (100mL)

1g de BSA

1g Ficoll 400

1g PVP 40

Dissolver em H₂O Mili-Q autoclavada

Solução de Esperma de salmão 4mg/mL (10mL)

40mg de espermidina

Completar com H₂O Mili-Q autoclavada

Solução de pré-hibridização (25mL)

SSC 20x----- 6,25mL

Denhardtts 50x----- 2,5mL

SDS 20%----- 0,625mL

Esperma de salmão 4mg/ml----- 120µL

H₂O----- 15,626mL

Solução de lavagem I (500mL)

SSC 1x

SDS 0,1%

Utilizar a solução a 65°C

Solução de lavagem II (500mL)

SSC 0,1x

SDS 0,1%

ANEXO II

Seqüências FASTA dos clones do retrotransposon contendo o gene da enzima transcriptase reversa utilizadas no trabalho.

A. duranensis

>Ad-Athila-1

```
tctaaatccagccatgaaggagggtggtgcagaaagaagtcactaaatfactagaggctgggattattatcctatttctgatagcccctgggtgagccctgtccaagtc  
gttcccaaaaaggagggaatgacagtggttcataatgaaaaaatgaactggttcctataaggacagtcacaggggtggcgcagtgtattgactacagaaggctcaat  
acagccactagaaaagatcatttctttaccattcatagaccaaatgctagaagactagctggccatgattactgcttttggatggctattcaggctacaaccaa  
attgcagtagatcctcaggaccaagagaaaaacagcattcacttgccttctggcgtgttgcctatagaggatgccttttggctgtgcaatgcacctgcaaccttca  
gagggtgcatgctctctatctctcagataggtaga gaaatttctggaagtcttcatggatgactttcagtataaggagactcattcagctcctgtcttaacacctagcac  
ttgcttgaaaagggtgccaagagactaacctggttttaactgggagaaatgctactttatggtgactgaaggaaatgtccttgggcacaaaattcaacaagggaata  
gagggtggataaggcaaaaggtaggtaattgaaaaattaccaccacctgccaatgtaaggcaatcagaagccttctggggcatgcagggttctacagaagggttat  
aaaggattttcaaaaattgcaaacctctgagtaacctgttagctgctgacacaccattgtgttgacacacagtgctgcaggcgtttagaccctgaaagcctaagc  
tggtcagcagaccagtcactctgcatcagattggacattgccattgaaactgatgtgta
```

>Ad-Athila-4

```
ctaaatccagccatgaaggagggtggtgcagaaagaggctactaaatfactagaggctgggattattatcctatttctgatagcccctgggtgagccctgtccaagtg  
tccccaaagaggaggcatgacagtggttcataatgaaaaaatgaactggttcctacaagaacagtcacaaggtggcgcagtgtattgactacagaaggctcaat  
acagccaccagaaaagatcatttctttaccattcatagaccaaatgctagaagactagctggcctgattactgcttttggatggctatttagttacaaccaa  
tgcaaaagatcctcaggaccaagagaaagacagcattcacttgccttctggcgtgttgcctacagaagaatgccttttggctgtgtaatgcacctgtaacttttagag  
atgcatgctatccatctctctgataggtagagaaagtcctggaagtctcatagatgacttctcagtataaggagactcattcagctcctgtcttaattatctagcactgt  
cctgaagggtgccaagagactaacctggtcttaactgggagaaatgctactttatggtgactgaaggaaatgtccttgggcacaaaattcaggcagaagaataaag  
gtggatcaagcaaaaggtagaagtaattgaaaaattaccaccacctaccatgcttaaggcaatcagaagccttctagggcctgcaggatttcaaaaagggtcataaag  
aacctttcaaaaattgcaaacctctagcaacctgctgctgacacgcaacttgggttggacacagagtgctgcaggcgtttagaccctgaaagcctaagcctgg  
tcacagcaccagtcactctgaccagactggacattaccattgaaactgatgtgta
```

>Ad-Athila-5

```
ctaaatccagccatgaaggagggtggtgcagaaagaggctactaaatfactagaggctgggattattatcctatctctgatagcccctgggtgagccctgtccaagtg  
tccccaaaaaggaaggcatgacagtggttcataatgaaaaaatgaactggttcctacaagaacagtcacaggggtggcgtatgtgtattgactacagaaggctcaat  
acagccaccagaaaggatcatttctttaccattcatagaccaaatgctagaagactagctggcctgattactgcttttggatgggtattcaggttacaaccaa  
ttgcagtagatcctcaggaccaagagaaaacagcatttacttgccttctggcgtgttgcctatagaggatgccttttggctgtgtaatgctctgcaaccttccaga  
gatgcatgttatccatctctctgataggtggagaaattcctggaagtcttcatggatgacttctcagtataaggagactcattcagctcctgtcttgaccacctgacact  
gtcctaaaaagggtccaagagactaacctgattttaaactgggagaaatgctactttatggtgactgaaggaaatgtccttgggcacaaaattcaagcagggggaata  
gagggtggataaggcaaaaggtaggtaattgaaaaattaccaccacctgccaatgtaaggcaatcagaagccttctggggcatgcaggattctatagaagggttata  
aaggattttgaaaatgcaaacctttagcaacctgctagctgctggacacaccattgtgttgacacacagtgctgcaggcatttagaccctgaaagcctaagc  
tggtcagcagcaccagtcatttctgaccagactggacattgccattgaaactgatgtgta
```

>Ad-Athila-6

```
ctaaatccagccatgaaggagggtggtgcagaaagaggctactaaatfactagaggctgggattattatcctatttctggtagcccctgggtgagccctgtccaagtg  
tccccaaaaaggaaggcatgacagtggttcataatgaaaaaatgaactggttcctacaagaacagtcacaggggtggcgtatgtgtattgactacagaaggctcaat  
acagccaccagaaaggatcatttctttaccattcatagaccaaatgctagaagactagctggcctgattactgcttttggatggctactcaggttacaaccaa  
attgcagtagatcctcaagaccaagagaaaaacagcatttacttgccttctggcgtgttgcctacagaggatgccttttggctgtgtaatgctctgcaaccttccag  
agatgcatgttatccatctctctgataggtggagaaattcctggaagtcttcatggatgacttctcagtataaggagactcattcagctcctgtcttaacacctgacact  
gtcctgaaaagggtccaagagactaacctggttttaactaggagaaatgctactttatggtgactgaaggaaatgtccttgggcacaaaattcaagcagggggaatag  
agggtgataaggcaaaaggtaggtaattgaaaaattaccaccacctgccaatgtaaggcaatcagaagccttctggggcatgcaggattctacagaagggttata  
aaggattttgaaaatgcaaacctttagtaacctgtagctgctgacacaccattgtgttgacacacagtgctgcaggcatttagaccctgaaagcctaagc  
gttcacagcaccagtcactctgaccagattggacattgccattgaaactgatgtgta
```

>Ad-Athila-8

ggagatgatagggtgcagaaggaggtcactaaactagaggctgggattatctctatctctgatagcccctgggtatgccctgtccaagtcgtccctaagaaggg
tggcatgacagtggttcataatgaaaaatgtactggtccaacaagaacggtacaggggtggcgtatgtgtattgattacagaaggctcaatacagctaccagaaag
gatcatttctttaccatcatagactagatgctagaaaattaacaggtcatgaatactactgctcttggatggatattcagggtataatcaaattgtatgatccccag
gatcaagaaaaatagctttcacatgtccatctggagatattgcatacagaaggatgcatttggctgtgcaatgcacctgtaacctttcaaaagatgctgtctctatctt
ctctgctatggtgaaaaatcttagaagtcttatggatgactttcaatatttggagactcattcagctcctgtcttgaccttagcactgttctaaagggtgccaaag
agactaacctggtttaaactgggagaaatgtcagttatggtgactgaaggaaattgctctgggcacaaaatcgaacaagggaatagaggtggatcaagctaaagg
tagaggtaattgaaaaataccaccactgccaatgttaaggcaatcagaagctttctggggcatgcaaggattctatagagggtttataaaagattttcaaaaatcgc
aaacctctgagtaactctgtagctgtgacacactattatctttgataaggagtgctgcaggcgtttaacactctgaaagtaagctggtcacaacaccagctatct
gcaccagattgaacattaccattgaactgatgtgtaa

>Ad-Athila-12

gtacgcacaagatcctattggagggtgacgctaagccaatggtcattcacaagggcgtgaatctagccatgaaggaggtggtgcaagaaggaggtcactaaatt
actagaggctgagattatctccatctctgatagaccctgggtaagccctgtccaagtcgtccctaagaagggtagcatgacagtggttcataatgaaaaatgaa
ctggttctacaagaacagttacaggatggcgtatgtgtattgattcgaaggctcaatacagctaccagaaagatcgtttcctttaccattcaaaagaccagatgct
agaagactagcaggctcatgaatactactgctcctggatagatattcagggtataatcaaattgcagtagatccccaggatcaagagaaaaacagcattcacatgac
atctagagattcacatacagaaggatgcatttggcctgtgcaatgcacctgcaacctttcaaaagggtgcatgtttctatttttggataggtggaaaaatcttgaa
gtcttcatggatgactttcagtttggagactcattcagctcctgtctgacacctgacactgtgcttaaaagggtccaataacactggttttaaaactgggaaaa
atgtcactttatggtgactgaaggaaattgcttgggcacaaaatcgcataaagggaatagaggttaagtagaggttaattgaaaaataccaccactgccaatgttaa
ggcaatcagaagcttctggggcatgcaaggattctatagagggtttataaaagattttcaaaaatgccaacactctgagtaactctgtagctactgacacgccattatct
ttgataaggagtgctgcaggcgttgagactctgaaagcctaagctggtcacagcaccagctatctctgcaccagactggacactaccattgaactgatgtgtaa

>Ad-Athila-13

ctaaatccagccatgaaggaggtggtgcagaaaagggtcactaaactagaggctgggattatctctctatctctgatagcccctgggtaagccctatccaagtc
gtccctaagaaggccatgacagtggttcataatgaaaaaatgaactggttctacaagaatagttacaggggtggcgtatgtgtattgattatagaaggctcagt
acagccaccagaaggataatctttaccattcatagacctgatgctagaagactggcaggtcatgaatactactgctcctagatggatattcagggtataatcaa
attgcagtagatccccaggatcaagagaaaaacagcattcacatgctaccggagatttgcatacagaaggatgccatttggctgtgcaatgcacctgcaaccttc
aaagggtcatgctctatcttcttaatatggtgaaaaatcttgaagcttcatggatgacttttagtatttggagactcattcagctcctgtctgacctatgacact
tgttctaaagaggtgccaagagactaacgggttttaactggaaaaatgtcactttatgtgactgaaggaaattgcttgggcacaaaatcgaacaagggaatag
agggtgatcaagctaaaggaaaaataccaccactgccaatgttaaggcaatcagaagcttctggggcatgtaggattttatagagggtttataaaaggatttcaaa
aattgccaacactctaaagtaactctgttagctgctgacatgctattatctttgataaggagtgctacagccattgagactctgaaagctaaattgctcacagccaatc
atctctgcaccaaacctggacattaccattgaactgatgtgtaa

>Ad-Athila-16

ctaaatccagccatgaaggaggtggtgcaaaaagggtcactaaactagaggctgggattatctctctatctctgatagcccctgggtgagccctgccaaagt
gtccccaaaaaggaggtcatgacagtggttcataatgaaaaaatgaactggttctacatgaacagtcacaggggtggcgtatgtgtattgactacagaaggctcaa
tatagccaccagaaggatcatttctttaccattcatagaccaatgctagaaagactagctggtcatgattactgcttttggatggctattcaggttacaaccaa
attgcagtagatctcaggaccaagagaaaaacagcattactgcccctctggcgtgttggctacagaaggatgccccttggctgtgtaattgctcctgcaaccttcag
agatgcatgctatctctctgatagggcggagaaatcttgaagctttatgtagacttctcagtatatggagactcattcagctcctgtcttaaccacctatcact
gtcctgaaaaagggtccaagagactaacctggttttaactgggaaaaatgtcactttatggtgactgaaggaaattgcttgggcacaaaatcgaacatgggaatag
agggtgataaggcaaggtagaggttaattgaaaaataccaccactgccaatgttaaggcaatcagaagcttctggggcatgcaaggattctacagaagggttataa
aggattttcgaaaatgcaaacctttgagtaacctgctagctgctgacacacactttgtgtttgacacacagtgctgcaggcatttgagaccctgaaagctaaagctg
gtcacagcaccagctctctgcaccagattggacattaccattgaactgatgtgtaa

>Ad-Athila-2

aaggacacacaagacagctctgggtggtcaatcagtgatcttaaggcattagcccagccagatgcatgcacaagatcttactggaggatgacgccaagccagtg
gttcaaccacaaaaggcgtgaatccagccatgaaaaagtggtgcagaaagggtcactaaactagaggctgggattatctctatcttctgacagcccctgg
gtgagcccctgtccaagtgccctaaagaaagggtgcatgagtggtttataatgaaaaaatgaactggttctacaagaacagttacaggggtggcgtatgtgtattg
attatagaaggctcaatacagccaccagaaaggatcatttctttaccattcatagaccagatgctagaagactagcaggctcatgaatactactgctcctggatgga
tattcagggtataatcaaattgcagtagatccccaggatcaggagaaaaacaacattcacatgccatctggagatttgcatacaggagataaccattggcctgtgca
atgcacctgcaactttcagaggtgcatgctcctaatctctgataggtgaaaaatcttggaaagcttcatgtagactttcagttatgagactcattcagctcct
gccttaaccatttagcactgttctgaaaagataccaagagactaacctggttttaaaactgggaaaaatgtcactttatggtgactgaagggtgctctgggcacaaa
attcgaacaagggaatagagggtgatcaagctaaaggtagaagtaattgaaaaataccaccactgccaatgttaaggaaatcagaagcttctggggcatgcaagg
attctatagagggtttataaaaggattttcaaaaatcgaacactctgagcaacctgctagctgctgacacgccatttatattgataaaagagtgctgcaggcattgag
actctgaaagctaaattggttacagccaatcatctctgcaccagactggacattaccattgaactgatgtgtaa

>Ad-Athila-3

tgtacgcacaagatcctattggaagataatgccaagccagtggttcaaccacagaggcgctaaatccaacctgaaggaggtggtgcagaaaaggctactaag
ttagtagaggtggtgattatctatctattctgatagccctgggtgagccctgtcacgtgtccccaaaaaggagcatgacagtggtcataatgaaaagaatga
actggtctcacaagaacagttataggggtgctatgttactacagaaggtcaatacagccaccagaaggatctttccctaccatcatagaccagatgct
agagagactagcaggctcatgaatactactgcttttgatggctatfcagggttacaacaaatgtatgtagatcctcaggaccaagaagaaacagcattcacatgctct
ctggagattttgcttacagaaggatccctttggtctgtgtattgcacctgcaaccttttagaggtgcatgctttctatctctctgataggtagagaagttctggaagtc
ttcatgtagacttctcagatatttgagactcattcagctcctgtctgtatcatttagcacttgcctgaaaagggtgccaagagaccaacctagtttaaacctggaaaaaat
tcactttatggtgactgaagggtgattgctattgggcataaaatfcaaacaaagggaatagaggtatgataagctaaagtgaagtaattgaaaaatfaccaccacctgcc
aatgtaaggcaatcagaagctttctggggcatgcaagattctacagaaaagttataaaggattttcaaaaattgcaaacctctgagtaattctgtagctgctgacatg
ccattatctttgataaggaggtttgcaggcgttgaaccttgaagctaaagctggtcacagcaccagtcattctgaccagactggacattaccattgaaactgatg
tgtga

>Ad-Athila-7

aaggacacacaagacagctctgggtggtccatcagtgatcttaaggcattagcccgccagatacatgaagaagatcctattggagggtgatgtaagccagtg
gtfcaaccacagtgacggctaaatccagccatgaaggaggtggtgcagaaaagggtcactaaatfactagaggctgggattatctatcttttgatagccctggg
ttagccctgtccaagtcacctaagaagggtggcatgacagtggtcataatgaaaaatgaactggtctcacaagaacagttacaagggtgacgtatgtgtattgatt
atagaaggctcaataatagctaccagaaaatgatctttctttaccattcatagaccagatgctagaaagactagcagggtcatgaatactactgcttctggatgaatattc
aggttataatcaaatgtagtagatcccaggatcaagagaaaacagcattcacatgctcattgagatattgcatacagaagaatgcccatttggatgtgcaatgca
cctgcaacctttcaaaaggtgctctctatttctctgataatgggaaaaatfctggaaagcttctcattgacttttcagtggttgaagactcattcagctcctgtctg
accatctagcactgttctaaagagatgccaagagactaacctggtttaaactgggagaaatgctcactttatggtgactgaaggattgctctgggcacaaaattctg
aacaagggaatagagggtgatcaagctaaaggtagagtaattgaaaaatfaccaccacctgccaatgtaaggcaatcagaagctttctggggcatacaggattcta
taggaggtttataaaagattttcaaaaatcgcacaaacctctgagcaatctgtagctgctgacacgccatttatctttgataaggactgtctcaagagttgagactcta
aaagctaaagctggtcacagcaccagtcattctgaccagactagacattaccattgaaactgatgtgtgaa

>Ad-Athila-9

aaggacacacaagacagctctgggtggtccataagtatcttaaggcattagcccgcaagatgcatgcacaagatcctattggaagataatgccaagccagtg
gtfcaaccacagaggaggtctaaatccagccatgaaggaggtggtgcagaaaagggtfactaaatfactagaggctgggattatctatcttttgatagccctgag
gtgagccctgtccaagttgtcccaagaaggagggcatgacagtggtcataatgaaaagaatgaactggtcgcacaagaacagttacagggtgctgctatgtgtat
tgactacagaaggctcaatacagccactagaaggatctttctttaccattcatagaccaaatgctagaaagactagctggtcatgattattactgcttttgatgg
ctattcaggttacaacaaatfcaatagatcctcaggaccaagagaaagacagcattcacatgctctctggagattttgcttacagaaggatgactttggtctgtgta
tgcaactgtaactttcagagatgcatgctatccatcttctttgataggtagagaagttctggaaagcttcatggatgacttctcagtatatggacacttattcagctcctg
tcttaatcatctagcactgtctgataaggtgccaagagactaacctggtcttaaacctgggagaaatgctcacttcatgggtgactgaaggattgctctgggcacaaaa
tttaggcaagggaatagagggtgatcaagcaaggtagaagtaattgaaaagttccaccacctaccaatgtaaggcaatcagaagctttctaggcatgcaagaa
ttacataaggttcataaaggactttcaaaaatfcaaaccttfaaaaacctctagctgctgacacgccattgtgttgacacaaagtgtctgcaaggcgtttgagac
cctgaaaagcaagctggtcacaacaccagtcattctgaccagattggacattaccattgaaactgatgtgtgaa

>Ad-Athila-10

aaggacacacaagacagctctgggtggtccataagtatcttaaggcattagcccgctagatgcatgcacaagatcctgtggaggataatgccaaccagtg
gtccaaccacagaggaggtctaaatccagccatgaaggaggtggtgcagaaaagggtcactaaatfactagaggctgggattatctatcttttgatagccctggg
gtgagccctgtacaagttgtcccaagaaggagggcatgacagtggtcataatgaaaaaatgaactggtctcacaagaacagtcacagggtggcgtatgtgtat
tgactacagaaggctcaatacagccaccagaaggatctttctttaccattcatagaccaaatgctagaaagactagctggtcatgattattactgcttttgatgg
ctattcaggttacaacaaatfcaatagatcctcaggaccaagagaaaacagcattcacatgctctctggagattttgcttacagaaggatgactttggtctgtgta
atgctcctacaacctttcagagatgcatgtattctgtctctctgataatgggagaaatfctggaaagcttctcattgacttctcagtatatggagactcattcagctcc
tgtcttaatcacctagcactgtctgaaaagggtgccaagagactaacctggtttaaactgggagaaatgctcactttatggtgactgaaggattgctctgggcacaaa
attfcaagcagggaatagagggtgataaggaaggttagagtaattgaaaaatfaccaccacctgccaatgtaaggcaatcagaagctttctggggcatgcaag
gattctacagaagggtttataaaggattttcgaaaaatfcaaacctttgagtaacctgctagctgctgacacaccatttatgtttgacacatagtgctgtaggcatttgag
acctgaaaagctaaagctggtcacagcaccagtcattctctacaccagattggacattgccatttgaactgatgtgtgaa

>Ad-Athila-11

aaggacacacaagacagctctgggtggtccataagtatcttaagggtattagcccgctagatgcatgcacaagatcctattggaggataatgccaaccagtg
tccaaccacagaggaggtctaaatcctgcatgaaggaggtggtgcagaaaagggtcactaaatfactagggctgggattatctatcttttgatagccctggg
tgagccctgtccaagttatcccaaaaaggagggcatgacagtggtcataatgaaaaaatgaactggtcctcacaaggacagtcacagggtggcgtatgtgtatt
gactacagaaggctcaatacagccaccagaaggatctttctttaccattcatagaccaaatgctagaaagactagctggccatgattattactgcttttgatgg
ctattcaggttacaacaaatfcaatagatcctcaggaccaagagaaaacagcattcacatgctctctggagattttgcttacagaaggatgacttttggtctgtgc
aatgcaacctgcaacctttcagaggtgcatgctctctatctctcagatattgtagaagaatttggaaagcttctcattgactttcagtatatggagactcatttaact
ctgtcttaatcacctagcactgtctgaaaagatgccaagagactaacctggtttaaactgggaaaaatgctcactttatggtgactgaaggattgctctgggtgtaaa
aattfcaagcaagggaatagagggtgataaggaaggttagagtaattgaaaaatfaccaccacctgccaatgtaaggcaatcagaagctttctggggcatgca
gggttctacagaagggtttataaaggattttcgaaaaatfcaaacctctgagtaacctgttagctgctgacacaccattgtgtttgacacacagtgctgcaaggcatt
gagacctgaaaagctaaagctggtcacagcaccagtcattctctacaccagattggacattgccatttgaactgatgtgtgaa

