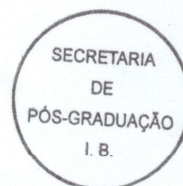


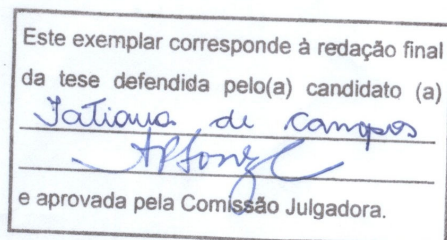
UNIVERSIDADE ESTADUAL DE CAMPINAS

Instituto de Biologia



Tatiana de Campos

DESENVOLVIMENTO DE MARCADORES MOLECULARES
MICROSSATÉLITES, MAPEAMENTO GENÉTICO E MAPEAMENTO DE
CARACTERÍSTICAS QUALITATIVAS EM FEIJOEIRO (*Phaseolus vulgaris* L.)



Tese apresentada ao Instituto de Biologia para obtenção do Título de Doutor em Genética e Biologia Molecular na área de Genética Vegetal e Melhoramento.

Orientadora: Profa. Dra. Anete Pereira de Souza

Co-orientadora: Profa. Dra. Luciana Lasry Benchimol

Campinas

2009

**FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DO INSTITUTO DE BIOLOGIA – UNICAMP**

C157d	Campos, Tatiana Desenvolvimento de marcadores microssatélites, mapeamento genético e mapeamento de características qualitativas em feijoeiro (<i>Phaseolus vulgaris</i> L.) / Tatiana Campos. – Campinas, SP: [s.n.], 2009. Orientadores: Anete Pereira de Souza, Luciana Lasry Benchimol. Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Biologia. 1. Feijoeiro. 2. <i>Phaseolus vulgaris</i>. 3. Mapeamento genético. 4. Marcadores microssatélites. I. Souza, Anete Pereira. II. Benchimol, Luciana Lasry. III. Universidade Estadual de Campinas. Instituto de Biologia. IV. Título. (scs/ib)
--------------	--

Título em inglês: Development of microsatellite markers, genetic mapping and qualitative characteristics mapping in common bean (*Phaseolus vulgaris* L.).

Palavras-chave em inglês: Common bean; *Phaseolus vulgaris*; Genetic mapping; Microsatellite markers.

Área de concentração: Genética Vegetal e Melhoramento.

Titulação: Doutora em Genética e Biologia Molecular.

Banca examinadora: Anete Pereira de Souza, Marie-Anne Van Sluys, Maria Imaculada Zucchi, Michel Georges Albert Vincentz, Sergio Augusto Morais Carbonell.

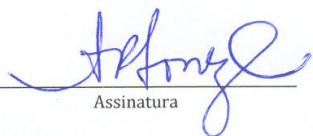
Data da defesa: 17/02/2009.

Programa de Pós-Graduação: Genética e Biologia Molecular.

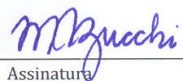
Campinas, 17 de Fevereiro de 2009.

BANCA EXAMINADORA

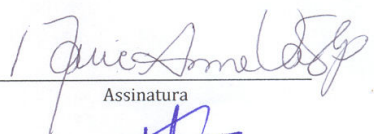
Profa. Dra. Anete Pereira de Souza (Orientadora)


Assinatura

Prof. Dra. Maria Imaculada Zucchi


Assinatura

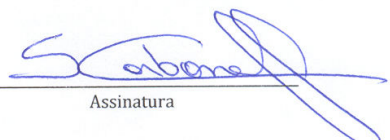
Profa. Dra. Marie-Anne Van Sluys


Assinatura

Prof. Dr. Michel Georges Vincentz


Assinatura

Prof. Dr. Sergio Augusto Morais Carbonell


Assinatura

Profa. Dra. Maria Elisa Ayres Guidetti Zagatto Paterniani

Assinatura

Prof. Dr. Sérgio Furtado dos Reis

Assinatura

Profa. Dra. Maria do Carmo Estanislau do Amaral

Assinatura

Dedico

Aos meus pais **Paulo** e **Yukie**,
À minha irmã **Raquel**,
Ao meu namorado **Thiago**,
pelo apoio e paciência em todos os momentos.

Ofereço

À minha orientadora **Anete**,
À minha co-orientadora **Luciana**,
pelos valiosos ensinamentos.

Agradeço

Aos meus pais Paulo e Yukie, por sempre acreditarem em mim e nos meus sonhos, por se orgulharem das minhas conquistas, pela preciosa educação, pela minha vida.

À minha irmã Raquel, por me ajudar a crescer, por fazer parte da minha vida.

Ao meu namorado Thiago, por ser um excelente companheiro e meu melhor amigo, por me fazer muito feliz e me completar.

À minha orientadora Dra. Anete Pereira de Souza, pela grande orientação, pelos conhecimentos passados com muita paciência e dedicação, pelo apoio e incentivo, por me ensinar a fazer pesquisa.

À minha co-orientadora Dr. Luciana Lasry Benchimol, por toda disponibilidade no esclarecimento das minhas dúvidas, por me ajudar a crescer.

Aos colegas do laboratório de Análises Genética e Molecular / CBMEG e Barracão, pelos conselhos e ajuda nos momentos de dúvidas, pelos momentos de descontração, pelo carinho durante estes anos de convivência.

Ao companheiro feijoeiro Danilo, pela valiosa e fundamental ajuda no desenvolvimento do projeto, pela união nas horas de dificuldade, pelo espírito de grupo, pela compreensão.

À Leticia e ao Vitor, pela atenciosa ajuda, pelo companheirismo de uma amizade em momentos de alegria e também pela força nos momentos de dificuldades.

À grande amiga Adna, pelo carinho e apoio de uma amizade sincera.

Ao professor Dr. Antônio Augusto Franco Garcia, por toda gentileza e disponibilidade no esclarecimento das minhas dúvidas, por me apresentar de forma clara a estatística como aliada.

Aos funcionários, Juverlande e Zildinha, aos técnicos, Georgina, Paulo, Felipe, Vânia, Paty, Deyse e demais técnicos que passaram pelo laboratório, por serem sempre prestativos e gentis, pela participação essencial no bom andamento dos experimentos.

Aos professores participantes da pré-banca, Dra. Maria Imaculada Zucchi e Dr. Sergio Furtado dos Reis, pela disponibilidade em participar da avaliação deste trabalho.

Aos professores participantes da banca, Dra. Maria Imaculada Zucchi, Dra. Marie-Anne Van Sluys, Dr. Michel Georges Vincentz, Dr. Sergio Augusto Morais Carbonell, pela análise da tese, comentários construtivos e tempo despendido.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico pela bolsa DD concedida (140310/2005-3).

À Fundação de Amparo à Pesquisa do Estado de São Paulo pelo auxílio financeiro do projeto (Proc. FAPESP 02/03225-9; 05/53819-0).

A todos que contribuíram no andamento deste trabalho e tornaram possível a sua realização.

Sumário

Lista de Tabelas	X
Lista de Figuras.....	XII
Prefácio	XIV
Resumo	XV
Abstract.....	XVII
1. Introdução.....	1
2. Revisão Bibliográfica	4
2.1. A Cultura do Feijoeiro.....	4
2.2 Evolução de Marcadores	10
2.2.1 Marcadores Microsatélites	13
2.2.2. Microsatélites na Cultura do Feijoeiro.....	17
2.3. Mapas Genéticos.....	21
2.3.1. Teste de Segregação.....	25
2.3.2. Teste de Ligação	26
2.3.3. Ordenação de Marcadores.....	27
2.3.4. Mapas Genéticos em Feijoeiro	28
2.4. Características Quantitativas e Qualitativas em Feijoeiro	33
3. Objetivo.....	38
3.1 Objetivos Específicos	39
4. Artigo I	

“Structure of genetic diversity among common bean (<i>Phaseolus vulgaris</i> L.) varieties of Mesoamerican and Andean origins using new developed microsatellite markers”	40
5. Artigo II	
“Development, characterization, and comparative analysis of polymorphism at common bean SSR loci isolated from genic and genomic sources”	57
6. Artigo III	
“Microsatellites for genetic studies and breeding programs in common bean”	70
7. Artigo IV	
“A linkage map for common bean (<i>Phaseolus vulgaris</i> L.) based on genic and genomic microsatellite markers”	75
8. Discussão	103
9. Conclusões	107
10. Perspectivas	109
11. Bibliografia	110

Lista de Tabelas

4 Artigo I

Table 1 Common bean (<i>P. vulgaris</i> L.) accessions used in the determination of the allelic variation of the microsatellites	43
---	----

Table 2 SSRs' annealing temperatures (Ta), primer sequences, predicted size of products lengths (base pairs, bp), number of alleles, range of the products (base pairs, bp) for each locus and polymorphic information content (PIC).....	46
---	----

5 Artigo II

Table 1 Number of cycles and annealing temperature (Ta) of PCR programs used to amplify common bean EST-SSR loci	60
--	----

Table 2 Frequency of individual SSR motif among 240 microsatellites isolated from the BEST database	61
---	----

Table 3 PvM loci derived from common bean expressed sequence tags (ESTs).....	62
---	----

Table 4 FJ loci derived from the enriched microsatellite library	64
--	----

6 Artigo III

Table 1 Primer sequences and characteristics of 20 common bean (<i>Phaseolus vulgaris</i>) microsatellite markers, obtained from variety IAC-UNA, tested on 14 accessions, comprising Mesoamerican and Andean gene pools	73
--	----

7 Artigo IV

Table 1 Description of the new mapped microsatellites, including primer sequences, repeat motif, allele size and annealing temperature (Ta)	83
---	----

Lista de Figuras

2 Revisão Bibliográfica

Figura 1 A)Localização dos 3 centros de origem de <i>Phaseolus vulgaris</i> : região Andina Norte e Sul e, região da Mesoamérica com destaque em B)para os países que formam a região	5
---	---

4 Artigo I

Figure 1 Number and characteristics of microsatellite repeats found after screening an enriched SSR library. The other categories include tetra-, penta and hexanucleotide motifs	45
---	----

Figure 2 Graph based on the BLASTX homology of the 123 sequences for the non-redundant database. The results are percentages over the total of similarity (27%) achieved from alignment. No hits represented 73% of the data	52
--	----

Figure 3 Amplification products for microsatellite SSR-IAC49 with common bean genotypes on a 6% denaturing polyacrilamide gel. (A,B) Amplification products for microsatellite (A) SSR-IAC62 and (B) SSR-IAC66 with common bean genotypes on 6% denaturing polyacrilamide gels. The numbers follow the order presented in Table 1. F1 is the hybrid material generated by IAC-UNA (19) X CAL143 (20) cross	52
--	----

Figure 4 UPGMA cluster analyses of genetic distances calculated from SSR data. Values at nodes represent the levels of bootstrap support (in percentages).....	53
--	----

Figure 5 Associations between 22 common bean genotypes revealed by principal coordinate analysis of Rogers' genetic distances	54
---	----

5 Artigo II

Figure 1 Distribution of bean EST-SSR loci according to their gene position..... 61

Figure 2 Display of a *Phaseolus vulgaris* microsatellite marker (PvM21) among 25 bean genotypes detected on silver-stained 6% polyacrilamide gels..... 66

Figure 3 Relationships among 25 bean genotypes. The unrooted tress ((A) from EST-SSR markers and (B) from genomic SSR markers) were constructed from the modified Rogers' distance matrix using UPGMA (unweighted paired group method using arithmetic averages) and Drawgram programs of the PHYLIP package (Felsenstein 2004)..... 67

6 Artigo IV

Figure 1 Locus FJ54 genotyped at polyacrilamide gel (6%) with 4 applications of 48 samples for time 88

Figure 2. Multiplex genotyping of locus SSR-IAC288 and SSR-IAC261 at automatic sequencer with two fluorescent primers: FAM (in black) and HEX 88

Figure 3 Genetic Map for IAC-UNA X CAL143 (UC). B1 to B11 are referent to linkage groups and, I to XI are the chromosome correspondence according to Pedrosa *et al.* (2003). Microsatellites underlined were anchored to the linkage groups. (*) Markers with segregation distortion to IAC-UNA and (**) to CAL143. (***) Phenotypic marker, *V* locus, linked at group B6. Arrows indicate possible position of *fin* and *Ct* loci. 90

Prefácio

Os resultados obtidos durante o desenvolvimento deste trabalho de tese estão apresentados na forma de artigos científicos. O primeiro artigo, *Structure of genetic diversity among common bean (Phaseolus vulgaris L.) varieties of Mesoamerican and Andean origins using new developed microsatellite markers*, publicado na revista *Genetic Resources and Crop Evolution* (54: 1747-1762; 2007), descreve o desenvolvimento de microssatélites a partir de uma biblioteca enriquecida, a caracterização e a aplicação destes locos para a análise de diversidade em 24 acessos. O segundo artigo, *Development, characterization, and comparative analysis of polymorphism at common bean SSR loci isolated from genic and genomic sources*, publicado na revista *Genome* (50: 266-277; 2007), apresenta a caracterização de novos locos microssatélites e a comparação destes com EST-SSRs. O terceiro artigo, *Microsatellites for genetic studies and breeding programs in common bean*, publicado na revista *Pesquisa Agropecuária Brasileira* (42: 589-592; 2007), descreve a validação de novos locos microssatélites. O quarto artigo, *Development of a new genetic map based on microsatellite markers in common bean (Phaseolus vulgaris L.)*, submetido à revista *Theoretical and Applied Genetics*, descreve a obtenção do mapa genético com os marcadores microssatélites desenvolvidos e também daqueles disponíveis na literatura, além do mapeamento de marcadores fenotípicos.

Resumo

O melhoramento genético do feijoeiro busca responder ou atender demandas específicas dos produtores. Um cultivar de feijoeiro deve atender às características de produtividade, de resistência às principais doenças da cultura, e de qualidade tecnológicas, tais como tempo de cozimento, qualidades nutricionais e tipo de caldo. A escolha de critérios racionais e eficientes para a identificação de linhagens superiores a serem utilizadas em cruzamentos facilita o trabalho do melhorista. Além disso, a escolha dos genitores com maior potencial genético para recombinação aumenta as chances de obtenção de cultivares mais produtivos e estáveis para o comércio, bem como facilita a seleção dos genótipos nos ensaios de competição.

O feijoeiro é uma das culturas de destaque no Brasil com potencial a ser explorado e requer esforços para o manejo adequado. Neste sentido, o investimento em técnicas moleculares pode ser associado ao auxílio no melhoramento clássico, como para a realização de programas de seleção assistida por marcadores. Os microssatélites são marcadores moleculares definidos por sequências repetitivas abundantes nos genomas de eucariotos, transferíveis e informativas. Os microssatélites são marcadores amplamente utilizados em estudos genéticos e no desenvolvimento de mapas genéticos.

Este trabalho teve como objetivo o desenvolvimento de marcadores moleculares do tipo microssatélites para feijoeiro para a construção de um mapa genético molecular. Foram desenvolvidos 488 novos microssatélites, sendo que 183 estão disponíveis em 3 publicações sobre caracterização destes locos e, 64 estão descritos no artigo referente ao mapa genético. Os demais locos

apresentaram-se monomórficos dentre os genótipos utilizados, mas também serão divulgados na forma de manuscrito.

O mapa genético foi estabelecido com base em uma população F_{10} segregante, composta de 380 linhagens endogâmicas, derivadas do cruzamento entre IAC-UNA e CAL143. Foram analisados no total 871 microssatélites, entre eles 265 (30,4%) foram polimórficos e 247 (28,4%) apresentaram padrão adequado de leitura de genotipagem. Para a construção do mapa, além dos marcadores moleculares, foram utilizados três marcadores fenotípicos: cor de flor, formato do ápice da vagem e hábito de crescimento. Foi possível mapear 198 microssatélites e os 3 marcadores fenotípicos. Dentre os marcadores mapeados, 131 tinham a posição até então desconhecida em grupos de ligação. O mapa resultante cobre 1865.9 cM, com uma distância média entre marcadores de 9.4 cM. A cobertura de mapa é considerada de saturação moderada e pode servir de base para o mapeamento de outras características fenotípicas e de QTLs.

Abstract

The common bean breeding programs are meant to attend specific demands of bean producers. A common bean genotype of commercial interest must present desirable productivity characteristics; resistance to the main diseases and technological quality, like cooking time, nutritional value and type of broth. The choice of efficient criteria for identification of superiors inbred lines to be used in crosses supports breeders work. Moreover, the use of genitors with higher genetic potential increases the chances to reach more stable and productive cultivars, as well as facilitates genotypes selection in competitive assays.

The common bean is one of the most important crops in Brazil and it has potential to be explored, but, the current productivity is low and requires efforts to improve the field performance. In this way, the search for molecular techniques can be associated to assist the classic breeding, like to perform marker-assisted selection. Microsatellites are repetitive sequences present in eukaryotes genomes, transferable, and informative. The microsatellite markers are widely used in genetic studies, and one of the main uses is to construct genetic maps.

The objective of the present work was to develop new microsatellite markers to common bean and to construct a genetic map. Up to now 488 microsatellites were developed, of which 183 are available in 3 articles about loci characterization and, 64 are described in genetic map article. The remaining loci were monomorphic for the genotypes used, and they will be described in an article too.

The genetic map was based on a mapping population F_{10} , formed by 380 recombinant inbred lines derived of IAC-UNAF and CAL143 crosses. We tested 871 microsatellites, of which 265 (30,4%) were polymorphic and 247 (28,4%) presented adequate genotyping standard. Beyond the molecular markers, we

evaluated three phenotypic markers: flower color, pod tip shape, and growth habit. It was possible to map 198 microsatellites and the 3 phenotypic markers. Amongst the mapped microsatellite markers, 131 have never been located before in any known linkage group. The resulting map covers a total of 1865.9 cM in length and average distance between markers was 9.4 cM. The coverage of the generated map is considered to have a moderate saturation, which makes it useful for mapping other qualitative traits and QTLs.

1. Introdução

A espécie *Phaseolus vulgaris* é uma leguminosa favorável para estudos genéticos por possuir uma série de características que facilitam tais estudos. As mais importantes são: apresentar o genoma relativamente pequeno (637 Mb; Arumuganathan & Earle, 1991), genética diplóide ($2n = 2x = 22$) e, pelo fato do feijoeiro apresentar ciclo de vida curto. O feijoeiro é uma planta herbácea e sua reprodução é por autofecundação. Devido ao mecanismo de cleistogamia, tem apenas 5% de fecundação cruzada, e por isso, encontra-se em homozigose.

O feijoeiro é uma das culturas mais importantes para o Brasil e, devido à produtividade atual ser considerada baixa, requer esforços para o melhoramento de seu desempenho em campo (Silva, 2000). Neste sentido, o investimento em técnicas moleculares associadas ao melhoramento clássico tem crescido.

Espera-se maior contribuição dos marcadores moleculares para seleção das características quantitativas, pelas dificuldades envolvidas no processo de seleção fenotípica desse tipo de caráter (Ferreira & Gratapaglia, 1998). Os marcadores moleculares podem ser utilizados como uma ferramenta para auxiliar no processo seletivo.

O poder de detecção da variabilidade existente diretamente ao nível de DNA foi alcançado com o desenvolvimento de técnicas em biologia molecular. Marcadores moleculares de DNA têm sido empregados na cultura do feijoeiro para estimar os níveis de diversidade e heterozigosidade entre diferentes genótipos, para mapeamento genético e de “QTLs”, e para realização de estudos de sintenia (Blair *et al.*, 2003; 2006a; Kwak *et al.*, 2008; Hougaard *et al.*, 2008), entre outros. O avanço do melhoramento genético pode ser assessorado pelo desenvolvimento de

novos marcadores, que constituem ferramentas para a execução de programas de seleção assistida (Oliveira *et al.*, 2008).

Os mapas genéticos são ferramentas que facilitam a seleção de genes de interesse agrônômico, pois através deles é possível acompanhar, pelo uso de marcadores, regiões do genoma que são selecionadas a cada geração melhorada. Através da utilização de um grande número de marcadores moleculares e de métodos estatísticos, é possível individualizar o efeito fenotípico de locos que contribuem para um QTL (*Quantitative Trait Loci*).

Uma das técnicas mais indicadas para estudar polimorfismos entre sequências de DNA é através dos microsatélites ou SSR (*Simple Sequence Repeats*, Tautz & Renz, 1984). Essa técnica baseia-se no uso de pares de iniciadores numa reação de PCR para detectar variações em locos de sequências repetitivas. A técnica de SSR revela polimorfismo em um loco devido a diferenças no número de vezes em que um motivo se repete naquele loco.

O conteúdo genético informativo de um loco SSR é bastante alto, por se tratar de sequências de alta taxa evolutiva. A sua natureza multialélica e codominante permite o estabelecimento da genotipagem de indivíduos dentro de populações ou entre genótipos relacionados, por isso, marcadores microsatélites têm sido usados para identificação individual, análise de diversidade (Powell *et al.*, 1996a) e estudos de evolução e de estrutura de populações de espécies relacionadas. Além destas aplicações, os marcadores microsatélites são amplamente utilizados no mapeamento genético e na análise de QTLs.

No início deste estudo existiam poucos locos microsatélites disponíveis na literatura. Como o trabalho de desenvolvimento de um mapa genético requer uma grande quantidade de marcadores, foi necessário centralizar esforços na busca por

novos locos. Neste contexto, o trabalho atual teve como objetivo central desenvolver marcadores microssatélites em grande número, permitindo assim a construção de um mapa genético molecular suficientemente saturado para assessorar os programas de melhoramento do feijoeiro.

2. Revisão Bibliográfica

2.1. A Cultura do Feijoeiro

O feijão pertence à ordem Rosales e a família Fabaceae (Leguminosae). A família das leguminosas representa a terceira maior família de plantas superiores e inclui mais de 19000 espécies (Lewis *et al.*, 2005). As plantas pertencentes a esta família são diversas quanto ao hábito de crescimento e ao potencial econômico explorado: existem árvores, arbustos, herbáceas anuais, ornamentais e culturas de importância agrônômica. As leguminosas têm um papel fundamental em ecossistemas naturais e na agricultura devido à capacidade de estabelecer simbiose com bactérias fixadoras de nitrogênio, tornando-as eficientes colonizadoras de ambientes pobres neste nutriente. Além disso, as leguminosas fornecem uma dieta rica em proteínas.

Duas espécies de feijão são muito consumidas no Brasil: *Phaseolus vulgaris*, o feijoeiro-comum, e *Vigna unguiculata*, vulgarmente conhecida por feijão de corda, feijão macassa, ou caupi (Vieira, 1988). O gênero *Phaseolus* compreende aproximadamente 55 espécies, das quais, apenas cinco são cultivadas: o feijoeiro comum (*P. vulgaris* L); o feijão de lima (*P. lunatus*); o feijão ayocote (*P. coccineus*); o feijão tepari (*P. acutifolius*); e o *P. polyanthus*. Existem mais de 29000 acessos de feijões domesticados e 1300 acessos selvagens em bancos de germoplasma (Broughton *et al.*, 2003).

O feijoeiro comum é uma leguminosa anual domesticada há mais de 10000 anos. A origem e a domesticação da cultura podem ser explicadas por Gepts (1986) através da análise de diferentes padrões eletroforéticos de faseolina, a proteína de

reserva das sementes. O autor sugere que existam três centros primários de diversidade genética: um Mesoamericano e dois Andinos (Norte e Sul), como mostra a Figura 1. O grupo Mesoamericano possui sementes pequenas e faseolina do tipo S. Os Andinos possuem sementes maiores e com faseolina tipos T, C, H e A. Dados de isoenzimas (Koenig & Gepts, 1989; Singh *et al.*, 1991a), DNA mitocondrial e RFLPs também diferenciam os dois *pools* gênicos, Andinos e Mesoamericanos, tanto em linhagens cultivadas como em selvagens (Becerra-Velazquez & Gepts, 1994; Khairallah *et al.*, 1992).

