

**UNIVERSIDADE FEDERAL DE PERNAMBUCO  
CENTRO DE CIÊNCIAS BIOLÓGICAS  
DEPARTAMENTO DE GENÉTICA**

**PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA  
TESE DE DOUTORADO**

**MAPEAMENTO CROMOSSÔMICO  
COMPARATIVO EM ESPÉCIES DE  
*Glycine* WILLD., *Phaseolus* L. E *Vigna*  
SAVI**

**KYRIA CILENE DE ANDRADE BORTOLETI**

**VIRTUS IMPAVIDA**

**RECIFE  
2010**

**UNIVERSIDADE FEDERAL DE PERNAMBUCO  
CENTRO DE CIÊNCIAS BIOLÓGICAS  
DEPARTAMENTO DE GENÉTICA**

**PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA  
TESE DE DOUTORADO**

**MAPEAMENTO CROMOSSÔMICO  
COMPARATIVO EM ESPÉCIES DE *Glycine*  
WILLD., *Phaseolus* L. E *Vigna* SAVI**

**KYRIA CILENE DE ANDRADE BORTOLETI**

Tese apresentada ao **Programa de Pós-Graduação em Genética** da Universidade Federal de Pernambuco como requisito para obtenção do grau de **Doutor em Genética**.

**Orientadora: Prof<sup>ª</sup> Dr<sup>ª</sup> Ana Maria Benko Iseppon**

**Coorientadora: Prof<sup>ª</sup> Dr<sup>ª</sup> Ana Christina Brasileiro Vidal**

**RECIFE  
2010**

**Bortoleti, Kyria Cilene de Andrade**

**Mapeamento cromossômico comparativo em espécies de *Glycine* Willd., *Phaseolus* L. e *Vigna* Savi / Kyria Cilene de Andrade Bortoleti. – Recife: O Autor, 2010.**

**154 folhas : il., fig., tab.**

**Orientadora: Ana Maria Benko Iseppon**

**Co-Orientadora: Ana Christina Brasileiro Vidal**

**Tese (doutorado) – Universidade Federal de Pernambuco.**

**CCB. Genética, 2010.**

**Inclui bibliografia e anexos**

**1. Genética 2. Citogenética 3. Cromossomos 4. Leguminosa  
I. Título.**

**572.8**

**CDD (22.ed.)**

**UFPE/CCB-2011-087**

UNIVERSIDADE FEDERAL DE PERNAMBUCO  
PROGRAMA DE PÓS-GRADUAÇÃO EM  
GENÉTICA E BIOLOGIA MOLECULAR

Parecer da comissão examinadora da tese de:  
KYRIA CILENE DE ANDRADE BORTOLETI

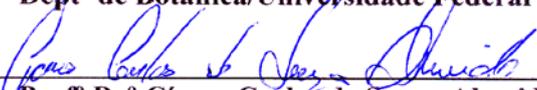
intitulada:

“MAPEAMENTO CROMOSSÔMICO COMPARATIVO EM ESPÉCIES DE *Glycine* WILLD.,  
*Phaseolus* L., E *Vigna* SAVI.”

A comissão examinadora considera o presente trabalho  
APROVADO

Portanto, cumpridas todas as exigências regimentais, KYRIA CILENE DE ANDRADE  
BORTOLETI faz jus ao grau de Doutora em Genética pela UFPE.  
Recife, 16/12/2010

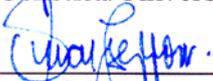
1º Examinador:   
Profª Drª Andrea Pedrosa Harand  
Deptº de Botânica/Universidade Federal de Pernambuco

2º Examinador:   
Profº Drº Cícero Carlos de Souza Almeida  
Campus Arapiraca/ Universidade Federal de Alagoas

3º Examinador:   
Profª Drª Neide Santos  
Deptº de Genética/Universidade Federal de Pernambuco

4º Examinador:   
Profº Drº Tercílio Calsa Júnior  
Deptº de Genética/Universidade Federal de Pernambuco

5º Examinador:   
Profª Drª Ana Maria Benko Iseppon  
Deptº de Genética/Universidade Federal de Pernambuco

Orientadora:   
Profª Drª Ana Maria Benko Iseppon  
Deptº de Genética/Universidade Federal de Pernambuco

Coorientadora:   
Profª Drª Ana Christina Brasileiro Vidal  
Deptº de Genética/Universidade Federal de Pernambuco

Coordenador:   
Profº Drº Antônio Carlos de Freitas  
Deptº de Genética/Universidade Federal de Pernambuco

## AGRADECIMENTOS

A **Deus** e **Meishu-Sama** pelo “*milagre da vida*”, ensinando-me a vivenciar a prática da gratidão, do amor ao próximo e do meu poder de escolha.

Aos meus pais, **Normar Bortoleti** e **Carlos Alberto Bezerra da Silva**, por todo amor e dedicação... “*Sempre ao meu lado, sem precisar explicações!*”

À minha mãe, **Ana Rosa Pereira de Andrade**, exemplo ímpar de fortaleza com uma leve doçura de ser... “*Exemplo de mãe, mulher e amiga!*”

Ao meu irmão, **Klávyo Cléber de Andrade Silva**, por acreditar que eu sou a melhor irmã do mundo... “*Na verdade, você é o melhor irmão do mundo. Obrigada pelo amor, cuidado (sempre segurando minhas mãos), companheirismo e respeito.*”

À **Maria Madalena Correia** (*Minha Dazinha*), **Maria Benedita Correia** (*Minha Ditinha*) e **Terezinha de Jesus Pereira** (Têê), alicerce da minha família; agradeço por todo amor, dedicação e infinitas orações... “*Simplesmente, a minha gratidão e meu amor são imensuráveis!*”

A todos os **meus familiares** pela credibilidade em minha pessoa.

À minha orientadora e amiga, **Ana Maria Benko Iseppon**. Após 10 anos de convivência, cada vez mais, admiro sua ampla atuação científica. Obrigada pela oportunidade, ensinamentos, compreensão e credibilidade em minha pessoa, desde a iniciação científica, auxiliando-me na construção da minha carreira profissional e formação pessoal. “*Tenho o maior orgulho de ser sua orientanda!*”.

À minha segunda orientadora e amiga, **Ana Christina Brasileiro Vidal**. Nos últimos quatro anos, você foi um exemplo de profissionalismo ético inquestionável. Não houve um dia, em que não estivesse presente, mesmo a 750 km de distância. Obrigada pela confiança, paciência, incentivo e amizade. “*Você ensinou-me a ser uma pessoa melhor!*”

Ao **Drº Natoniel Franklin de Melo**, pela acolhida em Petrolina/PE, disponibilizando-me a infraestrutura do Laboratório de Biotecnologia - Embrapa Semiárido (CPATSA), imprescindível para a continuidade deste trabalho. Não obstante à atenção profissional, agradeço pela grande amizade cultivada... *“A você, serei eternamente grata.”*

À **Profª Drª Andrea Pedrosa**, pela parceria iniciada com o projeto, bem como pela atenção prestada e ensinamentos.

À **Profª Neide Santos**, pelos ensinamentos profissionais e pessoais, resultando em uma admiração e verdadeira amizade... *“Nunca esquecerei nossas longas conversas...”*.

A todos que fazem ou fizeram parte da equipe do LGBV, **Ágatha, Alberto, Amanda, Amaro, Ana Félix, Ana Rafaela, Bruno Sérgio, Ana Carolina, Derovil, Diego Sotero, Diego Valério, Geyner, Gilda, Giovani, Hayana, Ivanildo, Karla, Lidiane, Luís Carlos, Neto, Nina, Pollyana, Rodrigo, Santelmo** e, em especial, à **Vanessa** e à **Claudete**... *“Com cada um de vocês, aprendi algo diferente!”*.

Aos colegas do Laboratório de Biotecnologia, **Adriano, Ângela, Arlindo, Eliene, Eiryane, Faubeanny, Ivanice, João, Socorro, Suellen, Teresa e Thiago**, sempre prestativos e acolhedores... *“Vocês me ensinaram a ser um pouco Petrolinense!”*

A **Ebenézer Bernardes (Bené)**, exemplo de *“melhor amigo”*. Nos momentos felizes e/ou árduos, você me ofereceu uma amizade sincera, segurança e paz. *“Ah, se todos no mundo fossem iguais a você...”*.

À **Analice Araújo (Aninha)**, enfatizando *“Sem você é que não tem direção...”*.

A **Santelmo Vasconcelos (Texugo, o mais lindo do laboratório)**, recebendo-me sempre com um abraço apertado e por estar, continuamente, à minha disposição... *“Laço eterno de amizade!”*

A **Emanuelle Vasconcelos (Manu)**, minha filha, amável e companheira espontaneamente... *“Obrigada pela confiança depositada em mim!”*

Às minhas amigas, **Cláudia Lins** e **Patrícia Cavalcanti**, por estarem sempre ao meu lado em momentos de prosperidade e adversidades... *“Amigas para Sempre!”*

A todos **os meus amigos** (e seus familiares) conquistados em **Petrolina/Juazeiro**, *“Vocês acolheram-me e ensinaram-me a viver a vida com perseverança, contudo, sem perder a leveza...”*.

À minha família baiana, **Tia Deíta**, **Ivanice** e **Tarcísio**, que me acolheram como filha e irmã, respectivamente.

À minha amiga **Eliene Matos**, exemplo de autoestima e fé em Deus... *“Com você, estou aprendendo que começar é para todos, continuar para os valentes e concluir para os heróis...”*.

À **Maria do Socorro**, apenas duas semanas de convivência foram necessárias para sermos melhores amigas... *“Incrível à nossa afinidade!”*

À **Universidade Federal do Vale do São Francisco**, em nome do **Magnífico Reitor José Weber Freire Macedo**, especialmente ao corpo docente e técnico do **Colegiado de Ciências Biológicas**... *“Estamos construindo nosso alicerce, Univasfianos!”*

À coordenação do **Programa de Pós-Graduação em Genética**, em nome da pessoa **Prof.º Antônio Carlos de Freitas**, bem como ao corpo docente, por terem contribuído para minha formação.

Aos **funcionários do Departamento de Genética**, em especial, à D<sup>a</sup> Zizi pela sabedoria repassada em todos os momentos de descontração.

À **CAPES** e ao **CNPq** pelo apoio financeiro oferecido para o desenvolvimento deste projeto.

A todos aqueles que, de alguma forma, auxiliaram-me na concretização de mais um sonho...

***Muito Obrigada!!!***

*A Deus;*

*Às “Ana’s”,*

*À minha mãe, Ana Rosa, meu porto seguro...*

*Às minhas “orientadoras”, Ana Benko e Ana Christina, por estarem sempre ao meu lado, fazendo-me acreditar que a mais tênue luz se sobressai em qualquer escuridão....*

Dedico.

*“... minha força e vitória tem um nome, Jesus...”.*

***Eliana Ribeiro***

*“Não importa o que o passado fez de mim. Importa o que farei com o que o passado fez de mim.”*

***Jean Paul Sartre***

## SUMÁRIO

	<b>Página</b>
<b>Lista de Abreviaturas</b> .....	i
<b>Lista de Figuras</b> .....	iv
<b>Lista de Tabelas</b> .....	vi
<b>Resumo</b> .....	vii
<b>Abstract</b> .....	viii
<b>1. Introdução</b> .....	1
<b>2. Objetivos</b> .....	4
<b>3. Revisão de Literatura</b> .....	6
3.1 Citogenética molecular.....	7
3.2 Sequências de DNA repetitivo.....	8
3.3 Classes de DNA repetitivo.....	9
3.3.1 Sequências repetitivas em tandem.....	9
3.3.1.1 DNA ribossômico.....	9
3.3.1.2 Microssatélites.....	10
3.3.2 Sequências repetitivas dispersas.....	14
3.4 Características taxonômicas e importância econômica da família Fabaceae Lindley.....	18
3.5 Características taxonômicas e citogenéticas do gênero <i>Glycine</i> Willd.....	20
3.6 Características taxonômicas e citogenéticas do gênero <i>Phaseolus</i> L.....	23
3.7 Características taxonômicas e citogenéticas do gênero <i>Vigna</i> SAVI.....	25
<b>4. Referências Bibliográficas</b> .....	29
<b>5. Capítulo I – Microssatélites: um importante constituinte do genoma de     leguminosas</b> .....	42
<b>6. Capítulo II - Ocorrência e distribuição dos retrotransposons <i>Ty1</i>-copia-like     e <i>Ty3</i>-gypsy-like em representantes dos gêneros <i>Glycine</i> Willd., <i>Phaseolus</i> L.     e <i>Vigna Savi</i></b> .....	77

<b>7. Conclusões</b> .....	108
<b>8. Anexos</b> .....	111
8.1 Informação Suplementar – Capítulo I.....	112
8.2 Instruções para Autores – <i>Annals of Botany</i> .....	133
8.3 Instruções para Autores – <i>Chromosome Research</i> .....	145
8.4 Resumo – 56º Congresso Brasileiro de Genética.....	151
8.5 Resumo – 56º Congresso Brasileiro de Genética.....	152
<b>9. Memorial da aluna</b> .....	153

## LISTA DE ABREVIATURAS

(AG) <sub>8,12</sub>	Sequência de oligonucleotídeos
(AT) <sub>8</sub>	Sequência de oligonucleotídeos
(AC) <sub>8, 12</sub>	Sequência de oligonucleotídeos
(AAC) <sub>5</sub>	Sequência de oligonucleotídeos
(AAG) <sub>5</sub>	Sequência de oligonucleotídeos
(ACC) <sub>5</sub>	Sequência de oligonucleotídeos
(GAA) <sub>7, 10</sub>	Sequência de oligonucleotídeos
(CTC) <sub>5</sub>	Sequência de oligonucleotídeos
(TGA) <sub>6</sub>	Sequência de oligonucleotídeos
(GATA) <sub>5, 7</sub>	Sequência de oligonucleotídeos
(GACA) <sub>4, 7</sub>	Sequência de oligonucleotídeos
APG	<i>The Angiosperm Phylogeny Group</i> O Grupo de Filogenia das Angiospermas
AT	Adenina/Timina
BAC	<i>Bacterial Artificial Chromosome</i> Cromossomo Artificial de Bactéria
BLAST	<i>Basic Local Alignment Search Tool</i>
BSA	<i>Bovine Serum Albumine</i> Albumina de Soro Bovino
CCB	Centro de Ciências Biológicas
CentGm-1	Sequência satélite centromérica de <i>Glycine max</i>
CentGm-2	Sequência satélite centromérica de <i>Glycine max</i>
CMA <sub>3</sub>	Cromomicina A <sub>3</sub>
CONAB	Companhia Nacional de Abastecimento
CPATSA	Centro de Pesquisa do Trópico Semi-Árido
DAF	<i>DNA Amplification Fingerprinting</i> Impressão Digital da Amplificação do DNA
DAPI	4',6-diamidino-2-fenilindol
DNA	Desoxiribonucleic Acid Ácido Desoxirribonucleico
DNAc	DNA complementar
DNAr	DNA ribossomal
DNArn	DNA ribossomal nuclear
DNAr 45S	DNA ribossomal 45S
DNAr 5S	DNA ribossomal 5S
DNAr 18S	Segmento 18S da Unidade de DNA ribossomal 45S
DNAr 5,8S	Segmento 5,8S da Unidade de DNA ribossomal 45S
DNAr 26S	Segmento 26S da Unidade de DNA ribossomal 45S
Domínio RT	Domínio da Transcriptase Reversa
<i>DraI</i>	Família de DNA repetitivo
Embrapa	Empresa Brasileira de Pesquisa Agropecuária
env	Envelope
EST	<i>Expressed Sequence Tags</i>

	Etiquetas de Sequências Expressas
ETS	<i>External Transcribed Spacer</i> Espaceador Transcrito Externo
FAO	<i>Food and Agriculture Organization</i> Organização para Agricultura e Alimentação
FISH	<i>Fluorescent In Situ Hybridization</i> Hibridização <i>In Situ</i> Fluorescente
FITC	Fluoresceína Isotiocianato
GC	Guanina/Citosina
<i>gag</i>	<i>Group-specific antigen</i> Antígeno grupo-específico
Gb	Giga bases
GISH	<i>Genomic In Situ Hybridization</i> Hibridização Genômica <i>In Situ</i>
H <sub>3</sub>	Proteína histônica 3
H <sub>4</sub>	Proteína histônica 4
HC	Heterocromatina Constitutiva
HIS	<i>In Situ Hybridization</i> Hibridização <i>In Situ</i>
ICTV	<i>International Committee on taxonomy of Viruses</i> Comitê Internacional em Taxonomia de Vírus
IGS	<i>Internal Gene Spacer</i> Espaceador Gênico Interno
ITS	<i>Internal Transcribed Spacer</i> Espaceador Transcrito Interno
kb	Kilo bases
LINE	<i>Long Interspersed Nucleotide Elements</i> Elementos Nucleares Interdispersos Longos
LGBV	Laboratório de Genética e Biotecnologia Vegetal
LTR	<i>Long Terminal Repeats</i> Longas Repetições Terminais
Mb	Mega bases
ORF	<i>Open Reading Frame</i> Quadro Aberto de Leitura
pb	Pares de bases
<i>pol</i>	<i>Polyprotein</i> <i>Poliproteína</i>
pHaS13	Sequências repetitivas
pHaS211	Sequências repetitivas
RNA	Ribonucleic Acid Ácido Ribonucleico
RNAi	RNA de interferência
RNAr	RNA ribossomal
RONs	Regiões Organizadoras de Nucléolos
RT	<i>Reverse Transcriptase</i> Transcriptase Reversa
RTvr-1	Retroelementos isolados do genoma de <i>V. radiata</i>

RTvr-2	Retroelementos isolados do genoma de <i>V. radiata</i>
SINEs	<i>Short Interspersed Nucleotide Elements</i> Elementos Nucleares Interdispersos Curtos
SNPs	<i>Single Nucleotide Polymorphism</i> Polimorfismos de um Único Nucleotídeo
sp	Espécie
ssp	Subespécie
SSRs	<i>Simple Sequence Repeats</i> Repetições de Sequências Únicas
TE	<i>Transposable Elements</i> Elementos Transponíveis
Tnt1-94	Retroelemento encontrado no genoma de <i>Nicotiana tabacum</i>
TRITC	Tetramethyl Rhodamine Iso-Thiocyanate
TSD	<i>Target Site Duplication</i> Duplicação do Sítio Alvo
Ty1- <i>copia</i>	Superfamília de retrotransposon LTR
Ty3- <i>gypsy</i>	Superfamília de retrotransposon LTR
UFPE	Universidade Federal de Pernambuco
VLP	<i>Virus-like Particle</i> Partícula Similar a Vírus

## LISTA DE FIGURAS

<b>Figura</b>	<b>Página</b>
 <i>Revisão de Literatura</i>	
<b>Figura 1.</b> Classificação dos microssatélites de acordo com a estrutura das unidades de repetições.....	11
<b>Figura 2.</b> Organização dos domínios gênicos das principais classes de retrotransposons LTRs (Ty1- <i>copia</i> e Ty3- <i>gypsy</i> ) (adaptado de Todorovska, 2007).....	17
 <i>Capítulo I</i>	
<b>Figura 1.</b> Padrão CMA <sub>3</sub> (A, D, G)/ DAPI (B, E, H) e distribuição dos sítios de DNAr 5S e 45S (C, F, I) em <i>Glycine max</i> (A-C), <i>Vigna unguiculata</i> (D-F, J) e <i>V. radiata</i> (G-I, J).....	54
<b>Figura 2.</b> Células metafásicas de <i>Glycine max</i> (A-B), <i>Phaseolus vulgaris</i> (C-D), <i>P. lunatus</i> (E-F), <i>Vigna unguiculata</i> (G-H) e <i>V. radiata</i> (I-J) mostrando a distribuição dos sítios do oligonucleotídeo com padrão SSR (AG) <sub>8</sub> (em verde) e DNAr 45S (pseudocoloridos em amarelo).....	56
<b>Figura 3.</b> Células metafásicas de <i>Glycine max</i> (A-B), <i>Phaseolus vulgaris</i> (C-D), <i>P. lunatus</i> (E-F), <i>Vigna unguiculata</i> (G-H) e <i>V. radiata</i> (I-J) mostrando a distribuição dos sítios do oligonucleotídeo com padrão SSR (AAG) <sub>5</sub> (em verde) e DNAr 45S (pseudocoloridos em amarelo).....	57
<b>Figura 4.</b> Metáfases de <i>Glycine max</i> (A-B), <i>Phaseolus vulgaris</i> (C-D), <i>P. lunatus</i> (E-F), <i>Vigna unguiculata</i> (G-H) e <i>V. radiata</i> (I-J) mostrando a distribuição dos sítios do oligonucleotídeo com padrão SSR (ACC) <sub>5</sub> (em verde) e DNAr 45S (pseudocoloridos em amarelo).....	58
<b>Figura 5.</b> Células metafásicas de <i>Glycine max</i> (A-B), <i>Phaseolus vulgaris</i> (C-D), <i>P. lunatus</i> (E-F), <i>Vigna unguiculata</i> (G-H) e <i>V. radiata</i> (I-J) mostrando a distribuição dos sítios do oligonucleotídeo com padrão SSR (CTC) <sub>5</sub> (em verde) e DNAr 45S (pseudocoloridos em amarelo).....	59
<b>Figura 6.</b> Células metafásicas de <i>Glycine max</i> (A-B), <i>Phaseolus vulgaris</i> (C-D), <i>P. lunatus</i> (E-F), <i>Vigna unguiculata</i> (G-H) e <i>V. radiata</i> (I-J) mostrando a distribuição dos sítios do oligonucleotídeo com padrão SSR (TGA) <sub>6</sub> (em verde) e DNAr 45S (pseudocoloridos em amarelo).....	61
<b>Figura 7.</b> Coeficiente de correlação de Pearson (r) entre a presença de microssatélites, genes e elementos transponíveis em cromossomos ou pseudomoléculas cromossômicas em <i>Glycine max</i> .....	64

**Figura 8.** Representação da distribuição *in silico* de [(AG)<sub>8</sub>]<sub>15</sub>, [(AAG)<sub>5</sub>]<sub>10</sub>, [(ACC)<sub>5</sub>]<sub>10</sub>, [(CTC)<sub>5</sub>]<sub>10</sub> e [(TGA)<sub>6</sub>]<sub>10</sub> em cada cromossomo ou pseudomolécula cromossômica de *Glycine max*, bem como da região centromérica e pericentromérica, sendo a primeira determinada pela disposição das sequências CentGm1 e CentGm-2 (<http://soybase.org/>)..... 66

**Figura 9.** Comparação dos mapas físico e cromossômico para o cromossomo 13 de *Glycine max*, evidenciando a localização pericentromérica da banda CMA<sub>3</sub><sup>+</sup> e DNAr 45S, bem como, a distribuição dos sítios de oligonucleotídeos associados a regiões eucromáticas e heterocromáticas..... 67

## Capítulo II

**Figura 1.** Amplificação por PCR do domínio RT do Ty1-*copia*-like (A) e Ty3-*gypsy*-like (B), a partir do DNA genômico de *Vigna unguiculata*, utilizando *primers* degenerados..... 88

**Figura 2.** Alinhamento múltiplo por Clustal W das sequências do domínio RT do retroelemento Ty1-*copia*-like, destacando-se a presença da sequência consenso SLYGLKQ (domínio laranja) e LYVDDM (domínio roxo), características dos retrotransposons do tipo *copia*..... 91

**Figura 3.** Alinhamento múltiplo por Clustal W das sequências do domínio RT do retroelemento Ty3-*gypsy*-like, destacando-se a presença da sequência consenso MCVDYR (domínio laranja) e MPFGV (domínio roxo), características dos retrotransposons do tipo *gypsy*..... 93

**Figura 4.** Árvore filogenética construída por máxima parcimônia utilizando sequências de nucleotídeos dos domínios RT dos retrotransposons Ty1-*copia*-like isolada de *Vigna unguiculata* (laranja) e depositadas no *GenBank* para outras espécies vegetais..... 94

**Figura 5.** Árvore filogenética construída por máxima parcimônia utilizando sequências de nucleotídeos dos domínios RT dos retrotransposons Ty3-*gypsy*-like isolada de *Vigna unguiculata* (laranja) e depositadas no *GenBank* para outras espécies vegetais..... 94

**Figura 6.** Células metafásicas de *Glycine max* (A-B), *Phaseolus vulgaris* (C-D), *P. lunatus* (E-F) *Vigna unguiculata* (G-H) e *V. radiata* (I-J) mostrando a distribuição dos sítios da sequência RT do retrotransposon Ty1-*copia*-like (em verde) e DNAr 45S (pseudocoloridos em amarelo)..... 97

**Figura 7.** Células metafásicas de *Glycine max* (A-B), *Phaseolus vulgaris* (C-D), *P. lunatus* (E-F) *Vigna unguiculata* (G-H) e *V. radiata* (I-J) mostrando a distribuição dos sítios da sequência RT do retrotransposon Ty3-*gypsy*-like (em verde) e DNAr 45S (pseudocoloridos em amarelo)..... 98

**LISTA DE TABELAS**

<b>Tabela</b>	<b>Página</b>
<b>Capítulo I</b>	
<b>Tabela 1.</b> Mapeamento <i>in silico</i> de oligonucleotídeos com padrão de SSR no genoma de <i>Glycine max</i> , indicando o número de sítios observados, o tamanho mínimo e máximo da sequência (em pb), bem como a densidade gênica da região e o número de genes e/ou elementos transponíveis associados.....	60
<b>Tabela 2.</b> Distribuição de classes de repetições dos oligonucleotídeos com padrão microssatélite mapeados nos cromossomos ou pseudomoléculas cromossômicas de <i>Glycine max</i> .....	65
<b>Capítulo II</b>	
<b>Tabela 1.</b> Identificação das sequências depositadas no <i>GenBank</i> que apresentaram similaridade com a sequência do domínio RT do retrotransposon Ty1- <i>copia</i> -like isolada do genoma de <i>Vigna unguiculata</i> , com seus respectivos números do acesso, descrição, cobertura (%), identidade (%) e tamanho (em pb).....	89
<b>Tabela 2.</b> Identificação das sequências depositadas no <i>GenBank</i> que apresentaram similaridade com a sequência do domínio RT do retrotransposon Ty3- <i>gypsy</i> -like isolada do genoma de <i>Vigna unguiculata</i> , com seus respectivos números do acesso, descrição, cobertura (%), identidade (%) e tamanho (em pb).....	90

## RESUMO

Um estudo genômico comparativo foi realizado mediante localização *in situ* de oligonucleotídeos com padrão de microssatélites [(AG)<sub>8</sub>, (AAG)<sub>5</sub>, (ACC)<sub>5</sub>, (CTC)<sub>5</sub> e (TGA)<sub>6</sub>] e dos domínios RT dos retrotransposons Ty1-*copia*-like e Ty3-*gypsy*-like, ao longo dos cromossomos de *Glycine max*, *Phaseolus vulgaris*, *P. lunatus*, *Vigna unguiculata* e *V. radiata*. Os oligonucleotídeos apresentaram-se em grande proporção nos genomas analisados, porém houve variação em quantidade, organização e distribuição ao longo dos cromossomos. Na análise genômica de soja, observou-se números e tamanhos (30 a 454 pb) de sítios de repetições variáveis localizados, principalmente, em regiões de alta a moderada densidade gênica, por vezes associados a genes e elementos transponíveis. A associação da FISH e análise *in silico* ressaltou uma distribuição diferencial não aleatória dos oligonucleotídeos, podendo estar preferencialmente associados à heterocromatina ou à eucromatina. Os domínios RT de Ty1-*copia*-like e Ty3-*gypsy*-like foram amplificados em *V. unguiculata*, sendo a sequência de Ty1-*copia*-like menos homogênea. As análises filogenéticas enfatizaram a origem monofilética dos referidos domínios RT. A FISH evidenciou a presença de sinais dispersos e pericentroméricos, utilizando ambos retroelementos, com algumas divergências interespecíficas. Em algumas espécies, tais marcações estavam associadas a DNAr 5S e 45S, bem como localizadas em regiões heterocromáticas, enfatizando uma distribuição preferencial nos genomas estudados. Os microssatélites e retrotransposons são apontados como importantes componentes dos genomas em espécies do clado Phaseoloids (tribo Phaseoleae), propiciando uma discussão sobre o potencial estrutural e funcional dessas sequências repetitivas.

**Palavras-chave:** Leguminosas, FISH, microssatélites e retroelementos LTR.

## ABSTRACT

A comparative genomic study was developed by physical mapping of microsatellite sequences [(AG)<sub>8</sub>, (AAG)<sub>5</sub>, (ACC)<sub>5</sub>, (CTC)<sub>5</sub> and (TGA)<sub>6</sub>] and reverse transcriptase (RT) domains of the retrotransposons Ty1-*copia*-like and Ty3-*gypsy*-like, along the *Glycine max*, *Phaseolus vulgaris*, *P. lunatus*, *Vigna unguiculata* and *V. radiata* genomes. Microsatellite sequences were highly distributed within the genomes of the analyzed species. However, there were variations in their quantities, and organization along the genomes. In genomic of soybean investigation, the visualization of a variable number of oligonucleotides repetitions sites with sizes ranging from 30 to 454 bp, located, mainly, in regions of high to moderate gene density, sometimes associated to genes and transposable elements. FISH and *in silico* analysis highlighted a differential and non-analogue distribution of different classes of the analyzed SSRs, which may be preferentially associated to heterochromatin or euchromatin. Additionally, the conserved domain of RT genes of Ty1-*copia*-like and Ty3-*gypsy*-like were amplified from the *V. unguiculata* genome, exhibiting significant heterogeneity levels, and the Ty1-*copia*-like sequences were less homogeneous in comparison to the former. Phylogenetic analysis of RT sequences emphasized the monophyletic origin of retrotransposons sequences in plant genomes. FISH analysis has shown the presence of dispersed and pericentromeric signals, using both retroelements, with some interspecific divergences. Such markers were associated, in some cases, to 5S and 45S rDNA loci, and located in heterochromatic regions, emphasizing a preferential distribution in the analyzed genomes. The microsatellites and retrotransposons were very important components for the phaseoloids (tribe Phaseoleae) clade evolution, leading to a discussion about the structural and functioning potential of these repetitive DNA sequences.

**Key-words:** Legume, FISH, Microsatellites and LTR Retroelements.

## ***Introdução***

---

---

## 1. INTRODUÇÃO

A família Fabaceae é a terceira maior família das Magnoliophyta, compreendendo 727 gêneros e aproximadamente 19.000 espécies (Lewis *et al.*, 2005). Apresenta grande diversidade, sendo subdividida em três subfamílias Papilionoideae, Mimosoideae e Caesalpinoideae (Judd *et al.*, 2002), distribuídas principalmente nas regiões tropicais e subtropicais do planeta (Joly, 2002). Economicamente, a subfamília Papilionoideae abrange a maioria das leguminosas cultivadas, como alfafa, amendoim, ervilha, feijão comum, feijão-caupi, grão-de-bico e soja, representando cerca de 27% da produção de plantas cultivadas no mundo (Graham e Vance, 2003).

Os gêneros *Glycine* Willd., *Phaseolus* L. e *Vigna* Savi pertencem à tribo Phaseoleae, estando agrupados em um clado monofilético (Phaseoloids) (Marechal *et al.*, 1978). Representantes desses gêneros apresentam grande importância econômica, destacando-se as espécies *Glycine max* (L.) Merr., *Phaseolus lunatus* (L.), *Phaseolus vulgaris* (L.) e *Vigna unguiculata* (L.) Walp., conhecidas popularmente como soja, feijão-fava, feijão comum e feijão-caupi, respectivamente.

Os grãos de feijão representam uma importante fonte proteica na dieta humana nos países em desenvolvimento das regiões tropicais e subtropicais. No Brasil, o feijão é um dos componentes básicos da dieta alimentar da população, destacando-se como importante fonte de proteína para as classes economicamente menos favorecidas (Wander, 2005). Por sua vez, a soja, além da importância proteica, é uma das principais fontes de óleo do mundo, tendo 85% do seu genoma sequenciado (Schmutz *et al.*, 2010).

As famílias de sequências repetitivas têm sido apontadas como o principal componente dos genomas eucarióticos, constituindo cerca de 70-90% em plantas superiores (Schmidt e Heslop-Harrison, 1998; Dimitri *et al.*, 2009), sendo citadas como fator responsável pela expansão e evolução dos genomas vegetais. Consideradas como elemento crítico para o entendimento da estrutura cromossômica e organização genômica, suas funções biológicas têm sido extensivamente discutidas. Por muito tempo, elas foram remetidas como *junk* DNA (DNA lixo) devido à limitada densidade gênica e baixo índice de recombinação (Jong *et al.*, 1999). Por outro lado, existem relatos afirmando que famílias de sequências repetitivas estão envolvidas na organização estrutural cromossômica e genômica, na expressão e recombinação gênica, assim como em diferentes aspectos do funcionamento celular (Henikoff, 2000; Dimitri *et al.*, 2009), tornando o conceito anterior obsoleto.

Tais particularidades, principalmente sua composição altamente repetitiva, tornam-nas um componente genômico com capacidade versátil de ser modificado, acumulado ou

eventualmente eliminado, em um intervalo de tempo evolucionário relativamente rápido, em comparação às sequências de cópia única (Winterfeld e Roser, 2007). Conseqüentemente, essas sequências podem apresentar uma distribuição genoma-específica, espécie-específica, bem como cromossomo-específica (Begum *et al.*, 2009).

As características peculiares das classes de DNA repetitivo, como DNAr (DNA ribossomal), microssatélites e retrotransposons, têm facilitado a sua utilização, mediante FISH (*Fluorescent In Situ Hybridization*; Hibridização *In Situ* Fluorescente), em estudos de relações evolutivas (Cuadrado e Jouve, 2007a; Moraes *et al.*, 2007; Nielen *et al.*, 2009), identificação cromossômica (Brasileiro-Vidal *et al.*, 2005; Chang *et al.*, 2008; Gaeta *et al.*, 2010), construção de mapas cromossômicos (Pedrosa *et al.*, 2002; Ohmido *et al.*, 2007) e na integração entre mapas genéticos e cromossômicos permitindo a elucidação da organização genômica entre espécies proximamente relacionadas.

Adicionalmente, o sequenciamento dos genomas tem permitido a investigação sobre a abundância, tipo e frequência de elementos repetitivos, cujas localizações permanecem como “lacunas”, devido à dificuldade em determinar sua correta posição, arranjo e número de cópias no genoma (Schlueter *et al.*, 2008; Schmutz *et al.*, 2010). Mais recentemente, a integração entre os mapas genéticos, cromossômicos e de sequenciamento tem fornecido um maior número de marcas e de informações sobre a organização e evolução dos genomas, facilitando um melhor entendimento da homeologia e conservação da sintenia entre as espécies (Findley *et al.*, 2010).

Considerando a existência de poucos trabalhos que caracterizem a organização cromossômica de sequências repetitivas em leguminosas cultivadas e a extensiva conservação genômica descrita para a subfamília Papilionoideae (Choi *et al.*, 2004; Mudge *et al.*, 2005; Zhu *et al.*, 2005), o presente trabalho visou à análise do padrão de bandeamento com os fluorocromos CMA<sub>3</sub>/DAPI e a distribuição cromossômica dos sítios de DNAr, de oligonucleotídeos com padrão de microssatélites e de retroelementos em algumas leguminosas mediante FISH, auxiliando a compreensão das propriedades evolutivas destes elementos repetitivos, permitindo inferências sobre as possíveis funções exercidas por tais sequências na organização e evolução genômica do grupo estudado.

## ***Objetivos***

---

## **2. OBJETIVOS**

### **2.1 OBJETIVO GERAL**

Realizar um estudo genômico comparativo entre os gêneros *Glycine*, *Phaseolus* e *Vigna* mediante a localização cromossômica de sequências de DNA repetitivo, a fim de construir mapas citogenéticos para *G. max*, *P. vulgaris*, *P. lunatus*, *V. unguiculata* e *V. radiata*, colaborando para o entendimento da organização e evolução genômica desses elementos repetitivos nas referidas espécies.

### **2.2 OBJETIVOS ESPECÍFICOS**

1. Elucidar a frequência e distribuição de cinco sequências com padrão de microssatélites, bem como dos retroelementos Ty1-*copia*-like e Ty3-*gypsy*-like, em espécies dos gêneros *Glycine*, *Phaseolus* e *Vigna*.

2. Comparar diferentes espécies dos gêneros *Glycine*, *Phaseolus* e *Vigna*, com base no padrão de sinais obtidos através da técnica de FISH, utilizando oligonucleotídeos com padrão de microssatélites e retrotransposons como sondas.

3. Associar, aos sinais de hibridização dos microssatélites e retrotransposons, a localização *in situ* de sondas de DNAr 5S e 45S e o padrão de bandas CMA<sub>3</sub>/DAPI, como marcas cromossômicas, a fim de auxiliar na identificação dos diferentes tipos cromossômicos.

## ***Revisão de Literatura***

---

### **3. REVISÃO DE LITERATURA**

#### **3.1 CITOGENÉTICA MOLECULAR**

Estudos citogenéticos têm revelado polimorfismos genéticos importantes na caracterização de espécies vegetais, através de marcadores citológicos que auxiliem na construção de mapas cromossômicos. Tais estudos têm contribuído para a caracterização de germoplasma, programas de melhoramento genético, entendimento das relações filogenéticas e identificação da origem de várias espécies (Benko-Iseppon, 2001; Pedrosa-Harand *et al.*, 2006; Moraes *et al.*, 2007; Vasconcelos *et al.*, 2010).

Métodos citomoleculares fornecem um grupo poderoso de ferramentas para análises da estrutura, organização e evolução genômica. A técnica de FISH tem permitido a localização de regiões cromossômicas específicas nas mais variadas espécies silvestres e cultivadas. O processo consiste basicamente na desnaturação e hibridização do ácido nucleico alvo com uma sonda adequada (DNA ou RNA conhecido e previamente marcado), possibilitando a localização *in situ* de sequências de ácidos nucleicos (DNA ou RNA) nos cromossomos, citoplasma, organelas ou tecidos (Guerra, 2004). Na hibridização *in situ* de sequências de DNA, muitos tipos de sequências têm sido visualizados através da FISH, incluindo cópias únicas ou com baixo número de repetições, até aquelas altamente repetitivas (Guerra, 2004).

Nos últimos anos, a FISH tem ampliado consideravelmente o nível de análise e o campo de atuação da citogenética vegetal. Esta técnica permite que várias sondas possam ser simultaneamente detectadas em diferentes cromossomos, possibilitando que seja determinada a ordem física de sequências e, por consequência, a construção de mapas cromossômicos. Em cromossomos mitóticos vegetais, calcula-se que duas sequências tenham que estar separadas por pelo menos 2 a 5 Mb para serem visualizadas como sinais distintos (Pedrosa-Harand e Guerra, 2004). Por outro lado, poucos resultados reprodutíveis têm sido relatados em plantas para alvos menores de 10 Kb (Jong *et al.*, 1999).

A construção de mapas cromossômicos tem sido importante para compreender a organização cromossômica e a evolução cariotípica, auxiliando na detecção de alterações cromossômicas estruturais, na identificação da posição linear de sequências, bem como no isolamento de genes mapeados (Pedrosa-Harand e Guerra, 2004). Adicionalmente, análises comparativas de mapas cromossômicos têm sido usadas para

estudar o nível de conservação da ordem dos genes em diferentes grupos taxonômicos (Mandaková e Lysak, 2008; Fonsêca *et al.*, 2010).

### **3.2 SEQUÊNCIAS DE DNA REPETITIVO**

Os elementos de DNA repetitivo consistem em sequências motivos com um tamanho variável de 2 a 10.000 pb, cujo número de cópias pode alternar de centenas a milhares de centenas, enfatizando sua grande representatividade genômica. Tais sequências têm sido apontadas como o principal componente dos genomas eucarióticos, constituindo cerca de 70-90% em plantas superiores (Schmidt e Heslop-Harrison, 1998; Dimitri *et al.*, 2009), sendo amplamente responsável pela variação no tamanho dos genomas (Jiang *et al.*, 2003).

As sequências repetitivas são diferenciadas pelo seu nível de homologia, distribuição entre as espécies e/ou organização física e genômica. Classicamente, suas unidades de repetições podem estar distribuídas em tandem ou dispersas ao longo do genoma. As sequências em tandem ocorrem em blocos de centenas a milhares de cópias, localizadas em um ou mais sítios de um dado genoma. Essa categoria inclui DNA codificante, como genes para RNAr (RNA ribossomal) e não codificante, como classes de DNA satélite (Schmidt e Heslop-Harrison, 1998). Por sua vez, o DNA repetitivo disperso é constituído por elementos de DNA localizados interdispersamente com outras sequências no genoma, como os elementos transponíveis, estando distribuídos ao longo dos cromossomos, embora regiões de deleção e amplificação possam ser encontradas (Kubis *et al.*, 1998).

As diferentes classes de DNA repetitivo refletem a grande importância evolutiva que estes elementos desempenham na organização genômica, cujas funções biológicas exercidas têm sido extensivamente discutidas. Por muito tempo, eles foram referidos como DNA lixo devido ao baixo índice de recombinação e à menor densidade gênica dessas regiões (Jong *et al.*, 1999). Por outro lado, Kubis *et al.* (1998) afirmaram que famílias de sequências repetitivas estão envolvidas na estabilização dos cromossomos, enquanto Winterfeld e Roser (2007) ressaltam a sua importância na expressão e recombinação gênica, assim como na organização estrutural do genoma. Conjuntamente, existem várias evidências que sugerem a influência desempenhada pelas sequências repetitivas em diferentes aspectos do funcionamento celular (Dimitri *et al.*, 2009; Gartenberg, 2009), levando o conceito anterior ao desuso.

### **3.3 CLASSES DE DNA REPETITIVO**

#### **3.3.1 SEQUÊNCIAS REPETITIVAS EM TANDEM**

As sequências repetidas em tandem referem-se a curtas unidades de repetições arranjadas em série, sendo distribuídas preferencialmente em posições pericentroméricas, subteloméricas, teloméricas e, mais raramente, nas regiões intercalares (Kubis *et al.*, 1998).

Geralmente, o DNA satélite encontra-se localizado em associação com a heterocromatina, sugerindo-se que estas sequências exerçam funções importantes na organização cromossômica em telômeros e centrômeros, onde auxilia na coesão entre as cromátides-irmãs e desempenha papel acessório na formação e/ou estabilidade do cinetócoro, atuando também na remodelagem da cromatina e na regulação da expressão gênica (Dimitri *et al.*, 2009; Gartenberg, 2009), conjuntamente, com outras sequências de DNA repetitivo constituintes desta forma de cromatina.

A heterocromatina apresenta características bem peculiares, como elevado nível de condensação, baixa densidade gênica e recombinação meiótica reduzida, replicação tardia durante a fase S, inativação transcricional, composição de DNA satélites altamente repetitivos e remanescentes de elementos transponíveis (Dimitri *et al.*, 2009; Gartenberg, 2009). Adicionalmente, encontra-se distribuída em blocos, devido a modificações pós-transcricionais nos nucleossomos, como metilação de H<sub>3</sub> e acetilação de H<sub>3</sub> e H<sub>4</sub> (Henikoff, 2000; Elgin e Grewal, 2005).

As particularidades supracitadas, principalmente sua composição altamente repetitiva, tornam-na um componente genômico com capacidade versátil de ser modificado, acumulado ou eventualmente eliminado, em um intervalo de tempo evolutivo relativamente rápido, em comparação com a eucromatina (Winterfeld e Roser, 2007). Consequentemente, as sequências de DNA não codificantes mostram níveis variáveis de abundância e de padrão de distribuição entre espécies proximamente relacionadas. Elas podem ser espécie-específicas, genoma-específicas e cromossomo-específicas, sendo consideradas úteis para estudos de questionamento taxonômico e relações filogenéticas entre espécies vegetais (Kubis *et al.*, 1998; Tóth *et al.*, 2000).

##### **3.3.1.1 DNA RIBOSSÔMICO**

Os genes de RNAr 5S e 45S apresentam-se altamente repetitivos e arranjados em tandem em sítios (loci) no genoma de diferentes espécies. Tal distribuição e o

número de sítios discretos propiciam sua extensa utilização na FISH (Pedrosa-Harand e Guerra, 2004), tornando os loci de DNAr marcadores valiosos para investigação da organização e evolução cromossômica.

O loco do DNAr 45S consiste de unidades de repetição com cerca de 9,2 kb, contendo os genes para DNAr 18S, 5,8S e 26S, espaçadores do tipo ETS (*External Transcribed Spacer*; Espaçador Transcrito Externo) e ITS (*Internal Transcribed Spacer*; Espaçador Transcrito Interno), bem como as regiões IGS (*Internal Gene Spacer*; Espaçador Gênico Interno) entre as unidades transcritas (Weider *et al.*, 2005). Por outro lado, o loco do DNAr 5S compreende uma sequência menor e relativamente menos conservada (Heslop-Harrison, 2000).

Os sítios de DNAr 45S correspondem às RONS (Regiões Organizadoras de Nucléolo), visíveis em mitose como constrições secundárias (Guerra, 2004). Frequentemente, estes loci encontram-se colocalizados com blocos de HC (heterocromatina constitutiva) CMA<sub>3</sub><sup>+</sup>, como relatado para *Crotalaria juncea* L. (Mondin *et al.*, 2007), *P. vulgaris* e *P. lunatus* (Almeida, 2006), *Ricinus communis* L. (Vasconcelos *et al.*, 2010), entre outras. Possivelmente, essa constituição preferencial está relacionada à composição dos genes ribossomais, bem como de seus espaçadores intercalares (Winterfeld e Roser, 2007). Segundo Eickbush e Eickbush (2007), os genes de RNAr são altamente conservados, explicando a uniformidade das sequências que compõem os arranjos de DNAr em cada organismo pela evolução em concerto, propiciando uma distribuição diferenciada em nível intra e interespecífico.

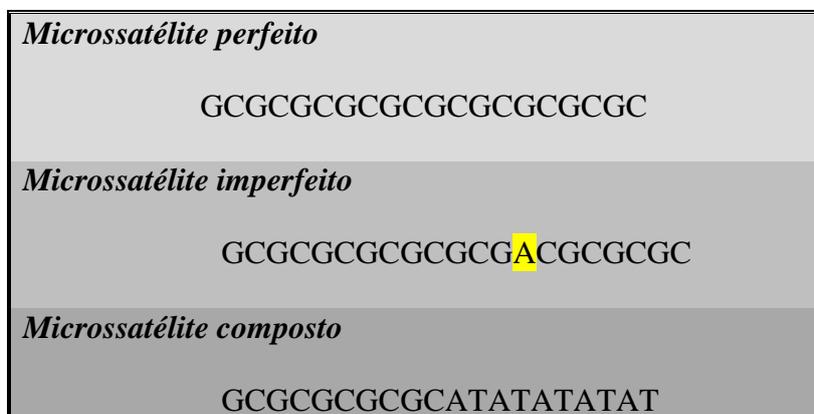
Variações no número e tamanho dos sítios de DNAr 5S e 45S têm sido frequentemente reportadas entre espécies de plantas pertencentes ao mesmo gênero (Marcon *et al.*, 2005; Hasterok *et al.*, 2006), assim como entre acessos de uma mesma espécie (Pedrosa-Harand *et al.*, 2006). Nesse sentido, o evento de *crossing-over* desigual, principalmente entre cromátides-irmãs, tem sido apontado como a principal força propulsora no processo evolutivo dos genes de RNAr. Tal recombinação aleatória facilitaria a formação de cromossomos, com diferentes números de unidades de repetições de DNAr (Eickbush e Eickbush, 2007).

### **3.3.1.2 MICROSSATÉLITES**

As SSRs (*Simple Sequence Repeats*; Repetições de Sequências Únicas), também conhecidas por microssatélites, consistem em pequenas unidades de repetições (1 a 6 pb) distribuídas em tandem, abundantes ao longo dos genomas procarióticos e

eucarióticos, sendo encontradas entre regiões codificantes de genes estruturais ou outras sequências repetitivas (Bhargava e Fuentes, 2009).

Os microssatélites podem ser caracterizados com base em parâmetros inerentes à sua estrutura e localização. Com relação à estrutura, essas sequências são classificadas em perfeitas, imperfeitas e compostas. Os microssatélites perfeitos são constituídos por extensões ininterruptas de um motivo particular. Por outro lado, as SSRs imperfeitas consistem em repetições interrompidas, ao longo de sua extensão, por nucleotídeos diferentes do motivo original. Por último, os microssatélites compostos possuem dois grupos de motivos diferentes adjacentes (Figura 1) (Weising *et al.*, 2005; Oliveira *et al.*, 2006).



**Figura 1.** Classificação dos microssatélites de acordo com a estrutura das unidades de repetições.

Considerando a sua localização no genoma, as SSRs são classificadas em genômicas ou gênicas. Nos microssatélites genômicos, a frequência e distribuição das sequências refletem a taxa de mutação básica do organismo ou da região genômica, na qual está localizada. Por outro lado, as SSRs gênicas têm sua disposição e assiduidade resultante da taxa de mutação dessas regiões codificantes associada a forças seletivas (Ellegren, 2004).

Apesar de pouco esclarecidas, as taxas mutacionais dessas repetições não são uniformes, variando entre os tipos e composição das repetições, o tipo de SSRs, assim como entre grupos taxonômicos (Primmer *et al.*, 1996). Além disso, são diretamente afetadas pelos motivos repetidos e por fatores como loco gênico, posição cromossômica, divisão celular, estrutura da cromatina, conteúdo GC (Guanina/Citosina) das sequências flanqueadoras e taxa local de mutação. Contudo, afirma-se que o

principal fator influenciador da mutabilidade está associado à extensão das repetições (Ellegren, 2004; Weising *et al.*, 2005).

O mecanismo de replicação *slippage* tem sido sugerido como o principal responsável pela formação e expansão dos arranjos de microssatélites, atuando juntamente com a funcionalidade de um grupo de enzimas e proteínas envolvidas no processamento do DNA e com a remodelagem da cromatina (Cuadrado e Schwarzacher, 1998; Bhargava e Fuentes, 2009). Segundo Bhargava e Fuentes (2009), os microssatélites também podem se originar via transposição de elementos móveis ou por microssatélites *de novo*. Neste último caso, os microssatélites surgem provavelmente de modo espontâneo, a partir ou internamente em sequências únicas dispostas em regiões eucromáticas. Estas regiões são caracterizadas pela alta proporção de genes expressos, bem como pela ocorrência de recombinação meiótica. Por consequência, as SSRs desempenhariam um papel fundamental na manutenção dos códons genéticos e da integridade proteica (Li *et al.*, 2004).

Os elementos transponíveis, por sua vez, são sequências que desempenham um papel fundamental na evolução genômica. Esses elementos móveis podem apresentar um ou mais sítios com predisposição para a formação de microssatélites, favorecendo sua dispersão e amplificação em diferentes regiões genômicas (Bhargava e Fuentes, 2009), inclusive na heterocromatina, uma vez que esses retroelementos podem estar direta ou distantemente associados a blocos heterocromáticos (Gartenberg, 2009). Todavia, o envolvimento dos retrotransposons na gênese dos microssatélites é bastante questionável, pois nem sempre uma alta densidade de elementos transponíveis está associada a uma elevada densidade de SSRs (Bhargava e Fuentes, 2009).

As informações contidas em bancos de dados, bem como o sequenciamento dos genomas em diferentes eucariotos têm permitido a acessibilidade à investigação da abundância, tipo e frequência dos motivos de repetições dos microssatélites em diferentes organismos. Análises destes genomas revelaram que a frequência dos microssatélites varia amplamente entre os genomas, ressaltando-se uma distribuição não aleatória destas sequências, as quais mostram propriedades únicas nas regiões genômicas com diferente funcionalidade (Tóth *et al.*, 2000). Dessa forma, observa-se uma seleção diferencial, em termos de número absoluto de loci de microssatélites e repetições preferenciais, por diferentes grupos taxonômicos.

Segundo Morgante *et al.* (2002) existe uma relação positiva linear entre a frequência de SSRs e a porcentagem de DNA cópia única em plantas, sugerindo que os

microssatélites seriam mais abundantes nestas regiões. Entretanto, isso contradiz o fato que os SSRs seriam elementos derivados de sequências repetitivas, estando o aumento da sua densidade associado ao tamanho do genoma das espécies.

Tóth *et al.* (2000) observaram que a abundância de vários tipos de microssatélites, assim como suas frequências em regiões intergênicas, éxons e íntrons eram divergentes entre os primatas, roedores, artrópodos, plantas, leveduras e outros fungos. Em plantas, observaram a predominância de mono (A e/ou T) e dinucleotídeos (AT) nas regiões intrônicas e intergênicas, enquanto que os trinucleotídeos (AAG) foram mais abundantes nos éxons. Mais especificamente, Kangfu *et al.* (1999) ressaltaram a predominância de repetições AT nos gêneros *Phaseolus* e *Vigna*, após analisarem 326 sequências de DNA depositadas no *GenBank*.

