

UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS BIOLÓGICAS
DEPARTAMENTO DE GENÉTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS
BIOLÓGICAS

TESE DE DOUTORADO

**DISTRIBUIÇÃO E PREVALÊNCIA DE
ELEMENTOS NÃO CODIFICANTES EM**
Glycine WILLD. E *Vigna* SAVI

POLLYANA KARLA DA SILVA

VIRTUS IMPAVIDA

RECIFE
2014

POLLYANA KARLA DA SILVA

Distribuição e Prevalência de Elementos Não Codificantes em *Glycine Willd.* e *Vigna Savi*

Tese para a obtenção do Grau de Doutor (a)
em Ciências Biológicas, área de concentração
“Biologia Molecular”, da Universidade Federal
de Pernambuco.

Orientadora: Prof^a Dr^a Ana Maria Benko-
Iseppon, Depto. de Genética, Centro de
Ciências Biológicas UFPE.

Coorientadora: Prof^a. Dr^a. Ana Christina
Brasileiro Vidal, Depto. de Genética, Centro
de Ciências Biológicas UFPE.

Recife
2014

Catálogo na Fonte:
Bibliotecário Bruno Márcio Gouveia, CRB-4/1788

Silva, Pollyana Karla da
Distribuição e prevalência de elementos não codificantes em *Glycine* Willd. *Vigna*
Savi / Pollyana Karla da Silva. – Recife: O Autor, 2014.

82 f.: il.

Orientadores: Ana Maria Benko-Iseppon, Ana Christina Brasileiro Vidal
Tese (doutorado) – Universidade Federal de Pernambuco. Centro de
Ciências Biológicas. Pós-graduação em Ciências Biológicas, 2014.

Inclui referências

1. Leguminosa 2. Citogenética I. Benko-Iseppon, Ana Maria II. Vidal,
Ana Christina Brasileiro (coorient.) III. Título.

583.74

CDD (22.ed.)

UFPE/CCB-2014-82

Pollyana Karla da Silva

**Distribuição e Prevalência de Elementos Não Codificantes em
*Glycine Willd. e Vigna Savi***

Comissão Julgadora da Tese para obtenção do grau de Doutor em
Ciências Biológicas

Prof^a. Dr^a. Ana Maria Benko- Iseppon

Departamento de Genética – Universidade Federal de Pernambuco
Presidente da Banca

Prof. Dr. Reginaldo de Carvalho

UFRPE - Universidade Federal Rural de Pernambuco

Dr^a. Valesca Pandolfi

Departamento de Genética – Universidade Federal de Pernambuco

Dr^a. Kyria Cilene de Andrade Bortoleti

UNIVASF – Universidade Federal Vale do São Francisco

Dr^a. Maria Betânia Melo de Oliveira

Departamento de Bioquímica – Universidade Federal Rural de Pernambuco

Recife, 30 de Junho, 2014

Dedicatória

***Dedico esse trabalho à minha mãe,
Maria Gorete Ribeiro,
pelo amor infinito e dedicação integral.***

Agradecimentos

A DEUS pela vida e por tudo que me ofereceu em cada momento dessa caminhada.

Aos meus pais pelo amor e por apoio que sem eles não conseguiria conquistar mais uma etapa importante em minha vida; em especial à minha MÃE por estar sempre comigo, principalmente nos momentos mais difíceis. Ela é a minha força, meu porto seguro e por ela cheguei até aqui... Mainha, eu te amo!

Aos meus irmãos e sobrinhos pelo amor, carinho, respeito e compreensão quando mais precisei.

À Professora Ana Benko por tudo!!! Por ter aceitado aquela que “caiu do céu” em sua equipe, pelo grande aprendizado a mim concedido, pela ajuda sempre necessária, pela paciência e pela confiança em mim depositada.

À Professora Ana Christina por tudo que fez por mim durante esse tempo. Ensinou-me, orientou-me, compreendeu-me, foi uma amiga e psicóloga, mas, sobretudo, esteve sempre comigo durante essa longa caminhada, até mesmo quando eu desanimava. Obrigada Professora por não desistir de mim.

Aos meus Padrinhos, Socorro e Welington por serem o que realmente lhe foram designados, segundos pais. Obrigada por tudo que sempre fizeram por mim durante toda minha trajetória.

Aos amigos do LGBV, Artur Wanderley, Bruna Piereck, Diego Sotero, Ebenézer Bernardes, Flávia Araújo, Geyner Alves, Igor Gominho, Hévila Mendes, João Pacífico, Karla Camila, Kyria Bortoleti, Lidiane Amorim, Marília Gabriela, Marx Oliveira, Mitalle Matos, Neto, Rafaela, Rodrigo Oliveira, Rosilda Cintra, Santelmo, Sheyla Lima, Silvany Araújo, Stephani Soares e Valesca Pandolfi por todos os momentos vividos, principalmente pelas boas e numerosas risadas compartilhadas. Momentos únicos que levarei comigo sempre.

Em especial, quero agradecer à **Vanessa** por ter sido o anjo que Deus colocou no meu caminho e esteve comigo em todos os momentos, fossem eles tristes ou alegres. À **Manuzinha** pela amizade, cumplicidade e ensinamentos na citogenética. E à **Anne** por ser esse ser de luz e paz que me acalmava a cada abraço diário.

À **Bruna** por ter aceitado a parceria e acima de tudo, o desafio. Obrigada por todo esforço e dedicação ao longo desses últimos dias. Como também a **João Pacífico** e **Sheyla** pela ajuda que só amigos que têm um bom coração podem fazer.

Aos professores e funcionários da Pós-Graduação que fizeram parte desse período de aprendizagem e de muito trabalho.

Meus agradecimentos também aos membros da banca de qualificação, Prof. Reginaldo Carvalho e Dra. Valesca Pandolfi, pela leitura cuidadosa e preciosas sugestões.

Um especial agradecimento à **FACEPE** (Fundação de Amparo à Ciência e Tecnologia de Pernambuco) pela concessão da bolsa de doutoramento, imprescindível para minha carreira profissional e ao **CNPq** (Conselho Nacional de Desenvolvimento Científico e Tecnológico) pelo precioso apoio financeiro no âmbito dos consórcios NordEST (feijão-caupi) e GENOSOJA.

Enfim, a todos que direta ou indiretamente contribuíram para a realização desse trabalho.

Muito obrigada!

Sumário

	Página
Lista de Abreviaturas	vii
Lista de Figuras.....	ix
Lista de Tabelas.....	xi
Lista de Quadros	xi
Resumo.....	xii
Abstract	xiii
1. Introdução	1
2. Objetivos	3
3. Revisão de Literatura	4
3.1 Família Fabaceae.....	4
3.2 Gênero <i>Glycine</i> Willd.....	6
3.2.1 Importância Econômica.....	8
3.2.2 Pools Gênicos.....	9
3.3 Gênero <i>Vigna</i> Savi.....	10
3.3.1 Importância Econômica.....	11
3.4 Citogenética Molecular.....	12
3.5 Microssatélites.....	14
3.6 Elementos transponíveis e retroelementos.....	16
4. Referências Bibliográficas.....	20
5. Capítulo I - Distribuição cromossômica de microssatélites em <i>Glycine soja</i> Sieb. & Zucc e <i>G. tomentella</i> Hayata	29
6. Capítulo II - Distribuição do retrotransposon <i>Ty1-copia</i> em espécies de <i>Vigna</i> Savi.....	47
7. Capítulo III - Identificação e caracterização de domínios da superfamília CACTA em <i>Glycine max</i>	61
8. Conclusões.....	80

Lista de Abreviaturas

(AAC) ₅	Sequência de oligonucleotídeos: Adenina, Adenina e Citosina, repetidos cinco vezes
(AAG) ₅	Sequência de oligonucleotídeos: Adenina, Adenina e Guanina repetidos cinco vezes
(ACC) ₅	Sequência de oligonucleotídeos: Adenina, Citosina e Citosina repetidos cinco vezes
(AG) ₈	Sequência de oligonucleotídeos: Adenina e Guanina repetidos oito vezes
(CTC) ₅	Sequência de oligonucleotídeos: Citosina, Timina e Citosina repetidos cinco vezes
(TGA) ₆	Sequência de oligonucleotídeos: Timina. Guanina e Adenina repetidos cinco vezes
A/T	Adenina/Timina
8-HQ	8-hidroxiquinoleína
BSA	<i>Bovine Serum Albumine</i> ; Albumina de Soro Bovino
CCB	Centro de Ciências Biológicas
CMA ₃	Cromomicina A ₃
DAPI	4',6-diamidino-2-fenilindol
DNA	<i>Desoxiribonucleic Acid</i> ; Ácido Desoxirribonucleico
Domínio RT	Domínio da Transcriptase Reversa
FISH	<i>Fluorescent In Situ Hybridization</i> ; Híbridização <i>In Situ</i> Fluorescente
FITC	Fluoresceína isotiocianato
G/C	Guanina/Citosina
GL	Grupo de Ligação
Kb	Kilo bases

LINE	<i>Long Interspersed Nucleotide Elements</i> ; Elementos Nucleotídicos Interdispersos Longos
LGBV	Laboratório de Genética e Biotecnologia Vegetal
LTRs	<i>Long Terminal Repeats</i> ; Longas Repetições Terminais
Mb	Mega bases
Pb	Pares de bases
Pg	Picogramas
ORFs	<i>Open Reading Frames</i>
RNA	Ribonucleic Acid; Ácido Ribonucleico
RNAr	RNA ribossomal
RONs	Regiões Organizadoras de Nucléolos
RT	<i>Reverse Transcriptase</i> ; Transcriptase Reversa
SINEs	<i>Short Interspersed Nucleotide Elements</i> ; Elementos Nucleotídicos Interdispersos Curtos
SNPs	<i>Single Nucleotide Polymorphism</i> ; Polimorfismos de um Único Nucleotídeo
sp	Espécie
ssp	Subespécie
SSC	Citrato de Sódio-Salino
SSRs	<i>Simple Sequence Repeats</i> ; Repetições de Sequências Simples
TEs	<i>Transposable Elements</i> ; Elementos Transponíveis
TIRs	<i>Terminal Inverted Repeats</i>
TRITC	Tetramethyl Rhodamine Iso-Thiocyanate
Ty1- <i>copia</i>	Superfamília de retrotransposon LTR
Ty3- <i>gypsy</i>	Superfamília de retrotransposon LTR

Lista de Figuras

Figura	Página
Revisão de Literatura	
Figura 1. Modelo de um cromossomo de planta com as diferentes classes de DNA repetitivo mostrando a distribuição genômica característica. Fonte: Schmidt e Heslop-Harrison (1998)	14
Figura 2. Organização dos domínios gênicos das principais classes de retrotransposons LTRs (Ty1- <i>copia</i> e Ty3- <i>gypsy</i>) [adaptado de Todorovska (2007) <i>apud</i> Bortoleti, 2010].	18
Capítulo I	
Figura 1. Coloração CMA/DAPI e hibridização <i>in situ</i> fluorescente dos oligonucleotídeos (AAC) ₅ , (AAG) ₅ , (ACC) ₅ , (AG) ₈ (CTC) ₅ , (TGA) ₆ , em cromossomos mitóticos de <i>Glycine soja</i> .	39
Figura 2. Coloração CMA/DAPI e hibridização <i>in situ</i> fluorescente dos oligonucleotídeos (AAC) ₅ , (AAG) ₅ , (ACC) ₅ , (AG) ₈ (CTC) ₅ , (TGA) ₆ , em cromossomos mitóticos de <i>Glycine tomentella</i> .	40
Capítulo II	
Figura 1. Coloração CMA/DAPI e hibridização <i>in situ</i> fluorescente do retrotransposon Ty1- <i>copia</i> -like, em cromossomos mitóticos de <i>Vigna umbellata</i> , <i>V. sesquipedalis</i> e <i>V. aconitifolia</i> .	54

Capítulo III

Figura 1. Imagem esquemática do elemento CACTA. 64

Figura 2. Fluxograma ilustrando as principais etapas metodológicas realizadas. 66

Figura 3. Gráfico representativo da quantidade de domínios relacionados à transposase identificados em CACTA. 68

Figura 4. Gráfico representativo do número absoluto das cinco principais categorias de domínios encontradas no genoma da soja. 71

Lista de Tabelas

Tabela	Página
Revisão de Literatura	
Tabela 1. Espécies citogeneticamente analisadas do gênero <i>Glycine</i> e seus respectivos números diploides. Fonte: adaptado de Bilyeu et al. (2010)	7
Capítulo I	
Tabela 1. Padrões prevalentes quanto à posição e distribuição de marcações cromossômicas observadas no presente estudo	38
Capítulo III	
Tabela 1. Domínios de transposases identificados na soja, incluindo domínios completos (-), domínio incompleto na região C-terminal (C), domínio incompleto na região N-terminal (N) e domínios incompletos nas duas regiões (NC).	69

Lista de Quadros

Quadro	Página
Capítulo III	
Quadro suplementar. Descrição e caracterização das 65 observadas no SoyTEdb (Du et al., 2010) e utilizadas na busca por homologia no presente trabalho.	75

Resumo

A genômica comparativa tornou-se uma importante área de pesquisa nos últimos anos, após a disponibilização de um número de genomas total ou parcialmente sequenciados, permitindo uma visão detalhada da organização de regiões gênicas ou não codificantes. A comparação dos genomas é de grande importância para a análise das regiões funcionalmente relevantes, permitindo também inferências sobre a evolução e os mecanismos que conduzem ao rearranjo de cariótipos, havendo importantes implicações práticas, principalmente quando são comparadas espécies cultivadas e seus parentes silvestres, potencialmente doadores de genes para o melhoramento. Este trabalho teve o objetivo de realizar um estudo citogenético e comparativo de espécies dos gêneros *Glycine* e *Vigna*. Para isso foram utilizadas três abordagens: Uma análise citogenômica comparativa mediante localização *in situ* de oligonucleotídeos com padrão de microssatélites [(AAC)₅, (AAG)₅, (ACC)₅, (AG)₈ (CTC)₅, (TGA)₆] ao longo dos cromossomos de *Glycine soja* e *G. tomentella*. Nesta etapa, foram observadas regiões de marcação mais intensa em alguns cromossomos, embora a maioria dos oligonucleotídeos tenha apresentado distribuição dispersa nos genomas analisados. Uma segunda abordagem avaliou a distribuição de domínios RT do retrotransposons Ty1-copia-like, nos cromossomos de *Vigna umbellata*, *V. sesquipedalis* e *V. aconitifolia*, apresentando uma distribuição dispersa e sinais proximais destes elementos nem determinados cromossomos. Em uma terceira abordagem, foi realizada uma investigação do elemento transponível CACTA no genoma da soja (*Glycine max*), mediante análise *in silico*. Neste estudo, foram identificados domínios transposase do elemento citado, compreendendo 10 Mb de Tnp1, 2,2 Mb de Tnp2, bem como domínios de outras transposases não relacionados diretamente ao elemento CACTA, mostrando que a diversidade e abrangência destes elementos é maior do que reportado até então. Considerando-se que microssatélites e retrotransposons são importantes componentes dos genomas em espécies da tribo Phaseoleae, os resultados aqui obtidos trazem interessantes evidências sobre o papel estrutural e funcional dessas sequências repetitivas.

Palavras-chave: *Glycine*, *Vigna*, elementos transponíveis, mapas cromossômicos comparativos, microssatélites.

Abstract

Comparative genomics has become an important research area in recent years, after the release of a number of fully or partially sequenced genomes, allowing a detailed view of the organization of gene or non-coding regions. The comparison of genomes is of great importance for the analysis of functionally relevant regions, also allowing inferences on the evolution of the mechanisms leading to the rearrangements within karyotypes, which present important practical implications, especially when cultivated species and their wild relatives are compared. This study aimed to conduct a cytogenetic and comparative study with species of the genera *Glycine* and *Vigna*. Accordingly, three approaches were used: A comparative cytogenomics analysis by means of in situ localization of oligonucleotides with microsatellite pattern [(AAC)₅, (AAG)₅, (ACC)₅, (AG)₈, (CTC)₅, (TGA)₆] along the chromosomes of *Glycine soja* and *G. tomentella*. At this stage of the study, regions of more intense staining were observed in some chromosomes, although most primers have presented scattered distribution in the analyzed genomes. A second approach evaluated the distribution of RT domains retrotransposons Ty1-copia-like, in the chromosomes of *Vigna umbellata*, *V. sesquipedalis* and *V. aconitifolia*, with a scattered distribution and proximal signals of these elements or certain chromosomes. In a third approach, the transposable element CACTA was searched throughout the genome of soybean (*Glycine max*) by means of in silico analysis. In this study, the transposase domains were identified comprising the TNP1 with 10 Mb and 2.2 Mb of the Tnp2, as well as other fields not directly related to the transposase of the CACTA element, showing that the diversity and extent of these elements is higher than reported hitherto. Given that microsatellites and retrotransposons are important components in the genomes of species of the Phaseoleae tribe, our findings bring interesting evidence about the structural and functional role of these repetitive sequences.

Keywords: *Glycine*, *Vigna*, transposable elements, comparative chromosome maps, microsatellites.

1. Introdução

A família Fabaceae, também conhecida como Leguminosae, compreende a segunda maior família de angiospermas (superada apenas pelas Orchidaceae), incluindo 730 gêneros e cerca de 19.400 espécies, apresentando uma extensa gama de hábitos, incluindo árvores, arbustos, lianas e ervas (APG III, 2009). Apresenta grande diversidade taxonômica, sendo subdividida em três subfamílias: Mimosoideae, Caesalpinoideae e Papilinoideae. Economicamente, a subfamília Papilinoideae abrange os legumes cultivados, como soja, ervilha, grão-de-bico, feijão comum, feijão-caupi, amendoim, ervilha e alfafa, representando cerca de 27% da produção de plantas cultivadas do mundo (Graham e Vance, 2003). Dentre estas podemos destacar a importância da soja e dos feijões para a economia nacional.

O gênero *Glycine* compreende dois subgêneros: *Glycine* Willd. (incluindo 26 espécies perenes) e *Soja* (Moench) F.J. Hermann, o mais importante economicamente, incluindo a soja cultivada [*G. max* (L.) Merr.] e seu suposto ancestral selvagem *G. soja* Sieb. & Zucc. (Clarindo *et al.*, 2007; Bilyeu *et al.*, 2010). A cultura da soja, *G. max* é a mais importante no planeta para a produção de óleo, usado principalmente como alimento humano, bem como na produção de biodiesel. Seus grãos são importantes para a fabricação da farinha de soja, rica em proteínas, que é utilizada principalmente na alimentação animal (Bilyeu *et al.*, 2010), além de outros produtos como o leite de soja, do tofu e de produtos fermentados como o molho de soja e a pasta de soja, entre outros (NSRL, 2012).

O gênero *Vigna* é composto por mais de 170 espécies, distribuídas principalmente nas áreas tropicais. É taxonomicamente dividido em sete subgêneros, dos quais apenas três (*Vigna*, *Ceratotropis* e *Plectotropis*) incluem espécies cultivadas (Choi *et al.*, 2013), as quais representam uma importante fonte proteica na dieta humana e animal de países em desenvolvimento de regiões tropicais e subtropicais. No Brasil, o feijão-caupi [*V. unguiculata* (L.) Walp.] é um dos componentes básicos da dieta alimentar da população, destacando-se como importante fonte de proteína para as classes economicamente menos favorecidas (Wander *et al.*, 2005).

O genoma dos eucariotos, incluindo o reino vegetal, é formado por uma grande quantidade de DNA não codificante, com ênfase para elementos transponíveis e microssatélites. Os elementos transponíveis (TEs) incluem

transposons e retrotransposons, sendo assim designados por sua capacidade de mudar sua posição no genoma, por vezes criando ou revertendo mutações e, desta forma, alterando o tamanho genômico. Apesar de seu caráter não codificante, tratam-se dos elementos mais abundantes, com inquestionável importância evolutiva (Wicker *et al.*, 2007)

Os retrotransposons são caracterizados por moverem-se via sequências de RNA. Com auxílio de uma transcriptase reversa e de uma transposase inserem uma cópia em outra região do genoma. Já os transposons baseiam-se na movimentação direta das sequências de DNA que, com ação de uma ou mais transposases (TNPs), são retirados e inseridos em outras regiões (Jiang *et al.*, 2004; Wicker *et al.*, 2007; Li *et al.*, 2009).

Microsatélites, também conhecidos como SSRs (do inglês *Simple Sequence Repeats*; Repetições de Sequências Simples) compreendem sequências repetitivas, presentes em qualquer região do genoma, codificantes ou não (Bhargava e Fuentes, 2010; Tóth *et al.*, 2000). Podem ser utilizados como marcadores genéticos, possuindo como atributos específicos a presença de repetições curtas (de dois a seis pares de bases) e maior susceptibilidade ao aparecimento de mutações, comparativamente ao restante do genoma (Oliveira *et al.*, 2006). Em vista dessa alta mutabilidade, acredita-se que os microsatélites podem desempenhar um papel significativo na evolução do genomas, pela criação e manutenção da variação da genética quantitativa (Tóth *et al.*, 2000).

A citogenética tem desempenhado um papel essencial em estudos da estrutura, comportamento e evolução cromossômica de inúmeras espécies de plantas. O advento da citogenética molecular, incluindo a hibridização *in situ* fluorescente, combinada com recente desenvolvimento de recursos genômicos, permitiu avanços no conhecimento da diversidade de cariótipos, genomas, de sua organização cromossômica e da evolução em plantas, incluindo leguminosas (Heslop-Harrison e Schwarzacher, 2011; Iwata *et al.*, 2013).