>Ad-Athila-14

aaggacacacaagacagctctgggtggtccatcagtgatctcaaggcattagcccagccagatgcatgcacaagatcctattggagggtgacgccaaagccagt
ggttcaaccacaaaggcggctgaatccatccatgaaggaggtggtgcagaaggaggtcactaaactagaggctgggattatctattctgacagcccctg
ggtgagccctcaagttgctcctaagaaaggtgcatgacagtggtcacaatgaaaaaatgaactggtcctacaagaacgttacagggtgctgatgtgat
tgattatagaaggctcaatacagccaccagaaaggaccattttcctaccatcatagatcaaatgctagaagactggcaggctgaatactactgcttctggtg
gatatttaggtataatcaaatgacagtagatccaaaagatcaagagaaaacggcattcacatgctccatctggagatattgcatacagaaggatgccattggtctgtgc
aatgcacctgcaacctttcagaggtgatgctctcaatttcttgatggtgaaaaagtctggaagtctcatgaatgactttcagatttggagactcattcaactcct
gtcttgaccatctagcactgttctaaagagatgccaagagactaacctggfctaaactgggaaaaatgctactttatggtgactgaaggaattgctctgggcacaaa
atctcgaacaaggggaatagagggtgatcaagctaaggtagaggaattgaaaaatfaccaccacctgccaatgtaaggcaatcagaagctttctggggcatgcag
gattctataggaggttataaaggattttcaaaaatcgcacaacctctgagcaacctgctagctgctgacacgccatttctttgataaggagtgtctgcaggcatttga
gactctgaaagctaaattggtcacagcacaatcatctctgcaccagactggacattaccattgaaactgatgtgtgaa

>Ad-Athila-15

aaggacacacaagacagctctgggtggtccatcagtgatcttaaggcattagcccagccagatgcatgcacaagatcctattggagggtgacgccaaaccagt
gtccaaccacaaagggtggtgaatccagccatgaaagaaggtggtgcagaaggaggtcactaaactagaggctgggattatctattctgacagcccctgg
gtgagccctgtccaagttgctcctaagaaagatggcatgacagtggtcacaatgaaaaaatgaactggtcctacaagaacgttacagggtgctgatgtgattg
attatagaaggctcaatacagccaccagaaaggatcatttctaccatcatagaccagatgctagaagactggcaggctgaatactactgcttctggtgatg
atattcagggtataatcaaatgcaatagatccaaaagatcaagagaaaatggcattcacatgctccattggagatattgcatacagaaggatgccattggtctgtgcaa
tgcacctgcaacctttcagaggtgcatgctctcaatttctctgatggtgaaaaatfctggaagtcttcatggtgactttcagatttggagactcattcagctcctg
cttgaccatttagcactgttctaaagagatgccaagagactaacctggfctaaactgggaaaaatgctacttcatggtgactgaaggaattgctctgggcacaaaat
ctcgaacaaggggaatagagggtgatcaagctaaggtagaggaattgaaaaatfaccaccacctgccaatgtaagcaatcagaagcttttggggcatgcaggat
tctataggaggttataaaggattttcaaaaatcgcacaacctctgagcaacctgctagctgctgacacgccatttctttgataaagagtgtctgtaggcatttga
ctgaaagctaaattggtcacagcacaatcatctctgcaccagactagacattaccattgaaactgatgtgtgaa

>Ad-Athila-19

ggagatgatagggtgcagaaggaggtcactaaactagaggctgggattatctattctgacagcccctgggtgagccctgtccaagttgctcctaagaaag
gtggcatgacagtggtcataatgaaaaaatgaactagttcctacaaggacgttacagggtgctgatgtgactgattacagaaggctcaatacagccaccaga
aaggatcattttcttaccatcatagaccagatgctagaagactagcaggctcatgaatactattgcttctggtgatattcagggtataatcaaatgacagtagatcc
ccaagatcaggagaaaacggcattcacatgccatctggagatattgcatacagaaggatgccatttggcctgtgcaatgcacctgcaacctttcagaggtgcatgct
ctcaatttctctgatggtgaaaaatctggaagtcttcatggtgactttcagatattggaagactcattcagctcctgcttaaccatttagcactgttctgaaaag
atgccaagagactaatctggtttaaactgggaaaaatgctactttatggtgactgaaggaattgctctgggcataaaaattcgaacaaggggatagagggtggccaa
gctaaggtagaagtaattgaaaaatfaccaccacctactaatgtaaggcaatcagaagcttttggggcatgcaggattctataggaggttataaaggattttcaaa
aatcaccaaaacctgagcaacctgctagctgacagacgccatttctttgataaagagtgtctgcaggcatttgaagactctgaaagctaaattggttacagccca
atcatctctgcaccagactggacattaccattgaaactgatgtgtgaa

>Ad-Athila-21

ctaatccagccatgaaggaggtgtacagaaagggtcactaaactagaggctgggattatctattctgatagcccctgggtgagccatgtccaagttg
tcccaagaaggaggcatgacagtggtcataatgaaaaaatgaactggtcctacaagaacaattatagggtgctgatgtgattgattacagaaggctcaata
caaccaccagaaggatcatttcttaccatcatagacaaaatgctagaagactaacctggcatgattattactacttttggatgctattcaggtfacaaccaat
gcagtagatcctaaaggactaagagaaaaacagcattcacttccccttctggcgtgtttgcttacagaagaatccctttggtctgtgtaatgcacctgcaacctttcagag
atgcatgctatccatctctctgatatagtagagaagttcctagaagttcctatggtgacttctcagtatatggagactcattcagctcctgtcttaatcatctagcattgt
cctgaaaagggtccaagagactaacctggcttgaactgggaaaaatgctactttatggtgactgaaggaattgctctgggcacaaaatfaccaggaagggaatg
agggtgatcaagcaagatagaagtaattgaaaaatfacaaccacctgccaatgtaaggcaatcagaagctttctagggcctgcaagatttacaagaagggtcataa
aggattttcaaaaatgcaaaccttaagcaacctgctagctgctgacacaccattgtgtttgacacagagtgtctgtaggcgtttgagaccctgaaagccaagctg
gtcacagcaccagctcatctctgcaccagattagacattaccattgaaactgatgtgtgaa

>Ad-Athila-22

aaggacacacaagacagctctgggtggtccataagtgatcttaaggcattagcccagcaagatgcatgcacaagatcctattggaagataatgccaagccagt
gttcaaccacagaggaggttaaatccagccatgaaggaggtggtgcagaagaagggtactaaactagaggctgggattatctattctgatagccctggtgagccatgtccaagttg
gtgagccctgtccaagttgctcctaagaaaggaggtgcatgacagtggtcacaatgaaaaaatgaactggtcctacaagaacgttacagggtgctgatgtgattg
tactacagaaggctcaatacagccactagaaggatcatttcttaccatcatagacaaaatgctagaagactagctggtcattgattactgcttttggatgg
ctattcaggtfacaaccaatfcaatagatcctcaggaccaagagagacagcattcacatgctctctggagatattgcttacagaaggatgactttgctctgtgtaa
tgcacctgtaactttcagagatgcatgctatccatcttcttgatggtgagaaagtctggaagtcttcatggtgacttctcagtatatggacacttattcagctcctg
tcttaatcatctagcactgtctgataagggtccaagagactaacctggctttaaactgggagaaatgctacttcatggtgactgaaggaattgctctgggcacaaa
tttaggcaaggggaatagagggtgatcaagcaaggtagaagtaattgaaaaatfaccaccacctaccatgtaaggcaatcagaagcttttagggcatgcagaat
ttacataaggttcataaaggactttcaaaaatfcaaaccttfaagaaacctctagctgctgacacgccatttgtgtttgacacagagtgtctgtaggcgtttgagaccctgaaagccaagctg
cctgaaagccaagctggtcacacaccagctcatctctgcaccagattggacattaccattgaaactgatgtgtgaa

>Ad-Athila-26

aaggacacacagacagctctgggtggtcatcagtgatcttaagggcattagctcagccagatgcatgcacaagatccaattggagggtgacccaagccaatgg
tccaaccataaaggcggctgaatccagccatgaaggaggtggtgcagaaggaggtcactaaactactagaggctggattatctatctgacagcccctggg
taagccctgtccaagtcgtccctaagaaggaggtgacagtggttcataatgaaaaaatgaactggtcctacaagaacagttacagggtggcgtatgtgattg
attatagaaggctcaatacagccaccagaaaggatcatttcttaccatcatagaccagatgctagaagactagcaggctcatgaatactactgctcctggtgga
tatttaggtataatcaaattgcagtagatcccaagatcaagagaaacggcattcacatgctcatctggagatttgcatacagaaggatgccatttggcctgtgcaa
agcacctgtaaccttccagagggtgcatgctctcaatttctctgatatggtggaaaaatttctggaagtcttcatggatgactttcagatttggagacttattcagctcctg
tcttaaccatctagcactgttctaaagagatgccaagagattaacctggtttaaactgggagaaggtgacactttatggtgactgaaggaattgctctgggcacaaaat
tcaacaagggaatagagggtgatcaagctaaaggtagaggtaattgaaaaattaccaccacctgccaatgtaaggcaatcagaagcttctggggcatgcagaat
tctataggaggttataaaggattttcaaaaatcgccaacctctgagcaacctgctagctactgacacgccatttctttgataacgagtgctgcaggcatttgaga
ctctgaaagctaattgtgcacaacaccaatcatctctgcaccaactggacattgccattgaaactgatgtgtgaa

>Ad-Athila-23

aaggacacacaagacagctctgggtggtcaatcagtgatcttaagggcattagcgcagccagatgcatgcacaagatcctattggaggatgacgtaagccagtg
gttcaaccacagaggcggctgaatccagccatgaaggaggtggtgcagaaggaggtcactaaactactgaggctgggattatctatctgtagactcctgg
atgagccctatccaagtcgtccctaagaagggtgcatgacagtggttcataatgaaaaaatgaactgattcctacaagaacagttacagggtggcgtatgtgatt
gattacagaaggctcaacacagctaccagaaaggatcatttcttaccatcatagaccagatgctagaagactagcaaggttgaatactactgctcctggatg
gatattcagggtataatcaaattgcagtagatcccaagatcaagagaaacaacttccatgctcatccggagatttgcatacagaaggatgctatttggctgtgca
atgcacctgcaaccttcaaaagggtgcatgctctcttctctgatatagtggaaaaatttctggaagtcttcatggatgactttcagatttggagactcattcagctcct
gtcttgaccatctagcactgttctaaagaggtgccaagagactaacctgtttaaactgggagaaggtgacactttatggtgactgaaggaattgctctgggcacaaa
atttcaacaagggaatagagggtgataaggcaaggtagaggttaattgaaaaattaccaccacctgccaatgtaaggcaatcagaagcttctggggcatgcag
gattctacagaagggtcataaaggattttcaaaaatcgcaaacctttagtaacctgctagctgctgacacaccatttgtttgacacacagtgctgcaggcatttga
gacctgaaagctaagctggtcacagcaccagtcctctgcaccagattggacattgccattgaaactgatgtgtgaa

>Ad-Athila-24

ctaatccagccatgaaggaggtggtgcagaaagaggtcactaagttactagaggctgggattatctatctgtagcccctgggtgagccctgtccaagtg
tccccagaaggaggtgcatgacagtggttcataatgaaaaatgaactggtcgtgcaagaacagttacagggtggcgtatgtgattgactacagaaggctcaata
cagccaccagaaaggatcatttcttaccatcatagaccagatgctagagagacagcaggtcatgaatattattgcttttggatggctattcaggttacaccaaat
tgcagtagatcctcaggaccaagagaaaacagcatttcatgctctctggagatttgcctacagaaggatgcttttggctgtgcaatcgccctgcaacctttaga
ggtgcatgctctctatctcttagataggttagagaaatttctggaagtcttcatgagatgactttcagatttggagactcattcaactcctgcttaacctctagcactgt
tctgaaaagatgccaagagaccaacctggtttaaactcgcaaaaatgacacttcatggtgactaaggggatttgccttgggcataaaaattcaagcaagggaataga
ggtggatcaagcaaggtagaggttaattgaaaaattaccaccacctaccaatgtaaggcaatcagaagcttctgggacatgcaggattctataagaaggttataaa
ggattttcaaaaatgcaaacctctgagtaacctgctagctgctgacacaccatttgtttgacacacagtgctgcaggcattttagaccctgaaagctaaagctgg
tcacagcaccagtcctctgcaccagactggacattaccattgaaactgatgtgtgaa

>Ad-Athila-18

aaggacacacaagacagctctgggtgatccatcagtgatcttaagggcattagcccagccagatgcatgcacaaaatttattggagggtgacgctaagccagtgg
tcaaccacaaaggcggctgaatccagccatgaaggaggtggtgcagaaggaggtcactaaactactagaggctgagattatctatctgtagcccagg
gtaagccctgtccaagtcgttctaagaagggtgcatgacagtggttcataatgaaaaatgaaactggtcctacagaacagttacagggtggcgtatgtgattaa
ttacagaaggctcaatacagccaccagaaaggatcatttcttaccatcatagactagatgctagaagactggcaggtcatgaatactaccgcttctggtgga
tattcagggtataatcaaattgcagtagatcccaagatcaagagaaaatgacattcacatgctcatccggagatttgcatacagaaggacgccattttagctgtgcaa
tgcacctgcaaccttcaaaagggtgcatgctctcaatttctctgatatggtggaaaaatttctggaagtcttcatgagatgactttcagatttggagactcattcagctcctg
tcttgaccatctagcactgttctaaagaggtgccaagagactaacctggtttaaactgggagaaggtgacactttatggtgactgaaggaattgctctgggcacaaaat
ttcgaacaagggaatagaaggtgatcaagctaaaggtagaggtaattgaaaaattaccaccacctgccaatgtaaggcaatcagaagcttctggggcatgcaggat
tctataggaggttataaaggattttcaaaaatcgccaacctctgagcaacctgctagctgctgacacaccatttctttgataaggagtgctgcaggcatttga
ctaggaagccaagctggtcacacaccagtcctctgcaccagactggacattaccattgaaactgatgtgtgaa

>Ad-Athila-17

aaggacacacaagacagctctgggtggtctatcagtgatcttaagggcattaccagccagatgcatgcataagatcctattggagggtgacgctaagccagtgg
tcaaccacagaggcactgaatccagccatgaaggagctggtgcagaaggaggtcactaaactactagaggctgagattatctatctgtagcccagg
taagccctgtccaagtcgtccctaagaagggtggtatgacagtggttcataatgaaaaaatgaactgattcctacaagaacagttatagggtggcgtatgtgattg
attatagaagactcaatacagccaccagaaagaatcatttcttaccatcatagaccagatactagaagactggcaggtcataatactactgcttctggtgga
tattcagggtataatcaaattgcagtagatcccaagatcaagagaaaacagcattcacatgctcatccggagatttgcatacagaaggatgccatttggcctgtgca
atgcacctgcaaccttcaaaagggtgcatctctatcttctctgatggtggaaaaatttctggaagtcttcatggatgactttcagatttggagactcattcagctcct
gtcttgaccatctagcactgttctaaagaggtgccaagagactaacctggtttaaactcgcaaaaatgacacttcatggtgactgaaggaattgctctgggcacaaa
atcttgaacaagggaatagagggtgatcaagctaaaggtagaggttaattgaaaaattaccaccacctgccaatgtaaggcaatcagaagcttctgaggcatgcagg
attctataagaggttataaaggattttcaaaaatcgccaacctctgagtaacctgctagctgctgacacaccatttctttgataaggagtgctgcaggcatttga
ctttgaaagctaagctggtcacagcaccagtcctctgcaccagactggacattaccattgaaactgatgtgtgaa

>Ad-Athila-27

catgctcccggccgcatggcggccgcgggaattcgattgtacgcacaagatcctattggaggatgatgctaagccagtagtcaaccacagaggcggctaaatc
cagccatgaaaagggtggtgcagaaagggtcactaagttactagagcgggggattattatctatttctgatagcccctgggtaagccctgtccactgttcccca
aaaaggagggcatgacagtggttcataatgaaaagaatgaactggttcctacaagaacagttacaggggtgctatgtgtattgactacagaaggctcaatacagcc
accagaaaagatcattttcttaccattcatagaccagatgctagagactagcagggcatgaatactactgcttttggatggctattcaggttacaaccaattgca
atagatcctcaggaccaagagaaaacagcattcacatgctctctggagtattgctacagaaggatgcccttctgtgtatgacacctgcaaccttcacaagtg
attctctctatctctgatgatgggaaaagtctcagaagcttcatggatgactctcagtttggagactcattcagctcctgtcttgatcattagcactgtcctgaa
aagggtgcaagagactaacctggtctfaaacctgggaaaaatgctactttatggtgactgaaggaaatgctcctgggcacaaaatttagaacaaggaaataggggtg
atcaagctaagggtgaggtaattgaaaaattaccaccactgccaatgtaaggcaatcagaagcttctggggcatgagcattctacagaaggttataaagggtttt
caaaaattgccaacctctgagtaattactagctgctgacacaccatttatcttgataaggagtgctgcaggcggttgagactctgaaggtaagctggtcacagca
ccagctatctctgaccagactggacattaccattgaaactgatgtgtgaa

>Ad-Athila-28

ctaatccagccatgaaggaggtggtgcagaaagggtcactaattactagaggctgggattattatctatttctaataagcccctgggtgagccctgtccaagttg
tccccaaaaaggagggcatgacagtggttcataatgaaaaaatgaactggttcctacaagaacagtcacaggggtgctatgtgtattaactacagaaggctcaat
acagccaccagaaggatcattttcttaccattcatagaccaatgctagaaagactagctggtcatgattactgcttttggatggctattcaggtacaccaa
attgcatgagccctcaggaccaagagaaaatagcattacttgccttctggcgtgtttgctacagaaggatgcccttggctgtgcaatgacacctgcaaccttca
aaggtgcatgctctctatctctcagatattgtagaataattctggaagtctcatggtgacttctcagttacggagactcattcagctcctgtcttaacacctgca
ctgtcctgaaaagatgccaagagactaacctagtfttaaacctgggagaaatgctactttatggtgactgaaggaaatgctcctgggcacaaaattcaagcaggggaa
tagaggtgataaggcaaaaggtagagtaattgaaaaattaccaccactgccaatgtaaggcaatcagaagcttctggggcatgagcattctacagaagggtc
ataaaggattttcgaaaattgcaaacctttgagtaacctgtagctgctgacacaccattgtgtttgacacacagtgctgcaggcatttgagacctgaaagctaa
gctgtcacagcaccagctatctctgaccagattggacattgccattgaaactgatgtgtgaaactact

>Ad-Athila-29

ctaatccagccatgaaggaggtggtgcagaaagggtcactaattactggaggctgggattattatctatttctgatagcccctgggtgagccctgtccaagttg
tccccaaaaaggagggcatgacagtggttcataataaaaaaatgaactggttcctacaagaacagtcacaggggtgctatgtgtattgactacagaaggctcaat
acagccaccagaaggatcattttcttaccattcatagaccaatgctggaaagactagctggtcatgattactgcttttggatggctattcaggttacaaccaa
attgtagtagatcccaggaccaagagaaaacagcattacttgccttctggcgtgtttgctacagaaggatgcccttggctgtgtaatgctcctgcaacctttag
agatgcatgctatccatcttctgacatggtgagaaatcctggaagtctcatggtgacttctcagttatggagactcattcagctcctgtcttaaccacctagc
ttgtcctgaaaagggtgccaagagactaacctggtfttaaacctgggagaaatgctactttatggtgactgaaggaaatgctcctggacacaaaattcaagcaggggaa
agaggtgataaggcaaaaggtagagtaattgaaaaataaccaccactgccaatgtaaggcaatcagaagcttctggggcatgagcattctacagaagggtc
ataaaggattttcgaaaattgcaaacctttgagcaacctgtagctgctgacacaccattgtgtttgacacacagtgctgcaggcatttgagacctgaaagttaa
actgtgtcacagcaccagctatctctgaccagattggacattaccattgaaactgatgtgtgaaatca

A. hypogaea

>Ah-Athila-3

aaggacacacaagacagctctcgggtggtacataagtgatcttaaggcattagcctagcaagatgcatgcacaagatcctattggaggatgatgctaagccagtg
gttcaaccacagagggtgctaaatccagccatgacggagggtggtgcagaaagggtcactaagttactagaggctgggattattatctatttctagataaccctgg
gtgagccctgtccactgttccccaaagaggagggcatgacagtggttcataatgaaaagaatgaactagttcctacaagaacagttacaggggtgctatgtgtatt
gactatagaaggctcaatacagccaccaaaaaggatcattttcttaccattcatagaccagatgctagagagactagcaggtcatgaatactactgcttttagatgg
ctattcaggttataaccaaattgcagtagatctcaggaccaagagaaaacagcattcacatgctctctggagtatttgcctacagaaggatgcccttggctgtgca
atgcacctgcaacctttagaggtgcatgctctctctctgatattgtagaataatttggagcttcatgtagacttctcagttttggagactcattcagctctt
gccttaaccttgcacttgtctgaaaagatgccaagagaccaacctagtctaaactgggaaaaatgctactttatggtgactgaagggaattgctcctgggcataaa
atctcaacaagggaatagaggtgcatcaagctaaagtgaagtaattgaaaaattaccaccactgccaatgtaaggcaatcagaagcttctcgggcagcagc
attctaccgaagggttataaaggattttcaaaaattcaaaaacctttgagtaattctgtagctcctgacacaccattatctttgataaggagtgctgcaggcggttgaga
ctctgaaagctaagctggtcacacaaccagctatctctgaaccggactggacattaccattgaaactgatgtgtgaa