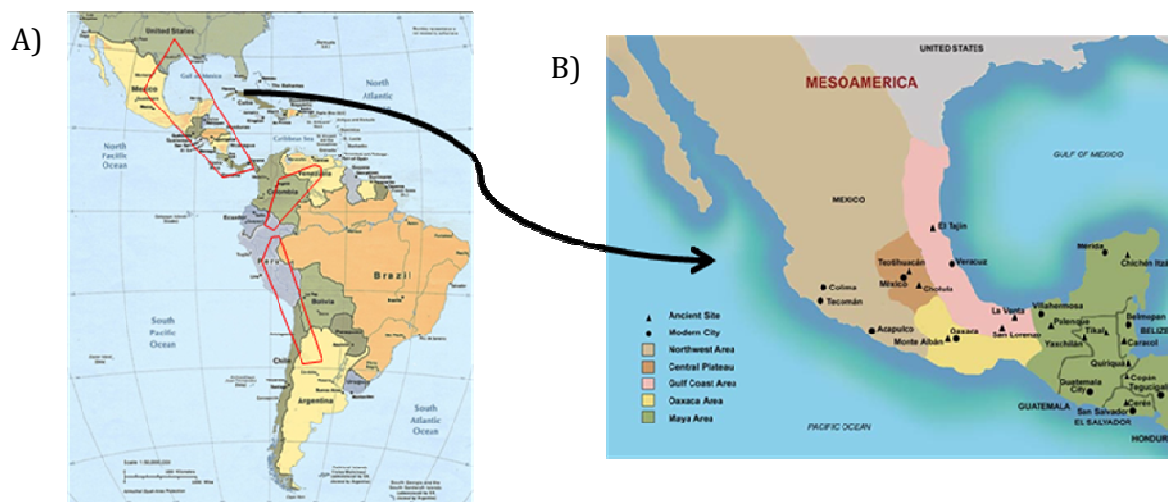


Figura 1. A) Localização dos 3 centros de origem de *Phaseolus vulgaris*: região Andina Norte e Sul e, região da Mesoamérica com destaque em B) para os países que formam a região.

Estima-se que a diversidade genética do feijoeiro cultivado seja menor em relação ao selvagem em função de um efeito de gargalo e do efeito fundador que ocorreu durante a domesticação da cultura, conhecida como síndrome da domesticação (Gepts & Bliss, 1986; Koinange *et al.*, 1996; Chacón *et al.*, 2005). As

diferenças fenotípicas entre cultivares selvagens e domesticados são resultados da seleção durante o período de domesticação para adaptação aos ambientes de cultivo. As diferenças ocorreram em características que facilitaram o manejo no cultivo, tais como na dormência de sementes, mecanismos de dispersão das sementes, hábito de crescimento e sensibilidade ao fotoperíodo. Em feijoeiro, os dois atributos mais importantes da síndrome da domesticação são: 1) a perda da capacidade de dispersão das sementes, condicionada pela presença de fibras nas vagens e, 2) a dormência nas sementes, condicionada pela impermeabilidade do tegumento da semente (Koinange *et al.*, 1996).

Segundo análises de polimorfismo de DNA de cloroplasto, há evidências de que houve um único evento de domesticação na região Andina, enquanto que ocorreram múltiplas domesticações e eventos de introgressão na Mesoamérica, resultantes de cruzamentos selvagens X cultivados (Chacón *et al.*, 2005). Genótipos selvagens podem ser usados como fonte de aumento de variabilidade para o feijoeiro cultivado, principalmente para o *pool* Andino, que apresenta baixa diversidade. Estima-se que menos de 5% da diversidade genética disponível tem sido usada globalmente (Broughton *et al.*, 2003).

O conceito de variabilidade genética é a base de toda capacidade de se promover seleção e, conseqüentemente do melhoramento genético. À medida que se processa um programa de seleção e este se perpetua por muitas gerações, há uma redução no nível de variabilidade genética (Kepler, 2000). Um importante pré-requisito para o uso de recursos genéticos é o amplo conhecimento da extensão e

distribuição da variabilidade genética das espécies cultivada e de seus parentes selvagens.

O feijoeiro representa a cultura mais difundida no mundo, principalmente na América Latina, África e Ásia, regiões onde consiste numa das principais fontes protéicas para cerca de 500 milhões de pessoas. Em muitos destes países, a proteína de origem vegetal chega a mais de 80% do total de proteínas da dieta humana (Brown *et al.*, 1990). O feijão é também fonte de calorias, fibras, minerais (Ca, Cu, Fe, Mg, Mn, Zn) e vitaminas. Trata-se da terceira leguminosa mais consumida no mundo, sendo superada apenas pela soja e o amendoim (Singh, 1999). Ele complementa uma dieta baseada em cereais e outros alimentos ricos em carboidratos, por fornecer muitos nutrientes. Além disso, o consumo regular de feijões ajuda a reduzir o colesterol e o risco de câncer.

Neste contexto, o Brasil é o maior produtor mundial do gênero *Phaseolus* (Vieira *et al.*, 1988) e o segundo maior produtor de feijão comum, mesmo apresentando produtividade média baixa, equivalente a 600 kg/ha (Silva, 2000). Essa baixa produtividade ocorre devido a vários fatores, dentre os quais: a falta de nutrientes no solo, em virtude de adubações inadequadas ou não realizadas; predominância de cultivos associados com outras culturas, sobretudo milho; ataque de pragas e incidência de doenças (Vieira, 1988). Existem mais de 200 doenças que atacam essa cultura (Bianchini *et al.*, 1989), causadas por fungos, vírus, bactérias e nematóides, limitando sua produção e reduzindo a qualidade fisiológica, sanitária, nutricional e comercial do produto.

Além dos principais motivos da baixa produtividade descritos anteriormente, destacam-se outras duas causas: as variedades plantadas e as sementes utilizadas. No Brasil, semeia-se grande número de variedades de feijão, das quais muitas nunca foram estudadas. Muitas dessas variedades apresentam menor capacidade de produção e, são susceptíveis a uma ou mais enfermidades. As sementes plantadas são geralmente produzidas pelo próprio agricultor e, por isso, nem sempre são de boa qualidade.

Doenças como a antracnose, mofo branco, crestamento bacteriano, mancha-de-alternaria e, mancha angular podem ser transmitidas pelas sementes, fato esse desconhecido por muitos produtores de feijão. Os pequenos agricultores são os principais produtores da cultura, e o que se observa usualmente são baixos níveis de tecnologia e insumos empregados e mão de obra familiar (Vieira, 1988).

No Brasil, o cultivo e colheita do feijoeiro se concentram em três safras tradicionais, segundo o zoneamento ecológico (época das águas, da seca e de inverno). Assim, outra causa provável das baixas produções brasileiras é o regime pluvial desfavorável. A maioria dos agricultores não realiza a irrigação da cultura. Dessa forma, as chuvas devido à escassez, má distribuição ou mesmo excesso, têm sido um motivo das baixas produções em algumas safras no Brasil (Vieira, 1988).

A partir do começo da década de 1970, a pesquisa de feijão no Brasil passou a receber mais atenção, com aumento dos recursos disponíveis e dos pesquisadores envolvidos. Depois das chamadas “crises de feijão”, em que por motivos climáticos houve escassez de feijão no mercado, fazendo com que produto atingisse altos preços

e o governo se viu obrigado a importá-lo. A partir deste momento, as autoridades concluíram que se tornara urgente a realização de investimentos em pesquisas, de forma a evitar futuras crises. O feijão havia então deixado de ser um produto abundante e consumido por todas as classes sociais (Vieira, 2004).

Em meio às atuais tendências de aumento da população e do consumo de feijão, pode ser esperado um aumento da demanda deste produto a níveis sem precedentes. Este aumento no consumo somente será suprido se novos cultivares, com níveis mais altos de produtividade, de resistência à doenças, tolerância à seca e com bons rendimentos em solos inférteis forem desenvolvidos (Guidolin, 2003). Além disso, o aumento da demanda por produtos alimentares livres de pesticidas e a necessidade de reduzir os custos de produção oferecem um grande desafio à comunidade de pesquisadores e aos produtores de feijão.

O melhoramento genético é fundamental no desenvolvimento da agricultura, gerando novos cultivares e variedades de interesse agrônômico. O aumento na eficiência de seleção, o melhor conhecimento e caracterização do germoplasma e a maximização dos ganhos genéticos têm sido objetivos de melhoristas de plantas do mundo inteiro. Com a cultura do feijoeiro não pode ser diferente. É notável a necessidade de se aprimorar estudos com a cultura e que haja interesse e cooperação entre os diferentes grupos de pesquisa para estudos genéticos que auxiliem o melhoramento e, que estes sejam realizados em maior número.

2.2 Evolução de Marcadores

Os primeiros marcadores utilizados em estudos genéticos foram os fenotípicos, os quais eram representados por caracteres morfológicos controlados por genes com segregação Mendeliana. Estes caracteres são de fácil identificação visual, como por exemplo, a deficiência clorofítica, nanismo, cor de pétalas ou morfologia foliar (Ferreira & Grattapaglia, 1995). Estes marcadores contribuíram significativamente para o desenvolvimento teórico da análise de ligação gênica e para a construção das primeiras versões de mapas genéticos. Os caracteres são de avaliação muito simples e de baixo custo. No entanto, esses marcadores existem em número limitado no genoma, e sua expressão gênica pode estar sujeita às variações do ambiente e à ação de outros genes (Borém, 1998).

Na década de 60, surgiram os marcadores de polimorfismos bioquímicos, denominados de isoenzimas. Esses marcadores são formas moleculares de uma mesma enzima que, apesar de desempenharem as mesmas funções catalíticas, podem apresentar propriedades cinéticas distintas, diferenciadas via eletroforese (Ferreira & Grattapaglia, 1995). As isoenzimas apresentam níveis de polimorfismos consideravelmente baixos, sendo pouco abundantes, o que reduz significativamente o potencial de utilização desses marcadores (Murphy *et al.*, 1990). Além disto, as isoenzimas podem estar sujeitas à influência do ambiente (Richards, 1997).

No início dos anos 80 descobriu-se a primeira classe de marcadores de polimorfismo ao nível do DNA, o RFLP (*Restriction Fragment Length Polymorphism*, Botstein *et al.*, 1980). Esse marcador se baseia na variação do tamanho de fragmentos de DNA obtidos com o uso de enzimas de restrição. O polimorfismo está baseado nos

diferentes tamanhos de fragmentos obtidos em função da variação na distribuição dos sítios de restrição. Tal variação é resultante de mutações pontuais que levam à criação ou eliminação de sítios de clivagem para uma mesma enzima, ao longo do genoma. Além disso, podem ocorrer inserções e deleções de bases, bem como translocações ou inversões das sequências de DNA em um dado fragmento delimitado por dois sítios de restrição. Essas variações alteram o tamanho do fragmento sem que tenha ocorrido modificação do sítio de restrição da enzima que delimita o fragmento (Botstein *et al.*, 1980). Os fragmentos produzidos por digestão com enzimas específicas são separados através de eletroforese em gel de agarose e visualizados pelo processo de *Southern Blot* (Southern, 1975). O marcador RFLP é considerado robusto, isto é, altamente reprodutível e estável, porém, é o mais caro e trabalhoso entre aqueles atualmente disponíveis devido à impossibilidade de automatização completa da técnica empregada (Garcia *et al.*, 2004).

Outro marcador amplamente utilizado na década de 90 é o RAPD (*Random Amplified Polymorphic DNA*, Williams *et al.*, 1990), o qual faz uso de oligonucleotídeos com sequência arbitrária de bases para a amplificação de regiões aleatórias do DNA genômico. Apesar dos marcadores RAPDs apresentarem a vantagem da rapidez na análise de um grande número de genótipos, existem grandes limitações, como a baixa reprodutibilidade dos resultados e a natureza dominante do marcador, não permitindo a distinção entre os locos heterozigotos e homozigotos dominantes. Tais fatos, limitam seu uso na estimação de parâmetros genéticos e no mapeamento genético molecular. Atualmente este marcador tem sido cada vez mais preterido frente a outros existentes devido às essas limitações (Garcia *et al.*, 2004).

Os marcadores AFLP (*Amplified Fragment Length Polymorphism*) são obtidos pela combinação da distribuição de sítios de restrição e amplificação aleatória de fragmentos de DNA. A técnica consiste em digerir o DNA com enzimas de restrição, ligar adaptadores aos fragmentos resultantes e re-amplificação de um conjunto selecionado de fragmentos digeridos e amplificados. A quantidade de informação gerada é bastante grande, possibilitando caracterizar dezenas de genótipos em poucas análises. Entretanto, o marcador AFLP apresenta as deficiências de um marcador dominante e o fato da técnica ser cara e muito sofisticada para grande parte dos laboratórios que trabalham com genotipagem de plantas (Zabeau & Vos, 1993).

O polimorfismo de nucleotídeos únicos ou SNPs (*Single Nucleotide Polymorphisms*) são variações pontuais no DNA primeiramente observadas com o sequenciamento do genoma humano (Brookes, 1999). Este marcador é bi-alélico e, corresponde a uma posição em que existe uma alternância dos nucleotídeos A, T, G, C. Os SNPs ocorrem em regiões codificantes e não codificantes nos genomas. Em regiões codificadoras, quando resultam em uma substituição de aminoácido na sequência protéica são denominados não sinônimos, podendo a substituição ser conservativa ou não conservativa em função dos aminoácidos envolvidos na troca. Nesses casos podem ocorrer modificações estruturais e funcionais na proteína. Quando essas substituições não alteram o aminoácido formado, elas são ditas sinônimas. Embora SNPs sinônimos não alterem a sequência protéica, eles podem modificar a estrutura e a estabilidade do RNA mensageiro (Kwok, 1999). Os SNPs podem ser identificados através de sequenciamento direto do DNA, de informações disponíveis em bancos de dados, ou através de métodos bioquímicos que investigam a presença de variantes. A

genotipagem não se baseia no comprimento dos alelos, como outros marcadores moleculares utilizados, e a distinção dos alelos podem ser automatizados (Giordano *et al.*, 1999). A grande vantagem dos SNPs reside na abundância de polimorfismos entre alelos de um determinado gene.

Os marcadores moleculares são classificados em dominantes e codominantes. Os marcadores codominantes permitem a visualização simultânea dos diferentes alelos de um mesmo loco, podendo-se identificar três classes genotípicas. Encontram-se nesta classe os RFLPs e os microssatélites. Os marcadores dominantes revelam o polimorfismo em dezenas de locos diferentes simultaneamente, como é o caso para os marcadores RAPDs e AFLPs.

2.2.1 Marcadores Microssatélites

Os marcadores microssatélites ou SSR (*Simple Sequence Repeat*, Tautz, 1989; Weber & May, 1989) são combinações de uma a seis bases repetidas em *tandem*. A natureza de sua variação é o número de repetições das unidades repetidas que o compõem. A técnica é baseada na amplificação das sequências repetidas, utilizando-se oligonucleotídeos iniciadores ancorados nas regiões flangeadoras às repetições. Apesar da variação no número de motivos, a sequência de bases adjacentes pode ser única no genoma e conservada entre indivíduos. Assim, indivíduos que apresentam variações nos motivos repetidos produzem diferentes tamanhos de produtos de amplificação.

Os marcadores microssatélites têm representado um importante avanço para estudos genéticos. Este marcador apresenta alto nível de polimorfismo, codominância, alta heterozigosidade e padrão de herança Mendeliana. Além disso, a técnica para sua obtenção é possível a partir de qualquer tipo de tecido em qualquer estágio de desenvolvimento. São marcadores baseados na amplificação via reação de polimerase em cadeia (*Polimerase Chain Reaction*, Mullis & Faloona, 1987), facilitando a transferência da técnica entre diferentes grupos de pesquisa. Este marcador encontra-se regularmente distribuído no genoma e, quando comparado a outros marcadores como os RFLPs, são menos custosos (Powell *et al.*, 1996a).

A presença dessas características faz com que os microssatélites sejam marcadores ideais para mapeamento genético de genomas, para identificação e discriminação de genótipos e estudos de genética de populações. Em função das vantagens que esse marcador representa, na última década ele tem se tornado o marcador genético mais popular (Schlotterer, 2000).

A hipervariabilidade de microssatélites não é totalmente entendida, podendo ser originada através de *crossing over* desigual ou deslizamento de sequências durante a replicação do DNA. Estes eventos podem resultar em variação do número de cópias da unidade básica de repetição (Schlotterer & Tautz, 1992).

Um fenômeno comum observado na genotipagem de locos microssatélites é a ocorrência de bandas fantasmas. Este padrão de dupla ou tripla banda pode ser derivado do deslizamento da *Taq*-polimerase durante a reação de amplificação (Levinson & Gutman, 1987; Schlotterer & Tautz, 1992).

A detecção dos microssatélites amplificados pode ser feita em gel de poliacrilamida ou em agarose de alta resolução e separadas por eletroforese. É necessária a utilização de géis adequados à separação eficiente dos alelos, pois esses podem diferir em apenas uma base. A visualização das bandas pode ser feita por coloração com brometo de etídeo, tratamento com nitrato de prata, auto-radiografia, ou pela utilização de fluoróforos.

O sistema fluorescente permite a precisa detecção dos alelos através de um sistema automatizado, um aspecto muito importante em análises de populações para evitar erros de genotipagem. Neste método, um dos iniciadores usados na amplificação tem a extremidade 5' marcada com fluorocromo. Entretanto, o custo da síntese de iniciadores fluorescentes é um fator limitante em muitos laboratórios, pois varia de 5 a 10 vezes mais que um *primer* comum. Os locos podem ser analisados individualmente ou em conjunto com outros SSRs de tamanhos diferentes, e neste caso, utiliza-se mais de um par de iniciadores na reação de PCR (Ferreira & Grattapaglia, 1995).

Na tentativa de reduzir os custos da análise por fluorescência, Oetting *et al.* (1995) propôs a estratégia via PCR de *multiplex* com 3 componentes: 1)iniciadores *forward* possuindo uma extensão de 19-pb na extremidade 5' com a sequência de um iniciador M13; 2)iniciador *reverse* sem modificações, e 3)iniciador M13 marcado com fluorescência. Os iniciadores estendidos com a cauda M13 contêm a sequência complementar do M13 marcado, gerando produtos fluorescentes que podem ser detectados em um sequenciador automático. A grande vantagem consiste em sintetizar apenas um iniciador M13 fluorescente universal, enquanto que pelo método

convencional para cada microsatélite a ser analisado seria necessário sintetizar um específico iniciador marcado.

Os marcadores microsatélites têm sido usados para a análise genética em espécies de mamíferos (Moore *et al.*, 1998), insetos (Harper *et al.*, 2003), aves (Primmer *et al.*, 1997), peixes (Rico, *et al.*, 1993), e plantas (Condit e Hubbell, 1991; Zhao e Kochert, 1992; 1993; Yang *et al.*, 1994; Yu *et al.*, 2000).

Os microsatélites podem ser obtidos convencionalmente através do isolamento e seqüenciamento de clones contendo possíveis motivos SSRs, seguidos de desenho de iniciadores para as regiões flanqueadoras aos motivos. Buscando a redução do custo e do tempo gasto, foram desenvolvidos processos de clonagem para a criação de bibliotecas enriquecidas em microsatélites (Kijas *et al.*, 1994; Billote *et al.*, 1999). A utilização da biblioteca enriquecida para desenvolvimento de microsatélites tem se mostrado mais eficiente e rápida do que a utilização da biblioteca genômica (Refseth *et al.*, 1997). O sequenciamento de ESTs (*Expressed Sequenced Tags*) permite que microsatélites sejam desenvolvidos a partir de sequências depositadas em bancos de dados públicos.

A maior limitação dos microsatélites é que eles precisam ser isolados para a maioria das espécies, uma vez que a transferência de iniciadores de uma espécie para outra nem sempre resulta em produtos de amplificação. Desta forma, quanto maior o número de microsatélites disponíveis para feijoeiro, mais saturado com marcadores poderá ser o mapa genético molecular da espécie. Uma vez desenvolvidos, os microsatélites são marcadores estáveis e de fácil manuseio.

2.2.2. Microsatélites na Cultura do Feijoeiro

O desenvolvimento de microsatélites em feijão iniciou-se recentemente (Yu *et al.*, 1999; Gaitán-Solís, *et al.*, 2002; Métais *et al.*, 2002; Blair *et al.*, 2003; Guerra-Sanz, 2004; Buso *et al.*, 2006). Yu *et al.* (1999) obtiveram 49 microsatélites para feijão comum a partir de pesquisa em bancos de sequências. Em outro estudo em banco de dados, foram desenvolvidos 18 microsatélites, e a avaliação dos locos foi feita num conjunto de acessos de *P. vulgaris*, além de dois acessos de *P. coccineus* (Guerra-Sanz, 2004). Nos estudos de Gaitán-Solís *et al.* (2002) foram isolados 68 microsatélites de sequências de origem genômica e testados em genótipos Andinos, Mesoamericanos e em outras espécies do gênero *Phaseolus*.

Dessa forma, em 2002, quando o grupo da Prof. Anete P. de Souza iniciou o desenvolvimento de marcadores microsatélites para feijoeiro, havia apenas 117 microsatélites disponíveis para a espécie. Esse número era muito pequeno para possibilitar a construção de um mapa genético com saturação média ou suficiente para o mapeamento preciso de características quantitativas de interesse econômico.

Em 2003, Yaish & de la Vega descreveram o isolamento de microsatélites (GA)_n utilizando a metodologia de bibliotecas enriquecidas e identificaram microsatélites ligados ao gene putativo MADS-box, cuja família gênica fornece recursos para estudos evolutivos em plantas (Winter *et al.* 1999; Theissen *et al.*, 2000). Bancos enriquecidos também foram utilizados por Buso *et al.* (2006), que

desenvolveram 20 marcadores microssatélites e 10 apresentaram padrão adequado para serem caracterizados em 85 genótipos.

O desenvolvimento de microssatélites a partir de BACs (*Bacterial Artificial Chromosome*) foi descrito por Caixeta *et al.* (2005). Neste trabalho, foram desenvolvidos 21 microssatélites a partir de 4 clones de BACs que foram subclonados, hibridizados e sequenciados. Foram geradas 62 sequências e 27 (43,6%) apresentaram motivos repetidos. Não houve relação entre a diferença no tamanho dos alelos com múltiplos do número de nucleotídeos que formavam a repetição.

Na cultura do feijoeiro já existem estudos relatando a geração e análise de sequências gênicas a partir de ESTs (*Expressed Sequence Tag*, Melotto *et al.*, 2005; Ramirez *et al.*, 2005; Hanai *et al.*, 2007). No trabalho de Melotto *et al.* (2005), as bibliotecas foram construídas com base na expressão diferencial dos genes ligados a resistência ao fungo *Colletotrichum lindemuthianum*, causador da antracnose. O acesso às sequências depositadas abre a perspectiva de sua utilização para localização de microssatélites gênicos, podendo reduzir custos com geração de novas bibliotecas enriquecidas. Estudos têm mostrado que os microssatélites de origem gênica se apresentam menos polimórficos que os de origem genômica (Blair *et al.*, 2003). Desta forma, uma biblioteca genômica enriquecida em microssatélites tenderia a possibilitar um acesso melhor ao polimorfismo do que microssatélites derivados de EST. Entretanto, não há até o momento evidências de alto conteúdo de polimorfismo associado aos locos de microssatélites de feijoeiro quer sejam eles gênicos ou genômicos. O baixo conteúdo de polimorfismo revelado em um conjunto de microssatélites poderia ser explicado por um conjunto de fatores: base genética

estreita para acessos cultivados, conforme relatado por Singh (1991b), tipo de reprodução, e pequeno tamanho dos motivos repetidos possivelmente insuficiente para acessar o polimorfismo dentro da espécie.

As repetições mais frequentes em plantas são motivos de AT (Powell *et al.*, 1996b). Assim, para analisar a informatividade deste tipo de repetição, microssatélites de repetições AT foram desenvolvidos recentemente para feijoeiro, e comparados quanto ao polimorfismo com outras repetições (Blair *et al.*, 2008). Os locos de dinucleotídeos AT apresentaram mais problemas de “bandas fantasmas”, e talvez sejam necessários ajustes de amplificação em outros laboratórios que venham a utilizar estes locos. Porém, os locos derivados de dinucleotídeos AT foram mais polimórficos que repetições GA e CA comparadas no estudo, o que demonstra o potencial do novo enriquecimento.

Além do desenvolvimento e validação dos marcadores microssatélites, a aplicação destes marcadores para estudos genéticos da cultura têm aumentado. Um estudo com microssatélites demonstrou recentemente declínio na frequência de alguns alelos em cultivares de feijoeiro cultivados nos dois últimos séculos (Maras *et al.*; 2006). Martinez-Castillo *et al.* (2007) encontraram que o fluxo gênico entre plantas domesticadas de *Phaseolus lunatus* para as populações selvagens é três vezes maior em relação ao sentido inverso. O fluxo gênico assimétrico encontrado pode criar uma redução drástica na diversidade genética de populações selvagens e até mesmo levar a extinção de populações locais. Consequentemente pode ocorrer a diminuição da diversidade genética do pool gênico de plantas domesticadas, reduzindo assim a disponibilidade de genes de interesse agrônômico.