Dessa forma, uma análise citogenética e genômica comparativa da prevalência e distribuição de elementos repetitivos em membros selecionados de *Glycine* e *Vigna* deve auxiliar no entendimento da evolução cariotípica desses grupos.

2. Objetivos:

2.1 Objetivo Geral

- Realizar um estudo comparativo em nível citogenético e genômico com espécies selecionadas dos gêneros *Glycine* e *Vigna* mediante a localização cromossômica ou diversidade genômica de sequências de DNA repetitivo, colaborando para o entendimento da organização e evolução genômica desses elementos repetitivos nos grupos em questão.

2.3 Objetivos Específicos

- Elucidar a prevalência e distribuição de seis sequências de microssatélites em *G. soja* e *G. tomentella*, discutindo comparando os padrões observados entre as espécies citadas.
- Identificar a distribuição cromossômica do retroelemento Ty1-*copia*-like em espécies selecionadas do gênero *Vigna*, discutindo os padrões observados e sua evolução.
- Localizar e identificar o padrão de bandas CMA₃/DAPI e associar aos sinais de hibridização dos microssatélites e retrotransposons como marcas cromossômicas.
- Identificar através de ferramentas *in silico* elementos da Superfamília CACTA no genoma da soja (*G. max*), inferindo sobre sua abundância e diversidade.
- Contribuir para o entendimento da macroevolução de elementos não codificantes nos táxons analisados.

3. Revisão de Literatura

3.1 Família Fabaceae

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 730 gêneros e aproximadamente 19.400 espécies (Lewis *et al.*, 2005). Seus representantes estão distribuídos nas regiões tropicais e subtropicais, ocupando o terceiro lugar em diversidade (Graham e Vance, 2003, Souza e Benko-Iseppon, 2004).

Compreende três subfamílias: Mimosoideae, que reúne cerca de 2.500 espécies; Caesalpinioideae, com cerca de 2.700, e Papilionoideae (ou Faboideae), com 12.615 espécies (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).

A família Fabaceae é classificada como a segunda maior família vegetal de importância econômica, no que tange à sua grande utilização para os mais diversos fins, perdendo apenas para a Poaceae (Singh *et al.*, 2007). Três gêneros que se destacam são *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). Estes gêneros se destacam pela sua importância na dieta alimentar e apresentam como espécies mais importantes *G. max* (L.) Merr. (soja), *P. vulgaris* L. (feijão comum), *P. lunatus* L. (feijão fava) e *V. unguiculata* (L.) Walp. (feijão-caupi) (Barroso *et al.*, 1991; Souza e Lorenzi, 2005).

Dentre as plantas cultivadas, de importância econômica e utilizadas pela população, o feijão-caupi destaca-se por possuir várias características vantajosas, sendo considerado o feijão mais adaptado às condições climáticas do Nordeste do Brasil (Freire-Filho *et al.*, 1999; 2005; 2011a). Constitui-se em importante fonte de proteínas para os agricultores de baixo poder aquisitivo,

contribuindo com 31% da proteína consumida no nordeste, com 24% de proteínas e 340 calorias para cada 100 g de sementes. Adicionalmente, esta cultura apresenta aminoácidos essenciais ausentes nas demais leguminosas, elevados níveis de ácido fólico, ferro, potássio, baixos níveis de fatores antinutricionais e flatulência, além de ter digestibilidade muito superior aos demais feijões (Ehlers e Hall, 1997; Silva e Freire-Filho, 1999; Maia *et al.*, 2000). 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).

Os feijões apresentam uma importante fonte proteica na dieta humana de países em desenvolvimento das regiões tropicais e subtropicais. No Brasil, é um dos componentes básicos da dieta alimentar da população, destacando-se como importante fonte de proteína para as classes economicamente menos favorecidas (Wander *et al.*, 2005). Por sua vez, a soja, além da importância proteica, é uma das principais fontes de óleo vegetal do mundo sendo 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).

Outras espécies da família Fabaceae são bastante utilizadas 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). Algumas espécies 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 explorada no primeiro ciclo econômico do Brasil. Outra importante aplicação é o uso na adubação verde, principalmente considerando a associação com bactérias fixadoras de nitrogênio (*Rhizobium* spp.) que produzem nódulos nas raízes de espécies como o 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).

3.2 Gênero *Glycine* Willd.

O gênero *Glycine* Willd. é membro da família Fabaceae/Leguminosae, Subfamília Papilionoideae e da tribo Phaseoleae, sendo subdividido em dois subgêneros, *Glycine* (perenes) e *Soja* (Moench) F.J.Herm, ambas espécies anuais (Bilyeu, *et al.*, 2010).

O subgênero *Glycine* é composto por 26 espécies nativas da Austrália e de ilhas do oceano Pacífico. Dessas, 21 apresentam cariótipo com $2n = 40$ cromossomos; duas espécies mostram $2n = 80$ (*G. dolichocarpa* Tateishi e Ohashi e *G. pescadrensis* Hayata); duas possuem $2n = 40$ ou 80 (*G. hirticaulis* Tind. & Craven e *G. tabacina*), enquanto uma espécie (*G. tomentella*) apresenta quatro citotipos diferenciados; diploide, tetraploide e dois aneuploides: $2n = 40$, 80 , 38 e 78 , respectivamente (Tabela 1). De um modo geral, as espécies do subgênero *Glycine* são distintas em sua morfologia, citologia e características genômicas. Além de crescerem em climas muito diferentes e de terem uma distribuição geográfica bastante ampla, tais espécies nunca foram domesticadas e permanecem como espécies perenes e selvagens (Chung e Singh, 2008).

No subgênero *Soja* estão incluídas apenas duas espécies: uma muito importante economicamente, a soja *G. max* com $2n = 40$ e a espécie selvagem, *G. soja*, com $2n = 40$, ambas tetraploides e anuais. Os genomas de *G. soja* e *G. max* são similares, sendo designados por G e G1, respectivamente (Singh *et al.*, 2007). A soja selvagem (*Glycine soja* Sieb. et Zucc.) tem como características uma alta reprodutividade, elevado número de vagens, alto teor de proteínas, capacidade de adaptação a diversos estresses, sendo reconhecida como o progenitor da soja cultivada [*Glycine max* (L.) Merr.]. O potencial da *G. soja* como fonte de diversidade genética para o melhoramento da soja cultivada é inquestionável, especialmente para a ampliação da base genética da soja (Wang *et al.*, 2013).

Tabela 1: Espécies citogeneticamente analisadas do gênero *Glycine* e seus respectivos números diploides. Fonte: adaptado de Bilyeu *et al.* (2010).

Subgênero/Espécie	Número Cromossômico (2n)
<i>Glycine</i> Willd.	
<i>G. albicans</i> Tindale and Craven	40
<i>G. aphyonota</i> B. Pfeil	40
<i>G. arenaria</i> Tindale	40
<i>G. argyrea</i> Tindale	40
<i>G. canescens</i> F. J. Hermann	40
<i>G. clandestina</i> Wendl.	40
<i>G. curvata</i> Tindale	40
<i>G. cyrotoloba</i> Tindale	40
<i>G. falcata</i> Benth.	40
<i>G. gracei</i> B.E. Pfeil and Craven	40
<i>G. latrobeana</i> (Meissn.) Benth.	40
<i>G. lactovirens</i> Tindale and Craven	40
<i>G. latifolia</i> (Benth.) Newell and Hymowitz	40
<i>G. microphylla</i> (Benth.) Tindale	40
<i>G. montis-douglas</i> B.E.Pfeil and Craven	40
<i>G. peratosa</i> B. E. Pfeil and Tindale	40
<i>G. pindanica</i> Tindale and Craven	40
<i>G. pullenii</i> B. Pfeil, Tindale and Craven	40
<i>G. rubiginosa</i> Tindale and B. E. Pfeil	40
<i>G. stenophita</i> B. Pfeil and Tindale	40
<i>G. syndetika</i> B.E. Pfeil and Craven	40
<i>G. pescadrensis</i> Hayata	80
<i>G. dolichocarpa</i> Tateishi and Ohashi	80
<i>G. hirticaulis</i> Tindale and Craven	40, 80
<i>G. tabacina</i> (Labill.) Benth	40, 80
<i>G. tomentella</i> Hayata	38, 40, 78, 80
<i>Soja</i> (Moench) F.J. Hermann	
<i>G. soja</i> Sieb. & Zucc.	40
<i>G. max</i> (L.) Merr.	40

A hipótese mais aceita sobre a origem geográfica do gênero *Glycine* foi proposta por Hymowitz (2004-2008) que, com base em publicações taxonômicas, citológicas e de sistemática molecular sobre o gênero *Glycine* e gêneros afins, propôs que um suposto ancestral diploide ($2n = 2x = 20$) do atual gênero *Glycine* teve origem no sudeste da Ásia. Esse ancestral teria migrado para o norte da China originando uma espécie perene selvagem ($2n = 4x = 40$,

desconhecida ou extinta) com subsequente evolução para a selvagem anual ($2n = 4x = 40$; *G. soja*) e finalmente para a soja cultivada ($2n = 4x = 40$; *G. max*) (Singh *et al.*, 2001; Bilyeu *et al.*, 2010).

3.2.1 Importância econômica

A soja *G. max* é uma leguminosa economicamente importante, cujas sementes contêm em média 40% de proteína e 20% de óleo. Suas plantas enriquecem o solo através de simbiose com bactérias fixadoras de nitrogênio. A produção de soja no mundo dobrou nos últimos 20 anos, sendo os EUA, Brasil, Argentina, China e Índia os maiores produtores (Chung e Singh, 2008). A soja hoje é responsável por 56,7% da produção mundial de óleo vegetal (USDA, 2013). Além disso, segundo a *Global Industry Analysts, Inc* (2013), tal cultura alcançará a movimentação anual de US\$ 43,2 bilhões de dólares, no ano de 2015, somente considerando derivados alimentícios da soja.

A produção de soja tem uma grande demanda e continua a crescer devido à sua utilização na fabricação de produtos alimentícios e industrializados (Wilson, 2008), sendo amplamente utilizada na alimentação humana e animal por ser fonte de proteínas de alta qualidade, aminoácidos essenciais, vitaminas, ácidos graxos essenciais (como o ômega 3), fitoquímicos e fármacos, além de auxiliar na redução do colesterol e no combate à diabetes, obesidade e certos tipos de câncer (Singh *et al.*, 2007). Tem sido também utilizada na fabricação de corantes e cosméticos, além da obtenção de biodiesel, tornando-se uma alternativa renovável e menos poluidora ao uso de combustíveis fósseis (Singh *et al.*, 2007).

No Brasil, a soja é a cultura que mais cresceu nas últimas três décadas e corresponde a 49% da área plantada em grãos (MAPA, 2013a). O complexo soja (grão, farelo e óleo) é o principal gerador de divisas cambiais, com negociações anuais que ultrapassam US\$ 20 bilhões (MAPA, 2013b). As previsões indicam que em 2019, a produção nacional deve representar 40% do comércio mundial do grão e 73% do óleo de soja (MAPA, 2013b). Nosso país é o segundo maior produtor de soja, superado apenas pelos Estados Unidos da

América do Norte, exportando cerca de 34,8 milhões de toneladas métricas anuais, seguido pela Argentina e pelo Paraguai (Soystats, 2013).

Com o *status* de principal cultura vegetal produzida no país e com índices elevados de crescimento previsto para os próximos anos, a soja é uma das leguminosas de maior interesse para a realização de pesquisas. Com isso, consideráveis investimentos foram e continuam sendo feitos por diversos países com o intuito de obter informações, incluindo o sequenciamento de todo o genoma para aplicação direta no melhoramento dessa espécie (Schmutz *et al.*, 2010).

3.2.2 Pools Gênicos

O conceito de pools gênicos primário (GP-1), secundário (GP-2) e terciário (GP-3) foi proposto por Harlan e de-Wet em 1971 e tem sido aplicado ao gênero *Glycine*, baseados no sucesso de hibridação entre e dentre as espécies, ou seja, na facilidade de troca de genes entre elas (Harlan e deWet, 1971; Borém e Vieira, 2005; Bilyeu *et al.*, 2010). O pool gênico primário (GP-1) de soja inclui todos os acessos de soja cultivada (*G. max*) e da espécie selvagem *G. soja*. O GP-1 consiste de espécies biológicas que, quando cruzadas, são capazes de produzir híbridos vigorosos, que exibem pareamento cromossômico meiótico normal e total fertilidade de sementes. A segregação dos genes na F1 é normal e a troca de genes é geralmente factível (Singh 2003; Bilyeu, *et al.*, 2010).

O *pool* gênico secundário inclui espécies pertencentes ao mesmo gênero, as quais podem ser cruzadas com GP-1, porém com dificuldade e com fertilidade apenas parcial em F1. Para soja, não há espécie conhecida em GP-2 (Chung e Singh, 2008).

Por sua vez, o GP-3 inclui cruzamentos entre espécies pouco aparentadas de outros gêneros ou espécies pouco relacionadas do mesmo gênero. Em geral, a hibridação é muito difícil e a esterilidade e/ou anomalia do híbrido são comuns, com necessidade de resgate de embriões e enxertia, entre outras medidas (Harlan e de Wet, 1971). Aparentemente, apenas *G. argyrea*, *G. canescens* e *G. tomentella* estão incluídas em GP-3. A produção de híbridos F1 dessas espécies com soja é possível mediante cultura de embriões, embora

sejam estéreis (Singh *et al.*, 1998; Chung e Singh, 2008). Somente para cruzamentos entre *G. max* e *G. tomentella* ($2n = 78$) foi desenvolvida uma metodologia para a produção de linhagens férteis derivadas, permitindo a introgressão transversal de genes úteis a partir de GP-3 (Singh e Chung, 2007a).

O *pool* gênico quaternário (GP-4) é formado por espécies que apresentam barreiras pré e pós-hibridização que inibem o desenvolvimento do embrião e geram o aborto prematuro. Raramente a letalidade das plântulas híbridas e a inviabilidade de sementes híbridas e de plantas F1 podem ser contornadas por cruzamentos-ponte dentro do gênero *Glycine* (Singh *et al.*, 2007a), observando-se que a maioria das 26 espécies do subgênero *Glycine* estão incluídas em GP-4. Entretanto, apenas algumas espécies de *Glycine* silvestres perenes foram hibridizadas com soja, como por exemplo, *G. clandestina*, *G. latifolia* e *G. tabacina* (Chung e Singh 2008).

3.3 Gênero *Vigna* Savi.

O gênero *Vigna* compreende cerca de 170 espécies, distribuídas nas áreas tropicais, sendo classificado como pertencente à ordem Fabales, família Fabaceae, subfamília Papilionioideae (Faboideae), tribo Phaseoleae e subtribo Phaseolina (Freire-Filho *et al.*, 2005; Choi *et al.*, 2013).

Usando taxonomia numérica, Marechal *et al.*, (1978) realizaram um amplo estudo neste gênero dividindo-o em sete subgêneros: *Vigna*, *Plectotropis*, *Ceratotropis*, *Lasiocarpa*, *Sigmoidotropis*, *Haydonia* e *Macrorhyca*. Destes, apenas três subgêneros (*Vigna*, *Ceratotropis* e *Plectotropis*) incluem plantas cultivadas (Verdcourt, 1970). O subgênero *Vigna* foi considerado endêmico da África e dividido em seis seções: *Comosae*, *Liebrehtsia*, *Macrodonatae*, *Reticulatae*, *Vigna* e *Catiang*, sendo a última formada por duas espécies botânicas reconhecidas, *V. unguiculata* e *V. nervosa* Markotter.

A espécie *V. unguiculata* foi subdividida em quatro subespécies: três subespécies selvagens, *V. dekindtiana* (Harms) Verd., *V. tenuis* (E. Mey.) M. M. & S. e *V. stenophylla* (Harvey) M. M. & S., e uma subespécie cultivada, *V. 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. Dessa forma, Westphal (1974) dividiram a subespécie *V. unguiculata* em quatro cultigrupos: *Unguiculata*, *Sesquipedalis*, *Biflora* e *Textilis* (Marechal *et al.*, 1978; Ng e Marechal 1985). Dentre esses cultigrupos, apenas *Unguiculata* e *Sesquipedalis* são cultivados no Brasil, compreendendo a quase totalidade das cultivares locais e melhoradas e o feijão-de-metro, respectivamente (Freire-Filho *et al.*, 2005).

Citogeneticamente, o gênero *Vigna* é diploide com $2n = 2x = 22$ (Sen e Bhowal 1960), apresentando cromossomos relativamente pequenos (1,6 - 3,7 μm) (Guerra *et al.*, 1996) e alta homogeneidade cariomorfológica (Venora e Saccardo, 1993; Venora e Padulosi, 1997; Venora *et al.*, 1999), características que dificultam as análises citológicas do grupo (Choi *et al.*, 2013).

O feijão-caupi apresenta um dos menores genomas entre as leguminosas (450 - 500 Mb) (Arumuganathan e Earle, 1991), apresentando cromossomos predominantemente submetacêntricos e metacêntricos, de difícil identificação devido ao tamanho e grau de condensação cromossômica (Bortoleti *et al.*, 2012).

3.3.1 Importância econômica

O gênero *Vigna* Savi, representado pelos grãos de feijão, representa 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 *et al.*, 2005).

Vigna unguiculata (L.) Walp. é uma cultura multifuncional e um dos alimentos mais importantes dentre as leguminosas forrageiras dos trópicos semiáridos (Timko e Singh, 2008). O Brasil é considerado o terceiro maior produtor de feijão-caupi, com destaque para as regiões Nordeste e Norte como maiores produtoras (EMBRAPA, 2011). A maior produção concentra-se na Região Nordeste, com 84% da área plantada e 68% da produção nacional, o

que gerou, nos últimos cinco anos, 1,2 milhão de empregos diretos (Castelletti e Costa, 2013). Nessa região a produção concentra-se nas áreas semiáridas, onde outras culturas leguminosas anuais, em razão da irregularidade das chuvas e das altas temperaturas, não se desenvolvem (Freire-Filho *et al.*, 2011a).

Feijões constituem a principal cultura de subsistência para a população mais carente, por seu alto valor nutritivo, devido ao alto conteúdo de proteína das sementes e pela geração de emprego e renda na zona rural (Frota *et al.*, 2008; Timko e Singh, 2008), podendo ser consumidos na forma de folhas jovens, brotos, vagens verdes e grãos (Broughton *et al.*, 2003) ou usados como adubo verde ou forrageira (Guedes, 2008).

Anualmente, grande parte da produção de feijão é comprometida pela ocorrência de diversos tipos de doenças e pragas, sendo necessária a utilização de defensivos agrícolas. A falta de recursos dos produtores, juntamente com os efeitos tóxicos causados pelos pesticidas, tem levado a comunidade científica a realizar projetos visando obter espécies mais tolerantes a diferentes tipos de estresse bióticos e/ou abióticos (Broughton *et al.*, 2003; Timko e Singh, 2008).

3.4 Citogenética Molecular

Estudos citogenéticos têm revelado polimorfismos genéticos importantes na caracterização de cultivares, através de marcadores citológicos que auxiliam na construção de mapas cromossômicos. Tais estudos têm contribuído com os programas de melhoramento genético, auxiliando na caracterização de germoplasma, no entendimento das relações filogenéticas e na identificação da origem de várias espécies (Benko-Iseppon, 2001; Carvalho e Guerra, 2002; Carvalho *et al.*, 2005; Pedrosa-Harand *et al.*, 2006).

Métodos citomoleculares fornecem poderosas ferramentas para análises da estrutura, da organização e da evolução genômica. A técnica de hibridização *in situ* fluorescente (FISH) tem sido aplicada na caracterização de espécies vegetais, localizando 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), permitindo a localização *in situ* de sequências de ácidos nucleicos (DNA ou RNA) nos cromossomos, citoplasma, organelas ou tecidos estudados (Guerra, 2004).

A hibridização *in situ* de sequências de DNA em cromossomos tem possibilitado a identificação de marcadores cromossômicos, revelando informações importantes sobre a distribuição de sequências de DNA ao longo dos cromossomos (Jong *et al.*, 1999). Muitos tipos de sequências de DNA têm sido visualizados através da FISH, incluindo cópias únicas ou com baixo número de repetições, até aquelas altamente repetitivas. As primeiras sequências detectadas foram as de DNA repetitivo, cujas unidades de repetição 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 e não codificante, como DNA telomérico (Leitch *et al.*, 1994).

Nos últimos anos, a FISH tem ampliado consideravelmente o nível de análise e o campo de atuação da citogenética vegetal. Esta técnica tem permitido a detecção simultânea de várias sondas no mesmo cromossomo ou em diferentes cromossomos, possibilitando que seja determinada a ordem física de sequências nos cromossomos, permitindo assim a construção de mapas físicos. Além disso, o poder de resolução espacial de sequências-alvo e o nível de sensibilidade de detecção têm sido aperfeiçoados. 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).

3.5 Microssatélites (SSRs)

As SSRs (*Repetições de Sequências Simples*), também conhecidas como 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 (Tóth, *et al.*, 2000; Bhargava e Fuentes, 2009). Sua distribuição, ao longo dos cromossomos, pode ser exemplificada pela Figura 1. São amplamente utilizados como marcadores moleculares por conta de uma característica particular na qual sofrem maiores taxas de mutações que o restante do genoma (Oliveira *et al.*, 2006).