>Ah-Athila-5

aaggacacacaagacagctctcgggtggtccataggtgaccttaaggcataagcccagctagatacatgcacaaaatctattggaggataatgctaaactagtg
ttcaaccacagagagcgttaatccagccatgaaggaggtggtacagaaagggtcaccaaattattggaggctgggattattatctatttctgatagcccctggg
tgagccctgttcaagttgtccccaaaaaggagggcatgacagtggttcataatgaaaagaatgaactggttcctacaagaatgtagcaggggtgctatgtattg
attacagaaggctcaatatagccaccaaaaaggatcattttcttaccattcatagaccagatgctagaagattggcaggtcatgattactgcttttagatggcta
ctcaggctataaccagattgtagtagatcccaggatcaagagaaaacagcattcacatgctcctctggagtgttgcctatagaagatgccatttgggctgtgta
gctcctgcaaccttccagagatgcatgctctctctgatatttctctgatattggtgaaaagatttctggaagtctcatgtagacttctatgtagatggagactcattcagctctg
tctgatcatctgaaactagtctgaaaagatgccaagagaccaacctggtfttaaacctgggaaaaatgctactttatgccaagagaccaacctgatattggtgactgaa
gggattgctcctgggcataaaattcacacaagggaatagaggtgcatcaagtaaaaatagaggttaattgaaaaattaccaccactgccaatgtaaggcaatcaga
agcttctggggcatgagcattctataggaggttataaaggattttcaaaaattgcaaacctctgagcaactgtagctgctgacacgccattgtgttgacaca
gagtgctgcaggcggttgaatgctgaaagcaagctggtcacacaaccagttatttctgaccagactggacattaccattgaaactgatgtgtgaa

>Ah-Athila-6

aaggacacacaagacagctctggatggcctcatcagtgatcttaagggcattagccagatgcatgcacaagatcctattggagggtgatgctaagccagtggttcaaccacagaggcggctgaatccagccatgaaggaggtggcgagaaggaggtcactaaactactagaggctgggattatccatttctgatagcccctgggtaagcctgttcaagtcgctcctaagaagggtggcatgacagtggttcataatgaaaaaatgaactggttctacaagaacagttacagggtggcgtatgtgtattgattacagaaggctcaatacaactaccagaaaggatcatttctttaccattcatagaccagatgctagaagattagcaggtcatgaatactactactccgggatggaatttcagggtataatcaaatgtagtagatcccaggatcaagagaaaacagtggttcacatgctcatctggagatttgcatacagaaggatgacatttggctgtgcaatgcacctgcaacctttcagaggtgcatgctctccatttctctgatatggtgaaaaattctggaggtctcatagatgactttcagatttggagactcattcaactcctgtcttgaccatctagaacttgttctaaagggtgccaagagaccaacctggtttaaactgggagaaatgtcactttatggtgactgaaagaattgtccttgggcataaaatttcgaagggaatagaggtggatcaagctaaggtagaggtaattgaaaaattaccaccactagcctaattgaaaggcaatcagaagctttctggggcatgcaggattctataggaggttataaaagattttcaaaaatcaccacacctctgagtaatctgctagctgctgacacgccatttatctttgataagggtgtctgtcagggcgtttgagactctaaaaactaagctgtgcacagcaccagtcctctgcaccagactggacattaccattgaaactgatgtgtgaa

>Ah-Athila-7

aaggacacacaagacagctctgggtggcctcatcagtgatcttaagggattagcccagccagatgtacgcacaagatcctattggagggtgatgctaagccagtggttcaaccacagaggcggctgaatccagccatgaaggaggtggcgagaaggaggtcactaaactactagaggctgggattatccatttctgatatcccctgggtaagtcctgttcaagtcgctcctaagaagggtggcatgacagtggttcataataaaaaatgaactgattctacaagaacagttacagggtggcgtatgtgtattaattacagaaggctcaatacagctaccagaaaggatcatttctttaccattcatagaccagatgctagaagactagcaggtcatgaatactactactcctggatggatatttaggtataatcaaatgtagtagatccccagatcaagagaaaacagcgttcataatggttcacatgtggagatttgcatacagaaggatgcaatttggctgtgcaatgcaacttccagaggtgcatgctctccatttctctgatatggtgaaaaattctggaggtttcagatttgcatacagaaggatgctcttcaatgtttggagactcattcagctctgtcttgacctagcacttgttctaaagggtgccaagagaccaacctagttttaaactcagagaaatgtcattttatggtgactgaaagaattgtccttgggcataaaatttcgaagaagggaatagaggtggatcaagctaagatagaggaattgaaaaattaccaccagcctaattgaaaggcaatcagaagctttctggggcatgcaggattctataggaggttataaaagattttcaaaaatcgcacacacctctgagtaatctgctagctgctgacacgccatttatctttgataaggaggtgtctacaggcattgaggctctgaagctaagctggtcacagcactagtcctctgcaccagactggacattaccattgaaactgatgtgtgaa

>Ah-Athila-8

aaggacacacaagacagctctgggtggctctataagtgatctcaagggcattagcccagctagatagcacaagatcctgttggaggataatgccaaaccagtggtccaaccacaaaggaggctaaatccagccatgaaggaggtggcgagaagaggtcactaaactactagaggctgggattatctatcttctgatagcccctgggtgagccctgtccaagttgtcccaaaaaggagggatgacagtggtttataatgaaaaaatgaactggttctacaagaacagtcacagggtggcgtatgtgtattgattacagaaggctcaatacagccaccagaaaggatcatttctttaccattcatagaccaaatgctagaagactagctgtgcatgattattactctttttggatggctattcaggtcacaacaaatgtagtagatcctcaagaccaagagaaaatgacattactgcccctctggcgtgtttgcttacaggagatgcccctttggctgtgcaatgcacctgcaacctttcagaggtgcatgctctctatcttctcagatattgtagaagaaattctggaggtctcatggacgacttctcagtatattggagactcattcagctctgtcttaatcactagcacttgtcctgaaaaagggtgccaagagactaacctggtttaaactgggagaaatgtcactttatggtgactgagggaattgtccttgggcacaaaatttcaagcaagggaatagaggtggataaggcaaggtagaggaattgaaaaattaccaccacctgccaatgtaaggcaatcagaagctttctaggcgtgcaggattctacagaagggttcataaaaggattttcaaaaattgcacacacctttagtaacctgctagttgctgacacaccatttgtttgacacacagtgctgcaggcctttgagacctgaaagctaagctggtcacagcaccagtcctctgcaccagattggacattgcaatttgaactgatgtgtgaa

>Ah-Athila-9

aaggacacacaagacagctctgggtggcctataagtgatcttaagggcattagcccagctagatgcatgcacaagatcctgttggaggataatgccaaaccaatggttcaaccacaaaggaggctaaatccagccatgaaggaggtggcgagaagaggtcactaaactactagaggctgggattatctatcttctgatagccccgggtgagccctgtccaagttgtcccaaaaaggagggatgacagtggttcataatgaaaaaatgaactggttctacaagaacagtcacaggatggcgtatgtgtattgactacagaaggctcaatacagccaccagaaaggatcatttctttaccattcatagaccaaatgctagaagactagctggcatgattattactgctttttggatggctattcaggtcacaacaaatgtagtagaccctcaggaccaagagaaaatgacattactgcccctctggcgtgtttgcttacagaagatgcccctttggctgtgcaatgcaacttcaacctttcaagggtgcatgctctctatcttctcagatattgtagaagaaattctggaggtctcatggatgacttctcagtatattggagactcattcagctctgtcttaatcactagcacttgtcctgaaaaagggtgccaagagactaacctggtttaaactgggagaaatgtcactttatggtgactgaaagaattgtccttgggcacaaaatttcaagcaagggaatagaggtggataaggcaaggtagaggaattgaaaaattaccaccacctgccaatgtaaggcaatcagaagctttctggggcatgcaggattctacagaagggttcataaaaggattttcaaaaattgcacacacctttagtaacctgctagttgctgacacaccatttgtttgacacacagtgctgcaggcctttgagacctgaaagctaagctggtcacagcaccagtcctctgcaccagattggacattgcaatttgaactgatgtgtgaa

>Ah-Athila-11

aaggacacacaaaacagctctgggtggcctcatcagtgatcttaagggcattagcccagccaaatgcatgcacaagatcttactggaggatgacccaagccagtggttcaaccacaaaggcggctgaatccagccatgaaagaaggtggcgagaagaggtcactaaactactagaggctgggattatctatcttctgatagccccctgggtgagccctgtccaagttgtccctaagaagggtggcatgacagtggttcataatgaaaaataaactggttctacaagaacagttacagggtggcgtatgtgtattgattatagaaggctcaatacagccaccagaaaggatcatttctttaccattcatagaacaaatgctagaagactagcaggtcatgaatactactgcttctggatggatattcagcttataatcaaatgtagtagatcccaggatcaagagaaaacacattcactatgctcatctggagatttgcatacagaaggatgcaatttggcctgtgcaatgcaacttcaacctttcagaggtgcatgctctcaatttctctgatatggtgaaaaattctggaggtctcatggatgactttcagatttggagactcattcagctctgccttaaccatttagcacttgttctgaaaaagatgcaaaagagactaacctggtttaaactgggaaaaatgtcacttcatggtgactgaaagaattgtccttgggcacaaaatctcgaacaagggtgataaggtggatcaagctaaggtagaggaattgaaaaattaccaccacctgccaatgtaaggcaatcagaagctttctggggcatgcaggattctataggaggttataaaaggattttcaaaaattgcacacacctttagtaacctgctagttgctgacacaccatttgtttgacacacagtgctgcaggcatttggactctgaaagctaatttggtcacagcaccatcctctgcaccagactggacattaccattgaaactgatgtgtgaa

>Ah-Athila-13

gtacgcacaaaatcctatggagggtgacgccaagcctgtggtccaaccacaaaggcggctgaaccagccatgaagaaagtgtgcaaaaggaggctactaa
attactagaggctgggattatctatctatctgtacagcccctgggtaagcctgtccaagctgcctcaaaaaggaggcatgacagtgttcacaatgaaaaaat
gaaactgttctacaagaacagttacagggtgctgtatgtattgattatagaaggctcaatacagccaccagaaaaggatcatttctttaccatcatagaccacaa
gctagaagactagcaggctatgaatactactgctcctggatggatattcaaggftataatcaaatgcaatagatcccaggatcaggagaaaacggcattcacctg
cccctctggagtattgcatacagaaggatgccattggcctgtgcaatgcacctacaactttcagagggtgcatgctcctcaatttctctgatattgtgaaaaattcctg
gaagtctcatggatgactttcagatattggagactcattcagctcctgccttaaccatttagcactgttctaaagagatgccaagagactaacctggtttaaactggg
agaagtgtcatttatggtagctgaaggatgtccttgggcacaaaatctcgaacaagggtataggtgatcaagctaaaggtagaggtaattgaaaaattaccac
cacctgccaatgttaacgcaatcagaagccttctgggcatgcaagattctatagagggttataaaggattttcaaaaattgccaacacctgagcaacctgctagct
actgacacgccatttatcttgataaagagtgtctgagggcgtttgagactctgaaagctaaattggttaaacaccaatcatctctgcaccagactggacattaccatt
gaactgatgttgaa

>Ah-Athila-1

ctaatccagccatgaaggaaagtgtgacagaaagggtcaccacaaactactagaggctgggattatctatctatctgtacagcccctgggtgagccctgtccaagtt
gtccccaaaaaggaggatgacagtgttcagaatgaaaaaatgaactggttcctacaagaacagttacaggggtggcatgtgtattgactacagaaggctca
atacagccaccagaaaggatcatttctttaccatcatagaccagatgctagaagactggcaggctcatgactatfactacttttggatggctactcaggctataacc
agattgcagtagatcccaggatcaagagaaaacaactacatgtccatctggaggtgttcttataagaagataccatttgggtgttaatacgttgcaaccttc
cagagatgcagctcctctatctctgatattgtggaaaaattctggaaactctcatggatgacttctcagtatatggagactcattcagcctgtcttgcacctgaa
acttgtctgaagagccaagaccactagtfttaactgggaaaaatgctacttcatgtgactgaagggtgtccttggcctatgcaaaaatcctcaacaaggga
atagaggtgatcaagaaaaatagaggtaattgaaaattaccacacctgccaatgtcaaggcaatcagaagccttctgggcatgcaaggattctatagagggtt
ataaaggattttcaaaaatcgaaaaaccttaagcaatctgctagctgctgacatgccatttgtttaaacaagagtgtctgcagcgtttgaaacgctgaaagccaag
ctggtcatagcaccagtatttccgcaccagactggacattaccattgaaactgatgttgaa

>Ah-Athila-2

ctatccagccatgaaggaaagtgtgacagaaagggtcactaagttactagaggctgggattatctatctatctgatagtcctgggtgagccctgtccaagttgtcc
ccaagaaaggaggatgacagtgttcataataaaaagaatgaactggtccacaaagaacagttacaggggtggcatatgtgtattgattacagaaggctcaatata
gccaccagaaaggatcatttctttaccatcatagaccacaaatgctagaagactagaaggctatgaatactactgcttttggatggctatfcaggctacaaccaa
gcaaggatcctcaagaccaagagaaaacagcattcacatgtccactggagatttcttacaagaagaatgcttttggctctacaatgcacctgcaaccttcagag
gtgcatgctctatcttctgatattgttagagaagttcttggaaactctcatggatgacttctcagatattggagactcattcagctcctgtcttgcattcatgacattgtc
ctgaaaagggtccaagaaactaacctgtctttaaactgggaaaaatgctactttatggtagtgaaggaaattgctcttgggcacaaaattcaagcaagggaatagag
gtagatcaagccaaggtagaggtaattgaaaaattaccgccaacctgccaatgtaaaggcaatcagaagccttctgggcatgcaaggattctatagagggttataaa
ggattttcaaaaatgcaaaaacctgtagcaactgctagctgctgacacaccatttgtttgacacagagtgtctgcagggtttgaaacctgaaagctaagctggt
cacagcaccagtcattctgcacctgactggacattaccattgaaactgatgttgaa

>Ah-Athila-4

gtacgcacaagatcctattggaggatgatccaacgcagtgttcaaccacagtggcggttaaatccagccatgaaggagggtgtatagaagaggctactaagtt
actagaggctgtaattatctatctatctgatagcccctaggtgagccctgtccaagttatcccaagaaggaggcatgacagtgttcataatgaaaaaatgaa
ctggttctacaagaacagttacaggggtgctgtatgtattgactacagaagactcaatacagccaccagaaaaggactatttctttaccatcatagaccatgctaga
gagactagcaggccatgaatactactacttttggatggctatfcagggttacaacaaatgtatgtagatcctcaggaccaagagaaaacagcattcacatgtcattg
gtgtattacctacagaatgatgccatttggctgtgcaatgcacctgcaaccttcagagggtgcatgctctctatctctgatattgtagaatttctgaaagtctc
atggatgacttctcagatattggagactcatttagctcattgacctttcagaaagggtccaagagatcaacctagttttaaactgggagaaatgca
ctttatggtgactgaaggaaatttcttgggcacaaaattcgaacaaaagaatagagggtgatcaagctaaggtagaggtaattgaaaaattaccaccacctacaat
gttaaggcaattagaagccttctgggcatgcaaggattctatagagggttataaaggattttcaaaaattgcaaaaacctgagcaatttctgtagctgctgatacacct
ttgctttgacacaaagtgcctgcaggcatttagaccctgaaagctaaactggtcacagcaccagtcatttctgcaccagactagacatttccattgaaactgatgtg
gaa

>Ah-Athila-10

gtacgcacaagatcctattggagggtgacgccaagcagtggtacaaccacaaaggcggctgaatccagccatgaaagaaagtgtgacagaaggaggctactaa
attactagaggctgggattatctatctatctgtacagcccctgggtaagcctgtccaagctgcctcaaaaaggaggcatgacagtgtttataatgaaaaaatg
aactggttctacaagaacagttacaggggtgctgtatgtattgattatagaaggctcaatacagccacaaaaggatcatttctttaccatcatagaccagatgc
tagaaagactagcagatcatgaatactactgcttctggatggatattcaaggftataatcaaatgcaatagatcccaggatcaggagaaaacggcattcacatgctc
atctggagtattgcatacagaaggatgccatttggcctgtgcaatgcacctgcaaccttcagagggtgcatgctcctcaatttctctgatattgtgaaaaatttctggaa
gtttatcggatgagtttctgatttggagactcattcagctcctgtctgacctatgacactgttctaaagagatgcaagagaccacctaagttttaaactgggagaa
atgtcatttatggtagctgaaggaaattgctcttgggcacaaaatctcgaacaaaaggaaatagagggtgatcaagctaaggtagaggtaattgaaaaattaccaccac
agccaatgtaaggcaatcagaagccttctgggcatgcaagggttctatagagggttataaaggacttctcaaaaattgccaacaccttaagcaatctgtagctgct
gacacgccatttatcttgataaaggagtgtctgagggcgtttgagacctgaaagctaaactggtcacagcaccagtcatttctgcaccagactggacattaccattg
aactgatgttgaa

>Ah-Athila-12

ctaatccagccatgaaggaggtggtgcagaaaagggtcaccagttactggaggtgggattatccattctgatagccccgggtgagccctgtccaagtt
gtcccaagaaggaggtgacaggtggtcataatgaaaaataaattggccctacaagaacagttacagagtgccgtatgtgtattgattacagaaggctcaat
atagccaccaggaaggatcatttctttaccattcatagaccaaatgttagaaagactagcaggtcatgaatactactgcttttgatggctatttaggtataaccaat
tgcagtagatcctcaggaccaagagaaaacagcattcacatgtcctcttggagtggttgcctacagaaggatgcttttgactgtgcaatgcacctgtaaccttcag
agatgcatgctctctatctctgatatggttagaaaaattctggaagtctttatggatgactttcagatttggagattcattcagctcctgacctgaccatttagcactgt
tctgaaaagatgccaagagactaacctagttaaaactgggaaaaatgtcactttatggtgactgaaggaaattgctctgggcataaaattcaacaagggaatagag
gtgatcaagctaaagtggagtaattgaaaaattaccactactgccaatgttaaggcaatcagaagctttctggggcatgcaaggattctatagagggttataaagg
attttcaaaaattgcaaacctctgagcaatctgctagctgctgacacgccatttgggtttgacacagagatctacagcgtttgaaactctgaagctaaagctggtca
cagaaccagttattttgaccagactggacattaccattgaaactgatgtgtgaa

>Ah-Athila-15

gtacgcacaagatcctattggaaggagacccaagcctgtggtccaaccacaaaaggcggctgaaccagccatgaaggaaagtgtgcaaaaaggaggtcactaa
attactaaaggctgggattatcttattctgacagccccgggtaagccctgtccaagctgctccataaaaggaggatgacagtggttcacaatgaaaaaat
gaaactggtctacaagaacagttacagggtgctgtatgtgtattgattatagaaggctcaatacagccaccagaaaggatcatttctttaccattcatagaccaat
gctagaagactagcaggtcatgaatactattgcttctggatggatattcaggttataatcaaattgtagatagatccgagatcaggagaaacggcattcactgc
ccatccggagatttgcatacagaaggatgccatttggcctgtgcaatgcacctgcaactttcagagggtgcatgctctcaatctctgatattggtgaaaaattctg
gaaagtctcatttattggtgactgaaggatcgtccttgggcacaaaatctcgaacaaggggatagagggtggaatcaggtaaggtaaggtaattgaaaaattacca
ccacctgccaatgttaaggcaatcagaagctttctggggcatgcaaggattctatagagggttataaaggattttcaaaaattgcaaacctctgagcaacctgctagc
cactgacacgccatttcttgataaagagtgctgcaaggcatttggactctaaagctaaattggttacagcacaatcatctctgaccagactggacattaccatt
tgaactgatgtgtgaa

>Ah-Athila-16

ctaatccagccatgaaggaggtggtgcagaaaaggtcactaaagtactagaggctgggattatcttattctgatagccccgggtaagccctgttcaagtcg
tcctaagaagggtgcatgacagcgggtcataatgaaaaaatgaactggtcctacaagaacagttactgggtgctgtatgtgtattgattatagaaggctcaatac
agctaccataaaggatcatttctttaccattcatagaccagatgctagaacgactggcaggtcatgaatactactgcttctggatgattatcagggtataatcaaatt
gcagtagatccacaagatcaagagaaaacgacattcacatgtctatatggagatttgcatacagaaggatgccatttggcctgtgcaatgcacctgcaacctttcaga
ggatgcatgctctcaatttctctgatattggtgaaaaattctggaaagtcttcatggatgactttccgtatttggagactcattcagctcctgcttggacctatgacttgt
tctaaagatgccaagagactaacctggtttaaattgggagaaatgtcactttatggtgactgaaggaaattgctctgggcacaaaatctcgaacaagggaataga
gggtgatcaagctaaagttagaggttaattgaaaaattaccaccactgccaatgttaaggcaatcagaagctttctggggcatgcaaggattttatagagggttataaag
gattttcaaaaatcgcacaacctctgagcaacctgctagctgctgacacgccatttcttgataaagaatgtctgcaaggcatttggactctgaaagctaaagttggtc
acaacccaatcatctctgaccagactggacattaccattgaaactgatgtgtgaa

>Ah-Athila-18

ggagatgatagggtgcagaaggaggtcactaaattactagatgctagaattatctccatttctgatagccccgggtaagccctatccaagtcacccataagaagg
tggtatgacagtggtcataatgaaaaaatgaactggtcctacaagaacagttacagggtagcgtatgtgtattgattacagaaggctcaatacagttaccaaaatg
gatcatttctttaccattcatagaccaaatgctagaagactagcaggtcatgaatactactgcttctatagatgattatcaggctataatcaaattgcagtagatccc
aagatcaagaagaaatgacattcacatgtccatccgacttattgcatacagaatgatgccatttggcctgtgcaatgcacctgcaacctttcaaagggtgcatgctctct
atcttctctgatattggtgaaaaattctggaagtcttcatggatgactttcagatttggagactcattcagctcctgcttggacctatgacttgttctaaaggaggtgc
aaaagactaacctggtttaaactggaaaaatgtcactttatggtgactgaaggaaattgctcttaggcacaaaatttgaacaagggaatagaggtgatcaagctaa
ggtagaggttaattgaaaaattaccaccactgctaatgttaaggcaatcagaagctttctggggcatgcaaggattctataggaggtttataaaggattttcaaaaatc
caaacctctgagtaactgctagctgctgataatgccaattatctttgataaaggaggtgctgcaaggcgtttgagactctgaaagctaaagctggtcacagcaccagtcact
ctgaccagactgaacattaccattgaaactgatgtgtgaa