Blair *et al* (2006a) utilizaram dois tipos de microssatélites, gênicos e genômicos. Foi observado que ambos os tipos foram consistentes para a separação dos genótipos dos centros de origem Andinos dos Mesoamericanos. Foi detectada menor diversidade genética no *pool* Mesoamericano, quando comparado ao Andino. Neste estudo, os microssatélites de origem gênica apresentaram-se menos polimórficos em número de alelos e no conteúdo de polimorfismo informativo (PIC).

Nos estudos de Zhang *et al.* (2008) foram avaliados acessos de feijoeiros da China para estimar a variabilidade genética e identidade do *pool* gênico. Esse estudo discute a possível existência de um centro secundário de diversidade para o feijoeiro na China.

O uso de ferramentas moleculares está se tornando mais efetiva para a análise de características cujo fenótipo é de difícil reconhecimento ou não pode ser acessado diretamente. Tal fato é decorrente da possibilidade de se monitorar diretamente a região genômica que possui o gene responsável pela característica de interesse. Assim, em cruzamentos onde se pretende transferir determinadas características de um genótipo a outro, é possível monitorar a introgressão do gene responsável pela característica desejada através de marcadores moleculares. Esse procedimento facilita a identificação dos indivíduos contendo o gene desejado, principalmente em situações onde o fenótipo da característica é por identificação indireta, como por exemplo, a análise de aminoácidos, enzimas e outros componentes celulares.

A utilização de microssatélites para auxiliar a recuperação do genótipo elite em etapas de retrocruzamento assistido já foi realizada com sucesso em feijoeiro (Oliveira *et al.*, 2008). Os resultados demonstraram a aplicação da tecnologia de

marcadores moleculares microssatélites associados ao melhoramento do feijoeiro para reduzir o número de retrocruzamentos necessários e consequentemente, o tempo para recuperação do genoma dos genitores recorrentes contendo as características de interesse.

O desenvolvimento e disponibilização de marcadores microssatélites para feijoeiro no momento, é um esforço conjunto da comunidade nacional e internacional, evidenciando assim, a importância da aplicação desta ferramenta para auxiliar o melhoramento do feijoeiro.

2.3. Mapas Genéticos

A disponibilidade de marcadores, aliada a métodos estatísticos complexos, tem permitido a construção de mapas genéticos para a maioria das espécies vegetais de interesse econômico, até mesmo para aquelas de longo ciclo de vida, como as florestais e as frutíferas (Carneiro e Vieira, 2002).

O objetivo principal de um trabalho de mapeamento é ordenar o genoma, de maneira regular, visando à localização de genes de interesse e à identificação de marcadores fortemente ligados a estes genes. O mapeamento visa também estimar a distância genética entre marcadores consecutivos. Mapas de ligação saturados são de extrema importância para assessorar os programas de melhoramento das diferentes espécies. Além disso, é possível maximizar os ganhos de produção de uma dada cultura através de programas de introgressão de genes responsáveis por características de interesse em cultivares comerciais e selecionar as plantas contendo o gene por seleção assistida por marcadores.

A seleção assistida por marcadores (SAM) fundamenta-se no conceito de que é possível inferir a presença de um gene a partir de um marcador fortemente ligado a este. Quando o marcador encontra-se distante da região de interesse, a possibilidade de ambos serem transmitidos aos indivíduos da progênie é reduzida devido aos eventos de recombinação. Assim, a existência de uma forte ligação entre a característica de interesse e o marcador é pré-requisito neste tipo de seleção (Kumar *et al.*, 1999). Com o uso de marcadores moleculares, a seleção assistida e a piramidação de genes (introgressão de vários genes que contribuem para a melhoria de uma mesma característica, por exemplo, resistência a uma doença específica) de interesse agrônômico pode ser realizada.

Estudos evolutivos também podem utilizar informações de mapas genéticos, pois os mapas fornecem informação sobre a estrutura, função e evolução dos genomas. É possível fazer a comparação de mapas entre diferentes espécies ou populações em diferentes níveis. Assim, pode-se efetuar a comparação quanto à homologia de genes, conservação da distância entre eles e a ordem de ligação dos genes nos cromossomos (Weeden *et al.*, 1992; Moore *et al.*, 1995; Hougaard *et al.*, 2008).

Os marcadores mais adequados para um trabalho de mapeamento genético apresentam como características, herança codominante, alto nível de polimorfismo, grande quantidade e distribuição uniforme pelo genoma. No mapeamento de características fenotípicas utilizando marcadores moleculares, avalia-se a segregação conjunta dos diferentes fenótipos associados a características de interesse, com os alelos e locos evidenciados pelos marcadores. A observação de co-segregação entre

alelo e fenótipos indica ligação física entre diferentes locos envolvidos, ou apenas associados a genes que controlam a característica. Se o marcador e genes responsáveis pela característica fenotípica segregarem sempre juntos, sugere-se que estejam ligados, ou seja, estão fisicamente próximos; se por outro lado, a segregação for independente, conclui-se que o marcador e o gene não estão ligados e, portanto, estão genética e possivelmente fisicamente distantes entre si (Schuster & Cruz, 2004).

Um dos fatores de fundamental importância para se obter dados consistentes em um trabalho de mapeamento é o tamanho da população estudada. A precisão na determinação das distâncias entre os locos no mapa e a capacidade de se determinar a sequência de marcadores nele estão diretamente relacionadas ao tamanho amostral (população que esta sendo genotipada). É fundamental que se estimem adequadamente distâncias entre os genes, estabelecendo-se o ordenamento correto e formando-se grupos de ligação que reflitam o número básico de cromossomos da espécie. Como os eventos de permutação ocorrem ao acaso ao longo do cromossomo, a probabilidade de recombinação é maior para locos que se encontram mais distantes entre si. Assim, quanto mais próximos estão dois genes, menor a frequência de permuta e, conseqüentemente, menor a distancia em cM.

Mapas genéticos são desenvolvidos com base em populações segregantes com o máximo de desequilíbrio de ligação para a obtenção das estimativas de recombinação. Uma população está em equilíbrio de ligação quando seus gametas são formados por combinações aleatórias dos alelos que constituem tal população. As frequências gaméticas são resultado do produto das frequências alélicas e, neste caso, elas são iguais. Quando a combinação dos alelos de diferentes locos não é aleatória, as

frequências gaméticas apresentam desvios em relação ao produto das frequências alélicas, e assim, a população está em desequilíbrio de ligação. O desequilíbrio de ligação pode ser decorrente de diversas causas de alteração na panmixia, como seleção e deriva genética, mas a principal destas é a ligação física dos locos. O desequilíbrio de ligação devido à ligação física dos locos é bastante elevado em populações de cruzamentos controlados; conseqüentemente, a capacidade de se detectar a ligação entre os dois locos fisicamente ligados também é elevada (Tanksley, 1993; Falconer & Mackay, 1996; Lynch & Walsh, 1998).

Com os dados de populações segregantes gerados por delineamentos específicos (F_2 , retrocruzamentos, linhagens recombinantes) e com indivíduos analisados por meio de marcadores moleculares, é possível a construção do mapa genético de ligação. Para se construir um mapa é necessário: (a) teste de segregação (teste do Qui-quadrado) para verificar o padrão de segregação Mendeliana de cada loco independentemente; (b) teste de ligação, para verificar o desequilíbrio de ligação entre os pares de marcadores que apresentaram segregação Mendeliana, formando um conjunto de marcadores para cada grupo de ligação e (c) ordenação dos locos a partir das distâncias entre eles, estimadas a partir das frequências de recombinação entre as marcas. Para formar os grupos de ligação é necessário testar estatisticamente se o valor da frequência de recombinação difere de 0,5 através do uso do teste de razão de verossimilhança (LR), ou outros testes equivalentes como o índice conhecido como *LOD-score* (*Logarithm of Odds*, Lynch & Walsh, 1998).

2.3.1. Teste de Segregação

A condução da população de mapeamento delineia diferentes tipos de segregação esperados. Mapas genéticos podem ser construídos empregando-se populações tipo F_2 , e, a proporção esperada dos alelos de uma dado loco, representado pelas diferentes bandas reveladas por um marcador é de 1:2:1; no caso de RILs (*Recombinant Inbred Lines*) avançadas, a proporção esperada entre os alelos de um loco é 1:1. A forma mais comum para avaliação de hipóteses genéticas é através do teste de Qui-quadrado, pois além de considerar desvios ocorridos entre valores previstos e observados, também é sensível ao tamanho da amostra em que são realizadas as mensurações:

$$\chi^2 = \frac{(fo - fe)^2}{fo}, \text{ onde:}$$

fo = frequência observada em cada uma das classes; fe = frequência esperada em cada uma das classes (proporção Mendeliana do modelo genético adotado).

Os valores obtidos do Qui-quadrado são comparados com valores tabelados para conclusão sobre a significância desse valor. Um aspecto que deve ser considerado é o nível de significância adotado. Por isso, a correção de Bonferroni (Lynch & Walsh, 1998), de acordo com a fórmula abaixo, é utilizada para resolver situações de erro tipo I, onde são descartadas hipóteses incorretamente:

$$\alpha = \frac{\gamma}{n}, \text{ em que:}$$

α = nível de significância; γ = probabilidade de se encontrar pelo menos um falso positivo (nível de significância global); n = número de testes independentes, correspondente ao número de marcadores utilizados. Assim, o nível de significância

para cada marcador pode ser obtido dividindo-se o nível de significância global pelo número de marcadores empregados.

2.3.2. Teste de Ligação

A frequência de recombinação (θ) traduz o número de permutações produzidas entre dois locos durante a meiose. A frequência de recombinantes esperada entre diferentes locos que não estão ligados é de 50% da progênie, devido à segregação independente. Porém, quando há ligação entre os locos, a frequência de recombinantes varia entre 0 e 50% (Griffiths *et al.*, 2000).

São feitos testes considerando-se duas marcas a cada vez, com H_0 considerando a frequência de recombinação igual a 0,5. O teste é feito para se verificar se o estimador θ é significativamente diferente de 0,5, e utiliza-se usualmente, o *LOD-score*.

Tal método consiste em calcular a probabilidade de se obter um conjunto de dados numa situação de segregação independente e, alternativamente, sob a ocorrência de ligação. Em seguida, calcula-se a razão entre as probabilidades dessas duas situações e, o logaritmo na base 10 dessa razão. Consideram-se locos ligados aqueles que apresentam valores de LOD superiores a 3, isto é, a probabilidade de que estejam ligados é 1000 vezes maior que a probabilidade dos locos não estarem ligados.

A frequência de recombinação é transformada em distância genética por funções de mapeamento. As duas funções de mapeamento genéticas mais utilizadas são as funções de Haldane (1919) e Kosambi (1944). Enquanto Haldane (1919) supõe

ausência de interferência, a segunda supõe interferência moderada. A interferência é causada pela possibilidade de ocorrência de mais de uma permutação em regiões adjacentes. A ocorrência de uma permuta afeta a ocorrência de outras em regiões adjacentes (Lynch & Walsh, 1998; Griffiths *et al.*, 2000). A relação entre a frequência de recombinação (distância genética) e o número de nucleotídeos entre dois marcadores (distância física) não é universal, podendo variar entre organismos e entre regiões do genoma.

Com o aumento do número de marcas utilizadas na construção dos mapas e aumento do número de indivíduos analisados, houve o desenvolvimento de algoritmos e programas computacionais para a construção de mapas genéticos, como o Mapmaker/EXP (Lander *et al.*, 1987; Lincoln *et al.*, 1992), JoinMap (Stam, 1993), e o OneMap (Margarido *et al.*, 2007).

2.3.3. Ordenação de Marcadores

Os marcadores são ordenados dentro de cada grupo de ligação pelo princípio da máxima verossimilhança. A melhor ordem de ligação das marcas de um conjunto de dados é a ordem que fornece a máxima verossimilhança para tal conjunto. Todas as ordens possíveis são testadas para um conjunto de marcas, e calcula-se a máxima verossimilhança para cada uma das ordens. O número de ordens possíveis é dado por $n!/2$, em que n é o número de marcas.

2.3.4. Mapas Genéticos em Feijoeiro

O uso de marcadores moleculares para acessar o perfil genético em feijoeiro tem sido amplamente empregado: para investigar a origem e a diversidade de cultivares (Gepts, 1998), a domesticação de *Phaseolus vulgaris* (Koinange *et al.*, 1996), o controle genético de importantes doenças (Nodari *et al.*, 1993; Blair *et al.*, 2006b), a construção e integração dos diferentes mapas já construídos para feijoeiro também para o mapeamento de QTLs (Blair *et al.* 2006b).

O feijoeiro cultivado apresenta uma base genética estreita devido ao processo de domesticação, como discutido anteriormente (Singh *et al.*, 1991a). Consequentemente há a necessidade ter disponível um grande número de marcadores para poder conduzir adequadamente um programa de mapeamento, pois muitos marcadores não apresentarão polimorfismo entre os genitores da população segregante e consequentemente, não serão úteis ao mapeamento. O número de marcadores polimórficos é geralmente muito baixo quando os genitores pertencem ao mesmo centro de origem.

Existem vários mapas genéticos de *Phaseolus* desenvolvidos nos últimos anos (Adam-Blondon *et al.*, 1994; Freire *et al.*, 1998.; Nodari *et al.*, 1993; Yu *et al.*, 2000). Eles podem ser distinguidos por várias características, tais como: genitores utilizados, padrão de segregação para a população estabelecida, características segregantes, tipos de marcadores ancorados, número de marcadores, tamanho (cM) do mapa.

O mapeamento genético do genoma do feijoeiro iniciou-se com o desenvolvimento um mapa de ligação baseado em marcadores morfológicos e isoenzimas (Basset, 1991). Os mapas de Vallejos *et al.* (1992), Gepts *et al.* (1993),

Nodari *et al.* (1993) e Adam-Blondom *et al.* (1994) foram baseados primariamente em marcadores RFLPs. Os estudos de Vallejos *et al.* (1992) e Nodari *et al.* (1993) representaram um grande avanço para o mapeamento do genoma do feijoeiro, pois cobriram 82% e 69% do genoma. Marcadores RAPDs (Bai *et al.*, 1997; Faleiro *et al.*, 2004) e AFLPs (Tar'an *et al.*, 2002) também têm sido utilizados para o desenvolvimento de mapas de ligação em feijoeiro.

O mapeamento comparativo é vantajoso por resultar em mapas de alta densidade e, conseqüentemente, há a possibilidade de que marcadores mais próximos a genes de interesse sejam identificados. A correspondência entre os mapas de ligação do feijoeiro obtidos por Gepts *et al.* (1993), por Freyre *et al.* (1998) e por Adam-Blondom *et al.* (1994) foi estudada por Freyre *et al.* (1998). Os grupos de ligação obtidos com diferentes genótipos representantes dos grupos Andinos e Mesoamericanos foram alinhados, mostrando que não ocorreram grandes rearranjos no genoma do feijoeiro. Desta forma, foi obtido um mapa integrado contendo 470 marcas RFLPs, 570 marcas RAPDs e 40 marcadores bioquímicos e morfológicos, cobrindo 1226 cM, com saturação média inferior a duas marcas por cM. Ainda busca-se a integração dos novos marcadores moleculares a este mapa consenso (Yu *et al.*, 2000). A população base que foi usada para o feijoeiro nesses trabalhos foi obtida a partir do cruzamento entre os genótipos BAT93 x Jalo EEP558 (BJ), uma população avançada até a geração F₁₀ contendo 75 RILs.

Em outro trabalho de mapeamento, Yu *et al.* (2000) testaram 37 microssatélites, dos quais 16 estavam ligados e 15 foram inseridos em grupos de ligação do mapa estabelecido por Freyre *et al.* (1998). Yu *et al.* (2000) observaram que

os locos que continham dinucleotídeos foram mais polimórficos e não encontraram relação entre o comprimento do motivo repetido do microssatélite e o nível do polimorfismo apresentado por ele. Devido à razoável saturação do mapa integrado do feijoeiro, a inclusão dos 15 microssatélites não aumentou de forma significativa a saturação, exceto para o grupo B4 do mapa de Freyre *et al.* (1998), onde havia uma lacuna, na qual foram inseridos três marcadores, melhorando assim a saturação nessa região.

O mapeamento do feijoeiro realizado por Tar'an *et al.* (2002) em 142 indivíduos da população OAC Seaforth e OAC95-4, empregando marcadores RAPD, RFLP, SSR e AFLP, objetivou identificar QTLs para diversos caracteres de importância agrônômica, entre eles o peso de 100 sementes e a produtividade de grãos. No mapeamento foram obtidos doze grupos de ligação cobrindo 1717 cM, com marcadores espaçados de 15 cM em média. Além de mapear QTLs, as informações obtidas foram integradas ao mapa desenvolvido por Freyre *et al.* (1998).

Nos estudos de Blair *et al.* (2003) foram utilizados 150 microssatélites para ancorar ao mapa consenso. Foi possível a localização da posição de 100 locos através do mapeamento comparativo em duas populações. Na população com 87 RILs, oriundas do cruzamento entre DOR364 X G19833, o polimorfismo foi de 56,0% e 78 microssatélites foram posicionados. Na população BAT93 X Jalo EEP558 o polimorfismo foi de 55,7% e 22 microssatélites foram integrados aos 11 grupos ligação. Neste estudo, o mapa obtido teve uma cobertura de 1720 cM, com uma média de distância de 19.5 cM entre os locos. Estes autores relataram que houve maior

número de microssatélites genômicos ligados ao mapa em relação aos microssatélites desenvolvidos a partir de sequências gênicas derivadas do *GenBank*.

Em uma população de mapeamento com 157 linhas geradas a partir de um genitor Andino cultivado e um acesso selvagem Mesoamericano (ICA Cerinza X G24404), Blair *et al.* (2006b) desenvolveram um mapa genético com 80 microssatélites e 4 marcadores SCAR, com uma cobertura de 869.5 cM e alta sintenia com o mapa obtido anteriormente (Blair *et al.*, 2003). Observou-se distorção da segregação nesta população, principalmente nos grupos B8 e B1. A observação de desvios de segregação em cruzamentos entre os dois *pools* gênicos tem sido freqüente (Freyre *et al.*, 1998; Johnson & Gepts, 2002). Foram observados desvio da segregação na população BAT 93 X Jalo EEP558 não somente com microssatélites, mas também com marcadores provindos do sequenciamento de genes conservados entre *Phaseolus vulgaris*, *Lotus japonicus*, *Medicago truncatula* e *Arachis* (Hougaard *et al.*, 2008). A transmissão preferencial de alelos de um dos genitores pode ser causada por uma vantagem de um alelo específico na viabilidade ou fertilidade. Assim, os desvios na segregação tem sido observados com alguns marcadores por eles representarem ou estarem em locos ligados a alelos acidentalmente selecionados durante as gerações de autofecundação utilizadas para o desenvolvimento da população.

Rodríguez-Suárez *et al.* (2006) desenvolveram um mapa genético utilizando marcadores RAPDs, RFLPs, SCARs e 1 marcador morfológico. A população de mapeamento continha 85 indivíduos na geração F₂. O mapa cobriu cerca de 1401.9 cM, com uma média de 7.1 cM de distância entre cada marcador. Este mapa apresentou marcadores ligados a genes de resistência à antracnose, vírus do mosaico

comum, vírus do mosaico dourado, mofo branco e ferrugem. Porém, o mapa não utilizou nenhum marcador do tipo microssatélite na sua construção. Isso diminuiu a transferência dos marcadores ligados aos genes de resistência a doenças a outras populações de feijoeiro, devido a dificuldade de utilização dos marcadores dominantes (RAPD), bem como de codominantes do tipo RFLP.

O último trabalho de saturação do mapa *core* resultou na identificação da posição de 106 microssatélites (Grisi *et al.*, 2007). A distância total do mapa foi de 1358 cM, com uma média de 7.2 cM entre as marcas. Os grupos de ligação puderam ser denominados de cromossomos, de acordo com um estudo de identificação dos grupos do mapa *core* com um mapa de cromossomos mitóticos, de C1 a C11 (Pedrosa *et al.*, 2003). Este mapa de ligação representa o mapa mais saturado em microssatélites até o momento. Porém, destaca-se o tamanho amostral da população referência, BAT 93 X Jalo EEP558, com 75 linhagens puras recombinantes, limitante na detecção de possíveis recombinantes. A precisão da distância entre locos em um mapa genético e a capacidade de se identificar a ordem correta deles nos grupos de ligação estão diretamente relacionadas com o tamanho da população (Bhering & Cruz, 2008).

Um importante avanço no estudo do genoma do feijoeiro foi a publicação do primeiro esboço de mapa físico (Schlueter *et al.*, 2008). O mapa físico corresponde a uma descrição mais próxima da realidade molecular devido ao processo de seqüenciamento e alinhamento do genoma. Ele inclui regiões idênticas entre indivíduos, o que não é possível em um mapa de ligação devido à ausência de polimorfismo. A ordem das referências deve ser a mesma para os dois mapas. Neste

trabalho de Schlueter *et al.* (2008) o sequenciamento das extremidades de BACs aliado a sequências obtidas por *shotgun* geraram uma cobertura de aproximadamente 10% do genoma. Os resultados obtidos indicaram que aproximadamente 49,2% do genoma contêm sequências repetitivas e 29,3% são genes. A análise de Blast no NCBI indicou que **apenas 6 microssatélites** disponíveis em publicações foram detectados na análise de sequências deste trabalho.

O mapeamento genético e o mapeamento físico são trabalhos que se complementam e associam informações, e são freqüentemente estudados por grupos de pesquisa correlacionados. Os marcadores microssatélites podem ser usados para ancorar um mapa genético ao mapa físico, sendo esta uma etapa fundamental para a clonagem de genes de interesse para o melhoramento da cultura (Grisi *et al.*, 2007).

2.4. Características Quantitativas e Qualitativas em Feijoeiro

A variação de caracteres quantitativos, como a produção e qualidade de grãos, bem como a tolerância a doenças são de grande interesse no melhoramento e apresentam influência ambiental. W. Johannsen (1909) demonstrou no início do século XX, através de experimentos com feijão, que variações quantitativas resultam da ação combinada da segregação de múltiplos genes e fatores ambientais. A genética de tais características foi estudada a princípio utilizando técnicas de estatística (médias de populações e variâncias, covariância, herdabilidade), se contrapondo à idéia de efeitos individuais de genes propagada por Mendel (Asins, 2002).

As características qualitativas, ao contrário das quantitativas, têm seus fenótipos bem definidos em classes distintas e sua distribuição é discreta. São características governadas por um ou poucos genes, sofrem pouca influência ambiental, e conseqüentemente são de alta herdabilidade. Caracterizam-se pela fácil identificação dos diferentes fenótipos e podem ser usadas na saturação de mapas genéticos como marcadores morfológicos.

A cultura do feijoeiro domesticada é morfológicamente diversa, com variação no hábito de crescimento, coloração das sementes e flores, tipo de vagem, e fenologia (Singh *et al.*, 1991a). A cor e padrão do tegumento da semente e das flores são características geralmente usadas para distinguir genótipos de muitas espécies de plantas superiores e, em geral são de características controladas por um ou poucos genes.

Em feijoeiro, a cor é controlada por um grupo bem definido de genes que regulam vias biossintéticas de flavonol e antocianina. A expressão da cor nas flores ou no tegumento das sementes é completamente dependente de múltiplos alelos presentes no loco *P* (Emerson, 1909). O alelo dominante em *P* potencializa a cor no tegumento da semente e nas flores. Estes tecidos são brancos em indivíduos em homozigose recessiva *pp*, onde indivíduos de flores quase brancas e sementes acinzentadas-brancas são homozigotas em *p^{gri}*, outro alelo em *P* (Basset, 1994). Em *P₋* indivíduos, uma série multialélica dominante em *V* controla cor das flores: *V₋* (violeta) > *v^{lae}₋* (rosa) > *vv* (branca) (Lamprech, 1935). O tegumento da semente de genótipos *V₋* contém pigmentos de antocianina, enquanto que em genótipos *v^{lae}₋* e *vv* há pigmentos flavonóides (Beninger *et al.*, 1999, 2000). Alelos em sete outros genes – *Gy*,

C, *R*, *J*, *G*, *B*, e *Rk* – interagem com *V* e também entre eles para determinar as várias cores encontradas em sementes de feijão.

O padrão rajado do tegumento é controlado pela ação de dois genes independentes, *C* e *T*. Todos os genes controladores de cor em tegumento requerem a presença do alelo dominante *P* para expressão fenotípica. *C* é um loco complexo consistindo de muitos genes proximamente ligados responsáveis por uma variedade de padrões incluindo salpicado, rajado e matizado. Esses padrões são geralmente distinguidos por uma cor de fundo clara com a sobreposição de um padrão de cor mais escuro, determinado pelo genótipo de genes de cores descrito. Fortemente ligado ao complexo loco *C* está o gene *R*, responsável por algumas cores de tegumento vermelhas (Prakken, 1974).