A frequência e distribuição de diferentes motivos SSRs têm sido alvo de estudos, principalmente em alguns genomas parcial ou completamente sequenciados, como em *P. vulgaris* (Schlueter *et al.*, 2008) e *G. max* (Hyten *et al.*, 2010a), auxiliando no entendimento da organização genômica de diferentes espécies. Concomitantemente, alguns estudos têm visado à distribuição e à organização cromossômica das SSRs mediante a FISH, sendo úteis na investigação de relações evolutivas (Cuadrado e Jouve, 2007a), identificação cromossômica (Chang *et al.*, 2008) e construção de mapas cromossômicos (Ohmido *et al.*, 2007).

Em termos gerais, a FISH tem revelado divergências em termos de quantidade e distribuição dos microssatélites entre os taxa, indicando uma distribuição genômica não aleatória e motivo dependente, como relatado em grão-de-bico (Gortner *et al.*, 1998), cevada (Cuadrado e Jouve, 2007a,b; Cuadrado *et al.*, 2008), trigo e centeio (Cuadrado *et al.*, 2008), entre outros. Em tomate, por exemplo, a heterocromatina pericentromérica é constituída por *clusters* de repetições GATA e AAC (Yang *et al.*, 2005).

Os oligonucleotídeos com padrão de microssatélites têm sido revelados como marcadores citológicos para a identificação cromossômica, sendo utilizados simultaneamente ou associadamente com outros tipos de DNA repetitivo (como os DNAr), permitindo a construção de mapas cromossômicos. Chang *et al.* (2008) descreveram o par cromossômico 2 da espécie *Solanum lycopersicum* L. pela presença de (GACA)<sub>4</sub> e DNAr 45S. Em leguminosas, Fuchs *et al.* (1998) estabeleceram marcadores físicos para cada um dos sete pares cromossômicos de ervilha através da disposição dos oligonucleotídeos (AG)<sub>12</sub>, (AC)<sub>12</sub>, (GAA)<sub>10</sub> e (GATA)<sub>7</sub> associados às sequências de DNAr 5S e 45S. Prontamente, estes estudos suportam a base da

identificação de segmentos e/ou cromossomos inteiros, auxiliando no entendimento da organização genômica, assim como da evolução, diversidade e relação cariotípica (Cuadrado *et al.*, 2008).

### **3.3.2 SEQUÊNCIAS REPETITIVAS DISPERSAS – ELEMENTOS TRANSPONÍVEIS**

Os elementos transponíveis consistem em sequências de DNA moderadamente repetidas que, ao contrário do DNA satélite, não estão arranjadas em tandem, mas sim dispersas no genoma (Kubis *et al.*, 1998). Tais elementos possuem a capacidade de movimentação ao longo dos genomas por transposição, de forma não aleatória, podendo causar mutações benéficas, neutras e/ou deletérias, alterações da estrutura e função dos genes, rearranjos cromossômicos e modificações na regulação gênica, impulsionando assim, a expansão e evolução da estrutura genômica em vegetais, seja em estado ativo e/ou inativo (Alix e Heslop-Harrison, 2004; Rajput e Upadhyaya, 2009).

O mecanismo de transposição requer a interação de proteínas codificadas pelos próprios elementos móveis. Os elementos autônomos possuem uma ORF (*Open Reading Frame*; Quadro Aberto de Leitura) que codifica os produtos necessários à transposição, como as enzimas transcriptase reversa ou transposase. Por sua vez, os elementos não autônomos não codificam proteínas, mas retêm sequências *cis*-regulatórias imprescindíveis para a transposição, bem como utilizam as enzimas codificadas pelos elementos autônomos (Wessler, 2006).

De acordo com sua organização estrutural e mecanismo de transposição, os TEs (*Transposable Elements*; Elementos transponíveis) são distribuídos em duas principais classes (I e II). Os elementos de classe I ou retrotransposons transpõem-se por um intermediário de RNA, via mecanismo do tipo *copy and paste*, o qual sofre transcrição reversa para sua reinserção no genoma. Os elementos de classe II ou transposons, por sua vez, codificam uma transposase, movimentando-se pelo genoma mediante um DNA intermediário, via mecanismo do tipo *cut and paste* (Todorovska, 2007; Du *et al.*, 2010).

Além desta classificação, os elementos transponíveis são agrupados em subclasses, superfamílias, famílias e subfamílias, com base na presença e/ou ausência de domínios específicos, similaridade nucleotídica e de aminoácidos, assim como pelo tamanho do TSD (*Target Site Duplication*; Duplicação do Sítio Alvo), curta sequência genômica duplicada no sítio de inserção dos TEs (Capy *et al.*, 1998).

Os retrotransposons são divididos em duas subclasses, as quais diferem em sua estrutura e ciclo de transposição, retroelementos não LTRs (*Long Terminal Repeats*; Longas Repetições Terminais) e LTRs (Friesen *et al.*, 2001; Hill *et al.*, 2005; Park *et al.*, 2007).

Os elementos não LTRs são divididos em duas superfamílias denominadas SINEs (*Short Interspersed Nucleotide Elements*; Elementos Nucleares Interdispersos Curtos) e LINEs (*Long Interspersed Nucleotide Elements*; Elementos Nucleares Interdispersos Longos). Os retroelementos LINEs são autônomos, possuindo uma região *gag* (*Group-specific antigen*; antígeno grupo-específico) e parte da região *pol* (RT e RNase H), bem como caudas poli-A; enquanto que os SINEs não apresentam nenhuma sequência codificadora, sendo elementos não autônomos, mas contêm um promotor interno para RNA polimerase e terminais ricos em adenina (Wessler, 2006).

Por sua vez, os retrotransposons LTR são flanqueados por duas LTRs (posições 5' e 3') e uma ORF internamente, a qual é normalmente constituída pelo gene *gag*, codificador de uma poliproteína, a qual compreende três proteínas correspondentes aos subcomponentes da VLP (*Virus-like Particle*; partícula similar a vírus), envolvida na maturação e empacotamento do RNA; e gene *pol* (*Polyprotein*; Poliproteína), responsável pela codificação de várias proteínas, como proteases, integrase, transcriptase reversa e RNase H, as quais são necessárias para a síntese e integração do DNAc (*Complementary DNA*; DNA complementar) em cromossomos hospedeiros (Chavanne *et al.*, 1998; Cheng *et al.*, 2009).

Considerando o ciclo de vida dos retroelementos, o RNA é inteiramente transcrito por uma RNA polimerase II codificada pelo hospedeiro sendo transportado para o citoplasma, onde é traduzido. Diferentes proteínas são resultantes do processo de tradução, entre elas a integrase e a transcriptase reversa. Esta última, utiliza o RNA transcrito como molde para a produção de uma molécula de DNAc, a qual é direcionada para o núcleo. Finalmente, o DNAc é integrado aos cromossomos dispersamente, pela ação da integrase, originando uma ou mais novas cópias do retrotransposon no genoma hospedeiro. Conseqüentemente, em seu processo de mobilização, os retroelementos não sofrem excisão e reinserção, permitindo que a sequência original permaneça no genoma hospedeiro, havendo um aumento exponencial do número de cópias do retroelemento (Kumar e Bennetzen, 1999).

Tal mecanismo replicativo parece estar relacionado à grande abundância dos retrotransposons, bem como às diferenças de tamanhos genômicos entre diferentes

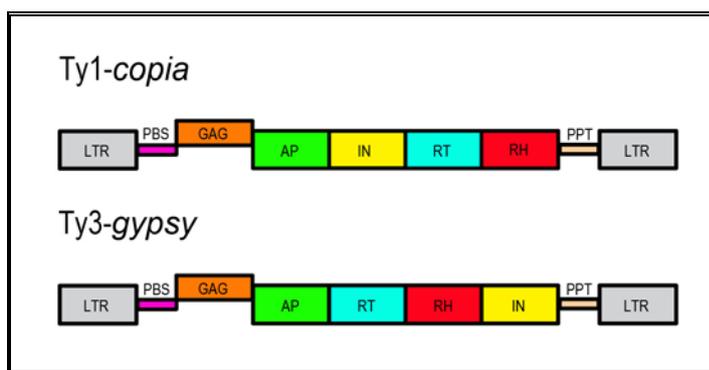
espécies vegetais, onde são caracterizados como componentes majoritários, perfazendo 50-90% dos genomas vegetais. Essa porcentagem enfatiza o papel central que os retroelementos desempenham na reestruturação evolutiva, sendo plausível a existência de mecanismos que controlem a atividade destes elementos na expansão dos genomas.

A inativação dos elementos móveis é decorrente de mecanismos de regulação controlados pelo hospedeiro, assim como pelo próprio elemento (Bennetzen, 2000). Considerando os elementos transponíveis autônomos, a regulação transcricional e pós-transcricional tem sido considerada como a forma mais efetiva de silenciamento dos elementos móveis, como por exemplo, a metilação. Este mecanismo controla a transcrição excessiva e mudanças estruturais ocasionadas por recombinação ectópica, principalmente em plantas, consistindo em um mecanismo de defesa genômica (Hirochika *et al.*, 2000).

Entre os retrotransposons LTRs, as superfamílias Ty1-*copia* e Ty3-*gypsy* são consideradas as mais abundantes e relatadas nos genomas de diferentes angiospermas, representando, por exemplo 25% e 75% dos genomas de arroz e milho, respectivamente (Du *et al.*, 2010).

Essas duas superfamílias de retrotransposons diferenciam-se pelo domínio conservado dos genes RT (*Reverse Transcriptase*; Transcriptase Reversa), característicos de cada retroelemento (Friesen *et al.*, 2001), bem como pela ordem gênica dos domínios RT e integrase na região *pol*. O Ty1-*copia* apresenta o domínio da integrase anterior (posição 5') ao da RT e RNase H, enquanto que, no Ty3-*gypsy*, tal domínio situa-se posterior (posição 3') ao RT e RNase H (Figura 2). Além dessa característica, os elementos Ty3-*gypsy-like* apresentam uma maior similaridade com os retrovírus (Kumar e Bennetzen, 1999; Hansen e Heslop-Harrison, 2004), pela presença de domínio *env* (Envelope), coincidente com a posição ocupada pelo gene codificante do envelope viral em retrovírus (Vershinin *et al.*, 2002), sendo estes classificados como *Gypsy-like* (Capy *et al.*, 1998).

Outra classificação foi proposta para os retrotransposons pelo ICTV (*Internacional Committee on Taxonomy of Viruses*; Comitê Internacional em Taxonomia de Vírus), sendo baseada nas similaridades da sua organização genômica com os retrovírus, bem como nas relações filogenéticas entre as transcriptases reversas. De acordo com esta classificação, os retrotransposons são subdivididos em duas ordens *Retrales* e *Retrovirales*, estando classificados nesta última, os retroelementos LTRs em duas famílias Pseudoviridae (Ty1-*copia*) e Metaviridae (Ty3-*Gypsy*) (Capy, 2005).



**Figura 2.** Organização dos domínios gênicos das principais classes de retrotransposons LTRs (Ty1-copia e Ty3-gypsy) (adaptado de Todorovska, 2007).

Vários estudos têm focado a caracterização do conteúdo, heterogeneidade, atividade e distribuição genômica geral dos retrotransposons em diferentes espécies vegetais, visando a uma melhor compreensão sobre a organização genômica. Os retroelementos Ty1-copia e Ty3-gypsy têm prevalecido em todos os principais grupos vegetais, como em *Helianthus* L. (Natali *et al.*, 2006), *Malus domestica* Borkh (Sun *et al.*, 2008) e *Pisum sativum* L. (Macas *et al.*, 2007). Entretanto, Dixit *et al.* (2006) relataram que a ocorrência de Ty3-gypsy tem sido pouco explorada, estando limitada a alguns grupos taxonômicos (Natali *et al.*, 2006; Park *et al.*, 2007).

Xiao *et al.* (2007) isolaram as sequências completas de dois elementos Ty1-copia-like (RTvr-1 e RTvr-2) presentes no genoma de *V. radiata*. Os retroelementos RTvr-1 e RTvr-2 apresentaram um domínio interno de 3991 e 6484 pb, respectivamente, com terminais LTR contendo, aproximadamente, 680 pb. O alinhamento destes retroelementos revelou a presença de uma sequência com 2500 pb, localizada entre a região 3' do domínio da integrase e 5' da transcriptase reversa, em RTvr-2. Com exceção deste domínio, a similaridade entre os dois retrotransposons foi igual a 92%, sugerindo que estes pertenciam à mesma família.

Conjuntamente, uma análise filogenética foi realizada visando investigar uma relação evolutiva entre os RTvrs isolados e outros retrotransposons do tipo LTR conhecidos, utilizando o domínio da transcriptase reversa. Os RTvrs agruparam-se com retrotransposons Ty1-copia-like identificados em diferentes espécies vegetais, contudo, estariam mais proximamente relacionados ao Tnt1-94, retrotransposon copia-like

proveniente de *Nicotiana tabacum*, do que a outras classes destes elementos observadas em outras leguminosas (Xiao *et al.*, 2007).

Diferentes estudos têm isolado e caracterizado o domínio RT de famílias de retrotransposons, observando a distribuição destes elementos no genoma, mediante FISH. Geralmente, os retroelementos apresentam uma distribuição dispersa, a qual reflete o seu mecanismo de transposição, sendo encontrados entre outras sequências repetitivas ou associados a regiões genômicas particulares (Kubis *et al.*, 1998; Heslop-Harrison, 2000).

Natali *et al.* (2006) analisaram a localização *in situ* das sequências repetitivas, pHaS13 e pHaS211, com similaridade aos genes *int* (retroelemento Ty3-*gypsy*) e RNase H (retrotransposon Ty1-*copia*), respectivamente, em espécies do gênero *Helianthus*. Apesar da distribuição dispersa ressaltada em todo o complemento cromossômico para ambos os elementos móveis, a sequência Ty3-*gypsy*-like apresentou uma localização preferencialmente centromérica em todas as espécies analisadas; enquanto que, em *Helianthus annuus* L., o fragmento Ty1-*copia*-like distribuiu-se nas regiões terminais dos cromossomos. Em *Copaifera langsdorffii* Desf., Gaeta *et al.* (2010) observaram *dots* e blocos de Ty3-*gypsy* em posições intercalares e terminais, por vezes associados a regiões ricas em GC.

Por sua vez, Nielen *et al.* (2009) isolaram e caracterizaram o retrotransposon Athila-like (Família Ty3-*gypsy*), o qual apresenta alta similaridade com *Arabidopsis thaliana*, do genoma do amendoim cultivado (*Arachis hypogaea* L.) e de seus parentais diploides *A. duranensis* Krapov. & W.C.Greg. (genoma AA) e *A. ipaënsis* Krapov. & W.C.Greg. (genoma BB). A análise com FISH mostrou que o elemento encontrava-se distribuído dispersamente em regiões eucromáticas dos cromossomos, possuindo maior intensidade nos cromossomos do genoma AA, em comparação aos do genoma BB.

### **3.4 CARACTERÍSTICAS TAXONÔMICAS E IMPORTÂNCIA ECONÔMICA DA FAMÍLIA FABACEAE LINDLEY**

A família Fabaceae compreende um grupo taxonômico de distribuição cosmopolita, com espécies nativas em todos os continentes, exceto na Antártida. É considerada como uma das maiores famílias das Magnoliophyta, constando de 727 gêneros e aproximadamente 19.000 espécies (Lewis *et al.*, 2005). Ocupa o terceiro lugar em diversidade, estando seus representantes distribuídos, principalmente nas regiões tropicais e subtropicais (Souza e Benko-Iseppon, 2004).

É constituída por três subfamílias Mimosoideae, Caesalpinioideae e Papilionoideae (ou Faboideae), as quais reúnem cerca de 2.500, 2.700 e 12.615 espécies, respectivamente (Judd *et al.*, 2002; Biondo *et al.*, 2005). São plantas de hábito variado, existindo árvores, ervas anuais ou perenes, arbustos e lianas, entre outros; apresentam nódulos radiculares contendo bactérias fixadoras de nitrogênio propiciando um eficiente metabolismo de nitrogênio (Barroso *et al.*, 1991; Joly, 2002; Souza e Lorenzi, 2005).

As folhas são compostas, com filotaxia alterna e estípulas e estípelas as quais, às vezes, transformam-se em espinhos. Apresentam inflorescência racemosa, cujas flores podem ser vistosas ou não, bissexuadas, actinomorfas ou zigomorfas, diclamídeas ou raramente monoclamídeas, usualmente com hipanto. Geralmente, o cálice e a corola são pentâmeros, dialissépalos ou gamossépalos; pétalas semelhantes entre si ou diferenciadas em carenas ou quilhas (inferiores), alas ou asas (medianas) e vexilo ou estandarte (a superior diferente de todas as outras). O androceu é representado por quatro a 10 estames, livres ou unidos entre si, anteras rimosas ou raramente poricidas. O gineceu é unicarpelar, apresentando ovário súpero com placentação marginal e uniloculado, por vezes divididos por falsos septos, com muitos óvulos. O fruto característico desta família é do tipo legume, definido como seco e deiscente por duas valvas; contudo, outros tipos são encontrados como drupa, sâmara, folículo, craspódio ou lomento (Barroso *et al.*, 1991; Joly, 2002; Souza e Lorenzi, 2005).

Esta família é classificada como a segunda maior família vegetal de importância econômica, perdendo apenas para a Poaceae, devido à sua grande utilização para os mais diversos fins (Singh *et al.*, 2007). Na dieta alimentar, destacam-se os gêneros *Glycine*, *Phaseolus* e *Vigna*, os quais pertencem à subfamília Papilionoideae, tribo Phaseoleae, subtribo Phaseolinae, estando agrupados em um clado monofilético (Phaseoloids) (Marechal *et al.*, 1978). Os referidos gêneros apresentam como espécies mais importantes, *G. max*, *P. vulgaris*, *P. lunatus* e *V. unguiculata*, conhecidas popularmente como soja, feijão comum, feijão-fava e feijão-caupi, respectivamente (Barroso *et al.*, 1991; Souza e Lorenzi, 2005).

Os grãos de feijão representam uma importante fonte proteica na dieta humana nos países em desenvolvimento das regiões tropicais e subtropicais. No Brasil, o feijão é um dos componentes básicos da dieta alimentar da população, destacando-se como importante fonte de proteína para as classes economicamente menos favorecidas (Wander, 2005). Além da importância proteica, a soja é uma das principais fontes de

óleo vegetal do mundo e a primeira leguminosa a ser completamente sequenciada. Estima-se que seu genoma servirá como referência para mais de 20.000 espécies de leguminosas e para o entendimento do mecanismo de fixação do nitrogênio atmosférico por simbiose (Schmutz *et al.*, 2010).

Na região Nordeste do Brasil, dentre as plantas cultivadas e usadas pela população, nenhuma tem tantas características vantajosas como o feijão-caupi (Freire-Filho *et al.*, 1999; 2005). Constitui-se em importante fonte de proteínas para os agricultores de baixo poder aquisitivo, contribuindo com 31% da proteína consumida no nordeste, encerrando 24% de proteínas e 340 calorias para cada 100 g de sementes. Adicionalmente, esta cultura apresenta aminoácidos essenciais ausentes nos demais, elevados níveis de ácido fólico, baixos níveis de fatores antinutricionais e flatulência, além de ter digestibilidade muito melhor que os demais feijões (Ehlers e Hall, 1997; Silva e Freire-Filho, 1999). Trata-se de uma das plantas mais versáteis que existem, sendo muito valiosa como forrageira, com uma capacidade ímpar de fertilizar solos pobres, mesmo em condições de calor, seca e salinidade (Ehlers e Hall, 1997).

Algumas espécies são bastante usadas para fins ornamentais, sendo a principal família utilizada na arborização urbana no Brasil, destacando-se a tipuana [*Tipuana tipu* (Benth.) Kuntze], o pau-ferro (*Caesalpinia ferrea* Mart.), o flamboyant [*Delonix regia* (Bojer ex Hook) Raf.] e a pata de vaca (*Bauhinia variegata* L.) (Souza e Lorenzi, 2005).

Por outro lado, inúmeras são as Fabaceae que produzem madeira de excelente qualidade, incluindo a cerejeira [*Amburana cearensis* (Arr. Cam.) A.C. Smith], o jatobá (*Hymenaea* spp.), o angelim (*Andira* spp.), o angico (*Anadenanthera* spp.) e o pau-brasil (*Caesalpinia echinata* Lam.), espécie nativa da Mata Atlântica responsável pelo primeiro ciclo econômico do Brasil. Outra importante aplicação é o uso na adubação verde, principalmente considerando a associação com bactérias fixadoras de nitrogênio (*Rhizobium*) que produzem pequenos nódulos nas raízes, como feijão-guandu [*Cajanus cajan* (L.) Millsp.], crotalária (*Crotalaria* spp.) e alfafa (*Medicago sativa* L.), sendo esta última também utilizada como forrageira (Barroso *et al.*, 1991; Souza e Lorenzi, 2005).

### **3.5 CARACTERÍSTICAS TAXONÔMICAS E CITOGENÉTICAS DO GÊNERO *Glycine* WILLD.**

A soja cultivada (*G. max*) apresenta uma provável origem monofilética a partir da soja selvagem, *G. soja* Sieb. & Zucc. É proveniente do sul da China, dispersando em seguida para outros países do leste da Ásia (Guo *et al.*, 2010). Atualmente, os Estados

Unidos possuem a maior área plantada e a maior produção, seguidos do Brasil, Argentina, China e Índia. De acordo com a estatística da FAO (*Food and Agriculture Organization*; Organização para Agricultura e Alimentação) (<http://faostat.fao.org/>), estes cinco países possuíram juntos em 2008, aproximadamente, 89% da área plantada e 91% da produção mundial.

O gênero *Glycine* é constituído por dois subgêneros: *Glycine* Willd., que inclui 26 espécies perenes selvagens; e *Soja* (Moench.) F.J. Herm., o qual é representado por *G. soja* e *G. max*. Ambas as espécies do subgênero *Soja* são anuais com  $2n = 40$  cromossomos, são reprodutivamente compatíveis, produzindo híbridos viáveis, vigorosos e férteis, capazes de trocar genes entre si (Singh e Hymowitz, 1999). Os genomas de *G. soja* e *G. max* são similares e têm sido designados por G e G<sub>1</sub>, respectivamente (Chung e Singh, 2008). Algumas linhagens diferem por apresentarem translocações recíprocas (Singh e Hymowitz, 1999) ou inversões paracêntricas (Palmer *et al.*, 2000).

O sequenciamento do genoma da soja foi do tipo *shotgun*, compreendendo 950 Mb. A maior parte das sequências do genoma está montada em 20 pseudomoléculas cromossômicas contendo 397 sequências *scaffolds* com posições ordenadas dentro dos 20 grupos de ligação da soja. Um conteúdo adicional de 17,7 Mb está presente em 1.148 sequências *scaffolds* não ancoradas, que são constituídas principalmente por DNA repetitivo e contém menos dos 450 genes previstos (Schmutz *et al.*, 2010). As posições dos *scaffolds* foram determinadas com extensos mapas genéticos, incluindo 4.991 SNPs (*Single Nucleotide Polymorphism*; Polimorfismos de um Único Nucleotídeo) e 874 SSRs (Song *et al.*, 2004; Choi *et al.*, 2007; Hyten *et al.*, 2010a, b).

De acordo com Schmutz *et al.* (2010), cerca de 57% das sequências genômicas da soja estão localizadas em regiões heterocromáticas com baixo índice de recombinação, as quais estão distribuídas próximas ao centrômero. Lin *et al.* (2005) sugerem que as sequências repetitivas de soja estejam restritas a grandes blocos de heterocromatina pericentromérica, constituída por DNA satélites com retroelementos interdispersos, sendo estes últimos caracterizados como pertencentes às classes *copy-like* e *gypsy-like* (Numberg *et al.*, 2006; Swaminathan *et al.*, 2007).

Por sua vez, Gill *et al.* (2009) caracterizaram e hibridizaram as sequências satélites CentGm-1 e CentGm-2, visualizando 14 e oito sítios centroméricos, respectivamente, afirmando serem constituintes das regiões centroméricas dos cromossomos de soja. Estas sequências repetitivas, contendo 91-92 bp por unidade,

representam uma fração dos *scaffolds* organizados em 19 das 20 pseudomoléculas cromossômicas, durante a construção do mapa físico de soja. Contudo, quatro cromossomos continham vários *clusters* de CentGm-1 e/ou CentGm-2 em diferentes posições cromossômicas, refletindo sua posição física correta ou a dificuldade de orientar essas regiões altamente repetitivas (Schmutz *et al.*, 2010).

A aloploidização parece ter desempenhado um papel muito importante na especiação dentro do gênero *Glycine*. Embora seus cromossomos se comportem citogeneticamente como diploides, as espécies com  $2n = 40$  apresentam provável origem aloploide (Chung e Singh, 2008; Gill *et al.*, 2009). Em soja, observou-se que, devido a várias rodadas de duplicação e diploidização no genoma, assim como rearranjos cromossômicos, múltiplas cópias de sequências de DNA (ou blocos) podem estar presentes em mais de dois cromossomos. Em média, 61,4% dos genes homólogos são encontrados em blocos envolvendo apenas dois cromossomos, apenas 5,63% ao longo de três cromossomos, e 21,53%, em quatro cromossomos.

A soja não é considerada uma planta modelo para estudos citogenéticos por apresentar cromossomos numerosos ( $2n = 40$ ), pequenos e de tamanhos similares (1,26–1,99  $\mu\text{m}$ , para cromossomos metafásicos) (Clarindo *et al.*, 2007). São seis pares de cromossomos metacêntricos e 14 submetacêntricos (Clarindo *et al.*, 2007), destacando-se apenas um par de cromossomos organizadores de nucléolo, o que também é observado para a maioria das espécies de *Glycine* com  $2n = 40$  (Singh *et al.*, 2001; Ohmido *et al.*, 2007).

A primeira descrição citológica para soja foi desenvolvida utilizando cromossomos paquitênicos, os quais foram numerados de 1 a 20, com base no comprimento cromossômico total, relação entre braços e proporções relativas de eucromatina e heterocromatina (Singh and Hymowitz, 1999). O uso de trissômicos primários de soja permitiu associar 11 dos 20 grupos de ligação moleculares (MLGs) a cromossomos específicos (1 = D1a+q, 3 = N, 5 = A1, 8 = A2, 9 = K, 13 = F, 14 = C1, 17 = D2, 18 = G, 19 = L and 20 = I) (Zou *et al.*, 2003).

Contudo, apesar da conclusão do sequenciamento do genoma da soja (Schmutz *et al.*, 2010), não foi possível a associação dos nove MLGs restantes a seus respectivos cromossomos em uma célula, utilizando trissômicos primários (Zou *et al.*, 2003). Para esses cromossomos, os números foram fornecidos em ordem decrescente com base no comprimento da distância genética estimada, disponível no banco de dados SoyBase (<http://soybase.org/LG2Xsome.php>). Recentemente, um coquetel de sondas [BAC

(*Bacterial Artificial Chromosome*; Cromossomo Artificial de Bactéria)] geneticamente ancorados foi utilizado na FISH de soja e *G. soja*, permitindo a identificação simultânea dos 20 pares cromossômicos em preparações mitóticas de soja e o estabelecimento da relação com suas respectivas pseudomoléculas sequenciadas. Além disso, esse estudo possibilitou a identificação de translocação recíproca entre os cromossomos 11 e 13 em dois acessos da soja selvagem (Findley *et al.*, 2010).

### 3.6 CARACTERÍSTICAS TAXONÔMICAS E CITOGENÉTICAS DO GÊNERO *Phaseolus* L.

O gênero *Phaseolus* é constituído por aproximadamente 75 espécies (Mercado-Ruaro *et al.*, 2009). Entre estas, destacam-se apenas cinco espécies cultivadas: *P. vulgaris*, *P. lunatus*, *P. coccineus* L., *P. acutifolius* A. Gray e *P. polyanthus* Greman (Vieira *et al.*, 1999; Mercado-Ruaro e Delgado-Salinas 2000).

Delgado-Salinas *et al.* (2006) analisaram a filogenia deste gênero, por meio de dados morfológicos, bioquímicos e moleculares (sequências DNArn ITS/5,8S e loci plastidiais trnK), confirmando sua origem monofilética. Estes autores observaram que as espécies do referido táxon podem ser divididas em dois clados principais. O clado A seria constituído pelos grupos Tuerckheimii, Pauciflorus e Pedicellatus. Por sua vez, o clado B era formado pelos grupos Filiformis, Leptostachyus, Lunatus, Polystachios e Vulgaris, este último agrupando as espécies *P. vulgaris*, *P. coccineus*, *P. acutifolius* e *P. polyanthus*. O táxon *P. lunatus* foi classificado no grupo Lunatus, sugerindo que, entre as espécies cultivadas, pode ser considerada a mais distante filogeneticamente.

O feijoeiro comum (*P. vulgaris*) é a espécie mais conhecida do gênero, existindo certa controvérsia quanto à sua origem. De Candolle (1883) *apud* Vieira *et al.* (1999) afirmou, em seu tratado sobre a origem das plantas cultivadas, que o feijoeiro apresentava origem desconhecida. Contudo, Gepts (1998) sugeriu que *P. vulgaris* tenha se originado na fronteira entre o Equador e o Nordeste do Peru, havendo dois centros de origem denominados como região Mesoamericana, estendendo-se do México ao norte da América do Sul, bem como a região Andina, compreendendo a região central e o sul da América do Sul.

Os primeiros estudos citogenéticos em *Phaseolus* foram restritos a contagens cromossômicas, devido ao pequeno tamanho dos cromossomos e a morfologia uniforme (Zheng *et al.*, 1991; Mercado-Ruaro e Delgado-Salinas, 2000), sendo determinado o número cromossômico  $2n = 22$  para as espécies *P. acutifolius*, *P. coccineus*, *P. lunatus*

e *P. vulgaris* (Karpetschenko, 1925; Mercado-Ruaro e Delgado-Salinas 2000). O feijão comum apresenta cromossomos metacêntricos e submetacêntricos, relativamente pequenos (1,5 a 3  $\mu\text{m}$ ) (Zheng *et al.*, 1991; Moscone *et al.*, 1999), incorporando um genoma estimado em 637 Mb ou 0,66 pg/1C (Aramuganathan e Earle, 1991).

A heterocromatina constitutiva, analisada através de bandeamento C, localiza-se preferencialmente nas regiões pericentroméricas da maioria dos cromossomos de *P. vulgaris*, ressaltando uma distribuição característica do gênero (Moscone *et al.*, 1999; Almeida, 2006). Contudo, existem relatos de blocos C<sup>+</sup> nas regiões teloméricas dos cromossomos de *P. lunatus* (Moscone *et al.*, 1999). Zheng *et al.* (1994) enfatizaram a presença de bandas CMA<sub>3</sub><sup>+</sup>/DAPI proximais no feijão comum, colocalizadas com as bandas C<sup>+</sup>, sugerindo uma composição de heterocromatina rica em GC. Tal dado foi corroborado pelos resultados obtidos por Almeida (2006) ao investigar, comparativamente, a distribuição da heterocromatina em *P. lunatus* e *P. vulgaris*. O complemento cromossômico de *P. lunatus* apresentou bandas CMA<sub>3</sub><sup>+</sup>/DAPI<sup>-</sup> localizadas nas RONS e regiões centroméricas. Em *P. vulgaris*, uma distribuição similar foi visualizada, entretanto, as bandas foram classificadas como CMA<sub>3</sub><sup>+</sup>/DAPI<sup>0</sup> (Almeida, 2006).

Com relação à distribuição do DNAr, Pedrosa-Harand *et al.* (2006) avaliaram a variação de sítios de DNAr 5S e 45S em 37 acessos selvagens e domesticados do feijão comum, representantes do *pool* gênico Mesoamericano e Andino. Os dois sítios de DNAr 5S mostraram-se bastante conservados dentro da espécie, enquanto houve uma grande variação no número de locos de DNAr 45S, tanto na sua posição quanto no número de repetições por loco. Observou-se 3-4 e 6-9 sítios, para os representantes Mesoamericano e Andino, respectivamente, sugerindo que mecanismos de recombinação ectópica entre regiões cromossômicas terminais seriam responsáveis pela gênese da variação observada. Em outras espécies cultivadas do gênero *Phaseolus* (*P. coccineus*, *P. acutifolius* e *P. lunatus*), Moscone *et al.* (1999) observaram uma variação tanto na distribuição dos sítios de DNAr 5S (1-3 sítios) quanto de DNAr 45S (1-3 sítios).

Os mapas genéticos têm sido correlacionados com os mapas cromossômicos mediante FISH de BACs, possibilitando identificar possíveis distorções das distâncias físicas encontradas nos mapas genéticos. Em *Phaseolus*, por exemplo, um mapa cromossômico foi recentemente concluído com sucesso para *P. vulgaris* (Pedrosa-Harand *et al.*, 2009; Fonsêca *et al.*, 2010).

Esses BACs foram utilizados como sondas em FISH na comparação entre os mapas de *P. vulgaris* e *P. lunatus*, revelando uma total conservação de sintenia e apenas três quebras de colinearidade (Almeida, 2006; Bonifácio, 2009). Por sua vez, entre *P. vulgaris* e *Phaseolus microcarpus* Mart., uma espécie selvagem e filogeneticamente distante, houve uma forte conservação de sintenia entre os cromossomos homeólogos e apenas quatro quebras de colinearidade, apesar da distância filogenética entre as duas espécies. Esses resultados demonstraram que em *Phaseolus* a evolução cromossômica parece estar relacionada a eventos raros e de pouca complexidade, como inversões, por exemplo, confirmando sua aparente estabilidade cariotípica postulada em estudos convencionais (Fonsêca, 2010).

### **3.7 CARACTERÍSTICAS TAXONÔMICAS E CITOGENÉTICAS DO GÊNERO *Vigna* SAVI.**

O gênero *Vigna* compreende cerca de 170 espécies (Freire-Filho *et al.*, 2005). Nas primeiras classificações taxonômicas descritas para o gênero *Vigna*, algumas espécies foram originalmente agrupadas em outros gêneros, como *Phaseolus* e *Dolichos* L. (Freire-Filho, 1988), sendo posteriormente alocadas no gênero *Vigna* a partir de análises morfológicas e taxonômicas (Marechal *et al.*, 1978).

Marechal *et al.* (1978), usando taxonomia numérica, realizaram um amplo estudo neste gênero dividindo-o em sete subgêneros: *Vigna*, *Plectotropis*, *Cestotropis*, *Lasiocarpa*, *Sigmoidotropis*, *Haydonia* e *Macrorhyca*. O subgênero *Vigna* foi considerado endêmico da África e dividido em seis seções: *Comosae*, *Liebrehtsia*, *Macrodontae*, *Reticulatae*, *Vigna* e *Catiang*, sendo a última formada por duas espécies botânicas reconhecidas, *V. unguiculata* e *Vigna nervosa* Markotter.

A espécie *V. unguiculata* foi diferenciada em quatro subespécies: três subespécies selvagens, *dekindtiana* (Harms) Verd., *tenuis* (E. Mey.) M. M. & S. e *stenophylla* (Harvey) M. M. & S., bem como uma subespécie cultivada, *unguiculata*, a qual reúne todo o feijão-caupi cultivado. A introdução das categorias cultigrupo e variedade, para distinguir unidades dentro das subespécies cultivada e selvagem, respectivamente, possibilitou a colocação de várias formas em uma mesma subespécie. Desse modo, Westphal (1974) *apud* Marechal *et al.* (1978) e Ng e Marechal (1985) dividiram a subespécie *unguiculata* em quatro cultigrupos: *Unguiculata*, *Sesquipedalis*, *Biflora* e *Textilis*. Dentre esses cultigrupos, apenas *Unguiculata* e *Sesquipedalis* são cultivados no Brasil, compreendendo a quase totalidade das cultivares locais e

melhoradas e ao feijão-de-metro, respectivamente (Freire-Filho *et al.*, 2005).

O feijão-caupi (*V. unguiculata*) tem vários nomes vulgares, sendo conhecido como feijão-de-macáçar ou de macaça, feijão-de-corda e feijão fradinho, entre outros. É uma planta herbácea, com ciclo de vida curto (entre 60 e 90 dias). Apresenta variação de porte, desde ereto até prostrado, e hábito de crescimento determinado a indeterminado (Freire-Filho *et al.*, 2005).

Em termos de origem e distribuição geográfica, o gênero *Vigna* possui distribuição pantropical, apresentando espécies na África, Índia, sudeste da Ásia, Américas Central e do Sul e na Austrália (Ghafoo *et al.*, 2001), com maior número de espécies endêmicas na África, sugerindo-se que a evolução e a dispersão deste gênero tenham ocorrido a partir deste continente (Freire-Filho, 1988).

O feijão-caupi (*V. unguiculata*) apresenta, como centro primário de diversidade, o oeste da África, mais precisamente, a Nigéria (Ng e Marechal, 1985). Todavia, Padulosi *et al.* (1997) afirmam que provavelmente a região do Transvaal, na República da África do Sul, seja a região de especiação desta cultura.

Considera-se que a dispersão do feijão-caupi tenha ocorrido por volta de 2.300 a.C (Ng e Marechal, 1985). No continente americano, acredita-se que a introdução do caupi tenha ocorrido a partir da Europa e do oeste da África, estando relacionada com a colonização espanhola e com o tráfico de escravos no século XVI (Steele, 1976).

Freire-Filho (1988) apresenta várias evidências de que o feijão-caupi foi introduzido na América Latina, no século XVI, pelos colonizadores espanhóis e portugueses. Contudo, Simon *et al.* (2007) analisando a diversidade genética e relações filogenéticas entre diferentes taxa de feijão-caupi por DAF (*DNA Amplification Fingerprinting*; Impressão Digital da Amplificação do DNA) contraria esta ideia. No Brasil, sua introdução provavelmente ocorreu pelo estado da Bahia, sendo levado pelos colonizadores para outras áreas da região Nordeste, assim como, do país. Tais afirmações levaram alguns autores a sugerir que o feijão-caupi teria uma base genética estreita no Brasil, em vista de seu melhoramento local ter ocorrido inicialmente com base em poucos genótipos introduzidos (Freire-Filho *et al.*, 1999; Freire-Filho *et al.*, 2005).

Atualmente, estudos citogenéticos encontram-se descritos para a caracterização cromossômica de espécies pertencentes ao gênero *Vigna*, no que se refere à utilização de técnicas de coloração convencional, bandeamento cromossômico, bem como FISH

com sondas de DNAr (Galasso *et al.*, 1995; Guerra *et al.*, 1996), DNA satélite (Galasso *et al.*, 1995) e retroelementos (Galasso *et al.*, 1997).

Diversos trabalhos envolvendo acessos de *Vigna*, pertencentes a diferentes espécies, enfatizaram o número básico  $x = 11$  e uma alta homogeneidade cariomorfológica para o citado gênero (Venora e Saccardo, 1993; Venora e Padulosi, 1997; Venora *et al.*, 1999). O primeiro relato de contagem cromossômica foi reportado por Karpechenko (1925), descrevendo o número cromossômico  $2n = 22$  para a espécie *V. unguiculata*.

O feijão-caupi apresenta um dos menores genomas entre as leguminosas (450-500 Mb), apresentando o nível diploide com  $2n = 22$  cromossomos (Arumuganathan e Earle, 1991), predominantemente submetacêntricos e metacêntricos, de difícil identificação devido ao tamanho e grau de condensação cromossômica.

O bandeamento C evidenciou um padrão pericentromérico em todo complemento cromossômico de *V. unguiculata* (Galasso *et al.*, 1992; Pignone *et al.*, 1995). Por sua vez, a análise com os fluorocromos CMA<sub>3</sub>/DAPI revelou a presença de bandas CMA<sub>3</sub><sup>+</sup>/DAPI nos pares cromossômicos 2, 3, 4 e 5, enquanto nenhuma banda DAPI<sup>+</sup> foi observada (Galasso *et al.*, 1993; 1995).

A FISH revelou a presença de quatro sítios maiores de DNAr 45S nas regiões subterminais dos cromossomos 2, 3, 4 e 5, assim como um sítio menor proximal no par cromossômico 10 ou 11 (Galasso *et al.*, 1995; Guerra *et al.*, 1996). Os sítios maiores estavam colocalizados com as bandas CMA<sub>3</sub><sup>+</sup> observadas por Galasso *et al.* (1993), indicando a presença de heterocromatina associada à RON. Os sítios de DNAr 5S foram localizados no braço longo do par cromossômico 4, portador de DNAr 45S, bem como no braço curto de um outro cromossomo não identificado.

Adicionalmente, Galasso *et al.* (1995) isolaram e caracterizaram uma sequência de DNA repetitivo (*Dra*I) do genoma de *V. unguiculata* com 488 nucleotídeos, rica em AT (Adenina/Timina), a qual hibridizou nas regiões pericentroméricas de todos os cromossomos do feijão-caupi, colocalizando com as bandas C, destacadas por Galasso *et al.* (1993). A referida sequência foi espécie-específica, não sendo detectada por *Southern Blot* em outras espécies de *Vigna* [*V. luteola* (Jacq.) Benth., *V. oblongifolia* A. Rich., *V. ambacensis* Welw. ex Baker e *V. vexillata* (L.) A. Rich.] e de leguminosas (*G. max*, *P. coccineus* L., *Vicia faba* L., *P. sativum* e *Lupinus luteus* L.).

Galasso *et al.* (1997) amplificaram os fragmentos do domínios RT de retrotransposons Ty1-*copia*-like a partir dos genomas de *V. unguiculata*, *V. unguiculata*

subsp. *dekindtiana*, *V. luteola* e *V. vexillata*, visualizando suas respectivas localizações *in situ*. Tal estudo revelou que estes retroelementos representavam uma grande fração do genoma das citadas espécies, estando dispersos uniformemente em todos os cromossomos, com exceção das regiões centroméricas, subteloméricas e das RONS.

## ***Referências Bibliográficas***

---

#### 4. REFERÊNCIAS BIBLIOGRÁFICAS

- Alix K and Heslop-Harrison JS (2004) The diversity of retroelements in diploid and allotetraploid Brassica species. *Plant Mol Biol* 54: 895-909.
- Almeida CCS (2006) Mapeamento físico e análise evolutiva em *Phaseolus vulgaris* L. e *P. lunatus* L., utilizando hibridização *in situ* fluorescente (FISH). Tese de Doutorado, Universidade Federal de Pernambuco. Recife, 104 pp.
- Arumuganathan K and Earle ED (1991) Nuclear DNA content of some important plant species. *Plant Mol Biol Rep* 9: 208-218.
- Barroso GM, Guimarães EF, Ichaso CLF, Costa CG, Peixoto AL and Lima HC (1991) Rubiaceae. In: *Sistemática de Angiospermas do Brasil*. Universidade Federal de Viçosa, Viçosa, pp 189-229.
- Bhargava A and Fuentes FF (2009) Mutational Dynamics of Microsatellites. *Mol Biotech* 44 (3): 250-266.
- Begum R, Alam SS, Menzel G and Schmidt T (2009) Comparative molecular cytogenetics of major repetitive sequence families of three *Dendrobium* species (Orchidaceae) from Bangladesh. *Ann Botany* 104: 863-772.
- Benko-Iseppon AM (2001) Estudos moleculares no caupi e em espécies relacionadas: Avanços e perspectivas. *Embrapa Documentos* 56: 327-332.
- Bennetzen JL (2000) Transposable element contributions to plant gene and genome evolution. *Plant Mol Biol* 42: 251-269.
- Biondo E, Miotto STS and Schifino-Wittmann MT (2005) Citogenética de espécies arbóreas da subfamília Caesalpinioideae - Leguminosae do sul do Brasil. *Ciênc Florestal* 15 (3): 241-248.
- Bonifácio EM (2009) Mapeamento comparativo entre o feijão comum e a fava. Trabalho de conclusão de curso (Ciências Biológicas), Departamento de Biologia, Universidade Federal Rural de Pernambuco. Recife, 64 pp.
- Brasileiro-Vidal AC, Cuadrado A, Brammer SP, Benko-Iseppon AM and Guerra M (2005) Molecular cytogenetic characterization of parental genomes in the partial amphidiploid *Triticum aestivum* x *Thinopyrum ponticum*. *Gen Mol Biol* 28: 308-313.
- Capy P, Bazin C, Higuët D and Langin T (1998) *Dynamics and Evolution of Transposable Elements*. Chapman and Hall, New York.

- Capy P (2005) Classification and nomenclature of retrotransposable elements. *Cytogenet Genome Res* 110: 457-461.
- Chang SB, Yang TJ, Datema E, van Vugt J, Vosman B, Kuipers A, Meznikova M, Szinay D, Lankhorst RK, Jacobsen E and de Jong H (2008) FISH mapping and molecular organization of the major repetitive sequences of tomato. *Chromosome Res* 16: 919–933.
- Chavanne F, Zhang D, Liaud M and Cerff R (1998) Structure and evolution of *Cyclops*: a novel giant retrotransposon of the *Ty3/Gypsy* family highly amplified in pea and other legume species. *Plant Mol Biol* 37: 363–375.
- Cheng X, Zhang D, Cheng Z, Keller B and Ling H (2009) A New Family of *Ty1-copia*-Like Retrotransposons Originated in the Tomato Genome by a Recent Horizontal Transfer Event. *Genetics* 181: 1183–1193.
- Choi HK, Kim D, Uhm T, Limpens E, Lim H, Mun JH, Kalo P, Penmetsa RV, Seres A and Kulikova O (2004) A sequence-based genetic map of *Medicago truncatula* and comparison of marker colinearity with *M. Sativa*. *Genetics* 166:1463-1502.
- Choi IY, Hyten DL, Matukumalli LK, Song Q, Chaky JM, Quigley CV, Chase K, Lark KG, Reiter RS, Yoon MS. *et al.* (2007) A soybean transcript map: gene distribution, haplotype and singlenucleotide polymorphism analysis. *Genetics* 176:685-696.
- Chung G and Singh RJ (2008) Broadening the genetic base of soybean: A multidisciplinary approach. *Critical Rev Plant Scienc* 27: 295-341.
- Clarindo WR, De Carvalho CR and Alves BMG (2007). Mitotic evidence for the tetraploid nature of *Glycine max* provided by high quality karyograms. *Plant Syst Evol* 265:101-107.
- Cuadrado A, Cardoso M and Jouve N (2008) Physical organisation of simple sequence repeats (SSRs) in Triticeae: structural, functional and evolutionary implications. *Cytogenet Genome Res* 120: 210–219.
- Cuadrado A and Jouve N (2007a) Similarities in the chromosomal distribution of AG and AC repeats within and between *Drosophila*, human and barley chromosomes. *Cytogenet Genome Res* 119: 91-99.
- Cuadrado A and Jouve N (2007b) The nonrandom distribution of long clusters of all possible classes of trinucleotide repeats in barley chromosomes. *Chromosome Res* 15: 711-720.

- Cuadrado A and Schwarzacher T (1998) The chromosomal organization of simple sequence repeats in wheat and rye genomes. *Chromosoma* 107: 587-594.
- Delgado-Salinas A, Bibler R and Lavin M (2006) Phylogeny of the genus *Phaseolus* (Leguminosae): A recent diversification in an ancient landscape. *Syst Bot* 31: 779–791.
- Dimitri P, Caizzi R, Giordano E, Accardo MC, Lattanzi G and Biamonti G (2009) Constitutive heterochromatin: a surprising variety of expressed sequences. *Chromosoma* 118: 419–435.
- Dixit A, Ma KH, Yu JW, Cho EG and Park YJ (2006) Reverse transcriptase domain sequences from Mungbean (*Vigna radiata*) LTR retrotransposons: Sequence characterization and phylogenetic analysis. *Plant Cell Rep* 25: 100–111.
- Du J, Tian Z, Hans CS, Laten HM, Cannon SB, Jackson SA, Shoemaker RC and Ma J (2010) Evolutionary conservation, diversity and specificity of LTR-retrotransposons in flowering plants: insights from genome-wide analysis and multi-specific comparison. *Plant J* 63: 584-598.
- Ehlers JD and Hall AE (1997) Cowpea (*Vigna unguiculata* L. Walp.). *Field Crops Res* 53: 187-204.
- Eickbush TH and Eickbush DG (2007) Finely orchestrated movements: evolution of the ribosomal RNA genes. *Genetics* 175: 477–485.
- Elgin SCR and Grewal SIS (2005) Heterochromatin: silence in golden. *Current Biology* 13: 895-898.
- Ellegren H (2004) Microsatellites: simple sequences with complex evolution. *Nature Rev Genet* 5: 435-445.
- Findley SD, Cannon S, Varala K, Du J, Ma J, Hudson ME, Birchler JA and Stacey G (2010) A fluorescence in situ hybridization system for karyotyping soybean. *Genetics* 185: 727-744.
- Freire-Filho FR (1988) Origem, evolução e domesticação do caupi [*Vigna unguiculata* (L.) Walp.]. In: Araujo JPP de and Watt EE (eds) O caupi no Brasil. Org. Embrapa-CNPAP: Goiânia, pp 25-46.
- Freire-Filho FR, Ribeiro VQ, Barreto PD and Santos CAF (1999) Melhoramento genético do caupi (*Vigna unguiculata* (L.) Walp.) na região nordeste. In: Queirós MA de, Goedert CO and Ramos SRR (eds) Recursos genéticos e melhoramento de plantas para o Nordeste Brasileiro. EMBRAPA CPATSA: Petrolina.

- Freire-Filho FR, Ribeiro VQ, Barreto PD and Santos CAF (2005) Melhoramento genético. In: Freire-Filho FR, Lima JAA and Ribeiro VQ (eds) Feijão-caupi, avanços tecnológicos. Embrapa Informação Tecnológica: Brasília, pp 28-92.
- Friesen N, Brandes A and Heslop-Harrison JS (2001) Diversity, origin, and distribution of retrotransposons (*gypsy* and *copia*) in conifers. *Mol Biol Evol* 18:1176-1188.
- Fonsêca AFA (2010) Evolução cariotípica no gênero *Phaseolus* L.: mapeamento comparativo entre *P. microcarpus* Mart. e o feijão comum (*P. vulgaris* L.). Dissertação de Mestrado, Universidade Federal de Pernambuco, Recife, 69p.
- Fonsêca A, Ferreira J, Santos TRB, Mosiolek M, Bellucci E, Kami J, Gepts P, Geffroy V, Schweizer D, Santos KGB. *et al.* (2010) Cytogenetic map of common bean (*Phaseolus vulgaris* L.). *Chromosome Research* 18:487-502.
- Fuchs J, Kuhne M and Schubert I (1998) Assignment of linkage groups to pea chromosomes after karyotyping and gene mapping by fluorescent *in situ* hybridization. *Chromosoma* 107: 272-276.
- Gaeta ML, Yuyama PM, Sartori D, Fungaro MHP and Vanzela ALL (2010) Occurrence and chromosome distribution of retroelements and NUPT sequences in *Copaifera langsdorffii* Desf. (Caesalpinioideae). *Chromosome Res* 18: 515-524.
- Galasso I, Harrison GE, Pignone D, Brandes A and Heslop-Harrison JS (1997) The distribution and organization of *Ty1-copia*-like retrotransposable elements in the genome of *Vigna unguiculata* (L.) Walp. (Cowpea) and its relatives. *Ann Bot* 80: 327–333.
- Galasso I, Pignone D and Perrino P (1992) Cytotaxonomic studies in *Vigna* I. General technique and *Vigna unguiculata* C-banding. *Caryologia* 45: 155-161.
- Galasso I, Pignone D and Perrino P (1993) Cytotaxonomic studies in *Vigna*. II. Heterochromatin characterization in *Vigna unguiculata* and three related wild species. *Caryologia* 46: 275-282.
- Galasso I, Schmidt T, Pignone D and Heslop-Harrison JS (1995) The molecular cytogenetics of *Vigna unguiculata* (L.) Walp: the physical organization and characterization of 18S-5,8S-25S rRNA genes, 5S rRNA genes, telomere-like sequences, and a family of centromeric repetitive DNA sequences. *Theor Appl Genet* 91: 928-935.
- Ghafoo A, Sharif A, Ahmad Z, Zahid MA and Rabbani MA (2001) Genetic diversity in blackgram (*Vigna mungo* L. Hepper). *Field Crop Res* 69: 183-190.