>Ah-Athila-19

ctaatccagccatgaaggaggtggtgcagaaaagggtcactaaattactagaggctgggattatcttattctgatagccccgggtgagccctgtccaagttg
tccccaaaaggaggtgacaggtggtcataatgaaaaatgaactggtcctacaagaacagtcacaggggtgctgtatgtgtattgattacagaaggctcaatac
cagccaccagaaaggatcatttctttaccattcatagaccaaatgctagaagactagctggtcatgattattactgcttttgatggctattcagggttacaccaaat
tgcagtagatcctcaggaccaagagaaaacagcatttactgccccttggcgtggttgcctacagaggagcgccttttgctgtgtaattgctcctgcaagctttcaga
gatgcatgttatcattctctgatattggtgagaaaattctggaagtcttcatggatgacttctcagtatatggagactcattcagctatgctttaatcacaacactt
gtcctgaaaagggtccaagagactaacctggtttaaactgggagaaatgtcactttatggtgactgaggaatttcttgggcacaaaatttcaagtaggggaatag
agggtgataaggcaaggtagaggttaattgaaaaattaccaccactgccaatgttaaggcaatcagaagctttctggggcatgtagattctacagaagggttataa
aggattttcaaaaattgcaaacctttgagtaacctgctagctgctgacacaccatttgggtttgacacacagtgctgcaaggcatttggaccctgaaagctaaagctg
gtcacagcaccagtcactctgaccagattggacattgccaattgaaactgatgtgtgaa

>Ah-Athila-21

aaggacacacaagacagctcttaggtggacaataagtgacctaaaggacattacccaacccgatgcatgcacaagatcctgctgaaggatgatccaacctgaa
ggaagtgtgcaaaaagtaagtcacaaaattatgggagccgagattatccatctctgacatccctgggtgagccctgccaaagttgtccaaaaaggagg
catgacagtggtcataatgataagaatgaactgattcctacaaagacaatcatagggtggtgatgtgcattgaatagaaagctcaataacgccacaaaaagga
tcaatctttaccattcattgatcaaatgctagagagactagcaggatgcttttattgttctcgatggtcattccggttacaatcaataacgccacaaaaagga
tcaaaaaatacagattacatgccatagcgtgtttgcttatagggtgcatctttggcctgtgcaatgcacctgccaccttcagagggtgatgctctctatgtct
ctgataggtggagaagttcctgaagtattcatggatgactcctctatctttggagattcactgaatcatgccttgaccttagccctagttccttaaacgatccaaga
cccaacttggtttaacttagaaaaatgtaactcctgactgaagggtgtctgggcatcgaattcaaacagggaatagaaagtgatcaagctaaaggtg
gaggaattgagcattaccaccactaattgcaaaaggatcagaagttctgagacatgacaggttataaaaggattctctaaagttgcaaa
actcctgtgcaacttactaacctgacactctattgtctttaccaagagttctgagggcctttgagactctgaaagtaagcttatcattgcccctgtaactctgac
ccaactgggacttaccattgaaactgatgtgaa

>Ah-Athila-22

ctaaatccagccatgaaggagtagtcagaaaaggctactaaagtactagaggctaggattatctattctgatagcccctgggtgagccctgtccaagttg
tcccaataaggaggatgacagtggtcataatgaaaaaatgaactggtcccacaagaatgttacagggtggcgtatgtgtattgactacagaaggctcaata
cagccaccagaaaggatctttcctttaccattcatagaccagatgctagagagactagcaggatgcaatattattgcttttggtgagctattcaggttacaacaaa
tgcagtagatcctcaggaccaagagaaaacagcatttacttccccttctggcgttttgcctacagggagatgcttttggtctgtgtaatgctctgcaactttcaga
gatgcatgttatccattctctgataggtggagaattcctgggaagcttcatggatgacttctcagttatggagactcattcagctcctgtcttaatacactagcatt
gtcctgaaaaggtccaagactaacctggttttaactgggagaatgtcactttatggtgactgaaggaaattgctctggcacaataatcaagcaggggaata
agggtggataaggcaaggtagagtaattgaaaaattaccaccactgccaatgtaaggcaatcagaagcttctgggcatgacaggttctacaaggtttat
aaaggattttcaaaaattgcaaacctttgagtaacctgtagctgctgacacacattgtgttgacacacagtgctgacggcattgagacctgaaagataagct
ggtcacagcaccagctctctgaccagattggacattgccattgaaactgatgtgaa

>Ah-Athila-23

aaggacacacaagacagctcttgattatccataagtgacctaaaggcataagcccggctagatgcatgcacaaaatcctattggaggatgatgctaagccagtg
ttcaaccacggaggcggctaaatccagccatgaaggaggtgtgtagaaaggctaccaatctggaggctgggattatctattctgatagcccctggg
tgagtctgtcaagttgtcccaagaagaaggcatgaaagtggtcataatgaaaaaatgaactggttctacaagaacgttacagggtagcgcattgtattga
ctacagaaggctcaatacagccaccagaaaggatctttcctttaccattcatagaccaaatgctagagagactagcaggatgcaatattactgcttttgatggct
attcaggttacaacaaattacagtagatcctcaggaccaagagaaaacagcattcacatgctctctgagatttgccttacagaagaatacctttgtctgtgtaag
cacctgcaactttcagagggtcatgcttctatctctctgataggttaaaaagttttagaaagcttcatggatgactttcagatttggagactcattcagctcctgct
tgatacttagcactgtctgaaaagatgtaagagaccaacctagtttaactggaaaaatgtcactttatgtgactgaaggaaattgtcctttggcataaaaattc
aaacaagggaatagaggtgataagcaagtaagtgaaatgaaaaattaccaccactgctaataatgaaggcaatcagaactttctgggcatgacaggttcta
caggaggtttataaaggattttcaaaaattgaaaaacccctgagcaatctgtagctgctgacacgccattgtgttgacacagagtgctgacaggcgtttgagact
tgaagctaaagctggtcacagcaccagttattctacaccagactggactttaccattgaaactgatgtgaa

>Ah-Athila-25

gtacgcacaagatcctattggaggataatgcaaacagtggtccaaccacagaggaggttaaatcctgcatgaaggaggtggtgcagaaaaggctactaaat
tactagaggctgggattatctattctgatagcccctgggtgagccctgtccaagttgtcccaaaaaggaggatgacaatggtcataatgaaaaaatgaa
ctgtccctataaggacagtcacagggtgctgattgactacagaaggctcaatacagccaccagaaaggatctttcctttaccattcatagacaaaatgc
tagaaaagactagctggtcattactgcttttgatggctatcaggctacaacaaattgcatgtagatcctcaggaccaagagaaaacatcattcactgtccct
ctggcgtgttgcctataggaagaatgcttttgctgtgcaatgcacctgcaaccgtcagagggtcatgctctctatctcagataggtgagaaattctggaag
cttcatgtagtacttctcagatataaggagactcattcagctcctgtcttaatacactagccctgtctgaaaaagggtccaaagagactaactggttttgactggga
aatgtcactttatggtgactgaaggaaattgtcttgggcacaaaattcaacaagggaatagaggtggataaggcacaaggtagaagtaattgaaaaattaccacc
ctgccaatgtaaggcaatcagaagctttctgggcatgacaggttctacagaagggtttataaaggactttcaaaaattgcaaacctctgagtaacctgttagctgct
gacacacattgtgttgacacacagtgctgacagcgtttgagaccctgaaagctaaagctggtcacagcaccagctctctgaccagattggacattgcccattg
aactgatgtgaa

>Ah-Athila-24

gtacgcacaagatcctattggaggatgtaaaccaagttgttcaaccacagaggcggctaaatccagccatgaaggaggtggtgcagaaaaggctaccaagt
tactggaggctgggattatctattctgatagcccctgggtgagccctgtccaagttgtcccaaaaaggaggatgacaatggtcataatgaaaaaatgaa
actggttctacaagaacagttatagggtggcgtatgtattgattacagaaggctcaatacagccaccaggaaggatcatttctttaccattcatagaccgatg
ctagaaaagactaacaggtcatgaatactactgcttttgatggctattcaggttaccgacaaaattacagtagatcctcaggaccaagagaaaacagcattcatgcc
ttctagaggtttgcctacagaaggatgcttttgctgtgcaatgcacctgcaaccttcagagggtcatgctctctatctctgataggttagagaattttgaaa
gtcttcatgtagtattcagctttggagactcattcagctcctgcttaaccatttagcactgtctgaaaaagatgccaagagaccaacctagtttaactgggaaa
aatgtcactttatggtgactgaagggtgtccttgggcataaaaattcaacaagggaatagaggtggatcaagctaaagttgaagtaattgaaaaattaccacc
tgccaatgtaaggcaatcagaagctttctgggcatgacaggttctacagaagggtttataaaggactttcaaaaattgcaaacctctgagcaacctgtagctgct
acatgccattgtgttgacacagagtgctgacagcgtttgaaactctgaaagctaaagttggtcacagcaccagttattctgaccagactggacattaccattgaa
tgatgtgtaa

>Ah-Athila-29

ctaatccagccatgaaggaggtggtgcagaaaagggtactaaactagaggctgggatttattctatctttgatagccccctgggtgagccctgtccaagttgt
tcccaaaaaggaggcatgacagtggtcataatgaaaaaatgaactggtcctacaaggacagtcacagggtggcgcgatgtgtattgactacagaaggctcaat
acagccaccagaaggatctttctttaccattcatagaccaatgctagaaagactagctggtcatgatttactgcttttgatggctatttaggctacaaccaa
ttgcagtagatcctcaggaccaagagaaaacagcattcactgcccctctggcgtgtttgctacaggaggatgcccctttgctgtgcaatgcacctgcaaccttcag
agggtcatgctctctatctctcagatgtgtagagaaaattctggaaagtctcatggatgactttcagtatatgaggattcatftagctcctgtcttaacacctagcact
gtcctgaaaagatgccaagagactaacctggtttaaactgggagaaatgacactttatggtgactgaaaggaattgctcctgggcacaaaattcaagcagggggaata
gaggtggataaggcagagggtggagtaatagaacattaccaccacctgccaatgftaaggcaatcagaagctttctggggcatgcaggattctacagaaggtttat
aaaggattttcgaaaatgcaaacctctgagtaacctgttagctgctgacacaccattgtgtttgacacacagtgctgcaggcatttgagaccctgaaagctaaagc
tggcagcagcaccagtcctctgcaccagattggacattgccaattgaaactgatgtgtaaatcact

>Ah-Athila-17

ggagctctcccatatggtcgacctgcagggcggccgcgaattcactagtattctaaatccagccatgaaggaggtggtgcagaaaagggtcactaagttactagag
gctgggatttattctatttctgatagccccctgggtgagccctatccaagttgtcccaagaaggagggcatgatagtggtcataataaaaagaatgaactggtcc
cacaagaacagttacaggggtggcgtatgtgtattgactacagaaggccaatacagccaccagaaggatctttctttaccattcatagaccaatgctagaaaag
actagctggctcatgatttactgcttttgatggctattcaggttacaaccaaattgtagtagatcctcaggaccaagagaagacagcattcactgcccctctggcgt
gtttgctacagaagaataaccttttgctgtgtaatgacctgcaacttttaaaagatgcatgctatccatctctctgatatgtagagaaggtcctggaagtctcatgg
atgacttctagatataggagactgctcagctcctgtcttaactatctagcactgtcttgaagaggtgccaagagactaacctggtcttaactgggagaaatgactct
catggtgactgaaaggaattgtcttgggcataaaattcagggcaaggaatagaggtggtacaaaggtagaagtaattgaaaaattaccaccacctaccaatg
taaggcaatcagaagctttctaggccatgcaggatfttacagaaggttcataaaaggactttcaaaaattgcaaaccttaagcaacctgctagctgttgacacgcca
ttgtgttgacacagagtgctgcaggcgtttgagaccctgaaatccaagctagtacagcaccagtaaatctctgcaccagattgaacattaccatttgaactgatgt
gaaatcgaattccgcggcccatggcggccgggagcatgcgacgtcgggc

>Ah-Athila-20

gcccagctgcgatgctcccggcccatggcggccgggaattcgattaaaggacacacaagacagctctgggtggtccatcagtgatcttaaggccattagc
ccagccagatgcatgcacaagatcttactggagatgacgccaagccagtgttcaaccacaaaggcggctgaatccagctatgaaagaagtggtgcagaaaaga
ggctactaaactagagggtgggatttattcttctgacagccccctgggtgagccctgtccatgtttgctcctaagaaagggtggcatgacagtgattcataatg
aaaaaatgaactggtcccacaagaacagttacagggtggcgtatgtcattgtttatagaaggctcaatacagccaccagaagatctttctttaccattcata
gaccagatgctagaagactagcaggtcatgaataactactgctcctggatggaatattcagggtataatcaaatgctgtagatccccatgatcaggaaacggca
ttactgcccactgtagatgctacacagaagatgccaattggcctgtgcaatgcacctgcaaccttcagagggtgcatgctcctcaatttctctgatgtggtgaa
aaattcctggaagtctcatggatgactttcagatggtgagactcattcagctcctgcttaaccatttagcactgttctgaaaagatgccaagagactaacctggttt
aaactgaaaaaatgctactttatggtgactgaaggaaatgctcctgggcacaaaattcgaacaagggaatagaggtggatcaagctaaagtaagtaagtaattgaaa
attgccaccacctgctaattgtaaggcaatcagaagctctctggggcatgcaggattctataggaggtttataaaggattttcaaaaatcgccaaacctctgagcaac
ctgctagctgctgacacgccatttatttataaaagatgctctgcaggcatttgagactctgaaagctaaattggtaacagcaccaatcatctttgaccaaactgaca
ttaccatttgaactgatgtgtaaatcactagtgaattcgcggccgctgcaggctgaccatattgggagagctccccacgcgtt

>Ah-Athila-26

cgacgtcgcagctcccggcccatggcggccgcgggaattcgattgtacgcacaagatcctattggagggtgacactaagccagtggtcaaccacagaggtg
gctgaatccagccatgaaggaggtggtgtagaaggaggtcactaaactagaggctaggatttattctatttctgatagccccctggtaacccctttcaagtcg
tccctaagaagggtgatgacagtggtcataatgaaaaaatgaactggtctacaagaacagttatagggtggcgtatgtgtattgattacagaaggctcaataca
gctaccagaaaaggatctttctttatcattcatagaccaatgctagaaagactagcaggtcatgaataactactgcttctggatggaatcaggttataatcaaatg
cagtagatccccaggatcaagaagaaacagcattcacatgccatccggagatttgcatacagaaggatgccaattgactctgtgcaatgcacctgcaaccttcaag
gtgcatgctctctatttctctgatatggtggaagatttctggaagtctcatggatgactttcagatggtggagactcattcagctcctgtcttgaccatctagcactgtc
ctaaagaggtgccaagagactaacctggtttaaactagagaaatgacactttatggtgaatgaaaggaattgctcctgggcacaaaattcgaacaagggaatagaa
gtggatcaagataaggtagaggtaattgaaaaattaccaccacctgccaatgftaaggcaatcagaagctttctggggcatgcaggattctataggaggtgtataaaa
gattttcaaaaatcgccaaacctctgagtaactgctgactgctgacacgccatttatttataaagagtgactacagggctttgaggctctaaagctaaagctggc
acagcaccagtcctctttactagactagacattaccattgaaactgatgtgtaaatcactagtgaattcgcggccgctgcaggctgaccatattgggagagctccc
aacgcgtt

A. ipaënsis

>Ai-Athila-1

ctaatccagccatgaaggaggtgcagaaggaggtcaccaagtactagagactgggattatctattctgatagcccctgggtgagccctgttcaagttg
tccccaaaaaggaggtatgacagtggttcataatgaaaagaatgaactggttctacaagaacagttacaggggtggcgtatgtattgactacagaaggctcaat
acagccaccagaaggatcattttcttaccattcatagaccagatgctagagagactagcaggctatgaatactactcttttggatggctattcaggtataacca
attgcagtagatcctcaggaccaagagaaaatgacattcacatgctcttctggagtggttcttacaaggagatgcccttggctgtgtaatgcacctgcaaccttca
gaggtgcatgctctctatctctctgatagtagaaattctggaaggcttcatggatgactttcagatattggagactcattcagctcttgcctaaccttagcact
tgtctgaaaagattccaagagatcaacctagttttaaactgggaaaaatgccactttatggtagtgaaggaaattgctctgggcataaaattcaacaagggaata
agggtgatcaagctaaagtggaaagtaattgaaaaattaccaccatctgtaattgtaaggcaatcagaagctttctggggcatgcaagattctacaggaggttataaa
ggattttcaaaaattgcaaaatccctgagtaattctgtagctgctgacacgccatttggttgacgcaagtgctgcaagggttgagactctgaaagctaaagctgg
tcacagcaccagttattctgcaccagactggacattaccattgaaactgatgtgtaa

>Ai-Athila-5

ctaatccagccatgaaggaggtgtacagaaagggtcactaaagtactagaggctgggattatctattctgatagcccctgggtgagccctgtccaagttg
ttcccagaaggaggatgacagtggttcataatgaaaagaatgaactggttctacaagaacagttacaggggtggcgtatgtattgactacagaaggctcaata
cagccacaaaaaggatcattttcttaccattcatagaccagatgctagagagactagcaggctatgaatactactcttttggatggctattcaggtataacca
ttgcagtagatcctcaggaccaagagaaaacagcattcacatgctcttctggagtagtttggtagacaaggatgccatttggctgtgcaataacctgcaaccttca
agggtcatgctctctatctctctgatagtagaaattctggaagcttcatggatgacttctcagatattggagactcattcagctcttgcctaaccttagcactt
gttctgaaaagggtccaagagacaacctggtttaaactgggaaaaatgcaactttatggtagtgaaggaaattgctctgggcataaaattcaacaagggaataga
gggtgatcaagctaaagttagagtaattgaaaaattaccaccctgccaatgtaaggcaatcagaagctttctggggcatgcaagattctacaggaggttataaa
ggattttcaaaaattgcaaaatccctgagtaattctgtagctgctgacacgccatttggttgacagagtgctgcatgctgttgaactctgaaagctaaagctgtca
cagcactagctctctgcaccagactggacgttaccattgaaactgatgtgtaa

>Ai-Athila-6

aaggacacacaagacagctctaaatccagccatgaaggaggtggtgcagaaagggtcaccaaaactgaggctgggatcattatctattctgatagcccct
gggtgagccctgttcaagttgctcctcaaaaaggaggtcatgacagtggttcataataaaaagaatgaactggttctacaagaacagttacaggggtggcgtatgt
attgactacagaaggctcaatacagccaccagaaaaggatcttttcttaccattcatagaccagatgctagaagactggcaggctcatgattactgcttttggat
ggctactcaggctataaccagatgtagtccccaggatcaagagaaaacagcattcacgtgctctatctggagtggttcttatagaaggatgccgttgggctgt
gtaatgtgctacaacctccagagatgcatgctctctatctctgatagtagaaattctggaagctttatggatgacttctcagatattggagactcattcagct
cctgtctgacacctgaaactggttcaaaaatgccaagagaccaacctggtttaaactgggaaaaatgcaactttatggtagtgaaggattgctcttgggcat
aaaattcaacaagaagtagaggtgatcaagaaaaatagagtaattgaaaaattaccaccctgccaatgtaaggcaatcagaagctttctggggcatgc
aggattctacagaagggttataaaggattttcgaaaatgcaaaacctctgagtaacctgttagctgctgacacaccatttggttgacacacagtgctcagggcatt
gagaccctgaaagctaaagtgtgcacagcaccagttatctctgcaccagattggacattgccattgaaactgatgtgtaa

>Ai-Athila-12

gtacgcacaagatcctattggaggtgatgccaagccagtggttaaaccacagagggtgctgaatccagccatgaaggaggtggtgcagaaagggtcactaagt
tactagaggctgggattatctattctgatagcccctgggtgagccctgtcagaagtttcccagaaggaggtcatgacagtggttcataatgaaaagaatga
actagttctacaagaatgtagcagggtggcgtatgtattgactacagaagactcaatacagccaccagaaaggatcattttcttaccattcatagaccagagg
ctagagagactagcaggctatgaatactactgcttttggatggctattcagattacaacaaatgcaagtagatcctcaggaccaagaaaaatgcaattacatgctc
ttctggagatttgcctacagaaagatgccatttggctgtgcaatgcacctgcaaccttcagagggtcatgctctatcttcttggatgtagaanaatttggaa
tcttcatggatgacttctagatttggagactcattcagctctgcttaccatctagacttctgtaaaagggtccaagaggctaacctgattttaaactggaaaa
gtcactttatgtgattgaagaaattgctcttgggcacaaaattcaaaataggaatgaggtggaatcaagctaaaggtagaggttaattgaaaaattaccaccacta
ccaatgtaaaagcaatcagaagctttctgggcatgcaagattctatagaggattataaaggattttcaaaatgcaaaacctctgagcaatctccttgaactgacac
gccatttatgttgcacaaaagtctgtaggcgtttagaccctgaaagctaaagtgtgcacagcaccagttattctgcaccagactggacattaccattgaaactgat
gtgtaa

>Ai-Athila-13

ggagatgataggtgcagaaggaggtcactaaatctagaggctgggattatctattctgacagcccctgggtaagccctgtccaagctgctccaaagg
gaggcatgacagtggttcataatgaaaaaatgaactggctcctacaagaacagttacaggggtggcgtatgtattgattatagaaggctcaatacagccaccaga
aaggatcattttcttaccattcatagaccagatgctagaagactagcaggctatgaatactactacttctagatgtagattcaagttataatcaaaatgcaagaaatc
cccagatcaggagaaaacggcattcacatgctcatctggagatttgcatacagaaggatgccatttggcctgtgtaatgcacctgcaaccttttagaggtgcatgct
cttaatttctctgataggtgaaaaattctggaagtttcatggatgactttcagatttggagactcattcagctctgctgacctctagcacttgtctaaagagat
gccaagagaccaacctgattttaaactgggaaaaatgcaactttatggtagtgaaggaaattgctctgggcacaaaatcgaacaaaggaaatagaggtgatcaa
gctaagtagaggttaattgaaaaattaccaccagccaatgtaaggcaatcagaagctttctggggcatgcaagattctatagagggttataaaggacttctcaa
aaatgccaaaccttaagcaatctgtagctgctgacacgccatttcttataaggaggtgctgacacatttgaaccctgaaagctaaagttggtcacagcacc
agctatctctgcaccagactggacattaccattgaaactgatgtgtaa

