

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

**ANÁLISE *IN SILICO* DE EST-SSR EM *Phaseolus vulgaris* E *Glycine max* E TRANSFERIBILIDADE DE MARCADORES PARA *Vigna unguiculata*.**

**JOSÉ RIBAMAR COSTA FERREIRA NETO**

**Recife, PE**  
**Julho, 2008**

**JOSÉ RIBAMAR COSTA FERREIRA NETO**

**ANÁLISE *IN SILICO* DE EST-SSR EM *Phaseolus vulgaris* E *Glycine max* E TRANSFERIBILIDADE DE MARCADORES PARA *Vigna unguiculata***

Dissertação apresentada ao Programa de **Pós- graduação em Genética, da Universidade Federal de Pernambuco**, como parte dos requisitos necessários para a obtenção do Grau de **Mestre em Genética**.

Orientador: **Prof. Dr. Éderson Akio Kido**, Depto. de Genética, Centro de Ciências Biológicas , UFPE.

Co-Orientadora: **Prof<sup>a</sup>. Dr<sup>a</sup>. Lauren Michelle Houllou-Kido**, Instituto Agrônomo de Pernambuco.

**Recife, PE**

**Julho, 2008**

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

**ANÁLISE *IN SILICO* DE EST-SSR EM *Phaseolus vulgaris* E *Glycine max* E TRANSFERIBILIDADE DE MARCADORES PARA *Vigna unguiculata*.**

**JOSÉ RIBAMAR COSTA FERREIRA NETO**

**Recife, PE**

**Julho, 2008**

**JOSÉ RIBAMAR COSTA FERREIRA NETO**

**ANÁLISE *IN SILICO* DE EST-SSR EM *Phaseolus vulgaris* E *Glycine max* E TRANSFERIBILIDADE DE MARCADORES PARA *Vigna unguiculata***

Dissertação apresentada ao Programa de **Pós- graduação em Genética, da Universidade Federal de Pernambuco**, como parte dos requisitos necessários para a obtenção do Grau de **Mestre em Genética**.

Orientador: **Prof. Dr. Éderson Akio Kido**, Depto. de Genética, Centro de Ciências Biológicas , UFPE.

Co-Orientadora: **Prof<sup>a</sup>. Dr<sup>a</sup>. Lauren Michelle Houllou-Kido**, Instituto Agrônômico de Pernambuco.

**Recife, PE**

**Julho, 2008**

Aos meus pais (Maria Alice de Oliveira Costa Ferreira e José Ribamar Costa Ferreira Júnior), aos meus avós [Raimunda Araújo de Oliveira Costa Ferreira e José Ribamar Costa Ferreira (*in memorian*)] e a meus irmãos (Marcell de Oliveira Costa Ferreira e Ana Alice de Oliveira Costa Ferreira), ***dedico***.

## AGRADECIMENTOS

A Deus e à Nossa Senhora pela ajuda e pela graça concedida;

Ao Santo Expedito, de quem sou devoto, por sempre olhar por mim nos em todos os momentos;

Aos meus pais e aos meus avós, por tudo;

Aos meus irmãos, pelos diálogos, atenção e força;

Ao meu orientador, Éderson Akio Kido, pela amizade ao longo desses dois anos;

À Prof<sup>ª</sup>. Iranise B.B. Torres, por ter me iniciado na Ciência;

À Prof<sup>ª</sup>. Ângela Célis de Almeida Lopes, pela amizade e por me direcionar à área da Genética Vegetal;

À Valesca Pandolfi, pela amizade e por seu altruísmo;

À Maria da Conceição Teixeira, pela atenção e carinho ao longo desses 25 longos anos;

Ao meu grande amigo, Mauro Sérgio Pavão Soares, pela amizade, atenção e pelo afilhado;

Aos *brothers* David Reis, Lourival Aragão, João Paulo Reis e Felipe Moura (*in memoriam*), pela amizade.

## SUMÁRIO

	Página
<b>Lista de Figuras</b>	07
<b>Lista de Tabelas</b>	08
<b>Lista de Abreviações</b>	09
<b>Resumo</b>	13
<b>1. Introdução</b>	14
<b>2. Revisão da Literatura</b>	16
2.1. Família Fabaceae	16
2.1.1. Taxonomia	16
2.1.2. Importância geral	16
2.2. Feijão-caupi	18
2.2.1. Taxonomia, centro de origem, centro de domesticação e introdução no Brasil.	18
2.2.2. Importância sócio-econômica global	19
2.2.3. Importância sócio-econômica no Brasil	21
2.2.4. Desenvolvimento científico atual e perspectivas para o feijão-caupi	22
2.3. Biotecnologia e marcadores genéticos	23
2.3.1. Marcadores morfológicos	24
2.2.3. Marcadores moleculares	24
2.3.2.1. Marcadores bioquímicos	24
2.3.2.2. Marcadores baseados em DNA	25
2.3.2.2.1. Marcadores de DNA baseados em hibridização	26
2.3.2.2.2. Marcadores de DNA baseados em PCR	27
a. Marcadores aleatórios	28
b. Marcadores sítios-específicos	28
2.3.2.2.3. Microsatélites	29
a. Estrutura e localização nos genomas	30
b. Gênese	30
c. Mutabilidade e evolução	31
d. Abundância e distribuição em plantas	34

e.	Potenciais funções	36
f.	Tipos de marcadores microssatélites	37
f.1.	Marcadores microssatélites genômicos	38
f.2.	EST-SSR (cDNA-SSR) ou microssatélites gênicos	39
f.2.1.	Transferabilidade de marcadores EST-SSRs	41
<b>3.</b>	<b>Referências Bibliográficas</b>	<b>44</b>
<b>4.</b>	<b>Manuscrito de Artigo Científico</b>	
	<b>Análise <i>in silico</i> de EST-SSRs de <i>Phaseolus vulgaris</i> e <i>Glycine max</i> e transferabilidade de marcadores para <i>Vigna unguiculata</i>.</b>	<b>65</b>
<b>5.</b>	<b>Informações Complementares</b>	<b>93</b>
<b>6.</b>	<b>Conclusões</b>	<b>100</b>
<b>7.</b>	<i>Abstract</i>	101
<b>8.</b>	<b>Anexos</b>	<b>102</b>
<b>8.1</b>	<b>Instruções para Autores - <i>Genetics and Molecular Biology</i></b>	<b>103</b>

## Lista de Figuras

	Figura	Página
<b>Manuscrito</b>		
<b>Fig 1.</b>	Polimorfismos em genótipos de <i>P. vulgaris</i> e <i>V. unguiculata</i> amplificados pelo <i>primer</i> CV539741 (P16). Os intervalos A a R contém genótipos de <i>V. unguiculata</i> , e A1 a J1 contém genótipos de <i>P. vulgaris</i> .	89
<b>Fig 2.</b>	Alinhamento (Clustal W) entre seqüência de <i>V. unguiculata</i> oriunda de produto de PCR (BR14Mulato-P17) e sua EST (CV531509) correspondente de <i>P. vulgaris</i> . Alinhamento contém: (1) seqüências de <i>primers</i> ( <i>left</i> e <i>right</i> ), indicadas pelas setas; (2) motivo SSR em <b>negrito e itálico</b> e; (3) posições dos nucleotídeos conservados indicados por asteriscos.	90

## Lista de Tabelas

	Tabela	Página
<b>Manuscrito</b>		
<b>Tab 1.</b>	Características dos motivos SSR encontrados em ESTs de <i>Phaseolus vulgaris</i> (P. v) e <i>Glycine max</i> (G. m).	85
<b>Tab 2.</b>	Posicionamento dos SSRs nas seqüências analisadas de <i>Phaseolus vulgaris</i> (P. v) e <i>Glycine max</i> (G. m).	86
<b>Tab 3.</b>	Motivos SSR em ESTs de <i>Phaseolus vulgaris</i> ou <i>Glycine max</i> e em <i>amplicons</i> de <i>Vigna unguiculata</i> , e valores de <i>e-value</i> em alinhamentos BLAST.	87
<b>Tab. 4</b>	<i>Primers</i> EST-SSR, respectivos motivos e repetições nas seqüências ESTs, com amplificação de monomorfismos ou polimorfismos em genótipos de P. vulgaris e <i>amplicons</i> de tamanho esperados (TE) ou não (S: sim; N: não), número de alelos e motivos (repetições) em <i>amplicons</i> seqüenciados (genótipo/ banda) e valores <i>e-value</i> entre as seqüências EST e <i>amplicons</i> (Blastn).	88
<b>Tab. 5</b>	<i>Primers</i> EST-SSR, respectivos motivos e repetições nas seqüências ESTs de P. vulgaris, com amplificação de monomorfismos ou polimorfismos em genótipos de V. unguiculata e <i>amplicons</i> de tamanho esperados (TE) ou não (S: sim; N: não), número de alelos e motivos (repetições) em <i>amplicons</i> seqüenciados (genótipo/ banda) e valores <i>e-value</i> entre as seqüências EST e <i>amplicons</i> (Blastn).	89

**Tab. 6** *Primers* EST-SSR, respectivos motivos e repetições nas seqüências ESTs, com amplificação de monomorfismos ou polimorfismos em genótipos de *G. max* e *amplicons* de tamanho esperados (TE) ou não (S: sim; N: não), número de alelos e motivos (repetições) em *amplicons* seqüenciados (genótipo/banda) e valores *e-value* entre as seqüências EST e *amplicons* (Blastn). 90

### Informações complementares

**Tab 7.** *Primers* EST-SSR, respectivos motivos e repetições nas seqüências ESTs de *G. max*, com amplificação de monomorfismos ou polimorfismos em genótipos de *V. unguiculata*, número de alelos e motivos (repetições) em *amplicons* seqüenciados e valores *e-value* entre as seqüências EST e *amplicons* (Blastn). 94

**Tab 8.** Relação de *primers* EST-SSR não funcionais em PCRs com DNAs de soja, motivos e repetições na seqüência original de *G. max* (ID TIGR) e tipo de seqüência (*singlet* ou *contig*) 94

**Tab 9.** *Primers* transferidos, mas com TE (tamanho esperado) não observado. BLASTn realizado entre *amplicon* de V.u (*Vigna unguiculata*) e *amplicon* de P.v (*Phaseolus vulgaris*). 95  
(S): sim; (N): não.

**Tab. 10** Especificações dos *primers* de *Phaseolus vulgaris* utilizados no presente trabalho, tamanho esperado do *amplicon*, além da localização e características dos SSRs flanqueados por eles. 96

<b>Tab 11.</b>	Especificações dos <i>primers</i> de <i>Glycine max</i> utilizados no presente trabalho, tamanho esperado do <i>amplicon</i> , além da localização e características dos SSRs flanqueados por eles.	97
<b>Tab 12.</b>	Análise dos <i>amplicons</i> produzidos nos diferentes indivíduos de <i>Phaseolus vulgaris</i> e <i>Vigna unguiculata</i> utilizados	98

## LISTA DE ABREVIACÕES

3' UTRs	<i>5' Untranslated Region (Regiao Não-Traduzida à 3' da ORF)</i>
5' UTRs	<i>5' Untranslated Region (Regiao Não-Traduzida à 5' da ORF)</i>
AAD	<i>Arbitrarily Amplified DNA (DNA Amplificado Arbitrariamente)</i>
ACGMs	<i>Amplified Consensus Genetic Markers</i>
AFLP	<i>Amplified Fragment Length Polymorphism (Polimorfismos de Comprimento de Fragmento Amplificado)</i>
AP-PCR	<i>Arbitrarily Primed PCR (PCR Iniciada Arbitrariamente)</i>
BLAST	<i>Basic Alignment Search Tool</i>
cDNA	<i>Complementary DNA (DNA complementar)</i>
cDNA-SSRs	Marcadores SSRs derivados de seqüências de cDNAs
DAF	<i>DNA Amplification Fingerprinting (Amplificação de Fingerprinting de DNA)</i>
DNA	<i>Desoxiribonucleic Acid (Ácido desoxirribonucléico)</i>
dNTP	Desorribonucleotídeo trifosfatado
EDTA	Ethylenediaminetetraacetic acid (Ácido Etileno Diamino Tetra Acético)
EMBL	<i>European Molecular Biology Laboratory</i>
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
EST	<i>Expressed Sequence Tags (Etiquetas de Seqüências Expressas)</i>
ESTPs	<i>Expressed Sequence Tags Polymorphisms (Polimorfismos de Etiquetas de Seqüências Expressas)</i>
EST-SSRs	Marcadores SSRs derivados de seqüências ESTs
gDNA	DNA genômico
gSSR	SSR genômico
Kb	Quilobase
kg.ha <sup>-1</sup>	Quilograma por hectare
M	Molar
Mb	Megabase
min	Minuto
mM	Milimolar
NCBI	<i>National Center for Biotechnology Information</i>
ng	Nanograma
° C	Graus Celsius
ORF	<i>Open Reading Frame (Fase ou Quadro Aberto de Leitura)</i>
p/v	Partes por volume

pb	Par de base
PCR	<i>Polymerase Chain Reaction</i> (Reação da Polimerase em Cadeia)
PIC	<i>Polymorphic Information Content</i> (Conteúdo de Informação Polimórfica)
RAPD	<i>Random Amplified Polymorphic DNA</i> (DNA Polimórfico Amplificado Randomicamente)
RFLP	<i>Restriction Fragment Length Polymorphism</i> (Polimorfismos em Tamanho de Fragmentos de Restrição)
RNA	<i>Ribonucleic Acid</i> (Ácido ribonucléico)
SCAR	<i>Sequence-Characterized Amplified Region</i> ou Regiões Amplificadas a partir de Sequências Caracterizadas
SSLP	<i>Simple Sequence Length Polymorphism</i> (Polimorfismos em Tamanho de Sequências Simples)
SSR	<i>Simple Sequence Repeats</i> (Sequências Simples Repetidas ou Microsatélites)
STR	<i>Short Tandem Repeats</i> (Pequenas Repetições Lineares)
STS	<i>Sequence Tagged Sites</i>
t.ha <sup>-1</sup>	Tonelada por hectare
TIGR	Antigo "The Institute for Genomic Research", atual "J. Craig Venter Institute"
U	Unidade
μL	Microlitro
μM	Micromolar
VNTR	<i>Variable Number Tandem Repeats</i> (Repetições Lineares em Número Variável ou mesmo Sequências Hipervariáveis)

## RESUMO

Bancos de dados de seqüências expressas representam uma fonte potencialmente valorosa para o desenvolvimento de marcadores moleculares. Neste trabalho, estudos foram realizados para o desenvolvimento de marcadores para aplicação no melhoramento de feijão-caupi, feijão-comum e soja. Foram analisadas 10.880 ESTs de feijão-comum e 116.965 ESTs de soja. Um total de 331 e 4553 SSRs foram identificados nas seqüências de feijão-comum e soja, respectivamente. Trinucleotídeos (45,62 %) foram as mais abundantes em feijão-comum, enquanto que os dinucleotídeos (49,67 %), em soja. A densidade média variou de um SSR a cada 22 Kb (feijão-comum) a um SSR a cada 14,68 Kb (soja). Repetições diméricas AG/GA / TC/CT e triméricas AAG/AGA/GAA / TTC/TCT/CTT foram as mais abundantes em ambas espécies. Um total de 153 e 1928 pares de *primers* EST-SSRs foram propostos para feijão-comum e para soja, respectivamente. Desses, sintetizaram-se 20 para soja, dos quais apenas nove amplificaram (sendo quatro de tamanho esperado e, desses, dois polimórficos) e dois foram transferíveis. Para feijão-comum sintetizaram-se 22, dos quais 15 amplificaram, sendo que 11 apresentaram tamanho esperado e 10 desses foram polimórficos (PIC médio de 0,50); 11 dos funcionais foram transferidos para feijão-caupi, sendo oito de tamanho esperado e cinco desses, polimórficos (PIC médio de 0,36). Análises detalhadas dos *amplicons* seqüenciados determinaram que as extensões dos motivos de SSRs foram variáveis e que as regiões flanqueadoras dessas repetições foram geralmente bem conservadas, confirmando o sucesso da transferibilidade entre feijão-comum e feijão-caupi.

**Palavras-chave:** microssatélite, mineração de dados, Fabaceae, leguminosa, feijão-caupi.

## 1. INTRODUÇÃO

Atualmente, os bancos de dados de domínio público como TIGR (*The Institute for Genomic Research*), NCBI (*National Center for Biotechnology Information*), EMBL (*European Molecular Biology Laboratory*), etc., tanto de seqüências de DNA quanto de proteínas, são amplamente acessados. Esses bancos recebem grandes aportes de seqüências, em decorrência principalmente dos inúmeros projetos de seqüenciamento genômico concluídos ou em andamento. Assim, o acúmulo e análise de seqüências expressas do tipo EST (*Expressed Sequence Tags* ou Etiquetas de Seqüências Expressas), obtidas pelo seqüenciamento de uma das extremidades de um determinado cDNA, fruto da conversão de um RNA poliadenilado, têm se tornado importantes componentes da pesquisa genômica. Nesses bancos, as principais culturas de importância econômica encontram-se bem representadas, pois se destacam em relação às demais quanto à disponibilidade de informações e de projetos de pesquisa, com envolvimento de inúmeros grupos de pesquisadores.

A família Fabaceae corresponde, dentre as plantas, à segunda em número de seqüências depositadas no banco de dados TIGR (seção *Plant Transcript Assemblies*; <http://plantta.tigr.org/>), só perdendo para a família Poaceae. Também é considerada a segunda família vegetal mais importante em termos agrícolas e nutricionais, ocupando parcela considerável das terras aráveis do planeta. É inegável a importância desse grupo vegetal na alimentação animal (grãos e forragem) e humana (grãos), sendo responsável por boa parte de toda proteína ingerida. Dentre as espécies da família, as mais bem representadas no banco supracitado estão *Glycine max* (L.) Merr. (soja), *Medicago truncatula* Gaertner e *Lotus japonicus* (Regel) K. Larsen. (ambas plantas-modelo e de clima temperado), e *Phaseolus vulgaris* L. (feijão-comum). Por sua vez, a espécie *Vigna unguiculata* (L.) Walp. (feijão-caupi), uma das mais cultivadas mundialmente em regiões pobres e semi-áridas, por ser ainda pouco estudada, não possui nenhuma informação depositada nesse banco. Na região Nordeste do Brasil, o feijão-caupi é uma das principais fontes de proteínas vegetal para populações menos favorecidas e também importante geradora de emprego e renda.

Atualmente, no campo da genômica de leguminosas, há uma necessidade de se integrar os conhecimentos obtidos em estudos envolvendo espécies-modelo (ou aquelas mais estudadas) com as questões biológicas e agronômicas das espécies cultivadas. Dessa forma, culturas com menor disponibilidade de informações poderiam se valer da disponibilidade de

dados de espécies correlacionadas. Neste contexto, certos tipos de marcadores moleculares, especialmente aqueles SSRs (*Simple Sequence Repeats* ou Sequências Simples Repetidas) ou microssatélites, poderiam ser detectados. Esses marcadores são um dos mais informativos devido a sua hipervariabilidade e natureza codominante. Marcadores moleculares de DNA, de modo geral, constituem ferramentas biotecnológicas utilizadas em análises genéticas, sendo de grande importância para o mapeamento e estudos de diversidade. Devido à sintenia entre espécies aparentadas, pares de *primers* desenvolvidos para flanquear SSRs, detectados em seqüências ESTs de uma determinada espécie, poderiam ser transferidos para outra, filogeneticamente próxima. Isso possibilitaria a minimização de custos necessários para o desenvolvimento desses marcadores (confeção de biblioteca específica e seqüenciamento dos clones para posterior proposição de *primers*) e uma democratização do progresso científico, a serem obtidas a partir de espécies mais estudadas e com maior abundância de dados. Este trabalho tem como objetivo a detecção, classificação e estudo de seqüências microssatélites (SSR) em seqüências ESTs de soja e feijão-comum, a proposição de *primers* flanqueadores para essas regiões, ensaios de validação desses EST-SSRs em suas respectivas espécies e teste de transferibilidade para feijão-caupi, visando o desenvolvimento de marcadores para futuras aplicações em estudos de diversidade e mapeamento genético das mesmas, bem como no mapeamento comparativo entre essas.

## **2. REVISÃO DA LITERATURA**

### **2.1. Família Fabaceae**

#### **2.1.1 Taxonomia**

A família Fabaceae constitui uma importante família vegetal, composta por espécies conhecidas comumente como leguminosas, sendo amplamente cultivadas (Singh, 2005). Inclui 727 gêneros e 19.325 espécies (Lewis *et al.*, 2005), variando de pequenos vegetais herbáceos a grandes árvores florestais e crescendo sob as mais diversas condições agroecológicas (Singh *et al.*, 2007). Tal família abrange espécies produtoras de grãos, oleaginosas, forrageiras, ornamentais, agroflorestais e medicinais (Lewis *et al.*, 2005). Dentre essas, ocupam posição de destaque (em virtude de suas características nutricionais, importância econômica ou científica) a soja (*Glycine max* (L.) Merr.), o amendoim (*Arachis hypogea* L.), o feijão-comum (*Phaseolus vulgaris* L.), o feijão-caupi [*Vigna unguiculata* (L.) Walp.] e as espécies-modelo *Medicago truncatula* Gaertner e *Lotus japonicus* (Regel) K. Larsen.

Análises taxonômicas, baseadas em caracteres morfológicos (particularmente estrutura floral), dividem as leguminosas em três subfamílias principais (Doyle e Luckow, 2003): (1) Papilionoidea, a mais numerosa (476 gêneros e cerca de 14.000 espécies) e economicamente dominante (Langer e Hill, 1991; Choi *et al.*, 2004a), com elevada diversidade (vegetais herbáceos a arbustos), ampla distribuição e da qual fazem parte todas as espécies supracitadas; (2) Caesalpinioidea (162 gêneros e cerca de 3.000 espécies) e (3) Mimosoidea (77 gêneros e ao redor de 3.000 espécies), ambas constituídas basicamente por árvores e arbustos de importâncias mais restritas (Langer e Hill, 1991; Singh *et al.*, 2007).

#### **2.1.2 Importância geral**

As Fabaceae representam a segunda família vegetal mais importante em termos agrícolas (Mun *et al.*, 2006; Singh *et al.*, 2007), ocupando aproximadamente 15% das terras aráveis do planeta e respondendo por cerca de 27% da produção agrícola mundial, além de aproximadamente 35% de todo o óleo vegetal processado globalmente (Vance *et al.*, 2000; Graham e Vance, 2003). Industrialmente, possui ainda ampla aplicabilidade, com emprego na confecção de plásticos biodegradáveis, biodiesel, cosméticos, tintas, etanol, produtos farmacêuticos, resinas, outros óleos, etc (Graham e Vance, 2003; The National Biodiesel

Board; Singh *et al.*, 2007). Aliada a essa importância agro-ecológica e financeiro-industrial, observa-se que a essa família apresenta potencial para desempenhar um importante papel social, contribuindo, sobretudo, para a melhoria da qualidade de vida de populações menos abonadas (Singh, 2005). O aumento no consumo desses vegetais pode reduzir problemas de má nutrição calórico-proteica, que afetam ao redor de 12 % dos habitantes do planeta (Singh *et al.*, 2007; World Hunger Education Service, 2007), uma vez que constituem uma fonte de proteínas mais acessível economicamente e amplamente difundida (Singh *et al.*, 2007). Segundo Vance *et al.* (2000), as leguminosas respondem por até 33% da dieta de proteína ingerida por humanos. Somado ao exposto, pesquisas têm relatado que o consumo moderado de certas leguminosas é útil no combate a doenças como Alzheimer, Parkinson e até mesmo certos tipos de câncer (Singh 2005, 2007).

Outra característica-chave desse grupo é a presença de nódulos radiculares contendo bactérias *Rhizobium*, ou gêneros relacionados, em relação de simbiose com a planta, e que auxiliam na fixação de nitrogênio no solo (Graham e Vance, 2003; Singh, 2005). Nesse processo, realizado por cerca de 20% dos representantes dessa família (Sprent e Sprent, 1990), ocorre a transformação do nitrogênio atmosférico (inerte) para sua forma biologicamente útil aos seres vivos (amônia, nitratos e nitrogênio orgânico) (Zahran, 1999; Lindemann e Glover, 2003). A disponibilidade desse elemento é de fundamental importância, pois influencia a produção em muitos ecossistemas naturais e agrícolas (Seastedt e Knapp, 1993; Vitousek *et al.*, 1997). Portanto, tal capacidade de fixação, confere a determinadas espécies dessa família valor estratégico, uma vez que o aumento nos níveis de proteína no vegetal e a diminuição das perdas de nitrogênio no solo são consequências óbvias desse processo (Zahran, 1999), além da redução de gastos inerentes ao processo de fertilização periódica dos solos em culturas não-leguminosas.

Cientificamente, as leguminosas também deram sua contribuição para o progresso da ciência. O desenvolvimento da genética moderna teve suas bases construídas a partir de estudos de hibridização em ervilhas (*Pisum sativum* L.), realizados por Gregor Mendel (Singh, 2005). Atualmente, as principais espécies-modelo são *M. truncatula* e *L. japonicus*, as quais são amplamente utilizadas em estudos genéticos, genômicos e fisiológicos (Choi *et al.*, 2004a; Gutierrez *et al.*, 2004; Cannon *et al.*, 2006; Mun *et al.*, 2006; Sato e Tabata, 2006; Starker *et al.*, 2006). Contrariamente, observa-se, também, um significativo número de espécies com disponibilidade de dados muito restrita, denominadas, tecnicamente, “órfãs”, provavelmente por possuírem importância mais regionalizada ou por serem mais consumidas

em regiões pobres, onde investimentos no setor agrícola são extremamente baixos (Nelson *et al.*, 2004).

Avanços na área da genômica têm fornecido um entendimento mais unificado da biologia vegetal, bem como um poderoso grupo de métodos, ferramentas moleculares e de bioinformática. Tais avanços fornecem oportunidades para uma eficiente transferência de informação (Naylor *et al.*, 2004), principalmente no campo da genômica de leguminosas, onde há uma necessidade urgente de se integrar os conhecimentos ganhos em estudos envolvendo espécies-modelo com questões biológicas e agronômicas de espécies cultivadas (Choi *et al.*, 2004a; Zhu *et al.*, 2005). Como consequência, investimentos relativamente pequenos na transferência de informação das espécies mais bem estudadas para espécies consideradas “órfãs” resultariam em maior produção, estabilidade da cultura e segurança alimentar em países menos desenvolvidos (Nelson *et al.*, 2004).

## **2.2 Feijão-caupi**

### **2.2.1 Taxonomia, centro de origem, centro de domesticação e introdução no Brasil.**

O feijão-caupi [*V. unguiculata* (L.) Walp.] é uma importante leguminosa utilizada como gênero alimentício em muitas nações em desenvolvimento devido ao seu excelente conteúdo nutricional, baixo custo e disponibilidade (Boeh-Ocansey, 1989; Uzogara e Ofuya, 1992; Prinyawiwatkul *et al.*, 1996). Essa espécie é também conhecida por outros nomes vulgares, como feijão-de-corda e feijão macassar (na região Nordeste), feijão miúdo (na região Sul), feijão catador e feijão gerutuba (em algumas regiões do Estado da Bahia e norte de Minas Gerais), e feijão fradinho (no Estado do Rio de Janeiro) (Freire Filho *et al.*, 1999). Trata-se de uma espécie diplóide ( $2n = 22$ ), dicotiledônea, pertencente à ordem Fabales, família Fabaceae, subfamília Papilionoidea, tribo Phaseolae, subtribo Phaseolina, gênero *Vigna* e seção Catiang (Frahm-Leliveld, 1965; Verdcourt, 1970, Maréchal *et al.*, 1978; Padulosi e Ng, 1997).

A espécie é originária da África, onde foi inicialmente domesticada (Zeven e de Wet, 1982), com centros de diversidade identificados na África e Ásia. Observações indicam serem os acessos asiáticos de feijão-caupi diversos e morfologicamente diferentes dos africanos, mas a ausência de parentes selvagens, que atuariam como possíveis progenitores no continente asiático, tem colocado a existência desse centro sob questionamento (Singh, 2005).

Em relação à sua domesticação, a exata região da realização desse processo tem sido objeto de especulações e discussões. Alguns autores (Ng e Maréchal, 1985; Vaillancourt e Weeden, 1992; Ng, 1995) sugeriram que o centro de domesticação do feijão-caupi poderia estar situado no oeste da África devido: (1) ao alto nível de diversidade morfológica dos acessos cultivados na região; (2) à existência de possíveis intermediários entre os feijões-caupi cultivado e selvagem; (3) às mais antigas evidências arqueológicas para tal espécie na região de Gana e (4) à identificação de acessos cultivados e selvagens com idênticos DNAs cloroplastidais, na Nigéria. Um segundo centro de domesticação indicado estaria situado no nordeste da África, pois: (1) há ausência de um acesso ecologicamente selvagem no oeste da África; (2) existe um elevado nível de diversidade morfológica para tal espécie na região que se estende da Etiópia à África do Sul (Baudoin e Maréchal, 1985) e (3) resultados alcançados em estudos de etnobotânica e ensaios com isoenzimas sinalizam positivamente para tal situação (Pasquet, 1996).

A introdução dessa leguminosa no continente americano ocorreu, provavelmente, no século XVI pelos colonizadores espanhóis e portugueses, em princípio nas colônias espanholas e, posteriormente, em nosso país, provavelmente pelo Estado da Bahia. A partir daí, se disseminou para as demais áreas da região Nordeste e outras regiões (Freire Filho *et al.*, 1999).

### **2.2.2. Importância sócio-econômica global**

Atualmente, o feijão-caupi é uma das 23 espécies cultivadas mais importantes (Cowpea Genomics, 2008), ocupando cerca de 14 milhões de hectares distribuídos pela Ásia, África, sul da Europa e Américas Central e do Sul (Henriet *et al.*, 1997; Mortimore *et al.*, 1997; Singh *et al.*, 1997; Van Ek *et al.*, 1997). Sua produção mundial é estimada em 4,5 milhões de toneladas anuais (Singh, 2005), sendo a África (70%), Américas do Sul e Central (19%) e Ásia (10%), os maiores produtores (Olufajo e Singh, 2002). O único país desenvolvido e fora dessa zona a produzir quantidades significativas de feijão-caupi corresponde aos Estados Unidos da América (Prinyawiwatkul *et al.*, 1996; Pasquet e Baudoin, 2001). No mais, a grande produção se encontra em países em desenvolvimento, onde é cultivado principalmente por lavradores de baixa renda e empregado tanto na alimentação humana quanto animal (Fery, 2002; Chen *et al.*, 2007).