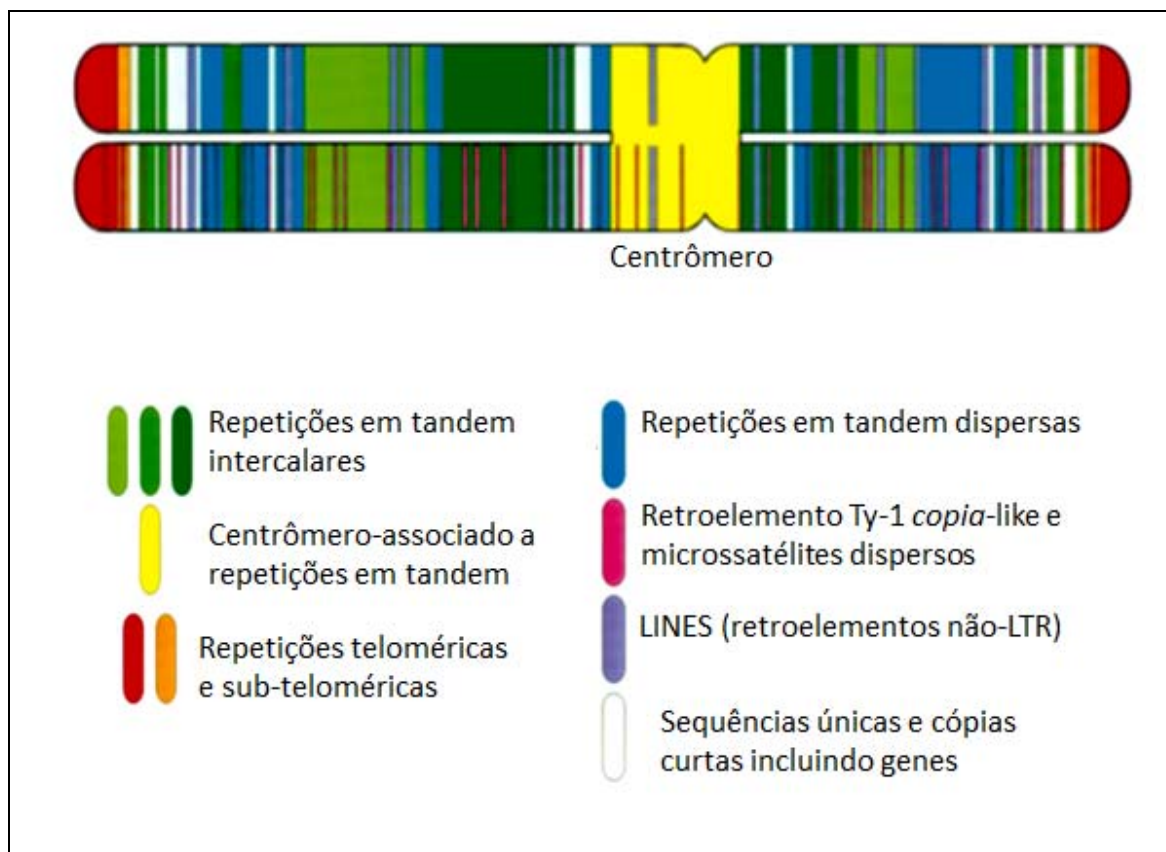


Figura 1: Modelo de um cromossomo de planta com as diferentes classes de DNA repetitivo mostrando a distribuição genômica característica. Os retroelementos estão localizados em todo o genoma com alguns clusters. Os microssatélites estão dispersos e presente como clusters (mostrado apenas na cromátide inferior e varia entre os motivos). Fonte: Schmidt e Heslop-Harrison (1998).

Os microssatélites são classificados de acordo com o tipo de sequência de repetição como perfeito, imperfeito, interrupto ou composto. Em um microssatélite perfeito, a sequência de repetição não é interrompida por qualquer base que não pertence ao motivo (por exemplo, TATATATATATATA). Em um microssatélite imperfeito, existe um par de bases entre o motivo repetido que não coincide com a sequência motivo (por exemplo, TATATATACTATATA). No caso de um microssatélite interrupto, existe uma pequena sequência dentro da sequência repetida que não coincide com a sequência do motivo (por exemplo, TATATACGTGTATATATATA). Por sua vez, em um microssatélite composto, a sequência contém duas sequências adjacentes distintas de repetição (por exemplo, TATATATATAGTGTGTGTGT) (Oliveira *et al.*, 2006).

As SSRs também podem ser classificadas, de acordo com sua localização no genoma, 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).

Embora os microssatélites tenham sido usados extensivamente em um número considerável de estudos, das mais variadas áreas da genética, a dinâmica de mutação destas regiões genômicas ainda não é bem compreendida. Vários mecanismos têm sido propostos para explicar a alta taxa de mutação de microssatélites, incluindo erros durante a recombinação,

crossing over desigual e *slippage* da enzima polimerase durante a replicação ou reparação do DNA (Oliveira *et al.*, 2006).

O *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). 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 de sequências únicas dispostas em regiões eucromáticas. Estas regiões eucromáticas 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; Bhargava e Fuentes, 2009).

3.6 Elementos transponíveis e Retroelementos

Os elementos transponíveis (TEs, *Transposable Elements*) são os mais abundantes componentes do genoma das plantas. Por exemplo, aproximadamente 40% do genoma do arroz e 80% de do genoma do milho é composto por TEs. Com base em mecanismos de transposição, TEs são geralmente classificados em dois tipos principais: transposons e retrotransposons (Du *et al.*, 2010)

Apesar de serem muitas vezes conhecidos como componentes do DNA, mais evidências demonstram que TEs não só contribuem para moldar e remodelar os genomas das plantas, mediante amplificação, recombinação e metilação, mas também têm um papel significativo na regulação da expressão de genes adjacentes e criando a matéria-prima para o a evolução de novos genes e novas funções genéticas (Du *et al.*, 2010).

Os TEs são distribuídos em duas classes principais (I e II), de acordo com sua organização estrutural e mecanismo de transposição. Os elementos

de classe I (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 (Transposons), codificam uma transposase, movimentando-se pelo genoma mediante um DNA intermediário, via mecanismo do tipo *cut and paste* (Todorovska, 2007; Du *et al.*, 2010). 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 retroelementos são componentes discretos do genoma de plantas, que replicam e se reintegram em sítios múltiplos através de um processo complexo que envolve a ativação da excisão, transcrição de RNA dependente de DNA, tradução de RNA em proteínas funcionais, síntese de DNA dependente de RNA e reintegração no genoma, gerando novas cópias de retroelementos (Heslop-Harrison, 2000). Apresentam-se frequentemente dispersos ao longo dos cromossomos, embora possam se associar a regiões particulares do genoma (Figura 1) (Schmidt e Heslop-Harrison, 1998). As principais classes de retroelementos incluem LINEs (*Long Interspersed Nuclear Elements*; Elementos Repetitivos Dispersos Longos), SINEs (*Short Interspersed Nuclear Elements*; Elementos Repetitivos Dispersos Curtos), elementos tipo *copia* e *gypsy*, bem como retrovírus (Heslop-Harrison, 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).

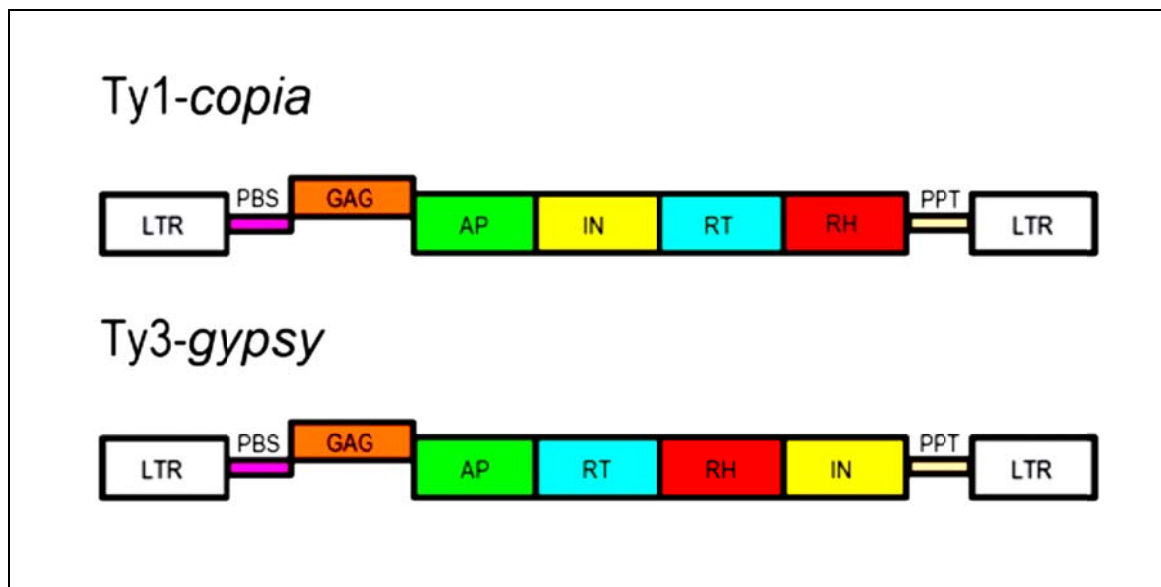


Figura 2: Organização dos domínios gênicos das principais classes de retrotransposons LTRs (Ty1-copia e Ty3-gypsy). Adaptado de Todorovska (2007) *apud* Bortoleti (2010).

Os retroelementos agem como elementos mutagênicos sendo considerados como fonte de diversidade. As sequências de retroelementos potencialmente ativas oferecem dados importantes sobre a evolução genômica e relações filogenéticas (Heslop-Harrison, 2000). Alguns tipos de retroelementos têm sido analisados no genoma de plantas. Em *Beta vulgaris* L., Schmidt *et al.* (1995) observaram a presença do retrotransposon Ty1-copia ao longo dos cromossomos. De modo semelhante, Galasso *et al.* (1997) estudaram a distribuição e localização dos retrotransposons Ty1-copia em espécies de *Vigna*, revelando que estes elementos representavam a maior fração do genoma e se encontravam dispersos uniformemente em todos os cromossomos, com exceção das regiões centroméricas, subteloméricas e das RONS (Regiões Organizadoras de Nucléolos).

O retroelemento Ty3-gypsy, que difere do grupo Ty1-copia pela ordem gênica no domínio interno, foi localizado nas regiões centroméricas em cromossomos de cereais sugerindo que este grupo de retroelementos pode

desempenhar importante função na evolução cariotípica e no funcionamento dos centrômeros (Presting *et al.*, 1998). Em grão-de-bico (*Cicer arietinum*), Stagginnus *et al.* (1999) observaram o agrupamento de elementos Ty3-gypsy na heterocromatina pericentromérica rica em AT em todos os cromossomos, sendo considerado um importante marcador nesta espécie.

Os transposons, por sua vez, são classificados em sete superfamílias: Tc1/Mariner; hAT; PiF; MiTE; CACTA; MULE, e Mutator, também denominados Pack-Mules. Em vegetais, esta última família é a mais abundante, embora as quantidades dos elementos da superfamília CACTA e Helitron também tenham sido consideradas relevantes (Jiang *et al.*, 2004; Wicker *et al.*, 2007; Li *et al.*, 2009).

Os elementos da superfamília CACTA são um dos que se apresentam mais relacionados a regiões ricas em genes, podendo exercer maior influência sobre eles. No entanto, nem sempre as alterações geradas são benéficas e os organismos por vezes reverterem sua influência através de mecanismos epigenéticos, especialmente por metilação, inibindo a atividade destes elementos (Cantu *et al.*, 2010; Saze *et al.*, 2012).

6. Referências Bibliográficas:

- Arumuganathan K and Earle ED (1991) Nuclear DNA content of some important plant species. *Plant Mol Biol Rep* 9: 208-218.
- APG III - Angiosperm Phylogeny Group (2009). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society* 161 (2): 105–121. doi:10.1111/j.1095-8339.2009.00996.x. Retrieved 4 February 2014.
- Bhargava A, Fuentes FF (2009) Mutational Dynamics of Microsatellites. *Molecular Biotechnology* 44:250-266.
- 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, 189-229 pp.
- Benko-Iseppon AM (2001) Estudos moleculares no caupi e em espécies relacionadas: Avanços e perspectivas. *Embrapa Documentos* 56:327-332.
- Bilyeu K, Ratnaparkhe M and Kole C (2010) Genetics, Genomics and Breeding of Soybean. Science Publishers, USA, 350 pp.
- Biondo E, Miotto STS and Schifino-Wittmann MT (2005) Citogenética de espécies arbóreas da subfamília Caesalpinioideae - Leguminosae do sul do Brasil. *Ciência Florestal* 15:241-248.
- Borém A and Vieira MLC (2005) Glossário de Biotecnologia. 1ª Edição Editora Suprema, Visconde do Rio Branco, 177 pp.
- Bortoleti KCA, Benko-Iseppon AM, Melo NF and Brasileiro-Vidal AC (2012) Chromatin differentiation between *Vigna radiata* (L.) R. Wilczek and *V. unguiculata* (L.) Walp. (Fabaceae). *Plant Syst Evol* 298:689-693.
- Bortoleti KCA (2010) Mapeamento cromossômico comparativo em *Glycine* Willd, *Phaseolus* L. e *Vigna* Savi. Tese de Doutorado, Universidade Federal de Pernambuco, Recife, 153 p.

- Broughton WJ, Hernandez G, Blair M, Beebe S, Gepts PE and Vanderleyden J (2003) Beans (*Phaseolus* spp.) - model food legumes. *Plant and Soil* 252:55-128.
- Cantu D, Vanzetti LS, Sumner A, Dubcovsky M, Matvienko M, Distelfeld A, Michelson RW, J (2010) Small RNAs, DNA Methylation and Transposable Elements in Wheat. *BMC Genomics* 11 (January): 408.
- Capy P, Bazin C, Higuete D and Langin T (1998) Dynamics and Evolution of Transposable Elements. Chapman and Hall, New York.
- Carvalho R, Soares Filho WS, Brasileiro-Vidal AC, Guerra M (2005) The relationships among lemons, limes and citron: a chromosomal comparison. *Cytogenet Genome Res* 109:276-82.
- Carvalho R and Guerra M (2002) Cytogenetics of *Manihot esculenta* Crantz (cassava) and eight related species. *Hereditas* 136:159-68.
- Castelletti CHM and Costa AF (2013) Feijão-caupi: alternativa sustentável para os sistemas produtivos. *Pesq agropec pernamb*, Recife, 18:1-2.
- Clarindo WR, De Carvalho CR and Alves BMG (2007) Mitotic evidence for the tetraploid nature of *Glycine max* provided by high quality karyograms. *Plant Systematics and Evolution* 265:101-107.
- Choi HW, Kim MY, Lee SH, Sultana S and Bang JW (2013) Molecular cytogenetic analysis of the *Vigna* species distributed in Korea. *Genes & Genomics* 35:257–264.
- Chung G and Singh RJ (2008) Broadening the genetic base of soybean: A multidisciplinary approach. *Critical Reviews in Plant Sciences* 27:295-34.
- Cuadrado A and Schwarzacher T (1998) The chromosomal organization of simple sequence repeats in wheat and rye genomes. *Chromosoma* 107: 587-594.
- 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.

- EMBRAPA (2011) Empresa Brasileira de Pesquisa Agropecuária, Meio Norte. Feijão-Caupi no Brasil. Produção, Melhoramento Genético, Avanços e Desafios. Teresina-PI, 84 p.
- Ehlers JD and Hall AE (1997) Cowpea (*Vigna unguiculata* L. Walp.). Field Crops Res 53:187-204.
- Ellegren H (2004) Microsatellites: simple sequences with complex evolution. Nature Reviews Genetics 5:435-445.
- Freire-Filho FR, Ribeiro VQ, Rocha MM, Silva KJD, Nogueira MSR and Rodrigues EV (2011a) Feijão-Caupi no Brasil. Produção, Melhoramento Genético, Avanços e Desafios. 1^a Edição, Jorimá Marques Ferreira Teresina, 84 p.
- Freire-Filho FR, Ribeiro VQ, Cardoso MJ, Azevedo JN, Ramos SRR, Rocha MM and Silva KJD (2011b) Coleção ativa de germoplasma de feijão-caupi (*Vigna unguiculata* (L.) Walp.) e de outras espécies do gênero *Vigna*, da Embrapa Meio-Norte, no período de 1976 a 2003 – Teresina. Embrapa Meio-Norte. 125 p.
- 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, 28-92 pp.
- Friesen N, Brandes A and Heslop-Harrison JS (2001) Diversity, origin, and distribution of retrotransposons (*gypsy* and *copia*) in conifers. Molecular Biology and Evolution 18:1176-1188.
- Frota KDG, Soares RAM and Areas JAG (2008) Chemical composition of cowpea (*Vigna unguiculata* L. Walp), BRS-Milenio cultivar. Ciênc Tecnol Aliment 28:470-6.

- 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. *Annals of Botany* 80:327–333.
- Global Industry Analysts, Inc, (2013): <http://goo.gl/qoph6a>. Acessado em 30 de Novembro, 2013.
- Guedes RE (2008) Bases para o Cultivo Orgânico de Feijão-Caupi [*Vigna unguiculata* L. (Walp.)] no Estado do Rio de Janeiro. Tese de doutorado, Universidade Federal Rural do Rio de Janeiro, Rio de Janeiro, 68 p.
- Guerra M (2004) Hibridização *in situ*: Princípios Básicos. In: Guerra M (ed). FISH: Conceitos e Aplicações na Citogenética. Sociedade Brasileira de Genética, Ribeirão Preto, 1-32 pp.
- 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. *Annals of Botany* 78:157-161.
- Graham PH and Vance CP (2003). Legumes: importance and constraints to greater use. *Plant Physiology* 131:872-877.
- Hansen CN and Heslop-Harrison JS (2004) Sequences and phylogenies of plant pararetroviruses, viruses and transposable elements. *Adv Bot Res* 41:165-193.
- Harlan JR, De Wet JMJ (1971) Toward a rational classification of cultivated plants. *Taxon* 20:509–517.
- 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, Schwarzacher T (2011) Organisation of the plant genome in chromosomes. *The Plant Journal* 66:18-33.
- Hill P, Burford D, Martin DM, Flavell AJ (2005) Retrotransposon populations of *Vicia* species with varying genome size. *Mol Gen Genomics* 273: 371–381.

- Hymowitz T (2004) Speciation and cytogenetics. In: HR Boerman, JE Specht (eds) Soybeans: Improvement, Production and Uses. Am Soc Agron, Madison, WI, USA, pp 97-136.
- Hymowitz T. (2008) The history of the soybean. In: LA Johson, PJ White, RGallway (eds) Soybeans Chemistry, Production, Processing, and Utilization. AOCS Press, Urbana, IL, USA, pp 1-31.
- Iwata A, Grenland CM and Jackson SA (2013) Cytogenetics of legumes in the Phaseoloid Clade. *The Plant Genome* 6:1-8
- Jiang N, Bao Z, Zhang X, Eddy SR, Wessler SR (2004) Pack-MULE Transposable Elements Mediate Gene Evolution in Plants. *Nature* 431 (7008): 569–73.
- Joly AB (2002) *Introdução à Taxonomia Vegetal*. 13ª edição, Companhia Editora Nacional, São Paulo, 777 pp.
- Jong JH, Fransz P and Zabel P (1999) High resolution FISH in plants – techniques and applications. *Trends in 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, 576 pp.
- Kumar A and Bennetzen JL (1999) Plant retrotransposons. *Annual Review of Genetics* 33:479-532.
- Leitch AR, Schwarzacher T, Jackson D and Leitch IL (1994) *In situ hybridization: a practical Guide*. Bios Scientific Publishers Ltd., UK, 118 pp.
- Lewis G, Schrirer B, Mackinder B and Lock M (2005) *Legumes of the world*, 1^s Kew : Royal Botanic Gardenst, London, 577 pp.
- Li Q, Li L, Dai JR, Li JS, Yan JB (2009) Identification and Characterization of CACTA Transposable Elements Capturing Gene Fragments in Maize. *Chinese Science Bulletin* 54 (4): 642–51

- Li YC, Abraham BK, Fahima T and Nevo E (2004) Microsatellites within genes: Structure, function, and evolution. *Molecular Biology and Evolution* 21:991-1007.
- Maia FMM, Oliveira JTA, Matos MRT, Moreira RA, Vasconcelos LM (2000) Proximate Composition, Amino Acid Content and Haemagglutinating and trypsin inhibiting actives of some Brazilian *Vigna unguiculata* (L.) Walp. Cultivares. *Journal of Science and Food Agriculture*, London, n. 80, p. 453-458, 2000.
- Ministério da Agricultura e Pecuária (MAPA) 2013 <http://goo.gl/QPrkHL> (28 de Novembro 2013).
- Ministério da Agricultura e Pecuária (MAPA) 2013a <http://goo.gl/0t1U9L> (28 de Novembro 2013).
- 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.
- Ng NQ and Maréchal R (1985) Cowpea taxonomy, origin germ plasm. In: Sinch SR, Rachie KO, eds. *Cowpea research, production and utilization*. Chichester, John Wiley, 11-21 pp
- National Soybean Research Laboratory (NSRL) (2012) "Soy Benefits". http://www.nsrl.uiuc.edu/soy_benefits.html (February 2014).
- Oliveira EJ, Pádua JG, Zucchi MI, Vencovsky R and Vierira MLC (2006) Origin, evolution and genome distribution of microsatellites. *Genetics and Molecular Biology* 29:294-307.
- 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 *Orobanchaceae* and *Phelipanche* (*Orobanchaceae*). *Gene* 387: 75–86.
- 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. Ribeirão Preto, Sociedade Brasileira de Genética 33-59 pp.