>Ai-Athila-14

aaggacacacaagacagctcttaggtggtccataagtgatcttaagggcattagcccagctaaatgcatgcacaagatcctattggaggatgatgctaagccagtggttcaaccacagaggcggctgaatccagccatgaaggaggtggtacagaagaggctactaagtactagaggctgggattatctattctgatagccctgggtgagccctgtccaagttgtccctaaagaagggggcatgacagtagtccataatgaaaagaatgaactgttctacaagaacagttacaggggtggcatatgtattgactacagaaggctcaatagaccaccagaaaggatctttctttaccattcataaacagatgctagagagactagcaagccatgaatactactgcttttggatggctattcaggttacaactaaattgcagtagatcctcaggaccaagagaaaacagcattcacatgctcatctggagtagtcttctacagaaggatgccatttggctgtgcaatgcacctgcaaccttcaagaggtgcatgctctctatttctctgatatggtagaaaatttctggaagcctcatggatgacttctcagatttggagactcattcagctcctgtcttgatcatctagcactgttctgaaaaaggtgccaaagagactaacctgggttfaaattgggagaaatgacatttatggtgactgaaggaaatgcttgggcacaaaaattcgaacaaaggaaatagagggtgataaagctaaagtagaggtaattgaaaattaccacatctgccaatgtaaggcaatcagaagctttctggggcatgcaggatctatagaaggttataaaggattttcaaaaatgcaaaacctctagcaatctcctgtagctgaca

>Ai-Athila-15

aaggacacacaagacagctctgggtggtccataggtgaccttaagggcataagcccagctagatgcatgcacaaaatctattggaggataacgctaaaccagtggttaaccacagaggcggctaaatccagccatgaagaaggtggtcagaagaggctaccacaaactactagaggctgggattatctattctgatagccctgggtgagccctgtccaagctgtcccaaaaaggggagcatgacagtgattcataatgaaaaaagaactggttctacaagaacagttacaggggtggcgcattgtgtatgactacagaaggctcaatacagccaccagaaaggatctttctttaccattcatagaccagatgctagaagactagcaggtcctgattactgcttttggacggctactcagctataaccagattgcatgtagctcctcaggatcaagagaaaacagcattcacatgctcatctggagtgctgcttatagaaggatgccatttggactatgtaatgcctaccctccagagatgcatgctctctatttctctgatattggtggaataattctggaaagcttcatggatgacttctcagtagatattggagactcattcagctcctgtctgacactgaaactgttctgaaaagatgcaagaaccaacttagtttaactgggaaaaatgctacttcatgggtgattgcaggggattgttcttgggcataaaatctcaacaagggaaatagagggtgcatcaagcaaaaatagaggtaattaaaaaattaccaccacctgccaatgtaaggcaatcagaagctttctggggcacgcagattctataggaggttataaaggattttcaaaaatgcaaaacctctagcaatctcctgtagctgaca

>Ai-Athila-16

ctaaatccagccatgaaggaggtagtgcaaaaagaggtcaccaagttactggaggtgggattatctccatctctgatagcccctaggtgagccctgtccaagttgtcccaagaaggaggtgacagtggttcataatgaaaaataaactggttctacaagaacagttataggggtggcgtatgctattgattatagaaggctcaat acagccaccaagaaggatcatttctttaccattcatagaccagatgctagaagactagcaggtcatgaatactactgcttttggatggttattcaggttataaccaaatgtagaagatctcaagaccaagagaaaacagcattcacatgctctctggaggtgttgcctacagaaggatgcttttggctgtgcaatgcactgcaacctttca gagatgcatgctctctatttctctgatattgtagaagaattttggaagcttcatggatgactttcaataattggagattcatttagctcctgccttgaccatttagcattgttctgaaaagatgccaagagactaacctagttttaaactgggaaaaatgctactttaggtgactgaaggaaatgcttgggcataaaaatcacaacaaagggaataga ggtgcatcaagctaaatggaagtaattgaaaatcaccaccctaccaatgtaaggcaatcagaagctttctgggcataagcattctatagagggttataaa ggattttcaaatattcaaaaacctctgacaaatctgtagctgctgacacgccatttgtgttgacacagagtgctgtagctggttgaactcaaaagctaaagctggtcacaacaccagttatttctgcaccagactggacattaccattgaaactgatgtgtga

>Ai-Athila-2

taaggacacacaagacagctctgggtggtccataagtgatcttaagggcattagcccagcaagatgcatgcacaagatcctattggaggatgatgctaagccagtggttcaaccacagaggcggctaaatccagccatgaaggaggtggtgcagaagaggctactaagtactagaggctgggattatctattctgatagccctgggtgagccctgtctatgttctctaaagaaggaggaatgacagtggttcataatgaaaagaatgaactggttctacaagaacagttacaggggtggcgtatgtattg actacagaaggctcaatagctaccagaaaggatctttctttaccattcatagaccagatgctagagagactagcaggtcataaatactattgcttttggatggct attcaggttacaacaaattgcagtagatcctcaggaccaagagaaaatgacattcacatgctctctggagtagtcttctgagatttgcctacagaaggatgccctttggctatgcaat gcacctgcaaccttcaagaggtgcatgctctctatttctctgatattgtagaagaatttctggaagcttcatggataactttcagtagctggaacacattcagctcctg tcttgaccatttagcactgttctgaaaagatgccaagagaccaactagttttaaactgggaaaaatgctactttaggtgactgaaggaaatgcttgggcataaaaatt tcaacaagggaatagagggtgattaaagctaaagtggaagtaattgaaaaatcaccaccctgccaatgtaaggcaatcaaaagctttctggggcatgcaggatt ctatagaagggttataaaggattttcaaaagattgcaaaacctctgagcaatctgtagctactaacacacacatttgtgttgacacagagtgctgcaggctttgaaact ctgaaaagctaaagctggtcacaacagcaggtatttctgaccagactagacattaccattgaaactgatgtgtga

>Ai-Athila-3

aaggacacacaagacagctctgggtggtccataagtgaccttaagggcataagcccagctagatgcatgcacaagatcctattggaggatgatgctaagccagtggttcaaccacaaaggcattaaatccagccatgaaggaggtggtgcagaagaggctaccacaaactgagaggctgggattatctattctgatagctcctgggtgagccctgtccaagttgtcccaaaaaggaggtgacaatggttcataatgaaaaaagaactggttctacaagaacattacaggggtggcgcattgtattgactacagaaggctcaatacagccacaaaaggatctttctttaccattcataaccagatgctaaaagactagcaggtcatgattttactgcttttggatagct atttaggttacaacaaattgcagtagatccccagatcaagaaaaactgattacatgctcatctggagtagtcttctcaaaaaggatgccattcgggctgtgtaat gcacctgcaaccttcaagaatgcatgctctctatttctctaatatgatgaaaaatttctggaagcttcatggatgacttctcagtagatattggagactcattcagctcctgt ctgtatcactgacactaatctgaaatgatgccaagagactaacctgggttgaactgggaaaaatgctactttaggtgactgaaggaaatgcttgggcataaaaatt tcaaaataggggatagagggtgatcaagctaaagtggaagtaattgaaaaatcaccaccctgccaatgtaaggcaatcaaaagctttctggggcatgcaggatt ctataggaggttataaaggattttcaaaaatgcaaaacctctgagcaatctgtagatgctgacacgccatttgtgttgacacagagtgctacaggcgtttgaaact ctgaaaagctaaagctggtcacaacactagttatttctgaccagactggacactaccattgaaactgatgtgtga

>Ai-Athila-4

aaggacacacaagacagctctgggtggtccataagtacacctaaggccataagcccagctagatgcatgcacaagatcctattggaggatgatgctaagccagtgttcaaccacagagggcgttaaatctagccatgaaggaaagtggtgagaagaggctaccacagctactggaggctgggattattatcctactctgatagccctgggtgagccctgtcaagttgtcccaaaaaaggaggcatgacagtggttcataatgaaaaaatgaactggttctacaagaacagttacaggggtgcatgtgtattgactacagaaggctcaatacagccaccagaaggatcatttccctaccattcatagaccaaatgctagagagactagcagatcatgagttactccttttgacaggctattcaggttacaaccaaatgcagtagatccccaggatgaagagaaaactgacattacatgttcatctggagatattgcttacagaagaatgccatttgggctgtgtaatgcacctgcaacctttccgagatgcatgctctctatttctctgatatggtggaaaatfttctggaaagtcttcatggatgacttctcagtatatggagactcattcagctcctgtcttgatcacctgacactagttctgaaaagatgccaagagactaacctggttttaactgggaaaaatgtaatttatggtggctgaaaggattgtccttgggcgtaaaatftcaacaagggaatagagggtgatcaagctaaagtggaaagtaaaaaattaccaccacctgccaatgtaaggcaatcagaagctttctgaggcatgcatgattataggaggtttataaaggattttcaaaaattgcaaacctctaaagcaatctgtagctgctgacacaccatttgtttgacacagagtgctgttagggcgttttaactttaaacctaagctggtcacagcaccagttattctacaccagactggacactaccatttgaactgatgtgtaa

>Ai-Athila-7

aaggacacacaagacagctctgggtggtccattagtgatctcaaggcattagcctggccagatgcatgcacaagatcctactagaggataatgctaaactagtgttcaaccacaatggcggctgaatccagccatgaaggagggtgtgacagaaaggctactaaactactagaggctgggattattatccatttctgatagccctaagtgagccctatccaagttgactcaaaaaaggaggcatgacagtggttcataatgaaaaaatgaattggttctacaagaatgttacaggggtggcgtatgtgtactgactatagaagactcaatacagccaccagaaggatcatttccctaccattcatagaccaaatgctagaaaggctagcaggtcatgactattactacttctggatggctactcaggctacaaccaaatgcagtagatccccaggatcaagaaaaatagcattcaaatgtccatctggagatattgcttacagaaggatgtccttggctatgtaatgaccagcaaccttccagatgctcttctctctgataagggtgaaaaatfttctggaaagtcttcatggatgacttttagtatttggagagcttccagctcctgtcttgatcatttagactgttctgaaaagatggcaagagactaacctggttttaactaggaaaaatgtcactttatagtactgaaaggattgtccttgggcataaaaatftgaacaaggcatagagggtgatcaagctaaagggtgaggtaatgaaaaattaccaccaccacaatgtaaggcaatcagaagctttctggatcagcaggattctataggaggtttataaagaattttcaaaaatgcaaacctctgagtaacctgtagctgctgatacaccatttgtttgacacagagtggttgcaggcctttgagactctaaaggctaaagctgttcacagcgtctgacatttctgcatcggactagacattaccatttgaactgatgtgtaa

>Ai-Athila-8

aaggacacacaagacagctctgggtggtccattagtgatcttaaggcattagcccagccagatgcatgcacaaaatcctattggaaggtagcgttagccagtgttcaaccacaaaagctgctgaaccagccatgaaggagggtgtgcagaaggaggtcactaaactactagaggctgggattattatcctatttctgttagccctgggtgaagccatgtccaagttgtccctaagaaaagtgcatgacagtggttcttaataaaaaatgaactggttctacaagaacattacaggggtggtatgtgtattgattatagaaggctcaatacagccaccagaaggatcatttccctaccattcatagaccagatgctggaagactagcaggacatgaatactactacttctggtatggatattcaggttataaaaaatftcagtagatccccaggatcaagagaaaacagcattcacatgtccatctggagatattgctacagaaggatgccatttggctgtgcaatgcacctgcaacctttcaagggtgcatgctctcaatfttctctgatatggtggaaaaatfttctggaaagtcttcatggatgactttcagatttggagactcattcagctcctgtcttgaccatctagcactgttctaaaaagggtccaagagactaacctggttttaactgggaaaaatgtcactttatggtgactgaaaggattgtccttgggcacaaaaatftgaacaagggaatagagggtgatcaagctaaaggtagaggtaattgaaaaattaccaccacctccaacgtaaggcaatcagaagctttctgggcatgaggatfttataggaggtttataaaggatttcaaacctctaagtaactgtagctgctgacacaccatttctttgataaggagtttctgcagggtattgagactctgaaagctaatggtcacagcacaatcatctctgaccaaactggatattaccatttgaactgatgtgtaa

>Ai-Athila-9

aaggacacacaagacagctctgggtggtccataagtatctcaaggcattagcccagccagatgcatgcacaagatcctattagaggatgatgctaagccagtgttcaaccacagagggcgtgaatccggccatgaaggagggtgtgcagaagaaggctactaaactactagaggctgggattattatcctatttctgatagccctgggtgagccctgtccaagttgtcccaagaagggaagcatgacagtggttcataataaaaaaatgaactggttctacaagaacagttataggggtgcatgtgtattgactacagaaggctcaatacagccaccagaaggatcatttccctaccattcatagaccagatgctagagagactagcaggctatgaatactactgcttttggatggctattcagaattataccaaactcaatacagtaaatctcaggaccaagagaaaacagcatttcatatgctcacttggagatttgcatacagaaggatgcttttggctgtgc aatgcacctgcaaccttcaaaagggtgaatgctctctatttctctgatatggtgagaaaaaatfttctggaaagtcttcatggataacttttagtatttggagactcattcagctcctgtcttgaccatctagcactgttctgaaaaagggtccaagagactaacctggttttaactgggaaaaatgtcactttatggtgactgaaaggattgtccttgggcacaaaaatftgaaacaagggaatagagggtgatcaagctaaaggtagaggtaattgaaaaattaccaccacctccaacgtaaggcaatcagaagctttctgggcatgcaagattctatagagggtttataaaggatttcaaaaatgcaaaagctctgagcaatctgtagctgctgacacccatttctttgacacagagtgctgcaggcgtttga aactctaaaagctaaagctgcatagcaccagctatttctgaccagactggacattaccatttgaactgata

>Ai-Athila-11

aaggacacacaagacagctctgggtggtccataagtaaccttaaggccataagcccagctagatgtagcacaagatcctattggaggatgatgctaagcagtaggttcaaccacagagggcgttaaatccagccatgaaggaaagtggtgcagaagaaggctaccacagttactggaggctgggattattatcctatttctgatagccctgggtgaagccatgtccaagttgtccgcaaaaaggaggatgactgtgttcataatgaaaaaatgaactggttctacaagaacagttatagggtagcgcagctgtattgactataggaggtcaatacagccaccacaaaaggatcatttccctaccattcatagaccagatgctagaaagactagcaggctatgattactgcttttggatggctattcaggttacaaccaaatgcagtagatccccaggatcaagagaaaacagtaattcacatgtccatctggagatattgcttacagaaggatgccatttgggctgtgtattgcacctgcaacctttcagagatgcatgctctctatttctctgatatggtggaaaaatfttctggaaagtcttcatggatgacttctaatatattggagactcattcagctcctgtcttgatcacctaacactggttctgaaaagatgccaagagactaacctggttttaactgggaaaaatgtcactttatggtgactgaaaggattgtccttgggcacaaatftgaaacaagggaatagagatgtagatcaagctaaaggtagaggtaattgaaaaattaccaccacctccaacgtaaggcaatcagaagctttctgggcatgcaagattctatagaaggtttataaaggattttcaaaaatgcaaaacctctgagcaatctgtagctgctgacacccatttctttgacacagagtgctgcaggcgtttgaaacctctaaaagctaaagctggtcacagcaccagttatttctgaccagactggacattaccatttgaactgatgtgtaa

>Ai-Athila-19

gacacacaagacagctcttgggtggtccataagtaaccttaagggcacaagcccaactagatgcatgcacaaaatcctattggaggatgatgctaagctagtgggtc
aaccacagaggcggcctaaatccagccatgaaggaaagtggtgcacaaaagaggtcaccaaaactctggaggctgggattatccattctgatagccctgggtg
agcctgttcaagttctcccaagaaggaggatgacagtggttcataatgaaaaaatgaactagtctctacaagaacagttacagggtgagcagatgtgtattgac
tacagaaggctcaatagccaccagaaagatcatttcttaccattcatagaccatagctagagagactagcaggatgaatattactgattttggatggctatt
caggttacaacaaattgcagtagatccccagaccaagagaaaacaacattcacatgccctctggagatttactaacagaagaatgcccttttggctatgtaatgc
acctgcaacctttcagagggtgcatgctctctctctctgatatggtagagaagttctggaagtcttcatggatgactttcagatattggagactcatttaactcctgct
tgaccatttagcactgttctgaaaagatgccaaagagaccaacctagttaaaactgggaaaaatgtcactttatggtagctgaaggaaattgccctgggcataaaatctc
aaacaagggaatagagggtgatcaagctaaagtggagtaattgaaaaataccaccacctgctaattgtaaggcaatcagaacattctggggcatgcaggattct
acaggaggttataaaggattttcaaaaattgcacaaacctgagcaatctcctgtagctgaca

>Ai-Athila-20

ctaaatccagccatgaaggaaagtggtgcagaaagaggtcaccaagttactggaggctgggattatctctattctgatagcccctgggtgatccctgtccaagtg
tccccaaaaaggaggatgacagtggttcataatgaaaaaatgaactagtctctacaagaacagttacagggtgagcagatgtgtattgactatagaaggctcaat
acagccaccagaagagatcattcttaccattcatagaccagatgctagaaagactagcaggatgattattactgcttttggatggctattcaagctacaacaa
attgtagtagatcccaggatcaagagaaaacagatcattcatgctcctctggagatttctctacagaaggatgccatttgggtgtgtaatgcacctgcaaccttca
gagatgcatgctctctatttctctgatatgggtgaaaaatcttggaaagttctcatgtagacttctcagatattggagactcattcagctcctgtcttgatcacctgacct
ggttctgaaaagatgccaaagagactaacctgggtttaaactggaaaaatgtcactttatggtgactgaaggattgctctgggcacaaagattcgaacaagggaata
aagggtgagatcaagcagaagtaattgaaaaatfaccaccacctcaatgtaaggcaatcagaagccttctggggcatgcaggattctataggaaggttataa
aggattttcaaaaattgcacaaaccttaagcaatctgtagctgctgacatgccatttatggttgacacagagtgctgcaggcgttgaacctgaaaagtaagctgg
tcacagcaccagttattctgaccagactggacattaccattgaaactgatgtgtgaa

>Ai-Athila-21

ctaaatccagccatgaaggaaagtggtgcagaaagaggtcaccaaaactggaggctgggattatctctattctgatagcccctgggtgagccctgttcaagtcg
tccccaaaaaggaggatgacagtaattcataatgaaaaaatgaactgttctctacaagaacagttacagggtgagcagatgtgtattgactacagaaggctcaat
acagccaccagaagagatcatttcttaccattcatagaccagatgctagaaagactagcaggatgattattactacttttggatggctactcaggctataaccag
attgtagtagatcccaggatcaagagaaaatagcattcatgctcctcgaaggttctctatagaaggatgccatttgggctatgtaatgcctgcaaccttca
gagatgcatgctctctatttcttctgatggtgaaaaatcttggaaagttctcatgtagacttctcagatattggagactcattcagctcctgtcttgatcacctgaaact
tgttctgaaaagatgccaaagagaccaacctagttaaaactaggaaggtgctcattctatggtgactgaaggaaattgctctgggcataaaggctcaaaataagggaata
gagggtgagatcaagcaaaaatagaggtaattgaaaaatfaccaccacctgcaatgtaaggcaatcagaagccttctggggcatgcaggattctataggaaggttata
aaggattttcaaaaattgcacaaaccttaagcaatctgtagctgctgacacgcatttgggttgacacagagtgctgcaggcgttgaacctgaaaagtaagctgg
gtcacagcaccagctatttctgaccagactggacattaccattgaaactgatgtgtgaa

>Ai-Athila-22

gtacgcacaagatcctattggaggatgatgccaagcagtagttcaaccactgaggcggcctaaatccagccatgaaggaggtagtgcagaaagaggtcaccaagt
tactagaggctgggattatctctattctgatagcccctgggtgagccccgtccaagttgtccccaaagaggaggtgatagtggttcataatgaaaaaataaa
ctggttctacaagaacatttacagggtgagcgtatgtgtattgactatagaaggctcaatacagccaccaggaagatcatttcttaccattcatagaccagatgct
agaagactagcaagtcattaatfctgcttttggatggctattcaggttacaacaaatgcagtagatcctcaggactaagagaaaacagcattcacatgcttct
tggagatattgctacagaaggatgcttcttggctgtgcaatgcacctgcaaccttcagagggtgcatgctctctctctctgatatggtagagaattcctggaagtg
ttcatgtagactttcagatatttggagactcattcatctctgcttggaccatttagcacttggctgaaaaggtgccaaagagactaacctagttaaaactggaaaaatg
tcaattatggtgactgaaggatgtccttgggcataaaatfcaacaagggaatagagggtgatcaagctaaagttagaggtaattgaaaaatfaccaccacctgccc
aatgtaaggcaatcagaagccttctggggcatgcaggattctatagagggttataagagattttcaaaaattgcacaaacctgagcaatctactagctgctgacac
gccatttgggttgacacagagtgctgttagggcgttgaacctgaaaagctaaagctggtcacaacaccagttatttctgaccagactggacattaccattgaaactgat
gtgtgaa

>Ai-Athila-24

gtacgcacaagatcctattggaggatgatgccaagcagtagttcaaccactgaggcggcctaaatccagccatgaaggaggtagtgcagaaagaggtcaccaagt
tactagaggctgggattatctctattctgatagcccctgggtgagccccgtccaagttgtccccaaagaggaggtgatagtggttcataatgaaaaaataaa
attggttctacaagaacagtcacagggtgagcgtatgtgtattgactacagaaggctcaatacagccaccagaaaggatcatttcttaccattcatagaccaaatg
ctagaaagactagctgcatgattactgcttttggatggctattcaggttacaacaaatgcagtagatcctcaagaccaagagaaaacaacatttacttgcctt
ctggcgtgttgcctacagaaggatgcttcttggatgtgtaatgctcctgcaaccttcagagatgcatgctatccatctctctgataggtggagaaattcttgaagtc
ttcatgtagacttctcagatatttggagactcattcagctcctgtcttaaccacctatcattgtcctgaaaaggtgccaaagagactaacctggtttaaactgggagaaa
gtcagttatggtgactgaaggaaattgcttggacacaaaatfcaacagggggaatagagggtgataaggcaaaaggtagaggtaatcagaagaaatfaccaccacc
tgcaatgtaaggcaatcagaagccttctggggcatgcaggattctacaaaagggttataaaggattttcgaaaaattcaaaaaccttggatgaactgctagctgctga
cacaccatttgggttgacacacagtgctgtagggcatttggagacctgaaaagctaaagctggtcacagcaccagtcattctgaccagattggacattgccattgaa
ctgatgtgtgaa