Devido à plasticidade fenotípica dos genótipos disponíveis (Freire Filho *et al.*, 1999; Dantas *et al.*, 2002; Xavier *et al.*, 2005), uma das principais características dessa espécie, a

cultura se constitui ainda em umas das mais cultivadas em sistemas agrícolas tradicionais aplicados em regiões semi-áridas (Olufajo e Singh, 2002; Singh, 2005), pois apresenta melhor adaptabilidade à seca e a altas temperaturas (Ehlers e Hall, 1997; Kuykendall *et al.*, 2000; Martins *et al.*, 2003).

Apesar de menos adaptada a solos frios (Dadson *et al.*, 2005; Cover Crop Database, 2008), o feijão-caupi é capaz de tolerar solos ácidos e alcalinos, além daqueles de baixa fertilidade (Elowad e Hall, 1987; Fery, 1990). Também é tolerante à baixa luminosidade, sendo compatível ao cultivo consorciado com milho, sorgo, milheto, cana-de-açúcar, algodão, entre outras culturas (Singh e Emechebe, 1998). Por apresentar essa plasticidade produtiva sob as mais diversas condições ambientais, tal leguminosa pode atenuar perdas econômicas de outras culturas, em caso de períodos desfavoráveis (Dadson *et al.*, 2005), sendo também importante componente da agricultura de subsistência (Singh *et al.*, 2003).

Adicionalmente, é utilizada como adubo verde, devido a sua eficiência no processo de fixação biológica do nitrogênio e capacidade de fertilização de solos (Elowad e Hall, 1987; Kuykendall *et al.*, 2000; Martins *et al.*, 2003), além de eficaz simbiose com micorrizas (Kwapata e Hall, 1985), o que contribuiria para a redução de custos de produção associados ao uso de fertilizantes (Martins *et al.*, 2003). Dados experimentais demonstraram que o teor de nitrogênio no solo aumenta cerca de 40,0 a 80,0 kg.ha<sup>-1</sup> quando da introdução dessa espécie em sistemas de rotação de culturas (Quin *et al.*, 1997).

Na culinária, o feijão-caupi é consumido principalmente na forma de grãos secos ou verdes (no sudeste dos Estados Unidos da América e Senegal, vagens verdes frescas (regiões úmidas da Ásia e Caribe), além de folhas (em muitas regiões do leste da África) (Ehlers e Hall, 1997). O perfil nutricional de seus grãos (Bressani, 1985) é similar ao do feijão-comum, com níveis mais elevados de ácido fólico e menores índices de fatores anti-nutricionais e de causadores de flatulência (Ehlers e Hall, 1997).

Na África, o feijão-caupi é importante fonte de proteínas, minerais e vitaminas, com participação diária na alimentação dos povos e influência positiva na saúde de populações menos abastadas (Singh, 2005) ao complementar a dieta básica rica em amido (macaxeira, inhame, milheto, milho ou sorgo). A adição dessa leguminosa, mesmo em pequenas quantidades, assegura um balanço nutricional destacado pela sinergia entre o seu alto conteúdo protéico e a alta quantidade de energia de alimentos ricos em amido (Singh, 2005). Adicionalmente, o comércio dessa espécie *in natura* ou na forma processada fornece aos produtores uma importante fonte de renda (Singh *et al.*, 2003; Singh, 2005).

A título de curiosidade, o feijão-caupi foi uma das poucas espécies escolhidas pela NASA (*National Aeronautical and Space Administration* ou Administração Nacional de Aeronáutica e Espaço – Estados Unidos da América) para estudos em estações espaciais, devido às suas características nutricionais superiores, versatilidade, adaptabilidade e potencial produtivo (Ehlers e Hall, 1997).

### **2.2.3 Importância sócio-econômica no Brasil**

O Brasil é o segundo maior produtor de feijão-caupi, atrás somente da Nigéria (Singh, 2005). A grande produção nacional ocorre na região Nordeste, onde aproximadamente 412 mil toneladas dessa leguminosa são produzidas anualmente, em uma área de cerca de 1,2 milhões de hectares, o que equivale a 94,4% da área total plantada com a cultura no Brasil e 87,7% da produção nacional (Silva, 2007). Estatísticas mais detalhadas ou acuradas não estão disponíveis, apesar da importância regional da cultura, pois atualmente todo feijão-caupi cultivado nacionalmente é computado conjuntamente com a produção de feijão-comum (*P. vulgaris*). Os Estados do Piauí e Ceará eram os detentores das maiores produções e extensões de área plantada, de acordo com o último Levantamento Sistemático da Produção Agrícola disponível (de 1993 a 1997).

Para o Nordeste brasileiro, a espécie é de grande valor estratégico, por ser possuidora de ampla variabilidade genética, ampla capacidade de adaptação, bom potencial produtivo, boa eficiência na fixação biológica de nitrogênio e excelente valor nutritivo (Freire Filho *et al.*, 1999). Admitindo-se que 1,0 hectare dessa leguminosa gere 1,5 empregos por ano, tal cultura seria responsável pela geração de 1,8 milhões de empregos diretos, além de abastecer a mesa de 27,5 milhões de nordestinos (Anuário Estatístico da Agricultura Brasileira, 1998; Freire Filho *et al.*, 1999). Adicionalmente, o valor de produção estimado para essa leguminosa alcançaria cerca de 249 milhões de dólares.ano<sup>-1</sup> (Caupi movimenta..., 2001). Segundo Freire Filho *et al.* (1999), esses dados são extremamente importantes, pois refletem a participação da cultura no contexto de geração de emprego, renda e produção de alimentos no país e a credenciaria para receber maior atenção por parte das políticas de abastecimento e de órgãos de fomento à pesquisa.

Nas áreas urbanas não metropolitanas da região Nordeste, o feijão-caupi contribui com 41,0 % de todo feijão consumido, constituindo-se em alimento básico para a população, exercendo a função de supridor das necessidades alimentares das camadas mais desprovidas (Dantas *et al.*, 2002). Suas propriedades nutricionais, relativamente superiores ao feijão-

comum, aliadas ao baixo custo de produção, fazem com que a cultura seja considerada importante em termos sociais e econômicos para a região (Araújo e Watt, 1988).

#### **2.2.4 Desenvolvimento científico atual e perspectivas para o feijão-caupi**

Apesar de ter sido uma das primeiras espécies investigadas por geneticistas (Fery, 1985), o feijão-caupi ainda possui disponibilidade de dados ainda restrita por receber pouca atenção do ponto de vista científico (Nelson *et al.*, 2004; Naylor *et al.*, 2004; Chen *et al.*, 2007). Comparada às outras culturas, essa leguminosa apresenta seu potencial genético pouco explorado, sendo considerada uma espécie subutilizada (Freire Filho *et al.*, 1999; Padulosi *et al.*, 2002; Chen *et al.*, 2007).

Dados de Bezerra (1997), em condições experimentais, revelaram uma produtividade de grãos secos acima de 3,0 t.ha<sup>-1</sup>, entretanto a expectativa é que seu potencial genético ultrapasse 6,0 t.ha<sup>-1</sup> (Freire Filho *et al.*, 1999). Apesar de pouco melhorada, a espécie apresenta ampla variabilidade genética para praticamente todos os caracteres de interesse agrônomo (EMBRAPA, 1990; Teófilo *et al.*, 1989, 1990; Freire Filho *et al.*, 1991, 1992). Por suas características de rusticidade, ampla adaptabilidade, precocidade e capacidade de produzir em ambientes desfavoráveis, a cultura é vista com grandes perspectivas para fazer frente à escassez de alimentos em regiões como o Nordeste brasileiro (Freire Filho *et al.*, 1999).

Segundo Freire Filho *et al.* (1999), faz-se necessário que os programas de melhoramento convencional busquem o apoio de ferramentas de biologia molecular para que se tornem mais ágeis e eficientes, visando competir em qualidade, oferta e custo de produção com outros feijões. No futuro, o feijão-caupi, pode vir a ser modelo para pesquisas avançadas, pois possui características vantajosas para o desenvolvimento de estudos: embora realize fecundação cruzada, a espécie é predominantemente de auto-fecundação (Ehlers e Hall, 1997); as gerações são curtas (cerca de três meses na região Nordeste do Brasil), facilitando a realização de experimentos; seu genoma, estimado em 620 Mb, é um dos menores dentre as leguminosas e dentre as plantas vasculares (Arumuganathan e Earke, 1991; Chen *et al.*, 2007); e há relatos na literatura de protocolos de transformação para essa espécie (Ikea *et al.*, 2003; Popelka, *et al.*, 2006). Assim, em um genoma estável, esse conjunto evitaria o “escape” de genes para parentes selvagens ou plantas circunvizinhas. Adicionalmente, alelos encontrados no feijão-caupi, que contribuem positivamente para a tolerância a solos secos e outros tipos de estresses abióticos, poderiam ser usados em outras leguminosas (Nelson *et al.*,

2004) ou a cultura poderia ser fonte potencial de expressão de genes exógenos e produção de substâncias bioativas (fármacos) ou de importância industrial.

### **2.3. Biotecnologia e marcadores genéticos**

A Biotecnologia, definida como “toda e qualquer técnica que usa organismos vivos, ou substâncias desses organismos, para fabricar ou modificar um produto, melhorar plantas e animais, ou desenvolver microorganismos para usos específicos” (Barnum, 1998), tem exercido um papel fundamental na ciência e sociedade. Na agricultura, tem sido empregada na cultura de tecidos, para rápida propagação de espécies vegetais; em diagnósticos, na detecção de agentes causadores de doenças; na modificação genética de genótipos, por meio da introdução de novas características, fazendo-se uso da técnica do DNA recombinante; e auxiliando programas de melhoramento convencional através do uso de marcadores moleculares (Presley, 1992).

O principal objetivo dos melhoristas é modificar positivamente as variedades existentes, as quais são deficientes em uma ou mais características (Kumar, 1999). O melhoramento convencional pode ser reduzido a duas etapas fundamentais (Manshardt, 2004): (1) geração de população de melhoramento (altamente variável para as características de interesse), obtida a partir de parentais rigorosamente escolhidos; (2) seleção de indivíduos recombinantes superiores na progênie segregante. Tais procedimentos são laboriosos e demorados, envolvendo diversos cruzamentos e gerações, além de uma cautelosa seleção fenotípica (Kumar, 1999).

Por sua vez, a tecnologia do DNA recombinante pode ajudar a vencer algumas limitações. Entretanto, técnicas de engenharia genética também possuem limitações (ex.: ausências de um número suficiente de genes clonados e de protocolos de transformação para muitas culturas) (Kumar, 1999). Com o advento da tecnologia dos marcadores moleculares, diversos tipos de marcadores de DNA e estratégias de melhoramento molecular têm sido disponibilizados aos melhoristas, visando vencer barreiras como a seleção com base em caracteres morfológicos e o tempo necessário para o desenvolvimento de uma nova cultivar, em relação ao melhoramento tradicional (Milach, 1998; Kumar, 1999; Phillips e Vasil, 2001; Andersen e Lübberstedt, 2003; Gupta e Rustgi, 2004). Para se entender o potencial dos marcadores moleculares, há de se primeiro entender que marcadores genéticos são localizações específicas em um cromossomo as quais são utilizadas como pontos de

referência para análises de genomas (Kumar, 1999) e podem ser de dois tipos: morfológicos e moleculares (Ferreira e Grattapaglia, 1998; Kumar, 1999; Weising *et al.*, 2005).

### **2.3.1. Marcadores morfológicos**

Até meados da década de 1960, os marcadores utilizados em estudos de genética e melhoramento vegetais eram associados a genes relacionados a caracteres morfológicos. A herança desses marcadores pode ser monitorada visualmente, sem necessidade de técnicas refinadas. Entretanto, sua expressão pode ser altamente influenciada pelo ambiente e por efeitos epistáticos e pleiotrópicos (Ferreira e Grattapaglia, 1998; Andersen e Lübberstedt, 2003). Adicionalmente, são considerados marcadores dominantes (incapazes de diferenciar organismos heterozigotos de homozigotos) e se apresentam em número limitado (e restritos às espécies-modelo), o que reduz a probabilidade de se encontrar associações entre esses marcadores e caracteres de importância econômica (Ferreira e Grattapaglia, 1998; Kumar *et al.*, 1999; Andersen e Lübberstedt, 2003).

Estes marcadores contribuíram significativamente para o desenvolvimento teórico da análise da ligação gênica e para a construção das primeiras versões de mapas genéticos (Ferreira e Grattapaglia, 1998).

### **2.3.2. Marcadores moleculares**

Marcador molecular, por sua vez, é todo e qualquer fenótipo molecular oriundo de um gene expresso, como no caso das isoenzimas, ou de um segmento específico de DNA, correspondente a regiões expressas, ou não, do genoma (Ferreira e Grattapaglia, 1998). São classificados em duas categorias principais (Kumar, 1999):

#### **2.3.2.1. Marcadores bioquímicos**

São proteínas que podem ser separadas por meio de eletroforese para identificação de alelos. As mais comumente usadas são as isoenzimas (Kumar, 1999). Estas desempenham a mesma atividade catalítica, entretanto podem exibir diferentes propriedades cinéticas e serem separadas por processos bioquímicos. Isto significa que as isoenzimas de um mesmo grupo diferem entre si na seqüência de aminoácidos, o que poderá influenciar, por conseguinte, na natureza de suas estruturas protéicas secundárias, terciárias e quaternárias (Ferreira e

Grattapaglia, 1998). O princípio básico da técnica reside no uso de eletroforese em gel de amido (Smithies, 1955) e na visualização do produto enzimático por métodos histoquímicos (Hunter e Markert, 1957). A premissa básica adotada ao se utilizar dados isoenzimáticos é que diferenças na mobilidade desses marcadores em um campo elétrico são resultantes de diferenças ao nível de seqüências de DNA que codificam tais enzimas. Assim se o padrão de bandas de dois indivíduos difere, assume-se que essas desigualdades possuam base genética e sejam herdáveis (Murphy *et al.*, 1990).

Esses marcadores são de natureza codominante, ou seja, com capacidade de distinção entre indivíduos homocigotos e heterocigotos (Ferreira e Grattapaglia, 1998; Kumar, 1999; Weising *et al.*, 2005) e correspondem às múltiplas formas moleculares da mesma enzima em uma dada espécie, como resultado da presença de dois ou mais alelos codificando cada uma das mesmas (Moss, 1982). Embora seja uma técnica barata, esses marcadores apresentam sérias limitações. Apesar de considerados seletivamente neutros, são de polimorfismo limitado, pois as enzimas possuem função metabólica (Ferreira e Grattapaglia, 1998; Weising *et al.*, 2005). Além disso, não permitem uma cobertura ampla do genoma, limitando as áreas de estudo e prejudicando a construção de mapas genéticos (Ferreira e Grattapaglia, 1998). Adicionalmente, as isoenzimas estão sujeitas as modificações pós-traducionais (Staub *et al.*, 1982) e podem, também, apresentar especificidade de formas estruturais em alguns tecidos ou ser a conformação tridimensional influenciada pelo estágio de desenvolvimento do organismo, além de gerarem polimorfismos em resposta a condições ambientais (influência das condições ambientais na atividade enzimática) (Ferreira e Grattapaglia, 1998). Tudo isso pode gerar falsos indicativos de semelhanças ou dissimilaridades entre os indivíduos sob análise.

### **2.3.2.2. Marcadores baseados em DNA**

Esses marcadores representam características de DNA que diferenciam dois ou mais organismos e são herdados geneticamente (Milach, 1998). Atualmente, os distintos tipos de marcadores moleculares baseados em DNA, diferenciam-se pela tecnologia empregada para revelar variabilidade e assim detectar diferenças entre organismos, além de custos, facilidades de uso, consistência e repetibilidade (Milach, 1998; Weising *et al.*, 2005). Uma grande variedade de marcadores moleculares baseados em DNA foi desenvolvida nas últimas décadas (Phillips e Vasil, 2001; Weising *et al.*, 2005). A imensa maioria utilizava como plataforma de desenvolvimento o DNA genômico, analisando, em grande parte,

polimorfismos oriundos de regiões não-codificantes (Gupta e Rustgi, 2004). Basicamente, são classificados em duas categorias: marcadores baseados em hibridização e baseados em PCR (Ferreira e Grattapaglia, 1999; Kumar, 1999; Weising *et al.*, 2005).

#### **2.3.2.2.1 Marcadores de DNA baseados em hibridização**

Polimorfismos baseados em hibridização incluem os marcadores RFLP (*Restriction Fragment Length Polymorphisms* ou Polimorfismos em Tamanho de Fragmentos de Restrição) e VNTR (*Variable Number of Tandem Repeats* ou Repetições Lineares em Número Variável ou mesmo seqüências hipervariáveis) como seus principais representantes (Milach, 1998; Ferreira e Grattapaglia, 1998; Kumar, 1999). Ambas as técnicas utilizam-se da homologia entre sondas específicas e o DNA previamente fragmentado por enzimas de restrição para análise de polimorfismos. Para isso, fragmentos de DNA são separados, via eletroforese, e posteriormente transferidos para membrana de *nylon* (ou nitrocelulose) via capilaridade ou vácuo (técnica de *Southern Blot*; Southern, 1975). Durante esse processo, a ordem dos fragmentos separados por meio da eletroforese é mantida na transferência. A seguir, os segmentos de DNA são fixados covalentemente na membrana (por alta temperatura em forno a vácuo ou luz ultravioleta). Por fim, as sondas são hibridizadas e o polimorfismo pode ser evidenciado por autoradiografia (Sambrook, 1989; Ferreira e Grattapaglia, 1998). A diferença entre os marcadores citados se encontra no tipo de sonda utilizada (Botstein *et al.*, 1980; Kumar *et al.*, 1999; Malacinski, 2003). Na técnica de RFLP, são usadas sondas homólogas a seqüências únicas do genoma. Na detecção de polimorfismos VNTR, as sondas são constituídas de seqüências repetitivas (mini ou microssatélites).

Os polimorfismos oriundos de análises RFLP são originados de substituições de base, inserções, deleções ou outros rearranjos (translocações, inversões) nos sítios de restrição ou entre esses (Parfitt e Arulsekhar, 1987; Dunham, 2004; Weising *et al.*, 2005). Nos marcadores VNTR, além do polimorfismo do comprimento dos fragmentos de restrição em cada loco hipervariável, também o polimorfismo oriundo do número de unidades repetitivas que constituem um loco VNTR, que são amplamente distribuídos pelo genoma, possibilitam a análise de diversas regiões simultaneamente (Ferreira e Grattapaglia, 1998; Lewin, 2004). Ambos constituem marcadores co-dominantes, altamente reprodutíveis (Weising *et al.*, 2005), mas com baixo potencial para automação por requerem pessoal qualificado para o trabalho com autoradiografia (para a revelação das bandas) e para o desenvolvimento e pré-avaliação de sondas antes do início das análises, sendo o processo demorado e laborioso (Ferreira e

Grattapaglia, 1998). Também exigem grande quantidade de DNA genômico relativamente puro e íntegro (Weising *et al.*, 2005). Além disso, marcadores RFLP revelam baixo conteúdo de informação por loco, em relação a outros marcadores, sendo a maioria de natureza dimórfica (enzima corta ou não corta) (Haezebroeck *et al.*, 2001), e heterozigosidade máxima de 0,5 (Genetic Markers, 2008). Quanto aos VNTR, estes apresentam um perfil de bandas altamente complexo, dificultando a identificação das bandas (alelos) de cada loco. Com isso, a determinação de genótipo e o teste de hipóteses de segregação mendeliana ficam comprometidos. Não são, assim, os mais adequados para o mapeamento genético ou testes de ligação gênica entre marcadores e genes de interesse (Ferreira e Grattapaglia, 1998).

#### **2.3.2.2.2. Marcadores de DNA baseados em PCR**

Praticamente, todos os sistemas de marcadores moleculares usados atualmente utilizam ao menos uma etapa realizada por PCR (*Polymerase Chain Reaction* ou Reação em Cadeia da Polimerase) (Weising *et al.*, 2005). Nessa técnica, originalmente descrita por Mullis e colaboradores em meados dos anos 1980, dois oligonucleotídeos iniciadores, orientados com suas extremidades 3' apontadas um para o outro, anelados a seqüências que lhe são complementares, são usados para demarcar a região que será amplificada no DNA molde desnaturado. A adição de nucleotídeos é realizada pela enzima DNA polimerase, sendo a reação composta de três passos principais (desnaturação, anelamento e extensão) que se repetem múltiplas vezes. A cada ciclo, a quantidade de DNA é dobrada, resultando em um aumento exponencial da seqüência alvo, a qual pode ser detectada em gel, após coloração com corante adequado.

Marcadores baseados em PCR se dividem em aleatórios e sítio específicos (Ferreira e Grattapaglia, 1998; Kumar, 1999; Weising *et al.*, 2005), de acordo com o tipo de *primer* usado, estringência das condições de PCR e métodos de detecção e separação dos fragmentos (Kumar, 1999).

##### **a. Marcadores aleatórios**

Inicialmente, a técnica de PCR apresentava uma limitação para a obtenção de marcadores distribuídos ao longo dos genomas: a construção de *primers* dependia do conhecimento prévio da identidade dos nucleotídeos que flanqueavam a seqüência de interesse (Ferreira e Grattapaglia, 1998, Phillips e Vasil, 2001). Logo, passos intensivos de

mão de obra, como clonagem e seqüenciamento, eram requeridos (Ferreira e Grattapaglia, 1998). Assim, métodos que utilizavam DNA amplificado arbitrariamente (AAD, *Arbitrarily Amplified DNA* ou DNA Amplificado Arbitrariamente) foram desenvolvidos (Caetano-Anollés, 2001). Nesse sentido, três grupos, independentemente, nos primórdios dos anos 1990, sugeriram protocolos com certas particularidades em relação aos demais (Ferreira e Grattapaglia, 1998; Caetano-Anollés, 2001; Phillips e Vasil, 2001; Weising *et al.*, 2005). Os representantes dessas vertentes são: RAPD (*Random Amplified Polymorphic DNA* ou DNA Polimórfico Amplificado Randomicamente; Williams *et al.*, 1990); DAF (*DNA Amplification Fingerprinting* ou Amplificação de *Fingerprinting* de DNA; Caetano-Anollés, 1991a, 1991b) e AP-PCR (*Arbitrarily Primed PCR* ou PCR Iniciada Arbitrariamente; Welsh e McClelland, 1990). Nesses casos, tipicamente, a PCR é realizada com o uso de *primers* únicos, sob condições de baixa estringência e sem conhecimento prévio a respeito da seqüência do DNA alvo (Weising *et al.*, 2005).

Essas técnicas que utilizam *primers* de seqüências arbitrárias têm em comum os seguintes pontos: (1) produção de um padrão de bandas composto por multilocos; (2) nenhum conhecimento prévio é requerido das seqüências do DNA genômico; e (3) *primers* usados universalmente em procariotos e eucariotos (Caetano-Anollés, 2001; Weising *et al.*, 2005). Embora *a priori* nada seja sabido sobre a identidade e o contexto da seqüência de um determinado produto de PCR obtido a partir das metodologias acima, a presença ou ausência de tal produto atua como caráter informativo (Weising *et al.*, 2005). A natureza dominante constitui uma das principais desvantagens desses marcadores baseados em AAD (Kumar, 1999; Phillips e Vasil, 2001; Weising *et al.*, 2005), o que limita sua utilização em estudos de genética de população e mapeamento (Weising *et al.*, 2005). Adicionalmente, marcadores RAPD são sensíveis a leves modificações nas reações de PCR, o que interfere na reprodutibilidade do padrão de bandas entre experimentos separados e de diferentes laboratórios (Ellsworth *et al.*, 1993; Muralidharan e Wakeland, 1993; Penner *et al.*, 1993). Uma revisão mais detalhada desses marcadores pode ser vista em: Ferreira e Grattapaglia, 1998; Caetano-Anollés, 2001; Phillips e Vasil, 2001; Weising *et al.*, 2005.

## **b. Marcadores sítio-específicos**

O acúmulo de informação oriunda do sequenciamento de determinados genes e genomas organelares abriu oportunidades para utilizar a PCR específica em análises genéticas mais detalhadas (Ferreira e Grattapaglia, 1998). Marcadores específicos baseados nessa

técnica correspondem àqueles derivados de *primers* constituídos por 18 a 24 bp, com temperatura de anelamento de 50 a 70°C e com algum conhecimento sobre a sequência alvo (Kumar, 1999). Atualmente, existem diversos sistemas de marcadores [como STS (*Sequence Tagged Sites* ou Sítios de Sequências Marcadas), ESTPs (*Expressed Sequence Tags Polymorphisms* ou Polimorfismos de Etiquetas de Sequências Expressas), ACGMs (*Amplified Consensus Genetic Markers* ou Marcadores Genéticos Amplificados a partir de Sequências Consenso), SCARs (*Sequence-Characterized Amplified Region* ou Regiões Amplificadas a partir de Sequências Caracterizadas)] que apresentam as características supracitadas, cada um com vantagens e desvantagens inerentes à informação produzida e processo de detecção. Detalhes em relação a marcadores desse grupo podem-se encontrados nos artigos de Brunel *et al.* (1999), Kumar (1999), Gupta e Rustgi (2004) e nos livros dos autores Ferreira e Gratappaglia (1998) e Weising *et al.* (2005), dentre outros. Nesta dissertação, especial atenção será dada aos marcadores moleculares específicos voltados às sequências microssatélites, os quais constituem objeto de estudos desta dissertação.

### **2.3.2.2.3. Microssatélites**

Sequências microssatélites (SSR – *Simple Sequence Repeats* ou Sequências Simples Repetidas; STR – *Short Tandem Repeats* ou Pequenas Repetições Lineares; SSLP – *Simple Sequence Length Polymorphism* ou Polimorfismos em Tamanho de Sequências Simples; ou mesmo VNTR) constituem uma classe de DNA repetitivo presente em organismos eucarióticos e procarióticos (Gur-Arie *et al.*, 2000; Tóth *et al.*, 2000). Atualmente, não há um denominador comum, no que se refere ao número mínimo de repetições necessárias para a caracterização de uma sequência repetitiva em microssatélite. Na prática, o limiar usado para a identificação dessas repetições em análises de sequências, deve ser especificado (Ellegren, 2004). A maioria dos dados de literatura relata que os microssatélites consistem de grupos de repetições (motivos), compostos de um a seis nucleotídeos (Tautz, 1989), dispostos linearmente em extensões de até 100 pares de bases (Powell *et al.*, 1996; Rackoczy-Trojanowska e Bolibok, 2004).

A presença de microssatélites em vegetais foi demonstrada inicialmente por ensaios de RFLP, usando-se sondas constituídas por unidades repetitivas (Weising *et al.*, 1989; Weising *et al.*, 1992). A estratégia de detecção dessas repetições, baseada na reação de PCR, foi disponibilizada inicialmente para humanos, no final da década de 1980 (Litt e Luty, 1989; Tautz, 1989; Weber e May, 1989). A partir dos anos 1990, tornou-se uma técnica largamente

usada em estudos de genética animal e vegetal (Weising *et al.*, 2005). Para um melhor entendimento da utilidade desses marcadores em análises genéticas, é necessário um entendimento aprofundado das características desses tipos de repetições.

### **a. Estrutura e localização nos genomas**

Seqüências microssatélites podem ser classificadas de acordo com parâmetros inerentes à sua estrutura e localização nos genomas.

Quanto à estrutura, podem ser classificados em (Peakall *et al.*, 1998; Rackoczy-Trojanowska e Bolibok, 2004; Weising *et al.*, 2005; Oliveira *et al.*, 2006):

- Perfeitos: microssatélites compostos por extensões ininterruptas de um motivo particular. Ex.: ATATATATATATAT
- Imperfeitos: microssatélites interrompidos, ao longo de sua extensão, por nucleotídeos diferentes daqueles do motivo original. Ex.: ATATATGATATATATA.
- Compostos: formados por dois grupos de motivos diferentes concatenados. Ex.: ATATATATACGCGCGCGCG.

Quanto à sua localização no genoma, os microssatélites podem ser genômicos ou gênicos. A distinção é devido às diferentes forças mutacionais e seletivas que operam em regiões genômicas ou codificantes. Microssatélites genômicos evoluem neutralmente (a frequência e distribuição refletem a taxa de mutação básica do organismo ou da região genômica em que está localizado) e aqueles associados às regiões codificantes, têm sua distribuição e frequência como produto da taxa de mutação local somada a forças seletivas (Ellegren, 2004).

### **b. Gênese**

Apesar de essas seqüências serem alvos de grandes estudos, o entendimento de sua origem ainda é controverso, sendo que a imensa maioria dos trabalhos, abordando evolução de microssatélites, é realizada a partir de análises de longas repetições (Zhu *et al.*, 2000). Atualmente, existem várias linhas de pesquisa que visam entender como tais regiões surgem nos genomas. Alguns autores sugerem haver um número mínimo de repetições, a serem geradas aleatoriamente por mutações, a partir do qual ocorre a expansão dos locos microssatélites via “deslizamento” (*Replication Slippage*) da enzima DNA polimerase (Messier *et al.*, 1996; Rose e Falush, 1998). Tal situação consiste de uma pausa na atividade

dessa enzima que se dissocia da fita molde durante a replicação de uma extensão constituída por unidades repetitivas. Durante a dissociação, apenas a porção terminal da fita recém sintetizada se separa do DNA molde e se anela, subseqüentemente, em outra unidade de repetição, numa posição diferente da anterior, ocasionando a adição ou subtração de um motivo repetitivo (Levinson e Gutman, 1987; Hile e Eckert, 2004).

Por sua vez, Zhu *et al.* (2000), analisando seqüências de DNA humano, oriundas de regiões codificantes, observaram que uma significativa parcela das inserções e um montante mais restrito de substituições de bases foram responsáveis pela duplicação de seqüências adjacentes, resultando em proto-microsatélites. Os autores também sugerem que os processos que levam à expansão e polimorfismos em microsatélites já estabelecidos (dentre esses, o deslizamento da DNA polimerase) também ocorrem, em menor freqüência, em regiões com pouca ou nenhuma repetição, sem a obrigatoriedade de um número mínimo de repetições para serem executados. Por sua vez, Weising *et al.* (2005) postularam que longas repetições podem surgir repentinamente pela substituição de bases que se localizam entre duas regiões repetitivas.