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. Theoretical and Applied Genetics 112:924-933.

Presting GG, Malysheva L, Fuchs J and Schubert I (1998) A TY3/GYPSY retrotransposon-like sequence localizes to the centromeric regions of cereal chromosomes. The Plant Journal 16:721-728.

Primmer CR, Ellegren H, Saino N and Moller AP (1996) Directional evolution in germline microsatellite mutations. Nature Genetics 13:391-393.

Saze H, Tsugane K, Kanno T, Nishimura T (2012) DNA Methylation in Plants: Relationship to Small RNAs and Histone Modifications, and Functions in Transposon Inactivation. Plant & Cell Physiology 53 (5): 766–84.

Schmidt T and Heslop-Harrison JS (1998) Genomes, genes and junk: the large-scale organization of plant chromosomes. Trends in Plant Science 3:195-199.

Schmidt T, Kubis S and Heslop-Harrison JS (1995) Analysis and chromosomal localization of retrotransposons in sugar beet (*Beta vulgaris* L.): LINEs and *Ty1-copia*-like elements as major components of the genome. Chromosome Research 3:335-345

Schmutz J, Cannon SB, Schlueter JA, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, *et al* (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178-183.

Sen NK and Vidyabhusan RV (1960) Tetraploid soybeans. Euphytica 9:317–322.

Singh RJ, Kim HH and Hymowitz T (2001) Distribution of rDNA loci in the genus *Glycine* Willd. Theoretical and Applied Genetics 103:212-218.

Singh RJ, Kollipara KP and Hymowitz T (1998) The genomes of *Glycine canescens* FJ Herm. and *G. tomentella* Hayata of Western Australia

- and their phylogenetic relationships in the genus *Glycine* Willd. *Genome* 41:669–679.
- Singh RJ, Chung GH and Nelson RL (2007) Landmark research in legumes. *Genome* 50:525-537.
- Singh RJ, Nelson RL and Chung GH (2007) Soybean (*Glycine max* (L.) Mer.). *In: Genetic Resources, Chromosome Engineering, and Crop Improvement Oilseed Crops* 4:13–50.
- Singh, R. J. (2003) *Plant Cytogenetics*, 2 ed. CRC Press, Inc., Boca Raton, FL.
- Silva SMS and Freire-Filho FR (1999) Proteínas de feijão-caupi (*Vigna unguiculata* (L.) Walp.): Caracterização e aplicação nutricional. *Embrapa Documentos* 44:520.
- 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, 639 pp.
- Souza MGC and Benko-Iseppon AM (2004) Cytogenetics and chromosome banding patterns in Caesalpinioideae and Papilionioideae species of Pará, Amazonas, Brazil. *Bot J Linn Soc* 144:181-191.
- Soystats 2013 <http://soystats.com/> (28 de Novembro 2013).
- Staginnus C, Winter P, Desel C, Schmidt T and Kahl G (1999) Molecular structure and chromosomal localization of major repetitive DNA families in the chickpea (*Cicer arietinum* L.) genome. *Plant Molecular Biology* 39:1037-1050.
- Timko MP and Singh BB (2008) Cowpea, a multifunctional legume. *In: Moore PH and Ming R (eds), Genomics of Tropical Crop Plants*, Springer Science Business Media, LLC, New York , 227-258 pp.
- Tóth G, Gáspári Z and Jurka J (2000) Microsatellites in different eukaryotic genomes: Survey and analysis. *Genome Research* 10:967-981
- Todorovska E (2007) Retrotransposons and their role in plant – genome evolution. *Biotechnol. & Biotechnol. EQ* 21: 294–305.
- USDA, 2013: <http://goo.gl/nwn1CF>; Acessado em 23 de Dezembro, 2013.

- 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.
- Wander AL, Didonet AD, Abreu AFB, Silva CC, Quintela ED, Fonseca JR, Silva JG, Stone LF, Ramalho MAP, Junior ML, *et al* (2005) Cultivo do feijão irrigado na região noroeste de Minas Gerais. Embrapa Arroz e Feijão.
- Wang W, He Q, Yang H, Xiang S, Zhao T and Gai J (2013) Development of a chromosome segment substitution line population with wild soybean (*Glycine soja* Sieb. et Zucc.) as donor parent. *Euphytica* 189:293-307.
- Weising K, Nybom H, Wolff K and Andkahl G (2005) DNA Fingerprinting in plants, principles, methods, and applications. 2nd ed., CRC Press, Boca Raton, 444 pp.
- Westphal E (1974) Pulses in Ethiopia: their taxonomy and agriculture significance. Agricultural research report 815. Center for Agricultural Publishing and Documentation, Wageningen 263 pp.
- Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, *et al.* (2007) A Unified Classification System for Eukaryotic Transposable Elements. *Nature Reviews. Genetics* 8 (12): 973–82.
- Wilson RF (2008) Soybean: market driven research needs. *Genetics and genomics of soybean* 2:3-15.

Capítulo I

Distribuição cromossômica de microssatélites em *Glycine soja* Sieb. & Zucc e *G. tomentella* Hayata

Pollyana Karla da Silva¹; Ana Maria Benko-Iseppon¹; Ricardo Vilela Abdelnoor²,
Marcelo Fernandes de Oliveira², Ana Christina Brasileiro-Vidal¹

¹Universidade Federal de Pernambuco, Departamento de Genética, Laboratório de Genética e Biotecnologia Vegetal, Av. da Engenharia, s/nº, CEP 50740-600, Recife – PE, Brasil.

²Embrapa Soja, Caixa Postal 231, CEP 86001-970, Londrina, PR, Brasil.

* Autor para correspondência: brasileirovidal.ac@gmail.com

Resumo

O presente trabalho teve como objetivo realizar um estudo citogenético comparativo entre *G. soja* e *G. tomentella*, mediante localização cromossômica de sondas de DNA repetitivo. Foram utilizados seis oligonucleotídeos com padrões de microssatélites [(AAC)₅, (AAG)₅, (ACC)₅, (AG)₈, (CTC)₅, (TGA)₆] em *G. soja* e *G. tomentella* como sondas na hibridização *in situ* fluorescente (FISH). A análise por FISH evidenciou que, em ambas as espécies, os microssatélites apresentaram uma marcação dispersa com predominância pericentromérica. No entanto, foram observadas variações em suas quantidades e sua distribuição ao longo dos cromossomos. Em *G. soja*, o oligonucleotídeo (ACC)₅ apresentou marcações dispersas, com marcas pericentroméricas mais intensas, além de marcações subterminais. O oligonucleotídeo (CTC)₅ apresentou cromossomos com marcações dispersas, e algumas marcações subterminais. Já a sonda (AG)₈ evidenciou marcações dispersas, pericentroméricas e terminais. O (TGA)₆ apresentou marcações dispersas, proximais e subterminais; enquanto (AAG)₅ e (AAC)₅ geraram sinais dispersos e proximais. Nas células analisadas de *G. tomentella*, foram observadas marcações dispersas, preferencialmente pericentroméricas para todos os oligonucleotídeos. À luz das informações disponíveis sobre o genoma de *G. max* e sabendo que as SSRs são normalmente distribuídas ao longo do genoma, propõe-se que, como em outros grupos, tais marcações se acumularam em áreas provavelmente pobres em genes mediante mecanismos de amplificação, com ênfase para o deslizamento da polimerase durante a replicação.

Palavras-chave: DNA repetitivo, Hibridização *in situ* fluorescente (FISH), *Glycine soja*, *G. tomentella*.

Abstract

This study aimed to perform a cytogenetic study comparing *G. soja* and *G. tomentella* by chromosomal location of repetitive DNA probes. Six oligonucleotides with microsatellite patterns [(AAC)₅, (AAG)₅, (ACC)₅, (AG)₈, (CTC)₅, (TG)₆] were used in *G. soja* and *G. tomentella* as probes in fluorescent in situ hybridization (FISH). The FISH analysis showed that, in both species, microsatellites showed a predominance dispersed with pericentromeric localization. However, variations in their amounts and distribution along the chromosomes were observed. In *G. soja*, the oligonucleotide (ACC)₅ had a dispersed distribution, with more intense pericentromeric marks, and subterminal markings. The oligonucleotide (CTC)₅ presented five chromosomes with scattered markings, and some subterminal markings. Already a probe (AG)₈ showed dispersed pericentromeric and terminal markings. The (TGA)₆ showed scattered, proximal and subterminal markings, while (AAG)₅ and (AAC)₅ generated and dispersed proximal signals. Cells analyzed from *G. tomentella* showed scattered signals mainly in the pericentromeric regions for all oligonucleotides. In the light of the information available about the genome of *G. max* and knowing that SSRs are normally distributed throughout the genome, we propose that, as in other groups, such markings have built probably in poor areas through mechanisms of gene amplification, with an emphasis on polymerase slippage during replication.

Keywords: Repetitive DNA, Fluorescence in situ hybridization (FISH), *Glycine soja*, *G. tomentella*.

Introdução

O gênero *Glycine* Willd. (família Fabaceae, subfamília Papilionoideae, tribo Phaseoleae) é constituído por dois subgêneros, *Glycine* Willd., que compreende 26 espécies perenes selvagens; e *Soja* (Moench.) F.J. Herm., constituído por *G. soja* Sieb. & Zucc. (soja selvagem) e *G. max* (L.) Merr. (soja cultivada). Considera-se que *G. max* apresente uma provável origem monofilética a partir de *G. soja* (Bilyeu, *et al.*, 2010; Guo, *et al.*, 2010). Ambas as espécies são anuais com $2n = 40$ cromossomos, 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 G₁, respectivamente (Chung e Singh, 2008), havendo algumas linhagens que diferem por apresentar translocações (Singh and Hymowitz, 1988; Findley *et al.*, 2010; 2011) ou inversões paracêntricas (Palmer *et al.*, 2000).

As espécies e respectivos acessos do subgênero *Soja* constituem o grupo gênico primário da soja, enquanto o grupo secundário não apresenta espécies associadas à cultura, e o terciário contém apenas três espécies (*G. argyrea*, *G. canescens* e *G. tomentella*), as quais podem ser hibridizadas com soja. Seus híbridos F₁ podem ser resgatados mediante cultivo de embriões, mas são estéreis e dificilmente passam do estágio de anfidiplóides (Chung e Singh, 2008). Dessa forma, apenas uma pequena fração da diversidade genética disponível no gênero tem sido efetivamente utilizada no melhoramento da soja. Em geral, os melhoristas se restringem a utilizar o material do grupo gênico primário, resultando em uma base genética estreita para a espécie (Mahmoud *et al.*, 2006; Guo *et al.*, 2010). Apesar dessa aparente limitação, o melhoramento genético da soja tem levado a progressos significativos em termos de produtividade nos últimos anos, aumentando sua produção. No Brasil, a soja é a cultura que mais cresceu nas últimas três décadas e corresponde a 49% da área plantada em grãos (MAPA, 2013).

A soja é uma das mais importantes fontes de proteína e de óleo do mundo, tratando-se da primeira leguminosa a ser completamente sequenciada. Seu genoma tem servido 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).

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* (significando armação ou esqueleto em português) 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 de 450 genes preditos (Schmutz *et al.*, 2010). As posições dos *scaffolds* foram determinadas com extensos mapas genéticos, incluindo 4.991 polimorfismos de um único nucleotídeo (SNPs) e 874 SSRs (*Simple Sequence Repeats*) (Song *et al.*, 2004; Choi *et al.*, 2007; Hyten *et al.*, 2010a, b).

Os SSRs, também conhecidos como microssatélites, são repetições curtas – um a seis pares de bases (pb) – presentes predominantemente nos genomas dos eucariontes, podendo ser encontradas em qualquer região do genoma, tanto em regiões codificantes como não codificantes (Tóth *et al.*, 2000; Bhargava e Fuentes, 2010). Tem sido amplamente utilizadas para geração de marcadores genéticos uma vez que possuem um atributo específico no qual sofrem maiores taxas de mutação que o restante do genoma (Oliveira *et al.*, 2006). Devido à sua alta mutabilidade, acredita-se que os microssatélites podem desempenhar um papel significativo na evolução do genoma pela criação e manutenção da variação da genética quantitativa (Tóth *et al.*, 2000).

Nesse sentido, o presente trabalho teve como objetivo realizar um estudo citogenético comparativo entre *G. soja* e *G. tomentella*, mediante localização cromossômica de sondas com padrão microssatélite, a fim de colaborar para a análise das diferenças estruturais do DNA repetitivo entre espécies dos dois subgêneros.

Material e Métodos

Sementes das espécies *G. soja* (Acesso PI 407236) e *G. tomentella* (Acesso PI 441008) foram cedidas pelo banco de germoplasma da Embrapa Soja (Empresa Brasileira de Pesquisa Agropecuária, Soja, Londrina). Estas foram escarificadas e postas para germinar em placas de Petri com papel filtro e uma camada de algodão umedecido. Pontas de raízes recém-germinadas foram pré-tratadas em solução de 8-hidroxiquinoleína (8HQ) a 2 mM por 4h e 30 min, à 18 °C, fixadas em metanol:ácido acético (3:1, v/v), por 2 a 24 h à temperatura ambiente e estocadas a -20 °C.

Preparo das lâminas

Os meristemas foram lavados duas vezes em água destilada por 5 min cada, sendo digeridos em uma solução de 2% (p/v) celulase 'Onozuka R-10' (Serva)/ 20% (v/v) pectinase (Sigma-Aldrich) em tampão citrato 10 mM por 2:30 h a 37 °C. Em seguida, as raízes foram lavadas em água destilada e colocadas em uma lâmina ligeiramente inclinada verticalmente. Algumas gotas de fixador (metanol: ácido acético; 3:1; v:v) gelado foram pingadas com uma pipeta Pasteur em cima das raízes para total remoção da água. Em seguida, procedeu-se à maceração mecânica das raízes e posterior gotejamento de fixador gelado sobre o material. Posteriormente, as lâminas foram secas com um auxílio de bombas de ar possibilitando a secagem das mesmas, sendo então mergulhadas em ácido acético 45% por 12 s e colocadas em uma superfície plana pré-aquecida (37 °C) por aproximadamente 3 min. As lâminas foram coradas com uma solução de 2 µg/mL de DAPI (4'-6'-diamidino-2-fenilindole)/ glicerol (1:1, v/v) e visualizadas em microscópio a fim de selecionar as melhores lâminas. As lâminas selecionadas foram descoradas em etanol:ácido acético (3:1, v/v) por 30 min, e mantidas etanol absoluto durante 1 h. Após secagem ao ar, as lâminas foram devidamente estocadas a -20 °C até o momento do procedimento de hibridização *in situ*.

Dupla coloração com fluorocromos CMA e DAPI

Para a coloração de CMA/DAPI, as lâminas foram armazenadas por três dias à temperatura ambiente, coradas com CMA (0,5 mg/mL) por 60 min e, em seguida, com DAPI (2 µg/mL), por 30 min. Posteriormente, as lâminas foram secas, montadas com meio de montagem de tampão McIlvaine's (pH 7.0):glicerol (1:1, v/v) e lutadas com esmalte. Após três dias, imagens das melhores células foram capturadas.

Obtenção de sondas

As sequências de oligonucleotídeos utilizados como sondas foram previamente preparadas e estocadas no Laboratório de Genética e Biotecnologia Vegetal da UFPE. Essas sequências descritas por Bortoleti (2010) pertencem às espécies *P. lunatus*, *P. vulgaris*, *V. unguiculata* e *V. radiata* e 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 para identificação de sequências de oligonucleotídeos com padrões frequentes em microsatélites. Os oligonucleotídeos 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.

Hibridização fluorescente *in situ* (FISH)

As lâminas foram pré-tratadas como descrito por Pedrosa *et al.* (2001). A desnaturação dos cromossomos e das sondas, os banhos pós-hibridização e a detecção foram efetuados de acordo com Heslop-Harrison *et al.* (1991), exceto pela lavagem de estringência, realizada em 0,1x SSC 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 70 °C e

hibridizadas de 18 a 48 h a 37 °C. As sondas marcadas com digoxigenina foram detectadas usando *sheep anti-digoxigenin* conjugado com fluoresceína isotiocianato (FITC; Roche) e amplificadas com *anti-sheep* conjugado com FITC (Roche) 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 em microscópio de epifluorescência Leica DMLB e imagens das melhores células capturadas com uma câmera Leica DFC 340FX, usando o programa Leica CW 4000. As imagens foram otimizadas para melhor brilho e contraste com o Adobe Photoshop CS4 (Adobe Systems Incorporated).

4. Resultados

Ambas as espécies *G. soja* ($2n = 40$) *G. tomentella* ($2n = 78$) apresentaram cromossomos metacêntricos e submetacêntricos e apenas um par de cromossomos satelitados. Na coloração por fluorocromos base-específicos CMA/DAPI em *G. soja* e *G. tomentella*, blocos CMA positivos foram observados nas regiões pericentroméricas de quase todos os cromossomos e nas regiões terminais do par de cromossomos satelitados de ambas as espécies. O padrão de distribuição dos seis microssatélites utilizados como sondas na FISH revelou uma prevalência de marcações dispersas ao longo da maioria dos cromossomos de *G. soja* e *G. tomentella*, bem como de sinais de hibridização nas regiões pericentroméricas de vários cromossomos, com particularidades para cada tipo de microssatélite [Figuras 1A'-F'; 2A'-F'; Tabela 1].

FISH em *Glycine soja*

Em *G. soja*, o oligonucleotídeo (AAC)₅ revelou marcações dispersas ao longo dos cromossomos, evidenciando marcações proximais e em alguns casos sinais subterminais (Figura 1A'). Por sua vez, a sonda (AAG)₅ revelou marcações intersticiais e proximais na maioria dos cromossomos (Figura 1B').

A hibridação das repetições (ACC)₅ revelou marcações dispersas em todos os cromossomos, com marcas pericentroméricas mais intensas, estando estas destacadas em alguns cromossomos, observando-se ainda marcações subterminais (Figura 1C').

No que diz respeito ao oligonucleotídeo (AG)₈, foram geradas marcações dispersas, predominantemente pericentroméricas, em alguns cromossomos apresentando marcações terminais (Figura 1D'), semelhante à repetição (CTC)₅ que também evidenciou marcações dispersas e predominantemente pericentroméricas, além de algumas marcações subterminais (Figura 1E').

O padrão (TGA)₆ associou-se de forma dispersa aos cromossomos, como todos os anteriores, observando-se ainda marcações intersticiais ao longo dos cromossomos, como também sinais de hibridizações proximais e subterminais (Figura 1F').

FISH em *Glycine tomentella*

Na espécie *G. tomentella*, foram observadas marcações dispersas, preferencialmente pericentroméricas ao longo dos cromossomos para todos os oligonucleotídeos utilizados como sonda [(AAC)₅, (AAG)₅, (ACC)₅, (AG)₈, (CTC)₅, (TGA)₆], nas células analisadas (Figura 2; Tabela 1). Nesta espécie o oligonucleotídeo (AAC)₅ revelou marcações preferencialmente intersticiais ao longo dos cromossomos (Figura 2A'), enquanto a sonda (AAG)₅ apresentou sinais pericentroméricos e intersticiais (Figura 2B'). Por sua vez, as repetições (ACC)₅ mostraram sinais de hibridização pericentroméricos, bem como alguns subterminais (Figura 2C').

A sonda (AG)₈ apresentou uma predominância de marcações pericentroméricas e em alguns cromossomo podemos observar marcações subterminais (Figura 2D'). No caso do (CTC)₅, assim como o (AG)₈ evidenciou uma maioria de marcações pericentroméricas e poucas marcações

subterminais (Figura 2 E'). O oligonucleotídeo (TGA)₆ apresentou marcações intersticiais, proximais e em alguns cromossomos marcações subterminais (Figura 2 F').

A presença de marcas preferencialmente pericentroméricas sugere uma associação dos microssatélites testados à heterocromatina constitutiva, embora marcações dispersas também tenham sido observadas em regiões eucromáticas.

Tabela 1. Padrões prevalentes quanto à posição e distribuição de marcações cromossômicas observadas no presente estudo.

Corante ou Sonda	Padrão Predominante	<i>G. soja</i> (2n = 40) (Nº pares) posição	<i>G. tomentella</i> (2n = 78) (Nº pares) posição
CMA ⁺	L	(Todos) Pc / (1) St	(Todos) Pc / (1) St
(AAC) ₅	D/L	(2) St / (10) Px	(3)Pc / (3) I
(AAG) ₅	D/L	(3) I / (6) Px	(4) Pc / (6) I
(ACC) ₅	D/L	(2) Pc / (1) St	(12) Px / (5) St
(AG) ₈	D/L	(15) Pc / (2) St	(6) I / (3) St
(CTC) ₅	D/L	(8) Pc / (2) St	(11) Pc / (4) Px
(TGA) ₆	D/L	(5) Px / (2) I / (1) St	(18) I

Legenda das abreviações: Padrão: **L** = marcação localizada; **D** = marcação dispersa; Posições: **St** = subterminal; **Pc** = pericentromérica; **I** = intersticial; **Px** = proximal.