>Ai-Athila-29

taaggacacacagacagctcttgggtggccataagtgatctcaaggacattagccaagcagatgatgcacaagatcctattggaggatgatgctaagccagtggttcaaccacagaggcggctgaaccagccatgaaggaggtgtacagaaagggtcactaagtactagaggctaggattatctattctgatagcccctgggtgagccgcaagtggtcccaagaaggagcattataaggtggcgtatgtgtattgactacagaaggctcaatacagccaccagaaatgatcatttcccttaccattcatagaccagatgctagagagactagcagaaaggatcatttcccttaccattcatagaccagatgctagagagactagcaggtcatgactactgcttttggatggctattcaggtfacaaccaaattacagtagatccacaggaccaagagaaaacggcattcacgtgctcatctgggtattcgcctacagaaggatgccatttggctgtgcaattcatctgcaaccttcagagatgatgctctctatttctctgatatggtgaaaaatftttggaagtctcatggatgacttcgcagatttggagactcattcagctcgtcttgaccatctaacactgttctgaaaaagggtccaagagactaacctggtftaaactaggagaaatgacacttatggtgactgaaaggaatttccctgggcacaaatfttcgaaacaaggatagagggtgatcaagctaaaggtagaggtaattgaaaaatftaccaccacctgccaatgtaaggcaatcagaagcttctgggcatgacagattctatagaaggttataaaagatttctgaaaaatgcaagcctctgagcaatctcctgtagctg

>Ai-Athila-30

aaggacacacaagacagctcttgggtggccatcagtgatcttaagggtattagcccagccaatacatgcacaagatcctattggagggtgacgctaagccagtggttcaaccacaaaggcggcttaatccagccatgaaggaggtgtgcagaaggaggtcactaaatfactagaggctgggattatttctattctgatagcccctaggtaagccctgtccaagctgtccctaaagagggtgcatgacagtggtcataatgaaaaaatgaaactggtcctacaagaacagttacagggtggcgtatgtgtattgattatataaggctcaatacagctaccagaaaggatcatttcccttacccttcatagacagatgtagaaagactagcaggtcatgaaactactgcttctggatggatgatacaggttataatcaaatgtagatgctcccaagatcaagaagaaacaacatfcacatgctccatccggagatfttgcatacagaaggatgccatttggctgtgcaatgcacctgcaaccttcaagggtgacgctcctcaatttctctgatatggtggaaaaatftatggaaagtctcatggatgactttcagatttggagactcattcagctcctatctgaccatctagcactgttctaaagaggtgccaagagactaacctggtftaaactgggagaaatgacactttaggtgattgaaaggaattgcttgggcacaaaatftgaaagagggaatagaaggtgacaaagctaaagtagaggttaattgaaaaatftaccaccacctaccaatgtaaggcaatcagaagcttctgggcatgacagattctataggaggttatcaaaagatttcaaaaatftccaacctctgagtaactgctgactgctgacacaccatttatcttggataaggagtgctgcaggcgttgagactctaaaagctaaagctgtgcacagcaccagctcatctctgaccagactggacattaccattgaaactaact

>Ai-Athila-17

gcatgctcccggccgcatggcggccgcccgaattcgattaaggacacacaagacagctcttgggtggccataagtatcttaagggtcattagcccagcaagatgcatgcacaagatcctattggaggacagtgctaaagccagtggttcaaccacagaggcggctaaatccagccatgaaggaggtgtgcagaagagggtcactaagttactggaggctgggattattatctatttctgatagcccctgggtgagccctgtctaaagctgacccaaaaaggaggtgacagtggtcataatgaaaaatgagaagtggttctacaagaatgtagcaggtgagcgtatgtgtattgactacagaaggctcaatacagctaccagaaaggatcatttcccttaccattcataaaccagatgctagaagactagcaggtcatgaaatfactatttggatggctatttaggttacaacaaatgtagatcctcaggaccaagagaaaacagcattcacatgctctctggagtggttgcctacagaaggatgcttggctgtgtaatgacacctgcaaccttcaaaagtgatgctctctatttctgatatggtagaaatftctgaaagttttatggatgactttcagatttagagactcattcagctctgcctgaccatttagcactgttcttaaaagataccaagagacaaactagtttagactgaaaaaatgctactttatggtgactgaaggaaftgctcttgggcataaaatftcaatcaaggggatagaggtagatcaagctaaagtagaggtaattgaaaaatftaccaccctgccaatgtaaggcaatcagaagcttctgggcatgacagattctataggaggtttataaaggattttcaaaaatftataaacctctgagtaactgctgactgctgacacgccatgtgttgacacagagtgctgcaggcgttgaactctgaaagctaaagctgtgcacagcaccagttatttctaccagactggacattaccattgaaactgatgtgaaatcactagtgaaatcgccggcctgcaggctgacatattgggagagctccaacgcgt

Seqüência consenso entre *A. duranensis* e *A. ipaënsis*

>Consensus/1-803 Percentage Identity Consensus

aaggaggtagggtcagaagaggcactaaatfactagaggctgggattattatctatttctgatagcccctgggtgagccctgtccaagttgtcccaagaaggaggcatgacagtggtcataatgaaaaaatgaaactggtcctacaagaacagttacagggtggcgtatgtgtattgactacagaaggctcaatacagccaccagaaaggatcatttcccttaccattcatagaccagatgctagaaagactagcaggtcatgaaactactgcttftggatggctattcaggttacaaccaaattgcagtagatctcaggaccaagagaaaacagcattcacatgctctctggagtattgctcacaagagatgccatttggctgtgcaatgcaactgcaaccttcagaggtgcatgctctatttctctgatatggtggaaaaatftctggaagtctcatggatgactttcagatttggagactcattcagctctgtctgaccatctagcactgttctgaaaagatgccaaagagactaacctggtttaaactgggaaaaatgcaactttatggtgactgaaagaaatgctcttgggcacaaaatftcaaacaaagggaatagagggtgatcaagctaaagtagagtaattgaaaaatftaccaccctgccaatgtaaggcaatcagaagcttctgggcatgacagattctataggaggtttataaaggattttcaaaaatftgcaaaaactctgagcaatctgtagctgctg

Seqüências traduzidas de clones representativos obtidos na árvore filogenética e publicadas por Wright e Voytas, 2001.

>Tat4-1

LGAGSIVEVKYPEWLANPVVVKKNGKLRVCIDFTDLNKACPKDSFPLPHIDRMVEATTGNELLSFMD
AFSGYNQIPMHKDDQEKTSFITDRGTYCYKVMPPFGLKNAGARYQRLVNQMFAPQLGKTMEVYIDDML
VKSKKSADHIEHLTACFETLNKYNMKNLPAKCSFGVTSGEFLGYIV

>Rire2

LAAGFIKEVLHPDWLANPVLVRKKTGQWRMCVDYTDLNKSCPDPFGLPRIDQVVDSTAGCELLSFLD
CYSGYHQIRLKESDCLKTSFITPFGAYCYVTMPFGLKNAGATYQRMIQRCFSTQIGRNVEAYVDDVVV
KTKQKDDLISDLEETFASIRAFRMKLNPEKCTFGVPSGKLLGFMV

>Vulgar

LAANFVREVHSEWLANVVMVPKDKSLRMCIDFKHVNVSPKDHFLPRIDQIVDSTAGCERLSFLD
AYSQYHQIRLYGPDELKTAFITPFGFCYITMPFGLKNAGATFMRMIQKCLDQIGRNVEAYMDDIVVK
SRKGSNLLTDLAETFANLRRYDIKLNPAKCSFGVPSGKLLDFV

>Cinful

LSAGVIREVKYPEWLANPVMVKKANGKWRMCIDFTDLNKACPKDEFPLPRIDSLVDATASSELMSLLD
CYSGYHQIWMKREDEPKTSFITPSGTYCYLRMPEGLKNAGGSFSRMTAKVLQSQIGRNVLTYVDDIIVK
STKQENHIADLQETFASFRQAGLKNPEKCVFGVKKGKFLGCLV

>f26h6

LDAGSIVEVRYPDWLRNPVVVKKNGKWRVCIDFTDLNKACPKDSFPLPHIDRLVEATAGNELLSFMD
AFSGYNQILMHQNDREKTVFITDQGTQCYKVMPPFGLKNAGATYPRLVNQMFTDQLDHSMEVYIDDML
VKSLRAEEHITHLRQCFQVLNRYNMKLNPSKCTFGVTSGEFLGYLV

>Grand1-4

VAAGFIREVLHPEWLANPVLVLKKNKVDWRMCVDYTDLNKHCPKDPFGLPRIDQVVDSTAGCSMLSF
LDCYSWYHQISLAKEDDEKTAFITPFGAFICYTSMFGLKNGRATYQRAIQTCLANHWGKRVEAYVDDV
VIKIENSENFIEDLQLVFNSLRRYRWKLNPEKCVFGVPAGKLLGFIV

>tob2-2

LDVGVVYPISDSSWTSPVQCVPKKGGMTVVKNKSNELIPTRITITGWRVCMDYRKLNVTKDHFPLPF
LDQMLDRLAGRAFYCFDDEYSQYQIPIAPEDPEKTTFTCPYGTFFVFSRMPFRLCNAPATFQRCMMAIFS
YMKDIFEVFMDDFSVVGHSFDECLNLDRLVLAHCEETNLVLNWEKCHFVVEEGIVLGHKI

>tob1

LDAGVIYPIYDSSZTSPVQCVPKKGGMTVVTVNEKSNELIPTRMVTGWRVCMDYRKLNVTRKDHFPFPF
LDQMLDRLACRAFYCFDZVSGYSQIFIAPZDHEKTTFTCPYGTFFAYKRMPPFGLCNALANFYRCMMAIF
TDMVKDYLVFMDDFSVVGDSFDDCLENLDKVLARYEETNLVLNWEKCHFMIIEGIVLGHKI

>tob4-1

LDVGVVYPISDSSCISPVQCVPKKGGMTVVANSQNGLIPTRIVTGWKCVCMDYRKLNVTRKDHFPPLPF
DQMLDRLAGRAFYCFDGYSGYNQIFITPEDQEKTTFTCPYGTFFAFSRMPFGLCNAPTTFZRYMMAIFT
DMVEDILEVFMDDFSVVGDSFDECLNLDRLVLAHCKETNLVLNWEKCHFVVEEGIVLGHKI

>tob5-3

LDVGVVYPIFDSSWTLPVQYVPPKGGMTVVTVNKNELIPTRITVTGWRVCMDYHKLNVTRKDHFPPLP
FLDQMLDRLAGCAFYCFDGYSGCNKILIAPKDQEKTTFTCTYGTFFVFSRMSFGLCNAPTTFZRCMMAI
FTYMVEDILEVFMDDFSVVGZDFDECLNLDRLVLAHCEANLVLNWEKCHFVVEEGIVLGHKI

>ricel

LYARIIYLVPISEWVSPVQVVPKGGMTAVANAQNELIPQRTVTGWRMCIDYRKLNVTRKDHFPPLPF
IDEMLERLANHSFFCFDGYSGYHQIPIHPEDQSKTTFTCPYGTFFAYRRMPFGLCNTPASFQRCMMSIFS
DMIEDIMEVFMDDFSVYGKTLGHCLQNLDKVLQRCQEKDLVLNWEKCHFVVEEGIVLGHRI

>rice2-10

LHARIIYPVPYSERVSPVQVVPKKGGMVAVANAQNELITQQTVTGWRMCIDYRKLNKATKKDHFPLPFI
IVEMLERLANHSFFCFLDGYFGYHQIPIHPEDZSKTTFTCPYGTIAYRRMSFGLCNAPASFQRCMMSIFS
DMIEDIMEVFMDDFSVYGKTFGHCLQNLDKVLQRCQEKDLVLNWEKZHFVREGIVLGHVRV

>rice2-17

LHAGIIYTPCSEWVSTVQVGPVKMGZMTVVANAQNKLIPOPTITGWRMCIDYRKLNKATKEDHFPLPFI
DEMLERMTNHSFFCFLDGYSGYHQIPIHPEDZSKTTFTCPYGTIAYRRMSFGLCNAPASFQRCMMSIFS
MEDIMKVFMDDFSUYGKTFGHCLZNLDKVLQRCQENDLVFNWEKCHFVREGIVLGHVRV

>rice5-2

LHAEIIPVPYREWVSPVZVMPKKGRTVIANANAQNELIPQRTVTGWRMCIDYMKLNKATKKDHFPLPFI
DEMLERLANHSFFRFLDGYSTRYDQIPIHPEDZSKTTFTCSYDTIAYRRMSFGLCNAPASFQRCMMSIFS
MIKDIMEVFMDDFSIYGKTSFGHCLQNLDKILQRCQEKDLVLNWEKCHFVREGIVLSHRV

>syc2-3

LKVZVIYPIZDRNWVSPVQVVPKIGITVVKZNDELVPVTSVQNGWRVCIDYRKLNVVTRKDHFLPFI
DQMLERLVGHSYYCFLDGYSSYFQIVITPEDZEKTTFTCPFGTFAYRCMPFGLCNAPTTFQRCMVSIFS
YIENIEVFMDDFIVYGDSFNNFLHNLTLVLQRCIETNLVLNWEKCHFVMEQGIVLGHVI

>syc4-2

LEVGVIIYLISNSNWVSPVQVAPKKTGITVVKQNDELVPVTHVQNGWVVCINRKLNVITCKDHFPLPFI
DKMLERLAGHSYYCFLDGYLGYFQIAITSEDQEKMIFKCPFGTFAYRHMPPFGLCNAPTTFZRCMVSIFS
YIENIEVFMDDFTVYGDSFDNCLHNLTLVIQRCIETNLVLNSZKCHFVMEQGIVLGHVV

>syc4-3

LDVGIYPIFYSNZVSPTQVVPKNSGVTVVKANDELIPNRLTIGWRVCINRKLNVTRKDHFLPFI
ZILERVAGHKFYFLYGYSTRYNZIEIAPEDZENTFTCPFGTFAYRRMSFGLCNALATFZRCMVSIFS
VEHFLEVFMDDFFVFGNSFDDCLHNLKVLNRCEEKNIILNZEKCHFVMSKRIVLGHVI

>syc4-7

LEVGVIIYPISDSNWVSPVQVVPKKTGITVVKQNDELVPVTRVQNGWQVCIDYIKLNVVTRKDHFLPFI
DQMFERLAGHSYYCFLDGYSCYFZIAITPEDQEKTTFTCPFGTFAYRCMPFGLCNAPATFQRCMVSIFS
YIENIEVFMDDFIVYEDSFDNCLHNLTLVFZRCIETNLVLNFEKCHVMVEZGIVLGHVV

>bar7

LEEGIIYPVAHSDWVSPVHCIPKKGGITVVPNDKDELIPZRIITGYRMVIDFRKLNKATRKHYPPLPIDQ
MLERLSKHHTHFCFLDGYTGFSQIPVAQSDQEKTTLTZHFGTFAIRMPFGLCNAPATFQRCMMAIFSDFC
EKIVNVFMDNFSVYGCSFDDCLNNVDRVLQRCQKDTNVVLNWEKCHFVMEGIVLGHKI

>bar2-4

LEAGIIYRVAHSDWLSRVHCVPKKGGITVVPNDKDELIPQRTITGYRMVIDFRKLNKATRKHYPPLPIDQ
QMRERLSKHHTHFCFLNGYFGFSQIPVAQSDQEKTTFTCPFGTFAYRRMTFGLCNAPASFQRCMMAIFPD
FCEKIVEVFMDDFSIYGSSFDDCLSNLDRVLQRCQKDTNLFLNWKCHFVMEGIVLGHKF

>bar2-12

LEAGIIYPVAHNDWVSPVHCVPKKGKITVVPNDKDELIPHRIITGYRMVIDFRKMNKATRKEHYPLPFS
QMLERLSKHHTHFCFLDGYSSFSQILVAQSDQEKTTFTYFPGTFAYRRMPFGLCNAPATFQRCMMAIFS
FCEKFVEVFMDDFSVYGSSFDDCLSNLDRVLQRCQKDTNLVLNWEKCHFVMEGIVLGHKI

>bar2-19

LDEGIIYHVAHSDWVSPVHVPKKGGITVVPNDKDELIPQRIITGYRMVIDFRKLNKATRKHYPPLPIDQ
HMLERLSKLTHFCFLDGYSSFSQIPVAQSDQEKTTFTCPFGTFAYRRMPFGLCNAPATFQRCMMAIFS
CENIVEVFMDDFSVYGSSFDDCLSNLDRVLQRCQKDTNLVLNWEKCHFVMEGIVLGHKI

>rye3-4

PEAGIIYPVADSQWVSHVHCVPKKGGMTVVPNDKHELIPQRIITGYRMVIDFRKLNKATKKDHYPLPFI
DQMLDRLSKHHTHFCFLDGYGFSQIPVSKGDQEKTTFTCPFGTFAYRRMPFGLCNAPATFQRCMMAIFS
DFZEKIVEVFMDDFSVYGTSFDDYLSNDRVLQRCQKDTNLVLNWEKCHFVMEGIVLGHKI

>rye4-4
LEAGIIPVADSQWVSHVHYVPKKGGMVVPNDKDELIPQRIVTGYRMVSDFRKLNKATKKDHYPLPF
IDQMLERLSKHTHFFFLDGYSGFSQIPVSKGDQEKTTFTCTFGTFAYRRMPFGLCNAPATFQRCMMAIFS
DFCEKIVEVFMDDFSVYGTSFDDCLSNLDRVLQRCEDTNLVLNCEKCHF MVNEGIVLGHKI

>rye5-2
LEAGINYPIADSQRVSHVHCVPKKGGMVVPKDKDEFIPQRIVTGYRMVIDFRKLNKATMKDHYPLPFI
DQMPDRLSKHTHFCFLDGYSGFSQIPLSKGDQEKTTFTCPFGTFAYRGMPPFGLCNAPATFQRCMIVIFS
VFEKIVEVFMDDFSVYGTSFDDCLSNLDRVLQRCEDTNLVLNWEKCHF MVNEGIVLGHKI

>rye5-4
LEAGIIPVADSQWVSPVHCVPKKGGMVVPNDKDELISQRIVTGYRMVIDFRKLNKATKKDQYPLPFI
DQMLERLSKHTHFCFLDGYSSFSQIPMSKGDQEKTTFTCPFGTFAYRRMPFGLCNASATFQTCMMAILY
DFCERIVDVFMDDFCIYETSFDDCLSNLDRVLQRCETNLVLNWEKSHF MVNEGIVLGHKI

>wheat3-1
LEVGIIPVADSQWVSPVHCVPKKGGMVVPNDKDELIPQRIITGYRMVIDFRKLNKATKRDHYPLPFID
QILERLCKHHTHYCFDGYPGFSQIPVSAKDQSKTTFTCPFGTFAYRCMPFGLCNAPATFQRCMMAIFSDF
CEKICEVFMDDFSVYGSSFDDCLSNLDRVLQRCETNLVLNWEKCHF MVNEGIVLGHKV

>wheat5-3
LEAGIIPVADSKWVIPVHVVPKKGGMVVPNDKDELIPQRTITGYRMVIDFRKLNKATKKYHYPLPFI
QMLERLSKHTHFCFLDGYSGFSQIPVSAKDQSKTTFTCPFGTFAYRRMPFGLCNAPATFQRYMMAILS
D FCEKICEVFMDDSSIYGSSFDDCLSNLDRVLQRCETNLVLNWEKCF MVNEGIVLGHKV

>wheat8-2
LEAGIIPVADSQWVSPVHCVLKKGGITVVPNDKDELIPQRIITGYRMVIDFRKLNKATKKDHYPLPFID
QMLERLCKHHTHYCFDGYSGFSQIPVSAKDQSKTTFTCPFGTFAYRRMPFDLCNAPATFQICMMAIFS
D FCEKICEVFMDDFSVYGSSYDDCLSNLNRVLQRCETNLVLNWEKCHF MVNEGIVLGHKV

>wheat8-5
LEAGIIPVAESQWVSPVHCVPKKGGMVVPNDKDELIPQRIITGYRMVIDFRKLNKATKKDHYPLPFID
QMLERLCKHHTHYCFDGYSGFSQIPVSAKDQSKTTFTCPFGTFAYRRMPFGLCNAPSTFQRCMMAIFS
D FCEKICEVFMDEFVSVYGSSFDDCLSNPDRVLQRCETNLVLNWEKCHF MVNEGIVLGHKV

>wheat8-11
LEAGITYPVADSEWVSPVHCVPKKGGMVVLNDKDELIPQIIITGYRMVIDFHKLNKATKKDHYPLPLID
QILERLSKHTHFCFLDGYTGFSQIPVSVKDQSKTTFTCPFGTFAYRLMPFGLCNAPSTFQRCMMAIFS
V FCEKICEVFMDDFSVYGSSFDDCLSNLDRVLQRCEDTSLILNWEKCHF MVNEGIVLGHKI

>pea1
LDSGMIYPISDSSWVSPVHVVPKKGGMVSVILNEKNELIPTRTGTGWRVCIDHRRNLNTATRKHDFPLPFID
QMLERLAGHEYCYCFDGYSGYNQIVVAPEDQEKTAFTCPYGFAYRRMPFGLCNAPATFQRCMITSIFS
D MLEKYMKVFMDDFSVFGSSFDNCLANLSVLQRCQETNLVLNWEKCHF MVNEGIVLGHKI

>pea8-1
LDAGMIYPISDSPWVSPVHVVPKKGGMVIRNDKDELIPTKVATGWRICIDYRQLNTATRKHDFPLPFM
DQMLERLSGQQYYCFDGYSGYNQIADVDPDHEKTAFTCPFGVFAYRKMPFGLCNAPATFQRCVLAIF
ADLIEKTMDFVFMDDFSVFGGTFSLCLANLKTVLERCVKTNLVLNWEKCHF MVNEGIVLGHKV