Um mecanismo totalmente distinto dos anteriores foi proposto por Nadir *et al.* (1996) que especularam que microsatélites ricos em adeninas foram gerados a partir da extensão da extremidade 3' de retrotranscritos, similarmente ao mecanismo de poliadenilação de RNAs mensageiros. Entretanto, apesar da referida associação (microsatélites / transposons) ter sido encontrada em um grande número de organismos, uma alta densidade de elementos transponíveis nem sempre coincide com uma alta densidade de microsatélites (Lin *et al.*, 1999). Por isso, o processo de retrotransposição como mecanismo gerador dessas repetições é mantido em questionamento (Oliveira *et al.*, 2006).

### **c. Mutabilidade e evolução**

Embora os microsatélites venham sendo extensivamente usados em um amplo espectro de estudos, nas mais variadas áreas da genética, a dinâmica mutacional dessas regiões é pouco entendida (Schötterer, 2000). Entretanto, é sabido que a taxa mutacional dessas repetições é muito maior do que em outras partes do genoma (Weber e Wong, 1993; Jarne e Lagoda, 1996; Primmer *et al.*, 1996), e que microsatélites associados a genes também apresentam uma maior taxa de mutação quando comparados a regiões não-repetitivas, sendo que compartilham dos mesmos processos mutacionais presentes em microsatélites genômicos (Li *et al.*, 2004).

Dois mecanismos mutacionais têm sido propostos para explicar a alta mutabilidade dessas regiões. O primeiro, denominado deslizamento da DNA polimerase (*Replication Slippage* ou *Slipped Strand Mismatching*; Levinson e Gutman, 1987), é o mais amplamente aceito (Schlötterer, 2000). Sugere que modificações na extensão dos microssatélites são causadas por um processo de “deslizamento” dessa enzima, conforme apresentando anteriormente. Apenas uma fração dessas mutações primárias não é corrigida pelo sistema de reparo de erros de pareamento (*Mismatch Repair System*), culminando, assim, em um evento mutacional (Strand *et al.*, 1993). É válido ressaltar que tal mecanismo está intimamente associado ao tamanho do microssatélite, ou seja, quanto maior a extensão da região repetitiva, maior a probabilidade de ocorrência desses “deslizamentos” enzimáticos (Levinson e Gutman, 1987; Kruglyak *et al.*, 1998; Harr e Schlotterer, 2000; Ellegren, 2004). O segundo mecanismo está associado a processos similares ao de recombinação, os quais envolvem *crossing-over* desigual ou conversão gênica. Entretanto, não há consenso sobre o assunto. Enquanto alguns relatos da literatura apontam forte correlação entre tais processos e a desestabilização de repetições em *tandem* (Jakupciak e Wells, 2000; Richard e Pâques, 2002), outros demonstram o contrário, assegurando que, apesar de haver associação entre regiões de microssatélites e sítios de recombinação (Majewski e Ott, 2000), essas repetições atuam no desenvolvimento do processo e não seriam produtos do mesmo (Trecó e Arnheim, 1986).

Dados da literatura têm revelado que as taxas mutacionais dessas repetições variam consideravelmente, dependendo do loco, das seqüências flanqueadoras, do organismo, da estrutura da cromatina, da taxa local de mutação de ponto e do contexto regional em que a seqüência repetitiva está submetida (Ellegren, 2004; Weising *et al.*, 2005). Contudo, é amplamente difundido que um dos principais fatores influenciadores da taxa de mutação dessas repetições está associado à extensão das mesmas (Weber, 1990; Edwards, 1992; Ellegren, 2004; Kumpatla e Mukhopadhyay, 2005; Weising *et al.*, 2005; Mun *et al.*, 2006). Entretanto, a mutabilidade dos locos contendo longas repetições é consideravelmente reduzida quando suas extensões são interrompidas por mutações de ponto ou por repetições variantes (Eichler *et al.*, 1994; Petes *et al.*, 1997). Isto se deve ao fato de que longas extensões de repetições puras fornecerem maiores oportunidades de ocorrência do deslizamento da DNA polimerase (Levinson e Gutman, 1987; Ellegren, 2004; Weising *et al.*, 2005). Valores de taxas mutacionais divulgados para humanos e vários animais, incluindo ratos, porcos, pássaros, peixes e moscas variam de  $5 \times 10^{-6}$  a  $1,5 \times 10^{-2}$  mutações por loco por gameta por geração (Dallas, 1992; Chakraborty *et al.*, 1997; Schlötterer *et al.*, 1998; Shimoda *et al.*, 1999; Brohede *et al.*, 2002) Em vegetais, apenas algumas poucas medidas diretas da taxa de

mutação dessas repetições foram realizadas (Weising *et al.*, 2005). Diwan e Cregan (1997) relataram a formação de novos alelos numa taxa de  $2 \times 10^{-4}$  em uma população de mapeamento em soja. Thuillet *et al.* (2002) determinaram a mesma frequência ( $2 \times 10^{-4}$ ) para 10 locos microssatélites de trigo duro, entretanto a taxa de mutação de determinados locos variou de zero a  $10^{-3}$ . Adicionalmente, altas taxas também têm sido divulgadas para longas repetições do tipo  $(TAA)_n$ , com n variando de 19 a 51, em populações endogâmicas de grão-de-bico (Udupa e Baum, 2001).

Nas últimas décadas, uma ampla variedade de modelos vem sendo apresentada à comunidade acadêmica e que visam explicar a dinâmica evolucionária dos microssatélites. Muitos deles derivados do “Modelo de Evolução Gradativa” (*Stepwise Mutation Model*), proposto por Ohta e Kimura (1973). Esse modelo postula, resumidamente, que alterações na extensão das seqüências repetitivas ocorrem devido à adição ou remoção de apenas uma unidade de repetição à estrutura do microssatélite, em uma taxa fixa (Shriver *et al.*, 1993; Valdes *et al.*, 1993; Kimmel e Chakraborty, 1996). Entretanto, há relatos de ações mutacionais ocasionando modificações (adição ou contração) em mais de uma unidade de repetição (Xu *et al.*, 2000; Harr *et al.*, 2002; Huang *et al.*, 2002). Para tal situação, Di Rienzo *et al.* (1994) propuseram uma extensão do modelo anterior, na qual sugeriram que, apesar da maioria dos eventos mutacionais nessas regiões ocorrerem pela adição ou subtração de um motivo repetitivo, mudanças mais bruscas em suas estruturas, isto é, envolvendo um número maior de repetições, podem ocorrer. Contudo, os modelos acima não levam a uma distribuição estacionária das extensões dessas repetições, o que não condiz com a realidade observada genomicamente (Ellegren, 2004). Para explicar tal situação, adaptações desses modelos têm sugerido um tamanho limite para essas regiões (Nauta e Weissing, 1996; Feldman *et al.*, 1997), apesar de não ser estabelecido um possível motivo para esta restrição (Sainudiin *et al.*, 2004), enquanto outros adotaram “tendências mutacionais”, nas quais longos alelos mutam, preferencialmente, a alelos menores (Ellegren, 2000).

Atualmente, segundo vários autores, o modelo de Kruglyak *et al.* (1998) é um dos mais atraentes (Ellegren, 2004; Sainudiin *et al.*, 2004). Neste modelo, a distribuição genômica e estacionária dos microssatélites resulta do balanço de duas forças antagônicas: a) mutações de comprimento, que apresentam frequência diretamente proporcional ao tamanho da região repetitiva e favorecem o alcance de extensões arbitrariamente grandes; b) mutações de ponto, as quais quebram longas repetições em repetições menores, diminuindo, assim, seu potencial polimórfico. O referido postulado tem sido bem recebido pela comunidade acadêmica, pois explica as diferentes distribuições dos microssatélites entre espécies e fornece uma elegante

solução ao fato de microssatélites não se expandirem em enormes grupos de unidades repetitivas (Ellegren, 2004).

#### **d. Abundância e distribuição em plantas**

Seqüências microssatélites são onipresentes nos genomas animal e vegetal, sendo relatadas, também, em procariotos (Field e Wills, 1996; Hancock, 1996). Alguns autores os consideram seqüências seletivamente neutras, distribuídas de forma aleatória ou “quase” aleatória nos genomas eucromáticos (Schlötterer e Wiehe, 1999; Schlötterer, 2000). Contudo, outra vertente de pesquisadores relata uma distribuição não uniforme ao longo dos genomas. Tóth *et al.* (2000) observaram uma presença diferencial de microssatélites entre regiões intrônicas, exônicas e intergênicas, sugerindo que o processo de deslizamento da DNA polimerase não é o único a influenciar sua distribuição. Segundo Morgante *et al.* (2002), esse fator é função de restrições seletivas aliadas à dinâmica e história evolutivas dos genomas. Também sugeriram que as diferentes frações genômicas estão sujeitas a forças seletivas distintas: seqüências codificantes estariam sob seleção negativa para todos os tipos de repetição, exceto para motivos triméricos; 5' UTRs estariam sob forte seleção positiva; 3'UTR sob seleção positiva moderada; DNA repetitivo sob pressão mutacional, que se eleva com o aumento da extensão do microssatélite; e o DNA pouco repetitivo, não-transcrito, estaria, presumivelmente, em equilíbrio. Adicionalmente, tem sido observada uma freqüente associação dessas repetições a regiões ricas em genes (Morgante *et al.*, 2002; La Rota *et al.*, 2005; Mun *et al.*, 2006).

Estudos iniciais indicaram uma menor abundância de microssatélites em plantas, em relação a animais (Lagercrantz *et al.*, 1993; Morgante e Olivieri, 1993). Análises mais recentes, entretanto, têm demonstrado que a ocorrência dessas repetições nesses genomas é maior que a divulgada anteriormente (Morgante *et al.*, 2002; Varshney *et al.*, 2002; Thiel *et al.*, 2003). Cardle *et al.* (2000) analisando repetições de mono a pentanucleotídeos em 27.000 Kb de seqüências de DNA genômico de *Arabidopsis thaliana* encontraram uma freqüência média de um microssatélite a cada 6,04 Kb, que é equivalente à encontrada em mamíferos (Beckman e Weber, 1992). Em ESTs, vários valores vêm sendo relatados, observando-se uma ampla margem de variação. Varshney *et al.* (2002) encontraram uma densidade média de um microssatélites a cada 6,0 Kb, em seqüências expressas de cereais, tendo analisado 75,2 Mb de cevada, 54,7 Mb de milho, 43,9 Mb de arroz, 3,7 Mb de centeio, 41,6 Mb de sorgo e 37,5 Mb de trigo. Para monocotiledôneas, de 1,5 a 8,2 % das ESTs apresentam microssatélites,

enquanto que em dicotiledôneas a frequência varia de 2,6 a 16,8 % (Kantety *et al.*, 2002; Thiel *et al.*, 2003; Kumpatla e Mukhopadhyay, 2005). Jayashree *et al.* (2006), por sua vez, analisando 302,4 Mb de seqüências ESTs de leguminosas (espécies *M. truncatula*, *L. japonicus* e *G. max*) encontraram uma taxa de um microsatélite a cada 3,5 Kb, sendo que 10,6 % das seqüências contidas em seu banco de dados apresentaram tais repetições.

É importante ressaltar que, além da variação inerente a cada espécie (Ellegren, 2004), a frequência global e a frequência individual de diferentes motivos e tipos de repetições dependem do critério de busca, do tamanho do banco de dados, do software utilizado, da extensão e da localização genômica em questão (íntrons, exons, regiões 5' ou 3' UTR, regiões intergênicas) (Tóth *et al.*, 2000; Morgante *et al.*, 2002; Kumpatla e Mukhopadhyay, 2005; Varshney *et al.*, 2005a; La Rota *et al.*, 2005).

Adicionalmente, um acúmulo taxon-específico dessas repetições, em decorrência de uma associação preferencial de determinados tipos específicos de microsatélites a certos grupos de organismos, tem sido observado (Katti *et al.*, 2001; Mun *et al.*, 2006; Tóth *et al.*, 2000). Exemplo desse fenômeno é a associação diferencial de trinucleotídeos ricos em “GC” a ESTs de monocotiledôneas, enquanto em dicotiledôneas observa-se maior presença de trinucleotídeos ricos em “AT” (Morgante *et al.*, 2002; Gao *et al.*, 2003; Kumpatla e Mukhopadhyay, 2005; Mun *et al.*, 2006). Segundo Morgante *et al.* (2002), esse fato ocorre em virtude de um maior conteúdo “GC” em monocotiledôneas, associado a um uso preferencial de códons ricos nessa composição de bases.

Em leguminosas, análises detalhadas do conteúdo de microsatélites em ESTs têm revelado que motivos compostos por trinucleotídeos e dinucleotídeos são os mais abundantes, sendo que as repetições diméricas localizam-se, predominantemente, em regiões transcritas não-traduzidas (UTRs), enquanto que nas ORFs, há um ancoramento preferencial de trinucleotídeos (Morgante *et al.*, 2002; Jayashree *et al.*, 2006; Mun *et al.*, 2006). Segundo Metzgar *et al.* (2000), a maior ocorrência de trinucleotídeos em regiões traduzidas pode estar relacionada à forte seleção negativa atuando sobre possíveis fontes de mutações que causem mudança na matriz de leitura. No que se refere à composição de bases, os trinucleotídeos de leguminosas seguem a tendência observada em dicotiledôneas, sendo ricos em “AT”. Mais especificamente, a repetição trimérica mais abundante é composta pelo motivo “AAG” e derivações (Morgante *et al.*, 2002; Kumpatla e Mukhopadhyay, 2005; Jayashree *et al.*, 2006; Mun *et al.*, 2006). No que tange aos dinucleotídeos, tem sido observada que aqueles constituídos por “AG”, e derivações, predominam entre as demais possibilidades de combinação (Morgante *et al.*, 2002; Kumpatla e Mukhopadhyay, 2005; Mun *et al.*, 2006).

## e. Potenciais funções

O impacto funcional dos microssatélites nos genomas eucariotos é pouco entendido (Weising *et al.*, 2005). A maioria de suas cópias é considerada DNA “egoísta”, pois são amplificadas e propagadas na ausência de pressões seletivas (Orgel e Crick, 1980). Adicionalmente, nas décadas passadas, essas repetições receberam a denominação de DNA lixo (“*junk DNA*”) devido a uma suposta ausência de funcionalidade que lhes foi atribuída à época (Rackoczy-Trojanowska e Bolibok, 2004; Kashi e King, 2006). Atualmente, várias linhas de evidência apontam para possíveis efeitos moleculares e fenotípicos da variação numérica de unidades repetitivas nas referidas estruturas (Hamada *et al.*, 1984; Young *et al.*, 2000; Hammock e Young, 2004; Li *et al.*, 2004; Kashi e King, 2006). Muitos relatos têm demonstrado que uma parcela significativa dos microssatélites está localizada em regiões transcritas, incluindo genes codificadores de proteína (Morgante *et al.*, 2002; Hammock e Young, 2004; Nicot *et al.*, 2004; Varshney *et al.*, 2005a). Em um ensaio pioneiro, Marcotte *et al.* (1999) analisaram 70.882 proteínas presentes no banco de dados SWISSPROT e estimaram que 14,0 % do total continham seqüências repetidas, sendo tal abundância três vezes maior em eucariotos que em procariotos. Em ESTs, os dados quanto à presença de microssatélites são amplamente divulgados e muito significativos (Thiel *et al.*, 2003; Eujayl *et al.*, 2004; Kumpatla e Mukhopadhyay, 2005), sendo que essas seqüências são detentoras de uma gama de funções, tais como enzimas metabólicas, proteínas estruturais e de armazenamento, fatores de transcrição, dentre outras. Essa presença marcante nas referidas seqüências e em proteínas sugere que algumas dessas repetições podem estar envolvidas em processos metabólicos e de evolução gênica (Li *et al.*, 2004).

Variação na extensão de microssatélites no interior de regiões codificantes deve ser crítica para a atividade gênica, pois expansões e contrações nessas estruturas afetam, diretamente, o produto gênico correspondente, podendo causar mudanças fenotípicas (Li *et al.*, 2004). Em bactérias, tem sido observada uma interessante associação entre os chamados locos de contingência e microssatélites (Bayliss *et al.*, 2001; Moxon *et al.*, 2006). Esses locos são regiões hipervariáveis que geram, aleatoriamente, ampla variabilidade genética e são detentores de grande importância para a sobrevivência desses organismos (Bayliss *et al.*, 2001). Apesar de alterações na extensão dos microssatélites poderem causar mudanças na matriz de leitura em regiões codificantes, tais alterações podem ser prontamente reversíveis por mutações subseqüentes. Deste modo, as bactérias tiram vantagens da propriedade supracitada para utilizar as referidas repetições em *tandem* como “interruptores”, ativando ou

“desligando” determinados tipos de genes que têm participação direta no processo de variação de fase (Bayliss *et al.*, 2001; Kashi e King, 2006; Moxon *et al.*, 2006). Em vegetais, estudos sobre a possível influência de microssatélites em regiões codificantes ainda são escassos (Li *et al.*, 2004). Contudo, Ayres *et al.* (1997) observaram associações entre variações no conteúdo de amilose de várias cultivares de arroz e polimorfismos em um sítio microssatélite intimamente ligado ao gene responsável pela produção da enzima sintetizadora de tal açúcar.

Alguns experimentos têm demonstrado que microssatélites localizados nas regiões 5'UTR, 3'UTR e em íntrons de determinados genes procarióticos e humanos podem regular a expressão de tais genes (Kashi *et al.*, 1997; Toutenhoofd *et al.*, 1998; Li *et al.*, 2004). Kashi e Soler (1999), por sua vez, relataram que repetições do tipo (CT)<sub>n</sub>, localizados em regiões 5'UTR de certos genes de *Arabidopsis*, exercem um importante papel na realização da transcrição anti-senso.

Sabendo que variações em regiões microssatélites possuem o potencial de afetar a expressão gênica de maneira qualitativa ou quantitativa (Li *et al.*, 2004), e o fato de que qualquer variável genômica que afete a função genética deva ter um papel evolucionário, Kashi e King (2006) propuseram o abandono da suposição de que essas repetições sejam DNA “lixo”. Adicionalmente, Fondon e Garner (2004) sugerem que microssatélites associados a genes são fontes primordiais de variação fenotípica para a evolução, fornecendo abundante diversidade e, deste modo, capacitando a rápida evolução de novas formas. Segundo Kashi e King (2006), a caracterização metafórica dos microssatélites como “pontos de modulação evolucionária” (King *et al.*, 1997; Trifonov, 2004) expressa o potencial de tais repetições para facilitar o eficiente ajuste adaptativo de características quantitativas. Adicionalmente, esses autores acreditam que os microssatélites não apenas atuam no fornecimento de significativa variabilidade adaptativa, como também sugerem que o fornecimento de tal variação pode ter desenvolvido a “função” dos microssatélites.

#### **f. Tipos de marcadores microssatélites**

Esses marcadores são classificados de acordo com o método de confecção e associação a determinadas regiões dos genomas, podendo ser de dois tipos: marcadores microssatélites genômicos e marcadores microssatélites gênicos, estes últimos também denominados cDNA-SSRs ou EST-SSRs (Blair *et al.*, 2003; Gupta *et al.*, 2003; Chagné *et al.*, 2004; Nicot *et al.*, 2004; Rakoczy-Trojanowska e Bolibok, 2004; Varshney *et al.*, 2005a; Mun *et al.*, 2006). Apesar da dicotomia classificatória, ambos os marcadores compartilham

características vantajosas para uso em análises de seqüências de DNA, tais como: (1) herança codominante, cuja propriedade de permitir a diferenciação de organismos homozigotos de heterozigotos maximiza o conteúdo informativo por loco; (2) reprodutibilidade; (3) abundância nos genomas eucarióticos; (4) requerimento de quantidades diminutas de DNA para realização de análises, e (5) fácil intercâmbio de dados entre laboratórios (Powell *et al.*, 1996; Kumpatla e Mukhopadhyay, 2005; Weising *et al.*, 2005). Entretanto, podem apresentar peculiaridades individuais.

### **f.1. Marcadores microssatélites genômicos**

Marcadores microssatélites derivados de seqüências genômicas foram extensivamente utilizados em diversas análises genéticas, como mapeamento, análises de diversidade, estudos populacionais, identificação de cultivares etc. (Peakall *et al.*, 1998; Yang *et al.*, 1994; Rongwen *et al.*, 1995). A principal característica desses marcadores é sua natureza hipervariável, a qual permite a diferenciação de indivíduos intimamente relacionados, sendo de suma importância para aplicação em análises de populações endogâmicas, populações recentemente separadas ou geograficamente próximas e em estudos de diversidade em espécies cultivadas cuja variabilidade genética é reduzida (Powell *et al.*, 1996).

Por um longo tempo, a única forma de obtenção dessas repetições era oriunda da construção e avaliação de bibliotecas de DNA genômico e posterior seqüenciamento de clones positivos e desenho de *primers* (Weising *et al.*, 2005). Tal estratégia, “baseada em bibliotecas não-enriquecidas”, apresentava baixo rendimento, além de outras limitações inerentes à técnica (Zane *et al.*, 2002; Rackoczy-Trojanowska, 2004; Weising *et al.*, 2005). Atualmente, um grande número de protocolos tem sido desenvolvido para explorar as vantajosas características dos microssatélites genômicos e contornar as dificuldades iniciais, sendo que muitos dessas metodologias apresentam somente suaves diferenças em relação às demais (Zane *et al.*, 2002). Uma das mais utilizadas é a confecção de bibliotecas “enriquecidas para determinados tipos de microssatélites” (Armour *et al.*, 1994; Fisher e Bachmann, 1998), podendo alcançar até 80 % dos clones positivos, aumentando a produção e eficiência do processo (Weising *et al.*, 2005).

Em outro contexto, tem sido proposta a obtenção desses marcadores sem a obrigatoriedade da construção de bibliotecas genômicas, utilizando-se clones de *amplicons*, oriundos de *primers* RAPD ou AFLP (*Amplified Fragment Length Polymorphism* ou Polimorfismos de Comprimento de Fragmento Amplificado), como substrato de análise

(Cifarelli *et al.*, 1995; Zane *et al.*, 2002). Entretanto, apesar de oferecer uma alternativa menos complexa ao processo laborioso e demorado de confecção e avaliação de bibliotecas, essa técnica ainda é pouco abordada (Zane *et al.*, 2002). Assim, a principal limitação ao uso desses marcadores está associada ao seu processo de desenvolvimento. A construção de bibliotecas genômicas (“não-enriquecidas” e “enriquecidas”) e a clonagem e seqüenciamento de *amplicons* são procedimentos caros, laboriosos, demorados e que requerem mão-de-obra qualificada para sua execução (Eujayl *et al.*, 2004; Nicot *et al.*, 2004; Gutierrez *et al.*, 2005; Kumpatla e Mukhopadhyay, 2005; Wesing *et al.*, 2005). Somado a esta limitação, observa-se que os *primers* deste tipo de marcador são, freqüentemente, espécie-específicos, apresentando reduzida aplicabilidade para estudos comparativos (Röder *et al.*, 1995; Sourdille *et al.*, 2001; Pashley *et al.*, 2006; Ellis e Burke, 2007).

## **f.2. EST-SSR (cDNA-SSR) ou microssatélites gênicos**

Devido à extensa disponibilização pública de clones de cDNA e de seqüências ESTs para as mais diversas espécies, grande ênfase tem sido dada ao desenvolvimento de marcadores moleculares obtidos a partir de regiões transcritas dos genomas (Gupta e Rustgi, 2004). Assim, marcadores microssatélites também vêm sendo desenvolvidos a partir dessa tendência, aproveitando-se dos inúmeros projetos de seqüenciamento baseados na produção de ESTs (Thiel *et al.*, 2003; Pashley *et al.*, 2006; Poncet *et al.*, 2006). Tais seqüências possuem cerca de 400 pares de bases, resultantes do seqüenciamento único de uma das extremidades de um cDNA, o qual foi obtido de um mRNA gerado em um dado momento, sob um determinado estímulo (Rounsley *et al.*, 1996; Kumpatla e Mukopadhyay, 2005; Poncet *et al.*, 2006).

A descoberta de microssatélites a partir de ESTs (denominados EST-SSR) é realizada através da mineração de dados, utilizando-se técnicas de bioinformática (Kumpatla e Mukopadhyay, 2005; Varshney *et al.*, 2005; Sharma *et al.*, 2007). Esse processo é rápido, acessível, economicamente viável e apresenta imenso potencial para automatização, aumentando, sobremaneira, a eficiência na geração desses marcadores (Gupta *et al.*, 2003; Kumpatla e Mukopadhyay, 2005; Varshney *et al.*, 2005). Essa abordagem se mostra mais representativa quando comparada à metodologia baseada em bibliotecas “enriquecidas para determinados tipos de microssatélites”, uma vez que são realizadas análises de todos os tipos e extensões de motivos e repetições existentes no sistema em questão (Leigh *et al.*, 2003). Comparativamente, além das vantagens técnicas sobre os marcadores microssatélites

genômicos, os EST-SSRs demonstram ser mais úteis na obtenção dos resultados, produzindo dados de melhor qualidade, isto é, bandas mais fortes ou com picos alélicos bem definidos e de mais fácil interpretação (Eujayl *et al.*, 2001; Leigh *et al.*, 2003; Woodhead *et al.*, 2005; Pahlsey *et al.*, 2006).

Microssatélites baseados em ESTs são poderosas ferramentas para análise de variação genética, estudos evolutivos, mapeamento e análise de características quantitativas (Cato *et al.*, 2001; Scott *et al.*, 2001; Holton *et al.*, 2002). De acordo com vários autores, a aplicação de EST-SSRs (ou marcadores oriundos de regiões codificantes, em geral) em estudos de avaliação de germoplasma acentua a resolução da análise, uma vez que o polimorfismo analisado advém de regiões transcritas, revelando, assim, a “verdadeira” diversidade genética (Maestri *et al.*, 2002; Yamanaka *et al.*, 2003; Gupta *et al.*, 2003). Desse modo, esses marcadores apresentam implicações diretas em programas de melhoramento vegetal, onde os parentais necessitam ser selecionados com base na distância genética (Gupta e Rustgi, 2004). Tais marcadores vêm sendo amplamente utilizados para essa finalidade em diversos vegetais, como trigo (Eujayl *et al.*, 2001), ervilha (Burstin *et al.*, 2001), cevada (Holton *et al.*, 2002; Thiel *et al.*, 2003), centeio (Hackauf e Wehling, 2002), etc.

Marcadores EST-SSRs, em virtude da influência de microssatélites na regulação gênica (Li *et al.*, 2004) e de sua associação física a regiões transcritas, apresentam potencial para o desenvolvimento de marcadores funcionais (FM, *Functional Markers*), ou seja, marcadores oriundos de sítios polimórficos que estão envolvidos diretamente na variação fenotípica de uma dada característica e que tem poder de diagnóstico (Andersen e Lübberstedt, 2003; Varshney *et al.*, 2005a). Tais marcadores são “bons tradutores” dos avanços da tecnologia genômica para espécies cultivadas (Bagge *et al.*, 2007), sendo superiores aos marcadores aleatórios e gene-específicos não associados a motivos funcionais (tecnicamente denominados GTMs ou *Gene Target Markers*) (Andersen e Lübberstedt, 2003).

Uma das restrições enfrentadas pelos EST-SSRs refere-se ao seu potencial polimórfico. Dados da literatura têm mostrado que esses marcadores tendem a apresentar um menor grau de polimorfismo, quando comparados à sua contraparte genômica, devido à maior conservação de seqüências de DNA localizadas em regiões transcritas, provavelmente devido à ação de forças seletivas (Scott *et al.*, 2000; Morgante *et al.*, 2002; Gupta *et al.*, 2003; Liewlaksaneeyanawin *et al.*, 2004; Chabane *et al.*, 2005; Varshney *et al.*, 2005a). Entretanto, alguns estudos têm divulgado um significativo grau de polimorfismo obtido a partir de análises com esses marcadores. Fraser *et al.* (2004) observaram que 93,5% dos EST-SSRs foram polimórficos e segregaram em uma população de mapeamento derivada de um

cruzamento entre dois genótipos diplóides de *Actinidia chinensis*. Outros trabalhos também demonstraram altos índices polimórficos oriundos desses marcadores (Eujayl *et al.*, 2004; Chagné *et al.*, 2004), havendo casos onde tais índices para microssatélites genômicos e EST-SSRs se equiparam (Leigh *et al.*, 2003; Poncet *et al.*, 2006).

Há diversos fatores indicando que, num futuro próximo, os EST-SSRs complementarão e ultrapassarão numericamente os marcadores microssatélites genômicos obtidos a partir de metodologias desenvolvidas em décadas passadas (Leigh *et al.*, 2003). Segundo Gupta e Rustgi (2004), marcadores oriundos de análises do transcriptoma se tornarão mais úteis do que os obtidos pelos métodos antigos. Entretanto, uma importante limitação ao uso desses marcadores está relacionada à disponibilidade de seqüências ESTs nos bancos de dados públicos, sendo que somente espécies de interesse científico ou que possuem importância econômica possuem um montante considerável de seqüências disponível para estudos (Eujayl *et al.*, 2004; Rakoczy-Trojanowska e Bolibok, 2004). Este fato ressalta a importância dos estudos de transferibilidade desses marcadores em espécies correlacionadas.

### **f.2.1. Transferibilidade de marcadores EST-SSRs**

Define-se transferibilidade de marcadores EST-SSRs como a propriedade de determinados pares de *primers*, desenvolvidos para flanquear regiões microssatélites em seqüências de uma dada espécie, também o realizarem em outra, filogeneticamente relacionada (Kuleung *et al.*, 2004). O intercâmbio dessa informação genética pode ser realizado devido à sintenia [conservação da ordem e conteúdo gênicos (Mudge *et al.*, 2005)] existente entre espécies aparentadas. Apesar de algumas exceções, decorrentes de rearranjos estruturais (Bennetzen e Ramakrishna, 2002; Song *et al.*, 2002; Zhu *et al.*, 2005), a sintenia é a principal característica dos genomas vegetais (Kumar, 1999), sendo essa inversamente proporcional à distância filogenética entre as espécies sob análise (Kumar, 1999; Kaló *et al.*, 2004). De acordo com alguns autores, a conservação do conteúdo e ordem dos genes, em diferentes espécies, são indicativos de que esses segmentos são vantajosos às plantas (portanto, não seriam submetidos a profundas modificações) além de, cientificamente, possuir caráter prático, atuando como ponte de ligação por onde transita a informação genética entre organismos afins, diminuindo a distância entre espécies bem estudadas e espécies com disponibilidade de dados mais restrita (Kumar, 1999; Zhu *et al.*, 2005).