O tegumento do feijão também pode ser parcialmente colorido, exemplo de um tegumento branco com cores secundárias. Sementes parcialmente coloridas requerem a presença do genótipo em homozigose recessiva *tt* no loco *T* (Emerson, 1909) e interações epistáticas entre três outros genes que controlam o tamanho e a forma da área colorida do tegumento: *zonal* (*Z*), *limiter* (*L*), e *bipunctata* (*Bip*) (Basset & McClean, 2000). O genótipo *tt* tem um efeito pleitrópico que determina flores brancas e é epistático para o loco *V* que origina flores violetas (Emerson, 1909). O loco *Z* descrito por Tschermak (1912) controla 3 padrões de pigmentação ao redor do hilo: *virgatus* (*z*), *piebald* (*z^{sel}/z*) e *sellatus* (*z^{sel}*).

Feijoeiros selvagens têm hábito de crescimento indeterminado e são trepadeiras típicas, crescendo sobre outras plantas que servem como suporte. Durante a domesticação do feijoeiro a disseminação de seus centros de domesticação,

o crescimento determinado foi selecionado em seus centros de domesticação por acelerar o florescimento e encurtar o período de floração (Cober & Taner, 1995). Em cultivares de crescimento indeterminado, o meristema terminal permanece em estado vegetativo, o qual controla a produção de folhas e hastes. Em cultivares com crescimento determinado, o meristema terminal troca de estado vegetativo a um estado reprodutivo, ou seja, é produzida uma inflorescência terminal. O loco *fin* é descrito como responsável pelo hábito de crescimento determinado ou indeterminado (Basset, 1997; Surface, 1916), sendo que o genótipo recessivo caracteriza indivíduos com crescimento determinado. Não existe um consenso na posição deste loco em estudos de mapeamento (McClellan *et al.*, 2002). Em estudos com diferentes populações e genitores de crescimento indeterminado, de origem Mesoamericana e Andina, foram obtidas posições em diferentes grupos de ligação (Koinange *et al.*, 1996; Bassett, 1997; Kwak, *et al.*, 2008). Sugere-se que a característica de hábito de crescimento evoluiu separadamente para cada centro de origem, e assim, podem ser genes múltiplos e presentes em diferentes grupos de ligação (Singh *et al.*, 1991b).

Em feijoeiro vários locos de maior efeito e QTL's já foram descritos para hábito de crescimento, sensibilidade ao fotoperíodo e tempo de floração (Norton, 1915, Wallace *et al.*, 1993; Jung *et al.*, 1996; Basset, 1997; McClellan *et al.*, 2002; Tar'an *et al.*, 2002). Mas a maioria dos estudos citados acima ainda se baseava em marcadores dominantes, como RAPD. Nestes casos, a informação contida fica restrita à população mapeada e utilizada em cada estudo, uma vez que a transferibilidade deste tipo de marcador é baixa entre genótipos diferentes.

Diante das informações apresentadas, observa-se que ainda existe a necessidade de estudos de características de herança simples, os caracteres chamados de qualitativos na cultura do feijoeiro. Frente à importância da cultura no contexto nacional e, ao potencial que pode ser explorado, torna-se evidente os avanços que podem ser alcançados no melhoramento genético do feijoeiro com a construção de mapas genéticos utilizando marcadores moleculares do tipo microssatélites, sejam eles genômicos ou funcionais.

3. Objetivo

O presente trabalho teve como objetivo principal a construção de um mapa genético para feijão (*Phaseolus vulgaris* L.) utilizando marcadores moleculares do tipo microssatélites, a partir de uma população de mapeamento derivada de duas linhagens comerciais divergentes, bem como, mapear características morfológicas de herança qualitativa.

3.1 Objetivos Específicos

- Desenvolvimento dos microssatélites:
 - Análise de sequências;
 - Desenho de iniciadores;
 - Caracterização dos microssatélites;
 - Otimização de condições de amplificação;
 - Avaliação do índice polimorfismo.

- Construção do mapa genético:
 - Genotipagem da população de mapeamento;
 - Localização dos microssatélites em grupos de ligação;
 - Mensurações fenotípicas de características qualitativas e quantitativas;
 - Mapeamento dos marcadores morfológicos.

4. Artigo I

“Structure of genetic diversity among common bean (*Phaseolus vulgaris* L.) varieties of Mesoamerican and Andean origins using new developed microsatellite markers”

Luciana Lasry Benchimol, **Tatiana de Campos**, Alisson Fernando Chioratto, Sérgio Augusto Moraes Carbonell, Carlos Augusto Colombo, Eduardo Fernandes Formighieri, Lígia Regina Lima Gouvêa e Anete Pereira de Souza.

Publicado na revista

Genetic Resources and Crop Evolution (54:1747-1762; 2007).

Structure of genetic diversity among common bean (*Phaseolus vulgaris* L.) varieties of Mesoamerican and Andean origins using new developed microsatellite markers

Luciana Lasry Benchimol · Tatiana de Campos ·
Sérgio Augusto Morais Carbonell · Carlos Augusto Colombo ·
Alisson Fernando Chioratto · Eduardo Fernandes Formighieri ·
Lígia Regina Lima Gouvêa · Anete Pereira de Souza

Received: 5 June 2006 / Accepted: 4 December 2006
© Springer Science+Business Media B.V. 2007

Abstract A common bean genomic library was constructed using the ‘IAC-UNA’ variety enriched for (CT) and (GT) for microsatellite motifs. From 1,209 sequenced clones, 714 showed microsatellites distributed over 471 simple and 243 compound motifs. GA/CT and GT/CA were the most frequent motifs found among these sequences. A total of 123 microsatellites has been characterized. Out of these, 87 were polymorphic (73.7%), 33 monomorphic (26.8%), and 3 (2.4%) did not amplify at all. In a sample of 20 common

bean materials selected from the Agronomic Institute Germplasm Bank, the number of alleles per locus varied 2–9, with an average of 2.82. The polymorphic information content (PIC) of each marker varied from 0.05 to 0.83, with a 0.45 average value. Cluster and principal coordinate analysis of the microsatellite data were consistent with the original assignment of the germplasm accessions into the Andean and Mesoamerican gene pools of common bean. Low polymorphism levels detected could be associated with the domestication process. These microsatellites could be a valuable resource for the bean community because of their use as new markers for genetic studies.

Electronic supplementary material The online version of this article (doi:10.1007/s10722-006-9184-3) contains supplementary material, which is available to authorized users.

L. L. Benchimol (✉) · S. A. M. Carbonell ·
C. A. Colombo · A. F. Chioratto · L. R. L. Gouvêa
Centro de Pesquisa e Desenvolvimento de Recursos
Genéticos Vegetais, Fazenda Santa Elisa, IAC,
CP. 28, Campinas 13012-970 SP, Brazil
e-mail: llasry@iac.sp.gov.br

T. de Campos · A. P. de Souza
Centro de Biologia Molecular e Engenharia Genética
(CBMEG), Departamento de Genética e Evolução,
Instituto de Biologia, UNICAMP, Campinas
13083-970 SP, Brazil

E. F. Formighieri
Laboratório de Genômica e Expressão, DGE,
Instituto de Biologia, UNICAMP, Campinas 13083-
970 SP, Brazil

Keywords Common bean · Molecular markers ·
Phaseolus vulgaris L. · Simple sequence repeats
(SSRs) · SSR Enriched libraries

Introduction

Common bean (*Phaseolus vulgaris* L.) is the primary source of protein in the human diet in some countries, such as Brazil. It consists of two major gene pools, a Mesoamerican and an Andean one, as determined by morphological and phaseolin seed protein attributes (Gepts 1988). From a genomic perspective, common bean has a small genome comparable to rice, estimated to be

about 450–650 million base pairs (Mb)/haploid (McClellan et al. 2004).

Common bean breeders have traditionally developed new cultivars by selection and adaptation of superior lines. Breeding programs based on previous knowledge of the genetic distances among potential progenitors to be crossed are of great importance. Molecular markers have been an important tool to characterize and determine genetic diversity among common beans (Vasconcelos et al. 1996; Métais et al. 2002). RFLPs were principally used as framework markers to develop molecular linkage maps in common beans (Nodari et al. 1993; Adam-Blondom et al. 1994). RAPDs have been extensively used, not only to develop linkage maps, but also to characterize genetic diversity (Beebe et al. 2000; Kelly et al. 2003). AFLPs have also proven useful to characterize germplasm (Pallottini et al. 2004) and develop low-density linkage maps (Ta'ran et al. 2002).

Microsatellites or Simple Sequence Repeats (SSRs) have been widely recognized as powerful and informative genetic markers in both animals and plants. SSRs consist of tandem repeated units of short nucleotide motifs that are 1–6 bp long. Di-, tri- and tetranucleotide repeats are the most common and widely distributed throughout genomes (Jarne and Lagoda 1996). Their great utility as genetic markers comes from their inherent variability that is derived from unusually high mutation rates for nucleotide sequences within SSR loci (Peakall et al. 1998).

In recent years, microsatellites for common beans have been developed from published sequences (Yu et al. 1999; Blair et al. 2003; Masi et al. 2003; Guerra-Sanz 2004) and from microsatellite-enriched libraries (Métais et al. 2002; Blair et al. 2003; Gaitán-Solis et al. 2002; Yaish et al. 2003). Recently, two major publications have added new data information to the *Phaseolus* database (Ramírez et al. 2005; Melotto et al. 2005). These publications reported EST sequencing in common bean, information which was deposited in a database at TIGR (Common Bean Gene Index—<http://www.tigr.org>). The number of microsatellites for common beans remains small, especially when compared to grasses, making it difficult to conduct

whole-genome or fine mapping. Additional microsatellite markers are needed to increase density on the linkage map (McClellan et al. 2004) especially for QTL mapping. They can also help in establishing marker-assisted selection programs and characterizing exotic germplasm. The present paper reports the use of newly developed microsatellite markers to evaluate the genetic divergence of common bean cultivars and assign them to common bean domestication centers.

Materials and methods

Plant material and DNA preparation

The variety IAC-UNA, developed by the Agronomic Institute (IAC, Campinas, São Paulo), was used for the enriched microsatellite library construction. It is a black-seeded variety that belongs to the Mesoamerican gene pool. It is resistant to anthracnose but susceptible to bean rust, *Fusarium oxysporum* Schlecht. f. sp. *phaseoli* [Kendrick and Snyder (Fop)] and angular leaf spot.

A total of 20 genotypes were selected in the IAC Core Germplasm Bank and used for the evaluation of the newly developed microsatellites (Table 1). The twenty *P. vulgaris* entries represented the Andean and Mesoamerican gene pools. Total genomic DNA from powdered lyophilised young leaf tissue was extracted using the CTAB method (Hoisington et al. 1994).

Construction of an enriched microsatellite library

An enriched library was constructed according to Billotte et al. (1999) for IAC-UNA. Six hundred nanograms of genomic DNA were digested with *RsaI* and adaptors (consisting of 21- and 25-mer primers) were ligated to the digested fragments. Selection was carried out using (CT)₈ and (GT)₈ biotinylated microsatellite primers and Streptavidine MagneSphere Paramagnetic Particles (Promega, São Paulo). The selected fragments were amplified by PCR using *Rsa21* primers and then cloned into the pGEM-T vector (Promega, São Paulo).

Table 1 Common bean (*P. vulgaris* L.) accessions used in the determination of the allelic variation of the microsatellites

No.	IAC accession	Gene pool
1	Sanilac	Mesoamerican
2	Bagajo	Andean
3	Baetão	Mesoamerican
4	Red Kidney	Andean
5	Cornell-49242	Mesoamerican
6	Porriño Sintético	Mesoamerican
7	Jamapa	Mesoamerican
8	Arc-1	Mesoamerican
9	G-4000	Mesoamerican
10	Flor de Mayo	Mesoamerican
11	Tu	Andean
12	Kaboon	Andean
13	Durango-222	Mesoamerican
14	Bayo	Andean
15	Goiano Precoce	Andean
16	Carioca Comum	Mesoamerican
17	Carioca ETE	Mesoamerican
18	Jabola (CB)	Andean
19	IAC-UNA	Mesoamerican
20	CAL-143	Andean

Plasmids were introduced into *XL-1* Blue cells; transformed cells were then plated onto agar plates containing 100 µg ml⁻¹ ampicillin and 50 µg ml⁻¹ X-galactosidase. Following incubation overnight at 37 °C, single colonies were transferred onto microplates for long-term storage at -70°C.

Detection and sequencing of microsatellite-containing clones

The genomic libraries were screened by picking 2 µl of frozen white colonies and amplifying them directly in a PCR reaction. Each amplification reaction contained 25 µl consisting of 1x reaction buffer, 2 mM MgCl₂, 0.5 µM of *RsaI* primer, 200 µM of total dNTP mixture, and 0.5U *Taq* DNA polymerase (Invitrogen, São Paulo). Amplifications were performed in a PTC-100 MJ Research thermocycler, programmed with a hot start of 4 min at 95°C; followed by 30 cycles of 30 s at 94°C, 45 s at 52°C, 1 min 30 s at 72°C, followed by 8 min at 72°C. PCR products were separated onto 3% agarose gels. Plasmid DNA was isolated according to Maniatis et al. (1982). Sequencing

was done with T₇ or SP₆ primer sites (all clones were sequenced in both directions) and the BigDye Terminator Cycle Sequencing Kit (Applied Biosystems, ABI 377 sequencer).

Analysis and edition of the generated sequences and primer design

Microsat Software (CIRAD, Risterucci et al. 2005) was used to excise adaptors and find possible *RsaI* sites inside the sequences. Reads were processed by the Phred version 0.000925.c base calling program (www.phrap.org; Ewing et al. 1998); and vector sequences, poly-A tail, and adapters were trimmed after cross-match analysis. Clustering was performed using CAP3 software with default parameters (Huang and Madan 1999). BLASTN and BLASTX program search utilities were used to identify similarities within known genes represented in the GenBank non-redundant database (Altschul et al. 1990, 1997). GO (Gene Ontology Consortium, <http://www.geneontology.org>) analysis was performed with the total amount of sequences with a stringency of 1 e⁻⁰⁵. Each sequence was blasted against the go.fasta sequences (downloaded from <http://www.geneontology.org/>) and the first hit was considered.

Only perfect and/or imperfect sequences containing to five or more repeated units were considered. To identify, count, and localize the SSR motif inside the sequences, a free software was used (SSRIT—Simple Sequence Repeat Identification Tool; Temnykh et al. 2001), available at <http://www.gramene.org/db/searches/ssr-tool>. Complementary primers to the single sequences, flanking the microsatellites, were designed with Primer Select software from the Lasergene program (DNASTar, Inc.). The stringency criteria adopted was based on GC content between 40 and 60%, melting temperature between 46 and 60°C, a salt concentration of 50 mM, and a product length between 150 and 300 pb. The primers were designed without extensive palindromes within a primer, and no dimmers. Microsatellites were synthesized by Imprint LTDA (São Paulo, Brazil).

Microsatellite primer characterization

The annealing temperature (T_a) of each microsatellite was identified using a temperature gradient in a PTC-200 thermocycler (MJ Research). PCRs were performed in a 25 μ l final volume containing 50 ng of template DNA, 0.2 μ M of each forward and reverse primer, 100 μ M of each dNTP, 2.0 mM $MgCl_2$, 10 mM Tris-HCl, 50 mM KCl, and 0.5 U *Taq* DNA Polymerase (Invitrogen, São Paulo). Reactions were performed using the following conditions: 1 min. at 94°C; then, 30 cycles of 1 min at 94°C, 1 min at T_a , 1 min at 72°C, followed by 5 min at 72°C. Amplification products were loaded on 6% w/v denatured polyacrylamide gels using a 10 bp ladder as a size standard and silver stained according to Creste et al. (2001).

Polymorphism analysis

SSR data scored for presence (1) or absence (0) of bands were transformed to genotypic data to identify locus and alleles. The polymorphism information content (PIC) value was calculated by the following formula $PIC = 1 - \sum_{i=1} f_i^2$, where f_i is the frequency of the i th allele (marker) for the i th SSR locus (Lynch and Walsh 1998).

Genetic distances (GD) were calculated from SSR data for all possible pairs of varieties using modified Rogers' genetic distance (Goodman and Stuber 1983). A genetic distance matrix was estimated using TFPGA vs. 1.3 (Mark Miller 1997). Cluster analysis was performed using UPGMA (Unweighted Pair-Group Method with Arithmetical Averages) with the NTSYS-pc computer package version 2.02E (Rohlf 1997). The stability of the clusters was also tested by a re-sampling procedure with 10,000 bootstraps using the BooD program (Coelho 2002). Principal coordinate analysis (PCO, Gower 1966) was performed the common bean MRD distance matrix and the first three principal coordinates were used to describe the dispersion of the 20 accessions according to their allele data.

Results

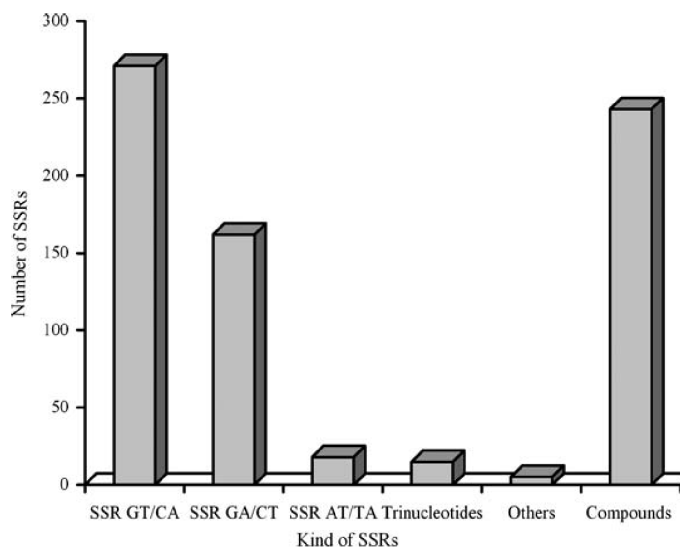
After screening of a total of 2,479 clones in the IAC-UNA library, 1,453 (58.6%) putative

positive colonies were isolated and 1,209 (48.8%) clones were sequenced. A total of 714 sequences contained microsatellites. The screening of the library showed that 471 SSRs were simple motifs (perfect and imperfect) and 243 were compound motifs. A total of 451 dinucleotides were observed and the maximum number of repeats among them was 37 (perfect GA). From Phred/Cap3 analysis, sequences were aligned in 478 contigs and 648 singlets. A total of 451 dinucleotides were observed and the maximum number of repeats among them was 37 (perfect GA). Trinucleotides were less frequent (2.1%) and showed a lesser number of repeat units (2–5 units). Other categories (tetra, penta, hexa-nucleotides) also appeared as simple repeats. Around 540 sequences were suitable for designing primers due to their sequence flanking repeats or complexity. SSRs containing GT/CA motif (38%) and GA/CT motif (22.7%) were most frequent among simple repeats (Fig. 1). The GA/CT motifs also appeared in the more extensive motifs in terms of the average number of nucleotide per motif (14 nucleotides in average) followed by compound motifs (12 nucleotides in average), which were represented by perfect and imperfect repeats. Many small microsatellites (fewer than five repeat units) were frequently found in all the analysed sequences, but they were all discarded.

Gene ontology (GO) analysis was performed with all aligned sequences (478 contigs and 648 singlets). At the biological processes level, the “cellular” category was the most expressed (35.57%). Following the cellular component level, most of the sequences presented products whose functions were supposed to be inside the “cell” (42.28%) and targeted to “organelles” (30.20%). At a molecular function level, the strongest category was “binding” (35.58%). Inside the binding category, most of the sequences seemed to be related to the “nucleic acid binding” function (38.71%), followed by “nucleotide binding” (7.53%).

BLASTN and BLASTX were used to align the 123 microsatellite sequences to the GeneBank database. The “no hits” category comprehended 73% of the data but 27% of the sequences showed some level of similarity with the alignment of non-redundant nucleotide database.

Fig. 1 Number and characteristics of microsatellite repeats found after screening an enriched SSR library. The other categories include tetra-, penta- and hexanucleotide motifs



From these, 33 unigenes presented similarity to protein domains with possible multiple functions. The most frequent category was the nucleic acid binding proteins (19% of the sequences similar) and followed by protein kinase related proteins and conserved genes of unknown function (15% each, Fig. 2). Three sequences (*SSR-IAC42*, *SSR-IAC55* and *SSR-IAC123*) were highly ($E\text{-value} < 10^{-32}$) homologous to chloroplast proteins. From BLASTN, nine sequences presented high homology with *Phaseolus vulgaris* microsatellites already deposited in NCBI: *SSR-IAC10* [~DQ185889.1 ($E < 10^{-105}$)], *SSR-IAC13* [~AJ416409.1 ($E < 10^{-68}$)], *SSR-IAC29* [~AF483864.1 ($E < 10^{-68}$)], *SSR-IAC39* [~AF483894.1 ($E < 10^{-15}$)], *SSR-IAC53* [~AF483894.1 ($E < 10^{-38}$)], *SSR-IAC72* [~AF483860.1 ($E < 10^{-18}$)], *SSR-IAC77* [~AF483853.1 ($E < 10^{-87}$)], *SSR-IAC78* [~AF483859.1 ($E < 10^{-21}$)], *SSR-IAC90* [~AF483854.1 ($E < 10^{-38}$)]; while two other sequences had low hits (*SSR-IAC16* and *SSR-IAC90*) with other microsatellites. Most of the hits found for bean unigenes showed significant similarities with *Arabidopsis thaliana* (L.) Heynh. and with genomic sequences of *Medicago truncatula* Gaertn.

Ultimately, a total of 123 microsatellites were characterized (GenBank accession numbers DQ469398 to DQ469520). After optimization of

their annealing temperatures, the microsatellites were evaluated in 6% acrylamide gels (Fig. 3). A total of 87 microsatellites were polymorphic (73.7%), 27 monomorphic (26.8%), and 3 (2.4%) did not amplify at all (Table 2). All amplified products obtained for common bean accessions presented the expected size range. The allele number ranged from 2 to 9, with an average of 2.82. Sequences were selected with five or more repeated motifs. PIC values varied from 0.05 to 0.83, with a 0.45 average value. The lowest PIC (0.05) was observed for microsatellite *SSR-IAC60*, which detained a core motif of five repetitive units, (TG)₅. The second smallest PIC value (0.10) was also found in a microsatellite (*SSR-IAC02*) with a five motif repeat. However, microsatellite *SSR-IAC27* which has the same number of repetitive motifs reached a PIC value above average (0.59). The highest PIC (0.83) was shown by microsatellite *SSR-IAC62*, which amplifies a (AG)₁₄ motif. Microsatellite *SSR-IAC47* containing a (GA)₂₀ motif, showed a PIC of 0.78.

Mesoamerican genetic distances ranged from 0.37 to 0.71, with a 0.57 average value; Andean genetic distances varied from 0.47 to 0.75 with a 0.63 average value. Genetic distances among pairs of cultivars obtained from different gene pools ranged from 0.51 to 0.82 with a 0.74 average value. Cluster analysis allocated

Table 2 SSRs' annealing temperatures (Ta), primer sequences, predicted size of product lengths (base pairs, bp), number of alleles, range of the products (base pairs, bp) for each locus and polymorphic information content (PIC).