- Gartenberg M (2009) Heterochromatin and the cohesion of sister chromatids. *Chromosome Res* 17: 229-238.
- Gepts P (1998) Origin and evolution of common bean: past events and recent trends. *Hort Science* 33:1124-1130.
- Gill N, Findley S, Walling JG, Hans C, Ma J, Doyle J, Stacey G and Jackson SA (2009) Molecular and Chromosomal Evidence for Allopolyploidy in Soybean. *Plant Physiol* 151: 1167–1174.
- Gortner G, Nenno M, Weising K, Zink D, Nagl W and Kahl G (1998) Chromosomal localization and distribution of simple sequence repeats and the Arabidopsis-type telomere sequence in the genome of *Cicer arietinum* L. *Chromosome Res* 6: 97-104.
- Graham PH and Vance CP (2003) Legumes: importance and constraints to greater use. *Plant Physiol* 131: 872-877.
- Guerra M (2004) Hibridização *in situ*: Princípios Básicos. In: Guerra M (ed) FISH: Conceitos e Aplicações na Citogenética. Sociedade Brasileira de Genética: Ribeirão Preto, pp. 1-32.
- Guerra M, Kenton A and Bennett MD (1996) rDNA Sites in Mitotic and Polytene Chromosomes of *Vigna unguiculata* (L.) Walp. and *Phaseolus coccineus* L. revealed by Fluorescent *in situ* Hybridization. *Ann Bot* 78: 157-161.
- Guo J, Wang Y, Song C, Zhou J, Qiu L, Huang H and Wang Y (2010) A single origin and moderate bottleneck during domestication of soybean (*Glycine max*): implications from microsatellites and nucleotide sequences. *Ann Bot* 106 (3): 505-514.
- Hansen CN and Heslop-Harrison JS (2004) Sequences and phylogenies of plant pararetroviruses, viruses and transposable elements. *Adv Bot Res* 41:165-193.
- Hasterok R, Wolny E, Hosiawa M, Kowalczyk M, Kulak-Ksiazczyk S, Ksiazczyk T, Heneen WK and Maluszynska J (2006) Comparative analysis of rDNA distribution in chromosomes of various species of Brassicaceae. *Ann Bot* 97: 205-216.
- Henikoff S (2000) Heterochromatin function in complex genomes. *Bioch et Biophys Acta* 1470: 01-08.
- Heslop-Harrison JS (2000) Comparative genome organization in plants: from sequence and markers to chromatin and chromosomes. *Plant Cell* 12: 617-635.
- Hill P, Burford D, Martin DM and Flavell AJ (2005) Retrotransposon populations of *Vicia* species with varying genome size. *Mol Gen Genomics* 273: 371–381.

- Hirochika H, Okamoto H and Kakutani T (2000) Silencing of retrotransposons in *Arabidopsis* and reactivation by the *ddm1* mutation. *Plant Cell* 12: 357-368.
- Hyten DL, Cannon SB, Song Q, Weeks N, Fickus EW, Shoemaker RC, Specht JE, Farmer AD, May GD and Cregan PB (2010a) High-throughput SNP discovery through deep resequencing of a reduced representation library to anchor and orient scaffolds in the soybean whole genome sequence. *BMC Genomics* 11: 38-45.
- Hyten DL, Choi I-Y, Song Q, Specht JE, Carter TE, Shoemaker RC, Hwang E-Y, Matukumalli LK and Cregan P (2010b) A high density integrated genetic linkage map of soybean and the development of a 1,536 Universal Soy Linkage Panel for QTL mapping. *Crop Science* 50: 960-968.
- Jiang J, Birchler JA, Parrott WA and Dawe RKA (2003) A molecular view of plant centromeres. *Trends Plant Science* 8: 570–575.
- Joly AB (2002) *Introdução à Taxonomia Vegetal*. Editora Nacional, São Paulo, Brasil. pp. 777.
- Jong JH, Fransz P and Zabel P (1999) High resolution FISH in plants – techniques and applications. *Trends Plant Science* 4: 258-263.
- Judd WS, Campbell CS, Kellogg EA, Stevens PF and Donoghue MJ (2002) *Plant Systematics: a phylogenetic approach*. 2nd edition. USA. pp. 576.
- Kangfu Y, Park SJ and Poysa V (1999) Abundance and variation of microsatellite DNA sequences in beans (*Phaseolus* and *Vigna*). *Genome* 42: 27-34.
- Karpechenko GD (1925) On the chromosomes of the Phaseolinae. *Bull Appl Botany Plant Breed* 14: 143-148.
- Kubis S, Schmidt T and Heslop-Harrison JS (1998) Repetitive DNA elements as a major component of plants genomes. *Ann Bot* 82: 45-55.
- Kumar A and Bennetzen JL (1999) Plant retrotransposons. *Annu Rev Genet* 33: 479-532.
- Lewis G, Schrirer B, Mackinder B and Lock M (2005) *Legumes of the world*. 1st edn. London: Kew.
- Li YC, Abraham BK, Fahima T and Nevo E (2004) Microsatellites within genes: Structure, function, and evolution. *Mol Biol Evol* 21: 991-1007.
- Lin JY, Jacobus BH, SanMiguel P, Walling JG, Yuan Y, Shoemaker RC, Young ND and Jackson SA (2005) Pericentromeric Regions of Soybean (*Glycine max* L. Merr.)

- Chromosomes Consist of Retroelements and Tandemly Repeated DNA and Are Structurally and Evolutionarily Labile. *Genetics* 170: 1221–1230.
- Macas J, Neumann P and Navrátilová A (2007) Repetitive DNA in the pea (*Pisum sativum* L.) genome: comprehensive characterization using 454 sequencing and comparison to soybean and *Medicago truncatula*. *BMC Genomics* 8: 427-442.
- Mandáková T and Lysak MA (2008) Chromosomal phylogeny and karyotype evolution in  $x=7$  crucifer species (Brassicaceae). *The Plant Cell* 20: 2559-2570.
- Marcon AB, Barros ICL and Guerra M (2005) Variation in chromosome numbers, CMA bands and 45S rDNA sites in species of *Selaginella* (Pteridophyta). *Ann Bot* 95:271-276.
- Marechal R, Mascherpa JM and Stainier F (1978) Étude taxonomique d'un groupe complexe d'espèces de genres *Phaseolus* et *Vigna* (Papilionaceae) sur la base de données morphologiques et polliniques, traitées par l'analyse informatique. *Boissiera* 28: 1-273.
- Mercado-Ruaro P and Delgado-Salinas A (2000) Cytogenetics studies in *Phaseolus vulgaris* L. (Fabaceae). *Genet Mol Biol* 23: 985-987.
- Mercado-Ruaro P, Delgado-Salinas A and Chiang F (2009) Taxonomic re-assessment of *Phaseolus dasycarpus* (Leguminosae): Systematic position, chromosome studies and re-description. *Brittonia* 61 (1): 8-13.
- Mondin M, Santos-Serejo JA and Aguiar-Perecin MLR (2007) Karyotype characterization of *Crotalaria juncea* (L.) by chromosome banding and physical mapping of 18S-5.8S-26S and 5S rRNA gene sites. *Genet Mol Biol* 30: 65-72.
- Moraes AP, Lemos RR, Brasileiro-Vidal AC, Soares Filho WS and Guerra M (2007) Chromosomal markers distinguish hybrids and non-hybrid accessions of Mandarin. *Cytogenet Genome Res* 119: 275-281.
- Morgante M, Hanafey M and Powell W (2002) Microsatellite are preferentially associated with nonrepetitive DNA in plant genomes. *Nature Genet* 30: 194-200.
- Moscone EA, Klein F, Lambrou M, Fuchs J and Schweizer D (1999). Quantitative karyotyping and dual-color FISH mapping of 5S and 18S-25S rDNA probes in the cultivated *Phaseolus* species (Leguminosae). *Genome* 42:1224-1233.
- Mudge J, Cannon SB, Kalo P, Oldroyd GED, Roe BA, Town CD and Young ND (2005) Highly syntenic regions in the genomes of soybean, *Medicago truncatula*, and *Arabidopsis thaliana*. *BMC Plant Biol* 5: 1-16.

- Natali L, Santini S, Giordani T, Minelli S, Maestrini P, Cionini PG and Cavallini A (2006) Distribution of *Ty3-gypsy*- and *Ty1-copia*-like DNA sequences in the genus *Helianthus* and other Asteraceae. *Genome* 49: 64–72.
- Ng NQ and Marechal R (1985) Cowpea taxonomy, origin and germplasm. In: Singh SR and Rachie KO (eds) *Cowpea research, production and utilization*. John Wiley: Chichester, pp. 11-21.
- Nielen S, Campos-Fonseca F, Guimarães P, Leal-Bertioli SC and Bertioli D (2009) Isolation and Characterization of Retrotransposons in Wild and Cultivated Peanut Species. *Food and Agric Organiz United Nations*: 439-442.
- Nunberg A, Bedell JA, Budiman MA, Citek RW, Clifton SW, Fulton L, Pape D, Cai Z, Joshi T, Nguyen H. *et al.* (2006) Survey sequencing of soybean elucidates the genome structure, composition and identifies novel repeats. *Func Plant Biol* 33: 765-773.
- Ohmido N, Sato S, Tabata S and Fukui K (2007) Chromosome maps of legumes. *Chromosome Res* 15: 97-103.
- Oliveira EJ, Pádua JG, Zucchi MI, Vencovsky R and Vieira MLC (2006) Origin, evolution and genome distribution of microsatellites. *Genet Mol Biol* 29: 294-307.
- Padulosi S, Ng QN and Perrino P (1997) Origin, taxonomy and morphology of *Vigna unguiculata* (L.) Walp. In: Singh BB and Raj M (eds) *Advances in cowpea research*. John Wiley: Chichester, pp. 35-47.
- Palmer RG, Sun H and Zhao LM (2000) Genetics and cytology of chromosome inversions in soybean germplasm. *Crop Science* 40: 683-687.
- Park JM, Schneeweiss GM and Weiss-Schneeweiss H (2007) Diversity and evolution of *Ty1-copia* and *Ty3-gypsy* retroelements in the non-photosynthetic flowering plants *Orobanche* and *Phelipanche* (Orobanchaceae). *Gene* 387: 75–86.
- Pedrosa A, Sandal N, Stougaard J, Schweizer D and Bachmair A (2002) Chromosomal map of the model legume *Lotus japonicus*. *Genetics* 161: 1661-1672.
- Pedrosa-Harand A, Almeida CCS, Mosiolek M, Blair MW, Schweizer D and Guerra M. (2006) Extensive ribosomal DNA amplification during Andean common bean (*Phaseolus vulgaris* L.) evolution. *Theor Appl Genet* 112: 924-933.
- Pedrosa-Harand A and Guerra M (2004) Contribuições da FISH para a Citogenética de Plantas. In: Guerra M (ed). *FISH: Conceitos e Aplicações na Citogenética*. Sociedade Brasileira de Genética: Ribeirão Preto, pp. 33-59.

- Pedrosa-Harand A, Kami J, Geffroy V, Gepts P and Schweizer D (2009) Cytogenetic mapping of common bean chromosomes reveals a less compartmentalized small-genome plant species. *Chromosome Res* 17: 405-417.
- Pignone D, Galasso I and Venora G (1995) Karyomorphological and heterochromatin similarities among four *Vigna* species. *Genet Resources Crop Evol* 42: 57-60.
- Primmer CR, Ellegren H, Saino N and Moller AP (1996) Directional evolution in germline microsatellite mutations. *Nature Genet* 13: 391-393.
- Rajput MJ and Upadhyaya EKC (2009) CARE1, a TY3-gypsy like LTR-retrotransposon in the food legume chickpea (*Cicer arietinum* L.). *Genetica* 136: 429-437.
- Schlueter JA, Goicoechea JL, Collura K, Gill N, Lin J-Y, Yu Y, Kudma D, Zuccolo A, Vallejos CE, Muñoz M. *et al.* (2008) BAC-end sequence analysis and draft physical map of the common bean (*Phaseolus vulgaris* L.) genome. *Tropical Plant Biol* 1:40-48.
- Schmidt T and Heslop-Harrison JS (1998) Genomes, genes and junk: the large-scale organization of plant chromosomes. *Trends Plant Science* 3: 195-199.
- Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J. *et al.* (2010) Genome sequence of the paleopolyploid soybean. *Nature* 463:178-183.
- Silva SMS and Freire-Filho FR (1999) Proteínas de feijão-caupi (*Vigna unguiculata* (L.) Walp.): Caracterização e aplicação nutricional. *Embrapa Documentos* 44: 5-20.
- Simon MV, Benko-Iseppon AM, Resende LV, Winter P and Kahl G (2007) Genetic diversity and phylogenetic relationships in *Vigna* Savi germplasm revealed by NA amplification fingerprinting. *Genome* 50: 538-547.
- Singh NK, Dalal V, Batra K, Singh BK, Chitra G, Singh A, Ghazi I, Yadav M, Pandit A, Dixit R. *et al.* (2007). Single-copy genes define a conserved order between rice and wheat for understanding differences caused by duplication, deletion, and transposition of genes. *Funct Integr Genomics* 7:17-35.
- Singh RJ and Hymowitz T (1999) Soybean genetic resources and crop improvement. *Genome* 42: 605-616.
- Singh RJ, Kim HH and Hymowitz T (2001) Distribution of rDNA loci in the genus *Glycine* Willd. *Theor Appl Genet* 103: 212-218.

- Song QJ, Marek LF, Shoemaker RC, Lark KG, Concibido VC, Delannay X, Specht JE and Cregan PB (2004) A new integrated genetic linkage map of the soybean. *Theor Appl Genet* 109:122-128.
- Souza MGC and Benko-Iseppon AM (2004) Cytogenetics and chromosome banding patterns in Caesalpinioideae and Papilionoideae species of Pará, Amazonas, Brazil. *Bot J Linnean Soc* 144:181-191.
- Souza VC and Lorenzi H (2005) Botânica sistemática: guia ilustrado para identificação das famílias de Angiospermas da flora brasileira, baseado em APG II. Nova Odessa, São Paulo, Brasil. pp 639.
- Steele WM (1976) Cowpeas: *Vigna unguiculata* (Leguminosae-Papilionatae). In: Simonds, NW (ed) *Evolution of Crop Plant*. Longmans: London, pp. 183-185.
- Sun H, Dai H, Zhao G, Ma Y, Ou C, Li H, Li L and Zhang Z (2008) Genome-wide Characterization of Long Terminal Repeat -retrotransposons in Apple Reveals the Differences in Heterogeneity and Copy Number between Ty1-*copia* and Ty3-*gypsy* Retrotransposons. *J Integ Plant Biol* 50: 1130–1139.
- Swaminathan K, Varala K and Hudson ME (2007) Global repeat discovery and estimation of genomic copy number in a large, complex genome using a high-throughput 454 sequence survey. *BMC Genomics* 8: 132-144.
- Todorovska E (2007) Retrotransposons and their role in plant – genome evolution. *Biotechnol. & Biotechnol. EQ* 21: 294–305.
- Tóth G, Gáspári Z and Jurka J (2000) Microsatellites in different eukaryotic genomes: Survey and analysis. *Genome Res* 10: 967-981.
- Vasconcelos S, Souza AA, Gusmão, CLS, Milani M, Benko-Iseppon AM and Brasileiro-Vidal AC (2010) Heterochromatin and rDNA 5S and 45S sites as reliable cytogenetic markers for Castor bean (*Ricinus communis*, Euphorbiaceae). *Micron* 41(7): 746-753.
- Venora G, Blangiforti S and Cremonini R (1999) Karyotype analysis of twelve species belonging to genus *Vigna*. *Cytologia* 64:117-127.
- Venora G and Padulosi S (1997) Karyotypic analysis of wild taxa of *V. unguiculata* (L.) Walpers. *Caryologia* 50:125-138.
- Venora G and Saccardo F (1993) Mitotic karyotype analysis in the *Vigna* genus by means of an image analyser. *Caryologia* 46: 139-149.

- Vershinin AV, Druka A, Alkhimova AG, Kleinhofs A and Heslop-Harrison JS (2002) LINEs and gypsy-like retrotransposons in *Hordeum* species. *Plant Mol Biol* 49: 1–14.
- Vieira C, Borem A and Ramalho MAP (1999) Melhoramento do feijão. In: Borem A (ed.) *Melhoramento de Espécies Cultivadas*. UFV: Viçosa, pp. 273-349.
- Wander AL (2005). Cultivo do feijão irrigado na região noroeste de Minas Gerais. EMBRAPA. Dez, 2005, n° 5 [Cited 04/May/2005]. Available from World Wide Web: <http://www.embrapa.br>.
- Weider LJ, Elser JJ, Crease TJ, Mateos M and Cotner JB (2005) The functional significance of ribosomal (r)DNA variation: impacts on the evolutionary ecology of organisms. *Annual Review of Ecology, Evolution, and Systematics* 36:219–242, 2005.
- Weising K, Nybom H, Wolff K and Kahl G (2005) DNA fingerprinting in plants: Principles, Methods, and Applications. 2nd edition. CRC Press, Boca Raton. pp.444.
- Wessler SR (2006) Transposable elements and the evolution of eukaryotic genomes. *PNAS* 47: 17600 - 17601
- Winterfeld G and Roser M (2007) Chromosomal localization and evolution of satellite DNAs and heterochromatin in grasses (Poaceae), especially tribe Aveneae. *Plant Syst Evol* 264: 75-100.
- Xiao W, Su Y, Sakamoto W and Sodmergen (2007) Isolation and characterization of Ty1-copia-like retrotransposons in mung bean (*Vigna radiata*). *J Plant Research* 120: 323-328.
- Yang TJ, Lee S, Chang SB, Yu Y, de Jong JH and Wing RA (2005) In depth sequence analysis of the centromeric region of tomato chromosome 12 centromeric region: identification of a large CAA block and characterization of centromeric retrotransposons. *Chromosoma* 114: 103- 117.
- Zheng J, Irifune K, Hirai K, Nakata M, Tanaka R and Morikawa H (1994) In situ Hybridization to metaphase chromosomes in six species of *Phaseolus* and *Vigna* using ribosomal DNA as the probe. *J Plant Res* 107: 365-369.
- Zheng JY, Nakata M, Uchiyama H, Morikawa H and Tanaka Y (1991) Giemsa C-banding patterns in several species of *Phaseolus* L. and *Vigna* Savi, Fabaceae. *Cytologia* 56: 459-466.

Zhu H, Choi HK, Cook AR and Shoemaker RC (2005) Bridging model and crop legumes through comparative genomics. *Plant Physiol* 137:1189-1196.

Zou JJ, Singh RJ, Lee J, Xu SJ, Cregan PB, Hymowitz T (2003) Assignment of molecular linkage groups to soybean chromosomes by primary trisomics. *Theor Appl Genet* 107: 745-750.

## ***Capítulo I***

---

### **Microssatélites: um importante componente do genoma de leguminosas**

**Artigo a ser submetido à revista *Annals of Botany***

---

**Type of article: Original Article**

**Microssatélites: um importante componente do genoma de leguminosas**

**Kyria Cilene de A. Bortoleti<sup>1,2</sup>, Ana Maria Benko-Iseppon<sup>2</sup>, Luís Carlos B. da Silva<sup>2</sup>, Ana Rafaela da S. Oliveira<sup>2</sup>, Ricardo V. Abdelnoor<sup>3</sup>, Ivanildo Ramalho<sup>4</sup>, Nataniel Franklin de Melo<sup>5</sup> and Ana Christina Brasileiro-Vidal<sup>2\*</sup>**

<sup>1</sup> Universidade Federal do Vale do São Francisco, Campus Ciências Agrárias, Rodovia BR 407, km 12 – Lote 543 – Projeto de Irrigação Senador Nilo Coelho, s/nº “C1”, CEP: 56304-917, Petrolina – PE, Brasil.

<sup>2</sup> Universidade Federal de Pernambuco, Departamento de Genética, Laboratório de Genética e Biotecnologia Vegetal, Av. da Engenharia, s/nº, CEP 50740-600, Recife – PE, Brasil.

<sup>3</sup> Embrapa Soja, Rod. Carlos João Strass, Distrito e Warta, Caixa Postal 231, CEP 86001-970, Londrina/PR, Brasil.

<sup>4</sup> Universidade Federal Rural de Pernambuco, Rua Dom Manoel de Medeiros, s/nº, Dois Irmãos, CEP: 52171-900, Recife/PE, Brasil

<sup>5</sup> Embrapa Semiárido, BR 428 Km 152, Zona Rural – Caixa Postal 23, CEP 56302-970 Petrolina/PE, Brasil

\* Autor para correspondência: brasileiro\_vidal@hotmail.com

## RESUMO

- **Introdução e Objetivos** Os microssatélites consistem em unidades de repetição (1 a 6 pb) distribuídas em tandem, organizadas de modo disperso ao longo dos genomas e/ou associadas a regiões heterocromáticas. O presente trabalho caracterizou a distribuição cromossômica e genômica de oligonucleotídeos com padrão de microssatélites, juntamente com DNAr 5S/45S e coloração com CMA<sub>3</sub>/DAPI, em alguns representantes de fabáceas visando elucidar a estrutura cromossômica e a organização genômica dos microssatélites nesse grupo taxonômico.
- **Métodos** Os oligonucleotídeos (AG)<sub>8</sub>, (AAG)<sub>5</sub>, (ACC)<sub>5</sub>, (CTC)<sub>5</sub> e (TGA)<sub>6</sub> foram localizados ao longo dos cromossomos de *Glycine max*, *Phaseolus vulgaris*, *P. lunatus*, *Vigna unguiculata* e *V. radiata* mediante o uso de FISH. Em soja, esta análise estendeu-se a uma investigação genômica, a partir de sequências disponibilizadas no banco de dados SoyBase (<http://www.soybase.org/>) usando ferramentas de bioinformática.
- **Resultados-chave** Análises de FISH com os oligonucleotídeos evidenciaram uma grande proporção dos microssatélites analisados nos genomas de fabáceas. No entanto, foram observadas variações em suas quantidades, organização (em tandem ou dispersa) e distribuição ao longo dos cromossomos e/ou genomas. Na análise genômica de soja, observou-se um número variável de sítios de repetições para os oligonucleotídeos, de tamanhos variáveis entre 30 a 454 pb, localizados, principalmente, em regiões de alta a moderada densidade gênica, por vezes associados a genes e elementos transponíveis. A associação de FISH e análise *in silico* ressaltou uma distribuição diferencial aparentemente não aleatória das diferentes classes de SSRs analisadas, podendo estar preferencialmente associadas à heterocromatina ou à eucromatina, com uma densidade de repetições variável.
- **Conclusões** O conhecimento do padrão de distribuição aparentemente não aleatória dos microssatélites analisados, principalmente de *G. max*, aponta-os como importantes componentes estruturais dos genomas em espécies do clado Phaseoloids (tribo Phaseoleae), bem como ressalta o poder de complementação resultante da integração da análise *in situ* e *in silico*, facilitando uma melhor compreensão da estrutura física e organização genômica, principalmente de regiões ricas em DNA repetitivo.

Palavras-chave: *Glycine max*, *Phaseolus vulgaris*, *Phaseolus lunatus*, *Vigna unguiculata*, *Vigna radiata*, Microssatélites, FISH.

## ABSTRACT

- **Background and Aims** Microsatellites consist of tandemly repeated units (1-6 bp), highly dispersed, distributed within the genomes and/or associated with heterochromatic regions. The present study analyzed the chromosomal and genomic distribution of oligonucleotides, together with rDNA 5S/45S and CMA<sub>3</sub>/DAPI staining in some representatives of Fabaceae to elucidate the chromosomal structure and genomic organization of microsatellites in this taxonomic group.
- **Methods** The microsatellites sequences (AG)<sub>8</sub>, (AAG)<sub>5</sub>, (ACC)<sub>5</sub>, (CTC)<sub>5</sub> and (TGA)<sub>6</sub> were distributed along the *Glycine max*, *Phaseolus vulgaris*, *P. lunatus*, *Vigna unguiculata* and *V. radiata* chromosomes by using FISH. In soybean, this analysis was extended by a genomic investigation for sequences available in the SoyBase database (<http://www.soybase.org>), using bioinformatic tools.
- **Key Results** Microsatellites sequences were highly distributed within the genomes of the Fabaceae. However, there were variations in their quantities, and organization (in tandem or dispersed) along the genomes. Genomic investigation in soybean allowed the visualization of a variable number of oligonucleotide repetitions sites with sizes ranging from 30 to 454 bp located, mainly, in regions of high to moderate gene density, sometimes associated to genes and transposable elements. FISH and *in silico* analysis highlighted a differential and apparently non-random distribution of different classes of the analyzed SSRs, which may be preferentially associated to heterochromatin or euchromatin, with a variable density of repetitions.
- **Conclusions** The patterns of microsatellite non-random distribution, mainly *G. max*, showed that these types of sequences were very important structural components for the phaseoloids (tribe Phaseoleae) clade evolution and emphasizes the power of complementation which may result from the combination of *in situ* and *in silico* analysis, facilitating a better understanding of the physical structure and genomic organization, mainly regarding repetitive DNA rich regions.

Keywords: *Glycine max*, *Phaseolus vulgaris*, *Phaseolus lunatus*, *Vigna unguiculata*, *Vigna radiata*, Microsatellite, FISH.

## INTRODUÇÃO

A família Fabaceae compreende um grupo taxonômico de distribuição cosmopolita, sendo a terceira maior família das Magnoliophyta. Apresenta-se subdividida em três subfamílias Papilionoideae, Mimosoideae e Caesalpinioideae (Biondo *et al.*, 2005), abrangendo 727 gêneros e aproximadamente 19.000 espécies (Lewis *et al.*, 2005). Economicamente, a subfamília Papilionoideae abrange a maioria das leguminosas cultivadas, como alfafa, amendoim, ervilha, feijão comum, feijão-caupi, grão-de-bico e soja, representando cerca de 27% da produção de plantas cultivadas mundialmente (Graham e Vance, 2003).

Os gêneros *Glycine* Willd., *Phaseolus* L. e *Vigna* Savi pertencem à tribo Phaseoleae, sendo agrupados em um clado monofilético (Phaseoloids) (Marechal *et al.*, 1978). Alguns representantes destes gêneros apresentam grande importância econômica, destacando-se as espécies *Glycine max* (L.) Merr., *Phaseolus lunatus* (L.), *Phaseolus vulgaris* (L.) e *Vigna unguiculata* (L.) Walp, conhecidas popularmente como soja, feijão-fava, feijão comum e feijão-caupi, respectivamente. Os grãos de feijão representam uma importante fonte proteica na dieta humana nos países em desenvolvimento das regiões tropicais e subtropicais (Wander, 2005). Por sua vez, a soja, além da importância proteica, é uma das principais fontes de óleo vegetal do mundo e a primeira leguminosa a ser completamente sequenciada (Schmultz *et al.*, 2010).

Cerca de 85% do genoma de *G. max*, estimado em 1,1 Gb, foi sequenciado por *shotgun*, sendo as sequências genômicas arranjadas em 20 pseudomoléculas cromossômicas, contendo 397 *scaffolds* com posições bem ordenadas. No entanto, o conteúdo restante corresponde às sequências *scaffolds* não ancoradas, as quais são constituídas por DNA repetitivo (Schmultz *et al.*, 2010), cuja posição tem sido determinada por mapas genéticos de SNPs (*Single Nucleotide Polymorphism*; Polimorfismos de um Único Nucleotídeo) e SSRs (*Simple Sequence Repeats*; Repetições de Sequências Únicas) (Song *et al.*, 2004; Choi *et al.*, 2007; Hyten *et al.*, 2010a, b).

As famílias de sequências repetitivas consistem no principal componente dos genomas vegetais (Schmidt e Heslop-Harrison, 1998; Heslop-Harrison, 2000), podendo constituir cerca de 70-90% em algumas espécies (Dimitri *et al.*, 2009), o que ressalta uma relação entre o conteúdo de DNA repetitivo e o tamanho do genoma (Tóth *et al.*,

2000). A função biológica exercida pelas citadas sequências tem sido extensivamente discutida. Por muito tempo, elas foram remetidas como *junk DNA* (DNA lixo) devido à limitada densidade gênica e baixo índice de recombinação (Jong *et al.*, 1999). Entretanto, existem várias evidências que sugerem a influência desempenhada pelas sequências repetitivas em diferentes mecanismos do funcionamento celular (Dimitri *et al.*, 2009; Gartenberg, 2009), tornando o conceito anterior obsoleto.

As SSRs, também conhecidas como microssatélites, consistem de pequenas unidades de repetições (1 a 6 pb) distribuídas em tandem, onipresentes ao longo dos genomas, sendo encontradas entre regiões codificantes de genes estruturais, entre outras sequências repetitivas, ou associadas a regiões heterocromáticas (Heslop-Harrison, 2000; Cuadrado e Jouve, 2010). Vários autores relatam os altos níveis de polimorfismos de sequências e número de cópias dos microssatélites presentes em diferentes espécies, citando a replicação por *slippage* e a recombinação desigual como fatores responsáveis pela origem e instabilidade dessas sequências (Kubis *et al.*, 1998; Bhargava e Fuentes, 2009).

Conseqüentemente, a rápida evolução das SSRs pode resultar em uma distribuição genoma-específica, espécie-específica, bem como cromossomo-específica (Begum *et al.*, 2009). A frequência e distribuição de diferentes motivos SSRs têm sido alvo de estudos, principalmente em alguns genomas parcial ou completamente sequenciados, como em *P. vulgaris* (Schlueter *et al.*, 2008) e *G. max* (Hyten *et al.*, 2010a), auxiliando no entendimento da organização genômica de diferentes espécies. Concomitantemente, alguns estudos têm visado ao entendimento da distribuição e organização cromossômica de SSRs mediante FISH (*Fluorescent In Situ Hybridization*; Hibridização *In Situ* Fluorescente), sendo úteis na investigação de relações evolutivas (Cuadrado e Jouve, 2007a), identificação cromossômica (Chang *et al.*, 2008) e construção de mapas cromossômicos (Ohmido *et al.*, 2007).

O presente trabalho caracterizou a distribuição cromossômica de bandas CMA<sub>3</sub><sup>+</sup>, DNAr e oligonucleotídeos com padrão de SSRs em alguns representantes de Phaseoleae, visando à elucidação da estrutura e organização cromossômica dessas sequências nesse grupo taxonômico. Em *G. max*, esta análise estendeu-se a uma investigação genômica a partir de sequências disponibilizadas no banco de dados SoyBase (<http://www.soybase.org/>), no intuito de compreender a evolução genômica dos microssatélites. Os resultados obtidos auxiliaram em um melhor entendimento das propriedades estruturais e funcionais das SSRs como importante constituinte genômico.

## MATERIAL E MÉTODOS

### *Análise citogenética*

#### **Material vegetal**

Sementes das espécies *Glycine max* cv. Conquista, *P. vulgaris* cv. BRS Esplendor, *P. lunatus* cv. Vermelhinha, *V. unguiculata* cv. BR14-Mulato e *Vigna radiata* foram obtidas do banco de germoplasma da Embrapa Soja (Empresa Brasileira de Pesquisa Agropecuária, Soja, Londrina), Embrapa-CNPAF (Embrapa Arroz e Feijão, Goiás) para feijão comum e feijão-fava, LGBV/UFPE (Laboratório de Genética e Biotecnologia Vegetal, Universidade Federal de Pernambuco, Recife) e Embrapa-CPATSA (Embrapa, Centro de Pesquisas do Trópico Semiárido, Petrolina), respectivamente.

#### **Pré-tratamento e fixação**

Para análise mitótica, pontas de raízes foram pré-tratadas com uma solução de 8-hidroxiquinoleína (8HQ) por 18 h, 1 h à temperatura ambiente e 17 h a 8 °C sendoposteriormente, fixadas em etanol absoluto:ácido acético (3:1, v/v) ou metanol:ácido acético (3:1, v/v) por 2 a 24 h à temperatura ambiente e estocadas a - 20°C.

#### **Preparação cromossômica**

Para a preparação das lâminas, os meristemas foram lavados duas vezes em água destilada, por 5 min, digeridos em solução enzimática contendo 2% (p/v) celulase 'Onozuka R-10' (Serva) e 20% (v/v) pectinase (Sigma-Aldrich) em tampão citrato 10 mM por 180 min a 37 °C. Em seguida, o material digerido foi lavado em água destilada e incubado em ácido acético 45% por 30 min a 37 °C. Uma ponta de raiz foi utilizada por lâmina, a qual foi macerada em ácido acético 45%, sendo posteriormente, submetida a nitrogênio líquido e secagem ao ar.

Outra metodologia utilizada seguiu o procedimento descrito por Carvalho e Saraiva (1993), com algumas modificações. Após digestão enzimática, as raízes foram lavadas em água destilada e colocadas em uma lâmina ligeiramente inclinada verticalmente. Algumas gotas de fixador (metanol: ácido acético, 3:1) gelado foram pingadas com uma pipeta Pasteur em cima das raízes para a total remoção da água. Em seguida, procedeu-se a maceração mecânica das raízes para posterior gotejamento de

fixador gelado sobre o material. Posteriormente, as lâminas foram ventiladas com um auxílio de bombas de ar para possibilitar a secagem das mesmas, mergulhadas em ácido acético 45% por 12 s e colocadas em uma superfície plana pré-aquecida (37 °C) por aproximadamente 3 min.

Em ambos os casos, as lâminas foram selecionadas mediante coloração com uma solução de 2 µg/mL de DAPI (4'-6'-diamidino-2-fenilindole)/glicerol (1:1, v/v). As preparações selecionadas foram descoradas em etanol absoluto:ácido acético (3:1, v/v) e etanol absoluto por 30 min e 1 h, respectivamente. Após secagem ao ar, as lâminas foram devidamente estocadas a -20 °C, até o momento do procedimento da coloração com os fluorocromos CMA<sub>3</sub> (Cromomicina A<sub>3</sub>)/DAPI e/ou FISH.

### **Coloração CMA<sub>3</sub>/DAPI**

A coloração com os fluorocromos foi realizada de acordo com Schweizer e Ambros (1994), sendo as melhores lâminas armazenadas por três dias, coradas com CMA<sub>3</sub> (0,5 mg/mL) por 1 h e DAPI (2 µg/mL) por 30 min, montadas em tampão McIlvaine-glicerol (1:1) e estocadas em câmara escura, por três dias. Após análise e captura das melhores células, as lâminas foram descoradas em etanol absoluto:ácido acético (3:1, v/v) e etanol absoluto por 30 min e 1 h, respectivamente, sendo estocadas a -20 °C para serem utilizadas na FISH.

### **Obtenção de sondas**

#### *Oligonucleotídeos com padrão SSR*

Sequências completas de nucleotídeos pertencentes às espécies *P. lunatus*, *P. vulgaris*, *V. unguiculata* e *V. radiata* foram obtidas em formato FASTA a partir do NCBI (*National Center for Biotechnology Information*; Centro Nacional para Informação Biotecnológica) (<http://www.ncbi.nlm.nih.gov>).

Posteriormente à anotação, os dados foram submetidos ao software SSRIT (*Simple Sequence Repeat Identification Tool*; Ferramenta de Identificação de Repetições de Sequências Únicas) disponível em [www.gramene.org/db/searches/ssrtool](http://www.gramene.org/db/searches/ssrtool) para identificação de sequências de oligonucleotídeos com padrões frequentes em microssatélites.

Cinco oligonucleotídeos (AG)<sub>8</sub>, (ACC)<sub>5</sub>, (AAG)<sub>5</sub>, (CTC)<sub>5</sub> e (TGA)<sub>6</sub> foram selecionados, sintetizados e marcados indiretamente com digoxigenina-11-dUTP, pela

metodologia de *end labeling* (DIG Oligonucleotide 3'-End Lab. Kit, 2nd generation, Roche), conforme as instruções do fabricante.

#### *Sondas de DNAr 5S e 45S*

As sondas R2, um fragmento de 6,5 Kb, contendo a unidade de repetição de DNAr 18S-5,8S-25S, oriunda de *Arabidopsis thaliana* (L.) Heynh. (Wanzenböck *et al.*, 1997), e D2, um fragmento de 400 pb contendo duas unidades de repetição de DNAr 5S, proveniente de *Lotus japonicus* (Regel) K. Larsen (Pedrosa *et al.*, 2002), foram utilizadas com o intuito de gerar sinais associados à marcação com os oligonucleotídeos, auxiliando na identificação cromossômica. As sondas de DNAr 5S e 45S foram marcadas com biotina-11-dUTP (Sigma) e digoxigenina-11-dUTP (Roche), respectivamente, por *nick translation* (Invitrogen).

#### **Hibridização *in situ* fluorescente (FISH)**

As lâminas, estocadas a -20 °C, foram pré-tratadas como descrito por Pedrosa *et al.* (2001). A desnaturação das sondas, os banhos pós-hibridização e a detecção foram efetuados de acordo com Heslop-Harrison *et al.* (1991), exceto pela lavagem de estringência a qual foi realizada em 0,1xSSC a 42 °C. As misturas de hibridização consistiram de formamida 50% (v/v), dextran sulfato 10% (p/v), 2x SSC e 2-5 ng/μL de sonda. As lâminas foram desnaturadas por 7 min a 78 °C, bem como hibridizadas por 18 a 48 h a 37 °C. Para soja, os cromossomos foram desnaturados com formamida 70% em 2xSSC a 70 °C por 7 min, sendo posteriormente, desidratados por 5 min em uma série alcoólica (70% e 100%) a -20 °C. As sondas marcadas com biotina foram detectadas com antibiotina obtida em camundongo (DAKO) em combinação com anticamundongo conjugado com TRITC (DAKO) em BSA 1% (p/v). Enquanto as sondas marcadas com digoxigenina foram detectadas usando antidigoxigenina de ovelha conjugado com fluoresceína isotiocianato (FITC; Boehringer Mannheim) e amplificadas com antiovelha conjugado com FITC (Serotec) também em BSA 1% (p/v). Todas as preparações foram montadas com 2 μg/mL de DAPI em Vectashield (Vector).

#### **Análise dos dados**

As células foram analisadas com um microscópio de epifluorescência Leica DMLB e imagens das melhores células foram capturadas com uma câmera Leica DFC 340FX, utilizando o programa CW 4000 da Leica. As imagens foram otimizadas para melhor brilho e contraste com o Adobe Photoshop CS3 (Adobe Systems Incorporated).

Para as imagens sobrepostas DAPI-DNAr 45S, a imagem do DAPI foi transformada para tons de cinza; em alguns casos, o DNAr foi pseudocolorido com amarelo e ambas as imagens foram sobrepostas, usando a ferramenta *lighten*.

O tamanho cromossômico relativo e a relação entre braços cromossômicos de cinco metáfases, coradas com DAPI, foram medidos por meio do programa UTHSCSA ImageTool, para elaboração do idiograma de *V. unguiculata* e *V. radiata*, usando o programa Adobe Illustrator CS3 (Adobe Systems Incorporated). Os cromossomos foram pareados e ordenados em tamanho decrescente, sendo identificados por letras.

### **Análise in silico**

#### **Mapeamento de repetições de oligonucleotídeos em *Glycine max***

As sequências genômicas de *G. max* da versão Glymal foram utilizadas diretamente do servidor web SoyBase (<http://www.soybase.org/>). O genoma da soja completo sem máscara foi investigado usando o programa BLASTn com os seguintes parâmetros: formato de saída de alinhamentos sem lacunas, matriz de comparação blossom62, valor W padrão, *e-value* 0,1 ou menor e filtro de baixa complexidade desligado. Os oligonucleotídeos [(AAG)<sub>5</sub>]<sub>10</sub>, [(ACC)<sub>5</sub>]<sub>10</sub>, [(AG)<sub>8</sub>]<sub>15</sub>, [(CTC)<sub>5</sub>]<sub>10</sub> e [(TGA)<sub>6</sub>]<sub>10</sub> foram usados como sondas, adotando como ponto de corte valores de identidade de pareamento inferior a 77%. Em seguida, foram anotadas as informações sobre o tamanho e posição de cada repetição nos cromossomos, incluindo a informação sobre genes e elementos transponíveis associados contendo as repetições de oligonucleotídeos. Devido à característica repetitiva das sondas utilizadas, o programa de alinhamento BLASTn criou o artifício de janelas deslizantes em regiões contínuas, sendo o tamanho da região em pb (pares de bases) calculado subtraindo a posição inicial da final do alinhamento para cada região. Coeficientes de correlação entre os microssatélites e elementos transponíveis (*Glyma transposable element models*) bem como entre microssatélites e genes foram calculados para cada cromossomo de soja usando a função estatística de correlação de Pearson.

#### **Representação da distribuição de repetições de oligonucleotídeos nas pseudomoléculas cromossômicas de *Glycine max***

Uma representação das pseudomoléculas cromossômicas foi construída para *G. max* mediante o programa Adobe Illustrator CS3 (Adobe Systems Incorporated), sendo utilizadas como parâmetro as medições dos tamanhos disponíveis no servidor web SoyBase

(<http://www.soybase.org/>), as quais variam entre 37,4 a 62,31 Mb. Considerando a relação 1 Mb igual a 1 mm, as repetições dos oligonucleotídeos foram posicionadas individual e longitudinalmente nos cromossomos ou pseudomoléculas cromossômicas, de acordo com as cores da legenda.

## RESULTADOS

### *Distribuição da heterocromatina e dos sítios de DNAr 5S e 45S*

O padrão de distribuição das bandas CMA<sub>3</sub>/DAPI e dos sítios de DNAr 5S e 45S foram mostrados apenas para *G. max*, *V. unguiculata* e *V. radiata* (Figura 1), uma vez que, os cariótipos de *P. lunatus* e *P. vulgaris* encontram-se bem descritos na literatura (Almeida, 2006; Pedrosa-Harand *et al.*, 2006; Fonsêca *et al.*, 2010). Os cromossomos de *V. unguiculata* e *V. radiata* foram identificados por letras (A – K) em ordem decrescente de tamanho (Figura 1J), não havendo relação com seus respectivos grupos de ligação.

Em *G. max* ( $2n = 40$ ), a coloração com os fluorocromos base-específicos CMA<sub>3</sub>/DAPI revelou um padrão de bandas CMA<sub>3</sub><sup>+</sup> (positiva)/DAPI<sup>-</sup> (negativo) nas regiões pericentroméricas de todos cromossomos. Entretanto, o par cromossômico 13 apresentou uma banda CMA<sub>3</sub><sup>+</sup>/DAPI<sup>-</sup> em toda extensão do braço curto, correspondente à marcação DNAr 45S. Adicionalmente, um par de DNAr 5S foi visualizado na região intersticial do cromossomo 19, adjacente à marcação pericentromérica CMA<sub>3</sub><sup>+</sup>/DAPI<sup>-</sup> deste par (Figura 1A-C).

Na espécie *V. unguiculata* ( $2n = 22$ ), foi observada uma elevada quantidade de heterocromatina rica em GC (guanina/citosina), mediante a presença de bandas CMA<sub>3</sub><sup>+</sup>, a qual foi classificada em três tipos CMA<sub>3</sub><sup>+</sup>/DAPI<sup>0</sup> (neutro), CMA<sub>3</sub><sup>+</sup>/DAPI<sup>-</sup> e CMA<sub>3</sub><sup>++</sup>/DAPI<sup>-</sup> (Figura 1D-E, J). As bandas CMA<sub>3</sub><sup>+</sup>/DAPI<sup>0</sup> foram encontradas com maior frequência, estando distribuídas nas regiões pericentroméricas de todo o complemento cromossômico do feijão-caupi. Além das regiões pericentroméricas, os pares cromossômicos B e G apresentaram fortes bandas subteloméricas CMA<sub>3</sub><sup>++</sup>/DAPI<sup>-</sup> colocalizadas com sítios de DNAr 45S observados na FISH (Figura 1F, J). Por outro lado, os cromossomos D, F e J apresentaram sítios de DNAr 45S menores colocalizados com bandas CMA<sub>3</sub><sup>+</sup>/DAPI<sup>-</sup>, distribuídas nas regiões subteloméricas em D e J, e proximal, no cromossomo F. Os sítios de DNAr 5S (CMA<sub>3</sub><sup>0</sup>/DAPI<sup>0</sup>) foram observados

nas regiões subterminais do braço curto dos cromossomos D (portador de DNAr 45S no braço longo) e I (Figura 1F, J).

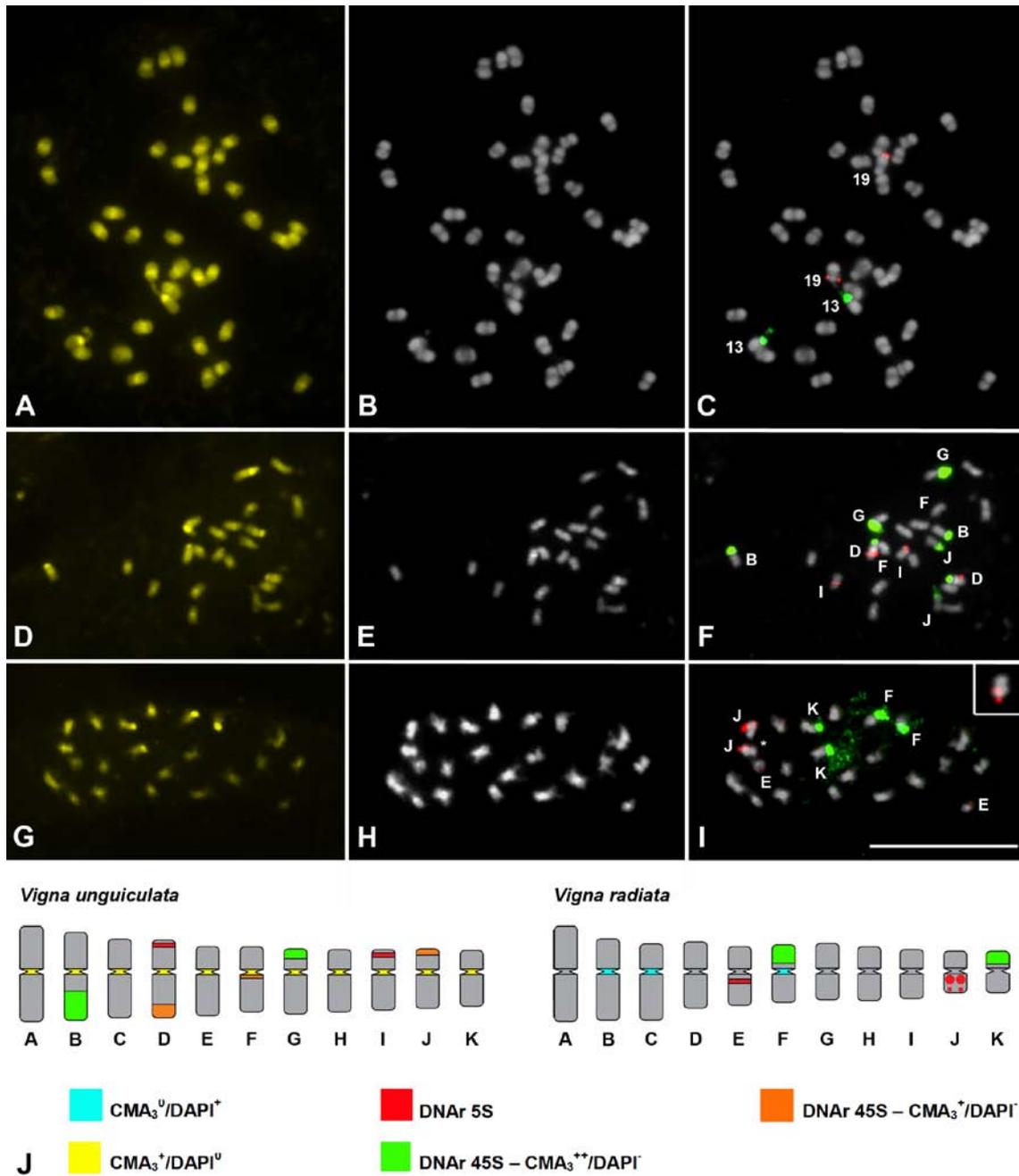
No complemento cromossômico de *V. radiata* ( $2n = 22$ ) foram observadas bandas heterocromáticas  $CMA_3^0/DAPI^+$  e  $CMA_3^{++}/DAPI^-$ , evidenciando regiões ricas em AT (adenina/timina) e GC, respectivamente (Figura 1G-H, J). O primeiro tipo de bandas foi localizado na região pericentromérica dos cromossomos B, C e F, enquanto que o segundo, em posição subtelomérica do braço curto dos pares F e K. Estes dois pares de bandas  $CMA_3^{++}/DAPI^-$  colocalizaram com os sítios de DNAr 45S encontrados para a espécie. *Vigna radiata* apresentou três pares de sítios DNAr 5S, localizados na região proximal do braço longo do cromossomo E, bem como a presença de bandas puntiformes (*dots*) na região proximal e subtelomérica do braço longo do cromossomo J (Figura 1I, J).

#### ***Localização de oligonucleotídeos com padrão SSR por FISH***

A FISH evidenciou distribuição dos microssatélites no complemento cromossômico de todas as espécies analisadas. Em termos gerais, sinais de hibridização foram notados nas regiões pericentroméricas, entretanto, uma marcação dispersa prevaleceu ao longo da maioria dos cromossomos de *G. max*, *P. vulgaris*, *P. lunatus*, *V. unguiculata* e *V. radiata*. De acordo com a composição dos oligonucleotídeos, diferenças foram observadas na abundância e na intensidade dos sítios, inter e intraespecificamente.

Uma ampla distribuição de marcações *in situ* foi observada para oligonucleotídeo  $(AG)_8$ , apresentando sinais com diferentes intensidades predominantemente nas regiões proximais de ambos os braços dos cromossomos de *G. max*, *P. vulgaris*, *P. lunatus*, *V. unguiculata* e *V. radiata* (Figura 2B, D, F, H, J). Entretanto, alguns cromossomos dessas espécies apresentaram sinais dispersos, em menor número e intensidade.

O oligonucleotídeo  $(AAG)_5$  revelou marcações bem evidentes e predominantes nas regiões proximais dos cromossomos de *G. max*, *V. unguiculata* e *V. radiata* (Figuras 3B, H, J). Contudo, sinais dispersos em todos os cromossomos de *P. vulgaris*, *V. unguiculata* e *V. radiata* foram notados (Figuras 3D, H, J). Em *P. lunatus*, marcação intensa foi restrita à região da constrição primária, sendo um possível constituinte do centrômero ou de regiões adjacentes (Figura 3F).

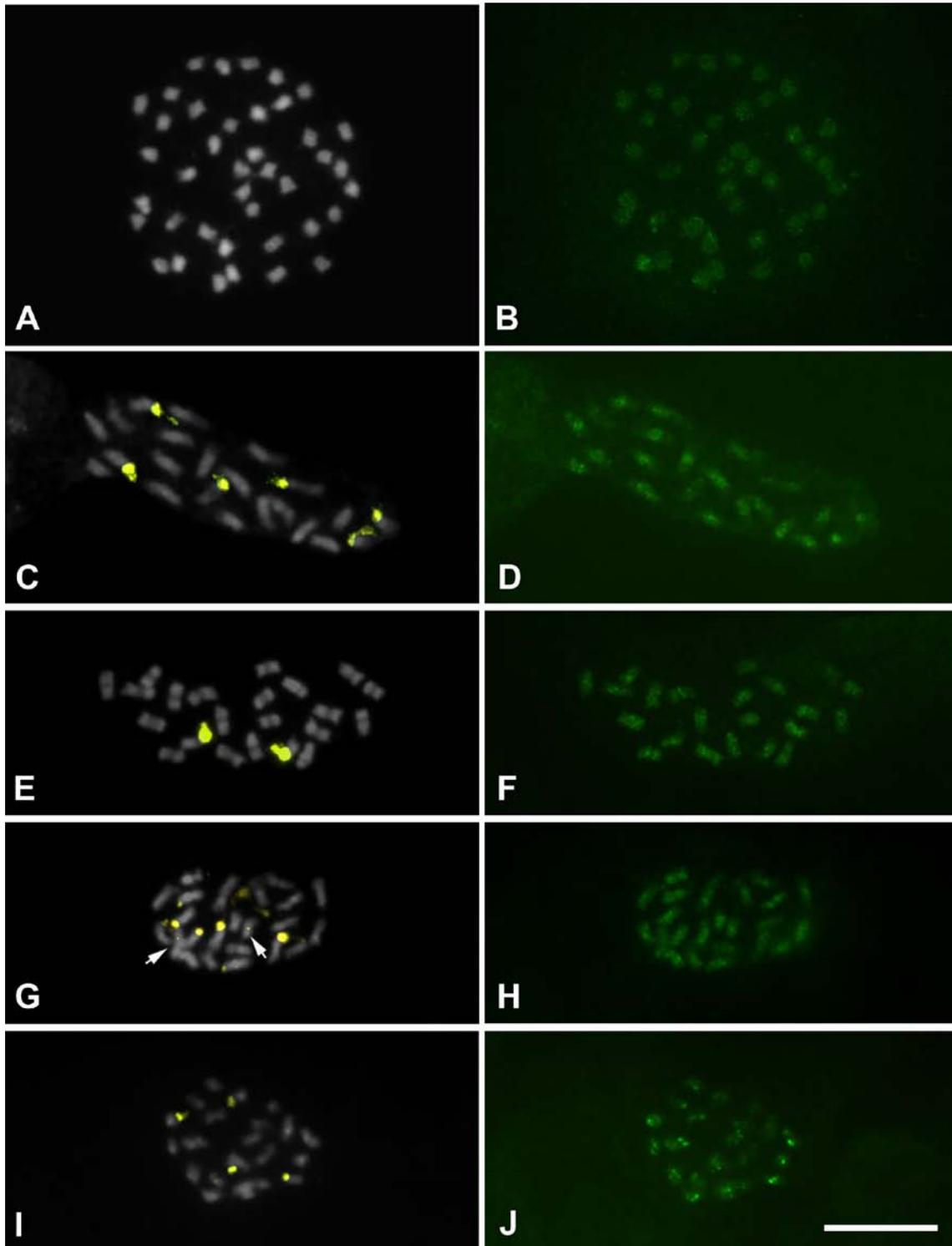


**Figura 1.** Padrão CMA<sub>3</sub> (A, D, G)/ DAPI (B, E, H) e distribuição dos sítios de DNAr 5S e 45S (C, F, I) em *Glycine max* (A-C), *Vigna unguiculata* (D-F, J) e *V. radiata* (G-I, J). Os números e letras indicam os cromossomos identificados pela distribuição dos sítios de DNAr 5S e 45S (vermelho e verde, respectivamente) em C, F e I. Idiogramas mostrando as marcas cromossômicas para *V. unguiculata* e *V. radiata* (J), em ordem decrescente de tamanho. Inserção indica o cromossomo J, portador de DNAr 5S, de *V. radiata* (I). Barra em I = 5 μm.