Vários estudos relatam extensiva sintonia genômica entre membros das famílias Solanaceae (Tanksley *et al.*, 1992; Livingstone *et al.*, 1999), Brassicaceae (Lagercrantz e Lydiate, 1996; Osborn *et al.*, 1997; Lagercrantz 1998; Lukens *et al.*, 2003), Poaceae (Devos e Gale, 1997; Gale e Devos, 1998; Keller e Feuillet, 2000), e outras. Aliás, o fenômeno de sintonia foi inicialmente demonstrado entre tomate e batata (Bonierbale *et al.*, 1988; Gebhardt *et al.*, 1991) e entre tomate e pimenta (Tanksley *et al.*, 1988). Em leguminosas, vários estudos relativos ao assunto têm sido divulgados. Comparações entre ervilha (*Pisum sativum* L.) e lentilha (*Lens culinaris* Medik.) (Weeden *et al.*, 1992); ervilha e grão-de-bico (*Cicer arietinum* L.) (Simon e Muehlbauer, 1997); feijão-mungo (*Vigna radiata* L.) e feijão-caupi (Menancio-Hautea *et al.*, 1993); feijão-mungo e lablab [*Lablab purpureus* (L.) Sweet] (Humphry *et al.*, 2002); alfafa (*Medicago sativa* L.) e ervilha (Kaló *et al.*, 2004); soja (*Glycine max*), *Arabidopsis* [*Arabidopsis thaliana* (L.) Heynh.] e *Medicago truncatula* (Mudge *et al.*, 2005); *Medicago truncatula* e *Lotus japonicus* (Cannon *et al.*, 2006), e entre diversas espécies pertencentes à subfamília Papilionoidea (Choi *et al.* 2004a, 2004b) comprovam a presença de tal fenômeno, destacando também rearranjos estruturais que podem fundamentar a variação no número de cromossomos entre essas espécies. Mais especificamente, Boutin *et al.* (1995) demonstraram que os mapas genéticos de feijão-mungo, feijão-comum e soja também exibiram grupos de ligação conservados, sendo que a sintonia era mais substancial entre as duas primeiras espécies. Um estudo mais recente, entretanto, usando *Arabidopsis thaliana* como espécie ligadora (âncora), revelou que segmentos homoeólogos dos cromossomos de soja apresentaram um maior grau de sintonia com os cromossomos de feijão-comum e feijão-mungo do que divulgado anteriormente (Lee *et al.*, 2001). O mesmo seria esperado entre soja, feijão-comum e feijão-caupi, visto que entre essa última espécie e feijão-mungo um alto grau de conservação foi observado (Menancio-Hautea *et al.*, 1993).

Segundo Varshney *et al.* (2005a), a principal característica dos EST-SSRs está associada ao seu potencial de transferibilidade. Como tais marcadores são derivados de regiões transcritas, espera-se que esses exibam maior mobilidade interespecífica quando comparados aos microssatélites genômicos (Scott *et al.*, 2000). Tal suposição vem sendo amplamente corroborada em estudos comparativos entre esses dois tipos de marcadores (Chagné *et al.*, 2004; Liewlaksaneeyanawin *et al.*, 2004; Gutierrez *et al.*, 2005; Pashley *et al.*, 2006). Um sério problema relativo ao uso interespecífico de microssatélites, está relacionado ao alto nível de alelos nulos, isto é, alelos não amplificados (Paetkau e Strobeck, 1995; Pemberton *et al.*, 1995; Pashley *et al.*, 2006). Indivíduos homozigotos para alelos nulos não

apresentam bandas no gel, enquanto que heterozigotos apresentam apenas uma banda, simulando um caso de homozigose (Weising *et al.*, 2005). Assim, esses “alelos” podem influenciar a interpretação de dados de herança, aumentando os níveis aparentes de endogamia (Kijas *et al.*, 1997; DeWoody *et al.*, 2006). Entretanto, tal ocorrência é menos freqüente em EST-SSRs, quando comparados aos microssatélites genômicos, devido à sua região de ancoragem (Leigh *et al.*, 2003; Liewlaksaneeyanawin *et al.*, 2004; Rungis *et al.*, 2004).

Atualmente, a transferibilidade desses marcadores vem sendo amplamente analisada. Em estudo recente, Poncet *et al.* (2006) obtiveram 94 % de sucesso amplificando pares de *primers* desenhados para *Coffea canephora* Pierre ex A. Froehner em outras espécies do gênero *Coffea*. Por sua vez, Pashley *et al.* (2006) obtiveram 73% de sucesso na transferência de pares de *primers* EST-SSRs de *Helianthus annuus* L. para *H. angustifolius* L. e *H. verticillatus* Small. Em gramíneas, estudos têm estimado que 44 a 60% dos EST-SSRs oriundos de trigo e cevada produziram *amplicons* em arroz (Holton *et al.*, 2002; Gao *et al.*, 2003; Thiel *et al.*, 2003; Yu *et al.*, 2004; Varshney *et al.*, 2005b). Saha *et al.*, 2004 desenvolveram pares de *primers* para *Festuca arundinacea* Schreb. e relataram que 59% desses amplificaram em arroz, enquanto que 71% foram funcionais em trigo. Similarmente, em leguminosas, 96% dos pares de *primers* desenhados para *Medicago truncatula* produziram *amplicons* em seis outras espécies pertencentes ao gênero *Medicago* (Eujayl *et al.*, 2004). Utilizar-se dessa característica para desenvolver marcadores EST-SSR para uso em culturas com disponibilidade de dados restrita, como é o caso do feijão-caupi, é primordial para alavancar o melhoramento dessa cultura importante para a população nordestina e também africana.

### 3. REFERÊNCIAS BIBLIOGRÁFICAS

- Andersen JR e Lübberstedt (2003) Functional markers in plants. *Trends in Plant Science* 8: 554-560.
- Anuário Estatístico ad Agricultura Brasileira. *Agrianual* 98. São Paulo: FNP Consultoria & Comércio, 1998. pp 247-253.
- Araújo JPP e Watt EE (1988) O caupi no Brasil. 1st edition. IITA/Embrapa, Brasília, 722 pp.
- Armour JAL, Neumann R, Gobert S e Jeffreys AJ (1994) Isolation of human simple repeat locos by hybridization selection. *Human Molecular Genetics* 3: 599-605.
- Arumuganathan K e Earle ED (1991) Nuclear DNA content of some important plant species. *Plant Molecular Biology Reporter* 9: 208-218.
- Ayres NM, McClung AM, Larkin PD, Bligh HFJ, Jones CA e Park WD (1997) Microsatellite and single nucleotide polymorphism differentiate apparent amylose classes in an extended pedigree of US rice germ plasm. *Theoretical and Applied Genetics* 94: 773-781.
- Bagge M, Xia X e Lübberstedt T (2007) Functional markers in wheat. *Current opinion in Plant Biology* 10: 211-216.
- Baudoin JP e Maréchal R (1985) Genetic diversity in *Vigna*. In: Singh SR, Rachie KO (eds) Cowpea research, production, and utilization. Wiley, New York, pp 3–11.
- Barnum SR (1998) *Biotechnology: An Introduction*. 1st edition. Wadsworth Publishing Company, Belmont-California, 360 pp.
- Bayliss CD, Field D e Moxon ER (2001) The simple sequence contingency locos of *Haemophilus influenzae* and *Neisseria meningitidis*. *The Journal of Clinic Investigation* 107: 657-666.
- Beckman JS e Weber JL (1992) Survey of human and rat microsatellites. *Genomics* 12: 627-631.
- Bennetzen JL e Ramakrishna W (2002) Numerous small rearrangements of gene content, order and orientation differentiate grass genomes. *Plant Molecular Biology* 48: 821-827.
- Bezerra AAC (1997) Variabilidade e diversidade genética em caupi (*Vigna unguiculata* (L.) Walp.) precoce, de crescimento determinado e porte ereto e semi-ereto, Universidade Federal de Pernambuco, Dissertação de Mestrado, Recife, 105 pp.
- Blair MW, Pedraza F, buendia HF, Gaitán-Solís, Beebe SE, Gepts P e Tohme J (2003) Development of genome-wide anchored microsatellite map for common bean (*Phaseolus vulgaris* L.). *Theoretical and Applied Genetics* 107: 1362-1374.

Boeh-Ocansey O (1989) Developments and challenges in Africa's food industry. *Food Technology*. 43: 84-92.

Bonierbale MW, Plaisted RL e Tanksley SD (1988) RFLP maps based on common sets of clones reveal modes of chromosome evolution in potato and tomato. *Genetics*: 120:1095-103.

Botstein D, White RL, Skolnick M e Davis RV (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* 32:314-331.

Boutin SR, Young ND, Olsen TC, Yu ZH, Shoemaker RC e Valejos CE (1995) Genome conservation among three legume genera detected with DNA markers. *Genome* 38: 928-937.

Bressani R (1985) Nutritive value of cowpea. In: Singh SR e Rachie KO (eds) *Cowpea Research, Production and Utilization*. 1st edition. Wiley, New York, pp 353-359.

Brohede J, Primmer CR, Möller A e Ellegren H (2002) Heterogeneity in the rate and pattern of germline mutation at individual microsatellite loci. *Nucleic Acids Research* 30: 1997-2003.

Brunel D, Froger N e Pelletier G (1999) Development of amplified consensus genetic markers (ACGM) in *Brassica napus* from *Arabidopsis thaliana* sequences of known biological function. *Genome* 42: 387-402.

Burstin J, Deniot G, Potier J, Weinachter C, Aubert G e Baranger A (2001) Microsatellite polymorphism in *Pisum sativum*. *Plant Breeding* 120: 311-317.

Caetano-Anollés G (2001) Plant genotyping using arbitrarily amplified DNA. In: Henry RJ (ed) *Plant Genotyping: The Fingerprinting of Plants*. 1st edition. CABI Publishing, Wallingford, pp 29-46.

Caetano-Anollés G, Bassam BJ e Gresshoff PM (1991a) DNA amplification fingerprinting: a strategy for genome analysis. *Plant Molecular Biology Reporter* 9: 294-307.

Caetano-Anollés G, Bassam BJ e Gresshoff PM (1991b) High resolution DNA amplification fingerprinting using very short arbitrary oligonucleotide primers. *Bio/Technology* 9: 553-557.

Cannon SB, Sterck L, Rombauts S, Sato S, Cheung F, Gouzy J, Wang X, Mudge J, Vasdewani J, Schiex T, Spannagl M, Monaghan E, Nicholson C, Humphray SJ, Schoof H, Mayer KFX, Rogers J, Quétier F, Oldroyd GE, Debelle F, Cook DR, Retzel EF, Roe BA, Town CD, Tabata S, Van der Peer Y e Young ND (2006) Legume genome evolution viewed through the *Medicago truncatula* and *Lotus japonicus* genomes. *Proceedings of National Academy of Sciences* 103: 14959-14964.

Cardle L, Ramsay L, Milbourne D, Macaulay M, Marshall D e Wagh R (2000) Computational and experimental characterization of physically clustered simple sequence repeats in plants. *Genetics* 156: 847-854.

Cato SA, Gardner RC, Kent J e Richardson TE (2001) A rapid PCR-based for genetically mapping ESTs. *Theoretical and Applied Genetics* 102: 296-306.

Caupi movimentada mais de US\$ 249 milhões no Brasil. *Embrapa Meio-Norte, Teresina*, v. 3, n. 3, p. 4-5, jul. 2001.

Chabane K, Ablett GA, Cordeiro GM, Valkoun J e Henry RJ (2005) EST versus genomic derived microsatellite markers for genotyping wild and cultivated barley. *Genetic Resources and Crop Evolution* 52: 903-909.

Chagné D, Chaumeil P, Ramboer A, Collada C, Guevara A, Cervera MT, Vendramin GG, Garcia V, Frigerio J-M, Echt C, Richardson T e Plomion C (2004) Cross-species transferability and mapping of genomic and cDNA SSRs in pines. *Theoretical and Applied Genetics* 109: 1204-1214.

Chakraborty R Kimmel M, Stivers L, Davidsen LJ e Deka R (1997) Relative mutation rates at di, tri and tetranucleotide microsatellite locos. *Proceedings of the National Academy of Sciences* 94: 1041-1046.

Chen X, Laudeman TW, Rushton PJ, Spraggins TA e Timko MP (2007) CGKB: an annotation knowledge for cowpea (*Vigna unguiculata* L.) methylation filtered genomic genespace sequences. *BMC Bioinformatics* 8: 129-137.

Choi H-K, Kim D, Uhm T, Limpens E, Lim H, Mun J, Kalo P, Penmetsa RV, Seres A, Kulikova O, Roe BA, Bisseling T, Kiss B e Cook DR (2004b) A sequence-based genetic map of *Medicago truncatula* and comparison of marker co-linearity with *Medicago sativa*. *Genetics* 166: 1463-1502.

Choi H-K, Mun J-H, Kim D-J, Zhu H, Baek J-M, Mudge J, Roe B, Ellis N, Doyle J, Kiss GB, Young ND e Cook DR (2004a) Estimating genome conservation between crop and model legume species. *Proceedings of National Academy of Sciences* 101: 15289-15294.

Cifarelli RA, Gallitelli M e Celline F (1995) Random amplified hybridization microsatellite (RAHM): isolation of a new class of microsatellite-containing DNA clones. *Nucleic Acids Research* 23: 3802-3803.

Coulibaly S, Pasquet RS, Papa R e Gepts R (2002) AFLP analysis of the phenetic organization and genetic diversity of *Vigna unguiculata* L. Walp. reveals extensive gene flow between wild and domesticated types. *Theoretical and Applied Genetics* 104: 358-366.

Cover Crop Database [http://www.sarep.ucdavis.edu/cgi-bin/CCrop.exe/show\\_crop\\_15](http://www.sarep.ucdavis.edu/cgi-bin/CCrop.exe/show_crop_15) (January 31, 2008)

Cowpea Genomics, <http://cowpeagenomics.med.virginia.edu/> (Janeiro, 2008).

Dadson RB, Hashem FM, Javaid I, Joshi J, Allen AI e Devine TE (2005) Effect of water stress on the yield of cowpea (*Vigna unguiculata* L. Walp.) genotypes in the Delmarva region of the United States. *Journal of Agronomy and Crop Science* 191: 210-217.

Dallas JF (1992) Estimation of microsatellite mutation rates in recombinant inbred strains of mouse. *Mammalian Genome* 3: 452-456.

Dantas JP, Marinho FJL, Ferreira MMM, Amorim MSN, Andrade SIO e Sales AL (2002) Avaliação de genótipos de caupi sob salinidade. *Revista Brasileira de Engenharia Agrícola Ambiental* 6: 425-430.

Devos KM e Gale MD (1997) Comparative genetics in the grasses. *Plant Molecular Biology* 35: 3-15.

DeWoody J, Nason JD e Hipkins VD (2006) Mitigating scoring errors in microsatellite data from wild populations. *Molecular Ecology Notes* 6: 951-957.

Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M e Freimer NB (1994) Mutational processes of simple sequence repeat locos in human population. *Proceedings of the National Academy of Sciences* 91: 3166–3170.

Diwan N e Cregan PB (1997) Automated sizing of fluorescent-labeled simple sequence repeat (SSR) markers to assay genetic variation in soybean. *Theoretical and Applied Genetics* 95: 723-733.

Doyle JJ e Luckow MA (2003) The rest of the iceberg. Legume diversity and evolution in a phylogenetic context. *Plant Physiology* 131: 900-910.

Dunham RA (2004) *Aquaculture and Fisheries Biotechnology: Genetic Approaches*. 1st edition. CABI Publishing, Wallingford, 384 pp.

Edwards A, Hammond HA, Jin L, Caskey CT, Chakraborty (1992) Genetic variation at five trimeric and tetrameric *tandem* repeat locos in four human population groups. *Genomics* 12: 241-253.

Ehlers JD e Hall AE (1997) Cowpea (*Vigna unguiculata* L. Walp.). *Field Crops Research* 53: 187-204.

Eichler EE, Holden JJA, Popovich BW, Reiss AL, Snow K, Thibodeau SN, Richards S, Ward PA e Nelson DL (1994) Length uninterrupted CGG repeats determines instability in the FMR1 gene. *Nature Genetics* 8: 88-94.

Ellegren H (2000) Heterogeneous mutation processes in human microsatellite DNA sequences. *Nature Genetics* 24: 400-402.

Ellegren H (2004) Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics* 5: 435-445.

Ellis JR e Burke JM (2007) EST-SSRs as a source for population genetic analyses. *Heredity* 99: 125-132.

Ellsworth DL, Rittenhouse KD and Honeycutt RL (1993) Artifactual variation in randomly amplified polymorphic DNA banding patterns. *Biotechniques* 14: 214-217.

Elowad HOA e Hall AE (1987) Influences of early and late nitrogen fertilization on yield and nitrogen fixation of cowpea under well-watered and dry field conditions. *Field Crops Research* 15: 229-244.

Embrapa. Centro Nacional de Pesquisa de Arroz e Feijão (Goiânia, GO). Catálogo descritivo de germoplasma de caupi (*Vigna unguiculata* (L.) Walp.). Goiânia, 1990. 16p. (EMBRAPA-CNPAF. Documentos, 31).

Eujayl I, Sledge MK, Wang L, May GD, Chekhovskiy K and Zwonitzer JC e Mian MAR (2004) *Medicago truncatula* EST-SSRs reveal cross-species genetic markers for *Medicago* spp. *Theoretical and Applied Genetics* 108: 414-422.

Eujayl I, Sorrels ME, Baum M, Wolters P e Powell W (2001) Assessment of genotypic variation among cultivated durum wheat based on EST-SSRs and genomic SSRs. *Euphytica* 119: 414-422.

Feldman MW, Bergman A, Pollock DD e Goldstein DB (1997) Microsatellite genetic distances with range constraints: analytic description and problems of estimation. *Genetics* 145: 207-216.

Ferreira ME and Grattapaglia D (1998) Introdução ao uso de marcadores moleculares em análise genética. 3ª edição. EMBRAPA-CENARGEN, Brasília, 220 pp.

Fery RL (1985) Improved cowpea cultivars for the horticultural industry in the USA. In: Singh SR and Rachie KO (eds) *Cowpea Research, Production, and Utilization*. 1st edition. Willey, New York, pp 127-135.

Fery RL, (1990) The cowpea: production, utilization, and research in United States. *Horticulture Review* 12: 197-222.

Fery RL (2002) New opportunities in *Vigna*. In: Janick J and Whipkey A (eds) *Trends in New Crops and New Uses*, ASHS Press, Alexandria, pp 424-428.

Field D e Wills C (1996) Long polymorphic microsatellites in simple organisms. *Proceedings of the Royal Society* 263: 209-215.

Fisher D e Bachmann K (1998) Microsatellite enrichment in organism with large genomes (*Allium cepa*). *Biotechniques* 24:796-802.

Fondon JW, III e Garner HR (2004) Molecular origins of rapid and continuous morphological evolution. *Theoretical and Applied Genetics* 101: 18058-18063.

Frahm-Leliveld JA (1965) Cytological data on some wild tropical *Vigna* species and cultivars from cowpea and asparagus bean. *Euphytica* 14: 251-270.

Fraser LG, Harvey CF, Crowhurst RN e Silva HN (2004) EST-derived microsatellites from *Actinidia* species and their potential for mapping. *Theoretical and Applied Genetics* 108: 1010-1016.

Freire Filho FR, Cardoso MJ, Ribeiro VQ, Athayde Sobrinho C e Silva PHS. Introdução, avaliação e utilização de germoplasma de feijão macassar (*Vigna unguiculata* (L.) Walp.). Relatório Técnico Anual da Unidade de Execução de Âmbito Estadual de Teresina 1990. Teresina, p 126-131, 1991.

Freire Filho FR, Cardoso MJ, Ribeiro VQ. Variabilidade Genética e capacidade de combinação em feijão macassar (*Vigna unguiculata* (L.) Walp.) In: Seminário de Pesquisa Agropecuária do Piauí, 6., 1990, Teresina. Anais... Teresina: Embrapa-UEPAE de Teresina, 1992. p. 219-227.

Freire Filho FR, Ribeiro VQ, Barreto PD and Santos CAF (1999) Melhoramento genético do caupi (*Vigna unguiculata* (L.) Walp.) na região Nordeste. In: Queiroz MA, Goedert CO; Ramos SRR (Org) Recursos genéticos e melhoramento de plantas para o nordeste brasileiro. 1st edition. Embrapa-CPATSA, Petrolina, pp.

Gale MD e Devos KM (1998) Comparative genetics in the grasses. *Proceedings of National Academy of Sciences* 95: 1971-1974.

Gao L, Tang J Li H e Jia J (2003) Analysis of microsatellites in major crops assessed by computational and experimental approaches. *Molecular Breeding* 12: 245-261.

Gebhardt C, Ritter E, Barone A, Debener T, Walkemeier B, Schachtschabel U, Kaufmann H, Thompson RD, Bonierbale MW, Ganai MW, Tanksley SD e Salamini F (1991) RFLP maps of potato and their alignment with the homoeologous tomato genome. *Theoretical and Applied Genetics* 83: 49-57.

Genetic Markers, <http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=hmg.section.1387> (Janeiro, 2008).

Graham PH e Vance CP (2003) Legumes: importance and constraints to greater use. *Plant Physiology* 131: 872-877.

Gupta PK e Rustgi S (2001) Molecular markers from the transcribed / expressed region of the genome in the higher plants. *Functional & Integrative Genomics* 4: 139-162.

Gupta PK, Rustgi S, Sharma S, Singh R, Kumar N e Balyan HS (2003) Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. *Molecular Genetics and Genomics* 270: 315-323.

Gur-Arie R, Cohen CJ, Eitan Y, Shelef L, Hallerman EM e Kashi Y (2000) Simple sequence repeats in *Escherichia coli*: abundance, distribution, composition, and polymorphism. *Genome Research* 10: 62-71.

Gutierrez MV, Patto MCV, Huguet T, Cubero JI, Moreno MT e Torrez AM (2005) Cross-species amplification of *Medicago truncatula* microsatellites across three major pulse crops. *Theoretical and Applied Genetics* 110: 1210-1217.

Hackauf B e Wehling P (2002) Identification of microsatellite polymorphism in an expressed portion of the rye genome. *Plant Breeding* 121: 17-25.

Haezebroeck V, Renaville R, Bertozzi C, Parmentier I, Pirard M e Portetelle D (2001) Molecular traceability of animals and their products. In: Renaville R e Burny A (eds) *Biotechnology in Animal Husbandry*. 1st edition. Kluwer Academic Publishers, Dordrech, pp 333-344.

Hamada H, Seidman M, Howard BH e Gorman CM (1984) Enhanced gene expression by the poly (dT-dG) poly (dC-dA) sequences. *Molecular and Cellular Biology* 4:2622-2630.

Hammock EAD e Young LJ (2004) Functional microsatellite polymorphism associated with divergent social structure in voles species. *Molecular Biology and Evolution* 21: 1057-1063.

Hancock JM (1996) Simple sequences in “minimal” genome. *Nature Genetics* 14: 14-15.

Harr B e Schlotterer C (2000) Long microsatellite alleles in *Drosophila melanogaster* have a downward mutation bias and short persistence times, which cause their genome-wide underrepresentation. *Genetics* 155: 1213-1220.

Harr B, Todorova J e Schlotterer C (2002) Mismatch repair-driven mutational bias in *D. Melanogaster*. *Molecular Cell* 10: 199-205.

Henriet J, van Ek GA, Blade SF e Singh BB (1997) Quantitative assessment of traditional cropping systems in the Sudan savanna of northern Nigeria. I. Rapid survey of prevalent cropping systems. *Samaru Journal of Agricultural Research*. 14: 37-45.

Hile SE e Eckert KA (2004) Positive correlation between DNA polymerase  $\alpha$ -primase pausing and mutagenesis within polypyrimidine / polypurine microsatellite sequences. *Journal of Molecular Biology* 335: 745-759.

Holton TA, Christopher JT, McClure L, Harker N e Henry RJ (2002) Identification and mapping of polymorphic SSR markers from expressed sequence of barley and wheat. *Genetics* 164: 655-664.

Huang Q-Y, Xu F-H, Shen H, Deng H-Y, Liu Y-J, Liu Y-Z, Li J-L, Recker RR e Deng H-W (2002) Mutation patterns at dinucleotide microsatellite locos in humans. *American Journal of Human Genetics* 70: 625–634.

Humphry ME, Konduri V, Lambrides CJ, Magner T, McIntyre CL, Aitken EAB e Liu CJ (2002) Development of a mungbean (*Vigna radiata*) RFLP linkage map and its comparison with lablab (*Lablab purpureus*) reveals a high level of colinearity between the two genomes. *Theoretical and Applied Genetics* 105: 106-166.

Hunter RL e Markert CL (1957) Histochemical demonstration of enzymes separated by zone electrophoresis in starch gels. *Science* 125: 1294-1295.

Ikea J, Ingelbrecht I, Uwaifo A e Thottappilly G (2003) Stable gene transformation in cowpea (*Vigna unguiculata* L. walp.) using particle gun method. *African Journal of Biotechnology* 2: 211–218.

Jakupciak JP e Wells RD (2000) Gene conversion (recombination) mediates expansions of CTG.CAG repeats. *Journal of Biological Chemistry* 275: 4003-4013.

Jarne P e Lagoda PJJ (1996) Microsatellites, from molecules to populations and back. *Trends in Ecology and Evolution* 11: 424-429.

Jayashree B, Punna R, Prasad P, Bantte K, Hash CT, Chandra S, Hoisington DA e Varshney RK (2006) A database of simple sequence repeats from cereal and legume expressed sequence tags mined in silico: survey and evaluation. *In Silico Biology* 6. <http://www.bioinfo.de/isb/2006/06/0054/>.

Kaló P, Seres A, Taylor SA Jakab J, Kevei Z, Kereszt A, Endre G, Ellis THN e Kiss GB (2004) Comparative mapping between *Medicago sativa* and *Pisum sativum*. *Molecular Genetics and Genomics* 272: 235-246.

Kantety RV, La Rota M, Matthews DE e Sorrells ME (2002) Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. *Plant Molecular Biology* 48: 501-510.

Kashi Y e King DG (2006) Simple sequence repeats as advantageous mutators in evolution. *Trends in genetics* 22: 253-259.

Kashi Y e Soller M (1999) Functional roles of microsatellites and minisatellites. In: Goldstein DB and Schlötterer C (eds) *Microsatellites: Evolution and Applications*. 1st edition. Oxford University Press, Oxford, pp 10-23.

Kashi Y D, King D e Soller M (1997) Simple sequence repeats as a source of quantitative genetic variation. *Trends in Genetics* 13: 74-78.

Katti MV, Ranjekar PK e Gupta VS (2001) Differential distribution of simple sequence repeats in eukaryotic genome sequences. *Molecular Biology and Evolution* 18: 1161-1167.

Keller B e Feuillet C (2000) Colinearity and gene density in grasses genomes. *Trends in Plant Science* 5: 246-251.

Kijas JMH, Thomas MR, Fowler JCS e Roose ML (1997) Integration of trinucleotide microsatellite into a linkage map of citrus. *Theoretical and Applied Genetics* 94: 701-706.

Kimmel M e Chakraborty R (1996) Measures of variation at DNA repeat locos under a general stepwise mutation model. *Theoretical Population Biology* 50: 345-367.

King DG, Soller M e Kashi Y (1997) Evolutionary tuning knobs. *Endeavour* 21: 36-40.

Kruglyak S, Durrett RT, Schug M D e Aquadro CF (1998) Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. *Proceedings of the National Academy of Sciences* 95: 10774-10778.

Kuleung C, Baezinger PS e Dweikat I (2004) Transferability of SSR markers among wheat, rye, and triticale. *Theoretical and Applied Genetics* 108: 1147-1150.

Kumar LS (1999) DNA markers in plant improvement: An overview. *Biotechnology Advances* 17: 143-182.

Kumapata SP e Mukhopadhyay S (2005) Mining and survey of simple sequence repeats in expressed sequence tags of dicotyledonous species. *Genome* 48: 985-998.

Kuykendall LD, Hashem FM, Dadson RB e Elkan GK (2000) Nitrogen fixation. In: Ledeborg J (ed) *Encyclopedia of Microbiology*, Vol 3. 2nd edition. Academic Press, New York, pp 329-404.

Kwapata MB e Hall AE (1985) Effects of moisture regime and phosphorus on mycorrhizal infection, nutrient uptake, and growth of cowpea (*Vigna unguiculata* (L.) Walp.). *Field Crop Research* 12: 241-250.

La Rota M, Kantety RV, Yu J-K e Sorrells ME (2005) Nonrandom distribution and frequencies of genomic and EST-derived microsatellite markers in rice, wheat, and barley. *BMC Genomics* 6: 23-35.

Lagercrantz U (1998) Comparative mapping between *Arabidopsis thaliana* and *Brassica nigra* indicates that *Brassica* genomes have evolved through extensive genome replication accompanied by chromosome fusions and frequent rearrangements. *Genetics* 150: 1217-1228.

Lagercrantz U e Lydiate DJ (1996) Comparative genome mapping in *Brassica*. *Genetics* 144: 1903-1910.

Lagercrantz U, Ellegren H e Andersson L (1993) The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. *Nucleic Acids Research* 21: 1111-1115.

Langer RHM e Hill GD (1991) *Agricultural Plants*. 2nd edition. Cambridge University Press, Cambridge, 401 pp.

Lee JM, Grant D, Vallejos CE e Shoemaker RC (2001) Genome organization in dicots. II. *Arabidopsis* as a bridging species to resolve genome duplication events among legumes. *Theoretical and Applied Genetics* 103: 765-773.

Leigh F, Lea V, Law J, Wolters P, Powell W e Donini P (2003) Assessment of EST-and genomic microsatellite markers for variety discrimination and genetic diversity studies in wheat. *Euphytica* 133: 359-366.

Levantamento Sistemático da Produção Agrícola, Rio de Janeiro, v.5, n.12, 1993, pp 65-71, IBGE.

Levantamento Sistemático da Produção Agrícola, Rio de Janeiro, v.6, n.12, 1994, pp 65-71, IBGE.

Levantamento Sistemático da Produção Agrícola, Rio de Janeiro, v.7, n.12, 1995, pp 65-70, IBGE.

Levantamento Sistemático da Produção Agrícola, Rio de Janeiro, v.8, n.12, 1996, pp 65-70, IBGE.

Levantamento Sistemático da Produção Agrícola, Rio de Janeiro, v.9, n.12, 1997, pp 65-70, IBGE.

Levinson G e Gutman GA (1987) Slipped-strand mispairing: A major mechanism for DNA sequence evolution. *Molecular Biology and Evolution* 4: 203-221.