SSR's code	Core motifs	Primer sequences	T _a (°C)	Predicted size (bp)	No. of alleles range (bp)	PIC
SSR-1AC01	(CT)8	TGCTCCCTTTTGTTGTT AAGGTCAGAAGAAGCAGAA	56	217	(2) 217–220	0.50
SSR-1AC02	(CA)5	ATGTCGCCCTCTTTTCA CATATTACAGGGTGGGCTTCT	56	288	(2) 250–290	0.10
SSR-1AC03	(AC)6	TCCCAAAATCAGACAGG TTTCAGATCCATCAGTAGTTTC	56	296	(3) 240–270	0.62
SSR-1AC04	(AG)25	GGGGTGGGATGAATGGA CAATCGGACCTGAACAATGAAA	50	241	(3) 240–250	0.61
SSR-1AC05	(TG)6 (GA)5 (AG)10 (ACA)5	TTGCAACAGCCTAAATAACCAT AGTCTCCAACTCTTCAAA	50	206	(2) 190–200	0.46
SSR-1AC06	(TA)7	CCGGTCTCTGCTGACG ATGTTCTGCTTTCGCTCCTT	56	187	(4) 174–190	0.68
SSR-1AC07	(CT)8	CTTGAGGGAGTGTAGATGTA TCAGGAGCCAAAGATCAAG	56	230	(2) 228–230	0.50
SSR-1AC08	(GT)7	CCCTCTAGTTTAAAGCCATCT GCAGGAAAATAATCGGTTGT	56	264	(2) 262–264	0.40
SSR-1AC09	(CA)9C(CA)2(TA)6	CTAGCCAGTTACATCAGACGA TCCCATTTGCCACTTC	56	191	(2) 191–193	0.66
SSR-1AC10	(GA)12 (AG)6 (AG)6	AGGAACATAAAGCCGAACCTGG GCCTCGCCGATCAACACTA	56	290	(8) 262–310	0.74
SSR-1AC11	(GA)24	TGATAAAAATGGTACACA TGATAAAAATGGTACACA	56	214	(6) 194–212	0.77
SSR-1AC12	(CA)6 (TA)4	CATTATATCTTCTCCCTTACG GAGCAACACCAAAAACACTACT	56	233	(2) 230–235	0.10
SSR-1AC13	(GA)10 A (GA)4 GG (GA)9	CCGCTGATTGGATATTAGAGTG AGCCCGTCTCTCGTTTAG	56	153	(3) 140–155	0.42
SSR-1AC14	(GT)7	GCTGCATGTTATCCACCTT TTGTTACTACCCCAACCATAC	56	221	(2) 221–223	0.47
SSR-1AC15	(TC)10	ATGCTCGACCTTCAATCCA CACTCGGCAAGCTCATTAACCA	56	217	(3) 190–230	0.10
SSR-1AC16	(GA)8	TGTAACGCCAGATTG GTTGCACTCCGACGAT	56	238	(2) 230–232	0.48
SSR-1AC17	(CA)5	AGAGGTTTCTGTGTGGTTAC ACGGTTGAATACTAGGGTTACT	56	179	(2) 179–183	Mono
SSR-1AC18	(GT)8	AACCTGTGCGCAGAAATGACT CACAACTCCCTTATGCTA	56	278	(2) 278–280	0.42
SSR-1AC19	(TC)6 TT (TC)3	AGAATGATGGTGTGAGAT CTTGGTGAATTGATAGACAT	56	173	(2) 153–175	0.26
SSR-1AC20	(GA)7 AA (GA)2	GCATCGGCAGTTCATCATT AACAAAACTACAGCCATCAGC	56	184	(3) 182–186	0.51

Table 2 continued

SSRs'code	Core motifs	Primer sequences	T_a (°C)	Predicted size (bp)	No. of alleles range (bp)	PIC
SSR-IAC21	(AC) ⁶	ACTAAATAGGAGCAGGAAGAG TAAACGAAATCAATAACAGGGT	56	234	(3) 232–236	0.65
SSR-IAC22	(TA) ⁸ (GA) ⁹	TGCAACCAACCAACCAACA GGGAAATGCGAGGCTTAGAA	56	143	(3) 138–143	0.52
SSR-IAC23	(TG) ⁷	TCGCCATTAGCCTAGAGAAGAA ACATAATATTGGCCGTAAACCTC	56	250	(2) 248–250	0.42
SSR-IAC24	(AC) ⁷ (AT) ⁶	TGGGAAATATTAGAGAACA AGCCACTGACCCCTTACAT	56	165	(3) 163–167	0.60
SSR-IAC25	(CA) ⁶ CAA(CA) ² CAA(CA) ³ CG(CA) ⁵	GAGCGTTTCATAATCAATA TTCATGCACAATAAATCACT	56	297	(3) 255–310	0.54
SSR-IAC26	(AG) ⁸	TGGATGGCAATAAATAAGCA TGTGGACTCAAAGGTGTCTC	56	148	(1) 148–188	mono
SSR-IAC27	(GT) ⁵	GGACCTGGGCAGAAAGAT AATCGATTGTTTGGTTTGAC	56	269	(3) 269–289	0.59
SSR-IAC28	(GT) ⁵ (TC) ¹⁰ (TA) ¹⁴	AAATTCAGTGTCTGTG AAGAGCTGTTAAGTTGAATA	56	289	(3) 280–294	0.66
SSR-IAC29	(GA) ²³	ACTTTTGTTCCTGCTGAT CTATTGGAGAAAGATGATGAGAG	56	230	(3) 230–232	mono
SSR-IAC30	(CA) ⁸	AATAGAAATACAAAGAGCCAAATG GGTGTCAAAATCAGAGGTAT	56	154	(3) 150–154	0.38
SSR-IAC31	(GT) ⁷	AAGCTTTGGGTTTCTCA TTCCACTATGCAATCAATCAC	56	249	(2) 247–249	0.22
SSR-IAC32	(TG) ⁷ (TA) ⁶	CAAATTAGGGGTTACAAAAG TCTAGATGGAACCTGACT	56	273	(2) 273–276	0.32
SSR-IAC33	(CA) ⁹	AACCTTAGCTCTCGCTGTGG CATTCATCTGGTATTGACTT	56	191	(2) 190–200	0.48
SSR-IAC34	(GA) ¹²	TTCCCTCTAGTTTGTGTT CTGACTGGGTATGAGATGAG	56	196	(4) 150–196	0.42
SSR-IAC35	(CT) ⁵	GTCCAACAATCATCCAACAGT ATAAGAAATCCAGGCAACA	56	269	(1) 269	mono
SSR-IAC36	(CA) ⁵	CTGTCGAGTGGAGGGGGATAA AAGGATGAATTGAGGCAGTGG	56	173	(1) 173	mono
SSR-IAC37	(TG) ⁷	TCGCCATTAGCCTAGAGAAGAA ACATAATATTGGCCGTAAACCTC	56	250	(1) 250	mono
SSR-IAC38	(TG) ⁵	TGACGGCAAAAGACACCA TAAGTAGCCAACCAATAAAA	56	255	(1) 255	mono
SSR-IAC39	(AC) ⁶	CTTGAATGCTTTAGATGTTTG GTTTGCACTCCGACGAT	56	220	(2) 198–220	mono
SSR-IAC40	(TG) ⁶	TCTGATCTCTGTACACTAA TAACTGGTCGAGATAAAAATGG	60	158	(1) 158	mono
SSR-IAC41	(AG) ⁵	AGACCCAGATGGAATAAG TTCTAATACITCCCACCCATCT	60	142	(1) 142	mono

Table 2 continued

SSR's code	Core motifs	Primer sequences	T _a (°C)	Predicted size (bp)	No. of alleles range (bp)	PIC
SSR-1AC42	(CT)14	ATTCATGTGACCTTATTT ATTGTTCCGCTCTGTATC	50–56	202	/	no amp.
SSR-1AC43	(TG)5	TGTTCTAATCCAGTTGA AGTCCACCCCTTTACA	50–56	153	/	no amp.
SSR-1AC44	(CA)7	GTGCGGGGAAGAAGACT TTGCATTTTAATATTTGGTTG	56	178	(2) 178–180	0.38
SSR-1AC45	(TG)5	CAGACAACAAATGAACAGA TTTTCAGCAGCTATGATTAT	56	201	(4) 200–238	0.44
SSR-1AC46	(CA)7	CCTTACATCTCAACTCCTAC TGATGTGACAAATAAAGAAG	56	253	(3) 253–265	0.50
SSR-1AC47	(GA)20	AAAGGGGTGTCTGAAGTT CAAGTTGGAAGAAGTGTGAG	56	306	(5) 290–340	0.78
SSR-1AC48	GTT(GT)6TGT	TGCACGTGGAACAGACA ACTAAGTGCCAAACAACCTATT	64.4	272	(2) 270–272	0.26
SSR-1AC49	(AG)9	GCCATCCATGACAGACAG GCTAATATAACACGCTAAAA	56	231	(2) 229–231	0.46
SSR-1AC50	(GT)7	ATGATATAACAACCTCACCATTT GTGCAACTCCACCATCT	56	169	(3) 156–160	0.54
SSR-1AC51	(GA)5CA(GA)9CA(GA)2	CCAGCAAAATAAACACCCCAA AACAGAGCAACGAAAAAAGG	56	221	(2) 242–252	0.47
SSR-1AC52	(GA)11	TGCATGTATGTAGCGGTTTA GTGGCTTTGTCTTTGTAGTCA	56	203	(9) 160–210	0.70
SSR-1AC53	(GA)9	ACGCATGAGTGATTGG CTGAAAAGGAGTGAGCA	56	175	(3) 175–179	0.38
SSR-1AC54	(AC)6CAAA(TA)3C(AT)5	CTTTGCTTGTGTGGAGAG CACCTGTGCTGACTTAG	56	152	(3) 156–160	0.58
SSR-1AC55	(GA)13	AACCCGTGAATCTTTGAGG ATTGATGGTGGATTITGAA	56	211	(4) 200–220	0.59
SSR-1AC56	(AC)8	CTGCACAACTCCCTTTATG ACTGTGCGAGAATGACTTGA	56	280	(2) 280–284	0.44
SSR-1AC57	(GT)5	CATGCTTGTGCTACTTTC TTTTTGTGAGCTTTATGTTAC	56	279	(3) 220–290	0.64
SSR-1AC58	(TG)10	CATGCACTCTGGTATTA AACTTTAGTCTCGGCTGTGGA	56	198	(2) 192–202	0.44
SSR-1AC59	(AC)7	CAAGTGACCGGAGAAGATTTT GTGCATCAGACGGCTCAAG	61	161	(4) 248–300	0.49
SSR-1AC60	(TG)5	CTCAAGTCAGCCAGCAAGAAA AGATTACGGACAAGAACTGAG	61	152	(2) 137–139	0.05
SSR-1AC61	(CA)6(AT)5	CACACGCACACGGCACAC TGGCAATGGAAGAAACAAAAT	58.6	157	(4) 145–175	0.34
SSR-1AC62	(AG)14	AACCCGTGAATCTTTGAGG ATTGATGGTGGATTITGAA	45.3	211	(9) 192–290	0.83

Table 2 continued

SSRs' code	Core motifs	Primer sequences	T_a (°C)	Predicted size (bp)	No. of alleles range (bp)	PIC
SSR-IAC63	(AC)6	TCGTAGCAGCTAAGATGGAAGA GTTTGTGAACTGTTGAAATGTG	59.8	210	(3) 208–212	0.47
SSR-IAC64	(AC)6	GCTTTACTTGCCCACTTGC TTCIAGCCAGATATTTCTCA	56	257	(3) 208–212	0.58
SSR-IAC65	(TG)5	AGTGATGAAATAGATGTCCTT GACTAGATGTTACCCCTCCTCA	60	285	(3) 280–300	0.51
SSR-IAC66	(GA)10	AATCACATCTTTAACCCCAACAG TTCACCTCCCTCCTATCTT	56	282	(7) 224–310	0.70
SSR-IAC67	(GT)7	GAAGCTGCGACGGAACATAG CCTAGTCCCTCCCATCCAG	56	110	(3) 115–110	0.46
SSR-IAC68	(CT)8	TGGAGGTAAACGCTTTTITG ATTTAACATGAACGACCACC	56	266	(2) 266–270	0.46
SSR-IAC69	(TG)8	TTTTAACATGCTCCCTCTAC GGTCCACAATCAAGCAGTCAA	61	283	(2) 258–275	0.29
SSR-IAC70	(AC)8	CTCCAGGAAGGTAATGTTGT AAATGGACTTGAGCACCTAAA	60	205	(02) 220–225	0.17
SSR-IAC71	(TG)7	TTCGGTGTGTAATCC GAATCCACTAGGTAATCAAAATCC	60	167	(02) 160–163	0.46
SSR-IAC72	(TG)7	ATCGGTTGAATGGCTTGAC ATTGCTTAAAGACTCTCTGTTC	45	262	(03) 203–210	0.73
SSR-IAC73	(AT)6 (GT)6	TTAGTTTCTCGTCAATGGA GCATAAGAAACCAAGAGCAT	60	227	(03) 238–250	0.64
SSR-IAC74	(CA)9 (TA)7	GGAATCGAAGTTTGAAGTGAGG AAATGACCAAGCCCAAGAAATGTT	60	271	(04) 260–294	0.50
SSR-IAC75	(GT)6 (TC)7	TGTGAGGTCAGAGGGTGT CGTGTGTTTATACGAATCA	60	282	(02) 340–350	0.46
SSR-IAC76	(TA)10 (TG)7 (TA)4	TTCATGGCCAATAATCAGG GAGAAATTCAGAGGGTAGATG	60	191	(03) 195–205	0.46
SSR-IAC77	(CA)6 (CT)4	CACGGTTGGAGAGATGATG ACCAATACAGGAAAGGAGTT	60	241	(02) 260–262	0.64
SSR-IAC78	(GT)8	GGCCATTGCACTCCGACGAT GGGGCTTTAGATGTTTGAGACG	60	215	(02) 215–218	0.63
SSR-IAC79	(GT)6	TGTTGCCATATGCTTCTAA CCTCCAACCGGTGTAACCT	60	179	(02) 190–193	0.30
SSR-IAC80	(TG)4 T (TG)2	TTTGGGAATATTAAGGCACATC CAAACTTAAATAATCGCAACT	60	187	(02) 212–212	0.58
SSR-IAC81	(TG)4 TT (TG)4 (AT)3	ATGGACCTATATGGCTTTGT GAACCTTTTAAATAATCTGAGT	60	250	(03) 273–277	0.50
SSR-IAC82	(GT)7	TGCGTATGTATGTTTCAGG TTGGGTCCATTTGGTGTCTCTAT	60	192	(02) 203–205	0.50
SSR-IAC83	(TC)11	ACCGAGATGACGTAGGAATG AGTTGAGAAGGGCGCAGTTGTTA	45	286	(05) 295–320	0.48

Table 2 continued

SSRs' code	Core motifs	Primer sequences	T _a (°C)	Predicted size (bp)	No. of alleles range (bp)	PIC
SSR-IAC84	(TG) ₆	TTGCACTCTTGTTGTTTATGGA CACAAATGACGACAGATGACAGA	60	154	(02) 162–164	0.30
SSR-IAC85	(TG) ₄ T (TG)	TCTTCCCTTCACACTCAAA AGGAATCTCAGAAATGCTCAATC	60	173	(02) 185–186	0.17
SSR-IAC86	(AC) ₅ (AT) ₂	AATGCGTACGTTTCTTAAT AGCACCGCTAIGGCATC	64.8	141	(02) 155–156	mono
SSR-IAC87	(AC) ₉	AACCTTAATTTGCTGCTGCA CTTCTCCCTATTTGICIT	63.5	242	(02) 250–257	0.55
SSR-IAC88	(CA) ₇ (AT) ₇	TCTTGGGTGCTCGTACTTA GTTTGGGATCTCGGGACCTT	60	274	(04) 275–290	mono
SSR-IAC89	(CA) ₂ C (CA) ₃	GCCTCCAGCGTCTCTTACTTG TCGGGATGCAGGAGGAC	60	222	(02) 164–165	mono
SSR-IAC90	(CT) ₅ (TC) ₆ e (CA) ₄ (TA) ₆	GCACATCTCTCTCCCTCCTAA GCGTGGCCTATTTTCATT	60	257	(05) 250–270	0.17
SSR-IAC91	(AC) ₃ (TC) ₂	AACCGCAAAGAGACACCA TGCAAGGTACACAAACACA	60	196	(05) 190–200	0.46
SSR-IAC92	(AC) ₆	GATAATCAGGTCAAAAGTT GTGGACAGGGACATAATCTAAT	60	211	(03) 115–222	0.76
SSR-IAC93	(AC) ₅	AGTTCGCTTGGATTCTA GGATTTGTTCTGCTTACC	60	251	(04) 260–268	0.18
SSR-IAC94	(AC) ₅	CGAGATGTCCTGCTTCA CTCCATTCTATATAGTTTCA	60	182	(02) 190–235	0.68
SSR-IAC95	(CA) ₇ (AT) ₇	AAACAACAGGGAATAACACAA TAAAGCACAAACGAAACATA	60	165	(03) 162–175	0.64
SSR-IAC96	(CA) ₅ (TA) ₂	AAGCGATAATCAATCCAACAT CTTACCCATCCTCATTTTCATT	60	285	(03) 283–295	0.62
SSR-IAC97	(AC) ₃ (TC) ₂	AACCTCATGCATCTCTTTATT GTGTTGGCTGCTGTCA	60	244	(04) 242–260	0.54
SSR-IAC98	(CT) ₈ (TA) ₃ (TG) ₈	ACATGGGCTACAGGGACAAT AGGGAGTTAACGGTATGGTATG	60	243	(02) 241–244	0.32
SSR-IAC99	(AAG) ₇	GATAGCTGAATGATTTGGTTTA TCCCTATTACTGCGACATT	60	176	(03) 174–180	0.42
SSR-IAC100	(AT) ₄ (GT) ₈	GTAACCGCAAAAGAGACACC TGCAAGGTACACAAACACA	60	274	(04) 274–282	0.72
SSR-IAC101	(AC) ₇	CGTTTAAATGCGTTGAA TTCGAGAGCAAGCAACCA	60	172	(02) 170–172	0.50
SSR-IAC102	(CT) ₇ (GTCA) (CT) ₈	GGTTGTTTCTGCCCTCCACTG GCTAATGCTGCACACCAAACT	60	201	(04) 225–233	0.49
SSR-IAC103	(AG) ₃ A (AG) G (AG) ₃	CCGTGAAGTGAACAAAGGTG GGCCATCATCCCCAACAG	60	127	(1) 130	mono
SSR-IAC104	(GT) ₄ (GTTG) ₂	GTTACTAGCAGGCGAGAGAA GCAGGGGACACTTACTACAGAG	60	235	(1) 246	mono

Table 2 continued

SSRs' code	Core motifs	Primer sequences	T_a (°C)	Predicted size (bp)	No. of alleles range (bp)	PlC
SSR-IAC105	(CA)5 (TA) (CA)3	GTGGACGAGGAGAAATTTTG AGACTTCCTCCCACTGATTG	68	248	(1) 268	mono
SSR-IAC106	(GT)6	TGCGTATTGTATGTTTCAGG TTGGGTCCATTTGGTGCTCT	60	192	(1) 213	mono
SSR-IAC107	(TG)3 (TA) (TG)3 (AG)3	CAATCTCGAGAGAAAT TAGAAAAGGGGATAAGTGA	62.3	273	(1) 298	mono
SSR-IAC108	(CT)3 (CA)3	TTCTGGTCTCTATGGTTGGT AATGAAATAATGCAGTGGTAGC	64.8	143	(1) 145	mono
SSR-IAC109	(GT)2 (AT) (GT)6	AGTCAGCCAGCAAGAAACACCA TGGGGAATATTTTGGCTGAA	64	138	(1) 144	mono
SSR-IAC110	(GT)6	GTGTGCTCTGAGGTGTGA GGGAATCCACTAAGTAATCAAA	64	190	(1) 177	mono
SSR-IAC111	(GT)4 e (CT)4	GTGCTGAGCCAAAGGAAGT AAACAGTAAAGCACCAAGAAA	60	272	(1) 270	mono
SSR-IAC112	(CA)6	CCAACATTCAGACACCATCCA TTTTTGCACTCTTGTTGTTTA	60	279	(1) 301	mono
SSR-IAC113	(CT)4 (AG)3	GGTGGTGGTTTCTCTCTC ACCTTAGCAACCTTAGT	66.5	272	(1) 300	mono
SSR-IAC114	(AG)8	GAGACGGAAGATAGAAAAAGA CTCTCGGCTCTTACTCTC		167	/	no amp.
SSR-IAC115	(TC)7	ATGCACAAGCGGTAAAAA TCGTGCTTGTCTTCTCTGTA	64.8	208	(1) 210	mono
SSR-IAC116	(GA)3 (GT)2	AGACATGTTGATACGGGAGAT CACCTTGACTTGCCTTTGAC	60	185	(1) 190	mono
SSR-IAC117	(CA)5	CATCAACAACAACAACATAACC ACACCTTTCTTTCTTTCTT	60	310	(1) 310	mono
SSR-IAC118	(TG)6	TTGGAACACCGGGGAATGGA GACGTTGGAGACAGGGGAGAG	60	143	(1) 150	mono
SSR-IAC119	(AC)7	TTTTTGCTGAAGGAGATAACA TCAAGTCAGCCAGCAAGAAACA	60	116	(1) 120	mono
SSR-IAC120	(AC)3 (TC)2	AGCTACATCCAGTCTCTCA AGTTTCGTTTCTGTGTTGTTA	60	199	(1) 207	mono
SSR-IAC121	(AG)2 (TG)3	TGCTTACGCTCCAGTCATTA GCAAAACACAGAAACGAAAC	60	279	(1) 281	mono
SSR-IAC122	(TG)8 (TA)3	TCCCGATTATAGTCTCATTT GGGACCTCTTCACTCTG	60	222	(1) 228	mono
SSR-IAC123	(AT)3 (CT)2	TGAAGCCCCCAAGGAGAAAT ATAGGCACAAAATACCCAAAGAA	60	295	(1) 295	mono

Fig. 2 Graph based on the BLASTX homology of the 123 sequences for the non-redundant database. The results are percentages over the total of similarity (27%) achieved from alignment. No hits represented 73% of the data

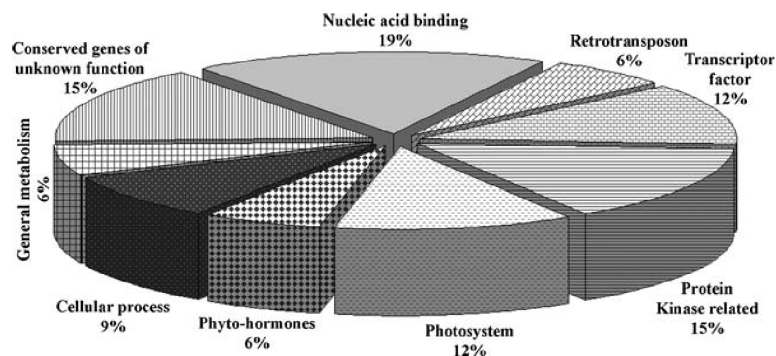
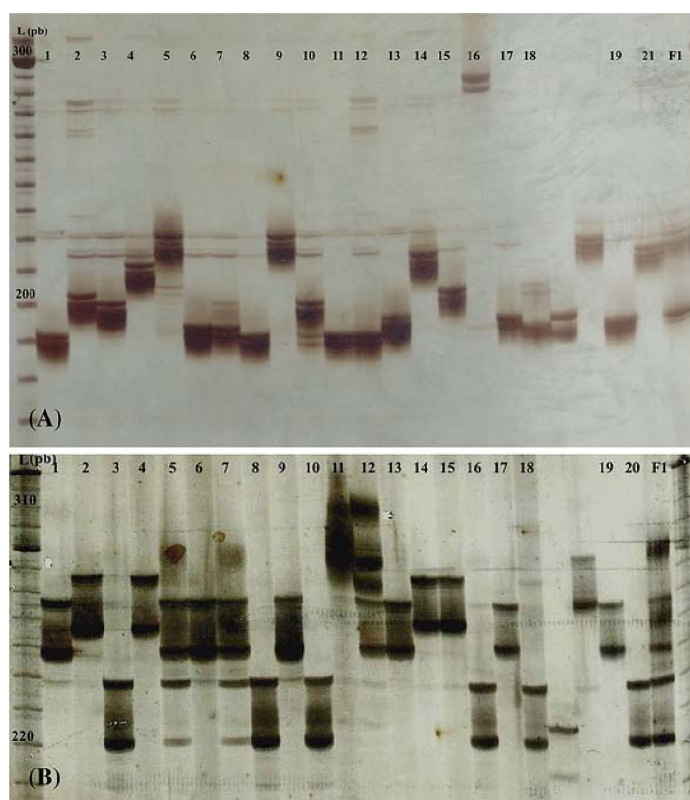


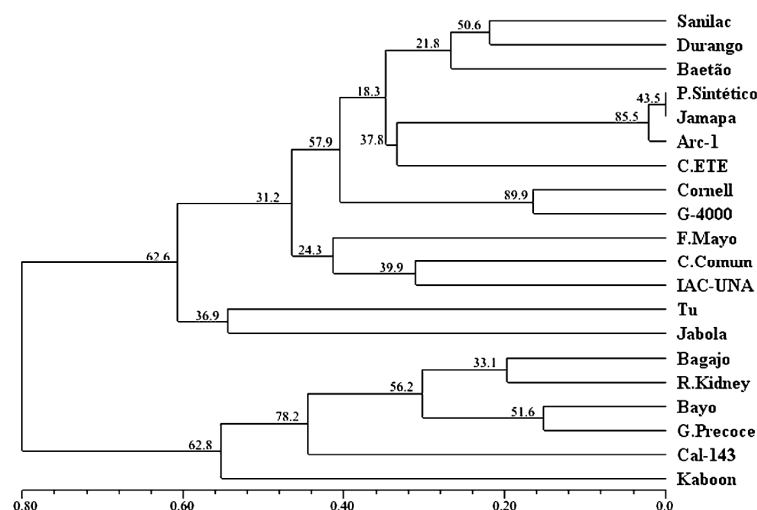
Fig. 3 Amplification products for microsatellite *SSR-IAC49* with common bean genotypes on a 6% denaturing polyacrylamide gel. **(a, b)** Amplification products for microsatellite **(A)** *SSR-IAC62* and **(B)** *SSR-IAC66* with common bean genotypes on 6% denaturing polyacrylamide gels. The numbers follow the order presented in Table 1. F₁ is the hybrid material generated by IAC-UNA (19) × CAL-143 (20) cross



genotypes in two distinct groups according to their domestication centers (Fig. 4). ‘Sanilac’, ‘Durango-222’, ‘Baetão’, ‘Porrillo Sintético’, ‘Jamapa’, ‘Arc-1’, ‘Carioca Eté’, ‘Cornell-49242’, ‘G-4000’, ‘Flor de Mayo’, ‘Carioca Comum’ and ‘IAC-UNA’ were grouped consistent with their Mesoamerican origin. In contrast, ‘Bagajo’, ‘Red Kidney’, ‘Bayo’, ‘Goiano

Precoce’, ‘CAL-143’ and ‘Kaboon’ grouped together according to their Andean origin with a bootstrap precision of 62.8%. The cultivar ‘Tu’ grouped with ‘Jabola’, both Andean materials, with a lower bootstrap node support (36.9%), but with a higher bootstrap support for the node that connects these two genotypes to Mesoamerican accessions (62.6%).