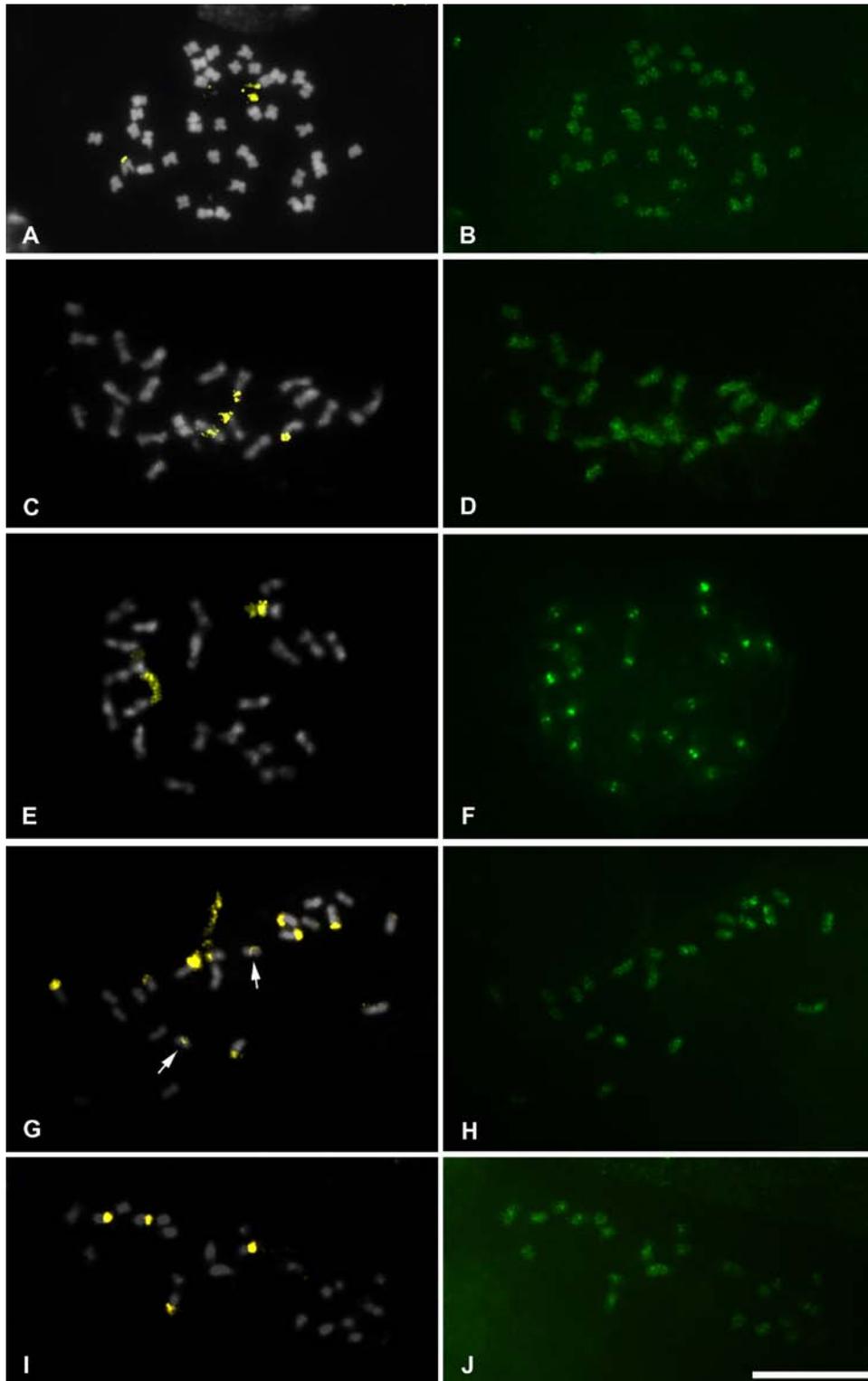
A sonda (ACC)<sub>5</sub> produziu sinais dispersos de diferentes intensidades nos cromossomos das cinco espécies estudadas (Figura 4). *Phaseolus vulgaris* apresentou uma menor distribuição de sítios (Figura 4D), em relação a *P. lunatus*, que apresentou um par cromossômico com marcações em quase sua extensão total (Figura 4F). Em *V. unguiculata*, ao menos, três pares com sítios de maior intensidade foram notados. Dentre estes, destaca-se o par J, portador de DNAr 45S, com sinais (ACC)<sub>5</sub> na região proximal (Figura 4G-H). Além dos sinais dispersos, sítios mais fortes distribuíram-se de modo proximal ou terminal em alguns cromossomos de *G. max* e *V. radiata* (Figuras 4B, J).

As repetições (CTC)<sub>5</sub> mostraram uma maior concentração nas regiões proximais de quase todos os cromossomos de *P. vulgaris* (Figura 5D), de modo semelhante ao evidenciado pelo (AG)<sub>8</sub>. Por sua vez, *P. lunatus* apresentou marcações dispersas em todos os cromossomos, também com sinais mais fortes pericentroméricos (Figura 5F). Em *G. max*, *V. unguiculata* e *V. radiata*, marcações proximais em cinco a seis pares cromossômicos destacaram-se em relação aos sítios dispersos (Figura 5B, H, J).

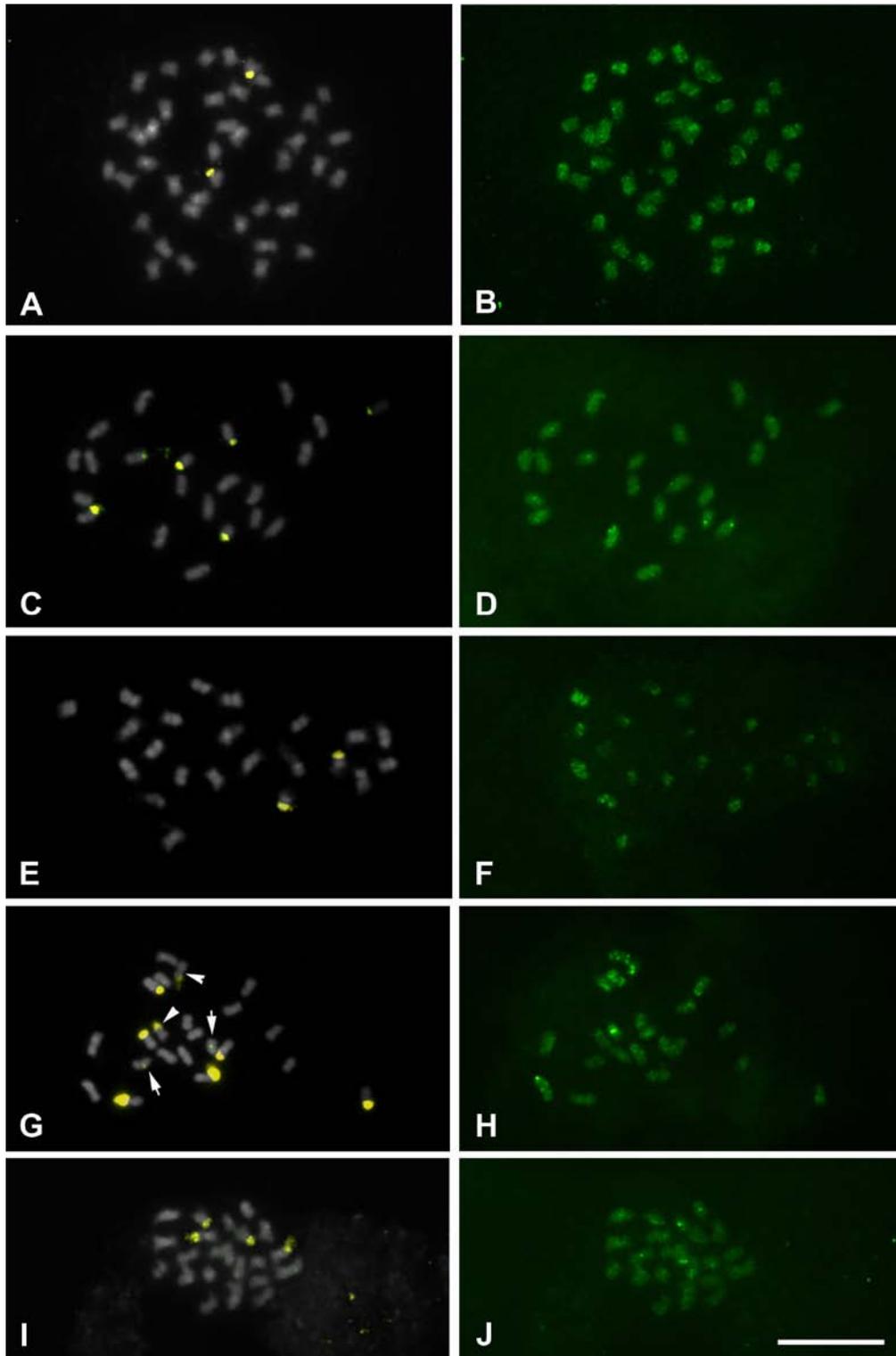
Em *G. max*, as repetições (TGA)<sub>6</sub> mostraram sinais de hibridização dispersos com diferentes intensidades, especialmente nas regiões subterminais de ambos os braços cromossômicos (Figura 6B). *Phaseolus vulgaris* revelou sítios de hibridização distribuídos pericentromericamente em alguns cromossomos, embora sinais dispersos com menor intensidade tenham prevalecido (Figura 6D). Similarmente, *P. lunatus* apresentou marcações dispersas discretas em todo complemento cromossômico, contudo, alguns sinais ligeiramente mais fortes foram evidenciados nas regiões proximais (Figura 6F). Um padrão amplamente disperso e de intensidade mais forte foi característico das espécies *V. unguiculata* e *V. radiata*, as quais apresentaram sinais pericentroméricos mais destacados na maioria dos seus cromossomos (Figura 6H, J).



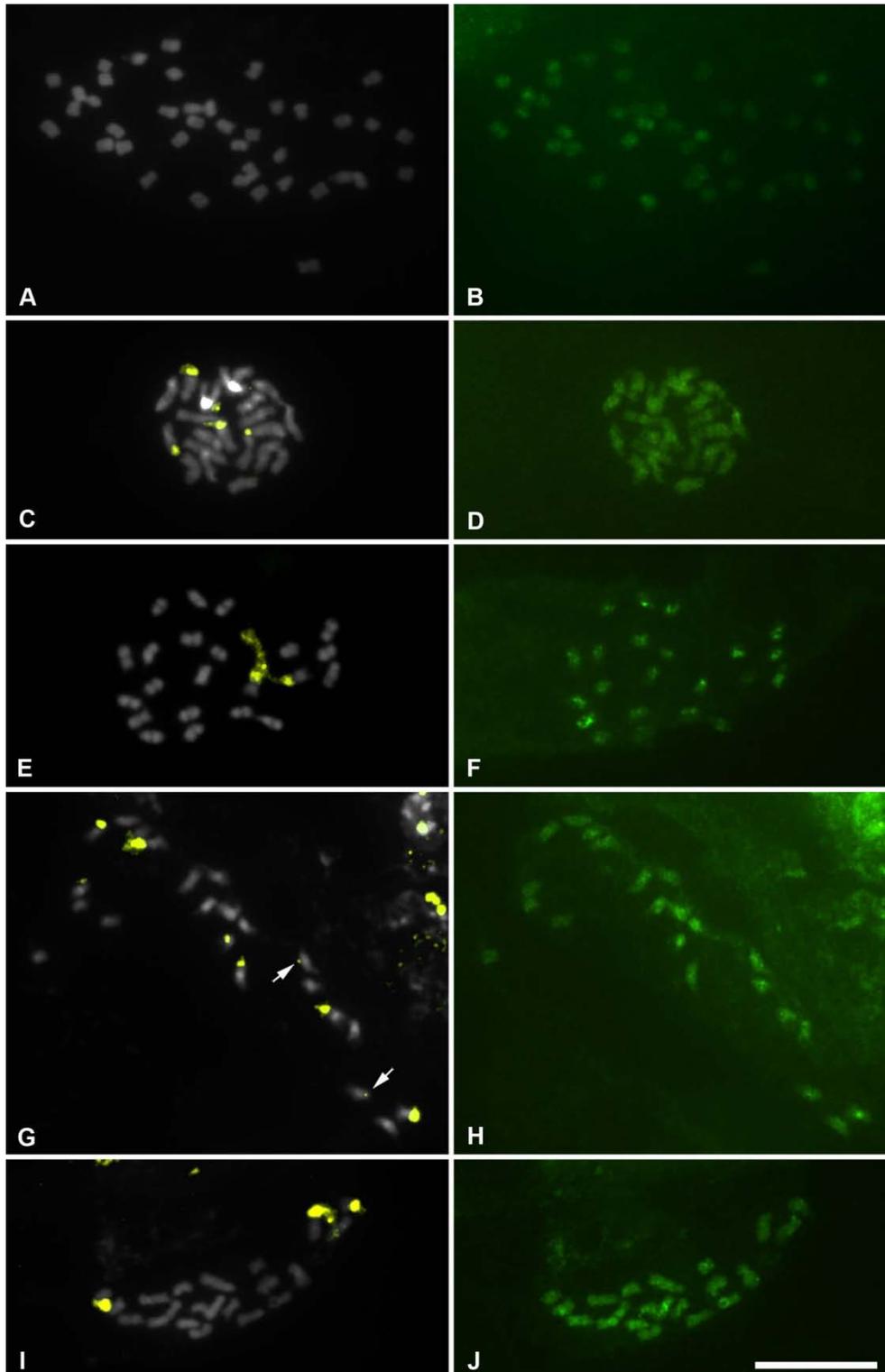
**Figura 2.** Células metafásicas de *Glycine max* (A-B), *Phaseolus vulgaris* (C-D), *P. lunatus* (E-F), *Vigna unguiculata* (G-H) e *V. radiata* (I-J) mostrando a distribuição dos sítios do oligonucleotídeo com padrão SSR  $(AG)_8$  (em verde) e DNAr 45S (pseudocoloridos em amarelo). Os cromossomos foram contracorados com DAPI e pseudocoloridos em cinza (A, C, E, G e I). Setas em G indicam sítios proximais de DNAr 45S em *V. unguiculata*. Barra em J = 5  $\mu$ m.



**Figura 3.** Células metafásicas de *Glycine max* (A-B), *Phaseolus vulgaris* (C-D), *P. lunatus* (E-F), *Vigna unguiculata* (G-H) e *V. radiata* (I-J) mostrando a distribuição dos sítios do oligonucleotídeo com padrão SSR (AAG)<sub>5</sub> (em verde) e DNAr 45S (pseudocoloridos em amarelo). Os cromossomos foram contracolorados com DAPI e pseudocoloridos em cinza (A, C, E, G e I). Setas em G indicam sítios proximais de DNAr 45S em *V. unguiculata*. Barra em J = 5 µm.



**Figura 4.** Metáfases de *Glycine max* (A-B), *Phaseolus vulgaris* (C-D), *P. lunatus* (E-F), *Vigna unguiculata* (G-H) e *V. radiata* (I-J) mostrando a distribuição dos sítios do oligonucleotídeo com padrão SSR (ACC)<sub>5</sub> (em verde) e DNAr 45S (pseudocoloridos em amarelo). Os cromossomos foram contracolorados com DAPI e pseudocoloridos em cinza (A, C, E, G e I). Setas e cabeças de setas em G apontam sítios proximais de DNAr 45S e cromossomo J em *V. unguiculata*, respectivamente. Barra em J = 5 µm.



**Figura 5.** Células metafásicas de *Glycine max* (A-B), *Phaseolus vulgaris* (C-D), *P. lunatus* (E-F), *Vigna unguiculata* (G-H) e *V. radiata* (I-J) mostrando a distribuição dos sítios do oligonucleotídeo com padrão SSR (CTC)<sub>5</sub> (em verde) e DNAr 45S (pseudocoloridos em amarelo). Os cromossomos foram contracolorados com DAPI e pseudocoloridos em cinza (A, C, E, G e I). Setas em G indicam sítios proximais de DNAr 45S. Barra em J = 5  $\mu$ m.

**Organização genômica de microssatélites em *Glycine max***

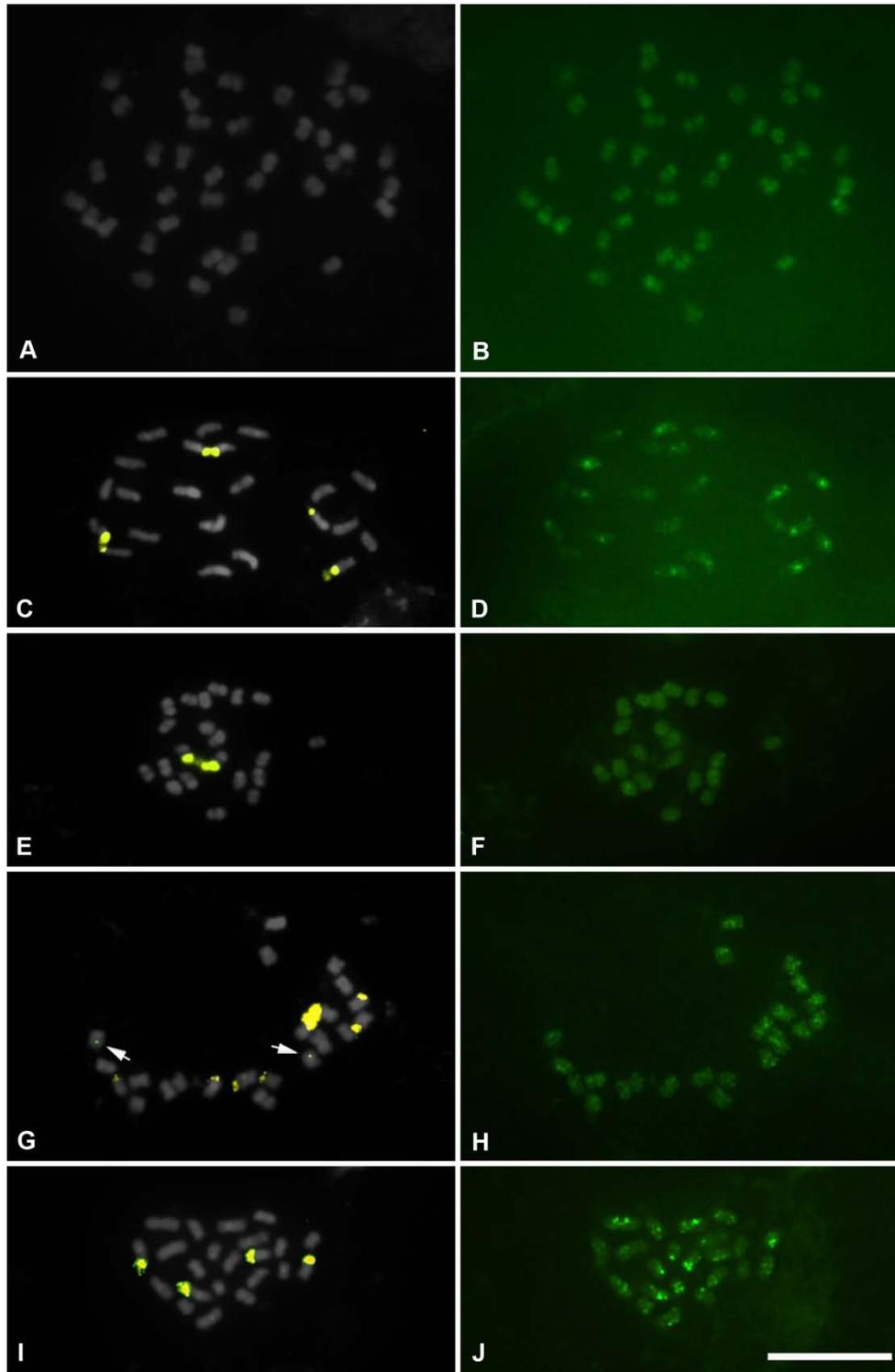
O mapeamento *in silico* de repetições dos oligonucleotídeos com padrão SSR em *G. max* revelou a presença de 92, 136, 83, 144 e 103 sítios de repetições para os oligonucleotídeos [(AG)<sub>8</sub>]<sub>15</sub>, [(AAG)<sub>5</sub>]<sub>10</sub>, [(ACC)<sub>5</sub>]<sub>10</sub>, [(CTC)<sub>5</sub>]<sub>10</sub> e [(TGA)<sub>6</sub>]<sub>10</sub>, respectivamente, com tamanhos variáveis de 30 a 454 pb, localizados, principalmente, em regiões de alta a moderada densidade gênica, por vezes associados a genes e elementos transponíveis (Tabela 1). Para cada sítio, uma a 96 unidades de repetição foram observadas, estando parcial ou totalmente sobrepostas. Contudo, a presença de repetições sobrepostas não estava diretamente relacionada a regiões com diferentes densidades gênicas.

**Tabela 1.** Mapeamento *in silico* de oligonucleotídeos com padrão de SSR no genoma de *Glycine max*, indicando o número de sítios observados, o tamanho mínimo e máximo da sequência (em pb), bem como a densidade gênica da região e o número de genes e/ou elementos transponíveis associados.

Oligos	Número de sítios/genoma	Tamanhos mínimo e máximo (pb) <sup>1</sup>	Localização <i>in silico</i>					
			Densidade gênica <sup>2</sup>				Número de genes associados	Número de elementos transponíveis associados
			A	M	B	N		
[(AG) <sub>8</sub> ] <sub>15</sub>	92	56 – 329	46	46	-	-	11	1
[(AAG) <sub>5</sub> ] <sub>10</sub>	136	44 – 454	44	78	8	6	29	5
[(ACC) <sub>5</sub> ] <sub>10</sub>	83	30 – 173	54	28	1	-	52	4
[(CTC) <sub>5</sub> ] <sub>10</sub>	144	37 – 407	38	92	4	10	15	4
[(TGA) <sub>6</sub> ] <sub>10</sub>	103	40 – 386	55	47	1	-	42	1

<sup>1</sup> O tamanho da região foi calculado subtraindo a posição inicial da posição final do alinhamento em cada região.

<sup>2</sup> A (Alta); M (Moderada); B (Baixa); N (ausência de genes), de acordo com o site [www.soybase.org](http://www.soybase.org), pesquisado em 02/04/2010.



**Figura 6.** Células metafásicas de *Glycine max* (A-B), *Phaseolus vulgaris* (C-D), *P. lunatus* (E-F), *Vigna unguiculata* (G-H) e *V. radiata* (I-J) mostrando a distribuição dos sítios do oligonucleotídeo com padrão SSR (TGA)<sub>6</sub> (em verde) e DNAr 45S (pseudocoloridos em amarelo). Os cromossomos foram contracorados com DAPI e pseudocoloridos em cinza (A, C, E, G e I). Setas em G indicam sítios proximais de DNAr 45S em *V. unguiculata*. Barra em J = 5 µm.

Os 92 sítios de repetições de [(AG)<sub>8</sub>]<sub>15</sub> apresentaram tamanho entre 56 a 329 pb, com 46 marcações distribuídas em regiões de alta densidade gênica, enquanto a outra metade esteve localizada em regiões de moderada densidade gênica. Tais sítios estavam associados a 11 genes e a uma sequência identificada como elemento transponível.

As repetições [(AAG)<sub>5</sub>]<sub>10</sub> variaram quanto a seu tamanho (entre 44 a 454 pb), estando 44, 78 e oito sítios localizados em regiões de alta, moderada e baixa densidade gênica, respectivamente, estando os últimos dispostos nas pseudomoléculas cromossômicas 2, 4, 7, 9 e 11 (Tabela 2). Entretanto, o pseudocromossomo 4 apresentou dois sítios distribuídos em uma distância de 11.004.576 pb, enquanto que o cromossomo 7 mostrou três sítios localizados em uma distância de 2.104.151 e 9.921.026 pb. Os seis sítios restantes estavam dispostos em locais com ausência de genes, nos cromossomos 3, 8, 11, 12 e 14, sendo observados dois sítios distantes por 8.178.930 pb, no par cromossômico 12. Cinco sequências típicas de elementos transponíveis e 29 genes foram observados em associação com os sítios de [(AAG)<sub>5</sub>]<sub>10</sub>.

Entre os sítios de [(ACC)<sub>5</sub>]<sub>10</sub>, cujo tamanho observado variou entre 30 a 173 bp, 54, 28 e um sítio de repetição foram encontrados em regiões de alta, moderada e baixa densidade gênica, respectivamente, estando o último localizado no cromossomo 7. Associados a esses sítios, foram anotados 52 genes e quatro elementos transponíveis.

O oligonucleotídeo [(CTC)<sub>5</sub>]<sub>10</sub> foi o mais frequente com 144 sítios de repetição, com tamanho entre 37 a 407 pb. Dez sítios foram observados em locais onde não havia a presença de genes, nos cromossomos 4, 6, 8, 11, 13, 14 e 16. Todavia, nos pares cromossômicos 8, 11 e 16, dois sítios de repetições foram anotados encontrando-se em uma distância de 3.914.815, 9.461.734, 8.317.511 pb, respectivamente. Por outro lado, quatro repetições apresentavam-se distribuídas em regiões de baixa densidade gênica, estando as últimas dispostas nos cromossomos 3, 15, 16 e 17. Entre essas repetições, notou-se a presença de 15 genes e quatro sequências de elementos transponíveis associados.

Com relação à distribuição de [(TGA)<sub>6</sub>]<sub>10</sub>, 103 sítios foram observados com tamanho variável entre 40 a 386 pb, estando 55 e 47 sítios localizados em regiões de alta e moderada densidade gênica, enquanto que, apenas um foi anotado em locais de baixa densidade gênica no cromossomo 19. Associados a esta distribuição, foram encontrados 42 genes e um elemento transponível.

Pressupondo a existência de uma correlação entre a presença de microssatélites encontrados na análise *in silico* e genes e/ou elementos transponíveis em soja, o

coeficiente de Pearson ( $r$ ) foi calculado (Figura 7). Os cromossomos ou pseudomoléculas cromossômicas 2, 4, 6, 8, 10, 12, 14 e 18 apresentaram uma correlação positiva entre a presença de microssatélites e genes, indicando que a abundância destas sequências repetitivas estaria diretamente relacionada com a ocorrência de genes. Por outro lado, uma correlação negativa foi observada para os cromossomos 1, 3, 7, 9, 11, 13, 15, 16, 17, 19 e 20, sugerindo que a presença dos microssatélites não estaria associada a genes. O coeficiente de Pearson foi nulo para o cromossomo 5, enfatizando a inexistência de uma correlação entre a ocorrência de oligonucleotídeos e sequências gênicas (Figura 7).

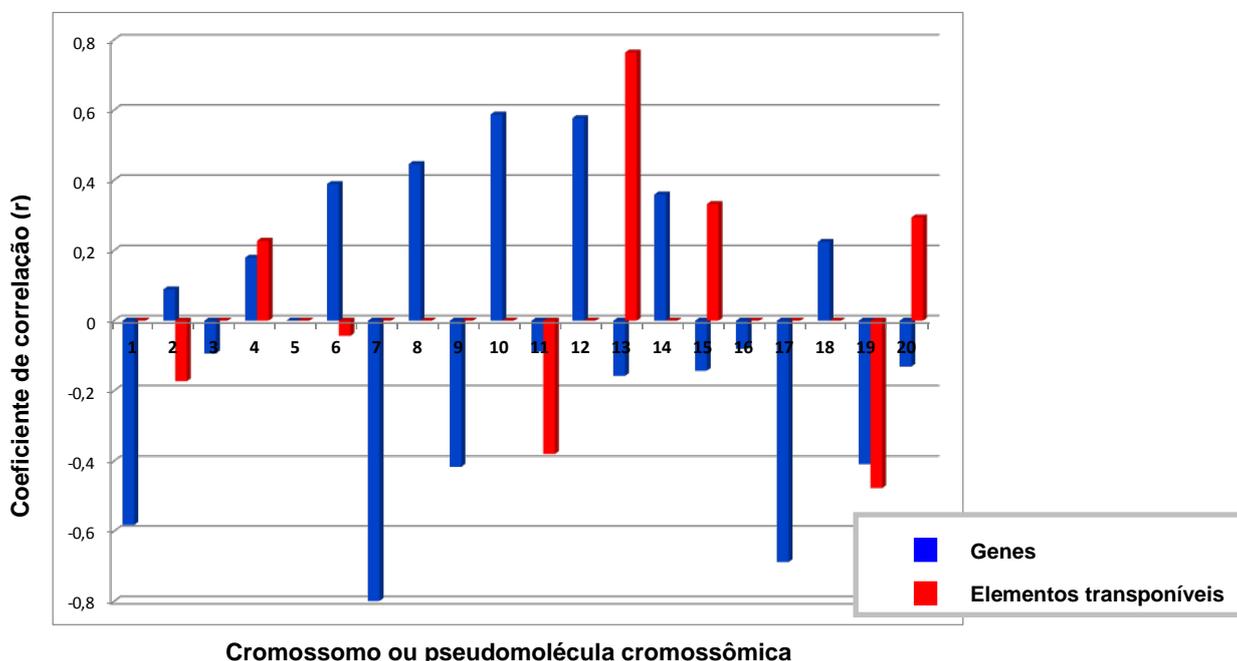
Em relação à associação de microssatélites e elementos transponíveis, uma correlação positiva foi notada para os cromossomos ou pseudomoléculas 4, 13, 15 e 20; enquanto que os de número 2, 6, 11 e 19 apresentaram correlação negativa. Por sua vez, para as pseudomoléculas 1, 3, 5, 7, 8, 9, 10, 12, 14, 16, 17 e 18, nenhuma correlação foi visualizada (Figura 7).

### ***Disposição dos sítios de oligonucleotídeos em pseudomoléculas cromossômicas de Glycine max***

O número de sítios de repetições dos microssatélites mapeados para os 20 cromossomos ou pseudomoléculas cromossômicas de soja está mostrado na Tabela 2. As sequências de oligonucleotídeos foram posicionadas em cada tipo cromossômico, de acordo com a sua localização genômica com base em dados do SoyBase (<http://soybase.org/>), resultando na construção de um padrão representativo da distribuição de [(AG)<sub>8</sub>]<sub>15</sub>, [(AAG)<sub>5</sub>]<sub>10</sub>, [(ACC)<sub>5</sub>]<sub>10</sub>, [(CTC)<sub>5</sub>]<sub>10</sub> e [(TGA)<sub>6</sub>]<sub>10</sub> para *G. max* (Figura 8). Para cada *cluster* de SSR identificado, as informações individuais envolvendo a posição exata no genoma, tamanho da repetição em pb, bem como os genes e elementos transponíveis associados, encontram-se disponível em anexo (Tabelas suplementares 1 - 5).

Algumas diferenças foram notadas na densidade das repetições entre os cromossomos. O oligonucleotídeo [(ACC)<sub>5</sub>]<sub>10</sub> mostrou o menor número de repetições para os pseudocromossomos 9 (uma repetição), 12 e 15 (duas repetições, cada), em relação às demais distribuições. Por sua vez, os microssatélites [(AAG)<sub>5</sub>]<sub>10</sub> e [(CTC)<sub>5</sub>]<sub>10</sub> apresentaram um maior número de unidades de repetição por cromossomo (Tabela 2).

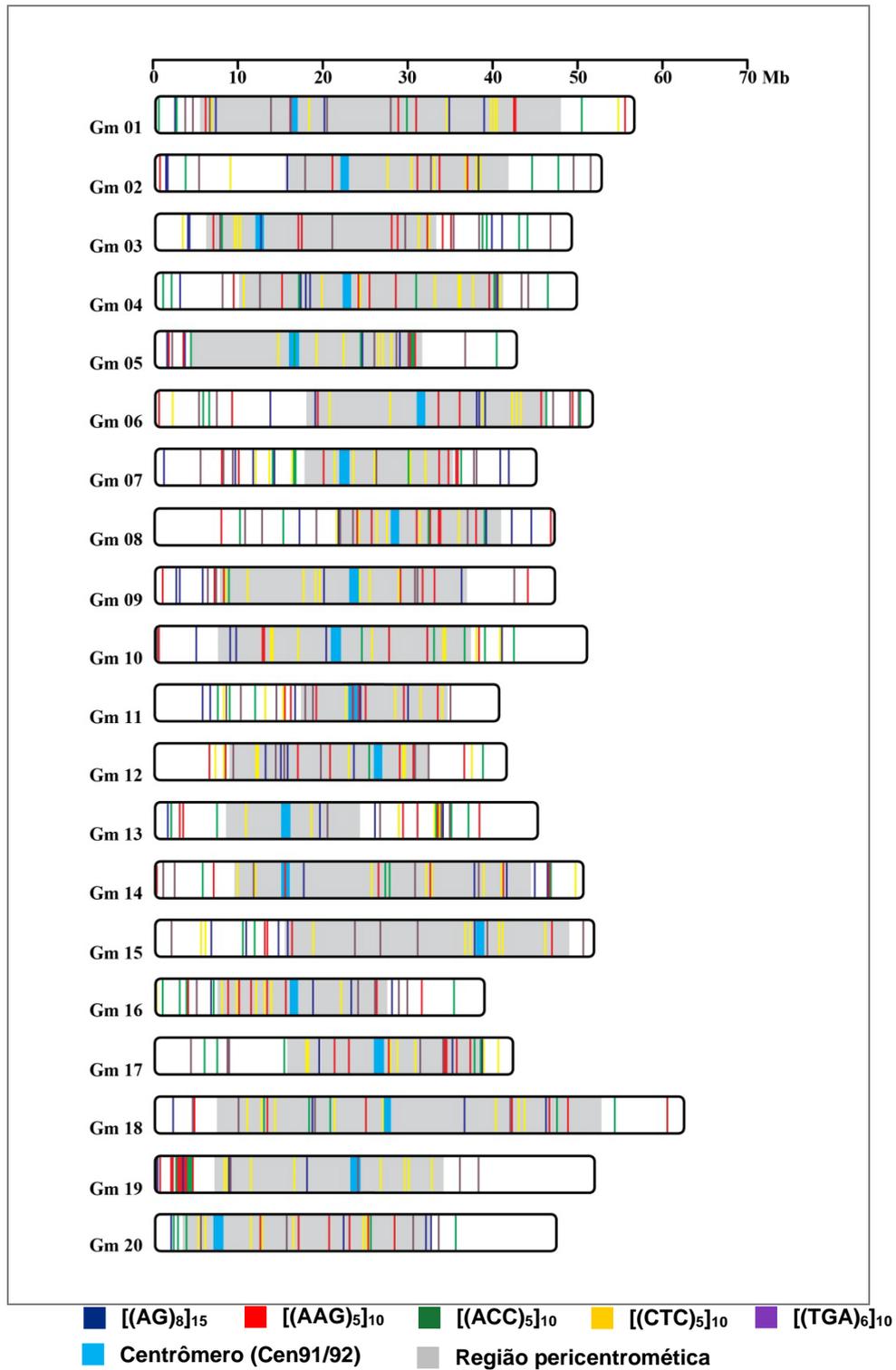
Uma análise mais detalhada foi realizada para o cromossomo 13, mediante a comparação dos mapas físico e cromossômico, visando confrontar a localização dos sítios dos microssatélites visualizada pela FISH e análise *in silico*. O referido cromossomo apresentou um tamanho relativo igual a 2,5  $\mu\text{m}$ , equivalente ao valor aproximado de 44,41 Mb citado no mapa físico, com a extensão total do braço curto constituída pelo DNAr 45S colocalizado com uma banda  $\text{CMA}_3^+$ , o qual se encontrou localizado na posição Gm13: 48186..1590000 (<http://www.soybase.org/>), envolvendo a região proximal. O mapa físico do cromossomo 13 permitiu a identificação de quatro sítios de oligonucleotídeos [(ACC)<sub>5</sub> – 13.2; (CTC)<sub>5</sub> - 13.1; (CTC)<sub>5</sub> – 13.2; (TGA)<sub>6</sub> – 13.1] localizados em alta e moderada densidade gênica e com tamanhos de 82 a 90 pb, na região pericentromérica (Gm:8550000..2417000). Comparativamente, esta região apresenta um tamanho próximo a 6 Mb, valor correspondente à heterocromatina pericentromérica  $\text{CMA}_3^+$  visualizada no mapa cromossômico (Figura 9), na qual foram observados diferentes sítios da FISH com os oligonucleotídeos.



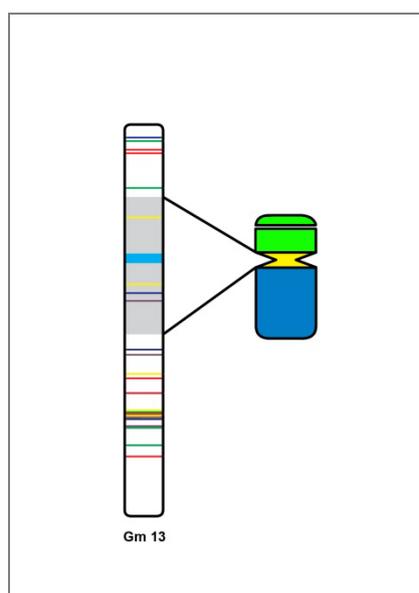
**Figura 7.** Coeficiente de correlação de Pearson (r) entre a presença de microssatélites, genes e elementos transponíveis em cromossomos ou pseudomoléculas cromossômicas em *Glycine max*.

**Tabela 2.** Distribuição de classes de repetições dos oligonucleotídeos com padrão de microssatélites mapeados nos cromossomos ou pseudomoléculas cromossômicas de *Glycine max*.

Oligos	Cromossomos ou pseudomoléculas cromossômicas (Gm)																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
[(AG) <sub>8</sub> ] <sub>15</sub>	5	4	5	5	4	5	6	5	5	5	5	4	4	5	5	4	4	4	4	4
[(AAG) <sub>5</sub> ] <sub>10</sub>	7	6	8	6	4	7	7	9	7	6	8	7	6	7	4	8	7	7	9	6
[(ACC) <sub>5</sub> ] <sub>10</sub>	4	3	5	6	6	4	5	4	1	5	3	2	5	4	2	5	5	5	4	5
[(CTC) <sub>5</sub> ] <sub>10</sub>	7	7	6	8	6	7	8	6	8	8	7	8	5	8	8	7	7	8	8	7
[(TGA) <sub>6</sub> ] <sub>10</sub>	6	5	6	5	5	5	6	6	5	4	6	6	4	5	6	5	5	4	5	4
Σ	29	25	30	30	25	28	32	30	26	28	29	27	24	29	25	29	28	28	30	26



**Figura 8.** Representação da distribuição *in silico* de  $[(AG)_8]_{15}$ ,  $[(AAG)_5]_{10}$ ,  $[(ACC)_5]_{10}$ ,  $[(CTC)_5]_{10}$  e  $[(TGA)_6]_{10}$  em cada cromossomo ou pseudomolécula cromossômica de *Glycine max*, bem como da região centromérica e pericentromérica determinada pela disposição descrita no site <http://soybase.org/>.



**Figura 9.** Comparação dos mapas físico e cromossômico para o cromossomo 13 de *Glycine max*, evidenciando a localização pericentromérica da banda CMA<sub>3</sub><sup>+</sup> e DNAr 45S, bem como a distribuição dos sítios de oligonucleotídeos associados a regiões eucromáticas e heterocromáticas.

## DISCUSSÃO

Os microssatélites constituem uma ampla fração do genoma de eucariotos, organizados de modo disperso ou participando da composição heterogênea da heterocromatina constitutiva (Cuadrado *et al.*, 2000; Yang *et al.*, 2005; Begum *et al.*, 2009). Uma grande proporção de sítios de (AG)<sub>8</sub>, (ACC)<sub>5</sub>, (AAG)<sub>5</sub>, (CTC)<sub>5</sub> e (TGA)<sub>6</sub> foi revelada nos genomas das espécies de leguminosas analisadas mediante a FISH com os oligonucleotídeos. No entanto, foram observadas variações em suas quantidades, sua organização (em tandem e/ou dispersa) e em sua distribuição ao longo dos cromossomos/genomas.

Supostamente, as marcações em tandem revelam a organização de SSRs em regiões de DNA repetitivo distribuídas nos cromossomos. Essas sequências correspondem a cerca de 100 a 1000 unidades de repetições, as quais estão associadas à heterocromatina constitutiva. Com relação às marcações dispersas, os loci de microssatélites, com extensão de 20 a 100 pb, podem estar distribuídos próximos um dos outros ao longo dos genomas, permitindo sua visualização como marcações fracas, entre sequências de cópias únicas e/ou outras sequências repetitivas (Cuadrado *et al.*, 2008).

Na FISH, um padrão de distribuição disperso foi notado para todos os microssatélites. Todavia, sinais com maiores intensidades foram visualizados nas regiões pericentroméricas, coincidindo com blocos de heterocromatina constitutiva para *G. max* (bandas CMA<sub>3</sub><sup>+</sup>), *V. unguiculata* (bandas CMA<sub>3</sub><sup>+</sup>) e *V. radiata* (bandas DAPI<sup>+</sup>), bem como para *P. vulgaris* e *P. lunatus*, cujos padrões de bandas CMA<sub>3</sub><sup>+</sup> foram descritos por Almeida (2006). Entre os trinucleotídeos, o motivo AAG tem sido apontado como a repetição trimérica mais abundante em leguminosas (Morgante *et al.*, 2002; Mun *et al.*, 2006), o que também foi notado pela distribuição dos sinais de hibridização nas espécies analisadas, especialmente na região pericentromérica de *P. lunatus*, caracterizando-o como um marcador citológico para essa espécie.

Em soja, as regiões heterocromáticas pericentroméricas detêm cerca de 57% das sequências genômicas, constituídas principalmente por sequências repetitivas (Lin *et al.*, 2005; Schmutz *et al.*, 2010), como sequências de DNA satélite com retroelementos interdispersos (Nunberg *et al.*, 2006; Swaminathan *et al.*, 2007; Findley *et al.*, 2010). Todavia, grande parte das regiões ricas em elementos repetitivos permanecem como “lacunas” nos projetos de sequenciamento, devido à dificuldade em determinar a sua

correta posição, arranjo e número de cópias no genoma, sendo uma fração dos constituintes dos *scaffolds*, que compreendem cerca de 17,7 Mb em soja (Schmutz *et al.*, 2010).

Gill *et al.* (2009) caracterizaram e hibridizaram as sequências satélites CentGm-1 e CentGm-2, visualizando 14 e oito sítios centroméricos, respectivamente, afirmando serem constituintes das regiões centroméricas dos cromossomos de soja. Estas sequências repetitivas, contendo 91-92 bp por unidade, representam uma fração dos *scaffolds* organizados em 19 das 20 pseudomoléculas cromossômicas, durante a construção do mapa físico de soja. Contudo, quatro cromossomos continham vários *clusters* de CentGm-1 e/ou CentGm-2 em diferentes posições cromossômicas, refletindo sua posição física correta ou a dificuldade de orientar essas regiões altamente repetitivas (Schmutz *et al.*, 2010).

No mapa cromossômico, a região pericentromérica dos cromossomos de soja foi caracterizada pela presença de bandas  $CMA_3^+$  bem como por sítios de FISH dos microssatélites  $(AG)_8$ ,  $(ACC)_5$ ,  $(AAG)_5$ ,  $(CTC)_5$  e  $(TGA)_6$ , ressaltando a composição heterogênea da heterocromatina. Como a FISH evidencia apenas sequências com comprimentos acima de 1-3 kb, fica evidente que as marcações de SSRs *in situ* estejam associadas às regiões de heterocromatina, sendo possivelmente uma fração constituinte dos 1.148 *scaffolds* existentes no sequenciamento da soja.

Considerando que as SSRs são componentes notáveis da heterocromatina constitutiva, sugere-se que estas sequências exerçam funções importantes na organização cromossômica, em nível de telômeros e centrômeros, onde podem auxiliar na coesão entre as cromátides-irmãs e desempenhar papel acessório na formação e/ou estabilidade do cinetócoro bem como atuar na remodelagem da cromatina e regulação da expressão gênica, conjuntamente com outras sequências de DNA repetitivo (Dimitri *et al.*, 2009; Gartenberg, 2009).

Uma comparação entre a análise *in situ* e *in silico*, bem como uma associação entre a heterocromatina  $CMA_3^+$  e o padrão de distribuição dos oligonucleotídeos puderam ser melhor detalhadas para o cromossomo 13, portador do sítio de DNAr 45S. O DNAr 45S ocupa a extensão total do braço curto, correspondente à região  $CMA_3^+$  e pericentromérica revelada pela análise *in silico*. Adicionalmente, observou-se uma grande quantidade de marcação dos oligonucleotídeos  $(AG)_8$ ,  $(ACC)_5$ ,  $(CTC)_5$ ,  $(CTC)_5$  e  $(TGA)_6$  na citada região, mediante à FISH. Na análise *in silico*, por sua vez, foram identificados quatro sítios dos referidos oligonucleotídeos na região pericentromérica,

estando localizados em regiões de alta e moderada densidade gênica e com tamanhos de 82 a 90 pb.

Tendo em vista as grandes regiões pericentroméricas marcadas por diferentes oligonucleotídeos após a FISH, tais sítios parecem estar associados a regiões de heterocromatina. Dessa forma, o tamanho dos sítios nas regiões pericentroméricas observado na análise *in silico* estaria subestimado em relação ao seu tamanho real, uma vez que a maioria das unidades de repetição em um sítio SSR alinhou de modo sobreposto e não adjacente. Outra possibilidade é que esses sítios de repetições corresponderiam aos grupos de microssatélites associados aos 21,6% dos genes localizados em regiões heterocromáticas do genoma da referida espécie, conforme descrito por Schmutz *et al.* (2010).

Por outro lado, oligonucleotídeos com padrão SSR também podem estar associados a regiões eucromáticas (Cuadrado e Jouve, 2002; Cuadrado e Jouve, 2007b). Em *G. max*, a presença de marcações (AG)<sub>8</sub> *in silico* corrobora com sua predominância em banco de ESTs (*Expressed Sequence Tags*; Etiquetas de Sequências Expressas) de soja (Shultz *et al.*, 2007; Shoemaker *et al.*, 2008), reforçando a hipótese que elas também estariam presentes entre sequências de cópia única. Por sua vez, os sinais presentes em *P. vulgaris*, *P. lunatus*, *V. unguiculata* e *V. radiata* condizem com a predominância deste dinucleotídeo em leguminosas (Morgante *et al.*, 2002; Mun *et al.*, 2006).

No presente trabalho, a distribuição eucromática dos oligonucleotídeos pôde ser observada nas pseudomoléculas cromossômicas de *G. max* mediante à análise *in silico*, a qual enfatizou a disposição dos microssatélites em regiões de considerável densidade gênica, fato anteriormente relatado por Song *et al.* (2004) e Choi *et al.* (2007). Segundo Morgante *et al.* (2002), existe uma relação positiva linear entre a frequência de SSRs e a porcentagem de DNA cópia única em plantas, sugerindo que os microssatélites seriam abundantes também nessas regiões. Uma correlação positiva para a associação de microssatélites com regiões de densidade gênica, bem como com a presença de elementos transponíveis também foi observada no presente trabalho. Dessa forma, considerando o tamanho aparentemente pequeno dos sítios encontrados para os oligonucleotídeos analisados *in silico*, as diferentes repetições das SSRs parecem estar organizadas na eucromatina e em parte da heterocromatina em grupos interdispersos ou associados a outras sequências repetitivas e/ou sequências únicas.

Na tentativa de identificar a relação entre a origem dos microssatélites e a distribuição em regiões de heterocromatina e/ou eucromatina observada nesta análise para *G. max*, Bhargava e Fuentes (2009) afirmaram que os microssatélites podem se originar “*de novo*” ou via transposição de elementos móveis. Provavelmente, os microssatélites *de novo* surgem espontaneamente, a partir ou internamente em sequências únicas dispostas em regiões eucromáticas, caracterizadas pela alta proporção de genes expressos bem como pela ocorrência de recombinação meiótica. Por conseguinte, estas SSRs desempenhariam um papel fundamental na manutenção dos códons genéticos e consequentemente da integridade proteica (Li *et al.*, 2004), uma vez que podem estar relacionadas à forte seleção negativa sobre fontes de mutação que causem mudança na matriz de leitura (Metzgar *et al.*, 2000).

Os elementos transponíveis também parecem estar envolvidos na gênese dos microssatélites. À medida que ocorre a transposição, esses elementos podem apresentar um ou mais sítios com predisposição à formação de microssatélites, favorecendo sua dispersão e amplificação em diferentes regiões genômicas (Bhargava e Fuentes, 2009), inclusive na heterocromatina, uma vez que esses retroelementos podem estar direta ou distantemente associados a blocos heterocromáticos (Gartenberg, 2009). Contudo, a origem de SSRs vinculada aos elementos transponíveis vem sendo questionada na literatura, pelo fato que regiões com alta densidade de elementos transponíveis nem sempre coincidem com locais ricos em microssatélites (Schlötterer, 2000).

Além disso, a expansão dos arranjos de microssatélites parece também ser fortemente influenciada pelo mecanismo de replicação do tipo *slippage*, atuando juntamente com a funcionalidade de um grupo de enzimas e proteínas envolvidas no processamento do DNA e com a remodelagem da cromatina (Kubis *et al.*, 1998; Bhargava e Fuentes, 2009). Por outro lado, apesar de pouco esclarecida, a taxa mutacional dessas repetições não é uniforme, variando entre os tipos e composição das repetições, o tipo de SSRs, assim como entre grupos taxonômicos (Primmer *et al.*, 1996). Dessa forma, a formação e expansão dos arranjos de microssatélites, bem como sua dinâmica mutacional podem ser considerados os principais responsáveis pela disposição diferencial das SSR, em termos de especificidade de repetição e abundância, entre os diferentes taxa analisados.

A localização dos microssatélites ressalta uma distribuição diferencial aparentemente não aleatória das diferentes classes de SSRs analisadas, caracterizadas como genoma-específicas e até mesmo cromossomo-específicas, podendo estar

preferencialmente associadas à heterocromatina ou à eucromatina, com uma densidade de repetições variável. Conseqüentemente, o conhecimento do padrão de distribuição dos microssatélites, principalmente de *G. max*, aponta-os como importantes componentes dos genomas em espécies do clado Phaseoloids (tribo Phaseoleae), auxiliando na compreensão das propriedades evolutivas de suas repetições. Por fim, a integração das análises *in situ* e *in silico* revelou-se como uma estratégia promissora para a ancoragem de microssatélites nos pseudocromossomos da soja, permitindo um melhor entendimento da localização física e organização genômica dessas sequências repetitivas, as quais são consideradas importantes elementos dos 17,7 Mb scaffolds não ancorados ao mapa físico desta espécie.

### AGRADECIMENTOS

Os autores agradecem ao CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) pelo apoio financeiro ao projeto.

### LITERATURA CITADA

- Almeida CCS. 2006. *Mapeamento físico e análise evolutiva em Phaseolus vulgaris L. e P. lunatus L., utilizando hibridização in situ fluorescente (FISH)*. Tese de Doutorado, Universidade Federal de Pernambuco, Recife.
- Begum R, Alam SS, Menzel G, Schmidt T. 2009. Comparative molecular cytogenetics of major repetitive sequence families of three *Dendrobium* species (Orchidaceae) from Bangladesh. *Annals of Botany* 104: 863-772.
- Bhargava A, Fuentes FF. 2009. Mutational Dynamics of Microsatellites. *Molecular Biotechnology* 44 (3): 250-266.
- Biondo E, Miotto STS, Schifino-Wittmann MT. 2005. Citogenética de espécies arbóreas da subfamília Caesalpinioideae - Leguminosae do sul do Brasil. *Ciência Florestal* 15 (3): 241-248.
- Carvalho CR, Saraiva LS. 1993. An air drying technique for maize chromosomes without enzymatic maceration. *Biotechnic and Histochemistry* 68 (3): 142-145.
- Chang SB, Yang TJ, Datema E. *et al.* 2008. FISH mapping and molecular organization of the major repetitive sequences of tomato. *Chromosome Research* 16: 919-933.