Lewin B (2004) *Genes VIII*. 1st edition. Pearson Prentice Hall, Upper Saddle River, 1056 pp.

Lewis G, Schirer B, Mackinder B e Lock M (eds) (2005) *Legumes of the world*. 1st edition. Kew, London, 592 pp.

Li Y-C, Abraham BK, Fahima T e Nevo E (2004) Microsatellites within genes: Structure, function, and evolution. *Molecular Biology and Evolution* 21: 991-1007.

Liewlaksaneeyanawin C, Ritland CE, EL-Kassaby YA e Ritland K (2004) Single-copy, species-transferable microsatellite markers developed from loblolly pine ESTs. *Theoretical and Applied Genetics* 109: 361-369.

Lin X, Kaul S, Rounsley S, Shea T, Benito MI, Town CD, Fujii CY, Mason T, Bowman CL, Barnstead M, Feldblyum TV, Buell CR, Ketchum KA, Lee J, Ronning CM, Koo HL, Moffat KS, Cronin LA, Shen M, Pai G, Van Aken S, Umayam L, Tallon LJ, Gill JE e Venter JC

(1999) Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. *Nature* 402:761-768.

Lindemann WC e Glover CR (2003) Nitrogen fixation by legumes. Cooperative Extension Service, College of Agriculture and Home Economics. University of New Mexico. Guide A-129.

Litt M e Luty JA (1989) A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *American Journal of Human Genetics* 44: 397-401.

Livingstone KD, Lackney VK, Blauth JR, van Wijk R e Jahn MK (1999) Genome mapping in *Capsicum* and the evolution of genome structure in the Solanaceae. *Genetics* 152: 1183-1202.

Lukens L, Zou F, Lydiate D, Parkin I e Osborn T (2003) Comparison of *Brassica oleracea* genetic map with genome of *Arabidopsis thaliana*. *Genetics* 164: 359-372.

Maestri E, Maleevschi A, Massari A e Marmioli N (2002) Genomic analysis of cultivated barley (*Hordeum vulgare*) using sequence-tagged molecular markers. Estimates of divergence based on RFLP and PCR markers derived from stress-responsive genes, and simple-sequence repeats (SSRs). *Molecular Genetics and Genomics* 267: 186-201.

Majewski J e Ott J (2000) GT repeats are associated with recombination on human chromosome 22. *Genome Research* 10: 1108-1114.

Malacinski GM (2003) *Essentials of Molecular Biology*. 4th edition. Jones and Bartlett Publishers, Boston, 516 pp.

Manshardt R (2004) Crop improvement by conventional breeding or genetic engineering: How different are they? *College of Tropical Agriculture and Human Resources, Bio-5*: 1-3

Marcotte EM, Pellegrini M, Yeates TO e Eisenberg D (1999) A census of protein repeats. *Journal of Molecular Biology* 293: 151-160.

Maréchal R, Mascherpa JM e Stainier F (1978) Etude taxonomique d'un groupe complexe d'espèces des genres *Phaseolus* et *Vigna* (Papilionaceae) sur la base de données morphologiques et polliniques, traitées par l'analyse informatique. *Biossiera Genève* 28: 1-273.

Martins LMV, Xavier GR, Rangel FW, Ribeiro JRA, Neves MCP, Morgado LB e Rumjanel NG (2003) Contributions of biological nitrogen fixation to cowpea: a strategy for improving grain yield in the semi-arid region of Brazil. *Biology and Fertility of Soils* 38: 333-339.

Menancio-Hautea D, Fatokun CA, Kumar L, Danesh D e Young ND (1993) Comparative genome analysis of mungbean (*Vigna radiata* (L.) Wilczek) and cowpea (*V. unguiculata* (L.) Walpers) using RFLP mapping data. *Theoretical and Applied Genetics* 86: 797-810.

Messier W (1996) The birth of microsatellites. *Nature* 381: 483.

Metzgar D, Bytof J and Wills C (2000) Selection against frameshift mutations limits microsatellite expansion in coding DNA. *Genome Research* 10: 72-80.

Milach SCK (1998) Marcadores de DNA. *Revista Biotecnologia* 5: 14-17.

Morgante M e Olivieri AM (1993) PCR-amplified microsatellites as markers in plant genetics. *The Plant Journal* 3: 175-182.

Morgante M, Hanafey M e Powell W (2002) Microsatellite are preferentially associated with nonrepetitive DNA in plant genomes. *Nature Genetics* 30: 194-200.

Mortimore MJ, Singh BB, Harris e Blade SB (1997) Cowpea in traditional cropping systems. In: Singh BB, Mohan Raj DR, Dashiell K e Jackai LEN (eds) *Advances in Cowpea Research*. 1st edition. IITA-JIRCAS, Ibadan, pp 99-113.

Moss DW (1982) *Izoenzimas*. 1st edition. Chapman & Hall, London, 204 pp.

Moxon R, Bayliss C e Hood D (2006) Bacterial contingency locos: The role of simple sequence DNA repeats in bacterial adaptation. *Annual Review of Genetics* 40: 307-333.

Mudge J, Cannon SB, Kalo P, Oldroyd GED, Roe BA, Town CD e Young ND (2005) Highly syntenic regions in the genomes of soybean, *Medicago truncatula*, and *Arabidopsis thaliana*. *BMC Plant Biology* 5: 15-30.

Mullis KB, Ferré F e Gibbs RA (eds) (1994) *The Polymerase Chain Reaction*. 1st edition. Birkhäuser, Basel, 458 pp.

Mun J-H, Kim D-J, Choi H-K, Gish J, Debellé F, Mudge J, Denny R, Endré G, Saurat O, Dudez A-M, Kiss GB, Roe B, Young ND e Cook DR (2006) Distribution of microsatellites in genome of *Medicago truncatula*: a resource of genetic markers that integrate genetic and physical maps. *Genetics* 172: 2541-2555.

Muralidharan K e Wakeland EK (1993) Concentration of primer and template qualitatively affects products in random-amplified polymorphic DNA PCR. *Biotechniques* 14: 362-364.

Murphy RW, Sites JWJ, Buth DG e Haufler CH (1990) Proteins I: isoenzyme electrophoresis. In: Hills DM e Moritz C (eds) *Molecular Systematics*. 1st edition. Sinauer Associates, Sunderland, pp 45-126.

Nadir E, Margalit H, Gallily T e Bensasson SA (1996) Microsatellite spreading in the human genome : Evolutionary mechanisms and structural implications. *Proceedings of the National Academy of Sciences* 93: 6470-6475.

Nauta MJ e Weissing FJ (1996) Constraints on allele size at microsatellite locos: implications for genetic differentiation. *Genetics* 143: 1021-1032

Naylor RN, Falcon WP, Goodman RM, Jahn MM, Sengooba T, Tefera H e Nelson RJ (2004) Biotechnology in the developing world: a case for increased investments in orphan crops. *Food Policy* 29: 15-44.

Nelson RJ, Naylor RL e Jahn MM (2004) The role of genomics in improvement of “orphan” crops. *Crop Science* 44: 1901-1904.

Nicot N, Chiquet V, Gandon B, Amilhat L, Legeai F, Leroy P, Bernard M e Sourdille P (2004) Study of simple sequence repeats (SSR) markers from wheat expressed tags (ESTs). *Theoretical and Applied Genetics* 109: 800-805.

Ng, NQ e Maréchal R (1985) Cowpea taxonomy, origin and germoplasm. In: Singh S, Rachie K (eds) *Cowpea research, production, and utilization*. Wiley, Chichester, UK, pp 11–21

Ng NQ (1995) Cowpea *Vigna unguiculata* (Leguminosea-Papilionoidaea). In: Smartt J, Simmonds N (eds) *Evolution of crop plants*. Longman, London, pp 326–332.

Ohta T e Kimura M (1973) A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genetical Research* 22: 201-204.

Oliveira EJ, Pádua JG, Zucchi MI, Vencovsky R e Vieira MLC (2006) Origin, evolution and genome distribution of microsatellites. *Genetics and Molecular Biology* 29: 294-307.

Olufajo OO e Singh BB. (2002) Advances in cowpea cropping systems research. In: *Challenges and opportunities for enhancing sustainable cowpea production. Proceedings of the World Cowpea Conference III, held at the International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria, 4–8 September 2000.* pp 267–277.

Orgel LE e Crick Francis (1980) Selfish DNA: The ultimate parasite. *Nature* 284: 604-607.

Osborn TC, Kole C, Parkin IAP, Sharpe AG, Kruiper M, Lydiate DJ e Trick M (1997) Comparison of flowering genes in *Brassica rapa*, *Brassica napus* and *Arabidopsis thaliana*. *Genetics* 146: 1123-1129.

Padulosi S, Hodgkin T, Williams JT e Haq N (2002) Underutilized crops: trends, challenges and opportunities in the 21st century. In: Engels JMM, Rao VR, Brown AHD and Jackson MT (eds) *Managing plant genetic resources*. CABI-IPGRI, Oxfordshire, pp 323-338.

Padulosi S e Ng NQ (1997) Origin, taxonomy, and morphology of *Vigna unguiculata* L. Walp. In: Singh BB, Mohan Raj DR, Dashiell KE and Jackai LEN (eds) *Advances in Cowpea Research*. 1st edition. IITA-JIRCAS, Ibadan, pp. 1–12.

Paetkau D e Strobeck C (1995) The molecular basis and evolutionary history of a microsatellite null alleles in bears. *Molecular Ecology* 4: 519-520.

Parfitt DE e Arulsekhar S (1987) Measurement and origin of genetic variation in tissue culture systems. In: Bonga JM e Durzan DJ (eds) Cell and Tissue Culture in Forestry, v. 1. 2nd edition. Martinus Nijhoff Publishers, Dordrecht, pp 286-297.

Pashley CH, Ellis JR, McCauley DE e Burke JM (2006) EST databases as a source for molecular markers: Lessons from the *Helianthus*. Journal of Heredity 97: 381-388.

Pasquet RS (1996) Cultivated cowpea (*Vigna unguiculata*): genetic organization and domestication. In: Pickersgill B, Lock J (eds) Advances in legumes. Systematics vol. 8: legumes of economic importance. Royal Botanic Gardens, Kew, pp 101-108.

Pasquet RS e Baudoin J-P (2001) Cowpea. In: Charrier A, Jacquot M, Hamon S e Niclas D (eds) Tropical Plant Breeding. 1st edition. Science Pub Inc, Enfield ,pp 177-198.

Peakall R, Gilmore S, Keys W, Morgante M e Rafalski A (1998) Cross-species amplification of soybean (*Glycine max*) simple sequence repeats (SSRs) within the genus and other legumes genera: Implications for transferability of SSRs in plants. Molecular Biology and Evolution 15: 1275-1287.

Pemberton JM, Slate J, Bancroft DR e Barrett JA (1995) Nonamplifying alleles at microsatellite locos: a caution for parentage and population studies. Molecular Ecology 4: 249-252.

Penner GA, Bush A, Wise R, Kim W, Domier L, Kasha K, Iarouche A, Scoles G, Molnar SJ, e Fedak G (1993) Reproducibility of random amplified polymorphic DNA (RAPD) analysis among laboratories. PCR Methods and Applications 2: 341-345.

Petes TD, Greenwell PW e Dominska M (1997) Stabilizations of microsatellite sequences by variant repeats in the yeast *Saccharomyces cerevisiae*. Genetics 146: 491-498.

Phillips RL e Vasil IK (2001) DNA-based markers in plants. 2nd edition. Kluwer Academic Publishers, Dordrecht, 529 pp.

Poncet V, Rondeau M, Tranchant C, Cayrel A, Hamon S, de Kochko A e Hamon P (2006) SSR mining in coffee tree EST database: potencial use of EST-SSRs as markers for the *Coffea* genus. Molecular Genetics and Genomics 276: 436-449.

Popelka JC, Gollasch S, Moore A, Molvig L, e Higgins TJV (2006) Genetic transformation of cowpea (*Vigna unguiculata* L.) and stable transmission of the transgenes to progeny. Plant Cell Reports 25: 304-312.

Powell W, Machray GC e Provan J (1996) Polymorphisms revealed by simple sequences repeats. Trends in Plant Science 1: 215-222.

Presley GJ (1992) Beyond Mendel's garden: Biotechnology in agriculture. In: Thottappilly G, Monti LM, Mohan Raj AW, Moore AW (eds) Biotechnology: Enhancing Research on Tropical Crops in Africa. 1st edition. CTA / IITA, Ibadan, Nigeria, pp 11-19.

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

Prinyawiwatkul W, Mcwatters KH, Beuchat LR e Phillips RD (1996) Cowpea flour: a potential ingredient in food products. *Critical Reviews in Food Science and Nutrition* 36: 413-436.

Quin FM (1997) Introduction. In: Singh BB, Mohan Raj DR, Dashiell KE and Jackai LEN (eds) *Advances in Cowpea Research*. 1st edition. IITA-JIRCAS, Ibadan, pp. ix-xv.

Rackoczy-Trojanowska M e Bolibok H (2004) Characteristics and comparison of three classes of microsatellite-based markers and their applications in plants. *Cellular & Molecular Biology Letters* 9: 221-238.

Richard GF e Paquês F (2000) Mini and microsatellite expansions: the recombination connection. *EMBO Reports* 11: 122-126.

Röder MS, Plaschke J, König SU, Börner A, Sorrells ME, Tanksley SD e Ganaw MW (1995) Absence, variability and chromosomal location of microsatellite in wheat. *Molecular Genetics and Genomics* 246: 327-333.

Rongwen J, Akkaya MS, Bhagwat AA, Lavi A e Cregan PB (1995) The use of microsatellites DNA markers for soybean genotype identification. *Theoretical and Applied Genetics* 90: 43-48.

Rose O e Falush D (1998) A threshold size for microsatellite expansion. *Molecular Biology and Evolution* 15: 613-615.

Rounsley SD, Glodek A, Sutton G, Adams MD, Somerville CR, Venter JC e Kerlavage AR (1996) The construction of *Arabidopsis* expressed sequence tags assemblies. *Plant Physiology* 112: 1177-1183.

Rungis D, Berube Y, Zhang J, Ralph S, Ritland CE, Ellis B, Douglas C, Bohlmann J e Ritland K (2004) Robust simple sequence repeat markers for spruce (*Picea* spp) from expressed sequence tags. *Theoretical and Applied Genetics* 109: 1283-1294.

Saha MC, Rouf Mian MA, Eujayl I, Zwonitzer JC, Wang L e May GD (2004) Tall fescue EST-SSR markers with transferability across several grasses species. *Theoretical and Applied Genetics* 109: 783-791.

Sainudiin R, Durrett RT, Aquadro CF e Nielsen R (2004) Microsatellite mutation models: insights from a comparison of humans and chimpanzees. *Genetics* 168: 383-395.

Sambrook J, Fritsch EF e Maniatis T (1989) *Molecular Cloning: laboratory manual*. 2nd edition. Cold Spring Harbor Laboratory Press, New York.

Sato S e Tabata S (2006) *Lotus japonicus* as a platform for legume research. *Current Opinion in Plant Biology* 9:128-132.

Schlötterer C (2000) Evolutionary dynamics of microsatellite DNA. *Chromosoma* 109:365-371.

Schlötterer C e Wiehe T (1999) Microsatellites, a neutral markers to infer selective sweeps. In: Goldstein DB e Schlötterer C (eds) *Microsatellites: Evolution and Applications*. 1st edition. Oxford University Press, Oxford, pp 238-247.

Schlötterer C, Ritter R, Harr B e Brem G (1998) High mutation rate of a long microsatellite allele in *Drosophila melanogaster* provides evidence for allele-specific mutation rates. *Molecular Biology and Evolution* 15: 1269-1274.

Scott KD, Eggler P, Seaton G, Rossetto M, Ablett EM, Lee LS e Henry RJ (2000) Analysis of SSRs derived from grape ESTs. *Theoretical and Applied Genetics* 100: 723-726.

Scott KD (2001) Microsatellite derived from ESTs, and their comparison with those derived by other methods. In: Henry RJ (eds) *Plant Genotyping: The DNA fingerprinting of Plants*. CABI Publishing, Oxon, pp 225-237.

Seastedt TR e Knapp AK (1993) Consequences of nonequilibrium resource availability across multiple time scales: the transient maxima hypothesis. *The American Naturalist* 141: 621-633.

Sharma PC, Grover e Kahl G (2007) Mining microsatellites in eukaryotic genomes. *Trends in Biotechnology* 25: 490-498.

Shimoda N, Knapik EW, Ziniti J Sim C, Yamada E, Kaplan S, Jackson D, de Sauvage F, Jacob H e Fishman MC (1999) Zebrafish genetic map with 2000 microsatellite markers. *Genomics* 58: 219-232.

Shriver MD, Jin L, Chakraborty R e Boerwinkle E (1993) VNTR allele frequency distributions under the stepwise mutation model: a computer simulation approach. *Genetics* 134: 983-993.

Silva KJD (2007) Feijão-caupi: uma opção à agricultura familiar e empresarial. [www.agrosoft.org.br/?q=node/27402](http://www.agrosoft.org.br/?q=node/27402)

Simon CJ e Muehlbauer FJ (1997) Construction of a chickpea linkage map and its comparison with maps pea and lentil. *Journal of Heredity* 88: 115-119.

Singh BB (2005) *Vigna unguiculata* L. Walp. In: Singh, RJ e Jahuar PP (eds) (2005) *Genetic Resources, Chromosome Engineering, and Crop Improvement. Grain Legumes*. Vol. 1. 1st edition. CRC Press, Boca Raton, pp 116-151.

Singh BB e Emechebe AM (1998) Increasing productivity of millet-cowpea intercropping systems. In: Emechebe AM, Ikwelle MC, Ajayi O, Aminu Kano M e Anaso AB (eds) Pearl Millet in Nigeria Agriculture: Production, Utilization and Research Priorities, Proceedings of the Pre-season Planning Meeting for the Nationally Coordinated Research Programme for Pearl Millet, Maiduguri, April 21–24, 1997. Lake Chad Research Institute, Maiduguri, Nigeria, pp 88–75.

Singh BB, Ajeibe HA, Tarawali SA, Fernandez-Rivera S e Abubakar M (2003) Improving the production and utilization of cowpea as food and fodder. *Field Crop Research* 84: 169-177.

Singh BB, Cambliss OI e Sharma B (1997) Recente advances in cowpea breeding. In: Singh BB, Mohan Raj DR, Dashiell K and Jackai LEN (eds). *Advances in cowpea research*. 1st edition. IITA-JIRCAS, Ibadan, pp 30-49.

Singh RJ (2005) Landmark research in grain legumes. In: Singh, RJ e Jahuar PP (eds) (2005) *Genetic Resources, Chromosome Engineering, and Crop Improvement. Grain Legumes, Vol. 1*. 1st edition. CRC Press, Boca Raton, pp 1-9.

Singh RJ, Chung GH e Nelson RL (2007) Landmark research in legumes. *Genome* 50: 525-537.

Singh, RJ (2007) Landmark research in oilseed crops. In: RJ Singh (ed) *Genetic Resources, Chromosome Engineering, and Crop Improvement Series: Oilseed Crops, Vol 4*. 1st edition. CRC Press, Boca Raton, pp 1-12.

Smithies O (1955) Zone electrophoresis in starch gels: group variation in the serum proteins of normal human adults. *Biochemical Journal* 61: 629-641.

Song R, Llaca V e Messing J (2002) Mosaic Organization of Orthologous Sequences in Grass Genomes. *Genome Research* 12: 1549-1555.

Sourdille P, Tavaud M, Charmet G e Bernard M (2001) Transferability of wheat microsatellite to diploid Triticea species carrying the A, B and D genomes.

Southern EM (1975) Detection of specific sequences among DNA fragments separated by electrophoresis. *Journal of Molecular Biology* 98: 503-517.

Sprent JL e Sprent P (1990) Nitrogen fixing organisms. Pure and applied aspects. 1st edition. Chapman & Hall, London, 256 pp.

Starker CG, Parra-Colmenares AL, Smith L, Mitra R and Long SR (2006) Nitrogen fixation mutants of *Medicago truncatula* fail to support plant and bacterial symbiotic gene expression. *Plant Physiology* 140: 671–680.

Staub JE, Kuhns LJ, May B e Grun P (1982) Stability of potato tuber isozymes under different storage regimes. *American Society for Horticultural Science* 107: 405-408.

Strand M, Prolla TA, Liskay RM e Petes TD (1993) Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature* 365: 274-276.

Tanksley SD, Bernatzky R, Lapitan NL e Prince JP (1988) Conservation of gene repertoire but not gene order in pepper and tomato. *Proceedings of National Academy of Sciences* 85: 6419-6423.

Tanksley SD, Ganai MW, Prince JP, de-Vicente MC, Bonierbale MW, Broun P, Fulton TM, Giovannoni JJ, Grandillo S, Martin GB, Messeguer R, Miller JC, Miller L, Paterson AH, Pineda O, Roder MS, Wing RA, Wu W e Young ND (1992) High density molecular linkage maps of the tomato and potato genomes. *Genetics* 132: 1141-1160.

Tautz D (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research* 17: 6463-6471.

Teófilo EM, Paiva JB e Vidal JJ. Renovação de estoque e caracterização de 94 cultivares de feijão-de-corda (*Vigna sinensis* (L.) Savi). In: Universidade Federal do Ceará (Fortaleza, CE). Centro de Ciências Agrárias. Relatório de Pesquisa 1988: criação difusão de novos cultivares de feijão-de-corda para o estado do Ceará, Fortaleza: UFC/CCA/FCPC, 1990. p.1-5.

Teófilo EM, Paiva JB e Vidal MJ. Estudo de caracterização e renovação de estoques de 143 cultivares de feijão-de-corda, *Vigna sinensis* (L.) Savi. In: Universidade Federal do Ceará (Fortaleza, CE). Centro de Ciências Agrárias. Relatório de Pesquisa 1987: criação difusão de novos cultivares de feijão-de-corda para o estado do Ceará, Fortaleza: UFC/CCA/FCPC, 1989. p. 1-18.

The National Biodiesel Board, <http://www.biodiesel.org/> (Maio, 2008)

Thiel T, Michalek W, Varshney RK e Graner A (2003) Exploiting EST databases for development and characterization of gene-derived SSR-markers in braley (*Hordeum vulgare* L.). *Theoretical and Applied Genetics* 106: 411-422.

Thuillet A-C, Bru D, David J, Roumet P, Santoni S, Sourdille P e Bataillon T (2002) Direct estimation of mutation rate for 10 microsatellite locos in durum wheat, *Triticum turgidum* (L.) Thell. Ssp. Durum. *Molecular Biology and Evolution* 19: 122-125.

Tóth G, Gáspári Z e Jurka J (2000) Microsatellites in different eukaryotic genomes: Survey and analysis. *Genome Research* 10: 967-981.

Toutenhoofd SL, Garcia F, Zacharias DA, Wilson RA e Strehler EE (1998) Minimum CAG repeat in the human calmodulin-1 gene 5' untranslated region is required for full expression. *Biochimica et Biophysica Acta* 1398:315-320.

Treco D e Arnheim N (1986) The evolutionarily conserved repetitive sequence d(TG. AC)<sub>n</sub> promotes reciprocal exchange and generates unusual recombinant tetrads during yeast meiosis. *Molecular and Cellular Biology* 6: 3934-3947.

Trifonov EN (2004) The tuning function of tandemly repeating sequences: a molecular devices for fast adaptation. In: Wasser SP (ed) *Evolutionary Theory and Processes*. 1st edition. Modern Horizons, Kluwer Academic Publishers, Dordrecht, pp 115-138.

Udupa SM e Baum M (2001) High mutation rate and mutational bias at (TAA)<sub>n</sub> microsatellite locos in chickpea (*Cicer arietinum* L.). *Molecular Genetics and Genomics* 265: 1097-1103.

Uzogara SG e Ofuya ZM (1992) Processing and utilization of cowpeas in developing countries: A review. *Journal Food Processing and Preservation* 16: 105-147.

Valdes AM, Slatkin M e Freimer NB (1993) Allele frequencies at microsatellite locos: the stepwise mutation model revisited. *Genetics* 133: 737-749.

Van Ek, GA, Henriet J, Blade SF e Singh BB (1997) Quantitative assessment of traditional cropping systems in the Sudan savanna of northern Nigeria. II. Management and productivity of major cropping sytems. *Samaru Journal of Agricultural Reseach*. 14: 47-60.

Vance CP, Gragam PH e Allan DL (2000) Biological nitrogen fixation. Phosphorus a critic future need. In: Pedrosa FO, Hungria M, Yates MG e Newton WE (eds) *Nitrogen fixation: from molecules to crop productivity*. 1st edition. Kluwer Academy Publishers, Dordrecht, pp 506-514.

Varshney RK, Grander A e Sorrells ME (2005a) Genic microsatellite markers in plants: features and applications. *Trends in Biotechnology* 23: 48-55.

Varshney RK, Sigmund R, Börner A, Korzun V, Stein N, Sorrells ME, Langridge P e Graner A (2005b) Interspecific transferability and comparative mapping of barley EST-SSR markers in wheat, rye and rice. *Plant Science* 168: 195-202.

Varshney RK, Thiel T, Stein N, Langridge P e Graner A (2002) In silico analysis on frequence and distributionof microsatellites in ESTs of some cereal species. *Cellular & Molecular Bioloy Letters* 7: 537-546.

Vaillancourt RE e Weeden NF (1992) Chloroplast DNA polymorphism suggests a Nigerian center of domestication for the cowpea, *Vigna unguiculata*, Leguminosae. *American Journal of Botany* 79: 1194–1199

Verdcourt B (1970) Studies in the Leguminosae – Papilionoidea for flora of tropical East Africa’: IV. *Kew Bulletin* 24: 507-569.

Vitousek PM, Aber JD, Howarth RW, Likens GE, Matson PA, Schindler DW, Schlesinger WH e Tilman DG (1997) Human alteration of the global nitrogen cycle: sources and consequences. *Ecological applications* 7: 737-750.

Weber JL (1990) Informativeness of human (dC-dA).(dG-dT) polymorphisms. *Genomics* 7: 524-530.

Weber JL e May PE (1989) Abundant class of human DNA polymorphisms wich can be typed usinh polymerase chain reaction. *American Journal of Human Genetics* 44: 388-396.

Weber JL. e Wong C (1993) Mutation of human short *tandem* repeats. *Human Molecular Genetics* 8: 1123-1128.

Weeden NF, Muehlbauer FJ e Ladizinsky G (1992) Extensive conservation of linkage relationships between pea and lentil genetic maps. *Journal of Heredity* 83: 123-129.

Weising K, Kaemmer D, Weigand F, Epplen JT e Kahl G (1992) Oligonucleotide fingerprinting reveals various probe-dependent levels of informativeness in chickpea *Cicer aeritinum*). *Genome* 35: 436-442.

Weising K, Nybom H, Wolff K e Kahl G (2005) DNA fingerprinting in plants: Principles, Methods, and Applications. 2nd edition. CRC Press, Boca Raton, 444 pp.

Weising K, Weigand F, Driesel A, Kahl G, Zischer H e Epplen JT (1989) Polymorphic simple GATA/GACA repeats in plant genomes. *Nucleic Acids Researchs* 17: 10128.

Welsh J and McClelland (1990) Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Research* 18: 7213-7218.

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA e Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research* 18: 6231-6235.

Woodhead M, Russel J, Squirrell J, Hollingsworth PM, Mackenzie K, Gibby M e Powell W (2005) Comparative analysis of population genetic structure in *Athyrium distentifolium* (Pteridophyta) using AFLPs and SSRs from anonimus and transcribed gene regions. *Molecular Ecology* 14 1681-1695.

World Hunger Education Service.  
<http://www.worldhunger.org/articles/Learn/world%20hunger%20facts%202002.htm> (Janeiro, 2008)

Xavier GR, Martins LMV, Rumjanek NG e Freire Filho FR (2005) Variabilidade genética em acessos de caupi analisada por meio de marcadores RAPD. *Pesquisa Agropecuária Brasileira* 40: 353-359.

- Xu X, Peng M, Fang Z e Xu X (2000) The direction of microsatellite mutation is dependent upon allele length. *Nature Genetics* 24: 396-399.
- Yamanaka S, Suzuki E, Tanaka M, Takeda Y, Watanabe JA e Watanabe KN (2003) Assessment of cytochrome P450 sequences offers a useful tool for determining genetic diversity in higher plants species. *Theoretical and Applied Genetics* 108: 1-9.
- Yang GP, Saghai Maroof MA, Xu CG, Zhang Q e Biyashev RM (1994) Comparative analysis of microsatellite DNA polymorphism in landraces and cultivars of rice. *Molecular and General Genetics* 245: 187-194.
- Young ET, Sloan JS e Van Riper K (2000) Trinucleotide repeats are clustered in regulatory genes in *Saccharomyces cerevisiae*. *Genetics* 154: 1053-1068.
- Yu JK, Dake TM, Singh S; Benscher D, Li W, Gill B e Sorrells ME (2004) Development and mapping of EST-derived simple sequence repeat (SSR) markers for hexaploid wheat. *Genome* 47: 805-818.
- Zahran HH (1999) Rhizobium-legume symbiosis and nitrogen fixation under severe conditions and in an arid climate. *Microbiology and Molecular Biology Reviews* 63: 968-989.
- Zane L, Bargelloni L e Patarnello T (2002) Strategies for microsatellite isolation: a review. *Molecular Ecology* 11: 1-16.
- Zeven AC e Wet JMJ (1982) Dictionary of cultivated plants and their regions of diversity: Excluding most ornamentals, forest trees and lower plants. 2nd edition. Centre for Agricultural Publishing and Documentation, Wageningen, 128 pp.
- Zhu H, Choi H-K, Cook DR e Schoemaker C (2005) Bridging model and crop legumes through comparative genomics. *Plant Physiology* 137: 1189-1196.
- Zhu Y, Strassmann JE e Queller D (2000) Insertions, substitutions and the origin of microsatellites. *Genetics Research* 76: 227-236.

## 4. MANUSCRITO DE ARTIGO CIENTÍFICO

# **ANÁLISE IN SILICO DE EST-SSR EM *Phaseolus vulgaris* E *Glycine max* E TRANSFERIBILIDADE DE MARCADORES PARA *Vigna unguiculata*.**

Manuscrito encaminhado à Revista

**Genetics and Molecular**

**Biology**

ISSN 1415-4757

Ribeirão Preto, Brasil

**ANÁLISE IN SILICO DE EST-SSR EM *Phaseolus vulgaris* E *Glycine max* E  
TRANSFERIBILIDADE DE MARCADORES PARA *Vigna unguiculata*.**

A ser submetido ao Jornal Genetics and Molecular Biology

Ferreira Neto JRC<sup>1</sup>; Houllou-Kido LM<sup>2</sup>; Costa AF<sup>2</sup>; Freire Filho FR<sup>3</sup>; Moraes RMA<sup>4</sup>; Benko-  
Iseppon AM<sup>1</sup>; Kido EA<sup>1</sup>.