Fig. 4 UPGMA cluster analyses of genetic distances calculated from SSR data. Values at nodes represent the levels of bootstrap support (in percentages)



Principal coordinate analysis (Fig. 5) accounted for 45% of the total variation on the three first principal coordinates. This analysis divided common bean materials into two main groups according to their Andean and Mesoamerican origins, reflecting the same group-association observed in the dendrogram. Three materials, 'Tu', 'Jabola' and 'Kaboon', all Andean, were clearly separated from the others. CAL-143 is more externally linked to its related group ('G. Precoce', 'Bayo', 'R. Kidney' and 'Bagajo').

Discussion

Results from the enriched library characterization are consistent with what has been described previously for SSR isolation from many plant species; however, GA repeats have usually been reported to be more abundant than CA repeats (Powell et al. 1996; Echt and May-Marquardt 1997; Maguire et al. 2000; Gaitán-Solís et al. 2002; Yaish et al. 2003). According to Powell et al. (1996), (AT)_n repeats usually represent the most common type of repeat in plants. This was shown by Yu et al. (1999) in common bean as well as they found that AT dinucleotides were more frequent than GA motifs in *P. vulgaris* L. and *Vigna unguiculata* (L.) Walp. However, they

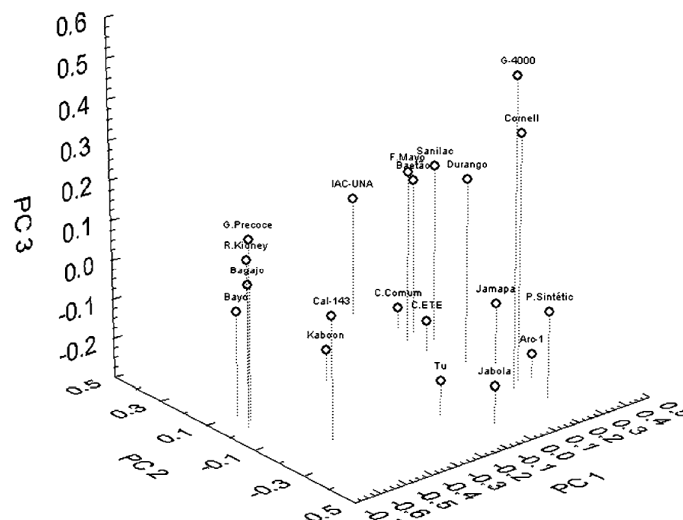
were rarely isolated in our study, probably because (AT)_n repeats are palindromic and therefore may have not been efficiently enriched during the capture process and because the library was enriched with CT and GT motifs.

The allele range encountered in our study is consistent with other common bean reports. Yu et al. (2000) found an allele range that varied from 2 to 10 for a 24 polymorphic SSR evaluation in 12 genotypes. Guerra-Sanz (2004) reported an allele variation from 2 to 7 alleles for 18 polymorphic SSR loci.

Regarding the similarities evaluated by GO for all alignments and BLASTX specifically performed for microsatellite sequences, the majority of the hits were associated to nucleic acid binding proteins. MADS-box proteins were found for SSR-IAC29, which detain (GA)₂₃ repetition. In fact, MADS-box is a highly conserved motif found in a family of transcriptional factors, which play an important role in developmental processes. A (GA)_n microsatellite linked to a putative MAD-box gene has already been isolated in the common bean (Yaish et al. 2003). Protein kinase related sequences (4.07%) found by common bean unigene alignment may be associated with disease resistance.

As reported by Gaitán-Solís et al. (2002), the efficiency of a given primer does not depend only on the number of patterns it generates. Our

Fig. 5 Associations between 22 common bean genotypes revealed by principal coordinate analysis of Rogers' genetic distances



results showed that composite motifs might also provide an informative polymorphic pattern for microsatellite assays. In fact, we found a high frequency of small repetitive units (usually four repetitive motives) and many compound motives in our enriched library and this may be a characteristic of the species. Our PIC values were about average and could be explained by common bean genetic diversity. Métais et al. (2002) published a range of 0.12–0.72 for PIC values, with an average of 0.44 when evaluating 15 polymorphic SSRs in 45 different bean lines, belonging to nine different quality types.

Cluster analysis allowed common bean germplasm to be divided into two main groups and this result was confirmed by coordinate analysis. The same grouping represented by the dendrogram was observed by Duarte et al. (1999) using RAPD markers: a clear separation between cultivars from Mesoamerican and Andean South American domestication centers. The upper dendrogram group included varieties and commercial cultivars ('Carioca Ete', 'Carioca Comum', 'IAC-UNA') from the Mesoamerican gene pool, all small seeded beans. In contrast, Andean genotypes seemed to cluster together. However, Andean genotypes showed greater mean genetic diversity than Mesoamerican accessions. In Brazil, most of the adapted cultivars belong to the Mesoamerican gene pool. Knowl-

edge of its divergence may have practical applications. As common bean organization into gene pools was originally based on phenotypic traits and biochemical markers (phaseolin and isoenzymes, Singh et al. 1991), only a part of bean genome information has been accessed. As more populations of wild and domesticated beans are examined by molecular markers, refinement of the current classification is expected. This way, a better look at 'Kabbon', 'Tu' and 'Jabola' classification should be considered due to the intrinsic variability shown by these materials with respect to the Andean and Mesoamerican gene pool diversity. These three accessions may results from hybridization between the two major gene pools.

Molecular markers could separate commercial bean lines according to their geographical origins (Métais et al. 2000), between red and black Mesoamerican beans (Beebe et al. 1995) or among different *Phaseolus* species (*P. vulgaris* x *P. coccineus* L., Sicard et al. 2005). The major difficulty of microsatellites is that they still need to be isolated *de novo* from most species examined for the first time, so cross-species amplification is not always possible. Yu et al. (2000) reported that SSR sequences are fairly abundant in the bean genome and distributed in a widespread manner. Once developed, microsatellites are ideal markers, as they are stable and easy to assay by polymerase chain reaction. Several important

genes such as the resistance genes may be linked to microsatellite motifs, which make them relevant for studies of germplasm characterization, mapping and marker assisted-selection.

Acknowledgements This work was supported by the Foundation for Research for the State of São Paulo (FAPESP), contract 02/03225-9. Dr. A. P. Souza received a fellowship from the National Council of Research and Development (CNPq). Dr. L.L. Benchimol received a post-graduate fellowship (02/00752) and T. Campos received under-graduate (03/13282-2) and post-graduate (140310/2005-3) fellowships from FAPESP and CNPq. We would like to thank Dr. Ange-Marie Risterrucci for helping with the construction of the enriched library and Dr. JP Jacquemoud-Collet for providing Microsat software (CIRAD, France); Dr. Dario A. Palmieri and Dr. Marco A. Takita for helping with the PHRED/CAP3 analysis (IAC, Cordeirópolis, S.P., Brazil) and Dr. Maria I. Zucchi for the bootstrap analysis (IAC, Campinas, S.P., Brazil).

References

- Adam-Blondom A, Sévignac M, Dron M (1994) A genetic map of common bean to localize specific resistance genes against anthracnose. *Genome* 37:915–924
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410
- Altschul SF, Madden TL, Schäffer A, Zhang J, Zhang Z, Miller W, Lipman DJ, (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402
- Beebe S, Skroch PW, Nienhuis J, Tivang J (1995) Genetic diversity among common bean breeding line developed for Central America. *Crop Sci* 35:1178–1183
- Beebe SE, Skroch P, Tohme J, Duque M, Pedraza F, Nienhuis J (2000) Structure of genetic diversity among common bean landraces of Middle-American origin based on correspondence analysis of RAPD. *Crop Sci* 40:264–273
- Billotte N, Lagoda P, Risterrucci AM, Baurens C (1999) Microsatellite-enriched libraries: applied methodology for the development of SSR markers in tropical crops. *Fruits* 54:277–288
- Blair MW, Pedraza F, Buendia HF, Gaitán-Solís E, Beebe SE, Gepts P, Tohme J (2003) Development of a genome-wide anchored microsatellite map for common bean (*Phaseolus vulgaris* L.). *Theor Appl Genet* 107:1362–1374
- Coelho ASG (2002) BooD: avaliação dos erros associados a estimativas de distâncias/similaridades genéticas através do procedimento de bootstrap com número variável de marcadores (software). Goiânia: UFG, Instituto de Ciências Biológicas, Laboratório de Genética Vegetal.
- Creste S, Tulmann A, Figueira A (2001) Detection of Single Sequence Repeat Polymorphism in denaturing Polyacrylamide Sequencing Gels by Silver Staining. *Plant Mol Biol Rep* 19:299–306
- Duarte JM, Dos Santos JB, Melo LC (1999) Genetic divergence among common bean cultivars from different races on RAPD markers. *Genet Mol Biol* 22(3):419–426
- Echt CS, May-Marquardt P (1997) Survey of microsatellite DNA in pine. *Genome* 40:9–17
- Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. *Genome Res* 8:175–185
- Gaitán-Solís E, Duque MC, Edwards KJ, Tohme J (2002) Microsatellite Repeats in Bean (*Phaseolus vulgaris*): Isolation, Characterization and Cross-Species Amplification in *Phaseolus* spp. *Crop Sci* 42:2128–2136
- Gepts P (1988) A Middle American and Andean gene pool. In: Gepts P (eds) Genetic resources of *Phaseolus* beans. Kluwer, Dordrecht, the Netherlands, pp 375–390
- Guerra-Sanz JM (2004) New SSR markers of *Phaseolus vulgaris* from sequence databases. *Plant Breeding* 123:87–89
- Hoisington D, Khairallah M, González-de-León D (1994) Laboratory Protocols: CIMMYT Applied Molecular Genetics Laboratory. 2nd Edition. Mexico, DF, CIMMYT
- Huang X, Madan A (1999) CAP3: A DNA Sequence Assembly Program. *Genome Res* 9:868–877
- Jarne P, Lagoda P (1996) Microsatellites from molecules to populations and back. *Trends Ecol Evol* 11:424–429
- Kelly JD, Gepts P, Miklas PN, Coyne DP (2003) Tagging and mapping of genes and QTL and molecular marker-assisted selection for traits of economic importance in bean and cowpea. *Field Crops Res* 82:135–154
- Lynch M, Walsh JB (1998) Genetics and analysis of quantitative traits. Sinauer Associates, Sunderland, MA. p 980
- Maguire TL, Edwards KJ, Saeger P, Henry R (2000) Characterization and analysis of microsatellite loci in mangrove species, *Avicennia marina* (Forsk.) Vierh. (Avicenniaceae). *Theor Appl Genet* 101:279–285
- Maniatis T, Fritsch EF, Sambrook J (1982) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Lab., Cold Spring Harbor, New York
- Masi P, Spagnoletti Zeulli PL, Donini P (2003) Development and analysis of multiplex microsatellite markers sets in common bean (*Phaseolus vulgaris* L.). *Mol Breeding* 11:303–313
- McClellan P, Gepts P, Kami J (2004) Genomics and genetic diversity in common bean. In: Wilson RF, Stalker HT, Brummer EC (eds) Legume Crops Genomics. AOCS Press, Champaign, Illinois, pp 60–82
- Melotto M, Monteiro-Vitorello CB, Bruschi AG, Camargo LEA (2005) Comparative bioinformatic analysis of genes expressed in common bean seedlings. *Genome* 48(3):562–570
- Métais I, Hamon B, Jalouzot R, Peltier D (2002) Structure and level of genetic diversity in various bean types

- evidenced with microsatellite markers isolated from a genomic enriched library. *Theor Appl Genet* 104:1346–1352
- Nodari RO, Tsai SM, Gilbertson RL, Gepts P (1993) Towards an integrated linkage map of common bean. II. Development of an RFLP-based linkage map. *Theor Appl Genet* 85:513–520
- Pallottini L, Garcia E, Kami J, Barcaccia G, Gepts P (2004) The genetic anatomy of a patented yellow bean. *Crop Sci* 44:968–977
- Peakall R, Gilmore S, Keys W, Morgante M, Rafalski A (1998) Cross-species amplification of soybean (*Glycine max*) Simple Sequence Repeats (SSRs) within the genus and other legume genera: implications for the transferability of SSRs plants. *Mol Biol Evol* 15(10):1275–1287
- Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A, (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for gemplasm analysis. *Mol Breeding* 2:225:238
- Ramírez M, Graham MA, Blanco-López L, Silvente S, Medrano-Soto A, Blair M, Risterucci AM, Duval MF, Rohde W, Billotte N (2005) Isolation and characterization of microsatellite loci from *Psidium guajava* L. *Mol Ecol Notes* 5(4):745–748
- Singh SP, Gepts P, Debouck DG (1991) Races of common bean (*Phaseolus vulgaris*, Fabaceae). *Econ Bot* 45:379–396
- Sicard D, Nanni L, Porfiri O, Bulfon D, Papa R (2005) Genetic diversity of *Phaseolus vulgaris* L. and *P. coccineus* L. landraces in central Italy. *Plant Breeding* 124:464–472
- Ta'ran B, Michaels Thomas E, Pauls KP (2002) Genetic mapping of agronomic traits in common bean. *Crop Sci* 42:544–556
- Temnykh S, De Clerk G, Lukashova A, Lipovich L, Cartinhour S, Mc Couch, S (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon association and genetic marker potential. *Genome Res* 11:1441–1452
- Vasconcelos MJV, Barros EG, Moreira MA and Vieira C (1996) Genetic diversity of common bean *Phaseolus vulgaris* L. determined by DNA-based molecular markers. *Brazilian J Genet* 19(3):447–451
- Yaish MWF, Pérez de la Vega M (2003) Isolation of (GA)_n microsatellite sequences and description of a predicted MADS-box sequence isolated from common bean (*Phaseolus vulgaris* L.). *Genet Mol Biol* 26(3):337–342
- Yu K, Park SJ, Poysa, V (1999) Abundance and variation of microsatellite DNA sequences in beans (*Phaseolus* and *Vigna*). *Genome* 42:27–34
- Yu K, Park J, Poysa V, Gepts P (2000) Integration of Simple Sequence Repeats (SSR) markers into a molecular linkage map of common bean (*Phaseolus vulgaris*). *J Hered* 91:429–434

5. Artigo II

“Development, characterization, and comparative analysis of polymorphism at common bean SSR loci isolated from genic and genomic sources”

Luiz Ricardo Hanai, **Tatiana de Campos**, Luis Eduardo Aranha Camargo, Luciana Lasry Benchimol, Anete Pereira de Souza, Maeli Melotto, Sérgio Augusto Morais Carbonell, Alisson Fernando Chioratto, Luciano Consoli, Eduardo Fernando Formigueri, Marcos Vinicius Bohrer Monteiro Siqueira, Siu Mui Tsai, e Maria Lucia Carneiro Vieira.

Publicado na revista
Genome (50:266-277; 2007).

Development, characterization, and comparative analysis of polymorphism at common bean SSR loci isolated from genic and genomic sources

Luiz Ricardo Hanai, Tatiana de Campos, Luis Eduardo Aranha Camargo, Luciana Lasry Benchimol, Anete Pereira de Souza, Maeli Melotto, Sérgio Augusto Moraes Carbonell, Alisson Fernando Chioratto, Luciano Consoli, Eduardo Fernando Formighieri, Marcos Vinicius Bohrer Monteiro Siqueira, Siu Mui Tsai, and Maria Lucia Carneiro Vieira

Abstract: Microsatellites or SSRs (single sequence repeats) have been used to construct and integrate genetic maps in crop species, including *Phaseolus vulgaris*. In the present study, 3 cDNA libraries generated by the Bean EST project (<http://lgm.esalq.usp.br/BEST/>), comprising a unigene collection of 3126 sequences and a genomic microsatellite-enriched library, were analyzed for the presence of SSRs. A total of 219 expressed sequence tags (ESTs) were found to carry 240 SSRs (named EST-SSR), whereas 714 genomic sequences contained 471 SSRs (named genomic-SSR). A subset of 80 SSRs, 40 EST-SSRs, and 40 genomic-SSRs were evaluated for molecular polymorphism in 23 genotypes of cultivated beans from the Mesoamerican and Andean genetic pools, including Brazilian cultivars and 2 related species. Of the common bean genotypes, 31 EST-SSR loci were polymorphic, yielding 2–12 alleles as compared with 26 polymorphic genomic-SSRs, accounting for 2–7 alleles. Cluster analysis from data using both genic and genomic-SSR revealed a clear separation between Andean and Mesoamerican beans. The usefulness of these loci for distinguishing bean genotypes and genetic mapping is discussed.

Key words: *Phaseolus vulgaris*, microsatellites, genetic markers, expressed sequence tags (EST).

Résumé : Des microsatellites ou SSRs (« simple sequence repeats ») ont été employés pour produire et intégrer des cartes génétiques chez plusieurs espèces cultivées, incluant le *Phaseolus vulgaris*. Dans le présent travail, trois banques d'ADNc générées par le « Bean EST project » (<http://lgm.esalq.usp.br/BEST/>) totalisant 3 126 unigènes et une banque génomique enrichie en microsatellites ont été analysées pour la présence de microsatellites. Au total, 219 ESTs comprenaient 240 SSRs (EST-SSR) tandis que 714 séquences génomiques comptaient 471 SSRs (SSR génomiques). Un sous-ensemble de 80 SSRs, 40 EST-SSRs et 40 SSRs génomiques, ont été évalués pour leur polymorphisme chez 23 génotypes de haricots cultivés provenant des ressources génétiques mésoaméricaines et andéennes, incluant des cultivars brésiliens, et deux espèces apparentées. Chez les génotypes du haricot, 31 EST-SSRs étaient polymorphes et révélaient entre 2 et 12 allèles tandis que 26 des SSR génomiques étaient polymorphes avec entre 2 et 7 allèles. Une analyse de groupement employant les données des deux types de SSR a montré une séparation claire des génotypes andéens et mésoaméricains. L'utilité de ces loci pour la discrimination de ces génotypes et pour la cartographie génétique est discutée.

Mots-clés : *Phaseolus vulgaris*, microsatellites, marqueurs génétiques, étiquettes de gènes exprimés (EST).

Introduction

Microsatellites or single sequence repeats (SSRs) are DNA regions composed of small motifs of 1 to 6 nucleotides repeated in tandem, which are widespread in both eukaryotic and prokaryotic genomes (Field and Wills 1996;

Tóth et al. 2000). Microsatellite markers were developed for a wide range of plant species, including many crops. The advantage of microsatellites as molecular markers is that they detect polymorphisms at highly variable loci (Morgante and Olivieri 1993; Powell et al. 1996; Hedrick 2001). Furthermore, they can be amplified by PCR, are codominant,

Received 9 August 2006. Accepted 6 December 2006. Published on the NRC Research Press Web site at genome.nrc.ca on 9 May 2007.

Corresponding Editor: F. Belzile.

L.R. Hanai, L.E.A. Camargo, M. Melotto, L. Consoli, M.V.B.M. Siqueira, and M.L.C. Vieira.¹ Universidade de São Paulo, Escola Superior de Agricultura "Luiz de Queiroz", Avenida Pádua Dias, 11, 13418-900, Piracicaba, Brazil.

T. de Campos, A.P. de Souza, and E.F. Formighieri. Universidade Estadual de Campinas, Instituto de Biologia, 13083-970, Campinas, Brazil.

L.L. Benchimol, S.A.M. Carbonell, and A.F. Chioratto. Instituto Agrônomo de Campinas, Centro de Pesquisa e Desenvolvimento de Recursos Genéticos Vegetais, Campinas, 13001-970, Brazil.

S.M. Tsai. Universidade de São Paulo, Centro de Energia Nuclear na Agricultura, Piracicaba, Brazil.

¹Corresponding author (e-mail: mlcvieir@esalq.usp.br).

and are distributed throughout the genome. Together, these characteristics make microsatellite loci one of the best genetic markers for mapping purposes (Oliveira et al. 2006).

Microsatellites were developed from both coding and non-coding regions of plant genomes (Brown et al. 1996; Scott 2001). Several resources can be used to search for SSR, including a variety of DNA libraries such as genomic, genomic-enriched for SSR, bacterial artificial chromosome and cDNA libraries, as well as public databases, including expressed sequence tags (ESTs) databases (Scott 2001; Broughton et al. 2003). EST-SSRs have some intrinsic advantages over anonymous markers because they can be easily found by in-silico analysis of EST databases and are associated with a coding gene. In addition, as primers are designed from highly conserved coding regions, i.e., from genes, across-species amplification is expected to be more successful than that obtained from genomic DNA (Varshney et al. 2005). Nevertheless, EST-SSRs are less polymorphic than genomic-SSR (Eujayl et al. 2001).

Beans are the most important grain legumes for human consumption in the developing world. Molecular markers have assisted common bean breeding programs in various ways (Jarne and Lagoda 1996): to investigate the origin and diversity of current cultivars (Gepts 1998), the domestication of *Phaseolus vulgaris* (Koinange et al. 1996), and the genetic control of resistance to important diseases (Nodari et al. 1993; Yu et al. 1998; Miklas et al. 2003). Several *Phaseolus* linkage maps have been developed in recent years (Freyre et al. 1998; Yu et al. 2000), and SSRs have been used mainly for constructing and integrating them (Yu et al. 2000; Tarlan et al. 2001; Blair et al. 2003).

In this study, we explored the Bean EST (BEST) database (Bean EST project; <http://lgm.esalq.usp.br/BEST/>) and a genomic library highly enriched for microsatellites screened for CT and GT sequences (Billotte et al. 1999), used to develop, characterize, and compare the SSR loci generated from both sources. Furthermore, we assessed 80 SSR markers to evaluate the diversity of a collection of *P. vulgaris* accessions, including South American genotypes.

Materials and methods

Plant material and DNA extraction

A set of 23 common bean inbred lines and 2 other related species were used to evaluate allelic variations at 80 SSR loci identified in both kinds of sequences, i.e., genomic and genic: 17 lines of Mesoamerican origin ('Brasil 2', 'Porcillo', 'Rio Tibagi', 'Mulatinho', 'Carioca Comum', 'Puebla 152', 'Great Northern', 'Flor de Mayo', 'Garbancillo', 'Apetito Blanco', 'Jamapa', 'Barbunya', 'IAC-UNA', 'Baetão', 'Sanilac', 'Tu', and 'Bat 93'), 6 lines of South American origin ('Pompadour', 'Antioquia 8', 'Jalo EEP 558', 'Jabola', 'CAL 143', and 'Red Kidney'), and 2 related species, *P. lunatus* ('87 JP 12') and *Vigna* sp. ('Fradinho Cruzeiro'). This set comprises a representative sample of cultivated dry beans and includes the parental lines of 3 mapping populations: 'Bat 93' × 'Jalo EEP558' (Nodari et al. 1993), 'Carioca' × 'Flor de Mayo' (Melo et al. 2002), and 'CAL 143' × 'IAC UNA' (Instituto Agronômico, Campinas, Brazil).