- Choi IY, Hyten DL, Matukumalli LK. *et al.* 2007. A soybean transcript map: gene distribution, haplotype and single nucleotide polymorphism analysis. *Genetics* 176: 685-696.
- Cuadrado A, Jouve N. 2002. Evolutionary trends of different repetitive DNA sequences during speciation in genus *Secale*. *Journal Heredity* 93: 339–345.
- Cuadrado A, Jouve N. 2007a. The nonrandom distribution of long clusters of all possible classes of trinucleotide repeats in barley chromosomes. *Chromosome Research* 15: 711-720.
- Cuadrado A, Jouve N. 2007b. Similarities in the chromosomal distribution of AG and AC repeats within and between *Drosophila*, human and barley chromosomes. *Cytogenetics and Genome Research* 119: 91-99.
- Cuadrado A, Jouve N. 2010. Chromosomal detection of simple sequence repeats (SSRs) using nondenaturing FISH (ND-FISH). *Chromosoma* 119: 495-503.
- Cuadrado A, Cardoso M, Jouve N. 2008. Physical organisation of simple sequence repeats (SSRs) in Triticeae: structural, functional and evolutionary implications. *Cytogenetics Genome Research* 120: 210–219.
- Cuadrado A, Schwarzacher T, Jouve N. 2000. Identification of different chromatin classes in wheat using *in situ* hybridization with simple sequence repeat oligonucleotides. *Theoretical Applied Genetics* 101: 711–717.
- Dimitri P, Caizzi R, Giordano E, Accardo MC, Lattanzi G, Biamonti G. 2009. Constitutive heterochromatin: a surprising variety of expressed sequences. *Chromosoma* 118: 419–435.
- Fonsêca A, Ferreira J, Santos TRB. *et al.* 2010. Cytogenetic map of common bean (*Phaseolus vulgaris* L.). *Chromosome Research* 18: 487-502.
- Findley SD, Cannon S, Varala K. *et al.* 2010. A fluorescence *in situ* hybridization system for karyotyping soybean. *Genetics* 185: 727-744.
- Gartenberg M. 2009. Heterochromatin and the cohesion of sister chromatids. *Chromosome Research* 17: 229-238.
- Gill N, Findley S, Walling JG. *et al.* 2009. Molecular and Chromosomal Evidence for Allopolyploidy in Soybean. *Plant Physiology* 151: 1167–1174.
- Graham PH, Vance CP. 2003. Legumes: importance and constraints to greater use. *Plant Physiology* 131: 872-877.
- Heslop-Harrison JS. (2000) Comparative genome organization in plants: from sequence and markers to chromatin and chromosomes. *Plant Cell* 12: 617-635.

- Heslop-Harrison JS, Schwazarcher T, Anamthawat-Jónsson K, Leitch AR, Shi M. 1991. *In situ* hybridization with automated chromosome denaturation. *Technique* 3:109-115.
- Hyten DL, Cannon SB, Song Q. *et al.* 2010a. High-throughput SNP discovery through deep resequencing of a reduced representation library to anchor and orient scaffolds in the soybean whole genome sequence. *BMC Genomics* 11: 38.
- Hyten DL, Choi I Y, Song Q. *et al.* 2010b. A high density integrated genetic linkage map of soybean and the development of a 1,536 Universal Soy Linkage Panel for QTL mapping. *Crop Science* 50: 960-968.
- Jong JH, Fransz P, Zabel P. 1999. High resolution FISH in plants – techniques and applications. *Trend Plant Science. Reviews* 4:258-263.
- Kubis S, Schmidt T, Heslop-Harrison JS. 1998. Repetitive DNA as a major component of plant genomes. *Annals of Botany* 82: 45-55.
- Lewis G, Schirer B, Mackinder B, Lock M. 2005. *Legumes of the world*, 1st edn. London: Kew.
- Li YC, Abraham BK, Fahima T, Nevo E. 2004. Microsatellites within genes: Structure, function, and evolution. *Molecular Biology and Evolution* 21: 991-1007.
- Lin JY, Jacobus BH, SanMiguel P. *et al.* 2005. Pericentromeric Regions of Soybean (*Glycine max* L. Merr.) Chromosomes Consist of Retroelements and Tandemly Repeated DNA and Are Structurally and Evolutionarily Labile. *Genetics* 170: 1221–1230.
- Marechal R, Mascherpa JM, Stainier F. 1978. Étude taxonomique d'un groupe complexe d'espèces de genres *Phaseolus* et *Vigna* (Papilionaceae) sur la base de données morphologiques et polliniques, traitées par l'analyse informatique. *Boissiera* 28:1-273.
- Metzgar D, Bytof J, Wills C. 2000. Selection against frameshift mutations limits microsatellite expansion in coding DNA. *Genome Research* 10: 72-80.
- Morgante M, Hanafey M, Powell W. 2002. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nature Genetics* 30: 194-200.
- Mun JH, Kim DJ, Choi HK. *et al.* 2006. Distribution of microsatellites in the genome of *Medicago truncatula*: A resource of genetic markers that integrate genetic and physical maps. *Genetics* 172: 2541-2555.

- Nunberg A, Bedell JA, Budiman MA. *et al.* 2006. Survey sequencing of soybean elucidates the genome structure, composition and identifies novel repeats. *Functional Plant Biology* 33: 765-773.
- Ohmido N, Sato S, Tabata S, Fukui K. 2007. Chromosome maps of legumes. *Chromosome Research* 15: 97-103.
- Pedrosa A, Jantsch MF, Moscone EA, Ambros PF, Schweizer D. 2001. Characterization of pericentromeric and sticky intercalary heterochromatin in *Ornithogalum longibracteatum* (Hyacinthaceae). *Chromosoma* 110: 203-213.
- Pedrosa A, Sandal N, Stougaard J, Schweizer D, Bachmair A. 2002. Chromosomal map of the model legume *Lotus japonicus*. *Genetics* 161:1661-1672.
- Pedrosa-Harand A, Almeida CCS, Mosiolek M, Blair MW, Schweizer D, Guerra M. 2006. Extensive ribosomal DNA amplification during Andean common bean (*Phaseolus vulgaris* L.) evolution. *Theoretical Applied Genetics* 112: 924-933.
- Primmer CR, Ellegren H, Saino N, Moller AP. 1996. Directional evolution in germline microsatellite mutations. *Nature Genetics* 13: 391-393.
- Schmidt T, Heslop-Harrison JS. 1998. Genomes, genes and junk: the large-scale organization of plant chromosomes. *Trends Plant Science* 3: 195- 199.
- Schlueter JA, Goicoechea JL, Collura K. *et al.* 2008. BAC-end sequence analysis and draft physical map of the common bean (*Phaseolus vulgaris* L.) genome. *Tropical Plant Biology* 1:40-48.
- Schlötterer C. 2000. Evolutionary dynamics of microsatellite DNA. *Chromosoma* 109: 365-371.
- Schmutz J, Cannon SB, Schlueter JA. *et al.* 2010. Genome sequence of the paleopolyploid soybean. *Nature* 463:178-183.
- Schweizer D, Ambros PF. 1994. Chromosome banding. In: Gosden JR, ed. *Methods in Molecular Biology*. Humana Press: Totowa, 97–112.
- Shoemaker KC, Grant D, Olson T. *et al.* 2008. Microsatellite discovery from BAC end sequences and genetic mapping to anchor the soybean physical and genetics maps. *Genome* 51: 294-302.
- Shultz JL, Kazi S, Bashir R, Afzal JA, Lightfoot DA. 2007. The development of BAC-end sequence-based microsatellite markers and placement in the physical and genetic maps of soybean. *Theoretical Applied Genetics* 114: 1081–1090.
- Song QJ, Marek LF, Shoemaker RC. *et al.* 2004. A new integrated genetic linkage map of the soybean. *Theoretical and Applied Genetics* 109: 122-128.

- Swaminathan K, Varala K, Hudson M E. 2007. Global repeat discovery and estimation of genomic copy number in a large, complex genome using a high-throughput 454 sequence survey. *BMC Genomics* 8: 132.
- Tóth G, Gáspári Z, Jurka J. 2000. Microsatellites in different eukaryotic genomes: Survey and analysis. *Genome Research* 10: 967-981.
- Wander AL (2005). Cultivo do feijão irrigado na região noroeste de Minas Gerais. EMBRAPA. Dez, 2005, n° 5 [Cited 04/May/2005]. Available from World Wide Web: <http://www.embrapa.br>.
- Wanzenböck EM, Schöfer C, Schweizer D, Bachmair A. 1997. Ribosomal transcription units integrated via T-DNA transformation associate with the nucleolus and do not require upstream repeat sequences for activity in *Arabidopsis thaliana*. *Plant Journal* 11: 1007-1016.
- Yang TJ, Lee S, Chang SB, Yu Y, de Jong JH, Wing RA. 2005. In depth sequence analysis of the centromeric region of tomato chromosome 12: Identification of a large CAA block and characterization of centromeric retrotransposons. *Chromosoma* 114: 103- 117.

## ***Capítulo II***

---

**Ocorrência e distribuição dos retrotransposons *Ty1-copia-like* e *Ty3-gypsy-like* em representantes dos gêneros *Glycine* Willd., *Phaseolus* L. e *Vigna* Savi.**

**Artigo a ser submetido à revista *Chromosome Research***

---

## **Ocorrência e distribuição dos retrotransposons Ty1-copia-like e Ty3-gypsy-like em representantes dos gêneros *Glycine* Willd., *Phaseolus* L. e *Vigna* Savi.**

Kyria Cilene de A. Bortoleti<sup>1,2</sup>, Ana Christina Brasileiro-Vidal<sup>2</sup>, Lidiane Lindinalva Barbosa Amorim<sup>2</sup>, Emanuele Varão Vasconcelos<sup>2</sup>, Valeska Pandolfi<sup>2</sup>, Ricardo Vilela Abdelnoor<sup>3</sup> e Ana Maria Benko-Iseppon<sup>2,4\*</sup>

1. Colegiado de Ciências Biológicas, Universidade Federal do Vale do São Francisco, Campus Ciências Agrárias, Rodovia BR 407, Km 12 – Lote 543 – Projeto de Irrigação Senador Nilo Coelho, s/nº “C1”, CEP: 56304-917, Petrolina, PE, Brasil.

2. Laboratório de Genética e Biotecnologia Vegetal, Departamento de Genética, Universidade Federal de Pernambuco, Av. da Engenharia, s/nº, CEP 50740-600, Recife, PE, Brasil.

3. Embrapa Soja, Rod. Carlos João Strass, Distrito e Warta, Caixa Postal 231, CEP 86001-970, Londrina, PR, Brasil.

4. Autor para correspondência: ana.benko.iseppon@pq.cnpq.br

Running title: Distribuição de retroelementos LTR em leguminosas

## Resumo

O domínio conservado dos genes RT (*Reverse Transcriptase*; Transcriptase Reversa) dos retrotransposons LTR Ty1-*copia*-like e Ty3-*gypsy*-like foram amplificados a partir do genoma de *Vigna unguiculata* usando *primers* degenerados e hibridizados *in situ* em alguns representantes dos gêneros *Glycine* L., *Phaseolus* L. e *Vigna* SAVI. As sequências RT correspondentes a ambos os grupos exibiram níveis significantes de heterogeneidade, sendo a sequência de Ty1-*copia*-like menos homogênea, em comparação ao Ty3-*gypsy*-like. As análises filogenéticas das sequências RT mostraram uma similaridade com retroelementos descritos para eudicotiledôneas e monocotiledôneas, enfatizando a origem monofilética dos domínios RT destes retrotransposons, assim como sua transmissão vertical durante os processos de especiação ocorrentes na evolução vegetal. A FISH (*Fluorescent In Situ Hybridization*; Hibridização *in situ* Fluorescente) evidenciou a presença de sinais dispersos e pericentroméricos, utilizando ambos retroelementos, em *Glycine max*, *Phaseolus vulgaris*, *P. lunatus*, *V. unguiculata* e *Vigna radiata*, com algumas divergências interespecíficas. Tais marcações estavam aparentemente colocalizadas com loci de DNAr 5S e 45S, bem comodistribuídas em regiões heterocromáticas pericentroméricas e subterminais, enfatizando uma distribuição preferencial nos genomas das leguminosas analisadas. A grande abundância das sequências RT e diferentes localizações cromossômicas foram atribuídas à alta taxa de mutação e à baixa pressão seletiva, associadas ao mecanismo de retrotransposição do tipo *copy and paste* específica destes retroelementos, afetando a evolução estrutural dos genomas vegetais, em especial, das leguminosas.

**Palavras-chave:** Ty1-*copia*-like, Ty3-*gypsy*-like, FISH, leguminosas, filogenia

## Abstract

The conserved domains of reverse transcriptase (RT) genes of *Ty1-copia* and *Ty3-gypsy* groups of long terminal repeat (LTR) retrotransposons were amplified from *Vigna unguiculata* genome using degenerate primers and located *in situ* in representatives of *Glycine* L., *Phaseolus* L and *Vigna* Savi genus. The RT sequences exhibited significant heterogeneity levels, and the *Ty1-copia*-like sequences were less homogeneous in comparison to the former. Phylogenetic analysis of RT sequences showed a similarity to retroelements described for rosids, emphasizing the monophyletic origin of retrotransposon sequences in plant genomes, as well as its vertical transmissions during the speciation processes occurred in plant evolution. FISH analysis has showed the presence of dispersed and pericentromeric signals, using both retroelements, in *Glycine max*, *Phaseolus vulgaris*, *P. lunatus*, *V. unguiculata* and *Vigna radiata*, with some interspecific divergences. Such marks were apparently associated, in some cases, to 5s and 45S rDNA loci, and located in pericentromeric and subterminal heterochromatic regions, emphasizing a preferential distribution in the analyzed genomes. The large occurrence of RT sequences within genomes and their differential distribution along chromosomes should be due to the high mutational rates and low selective pressure associated to the *copy and paste* retrotransposition mechanism, affecting the structural evolution of plant genomes, and especially of these legumes.

**Keywords:** *Ty1-copia*-like, *Ty3-gypsy*-like, FISH, Leguminosae, phylogeny

### **Abreviações:**

LTRs = *Long Terminal Repeats*; Longas Repetições Terminais

gene *gag* = *Group-specific antigen*; antígeno grupo-específico

VLP = *Virus-like Particle*; partícula similar a vírus

gene *pol* = *Polyprotein*; Poliproteína

### **Introdução**

Os retrotransposons representam uma importante fração do DNA moderadamente repetitivo, apresentando um mecanismo de transposição do tipo *copy and paste* via RNA intermediário ao longo dos genomas, o qual é apontado como principal fator responsável pela expansão e evolução dos genomas vegetais, juntamente com a poliploidização (Todorovska, 2007; Du *et al.*, 2010a).

Estes retroelementos são divididos em duas subclasses, as quais diferem em sua estrutura e ciclo de transposição, incluindo retroelementos LTRs (*Long Terminal Repeats*; Longas Repetições Terminais) e Não-LTRs (Friesen *et al.*, 2001; Park *et al.*, 2007). Os retrotransposons LTR são flanqueados por duas LTRs (posições 5' e 3') e uma ORF internamente, a qual é normalmente constituída pelos genes *gag* (*Group-specific antigen*; antígeno grupo-específico) e *pol* (*Polyprotein*; Poliproteína).

O gene *gag* é codificador de uma poliproteína, a qual compreende três proteínas correspondentes aos subcomponentes da VLP (*Virus-like Particle*; partícula similar a vírus), a qual está envolvida na maturação e empacotamento do RNA. Por sua vez, o gene *pol* é responsável pela codificação de várias proteínas, como protease, integrase, RT (*Reverse Transcriptase*; Transcriptase Reversa) e RNase H, as quais são necessárias para a síntese e integração do DNAc (*Complementary DNA*; DNA complementar) em cromossomos hospedeiros (Chavanne *et al.*, 1998; Cheng *et al.*, 2009).

Entre os retrotransposons LTRs, os elementos Ty1-*copia* (Pseudoviridae) e Ty3-*gypsy* (Metaviridae) são considerados os mais abundantes e relatados nos genomas de diferentes angiospermas, representando, 25% e 75% dos genomas de arroz e milho, respectivamente (Du *et al.*, 2010a). Estas duas classes de retrotransposons diferenciam-se pelo domínio conservado dos genes RT, característico de cada retroelemento (Friesen *et al.*, 2001), bem como pela ordem gênica dos domínios RT e integrase na região *pol*. O Ty1-*copia* apresenta o domínio da integrase anterior (posição 5') ao da RT e RNaseH, enquanto que no Ty3-*gypsy*, tal domínio situa-se posterior (posição 3') ao RT e RNaseH. Adicionalmente, o Ty3-*gypsy* apresenta uma maior similaridade com os

retrovírus (Kumar e Bennetzen, 1999; Hansen e Heslop-Harrison, 2004), pela presença de um domínio *env* (Envelope), coincidente com a posição ocupada pelo gene codificante do envelope viral em retrovírus (Vershinin *et al.*, 2002), sendo estes classificados como *Gypsy-like* (Capy *et al.*, 1998).

Diferentes estudos têm isolado e caracterizado o domínio RT de famílias de retrotransposons, observando a distribuição destes elementos no genoma. Os retroelementos Ty1-*copia* e Ty3-*gypsy* têm prevalecido nos principais grupos vegetais, como em *Helianthus* L. (Natali *et al.*, 2006; Ungerer *et al.*, 2009), *Malus domestica* Borkh (Sun *et al.*, 2008), *Pisum sativum* L. (Macas *et al.*, 2007). Entretanto, Dixit *et al.* (2006) relataram que a ocorrência de Ty3-*gypsy* tem sido pouco identificada, estando limitada a alguns grupos taxonômicos (Natali *et al.*, 2006; Park *et al.*, 2007).

A maioria dos retrotransposons é transcricionalmente inativa durante o crescimento e desenvolvimento vegetal. Contudo, o aumento dos níveis transcricionais desses elementos é relatado em resposta a vários estímulos ambientais, mutagênese e evolução genômica (Kumar e Bennetzen, 1999; Xiao *et al.*, 2004). O modo replicativo da transposição dos retroelementos resulta em seu acúmulo no genoma, podendo também gerar mutações pela inserção de novas cópias dentro ou próximas a genes e causar rearranjos cromossômicos servindo como sítios para recombinação ectópica (Bennetzen, 2000). Supostamente, estas mutações originam distintas famílias de retroelementos, as quais se diferenciam pelas suas atividades e distribuições nos cromossomos e/ou porções cromossômicas (Natali *et al.*, 2006).

Desta forma, a elevada abundância e taxa de movimentação genômica tem permitido a utilização dos retroelementos na FISH (*Fluorescent In Situ Hybridization*; Hibridização *in situ* Fluorescente), em espécies proximamente relacionadas, contribuindo para o melhor entendimento de relações filogenéticas (Nielen *et al.*, 2009). Os gêneros *Glycine* Willd., *Phaseolus* L. e *Vigna* Savi pertencem à tribo Phaseoleae, agrupados em um clado monofilético (Phaseoloids), assim como apresentam ampla distribuição mundial (Marechal *et al.*, 1978). Representantes destes gêneros apresentam grande importância econômica, destacando-se as espécies *G. max* (L.) Merr., *P. vulgaris* (L.), *P. lunatus* (L.) e *V. unguiculata* (L.) Walp conhecidas popularmente como soja, feijão comum, feijão-fava e feijão-caupi, respectivamente.

Considerando a existência de poucos trabalhos que caracterizem a organização genômica de elementos retrotransponíveis em leguminosas cultivadas, o presente trabalho caracterizou sequências RT dos elementos Ty1-*copia*-like e Ty3-*gypsy*-like do

genoma de *V. unguiculata*, bem como, realizou uma abordagem comparativa da distribuição cromossômica dos citados retroelementos em espécies dos gêneros *Glycine*, *Phaseolus* e *Vigna* visando uma melhor compreensão sobre o mecanismo de evolução genômica desses retroelementos na tribo Phaseoleae.

## Material e métodos

### Material vegetal

Sementes das espécies *Glycine max* cv. Conquista, *P. vulgaris* cv. BRS Esplendor, *P. lunatus* cv. Vermelhinha, *V. unguiculata* cv. BR14-Mulato e *Vigna radiata* foram obtidas do banco de germoplasma da Embrapa Soja (Empresa Brasileira de Pesquisa Agropecuária, Soja, Londrina), Embrapa-CNPAF (Embrapa Arroz e Feijão, Goiás) para feijão comum e feijão-fava, LGBV/UFPE (Laboratório de Genética e Biotecnologia Vegetal, Universidade Federal de Pernambuco, Recife) e Embrapa-CPATSA (Embrapa, Centro de Pesquisas do Trópico Semiárido, Petrolina), respectivamente.

### Análise molecular

Isolamento e caracterização do domínio da transcriptase reversa (RT) dos retrotransposons Ty1-*copia*-like e Ty3-*gypsy*-like

A extração do DNA genômico foi realizada a partir de folhas jovens de *V. unguiculata* BR14-Mulato utilizando o método de CTAB, segundo descrito por Weising *et al.* (1995). Posteriormente, a obtenção da região genômica correspondente à transcriptase reversa do retrotransposon Ty1-*copia*-like foi realizado pela amplificação de DNA por PCR (*Polymerase Chain Reaction*; Reação da Polimerase em Cadeia) utilizando os *primers* degenerados 5'-GAGAATTCACNGCNTTYTNCAYGG-3' e 5'-GAGGATCCATRTCRTCACRTANAR-3' (Xiao *et al.*, 2004), desenhados para os motivos conservados 5'-TAFLHG-3' e 5'-YVDDML-3' da referida enzima. Por sua vez, o elemento Ty3-*gypsy*-like foi obtido através de amplificação com os *primers* GyRT1 (RMCVDYR, 5'-MRNATGTGYGTNGAYTAYMG-3') e GyRT4 (YAKLSKC, 5'-RCAYTTNSWNARYTTNGCR-3') (Friesen *et al.*, 2001).

As reações de PCR ocorreram em termociclador *Techne TC-412* (Barloworld Scientific). O volume final de 25 µL da reação foi constituído por 25 a 50 ng de DNA genômico de *V. unguiculata*, 50 pmol de cada iniciador, 0,16 mM de cada dNTP (Fermentas), 1U da enzima Taq DNA Polimerase (Fermentas) e 3,5 mM de MgCl<sub>2</sub>, na

presença do tampão da enzima. Para amplificação do retroelemento Ty1-*copia*-like, os ciclos de PCR ocorreram nas seguintes condições: desnaturação inicial a 94°C por 3 min, seguida por 35 ciclos constituídos por desnaturação a 94 °C por 30 s, anelamento a 50 °C por 30 s e extensão a 72 °C por 30 s. Por sua vez, para o retrotransposon Ty3-*gypsy*-like, os ciclos de PCR consistiram em desnaturação inicial a 94°C por 3 min, seguida por 40 ciclos de desnaturação a 94 °C por 1 min, etapa de ligação dos iniciadores a 52,7 °C por 1 min e extensão a 72 °C por 1 min.

Os produtos de PCR foram examinados em gel de agarose 1,8% corado com brometo de etídeo (0,5 mg/mL), os quais foram visualizados sob luz ultravioleta em transiluminador. Os fragmentos de interesse amplificados foram excisados do gel de agarose com auxílio de uma lâmina descartável, sendo purificados mediante o kit comercial GFX PCR DNA e Gel Band Purification Kit (GE Healthcare, USA), de acordo com especificações fornecidas pelo fabricante.

#### Clonagem e sequenciamento dos fragmentos RT

Os fragmentos purificados foram clonados no vetor pGEM® -T Easy (Promega), o qual foi inserido em *Escherichia coli* (cepa DH10B) por eletroporação, utilizando o BioRad - modelo Gene Pulser II.

A presença de insertos de interesse nos plasmídeos recombinantes foi verificada por PCR de colônia, utilizando os *primers* T7 e SP6 próprios para o vetor pGEM-T Easy. Clones individuais foram selecionados, sendo os DNA plasmidiais extraídos utilizando o protocolo do Qiagen Plasmid Mini Kit (Qiagen), com algumas adaptações. Para as reações de sequenciamento dos clones recombinantes foi utilizado o *kit BigDye Terminator* versão 3.1 (*Applied Biosystem*), os *primers* M13 (Fw): 5'-GTAAAACGACGGCCAGTG -3' e M13 (Rv): 5'- GGAAACAGCTATGACCATG-3' e o equipamento automático *ABI 3130 xl* (*Applied Biosystem*).

As sequências foram analisadas mediante o *software* BioEdit 7 (Hall, 1999), visando à construção de sequências consenso, as quais foram comparadas com os dados depositados no *GenBank* utilizando o programa BLAST (*Basic Local Alignment Search Tool*) (Altschul *et al.*, 1997) ([www.ncbi.nlm.nih.gov/BLAST](http://www.ncbi.nlm.nih.gov/BLAST)), sendo a seguir analisadas com o programa ORF Finder para identificação de domínios conservados. As análises subsequentes de nucleotídeos e proteínas foram realizadas utilizando o programa CLUSTAL W (Thompson *et al.*, 1994), sendo necessários alguns ajustes manuais nos alinhamentos.

### Análise filogenética

As análises filogenéticas foram conduzidas utilizando o programa MEGA 4.0 (Tamura *et al.*, 2007), sendo a inferência das árvores realizada por Máxima Parcimônia (MP) com o método de pesquisa Close-Neighbor-Interchange (CNI) e opção *random addition* (10 amostragens), bem como teste de *bootstrap* com 2.000 repetições.

### Estudos citogenéticos

#### Pré-tratamento e fixação

A análise mitótica foi realizada utilizando pontas de raízes pré-tratadas com uma solução de 8-hidroxiquinoleína (8HQ) por 18 h, 1 h à temperatura ambiente e 17 h a 8 °C, fixadas em etanol absoluto:ácido acético (3:1, v/v) ou metanol:ácido acético (3:1, v/v) por 2 a 24 h à temperatura ambiente e estocadas a -20 °C.

#### Preparação cromossômica

Para a preparação das lâminas, os meristemas foram lavados duas vezes em água destilada, por 5 min, digeridos em solução enzimática contendo 2% (p/v) celulase 'Onozuka R-10' (Serva) e 20% (v/v) pectinase (Sigma-Aldrich) em tampão citrato 10 mM por 180 min a 37 °C. Em seguida, o material digerido foi lavado em água destilada e incubado em ácido acético 45% por 30 min a 37 °C. Uma ponta de raiz foi utilizada por lâmina, a qual foi macerada em ácido acético 45%, sendo posteriormente submetida ao nitrogênio líquido para remoção da lamínula com posterior secagem ao ar. Outra metodologia utilizada seguiu o procedimento descrito por Carvalho e Saraiva (1993), com algumas modificações (Bortoleti *et al.*, em preparação).

Em ambos os casos, as lâminas foram selecionadas mediante coloração com uma solução de 2 µg/mL de DAPI (4'-6'-diamidino-2-fenilindole)/ glicerol (1:1, v/v). As preparações selecionadas foram descoradas em etanol absoluto:ácido acético (3:1, v/v) e etanol absoluto por 30 min e 1 h, respectivamente. Após secagem ao ar, as lâminas foram devidamente estocadas a -20 °C, até o momento do procedimento da FISH.

#### Marcação e obtenção das sondas de retrotransposons e DNAr 5S e 45S

Os fragmentos Ty1-*copia*-like e Ty3-*gypsy*-like isolados do genoma de *V. unguiculata* foram marcados com digoxigenina-11-dUTP (Roche) por *nick translation* (Invitrogen). As sondas R2, um fragmento de 6,5 Kb, contendo a unidade de repetição de DNAr 18S-5,8S-25S, oriunda de *Arabidopsis thaliana* (L). Heynh (Wanzenböck *et*

*al.*, 1997), e D2, um fragmento de 400 pb contendo duas unidades de repetição de DNAr 5S, proveniente de *Lotus japonicus* (Regel) K.Larsen (Pedrosa *et al.*, 2002), também foram utilizadas como marcadores cromossômicos. As sondas de DNAr 5S e 45S foram marcadas com biotina-11-dUTP (Sigma) e digoxigenina-11-dUTP, respectivamente, por *nick translation*.

#### Hibridização *in situ* Fluorescente (FISH)

As lâminas, estocadas a -20 °C, foram pré-tratadas como descrito por Pedrosa *et al.* (2001). A desnaturação das sondas, os banhos pós-hibridização e a detecção foram efetuados de acordo com Heslop-Harrison *et al.* (1991), exceto pela lavagem de estringência a qual foi realizada em 0,1xSSC a 42 °C. As misturas de hibridização consistiram de formamida 50% (v/v), dextran sulfato 10% (p/v), 2x SSC e 2-5 ng/μL de sonda. As lâminas foram desnaturadas por 7 min a 78 °C, bem como hibridizadas por 18 a 48 h a 37 °C. Para soja, os cromossomos foram desnaturados com formamida 70% em 2xSSC a 70 °C por 7 min, sendo posteriormente, desidratados por 5 min em uma série alcoólica (70% e 100%) a -20 °C. As sondas marcadas com biotina foram detectadas com antibiotina obtida em camundongo (DAKO) em combinação com anticamundongo conjugado com TRITC (DAKO) em BSA 1% (p/v). Enquanto as sondas marcadas com digoxigenina foram detectadas usando antidigoxigenina de ovelha conjugado com fluoresceína isotiocianato (FITC; Boehringer Mannheim) e amplificadas com antiovelha conjugado com FITC (Serotec) também em BSA 1% (p/v). Todas as preparações foram montadas com 2 μg/mL de DAPI em Vectashield (Vector).

#### Análise dos dados

As células foram analisadas com um microscópio de epifluorescência Leica DMLB e imagens das melhores células foram capturadas com uma câmera Leica DFC 340FX, utilizando o programa CW 4000 da Leica. As imagens foram otimizadas para melhor brilho e contraste com o Adobe Photoshop CS3 (Adobe Systems Incorporated). Para as imagens sobrepostas DAPI-DNAr 45S, a imagem do DAPI foi transformada para tons de cinza, o DNAr foi pseudocolorido com amarelo e ambas as imagens foram sobrepostas, usando a ferramenta *lighten*.

A identificação cromossômica de *P. vulgaris* e *P. lunatus* seguiu-se a nomenclatura determinada por Pedrosa-Harand *et al.* (2008) e Almeida (2006),

respectivamente. Para *V. unguiculata* e *V. radiata*, foram utilizadas as identificações descritas por Bortoleti *et al.* (em preparação).

## Resultados

Isolamento do domínio RT de sequências Ty1-*copia*-like e Ty3-*gypsy*-like do genoma de *V. unguiculata*

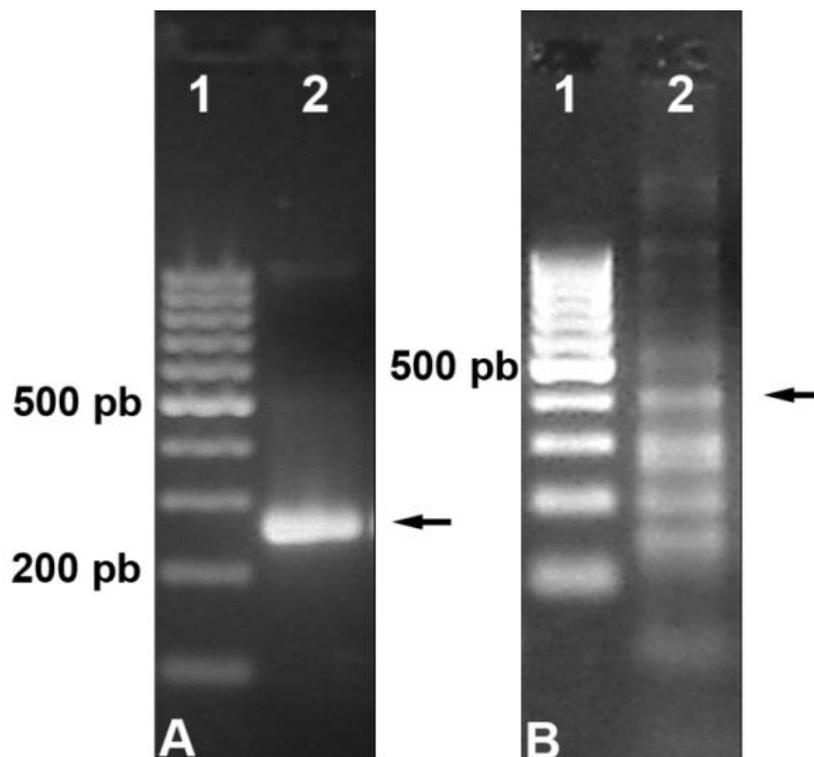
A PCR com os *primers* degenerados para o domínio RT do retrotransposon Ty1-*copia*-like produziu apenas um fragmento com o comprimento de aproximadamente 270 pb, o qual foi purificado e clonado (Figura 1A). Os clones foram sequenciados, apresentando um inserto com o tamanho de 280 pb de *V. unguiculata*. Contudo, apenas um clone, denominado *Vung* Ty1 clone 3, apresentou o domínio completo da transcriptase reversa. Essa sequência mostrou similaridade com regiões do domínio RT do retroelemento Ty1-*copia*-like encontradas em diferentes monocotiledôneas e dicotiledôneas, a partir do BLASTn realizado com sequências depositadas no *GenBank* (Tabela 1).

Por sua vez, a reação inespecífica com os *primers* degenerados GyRT1 e GyRT4 produziu bandas com tamanhos variáveis (Figura 1B), entre as quais o fragmento correspondente a aproximadamente 430 pb foi isolado e purificado. Após sequenciamento, dois clones mostraram insertos com o tamanho de 421 pb e 417 pb, *Vung* Ty3 clone 1 e *Vung* Ty3 clone 2, os quais apresentaram domínios conservados. Similarmente ao Ty1-*copia*-like, as respectivas sequências mostraram similaridade com fragmentos do domínio RT do retroelemento Ty3-*gypsy*-like caracterizados em diferentes espécies vegetais (Tabela 2).

O alinhamento múltiplo das sequências nucleotídicas dos fragmentos RT dos retrotransposons isolados do genoma de *V. unguiculata* com as sequências de DNA depositadas no *GenBank* para algumas espécies vegetais (ver Tabela 1 e 2) evidenciou que apenas 27,2% e 31,1% dos sítios compartilhados de Ty1-*copia* e Ty3-*gypsy*, respectivamente, estavam conservados, indicando um baixo nível de homologia entre elas.

Por sua vez, o alinhamento múltiplo das sequências de aminoácidos permitiu a identificação de motivos conservados característicos para os domínios RT dos retroelementos em análise. Para o retrotransposon Ty1-*copia*-like foi visualizado a presença dos motivos SLYGLKQA e LYVDDM, na porção central e terminal da sequência isolada de *V. unguiculata*, respectivamente (Figura 2). Enquanto que o

motivo SLYGLKGA também foi conservado nas múltiplas sequências alinhadas ao domínio RT isolado no presente estudo.



**Figura 1.** Amplificação por PCR do domínio RT do Ty1-*copia*-like (A) e Ty3-*gypsy*-like (B), a partir do DNA genômico de *Vigna unguiculata*, utilizando *primers* degenerados. As linhas 1 correspondem ao marcador de peso molecular 100 pb (Fermentas), enquanto que, as linhas 2 representam o produto da PCR para os respectivos domínios RT. Setas indicam fragmentos isolados e purificados.

**Tabela 1:** Identificação das sequências depositadas no *GenBank* que apresentaram similaridade com a sequência do domínio RT do retrotransposon Ty1-*copia*-like isolada do genoma de *Vigna unguiculata*, com seus respectivos números do acesso, descrição, cobertura (%), identidade (%) e tamanho (em pb).

Número do acesso	Descrição	Query Coverage (%)	Identidade (%)	Tamanho da sequência (em pb)
AJ290735.1	<i>Ginkgo biloba</i> copia like retroelement partial RT gene clone pGity2	94	72	264
AJ224364.1	<i>Picea abies</i> DNA for reverse transcriptase gene of Ty1- <i>copia</i> like retroelements	94	72	264
GU976509.1	<i>Eleocharis interstincta</i> clone copEint17 retrotransposon Ty1- <i>copia</i> reverse transcriptase gene, partial cds	94	73	266
GU976679.1	<i>Eleocharis palustris</i> subsp. <i>Waltersii</i> clone copEwal19 retrotransposon Ty1- <i>copia</i> reverse transcriptase gene partial cds	94	71	266
EF623926.1	<i>Malus x domestica</i> isolate CoRTFJ-9 retrotransposon Ty1- <i>copia</i> reverse transcriptase gene partial cds	94	71	266
EF623935.1	<i>Malus x domestica</i> isolate CoRTFJ-19 retrotransposon Ty1- <i>copia</i> reverse transcriptase gene partial cds	94	71	266
AJ411778	<i>Cicer arietinum</i> Ty1- <i>copia</i> like retrotransposon and partial RT gene for reverse transcriptase, clone TyCa9	94	85	267
AY684661	<i>Vigna radiata</i> clone VRC-39 copia-like retrotransposon reverse transcriptase gene, partial cds	90	82	279
AY684660	<i>Vigna radiata</i> clone VRC-38 copia-like retrotransposon reverse transcriptase gene, partial cds	90	84	279
DQ100159.1	<i>Camellia sinensis</i> transposon Ty1- <i>copia</i> -like reverse transcriptase gene, partial cds	94	70	267
EF581551.1	<i>Hypochoeris clarionoides</i> clone pHclty7 retrotransposon Ty1- <i>copia</i> -like reverse transcriptase (RT) gene, partial sequence	94	75	267
EF581724.1	<i>Hypochoeris sessiliflora</i> clone pHsety24 retrotransposon Ty1- <i>copia</i> -like reverse transcriptase (RT) gene, partial sequence	94	75	267
DQ376258.1	<i>Orobancha crenata</i> clone pOcrEty7 Ty1- <i>copia</i> -like retrotransposon reverse transcriptase gene, partial cds	94	73	264
DQ376293.1	<i>Orobancha owerinii</i> clone pOowty1 Ty1- <i>copia</i> -like retrotransposon reverse transcriptase gene, partial cds	94%	73%	264
DQ376258.1	<i>Orobancha crenata</i> clone pOcrGty19 Ty1- <i>copia</i> -like retrotransposon reverse transcriptase gene, partial cds	94%	73%	264

**Tabela 2:** Identificação das sequências depositadas no *GenBank* que apresentaram similaridade com a sequência do domínio RT do retrotransposon Ty3-gypsy-like isolada do genoma de *Vigna unguiculata*, com seus respectivos números do acesso, descrição, cobertura (%), identidade (%) e tamanho (em pb).

Número do acesso	Descrição	Query Coverage (%)	Identidade (%)	Tamanho da sequência (em pb)
AJ290595.1	<i>Picea abies</i> Ty3-gypsy like retroelement partial RT gene, clone pPa15ty	82%	66%	414
AJ290628.1	<i>Pinus sibirica</i> Ty3-gypsy like retroelement partial RT gene, clone pPsbgy6	86%	66%	409
GU976904.1	<i>Eleocharis palustris</i> clone gypEpal5 retrotransposon Ty3-gypsy RT gene, partial sequence	32%	70%	430
GU976843.1	<i>Eleocharis macrostachya</i> clone gypEmac60 retrotransposon Ty3-gypsy RT gene, partial cds	32%	70%	431
GU976942.1	<i>Eleocharis quinqueflora</i> clone gypEqui1 retrotransposon Ty3-gypsy RT gene, partial cds	81%	64%	431
GU976714.1	<i>Eleocharis acicularis</i> clone gypEaci8 retrotransposon Ty3-gypsy RT gene, partial cds	80%	66%	431
AY683011.1	<i>Vigna radiata</i> clone VRG-13 gypsy-like retrotransposon, partial sequence	98%	74%	416
AJ006003.1	<i>Cicer arietinum</i> gypsy like element CaRep2, clone pCaEr638 partial sequence	98%	70%	537
GQ367287.1	<i>Helianthus petiolaris</i> isolate 99XPET10 retrotransposon Ty3/gypsy-like RT-like gene, partial sequence	87%	68%	481
GQ367082.1	<i>Helianthus deserticola</i> isolate 82x7Edes retrotransposon Ty3/gypsy-like RT-like gene, partial sequence	98%	70%	481
DQ381021.1	<i>Orobanche crenata</i> clone pOcrGgy15 Ty3 gypsy retrotransposon	98%	68%	431
DQ381101.1	<i>Orobanche ramosa</i> clone pPraTgy14 Ty3 gypsy retrotransposon RT gene, partial cd	99%	67%	431

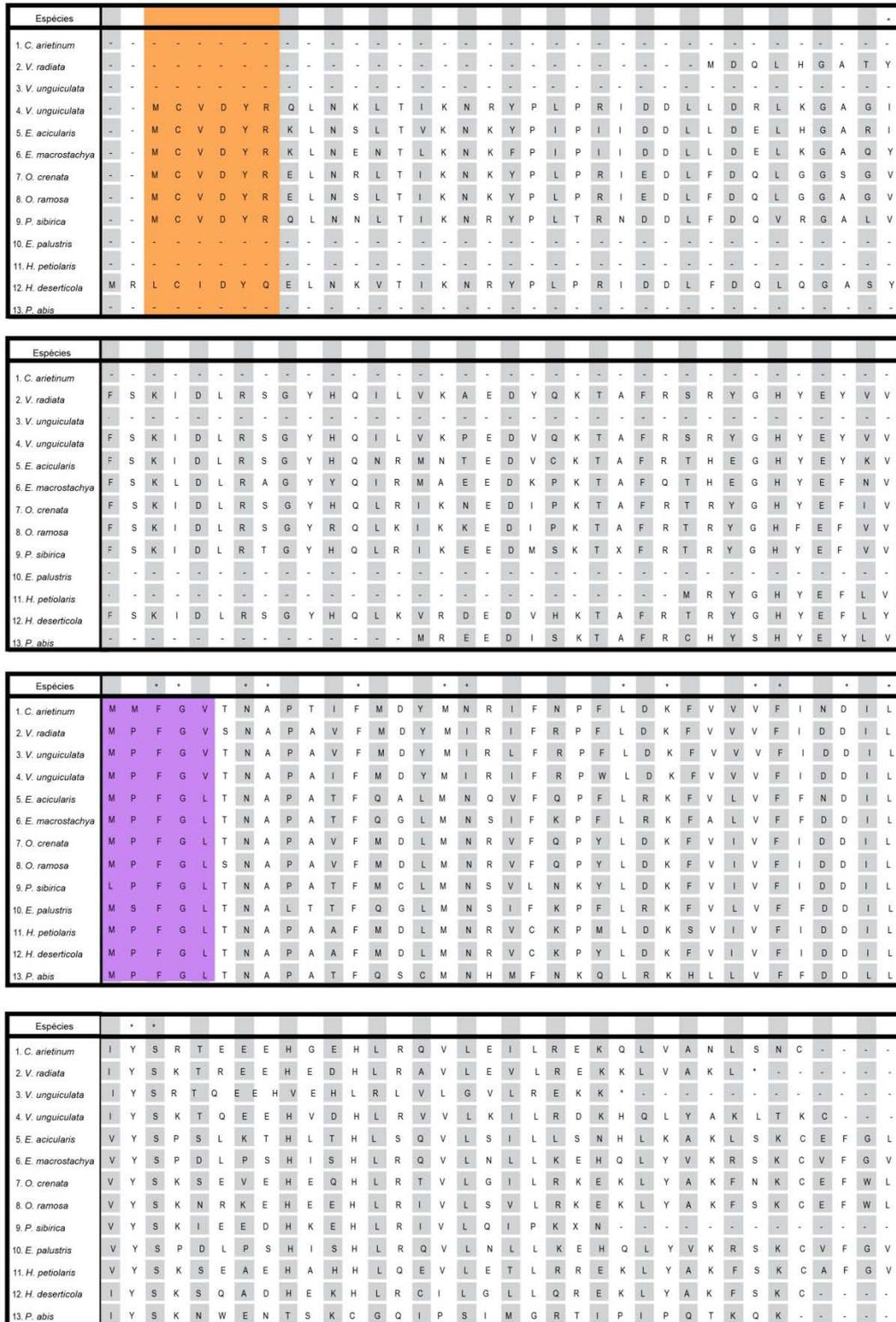


Considerando o retroelemento Ty3-gypsy-like, todas as sequências alinhadas apresentaram o motivo conservado MPFGV na porção central, inclusive *V. unguiculata* Ty3 clone 1 e *V. unguiculata* Ty3 clone 2. Contudo, a presença do motivo MCVDYR foi adicionalmente notada na porção inicial das sequências de *V. unguiculata* (Ty3 clone 2), *Eleocharis acicularis* (L.) Roem. & Schult., *E. macrostachya* (Britton), *Orobanche crenata* Forssk., *O. ramosa* L. e *Pinus sibirica* Du Tour (Figura 3).

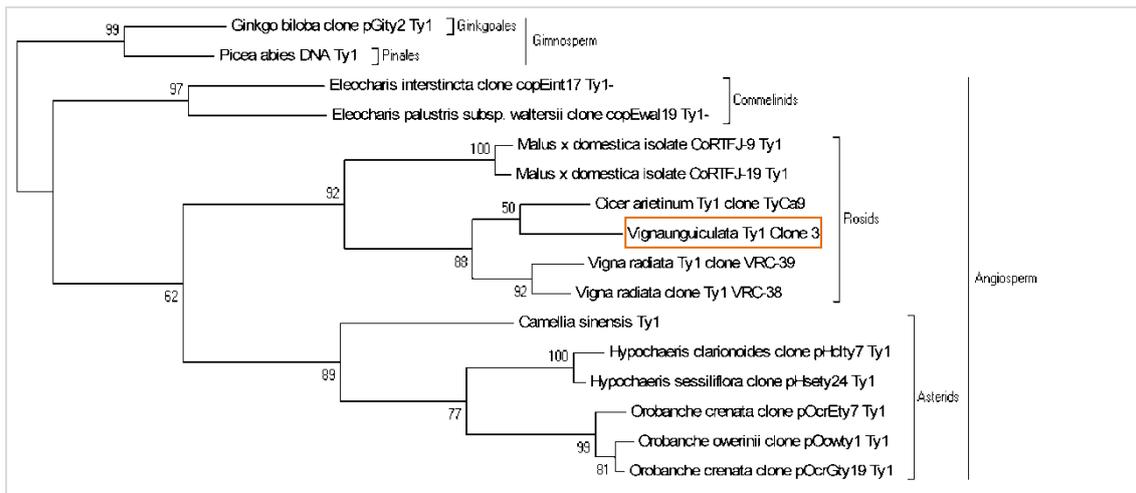
As análises filogenéticas mostraram que as sequências RT isoladas, como Ty1-copia-like e Ty3-gypsy-like, do genoma de *V. unguiculata* agruparam-se com retroelementos previamente descritos para as rosídeas (Figuras 4 e 5), cujas informações das sequências estão listadas nas Tabelas 1 e 2.

O dendrograma obtido para a sequência RT do Ty1-copia-like permitiu a observação de clados relativamente bem estruturados, com valores de *bootstrap* acima de 60, apresentando as sequências de *Ginkgo biloba* L. e *Picea abies* H.Karst., como grupo externo. A sequência RT de *V. unguiculata* (*Vigna unguiculata* Ty1 Clone 3) apresentou maiores valores de identidade com as leguminosas *Cicer arietinum* L. e *V. radiata* (Tabela 1), as quais, agruparam-se juntamente com *Malus x domestica*, mostrando um valor de *bootstrap* consistente (92) (Figura 4). Por fim, as sequências provenientes das espécies *Eleocharis interstincta* (Vahl) Roem. & Schult. e *E. palustris* (L.) Roem. & Schult. (monocotiledôneas) formaram um grupo irmão das eucotiledôneas altamente consistente (*bootstrap* 97), em uma posição mais basal.

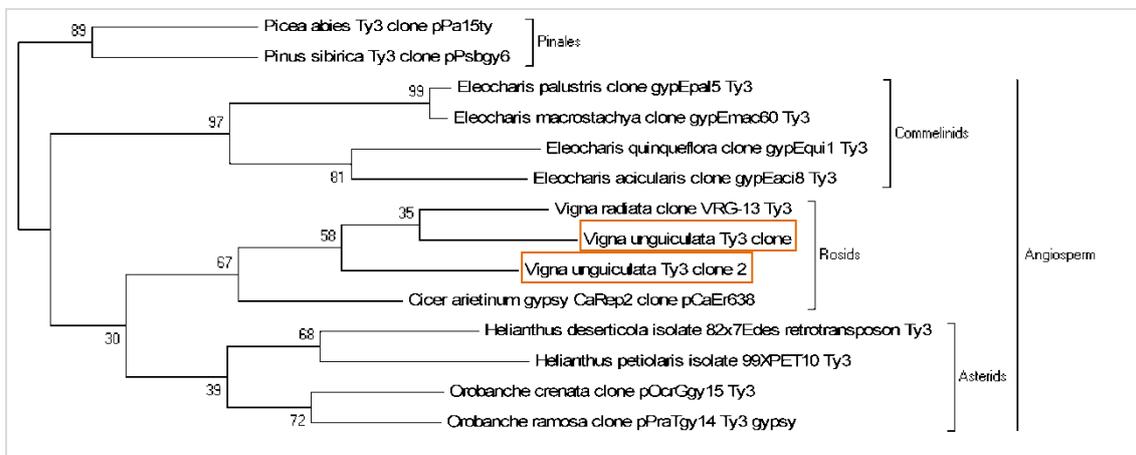
Considerando a árvore construída para as sequências RT do Ty3-gypsy-like, o grupo externo foi representado pelas sequências das espécies *P. abies* e *P. sibirica* (ordem Pinales), enquanto as sequências pertencentes ao gênero *Eleocharis* R. Br. constituíram o grupo irmão das eudicotiledôneas, suportado pelo valor de *bootstrap* igual a 97 (Figura 5). Neste dendrograma, as sequências *V. unguiculata* Ty3 clone 1 e *V. unguiculata* Ty3 clone 2 apresentaram-se mais proximamente relacionadas com *C. arietinum* e *V. radiata*, sendo agrupadas com as asterídeas por um *bootstrap* 30.



**Figura 3.** Alinhamento múltiplo por Clustal W das sequências do domínio RT do retroelemento Ty3-gypsy-like, destacando-se a presença da sequência consenso MCVDYR (domínio laranja) e MPFGV (domínio roxo), características dos retrotransposons do tipo gypsy.



**Figura 4:** Árvore filogenética construída por máxima parcimônia utilizando seqüências de nucleotídeos dos domínios RT dos retrotransposons Ty1-*copia*-like isolada de *Vigna unguiculata* (laranja) e depositadas no *GenBank* para outras espécies vegetais. Os números em cada nó correspondem aos valores percentuais de *bootstrap* relativos a 2000 réplicas.



**Figura 5:** Árvore filogenética construída por máxima parcimônia utilizando seqüências de nucleotídeos dos domínios RT dos retrotransposons Ty3-*gypsy*-like isolada de *Vigna unguiculata* (laranja) e depositadas no *GenBank* para outras espécies vegetais. Os números em cada nó correspondem aos valores percentuais de *bootstrap* relativos a 2000 réplicas.

## Distribuições das sequências RT de retroelementos Ty1-*copia*-like e Ty3-*gypsy*-like mediante FISH

A FISH com os domínios RT dos retroelementos Ty1-*copia*-like (*Vung* Ty1 clone 3) e Ty3-*gypsy*-like (*Vung* Ty3 clone 2) ressaltou marcações dispersas e pericentroméricas ao longo dos complementos cromossômicos de *G. max*, *P. vulgaris*, *P. lunatus*, *V. unguiculata* e *V. radiata*, estando ou não associadas a sítios de DNAr 5S e 45S (Figuras 6 e 7).

As espécies *G. max*, *P. vulgaris* e *P. lunatus* apresentaram padrões similares de distribuição do retroelemento Ty1-*copia*-like, com marcações dispersas ao longo dos pares cromossômicos. Contudo, na maioria dos cromossomos, foram visualizados sinais mais evidentes nas regiões pericentroméricas (Figura 6B, D, F). Nas referidas espécies, observou-se uma aparente colocalização dos sítios de DNAr 5S no par cromossômico 19 em soja (Figura 6B), 6 e 10 no feijão comum (Figura 6D), assim como no cromossomo 10 em feijão-fava (Figura 6F). Por sua vez, os sinais de DNAr 45S encontraram-se adjacentes às marcações do Ty1-*copia*-like nos taxa supracitados (Figura 6A, C e E).