<sup>1</sup> Departamento de Genética, Universidade Federal de Pernambuco, Recife, Brasil

<sup>2</sup> Instituto Agrônômico de Pernambuco, Recife, Brasil

<sup>3</sup> EMBRAPA MEIO-NORTE, Teresina, Brasil.

<sup>4</sup> EMBRAPA TRIGO, Passo Fundo, RS, Brasil.

Autor para correspondência:

Éderson Akio Kido

Laboratório de Genética Molecular / UFPE

Av. Prof. Moraes Rego, 1235 - Cidade Universitária,

Recife - PE – Brasil

CEP: 50670-901

ederson.kido@gmail.com

## RESUMO

Bancos de dados de seqüências expressas representam uma fonte potencialmente valiosa para o desenvolvimento de marcadores moleculares. Neste trabalho, estudos foram realizados para o desenvolvimento de marcadores para aplicação no melhoramento de feijão-caupi, feijão-comum e soja. Foram analisadas 10.880 ESTs de feijão-comum e 116.965 ESTs de soja. Um total de 331 e 4553 SSRs foram identificados nas seqüências de feijão-comum e soja, respectivamente. Trinucleotídeos (45,62 %) foram as mais abundantes em feijão-comum, enquanto que os dinucleotídeos (49,67 %), em soja. A densidade média variou de um SSR a cada 22 Kb (feijão-comum) a um SSR a cada 14,68 Kb (soja). Repetições diméricas AG/GA / TC/CT e triméricas AAG/AGA/GAA / TTC/TCT/CTT foram as mais abundantes em ambas espécies. Um total de 153 e 1928 pares de *primers* EST-SSRs foram propostos para feijão-comum e para soja, respectivamente. Desses, sintetizaram-se 20 para soja, dos quais apenas nove amplificaram (sendo quatro de tamanho esperado e, desses, dois polimórficos) e dois foram transferíveis. Para feijão-comum sintetizaram-se 22, dos quais 15 amplificaram, sendo que 11 apresentaram tamanho esperado e 10 desses foram polimórficos (PIC médio de 0,50); 11 dos funcionais foram transferidos para feijão-caupi, sendo oito de tamanho esperado e cinco desses, polimórficos (PIC médio de 0,36). Análises detalhadas dos *amplicons* seqüenciados determinaram que as extensões dos motivos de SSRs foram variáveis e que as regiões flanqueadoras dessas repetições foram geralmente bem conservadas, confirmando o sucesso da transferibilidade entre feijão-comum e feijão-caupi.

**Palavras-chave:** microssatélite, mineração de dados, Fabaceae, leguminosa, feijão-caupi.

## 1. INTRODUÇÃO

A família Fabaceae é a segunda mais importante em termos de área cultivada, contribuição na dieta humana e valor econômico. Nesta família, o feijão-comum (*Phaseolus vulgaris* L.) e a soja [*Glycine max* (L.) Merr.] são os representantes mais importantes a nível global. Programas de melhoramento para ambas as espécies são bem estabelecidos e a seleção assistida por marcadores tem sido divulgada (Alzate-Marin *et al.*, 2005). Por sua vez, o feijão-caupi [*Vigna unguiculata* (L.) Walp.] é o principal componente de sistemas agrícolas tradicionais na África, Ásia e Américas Central e do Sul (Olufajo e Singh, 2002), sendo uma das principais fontes de proteínas na alimentação diária de povos menos favorecidos (Singh, 2005). No Brasil, é um dos principais componentes da dieta alimentar e gerador de emprego e renda, principalmente na região Nordeste (Freire Filho *et al.*, 1999). Sua rusticidade e plasticidade fenotípica permitem o cultivo em condições adversas de clima e solo, que associados ao potencial para fornecimento de grãos, ampla variabilidade genética e boa capacidade de fixação de nitrogênio atmosférico, diminui a dependência quase exclusiva do feijão-comum e lhe confere certo valor estratégico (Freire Filho *et al.*, 1999). Apesar dessa importância, o feijão-caupi pode ser considerado uma cultura “órfã” em relação às duas espécies anteriores, devido à menor disponibilidade de recursos para pesquisas (Nelson *et al.*, 2003; Naylor *et al.*, 2004).

No campo da genômica de leguminosas há necessidade de se integrar conhecimentos derivados de estudos envolvendo espécies-modelo (ou aquelas mais estudadas) na resolução de questões biológicas e agrônômicas de espécies cultivadas (Choi *et al.*, 2004a; Zhu *et al.*, 2005). Assim, investimentos relativamente pequenos na transferência de informação dessas espécies para aquelas “órfãs” resultariam em progressos mais rápidos nos sistemas de produção, principalmente em países menos desenvolvidos (Nelson *et al.*, 2004). Marcadores moleculares são amplamente utilizados nas mais diversas finalidades: estudos de diversidade, filogenia, caracterização de germoplasma, construção de mapas genéticos, estudos de genética comparativa, clonagem posicional etc. (Kumar, 1999; Choi *et al.*, 2004b; Mun *et al.*, 2006; Simon *et al.*, 2007). Em muitos estudos, os marcadores utilizados derivam de microssatélites ou seqüências simples repetidas (SSR). Essas seqüências constituem-se de motivos de 1 a 6 nucleotídeos repetidos em *tandem* (Tóth *et al.*, 2000) e podem estar relacionadas à regulação da expressão gênica e evolução (Ayres *et al.*, 1997; Li *et al.*, 2004; Trifonov, 2004; Kashi e King, 2006). Sua ampla utilização advém das vantagens de sua natureza hipervariável, herança co-dominante, ampla distribuição nos genomas e detecção via PCR (Powell *et al.*,

1996), podendo ser até 10 vezes mais polimórficos que outros marcadores (Rackoczy-Trojabowska e Bolibok, 2004). Sua principal desvantagem é a necessidade de se conhecer previamente as seqüências flanqueadoras dos microssatélites para desenhos de *primers*, o que envolveria a criação de bibliotecas genômicas, hibridizações com oligonucleotídeos repetidos em *tandem* e seqüenciamento de clones (Zane *et al.*, 2002; Weising *et al.*, 2005), sendo um processo caro e laborioso. Com a crescente disponibilidade de seqüências expressas (EST, *Expressed Sequence Tags* ou Etiquetas de Seqüências Expressas) em bancos de dados públicos, a busca de seqüências microssatélites via mineração de dados se tornou em um processo rápido, acessível e economicamente viável, com potencial para automatização e geração de novos marcadores (Gupta *et al.*, 2003; Kumpatla e Mukopadhyay, 2005; Varshney *et al.*, 2005). Vários ensaios demonstram uma abundância significativa dessas repetições em seqüências EST (Morgante *et al.*, 2002; Thiel *et al.*, 2003; Kumpatla e Mukhopadhyay, 2005). Entretanto, tal metodologia é limitada às espécies bem representadas nesses bancos. Nesse contexto, a transferibilidade de marcadores EST-SSR tem sido estudada em diferentes espécies vegetais (Gao *et al.*, 2003; Eujayl *et al.*, 2004; Yu *et al.*, 2004; Gutierrez *et al.*, 2005; Poncet *et al.*, 2006). Isto é possível em decorrência da sintenia observada entre espécies aparentadas. Estudos de genética comparativa indicam um significativo grau de sintenia entre leguminosas (Frugoli *et al.*, 1999; Humphry *et al.*, 2002; Choi *et al.*, 2004a; Mudge *et al.*, 2005; Cannon *et al.*, 2006). Dados de mapeamento comparativo entre feijão-mungo [*Vigna radiata* (L.) Wilczek], feijão-comum e soja indicaram que as duas primeiras espécies compartilhavam grupos de ligação altamente conservados, e a sintenia de ambos com soja estava limitada a curtos blocos de ligação (Boutin *et al.*, 1995). O mesmo seria esperado com feijão-caupi, visto que entre esse e feijão-mungo um alto grau de conservação foi observado (Menancio-Hautea *et al.*, 1993). Em estudo tendo *Arabidopsis thaliana* como espécie ligadora, entretanto, relatos de segmentos homoeólogos de soja apresentando maior grau de sintenia com feijão-comum e feijão-mungo do que previamente divulgado, foram descritos (Lee *et al.*, 2001). Deste modo, os objetivos deste trabalho foram: (1) a identificação, classificação e estudo de SSRs em EST/cDNA de feijão-comum e soja depositadas no banco de dados TIGR; (2) a proposição de *primers* flanqueadores dessas regiões com testes em amplificações com cultivares das espécies em análise; e (3) a análise de transferibilidade para cultivares de feijão-caupi.

## 2. MATERIAL E MÉTODO

### 2.1. Material Vegetal e Isolamento do DNA

DNAs genômicos foram extraídos de folhas jovens, usando-se o método CTAB descrito por Doyle e Doyle (1987). Os materiais vegetais abrangeram genótipos de feijão-comum (Esplendor, Boreal, Jaraguá, Princesa, Gurgutuba, Luminoso, Galante, Alvorada, IPA 7419 e Harmonia), cedidos pelo Instituto Agrônomo de Pernambuco – IPA (PE, Brasil) e adquirido comercialmente (Macarrão Trepador); de genótipos de soja (Fepagro 25, BRS 244 RR, BRS 232, CD 214 RR, BRS Pala, BRS 155, CD 217, BRS 258, Fundacep 45, BRS 257, BRS Febana, BRS 153, BRS Pampa, BRS 216, BRS 213), cedidos pela EMBRAPA Trigo (Passo Fundo, RS, Brasil); e genótipos de feijão-caupi, obtidos de diferentes instituições: Embrapa Meio-Norte (Teresina, PI) (cultivares TE97G12, TE9612G, TE9764, TE9669, BR14 Mulato e IT85F); IPA (cultivares Vita 7, Patativa, Pitiúba, IPA 206, IPA 207, BRS-Guariba, BRS-Gurguéia, BRS-Paraguaçu e Fradinho) e Universidade Federal Rural de Pernambuco (Feijão-de-vagem e Verde ligeiro).

### 2.2. Análise *in silico* de EST-SSRs em *Phaseolus vulgaris* e *Glycine max*

Seqüências EST e de cDNAs (total ou parcial) de feijão-comum (10.880) e de soja (116.965) do banco de dados TIGR (seção Plant Transcript Assemblies; <http://plantta.tigr.org/>), foram utilizadas neste estudo. Trata-se de banco não-redundante, constituído por *singlets* e *contigs*, composto de seqüências expressas coletadas do GenBank Nucleotide (NCBI), tendo-se excluídas as seqüências gênicas preditas computacionalmente (Childs *et al.*, 2007).

Os motivos SSR perfeitos, definidos com um número mínimo de 9, 6, 5, 5, 4 repetições, respectivamente, para di, tri, tetra, penta e hexanucleotídeos foram identificados com a ajuda do software TRA (Bilgen *et al.*, 2004). As combinações de bases para os diferentes motivos foram agrupadas em classes únicas com base na complementaridade de bases do DNA (Jurka e Pethiyagoda, 1995). Motivos diméricos, por exemplo, foram agrupados em 4 classes únicas de repetição: (1) AT/TA; (2) AG/GA / TC/CT; (3) AC/CA / TG/GT; e (4) GC/CG. No geral, foram consideradas 4, 10, 33, 102 e 305 classes únicas de repetições para di, tri, tetra, penta e hexanucleotídeos, respectivamente.

### 2.3. Desenho de *primers*, amplificação via PCR e análise de produtos

*Primers* foram desenhados com a ferramenta Primer3 (<http://fokker.wi.mit.edu/primer3/input.htm>) e uso de parâmetros padrões, exceto para variação de conteúdo GC (40 a 70%, com ótimo de 50%) e extensão (18 a 24 pb, com ótimo de 20 pb). O tamanho esperado do *amplicon* foi de, no máximo, 500 pb, visando diminuir o risco de amplificação de íntrons. Análises BLASTn (Altschul *et al.*, 1990) foram realizadas contra o banco de dados GenBank, usando-se os *primers* como *queries* para verificar possíveis similaridades com seqüências diferentes das originais. Uma amostra de *primers* foi selecionada com base na extensão da região repetitiva flanqueada e utilizada em reações de PCR. Essas reações apresentaram volume total de 20 µL [20 ng de DNA genômico, 1x PCR buffer (Fermentas), 1,5 mM de MgCl<sub>2</sub>, 100 µM de dNTPs, 0,15 µM de cada primer e 0,5 U de Taq DNA polimerase (Fermentas)], e desnaturação inicial de cinco min, a 95° C, seguida de 30 ciclos de um min a 95° C, um min a 54° C e um min a 72° C, com extensão final de 10 min a 72° C. Ao final, as amostras foram misturadas com igual volume de tampão de carregamento (formamida contendo 10 mM de EDTA e traços de xileno cianol e azul de bromofenol) e desnaturadas (95°C, por cinco min). Os fragmentos amplificados foram separados em gel de acrilamida 7 % (p/v) em condição desnaturante (7 M de uréia), sendo a visualização dos *amplicons* realizada após coloração com solução de nitrato de prata, conforme descrito por Bassam *et al.* (1991).

Para confirmar a presença dos motivos repetitivos nos *amplicons* de feijão-caupi, e analisar as origens dos polimorfismos nas três espécies estudadas, foram seqüenciados alguns produtos de PCR de tamanhos esperados, com os *primers* forwards correspondentes. As reações de PCR foram realizadas conforme descrito acima e os *amplicons*, após eletroforese em gel de agarose 1% (p/v) [em tampão TAE (40 mM de Tris-Acetato e 1 mM de EDTA, pH = 8,3), por 2 h a 90 V] e coloração com Brometo de Etídio, foram excisados do gel e purificados [kit *QIAquick Gel Extraction* (Quiagen)]. O seqüenciamento foi realizado usando-se *BigDye Terminator Cycle sequencing v 3.1* kit (PE Biosystems, Foster City, Calif.) em seqüenciador ABI Prism 3100 (Applied Biosystems).

### 2.4. Análise dos marcadores

Polimorfismos foram avaliados de acordo com a presença ou ausência de um determinado loco SSR. O conteúdo informativo de um marcador foi estimado usando-se o

método do Conteúdo de Informação Polimórfica (PIC, *Polymorphic Information Content*), descrito por Bolstein *et al.*, (1980) e modificado por Anderson *et al.*, (1993) para espécies de autofecundação, onde  $P_{ij}$  é a frequência do  $i$ -ésimo alelo do marcador “ $i$ ” e “ $n$ ” é o número de alelos de tal marcador.

$$\text{PIC} = 1 - \sum_{j=1}^n P_{ij}^2$$

Análises BLAST2seq (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) foram realizadas para verificar a identidade dos *amplicons* seqüenciados de feijão-caupi e ESTs de soja ou feijão-comum correspondentes. Adicionalmente, alinhamentos entre os *amplicons* e as ESTs correspondentes foram feitos para verificar a conservação das seqüências nucleotídicas, com uso da ferramenta *on line* Clustal W (<http://www.ebi.ac.uk/Tools/clustalw2/index.html>)

### 3. RESULTADOS E DISCUSSÃO

#### 3.1. Análise *in silico* de EST- SSR

As análises *in silico* em 10.880 seqüências de feijão-comum (aproximadamente 7,3 Mb), detectaram 331 SSRs em 304 ESTs/cDNAs (presença em cerca de 2,8 % das seqüências). Em 116.965 seqüências de soja, equivalentes a 66 Mb, foram detectadas 4.553 SSRs em 4.299 ESTs/cDNAs (3,7 % das seqüências analisadas). Mesmo com uso de critérios mais rigorosos, esses valores estão de acordo com os obtidos por Kumpatla e Mukhopadhyay (2005) que obtiveram de 2,6 a 16,8 % para 49 grupos de dicotiledôneas analisadas. As densidades médias estimadas em feijão-comum e soja (um SSR a cada 22,0 Kb e a cada 14,7 Kb, respectivamente) foram menores que observadas por Jayashree *et al.* (2006) também em leguminosas (SSR a cada 3,5 Kb). No entanto, esses autores utilizaram um tamanho mínimo de 12 pb para caracterizar uma extensão repetitiva de nucleotídeos em um SSR. Além da variação interespecífica inerente a cada grupo vegetal, diversos fatores podem afetar a frequência global e individual de motivos e tipos de repetições. Dentre esses, destacam-se, além do já citado, o tamanho e redundância do banco de dados, o software utilizado, os critérios de busca (SSRs perfeitos ou imperfeitos, ou ambos) e a localização genômica em questão (íntrons, exons, regiões 5' ou 3' UTR, regiões intergênicas, etc) (Kumpatla e Mukhopadhyay, 2005; Varshney *et al.*, 2005; Morgante *et al.*, 2002; Tóth *et al.*, 2000; La Rota *et al.*, 2005).

Motivos de tetranucleotídeos e de pentanucleotídeos foram os menos representados em ambas as espécies (Tab. 1). Em relação aos mais abundantes, os motivos triméricos (45,6 %), seguidos dos diméricos (40,8 %), foram os mais frequentes em feijão-comum (Tab. 1). A predominância dessas repetições em seqüências expressas de vegetais tem sido amplamente relatada (Varshney *et al.*, 2002; Gao *et al.*, 2003; Thiel *et al.* 2003; Mun *et al.*, 2006; Poncet *et al.*, 2006). Em soja, as repetições de dinucleotídeos (49,7 %) foram mais abundantes que as de trinucleotídeos (39,6 %). Uma menor abrangência de repetições não-triméricas nessas seqüências seria esperada, uma vez que expansões ou retrações das mesmas causariam mutações na matriz de leitura em regiões codificantes e estariam sendo selecionadas negativamente (Metzgar *et al.*, 2000). Tal abundância de dinucleotídeos pode estar relacionada a uma característica peculiar do banco de dados de soja. Morgante *et al.* (2002), investigando SSRs em diferentes frações genômicas de *Arabidopsis*, observaram que a frequência relativa de trinucleotídeos duplicou em ORFs, quando comparada com frações não-codificantes, possivelmente como resultado de pressão de seleção (Metzgar *et al.*, 2000) ou mesmo, à seleção positiva para formação de extensões formadas por repetições de determinados aminoácidos, enquanto que repetições não-triméricas apresentaram frequência reduzida nessas regiões. Adicionalmente, sugeriram que diferentes pressões seletivas atuariam nas diferentes regiões de uma unidade transcricional (5'UTR, ORF e 3'UTR) e que as UTRs estariam sob pressão seletiva positiva para tais repetições, sendo a abundância de SSRs nessas áreas devido à massiva presença de dinucleotídeos e trinucleotídeos na 5'UTR e tetranucleotídeos e trinucleotídeos na 3'UTR. Assim, a abundância de dinucleotídeos em nossos resultados sugere, então, uma maior cobertura de regiões 5'UTR em relação às ORFs, nas seqüências analisadas.

Os motivos triméricos mais abundantes foram AAG/AGA/GAA / TTC/TCT/CTT em ambas as espécies (com 44,4 %, em feijão-comum e 27,2 %, em soja; Tab. 1). Estes são os mais comuns em leguminosas e dicotiledôneas, em geral (Kumpatla e Mukhopadhyay, 2005; Jayshree *et al.*, 2006). Adicionalmente, foi observada uma significativa escassez (apenas 7,9 % em feijão-comum e 13,8 %, em soja) de motivos triméricos ricos em GC (dados não apresentados), concordando com a tendência observada na maioria das dicotiledôneas (Morgante *et al.*, 2002; Kumpatla e Mukhopadhyay, 2005), provavelmente devido ao baixo conteúdo GC dessas espécies e/ou ao uso preferencial de códons ricos em AT em espécies desse grupo vegetal (Morgante *et al.*, 2002). Quanto aos motivos diméricos, os mais frequentes foram AG/GA / TC/CT nas duas espécies estudadas (com 43,7 % em feijão-comum e 64,9 %, em soja) (Tab. 1), concordando com os relatos que citam serem estes os

diméricos mais frequentemente observados em EST de vegetais (Morgante *et al.*, 2002; Kantety *et al.*, 2002; Varshney *et al.*, 2002; Thiel *et al.*, 2003; Kumpatla e Mukhopadhyay, 2005; Jayashree *et al.*, 2006).

A extensão média (EM) estimada de todos os motivos SSR detectados foi de 27 pb em feijão-comum e 30 pb, em soja, sendo os de motivos triméricos os menores e os diméricos, os maiores nos referidos vegetais (Tab. 1). Tal fato pode estar associada à localização dos motivos SSR nas ESTs analisadas e à existência de uma relação linear entre mutabilidade e extensão desses motivos (Ellegren, 2004). Dos motivos triméricos, 54,4 % em feijão-comum e 53,5 % em soja, foram localizados em ORFs (Tab. 2). Essas repetições triméricas apresentariam um tamanho reduzido devido à ação de forças seletivas que visam preservar a estrutura do transcrito, uma vez que variações de SSR dentro de genes influenciam a atividade do mesmo, pois expansões ou contrações dessas regiões afetariam diretamente o produto gênico correspondente, podendo gerar modificações fenotípicas. SSRs com motivos diméricos, por sua vez, foram localizados em sua maioria [68,8 % (feijão-comum) e 79,0 % (soja)] nas regiões não traduzidas (5' e 3'UTR), estando sob seleção positiva para SSRs (Morgante *et al.*, 2002), o que permitiria uma maior extensão das mesmas.

### 3.2. Análises de marcadores EST-SSR em *P. vulgaris*

As 331 seqüências simples repetidas identificadas *in silico* possibilitaram a proposição de 153 pares de *primers* flanqueando tais regiões (para 46,2 % dos motivos SSR). Todos os pares de *primers* desenhados no presente trabalho estão disponíveis *on line* em: <http://www.vigna.ufpe.br/media/neto/>. A discrepância entre o número de SSRs e pares de *primers* decorre do fato de muitos motivos SSR se encontrarem próximos a uma das extremidades, impossibilitando o desenho de *primers* apropriados. De uma amostra de 22 pares de *primers* (abrangendo tipos de repetições distintos) testados em reações de PCR com um painel de DNAs de diferentes genótipos de feijão-comum, 15 (68,2 %) pares amplificaram bandas, 11 dos quais de fragmentos com tamanhos esperados, sendo 10 desses com geração de polimorfismos (Fig 1). O Conteúdo de Informação Polimórfica (PIC) médio desse último grupo de marcadores foi de 0,50 (variando de 0,18 a 0,72). Esses dados vão de encontro à maioria dos relatos da literatura, os quais sugerem que a natureza conservada de EST-SSR promove a transferibilidade, contudo reduz o caráter polimórfico e, por conseqüência, sua informatividade (Cho *et al.*, 2000; Scott *et al.*, 2000; Rungis *et al.*, 2004), sendo o nível de polimorfismo desses marcadores menor do que o

de SSR oriundo de bibliotecas genômicas (*gSSRs*) (Fraser *et al.*, 2001; Gupta *et al.*, 2003; Chabane *et al.*, 2005). Entretanto, alguns autores têm divulgado níveis de polimorfismo para EST-SSR equiparados aos dos *gSSRs*. Fraser *et al.* (2004) relataram que 93,5 % dos EST-SSRs foram polimórficos e segregaram em uma população de mapeamento derivada do cruzamento de dois genótipos diplóides de *Actinidia chinensis*. Por sua vez, Poncet *et al.* (2006) observaram que 75% dos EST-SSRs foram polimórficos em indivíduos da espécie *Coffea liberica* (PIC médio de 0,57), enquanto que 86 % o foram em indivíduos de *C. canephora* (PIC médio de 0,59). Em estudo com espécies do gênero *Medicago*, Eujayl *et al.* (2004) sugeriram que os altos índices polimórficos (PIC médio de 0,66 e cerca de 70% dos pares de *primers* sendo polimórficos entre os materiais testados) em relação aos obtidos em ensaios utilizando-se espécies altamente domesticadas, como arroz e trigo (Cho *et al.*, 2000; Eujayl *et al.*, 2002), eram devido, em parte, à grande variação existente dentro desse grupo. Entretanto, Leigh *et al.* (2003) analisando genótipos de trigo, em outro ensaio, obtiveram 96,9 % de polimorfismo para EST-SSRs testados, equiparando-os aos 97,0 % dos *gSSRs*, obtidos no mesmo ensaio.

### 3.3 Análise de marcadores EST-SSR em *G. max*

Dos 4.553 motivos SSR detectados em ESTs de soja, um total de 1.928 pares de *primers* flanqueadores desses motivos foram possíveis de serem desenhados, abrangendo cerca de 42,3% dos motivos detectados. Desse total, 1.219 pares (63,2 %) foram desenhados com base em seqüências EST com anotações disponíveis, o que pode favorecer o mapeamento de marcadores funcionais, quando usado em progênie para essa finalidade. Dos 20 pares de *primers* sintetizados, nove (45,0 %) amplificaram DNAs de genótipos distintos de soja, mas somente quatro pares (20,0 %) amplificaram bandas de tamanho esperado e, desses, apenas dois (10,0 %) foram polimórficos. O elevado percentual de pares de *primers* não-funcionais (55,0 %) vai de encontro aos dados da literatura que relatam uma taxa média de não sucesso em ampliações de 10,0 – 40,0 % (Thiel *et al.*, 2003; Kota *et al.*, 2001; Yu *et al.*, 2004; Saha *et al.*, 2004; Cordeiro *et al.*, 2001; Gupta *et al.*, 2003). As possíveis causas para tal acontecimento podem ser: (1) uso de seqüências com conteúdo informativo questionável (qualidade de bases não é passível de verificação em bancos públicos de seqüências); (2) pares de *primers* derivados de clones quiméricos de cDNA; (3) um ou ambos os *primers* amplificando ao longo de uma região que sofre *splicing* alternativo; ou (4) a presença de longos íntrons na seqüência de DNA genômico. A qualidade das ESTs é um fator

importante a ser levado em consideração, visto que tal informação não está disponível nessas bases de dados. Sreenivasulu *et al.* (2002), analisando banco de dados de ESTs de cereais, estimaram em até 10 %, as seqüências com mais de três ambigüidades a cada 25 bases. Um dado revelador foi o fato de que oito dos 11 *primers* não-funcionais em soja foram desenhados a partir de seqüências denominadas “TA” (que representam *contigs*). Por outro lado, daqueles que amplificaram (nove), quatro foram de seqüências TA (três deles com tamanhos esperados de amplicons e com motivos SSR presentes em alelos relacionados). Aquele que amplificou fragmento de tamanho não esperado [(P19) TA69560\_3847] não demonstrou o motivo SSR após o sequenciamento (dados não apresentados). A eficiência observada para primers desenhados a partir de *contigs*, portanto, foi de 33,4 %. Para *singlets*, essa eficiência foi de 62,5 % (cinco em oito).

Dos 15 pares de *primers* funcionais em *P. vulgaris*, somente dois foram desenhados a partir de *contigs*, e esses amplificaram produtos de tamanhos não esperados, mas com os motivos conservados. O grande número de pares de *primers* [cinco dos nove funcionais (55,5 %)] amplificando bandas de tamanho não esperado, em soja, foi provavelmente resultante também da amplificação de *introns* ou inserções na seqüência genômica correspondente.

A utilização de *singlets* somente faz sentido em banco não redundante, como esse analisado, enquanto que a clusterização de seqüências com certa ambigüidade pode interferir na eficiência dos *primers* desenhados. Deve ser ressaltado que os *primers* sintetizados foram selecionados em virtude da extensão dos motivos SSR que flanqueavam e não em função de serem oriundos de *singlets* ou *contigs*.

### 3.4 Transferibilidade de marcadores EST-SSR para feijão-caupi

Análises de transferibilidade envolveram: (I) 15 pares de *primers* funcionais com diferentes genótipos de feijão-comum; e (II) nove pares de *primers* funcionais com diferentes genótipos de soja. Do grupo I, 11 (73,3 %) pares amplificaram *amplicons* nos vários genótipos de feijão-caupi, sendo que oito primers amplificaram fragmentos de tamanhos esperados e destes, cinco amplificaram polimorfismos (Fig. 1). Com base nos resultados, o grupo de feijão-caupi analisado apresentou-se menos heterogêneo que o grupo de feijão-comum. O índice PIC médio dos marcadores transferidos, aqueles polimórficos e que com bandas de tamanho esperado, foi de 0,36, variando 0,12 a 0,72, enquanto que os mesmos, em feijão-comum, exibiram valores que oscilaram de 0,18 a 0,72, com PIC médio de 0,52.

Dos quatro pares de *primers* que amplificaram bandas de tamanho não esperado em feijão-comum, três também o fizeram em feijão-caupi (o outro não amplificou) e produziram bandas de mesmo tamanho em ambas as espécies. Análises do seqüenciamento desses amplicons (feijão-caupi e feijão-comum) revelaram significativas similaridades entre ambos e conservação da presença do microssatélite, sendo o tamanho não esperado produzido provavelmente devido à amplificação de um *íntron* (dado não apresentado). O percentual de transferibilidade obtido no presente trabalho foi mais significativo (73,3 %, de feijão-comum para feijão-caupi) do que o obtido por Gutierrez *et al.* (2005) com transferibilidade de marcadores EST-SSR de *Medicago truncatula* para: feijão-fava (*Vicia faba*), em 40,0 % dos casos; duas espécies de grão de bico (*Cicer arietinum* e *Cicer reticulatum*) em 36,3 % e em ervilha (*Pisum sativum*), em 37,6 % dos casos. Como sintenia está intimamente relacionada à distância filogenética entre as espécies consideradas (Kaló *et al.*, 2004) e a transferibilidade ocorre primariamente em função da sintenia, uma possível causa seria o maior grau de parentesco entre as duas espécies estudadas no presente trabalho (Doyle e Luckow, 2003).

Em relação ao grupo II, apenas dois pares de *primers* amplificaram em diferentes genótipos de feijão-caupi e ambos produziram bandas únicas de tamanhos não esperados.