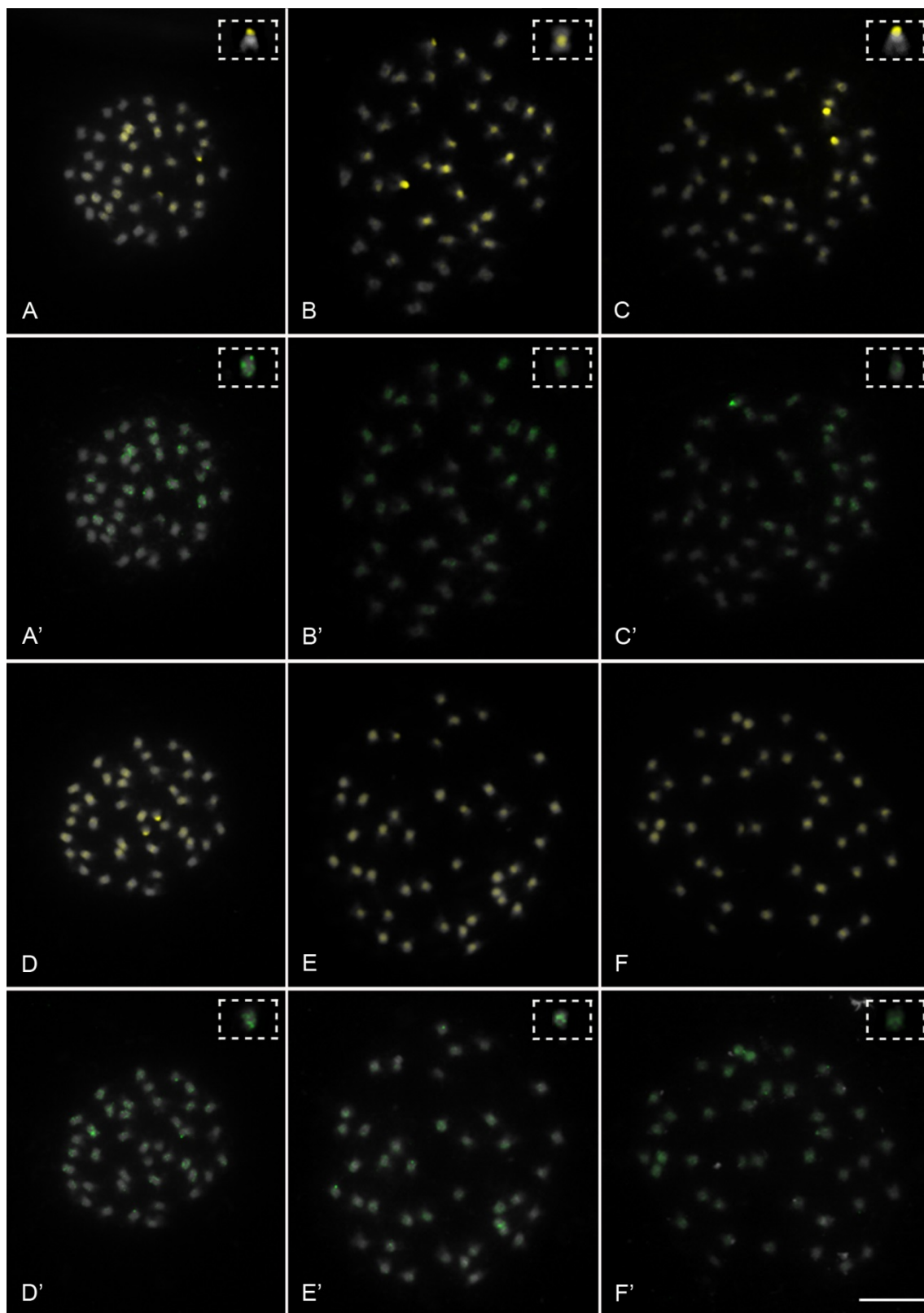


Figura 1: Coloração CMA/DAPI e hibridização *in situ* fluorescente dos oligonucleotídeos (AAC)₅, (AAG)₅, (ACC)₅, (AG)₈ (CTC)₅, (TGA)₆, em cromossomos mitóticos de *Glycine soja* (2n = 40). Cromossomos contraincolorados com DAPI (pseudocolorido em cinza), com blocos CMA positivos (amarelo) nas regiões pericentroméricas dos cromossomos e nas regiões terminais dos cromossomos satelitados (A-F). Marcações dispersas e pericentroméricas dos oligonucleotídeos (AAC)₅ (A'), (AAG)₅ (B'), (ACC)₅ (C'), (AG)₈ (D'), (CTC)₅ (E'), (TGA)₆ (F'), em verde. Os inserts evidenciam exemplos de cromossomo satelitado (A), marcação pericentromérica (B), dispersa (A'), intersticial (B'), pericentromérica (C'), dispersa (D'), pericentromérica (E'), proximal (F'). A barra em F' corresponde a 10 μ m.

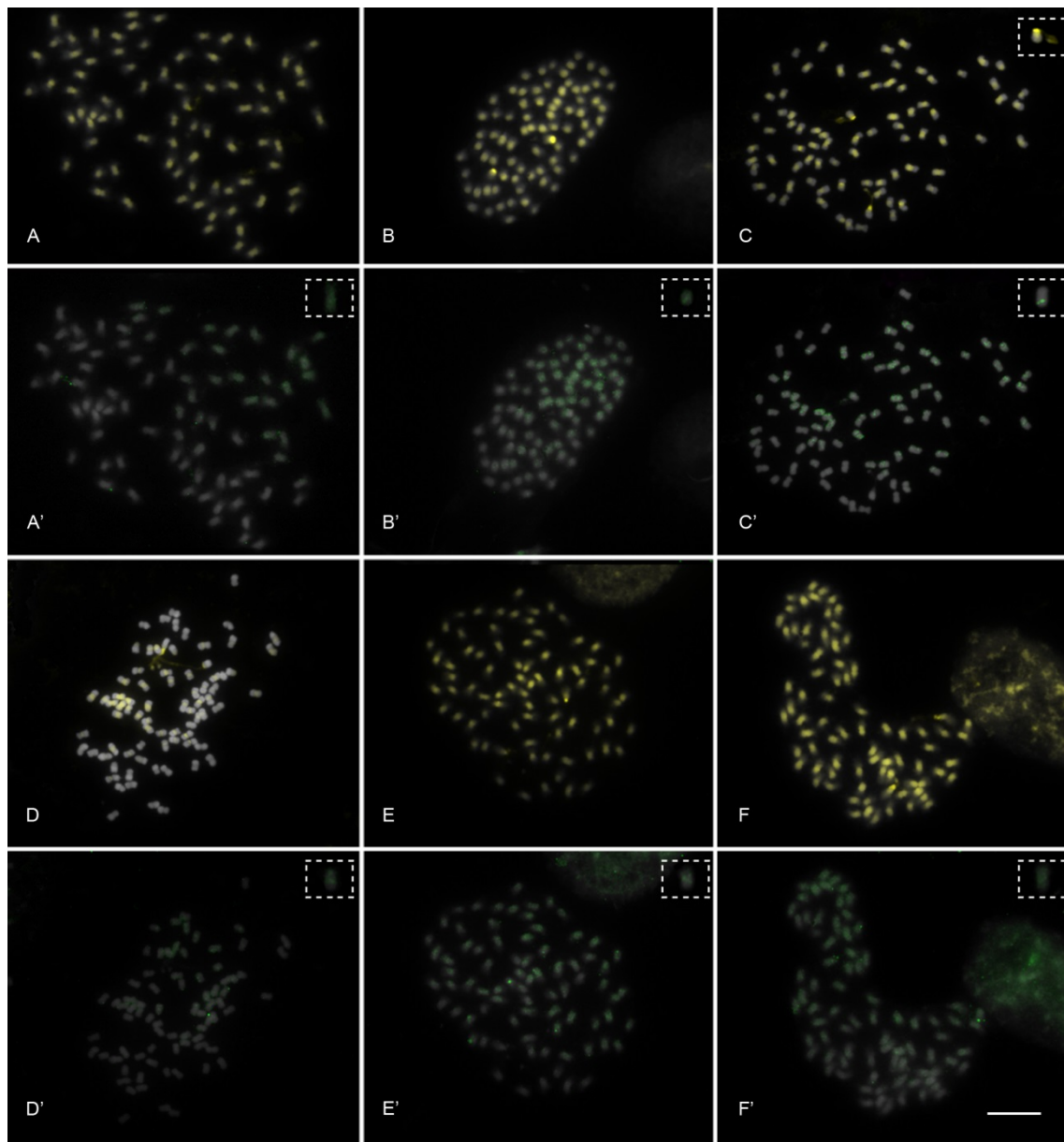


Figura 2: Coloração CMA/DAPI e hibridização *in situ* fluorescente dos oligonucleotídeos (AAC)₅, (AAG)₅, (ACC)₅, (AG)₈, (CTC)₅, (TGA)₆, em cromossomos mitóticos de *Glycine tomentella*. Metáfases contendo $2n = 78$ cromossomos contrainformados com DAPI (pseudocolorido em cinza), com blocos CMA positivos (amarelo) nas regiões pericentroméricas dos cromossomos e nas regiões terminais dos cromossomos satelitados (A-F) em amarelo. Marcações dispersas e pericentroméricas dos oligonucleotídeos (AAC)₅ (A'), (AAG)₅ (B'), (ACC)₅ (C'), (AG)₈ (D'), (CTC)₅ (E'), (TGA)₆ (F') em verde. Os inserts evidenciam exemplos de cromossomo satelitado (C), marcação intersticial (A'), pericentromérica (B'), subterminal (C'), subterminal (D'), pericentromérica (E'), proximal (F'). A barra em F' corresponde a 10 μ m.

Discussão

O presente estudo avaliou pela primeira vez a prevalência e distribuição cromossômica de alguns destes elementos em duas espécies silvestres do gênero *Glycine*, confirmando sua abundância nas diversas regiões cromossômicas dos táxons estudados. Mediante a hibridização *in situ* fluorescente, os oligonucleotídeos (AAC)₅, (AAG)₅, (ACC)₅, (AG)₈, (CTC)₅, (TGA)₆, mostraram de um modo geral marcações dispersas nos genomas das espécies analisadas; embora tenham sido observadas diferenças em suas quantidades e distribuição ao longo dos cromossomos.

Os microssatélites (SSR) são considerados uma classe de sequências de DNA repetitivo encontrada nos genomas dos eucariotos. Tratam-se de elementos abundantes gerados a partir da dispersão de pequenas unidades de repetição (de um a seis pb) organizados *em tandem* sendo encontrados ao longo dos genomas e mais frequentemente associados às regiões heterocromáticas (Cuadrado *et al.*, 2000, Oliveira *et al.*, 2006). Uma grande distribuição dessas sequências nos genomas tem potencializado o uso de microssatélites como marcadores moleculares como ferramenta na identificação de indivíduos, espécies e variedades, em estudos populacionais e evolutivos, bem como, na geração de mapas genéticos, como em *G. max* (Guo *et al* 2010, Hyten *et al* 2010a), *G. soja* (He *et al*, 2012), *Cicer arietinum* (Lichtenzweig *et al.*, 2005), *Phaseolus vulgaris* (Yu *et al.*, 2000; Buso *et al.*, 2006), *Vigna unguiculata* (Diouf e Hilu, 2005) e *Medicago truncatula* (Mun *et al.*, 2006), entre outros. Tais oligonucleotídeos sintéticos com padrão de microssatélites têm sido utilizados na FISH, auxiliando no entendimento da organização genômica, inferindo sobre mecanismos de dispersão, amplificação e evolução, confirmando a ampla distribuição destes elementos na maioria das espécies analisadas (Cuadrado *et al.*, 2000; Heslop-Harrison, 2000).

Mediante FISH, Bortoleti (2010) observou que os microssatélites estão presentes em grande proporção nos genomas das espécies *G. max*, *P. vulgaris*, *P. lunatus*, *V. unguiculata*, *V. radiata*, evidenciando variações em suas quantidades, organização (tandem ou dispersa) e distribuição ao longo dos cromossomos, de modo semelhante ao observado no presente estudo. Também em análise das sequências genômicas da soja pela citada autora,

foram detectados 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 (Bortoleti 2010). Este achado contraria evidências comparativas entre os sinais de FISH e a análise das sequências componentes dos *scaffolds* do genoma da soja, que indicam um número muito maior de repetições não evidentes devido às metodologias de alinhamento usadas na montagem de genomas (Belarmino *et al.*, 2012).

Em marcações dispersas, os loci de microssatélites, correspondem a sequências com extensão de 20 a 100 pb, os quais podem estar distribuídos próximos um dos outros ao longo dos genomas, permitindo sua visualização como marcações fracas, entre sequências de cópia única e/ou outras sequências repetitivas. No caso de haver marcação em tandem revela a organização de SSRs em regiões de DNA repetitivo. Essas sequências correspondem a cerca de 100 a 1000 unidades de repetições, as quais estão associadas à heterocromatina constitutiva (Cuadrado *et al.*, 2008).

Estudos com a associação da FISH e análise *in silico* em *G. max* ressaltaram 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 (Bortoleti 2010).

Na 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), tais como sequências de DNA satélite com retroelementos interdispersos (Nunberg *et al.*, 2006; Swaminathan *et al.*, 2007; Findley *et al.*, 2010). Assim, 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).

Sabendo 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).

Agradecimentos

Os autores agradecem ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), no âmbito do Consórcio Brasileiro Genosoja da Empresa Brasileira de Pesquisa Agropecuária, Soja (Embrapa Soja) e à Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (Facepe) pelo apoio financeiro ao projeto.

6. Referências bibliográficas

- Arumuganathan K and Earle ED (1991) Nuclear DNA content of some important plant species. *Plant Mol Biol Rep* 9: 208-218.
- Belarmino LC, Oliveira ARS, Brasileiro-Vidal AC, Bortoleti KCA, Bezerra-Neto JP, Abdelnoor RV, Benko-Iseppon AM (2012). Mining plant genome browsers as a mean for efficient connection of physical, genetic and cytogenetic mapping: an example using soybean. *Genetics and Molecular Biology* 35: 335-347.
- Bhargava A and Fuentes FF (2010) Mutational Dynamics of Microsatellites. *Molecular Biotechnology* 44:250-266.
- Bilyeu K, Ratnaparkhe M and Kole C (2010) Genetics, Genomics and Breeding of Soybean. Science Publishers, USA, 350 pp.
- Bortoleti KCA (2010) Mapeamento cromossômico comparativo em espécies de *Glycine* Willd., *Phaseolus* L. e *Vigna* Savi. Tese de Doutorado, UFPE. Recife.
- Buso GSC, Amaral ZPS, Brondani RPV and Ferreira ME (2006) Microsatellite markers for the common bean *Phaseolus vulgaris*. *Mol Ecol Notes* 6:252-254777.

- Choi IY, Hyten, DL, Matukumalli LK, Song Q, Chaky JM, Quigley CV, Chase K, Lark KG, Reiter RS, Mun-Sup Y, *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 Reviews in Plant Sciences* 27:295-34.
- Cuadrado A, Schwarzacher T and Jouve N (2000) Identification of different chromatin classes in wheat using *in situ* hybridization with simple sequence repeat oligonucleotides. *Theor Appl Genet* 101:711–717.
- Cuadrado A, Cardoso M and Jouve N (2008) Physical organisation of simple sequence repeats (SSRs) in Triticeae: structural, functional and evolutionary implications. *Cytogenetics and Genome Research* 120:210–219.
- 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.
- Diouf D and Hilu KW (2005) Microsatellites and RAPD markers to study genetic relationships among cowpea breeding lines and local varieties in Senegal. *Genet. Resour. Crop Evol.* 52:1057-1067.
- Findley SD, Pappas AL, Cui L, Birchler JA, Palmer RG, Stacey G (2011) Fluorescence In Situ Hybridization–Based Karyotyping of Soybean Translocation Lines. *G3 Genes Genome Genetics*. V 1. 117-129.
- 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.
- Gartenberg M. (2009) Heterochromatin and the cohesion of sister chromatids. *Chromosome Research* 17:229-238.
- 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. *Annals of Botany* 106:505–514.
- He S, Wang Y, Volis S, Li D, Yi T (2012) Genetic Diversity and Population Structure: Implications for Conservation of Wild Soybean (*Glycine soja*

- Sieb. et Zucc) Based on Nuclear and Chloroplast Microsatellite Variation. International Journal of Molecular Sciences. 13 12608-12628.
- 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 and Shi M (1991) *In situ* hybridization with automated chromosome denaturation. Technique 3:109-115.
- 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
- Hyten DL, Choi IY, Song Q, Specht JE, Carter TE, Shoemaker RC, Hwang EY, Matukumalli LK and Cregan PB (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.
- Lichtenzweig J, Scheuring C, Dodge J, Abbo S and Zhang AB (2005) Construction of BAC and BIBAC libraries and their applications for generation of SSR markers for genome analysis of chickpea, *Cicer arietinum* L. Theor Appl Genet 110:492-510.
- 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.
- Mahmoud AA, Natarajan SS, Bennett JO, Mawhinney TP, Wiebold WJ and Krishnan HB (2006) Effect of six decades of selective breeding on soybean protein composition and quality: a biochemical and molecular analysis. Journal of Agricultural and Food Chemistry 54:3916-3922.
- Ministério da Agricultura e Pecuária (MAPA) <http://goo.gl/QPrkHL> (28 de Novembro 2013).
- Mun JH, Kim DJ, Choi HK, Gish J, Debellé F, Mudge J, Denny R, Endré G, Saurat O, Dutez AM, 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, 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. *Functional Plant Biology* 33:765-773.
- Oliveira EJ, Pádua JG, Zucchi MI, Vencovsky R and Vierira MLC (2006) Origin, evolution and genome distribution of microsatellites. *Genetics and Molecular Biology* 29:294-307.
- Palmer RG, Sun H and Zhao LM (2000) Genetics and cytology of chromosome inversions in soybean germplasm. *Crop Science* 40:683-687.
- Pedrosa A, Jantsch MF, Moscone EA, Ambros PF and Schweizer D (2001) Characterization of pericentromeric and sticky intercalary heterochromatin in *Ornithogalum longibracteatum* (Hyacinthaceae). *Chromosoma* 110:203-213.
- Schmutz J, Cannon SB, Schlueter JA, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, *et al* (2010) Genome sequence of the palaeopolyploid soybean. *Nature* 463:178-183.
- Singh RJ and Hymowitz T (1988) The genomic relationship between *Glycine max* (L.) Merr. and *G. soja* Sieb. and Zucc. as revealed by pachytene chromosome analysis. *Theor Appl Genet* 76:705-711.
- Singh RJ and Hymowitz T (1999) Soybean genetic resources and crop improvement. *Genome* 42:605-616.
- Song QJ, Marek L F, 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.
- 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.
- Tóth G, Gáspári Z and Jurka J (2000) Microsatellites in different eukaryotic genomes: Survey and analysis. *Genome Research* 10:967-981.
- Yu K, Park SJ, Poysa V and Gepts P (2000) Integration of simple sequence repeat (SSR) markers into a molecular linkage map of common bean (*Phaseolus vulgaris* L.). *The Journal of Heredity* 91:429-434.

Capítulo II

Distribuição do retrotransposon *Ty1-copia* em espécies de *Vigna* Savi

Pollyana Karla da Silva¹; Ana Christina Brasileiro-Vidal¹; Lidiane L. Barbosa Amorim; Ana Maria Benko-Iseppon¹

¹Universidade Federal de Pernambuco, Departamento de Genética, Laboratório de Genética e Biotecnologia Vegetal, Av. da Engenharia, s/nº, CEP 50740-600, Recife – PE, Brasil.

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

Resumo

A caracterização da organização e distribuição de Elementos Transponíveis (TEs) nos genomas de espécies vegetais tem ajudado ainda mais na compreensão dos fatores que têm contribuído para o funcionamento e remodelação dos genomas. O presente trabalho teve como objetivo identificar, mediante Hibridização *in situ* Fluorescente (FISH) a distribuição do retroelemento Ty1-*copia*-like em cromossomos de espécies de *Vigna aconitifolia*, *V. sesquipedalis* e *V. umbellata*. Na coloração CMA/DAPI as espécies apresentaram variação no número de pares de cromossomos satelitados. *Vigna aconitifolia* apresentou um par, *V. sesquipedalis*, três pares e *V. umbellata* um par todos nas regiões terminais. O padrão de distribuição do retrotransposon Ty1-*copia*-like utilizado como sonda na FISH revelou marcações dispersas ao longo de alguns cromossomos nas três espécies, bem como sinais de hibridização em regiões proximais de alguns dos cromossomos, com particularidades para cada espécie. Houve uma maior abundância do retroelemento Ty1-*copia*-like em regiões preferenciais, como heterocromatina. Provavelmente, as diferentes localizações cromossômicas devem ser atribuídas à alta taxa de mutação e à baixa pressão seletiva, associadas ao mecanismo de retrotransposição.