Em *V. unguiculata* e *V. radiata*, pequenos *dots* bem definidos foram visualizados nas regiões intercalares e subteloméricas dos cromossomos (Figura 6H e J). Adicionalmente, marcações proximais foram notadas em menor quantidade em ambas as espécies. No feijão-caupi, os pequenos sinais de Ty1-*copia*-like colocalizaram aparentemente com os sítios de DNAr 5S e 45S (Figura 6G-H), enquanto que, em *V. radiata*, apenas os sinais de DNAr 45S estavam colocalizados com as marcações do retroelemento, como observado nos pares cromossômicos F e K (Figura 6I - J).

Similarmente ao notado para a sequência Ty1-*copia*-like, uma grande abundância de marcações pericentroméricas do retrotransposon Ty3-*gypsy*-like foi encontrada, juntamente com sinais dispersos, ao longo dos cromossomos de *G. max*, estando adjacentes com sítios de DNAr 5S e 45S (Figura 7A-B).

Em *P. vulgaris* e *P. lunatus*, uma disposição pericentromérica foi evidente para os sítios de Ty3-*gypsy*-like; entretanto, alguns sinais foram visualizados em posição intercalar e terminal (Figura 7D e F). Os sítios de DNAr 45S mostraram-se colocalizados aparentemente com as marcações deste retroelemento nos pares cromossômicos 6, 9 e 10, bem como no cromossômico 6 em *P. vulgaris* e *P. lunatus*, respectivamente (Figura 7C e E). Adicionalmente, o par cromossômico 10 do feijão-fava apresentou uma marcação bem característica envolvendo a extensão de seu braço

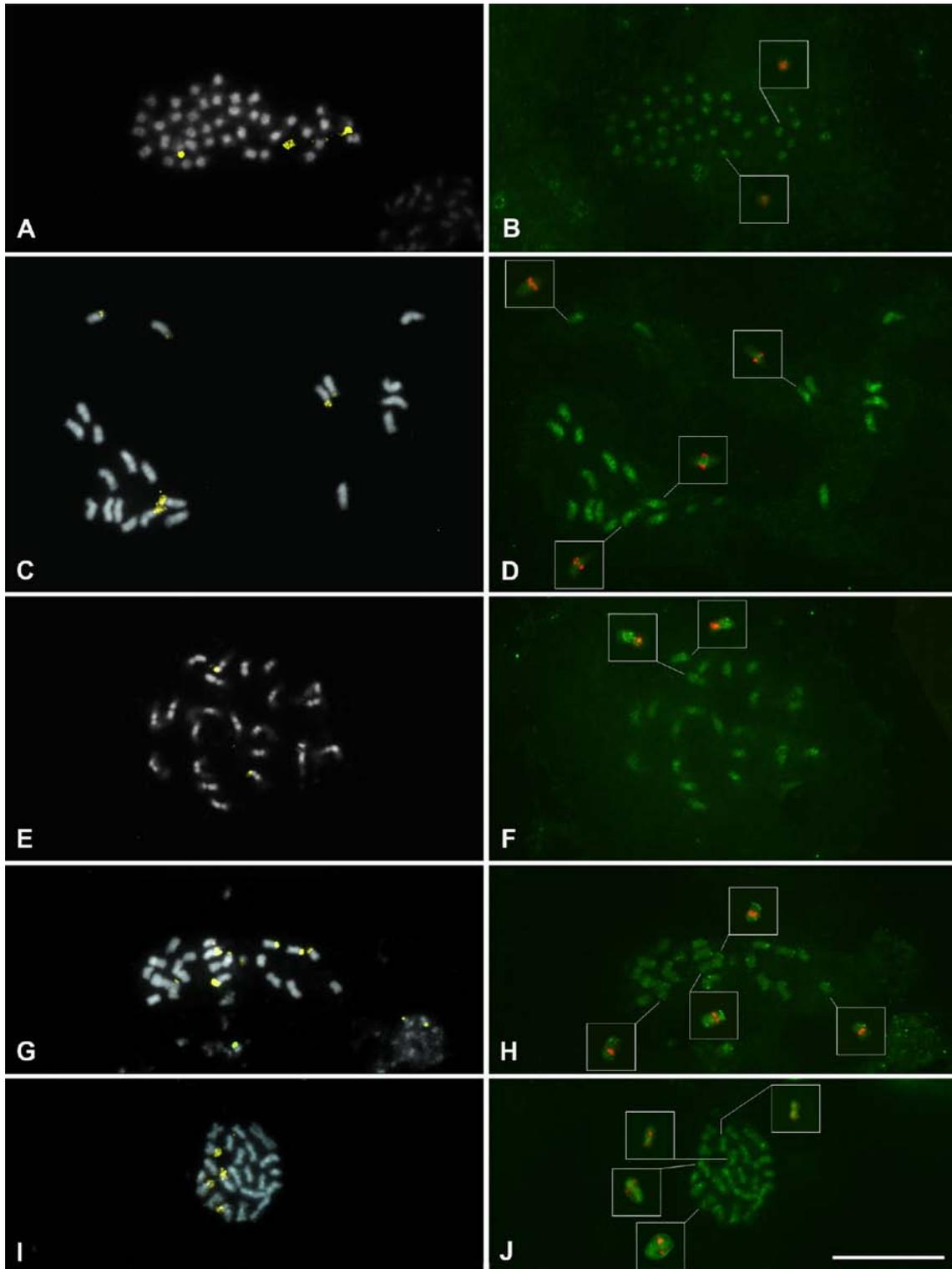
longo (Figura 7F). Com relação às espécies *V. unguiculata* e *V. radiata*, destacam-se marcações dispersas ao longo dos cromossomos (Fig. 7H; 7J), aparentemente colocalizadas com os sítios de DNAr 45S (dados não mostrados).

## Discussão

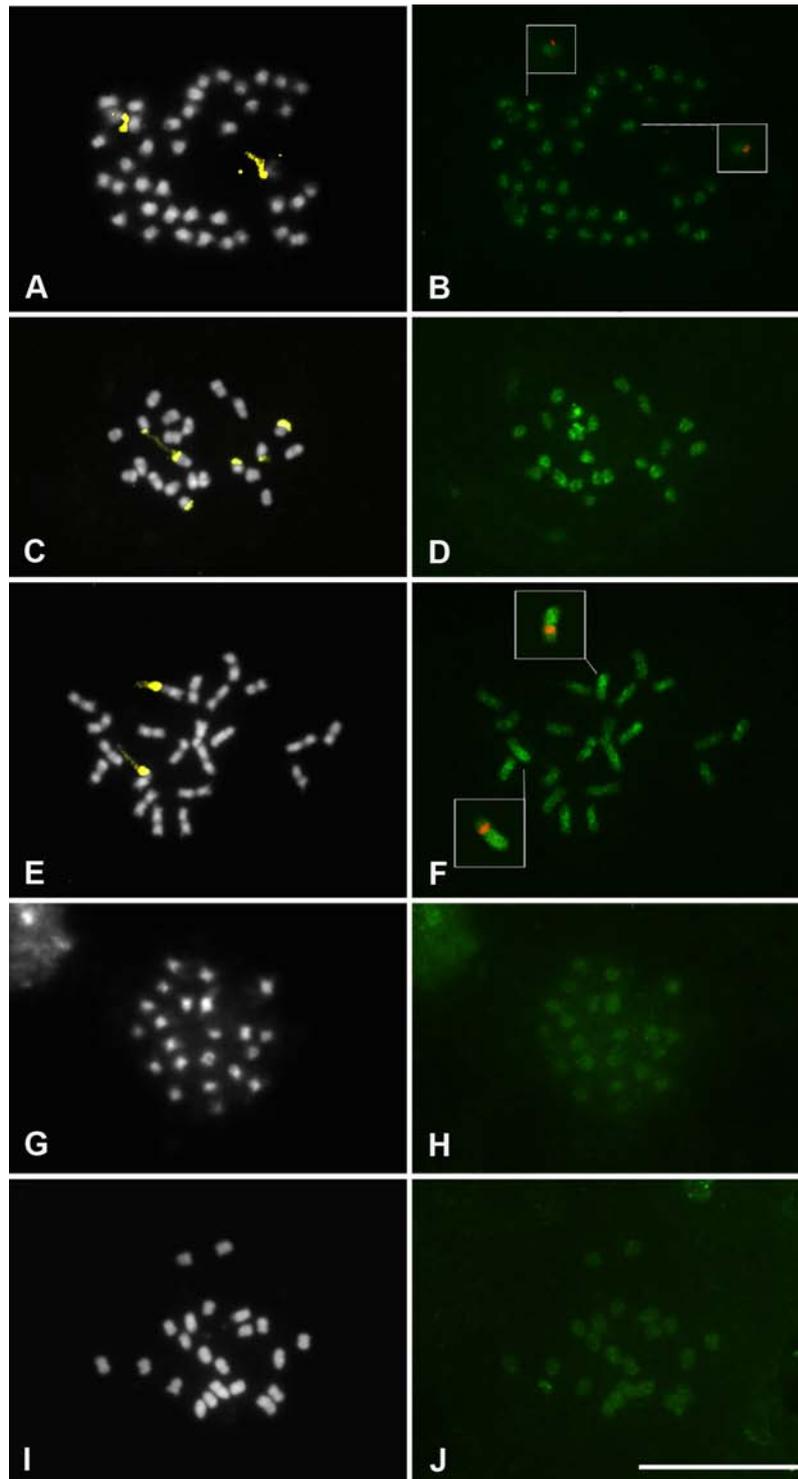
O isolamento e amplificação dos domínios RT dos retroelementos Ty1-*copia*-like e Ty3-*gypsy*-like do genoma de *V. unguiculata*, caracterizados pela presença de motivos específicos (Flavell *et al.*, 1992; Sun *et al.*, 2008), confirmaram a especificidade dos *primers* degenerados para esta finalidade, ressaltando o elevado nível de conservação do domínio RT em diferentes organismos vegetais, como relatado por Doolittle *et al.* (1989).

Entretanto, uma heterogeneidade significativa foi evidenciada entre as sequências isoladas no presente trabalho em relação às demais sequências depositadas no *GenBank*, para ambos os retroelementos. As sequências RT de Ty3-*gypsy*-like apresentaram-se mais conservadas, em comparação às sequências RT de Ty1-*copia*-like, conforme observado em espécies dos gêneros *Orobancha* L. e *Phelipanche* Pomel (Park *et al.*, 2007) e *M. domestica* (Sun *et al.*, 2008). Tal dado sugere que a elevada heterogeneidade dos elementos Ty1-*copia*-like é decorrente da ocorrência de eventos mais antigos de transposição das citadas sequências, resultando em uma ampla distribuição de diferentes famílias Ty1-*copia* entre as espécies hospedeiras (Hansen e Heslop-Harrison, 2004; Hill *et al.*, 2005). Além disso, deve-se levar em conta que as regiões não conservadas (extra-domínios) de tais sequências são mais susceptíveis ao estabelecimento de novas mutações, fato visível pelos valores baixos de bootstrap especialmente quando se comparam classes diferentes de plantas.

As análises filogenéticas dos RTs enfatizaram uma maior proximidade entre as sequências isoladas dos retrotransposons Ty1-*copia*-like e Ty3-*gypsy*-like de *V. unguiculata* com outras sequências pertencentes aos genomas de leguminosas, como *C. arietinum* e *V. radiata*, indicando que estes elementos existiam nos ancestrais do referido grupo, enquanto que os níveis de divergências entre elas permitem evidenciar uma amplificação recente de certos grupos de retroelementos durante a evolução e diversificação das leguminosas, conforme relatado por Dixit *et al.* (2006).



**Figura 6.** Células metafásicas de *Glycine max* (A-B), *Phaseolus vulgaris* (C-D), *P. lunatus* (E-F) *Vigna unguiculata* (G-H) e *V. radiata* (I-J) mostrando a distribuição dos sítios da sequência RT do retrotransposon Ty1-*copia*-like (em verde) e DNAr 5S (pseudocoloridos em amarelo). Os cromossomos foram contracolorados com DAPI e pseudocoloridos em cinza (A, C, E, G e I). Inserções indicam os cromossomos portadores de DNAr 5S (B, D, F, H e J). Barra em J = 5  $\mu$ m.



**Figura 7.** Células metafásicas de *Glycine max* (A-B), *Phaseolus vulgaris* (C-D), *P. lunatus* (E-F) *Vigna unguiculata* (G-H) e *V. radiata* (I-J) mostrando a distribuição dos sítios da sequência RT do retrotransposon Ty3-gypsy-like (em verde) e DNAr 45S (pseudocoloridos em amarelo). Os cromossomos foram contracorados com DAPI e pseudocoloridos em cinza (A, C, E, G e I). Inserções indicam os cromossomos portadores de DNAr 5S (B, D, F, H e J). Barra em J = 5  $\mu$ m.

No sentido mais amplo, as sequências RT isoladas apresentaram maior similaridade com retroelementos de eudicotiledôneas do que de monocotiledôneas, resultando em um agrupamento similar à filogenia pré-estabelecida para as espermatófitas, segundo o APG III (2009). Conseqüentemente, a topologia das árvores ressaltou a origem monofilética dos domínios RT dos retroelementos (Friesen *et al.*, 2001; Hansen e Heslop-Harrison, 2004), bem como sua provável transferência vertical (Kumar e Bennetzen, 1999; Kumekawa *et al.*, 1999).

A transferência vertical suporta a ubiquidade dos retrotransposons Ty1-*copia*-like e Ty3-*gypsy*-like nos genomas vegetais, uma vez que postula a existência de um ancestral comum com retroelementos similares, os quais foram transmitidos durante os eventos de especiação subsequentes, da linhagem parental para a progênie, juntamente com outros componentes do genoma. Tal mecanismo de transferência tem sido relatado para elementos Ty1-*copia*-like em briófitas, pteridófitas, angiospermas e gimnospermas (Flavell *et al.*, 1992; Voytas *et al.*, 1992; Kumar *et al.*, 1997).

Por outro lado, alguns estudos sugerem que a transmissão horizontal tem desempenhado um papel importante na evolução e dispersão dos retrotransposons entre diferentes espécies vegetais (Flavell *et al.* 1992; Cheng *et al.*, 2009). Todavia, a identificação de sequências semelhantes dos retroelementos Ty1-*copia*-like e Ty3-*gypsy*-like em diferentes espécies seria necessária para explicar como esses elementos movem-se entre espécies isoladas reprodutivamente.

Adicionalmente, a conservação do domínio RT permitiu a utilização deste fragmento dos retrotransposons, Ty1-*copia*-like e Ty3-*gypsy*-like, como sonda na FISH auxiliando na localização cromossômica desses elementos nos genomas das espécies analisadas. A análise da distribuição cromossômica mostra-se de grande valia para o entendimento do processo de expansão dos retrotransposons em diferentes espécies proximamente relacionadas, uma vez que esses elementos constituem uma fração bem representativa dos genomas vegetais (Lamb e Birchler, 2006).

Geralmente os retrotransposons apresentam uma organização cromossômica dispersa, a qual reflete o modo de amplificação e inserção desses elementos, sendo encontrados entre outras sequências repetitivas ou associados a regiões genômicas particulares (Kubis *et al.*, 1998; Heslop-Harrison, 2000). No presente trabalho, a predominância de sinais dispersos e pericentroméricos revelaram a ocorrência marcante dos elementos móveis Ty1-*copia*-like e Ty3-*gypsy*-like, ao longo dos cromossomos de *G. max*, *P. vulgaris*, *P. lunatus*, *V. unguiculata* e *V. radiata*.

Em todas as espécies analisadas, um padrão de distribuição disperso foi notado para o retroelemento Ty1-*copia*-like, fato comum descrito anteriormente para várias espécies, como *Beta vulgaris* L. (Weber *et al.*, 2010), *Copaifera langsdorffii* Desf. (Gaeta *et al.*, 2010) e *Helianthus annuus* L. (Natali *et al.*, 2006), enfatizando a afirmação que este grupo corresponde aos mais bem caracterizados retrotransposons em vegetais (Todorovska, 2007). Por outro lado, esses sinais dispersos mostraram uma clara associação com os loci de DNAr, bem como com as sequências de DNA satélite constituintes da heterocromatina terminal e pericentromérica, presentes nos genomas das leguminosas estudadas, sugerindo uma localização preferencial entre tais sequências.

Entretanto, tais dados contradizem a afirmação de que os elementos Ty1-*copia*, geralmente estão presentes em ambos os braços cromossômicos, com a exclusão dos centrômeros, telômeros e loci de DNAr 45S (Weber *et al.*, 2010), como observado anteriormente, em *V. unguiculata* por Galasso *et al.* (1997). Provavelmente, esses autores amplificaram e hibridizaram o domínio RT pertencente a uma família de Ty1-*copia* diferente da amplificada no presente trabalho, o que justificaria as divergências em termos de sequência de DNA e/ou organização genômica, salientando que é possível a presença de diferentes famílias no genoma de um organismo, tratando-se de uma característica geral das angiospermas (Fregonezi *et al.*, 2007).

Em *G. max*, a localização pericentromérica dos sítios do Ty1-*copia*-like e Ty3-*gypsy*-like foi bem característica, assim como coerente com os dados resultantes do sequenciamento genômico da referida espécie. Aproximadamente 42% do genoma da soja é constituído por retrotransposons do tipo LTR, compreendendo mais de 32.000 elementos, os quais estão depositados no banco de dados SoyTEdb (Du *et al.*, 2010b). Entre esses, foi estimada a presença de 510 famílias diferentes de retrotransposons distribuídas pericentromericamente, contendo 14106 elementos completos cujo tamanho variou de 1 a 21 Kb, sendo 31% e 69% classificados como *copia*-like e *gypsy*-like, respectivamente (Du *et al.*, 2010a; Schmutz *et al.*, 2010).

No presente trabalho, os sinais Ty3-*gypsy*-like visualizados, também, na região proximal dos cromossomos das espécies analisadas, indicam que as sequências desse retroelemento não estão agrupadas, mas dispersas, conforme observado para o Ty1-*copia*-like. Supostamente, estes sítios dos retroelementos LTR fazem parte da heterocromatina pericentromérica ressaltada pelo padrão de bandeamento CMA<sub>3</sub>/DAPI em *G. max* (CMA<sub>3</sub><sup>+</sup>), *V. unguiculata* (CMA<sub>3</sub><sup>+</sup>), *V. radiata* (DAPI<sup>+</sup>) (Bortoleti *et al.*, em

preparação), bem como em *P. vulgaris* e *P. lunatus* (CMA<sub>3</sub><sup>+</sup>) descrito por Almeida (2006).

A distribuição dos retroelementos Ty3-gypsy-like tem sido menos estudada, em comparação ao Ty1-copia-like (Natali *et al.*, 2006, Fregonezi *et al.*, 2007). Por acumularem-se preferencialmente nas regiões centroméricas (Weber *et al.*, 2010), a maioria destas sequências vem sendo denominada CRs (retrotransposons centroméricos) em diferentes espécies vegetais, como por exemplo, *Brassica* L. (Lim *et al.*, 2007), *G. max* (Tek *et al.*, 2010), *Lotus* L. e *Medicago* L. (Gorinsěk *et al.*, 2004), apresentando um elevado nível de conservação o que sugere a ocorrência de mecanismos de transposição recentes deste retroelemento, bem como seu papel essencial na manutenção e funcionamento do centrômero (Nagaki *et al.*, 2003).

A presença de retrotransposons em regiões heterocromáticas tem sido relatada (Dimitri *et al.*, 1999), sugerindo que esta classe de cromatina pode tolerar o acúmulo de elementos funcionais ou degenerados, uma vez que, geralmente, caracteriza-se como uma região genômica silenciada, acompanhada por metilação de DNA e modificações de histonas (Natali *et al.*, 2006; Weber *et al.*, 2010). Adicionalmente, o envolvimento da pressão seletiva contra a inserção dos retroelementos em regiões codificantes favorece esta localização preferencial na heterocromatina. A inserção em genes não é desejada, pois pode causar efeitos deletérios para a estrutura e o funcionamento de genes, modificações no padrão espacial e temporal da regulação gênica, bem como a origem de diferentes rearranjos cromossômicos (Rajput e Upadhyaya, 2009). Contudo, vale salientar que a associação preferencial heterocromática não é obrigatória; por exemplo, em *Arachis hypogaea* L., foi relatada a presença do retroelemento Ty3-gypsy-like (tipo *Athila*) em regiões eucromáticas, estando tais elementos ausentes nas regiões teloméricas e centroméricas (Nielen *et al.*, 2009).

A aparente colocalização de sítios de retroelementos e de DNAr 5S e 45S mostrou-se uma característica peculiar nas leguminosas estudadas, fato também observado em *Hordeum spontaneum* (K.Koch) Thell (Belyayev *et al.*, 2005), *Cestrum strigilatum* Ruiz & Pav. e *C. intermedium* Sendtn. (Fregonezi *et al.*, 2007) para o DNAr 45S. Os loci de DNAr têm sido apontados também como sítios preferenciais para a inserção de elementos móveis, os quais são eliminados pelo processo de *crossing-over* desigual entre cromátides-irmãs, conforme o modelo de evolução em concerto característico dos genes de RNAr, evitando alteração ou inativação dos genes de RNAr. Assim sendo, a evolução em concerto permite que os loci de DNAr funcionem

homogeneamente em face ao ataque contínuo de inserções destes elementos, além de favorecer a ativação transposicional, visto que facilita a interação dos retroelementos com vários sítios alvos (Eickbush e Eickbush, 2007).

Em suma, uma grande distribuição dos fragmentos RT dos retroelementos Ty1-*copia*-like e Ty3-*gypsy*-like foi notada em regiões preferenciais, como heterocromatina e loci de DNAr. Contudo, regiões de deleção ou amplificação dos elementos refletem uma distribuição diferenciada entre as espécies, como observado nas leguminosas estudadas. Provavelmente, as diferentes localizações cromossômicas, bem como a grande ubiquidade dos retrotransposons devem ser atribuídas à alta taxa de mutação e à baixa pressão seletiva, associadas ao mecanismo de retrotransposição, gerando diferentes famílias destes elementos, criando uma significativa diversidade em termos de estrutura populacional destas sequências, afetando a evolução estrutural dos genomas vegetais, especialmente de plantas superiores.

### **Agradecimentos**

Os autores agradecem a Anaregina Sousa de Araújo e Semiramis Jamil Hadad do Monte pertencente ao quadro da Universidade Federal do Piauí, por disponibilizar o sequenciador automático, contribuindo para o desenvolvimento deste trabalho, bem como ao CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) pelo apoio financeiro ao projeto.

### **Referências Bibliográficas**

- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25 (17): 3389-3402.
- Almeida CCS (2006) Mapeamento físico e análise evolutiva em *Phaseolus vulgaris* L. e *P. lunatus* L., utilizando hibridização *in situ* fluorescente (FISH). Tese de Doutorado, Universidade Federal de Pernambuco, Recife.
- APG III (2009) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Bot J Linnean Soc* 161: 105–121.
- Belyayev A, Raskina O, Nevo E (2005) Variability of the chromosomal distribution of Ty3-*gypsy* retrotransposons in the populations of two wild Triticeae species. *Cytogenet Genome Res* 109: 43–49 (2005).

- Bennetzen JL (2000) Transposable element contributions to plant gene and genome evolution. *Plant Mol Biol* 42: 251–269.
- Capy P, Bazin C, Higuete D, Langin T (1998) Dynamics and Evolution of Transposable Elements. New York: Chapman and Hall.
- Carvalho CR, Saraiva LS (1993) An air drying technique for maize chromosomes without enzymatic maceration. *Biotech Histochem* 68: 142-145.
- Chavanne F, Zhang D, Liaud M, Cerff R (1998) Structure and evolution of *Cyclops*: a novel giant retrotransposon of the *Ty3/Gypsy* family highly amplified in pea and other legume species. *Plant Mol Biol* 37: 363–375.
- Cheng X, Zhang D, Cheng Z, Keller B, Ling H (2009) A New Family of *Ty1-copia*-Like Retrotransposons Originated in the Tomato Genome by a Recent Horizontal Transfer Event. *Genetics* 181: 1183–1193.
- Dimitri P, Junakovic N (1999) Revising the selfish DNA hypothesis. New evidence on accumulation of transposable elements in heterocromatin. *Trends Genet* 15: 123-124.
- Dixit A, Ma KH, Yu JW, Cho EG, Park YJ (2006) Reverse transcriptase domain sequences from Mungbean (*Vigna radiata*) LTR retrotransposons: Sequence characterization and phylogenetic analysis. *Plant Cell Rep* 25: 100–111.
- Doolittle RF, Feng DF, Johnson MS, McClure MA (1989). Origins and evolutionary relationships of retroviruses. *Quarterly Rev Biol* 64: 1–29.
- Du J, Tian Z, Hans CS, Laten HM, Cannon SB, Jackson SA, Shoemaker RC, Ma J (2010a) Evolutionary conservation, diversity and specificity of LTR-retrotransposons in flowering plants: insights from genome-wide analysis and multi-specific comparison. *Plant J* 63: 584-598.
- Du J, Grant D, Tian Z, Nelson RT, Zhu L, Shoemaker RC, Ma J (2010b) SoyTEdb: a comprehensive database of transposable elements in the soybean genome. *BMC Genomics* 11: 113-119.
- Eickbush TH, Eickbush DG (2007) Finely Orchestrated Movements: Evolution of the Ribosomal RNA Genes. *Genetics* 175: 477–485.
- Flavell AJ, Dunbar E, Anderson R, Pearce SR, Hartley R, Kumar A (1992) *Ty1-copia* group retrotransposons are ubiquitous and heterogeneous in higher plants. *Nucleic Acids Res* 20: 3639–3644.
- Fregonezi JN, Vilas-Boas LA, Fungaro MHP, Gaeta ML, Vanzela ALL (2007) Distribution of a *Ty3/gypsy*-like retroelement on the A and B-chromosomes of

- Cestrum strigilatum* Ruiz & Pav. and *Cestrum intermedium* Sendtn. (Solanaceae). Genet Mol Biol 30: 599-604.
- Friesen N, Brandes A and Heslop-Harrison JS (2001) Diversity, origin, and distribution of retrotransposons (*gypsy* and *copia*) in conifers. Mol Biol Evol 18:1176-1188.
- Gaeta ML, Yuyama PM, Sartori D, Fungaro MHP, Vanzela ALL (2010) Occurrence and chromosome distribution of retroelements and NUPT sequences in *Copaifera langsdorffii* Desf. (Caesalpinioideae). Chromosome Res 18: 515-524.
- Galasso I, Harrison GE, Pignone D, Brandes A, Heslop-Harrison JS (1997) The distribution and organization of *Ty1-copia*-like retrotransposable elements in the genome of *Vigna unguiculata* (L.) Walp. (Cowpea) and its relatives. Ann Bot 80: 327–333.
- Gorinsěk B, Gubenšek F, KordiškD (2004) Evolutionary genomics of chromoviruses in eukaryotes. Mol Biol Evol 21:781–798.
- Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nuclei Acids Symp Ser 41: 95–98.
- Hansen CN, Heslop-Harrison JS (2004) Sequences and phylogenies of plant pararetroviruses, viruses and transposable elements. Adv Bot Res 41:165-193.
- Heslop-Harrison JS (2000) Comparative genome organization in plants: from sequence and markers to chromatin and chromosomes. Plant Cell 12: 617-635.
- Heslop-Harrison JS, Schwazarcher T, Anamthawat-Jónsson K, Leitch AR, Shi M (1991) *In situ* hybridization with automated chromosome desnaturation. Technique 3:109-115.
- Hill P, Burford D, Martin DM, Flavell AJ (2005) Retrotransposon populations of *Vicia* species with varying genome size. Mol Gen Genomics 273: 371–381.
- Kubis S, Schmidt T, Heslop-Harrison JS (1998) Repetitive DNA elements as a major component of plants genomes. Ann Bot 82: 45-55.
- Kumar A, Bennetzen JL (1999) Plant retrotransposons. Annu Rev Genet 33: 479-532.
- Kumar A, Pearce SR, McLean K, Harrison G, Heslop-Harrison JS, Waugh R, Flavell AJ (1997) The *Ty1-copia* group of retrotransposons in plants: genomic organisation, evolution, and use as molecular markers. Genetica 100: 205–217.
- Kumekawa N, Ohtsubo H, Horiuchi T, Ohtsubo E (1999) Identification and characterization of novel retrotransposons of the *gypsy* type in rice. Mol Gen Genet 260: 593-602.

- Lamb JC, Birchler JA (2006). Retroelement Genome Painting: Cytological Visualization of Retroelement Expansions in the Genera *Zea* and *Tripsacum*. *Genetics* 173: 1007–1021.
- Lim K, Yang T, Hwang Y, Kim JS, Park J, Kwon S, Kim J, Choi B, Lim M, Jin M, Kim H, DeJong H, Brancroft I, Lim Y, Park B (2007) Characterization of the centromere and peri-centromere retrotransposons in *Brassica rapa* and their distribution in related Brassica species. *Plant J* 49:173–183.
- Macas J, Neumann P, Navrátilová A (2007) Repetitive DNA in the pea (*Pisum sativum* L.) genome: comprehensive characterization using 454 sequencing and comparison to soybean and *Medicago truncatula*. *BMC Genomics* 8: 427-442.
- Marechal R, Mascherpa JM, Stainier F (1978) Étude taxonomique d'un groupe complexe d'espèces de genres *Phaseolus* et *Vigna* (*Papilionaceae*) sur la base de données morphologiques et polliniques, traitées par l'analyse informatique. *Boissiera* 28:1-273.
- Nagaki K, Song J, Stupar R, Parokonny AS, Yuan Q, Ouyang S, Liu J, Hsiao J, Jones KM, Dawe RK, Buell CR, Jiang J (2003) Molecular and cytological analyses of large tracks of centromeric DNA reveal the structure and evolutionary dynamics of maize centromeres. *Genetics* 163: 759–770.
- Natali L, Santini S, Giordani T, Minelli S, Maestrini P, Cionini PG, Cavallini A (2006) Distribution of *Ty3-gypsy*- and *Ty1-copia*-like DNA sequences in the genus *Helianthus* and other Asteraceae. *Genome* 49: 64–72.
- Nielen S, Campos-Fonseca F, Guimarães P, Leal-Bertioli SC, Bertioli D (2009) Isolation and Characterization of Retrotransposons in Wild and Cultivated Peanut Species. *Food and Agric Organiz United Nations*: 439-442.
- Park JM, Schneeweiss GM, Weiss-Schneeweiss H (2007) Diversity and evolution of *Ty1-copia* and *Ty3-gypsy* retroelements in the non-photosynthetic flowering plants *Orobanche* and *Phelipanche* (*Orobanchaceae*). *Gene* 387: 75–86.
- Pedrosa A, Jantsch MF, Moscone EA, Ambros PF, Schweizer D (2001) Characterisation of pericentromeric and sticky intercalary heterochromatin in *Ornithogalum longibracteatum* (*Hyacinthaceae*). *Chromosoma* 110: 203-213.
- Pedrosa A, Sandal N, Stougaard J, Schweizer D, Bachmair A (2002) Chromosomal map of the model legume *Lotus japonicus*. *Genetics* 161: 1661-1672.
- Pedrosa-Harand A, Porch T, Gepts P (2008) Standard nomenclature for common bean chromosomes and linkage groups. *Annu Rep Bean Improv Coop* 51: 106-107.

- Rajput MJ, Upadhyaya EKC (2009) CARE1, a TY3-gypsy like LTR-retrotransposon in the food legume chickpea (*Cicer arietinum* L.). *Genetica* 136: 429–437.
- Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang X, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA (2010) Genome sequence of the paleopolyploid soybean. *Nature* 463:178-183.
- Sun H, Dai H, Zhao G, Ma Y, Ou C, Li H, Li L, Zhang Z (2008) Genome-wide Characterization of Long Terminal Repeat -retrotransposons in Apple Reveals the Differences in Heterogeneity and Copy Number between Ty1-*copia* and Ty3-*gypsy* Retrotransposons. *J Integ Plant Biol* 50: 1130–1139.
- Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24: 1596-1599.
- Tek LA, Kashihara K, Murata M, Nagaki K (2010) Functional centromeres in soybean include two distinct tandem repeats and a retrotransposon. *Chromosome Res* 18: 337-347.
- Thompson JD, Higgins DG, Gibson TJ (1994) Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22: 4673–4680.
- Todorovska E (2007) Retrotransposons and their role in plant – genome evolution. *Biotechnol. & Biotechnol. EQ* 21: 294–305.
- Ungerer MC, Strakosh SC, Stimpson KM (2009) Proliferation of Ty3/gypsy-like retrotransposons in hybrid sunflower taxa inferred from phylogenetic data. *BMC Biology* 7: 40-52.
- Vershinin AV, Druka A, Alkhimova AG, Kleinhofs A, Heslop-Harrison JS (2002) LINEs and gypsy-like retrotransposons in *Hordeum* species. *Plant Mol Biol* 49: 1–14.
- Voytas DF, Cummings MP, Konieczny A, Ausubel FM, Rodermel SR (1992) *Copia*-like retrotransposons are ubiquitous among plants. *Proc Natl Acad Sci (USA)* 89: 7124–7128.

- Wanzenböck EM, Schöfer C, Schweizer D, Bachmair A (1997) Ribosomal transcription units integrated via T-DNA transformation associate with the nucleolus and do not require upstream repeat sequences for activity in *Arabidopsis thaliana*. *Plant J* 11: 1007-1016.
- Weber B, Wenke T, Frömmel U, Schmidt T, Heitkam T (2010) The Ty1-*copia* families SALIRE and Cotzilla populating the *Beta vulgaris* genome show remarkable differences in abundance, chromosomal distribution, and age. *Chromosome Res* 18: 247–263.
- Weising K, Nybom H, Wolf K, Meyer W (1995) DNA fingerprinting in plants and fungi. Boca Raton: CRC Press, 322p.
- Xiao WM, Sakamoto W, Sodmergen (2004) Isolation and characterization of Ty1/*copia*-like reverse transcriptase sequences from mung bean. *Acta Bot Sin* 46: 582–587.

## **Conclusões**

## 7. CONCLUSÕES

1. Os microssatélites (AG)<sub>8</sub>, (ACC)<sub>5</sub>, (AAG)<sub>5</sub>, (CTC)<sub>5</sub> e (TGA)<sub>6</sub> apresentam-se como importantes constituintes da fração repetitiva dos genomas de *G. max*, *P. vulgaris*, *P. lunatus*, *V. unguiculata* e *V. radiata*, havendo variações interespecíficas em suas quantidades, organização (em tandem ou dispersa) e distribuição ao longo dos cromossomos.

2. A utilização de marcadores cromossômicos para a região pericentromérica dos cromossomos de soja, como os sítios de FISH dos microssatélites (ACC)<sub>5</sub>, (AAG)<sub>5</sub>, (CTC)<sub>5</sub> e (TGA)<sub>6</sub> e a presença de bandas CMA<sub>3</sub><sup>+</sup>, evidencia que as marcações de SSRs *in situ* estejam associadas às regiões de heterocromatina, sendo possivelmente uma fração constituinte dos 1.148 *scaffolds* existentes no sequenciamento da soja, não ancorados a plataforma do SoyBase.

3. A análise comparativa dos mapas físico de contigs e citogenético de *G. max* ressaltou uma distribuição diferencial não aleatória das diferentes classes de SSRs analisadas, estando estas preferencialmente associadas à heterocromatina ou à eucromatina, enfatizando assim o poder complementar da integração das análises *in silico* e *in situ*, o que facilita um melhor entendimento da estrutura física e organização genômica.

4. As sequências RT do retroelemento Ty1-*copia*-like apresentaram-se mais heterogêneas, em comparação ao Ty3-*gypsy*-like, quando comparadas com sequências de outras espécies, sugerindo que esses elementos sejam constituintes mais antigos dos genomas das espécies hospedeiras.

5. As análises filogenéticas das sequências RT dos retrotransposons Ty1-*copia*-like e Ty3-*gypsy*-like mostraram uma similaridade com retroelementos descritos para rosídeas, enfatizando a origem monofilética dos domínios RT destes retrotransposons, assim como sua provável transmissão vertical durante os processos de especiação ocorrentes na evolução vegetal.

6. A grande ubiquidade dos retrotransposons Ty1-*copia*-like e Ty3-*gypsy*-like, bem como suas diferentes localizações cromossômicas, podem ser atribuídas à alta taxa de mutação e baixa pressão seletiva, associadas ao mecanismo de transposição do tipo *copy and paste*, característico desta classe de retroelementos.

***Anexos***

---

## 8. ANEXOS

### 8.1 INFORMAÇÃO SUPLEMENTAR – CAPÍTULO I

**Tabela 1.** Repetições de [(AG)<sub>8</sub>]<sub>15</sub> identificadas contendo as informações individuais sobre a posição exata no genoma, tamanho da repetição em pb, bem como os genes e elementos transponíveis associados.

Código	Região (AG) <sub>8</sub>	Tamanho (bp)	Densidade gênica	Gene models	Gene models (transposon-like)
(AG) <sub>8</sub> -1.1	Gm01:2578634..2578735	101	Moderada	-	-
(AG) <sub>8</sub> -1.2	Gm01:7220426..7220548	122	Alta	-	-
(AG) <sub>8</sub> -1.3	Gm01:20044064..20044191	127	Moderada	-	-
(AG) <sub>8</sub> -1.4	Gm01:34721577..34721672	95	Moderada	-	-
(AG) <sub>8</sub> -1.5	Gm01:38773016..38773141	125	Moderada	-	-
(AG) <sub>8</sub> -2.1	Gm02:1370407..1370736	329	Alta	-	-
(AG) <sub>8</sub> -2.2	Gm02:1617173..1617016	-157	Alta	-	-
(AG) <sub>8</sub> -2.3	Gm02:15741647..15741759	112	Alta	Glyma02g17490 (Beta-glucosidase, lactase phlorizinhidrolase, and related proteins / Glycosyl hydrolase family 1)	-
(AG) <sub>8</sub> -2.4	Gm02:38267342..38267218	-124	Moderada	-	-
(AG) <sub>8</sub> -3.1	Gm03:3912435..3912556	121	Moderada	-	-
(AG) <sub>8</sub> -3.2	Gm03:3932645..3932808	163	Moderada	-	-
(AG) <sub>8</sub> -3.3	Gm03:12582228..12582044	-184	Moderada	-	-
(AG) <sub>8</sub> -3.4	Gm03:39809872..39810010	138	Alta	-	-
(AG) <sub>8</sub> -3.5	Gm03:41098368..41098215	-153	Alta	-	-
(AG) <sub>8</sub> -4.1	Gm04:3128321..3128428	107	Alta	-	-
(AG) <sub>8</sub> -4.2	Gm04:17342139..17342220	81	Moderada	-	-
(AG) <sub>8</sub> -4.3	Gm04:17780728..17780861	133	Moderada	-	-
(AG) <sub>8</sub> -4.4	Gm04:18312866..18312726	-140	Moderada	-	-
(AG) <sub>8</sub> -4.5	Gm04:40449464..40449379	-85	Alta	-	-
(AG) <sub>8</sub> -5.1	Gm05:1704036..1703924	-112	Alta	-	-
(AG) <sub>8</sub> -5.2	Gm05:3750549..3750549	-147	Moderada	-	-
(AG) <sub>8</sub> -5.3	Gm05:24631461..24631362	-99	Moderada	-	-
(AG) <sub>8</sub> -5.4	Gm05:28964812..28964649	-163	Moderada	-	-
(AG) <sub>8</sub> -6.1	Gm06:13690329..13690203	-126	Alta	-	-
(AG) <sub>8</sub> -6.2	Gm06:19266574..19266714	140	Moderada	-	-
(AG) <sub>8</sub> -6.3	Gm06:38040199..38040026	-173	Alta	-	-
(AG) <sub>8</sub> -6.4	Gm06:38274323..38274176	-147	Moderada	Glyma06g36190 (Serine/Threonine-Protein Kinase WNK (WITH NO LYSINE)-RELATED)	-
(AG) <sub>8</sub> -6.5	Gm06:38937436..38937565	129	Moderada	-	-
(AG) <sub>8</sub> -7.1	Gm07:1092569..1092699	130	Alta	-	-
(AG) <sub>8</sub> -7.2	Gm07:9583953..9584075	122	Alta	Glyma07g11420 (RAS-RELATED GTPASE / GTPase Rab11/YPT3, small G protein superfamily / Miro-like protein)	-
(AG) <sub>8</sub> -7.3	Gm07:11720225..11720050	-175	Moderada	-	-
(AG) <sub>8</sub> -7.4	Gm07:14231705..14231556	-149	Moderada	-	-

(AG) <sub>8</sub> -7.5	Gm07:40793910..40794055	145	Moderada	-	-
(AG) <sub>8</sub> -7.6	Gm07:41855660..41855529	-131	Alta	-	-
(AG) <sub>8</sub> -8.1	Gm08:17354904..17354786	-118	Alta	-	-
(AG) <sub>8</sub> -8.2	Gm08:21881189..21881347	158	Alta	-	-
(AG) <sub>8</sub> -8.3	Gm08:39215303..39215185	-118	Alta	-	-
(AG) <sub>8</sub> -8.4	Gm08:42239209..42239327	118	Alta	Glyma08g42250 (Glycosyl hydrolases family 16)	-
(AG) <sub>8</sub> -8.5	Gm08:44522145..44522268	123	Alta	-	-
(AG) <sub>8</sub> -9.1	Gm09:2636865..2637026	161	Moderada	-	-
(AG) <sub>8</sub> -9.2	Gm09:3037293..3037197	-96	Alta	-	-
(AG) <sub>8</sub> -9.3	Gm09:5755495..5755585	90	Alta	-	-
(AG) <sub>8</sub> -9.4	Gm09:20035280..20035195	-85	Moderada	-	-
(AG) <sub>8</sub> -9.5	Gm09:36198825..36198919	94	Alta	Glyma09g29260 (ACTIN-RELATED)	-
(AG) <sub>8</sub> -10.1	Gm10:4972612..4972721	109	Alta	-	-
(AG) <sub>8</sub> -10.2	Gm10:8904910..8905033	123	Moderada	-	-
(AG) <sub>8</sub> -10.3	Gm10:9691175..9691080	-95	Moderada	-	-
(AG) <sub>8</sub> -10.4	Gm10:20259386..20259278	-108	Moderada	-	-
(AG) <sub>8</sub> -10.5	Gm10:41132363..41132461	98	Alta	-	-
(AG) <sub>8</sub> -11.1	Gm11:5780501..5780400	-101	Alta	-	-
(AG) <sub>8</sub> -11.2	Gm11:6737436..6737295	-141	Alta	-	-
(AG) <sub>8</sub> -11.3	Gm11:16725604..16725727	123	Alta	-	-
(AG) <sub>8</sub> -11.4	Gm11:24368060..24367902	-158	Moderada	-	-
(AG) <sub>8</sub> -11.5	Gm11:30105558..30105672	114	Moderada	-	-
(AG) <sub>8</sub> -12.1	Gm12:13321676..13321596	-80	Moderada	-	-
(AG) <sub>8</sub> -12.2	Gm12:14970273..14970273	56	Alta	-	-
(AG) <sub>8</sub> -12.3	Gm12:15780128..15780056	-72	Moderada	-	-
(AG) <sub>8</sub> -12.4	Gm12:23658700..23658561	-139	Moderada	-	-
(AG) <sub>8</sub> -13.1	Gm13:1599891..1599793	-98	Alta	-	-
(AG) <sub>8</sub> -13.2	Gm13:19473209..19473053	-156	Moderada	Glyma13g15590 (Apoptotic ATPase / Leucine Rich Repeat)	-
(AG) <sub>8</sub> -13.3	Gm13:26097377..26097243	-134	Alta	-	-
(AG) <sub>8</sub> -13.4	Gm13:34028182..34028315	133	Alta	-	-
(AG) <sub>8</sub> -14.1	Gm14:17591919..17591825	-94	Moderada	-	-
(AG) <sub>8</sub> -14.2	Gm14:37718099..37717977	-127	Moderada	Glyma14g30940 (STROMAL CELL PROTEIN/NODULIN MTN3-RELATED / Multitransmembrane protein / MtN3/saliva family)	-
(AG) <sub>8</sub> -14.3	Gm14:41473529..41473529	-139	Moderada	-	-
(AG) <sub>8</sub> -14.4	Gm14:44794548..44794657	109	Alta	Glyma14g35730 (Mitochondrial tricarboxylate/dicarbox ylate carrier proteins)	-
(AG) <sub>8</sub> -14.5	Gm14:46350211..46350104	-107	Alta	-	-
(AG) <sub>8</sub> -15.1	Gm15:6725015..6724903	112	Alta	-	-
(AG) <sub>8</sub> -15.2	Gm15:10792603..10792730	127	Alta	-	-
(AG) <sub>8</sub> -15.3	Gm15:14621921..14622170	249	Moderada	-	-
(AG) <sub>8</sub> -15.4	Gm15:15706833..15706724	-109	Alta	-	-
(AG) <sub>8</sub> -15.5	Gm15:37701239..37701361	122	Moderada	Glyma15g33780	-
(AG) <sub>8</sub> -16.1	Gm16:6729147..6729070	-77	Moderada	-	-
(AG) <sub>8</sub> -16.2	Gm16:18711308..18711168	-140	Moderada	-	-
(AG) <sub>8</sub> -16.3	Gm16:23188516..23188599	83	Moderada	-	-
(AG) <sub>8</sub> -16.4	Gm16:28049147..28049300	153	Moderada	-	-
(AG) <sub>8</sub> -17.1	Gm17:19474110..19474202	92	Alta	Glyma15g21240 (CYTOSOLIC PURINE 5-	-

				NUCLEOTIDASE-RELATED / IMP-GMP specific 5'-nucleotidase / 5' nucleotidase family)	
(AG) <sub>8</sub> -17.2	Gm17:35271386..35271217	-169	Alta	Glyma17g20450 (CYTOSOLIC PURINE 5-NUCLEOTIDASE-RELATED / IMP-GMP specific 5'-nucleotidase / 5' nucleotidase family)	-
(AG) <sub>8</sub> -17.3	Gm17:38588799..38588675	-124	Alta	-	-
(AG) <sub>8</sub> -17.4	Gm17:38621858..38621949	91	Alta	-	-
(AG) <sub>8</sub> -18.1	Gm18:2275280..2275034	-246	Alta	-	-
(AG) <sub>8</sub> -18.2	Gm18:18728696..18728864	168	Moderada	-	-
(AG) <sub>8</sub> -18.3	Gm18:36619321..36619070	-251	Moderada	-	-
(AG) <sub>8</sub> -18.4	Gm18:46165176..46165045	-131	Moderada	-	-
(AG) <sub>8</sub> -19.1	Gm19:187636..187505	-131	Alta	-	-
(AG) <sub>8</sub> -19.2	Gm19:3601753..3601833	80	Alta	-	-
(AG) <sub>8</sub> -19.3	Gm19:8810999..8810905	-94	Alta	-	-
(AG) <sub>8</sub> -19.4	Gm19:17986912..17986767	-145	Moderada	-	Glyma19g14910
(AG) <sub>8</sub> -20.1	Gm20:21300838..21300838	89	Moderada	-	-
(AG) <sub>8</sub> -20.2	Gm20:22272312..22272170	-142	Moderada	-	-
(AG) <sub>8</sub> -20.3	Gm20:31931793..31931921	128	Alta	-	-
(AG) <sub>8</sub> -20.4	Gm20:32591157..32591249	92	Alta	-	-

**Tabela 2.** Repetições de [(AAG)<sub>5</sub>]<sub>10</sub> identificadas contendo as informações individuais sobre a posição exata no genoma, tamanho da repetição em pb, bem como os genes e elementos transponíveis associados.