### 3.5 Análise do seqüenciamento de amplicons

Para confirmar a identidade dos fragmentos amplificados e a possível origem dos polimorfismos, foram seqüenciados 38 alelos de feijão-comum e de feijão-caupi e 10 alelos amplificados com os *primers* de soja (dois alelos de feijão-caupi e oito de soja). Conforme observado em estudo prévio realizado por Choumane *et al.* (2004), a amplificação de um loco particular em um genoma, com pares de *primers* desenhados para outra espécie, depende não apenas da distância evolutiva entre essas, mas também da taxa de evolução das seqüências genômicas para os quais os *primers* foram desenhados. Também ressaltaram que a amplificação por esses *primers* de uma determinada região não implica na similaridade ou identidade da seqüência amplificada com a seqüência matriz. O alinhamento das seqüências dos alelos de feijão-caupi, oriundos de amplificações com primers de soja, com as ESTs correspondentes exibiram *e-values* considerados altos (Tabela 3), sugerindo a extensão de uma região não similar à seqüência matriz. Por sua vez, todas as seqüências de feijão-caupi analisadas apresentaram similaridades significativas com suas respectivas ESTs de feijão-comum, com *e-values* variando de  $e^{-21}$  a  $e^{-63}$  (Tabela 3). No geral, os polimorfismos em todas as seqüências analisadas de amplicons de tamanho esperado, decorreram, principalmente, de

variações nas extensões dos motivos SSR (Tabelas 4,5,6), ocorrendo apenas um caso em feijão-caupi (IPA 7419-P18) (Tabela 5) onde, apesar do alelo ser diferente dos demais, o motivo original foi mantido, sendo esse polimorfismo possivelmente resultado de inserções em regiões não-repetitivas. Adicionalmente, um fragmento amplificado com o primer (P16)CV539741 em uma cultivar de feijão-comum (Jaraguá) representou um alelo onde não se observou a presença do motivo SSR (Tabela 4), sendo o polimorfismo, possivelmente, oriundo da perda dessa região repetitiva. Em feijão-caupi, esse mesmo *primer* amplificou um alelo que exibiu um motivo diferente do original, TCTA ao invés de TCT, na cultivar IT85F, entretanto a seqüência do amplicom apresentou significativa similaridade com a sua seqüência matriz (Tabela 5). Tal fato pode estar associado aos complexos mecanismos de evolução dos microssatélites.

Análises detalhadas das seqüências, realizadas pelo *software* Clustal W, determinaram que as regiões flangeadoras dessas repetições foram geralmente bem conservadas (Fig 2.), confirmando o sucesso da transferibilidade entre feijão-comum e feijão-caupi.

#### 4. CONCLUSÕES

Motivos SSR podem ser facilmente detectados em seqüências EST de bancos de dados públicos, e para uma boa parcela desses EST-SSRs identificados, pares de *primers* podem ser propostos, de forma mais acessível do que para SSR genômicos. Marcadores EST-SSR desenvolvidos para *P. vulgaris* apresentaram boa eficiência em reações de PCR e foram facilmente transferidos para *V. unguiculata*, espécie considerada órfã e para a qual não havia *primers* EST-SSRs disponíveis. Esses marcadores podem ser úteis em análises genéticas de ambas as espécies, bem como em mapeamento comparativo. Por outro lado, marcadores EST-SSRs desenvolvidos para *G. max* não apresentaram a mesma eficiência em reações com DNAs de genótipos brasileiros de soja, e nem quando transferidos para *V. unguiculata*. Adicionalmente, nossos dados indicam que o desenho de *primers*, para desenvolvimento de marcadores a partir de ESTs depositadas em banco de dados não-redundantes, deve ser, preferencialmente, executado a partir de *singletons*, afim de aumentar a eficiência do processo, uma vez que possíveis ambigüidades em *contigs* podem gerar seqüências não correspondentes ao que é observado genomicamente.

## 5. REFERÊNCIAS BIBLIOGRÁFICAS

- Altschul SF, Gish W, Miller W, Myers EW e Lipman DJ (1990) Basic Local Alignment Search Tool. *Journal of Molecular Biology* 215: 403-410.
- Alzate-Marin AL, Cervigni GDL, Moreira MA e Barros EG (2005) Seleção assistida por marcadores moleculares visando ao desenvolvimento de plantas resistentes a doenças, com ênfase em feijoeiro e soja. *Fitopatologia Brasileira* 30: 333-342.
- Anderson JA, Churchill GA, Sutriquet JE, Tanksley SD e Sorrells ME (1993) Optimizing parental selection from genetic linkage maps. *Genome* 36: 181-186.
- Ayres NM, McClung AM, Larkin PD, Bligh HFJ, Jones CA e Park WD (1997) Microsatellite and single nucleotide polymorphism differentiate apparent amylose classes in an extended pedigree of US rice germ plasm. *Theoretical and Applied Genetics* 94: 773-781.
- Bassam BJ, Caetano-Anollés G e Gresshoff PM (1991) Fast and sensitive silver staining of DNA in polyacrylamide gels. *Analytical Biochemistry* 196: 80-83.
- Bilgen M, Karaca M e Ince AG (2004) A software program combining sequence motif searches with keywords for finding repeats containing DNA sequences. *Bioinformatics* 20: 3379-3386.
- Bolstein D, White RL, Skolnick M e Davis RW (1980) Construction of genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* 32: 314-331.
- Boutin SR, Young ND, Olsen TC, Yu ZH, Shoemaker RC e Valejos CE (1995) Genome conservation among three legume genera detected with DNA markers. *Genome* 38: 928-937.
- Cannon SB, Sterck L, Rombauts S, Sato S, Cheung F, Gouzy J, Wang X, Mudge J, Vasdewani J, Schiex T, Spannagl M, Monaghan E, Nicholson C, Humphray SJ, Schoof H, Mayer KFX, Rogers J, Quétier F, Oldroyd GE, Debélle F, Cook DR, Retzel EF, Roe BA, Town CD, Tabata S, Van der Peer Y e Young ND (2006) Legume genome evolution viewed through the *Medicago truncatula* and *Lotus japonicus* genomes. *Proceedings of National Academy of Sciences* 103: 14959-14964.
- Chabane K, Ablett GA, Cordeiro GM, Valkoun J e Henry RJ (2005) EST versus genomic derived microsatellite markers for genotyping wild and cultivated barley. *Genetic Resources and Crop Evolution* 52: 903-909.

Childs KL, Hamilton JP, Zhu W, Ly E, Cheung F, Wu H, Rabinowicz PD, Town CD, Buell CR e Chan AP (2007) The TIGR Plant Transcript Assemblies database. *Nucleic Acids Resesarch* 35 (Database issue): D846-D851.

Cho YG, Ishii T, Temnykh S, Lipovich XCL, McCouch SR, Park WD, Ayres N and Cartinhour S (2000) Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.).

Choi H-K, Kim D, Uhm T, Limpens E, Lim H, Mun J, Kalo P, Penmetsa RV, Seres A, Kulikova O, Roe BA, Bisseling T, Kiss B e Cook DR (2004b) A sequence-based genetic map of *Medicago truncatula* and comparison of marker co-linearity with *Medicago sativa*. *Genetics* 166: 1463-1502.

Choi H-K, Mun J-H, Kim D-J, Zhu H, Baek J-M, Mudge J, Roe B, Ellis N, Doyle J, Kiss GB, Young ND e Cook DR (2004a) Estimating genome conservation between crop and model legume species. *Proceedings of National Academy of Sciences* 101: 15289-15294.

Choumane W, Winter P, Baum M and Kahl G (2004) Conservation of microsatélite flanking sequences in different taxa of Leguminosae. *Euphytica* 138: 239-245.

Cordeiro GM, Casu R, McIntyre CL, Manners JM and Henry RJ (2001) Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to erianthus and sorghum. *Plant Science* 160: 1115-1123.

Doyle JJ e Doyle JL (1987) A rapid DNA isolation procedure from small quantities of fresh leaf tissues. *Phytochemical Bulletin* 19: 11-15.

Doyle JJ e Luckow MA (2003) The rest of the iceberg. Legume diversity and evolution in a phylogenetic context. *Plant Physiology* 131: 900-910.

Ellegren H (2004) Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics* 5: 435-445.

Eujayl I, Sledge MK, Wang L, May GD, Chekhovskiy K and Zwonitzer JC e Mian MAR (2004) *Medicago truncatula* EST-SSRs reveal cross-species genetic markers for *Medicago* spp. *Theoretical and Applied Genetics* 108: 414-422.

Eujayl I, Sorrels ME, Wolters P, Baum M e Powell W (2002) Isolation of EST-derived microsatélite larkers for genotyping the A and B genomes of wheat. *Theoretical and Applied Genetics* 104: 399-407.

Fraser LG (2004) EST-derived microsatellites from *Actinidia* species and their potential for mapping. *Theoretical and Applied Genetics* 108: 1010-1016.

Freire Filho FR, Ribeiro VQ, Barreto PD and Santos CAF (1999) Melhoramento genético

do caupi (*Vigna unguiculata* (L.) Walp.) na região Nordeste. In: Queiroz MA, Goedert CO; Ramos SRR (Org) Recursos genéticos e melhoramento de plantas para o nordeste brasileiro. 1st edition. Embrapa-CPATSA, Petrolina, pp.

Gao L, Tang J Li H e Jia J (2003) Analysis of microsatellites in major crops assessed by computational and experimental approaches. *Molecular Breeding* 12: 245-261.

Gao L, Tang J Li H e Jia J (2003) Analysis of microsatellites in major crops assessed by computational and experimental approaches. *Molecular Breeding* 12: 245-261.

Gupta PK, Rustgi S, Sharma S, Singh R, Kumar N e Balyan HS (2003) Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. *Molecular Genetics and Genomics* 270: 315-323.

Gupta PK, Rustgi S, Sharma S, Singh R, Kumar N e Balyan HS (2003) Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. *Molecular Genetics and Genomics* 270: 315-323.

Gutierrez MV, Patto MCV, Huguet T, Cubero JI, Moreno MT e Torrez AM (2005) Cross-species amplification of *Medicago truncatula* microsatellites across three major pulse crops. *Theoretical and Applied Genetics* 110: 1210-1217.

Gutierrez MV, Patto MCV, Huguet T, Cubero JI, Moreno MT e Torrez AM (2005) Cross-species amplification of *Medicago truncatula* microsatellites across three major pulse crops. *Theoretical and Applied Genetics* 110: 1210-1217.

Humphry ME, Konduri V, Lambrides CJ, Magner T, McIntyre CL, Aitken EAB e Liu CJ (2002) Development of a mungbean (*Vigna radiata*) RFLP linkage map and its comparison with lablab (*Lablab purpureus*) reveals a high level of colinearity between the two genomes. *Theoretical and Applied Genetics* 105: 106-166.

Jayashree B, Punna R, Prasad P, Bantte K, Hash CT, Chandra S, Hoisington DA e Varshney RK (2006) A database of simple sequence repeats from cereal and legume expressed sequence tags mined in silico: survey and evaluation. *In Silico Biology* 6. <http://www.bioinfo.de/isb/2006/06/0054/>.

Jurka, J e Pethiyagoda C (1995) Sequences from primates: Compilation and analysis. *Journal of Molecular Evolution* 40: 120-126.

Kaló P, Seres A, Taylor SA Jakab J, Kevei Z, Kereszt A, Endre G, Ellis THN e Kiss GB (2004) Comparative mapping between *Medicago sativa* and *Pisum sativum*. *Molecular Genetics and Genomics* 272: 235-246.

Kantety RV, La Rota M, Matthews DE e Sorrells ME (2002) Data mining for simple

sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. *Plant Molecular Biology* 48: 501-510.

Kashi Y e King DG (2006) Simple sequence repeats as advantageous mutators in evolution. *Trends in genetics* 22: 253-259.

Kota R., Varshney RK, Thiel T, Dehmer KJ and Graner A (2001) Generation and comparison of EST-derived SSRs and SNPs in barley (*Hordeum vulgare L.*). *Hereditas* 135: 145-15

Kumar LS (1999) DNA markers in plant improvement: An overview. *Biotechnology Advances* 17: 143-182.

Kumpatla SP e Mukhopadhyay S (2005) Mining and survey of simple sequence repeats in expressed sequence tags of dicotyledonous species. *Genome* 48: 985-998.

La Rota M, Kantety RV, Yu J-K e Sorrells ME (2005) Nonrandom distribution and frequencies of genomic and EST-derived microsatellite markers in rice, wheat, and barley. *BMC Genomics* 6: 23-35.

Lee JM, Grant D, Vallejos CE e Shoemaker RC (2001) Genome organization in dicots. II. *Arabidopsis* as a bridging species to resolve genome duplication events among legumes. *Theoretical and Applied Genetics* 103: 765-773.

Leigh F, Lea V, Law J, Wolters P, Powell W e Donini P (2003) Assessment of EST-and genomic microsatellite markers for variety discrimination and genetic diversity studies in wheat. *Euphytica* 133: 359-366.

Li Y-C, Abraham BK, Fahima T e Nevo E (2004) Microsatellites within genes: Structure, function, and evolution. *Molecular Biology and Evolution* 21: 991-1007.

Liewlaksaneeyanawin C, Ritland CE, EL-Kassaby YA e Ritland K (2004) Single-copy, species-transferable microsatellite markers developed from lobolly pine ESTs. *Theoretical and Applied Genetics* 109: 361-369.

Menancio-Hautea D, Fatokun CA, Kumar L, Danesh D e Young ND (1993) Comparative genome analysis of mungbean (*Vigna radiata (L.) Wilczek*) and cowpea (*V. unguiculata (L.) Walpers*) using RFPL mapping data. *Theoretical and Applied Genetics* 86: 797-810.

Metzgar D, Bytof J and Wills C (2002) Selection against frameshift mutations limits microsatellite expansion in coding DNA. *Genome Research* 10: 72-80.

Morgante M, Hanafey M e Powell W (2002) Microsatellite are preferentially associated with nonrepetitive DNA in plant genomes. *Nature Genetics* 30: 194-200.

Mudge J, Cannon SB, Kalo P, Oldroyd GED, Roe BA, Town CD e Young ND (2005)

Highly syntenic regions in the genomes of soybean, *Medicago truncatula*, and *Arabidopsis thaliana*. *BMC Plant Biology* 5: 15-30.

Naylor RN, Falcon WP, Goodman RM, Jahn MM, Sengooba T, Tefera H e Nelson RJ (2004) Biotechnology in the developing world: a case for increased investments in orphan crops. *Food Policy* 29: 15-44.

Nelson RJ, Naylor RL e Jahn MM (2004) The role of genomics in improvement of “orphan” crops. *Crop Science* 44: 1901-1904.

Olufajo OO e Singh BB. (2002) Advances in cowpea cropping systems research. In: Challenges and opportunities for enhancing sustainable cowpea production. Proceedings of the World Cowpea Conference III, held at the International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria, 4–8 September 2000. pp 267–277.

Poncet V, Rondeau M, Tranchant C, Cayrel A, Hamon S, de Kochko A e Hamon P (2006) SSR mining in coffee tree EST database: potencial use of EST-SSRs as markers for the *Coffea* genus. *Molecular Genetics and Genomics* 276: 436-449.

Powell W, Machray GC e Provan J (1996) Polymorphisms revealed by simple sequence repeats. *Trends in Plant Science* 1: 215-222.

Rackoczy-Trojanowska M e Bolibok H (2004) Characteristics and comparison of three classes of microsatellite-based markers and their applications in plants. *Cellular & Molecular Biology Letters* 9: 221-238.

Rungis D, Bérubé Y, Zhang J, Ralph S, Ritland CE , Ellis BE, Douglas C, Bohlmann J e Ritland K, (2004) Robust simple sequence repeats markers for spruce (*Picea* spp.) from expressed sequence tags. *Theoretical and Applied Genetics* 109: 1283-1294.

Saha MC, Rouf Mian MA, Eujayl I, Zwonitzer JC, Wang L e May GD (2004) Tall fescue EST-SSR markers with transferability across several grasses species. *Theoretical and Applied Genetics* 109: 783-791.

Scott KD, Eggler P, Seaton G, Rossetto M, Ablett EM, Lee LS e Henry RJ (2000) Analysis of SSRs derived from grape ESTs. *Theoretical and Applied Genetics* 100: 697-712.

Simon MV, Benko-Iseppon AM, Resende LV, Winter P e Kahl G (2007) Genetic diversity and phylogenetic relationship in *Vigna savi* germoplasm revealed by DNA amplification fingerprint. *Genome* 50: 538-547.

Singh BB (2005) *Vigna unguiculata* L. Walp. In: Singh, RJ e Jahuar PP (eds) (2005) Genetic Resources, Chromosome Engineering, and Crop Improvement. Grain Legumes.

Vol. 1. 1st edition. CRC Press, Boca Raton, pp 116-151.

Sreenivasulu N, Ksihor PBK, Varshney RK and Altschmied L (2002) Mining functional information from cereal genomes – the utility of expressed sequence tags. *Current Science* 83: 965-973.

Thiel T, Michalek W, Varshney RK e Graner A (2003) Exploiting EST databases for development and characterization of gene-derived SSR-markers in braley (*Hordeum vulgare* L.). *Theoretical and Applied Genetics* 106: 411-422.

Tóth G, Gáspári Z e Jurka J (2000) Microsatellites in different eukaryotic genomes: Survey and analysis. *Genome Research* 10: 967-981.

Trifonov EN (2004) The tuning function of tandemly repeating sequences: a molecular devices for fast adaptation. In: Wasser SP (ed) *Evolutionary Theory and Processes*. 1st edition. Modern Horizons, Kluwer Academic Publishers, Dordrecht, pp 115-138.

Varshney RK, Graner A e Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. *Trends in Biotechnology* 23: 48-55.

Yu JK, Dake TM, Singh S; Benscher D, Li W, Gill B e Sorrells ME (2004) Development and mapping of EST-derived simple sequence repeat (SSR) markers for hexaploid wheat. *Genome* 47: 805-818.

Zane L, Bargelloni L e Patarnello T (2002) Strategies for microsatellite isolation: a review. *Molecular Ecology* 11: 1-16.

Zhu H, Choi H-K, Cook DR e Schoemaker C (2005) Bridging model and crop legumes through comparative genomics. *Plant Physiology* 137: 1189-1196.

**Tabela 1.** Características dos motivos SSR encontrados em ESTs de *Phaseolus vulgaris* (P.v) e *Glycine max* (G.m).

CARACTERÍSTICAS	MOTIVO									
	DIMÉRICO		TRIMÉRICO		TETRAMÉRICO		PENTAMÉRICO		HEXAMÉRICO	
	P. v	G. m	P. v	G. m	P. v	G. m	P. v	G. m	P. v	G. m
Frequencia Absoluta	135	2247	151	1793	13	159	2	40	30	285
Freq. relativa (%)	(40,8)	(49,7)	(45,6)	(39,6)	(3,6)	(3,5)	(0,6)	(09)	(9,1)	(6,3)
Extensão Média (pb)	39,0	38,2	19,9	21,1	21,5	22,2	25,0	26,7	25,2	25,8
Motivo mais freqüente	<u>AG/GA</u> TC/CT	<u>AG/GA</u> TC/CT	<u>AAG/AGA/GAA</u> TTC/TCT/CTT	<u>AAG/AGA/GAA</u> TTC/TCT/CTT	uniforme	CTTT e derivações	TTCTT e derivações	CTCTT, GTTTG e derivações	uniforme	AACCCT e derivações
Freq. motivo (%)	43,7	64,9	44,4	27,2	-	27,7	100,0	20,0 cada	-	6,7

**Tabela 2.** Posicionamento dos SSRs nas seqüências analisadas de *Phaseolus vulgaris* (P. v) e *Glycine max* (G. m).

MOTIVO SSR	DIMÉRICO		TRIMÉRICO		TETRAMÉRICO		PENTAMÉRICO		HEXAMÉRICO	
	P.v	G. m	P. v	G. m	P. v	G. m	P. v	G. m	P. v	G. m
5' UTR (%)	68,8	51,0	31,3	31,0	50,0	45,0	50,0	61,5	29,0	27,0
ORF (%)	16,0	21,0	54,4	53,5	0	18,0	0	7,5	51,6	58,5
3' UTR (%)	15,2	28,0	14,3	15,5	50,0	37,0	50,0	31,0	19,4	14,5

**Tabela 3.** Motivos SSR em ESTs de *Phaseolus vulgaris* ou *Glycine max* e em amplicons de *Vigna unguiculata*, e valores de *e-value* em alinhamentos BLAST.

Primer / EST	Espécie	Motivo (rep)	<i>Vigna unguiculata</i>		<i>e-value</i> *
			Genótipo-banda	Motivo (rep)	
(P4) Pcv538096	<i>P. vulgaris</i>	AT (16)	IT85F-P4	AT (17)	e <sup>-21</sup>
(P6) Pcb544086	<i>P. vulgaris</i>	CT (25)	BR14Mulato-P6	CT (26)	e <sup>-21</sup>
(16) Pcv539740	<i>P. vulgaris</i>	TCT (10)	IT85F-P16	TCTA (4)	e <sup>-35</sup>
(P17) Pcv531509	<i>P. vulgaris</i>	TCA (10)	BR14Mulato-P17	TCA (6)	e <sup>-63</sup>
(P21) Pcv533580	<i>P. vulgaris</i>	CTCCTG (7)	BR14Mulato-P21	CTCCTG (5)	e <sup>-25</sup>
(P6S) Pbq253715	<i>G. max</i>	AAT (16)	BR14Mulato-P6S	-	e <sup>-03</sup>
(P7S) Pbu544951	<i>G. max</i>	TA (27)	BR14Mulato-P7S	-	e <sup>-01</sup>

*e-value*: alinhamento (BLAST2SEQ) entre EST de *P. vulgaris* ou *G. max* contra amplicon de *V. unguiculata* correspondente.

**Tabela 4.** Primers EST-SSR, respectivos motivos e repetições nas seqüências ESTs, com amplificação de monomorfismos ou polimorfismos em genótipos de *P. vulgaris* e amplicons de tamanho esperados (TE) ou não (S: sim; N: não), número de alelos e motivos (repetições) em amplicons seqüenciados (genótipo/ banda) e valores *e-value* entre as seqüências EST e amplicons (Blastn).

<i>Primer</i>	<b>EST Motivo (rep)</b>	<b>Amplificação em <i>P. vulgaris</i></b>	<b>TE</b>	<b>Nº Alelos</b>	<b>Genótipo – banda</b>	<b>Amplicon Motivo (rep)</b>	<b><i>e-value</i></b>
P12 (CV543440)	CTT (8)	Polimorfismos	Não	1	-	-	-
P5 (CB544086)	TA (19)	Polimorfismos	Não	3	Jaraguá-P5	TA (8)	e <sup>-42</sup>
P11 (TA3919)	TAA (8)	Polimorfismos	Não	3	Princesa-P11	TAA (5)	e <sup>-111</sup>
					Jaraguá-P11	TAA (8)	e <sup>-112</sup>
					Luminoso-P11	TAA (9)	e <sup>-113</sup>
P22 (TA3994)	ATGAAC (6)	Polimorfismos	Não	3	Esplendor-P22	ATGAAC (8)	e <sup>-87</sup>
					Galante-P22	ATGAAC (2)	e <sup>-37</sup>
P2 (CV543232)	GA (18)	Polimorfismos	Sim	3	-	-	-
P6 (CB539191)	CT (25)	Polimorfismos	Sim	5	-	-	-
P4 (CV538096)	AT (16)	Polimorfismos	Sim	3	Boreal-P4	AT (16)	e <sup>-58</sup>
					Jaraguá-P4	AT (20)	e <sup>-60</sup>
P8 (CV543035)	TAT (7)	Monomorfismo	Sim	1	Jaraguá-P8	TAT (7)	e <sup>-41</sup>
P13 (CV533448)	AAG (10)	Polimorfismos	Sim	2	Jaraguá-P13	AAG (9)	e <sup>-104</sup>
					Boreal-P13	AAG (10)	e <sup>-149</sup>
P16 (CV539741)	TCT (10)	Polimorfismos	Sim	3	Esplendor-P16	TCT (9)	e <sup>-33</sup>
					Jaraguá-P16	Ausente	e <sup>-22</sup>
					Gurgutuba-P16	TCT (14)	e <sup>-31</sup>
P17 (CV531509)	TCA (10)	Polimorfismos	Sim	3	Jaraguá-P17	TCA (11)	e <sup>-87</sup>
					IPA7419-P17	TCA (9)	e <sup>-27</sup>
					Harmonia-P17	TCA (6)	e <sup>-41</sup>
					Esplendor-P18	AATG (6)	e <sup>-105</sup>
P18 (CB542646)	AATG (6)	Polimorfismos	Sim	3	Gurgutuba-P18	AATG (6)	e <sup>-92</sup>
					Esplendor-P19	ATGA (5)	e <sup>-102</sup>
P19 (57022)	ATGA (7)	Polimorfismos	Sim	2	Gurgutuba-P19	ATGA (7)	e <sup>-27</sup>
					Gurgutuba-P20	TTTC (7)	e <sup>-30</sup>
P20 (CB540660)	TTTC (7)	Polimorfismos	Sim	3	Alvorada-P20	TTTC (5)	e <sup>-27</sup>
					Esplendor-P21	CTCCTG (3)	e <sup>-31</sup>
P21 (CV533580)	CTCCTG (7)	Polimorfismos	Sim	2	Esplendor-P21	CTCCTG (3)	e <sup>-31</sup>

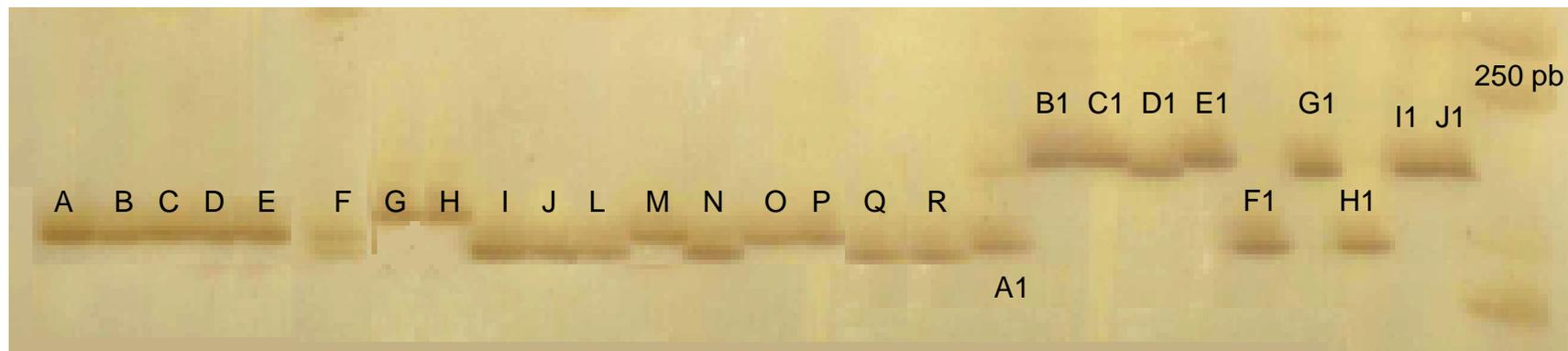
**Tabela 5.** *Primers* EST-SSR, respectivos motivos e repetições nas seqüências ESTs de *P. vulgaris*, com amplificação de monomorfismos ou polimorfismos em genótipos de *V. unguiculata* e *amplicons* de tamanho esperados (TE) ou não (S: sim; N: não), número de alelos e motivos (repetições) em *amplicons* seqüenciados (genótipo/ banda) e valores *e-value* entre as seqüências EST e *amplicons* (Blastn).

<i>Primer</i>	EST Motivo (rep)	Amplificação em <i>V. unguiculata</i>	TE	Nº Alelos	Genótipo – banda	<i>Amplicon</i> Motivo (rep)	<i>e-value</i>
P5 (CB544086)	TA (19)	Monomorfismo	Não	1	BR14Mulato-P5	TA (15)	$e^{-28}$
P8 (CV543035)	TAT (7)	Monomorfismo	Sim	1	BR14Mulato-P8	TAT (5)	$e^{-41}$
P11 (TA3919)	TAA (8)	Monomorfismo	Não	1	BR14Mulato-P11	TAA (8)	$e^{-75}$
P13 (CV533448)	AAG (10)	Monomorfismo	Sim	1	BR14Mulato-P13	AAG (10)	$e^{-23}$
P19 (57022)	ATGA (7)	Monomorfismo	Sim	1	IPA207-P19	ATGA (7)	$e^{-24}$
P22 (TA3994)	ATGAAC (6)	Monomorfismo	Não	1	BR14Mulato-P22	ATGAAC (4)	$e^{-79}$
P4 (CV538096)	AT (16)	Polimorfismos	Sim	2	BR14Mulato-P4	AT (14)	$e^{-30}$
					IT85F-P4	AT (17)	$e^{-21}$
P6 (CB539191)	CT (25)	Polimorfismos	Sim	4	BR14Mulato-P6	CT (26)	$e^{-21}$
P16 (CV539741)	TCT (10)	Polimorfismos	Sim	3	IPA207-P16	TCT (12)	$e^{-28}$
					IT85F-P16	TCTA (4)	$e^{-35}$
P18 (CB542646)	AATG (6)	Polimorfismos	Sim	3	IPA7419-P18	AATG (6)	$e^{-63}$
P21 (CV533580)	CTCCTG (7)	Polimorfismos	Sim	2	BR14Mulato-P21	CTCCTG (5)	$e^{-25}$

**Tabela 6.** *Primers* EST-SSR, respectivos motivos e repetições nas seqüências ESTs, com amplificação de monomorfismos ou polimorfismos em genótipos de *G. max* e *amplicons* de tamanho esperados (TE) ou não (S: sim; N: não), número de alelos e motivos (repetições) em *amplicons* seqüenciados (genótipo/ banda) e valores *e-value* entre as seqüências EST e *amplicons* (Blastn).

<i>Primer</i>	EST Motivo (rep)	Amplificação em <i>G. max</i>	TE	Nº Alelos	Genótipo/banda	<i>Amplicon</i> Motivo (rep)	<i>e-value</i>
P4 (BI974060)	AAT (20)	Monomorfismo	Não	1	-	-	-
P7 (BU544951)	TA (27)	Monomorfismo	Não	2	-	-	-
P19 (TA69560_3847)	CTT (13)	Monomorfismo	Não	1	Fepagro25-P19	ausente	1.8
P3 (BG882912)	TTTC (8)	Polimorfismos	Não	3	-	-	-
P6 (BQ253715)	AAT (16)	Polimorfismos	Não	2	Fepagro25-P6	AAT (20)	e <sup>-93</sup>
P8 (CA800521)	AT (31)	Monomorfismo	Sim	1	Fepagro25-P8	TA (15)	e <sup>-40</sup>
P13 (TA44541_3847)	AAACCT (5)	Monomorfismo	Sim	1	Fepagro25-P13	AAACCT (5)	e <sup>-41</sup>
P14 (TA50965_3847)	CTTT (7)	Polimorfismos	Sim	2	Fepagro25-P14	CTTT (5)	e <sup>-109</sup>
P20 (TA72884_3847)	CCTAAC (9)	Polimorfismos	Sim	5	Fepagro25-P20	CCTAAC (4)	e <sup>-42</sup>
					BR232-P20	CCTAAC (6)	e <sup>-39</sup>
					CD217-P20	CCTAAC (5)	e <sup>-43</sup>

**Figura 1.** Polimorfismos em genótipos de *P. vulgaris* e *V. unguiculata* amplificados pelo *primer* CV539741 (P16). Os intervalos A a R contém genótipos de *V. unguiculata*, e A1 a J1 contém genótipos de *P. vulgaris*.