>pea9-1
LDAGMIYPISDSPWVSPVHVVPKKGGMVIRNDKDELIPTKVETGWRMCIDYRRLNTATRKHDFPLPFM
DQMLERLSGQQYYCFDGYSGYNQIADVDPADHEKTAFTCPFGVFAYRKMPFGLCNAPATFQRCVQAF
VDLIEKTMDFVFMDDFSVFGGFSLSLCLANLKTVLERCVKTNLVLNWEKCHF MVNEGIVLGHKV

>will2
LEAGLIYPISDSAWVSPVQVVPKKEGKTVIKDEKDELISTRITGWRMCIDYQKLNDAATRKHDFPLPFM
DQMLERLAGQSYCYCFDGYSGYNQIDVDPKDQEKTAFTYFPGVFAYRRMPFGLCNAPATFQRCMMTIF
SDMVEKZIEVFMDDFSIFGSPFEGCLSNLERVLKRREESKLVNWEKCHF MVNEGIVLGHKI

>will3

LEAGLIYPISDSSWVSPVQVVPKKGGMVVVKNDRNELIPTRRVTRWRMCIDYRKLNEATRKHDFPLPFMDQMLERLARQSFYRFLDGYSGYNQIAVDPQDQEKTAFTCPFGVFAYRRMPFGLCNASTTFQRMMMAIFDDMVEKCIIEVFMDDFSFFGASFGNCLANLEKVLQRCEKSNLVLNWEKCHF MVQEGIVLGHKI

>will8-2

LEVGLIYPISDSAWVSSNZVVPKKGGMVTIHNDKNDLIPTQTIRWQMCIDYHKLNDVTKKDHFPLPFMDQMLERLAGQAFYCFLDGYSGYNQIAVHLKDQEKTTIICPFGVFAYRQMSFELCNAPTTFZRFMMAIFA DLVEKCIIEVFMNDFSIFGSSFYHCLSNLELVLQRCAETNLLMNWEKCHF MVQEGIVLGHKI

>will8-3

LEVGLIYLISDSAWVSLVQVAPKKCGMTVVQNERNDLIPTRTVTGZRMCIDYCKLNEATRKHDFPLPFMDQMLERLAGQAYYCFLDYRYSYGNQIAVDPDQEKMAFTCPFGVFAYRRMSFRLCNAPATFQRCMLAIFSDMVEKSIEVFMDEFSIFGPLFDSCLRNLNEMVLQRVCZTNLVLNZEKCHF MVREGIVMGHNI

>L852

LEAGLIYPISNSTZVSPVQVVPKKGGMVVQNEKNDLIPTRTVTSWRICIDYRKLNEATRKHDFPLPFMDQMLERLAGQAYYCFLDGYSRYNQIAVDPDQEKTTFTCPFGVFAYRRMPFGLCNAPATFQRCMLAIFSDMVEKNIEVFMDDFSVFGPSFDSCLRNLNEMVLZRCVETNLVLNWEKCHF MVREGIVLSHKI

>L858-2

LEVGLIYLISDSAWVSLVQVAPKKCGMTVVQNERNDLIPTRTVTGZRMCIDYCKLNEATRKHDFPLPFMDQMLERLAGQAYYCFLDYRYSYGNQIAVDPDQEKMAFTCPFGVFAYRRMSFRLCNAPATFQRCVLAIFSDMVEKSIEVFMDEFSIFGPLFDSCLRNLNEMVLQRVCZTNLVLNZEKCHF MVREGIVMDHNI

>L859-2

LEVGLIYPISDSAWVSPVLVVSKEGGMVTIRNEKNDLIPTRTVTSWKLCIDYRKLNEATRKHDFPLPFMDQMLERLAGHAYYCFLDAYFGYNQIVDPKDQEKMAFTCPFGVFAYRRIPFGLCNAPTTFQMCMLAIFA DIVEKSIEVFMDDFSVFPVPSLESCLKLEMVLQRVCVETNLVLNWEKCHF MVREGIVLGHKI

>L859-3

LEAGLIYLISDSSZVSPVHVALKKGGMTVIKNDRDELIPTRIVTGWRMGIDYKKLNEATRKHDFPLPFMDQMLERLAGQSSYYLLDGYSGYNQIAVDPDQEKTAFTCPFGVFAYRRMSFGLCNAPTTFQRMMMAIFADMVKKCIIEVFMDDFSVFGASFENCLANLEKVLQRYEESNLVLNWEKCHF MVQEGIMLGHKI

>L859-6

LEAGLIYPISDSAWVSPMQVVPKKGGMVTIKNDKDELISTRITVTGWRMCIDYRKLNDATRKHDFPLPFMGHMLERLVGQSYCFLDGYSGYNQIVDPKDQEKTAFTYFPGVFAYQCMFGLCNAPATFQRMMMAIFSDMVEICIEVFMDDFSIFGPSFEGCLSNLEKVLKRCEESNLVLNWKCHF MVQEGIMLGHKI

>hark2

LEADLIYPISDSTWVSPVQVVPKKGGMVTIKNDKDELISTRITVTGWRMCIDYRKLNDATQKDHYSLPFMDQMLERLAGQSYCFLDGYSGYNQIVDPKDQEKTAFTCLFGVFAYKRMHFGLCNAPTTQRCMMMAIFTIFSGIVEKCIELFMDDFSIFGPSFEGYLSNLERVLQRCEESNLVLNWEKCHF MVQEGIVLGHKI

>hark5-1

LEAGLIYPISDSAWVSPVQVVPKKEGKTVIKDEKDELISTRITITGWRMCIDYQKLNDATRKHDFPLPFMDQMLERLAGQSYCFLDGYSGYNQIDVDPKDQEKTAFTYFPGVFAYRRMPFGLCNAPATFQRMMMAIFTIFSDMVEKZIEVFMDDVSIFGPSFEGCLSNLERVLKRREESKLVLNWEKCHF MVQEGIVLGHKI

>hark5

LEAGLIYLISDSAWVSPVHVVPKKGGMVTVRNEKNDLILTRITVTGWRMCIDYRKLNDAIKDHDFPLPFMDQMLERLASQSFYFLDEYSRYNQIAIHPKDQEKIAFTCPFGVFAYRRMPFELCNAPATFZRHMLAIFANMVEKCIIEVFIIDFSVFGPSFVCCLTNLELVLKYCEETNLVLNWEKCHF MVQEGIMLGHKI

>sorg4-3

YHAGIIPVPHSEWVSPVQVVPKKGGMVTVRNEKNEIPQRIVTGWRMCIDYQKLNTATKKNDFPLPFIDEMLERLANHSFFCFLDGYSGYHQIPIHPDDQEKTTFTCPYGYAZRRMSFGLCNAPASFQRMMMSIFSDMIEKIMEVFMDDFTVYGKTFDHCLNDRVLQRCEEKHLILNWEKCHF MVQEGIVLGHKV

>sorg5-2

LHAGIIPVPHSEWVSPVQVVPKKGGMVTIINEKNELIPQRTVTGWQMCIDYRKLNKATRKHDFPLPFID
EMLERLANHSFFCFLDGYSGYHQIPIHPDDQSKTTFTCPYGTAYARRMSFGLCNAPASFQRCMMSIFS
MIEEIMEVFMDDFSVYGKAFDSCLENLQVLSCEEKHLILNWEKCHFVREGIVLGHV

>sorg5-4

LHTRIIYLVPHSEWVSTVQVVPKKGGMVVRNEKNEFIPQQTVTGWRMCIDYQKLNKATRKHDFPLPFI
DEMLZWLTNHSFFCFLDGYSGYHQIPIHDDQSKTTFTZPYGTAYARRMSFRLCNAPASFQRCMMSIFS
NMIEKIMEVFTDDFTVYGKTFDDCLENLQVLSCEGKHLIVNZEKCHFVREGIVLGHV

>sorg5-5

LHAEIIVPHSEWVSPVQVVPKKGGMIVVTNEKNELIPQRTVTGWRMCIDYRKLNKATRKHDFPLPFI
DEMLERLANHSFFCFLDGYSGYHQIPIHDDQSKTTFTCPHGTAYARRMSFGLCNAPASFQRCMMSVFS
NMIEEIMEVFMDDFSVYGKTFDSCLENLDRVLRCEEKYLVLNWKKCHFVREGIVLGHV

>sorg5-6

YHAGIIPVPHSEWVSPVQVVPKKEGGMVVRNEKNEKELIPQIVTRWRMCIDYRKLNKATKKDHDFPLPFI
DEMLEWLANHSFFCFLDGYSGYHQIPIHPDDQEKTTFTCPYZTYAYZRRMSFGLCNALASFQRCMMSIFS
DMIEKIMEVFMDDFTVYGKTFDHCLNLDRLVLRCEENHLILNWEKCHFVREGIVLGHV

>sorg5-8

LHAGIIVLVPHSEWVSPVQVVPKKGGMVTIIMNEKNELIPQRTVTVWRMCIDYRKLNKATREDHFPLPFI
DEMLEWLANHSFFCFLDGYSGYHQIPIHPDDQSKTTFTCPYGTAYARRMSFGLCNALASFQRCMMSIFS
DMIEEIMEVFMDDFSVYGKTFDSCLENLQVLSCEEKHLVLNWEKCHFVREGIVLGHV

>tom4-4

LDTGIVZPISDNKZVSPVQCEPKKGDITVITNEKNELIPTMIVTZWRICMDYRKLNEATRKHDFPLPFI
MLDRLAGEZYCFDNGYLRYNQIVISPKDZEKTTFTCPYGTAYAFKIPFGLZNASATFQZCMMMAIFHDM
VEDFVEIFMNDFSVFGDSFDMCLENLDSVLASCEETNLFLNWEZZQFLVKEGIMLGHV

>tom4-10

LDARVIYPIADSSWVCLVQCVPKKGGMTVVPNEKNELVVRMPVTGWRVCMYRKLNSZTEKDYFHM
PFMDQMLDRLAGKGYWCFLDGYSGYNQISIAPEDQEKTTFTCPYGTAFARRMSFGLCNAPATFQRWM
MSIFSMMEDTIEVFMDDFSVVGDSFERCLSNLSEVLKRCEDCNLVLNWEKCHFVKEGIVLGHRI

>tom10-4

LDAGIYVPIISDNKWSLVQCVPKKGGMAMITNENNEFIPTSTVTRWRICMNYTKLNEATRKNHYPILFI
DYMLDRLAGQEYYCFLDYSGYNZILIAPEDEKTTFTCPYGTAYAFKRIPFGLCNALSINFQRCMMTIFH
DMVEYFEDIFMDDFLVWESFDRCLNLRLLARCEQTNLVLNWEKCHFLVKEGNFSGHKV

>tom10-16

LDAGVIYPIISDSSWVCPICVVPKKGGMVVPNKKNELVLMRPVTGGWVCMYRKLNAWTEKDHDFPM
PFMDQMLDRLAEGWYCFDGYSGYNZISIAPEDQEKTTFTCPYGTAFALKRMSFGLCNAPATFHRCKM
LIFDMVDDTIDAFMDDFSLVGESFERCLNHLSDVLKRCEDCNLVLNWEKCHFVKKGIVLGHRI

>cot5-3

LDDGMIYPIISNSNWVSPVHIVPKKTSATVIENSAGEIVPTRVQNGWRVCIDYRKLNSLTRKHDFPLPFID
QMLERLAGKSHYCCLDGYZGFFQIPVAPEDQEKTMFTCPFGTFYRRMPFGLCNAPASFHRCMVSIFS
YVDKIIIEVFMDDFTVYGESFEVSLTNLAKILERCLEFNVLNVEKCHFVMDKGLVLGHII

>cot8-6

LDSGMIYPIISDNNWVSPVHIVPKKTGVTVIENSAGEMVPTZVRNGRRVCIDYRKLNSLTRKHDFPLPFID
QMLEHLARKSHYCCLDGYSGLFQIPVAPEDQEKTMFTCPFGTFAYRRMSFRLCNAPTMFQRCMISIF
DYVKKIIEVFMDEFTVYSESEFEVYLSNLEKFLERCLEFNVLNVEKCHFVMDKGLVLGHII

>cot8-7

LDAGMIYSIFSDWVSWVHIVPKKTGVTVVKNSGELVPTRVQNRWRVCIDYRKLNAATRNDHFPLP
FIDQMLERLANKTHYCCLDGYSGLFQIPVAPEDQDKTTFTCPFGTFAYRRMSFGLCNAPATFQRCMVIS
FSDYVEKIIIEFFMDDFTVYGNFNECLDNLAKILQRCLEFNVLNVEKCHFVMDKGLLILGHIV

>pot5-1

LDARIVYPISDSKWWSPVKCVPKKGRMTVL TNEKNEVIPTRTVTGZRICMDYMKLN DATRKDHYPV PFI
DKILDRLAGHEYCYCFLGVYSGYNQIVIAIEDZVKTTFTCSYGT YAFKHMPFGLCNALATFQRCMLAIFH
DMVEDFVEVFMDDFLVFGESFELCLTNFDRFLARCEETNLVINZZKCHFLVREGIVLGHKI

>pot8-3

LDVGIVYPISESKWWSPVZCVPKKRGMPVITNEKNELIPTRTVTGWRICMDYRKLNEATRKHCPVPFID
QMLDRLVGQEYYCFLEGYSYGNQIVIAPEDEKTTFTCLYGT YAFKZLPFGLCNAPATFQRZMMAIFHD
MVEDFVEIFMDDFSVFRESFDRCLLENWDRV LARCEETNLILNWKKCHFLVNEGIVLGHKV

>pot8-4

LDAGIVYPISDSKWWSPVQCVPKKGGMTVITNEKNELIPTRTVTGWRICMDYRKLNEATRKHYPV PFI
DQMLDRLAGQEYYCFLDGYSYGNQIVIAPEDEKTTFTCLYGT YVSKRMSFGLCNAPSIFQRCMMAIF
HDKVEDFMEIFMDDFSVFGESFDRCLLENLDRV LARCEETNFVLNWEKCHFLVKEGIVLGHKV

>pot8-5

SDAGIVYPIYDIKWISPVHCVPKKGGMTIITNEKKELISARTVIEWHIZMDYRRLNEATRKEHYPV PFI
MLDRFIGQEYYCFLDGYSYGNQIVIAPEDEKTTFTSLYGT YAFKRMSFGPCNAPTTFQRCMTAIFHDM
VKYFVEIFMDEFLVFGESFDTCLELDNVLARCEETNPVLNWEKCHFLVKKGIVLGHKV

>pot8-8

LDAGIVYQISDSKGVYPIZFPVKKCSMTVITNEKNELIPTRTVTGWRICMDYMKLN EATRKHYPHIFID
QMLDKLAEZKYCYFLACYSRYNQFLIAPQDEETTTFTCPYGT YAFKRMSFGLCNAPTTFQRCIRAI FHD
MVEDFVEIFMDDFSVFGZSFERCLLENFDRV LAVEETNFFLNWEKCHFLVKEGIVLGHKV

>pot8-10

LDTGIVYPISDNKWASPVQCVPKKGGMTVVITNEKNELIPTRTVTGWRICMDYRKLNEATRKHYSVPFI
IDQMLDRLAGQEYYCFLDGYSRYNZIVIAPEDEKTTFTCPYGT YAFKRLPFGLCNAPTLFQRCMMAIF
HDMVEDFVKVYMDDFSVFGESFELCLSNRDRV LTRCEETNLVLNWEKCHFLVREGIMLGQKI

>oat6-1

MDAGIYPIADSEWVSHVHCVPKKGGITVVPNDNDELIPQRIVVGYRMCIDFRKVNKVTKKDHYPLPFI
DQMLERFSKKT HFCFLDGYSGFSQIVVKQDQDEKTTFTCPYGT YAYRCMPFGLCNAPSTFLRCMSAIFH
GFCEEIVEVFMDDFSVYGT SFDNCLHNLDKVLQRCEETNLVLNWEKCHFMVNEGIVLGHKV

>oat6-7

LDAGIYPIADSQWVSLVHCVPKKGGITVVPNEDNELIPQRVVVYRMCIDFRRINKVTRKDHYPLPFID
QMLERLSKKT HFCFLDGHSQIVVKAQDQDEKTTFTCPYGT YDYRRMPFGLCNAPATFQRCMSAIFH
GFCEEIVEVFMDDFSVYGT SFDNCLHNLDKFLQRFEETNLVLNWEKCHFMVNEGIVLGHKI

>oat6-8

LDVGIIYPIADSEWVSLVHCVPKKGGITVVPNDNDELIPQRIVVGYRMCIDFRKVNKVTKKDHYPLPFI
QMLERLSKKT HFCFLDGYSQIAVKQDQDEKTTFTCPYGT FAYRRMPIGLCNAPATFQRCMSAIFHG
FCEEIVEVFMDDFSVYGT SFDNCLHNLDKVLQRCEETNIVLNWEKHFHFMVNEGIVLGHKV

>hau7-2

LDAGVIYPISDSTWVSPVHCVPKKGGMTVVKNSKDELIPTRTITGHRKCIDYRKLNAASRKDHFP LPFID
QMLERLANHPYYCFLDGYSGFFQIPIHPNDQEKTTFTCPYGT FAYKRMPFGLCNAPATFQRCMTSISDL
IEEMVEVFMDDFSVYGSFSSCLLNLCRVLKRCEETNLVLNWEKCHFMVREGIVLGHKI

>hau8-3

LEAGIYPISSWVSPVHVVPKKEGVTEVKNEIDELIPTRTITGHWMCIDYRKLNAATRKHDFPLPFIDH
MLERLANHKYYCFLDGYSGFFQIPIHPDDQEKZTFTCPYGT VAYRRMPFGLCNAPATFQRGMMSIFTD
MIEDIMEVFMDDFSVYGSFLFEDCLNHCKVLARCEEKHLVLNWEKCHFRVQDGI VLGHRI

>hau8-4

LDAGVIYPISDRTWVSPVHCVPKKGGMTVVKNEKDELIPTRTITGHRMCIDYRKLNAASRKDHFP LPFID
QMLERLANHPYYCLLDGYSGFFQIPIHPNDQEKTTFTCPYGT FAYKRMPFGLCNAPATFQRCMTSIFSDL
IEEMVEVFMDDFSVYGPSFSSCLLNLCRVLTRCEETNLVLNWEKCHFMVKEGIVLGHKI

>TfcII-1

LAADIIYPISDSRWVSPIHVVPKKTGTTLVKTSDDDELVPQRVANSWRVCVDYRRLNANTRKDHFPPLPFID
SMLERLAGNVYYCFLDGYSGYNQIHIAPEDQEKTTFCTPFGTFAYRRMPFGLCNAPATFQRCMVSIFQD
MNERYLEVFMDDFSVHGDSFQTCLDNLESVLSRCVEKNLVLNWEKCHF MVQHGIVLGHII

>TfcII-2

LAADIIYPISDSRWVSPIHVVPKKTGTTLVKTSDDDELVPQRVANSWRVCVDYRRLNANTRKDHFPPLPFID
SMLERLAGNMYCFLDGYSGYNQIHIAPEDQEKTTFCTPFGTFAYRRMPFGLCNAPATFQRCMVSIFQD
MNERYLEVFMDDFSVHGDSFQTCLDNLESVLSRCVEKNLVLNWEKCHF MVQHGIVLGHII

>TfcII-3

LAADIIYLISDSRWVSPIHVVPKKTGTTLVKTSDDDELVPQRVANSWRVCVDYRZLNANTRKDHFPPLPFID
SMLERLAGNMYCFLDGYSGYNQIHIAPEDQEKTTFCTPFGTFAYRRMPFGLCNAPATFQRCMVSIFQD
MNERYLEVFMDDFSVHGDSFQTCLDNLESVLSRCVEKNLVLNWEKCHF MVQHGIVLGHII

>La8-1

LDAGIYPISDSTWVSPVHVVPKKGVTVVKNDNDELIPTRTITGHRMCIDYRKLNAASRKDHFPSPFID
QMLERLANHHTHYCFLDGYSGFFQIPIHPNDQEKTTFCTPFGTFAYRRMPFGLCNAPATFQRCMMSIFSD
LIENVVEVLMDDFSVYGDSFASCLSNLCRVLKRCEETNLVLNWEKCHF MVRDGIVLGHKI

>La9-5

LEAGIYPISDSSWVSPVHVVPKKGVTIVKNEEDEMIHTRTITGHRMCIDYZKLNAATRKHDFSVPFIDQ
MLERLANHKYYFFLDGYSGFFQIPIHPDDZEKMTFTCPYGTFA YRRMPFGLCNAPTTFQIGMMSIFTNMI
EDIMEVFMDEF SVYGSSFEDSLENLCKVLARCEEEHL VFNWEKCHLMVQDGIILGHKI

>La9-6

LDAGIYPISDSTWVSPVHVVPKKGVTVVKNDNDELIPTRTITGHRMCIDYRKLNAASRKDHFPPLPFID
QMLERLANHHTHYCFLDGYSGFFQIPIHPNDQEKTTFCTPFGTFAYRRMPFGLCNAPATFQRCMMSIFSD
LIENVVEVLMDDFSVYGDSFASCLSNLCRVLKRCEETNLVLNWEKCHF MVRDGIVLGHKI

>La9-7

LDAGIYPISDSTWVSPVHVVPKKGVTVVKNDIDELIPTRTIMGHRMCIDYRKLNAASRKDQFPLLFID
QMLERLANHHTHYCFLGGYSGFFQIPIHPNDQEKTTFCTPYGTFA YRRMPFGLCNAPATFQRCMMSIFSD
LIENVVKVFMDDFSVYGDSFASCLSNLCRVLKRCZETNLVLNWEKZHF MVRDGIVLGHKI

>La9-9

LDAGIYPISDSTWVSPVHVVPKKGVTVVKNDIDELIPTRTIMGHRMCIDYRKLNAASRKDQFPLLFID
QMLERLANHHTHYCFLDGYSGFFQIPIHPNDQEETTFTCPYGTFA YRRMPFGLCNAPATFQRCMMSIFSD
LIENVVKVFMDDFSVYGDSFASCLSNLCRVLKRCZETNLVLNWEKZHF MVRDGIVLGHKI