Código	Região (AAG) <sub>5</sub>	Tamanho (bp)	Densidade gênica	Gene models	Gene models (transposon-like)
(AAG) <sub>5</sub> -1.1	Gm01:6146685..6146791	106	Moderada	-	-
(AAG) <sub>5</sub> -1.2	Gm01:15981319..15981272	-47	Moderada	-	-
(AAG) <sub>5</sub> -1.3	Gm01:28716889..28717046	157	Moderada	-	-
(AAG) <sub>5</sub> -1.4	Gm01:30816228..30816154	-74	Moderada	-	-
(AAG) <sub>5</sub> -1.5	Gm01:42278590..42278516	-74	Moderada	-	-
(AAG) <sub>5</sub> -1.6	Gm01:42451856..42451747	-109	Moderada	-	-
(AAG) <sub>5</sub> -1.7	Gm01:55427375..55427317	-58	Moderada	Glyma01g44970 (Beta Catenin-related armadillo repeat-containing/HEAT repeat)	-
(AAG) <sub>5</sub> -2.1	Gm02: 608692.. 608692	-83	Alta	-	-
(AAG) <sub>5</sub> -2.2	Gm02: 688006.. 688006	117	Alta	Glyma02g00890/ Glyma02g00900	-
(AAG) <sub>5</sub> -2.3	Gm02:21153681..21153635	-46	Baixa	-	-
(AAG) <sub>5</sub> -2.4	Gm02:31111474..31111568	94	Moderada	-	-
(AAG) <sub>5</sub> -2.5	Gm02:33659668..33659774	106	Moderada	-	-
(AAG) <sub>5</sub> -2.6	Gm02:38906637..38906558	-79	Moderada	-	-
(AAG) <sub>5</sub> -3.1	Gm03:6971872..6971922	50	Moderada	-	-
(AAG) <sub>5</sub> -3.2	Gm03:16982990..16982884	-106	Ausência de genes	-	-
(AAG) <sub>5</sub> -3.3	Gm03:17385685..17385558	-127	Moderada	-	-
(AAG) <sub>5</sub> -3.4	Gm03:28029837..28029891	54	Moderada	-	-
(AAG) <sub>5</sub> -3.5	Gm03:28699785..28699680	-105	Alta	-	-
(AAG) <sub>5</sub> -3.6	Gm03:32205130..32205183	53	Moderada	-	-
(AAG) <sub>5</sub> -3.7	Gm03:34061615..34061665	50	Alta	-	-
(AAG) <sub>5</sub> -3.8	Gm03:34979890..34979766	-124	Moderada	Glyma03g27290 (RNA-binding protein musashi/mRNA cleavage and polyadenylation factor I complex, subunit HRP1)	-
(AAG) <sub>5</sub> -4.1	Gm04:9317887.. 9317823	-64	Alta	-	-
(AAG) <sub>5</sub> -4.2	Gm04:15026628..15026574	-54	Moderada	-	-
(AAG) <sub>5</sub> -4.3	Gm04:23901217..23901111	-106	Moderada	-	-
(AAG) <sub>5</sub> -4.4	Gm04:25350971..25350872	-99	Moderada	-	-
(AAG) <sub>5</sub> -4.5	Gm04:28368843..28368759	-84	Baixa	-	-
(AAG) <sub>5</sub> -4.6	Gm04:39373516..39373419	-97	Baixa	-	-
(AAG) <sub>5</sub> -5.1	Gm05:1721347..1721467	120	Alta	Glyma05g02340 (U3 small nucleolar ribonucleoprotein (snoRNP) subunit - Mpp10p)	-
(AAG) <sub>5</sub> -5.2	Gm05:3571626..3571555	-71	Alta	-	-

(AAG) <sub>5</sub> -5.3	Gm05:30350245..30350297	57	Moderada	-	-
(AAG) <sub>5</sub> -5.4	Gm05:30825566..30825505	-143	Alta	-	-
(AAG) <sub>5</sub> -6.1	Gm06: 639371.. 639166	-205	Alta	Glyma06g01070 (Nuclear Receptor Binding Set )	-
(AAG) <sub>5</sub> -6.2	Gm06: 9220023.. 9220126	103	Alta	-	-
(AAG) <sub>5</sub> -6.3	Gm06:19272159..19272215	56	Moderada	-	-
(AAG) <sub>5</sub> -6.4	Gm06:33546748..33546603	-145	Alta	-	-
(AAG) <sub>5</sub> -6.5	Gm06:36061990..36061918	-72	Moderada	-	-
(AAG) <sub>5</sub> -6.6	Gm06:45592332..45592280	-52	Moderada	-	-
(AAG) <sub>5</sub> -6.7	Gm06:49344483..49344430	-53	Alta	-	-
(AAG) <sub>5</sub> -7.1	Gm07:8008453..8008366	-87	Baixa	-	-
(AAG) <sub>5</sub> -7.2	Gm07:10112517..10112593	76	Baixa	Glyma07g11850	-
(AAG) <sub>5</sub> -7.3	Gm07:20033619..20033666	47	Baixa	-	-
(AAG) <sub>5</sub> -7.4	Gm07:33574977..33575146	169	Moderada	-	-
(AAG) <sub>5</sub> -7.5	Gm07:34717074..34717171	97	Moderada	Glyma07g29700	-
(AAG) <sub>5</sub> -7.6	Gm07:35754634..35754748	114	Alta	-	-
(AAG) <sub>5</sub> -7.7	Gm07:35785253..35785335	82	Alta	Glyma07g30840 (Putative transcription factor 5qNCA, contains JmjC domain)	-
(AAG) <sub>5</sub> -8.1	Gm08:7901288..7901380	92	Alta	Glyma08g10860	-
(AAG) <sub>5</sub> -8.2	Gm08:23896817..23896941	124	Moderada	-	-
(AAG) <sub>5</sub> -8.3	Gm08:25705785..25705664	-121	Moderada	-	-
(AAG) <sub>5</sub> -8.4	Gm08:31125151..31125275	124	Moderada	-	-
(AAG) <sub>5</sub> -8.5	Gm08:32602409..32602533	124	Ausência de genes	-	-
(AAG) <sub>5</sub> -8.6	Gm08:33649211..33649145	-66	Moderada	-	-
(AAG) <sub>5</sub> -8.7	Gm08:33654020..33653948	-72	Moderada	-	-
(AAG) <sub>5</sub> -8.8	Gm08:38107984..38107916	-68	Moderada	-	-
(AAG) <sub>5</sub> -8.9	Gm08:46813623..46813714	91	Alta	-	Glyma08g48 110
(AAG) <sub>5</sub> -9.1	Gm09:1040692..1040629	-63	Alta	Glyma09g01620 (Protein involved in Snf1 protein kinase complex assembly / 5'- AMP-activated protein kinase, beta subunit, complex- interacting region)	-
(AAG) <sub>5</sub> -9.2	Gm09:7277766..7277657	-109	Alta	-	-
(AAG) <sub>5</sub> -9.3	Gm09:8295605..8295504	-101	Moderada	-	-
(AAG) <sub>5</sub> -9.4	Gm09:28868155..28868210	55	Moderada	-	-
(AAG) <sub>5</sub> -9.5	Gm09:31630051..31629990	-61	Moderada	-	-
(AAG) <sub>5</sub> -9.6	Gm09:32943593..32943669	76	Alta	-	-
(AAG) <sub>5</sub> -9.7	Gm09:43929049..43928892	157	Baixa	-	-
(AAG) <sub>5</sub> -10.1	Gm10:473625..473731	106	Alta	-	-
(AAG) <sub>5</sub> -10.2	Gm10:12792444..12792381	-63	Moderada	-	-
(AAG) <sub>5</sub> -10.3	Gm10:12957886..12957941	55	Moderada	-	-
(AAG) <sub>5</sub> -10.4	Gm10:27675105..27675168	63	Moderada	-	-
(AAG) <sub>5</sub> -10.5	Gm10:32171343..32171432	89	Moderada	-	-
(AAG) <sub>5</sub> -10.6	Gm10:38324155..38324209	54	Alta	-	-
(AAG) <sub>5</sub> -11.1	Gm11:15548297..15548457	160	Alta	Glyma11g18940	-

				(Heterogeneous nuclear ribonucleoprotein R (RRM superfamily / RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain))	
				Glyma1g19660 (Vacuolar assembly/sorting protein VPS41 / Region in Clathrin and VPS)	
(AAG) <sub>5</sub> -11.2	Gm11:16426534..16426448	-86	Moderada		-
(AAG) <sub>5</sub> -11.3	Gm11:19169185..19169243	58	Moderada	-	-
(AAG) <sub>5</sub> -11.4	Gm11:23532975..23533081	106	Ausência de genes	-	-
(AAG) <sub>5</sub> -11.5	Gm11:24189963..24189896	-67	Moderada	-	-
(AAG) <sub>5</sub> -11.6	Gm11:25091266..25091364	98	Moderada	-	-
(AAG) <sub>5</sub> -11.7	Gm11:29465299..29465193	-106	Baixa	-	-
(AAG) <sub>5</sub> -11.8	Gm11:33526774..33526707	-67	Moderada	-	-
(AAG) <sub>5</sub> -12.1	Gm12:6586468.. 6586560	92	Alta	Glyma12g08840 (Vacuolar Protein Sorting VPS41)	-
(AAG) <sub>5</sub> -12.2	Gm12:8506946.. 8506750	-196	Alta	-	-
(AAG) <sub>5</sub> -12.3	Gm12:16964584..16964505	-79	Moderada	-	-
(AAG) <sub>5</sub> -12.4	Gm12:20828321..20828427	106	Ausência de genes	-	-
(AAG) <sub>5</sub> -12.5	Gm12:29007194..29007251	57	Ausência de genes	-	-
(AAG) <sub>5</sub> -12.6	Gm12:30707234..30707171	-63	Moderada	-	-
(AAG) <sub>5</sub> -12.7	Gm12:36624950..36625076	126	Alta	Glyma12g33290	-
(AAG) <sub>5</sub> -13.1	Gm13:3014687..3014587	-100	Moderada	-	-
(AAG) <sub>5</sub> -13.2	Gm13:3434191..3434251	60	Moderada	-	-
(AAG) <sub>5</sub> -13.3	Gm13:29264742..29264805	63	Alta	Glyma13g26040 (Annexin)	-
(AAG) <sub>5</sub> -13.4	Gm13: 31155403.. 31155500	97	Alta	Glyma13g28060 (F-box domain)	-
(AAG) <sub>5</sub> -13.5	Gm13:33503354..33503808	454	Alta	-	Glyma13g31030 Glyma13g31040
(AAG) <sub>5</sub> -13.6	Gm13:38302395..38302266	-129	Alta	-	-
(AAG) <sub>5</sub> -14.1	Gm14:334373..334481	108	Moderada	Glyma14g00680	-
(AAG) <sub>5</sub> -14.2	Gm14: 7029742.. 7029818	76	Alta	Glyma14g09040 (LIPIN)	-
(AAG) <sub>5</sub> -14.3	Gm14:15419400..15419506	106	Ausência de genes	-	-
(AAG) <sub>5</sub> -14.4	Gm14:26367001..26366937	-64	Moderada	-	-
(AAG) <sub>5</sub> -14.5	Gm14:32482967..32483091	124	Moderada	-	-
(AAG) <sub>5</sub> -14.6	Gm14:41128312..41128365	53	Moderada	-	-
(AAG) <sub>5</sub> -14.7	Gm14:46504793..46504856	63	Alta	Glyma14g37230	-
(AAG) <sub>5</sub> -15.1	Gm15:13001161..13000995	-166	Alta	-	Glyma15g16700
(AAG) <sub>5</sub> -15.2	Gm15:13205527..13205610	83	Moderada	Glyma15g16970	-
(AAG) <sub>5</sub> -15.3	Gm15:16220882..16220816	-66	Moderada	-	-
(AAG) <sub>5</sub> -15.4	Gm15:46770377..46770313	-64	Moderada	-	-

(AAG) <sub>5</sub> -16.1	Gm16:4011030..4011080	50	Alta	Glyma16g04730 (Ricin-type beta-trefoil lectin domain)	-
(AAG) <sub>5</sub> -16.2	Gm16:8728880..8728962	82	Moderada	-	-
(AAG) <sub>5</sub> -16.3	Gm16:10047135..10047038	-97	Moderada	-	-
(AAG) <sub>5</sub> -16.4	Gm16:11383936..11384042	106	Moderada	-	-
(AAG) <sub>5</sub> -16.5	Gm16:13274990..13275114	124	Moderada	-	-
(AAG) <sub>5</sub> -16.6	Gm16:15518880..15519007	127	Moderada	-	-
(AAG) <sub>5</sub> -16.7	Gm16:26249796..26249840	44	Moderada	-	-
(AAG) <sub>5</sub> -16.8	Gm16:31491284..31491226	-58	Alta	Glyma16g27470 (Peroxisomal membrane protein 2, PXMP2/ Peroxisomal membrane protein MPV17 and related proteins)	-
(AAG) <sub>5</sub> -17.1	Gm17:21260883..21260810	-73	Moderada	Glyma17g21790	-
(AAG) <sub>5</sub> -17.2	Gm17:22926555..22926682	127	Moderada	-	-
(AAG) <sub>5</sub> -17.3	Gm17:27714249..27714338	89	Moderada	-	-
(AAG) <sub>5</sub> -17.4	Gm17:34328010..34327732	-278	Alta	-	-
(AAG) <sub>5</sub> -17.5	Gm17:34461703..34461758	55	Alta	-	-
(AAG) <sub>5</sub> -17.6	Gm17:35665245..35665173	-72	Moderada	-	-
(AAG) <sub>5</sub> -17.7	Gm17:37330517..37330626	109	Moderada	-	-
(AAG) <sub>5</sub> -18.1	Gm18:4626393..4626509	116	Alta	Glyma18g06030 (Metalloprotease / N-arginine dibasic convertase NRD1 and related Zn <sup>2+</sup> -dependent endopeptidases, insulinase superfamily / Peptidase M16 inactive domain)	-
(AAG) <sub>5</sub> -18.2	Gm18:13446144..13446213	69	Moderada	-	-
(AAG) <sub>5</sub> -18.3	Gm18:24867984..24867933	-51	Moderada	-	-
(AAG) <sub>5</sub> -18.4	Gm18:42208419..42208498	79	Moderada	Glyma18g35940 (GPI transamidase complex, GPI17/PIG-S component, involved in glycosylphosphatidylinositol anchor biosynthesis)	-
(AAG) <sub>5</sub> -18.5	Gm18:46650323..46650372	49	Moderada	-	-
(AAG) <sub>5</sub> -18.6	Gm18:48777914..48777848	-66	Alta	-	-
(AAG) <sub>5</sub> -18.7	Gm18:60517761..60517701	-60	Alta	Glyma18g51700 (Leucine-rich transmembrane proteins) KOG / Apoptotic ATPase) / NB-ARC domain)	-
(AAG) <sub>5</sub> -19.1	Gm19:7143444..7143373	-71	Alta	-	-
(AAG) <sub>5</sub> -19.2	Gm19:20623667..20623593	-74	Moderada	-	-

(AAG) <sub>5</sub> -19.3	Gm19:20895501..20895568	67	Moderada	-	-
(AAG) <sub>5</sub> -19.4	Gm19:28768042..28768097	55	Moderada	-	-
(AAG) <sub>5</sub> -19.5	Gm19:29093280..29093428	148	Moderada	-	-
(AAG) <sub>5</sub> -19.6	Gm19:30091631..30091576	-55	Alta	-	-
(AAG) <sub>5</sub> -19.7	Gm19:37903031..37902910	-121	Alta	Glyma19g30250 (RNA-binding protein musashi/mRNA cleavage and polyadenylation factor I complex, subunit HRP1 / RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)	-
(AAG) <sub>5</sub> -19.8	Gm19:38388372..38388473	101	Alta	-	-
(AAG) <sub>5</sub> -19.9	Gm19:46503099..46503040	-59	Alta	-	-
(AAG) <sub>5</sub> -20.1	Gm20:12656109..12656028	-81	Moderada	Glyma20g08920 (UBA/TS-N domain)	-
(AAG) <sub>5</sub> -20.2	Gm20:17045652..17045708	56	Moderada	-	-
(AAG) <sub>5</sub> -20.3	Gm20:20585241..20585135	-106	Moderada	-	-
(AAG) <sub>5</sub> -20.4	Gm20:23088914..23088999	85	Moderada	Glyma20g16560 gene (Low confidence)	-
(AAG) <sub>5</sub> -20.5	Gm20:25156628..25156486	-142	Moderada	-	Glyma20g17 970
(AAG) <sub>5</sub> -20.6	Gm20:28326901..28326970	69	Moderada	-	-

**Tabela 3.** Repetições de [(ACC)<sub>5</sub>]<sub>10</sub> identificadas contendo as informações individuais sobre a posição exata no genoma, tamanho da repetição em pb, bem como os genes e elementos transponíveis associados.

Código	Região (ACC) <sub>5</sub>	Tamanho (bp)	Densidade gênica	Gene models	Gene models (transposon-like)
(ACC) <sub>5</sub> -1.1	Gm01:560347..560263	-84	Alta	Glyma01g00890 Serine/threonine protein kinase / Plant-type	-
(ACC) <sub>5</sub> -1.2	Gm01:2585786..2585935	149	Moderada	Glyma01g03070	-
(ACC) <sub>5</sub> -1.3	Gm01:29703168..29703109	-59	Moderada	Glyma01g23180 Serine/threonine protein kinase	-
(ACC) <sub>5</sub> -1.4	Gm01:50272466..50272566	100	Alta	Glyma01g38100	-
(ACC) <sub>5</sub> -2.1	Gm02:3672690..3672578	-112	Alta	Glyma02g0451	-
(ACC) <sub>5</sub> -2.2	Gm02:44520758..44520856	98	Alta	Glyma02g39270 SelenoproteinT	-
(ACC) <sub>5</sub> -2.3	Gm02:47647417..47647305	-112	Alta	Glyma02g42600 Collagen.alpha1 (X) CHAIN	Glyma02g42610
(ACC) <sub>5</sub> -3.1	Gm03:8152020..8151931	-89	Alta	Glyma03g30790	-
(ACC) <sub>5</sub> -3.2	Gm03:38655204..38655283	79	Alta	Glyma03g30790 (HEPATOMA-DERIVED GROWTH FACTOR-RELATED / PWWP domain)	-
(ACC) <sub>5</sub> -3.3	Gm03:39233038..39232964	-74	Alta	-	-
(ACC) <sub>5</sub> -3.4	Gm03:42911398..42911455	57	Alta	Glyma03g35810 (Protein of unknown function, DUF623)	-
(ACC) <sub>5</sub> -3.5	Gm03:43918498..43918452	-46	Alta	Glyma03g37230 (F-box domain)	-
(ACC) <sub>5</sub> -4.1	Gm04:913799..913904	105	Alta	-	Glyma04g01420
(ACC) <sub>5</sub> -4.2	Gm04:2137687..2137654	-33	Alta	Glyma04g02960 (B-box zinc finger)	-
(ACC) <sub>5</sub> -4.3	Gm04:16941660..16941711	51	Moderada	-	-
(ACC) <sub>5</sub> -4.4	Gm04:30836359..30836293	-66	Moderada	-	-
(ACC) <sub>5</sub> -4.5	Gm04:39887458..39887409	-49	Alta	-	-
(ACC) <sub>5</sub> -4.6	Gm04:46334749..46334707	-42	Alta	Glyma04g40220 (Uncharacterized conserved protein / Zinc finger, C3HC4 type (RING finger))	-
(ACC) <sub>5</sub> -5.1	Gm05:4382289..4382245	-44	Moderada	-	-
(ACC) <sub>5</sub> -5.2	Gm05:16628652..16628596	-56	Moderada	-	-
(ACC) <sub>5</sub> -5.3	Gm05:24525612..24525671	59	Moderada	-	-
(ACC) <sub>5</sub> -5.4	Gm05:30370555..30370479	-76	Moderada	-	-
(ACC) <sub>5</sub> -5.5	Gm05:30605856..30605801	-55	Alta	-	-
(ACC) <sub>5</sub> -5.6	Gm05:40424247..40424205	-42	Alta	Glyma05g36590	-
(ACC) <sub>5</sub> -6.1	Gm06:5772776..5772663	-113	Alta	Glyma06g07850 (ZINC FINGER FYVE DOMAIN CONTAINING PROTEIN / Predicted Zn <sup>2+</sup> -binding protein,	-

				contains FYVE domain / Tetratricopeptide repeat)	
(ACC) <sub>5</sub> -6.2	Gm06:8534521..8534455	-66	Alta	Glyma06g11160	-
(ACC) <sub>5</sub> -6.3	Gm06:46235625..46235761	136	Alta	Glyma06g42930	Glyma06g42940
(ACC) <sub>5</sub> -6.4	Gm06:50177728..50177680	-48	Alta	-	-
(ACC) <sub>5</sub> -7.1	Gm07:14236909..14236972	63	Moderada	Glyma07g14610 (SBP domain)	-
(ACC) <sub>5</sub> -7.2	Gm07:16296232..16296336	104	Moderada	-	-
(ACC) <sub>5</sub> -7.3	Gm07:16306420..16306495	75	Alta	Glyma07g16590	-
(ACC) <sub>5</sub> -7.4	Gm07:30099365..30099310	-55	Baixa	-	-
(ACC) <sub>5</sub> -7.5	Gm07:36225576..36225620	44	Moderada	Glyma07g31250	-
(ACC) <sub>5</sub> -8.1	Gm08:10172364..10172364	-119	Alta	Glyma08g13980 (Plant specific eukaryotic initiation factor 4B)	-
(ACC) <sub>5</sub> -8.2	Gm08:15335039..15335121	82	Alta	-	-
(ACC) <sub>5</sub> -8.3	Gm08:32654074..32654008	-66	Moderada	-	-
(ACC) <sub>5</sub> -8.4	Gm08:39013259..39013152	-107	Alta	Glyma08g39480 (Serine/threonine protein kinase / Protein tyrosine kinase)	-
(ACC) <sub>5</sub> -9.1	Gm09:8760764..8760764	136	Moderada	Glyma09g09190	-
(ACC) <sub>5</sub> -10.1	Gm10:24614555..24614491	-64	Moderada	Glyma10g19710	-
(ACC) <sub>5</sub> -10.2	Gm10:33055035..33055035	101	Moderada	Glyma10g25350	-
(ACC) <sub>5</sub> -10.3	Gm10:36634211..36634104	-107	Alta	Glyma10g27810 (SecA DEAD-like domain)	-
(ACC) <sub>5</sub> -10.4	Gm10:38908222..38908282	60	Alta	-	-
(ACC) <sub>5</sub> -10.5	Gm10:42374425..42374350	-75	Alta	Glyma10g34200	-
(ACC) <sub>5</sub> -11.1	Gm11:7667087..7666914	-173	Alta	Glyma11g10790 (Apoptosis-promoting RNA-binding protein TIA-1/TIAR (RRM superfamily) / RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain))	-
(ACC) <sub>5</sub> -11.2	Gm11:8908168..8908244	76	Alta	-	Glyma11g12470
(ACC) <sub>5</sub> -11.3	Gm11:11932695..11932751	56	Moderada	Glyma11g16210 (N-TERMINAL ACETYLTRANSFERAS E) / Subunit of the major N alpha-acetyltransferase / Acetyltransferase (GNAT) family)	-
(ACC) <sub>5</sub> -12.1	Gm12:25365900..25365838	-79	Moderada	Glyma12g23110 (WD40 repeat-containing protein / WD domain 44, G-beta repeat)	-
(ACC) <sub>5</sub> -12.2	Gm12:38840126..38840168	42	Alta	Glyma12g35720 (TCP family transcription factor)	-
(ACC) <sub>5</sub> -13.1	Gm13:2107944..2107848	-96	Alta	-	-
(ACC) <sub>5</sub> -13.2	Gm13:7393562..7393652	90	Alta	-	-
(ACC) <sub>5</sub> -13.3	Gm13:33193643..33193585	-58	Alta	Glyma13g30630	-
(ACC) <sub>5</sub> -13.4	Gm13:34994437..34994501	64	Alta	-	-

(ACC) <sub>5</sub> -13.5	Gm13:36907670..36907620	-57	Moderada	Glyma13g35500 (WD40 repeat-containing protein / WD domain 44, G-beta repeat)	-
(ACC) <sub>5</sub> -14.1	Gm14:5710389..5710466	77	Alta	Glyma14g07550	-
(ACC) <sub>5</sub> -14.2	Gm14:27250565..27250611	46	Alta	Glyma14g22930 (PROTEIN./Zinc finger, C3HC4 type (RING finger))	-
(ACC) <sub>5</sub> -14.3	Gm14:27707675..27707762	87	Moderada	-	-
(ACC) <sub>5</sub> -14.4	Gm14:46674723..46674839	116	Alta	Glyma14g37390 (Selenoprotein T)	-
(ACC) <sub>5</sub> -15.1	Gm15:10414999..10415010	-147	Alta	Glyma15g13800	-
(ACC) <sub>5</sub> -15.2	Gm15:11811269..11811174	-129	Alta	Glyma15g15390	-
(ACC) <sub>5</sub> -16.1	Gm16:1134893..1134937	44	Alta	-	-
(ACC) <sub>5</sub> -16.2	Gm16:3070136..3070188	52	Alta	Glyma16g03650 (Serine/threonine protein kinase) / Protein tyrosine kinase)	-
(ACC) <sub>5</sub> -16.3	Gm16:3859197..3859265	68	Alta	Glyma16g04570 (Nucleolar protein family a member 1 (SNORNP PROTEIN GAR1) / H/ACA small nucleolar RNP component GAR1 / Gar1 protein RNA binding region)	-
(ACC) <sub>5</sub> -16.4	Gm16:6960258..6960313	55	Alta	-	-
(ACC) <sub>5</sub> -16.5	Gm16:35357840..35357810	-30	Alta	-	-
(ACC) <sub>5</sub> -17.1	Gm17:6102936..6102974	38	Alta	Glyma16g32180 (Putative methyltransferase)	-
(ACC) <sub>5</sub> -17.2	Gm17:7559708..7559666	-42	Alta	Glyma17g10100 (F-BOX AND WD40 DOMAIN PROTEIN / Microtubule binding protein YTM1 (contains WD40 repeats) / WD domain, G-beta repeat)	-
(ACC) <sub>5</sub> -17.3	Gm17:15472694..15472762	68	Alta	-	-
(ACC) <sub>5</sub> -17.4	Gm17:37842193..37842158	-46	Alta	Glyma17g33930 (FORMIN-RELATED / Transcription factor Abd-B, contains HOX domain / Formin Homology 2 Domain)	-
(ACC) <sub>5</sub> -17.5	Gm17:38551494..38551542	48	Alta	Glyma17g34550 (Glutaredoxin and related proteins)	-
(ACC) <sub>5</sub> -18.1	Gm18:13074001..13073897	-104	Moderada	-	-
(ACC) <sub>5</sub> -18.2	Gm18:18301370..18301276	-94	Moderada	Glyma18g17220 (SIN3B-RELATED)	-
(ACC) <sub>5</sub> -18.3	Gm18:20899318..20899392	74	Moderada	-	-
(ACC) <sub>5</sub> -18.4	Gm18:47502919..47502833	-93	Moderada	-	-
(ACC) <sub>5</sub> -18.5	Gm18:54283035..54283128	93	Moderada	-	-
(ACC) <sub>5</sub> -19.1	Gm19:26750947..26750905	-43	Moderada	-	-
(ACC) <sub>5</sub> -19.2	Gm19:38606358..38606395	37	Alta	Glyma19g30910 (CIRCADIAN)	-

				PROTEIN CLOCK/ARNT/BMAL/ PAS / Helix-loop-helix DNA-binding domain)	
(ACC) <sub>5</sub> -19.3	Gm19:38950978..38951010	32	Alta	Glyma19g31180 (ANKYRIN REPEAT- CONTAINING / BTB/POZ domain)	-
(ACC) <sub>5</sub> -19.4	Gm19:41233914..41233999	85	Moderada	Glyma19g33640 (HEPATOMA- DERIVED GROWTH FACTOR-RELATED / PWWP domain)	-
(ACC) <sub>5</sub> -20.1	Gm20:2359987..2360017	30	Alta	Glyma20g02780	-
(ACC) <sub>5</sub> -20.2	Gm20:2782423..2782489	66	Alta	Glyma20g03040 Asparagine synthase) Glyma20g03050 (JUMONJI DOMAIN CONTAINING PROTEIN)	-
(ACC) <sub>5</sub> -20.3	Gm20:3845919..3845848	-71	Moderada	-	-
(ACC) <sub>5</sub> -20.4	Gm20:25554008..25554096	88	Moderada	-	-
(ACC) <sub>5</sub> -20.5	Gm20:35551049..35550998	-51	Alta	Glyma20g25950	-

**Tabela 4.** Repetições de [(CTC)<sub>5</sub>]<sub>10</sub> identificadas contendo as informações individuais sobre a posição exata no genoma, tamanho da repetição em pb, bem como os genes e elementos transponíveis associados.

Código	Região (CTC) <sub>5</sub>	Tamanho (bp)	Densidade gênica	Gene models	Gene models (transposon-like)
(CTC) <sub>5</sub> -1.1	Gm01:6640016..6639955	-61	Alta	-	-
(CTC) <sub>5</sub> -1.2	Gm01:18163685..18163743	58	Moderada	-	-
(CTC) <sub>5</sub> -1.3	Gm01:34386945..34386883	-62	Moderada	-	-
(CTC) <sub>5</sub> -1.4	Gm01:39570056..39569998	-58	Moderada	-	-
(CTC) <sub>5</sub> -1.5	Gm01:39997119..39997034	-85	Alta	-	-
(CTC) <sub>5</sub> -1.6	Gm01:40315593..40315523	-70	Moderada	-	-
(CTC) <sub>5</sub> -1.7	Gm01:54603443..54603522	79	Alta	Glyma01g43640 (Thump Domain Containing Protein 1- Related)	-
(CTC) <sub>5</sub> -2.1	Gm02:8867372..8867433	61	Moderada	-	-
(CTC) <sub>5</sub> -2.2	Gm02:27468198..27468286	88	Alta	-	-
(CTC) <sub>5</sub> -2.3	Gm02:30325216..30325271	55	Moderada	-	-
(CTC) <sub>5</sub> -2.4	Gm02:33097271..33097198	-73	Moderada	-	-
(CTC) <sub>5</sub> -2.5	Gm02:36687723..36687638	-85	Moderada	-	-
(CTC) <sub>5</sub> -2.6	Gm02:37935393..37935451	58	Moderada	-	-
(CTC) <sub>5</sub> -2.7	Gm02:38405619..38405534	-85	Moderada	-	-
(CTC) <sub>5</sub> -3.1	Gm03:3405566..3405505	-61	Alta	-	-
(CTC) <sub>5</sub> -3.2	Gm03:9541893..9541840	-53	Alta	-	-
(CTC) <sub>5</sub> -3.3	Gm03:9805024..9804948	-76	Baixa	-	-
(CTC) <sub>5</sub> -3.4	Gm03:10244743..10244801	58	Moderada	-	-
(CTC) <sub>5</sub> -3.5	Gm03:31175497..31175424	-73	Moderada	-	-
(CTC) <sub>5</sub> -3.6	Gm03:32305755..32305688	-67	Moderada	-	-
(CTC) <sub>5</sub> -4.1	Gm04:10465584..10465657	73	Moderada	-	-
(CTC) <sub>5</sub> -4.2	Gm04:19663760..19663842	82	Moderada	-	-
(CTC) <sub>5</sub> -4.3	Gm04:24141145..24141087	-58	Ausência de genes	-	-
(CTC) <sub>5</sub> -4.4	Gm04:33121499..33121569	70	Moderada	-	-
(CTC) <sub>5</sub> -4.5	Gm04:35905571..35905504	-67	Moderada	-	-
(CTC) <sub>5</sub> -4.6	Gm04:36055521..36055618	97	Moderada	-	-
(CTC) <sub>5</sub> -4.7	Gm04:37527008..37527067	59	Moderada	-	-
(CTC) <sub>5</sub> -4.8	Gm04:40561382..40561321	-61	Moderada	-	-
(CTC) <sub>5</sub> -5.1	Gm05:14712040..14711955	-85	Moderada	-	-
(CTC) <sub>5</sub> -5.2	Gm05:19227153..19227214	61	Moderada	-	-
(CTC) <sub>5</sub> -5.3	Gm05:22366873..22366943	70	Moderada	-	-
(CTC) <sub>5</sub> -5.4	Gm05:26487430..26487330	-100	Moderada	-	-
(CTC) <sub>5</sub> -5.5	Gm05:26926530..26926615	85	Moderada	-	-
(CTC) <sub>5</sub> -5.6	Gm05:28043907..28043819	-88	Moderada	-	-
(CTC) <sub>5</sub> -6.1	Gm06:2062542..2062352	-190	Alta	Glyma06g02960	-
(CTC) <sub>5</sub> -6.2	Gm06:20649064..20649146	82	Moderada	-	-
(CTC) <sub>5</sub> -6.3	Gm06:27802764..27802825	61	Ausência de genes	-	-
(CTC) <sub>5</sub> -6.4	Gm06:38725587..38725535	-52	Moderada	-	-
(CTC) <sub>5</sub> -6.5	Gm06:42182527..42182463	-64	Alta	-	-
(CTC) <sub>5</sub> -6.6	Gm06:42714151..42714096	-55	Moderada	-	-
(CTC) <sub>5</sub> -6.7	Gm06:43217790..43217699	-91	Moderada	-	-
(CTC) <sub>5</sub> -7.1	Gm07:12015936..12015994	58	Moderada	-	-
(CTC) <sub>5</sub> -7.2	Gm07:13647697..13647642	-55	Alta	-	-
(CTC) <sub>5</sub> -7.3	Gm07:16296234..16296325	91	Moderada	-	-
(CTC) <sub>5</sub> -7.4	Gm07:21230208..21230155	-53	Alta	-	-
(CTC) <sub>5</sub> -7.5	Gm07:23461475..23461417	-58	Moderada	-	-
(CTC) <sub>5</sub> -7.6	Gm07:25973403..25973403	64	Moderada	-	-
(CTC) <sub>5</sub> -7.7	Gm07:30175473..30175411	62	Moderada	-	-

(CTC) <sub>5</sub> -7.8	Gm07:32081469..32081536	67	Moderada	-	-
(CTC) <sub>5</sub> -8.1	Gm08:21572654..21572614	-40	Moderada	-	-
(CTC) <sub>5</sub> -8.2	Gm08:23979116..23979024	-92	Moderada	-	-
(CTC) <sub>5</sub> -8.3	Gm08:26263022..26263062	40	Moderada	-	-
(CTC) <sub>5</sub> -8.4	Gm08:27515170..27515109	-61	Ausência de genes	-	-
(CTC) <sub>5</sub> -8.5	Gm08:31429985..31430040	55	Ausência de genes	-	-
(CTC) <sub>5</sub> -8.6	Gm08:36059230..36059193	-37	Moderada	-	-
(CTC) <sub>5</sub> -9.1	Gm09:8390509..8390445	-64	Alta	-	-
(CTC) <sub>5</sub> -9.2	Gm09:10900860..10900796	-64	Moderada	-	-
(CTC) <sub>5</sub> -9.3	Gm09:17621788..17621721	-67	Moderada	-	-
(CTC) <sub>5</sub> -9.4	Gm09:19129292..19129356	64	Moderada	-	-
(CTC) <sub>5</sub> -9.5	Gm09:19477972..19477878	-94	Alta	-	-
(CTC) <sub>5</sub> -9.6	Gm09:24263864..24263809	-55	Moderada	-	-
(CTC) <sub>5</sub> -9.7	Gm09:25428267..25428212	-55	Moderada	-	-
(CTC) <sub>5</sub> -9.8	Gm09:28816010..28815948	-62	Moderada	-	-
(CTC) <sub>5</sub> -10.1	Gm10:13897367..13897300	-67	Moderada	-	-
(CTC) <sub>5</sub> -10.2	Gm10:13967023..13966947	-76	Moderada	-	-
(CTC) <sub>5</sub> -10.3	Gm10:17124986..17124928	-58	Alta	-	-
(CTC) <sub>5</sub> -10.4	Gm10:25613374..25613304	-70	Moderada	-	-
(CTC) <sub>5</sub> -10.5	Gm10:34155376..34155437	61	Moderada	-	-
(CTC) <sub>5</sub> -10.6	Gm10:34309386..34309447	61	Alta	Glyma10g26200 (TNPI/EN/SPM transposase)	-
(CTC) <sub>5</sub> -10.7	Gm10:37947372..37947463	91	Alta	Glyma10g29040 (SPLICING FACTOR, ARGININE/SERINE-RICH 2, RNAP C-TERM INTERACTING PROTEIN / Topoisomerase I-binding arginine-serine-rich protein / P-11 zinc finger)	-
(CTC) <sub>5</sub> -10.8	Gm10:40859202..40859130	-72	Alta	Glyma10g32370 (Prenylated rab acceptor 1 / PRA1 family protein)	-
(CTC) <sub>5</sub> -11.1	Gm11:8340848..8340955	107	Alta	Glyma11g11690 (FORMIN-RELATED / Frigida-like protein)	-
(CTC) <sub>5</sub> -11.2	Gm11:13227179..13227222	43	Ausência de genes	-	-
(CTC) <sub>5</sub> -11.3	Gm11:15548457..15548324	-133	Alta	Glyma11g18940 (RNA-BINDING PROTEIN / Heterogeneous nuclear ribonucleoprotein R (RRM superfamily) / RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain))	-
(CTC) <sub>5</sub> -11.4	Gm11:22689017..22688956	-61	Ausência de genes	-	-
(CTC) <sub>5</sub> -11.5	Gm11:28471109..28471036	-73	Moderada	-	-
(CTC) <sub>5</sub> -11.6	Gm11:31555254..31555191	-63	Alta	-	Glyma11g30560
(CTC) <sub>5</sub> -11.7	Gm11:34150059..34150109	50	Alta	-	-
(CTC) <sub>5</sub> -12.1	Gm12:7322562..7322707	145	Alta	Glyma12g09530	-

					(RNA-Binding Protein / Heterogeneous nuclear ribonucleoprotein R (RRM superfamily / RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain))	
(CTC) <sub>5</sub> -12.2	Gm12:8349105..8349193	88	Moderada	-	-	
(CTC) <sub>5</sub> -12.3	Gm12:12260915..12261003	88	Alta	-	-	
(CTC) <sub>5</sub> -12.4	Gm12:12332917..12332839	-78	Moderada	-	-	
(CTC) <sub>5</sub> -12.5	Gm12:23088356..23088408	52	Moderada	-	-	
(CTC) <sub>5</sub> -12.6	Gm12:29434516..29434580	64	Moderada	-	-	
(CTC) <sub>5</sub> -12.7	Gm12:29630745..29630683	-62	Moderada	-	-	
(CTC) <sub>5</sub> -12.8	Gm12:37501314..37501410	96	Alta	Glyma12g34350 (Formin Homology 2 Domain)	-	
(CTC) <sub>5</sub> -13.1	Gm13:10774587..10774669	82	Moderada	-	-	
(CTC) <sub>5</sub> -13.2	Gm13:18511209..18511139	-70	Ausência de genes	-	-	
(CTC) <sub>5</sub> -13.3	Gm13:28778466..28778873	407	Alta	Glyma13g25550 (Plant invertase/pectin methylesterase inhibitor)	-	
(CTC) <sub>5</sub> -13.4	Gm13:32959159..32959088	-71	Alta	Glyma13g30280 (Transcription factor GT-2 and related proteins, contains trihelix DNA-binding/SANT domain)	-	
(CTC) <sub>5</sub> -13.5	Gm13:33503832..33503644	-188	Alta	-	Glyma13g31030 Glyma13g31040	
(CTC) <sub>5</sub> -14.1	Gm14:9778741..9778662	-79	Moderada	-	-	
(CTC) <sub>5</sub> -14.2	Gm14:11894598..11894665	67	Moderada	-	-	
(CTC) <sub>5</sub> -14.3	Gm14:25650464..25650403	-61	Moderada	-	-	
(CTC) <sub>5</sub> -14.4	Gm14:31969419..31969361	-58	Moderada	-	-	
(CTC) <sub>5</sub> -14.5	Gm14:32702214..32702150	-64	Moderada	-	-	
(CTC) <sub>5</sub> -14.6	Gm14:38842609..38842540	-69	Moderada	-	-	
(CTC) <sub>5</sub> -14.7	Gm14:40914428..40914498	70	Ausência de genes	-	-	
(CTC) <sub>5</sub> -14.8	Gm14:49618357..49618240	-117	Moderada	Glyma14g40640	-	
(CTC) <sub>5</sub> -15.1	Gm15:5465158..5465226	68	Alta	Glyma15g07760 (Splicing factor 3b, subunit 2 / PSP)	-	
(CTC) <sub>5</sub> -15.2	Gm15:5889012..5889127	115	Alta	Glyma15g08330	Glyma15g08340	
(CTC) <sub>5</sub> -15.3	Gm15:18740876..18740809	-67	Alta	-	-	
(CTC) <sub>5</sub> -15.4	Gm15:36591821..36591888	67	Moderada	-	-	
(CTC) <sub>5</sub> -15.5	Gm15:37247312..37247376	64	Moderada	-	-	
(CTC) <sub>5</sub> -15.6	Gm15:40572795..40572853	58	Moderada	-	-	
(CTC) <sub>5</sub> -15.7	Gm15:41010032..41009959	-73	Moderada	-	-	
(CTC) <sub>5</sub> -15.8	Gm15:45870934..45870855	-79	Baixa	-	-	
(CTC) <sub>5</sub> -16.1	Gm16:168155..168069	-86	Alta	Glyma16g00470	-	
(CTC) <sub>5</sub> -16.2	Gm16:8155187..8155230	43	Alta	-	-	
(CTC) <sub>5</sub> -16.3	Gm16:9732703..9732750	47	Alta	-	-	
(CTC) <sub>5</sub> -16.4	Gm16:12058275..12058336	61	Baixa	-	-	
(CTC) <sub>5</sub> -16.5	Gm16:12887773..12887843	70	Moderada	-	-	
(CTC) <sub>5</sub> -16.6	Gm16:13787256..13787214	-42	Ausência de genes	-	-	

(CTC) <sub>5</sub> -16.7	Gm16:22104846..22104767	-79	Ausência de genes	-	-
(CTC) <sub>5</sub> -17.1	Gm17:17966748..17966809	61	Baixa	-	-
(CTC) <sub>5</sub> -17.2	Gm17:18281331..18281276	-55	Alta	-	-
(CTC) <sub>5</sub> -17.3	Gm17:25784855..25784791	-64	Moderada	-	-
(CTC) <sub>5</sub> -17.4	Gm17:27714339..27714339	-89	Moderada	-	-
(CTC) <sub>5</sub> -17.5	Gm17:28708482..28708546	64	Moderada	-	-
(CTC) <sub>5</sub> -17.6	Gm17:30782492..30782559	67	Moderada	-	-
(CTC) <sub>5</sub> -17.7	Gm17:40574803..40574742	-61	Alta	-	-
(CTC) <sub>5</sub> -18.1	Gm18:11083418..11083485	67	Moderada	-	-
(CTC) <sub>5</sub> -18.2	Gm18:12720889..12720950	61	Moderada	-	-
(CTC) <sub>5</sub> -18.3	Gm18:14311513..14311586	73	Moderada	-	-
(CTC) <sub>5</sub> -18.4	Gm18:21331013..21330931	-82	Moderada	-	-
(CTC) <sub>5</sub> -18.5	Gm18:27111402..27111482	80	Moderada	-	-
(CTC) <sub>5</sub> -18.6	Gm18:40349468..40349547	79	Moderada	-	-
(CTC) <sub>5</sub> -18.7	Gm18:42895817..42895717	-100	Alta	-	-
(CTC) <sub>5</sub> -18.8	Gm18:43663613..43663546	-67	Moderada	-	-
(CTC) <sub>5</sub> -19.1	Gm19:8322051..8322139	88	Alta	-	-
(CTC) <sub>5</sub> -19.2	Gm19:8701809..8701757	-52	Alta	-	-
(CTC) <sub>5</sub> -19.3	Gm19:11451791..11451730	-61	Alta	-	-
(CTC) <sub>5</sub> -19.4	Gm19:16551674..16551729	55	Moderada	-	-
(CTC) <sub>5</sub> -19.5	Gm19:26668905..26668832	-73	Moderada	-	-
(CTC) <sub>5</sub> -19.6	Gm19:29516705..29516775	70	Moderada	-	-
(CTC) <sub>5</sub> -19.7	Gm19:29865648..29865566	-82	Moderada	-	-
(CTC) <sub>5</sub> -19.8	Gm19:32694382..32694327	-55	Moderada	-	-
(CTC) <sub>5</sub> -20.1	Gm20:5252009..5252076	67	Moderada	-	-
(CTC) <sub>5</sub> -20.2	Gm20:6136712..6136660	-52	Moderada	-	-
(CTC) <sub>5</sub> -20.3	Gm20:11379140..11379204	64	Moderada	-	-
(CTC) <sub>5</sub> -20.4	Gm20:12706101..12706040	-61	Moderada	-	-
(CTC) <sub>5</sub> -20.5	Gm20:16453051..16452993	-58	Moderada	-	-
(CTC) <sub>5</sub> -20.6	Gm20:24624812..24624747	-65	Moderada	-	-
(CTC) <sub>5</sub> -20.7	Gm20:25110589..25110513	-76	Moderada	-	-
(CTC) <sub>5</sub> -20.1	Gm20:5252009..5252076	67	Moderada	-	-
(CTC) <sub>5</sub> -20.2	Gm20:6136712..6136660	-52	Moderada	-	-
(CTC) <sub>5</sub> -20.3	Gm20:11379140..11379204	64	Moderada	-	-
(CTC) <sub>5</sub> -20.4	Gm20:12706101..12706040	-61	Moderada	-	-
(CTC) <sub>5</sub> -20.5	Gm20:16453051..16452993	-58	Moderada	-	-
(CTC) <sub>5</sub> -20.6	Gm20:24624812..24624747	-65	Moderada	-	-
(CTC) <sub>5</sub> -20.7	Gm20:25110589..25110513	-76	Moderada	-	-

**Tabela 5.** Repetições de [(TGA)<sub>6</sub>]<sub>10</sub> identificadas contendo as informações individuais sobre a posição exata no genoma, tamanho da repetição em pb, bem como os genes e elementos transponíveis associados.

Código	Região (TGA) <sub>6</sub>	Tamanho (bp)	Densidade gênica	Gene models	Gene models (transposon-like)
(TGA) <sub>6</sub> -1.1	Gm01:3717766..3717863	97	Alta	Glyma01g04230	-
(TGA) <sub>6</sub> -1.2	Gm01:4580648..4580805	157	Moderada	-	-
(TGA) <sub>6</sub> -1.3	Gm01:6607331..6607254	-77	Alta	-	-
(TGA) <sub>6</sub> -1.4	Gm01:13753446..13753403	-43	Moderada	-	-
(TGA) <sub>6</sub> -1.5	Gm01:20364973..20365079	106	Moderada	-	-
(TGA) <sub>6</sub> -1.6	Gm01:27906665..27906732	67	Moderada	-	-
(TGA) <sub>6</sub> -2.1	Gm02:5329430..5329497	67	Alta	Glyma02g06720 (Lim domain / Adaptor protein Enigma and related PDZ-LIM proteins / Ubiquitin interaction motif)	-
(TGA) <sub>6</sub> -2.2	Gm02:17805341..17805498	157	Moderada	-	-
(TGA) <sub>6</sub> -2.3	Gm02:32615604..32615743	139	Moderada	-	-
(TGA) <sub>6</sub> -2.4	Gm02:49418577..49418577	-102	Alta	Glyma02g45000 ATP-dependent – helicase smarca (SWI/SNF-related matrix-associated actin-dependent regulator of cromatina / Chromodomain-helicase DNA-binding protein/ 'chromo' (CHRromatin Organisation Modifier) domain)	-
(TGA) <sub>6</sub> -2.5	Gm02:51435040..51434973	-67	Alta	Glyma02g48040 (UNCHARACTERIZED / Protein of unknown function (DUF630))	-
(TGA) <sub>6</sub> -3.1	Gm03:7772144..7772098	-46	Moderada	-	-
(TGA) <sub>6</sub> -3.2	Gm03:21010807..21010752	-55	Moderada	-	-
(TGA) <sub>6</sub> -3.3	Gm03:29603466..29603638	172	Moderada	-	-
(TGA) <sub>6</sub> -3.4	Gm03:35305608..35305532	-76	Alta	Glyma03g27570 (NASCENT POLYPEPTIDE ASSOCIATED COMPLEX ALPHA SUBUNIT-RELATED / Transcription factor containing NAC and TS-N domains/ NAC domain)	-
(TGA) <sub>6</sub> -3.5	Gm03:38338037..38338125	88	Alta	-	-
(TGA) <sub>6</sub> -3.6	Gm03:46673187..46673103	-84	Alta	-	-
(TGA) <sub>6</sub> -4.1	Gm04:8044191..8044126	-65	Alta	Glyma04g09780 Rho-GTPase activating protein / RhoGAP domain)	-
(TGA) <sub>6</sub> -4.2	Gm04:12448533..12448436	-97	Moderada	-	-
(TGA) <sub>6</sub> -4.3	Gm04:40344633..40344554	-79	Moderada	-	-
(TGA) <sub>6</sub> -4.4	Gm04:43163324..43163410	86	Alta	Glyma04g36670	-

				(U2 SNRNP AUXILIARY FACTOR, SMALL SUBUNIT / CCCH-type Zn-finger protein / Zinc finger C- x8-C-x5-C-x3-H type (and similar))	
(TGA) <sub>6-4.5</sub>	Gm04:43855200..43855276	76	Alta	Glyma04g37460 Glyma05g02920	-
				(U2 SNRNP AUXILIARY FACTOR, SMALL SUBUNIT / CCCH-type Zn-finger protein / Zinc finger C- x8-C-x5-C-x3-H type (and similar))	
(TGA) <sub>6-5.1</sub>	Gm05:2243455..2243545	90	Alta		-
(TGA) <sub>6-5.2</sub>	Gm05:25886940..25886813	-127	Alta	-	-
(TGA) <sub>6-5.3</sub>	Gm05:28590509..28590610	101	Moderada	-	-
(TGA) <sub>6-5.4</sub>	Gm05:29926216..29926137	-79	Moderada	-	-
(TGA) <sub>6-5.5</sub>	Gm05:36724626..36724505	-121	Alta	Glyma05g31610	-
(TGA) <sub>6-6.1</sub>	Gm06:5342857..5342785	-72	Alta	Glyma06g07330 Glyma06g09860	-
(TGA) <sub>6-6.2</sub>	Gm06:7392889..7392824	-65	Alta	(Rho-GTPase activating protein / RhoGAP domain)	-
(TGA) <sub>6-6.3</sub>	Gm06:46942079..46941976	-103	Alta	-	-
(TGA) <sub>6-6.4</sub>	Gm06:49237155..49237250	95	Alta	-	-
(TGA) <sub>6-6.5</sub>	Gm06:50129615..50129732	117	Alta	-	-
				Glyma07g06810 (BOLA-LIKE PROTEIN-RELATED / Stress-induced protein UVI31+ / Fe-S metabolism associated domain)	
(TGA) <sub>6-7.1</sub>	Gm07:5484535..5484475	-60	Alta		-
(TGA) <sub>6-7.2</sub>	Gm07:8243903..8244048	145	Alta	-	-
(TGA) <sub>6-7.3</sub>	Gm07:9320869..9320781	-88	Alta	-	-
(TGA) <sub>6-7.4</sub>	Gm07:26197875..26197933	58	Moderada	-	-
(TGA) <sub>6-7.5</sub>	Gm07:37719526..37719431	-95	Alta	-	-
(TGA) <sub>6-7.6</sub>	Gm07:37898390..37898321	-69	Alta	Glyma07g33010 (SOMETHING ABOUT SILENCING PROTEIN SAS10 / Disrupter of silencing SAS10)	-
(TGA) <sub>6-8.1</sub>	Gm08:10795764..10795646	-118	Alta	Glyma08g14830	-
(TGA) <sub>6-8.2</sub>	Gm08:12776519..12776442	-77	Moderada	Glyma08g17270 (ASPARTYL PROTEASES / Aspartyl protease)	-
(TGA) <sub>6-8.3</sub>	Gm08:19274382..19274336	-46	Moderada	-	-
(TGA) <sub>6-8.4</sub>	Gm08:22091034..22090978	-56	Alta	Glyma08g27770 (F-box domain)	-
(TGA) <sub>6-8.5</sub>	Gm08:23550932..23551053	121	Alta	-	-
(TGA) <sub>6-8.6</sub>	Gm08:37141157..37141238	81	Moderada	-	-
				Glyma09g07410 (NUCLEOSOME ASSEMBLY PROTEIN / Nucleosome assembly protein NAP-1 / Nucleosome assembly protein (NAP))	
(TGA) <sub>6-9.1</sub>	Gm09:6287405..6287317	-88	Moderada		-

(TGA) <sub>6</sub> -9.2	Gm09:7277766..7277656	-110	Alta	-	-
(TGA) <sub>6</sub> -9.3	Gm09:30664066..30663947	-119	Moderada	-	-
(TGA) <sub>6</sub> -9.4	Gm09:31116564..31116742	178	Alta	-	-
(TGA) <sub>6</sub> -9.5	Gm09:42363403..42363479	76	Alta	Glyma09g36690 (SERINE/THREONINE PROTEIN KINASE / Microtubule-associated serine/threonine kinase and related proteins / Protein tyrosine kinase)	-
(TGA) <sub>6</sub> -10.1	Gm10:447967..447967	74	Alta	-	-
(TGA) <sub>6</sub> -10.2	Gm10:28598997..28599383	386	Moderada	-	-
(TGA) <sub>6</sub> -10.3	Gm10:47768712..47768646	-66	Alta	-	-
(TGA) <sub>6</sub> -10.4	Gm10:49091577..49091409	-168	Moderada	-	-
(TGA) <sub>6</sub> -11.1	Gm11:8631678..8631737	59	Alta	Glyma11g12090 Glyma11g14370 (CATION EFFLUX PROTEIN/ ZINC TRANSPORTER / Zn <sup>2+</sup> transporter / Cation efflux family)	-
(TGA) <sub>6</sub> -11.2	Gm11:10279513..10279513	66	Alta	-	-
(TGA) <sub>6</sub> -11.3	Gm11:14546844..14546891	47	Moderada	-	-
(TGA) <sub>6</sub> -11.4	Gm11:17927713..17927756	43	Moderada	-	-
(TGA) <sub>6</sub> -11.5	Gm11:18778275..18778275	88	Moderada	-	-
(TGA) <sub>6</sub> -11.6	Gm11:35095808..35095719	-89	Moderada	-	-
(TGA) <sub>6</sub> -12.1	Gm12:9373190..9373126	-64	Moderada	-	-
(TGA) <sub>6</sub> -12.2	Gm12:14388896..14388802	-94	Moderada	-	-
(TGA) <sub>6</sub> -12.3	Gm12:15430520..15430607	87	Moderada	-	-
(TGA) <sub>6</sub> -12.4	Gm12:19686546..19686595	49	Moderada	-	-
(TGA) <sub>6</sub> -12.5	Gm12:30834149..30834215	66	Moderada	Glyma12g27320 Glyma12g28970 (TCP family transcription factor)	-
(TGA) <sub>6</sub> -12.6	Gm12:32363006..32363046	40	Alta	Glyma13g16500 (NUCLEOSOME ASSEMBLY PROTEIN / Nucleosome assembly protein NAP-1 / Nucleosome assembly protein (NAP))	-
(TGA) <sub>6</sub> -13.1	Gm13:20452807..20452713	-94	Alta	Glyma13g23200	-
(TGA) <sub>6</sub> -13.2	Gm13:26654346..26654102	-244	Alta	Glyma13g31190	-
(TGA) <sub>6</sub> -13.3	Gm13:33632763..33632817	54	Alta	Glyma13g32870 Glyma13g32880 (Protein of unknown function, DUF593)	-
(TGA) <sub>6</sub> -13.4	Gm13:34920397..34920457	60	Alta	Glyma14g01890	-
(TGA) <sub>6</sub> -14.1	Gm14:1092103..1092148	45	Alta	Glyma14g03780 (ATP-DEPENDENT HELICASE SMARCA (SWI/SNF-RELATED MATRIX- ASSOCIATED ACTIN- DEPENDENT REGULATOR OF CHROMATIN A)- RELATED / Chromodomain-helicase DNA-binding protein / 'chromo' (CHRromatin Organisation Modifier)	-
(TGA) <sub>6</sub> -14.2	Gm14:2447515..2447611	96	Alta	-	-

				domain)	
(TGA) <sub>6</sub> -14.3	Gm14:11747231..11747330	99	Alta	-	-
(TGA) <sub>6</sub> -14.4	Gm14:30679380..30679597	217	Moderada	-	-
(TGA) <sub>6</sub> -14.5	Gm14:38197220..38197220	-104	Moderada	-	-
(TGA) <sub>6</sub> -15.1	Gm15:2001015..2000914	-101	Alta	Glyma15g02860 (Nucleosome assembly protein NAP-1 (NAP))	-
(TGA) <sub>6</sub> -15.2	Gm15:23606194..23606092	-102	Moderada	-	-
(TGA) <sub>6</sub> -15.3	Gm15:26582116..26582173	57	Moderada	-	-
(TGA) <sub>6</sub> -15.4	Gm15:30994342..30994414	72	Moderada	-	-
(TGA) <sub>6</sub> -15.5	Gm15:39202734..39202635	-99	Moderada	-	-
(TGA) <sub>6</sub> -15.6	Gm15:50451089..50451142	53	Moderada	-	-
(TGA) <sub>6</sub> -16.1	Gm16:5227532..5227590	58	Moderada	Glyma16g05840 (TCP family transcription factor)	-
(TGA) <sub>6</sub> -16.2	Gm16:24119128..24119128	117	Moderada	-	-
(TGA) <sub>6</sub> -16.3	Gm16:25972856..25972950	94	Alta	Glyma16g22480	-
(TGA) <sub>6</sub> -16.4	Gm16:28818478..28818554	76	Alta	-	-
(TGA) <sub>6</sub> -16.5	Gm16:29804953..29805017	64	Moderada	Glyma16g25750 (LIM DOMAIN / Adaptor protein Enigma and related PDZ-LIM proteins / Ubiquitin interaction motif)	-
(TGA) <sub>6</sub> -17.1	Gm17:4367550..4367644	94	Alta	Glyma17g06170 (Nucleosome assembly protein (NAP-1))	-
(TGA) <sub>6</sub> -17.2	Gm17:8726555..8726660	105	Alta	Glyma17g11610	-
(TGA) <sub>6</sub> -17.3	Gm17:8735388..8735486	98	Alta	Glyma17g11630	-
(TGA) <sub>6</sub> -17.4	Gm17:31399671..31399734	63	Moderada	Glyma17g29170	-
(TGA) <sub>6</sub> -17.5	Gm17:34165281..34165410	129	Moderada	-	-
(TGA) <sub>6</sub> -18.1	Gm18:4626393..4626499	106	Alta	Glyma18g06030 (METALLOPROTEASE / N-arginine dibasic convertase NRD1 and related Zn <sup>2+</sup> -dependent endopeptidases, insulinase superfamily / Peptidase M16 inactive domain)	-
(TGA) <sub>6</sub> -18.2	Gm18:9884486..9884440	-46	Alta	-	-
(TGA) <sub>6</sub> -18.3	Gm18:18905678..18905605	-73	Moderada	-	-
(TGA) <sub>6</sub> -18.4	Gm18:42152396..42152396	88	Alta	-	-
(TGA) <sub>6</sub> -19.1	Gm19:178355..178438	83	Alta	Glyma19g00380 ( serine/threonine protein kinase)	-
(TGA) <sub>6</sub> -19.2	Gm19:8780981..8781073	92	Moderada	-	-
(TGA) <sub>6</sub> -19.3	Gm19:23904735..23904796	61	Baixa	-	-
(TGA) <sub>6</sub> -19.4	Gm19:35944592..35944637	45	Moderada	-	-
(TGA) <sub>6</sub> -19.5	Gm19:38160344..38160271	-73	Alta	Glyma19g30540 (NASCENT POLYPEPTIDE ASSOCIATED COMPLEX ALPHA SUBUNIT-RELATED / Transcription factor containing NAC and TS- N domains / NAC domain)	-
(TGA) <sub>6</sub> -20.1	Gm20:5499649..5499649	-52	Moderada	-	-
(TGA) <sub>6</sub> -20.2	Gm20:15642188..15642118	-70	Alta	-	-

<b>(TGA)<sub>6</sub>-20.3</b>	<b>Gm20:30489848..30489972</b>	124	Moderada	-	Glyma20g21290
				Glyma20g23790 (RING FINGER PROTEIN 6/12/38 / Anaphase-promoting complex (APC), subunit 11 (PHD-finger)	
<b>(TGA)<sub>6</sub>-20.4</b>	<b>Gm20:33536236..33536126</b>	-110	Alta		-

---

## 8.2 INSTRUÇÕES PARA AUTORES – ANNALS OF BOTANY



### Information for Authors

#### INTRODUCTION

#### PREPARING THE ARTICLE FILE

#### PREPARING TABLE FILES, FIGURE FILES, SUPPLEMENTARY INFORMATION FILES and VIDEO FILES

#### THE REVIEW PROCESS

#### FORMATTING AND SUBMITTING A REVISED PAPER

#### ACCEPTANCE, PROOFS, PRODUCTION AND PUBLICATION

#### FORMAL STATEMENT

**New for 2010 – Please note that the journal now encourages authors to complete their copyright licence to publish form online**

### INTRODUCTION

#### Scope of the journal

*Annals of Botany* is published for the Annals of Botany Company by Oxford University Press. Experimental, theoretical and applied papers on all aspects of plant science are welcome. The submitted manuscript or its essential content must not have been published previously or be under consideration for publication elsewhere. To merit publication in *Annals of Botany*, contributions should be substantial, written in clear English and combine originality of content with potential general interest. Submission of manuscripts that report small incremental advances or are of geographically local interest only is discouraged unless the implications of the findings are wide-reaching. Agronomic papers are expected to contain a substantial amount of basic plant biology. In general, a paper is unlikely to be accepted unless the referees and editors involved in its evaluation are enthusiastic about the science. The Covering Letter is an essential part of all submissions. It should include an ~60 word summary of the scientific strengths of the paper that the author(s) believe qualify it for consideration by *Annals of Botany*.