Palavras-chave: CMA; Heterocromatina; Retrotransposon; Ty1-*copia*; *Vigna*

Abstract

The characterization of the organization and distribution of Transposable Elements (TEs) in the genomes of plant species have helped to improve the understanding of the factors that have contributed to the operation and remodeling of genomes. This study aimed to identify, by Fluorescent in situ Hybridization (FISH) the distribution of the retroelement Ty1-*copia*-like in the chromosomes of the species *Vigna aconitifolia*, *V. sesquipedalis* and *V. umbellata*. In CMA/DAPI staining, species showed variation in the number of pairs of satellited chromosomes. *Vigna aconitifolia* presented a pair, *V. sesquipedalis*, three pairs and *V. umbellata* a pair all in terminal regions. The pattern of distribution of the retrotransposon Ty1-*copia*-like used as a probe in FISH revealed scattered distribution in some chromosomes of the three species, with hybridization signals in the proximal regions of some chromosomes, with special features for each species. There was a greater abundance of the retroelement Ty1-*copia*-like in some regions such as heterochromatin. Probably, the different chromosomal locations should be attributed to the high mutation rate and low selective pressure associated with the retrotransposition mechanism.

Keywords: CMA; heterochromatin; retrotransposon; Ty1-*copia*; *Vigna*

Introdução

O gênero *Vigna* compreende cerca de 170 espécies, distribuídas em áreas tropicais, sendo classificado como pertencente à ordem Fabales, família Fabaceae, subfamília Papilionioideae (Faboideae), tribo Phaseoleae e subtribo Phaseolina. (Freire-Filho *et al.*, 2005; Choi *et al.*, 2013). Os grãos de feijão destacam-se pela importância econômica dos seus representantes, constituindo a principal fonte de proteína vegetal na dieta humana dos países em desenvolvimento nas regiões tropicais e subtropicais, os quais são responsáveis por 86,7% do consumo mundial (Wander *et al.*, 2005).

Citogeneticamente, o gênero *Vigna* é diploide com $2n = 2x = 22$ (Sen e Bhowal 1960), apresentando cromossomos relativamente pequenos (1.6 - 3.7 μm) (Guerra *et al.*, 1996) e alta homogeneidade cariomorfológica (Venora e Saccardo, 1993; Venora e Padulosi, 1997; Venora *et al.*, 1999), características que dificultam as análises citológicas do grupo (Choi *et al.*, 2013). A espécie *Vigna unguiculata* (L.) Walp. (feijão-caupi) apresenta importância econômica e exibe um dos menores genomas entre as leguminosas (450 - 500 Mb) (Arumuganathan e Earle, 1991), com cromossomos predominantemente submetacêntricos e metacêntricos, de difícil identificação devido ao tamanho e grau de condensação cromossômica (Bortoleti, 2010).

Assim como todos eucariotos, seu genoma possui uma quantidade significativa de retroelementos. Estes, também denominados retrotransposons, são elementos transponíveis que apresentam um mecanismo de transposição do tipo “*copy and paste*” via RNA intermediário ao longo dos genomas. Este mecanismo envolve replicação e reintegração em sítios múltiplos através de um processo complexo constituído por: ativação da excisão, transcrição de RNA dependente de DNA, tradução de RNA em proteínas funcionais, síntese de DNA dependente de RNA e reintegração no genoma, gerando novas cópias de retroelementos (Heslop-Harrison, 2000).

Os retrotransposons representam uma importante fração do DNA moderadamente repetitivo, sendo seu mecanismo de transposição de modo replicativo apontado como principal fator responsável pela expansão e evolução dos genomas vegetais, juntamente com a poliploidização

(Todorovska, 2007; Du *et al.*, 2010). Além de seu acúmulo no genoma, a transposição dos retroelementos pode gerar mutações pela inserção de novas cópias dentro ou próximas a genes e causar rearranjos cromossômicos servindo como sítios para recombinação ectópica (Bennetzen, 2000). Acredita-se que estas mutações originam distintas famílias de retroelementos, as quais se diferenciam por suas atividades e distribuições nos cromossomos e/ou porções cromossômicas (Natali *et al.*, 2006).

Desta forma, a caracterização da organização e distribuição de TEs nos genomas pode incrementar nossa compreensão a respeito dos fatores que têm contribuído para o funcionamento e remodelagem dos genomas vegetais (Herrera *et al.*, 2013). Além disso, a elevada abundância e taxa de movimentação genômica têm permitido a utilização dos retroelementos na FISH (*Fluorescent In Situ Hybridization*; Hibridização *in situ* Fluorescente) para a caracterização de espécies proximamente relacionadas, contribuindo para o melhor entendimento das relações filogenéticas (Nielen *et al.*, 2009).

Devido ao fato de existir poucos trabalhos que caracterizem a distribuição citogenética de elementos retrotransponíveis em leguminosas cultivadas, o presente trabalho teve como objetivo identificar a distribuição do retroelemento Ty1-*copia*-like em espécies do gênero *Vigna* mediante FISH.

Material e Métodos

Sementes das espécies de *V. aconitifolia*, *V. sesquipedalis* e *V. umbellata* foram postas para germinar em placas de Petri com papel filtro e uma camada de algodão, ambos umedecidos. Pontas de raízes recém-germinadas foram pré-tratadas com uma solução de 8-hidroxiquinoleína (8HQ) a 2 mM por 4:30 h, à temperatura de 18 °C, fixadas em metanol:ácido acético (3:1, v/v) por 2 a 24 h à temperatura ambiente e estocadas a -20°C.

Preparo das lâminas

Os meristemas foram lavados duas vezes em água destilada por 5 min cada e digeridos em uma solução de 2% (p/v) celulase 'Onozuka R-10' (Serva)/ 20% (v/v) pectinase (Sigma-Aldrich) em tampão citrato 10 mM por 2:30 h a

37°C. Em seguida, as raízes foram lavadas em água destilada e colocadas em uma lâmina ligeiramente inclinada verticalmente. Algumas gotas de fixador (metanol: ácido acético; 3:1; v:v) gelado foram pingadas com uma pipeta Pasteur em cima das raízes para total remoção da água. Em seguida, procedeu-se a maceração mecânica das raízes e posterior gotejamento de fixador gelado sobre o material. Posteriormente, as lâminas foram ventiladas com um auxílio de bombas de ar possibilitando 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. As lâminas foram coradas com uma solução de 2 µg/mL de DAPI (4'-6'-diamidino-2-fenilindole)/ glicerol (1:1, v/v), a fim de selecionar as melhores lâminas. As lâminas selecionadas foram descoradas em etanol:ácido acético (3:1, v/v) por 30 min e mantidas etanol absoluto durante 1 h. Após secagem ao ar, as lâminas foram devidamente estocadas a -20°C, até o momento do procedimento de hibridização *in situ*.

Dupla coloração com fluorocromos CMA e DAPI

Na coloração CMA/DAPI as lâminas foram armazenadas por três dias a temperatura ambiente, coradas com CMA (0,5 mg/mL) por 60 min e, em seguida, com DAPI (2 µg/mL), por 30 min. Posteriormente, as lâminas foram secas, montadas com meio de montagem de tampão McIlvaine's (pH 7.0):glicerol (1:1, v/v) e lutadas com esmalte. Após três dias, imagens das melhores células foram capturadas.

Hibridização fluorescente *in situ* (FISH)

As lâminas foram pré-tratadas como descrito por Pedrosa *et al.* (2001). A desnaturação dos cromossomos e das sondas, os banhos pós-hibridização e a detecção foram efetuados de acordo com Heslop-Harrison *et al.* (1991), exceto pela lavagem de estringência, realizada em 0,1x SSC 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 70°C e hibridizadas de 18 a 48 h a 37°C. A sonda utilizada do retrotransposon do

domínio RT de Ty1- *copia*-like foi previamente preparada e estocada no Laboratório de Genética e Biotecnologia Vegetal da UFPE, seguindo a metodologia descrita por Bortoleti (2010). A sonda foi marcada com digoxigenina e detectada usando *sheep anti-digoxigenin* conjugado com fluoresceína isotiocianato (FITC; Roche) e amplificadas com *anti-sheep* conjugado com FITC (Roche) 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 em microscópio de epifluorescência Leica DMLB e imagens das melhores células capturadas com uma câmera Leica DFC 340FX, usando o programa Leica CW 4000. As imagens foram otimizadas para melhor brilho e contraste com o Adobe Photoshop CS4 (Adobe Systems Incorporated).

Resultados e Discussão

A análise da distribuição cromossômica torna-se importante 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). As espécies *V. aconitifolia*, *V. sesquipedalis* e *V. umbellata* apresentaram variação no número de pares de cromossomos satelitados. Na coloração por fluorocromos base-específicos CMA/DAPI, *V. aconitifolia* apresentou apenas um par portador de uma banda CMA positiva (CMA⁺) terminal e cinco pares com marcações pericentroméricas visíveis. Em *V. sesquipedalis*, blocos CMA⁺ foram observados em regiões pericentroméricas de alguns cromossomos (dois pares) e três pares nas regiões terminais, sendo o terceiro par com sinal muito fraco, visualizado melhor em outras células (dados não mostrados). Na espécie *V. umbellata*, também foram observados sinais pericentroméricos em dois pares e um par com marcação terminal nos cromossomos satelitados (Figura 1A-C).

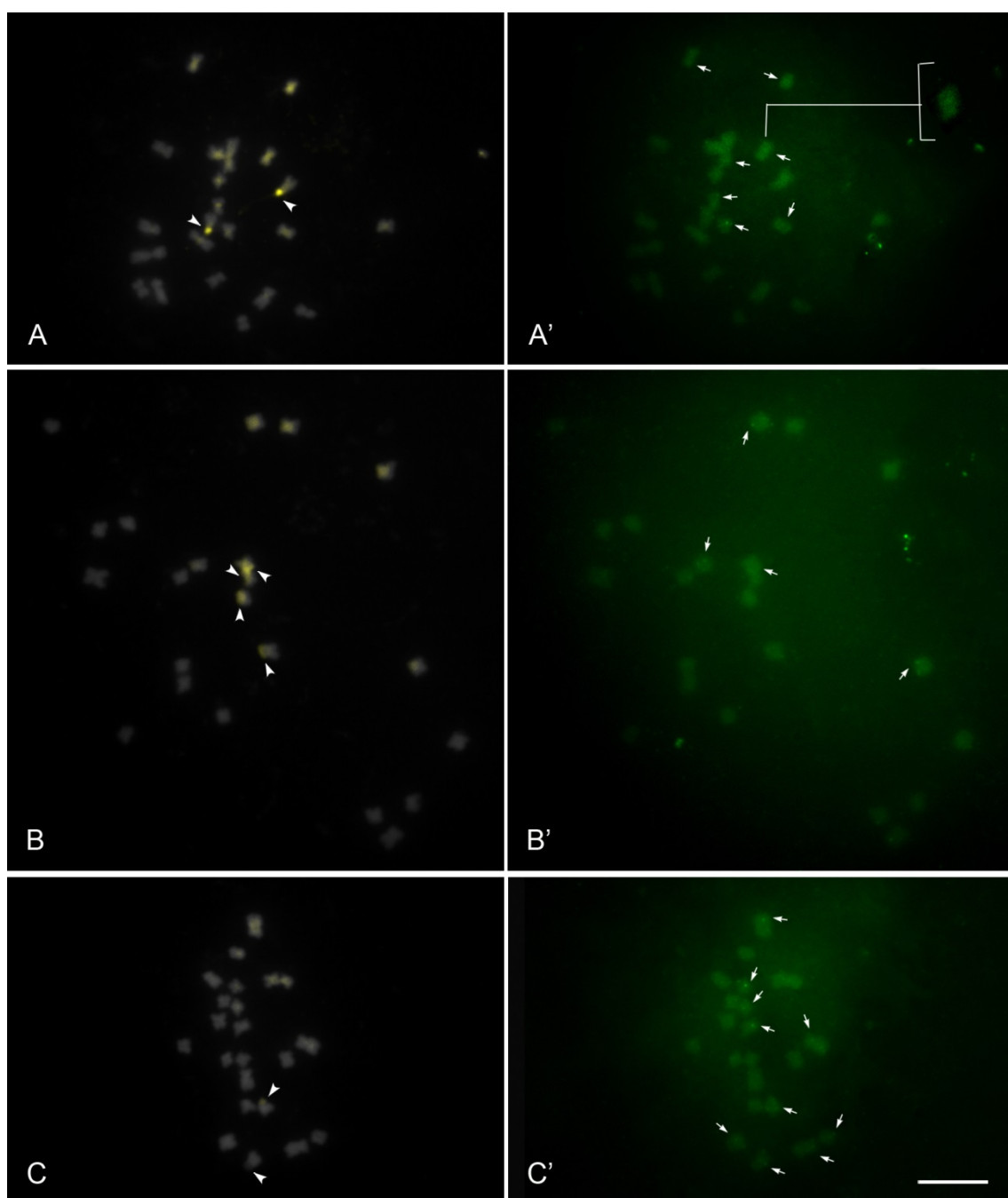


Figura 1: Coloração CMA/DAPI e hibridização *in situ* fluorescente do retrotransposon Ty1-*copia*-like em cromossomos mitóticos de *Vigna aconitifolia* (A, A'), *V. sesquipedalis* (B, B') e *V. umbellata* (C, C'). Metáfases contendo $2n = 22$ cromossomos contraincoradas com DAPI (pseudocolorido em cinza), com blocos CMA positivos (amarelo) nas regiões pericentroméricas dos cromossomos e nas regiões terminais dos cromossomos satelitados (A-C). Marcações da sonda Ty1-*copia* (A'-C'), em verde. As cabeças de seta em A-C apontam bandas CMA⁺, dos cromossomos satelitados e as setas em A'-C' mostram marcações com Ty1-*copia* mais evidentes. A barra em F' corresponde a 10 μ m.

Após hibridação do retrotransposon Ty1-*copia*-like utilizado como sonda na FISH observaram-se marcações dispersas ao longo dos cromossomos de *V. aconitifolia*, *V. sesquipedalis* e *V. umbellata*, bem como sinais de hibridização nas regiões proximais dos cromossomos, com particularidades para cada espécie Figura 1 (A' - C').

Em *V. aconitifolia*, foram observadas marcações dispersas, com sítios mais evidentes na região proximal e/ou subterminal ao menos sete cromossomos indicados pelas setas (Figura 1A'). Na espécie *V. sesquipedalis*, foram evidenciadas marcações subterminais em ao menos dois pares cromossomos em um ou ambos os braços, indicados pelas setas (Figura 1B'). Os sinais de hibridização em *V. umbellata* foram dispersos discretos, porém marcações proximais e subterminais foram evidenciadas, em ao menos 10 cromossomos apontados por setas (Figura 1C').

No presente estudo, os sinais dispersos e proximais encontrados revelaram a ocorrência do elemento móvel Ty1-*copia*-like ao longo dos cromossomos de *V. aconitifolia*, *V. sesquipedalis* e *V. umbellata*. Em outras espécies de *Vigna* (*V. unguiculata* e *V. radiata*), pequenos *dots* bem definidos foram visualizados nas regiões intercalares e subteloméricas dos cromossomos, como também marcações proximais em menor quantidade (Bortoleti, 2010). Esses pequenos *dots* também foram observados em regiões similares nas espécies de *Vigna* analisadas neste trabalho.

Os retrotransposons apresentam normalmente 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; Bortoleti, 2010). Em resultados similares, utilizando o retrotransposon Ty1-*copia*-like como sonda, o padrão de distribuição disperso também foi descrito por Bortoleti (2010) nas espécies de *G. max*, *P. vulgaris*, *P. lunatus*, *V. unguiculata* e *V. radiata*; como também descrito para as espécies *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).

Resultados semelhantes aos observados em *Vigna* no presente estudo e por (Bortoleti, 2010) foram também encontrados na FISH realizada com o objetivo de localizar sequências de Ty1-*copia*-like no genoma de *Coffea arabica* e suas espécies ancestrais, revelando uma distribuição dispersa ao longo dos cromossomos, como algumas áreas maior concentração, fornecendo novas evidências sobre o retroelemento no genoma do café (Herrera et al., 2013)

Trabalhos que visam caracterizar as sequências de Ty1-*copia*-like RT revelaram uma grande heterogeneidade (Hafez e Zaki, 2005; Bortoleti 2010). Tais dados sugerem que a citada 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). Em um estudo de análise molecular com algodão egípcio (*Gossypium barbadense* L.), Hafez e Zaki (2005) verificaram que o grupo dos retrotransposons Ty1-*copia* ultrapassa os limites da espécie, sugerindo que existia no início da evolução das plantas, e se divergiu em sequências heterogêneas antes da divergência de angiospermas atuais.

Nas espécies de *Vigna* aqui analisadas (*V. aconitifolia*, *V. sesquipedalis* e *V. umbellata*), um padrão de distribuição disperso foi notado para o retroelemento Ty1-*copia*-like, mostrando uma clara associação 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. Dimitri et al. (1999) relataram a presença de retrotransposons em regiões heterocromáticas, 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).

Em suma, uma distribuição significativa do retroelemento Ty1-*copia*-like foi notada em regiões preferenciais, incluindo as heterocromáticas. Porém, ocorre uma distribuição diferenciada entre as espécies estudadas. Provavelmente, as diferentes localizações cromossômicas 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 ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e à Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (Facepe) pelo apoio financeiro ao projeto.

Referências Bibliográficas

- Arumuganathan K and Earle ED (1991) Nuclear DNA content of some important plant species. *Plant Mol Biol Rep* 9: 208-218.
- Bennetzen JL (2000) Transposable element contributions to plant gene and genome evolution. *Plant Mol Biol* 42: 251–269.
- Bortoleti KCA (2010) Mapeamento cromossômico comparativo em *Glycine* Willd, *Phaseolus* L. e *Vigna* Savi. Tese de Doutorado, Universidade Federal de Pernambuco, Recife, 153 p.
- Choi HW, Kim MY, Lee SH, Sultana S and Bang JW (2013) Molecular cytogenetic analysis of the *Vigna* species distributed in Korea. *Genes & Genomics* 35:257–264.
- Dimitri P, Junakovic N (1999) Revising the selfish DNA hypothesis. New evidence on accumulation of transposable elements in heterocromatin. *Trends Genet* 15: 123-124.
- Du J, Tian Z, Hans CS, Laten HM, Cannon SB, Jackson SA, Shoemaker RC, 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.

- 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, 28-92 pp.
- 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.
- 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. Annals of Botany 78:157-161.
- Hafez EE, Zaki EA. (2005) Phylogenetic and molecular evolutionary analyses of Ty1-copia group retrotransposons in cultivated Egyptian cotton, *Gossypium barbadense* L. African Journal of Biotechnology Vol. 4 (11), pp. 1275-1280.
- Herrera JC, Camayo G, De-La-Torre G, Galeano N, Salcedo E, Rivera LF, Duran A. (2013) Identification and chromosomal distribution of copia-like retrotransposon sequences in the coffee (*Coffea* L.) genome. Agronomía Colombiana 31(3), 269-278.
- 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 and Shi M (1991) *In situ* hybridization with automated chromosome denaturation. 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.
- Lamb JC, Birchler JA (2006). Retroelement Genome Painting: Cytological Visualization of Retroelement Expansions in the Genera *Zea* and *Tripsacum*. *Genetics* 173: 1007–1021.
- 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.
- Pedrosa A, Jantsch MF, Moscone EA, Ambros PF and Schweizer D (2001) Characterization of pericentromeric and sticky intercalary heterochromatin in *Ornithogalum longibracteatum* (Hyacinthaceae). *Chromosoma* 110:203-213.
- Sen NK and Vidyabhusan RV (1960) Tetraploid soybeans. *Euphytica* 9:317–322.
- Todorovska E (2007) Retrotransposons and their role in plant – genome evolution. *Biotechnol. & Biotechnol. EQ* 21: 294–305.
- 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.

- Wander AL, Didonet AD, Abreu AFB, Silva CC, Quintela ED, Fonseca JR, Silva JG, Stone LF, Ramalho MAP, Junior ML, *et al* (2005) Cultivo do feijão irrigado na região noroeste de Minas Gerais. Embrapa Arroz e Feijão.
- 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.

Capítulo III

Identificação e Caracterização de domínios da superfamília CACTA em *Glycine max*

Bruna Piareck Moura¹, Pollyana Karla da Silva¹, João Pacifico Bezerra Neto¹,
Sheyla Silva¹, Ana Christina Brasileiro-Vidal¹ e Ana Maria Benko-Iseppon^{1*}

¹ Universidade Federal de Pernambuco, Departamento de Genética,
Laboratório de Genética e Biotecnologia Vegetal, Av. da Engenharia, s/nº, CEP
50740-600, Recife – PE, Brasil.

* Autor para correspondência: ana.iseppon@gmail.com

Resumo

Elementos transponíveis (TEs) são elementos móveis encontrados na maioria dos genomas eucariotos. Matéria-prima de variabilidade genética e propulsores evolutivos, eles exercem influência tanto na estrutura quanto na expressão gênica, tornando sua identificação etapa crucial para compreensão de sua influência na evolução de genes e genomas e no entendimento de seus mecanismos de ação. Em soja, 65 sequências de CACTA foram descritas até o momento. Por tratar-se de elemento conhecido por sua interação com regiões ricas em genes com abrangência em diversos genomas, uma nova busca, baseada em homologia foi realizada, identificando-se duas enzimas necessárias para a transposição deste elemento: Tnp1 e Tnp2, totalizando 2,2 e 10 Mb, respectivamente. A busca revelou também domínios de outras transposases e domínios não relacionados diretamente ao elemento CACTA. Os resultados revelam uma destacada abundância e diversidade desse elemento em soja.