Total genomic DNA was extracted from leaf tissues using

the cetyltrimethylammonium bromide extraction method as described by Doyle and Doyle (1987). DNA was resuspended in Tris-EDTA buffer (10 mmol/L Tris-HCl, pH 8.0; 1 mmol/L EDTA, pH 8.0), and DNA concentrations were estimated by electrophoresis on ethidium-bromide-stained agarose gels using appropriate molecular mass standards.

EST-SSR and genomic-SSR analyses, and primer designing

A unigene collection of 3126 sequences of *P. vulgaris* EST isolated from seedlings inoculated and not inoculated with the fungal pathogen *Colletotrichum lindemuthianum* was used as a source of EST-SSR (Melotto et al. 2005). We searched the collection for the presence of perfect SSR using SSRIT software (Temnykh et al. 2001). The minimum length, i.e., 5 repeat units for di-, tri-, and tetranucleotides, was used as a criterion for SSR mining. SSR motifs were found in 219 unigene sequences. We grouped the dinucleotide SSRs into 4 motifs (AT, AG, AC, and CG) and the trinucleotides into 10 motifs (GAA, AAT, CAT, AAC, TGG, AGG, TAG, TGC, CGG, and ACG). The tetranucleotides were as few as 7 and were grouped in 5 motifs (AAAG, AATC, AATG, AATT, and TGCA). Each motif included all the complementary and permuted sequences for a given simple sequence. These sequences were further classified according to the position of the SSRs within the gene, that is, in the 3' untranslated region (UTR), 5'UTR, or coding sequence (CDS). This analysis was performed by searching the putative start and stop codons in each EST after determining the frame of the transcribed EST. For this purpose, the BLASTx algorithm was used to obtain the best amino acid alignment of the EST with respect to similar sequences deposited in the nonredundant GenBank database.

For primer design, only the EST-SSRs that contained enough nonrepetitive flanking regions was used. PCR primers between 16 and 24 bp in length were designed to flank the SSRs using Primer3 software (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). To standardize the amplification conditions and to decrease the probability of generating spurious PCR fragments, 2 criteria were used to design the primers: a melting temperature ranging from 57 to 63 °C and a maximum difference of 3 °C between the estimated optimal melting temperature for the forward and reverse primer annealing.

An enriched-microsatellite library was screened for 2 repeated sequences, CT and GT. It was constructed using the Brazilian variety 'IAC-UNA', following the protocol of Billotte et al. (1999). Sequencing was performed with T₇ or SP₆ primer sites (all clones were sequenced in both directions) using the Bigdye Terminator Cycle Sequencing Kit (Applied Biosystems, ABI 377 sequencer). Readings were processed by the Phred v. 0.000925.c base-calling program (<http://www.phrap.org>; Ewing et al. 1998), and poly-A/T tails, vector, and adapter sequences were trimmed using cross-match analyses. Clustering was performed using CAP3 software with default parameters (Huang and Madan 1999). Primer designing for genomic-SSR was carried out by means of PrimerSelect (DNASStar Inc.) according to the following conditions: expected size of the amplified product ranging from 150 to 350 pb; GC content between 40% and

60%; melting temperature (T_m) from 45 to 60 °C; primer length ranging from 18 to 22 bp; and no hairpin or dimer formation.

EST-SSR and genomic-SSR assays

The PCR for EST-SSR assays was performed in a final volume of 20 µL in 96-well plates using a Gen Amp[®] PCR System 9700 thermocycler (Applied Biosystems Inc.). Approximately 20 ng of template DNA was mixed in a solution containing 0.25 µmol/L of both forward and reverse primer, 200 µmol/L of each deoxynucleoside triphosphate (dNTP), 2.5 mmol/L MgCl₂, 10 mmol/L Tris-HCl, 50 mmol/L KCl, and 0.5 U Taq DNA polymerase (Invitrogen). Amplification profiles are shown in Table 1.

For the genomic SSR assay, the annealing temperature (T_a) of each microsatellite sequence was identified using a temperature gradient on a PTC-200 thermocycler (MJ Research Inc.). PCRs were performed in a 25 µL final volume containing 50 ng of template DNA, 0.2 µmol/L of each forward and reverse primer, 100 µmol/L of each dNTP, 2.0 mmol/L MgCl₂, 10 mmol/L Tris-HCl, 50 mmol/L KCl, and 0.5 U Taq DNA polymerase (Invitrogen). For the reactions the following conditions were used: 1 min at 94 °C; then, 30 cycles of 1 min at 94 °C, 1 min at T_a , and 1 min at 72 °C, ending with a final extension for 5 min at 72 °C.

Five microlitres of the amplification products, previously denatured with 1× loading buffer (0.2% each bromophenol blue and xylene cyanol, 10 mmol/L EDTA, pH 8.0, and 95% formamide) at 95 °C for 5 min, was loaded on 6% w/v denatured polyacrylamide gels using a 10 bp ladder as a size standard. Silver staining was performed according to Creste et al. (2001).

Data analysis

The polymorphic information content (PIC) was used to measure the informativeness of the 80 SSR loci, according to the equation of Anderson et al. (1993)

$$PIC = 1 - \sum_{i=1}^k P_i^2$$

where k is the total number of alleles detected for a SSR and P_i the frequency of the i th allele detected in the 25 bean genotypes.

SSR data were scored visually, and genetic distances among the genotypes were estimated separately for EST-SSRs and genomic-SSRs by using Wright's distance (Wright 1978) modified from Roger's distance. The genetic distance matrix was estimated using TFPGA v. 1.3 (Miller 1997). Cluster analyses were performed according to the UPGMA method (unweighted pair-group method with arithmetical averages). Trees were drawn using Drawgram with PHYLIP computer package v. 3.6 (Felsenstein 2004). The stability of the clusters was tested by a resampling procedure with 10 000 bootstraps using the BootD program (Coelho 2002).

Results

Frequency and distribution of EST-SSR

Previous studies published by our group (Melotto et al. 2005) identified a set of 240 potential SSR markers in 219

Table 1. Number of cycles and annealing temperature (T_a) of PCR programs used to amplify common bean EST-SSR loci.

	PCR program		
	1	2	3
Stage 1			
Cycles	12	10	10
T_a (°C)	65–53 ^a	55–50 ^b	55–45 ^a
Stage 2			
Cycles	25	25	25
T_a (°C)	53	50	45

Note: An initial denaturing step was conducted at 94 °C for 2 min before stage 1 in all 3 programs. At both stages 1 and 2, denaturing and elongation conditions were held constant at 94 °C for 30 s and 72 °C for 1 min, respectively. The annealing duration in all programs was 45 s and 1 min at stages 1 and 2, respectively.

^aAnnealing temperature was decreased by 1 °C after each cycle.

^bAnnealing temperature was decreased by 0.5 °C after each cycle.

unigene sequences: 137 (57%) were composed of dinucleotide motifs, 96 (40%) of trinucleotide motifs, and 7 (3%) of tetranucleotide motifs. The average length of the repeats was 12.6 bp for dinucleotides, 16.6 bp for trinucleotides, and 21.7 bp for tetranucleotides. The dinucleotide motifs were the most abundant: AG with 80 occurrences (33.3%) followed by AT with 37 occurrences (15.4%). The CG motif never occurred in the sequences (Table 2). The most frequent trinucleotide motif was GAA, with 34 occurrences (14.2%), followed by CAT and TGG with 18 (7.5%) and 10 occurrences (4.2%), respectively.

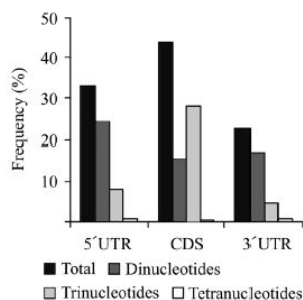
EST-SSR loci were mostly found in UTR (33% and 23% at 5' and 3' ends, respectively). Dinucleotides were found predominantly in UTR (73%), and trinucleotides were preferentially found in coding regions (71%) (Fig. 1).

Here, we designed primers for 156 SSRs. The remaining 86 either had an unsuitable GC content or were not flanked by nonrepeated sequences because they were located at the EST ends. PCR assays were conducted for a subset of 46 loci. Four primer pairs did not yield a PCR product, and 2 primer pairs produced large fragments, probably because of the presence of introns in the amplicon causing problems, as large fragments cannot be accurately scored for small size differences. The remaining 40 EST-SSR markers, which produced fragments of suitable size, were used for further analyses.

A total of 2479 clones were screened from the 'IAC-UNA' library, and 1453 (58.6%) putative positive colonies were isolated, of which 1209 (48.8%) were sequenced. A total of 714 sequences contained microsatellites. The screening of the library showed that 471 sequences contained simple motifs (perfect and imperfect) and 243 composite ones. From Phred/Cap3 analyses, 478 contigs and 648 singlets were obtained. Around 540 sequences were suitable for designing primers as a result of their sequence flanking repeats or complexity. SSR containing GT/CA (38%) or

Table 2. Frequency of individual SSR motifs among 240 microsatellites isolated from the BEST database.

SSR type	Motif	No. of repeats							Total/motif	
		5	6	7	8	9	10–15	16–25	No.	(%)
Dinucleotides	AG	42	15	6	4	5	6	2	80	(33.3)
	AT	20	8	3		1	4	1	37	(15.3)
	AC	17	1	1		1			20	(8.3)
	CG								0	(0.0)
Trinucleotides	GAA	18	11	4			1		34	(14.2)
	CAT	11	6	1					18	(7.5)
	TGG	6	3	1					10	(4.2)
	AAC	6	2						8	(3.3)
	TGC	6	1						7	(2.9)
	AAT	4	1	1					6	(2.5)
	CGG	4							4	(1.7)
	TAG	2	1						3	(1.3)
	AGG	2		1					3	(1.3)
	ACG	3							3	(1.3)
	AAAG	1		1					2	(0.8)
Tetranucleotides	AATC	1							1	(0.4)
	AATG	1	1						2	(0.8)
	AATT	1							1	(0.4)
	TGCA	1							1	(0.4)
	Total	146	50	19	4	7	11	3	240	(100)

Fig. 1. Distribution of bean EST-SSR loci according to their gene position.

GA/CT (22.7%) motifs were the most frequent among simple repeats.

Polymorphism of SSR loci in *P. vulgaris* genotypes

A set of 80 SSR loci, 40 EST-SSRs (Table 3), and 40 genomic-SSRs (Table 4, Genbank accession numbers EF136464 to EF136503) was used to study the allelic diversity in the selected common bean genotypes. Out of those 40 EST-SSR markers, 11 had dinucleotide repeats, 23 had trinucleotide, and 6 had tetranucleotide repeats. A total of 31 EST-SSRs were polymorphic among the 25 common bean genotypes (Table 3). The average number of alleles was 2.7 for the 40 EST-derived loci but 3.1 for the polymorphic ones. Of all the EST-SSR markers, the locus PvM21 revealed the most alleles, i.e., 12 (Fig. 2). For the genomic SSRs, 15 loci had simple dinucleotide repeats and 25 were composite ones. A total of 26 loci were polymorphic among the genomic-SSRs. The greatest number of alleles per locus (7) was detected at the locus FJ55, which amplified 3 dis-

tinct microsatellites: 1 composite between 2 simple ones ((TC)₅-(TC)₁₀(CT)₆-(CT)₁₅). The average number of alleles was 2.4 for the 40 genomic-SSR-derived markers (2 to 7) but 3.2 when only polymorphic markers were considered.

To measure the informativeness of SSR markers, PIC was calculated for each of the 80 loci using the 25 genotypes. PIC values ranged from 0.08 to 0.86, with an average of 0.44 for EST-SSR and from 0.08 to 0.81 with an average of 0.45 for genomic-SSR. Data from genic and genomic SSRs were used for phenetic analysis among the 25 bean lines. The dendrogram (Fig. 3A) obtained from data on 31 polymorphic EST-SSRs first clustered a Mesoamerican *Vigna* species and 18 bean lines, 17 of them belonging to the gene pool of Mesoamerican origin and 'Fradinho Cruzeiro'. The second cluster in the dendrogram comprised 6 lines of Andean common beans and the cultivar '87 JP 12' of *P. lunatus*. Cluster analyses from data on the 26 genomic SSRs and genic SSRs also revealed a clear difference between Andean and Mesoamerican beans (Fig. 3).

Discussion

SSR in expressed sequences

Microsatellites are usually considered as evolutionary neutral DNA markers and selectively neutral sequences randomly or nearly randomly distributed over the euchromatic genome (Schlotterer and Wiehe 1999). However, reports have repeatedly shown that SSRs also exist within genes or functional sequences (Li et al. 2004). Yaish and Pérez de la Vega (2003) isolated a GA SSR that was immediately upstream of the methionine codon of the MADS-box in common bean. Hu et al. (2004) found 394 SSRs in more than 4000 sweet potato ESTs, and recently Peng and Lapitan (2005) identified more than 40 000 SSRs in nearly half a million wheat ESTs. Melotto et al. (2005) identified 240

270

Genome Vol. 50, 2007

Table 3. PvM loci derived from common bean expressed sequence tags (ESTs).

SSR code	Bean unigene	Forward and reverse primer sequences (5'→3')	Repeat motif	Gene location	Predicted/ observed length (bp) ^a	PCR program	PIC	Na
PvM01	TC34	GCTCTAGCCTCAATCTTTGG GGTTCTGCTTCCACATT	(TTC) ₇	5'UTR	240/240	1	0.60	4
PvM02	TC116	CGCCATTGGATTGGATT AGGCGTGGAAAGTGGAGTG	(CTT) ₆	CDS	191/191	1	0.73	6
PvM03	TC116	CCGCTCTCTCTCTCTCTC CGGCGAGTCATCTTTCC	(TTC) ₆	CDS	167/167	1	0.62	4
PvM04	TC323	GGTTCTCTCTCTCTCTGCT GCGCGTCTTTTGGTAGT	(TTC) ₁₀	5'UTR	210/210	2	0.65	3
PvM06	TC103	TTGCCAACAAATGGAAATGA TTGCACAATGAGAAGGAGA	(TAC) ₆	3'UTR	190/170	2	0.15	3
PvM07	TC127	CTGGAATGGCTTACAGATGC CCAAGATCAGCACAGTGAA	(ATG) ₆	CDS	213/213	2	0.46	2
PvM09	PVEPSE3030D08	GCTTGACGCTTTGAACC CGTGGCAITGATTCCTC	(AAG) ₅	CDS	216/216	1	mono	1
PvM11	PVEPSE3030J03	TGGGATTTTGAGCGTGTG TCCACCGACTTACCGAACA	(AGA) ₆	n.d.	154/154	2	0.46	3
PvM12	PVEPSE3030L20	CCAGTGTGTCAACAATGGA AGGCTTAGGGAATGGCTCA	(TGC) ₅	CDS	191/191	2	0.38	3
PvM13a	PVEPSE3030M09	GAGAAGCCGACAGAGAGA AGATGCCGCGAACAGAAC	(GAA) ₆	CDS	216/216	2	0.71	4
PvM13b	PVEPSE3030M09	GAGAAGCCGACAGAGAGA AGATGCCGCGAACAGAAC	(GAA) ₆	CDS	216/240	2	0.73	4
PvM14	PVEPLE1004G06	AAAGCCACAAAACCAATAAAC GCCTCCGAAAGATTCAACG	(AATC) ₅	n.d.	204/204	2	0.61	3
PvM15	PVEPSE2010E01	GGTGGAGCGAGTGGTGT TTCACCTTCCCAAGACCTGA	(TGC) ₅	n.d.	185/185	2	0.46	2
PvM16	PVEPSE2022E07	ATATCGCCACCCACTCCA AACCAGAAAGGCAAAACG	(TTTC) ₇	CDS	179/179	2	0.08	2
PvM17	PVEPSE2030E02	AGGATGGGTTCCGTGCTT GGCAGCTTCCACATCGTC	(ATGA) ₅	CDS	213/213	2	0.40	2
PvM18	PVEPSE3003E05	CAACCATCTCTGTGGCAAG TGAGTTTGGGAGAGTGAGTTT	(AATT) ₅	3'UTR	202/102	2	0.58	3
PvM20	PVEPSE1004B07	CGCTTTGATAAACCGTTGC TGTTGGTGCTAACGATGGAA	(CT) ₇	3'UTR	222/340	2	mono	1
PvM21	TC449	CCCATCTTCCATTCCACAG AAGCCACAGAAACTGAGCA	(AT) ₁₄	CDS	251/251	2	0.86	12
PvM22	PVEPSE2004H10	ACTCTACAATGGCGGAATC GGCGTTTCTCCTCTCTCTT	(TC) ₅	CDS	222/222	2	0.40	3
PvM24	PVEPSE2009B07	CAGAAAAATGGCACAAAG AAAGTCAACGCACACACC	(TG) ₅	3'UTR	345/ 345	2	mono	1
PvM25	TC54	GAAGACGTGAAGGAGCCACT	(TGA) ₇	3'UTR	183/183	2	0.08	2

Table 3 (continued).

SSR code	Bean unigene	Forward and reverse primer sequences (5'→3')	Repeat motif	Gene location	Predicted/observed length (bp) ^a	PCR program	PIC	N _{af} ^a
PvM26	TC80	CAC TTTCAGAGAGAACGCAAT TTCTCCACCCATTAACTGC	(TCA) ₇	CDS	206/206	2	0.08	2
PvM27	TC159	CTCGGACACAAGTCTCCAC GGAGTTGAAAAGGTGGCTA	(TGC) ₅	CDS	192/192	2	mono	1
PvM28	TC390	TGATACAGAAAGCAGGAGAACAA GGAGAGGAAACACGTAATGGA	(TCA) ₅	n.d.	189/189	2	0.37	2
PvM29	TC487	TCGACACAAGAAAGAAAGAAAGAA CAGCAGAACACAGAGTGATTTC	(AAG) ₅	5'UTR	275/275	2	0.08	2
PvM30	TC527	GGTGGCAGGTACGAAAGAGT GCAGACCCCTGTCAACAACAA	(GTT) ₅	3'UTR	222/222	2	0.55	5
PvM32	PVEPSE3030F06	GCCATGAAGCAAGGTGAAGA GAGAAGCCAGGAACCTTA	(ATC) ₅	n.d.	159/159	3	mono	1
PvM33	PVEPSE3029G08	GCATCCCTTTTGTGTC GCCCTCTTTCCAGCAG	(TTG) ₅	CDS	220/220	3	0.08	2
PvM34	PVEPSE3027C10	GTCTATTGTGTTTCTTTTCTC CTTCTCCCTCTTCCCTCT	(GAG) ₅	n.d.	273/273	3	0.58	3
PvM36	PVEPSE3011F10	GGACAGTGATAGTTTCAGATTTC GCCCCCTATTCCCTACCC	(ATC) ₅	CDS	201/201	3	0.47	2
PvM37	PVEPSE3011A04	CATCTTCTTCTCACCCCTTA GGACAGGGGAATCTCACTA	(GAG) ₇	CDS	158/158	3	0.37	2
PvM39	PVEPSE3006G01	CGTGTGCTGGTGGTG ATCTCTTTCGGGAGAGA	(GAT) ₆	CDS	133/133	3	mono	1
PvM40	PVEPSE3001F10	ATTGTGCTTCTCAACG CTCTCCAATTTTGTGTCAG	(CTG) ₆	CDS	202/202	3	0.55	3
PvM41	PVEPSE2029D10	GAACCATCAGAAACTAC AAATGTCCCAAGCCACAC	(CT) ₅	5'UTR	174/300	3	mono	1
PvM42	PVEPSE2027C10	TTCCGTTTACCTTACCCTTC CCATCGTCTTTCTTCAACA	(AG) ₅	3'UTR	255/255	3	0.08	2
PvM44	PVEPSE2024H04	CACGTGCTTGGGCTATC CTCTTACCAGCAGCATCA	(TC) ₈	5'UTR	182/600	3	mono	1
PvM45	PVEPSE2024D09	CTCTTCTCTCTTCTCTCT GAGGGTAGTTCTTGGA	(CT) ₂₄	5'UTR	232/232	3	0.66	4
PvM46	PVEPSE2024B07	GACCAAGTCTCTTTC CGCTTATGTTGGTGGAG	(GA) ₅	CDS	246/246	3	0.47	2
PvM48	PVEPSE2021H06	CGTGGAGGAGAGAGGAGAGA GCAATGGTGGGTGATG	(CT) ₆	n.d.	122/122	3	mono	1
PvM49	PVEPSE2027H04	AGGAGGTGAGGTGGTGA ATCTCTGCTCGGAGTGT	(TC) ₆	5'UTR	197/197	3	0.27	2
		TATCCTTTCGCCCAATCCATA						

Note: Na, number of alleles amplified using 25 bean genotypes; n.d., not determined, because either no hit or low hit in BLASTx.

^aAt least 1 genotype amplified the observed size.

Table 4. FJ loci derived from the enriched-microsatellite library.

SSR code	Forward and reverse primer sequence (5' → 3')	Repeat motif	Predicted length (bp)	Range (bp)	Ta (°C)	PIC	Na
FJ 21/EF136464	TAATACTTGTATCTCCCTTGTG GGCCATGGAAAACCTTGTG	(GT) ₄ G (GT) ₃	293	310	45.0	mono	1
FJ22/EF136465	CACCGCTTCTACCTCCGTCC GGGGCCGCATAGTGATTCTTCT	(CT) ₃ (CA) ₂	276	293	64.8	mono	1
FJ 23/EF136466	TAAATAGCCCTAAAACAAATCC ATATAGTGGCTTCCTTCAAG	(AG) ₃ (TA) ₂	200	200	60.0	mono	1
FJ 24/EF136467	TGCAATCCAAACGAGTGAAG AAATACAGTTGAAATCCTCCAC	(GT) ₃ AT (GT) ₃	290	310	45.0	mono	1
FJ 25/EF136468	AAGCTTTTGGGTTTCTCA TTATCCCAAGGGCTACATTC	(AC) ₇	291	310	45.0	mono	1
FJ 26/EF136469	ATACCAATTTAATGTTCTTTC GAGGTATACGTAGTTTGGAGTT	(CT) ₆	168	190–193	60.0	0.08	2
FJ 27/EF136470	ATGAAAGCGATAGTGATGAGAT CCAGTGGGTAATGAGTTG	(GA) ₁₇	246	230–270	60.0	0.66	4
FJ 28/EF136471	GCGCATTCATCTCATCCATTAT CCAGGCAGAGCAAGCAAGTT	(AC) ₁₁	192	199–207	60.0	0.53	3
FJ 29/EF136472	CATGTTGGCTTAGTGATGTTTC CATGTTGGCTTAGTGATGTTTC	(TA) ₅ (TG) ₈	237	240–269	60.0	0.80	6
FJ 30/EF136473	TTCTCTTCTTCTTCCTTCCTC TTTTCTCTTCCACCTTCTCTA	(GT) ₆ ; (AG) ₅	251	255	60.0	mono	1
FJ 31/EF136474	TTTTTAACATGCTCCCTCCTAC GGTCCACAATCAAGCAGTCAAG	(AC) ₉ (AT) ₄	289	308–312	60.0	0.21	2
FJ 32/EF136475	AGGGCTAGGAAACACCGAAAAAC CCCGCCCATCAGAAAAAG	(TG) ₆	159	159	60.0	mono	1
FJ 33/EF136476	GTGTGATGAGCCAAAGGAAGT GAAACAAAGCTCAAAACAGTAA	(GT) ₆	268	290	60.0	mono	1
FJ 34/EF136477	CTTCTGGGCTTGTTTCA CTATATCAGCCTAGTTATTTTCA	(GT) ₆	271	287–290	60.0	0.44	2
FJ 35/EF136478	TAATCTTAAACAAAAACCACA GGGGATCGAATGAAAAGACT	(CA) ₁₀ (TA) ₁₁	264	300–340	60.0	0.44	2
FJ 36/EF136479	TGAATTTTAAAGTGGCAAGAT CCAAACACAATAGAAATGATAG	(GT) ₅	178	188	45.0	0	1
FJ 37/EF136480	TTATCGCTAGATTCTTGTGTTA GGATGCATCTTTAGGAGTGA	(TA) ₂ (GT) ₆	281	140–152	60.0	0.73	4
FJ 38/EF136481	TTCTCTTCTTCTTCCTTCCTC TTTTCTCTTCCACCTTCTCTA	(GT) ₆	251	260	60.0	mono	1
FJ 39/EF136482	TGCTTACGCTCCAGTCATT TAGCAAACACAGAAACGAAACC	(CA) ₃ (CT) ₂	280	262	60.0	mono	1
FJ 40/EF136483	TCAACATTCTTGGTGTGGTT AATGTTTATATTGGTTCACCTCA	(GT) ₅	288	260	45.0	mono	1
FJ 41/EF136484	CATTCTAGTGGTGTGTTGTGC AATCGGTTGTTTCGCATAA	(CT) ₂ (GT) ₃ ; (AAG) ₂ (GT) ₄	268	278–280	60.0	0.21	2
FJ 42/EF136485	GGAGCGTAAAACACCTT TCGAAAGCTTGAAGACCTTG	(AC) ₅ ; (AC) ₇	202	215–231	60.0	0.15	3
FJ 43/EF136486	GTGTGATGAGCCAAAGGAAGT GAAACAAAGCTCAAAACAGTAA	(TG) ₅	268	300	52.7	mono	1
FJ 44/EF136487	TATCAGCCTAGTTATTTTCAAG CATACTTTTCTTATTTTCTGGA	(CA) ₆	255	260–263	53.7	0.37	2
FJ 45/EF136488	TTGATCATGAGTTCAGAGTCCA GATTGATAGAGATTGCCATAAGTC	(CA) ₆	226	239–240	60.0	0.21	2
FJ 46/EF136489	AGTGGACCACTTCTTCTTTC CAATTTACCAGCTGGACTTTAG	(CT) ₉	263	290–310	60.0	0.29	4
FJ 47/EF136490	ATGCTGGAATGACTGTGATGGA AATGGGTGAAATAAGTGGGTTG	(TC) ₂₀	248	278–312	60.0	0.76	6
FJ 48/EF136491	ATTGAAGTGAGTTATTTTGTGAAG	(TG) ₅	229	250–260	60.0	0.37	2

Table 4 (concluded).