Authors whose first language is not English are well advised to have the manuscript, including Abstract, and the Covering Letter, checked by a native English speaker prior to submission. Many professional language-editing services are available on the internet. One such service is provided by Rescript Co of New Zealand, which includes help with manuscript revisions. It is contactable on <http://www.rescript.co.nz>. Other services are smart English, which can be contacted via <http://www.smartenglish.co.uk/>, and proofproper at [www.proofproper.co.uk](http://www.proofproper.co.uk). The *Annals of Botany* takes no responsibility for the quality of the service offered by these companies and use of such a

service does not guarantee that any manuscript will be accepted for publication. We welcome feedback about these services, and suggestions for any others.

## **Charges**

Authors pay no fees or page charges unless electing for our Open Access scheme (see below for details). The corresponding author receives a free copy of the issue of the Journal in which their paper appears and a unique URL that gives access to a PDF (Portable Document Format) file of their article. In addition, 25 offprints of the article will be supplied without charge. Additional offprints can be ordered using the [offprint order form](#) sent out with the proofs. Orders from the UK will be subject to a 17.5% VAT charge. For orders from elsewhere in the EU you or your institution should account for VAT by way of a reverse charge. Please provide us with your or your institution's VAT number. Colour photographs and graphics are also printed without charge where their use enhances scientific content or clarity.

## **Open Access**

*Annals of Botany* authors have the option to make their accepted paper freely available online immediately upon publication, under the '[Oxford Open](#)' initiative. Authors can choose this open access option when completing the online Licence to Publish. Here, authors are asked to indicate whether or not they wish to have their paper made freely available immediately online. There is a charge, which varies depending on circumstances (see <http://www.oxfordjournals.org/oxfordopen/charges.html>) but it is considerably cheaper for authors whose university or institution subscribes to the Journal, and for authors in developing countries. Orders from the UK will be subject to a 17.5% VAT charge. For orders from elsewhere in the EU you or your institution should account for VAT by way of a reverse charge. Please provide us with your or your institution's VAT number. If you do not select the Open Access option, your paper will be published with standard subscription-based access and you will not be charged.

## **Types of article**

Standard research papers ('ORIGINAL ARTICLES') and 'TECHNICAL ARTICLES' should not normally exceed ten printed pages (each printed page in the journal holds approximately 1000 words or 40–50 references). A 'REVIEW' submitted speculatively should have fewer than 24 printed pages. 'SHORT COMMUNICATIONS' should not exceed six printed pages. Short opinion papers ('VIEWPOINT') up to 6 pages long will also be considered. 'INVITED REVIEWS' (up to 24 pages) and 'BOTANICAL BRIEFINGS' (up to 6 pages) are published by invitation only. The journal also publishes book reviews ([Publishers' Books for Review](#)).

## **Summary of submission processes**

Submission management and evaluation of submitted manuscripts will involve the Journal's online manuscript submission system. The manuscript text should be prepared in English (see **PREPARING THE ARTICLE FILE** below for details) and submitted online starting from our login page. Figures, tables and other types of content should be organized into separate files for submission (see **PREPARING TABLE** and **FIGURE FILES, SUPPLEMENTARY INFORMATION FILES** and **VIDEO FILES** below for details). If you are using the online submission system for the first time please go to

the login page and generate a login name and password after clicking on the “**First time authors only should register here**” link. If you are already registered but need to be reminded of your login name or password please go to the login page and click on “**Unknown/Forgotten password?**”. There is extensive guidance available throughout the submission process. To make use of this guidance please click on the “Author Instructions” link or the “Tips” link situated at the top of every screen. In addition, there are frequent context-sensitive help points throughout the site that can be opened by clicking on the following symbol ?.

If you are unable to access our web-based submission system, please contact the Editorial Office for alternative methods of submitting your paper.

[Back to Start](#)

### **Preparing a covering letter**

Preparing a covering letter

Each submission should be accompanied by a **Covering Letter** formatted in MS Word (file type DOC) or in Rich Text Format (file type RTF). The letter should include contact details of the corresponding author, the title and authorship of the paper, and should state if the paper is a first submission, revision or a resubmission. It must also include an ~60 word summary of the scientific strengths of the paper that the author(s) believe qualify it for consideration by *Annals of Botany*. The manuscript reference number must be given if the paper is a revision or resubmission. If the paper is a revised or resubmitted manuscript, the letter should explain what changes have been made to the manuscript and where changes requested by the Handling Editor and referees have not been carried out. Any other information to which authors wish to draw the Chief Editor’s attention should also be included in this letter.

## **PREPARING THE ARTICLE FILE**

(Always consult a recent issue of *Annals of Botany* for layout and style)

Text should be typed using size 12 Times New Roman or Courier, double-spaced throughout and with an approx. 25 mm margin. All pages should be numbered sequentially. Each line of the text should also be numbered, with the top line of each page being line 1. The article file should be in PC-compatible Microsoft Word - file type DOC [please make sure the "Language" is "English (U.K)" via Tools→ Language → Set Language]. RTF files are also acceptable. Please do not use the Windows Vista DOCX format: if you have created the text in this format, please save the files as RTF before submitting them. Please do *not* submit PDFs, desktop publishing files or LaTeX files. The article file should *include* a list of any figure legends but *exclude* tables and any figures themselves – these should be submitted separately. Please do *not* embed tables and images in the article file. Instead, tables and figures should each be allocated separate electronic files on your computer for later uploading as explained below under **PREPARING TABLE and FIGURE FILES, SUPPLEMENTARY INFORMATION FILES AND VIDEO FILES.**

It is NOT journal style to have footnotes within articles. Any such notes must be incorporated into the main text, for example within brackets or as a separate paragraph.

The **first page** should state the type of article (e.g. Original Article, Technical Article) and provide a concise and informative full **title** followed by the names of all authors. Where necessary, each name should be followed by an identifying superscript number (<sup>1, 2, 3</sup> etc.) associated with the appropriate institutional address to be entered further down the page. For papers with more than one author, the corresponding author's name should be followed by a superscript asterisk\*. The institutional address(es) of each author should be listed next, each address being preceded by the relevant superscript number where appropriate. A running title of not more than 75 characters, including spaces, should also be provided, followed by the e-mail address of the corresponding author. Please follow the layout used for the first page of papers published in *Annals of Botany*.

The **second page** should contain a structured **Abstract** not exceeding 300 words made up of bulleted headings. For 'ORIGINAL ARTICLES' these heading will normally be as follows:

- *Background and Aims*
- *Methods*
- *Key Results*
- *Conclusions*

Alternative bulleted headings, such as '*Background*', '*Scope*' and '*Conclusions*', are also acceptable for 'REVIEWS', 'INVITED REVIEWS', 'BOTANICAL BRIEFINGS', 'TECHNICAL ARTICLES' papers and 'VIEWPOINT' papers.

The Abstract should be followed by between three and 12 **Key words** that include the complete botanical name(s) of any relevant plant material. If many species are involved, species groups should be listed instead. Note that essential words in the title should be repeated in the key words since these, rather than the title, are used in some electronic searches. **Title**, **Abstract** and **Key words** should be self-explanatory without reference to the remainder of the paper.

The **third and subsequent pages** should comprise the remaining contents of the article text. 'ORIGINAL ARTICLES' and 'SHORT COMMUNICATIONS' will usually have the structure INTRODUCTION, MATERIALS AND METHODS, RESULTS, DISCUSSION, ACKNOWLEDGEMENTS and LITERATURE CITED followed by a list of captions to any figures.

The RESULTS section should not include extensive discussion and data should not be repeated in both graphical and tabular form. The DISCUSSION section should avoid extensive repetition of the RESULTS and *must* finish with some conclusions.

**Abbreviations** are discouraged *except* for units of measurement, standard chemical symbols (e.g. S, Na), names of chemicals (e.g. ATP, Mes, Hepes, NaCl, O<sub>2</sub>), procedures (e.g. PCR, PAGE, RFLP), molecular terminology (e.g. bp, SDS) or statistical terms (e.g. ANOVA, s.d., s.e., *n*, *F*, *t*-test and *r*<sup>2</sup>) where *these are in general use*. Other abbreviations should be spelled out at first mention and all terms must be written out in full when used to start a sentence. Abbreviations of scientific terms should not be followed by a full stop. Use the minus index to indicate 'per' (e.g. m<sup>-3</sup>, L<sup>-1</sup>, h<sup>-1</sup>) except in such cases as 'per plant' or 'per pot'. If you decide that a list of abbreviations would help the reader, this should be included as an Appendix.

**Units of Measurement.** Use the *Système international d'unités* (SI) wherever possible. If non-SI units have to be used, the SI equivalent should be added in parentheses at first mention. For units of volume, expressions based on the cubic metre (e.g.  $5 \times 10^{-9} \text{ m}^3$ ,  $5 \times 10^{-6} \text{ m}^3$  or  $5 \times 10^{-3} \text{ m}^3$ ) or the litre (e.g. 5  $\mu\text{L}$ , 5 mL, 5 L) are acceptable, but one or other system should be used consistently throughout the manuscript. Typical expressions of concentrations might be 5  $\text{mmol m}^{-3}$ , 5  $\mu\text{M}$  (for 5  $\mu\text{mol L}^{-1}$ ), or 25  $\text{mg L}^{-1}$ . The Dalton (Da), or more conveniently the kDa, is a permitted non-S

**Names of plants** must be written out in full (Genus, species) in the abstract and again in the main text for every organism at first mention (but the genus is only needed for the first species in a list within the same genus, e.g. *Lolium annuum*, *L. arenarium*). The authority (e.g. L., Mill., Benth.) is *not* required unless it is controversial. Guidance for naming plants correctly is given in The International Plant Names Index and in *The Plant Book: a Portable Dictionary of the Vascular Plants* (1997) by D.J. Mabberley (Cambridge: Cambridge University Press. ISBN 0521-414210-0). After first mention, the generic name may be abbreviated to its initial (e.g. *A. thaliana*) except where its use causes confusion.

Any cultivar or variety should be added to the full scientific name e.g. *Solanum lycopersicum* 'MoneyMaker' following the appropriate international code of practice. For guidance, refer to the ISHS International Code of Nomenclature for Cultivated Plants (2004) edited by C.D. Brickell, B. R. Baum, W. L. A. Hetterscheid, A. C. Leslie, J. McNeill, P. Trehane, F. Vrugtman, J. H. Wiersema (ISBN 3-906166-16-3).

Once defined in full, plants may also be referred to using vernacular or quasi-scientific names without italics or uppercase letters (e.g. arabidopsis, dahlia, chrysanthemum, rumex, soybean, tomato). This is often more convenient.

Items of **Specialized Equipment** mentioned in MATERIALS AND METHODS should be accompanied by details of the model, manufacturer, and city and country of origin.

**Numbers** up to and including ten should be written out unless they are measurements. All numbers above ten should be in numerals except at the start of sentences. **Dates** should be in the form of 10 Jan. 1999, and **Clock Time** in the form of 1600 h.

**Mathematical equations** must be in proper symbolic form; word equations are not acceptable. Each quantity should be defined with a unique *single character* or symbol together with a descriptive subscript if necessary. Each subscript should also be a *single character* if possible, but a short word is permissible. For example, a relationship between plant dry mass and fresh mass should appear as  $M_d = 0.006M_f^{1.461}$ , where  $M_d$  is plant dry mass and  $M_f$  is plant fresh mass; and not as  $DM = 0.006FM^{1.461}$ .

The meaning of terms used in equations should be explained when they first appear. Standard conventions for use of *italics* only for variables should be followed: normal (Roman) font should be used for letters that are identifiers. Thus in the above example,  $M$  is the *variable quantity* of mass, the subscripts d and f are identifiers for dry and fresh respectively.

*Special note regarding 'Equation Editor' and other software for presentation of mathematics.* Symbols and equations that are imported into Word documents as embedded objects from other software packages are generally incompatible with

typesetting software and have to be re-keyed as part of the proof-making process. It is therefore **strongly advisable** to type symbols and equations directly into MS Word wherever possible. Importing from other software should ideally be confined to situations where it is essential, such as two-line equations (i.e. where numerators and denominators cannot be set clearly on a single line using '/') and to symbols that are not available in Word fonts. This will minimize the risk of errors associated with rekeying by copyeditors.

**Summary statistics** should be accompanied by the number of replicates and a measure of variation such as standard error or least significance difference. Analysis of variance is often appropriate where several treatments are involved. Presentation of an abridged ANOVA table is permissible when its use illustrates critical features of the experiment.

**Chemical, biochemical and molecular biological nomenclature** should be based on rules of the International Union of Pure and Applied Chemistry (IUPAC) and the International Union of Biochemistry and Molecular Biology (IUBMB). Chapter 16 of *Scientific Style and Format. The CBE Manual for Authors, Editors, and Publishers 6th edn.*, by Edward J. Huth (Cambridge: Cambridge University Press. ISBN 0-521-47154-0) gives guidelines.

**Sequence information.** Before novel sequences for proteins or nucleotides can be published, authors are required to deposit their data with one of the principal databases comprising the International Nucleotide Sequence Database Collaboration: EMBL Nucleotide Sequence Database, GenBank, or the DNA Data Bank of Japan and to include an accession number in the paper. Sequence matrices should only be included if alignment information is critical to the message of the paper. Such matrices can be in colour but should not occupy more than one printed page. Larger matrices will only be printed by special agreement but may more readily be published electronically as **Supplementary Information** (see below).

**Gene nomenclature.** Species-specific rules on plant gene nomenclature are available for:

maize;  
rice;  
wheat and

arabidopsis.

Otherwise, *Annals of Botany* adopts the following conventions for abbreviations: each gene abbreviation is preceded by letters identifying the species of origin. Lower-case italics should be used for mutant genes (e.g. *Rp-etr1*); upper-case italics (e.g. *Le-ACOI*) for wild-type genes; upright lower-case for proteins of mutated genes (e.g. Le-adh1); and upright upper-case for proteins of wild-type genes (e.g. At-MYB2). It may often be helpful to readers if the names of genes or gene families are spelled out in full at first mention.

**Citations in the text.** These should take the form of Felle (2005) or Jacobsen and Forbes (1999) or (Williamson and Watanabe, 1987; Rodrigues, 2002*a, b*) and be ordered chronologically. Papers by three or more authors, even on first mention, should be abbreviated to the name of the first author followed by et al. (e.g. Zhang *et al.*, 2005).

If two different authors have the same last name, give their initials (e.g. NH Kawano, 2003) to avoid confusion. Only refer to papers as 'in press' if they have been accepted for publication in a named journal, otherwise use the terms 'unpubl. res.', giving the initials and location of the person concerned. (e.g. H Gautier, INRA, Lusignan, France, unpubl. res.) or 'pers. comm.' (e.g. WT Jones, University of Oxford, UK, 'pers. comm.')

The **LITERATURE CITED** should be arranged alphabetically based on the surname of the first or sole author. Where the same sole author or same first author has two or more papers listed, these papers should be grouped in year order. Where such an author has more than one paper *in the same year*, these should be ordered with single authored papers first followed by two-author papers (ordered first alphabetically based on the second author's surname, then by year), and then any three-or-more-author papers (in year order only). Italicized letters 'a', 'b', 'c', etc., should be added to the date of papers with the same first authorship and year.

For papers with *six* authors or fewer, please give the names of *all* the authors. For papers with *seven* authors or more, please give the names of the *first three* authors only, followed by *et al.*

Each entry must conform to one of the following styles according to the type of publication.

#### *Books*

**Öpik H, Rolfe S. 2005.** *The physiology of flowering plants. Physicochemical and environmental plant physiology*, 4th edn. Cambridge: Cambridge University Press.

#### *Chapters in books*

**Scandalios JG. 2001.** Molecular responses to oxidative stress. In: Hawkesford MJ, Buchner P, eds. *Molecular analysis of plant adaptation to the environment*. Dordrecht: Kluwer, 181-208.

#### *Research papers*

**Popper ZA, Fry SC. 2003.** Primary cell wall composition of bryophytes and charophytes. *Annals of Botany* **91**: 1–12.

#### *Theses*

**Tholen D. 2005.** *Growth and photosynthesis in ethylene-insensitive plants*. PhD Thesis, University of Utrecht, The Netherlands.

#### *Anonymous sources*

**Anonymous. Year.** *Title of booklet, leaflet, report, etc.* City: Publisher or other source, Country.

Online references should be structured as: **Author(s) name, author(s) initial(s). year.** *Full title of article.* Full URL. Date of last successful access (e.g. 12 Jan. 2003)

**Acknowledgements.** In the ACKNOWLEDGEMENTS, please be brief. 'We thank . . .' (not 'The present authors would like to express their thanks to . . .').

**Funding information.** Details of all funding sources for the work in question should be given in a separate section entitled 'Funding'. This should appear before the 'Acknowledgements' section.

The following rules should be followed:

- The sentence should begin: 'This work was supported by ...'
- The full official funding agency name should be given, i.e. 'the National Cancer Institute at the National Institutes of Health' or simply 'National Institutes of Health' not 'NCI' (one of the 27 subinstitutions) or 'NCI at NIH' (full RIN-approved list of UK funding agencies)
- Grant numbers should be complete and accurate and provided in brackets as follows: '[grant number ABX CDXXXXXX]'
- Multiple grant numbers should be separated by a comma as follows: '[grant numbers ABX CDXXXXXX, EFX GHXXXXXX]'
- Agencies should be separated by a semi-colon (plus 'and' before the last funding agency)
- Where individuals need to be specified for certain sources of funding the following text should be added after the relevant agency or grant number 'to [author initials]'

An example is given here: 'This work was supported by the National Institutes of Health [P50 CA098252 and CA118790 to R.B.S.R.] and the Alcohol & Education Research Council [HFY GR667789].

Oxford Journals will deposit all NIH-funded articles in PubMed Central. See [http://www.oxfordjournals.org/for\\_authors/repositories.html](http://www.oxfordjournals.org/for_authors/repositories.html) for details. Authors must ensure that manuscripts are clearly indicated as NIH-funded using the guidelines above.

### **Appendix.**

If elaborate use is made of units, symbols and abbreviations, or a detailed explanation of one facet of the paper seems in order, further details may be included in a separate APPENDIX placed after the LITERATURE CITED. For more detail and information on types of files required for text, graphics and tables etc., please see the next section.

[Back to Start](#)

## **PREPARING TABLE FILES, FIGURE FILES, SUPPLEMENTARY INFORMATION FILES AND VIDEO FILES**

Each table, figure, video and set of supplementary information should be prepared as a separate file on your computer in preparation for online submission. Towards the bottom of the first submission screen of the online submission system, you should enter the appropriate number of files you have in each category. This creates the spaces (boxes) that will accommodate the files when they are uploaded later. The files are

categorized as 'Colour Figures', 'Black and White Figures', 'Tables', 'Supplemental Material' and 'Video'.

**Tables.** The best guide for laying out tables and diagrams are papers in a recent issue of *Annals of Botany*. Each table should have a separate file, a complete caption at the top and be numbered Table 1, Table 2 etc. according to the order in which they are first mentioned in the text. When preparing tables, adopt the 'Tables' set-up in MS Word, using one cell for each datum cluster (e.g.  $12.2 \pm 1.65$ ) and avoid the use of the 'return' key. Please do not use MS Excel for submitting tables. These can easily be copied into MS Word files prior to submission.

**Figures.** All images (e.g. line diagrams, drawings, graphs, photographs, plates) are considered to be 'Figures'. Each figure should be in a separate file and be numbered (Fig. 1, Fig. 2 etc.) according to the order in which they are first mentioned in the text. Electron and light **photomicrographs** should have internal scale markers. **Colour images** are encouraged and printed without charge where they enhance significantly the clarity of the scientific information. **Line diagrams** will normally be black on white and boxed with inward scale markings. Use of colour in line diagrams may sometimes be agreed where this enhances clarity significantly. Use open and/or closed circles, squares and triangles for symbols in line graphs. Height and width should be chosen for either single (8.4 cm wide) or double (up to 17.3 cm wide) column reproduction. Grouping of related graphics into a single figure is strongly encouraged. When a block of illustrative material consists of several parts, each part should be labelled A, B, C, etc. and not treated as separate figures. Note that graphs and diagrams of finally accepted papers are normally redrawn by the publisher to ensure a consistent house style and should be inspected carefully by authors at the proof stage.

Simple black and white **line drawings and graphs** should be supplied as approx. 300 dpi JPG files or MS PowerPoint files. The publisher will almost always redraw all such material if the paper is accepted. More complicated drawings, such as detailed botanical illustrations will not be redrawn and should be supplied as 600 dpi JPG files. For continuous tone images (e.g. **photographs**), please supply JPG files at 300 dpi (or 600 dpi if the image is a mix of pictures and text and/or has thin lines). Keeping total files sizes down will lessen up- and downloading times. To help achieve this *all images should be submitted at approximately the physical size they would appear in the Journal*. Scaling, sizing and cropping are best carried out within image handling programs such as Adobe PhotoShop or Corel PhotoPaint. Please do not supply photographic images as PowerPoint files as these are generally of poor resolution. Note that PDF files are not acceptable. Also, please ensure that images that do NOT contain colour are saved as 'grayscale' and that any layers have been flattened – taking these steps can make the file size up to 10 times smaller. Note that a JPG file should not be repeatedly saved as this reduces quality.

Large amounts of additional information can be submitted for publication electronically as **Supplementary Information** provided that it is not essential for a basic understanding of the main paper. Supplementary material will be refereed along with the core paper. At appropriate positions in the main text authors should indicate what details are being made available, followed by the words [**Supplementary Information**] in bold and between square brackets. The online submission system provides space for supplementary information to be uploaded in "Supplemental Material" files. The appropriate number of these types of file can be selected towards the bottom of the first

submission screen. Similarly, if you are including a **video** you should enter [**Supplementary Information - Video**] in bold and between square brackets at the appropriate place(s) in the text. A video can be uploaded after selecting a “Video” file on the first submission screen. The movie should be created in a widely available program such as Windows MediaPlayer. A short paragraph describing the contents of any Supplementary Information or Video should also be inserted in the main text immediately before ACKNOWLEDGEMENTS.

[Back to Start](#)

## THE REVIEW PROCESS

The corresponding author and all co-authors receive an acknowledgment of receipt of the manuscript and a manuscript reference number by e-mail. The corresponding author is informed when a Handling Editor has been assigned to the paper. Manuscripts considered suitable for peer review are sent to at least two outside referees. We give referees a target of two weeks for the return of their reports. Currently less than 25 % of submitted papers are accepted. Authors are asked to revise provisionally accepted articles within four weeks. To view the make-up of the Editorial Board click on [View full editorial board](#).

[Back to Start](#)

## FORMATTING AND SUBMITTING A REVISED PAPER

The technical requirements for the Article, Table and Figure Files etc. are as described above for the first submission. If the technical requirements are not met, the paper will be sent back to the author until satisfactory files are provided. Revised papers are checked by a member of the Editorial Board and may be subject to a further round of refereeing.

[Back to Start](#)

## ACCEPTANCE, PROOFS, PRODUCTION AND PUBLICATION

When a paper is finally accepted you will be asked to supply some additional material for our **ContentSnapshot** feature. Each **ContentSnapshot** comprises a thumbnail image relevant to the paper and a short summary of its principal findings. For this, you will be asked to prepare a suitable Snapshot Image file (in colour) for the thumbnail illustration and also a short summary title and text (up to 60 words) to associate with the image. Examples of **ContentSnapshots** can be found at <http://aob.oxfordjournals.org/content/106/4/i.full.pdf+html>.

You will also be invited to submit an eye-catching front **cover picture** and about 60 words of text for possible printing on the inside cover of the issue in which your article would appear. The technical requirements are similar to those for manuscript photographs. The picture should be sharp, of good contrast and be related to the content of the submitted paper; however, it need not be duplicated in the paper itself. The image should be sent as a TIFF, JPG or GIF file at 300 dpi, size approx. 10 × 10 cm. Authors

of selected material will receive a copy of the cover illustration and a complimentary copy of the relevant issue of the Journal.

After acceptance you will also be asked to complete the online Licence to Publish form. This form also offers the opportunity to choose to have the full text and PDF versions of the paper made available to non-subscribers online from the time of first publication (**Open Access**). There is a charge for this, which varies depending on circumstances (see <http://www.oxfordjournals.org/oxfordopen/charges.html>) but it is considerably cheaper for authors whose university or institution subscribes to the Journal, and for authors in developing countries. If you do not select the Open Access option, your paper will be published with standard subscription-based access and no charge will be made.

Information about the **New Creative Commons licence** can be found [here](#).

Once a satisfactorily revised version has been received and accepted, the title of the paper, authorship and hyperlinked e-mail address of the corresponding author will be posted on the *Annals of Botany* website under **AOBFirstAlert**. This feature is accessible by subscribers and non-subscribers from the [Journal's home page](#). Corresponding authors will receive PDF **proofs** by e-mail attachment approximately 4–6 weeks after acceptance. Corrected proofs should be returned within 24 h. Adobe Acrobat Reader will be needed to read the PDF proof and is downloadable without charge from: <http://www.adobe.com/products/acrobat/readstep2.html>. Authors should pay special attention to **diagrams, figures** and to **equations** since these items are usually re-keyed or redrawn by the publisher. At this stage, authors will also be invited to order offprints and extra single copies of the issue in which the article will appear.

### **Publication and printing process**

Once corrected proofs have been received and checked, the paper is posted on the website approximately six weeks ahead of print under **AOBPreview**. Each article is identified by a unique DOI (Digital Object Identifier), a code that can be used in bibliographic referencing and searching. The DOI and date of electronic publication in **AOBPreview** are also printed in the normal fully paginated monthly issue that is published about six weeks later. The paper will appear online and in print during the week preceding the start of the month of issue. The dates of submission, first return for revision, final acceptance and date of electronic publication of each article are printed on each paper.

The corresponding author will receive a free copy of the printed issue in which their paper appears and a free URL that gives access to the article online and to a downloadable PDF. In addition, 25 free printed copies of the article will be supplied. These items are normally dispatched within seven days of publication of the printed journal. The corresponding author is responsible for distributing this URL to any co-authors.

### **Post-publication services**

Monthly alerts that supply the Journal's current **Table of Contents** can be requested by clicking on [Email table of contents](#) or by using an RSS feed. For more details on the latter click on [XML RSS feed](#). Readers can also be alerted to related papers in *Annals of*

*Botany* and a wide range of other journals using the High Wire 'CiteTrack' alerting system. To access this click on [CiteTrack](#).

[Back to Start](#)

## **FORMAL STATEMENT**

Upon receipt of accepted manuscripts at Oxford Journals authors will be invited to complete an online copyright licence to publish form. Please note that by submitting an article for publication you confirm that you are the corresponding/submitting author and that Oxford University Press ("OUP") may retain your email address for the purpose of communicating with you about the article. You agree to notify OUP immediately if your details change. If your article is accepted for publication OUP will contact you using the email address you have used in the registration process. Please note that OUP does not retain copies of rejected articles.

Authors or their employers retain copyright on articles published in *Annals of Botany*. However, it is a condition of publication in the Journal that authors or their employers grant an exclusive licence to the Annals of Botany Company by completing and signing the Licence to Publish. This ensures that requests from third parties to reproduce articles are handled efficiently and consistently and allows the article to be disseminated as widely as possible. The Licence permits authors to use their own material in other publications provided that the Journal is acknowledged as the original place of publication and that the Annals of Botany Company is notified in writing

Papers are published on the understanding that the work is free of plagiarism, that all authors have agreed to publication in *Annals of Botany* and that those contributing substantially to the work have been appropriately acknowledged or given co-authorship. The official publication date is the date on which the paper is first posted electronically on the website. This date will normally be when the paper appears in *AOBPreview*. If a paper is not posted in *AOBPreview*, the date of publication is the date of first appearance in a fully paginated print or electronic monthly issue. For more about rights and permissions click on this link [Rights](#)

### **Author Self-Archiving/Public Access policy**

For information about this journal's policy, please visit our [Author Self-Archiving policy page](#).

[Back to Start](#)

### **8.3 INSTRUÇÕES PARA AUTORES – CHROMOSOME RESEARCH**

## **Chromosome Research**

### **Instructions for authors**

#### **Manuscript Submission**

There are no page charges or charges for the publication of colour illustrations or administration charges for papers published in Chromosome Research.

It is understood that papers submitted for publication have not been published previously and are not simultaneously offered to any other journal. Before submission, the submitting author must ensure that the manuscript has been seen and approved by all other named authors.

#### **How to Submit your Manuscript**

Authors should submit their manuscripts online. Please follow the hyperlink “Submit online” on the right of the [Chromosome Research web page](#), or follow this link:

<http://www.editorialmanager.com/chromres/>

Upload all of your manuscript files following the instructions given on the screen. A decision with regard to acceptance for publication will normally be given within 21 days of receipt of the manuscript.

When submitting your manuscript, please choose one of the Associate Editors as your preferred handling editor. If you do not have a preference, or leave this option open, your submission will be handled by the Editor-in-Chief.

#### **Editor-in-Chief:**

Professor Herbert Macgregor  
School of Biosciences  
University of Exeter, UK

#### **Associate Editors:**

Professor Wendy Bickmore  
MRC Human Genetics  
Unit Western General Hospital NHS Trust  
Edinburgh, Scotland, UK

Dr. Mary E. Delany  
Department of Animal Science  
University of California  
Davis, CA USA

Professor. J.S. (Pat) Heslop-Harrison

Department of Biology  
University of Leicester  
Leicester, UK

Dr. Dean Jackson  
Faculty of Life Sciences  
The University of Manchester  
Manchester, UK

Dr. Jiming Jiang  
Department of Horticulture  
University of Wisconsin -Madison  
Madison, WI, USA

Dr. Hans de Jong  
Laboratory of Genetics  
Wageningen Universiteit  
Wageningen, the Netherlands

Professor Hans J. Lipps  
Institut für Zellbiologie  
Universität Witten/Herdecke  
Witten, Germany

Dr. Yoichi Matsuda  
Department of Applied Molecular Biosciences  
Nagoya University  
Nagoya, Japan

Dr. Irina Solovei  
School of Biosciences  
University of Exeter  
Exeter, UK

Dr. Beth A Sullivan  
Duke University  
Durham, NC, USA

Dr. Walther Traut  
Institut fuer Biologie  
Universitaet Luebeck  
Luebeck, Germany

Dr. Fengtang Yang  
The Wellcome Trust Sanger Institute  
Wellcome Trust Genome Campus  
Hinxton, Cambridge, UK

### **Style and Presentation**

The manuscript should be typed with double spacing throughout, allowing for ample margins. The title page should show the paper title, names and addresses of all authors, a short running title, and fax and telephone numbers and the e-mail address for the corresponding author. The text of the paper should be arranged in the following sequence: Introduction, Materials and Methods, Results, Discussion, Acknowledgements and References. Informative legends should be provided for all illustrations and should be grouped together at the end of the paper, along with all tables. Subheadings may be inserted in the main text, but should not be numbered or lettered. Manuscripts should be written in clear, grammatical, idiomatic English. Spelling should conform to Webster's International Dictionary or The Concise Oxford English Dictionary and data should be presented simply and concisely, using Systeme International (SI) units.

### **Abbreviations**

All abbreviations not obvious to the general reader should be defined the first time they are used. A complete alphabetically arranged list of all abbreviations used, including the definitions, should be included in the manuscript.

### **References**

References should be cited in the text using the Harvard (name–date) system. Where there are three or more authors, only the first author's name should appear, followed by et al. Where several references are cited at the same point in the text, these should be arranged in chronological order. The reference list should be typed with double spacing and arranged in alphabetical order. References should include: names and initials of all authors (unless there are more than six authors, when only the first three authors should be given, followed by et al.); year of publication; full title of the article; source using abbreviations for journals as shown in Index Medicus; volume number; and first and last page numbers. Abstracts should be identified as such. For citations from books, the chapter title should be followed by the names and initials of all editors, the title of the book, edition, place of publication, publisher and first and last page numbers.

Examples: Thomas HM, Harper JA, Morgan WG (2001) Gross chromosome rearrangements are occurring in an accession of the grass *Lolium rigidum*. *Chromosome Res* 9: 585–590.

Ohno S (2001) The one-to-four rule and paralogues of sex determining genes. In: Scherer G, Schmid M, eds. *Genes and Mechanisms in Vertebrate Sex Determination*. Birkhauser Verlag, pp 1–10.

Engel E, Antonarakis SE (2002) *Genomic Imprinting and Uniparental Disomy in Medicine*. New York: Wiley–Liss.

Only accepted papers should be referenced; all other material should be referred to in the text as 'in preparation', 'personal communication' 'unpublished observations' and should not be included in the reference list.

### **Citing Internet References**

World Wide Web: All references should include the same information that would be provided for a printed source (or as much of that information as possible). The Web

information is then placed at the end of the reference. It is important to use "Retrieved from" and the date because documents on the Web may change in content, move, or be removed from a site altogether. To cite a Web site in text (but not a specific document), it is sufficient to give the address (e.g., <http://www.apa.org>) there and no reference entry is needed. However, when citing a particular web page a citation in the text (e.g. Gaten 2000) and an entry in the reference list will be required.

For example: Gaten E. (2000) Internet references. Retrieved from <http://www.le.ac.uk/biology/teach/mod300/ecitations.html> 19/9/2000

E-mail: E-mail communications from individuals should be cited as personal communications. The format in the text (personal communications are not cited in the reference list) is as follows:

(E. Gaten personal communication, March 28, 2001). It is possible to send an e-mail note disguised as someone else. Authors— not journal editors or copy editors — are responsible for the accuracy of all references, which includes verifying the ~~source~~ <sup>source</sup> of e communications before citing them as personal communications in manuscripts.

One of the most comprehensive guides to citing internet references is provided by the American Psychological Association: <http://www.apastyle.org/electref.html>

### **Keywords**

Please provide 4 to 6 keywords which can be used for indexing purposes. Key words should be chosen carefully, they are essential to electronic search tools.

### **Tables and Illustrations**

All tables and illustrations should be referred to in the text, with appropriate locations indicated in the text margin. Tables should present new information and not duplicate data included in the text. Every table should have a descriptive title and, if numerical measurements are given, the units should be included in the column heading. Line drawings should be supplied in a form suitable for high-quality reproduction. Axes should be labelled clearly; other lettering should be kept to a minimum. Avoid the use of fine tints, especially as background to text.

### **Electronic Figure Submission**

Supply all figures electronically. For vector graphics, the preferred format is EPS; for halftones, please use TIFF format. MS Office files are also acceptable. Vector graphics containing fonts must have the fonts embedded in the files. Name your figure files with "Fig" and the figure number, e.g., Fig1.eps. Color illustrations should be submitted as RGB (8 bits per channel) TIFF. Photographs and photomicrographs should be cropped as close as possible to the area of interest and should be submitted as high resolution (>300dpi) as electronic files together with the manuscript. All photomicrographs must be accompanied by a scale bar with the equivalent measurement stated in the figure legend.

### **Colour figures**

There are no charges for the publication of colour illustrations, either in the online or in the printed version of Chromosome Research.

### **Electronic Supplementary Material**

Electronic Supplementary Material (ESM) for a paper will be published in the electronic edition of this journal provided the material is submitted in electronic form together with the manuscript and accepted after peer review. Electronic supplementary material will be published as received from the author without any conversion, editing, or reformatting.

After acceptance for publication, ESM will be published as received from the author in the online version only. Reference will be given in the printed version.

ESM may consist of:

- Information that cannot be printed: animations, video clips, sound recordings (use PDF, QuickTime, .avi, .mpeg, animated GIFs, or any other common file format) Submit your text or presentation material in PDF format; .doc or .ppt files are not suitable for longterm storage.
  - Information that is more convenient in electronic form: sequences, spectral data, etc.
  - Large quantities of original data that relate to the paper, e.g. additional tables, large numbers of illustrations (colour and black & white), etc. Submission
- 
- Supply all supplementary material in standard file formats.
  - To accommodate user downloads, please keep in mind that largesized files may require very long download times and that some users may experience other problems during downloading.

Numbering If supplying any supplementary material, the text must make specific mention of the material as a citation, similar to that of figures and tables (e.g., ". . . as shown in supplement S3").

### **Legends**

Legends must be brief, selfsufficient explanations of the ESM. ESM is to be numbered and referred to as S1, S2, etc.

### **Permission and Copyright**

It is the responsibility of the author to obtain written permission for a quotation from unpublished material, or for all quotations in excess of 250 words in one extract or 500 words in total from any work still in copyright, and for the reprinting of figures or tables from unpublished or copyrighted material. Copyright in articles published in this journal is the property of Springer to the extent transferable. Authors are themselves responsible for obtaining permission to reproduce copyright material from other sources. There are no page charges or administration charges for papers published in this journal.

### **Open access publishing: Springer Open Choice™**

In addition to the normal publication process (whereby an article is submitted to the journal and access to that article is granted to customers who have purchased a subscription), Springer now provides an alternative publishing option: Springer Open Choice™. A Springer Open Choice™ article receives all the benefits of a regular subscriptionbased article, but in addition is made available publicly (free of charge) through Springer's online platform SpringerLink. To publish via Springer Open Choice™, upon acceptance please visit [www.springer.com/openchoice](http://www.springer.com/openchoice) to complete the relevant order form and provide the required payment information. Payment must be received in full before publication or articles will publish as regular subscriptionmodel

articles. We regret that Springer Open Choice™ cannot be ordered for published articles.

Please note that funding agencies with mandatory open access or public access (i.e. USA NIH and the UK Wellcome Trust) and several other institutes (The Max Planck in Germany, the Dutch UKB and others) have special arrangements for Open Access publishing with Springer. Please consult your funding agency's representative for specific details.

**Proofs**

Proofs will be sent to the corresponding author by email. Your response, with or without corrections, should be sent within 72 hours.

**Offprints**

Authors may order additional offprints on the form which accompanies the proofs.

**Additional Information**

Max Haring, PhD Publishing editor for Chromosome Research Springer Science & Business Media

P.O. Box 17 3300 AA Dordrecht The Netherlands Email: [Max.Haring@springer.com](mailto:Max.Haring@springer.com)

<http://www.springer.com/journal/10577>



Chromosome Research

Cytogenetics, genomics, chromatin and the nucleus

Editor-in-Chief: H. C. Macgregor

ISSN: 0967-3849 (print version)

ISSN: 1573-6849 (electronic version)

Journal no. 10577

## 8.4 RESUMO – 56º CONGRESSO BRASILEIRO DE GENÉTICA

## Distribuição de microssatélites no genoma de leguminosas mediante hibridização *in situ* fluorescente

Bortoleti, KCA<sup>1,2</sup>; Benko-Iseppon, AM<sup>1</sup>; Belardino, LCS<sup>1</sup>; Oliveira, ARS<sup>1</sup>; Abdelnoor, RV<sup>3</sup>; Nascimento IR<sup>4</sup>; Brasileiro-Vidal, AC<sup>1</sup>

<sup>1</sup>Laboratório de Genética e Biotecnologia Vegetal, Departamento de Genética, Universidade Federal de Pernambuco  
<sup>2</sup>Colegiado de Ciências Biológicas, Universidade Federal do Vale do São Francisco  
<sup>3</sup>Embrapa Soja, <sup>4</sup>Universidade Federal Rural de Pernambuco  
brasileiro\_vidal@hotmail.com

**Palavras-chave:** FISH, oligonucleotídeos, Leguminosas, análise *in silico*, DNA repetitivo

Os microssatélites consistem em unidades de repetição (1 a 5pb) distribuídas em tandem, que podem estar dispersas ao longo dos genomas ou associadas a regiões heterocromáticas. Tais características têm permitido sua utilização na construção de mapas cromossômicos mediante hibridização *in situ* fluorescente (FISH), auxiliando no entendimento da organização genômica entre espécies proximamente relacionadas. O presente trabalho caracterizou a distribuição de oligonucleotídeos sintéticos (AG)<sub>n</sub>, (AAG)<sub>n</sub>, (ACC)<sub>n</sub>, (CTC)<sub>n</sub> e (TGA)<sub>n</sub> em *Phaseolus vulgaris*, *P. lunatus*, *Vigna unguiculata*, *V. radiata* e *Glycine max*, identificando marcas cromossômicas para a comparação destas leguminosas. Lâminas foram hibridizadas com sondas de oligonucleotídeos e reibridizadas com DNAr 5S e 45S, provenientes de *Lotus japonicus* e *Arabidopsis thaliana*, respectivamente. Em soja, esta análise estendeu-se a uma investigação genômica, a partir de sequências disponibilizadas no banco de dados SoyBase (<http://www.soybase.org/>), usando o programa BLASTN com os seguintes parâmetros: formato de saída de alinhamentos sem lacunas, matriz de comparação blossom62, valor W padrão, *e-value* 0.1 ou menor e filtro de baixa complexidade desligado. Os oligonucleotídeos [(AAG)<sub>n</sub>]<sub>10</sub>, [(AAC)<sub>n</sub>]<sub>10</sub>, [(AG)<sub>n</sub>]<sub>10</sub>, [(CTC)<sub>n</sub>]<sub>10</sub> e [(TGA)<sub>n</sub>]<sub>10</sub> foram utilizados como sondas, adotando como ponto de corte valores de identidade de pareamento inferiores a 77%, objetivando caracterizar número, tamanho e localização destas unidades de repetições no genoma. A FISH com oligonucleotídeos evidenciaram marcações pericentroméricas e proximais, embora um padrão de distribuição disperso tenha prevalecido ao longo dos cromossomos das espécies analisadas, com exceção de (AG)<sub>n</sub> em soja que não revelou marcações visíveis. O oligonucleotídeo (AAG)<sub>n</sub> revelou uma marcação pericentromérica evidente em *P. lunatus*, tratando-se de um marcador cromossômico para tal espécie. Na hibridização com (ACC)<sub>n</sub> notou-se a presença de seis marcações fortes em *V. unguiculata*, estando uma delas localizada no par cromossômico 10, portador do sítio de DNAr 45S. O microssatélite (TGA)<sub>n</sub> revelou uma marcação bem evidente no cromossomo 10 de *P. lunatus*, identificado pela presença de DNAr 5S. Algumas divergências observadas no padrão de distribuição e na intensidade dos sinais entre os genomas das espécies confirmam a hipótese de que as sequências de DNA repetitivo apresentam uma composição heterogênea. Na análise genômica de soja, foram observados 92, 84, 83, 142 e 103 sítios de repetições para os oligonucleotídeos (AG)<sub>n</sub>, (AAG)<sub>n</sub>, (ACC)<sub>n</sub>, (CTC)<sub>n</sub> e (TGA)<sub>n</sub>, respectivamente, de tamanhos variáveis entre 30 a 454 pb, localizados, principalmente, em regiões de alta a moderada densidade gênica, por vezes associados a genes e elementos transponíveis. Considerando o pequeno tamanho, fica evidente que estas sequências não correspondem aos sítios observados pela FISH, uma vez que esta técnica evidencia apenas sequências com comprimentos acima de 1-3kb. Sugere-se que as marcações localizadas por FISH estejam associadas a regiões de heterocromatina e que sejam constituintes dos mais de 1.000 *scaffolds* existentes no sequenciamento da soja, não sendo, por essa razão, detectadas pela análise *in silico*. Apoio financeiro: CNPq, Embrapa, UFPE

## 8.5 RESUMO – 56º CONGRESSO BRASILEIRO DE GENÉTICA

### Ocorrência e distribuição cromossômica de retroelemento *Ty1-copia*-like em espécies dos gêneros *Vigna Savi* e *Phaseolus L.* mediante hibridização *in situ* fluorescente

Bortoleti, KCA<sup>1,2</sup>; Brasileiro-Vidal, AC<sup>1</sup>; Amorim, LLB<sup>1</sup>; Vasconcelos, EV<sup>1</sup>; Benko-Iseppon, AM<sup>1</sup>

<sup>1</sup>Laboratório de Genética e Biotecnologia Vegetal, Departamento de Genética, Universidade Federal de Pernambuco  
<sup>2</sup>Colegiado de Ciências Biológicas, Universidade Federal do Vale do São Francisco  
ana.benko.iseppon@pq.cnpq.br

**Palavras-chave:** FISH, *Ty1-copia*-like, feijão-caupi, feijão comum, marcador cromossômico

Os gêneros *Vigna Savi* e *Phaseolus L.* (Fabaceae) têm sido bastante utilizados em estudos citomoleculares visando à construção de mapas cromossômicos através da hibridização *in situ* fluorescente (FISH). Diferentes classes de DNA repetitivo vêm sendo caracterizadas, entre elas, os retrotransposons, os quais possuem um mecanismo de amplificação e dispersão, via RNA intermediário, resultando geralmente em uma localização dispersa ao longo dos cromossomos. Nesse sentido, o presente trabalho objetivou caracterizar a distribuição cromossômica de um retroelemento *Ty1-copia*-like em *V. unguiculata* cv. BR14-Mulato, *V. radiata*, *P. vulgaris* cv. BRS Esplendor e *P. lunatus* cv. Vermelhinha, visando identificar marcas cromossômicas que auxiliem na construção de mapas físicos comparativos para essas espécies. Sequência do retrotransposon *Ty1-copia*-like (domínio RT) foi amplificada a partir do DNA genômico de *V. unguiculata* utilizando os primers degenerados (5'-GAGAATTCACNGCNTTYYTNCAYGG-3'/5'-GAGGATCCATRTCRTCACRTANAR-3'), descritos anteriormente na literatura para *V. radiata*. Obteve-se um fragmento com aproximadamente 270 pb, o qual foi purificado pelo kit GFX™ PCR DNA and Gel Band Purification Kit (GE Healthcare). Posteriormente, esse fragmento foi reamplificado, marcado por digoxigenina-11-dUTP por PCR e utilizado como sonda na FISH. As lâminas foram rehibridizadas com DNAr 5S (D2) e 45S (R2), provenientes de *Lotus japonicus* e *Arabidopsis thaliana*, respectivamente, para identificação cromossômica. A FISH evidenciou um padrão de distribuição disperso ao longo dos cromossomos das quatro espécies analisadas; entretanto, algumas marcações mais fortes foram observadas em *V. unguiculata*, *P. vulgaris* e *P. lunatus*. Em *V. unguiculata*, sinais mais evidentes foram notados nas regiões subterminais, próximos aos sítios de DNAr 45S. Por sua vez, *V. radiata* apresentou uma marcação mais dispersa em todo seu complemento cromossômico. Em *P. vulgaris* e *P. lunatus*, a distribuição de *Ty1-copia*-like evidenciou sítios proximais mais fortes. Todavia, na última espécie, tais sinais mostraram-se em maior intensidade, destacando-se o cromossomo 10, portador do sítio DNAr 5S. Possivelmente, essas diferenças estão relacionadas com o número de repetições do retroelemento *Ty1-copia*-like nos genomas das referidas espécies. Os resultados observados indicam uma homologia entre domínios conservados de *Ty1-copia* nas diferentes espécies analisadas, identificando marcadores cromossômicos úteis em estudos de evolução genômica e relações filogenéticas na subfamília Papilionoideae  
Apoio financeiro: CNPq e UFPE

***Memorial da Aluna***

---

## 9. MEMORIAL DA ALUNA

Kyria Cilene de Andrade Bortoleti [kyriabortoleti@gmail.com], nascida em 02 de julho de 1980, Bacharel em Ciências Biológicas, título obtido em 06 de outubro de 2003, pela Universidade Federal de Pernambuco, com o trabalho de conclusão de curso intitulado “Variabilidade citogenética em plantas nativas e cultivadas de *Genipa americana* L. (Rubiaceae)”. Mestre em Genética e Evolução, pela Pós-Graduação em Genética e Biologia Molecular vinculada à Universidade Federal de Pernambuco, com colação de grau em 12 de dezembro de 2006, título obtido pela a apresentação da dissertação intitulada “Estudos citogenéticos em populações remanescentes de *Genipa americana* L. na Mata Atlântica de Pernambuco”. Na qualidade de doutoranda e primeira autora, publicou os resultados em resumos científicos intitulados: “Distribuição de microssatélites no genoma de leguminosas mediante hibridização in situ fluorescente” e “Ocorrência e distribuição cromossômica de retroelementos Ty1-*copia*-like em espécies dos gêneros *Vigna* Savi e *Phaseolus* L. mediante hibridização in situ fluorescente”, no 56º Congresso Brasileiro de Genética. Atua como professora assistente do Colegiado de Ciências Biológicas da Universidade Federal do Vale do São Francisco, com experiência nas áreas de concentração em Morfologia e Genética e Biologia Molecular. Coordena projetos de monitoria, desempenhando também atividades administrativas, sendo representante titular do Colegiado de Ciências Biológicas na Comissão Gestora do Programa Integrado de Desenvolvimento e Capacitação Estudantil (Pró-Reitoria de Ensino) e da Câmara de Integração (Pró-Reitoria de Integração).