Palavras-Chave: CACTA, Domínios Tnp1 e Tnp2, Elementos transponíveis Soja.

Abstract

Transposable elements (TEs) are mobile elements found in most eukaryotic genomes. Raw material of genetic variability and evolutionary drivers, they influence both gene structure and expression, making their identification a crucial step for understanding their influence on the evolution of genes and genomes as well as their acting mechanisms. In soybean 65 CACTA elements were described so far. Because its interaction with regions rich in genes with coverage in various genomes, a new search based on homology was carried out, and identified two enzymes necessary for transposition of this element: Tnp1 and Tnp2, totalizing 2.2 Mb and 10 Mb, respectively. The search also revealed areas of transposases and other areas not directly related to CACTA element. The results revealed a prominent abundance and diversity of this element in soybean.

Key-words: CACTA, Soybean, transposable elements, Tnp1 and Tnp2 domains.

Introdução

Elementos transponíveis (TEs) são sequências genômicas com capacidade de se movimentar dentro de um mesmo cromossomo ou entre diferentes cromossomos de um genoma. Apresentam como mecanismo básico sua excisão e inserção em regiões diferentes daquela em que se encontram originariamente, podendo alterar a informação e a estrutura genética de diversas formas. Descobertos em milho por Barbara McClintock nos anos 1950, TEs são hoje descritos como um dos componentes mais abundantes dos genomas eucariotos, compreendendo poderosa força evolutiva e geradora de diversidade. Diversificados e dispersos em todos os reinos, são especialmente abundantes em plantas, embora sejam observados em quantidades generosas também em metazoários e fungos (Biémont e Vieira 2006; Wicker et al., 2007).

Tais elementos são classificados em duas grandes classes: Classe I (retrotransposons) e Classe II (transposons de DNA ou transposons). Cada uma destas classes permite sua classificação em níveis hierárquicos menores: subclasse, ordem, superfamília, família e subfamília devido à sua diversidade. Os retrotransposons são caracterizados por moverem-se via sequências de RNA, com auxílio de uma transcriptase reversa e uma transposase, inserindo uma cópia em outra região do genoma. A Classe dos transposons, ao contrário, baseia a sua dinâmica na movimentação direta das sequências de DNA que, com ação de uma ou mais transposases (TNPs), são retiradas e inseridas em outras regiões. Transposons podem com frequência carregar fragmentos de diversas sequências, dependendo da superfamília, podendo em casos raros também duplicar-se. Os mais conhecidos por esse feito são os transposons da superfamília Mutator, conhecidos como Pack-Mules, no entanto elementos da superfamília CACTA e Helitron também já foram observados capturando sequências (Jiang et al., 2004; Wicker et al., 2007; Li et al., 2009).

Os impactos acarretados pela mobilidade desses elementos tornam-se ainda maiores quando acompanhados pela captura de fragmentos, gerando sequências quiméricas, que podem acarretar na formação de novos quadros abertos de leitura (ORFs – *Open Reading Frames*). Caso os novos produtos gênicos sejam funcionais e vantajosos, podem em longo prazo se fixar e se estabelecer evolutivamente na espécie. Além da sua mobilidade, o impacto

evolutivo desses elementos pode ser ainda mais abrangente, interferindo em mecanismos epigenéticos. Tal ação foi observada a partir de estudos demonstrando que a localização e/ou a quantidade de DNA de transposons em um organismo pode estar relacionada com a maior ou menor expressão de alguns genes, observando-se que determinadas superfamílias podem estar mais presentes em plantas que apresentam maior resistência a certos tipos de estresse (Jiang et al., 2004; Li et al., 2009; Cantu et al., 2010; Chénais et al., 2012).

Os elementos da superfamília CACTA destacam-se entre os mais relacionados a regiões ricas em genes, podendo exercer maior influência sobre sua estrutura ou expressão. No entanto, nem sempre as alterações geradas são benéficas e os organismos por vezes silenciam tais regiões possivelmente através de mecanismos que minimizam fatores deletérios. Tal ação epigenética ocorre especialmente por metilação, inibindo a atividade desses elementos (Cantu et al., 2010; Saze et al., 2012). São denominados CACTA por possuírem um motivo conservado de cinco nucleotídeos (nt) 'CACTA' em suas repetições terminais invertidas (TIRs, *Terminal Inverted Repeats*), embora sejam também conhecidos por En/Spm, nome dos primeiros elementos descobertos por Peterson e McClintock de forma independente. Esta superfamília se caracteriza por apresentar TIRs com até 30 nt, duplicações de sítio alvo de três nt e regiões subterminais (Sub-TR) relativamente bem conservadas. As enzimas Tnp1 e Tnp2, homólogas das enzimas TnpA e TnpD, são duas transposases consideradas essenciais à transposição nesta superfamília de elementos, os quais tornam-se autônomos apenas na presença de ambas as enzimas (Figura 1) (Masson et al., 1991; Saze et al., 2012; Takahashi et al., 2012).



Figura 1: Imagem esquemática do elemento CACTA. As ampulhetas laterais representam as duplicações de sítio alvo. Setas em preto representam as Repetições Terminais Invertidas (TIRs, *Terminal Inverted Repeats*). Cabeças de seta prateadas representam regiões subterminais. No centro as duas transposases, Tnp1 e Tnp2. Esquema dos autores.

Tendo em vista tantos aspectos relacionados a esses elementos e sua relação com os genomas, a identificação, caracterização e quantificação de seus elementos é essencial para compreensão das dimensões de sua influência nos genomas das espécies. Para a Soja, foram descritos até o momento 65 elementos da superfamília CACTA, sabendo que este grupo de sequência está entre os mais abrangentes da Classe II. O presente trabalho reavaliou a prevalência e abundância de tais elementos, valendo-se de novas ferramentas bioinformáticas para a identificação de elementos desta família de forma mais abrangente e eficiente, atualizando sobre a quantidade destes em soja.

Metodologia

Busca por sondas

Foi realizado um levantamento bibliográfico com o objetivo de identificar sequências da Classe II, da superfamília CACTA, para serem utilizadas como sondas (*seed sequences*). Assim, o banco de sequências que inclui todas as superfamílias identificadas no genoma da soja (<http://www.soybase.org/soytedb/>) foi utilizado para a obtenção das sondas do presente trabalho, totalizando 65 sequências referentes à superfamília CACTA (Du et al., 2010), listadas na Tabela Suplementar 1 (Tabela S1).

Mineração de Sequências Candidatas

As sondas selecionadas foram ancoradas contra o banco de dados de sequências genômicas acesso de restrito Soybean Genome Project (<http://bioinfo03.ibi.unicamp.br/soja/>), mediante ferramenta BLASTn, utilizando como ponto de corte (*cut-off*) o e-value de e^{-4} , contra '*Chromosomes and scaffold sequences*' e resultado em '*Hit table*'.

Recuperação e tradução das sequências

Com as ferramentas gotDNA+.java (Java script), utilizando as coordenadas obtidas após ancoragem contra o genoma de soja, as sequências alvo foram recuperadas e salvas em um banco de dados local. Para a tradução das sequências nos seis frames, foi utilizado o script

“seqs_processor_and_translator_bin_V118_AGCT.py”, disponível em linguagem Python (The Compositae Genome Project, <http://cgpdb.ucdavis.edu>).

Caracterização

As sequências foram caracterizadas mediante ferramenta CD-Search Batch disponível no NCBI, alterando o número de ‘*Maximum number of hits*’ para 1000 e resultados curados manualmente, avaliando-se a presença do domínio das transposases e sua integridade. Adicionalmente foi observada a possível presença de domínios de outras proteínas por meio de buscas baseadas em homologia de sequências. As etapas dos procedimentos adotados e seus resultados estão ilustrados de forma esquemática na Figura 2.

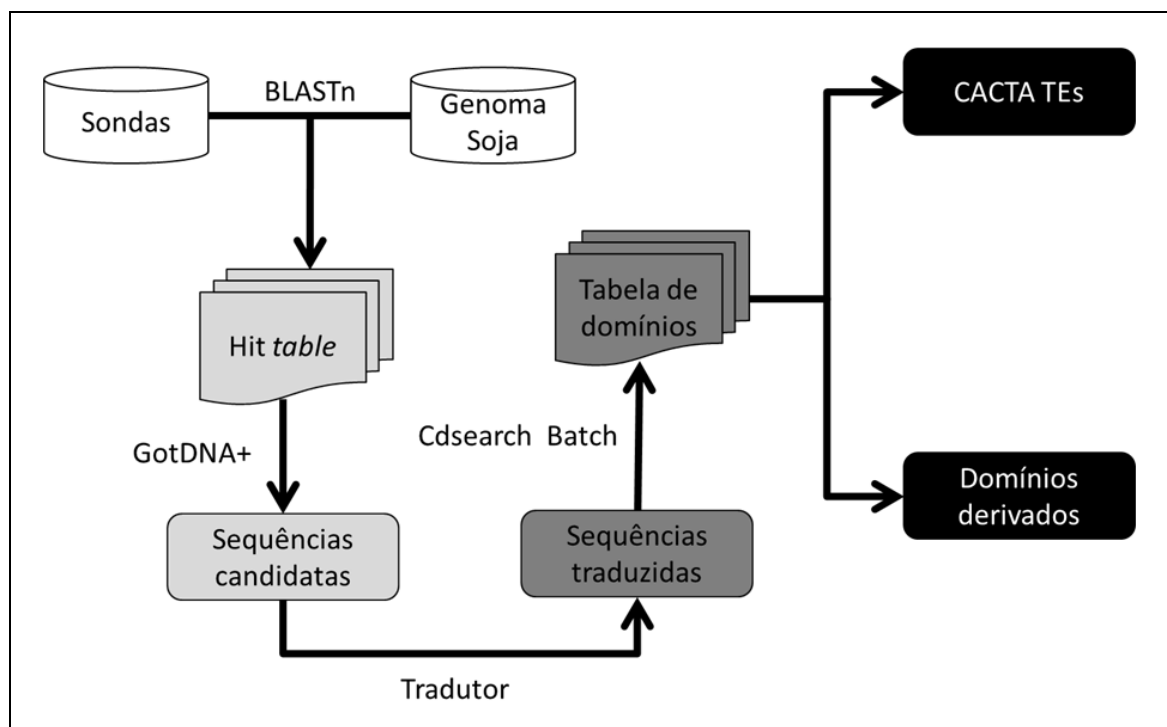


Figura 2. Fluxograma ilustrando as principais etapas metodológicas realizadas. Em branco encontram-se as etapas de identificação, seguidas pela tradução dos candidatos em cinza claro. Na cor cinza escuro encontram-se as etapas de caracterização e destacada em preto, a análise visual dos resultados finais permitindo sua classificação em dois grupos.

Resultados e Discussão

As análises propostas identificaram 870.850 sequências que ao serem caracterizadas permitiram a descrição de aproximadamente 380 domínios diferentes. Dentre estes se destacaram 15,21 Mb de sequências relacionadas às transposases, com 2,2 Mb distribuídos em 19.293 sequências continham o domínio da Tnp1, sendo 11.310 completos (Figura 3). A transposase Tnp1 foi caracterizada em *Antirrhinum majus* L. (família Plantaginaceae, Dicotyledoneae) no elemento *tam1*, sendo homóloga a TnpA, transposase característica do elemento *Spm*, uma das transposases essenciais à mobilidade do elemento CACTA, responsável por reconhecer as regiões subterminais e capaz de desmetilar sua própria região promotora mostrando-se como participante ativa da regulação desta superfamília (Masson et al., 1991; Nacken et al., 1991; Saze et al., 2012).

Juntamente com a TnpA, a TnpD (homóloga da Tnp2) é também crucial para a mobilidade desta superfamília e, estando ambos os domínios completos e em tandem no genoma da soja, é possível inferir que se trata de um elemento autônomo e completo (Masson et al., 1991). No presente trabalho, a Tnp2 foi observada em 10,6 Mb do genoma, compreendendo 90.876 domínios, dos quais 25.037 domínios completos (Tabela 1), destacando-se como o domínio relacionado à mobilidade de CACTA mais observado em soja.

Ambas as proteínas foram observadas em todos os cromossomos, sendo ainda necessárias análises adicionais sobre sua distribuição na estrutura dos elementos para inferir sobre a autonomia (ou não) dos referidos TEs. Adicionalmente, observou-se que 0.07 Mb do genoma continha sequências com o domínio Transposase_24 (Tnp24), referente à família de transposons CACTA-like subfamília Ptta, com 1.385 sequências, sendo 19 delas com domínio completo (Tabela 1) (Masson et al., 1991; Nacken et al., 1991).

Além das sequências descritas anteriormente, foram encontradas nove sequências relativas ao domínio DDE_Tnp, distribuídas nos cromossomos 11, 18 e 14, somando 0,0028 Mb, com apenas uma sequência completa localizada no cromossomo 11 de soja. Um total de oito sequências corresponderam à transposase DDE_Tnp_IS66 de *Agrobacterium tumefaciens*, possivelmente fixada após transferência horizontal de genes. O motivo DDE está presente em

diversas famílias de transposases, responsável por catalisar a clivagem do DNA, sendo caracterizado pela presença de dois resíduos de ácido aspártico (D) e um de ácido glutâmico (E). Outro domínio presente é o Dimer_Tnp_hAT, encontrado em todos os cromossomos de soja. Embora seja um domínio característico da transposase de outra superfamília da Classe II, foi observado com e-value variando entre $9,26e$ e $7,33e^{-22}$, demonstrando uma similaridade significativa, com suas 80 sequências, somando 0,1Mb e 55 domínios completos (Tabela 1) (Rubin et al., 2001; Yuan e Wessler 2011).

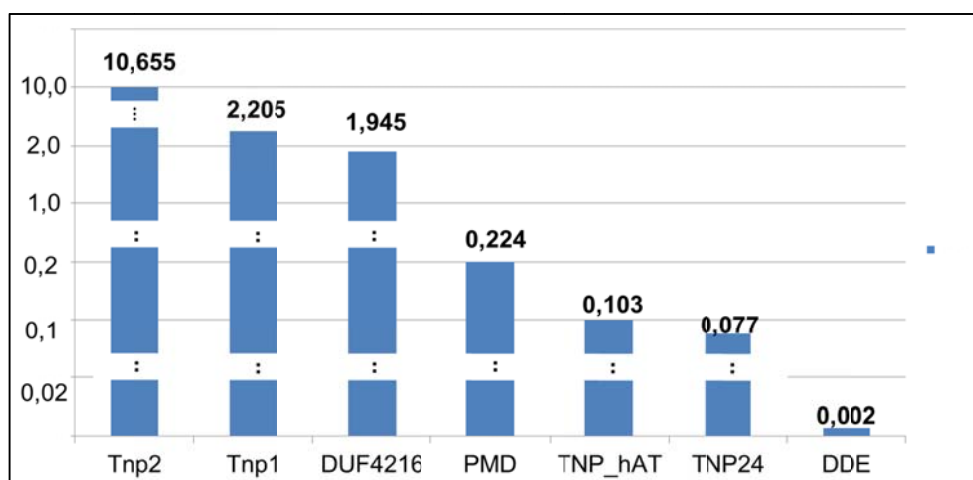


Figura 3: Gráfico representativo da quantidade de domínios relacionados à transposase identificados em CACTA. Valores em Megabases (Mb) no genoma de soja (*Glycine max*).

Identificados a partir de sondas estritamente referentes ao elemento CACTA, a presença destes dois domínios (DDE_tnp e Dimer_Tnp_hAT) pode indicar a captura de elementos que formam mosaicos com elementos CACTA, dado que o motivo DDE, embora representativo para diversas superfamílias, é ambíguo para a superfamília CACTA, não sendo característico para a mesma, face à sua maior associação a sequências da superfamília hAT. Outra justificativa possível é que estes representem transposases com origem comum. No entanto, análises mais acuradas devem ser realizadas para maior segurança sobre o papel das associações observadas desses elementos na soja (Yuan e Wessler 2011).

Tabela 1: Domínios de transposases identificados na soja, incluindo domínios completos (-), domínio incompleto na região C-terminal (C), domínio incompleto na região N-terminal (N) e domínios incompletos nas duas regiões (NC).

Região	TNP2		TNP1		TNP_24		DDE		Dimer_Tnp_hAT		DUF4216		PMD	
	Nº absoluto	Mb	Nº absoluto	Mb	Nº absoluto	Mb	Nº absoluto	Mb	Nº absoluto	Mb	Nº absoluto	Mb	Nº absoluto	Mb
-	25037	5,1381	11310	0,7372	19	0,0033	1	0,00031	55	0,0440	23071	1,6245	82	0,0323
C	32709	3,3193	81	0,0039	199	0,0106	0	0	8	0,0043	3210	0,1470	762	0,1130
N	19072	1,4880	7868	0,0039	660	0,0471	0	0	15	0,0056	5428	0,1712	623	0,0692
NC	14058	0,7095	33	0,3598	507	0,0165	8	0,0025	2	0,0005	124	0,00307	122	0,0097
Total	90876	10,6550	19293	2,2055	1385	0,0776	9	0,00281	80	0,1030	31833	19,0189	1589	0,2243

Em nosso estudo, o domínio PMD (*Plant Mobile Domain*) também foi identificado no genoma de soja em 0,2 Mb de sequências, incluindo 1.589 domínios, dos quais 82 encontravam-se completos (Tabela 1). Babu et al. (2006) reportaram a existência de uma relação entre o domínio de ligação ao DNA (DBD, *DNA Binding Domain*), fatores de transcrição e elementos transponíveis, observando-se que esta última categoria apresentava similaridade com sequências de Mutator-like e de TnpA. Tal similaridade com a TNPA justifica a identificação, dado que a metodologia de busca foi baseada em homologia de sequência. O referido artigo aborda a associação das transposases com DBD como um dos responsáveis pela neofuncionalização de fatores de transcrição. Concordando com isso Aziz et al. (2010) sugerem a importância da transposase como propulsora da variabilidade e da evolução dos organismos independentemente dos elementos transponíveis (Babu et al., 2006).

Aziz et al. (2010) constataram que as transposases são os elementos mais prevalentes nos genomas de eucariotos, tanto por sua abrangência em todos os metagenomas estudados (considerando-se como metagenoma o coletivo dos genomas em um ecossistema específico), como pela abundância em que são encontrados nos metagenomas disponíveis e nos genomas sequenciados. Ainda de acordo com os autores não existem evidências para acreditar em uma função essencial para subsistência de tais elementos nos organismos, apesar de sua abundância e distribuição. Porém há diversas evidências de que esses elementos são ricas fontes de diversidade genética (Aziz et al., 2010). Por exemplo, o domínio DUF4216 (que não tem função conhecida) foi observado na região C-terminal de algumas proteínas da soja que carregam o domínio Tnp1. O domínio DUF4216 ocupa 1,9 Mb do genoma da soja, com 31.833 sequências, sendo 23.071 com domínio completo. Tais proteínas podem derivar de genes associados a uma transposase, adquirindo uma nova função ainda desconhecida, de maneira semelhante ao descrito por Babu et al. (2006). Além disso, o citado domínio chama atenção por sua abrangência e distribuição dispersa em todos os cromossomos em soja (Babu et al., 2006; Aziz et al., 2010).

Apesar da óbvia associação das transposases aos elementos transponíveis, em especial aos elementos de Classe II, elas já foram

encontradas influenciando uma variada gama de genes, sem necessariamente transpô-los. Considera-se que tais transposases podem atuar como potencializadoras da expressão de alguns genes, podendo ainda desmetilar regiões promotoras, dessa forma possibilitando ou até modificando a expressão gênica (Lin et al., 2007; Aziz et al., 2010). No presente estudo, não apenas domínios relacionados a transposases foram identificados, mas um total de 342 domínios diferentes foram identificados, variando em número de sequências entre 1 e 1601 sequências por domínio, com destaque para os cinco tipos de domínios mais observados: *cbiO*, *gluc_glyc_Psyn*, *bacterial_phosphagen_kin*, *ccmA* e *CydD* (Figura 4). Além destes, 29 domínios com função desconhecida foram anotados, com número de sequências variando entre 10 e 46.324.

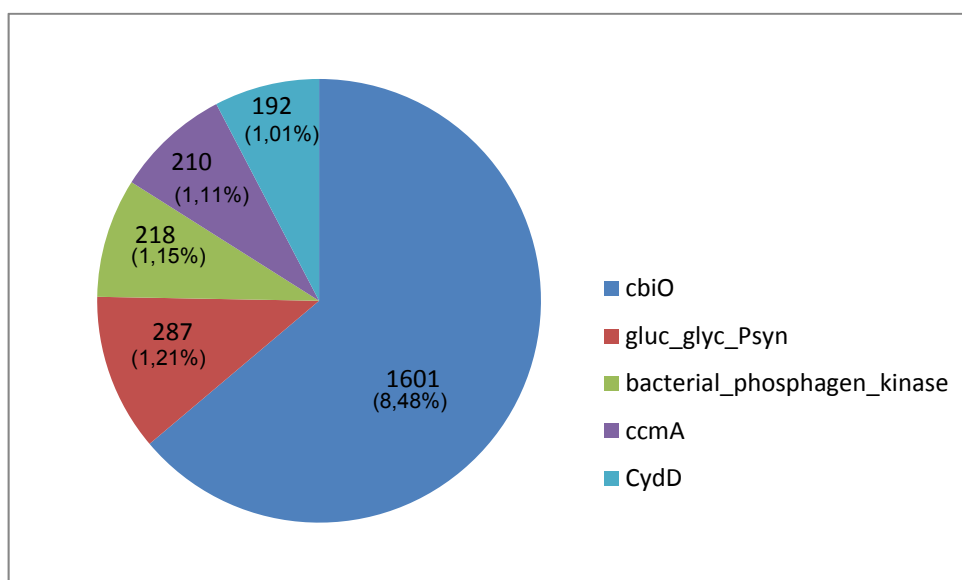


Figura 4: Gráfico representativo do número absoluto das cinco principais categorias de domínios encontradas no genoma da soja dentre aquelas não relacionadas à transposase, bem como sua porcentagem em relação ao total de sequências não relacionadas.