**Figura 2.** Alinhamento (Clustal W) entre seqüência de *V. unguiculata* oriunda de produto de PCR (BR14Mulato-P17) e sua EST (CV531509) correspondente de *P. vulgaris*. Alinhamento contém: (1) seqüências de *primers* (*left* e *right*), indicadas pelas setas; (2) motivo SSR em negrito e itálico e; (3) posições dos nucleotídeos conservados indicados por asteriscos.



## INFORMAÇÕES COMPLEMENTARES

As tabelas seguintes apresentam um resumo de todas as análises realizadas no presente trabalho, além de exibirem as informações classificadas como “dados não apresentados” no manuscrito científico.

Nove pares de *primers* foram funcionais em *Glycine Max*, sendo que dois desses foram transferidos para *V. unguiculata*. A tabela sete apresenta o resultado obtido para essa atividade, além de análises de similaridade entre o *amplicon* de *V. unguiculata* seqüenciado e sua EST (de *G.max*) de correspondente.

A tabela oito apresenta os *primers* não-funcionais em soja, destacando o tipo de seqüência (*singlet* ou *contig*) em que os *primers* foram propostos.

A tabela nove apresenta análises de similaridade entre *amplicons* de *V. unguiculata* e *P. vulgaris* de tamanho não esperado, sugerindo que tais regiões seriam homólogas.

As tabelas dez e onze apresentam informações gerais sobre todos os *primers* utilizados no presente trabalho bem como dados sobre os SSRs por eles flanqueados.

A tabela doze apresenta a análise dos resultados (scoreamento de bandas) das amplificações realizadas com os *primers* oriundos de ESTs de *Phaseolus vulgaris*.

**Tabela 7.** *Primers* EST-SSR, respectivos motivos e repetições nas seqüências ESTs de *G. max*, com amplificação de monomorfismos ou polimorfismos em genótipos de *V. unguiculata*, número de alelos e motivos (repetições) em *amplicons* seqüenciados e valores *e-value* entre as seqüências EST e *amplicons* (Blastn).

<i>Primer</i>	EST Motivo (rep)	Amplificação em <i>V. unguiculata</i>	Nº Alelos	Genótipo – banda	<i>Amplicon</i> Motivo (rep)	<i>e-value</i>
P6 (BQ253715)	AAT (16)	Monomorfismo	1	BR14-P6	Ausente	0.001
P7 (BU544951)	TA (27)	Monomorfismo	2	BR14-P7	Ausente	0.1

**Tabela 8.** Relação de *primers* EST-SSR não funcionais em PCRs com DNAs de soja, motivos e repetições na seqüência original de *G. max* (ID TIGR) e tipo de seqüência (*singlet* ou *contig*).

Primer (ID TIGR)	Motivo (rep)	Seqüência
P1 (BE331188)	TTTC (7X)	<i>singlet</i>
P2 (BF067601)	CTCTT (5X)	<i>singlet</i>
P5 (BI974194)	CTGGG (6X)	<i>singlet</i>
P9 (TA40585_3847)	CT (36X)	<i>contig</i>
P10 (TA42316_3847)	GA (31X)	<i>contig</i>
P11 (TA42973_3847)	TC (36X)	<i>contig</i>
P12 (TA44398_3847)	AG (53X)	<i>contig</i>
P15 (TA55178_3847)	TTA (15X)	<i>contig</i>
P16 (TA58735_3847)	TTAA (7X)	<i>contig</i>
P17 (TA63397_3847)	ATT (21X)	<i>contig</i>
P18 (TA63471_3847)	TA (45X)	<i>contig</i>

**Tabela 9.** Primers transferidos, mas com TE (tamanho esperado) não observado. BLASTn realizado entre *amplicon* de V.u (*Vigna uniuiculata*) e *amplicon* de P.v (*Phaseolus vulgaris*). (S): sim; (N): não.

<b>Primers</b>	<b>Amplificação em. P.v e V.u</b>	<b>TE</b>	<b>Blast entre:</b>	<b>e-value</b>
P5 (CB544086)	S	N	Jaraguá-P11 e BR14-P11	e-73
P11 (TA3919)	S	N	Luminoso-P11 e BR14-P11	e-26
P22 (TA3994)	S	N	Esplendor-P22 e BR14-P22	e-48

**Tabela 10.** Especificações dos *primers* de *Phaseolus vulgaris* utilizados no presente trabalho, tamanho esperado do *amplicon*, além da localização e características dos SSRs flanqueados por eles.

<i>Primer</i>	Motivo (Nº de repetições)	Localização do SSR	<i>Left primer</i>	<i>Right primer</i>	<i>Amplicon</i> (pb)	T <sub>m</sub>
P1 (CV530714)	AG(14)	5' UTR	GCGTCCGAGAACCTTTTAATC	AGCCATCAACGAAATCCAAC	346	60
P2 (CV543232)	GA(18)	ORF	AAATTTGGGGGAAATTGAGG	TGCATCATGCCATGTTTCTC	306	60
P3 (TA4786)	GA(10)	5' UTR	ACGCGTCCGGTTCACAGT	AGAACCGCAGCTTCTCAAAA	200	60
P4 (CV538096)	AT(16)	5' UTR	TGCTGACCCTTCAACACGTA	CAGGCAAAGCAGACATGAAA	355	60
P5 (CB544086)	TA(19)	ORF	TTCACAAGATTGGGGAGACC	AAGTCACAGCCCTCACTGCT	337	60
P6 (CB539191)	CT(25)	ORF	TGCTCCATTACAATTACCAAGG	AGAAAGGGGTGAGGGAAAAA	253	60
P7 (TA3313)	TC(17)	ORF	TAGGGTTTTGGTGGGAAGT	AGGCTGACGGAAAATGTGAC	379	60
P8 (CV543035)	TAT(7)	ORF	TCGGATCATTCCATCTACCC	CAAAAATTCCCCAGAGGAT	193	60
P9 (TA3145)	ATT(7)	3' UTR	AGAGAGAGATGCAGGGGTTG	GAATTCGGCACGAGGATAAA	389	60
P10 (CV533580)	TTA(8)	5' UTR	GCGTGCAGTATGAAGCTGAG	AGCTGTGCAGAAGATGCAGA	373	60
P11 (TA3919)	TAA(8)	3' UTR	GTCAAAGCAGCTCCAAAGG	ATTGCAGCATAGCCCTACCA	332	60
P12 (CV543440)	CTT(8)	ORF	CACCCATTCTCTCTCAGCA	AACACGTCTTGGCGAAGTTT	355	60
P13 (CV533448)	AAG(10)	ORF	GGCAATGATGGATTTGAAGG	GCAGGTGGTGGTTTTGAAGT	396	60
P14 (CB280641)	TTC(10)	3' UTR	TAATTCTCGCAAACCCTTCG	GGAAGCTTTGCATCTGGAAG	380	60
P15 (CV536084)	AGA(8)	5' UTR	TGAACATCGGAATCGAATAGG	GGATTGCTGTTCCCATAGA	315	60
P16 (CV539741)	TCT(10)	5' UTR	AGGCATTCTCACAGCACAGA	AAAGGGGTGGAAGCTTGAGT	178	60
P17 (CV531509)	TCA(10)	ORF	ATCAGTGTGTTTCGCTGTGC	CTCAGGGAACCCTTCATTCA	354	60
P18 (CB542646)	AATG(6)	3' UTR	CAAGGAAGGCGAAGATGAAG	CCATAATCCATCATGCACCA	348	60
P19 (X57022)	ATGA(7)	3' UTR	ACCTTCTTCGGAAAGGATGG	ATGCACTGCACTTGACGAAC	353	60
P20 (CB540660)	TTTC(7)	5' UTR	GCCGTACACACAACAAAAC	TTTCCATCCACCTGAAGAGG	248	60
P21 (CV533580)	CTCCTG(7)	ORF	GCGTGCAGTATGAAGCTGAG	AGCTGTGCAGAAGATGCAGA	156	60
P22 (TA3994)	ATGAAC(6)	ORF	GGTGATTCAGGTGTGGGTTC	GCTGCTTCGTACACCTCCTC	375	60

**T<sub>m</sub>**: temperatura de *melting* do par de primers em questão; **5'UTR**: região não-traduzida 5'; **3' UTR**: região não-traduzida 3'.

**Tabela 11.** Especificações dos *primers* de *Glycine max* utilizados no presente trabalho, tamanho esperado do *amplicon*, além da localização e características dos SSRs flanqueados por eles.

<i>Primer</i>	Motivo (N° de repetições)	Localização do SSR	<i>Left primer</i>	<i>Right primer</i>	<i>Amplicon</i> (pb)	T <sub>m</sub>
(P1) BE331188	TTTC(7)	5' UTR	AGCGGCACCAGGAAAGTT	CCCGAATTCCATCATGTACC	299	60
(P2) BF067601	CTCTT(5)	5' UTR	TTACGAGTTCGACCCCTCCT	GTAGGCTCACATTCCCAGAG	381	60
(P3) BG882912	TTTC(8)	5' UTR	CCTCAATTGGGGTGTGAAAG	CCCGAATTCCATCATGTACC	307	60
(P4) BI974060	AAT(20)	ORF	AACTTCATGGGCTGATTTGG	GTGTTTTCTGCTGCTCCTC	351	60
(P5) BI974194	CTGGG(6)	ORF	GTGCTCTGCTCTGCTCCTTT	CAAGACAACCTGAACGCTGA	389	60
(P6) BQ253715	AAT(16)	ORF	CAAAAGATGCCTTCCCTCAA	CGGAAGGTGTAGAAGGGTTG	358	60
(P7) BU544951	TA(27)	ORF	GATGCTGGTGGGAGATGTTC	GCAGGATGACACAAACATGC	208	60
(P8) CA800521	AT(31)	ORF	TGATCCTCCACATGGCTACA	GATGGACAGAAGCAAACACG	272	60
(P9) TA40585	CT(36)	5' UTR	TAAATGGTGGTTGCCTCCTC	TCTCTTCCATGCCCTCCTAA	295	60
(P10) TA42316	GA(31)	5' UTR	GCATGGGAGTCTTGTTGTT	TATGAGGTGGTGAAGGCACA	233	60
(P11) TA42973	TC(36)	ORF	GACAAAACGCCGAAACTTTG	GCTCAGCAAGAAGGACCAAG	269	60
(P12) TA44398	AG(53)	5' UTR	TTGGGCAAGACAAGAACACC	CATTCGAGGTCCAATCCAGT	400	60
(P13) TA44541	AAACCT(5)	5' UTR	CGTCCCTTTCTGAGCGTAAC	GGGGATTTGGGAAATTCAGT	175	60
(P14) TA50965	CTTT(7)	5' UTR	CGAGCCTCACGCTTAAACTC	TGAAGGGGACAAAGATGAGG	354	60
(P15) TA55178	TTA(15)	3' UTR	TTGCTATGCAACGACATGGT	GCTTCCCCTGAACCGAGTA	275	60
(P16) TA58735	TTAA(7)	3' UTR	CTTCTCCCTTGGGTTCTTCC	CATCCTAACCGGTCAAAAACA	129	60
(P17) TA63397	ATT(21)	5' UTR	CCGGGTTTTATGGGAAGAGT	GCTTAGCGGTATCCTCAACG	351	60
(P18) TA63471	TA(45)	3' UTR	CTGGGACTTGATGTGCTTGA	GAGAGAGAAAATGTGTAGTGGTG	330	60
(P19) TA69560	CTT(13)	5' UTR	TCCTTTGCTTCTTCCAACA	TTGCCAAGACTTAGCCACCT	214	60
(P20) TA72884	CCTAAC(9)	ORF	GGCACAGTCTCAGAGCAACA	GCGTAGCCCACTGACTCTTC	171	60

**T<sub>m</sub>**: temperatura de *melting* do par de primers em questão; **5' UTR**: região não-traduzida 5'; **3' UTR**: região não-traduzida 3'.

**Tabela 12.** Análise dos *amplicons* produzidos nos diferentes indivíduos de *Phaseolus vulgaris* e *Vigna unguiculata* utilizados.

Primers	Alelos	MT	Es	Bo	Já	Pr	Gu	Lu	Ga	Al	IPA	Há	TE1	TE2	TE3	TE4	Br	IT	IPAA	IPAb	Pi	Vi	Pa	BRs	Fra	Br17	BRsP	Fv	VI
P1 (CV530714)	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
P2 (CV543232)	1	1	1	1	1	0	0	0	0	0	0	0	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	2	0	0	0	0	1	1	1	1	0	0	0	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	3	0	0	0	0	0	0	0	0	1	1	1	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
P3 (TA4786)	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
P4 (CV538096)	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0
	3	0	1	1	0	0	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	0	1
P5 (CB544086)	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
P6 (CB539191)	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	1	1	0	0	1	0	0	1	0	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0
	3	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0
	4	0	0	0	0	0	0	1	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	1	0	0	1	0	0
	5	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	1	1	0	1	0	0	1	0	0	1
P7 (TA3313)	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
P8 (CV543035)	1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	2	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P9 (TA3145)	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
P10 (CV533580)	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
P11 (TA3919)	1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	2	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	4	1	1	1	1	0	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P12 (CV543440)	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
P13 (CV533448)	1	0	0	1	0	0	0	1	0	0	0	1	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	2	1	1	0	1	1	1	0	1	1	1	0	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
P14 (CB280641)	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x

Cont.

Primers	Alelos	Mt	Es	Bo	Ja	Pr	Gu	Lu	Ga	Al	IPA	Ha	TE1	TE2	TE3	TE4	Br	IT	IPAA	IPAb	Pi	Vi	Pa	BRs	Fra	BR17	BRsP	Fv	VI	
P15 (CV536084)	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	
P16 (CV539741)	1	1	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	1	0	0	1	1	1	0	1	0	0	0	1	
	2	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0	1	0	1	1	1	0	
	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	
	4	0	0	0	0	1	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P17 (CV531509)	1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	
	2	0	1	1	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	3	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	4	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
P18 (CB542646)	1	1	0	0	0	0	0	0	0	0	0	0	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	
	2	0	0	0	1	1	1	1	1	1	1	0	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	
	3	0	1	1	0	0	0	0	0	0	0	1	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	
P19 (X57022)	1	1	1	1	1	1	0	0	1	1	1	1	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	
	2	0	0	0	0	0	1	1	0	0	0	0	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	
P20 (CB540660)	1	1	1	1	1	1	0	0	0	1	1	0	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	
	2	0	0	0	0	0	0	0	1	0	0	0	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	
	3	0	0	0	0	0	1	1	0	0	0	1	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	
P21 (CV533580)	1	1	1	0	1	1	1	1	1	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
	2	0	0	1	0	0	0	0	0	0	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	
P22 (TA3994)	1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
	2	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	3	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	4	1	1	1	1	1	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

**Legenda:** Mt (Macarrão trepador), Es (Esplendor), Bo (Boreal), Ja (Jaraguá), Pr (Princesa), Gu (Gurgutuba), Lu (Luminoso), Ga (Galante), Al (Avorada), IPA (IPA 7419), Ha (Harmonia), TE1 (TE97G12), TE2 (TE9612G), TE3 (TE9764), Br (BR14 mulato), IT (IT85F), IPAA(IPA206), IPAb (IPA207), Pi (Pitiúba), Vi (Vita 7), Pa (Patativa), BRs (BRS Guariba), Fra (Fradinho), BR17 (BR17 Gurguéia), BRsP (BRS Paraguaçu), FV (Feijão de vagem) e VI (Verde ligeiro).

## 6. CONCLUSÕES

Motivos SSR podem ser facilmente detectados em seqüências EST de bancos de dados públicos, e para uma boa parcela desses EST-SSRs identificados, pares de *primers* podem ser propostos, de forma mais acessível do que para SSR genômicos. Marcadores EST-SSR desenvolvidos para *P. vulgaris* apresentaram boa eficiência em reações de PCR e foram facilmente transferidos para *V. unguiculata*, espécie para a qual não havia *primers* EST-SSRs disponíveis. Esses marcadores podem ser úteis em análises genéticas de ambas as espécies, bem como em mapeamento comparativo. Por outro lado, marcadores EST-SSR desenvolvidos para *G. max* não apresentaram a mesma eficiência em reações com DNAs de genótipos brasileiros de soja, e nem quando transferidos para *V. unguiculata*.

## 7. ABSTRACT

Expressed Sequence Tags (EST) database is a valuable source of information for molecular markers development. In the present work, simple sequence repeats (SSR) have been detected in *Phaseolus vulgaris* and *Glycine max* ESTs, aiming the development of EST-SSR markers for cowpea (*Vigna unguiculata*) improvement. A total of 10,880 common bean ESTs and 116,965 soy ESTs were analyzed and the estimated averages frequencies were one SSR in every 22.0 kb for common bean EST and one in every 14.7 Kb for soy EST. In common bean ESTs, 331 SSRs were identified and the trimeric motifs were the most abundant (45.6%). In soy ESTs, 4,553 SSRs were detected and the dimeric motifs were the most prevalent (49.7%). In both species, the AG/GA / TC/CT dinucleotide motifs and the AAG/AGA/GAA / TTC/TCT/CTT trinucleotide motifs were the most abundant. A total of 153 common bean EST-SSR primers pairs and 1,928 soy primers pairs were designed. Nine of a 20 soy EST-SSR primers pairs amplified soy DNAs. After transferability study to *V. unguiculata*, two of these primers pairs amplified but a non-expected size bands with no conserved original motif were observed. On the other hand, 15 of 22 common bean EST-SSR primers pairs amplified in PCR reactions with DNAs from 11 common bean genotypes and 10 of them showed polymorphism (estimated average PIC of 0.50). With cowpea DNAs (17), 11 of these 15 primers pairs amplified several amplicons. Five of them showed polymorphism (estimated average PIC of 0.36). Some PCR products were sequenced and the alignments with the correspondent EST showed in general that the SSR motif extensions modified but the sequences ends were conserved between species. These results confirmed the success of the EST-SSR markers transferability between common beans and cowpea.

**KEYWORDS:** microsatellites, data mining, Fabaceae, leguminous, cowpea.

## **8. ANEXO**

## **8.1. Anexo 1**

### **INSTRUÇÕES PARA AUTORES**

**Revista**

***GENETICS AND MOLECULAR BIOLOGY***

**ISSN 1415-4757**

**Ribeirão Preto, Brasil**

## INSTRUCTIONS TO AUTHORS

## SCOPE AND POLICY

**Genetics and Molecular Biology** (formerly named Revista Brasileira de Genética/Brazilian Journal of Genetics - ISSN 0100-8455) is published by the Sociedade Brasileira de Genética (Brazilian Society of Genetics).

The Journal considers contributions that present the results of original research in genetics, evolution and related scientific disciplines.

Although **Genetics and Molecular Biology** is an official publication of the Brazilian Society of Genetics, contributors are not required to be members of the Society.

It is a fundamental condition that submitted manuscripts have not been and will not be published elsewhere. With the acceptance of a manuscript for publication, the publishers acquire full and exclusive copyright for all languages and countries.

Manuscripts considered in conformity with the scope of the journal as judged by the Editor in conjunction with the Editorial Board are reviewed by the Associate Editors and two or more external reviewers. Acceptance by the Editor is based on the quality of the work as substantial contribution to the field and on the overall presentation of the manuscript.

## SUBMISSION OF PAPERS

1. Manuscripts should be submitted to:

Angela M. Vianna-Morgante, Editor-in-Chief  
Genetics and Molecular Biology

Postal Address:

Rua Capitão Adelmio Norberto da Silva, 736  
14025-670 Ribeirão Preto, SP – Brazil

Electronic Address:

editor@gmb.org.br

2.A submission package sent to the Editorial Office must contain **New**

a) A cover letter stating that all authors have approved the submission of the manuscript. The letter must be signed by the Corresponding Author and must inform the e-mail addresses of all other authors so that they can be contacted by the Editorial Office for confirmation of the submission.

b) An electronic copy of the text, tables and figures, including supplementary material to be published online only. Formats for text are Word or RTF in Windows platform. Images in TIFF or JPEG formats should be sent in separate files (For Figures, see detailed instructions in 3.1.h). Mailed CD-ROMs must be labeled with the first author's last name, platform and software. Failure to adhere to these guidelines can delay the handling of your contribution and manuscripts may be returned before being reviewed.

c) Manuscripts including photos or any other identifiable data of human subjects must be accompanied by a copy of the signed consent by the individual or his/her guardian.

## 3. Categories of Contribution

## 3.1. Research Articles

Manuscripts must be written in English in double-spaced, 12-point type throughout; formatted to A4 paper with 2.5 cm margins; marked with

consecutive line and page numbers, beginning with the cover page.

The following elements must start on a new page and be ordered as they are listed below:

**a) The title page** must contain: a concise and informative title; the authors' names (first name at full length); the authors' institutional affiliation, including department, institution, city, state or province and country; different affiliations indicated with superscript numbers; a short running title of about 35 characters, including spaces; up to five key words; the corresponding author's name, postal address, phone and fax numbers and email address. The corresponding author is the person responsible for checking the page proofs, arranging for the payment of color illustrations and author's alteration charges.

**b) The Abstract** must be a single paragraph that does not exceed 200 words and summarizes the main results and conclusions of the study. It should not contain references.

**c) The text** must be as succinct as possible. *Text citations*: articles should be referred to by authors' surnames and date of publication; citations with two authors must include both names; in citations with three or more authors, name the first author and use *et al.* List two or more references in the same citation in chronological order, separated by semi-colons. When two or more works in a citation were published in the same year, list them alphabetically by the first author surname. For two or more works by the same author(s) in a citation, list them chronologically, with the years separated by commas. (Example: Freire-Maia *et al.*, 1966a, 1966b, 2000). Only articles that are published or in

press should be cited. In the case of personal communications or unpublished results, all contributors must be listed by initials and last name (*et al.* should not be used). *Numbers*: In the text, numbers nine or less must be written out except as part of a date, a fraction or decimal, a percentage, or a unit of measurement. Use Arabic numerals for numbers larger than nine. Avoid starting a sentence with a number. *Binomial Names*: Latin names of genera, species and intraspecific taxa in the text must be printed in italics; names of orders and families should appear in the Title and also when first mentioned in the text. URLs for programs, data or other sources should be listed in the Internet Resources Section, immediately after the References Section, not in the text. URLs for citations of publications in electronic journals should appear in the reference section

The text includes the following elements:

**Introduction** –Description of the background that led to the study.

**Material (or Subjects) and Methods** – Details relevant to the conduct of the study. Statistical methods should be explained at the end of this section.

**Results** – Undue repetition in text and tables should be avoided. Comment on significance of results is appropriate but broader discussion should be part of the Discussion section.

**Discussion** – The findings of the study should be placed in context of relevant published data. Ideas presented in other publications should not be discussed solely to make an exhaustive presentation.

Some manuscripts may require different formats appropriate to their content.

**d) The Acknowledgments** must be a single paragraph that immediately follows the discussion and includes references to grant support.

**e) The References Section:** references must be ordered alphabetically by the first author surname; references with the same first author should be ordered as follows: first, as single author in chronological order; next, with only one more co-author in alphabetical order by the second author; and finally followed by references with more than two co-authors, in chronological order, independent of the second author surnames. In references with more than 10 authors only the first ten should be listed, followed by *et al.* Use standard abbreviations for journal titles as suggested by ISI Web of Knowledge ([http://apps.isiknowledge.com/WoS/help/C\\_abrvjt.html](http://apps.isiknowledge.com/WoS/help/C_abrvjt.html)) or PubMed (<http://www.ncbi.nlm.nih.gov>).

Only articles that are published or in press should be included in this section. Works submitted for publication but not yet accepted, personal communications and unpublished data must be cited within the text. "Personal communication" refers to individuals other than the authors of the manuscript being submitted; "unpublished data" refers to data produced by at least one of the authors of the manuscript being submitted.

*Sample journal article citation:*

Breuer ME and Pavan C (1955) Behaviour of polytene chromosomes of *Rhynchosciara angela* at different stages of larval development. *Chromosoma* 7:371-386.

Yonenaga-Yassuda Y, Rodrigues MT and Pellegrino KCM (2005) Chromosomal banding patterns in the eyelid-less microteiid lizard radiation: The X<sub>1</sub>X<sub>1</sub>X<sub>2</sub>X<sub>2</sub>:X<sub>1</sub>X<sub>2</sub>Y sex chromosome system in *Calyptommatus* and the karyotypes of *Psilophthalmus* and *Tretioscincus* (Squamata, Gymnophthalmidae). *Genet Mol Biol* 28:700-709.

*Sample book citation:*

Dobzhansky T (1951) *Genetics and Origin of Species*. 3rd edition. Columbia University Press, New York, 364 pp.

*Sample chapter-in-book citation:*

Crawford DC and Howard-Peebles PN (2005) Fragile X: From cytogenetics to molecular genetics. In: Gersen SL and Keagle MB (eds) *The Principles of Clinical Cytogenetics*. 2nd edition. Humana Press, New Jersey, pp 495-513.

*Sample Electronic Article citation:*

Simin K, Wu H, Lu L, Pinkel D, Albertson D, Cardiff RD and Van Dyke T (2004) pRb inactivation in mammary cells reveals common mechanisms for tumor initiation and progression in divergent epithelia. *Plos Biol* 2:194-205. <http://www.plosbiology.org> .

## **f) Internet Resources Section**

this section should contain a list of URLs referring to data presented in the text, software programs and other Internet resources used during data processing. Date of consultation must be stated.

*Sample Internet Resource citation :*

Online Mendelian Inheritance in Man (OMIM),

<http://www.ncbi.nlm.nih.gov/OMIM>  
(September 4, 2005)

LEM Software,  
[http://dir.niehs.nih.gov/dirbb/weinbergfiles/hybrid\\_design.htm](http://dir.niehs.nih.gov/dirbb/weinbergfiles/hybrid_design.htm) (September 4, 2005)

**g) Tables:** must be inserted at the end of the main text file, each table starting on a new page. A concise title should be provided above the table. Tables must be numbered consecutively in Arabic numerals. Each column must have a title in the box head. Footnotes typed directly below the table should be indicated in lowercase superscript numbers.

**h) Figures** must be numbered consecutively in Arabic numerals. Legends should be typed on a new page that immediately follows the tables. Images should be in TIFF or JPEG format and provided in separate files. Identify each illustration by the first author name and the number of the respective figure. Figures in Word format cannot be published. Journal quality reproduction will require grayscale and color at resolution yielding 300 dpi. Authors should submit bitmapped line art at resolution yielding 600-1200 dpi. These resolutions refer to the output size of the file; if it is anticipated that images will be enlarged or reduced, the resolutions should be adjusted accordingly. Scanned figures should not be submitted. Color illustration can be accepted, but authors may be asked to defray the cost.

**i) Nomenclature** should adhere to current international standards.

**j) Sequences** may appear in text or in figure. DNA, RNA and protein sequences equal to or greater than 50 units must be entered into public databases. The accession number must

be provided and released to the general public together with publication of the article. Long sequences requiring more than two pages to reproduce will not be published unless the Editorial decision is that the publication is necessary. Complete mtDNA sequence will not be published.

**k) Data access:** reference should be made to availability of detailed data and materials used for reported studies.

**l) Ethical issues:** Reports of experiments on live vertebrates must include a brief statement that the institutional review board approved the work and the protocol number must be provided. For experiments involving human subjects, authors must also include a statement that informed consent was obtained from all subjects. If photos or any other identifiable data are included, a copy of the signed consent must accompany the manuscript.

**m) Supplementary Material:** Data that the authors consider of importance for completeness of a study, but which are too extensive to be included in the published version, can be submitted as Supplementary Material. This material will be made available together with the electronic version. In case a manuscript contains such material, it should be appropriately identified within the text. Supplementary material in tables should be identified as Table S1, Table S2, etc., in case of figures they should be named accordingly, Figure S1, Figure S2. For online access, supplementary material should be in PDF, JPEG or GIFF formats.

In addition, a list of this material should be presented at the end of the manuscript text file, containing the

following statement: *Supplementary material - the following online material is available for this article:*

- *Table S1* – < short title >

- *Figure S1* – < short title >

*This material is available as part of the online article from <http://www.scielo.br/gmb>*

### 3.2 Short Communications

Present brief observations that do not warrant full-length articles. They should not be considered preliminary communications. They should be 15 or fewer typed pages in double spaced 12-point type, including literature cited. They should include an Abstract no longer than five percent of the paper's length and no further subdivision with introduction, material and methods, results and discussion in a single section. Up to two tables and two figures may be submitted. The title page and reference section format is that of full-length article.

### 3.3 Letters to the Editor

Relate or respond to recent published items in the journal. Discussions of political, social and ethical issues of interest to geneticists are also welcome in this form.

### 3.4 Review Articles

Review Articles are welcome.

### 3.5 Book Reviews

Publishers are invited to submit books on Genetics, Evolution and related disciplines, for review in the journal. Aspiring reviewers may propose writing a review.

### 3.6 History, Story and Memories

Accounts on historical aspects of Genetics relating to Brazil.

### 4. Proofs and Copyright Transfer

Page proofs will be sent to the corresponding author. Changes made to page proofs, apart from printer's errors, will be charged to the authors. Notes added in proof require Editorial approval. A form of consent to publish and transfer of copyright will have to be signed by the corresponding author, also on behalf of any co-authors.

### 5. Reprints

Reprints are free of charge and provided as a pdf-file.

### *Subscription*

Membership to the Brazilian Society of Genetics entitles subscription to Genetics and Molecular Biology.

For nonmembers and institutions, the annual subscriptions rates (four issues/year) are:

#### **Brazil and other South American Countries (Air Mail):**

Institutional - R\$ 500,00

Personal: R\$ 150,00.

#### **Other Countries (Air Mail):**

Institutional - US\$ 300.00

Personal - US\$ 100.00.