>favabean1

LDAGMIYPISDSSWVSPIHVMPKKGTTVIKNEKNELIPTRTVTGWZVCIDYRRLNLATRKHDFPLPFID
QMLERLADHEYYCFLDGYSGYNQIAVVPEDQEKTTFCTPFGIFS YRRIPFGLCNAPATFQRCMQSIFAD
MLEKYMEVFMDDFSVFGKSFNCLSNLALVLERCQESNLILNWEKCHF MVREGIVLGHKI

>Cyclops-2

LDARMIYPISDSPWVSPVHVVPKKGNTVIRNDKDELIPTKVATGW RMCIEYRRLNTATRKHDFPLPFM
DQMLERLSGQQYYCFLDGYSGYNQIAVDPADHZKTAFTCPFGVFA YRKMSFGLCNAPTTFQRCVQAIF
ADLNEKTMEVFMDDFSVFGVSFSLCLANLKT VLERCVKTNLVLNWZKCHF MVTEGIVLGHKV

>Athila1-1

LDAGIYPISDSTWVSPVHVVPKKGVTVVKNDNDELIPTRTITGHRMCIDYRKLNAASRKDHFPPLPFID
QMLERLANHHTHYCFLDGYSGFFQIPIHPNDQEKTTFCTPYGTFA YRRMPFGLCNAPATFQRCMMSIFSD
LIENVVEVFMDDFSVYGDSFASCLSNLCRVLKRCEETNLVLNWEKCHCMV RDGIVLGHKI

>f23m2

LEAGIYSISDSSWVSPVHVVPKKEGVTEVKNEIDELIPTRTITGHWMCIDYRKLNAATRKHDFPLPFIDH
MLERLANHKYYCFLDGYSGFFQILIHPPDDQEKZTFTCPYGTVA YRRMPFGLCNAPATFQRCMMSIFTD
MIEDIMEVFMDDFSVYGS LFEDCLENLYKVLARCEEKHLVLNWEKCHFRVQDGIIVLGHRI

>t17a11

LEAGIYIPILDSSZVSPVHVVPKKGGMVTVVKNKDELIPTRTIIGHRMCIDYRKLNAATKKDHFALPFIDQ
MLERLENHKYYCFLDGYSGFFQIPIHPDDQEKTTFYQRMPSFLCNAPATFQRGMMSIFDMIEDIMEVVFV
DDFSVYGSSFEDCLENLYKVLARCEEKHLVLNWEKCHF MVQNGIVLGHRI

>t13d4

LDADIIYIPISDSTWVSPVHCVPKKGGMVTVVKNLDELIPTRTIIGHRMCIDYRKLNAATRKHDFPLPFIDQ
MLERLANHVYYCFLDGYSGFFQIAIHPNDQEKTTFYQRMPSFLCNAPATFQRSMTSIFSDFI
EEIMEVFMDDFSVYGSSSCLLNLCRVLERCEETNLVLNWEKCHF MVQEGIVLGHKI

>t24g23

LDADIIYIPISDCTWVSAVHCVPKKGGMVTVVKNLDELIPTRTITGHRMCIDYRKLNAATRKHDFPLPFIDQ
QMLERLANHVYYCFLDGYSGFFQIAIHPNDQEKTTFNCPYGTFA YKRMPSFLCNAPATFQRSMTSIFSD
FIEIMEVFMDDFSVYGSSSCLLNLCRVLERCEETNLVLNWEKCHF MVQEGIVLGHKI

>t16i21

LDVDIIYIPISDSTWVSPVHCVSKKGGMIVVKNLDELIPARTITGHKMCIDYRKLNAATRKHDFPLPFIDQ
QMLERLANHVYYCFLDGYSGFFQIAIHPNDQEKTTFNCPYGTFA YKRMSFGLCNAPATFQRSMTSIFSD
FIEIMEVFMDDFSVYGSSSCLLNLCRVLERCEETNLVLNWEKCHF MVHEGIVLGHKI

>Athila6-1

LDAGVIYIPISDSTWVSPVHYVPKKGGMVTVVKNKDELIPTRTTTGHRCMCIDYRKLNAASRKDHFPLPFI
DQMLERLANHPYYCFLDGYSGFFQIPIHPNDQEKTTFYQRMPSFLCNAPATFQRCMTSIFSD
LIEEMVEVFMDDFSVYGSSSCLLNLCRVLKRCEETNLVLNWEKCHF MVREGIVLGHKI

>f18p14

LDASVIYIPISDSTWVSPVHCVPKKGGMVTVVKNKDELIPTRTITGHRICIDYKKNLAATRKHDFPLPFIDQ
MLKRLANHPYYCFLDGYSGFFQIPIHPNDQEKTTFYQRMPSFLCNAPATFQRYMSTIFSDLI
KEMVEVFMDDFSVYGSSSCLLNLCRVLKRCEETNLVLNWEKCHF MVREGIVLGHKI

>f7b19

LDAGVIYIPISDSIWVSPMHCVPKKGGMIVVKNZKDELIPTRAITSHRVCIDYRKLNAASKKHDFPLPFIDQ
QMLERLANHPYYCFLDGYSGFFQIPIHPNDQZKTTFTCPYGTFA YKRMPSFLCNAPATFQRCMTFIFLDL
IEEIVEVFMDDFSVYGSPSCLLNLCRVLTRCKETNLVHNWEKCHF MVKEGIVLVHKI

>t13e11

VDAGVKYPISESTWISLVHCVPKKGGMVTVVKNKDELIPTRTITWHRMCIDYRKLNVASRKDHFPLPFIDQ
QMLERLANHPYYCFFDEYSGFFQIPIHPNDQEKTTFYQRMPSFLCNAPATFQRCMTSIFSDL
IEEIVEVFMDDFSVYGSPSCLLNLCRVLTRCKETNLVLNWEKCHF MVKEGIVLGHKI

>fababean2

LDAGMIYPISESSZVSPVHVVPKKGGMVTVVKNKDELIPTRTITWHRMCIDYRKLNAATRKHDFPLPFIDQML
ERLADHEYYCFLDGYSGYNQIAVVPEDQEKTTFYQRMPSFLCNAPATFQRCMTSIFSDL
YMEVFMDDFSVFGSFDNCLNLSLFLRKFPEHLLNWEKCHF MVREGIVLGHKI

>Cyclops-1

LDARMIYPISDSPWVSPVHVVPKKGGMVTVVKNKDELIPTRTATGWRMCIEYRRLNTATRKHDFPLPFM
DQMLERLSGQQYYCFLDGYSGYNQIAVDPADHZKTAFTCPFGVFAYRKMPSFLCNAPATFQRCVQAIF
ADLIEKTMVEVFMDDFSVFGSFDNCLNLSLFLRKFPEHLLNWEKCHF MVTEGIVLGYKV

>Calypso1-1

LEAGFIYPISDSAWVSPVQVVPKKGGMVTVVKNKDELIPTRTATGWRMCIDYRKLNEATQKHDFPLPF
MDZMLERLAGQAYYCFLDGYSGYNQIAVDPADHZKTAFTCPFGVFAYRKMPSFLCNALAIQRCMLAI
FSDMVEKSIEVFMDDFWIFGSPFDNYLRNLEMVLRQCVZTNLVLNWEKCHF MVREGIVLSHKI

>Calypso1-2

LEAGLIYPPSNSAWVSPVQVVPKKGEMTVVRNEKNDLIPRRTITGWRMCINYRKLNEATRKHDFPLLF
MDQMLERLVGQAYYYFLDGYSGYNQIAVDPDQEKAAFTCPFGVFAFYRRMPFGLCNAPATFQRFMLA
IFSDMVZKSIEVFMDDFWVFGPSFNSLRNLEMVLZSZVETNLVLNWEKCHF MVQEGIVLGHKI

>Calypso2-1

ZEAGLIYPPSSDSAWVSLVQVVPKKGGM TVIKNDKDELISIRTVTGWRMCIDYRKLNDATRKHDFPLPF
MDQMLERLVGZSYCYFLDEYSGYNZIVVDPKDQEKTAFTYFPFGVFAFYRHPFGLCNAPATFQRCIMAI
FSDMVEKCIIEVFMDDFSIFGSPFKGCLLNLERVLQRCEESNLVLNWEKCHF MVQEGIVLGHKI

>Calypso2-2

LEAGLIHPISDSAWVSPVQVVLKKGGM TVIKNDKDELISTRVTVTGWRMCIDYRKLNNATWKDHYPLPF
MDHMLERLARQSYCYFLDGYSSYNZIAIDIKDQEKTTFTFPFGVFAFYRCMPFGLCNALATFQRCMMAIF
SDMVEKCIIEVFMDDFSVFGSPFDGCLSNLERVZRCESNLVLNWEKCHF MVQEGIVLGHKI

>Calypso5-1

LEAGLVYPISDSAWVSPVQVVPKKGGM TVIRNEKNDLIPTRTVTGWRMCIDYRKLNDATRKHDFPLPF
MDQMLERLAGQSFYCYFLDGYSGYDQIVVDLNDQEKTAFTTRPFGVFAFYRRMPFNLCNGPATFQRCMMA
IFADMVEKYIEVFMDDFFVFGSSFDCLANLERILQZCEETNLVLNWFHFMVQERIVLGHKI

>Calypso6-1

LEAGFIYPISDNSWVSLVQVVPKKGGM AVIRNDRNELIPTRTVTGWRMCIDYRKLNEATRKHDFPLPF
MDQMLERLAGQSFYCYFLDGYSGYNPSWITCPFGVFAFYRRMSFGLCNVLA TFQRCMMAIFAVMVEKCF
EVFMDDFSVFGASFENCLANLEKVLQRCEETNLVLNREKCNFMVQEGIVLGHKI

>Calypso1-3

LEVGLIYPISDNAWVSPVQVVPKKGGM TVVQNERNDLIPTRTVTGWRMCIDYHKLNEATRKHDFPLPF
MDQMLERLAGQAYCYFLDGYSGYNQIAVDPIDQEKTVFTCPFGVFAFYRRMSFGLCNVPATFQRCMLTI
FSDMVEKSIEVFMDDFSVFGSPFSDCLRNLEMVLQRCVETNLVLNWEKCHF MVREGIVLGHKI

>Athila4-1

LDAGVIYPISDSTWVSPVHCVPKKGGM TVVKNEKDELIPTRTITGHRMCIDYRKLNAASRKDHFPLPFID
QMLERLANHPYYCYFLDGYSGFFQIPIHPNDQEKTTFTCPYGTFAFKRMPFGLCNAPATFQRCMTSIFSDL
IEEMVEVFMDDFSVYGPSFSSCLLNLRVLRCEETNLVLNWEKCHF MVKEGIVLDHKI

>Athila4-2

LDAGVIYPISDSTWVSPVHCVPKKGGM TVVKNEKDELIPTRTITGHRMCIDYRKLNAASRKDHFPLPFID
QMLERLAKYPPYYCYFLDGYSGFFQIPIHPNDQEKTTFRCPYGTFAFKRMPFGLCNAPATFRRCMTSIFSDL
IEEMVEVFMDDFSVYGPSFSSCLLNLRVLRCEETNLVLNWEKCHF MVKEGIVLGHKI

>Athila4-3

LDAGVIYPISDSTWVSPVHCVPKKDGMI VVKNEKDELIPTRTITGHRMCIDYRKLNAASRKDHFPLPFID
QMLERLANHPYYCYFLDGYSGFFQIPIHPNDQEKTTFTCPYGTFAFKRMPFGLCNAPATFQRCMTSIFSDL
IKEMVEVFMDDFSVYGPSFSSCLLNLRVLRCEETNLVLNWEKCHF MVKEGIVLGHKI

>Athila4-4

LDAGVIYPISDSTWVSPVHCVPKKDGMT VVKNEKDELIPTRTITGHRMCIDYRKLNAASRKDHFPLPFID
QMLERLANHPYYCYFLDGYNGFFQIPIHPNDQEKTTFTCPYGTFAFKRMPFGLCNAPATFQRCMTSIFSD
LIEEMVEVFMDDFSVYGPSFSSCLLNLRVLRCEETNLVLNWEKCHF MVKEGIVLGHKI

>Diaspora

LQAGIYPISDSQWVSPVQVVPKKTGLTVIKNEKEELIPTRVQNNWRVCIHYRRLNQVTKKDHFPLPFID
QILECLAGKSHYCYFLDGFSGYMQITIALEDQEKTTFTCLFGTFAYRRMSFGLCNAPGTFQRCMISIFSDL
ENCIEEFMDDFTVYGPSFSDGCLDSLEKVLNRRRIETNLVLNFEKCHF MVQEGIVLGHII

>Baggy2

LEAGIYPVAHSDWVSPVHCVPKKGGITVVPNDKNELIPHRIVTGYRMVIDFRKLNKATRKHDFPLPFID
KMLERLSKNTHFCFLDGYSQISAAQSDQEKTTFTCPFGTFAYRRMPFGLCNAPATFQRCMMAISSD
FCEKIIIEVFMDDFSVYGSSFDDYLSNLDRLVLRCKDTNLVLNWEKCHF MVNEGIVLGHKI

>RiceAthila

LHARIYLVYPYSEWVSPVQVVPKKGGMTAVANAQNELIPQRTVTGWRMCIDYRKLNKATKKNHFPLPF
VDEMLERLANHSFFCFLDGYSYHQIPIHPENZSKTTFTCPYGTAYCRMSFGLCNAPASFQRCMMSIFS
DMIEDIMEVFMNDFS VYRKTFGHCLQNLDKVLERYQEKDLVLNWEKCHF MVCEGIVLGHRV

>Medicago

LIPTRTVTGWRMCIDYRRLNTATRKHDFPLPFMDQMIERLAGQAFYCFDLGYSYGNQIAVAPEDQEKIA
FTCPFGVFAYRRMPFGLCGAPATFQRRMLSIFSDMIEKNIKVMDDFSVFGKSFQCLFHLNAVLKRCT
ETNLILNWEKCHF MVTEGIVLGHKIS

>Ad-Athila-16

LEAGIHPISDSPWVSPAQVVPKKGGMTVVHNEKNELVPTTGTGWRMCIDYRRLNIATRKHDFPLPFID
QMLERLAGHDYYCFDLGYSYGNQIAVDPQDQEKTAFTCPSGVFAYRRMPFGLCNAPATFQRCMLSIFS
DMAEKFLEVFMDDFSVYGDSFSSCLNHLVLRKRCQETNLVLNWEKCHF MVTEGIVLGHKI

>Ai-Athila-25

LKAGIYPISDSPWVSPVHVVPKKGGMTVVHNEKNELVPTRTVTGWRMCIDYRRLNTATRKHDFSLQFI
DQMLERLAGHEYYCFDLGYSYGNQIAVDPQDQEKTAFTCRSGVFAYRRMPFGLCNAPATFQRCMLSIF
SDMVEKFLEVFMDDFSVFEDSFISCLNHLALVLRKRCETNLVLNWEKCHF MVTEGIVLGHKI

>Ai-Athila-12

LEAGIYPISDSPWVSPVQVFPKKGGMTVVHNEKNELVPTRTVTGWRMCIDYRRLNTATRKHDFPLPFID
QRLERLAGHEYYCFDLGYSYGNQIAVDPQDQEKIAFTCPSGVFAYRKMPPFGLCNAPATFQRCMLSIFD
MVEKFLEVFMDDFLVFGDSFSSCLDHLVLRKRCQEANLILNWKKCHF MVIEGIVLGHKI

>Ah-Athila-1

LEAGIYPISDSPWVSPVQVVPKKGGMTVVQNEKNELVPTRTVTGWHMCIDYRRLNTATRKHDFPLPFI
DQMLERLAGHDYYYFLDGYSYGNQIAVDPQDQEKTTFTCPSGVFAYRRIPFGLCNTLATFQRCMLSIFS
DMVEKFLEVFMDDFSVYGDSFSTCLDHLKLVLKRCQETNLVLNWEKCHF MVTEGIVLGHKI

>Ah-Athila-10

LEAGIYPISDSPWVSPVQVVPKKGGMTVVYNEKNELVPTRTVTGCRMCIDYRRLNTAIKDHDFPLPFID
QMLERLADHEYYCFDLGYSYGNQIAVDPQDQEKTAFTCPSGVFAYRRMPFGLCNAPATFQRCMLSIFS
DMVENFLEVFMDEFVFGDSFSSCLDHLALVLRKRCQETNLVLNWEKCHF MVTEGIVLGHKI

>Arachis

LEAGIYPISDSPWVSPVQVVPKKGGMTVVHNEKNELVPTRTVTGWRMCIDYRRLNTATRKHDFPLPFI
DQMLERLVGHEYYCFDLGYSYGNQIAVDPDQEKIAFTCPSGVFAYRRMPFGLCNAPVTFQRCMLSIFSD
IVEKFLEVFMDDFSVYGDSFNSCLNHLALVLRKRCQETNLVLNWEKCHF MVTEGIVLGHKI

Seqüência obtida para o retrotransposon até o momento

>Ah-GW-2b.T1 All-reps.0.302

aaggggctcaaaagctcattgatacagttgctagaaatcagcatatgacttaagtagtgagtcctccatgaaagaagaggctaaagtagtatccactgaacttaatcc
tcagaacaggttftgaactcaatcagcaattgctctttataacaaaatagttagcagaatttaaagagatgctccaagacactaaaaaggctaacaagaatatggaa
gcacaattgaatcagacaagacagcaactatctaaacagataacagaagagtgccaagctgttcaaataaagagtggaagacattgaatgtctcaactcaagca
gcagaanaagcaagaaaggaacaactgacagaggatgaccaaccactcccaaaatccctctgaggacagtaagagcccagagaggaaatcttgccattcaa
acgccagaaaagggtaaaaatggcattaaacgcccaaccataccagttatggcgttcaaacgccaggatattagacgccattccaatcctggcgttcaatga
caatgagggatcagacacctgcaagtgtcaataacaacccccctaaagcaggcttctcaaccacttctgtaagaagtaaacctgcagcaactaaggftgaagaatata
aagccaagatgccattatcctcagaaactccgccaagcgaataggataaacaattggccgcttgcagactatctcatgactcttgaataaagattccgfttgcagag
gcacttgagcaataacctcttattctaaagttcatgaaagagatcttaagccataagaaggattggagagaaactgaaaatcttctcactgaagaatgcatgagcag
cattctcaaaagcttaccaaagaagcttaaggatcctggaaagctttacgatacattgacattagaggatgattgaccaaagacagctctatgtgacttgggcaagt
atcaacctaaactgcatccactatcaaaaagcttggttgactgaagaagttaaaccaaccggatagcctcaactgctgatgactcattaaaatccatcaggc
gtaattgaggacatgattgtcaaaagttggccattcgccttccactaactttgtagtgtagaaatggaggagcacaagagtgcaactctcattctaggaagacctt
cctagcactggagcaaccctcattaatgtcaaaaaggggaggtgaccttgagagtaaatggatgagtttaagttgaatgtgtcaaaagatagcagcatccag
acacctcaaaagactgcatgacattgataattgactcttggggagaggtcaaatgactgagagctcaaatcagagctagaggacattttaaagatgttcag
cctgatctggaaagaccagagaaaacagaagaacctcgaagactcctcaggaagaggagaatcctcaaaacccgagctcaaacactaccaccatctctgaaa
tatgcattctgggagaagatgacatttttccggtgacataagctctgctttaaaccactggaaagaggaagcactaattcaagtgctaaaggacacacaagcagctc
tgggtggtccataaagtgaacttaaggcataagcccagctaaatgtacgcacaagatcctattggaggatgagccaagcaggttcaaccacagagggccta
aatccagccatgaaggaggtggtgcagaaaagaggtcactaaactagaggctgggattatctctattctgatagccctgggtgagccctgtccaagttgtcc
ccaagaaggagcatgacagtggtcataatgaaaaaatgaactggtcctacaagaacagttacagggtgctgtagtattgattacagaaggctcaataca
gccaccagaaaaggatcatttcttaccattcatagatcagatgctagaagactagtggtcatgaaatfactgcttcttgatgattcaggtataatcaaatg
cagtagatccttaagaccaagagaaaatgacattcacttggccatctggcgtgttgcatacaggaggatgccattggctctatgtaatgctcctgtaaccttccagagat
gcatgctatctatttctctgatagtgaggaaattctagaagcttcatggatgactctcagtatatggagactcattcaactcctgtcttaaccactagcacttgcct
gaaaaggtccaagagactaacctggttttaactgggagaaatgtcactttatggtagtgaaggaattgctctgggcacaaaatttcaacaagggaatagagg
ggatcaagctaaggtagaggaattgaaaaattaccatcttggcaatgtaaggaatcagaagcttctgggcatgcatgattctatagagggtttataaaagatt
ttcaaaaatgccagacctctgagtaatctgctagctgctgacacaccattatgttggatagagattctgcaggcatttgagactctgaaagctaagctggtcacag
caccagtcactctccaccagactggacattaccattgaaactaatgtgtgatgccagtgatcatgccattggcagtggtgggacagaggcataacaagctctgcat
gtcatttattatgctagccgtgtttaaatgatgcacagaaaaattacacaaccacagaaaaagaattactgcaagtggtttatgccattgacaagtttagatcctatttag
aggatcaaaagtgattgtgtactgacctgacctgctctttaaacttactcacaagcaggattcaaaacccaggtcataagatgggtgttcttctgcaagagttg
atatagaaataagagacagaaaaggacagagaaccaggtagctgatcttgcctcaaatagaaccagtagctggagcgtccctcccttactgagatcctgag
acttccagatgagcaactcttggcattcaggaagctccatggttgcagatattgcaaatataaagatgtgaggttcataccaggagtagcagagagtgcaaaa
gaaagaaatatttcaagatgccaagtattacctatgggatgagccatattctttaaagagatgtgcagatggaatgatccgcagatgtgtaccagagagggaagc
aaaggatcctatggcattgcatgacagatgggggacatttggagtgagcgaacagccactaaggctcctcaatgtggcttctactggcctactctctatag
ggatgccgggagttgtgctgaactgtgacagttgccaagagctggtaacttgcctcatggttacccatgctcacaagggatcttagagattgaattgttggat
gtatgggtattgacttcatgggtccctccgaccatcataatcaaacactataatcttggcagatgactacgtataatgggtagaagcaattgccacacacaat
aatgataccaagaccgtctgaaatcctcagaacacatcttcagcaggttgggtccagagtagtaataca