Assim como a influência das transposases não se resume aos elementos transponíveis, TEs também podem alterar a expressão em nível pós-transcricional. Estudos recentes mostram que parte considerável (aproximadamente 1/6) dos pequenos RNAs (sRNAs, *small RNAs*) conhecidos na atualidade deriva de transposons. Portanto, tal influência vai além do controle epigenético da transposição, influenciando outros genes que

apresentem similaridade de sequência, sendo inclusive possível prever alvos de microRNAs (miRNAs) cujo elemento transponível de origem seja conhecido. Portanto, todas as sequências identificadas apresentando homologia com TEs podem ser não apenas fragmentos capturados por eles com capacidade de gerar novas ORFs, mas também potenciais alvos de sRNAs, enfatizando a importância de desenvolvimento de ferramentas sensíveis para catalogação desses elementos (Lin et al., 2007; Cantu et al., 2010; Saze et al., 2012).

Destaca-se que o presente estudo permitiu a detecção de um número muito maior de TEs da superfamília CACTA e seus fragmentos do que o observado em estudos anteriores. A caracterização de toda a diversidade de elementos, especialmente para grupos como CACTA, conhecido por sua relação com regiões ricas em genes, apresenta-se como o primeiro passo para o entendimento de sua evolução e papel funcional.

Referências Bibliográficas

- Aziz RK, Breitbart M, Edwards RA (2010) Transposases Are the Most Abundant, Most Ubiquitous Genes in Nature. *Nucleic Acids Research* 38 (13): 4207–17.
- Babu MM, Iyer LM, Balaji S, Aravind L (2006) The Natural History of the WRKY-GCM1 Zinc Fingers and the Relationship between Transcription Factors and Transposons. *Nucleic Acids Research* 34 (22): 6505–20.
- Biémont C and Vieira C (2006) Genetics: Junk DNA as an Evolutionary Force. *Nature* 443 (7111): 521–24.
- Cantu D, Vanzetti LS, Sumner A, Dubcovsky M, Matvienko M, Distelfeld A, Michelmore RW, J (2010) Small RNAs, DNA Methylation and Transposable Elements in Wheat. *BMC Genomics* 11 (January): 408.

- Chénais B, Caruso A, Hiard S, Casse N(2012) The Impact of Transposable Elements on Eukaryotic Genomes: From Genome Size Increase to Genetic Adaptation to Stressful Environments. *Gene* 509 (1): 7–15.
- Du J, Grant D, Tian Z, Nelson RT, Zhu L, Shoemaker RC, Ma J (2010) SoyTEdb: A Comprehensive Database of Transposable Elements in the Soybean Genome. *BMC Genomics* 11 (January): 113.
- Jiang N, Bao Z, Zhang X, Eddy SR, Wessler SR (2004) Pack-MULE Transposable Elements Mediate Gene Evolution in Plants. *Nature* 431 (7008): 569–73.
- Li Q, Li L, Dai JR, Li JS, Yan JB (2009) Identification and Characterization of CACTA Transposable Elements Capturing Gene Fragments in Maize. *Chinese Science Bulletin* 54 (4): 642–51.
- Lin R, Ding L, Casola C, Ripol DRI, Feschotte C, Wang H (2007) Transposase-Derived Transcription Factors Regulate Light Signaling in *Arabidopsis*. *Science (New York, N.Y.)* 318 (5854): 1302–5.
- Masson P, Strem M, Fedoroff N (1991) The tnpA and tnpD Gene Products of the Spm Element Are Required for Transposition in Tobacco. *The Plant Cell* 3 (1): 73–85.
- Nacken WK¹, Piotrowiak R, Saedler H, Sommer H (1991) The Transposable Element Tam1 from *Antirrhinum Majus* Shows Structural Homology to the Maize Transposon En/Spm and Has No Sequence Specificity of Insertion. *MGG Molecular & General Genetics* 228 (1-2): 201–8.
- Rubin E, Lithwick G, Levy AA (2001) Structure and Evolution of the hAT Transposon Superfamily. *Genetics* 158 (3): 949–57.
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1461711&tool=pmcentrez&rendertype=abstract>.

- Saze H, Tsugane K, Kanno T, Nishimura T (2012) DNA Methylation in Plants: Relationship to Small RNAs and Histone Modifications, and Functions in Transposon Inactivation. *Plant & Cell Physiology* 53 (5): 766–84.
- Takahashi R, Morita Y, Nakayama M, Kanazawa A, and Abe J. (2012) An Active CACTA-Family Transposable Element Is Responsible for Flower Variegation in Wild Soybean. *The Plant Genome Journal* 5 (2): 62.
- Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, et al. (2007) A Unified Classification System for Eukaryotic Transposable Elements. *Nature Reviews. Genetics* 8 (12): 973–82.
- Yuan YW and Wessler SR (2011) The Catalytic Domain of All Eukaryotic Cut-and-Paste Transposase Superfamilies. *Proceedings of the National Academy of Sciences of the United States of America* 108 (19): 7884–89.

Quadro Suplementar: Descrição e caracterização das 65 sequências observadas no SoyTEdb (Du et al., 2010) e utilizadas na busca por homologia no presente trabalho.

	Sonda	5 nt (TIR)	Chr	Tam (nt)	Domínios			
					-	C	N	NC
1	>DTC_uuu_Gm1-1	CACTA	Gm0 1	13825	Tnp2 / DUF4218 / DUF4216 / Tnp_assoc	TR_nLTR / SBP	Tnp1	0
2	>DTC_uuu_Gm1-2	CACT G	Gm0 1	11757	DUF4218 / DUF4216	Tnp2	Tnp2 / Tnp_assoc	Tnp2
3	>DTC_uuu_Gm1-3	CACTA	Gm0 1	12898	Tnp1 / Tnp2 / zf-RVT	zf-RVT	0	0
4	>DTC_uuu_Gm1-4	CACTA	Gm0 1	13558	Tnp2 / DUF4216 / Tnp_assoc	ABC / PRK / ccmA / FepC / Ttg2A / TauB / cbiO / GlnQ / metN / etc	type_I_sec_LssB / CydC / type_I_sec_PrtD / DUF4218 / etc	MdlB / PRK / CydD / ATM1 / ABC / chvA / rim_protein / etc
5	>DTC_uuu_Gm1-5	CACTA	Gm0 1	9088	Tnp2 / Tnp_assoc / DUF4216	DUF4218	DUF4218 / Tnp1	0
6	>DTC_uuu_Gm2-1	CACTA	Gm0 2	14366	Tnp2 / DUF4218 / DUF4216 / Tnp_assoc	0	chaperonin_like / PMI_type1 / Fab1_TCP	TCP1 / ABD superfamily / thermosome_arch / ManA / etc
7	>DTC_uuu_Gm2-2	CACTA	Gm0 2	5719	0	Tnp2 / DUF4218	Tnp2 / Tnp24	0
8	>DTC_uuu_Gm3-1	CACTA	Gm0 3	12406	DUF4218 / DUF4216 / Tnp_assoc	Tnp2	Tnp1 / Tnp2	0
9	>DTC_uuu_Gm3-2	CACTA	Gm0 3	9989	Tnp2 / DUF4216 / Tnp_assoc	DUF4218	Tnp24 / DUF4218	0
10	>DTC_uuu_Gm3-3	CACT G	Gm0 3	5683	Tnp2 / DUF4218 / DUF4216	0	0	0
11	>DTC_uuu_Gm3-4	CACTA	Gm0 3	12601	Tnp2 / DUF4216	DUF4218	Tnp24 / Tnp_assoc / DUF4218	Tnp24
12	>DTC_uuu_Gm4-1	CACTA	Gm0 4	14305	DUF4218 / DUF4216 / Tnp_assoc / RT_nLTR	Tnp2 / MIP	Tnp2	0

	Sonda	5 nt (TIR)	Chr	Tam (nt)	Domínios			
					-	C	N	NC
13	>DTC_uuu_Gm4-2	CACTA	Gm04	16996	Tnp1 / Tnp2 / DUF4218 / zf-RVT / L1-EM / EEP / etc	RT	Nape_like_AP-endo	0
14	>DTC_uuu_Gm4-3	CACTA	Gm04	10124	Tnp2 / Tnp_assoc	0	0	PLN02874 / PRK05617
15	>DTC_uuu_Gm5-1	CACTA	Gm05	8452	Tnp2	DUF4218	Tnp_assoc	0
16	>DTC_uuu_Gm5-2	CACTA	Gm05	11516	Tnp2	DUF4218 / DUF4216 / Tnp_assoc / GH16_XET / etc	DUF4218 / Tnp_assoc	0
17	>DTC_uuu_Gm5-3	CACTA	Gm05	13169	Tnp2 / Tnp_assoc / DUF4216	DUF4218	DUF4218 / Tnp24	0
18	>DTC_uuu_Gm6-1	CACTG	Gm06	10562	Tnp2	0	Plant_zn_clust	0
19	>DTC_uuu_Gm6-2	CACTA	Gm06	13311	Tnp2 / Tnp_assoc / DUF4216	DUF4218	DUF4218 / Tnp24	Tnp24
20	>DTC_uuu_Gm6-3	CACTA	Gm06	18061	Tnp2 / DUF4216 / Pkinase	DUF4218 / pknD	DUF4218 / Tnp_assoc / Pkc / PTKc / MIP / GlpF / etc	Pkc / STKc / SPS1 /
21	>DTC_uuu_Gm7-1	CACTA	Gm07	18035	Tnp1 / Tnp2 / zf-RVT / DUF4216	DUF4218	RT_nLTR / RVT / DUF4218	
22	>DTC_uuu_Gm8-1	CACTA	Gm08	13585	Tnp2 / DUF4218 / DUF4216 / Retrotrans_gag / RT_LTR / etc	0	Tnp2	Tnp2
23	>DTC_uuu_Gm9-1	CACTA	Gm09	4593	DUF4218 / DUF4216	0	Tnp2	Glyco_hydro_35 / PLN03059
24	>DTC_uuu_Gm9-2	CACTA	Gm09	8376	Tnp2 / DUF4216	DUF4218	DUF4218 / Tnp24	0
25	>DTC_uuu_Gm9-3	CACTA	Gm09	12556	Tnp2 / Tnp_assoc / DUF4216	DUF4218 / Tnp24	Tnp24	0
26	>DTC_uuu_Gm9-4	CACTA	Gm09	12795	Tnp2 / Tnp_assoc / DUF4216 / DUF4218	0	Tnp2	0

	Sonda	5 nt (TIR)	Chr	Tam (nt)	Domínios			
					-	C	N	NC
27	>DTC_uuu_Gm9-5	CACTA	Gm09	12797	Tnp_assoc / DUF4216	Tnp2	Tnp2 / DUF3464	Tnp2
28	>DTC_uuu_Gm10-1	CACTA	Gm10	13345	Tnp1 / Tnp2 / zf-RVT	0	Pectinesterase / Tnp_assoc	APC10
29	>DTC_uuu_Gm10-2	CACTA	Gm10	12661	Tnp2 / Tnp_assoc / DUF4216	DUF4218	Tnp_assoc / DUF4218 / FBA_1	0
30	>DTC_uuu_Gm10-3	CACTG	Gm10	12336	Tnp2 / Tnp_assoc / DUF4216 / DUF4218	0	PMI / PTZ / PRK / 26Sp45	PRK / PLN / ABD / ManA
31	>DTC_uuu_Gm10-4	CACTA	Gm10	8804	Tnp2 / Tnp_assoc	0	Tnp24	0
32	>DTC_uuu_Gm11-1	CACTA	Gm11	11962	Tnp2 / DUF4216	0	DUF4218 / Tnp24 / Tnp_assoc	DUF4218
33	>DTC_uuu_Gm12-1	CACTA	Gm12	11814	Tnp2 / DUF4216	Tnp_assoc / DUF4218 / Tubulin_C / FtsZ /	DUF4218 / Tnp24 / FtsZ / Tubulin	Tubulin_C / PRK
34	>DTC_uuu_Gm12-2	CACTA	Gm12	14477	Tnp2 / Tnp_assoc / DUF4216	STKc / TOMM_kin	Tnp2 / PTZ / Glyco_hydro	STKc / PKc / PLN / PTZ / SPS / Pkinase / Glyco_hydro
35	>DTC_uuu_Gm12-3	CACTA	Gm12	11429	Tnp1 / Tnp2 / DUF4218 / DUF4216	0	PP2Cc	PP2Cc / Tnp2
36	>DTC_uuu_Gm12-4	CACTA	Gm12	17757	Tnp1 / Tnp2 / DUF4218 / DUF4216 / zf-RVT	0	RT_nLTR / RVT	0
37	>DTC_uuu_Gm12-5	CACTA	Gm12	8863	Tnp_assoc / DUF4218 / DUF42186	DUF4218	Tnp2	0
38	>DTC_uuu_Gm12-6	CACTG	Gm12	298	0	0	0	0
39	>DTC_uuu_Gm12-7	CACTG	Gm12	5396	Tnp2 / Tnp_assoc	0	0	ABC / AarF / APH / UniB
40	>DTC_uuu_Gm14-1	CACTA	Gm14	9896	Tnp2 / Tnp_assoc / DUF4216	DUF4218	DUF4218 / Tnp24	0

	Sonda	5 nt (TIR)	Chr	Tam (nt)	Domínios			
					-	C	N	NC
41	>DTC_uuu_Gm14-2	CACTA	Gm14	9073	Tnp24 / Tnp_assoc	Tnp2	Tnp2	double_bond_reductase_like / MDR
42	>DTC_uuu_Gm14-3	CACTA	Gm14	11728	Tnp1 / DUF4218 / DUF4216	Tnp2	Tnp2 / Tnp_assoc	0
43	>DTC_uuu_Gm15-1	CACTA	Gm15	14703	Tnp_aassoc / DUF4218 / DUF4216 / RT_like	Tnp2 / zf-RVT	Tnp1 / Tnp2 / RT_nLTR / RT_like / RVT	0
44	>DTC_uuu_Gm15-2	CACTA	Gm15	15853	Pkc / Tnp1 / Tnp2 / Tnp_assoc / STKc_Nek	PTKc / STKc / PKc / TyrKc / SPS1 / PTZ / TOMM / etc	PTKc / Malectin	PLN / STKc / PKc / PHA / PTZ
45	>DTC_uuu_Gm16-1	CACTA	Gm16	12954	DUF4216	Tnp2 / DUF4218	Tnp1 / Tnp2 / DUF4218	0
46	>DTC_uuu_Gm16-2	CACTA	Gm16	18165	Tnp2 / Tnp_aassoc / DUF4218 / DUF4216	0	Tnp1 / Rnase_H_Ty / Rnase_H_like	0
47	>DTC_uuu_Gm17-1	CACTA	Gm17	12842	Tnp2 / DUF4218 / DUF4216	0	Methyltransf_29 / AdoMet_MTases superfamily / Tnp_assoc	0
48	>DTC_uuu_Gm17-2	CACTA	Gm17	8127	Tnp2 / DUF4216	DUF4218	Tnp24 / Tnp_aassoc / DUF4218 / RPN1	0
49	>DTC_uuu_Gm17-3	CACTA	Gm17	13312	Tnp2 / Tnp_aassoc / DUF4216	DUF4218	Tnp24 / DUF4218	0
50	>DTC_uuu_Gm17-4	CACTA	Gm17	9731	Tnp2 / Tnp_aassoc / DUF4216	DUF4218	Tnp24 / DUF4218	0
51	>DTC_uuu_Gm17-5	CACTA	Gm17	11439	Tnp1 / Tnp2 / Tnp_assoc / EFh / DUF4216	EFh / PTZ / DUF4218	ftsH_fam / HflB / PTZ / FRQ1 / DUF4218	Peptidase_M41
52	>DTC_uuu_Gm18-1	CACTA	Gm18	7594	Tnp2 / LolD / nickel_nikD / urea_trans_UrtE / etc	ABC / PotA / CcmA / MalK / OpuBA / TauB / metN / etc	ABC / MsbA / PRK / SunT / CydD / NHLM / MRP / etc	rim_protein / PRK / PLN / SapD / znuC / dppD / etc
53	>DTC_uuu_Gm18-2	CACTA	Gm18	10728	Tnp2 / Tnp_assoc / EFh / DUF4216	0	0	0
54	>DTC_uuu_Gm18-3	CACTA	Gm18	10473	Tnp2 / Tnp_assoc	0	0	0

	Sonda	5 nt (TIR)	Chr	Tam (nt)	Domínios			
					-	C	N	NC
55	>DTC_uuu_Gm18-4	CACTA	Gm18	10160	Tnp2 / Tnp_assoc	DUF4216	Tnp1	0
56	>DTC_uuu_Gm18-5	CACTG	Gm18	9537	Tnp2	PTZ / Ribosomal_S17	0	ALDH
57	>DTC_uuu_Gm18-6	CACTG	Gm18	4002	Tnp2	Tnp_assoc	Tnp2 / STKc_CdkB_plant / PKc_like	Tnp2 / STKc / Pkc / PLN / PTZ
58	>DTC_uuu_Gm19-1	CACTA	Gm19	11463	Tnp2 / Tnp_assoc / DUF4216	PHA / DUF4218 / PHA / FAP	PHA / PRK / DUF4218	PHA / PRK / rad23
59	>DTC_uuu_Gm19-2	CACTA	Gm19	12157	Tnp2 / Tnp_assoc / DUF4216	DUF4218 / FtsZ_C / Tubulin_C	KISc_CEMP / Motor_dom / DUF4218 / Tubulin / Tnp24 / etc	KIP1 / Tubulin_C / PRK
60	>DTC_uuu_Gm19-3	CACTA	Gm19	11753	Tnp2 / Tnp_assoc / DUF4218 / DUF4216	0	WRKY	0
61	>DTC_uuu_Gm19-4	CACTA	Gm19	5039	Tnp_assoc	Tnp2	0	0
62	>DTC_uuu_Gm20-1	CACTG	Gm20	9072	Tnp_assoc / DUF4216 / Dimer_Tnp_hAT	Tnp2	0	M20_IAA_Hyd / Zinc_peptidase_like
63	>DTC_uuu_Gm20-2	CACTA	Gm20	10751	DUF4216	Tnp2 / PMD / DUF4218	PMD / DUF4218 / Tnp_assoc	0
64	>DTC_uuu_Gm20-3	CACTA	Gm20	13328	Tnp2 / Tnp_assoc / DUF4216	DUF4218	DUF4218 / Tnp24	0
65	>DTC_uuu_Gm20-4	CACTA	Gm20	11058	Tnp2 / Tnp_assoc / DUF4216	DUF4218 / RRM	DUF4218 / Tnp24 / Sulfate_transp	0

Legenda: (-) domínios completos, (C) domínio incompleto na região C-terminal, (N) domínio incompleto na região N-terminal, (NC) domínio incompletos nas duas regiões.

Conclusões:

1. Os microssatélites analisados [(AAC)₅, (AAG)₅, (ACC)₅, (AG)₈, (CTC)₅ e (TGA)₆] compreendem componentes importantes da fração repetitiva dos genomas de *Glycine soja* e *G. tomentella*, havendo particulares para cada espécie. Observa-se, porém, uma predominância de marcação dispersa, ao longo de um ou ambos os braços cromossômicos e, menos frequentemente em tandem em regiões pericentroméricas ou terminais.
2. Análises com os fluorocromos CMA/DAPI nas espécies de *Vigna* analisadas revelam variação no número de pares de cromossomos satelitados, bem como naqueles com marcações pericentroméricas ou terminais, indicando rearranjos envolvendo as regiões CMA positivas, que podem fornecer marcadores cromossômicos úteis para estudos de macrossintenia, especialmente em conjunto com sondas marcadas por FISH.
3. O retroelemento Ty1-*copia*-like encontra-se presente e disperso em vários cromossomos de *V. aconitifolia*, *V. sesquipedalis* e *V. umbellata*. Sua presença em regiões proximais e CMA positivas de alguns dos cromossomos indica uma possível associação com a heterocromatina constitutiva.
4. A análise *in silico* dos domínios de transposases Tnp1 e Tnp2 do elemento CACTA revelou uma significativa abundância do domínio Tnp2 (90.876 domínios anotados, cobrindo 10 Mb do genoma, sendo 27,5% completos), enquanto Tnp1 mostrou-se menos prevalente (19.293 domínios; 2,2 Mb, sendo 58,6% completos). A associação de domínios truncados nos terminais C e N (ou em ambos), em grande parte das sequências, indica alta incidência de eventos progressos de transposição, merecendo análises complementares.

5. No conjunto, os elementos não codificantes avaliados apresentam potencial para uso em análises de macrossintenia em Phaseolidae, colaborando para o entendimento de possíveis mecanismos que influenciaram a evolução do grupo.