SSR code	Forward and reverse primer sequence (5'→3')	Repeat motif	Predicted length (bp)	Range (bp)	Ta (°C)	PIC	Na
FJ 49/EF136492	CATTCTTTATGGTTTTCCTTAG ATGAGAAAATAACGTGGAGGAA GACGGGATCGGTTTACTGAG	(AG) ₁₀	264	275–280	60.0	0.40	2
FJ 50/EF136493	CATTTTCATGTGATTCTCTTGT ATTGTGCTCTCCTTTGATAA	(CA) ₅	226	228–230	45.0	0.40	2
FJ 51/EF136494	AAGATTTTCCAGCGACCTC TTCACGGATCCAAAGTATTCTC	(CA) ₅	212	213–238	60.0	0.74	6
FJ 52/EF136495	GCTCGTGAATCAAGGTAA CCGAATCGGTAATAGTAAATCC	(CA) ₅	158	158–162	60.0	0.08	2
FJ 53/EF136496	CCGCGGACTACTGTAATCTATG GCAGCAAAACAAAATCAAATC	(CT) ₁₀	174	195–218	45.0	0.65	5
FJ 54/EF136497	AGCATGCCCTGTTCTTGTTCTA TGTGAATTATGGATGTGTTTGT	(TC) ₁₀	248	282–292	60.0	0.57	3
FJ 55/EF136498	CAAACACTGACACACAAACGA AGAGAAATCAAAGCCGAAGGA	(TC) ₅ ; (TC) ₁₀ (CT) ₆ ; (CT) ₁₅	180	144–230	60.0	0.77	7
FJ 56/EF136499	ATTATGCTCTTCTTCTTTGTTT GAAAATATTAGAAAATGCCGC	(CA) ₅	263	266–282	60.0	0.42	3
FJ 57/EF136500	CCTAGGTTCTGCGCTGCTTTTG TCCCCCTCTCATCATCCATAA	(AG) ₅ ; (AGG) ₅	263	227	45.0	mono	1
FJ 58/EF136501	GAGCTTCGGTTCTTTTCTTTG GCAGCCTCCCACATAATAGTAA	(TC) ₈	186	195–198	60.0	0.48	2
FJ 59/EF136502	GAGCGTGTTCGAAGCGAGTT ACCCAGCCGAGACGACAG	(AG) ₅ (GT) ₈ ; (GA) ₅	201	237–240	60.0	0.61	3
FJ 60/EF136503	GGAGGGCATGAGAAGACGACT ACAAGACACGGGCAAAATAAAT	(CT) ₅	134	147–150	45.0	0.40	2

Note: Na, number of alleles amplified using 25 bean genotypes.

SSRs in 5255 common bean ESTs, thus indicating that SSRs also exist in the expressed genome of this crop species.

In the Bean EST database, SSRs were found in 7% of the total unigene sequences (Melotto et al. 2005). The SSR frequency in the Bean EST database is similar to that found in soybeans (5.2%) but lower than in sweet potatoes (9.1%) (Cardle et al. 2000). However, this comparison should be interpreted with caution, as the method for identifying microsatellites may well be different.

Earlier studies on the abundance of different SSR motifs in plant databases indicated that AT was the most common dinucleotide repeat (Lagercrantz et al. 1993). However, our data showed that in the common bean the most frequent dinucleotide repeat was AG. In *Arabidopsis thaliana* and soybean, in fact, expressed sequences had shown a higher frequency of AG repeats than AT repeats (Morgante et al. 2002).

Among the trinucleotidic microsatellites, GAA, CAT, and TGG motifs were the most abundant, accounting for 35.4%, 18.8%, and 10.4% of occurrences, respectively. Wang et al. (1994) identified AAT, AAC, and TGC as the most common in plants, in contrast to our data. However, differences also exist among plant species in terms of the prevalence of particular microsatellite motifs. In cereals, for example, the CGG motif is the most common (Cordeiro et al. 2001; Varshney et al. 2002), whereas in soybean (Gao et al. 2003) and *Arabidopsis* (Cardle et al. 2000) the GAA motif prevails. Our research supports the proposal that the motif

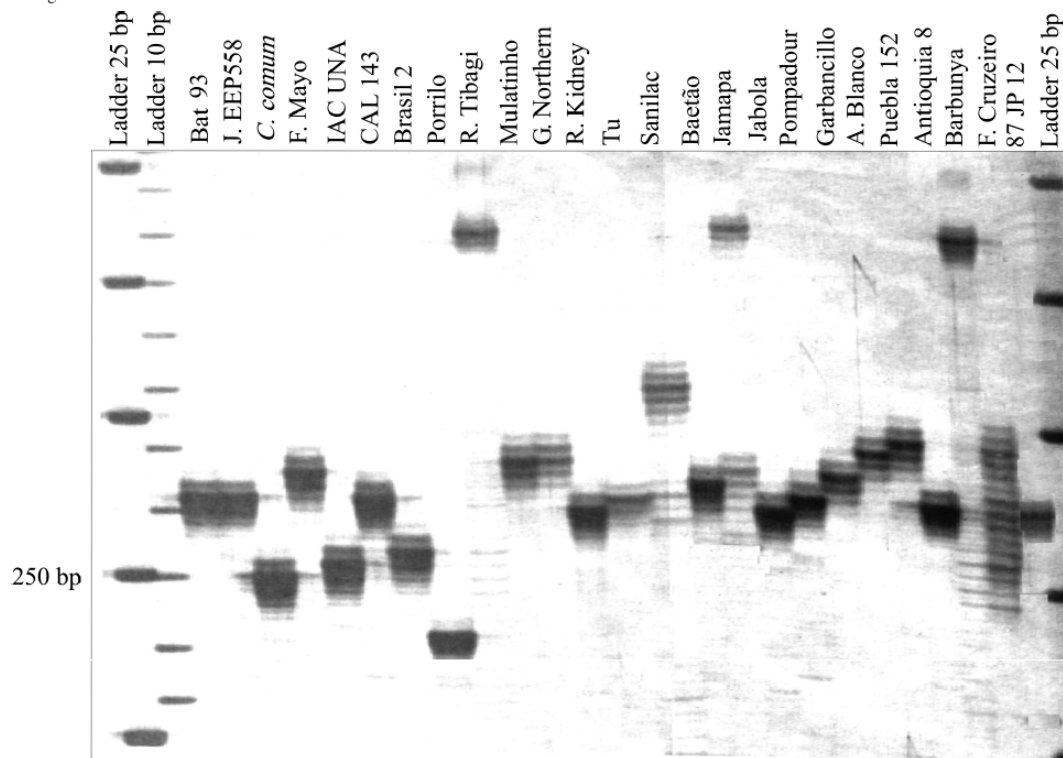
GAA is more prevalent among dicots, as previously reported.

The trinucleotide repeats are seen as the most abundant class of microsatellites in the EST sequences of several species, including the dicots *A. thaliana* and soybean (Morgante et al. 2002; Gao et al. 2003). We found dinucleotide microsatellites as the most abundant in our EST libraries. This can be explained by the different stringencies used for the SSR detection. We used a minimum of 10 bp for the dinucleotides and 15 bp for the trinucleotides. Furthermore, the dinucleotide microsatellites were found mainly in noncoding regions 5'UTR and 3'UTR (Fig. 1). Trinucleotides, in turn, were most abundant in coding portions of the EST (Fig. 1). The preferential occurrence of this type of SSR in coding regions may result from the fact that they do not cause a disruption in the triplet reading frame.

Of the ESTs further characterized in this study, 27 were homologous to known genes, as indicated by previous BLASTx alignments carried out by our group (Melotto et al. 2005). The PVEPSE3003E05 EST is similar to a disease resistance protein family (LRR) from *A. thaliana*. This EST contains a (AATT)₅ microsatellite in the 3'UTR. Another example is the TC127 consensus sequence of a cluster of ESTs, which carries the microsatellite (ATG)₆ and is similar to a salt-tolerance protein from *A. thaliana*.

The presence of SSR sequences in known gene sequences provides an opportunity to study the correlation between repeat numbers and the functional aspects of genes them-

Fig. 2. Display of a *Phaseolus vulgaris* microsatellite marker (PvM21) among 25 bean genotypes detected on silver-stained 6% polyacrylamide gels.



selves. EST-SSR combines the advantages of microsatellite variability with the information potentially carried by expressed sequences. Then, once EST-SSR loci have been mapped, they automatically yield information as to the location of the genes that carry them, thus representing an efficient way of mapping cDNA sequences and increasing the density of gene markers on linkage maps.

SSR polymorphism

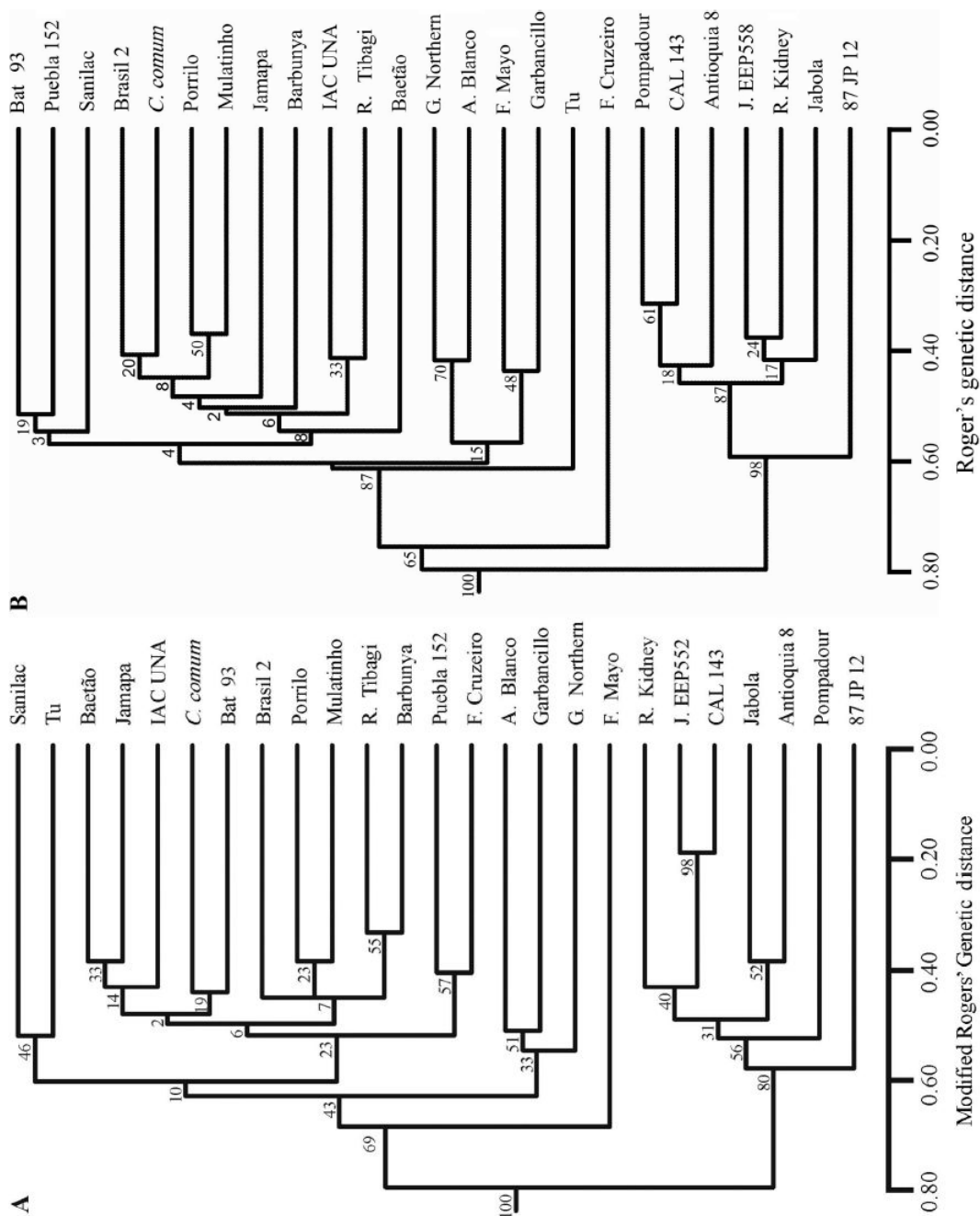
The SSR markers derived from EST are considered less powerful in the discrimination of genotypes than other sources. Eujayl et al. (2001) compared EST- and genomic-SSR markers to investigate genotypic variation of 64 durum wheat lines, landraces, and varieties, obtaining 25% polymorphic loci among 137 EST-SSR markers and 50% among 108 genomic-SSR markers, with an average of 4.1 and 5.0 alleles per locus, respectively. Interestingly, we found similar results at both EST-SSR and genomic-SSR loci. For the EST-SSR markers we found an average of 2.7 alleles per locus and an average PIC value of 0.44, and for the genomic-SSR 2.4 alleles per locus were found, with a mean PIC value of 0.45.

Our EST-SSR data agree with previous findings from common-bean-sequence databases: by using 12 SSR loci de-

veloped from the GenBank database to assess the allelic variation in 12 common bean inbred lines, Yu et al. (1999) found 2.9 alleles per locus. Guerra-Sanz (2004) detected an average of 3.1 alleles per locus at 20 loci derived from public databases assessed in 12 *Phaseolus* accessions. However, the genomic SSRs had been more informative than in the present study. Gaitán-Solís et al. (2002) used 68 primer pairs to investigate the polymorphism among 21 *P. vulgaris* genotypes and found 6 alleles, on average, per primer pair. A similar result was described by Buso et al. (2006), who studied 10 loci isolated from a genomic library enriched for AG in 85 bean accessions. Average values of 7 alleles per locus and PIC of 0.59 were found.

In general, cluster analyses based on SSR data grouped our samples in accordance with the seed size (Fig. 3A and B) in the same way as was previously reported for grouping of common bean races (Singh 2001). The analysis based on EST-derived genetic distances (Fig. 3A) showed a clear distinction between Mesoamerican and Andean materials, both grouped separately with bootstrap support values of 69% and 80%, respectively. 'Flor de Mayo' appeared as an outlier, externally linked to other Mesoamerican beans. 'Bat 93', one of the parents of the bean core map population grouped together with 'Carioca Comum', 'IAC-UNA', 'Ja-

Fig. 3. Relationships among 25 bean genotypes. The unrooted trees ((A) from EST-SSR markers and (B) from genomic SSR markers) were constructed from the modified Rogers' distance matrix using the UPGMA (unweighted paired group method using arithmetic averages) and Drawgram programs of the PHYLIP package (Felsenstein 2004).



mapa', and 'Baetão'. With the exception of 'Tu', which is Andean, the first group consisted of Mesoamerican small-seeded beans, and the second group included Andean large-seeded beans. The Andean 'CAL-143' grouped together with 'J. EEP552', the other parent of the core map population with 98% bootstrap support. 'Fradinho Cruzeiro', a *Vigna* spp. accession, was internally linked to other Mesoamerican beans, as indicated by its small-seeded feature. In addition, the '87 JP 12', which is a *P. lunatus* L., was externally linked to Andean beans.

When the analysis based on genomic-derived genetic distances was performed, the 2-group pattern was maintained (Fig. 3B), though some modifications were noted, such as the clustering of 'Flor the Mayo' with 'Garbancillo', another 'Jalisco' genotype. Interesting groups were formed, such as that between 'Rio Tibagi' and 'IAC-UNA', improved cultivars that have black tegument and medium sized seeds. Some of the white Mesoamerican seeded beans tended to cluster together, e.g., 'Apetito Blanco', 'Garbancillo', and 'Great Northern'. 'Fradinho Cruzeiro' once more grouped with other Mesoamerican small-seeded beans.

Overall, microsatellites were useful for distinguishing Andean and Mesoamerican genotypes, for uncovering seed-size relations, and for separating cultivars from wild accessions. The ability to differentiate Andean and Mesoamerican gene pools was shared by both types of marker developed here, and dendrogram comparison showed only slight differences. 'Jalo EEP558' and 'Bat 93', common beans used as parents of the core mapping population (Nodari et al. 1993), were well differentiated by both sources of microsatellite loci.

In conclusion, we have developed a new set of microsatellite markers based on EST and genomic sequences and showed that the proportion of polymorphic SSRs in expressed regions of the bean genome is high, although the average number of alleles from EST-SSR and genomic-SSR were practically the same. Moreover, microsatellites developed from enriched-genomic libraries seemed to be more informative than EST-derived ones. The increasing availability of bean sequences from EST projects, as the one coordinated by the international consortium *Phaseomics* (Broughton et al. 2003), are valuable sources for SSR marker development. The markers developed here will certainly be useful for genetic mapping and marker-assisted selection programs. They are currently being mapped in an important recombinant inbred line population, 'Bat 93' × 'Jalo EEP558', and 'IAC-UNA' × 'CAL-143', which is part of the genetic material developed in Brazil, the world's main bean producer.

Acknowledgments

The EST data were based on projects supported by FAPESP, Fundação de Amparo à Pesquisa do Estado de São Paulo, Brazil (grants 00/09049-2 and 02/09049-8). This study was supported by FAPESP (Grants 04/13547-9; 02/03225-9; 02/00752-8) and CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico, Brazil (140310/2005-3). We sincerely thank C.A. Oliveira for his technical support, and Ricardo Vieira Cesar for proofreading the manuscript.

References

- Anderson, J.A., Churchill, G.A., Autrique, J.E., Tanksley, S.D., and Sorrells, M.E. 1993. Optimizing parental selection for genetic linkage maps. *Genome*, **36**: 181–186.
- Billotte, N., Lagoda, P.J.L., Risterucci, A.M., and Baurens, C. 1999. Microsatellite-enriched libraries: applied methodology for the development of SSR markers in tropical crops. *Fruits*, **54**: 277–288.
- Blair, M.W., Pedraza, F., Buendia, H.F., Gaitán-Solís, E., Beebe, S.E., Gepts, P., and Tohme, J. 2003. Development of a genome-wide anchored microsatellite map for common bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* **107**: 1362–1374. doi:10.1007/s00122-003-1398-6. PMID:14504741.
- Broughton, W.J., Hernández, G., Blair, M., Beebe, S., Gepts, P., and Vanderleyden, J. 2003. Beans (*Phaseolus* spp.) – model food legumes. *Plant Soil*, **252**: 55–128. doi:10.1023/A:1024146710611.
- Brown, S.M., Hopkins, M.S., Mitchell, S.E., Senior, M.L., Wang, T.Y., Duncan, R.R., et al. 1996. Multiple methods for the identification of polymorphic simple sequence repeats (SSRs) in sorghum [*Sorghum bicolor* (L.) Moench.]. *Theor. Appl. Genet.* **93**: 190–198. doi:10.1007/s001220050265.
- Buso, G.S.C., Amaral, Z.P.S., Brondani, R.P.V., and Ferreira, M.E. 2006. Microsatellite markers for the common bean *Phaseolus vulgaris*. *Mol. Ecol. Notes*, **6**: 252. doi:10.1111/j.1471-8286.2006.01210.x.
- Cardle, L., Ramsay, L., Milbourne, D., Macaulay, M., Marshall, D., and Waugh, R. 2000. Computational and experimental characterization of physically clustered simple sequence repeats in plants. *Genetics*, **156**: 847–854. PMID:11014830.
- Coelho, A.S.G. 2002. BootD: avaliação dos erros associados a estimativas de distâncias/similaridades genéticas através do procedimento de bootstrap com número variável de marcadores. Computer software distributed by the author. Laboratório de Genética Vegetal, Instituto de Ciências Biológicas, Universidade Federal de Goiás, Goiânia, Brasil. [In Portuguese.]
- Cordeiro, G.M., Casu, R., McIntyre, C.L., Manners, J.M., and Henry, R.J. 2001. Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to *erianthus* and *sorghum*. *Plant Sci.* **160**: 1115–1123. doi:10.1016/S0168-9452(01)00365-X. PMID:11337068.
- Creste, S., Tulmann-Neto, A., and Figueira, A. 2001. Detection of single sequence repeat polymorphisms in denaturing polyacrylamide sequencing gels by silver staining. *Plant Mol. Biol. Rep.* **19**: 299–306.
- Doyle, J.J., and Doyle, J.L. 1987. A rapid DNA isolation procedure from small quantities of fresh leaf tissue. *Phytochem. Bull.* **19**: 11–15.
- Eujayl, I., Sorrells, M.E., Baum, M., Wolters, P., and Powell, W. 2001. Assessment of genotype variation among cultivated durum wheat based on EST-SSRs and genomic SSRs. *Euphytica*, **119**: 39–43. doi:10.1023/A:1017537720475.
- Ewing, B., Hillier, L., Wendt, M.C., and Green, P. 1998. Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. *Genome Res.* **8**: 175–185. PMID:9521921.
- Felsenstein, J. 2004. PHYLIP (Phylogeny Inference Package) v. 3.6 [computer program]. Department of Genetics, University of Washington, Seattle. [Distributed by the author.]
- Field, D., and Wills, C. 1996. Long, polymorphic microsatellites in simple organisms. *Proc. Biol. Sci.* **263**: 209–215. doi:10.1098/rspb.1996.0033. PMID:8728984.
- Freyre, R., Skroch, P., Geffroy, V., Adam-Blondon, A.-F., Shimo-hamadali, A., Johnson, W., et al. 1998. Towards an integrated linkage map of common bean. 4. Development of a core map

- and alignment of RFLP maps. *Theor. Appl. Genet.* **97**: 847–856. doi:10.1007/s001220050964.
- Gaitán-Solís, E., Duque, M.C., Edwards, K.J., and Tohme, J. 2002. Microsatellite Repeats in Common Bean (*Phaseolus vulgaris*): isolation, characterization, and cross-species amplification in *Phaseolus* spp. *Crop Sci.* **42**: 2128–2136.
- Gao, L., Tang, J., Li, H., and Jia, J. 2003. Analysis of microsatellites in major crops assessed by computational and experimental approaches. *Mol. Breed.* **12**: 245–261. doi:10.1023/A:1026346121217.
- Gepts, P. 1998. Origin and evolution of common bean: past events and recent trends. *HortScience*, **33**: 1124–1130.
- Guerra-Sanz, J.M. 2004. New SSR markers of *Phaseolus vulgaris* from sequence databases. *Plant Breed.* **123**: 87–89. doi:10.1046/j.0179-9541.2003.00929.x.
- Hedrick, P.W. 2001. Conservation genetics: where are we now? *Trends Ecol. Evol.* **16**: 629–636.
- Hu, J., Nakatani, M., Mizuno, K., and Fujimura, T. 2004. Development and characterization of microsatellite markers in sweetpotato. *Breed. Sci.* **54**: 177–188. doi:10.1270/jsbbs.54.177.
- Huang, X., and Madan, A. 1999. CAP3: A DNA sequence assembly program. *Genome Res.* **9**: 868–877. doi:10.1101/gr.9.9.868. PMID:10508846.
- Jarne, P., and Lagoda, P.J.L. 1996. Microsatellites, from molecules to populations and back. *Trends Ecol. Evol.* **11**: 424–429. doi:10.1016/0169-5347(96)10049-5.
- Koinange, E.M.K., Singh, S.P., and Gepts, P. 1996. Genetic control of the domestication syndrome in common-bean. *Crop Sci.* **36**: 1037–1045.
- Lagercrantz, U., Ellegren, H., and Andersson, L. 1993. The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. *Nucleic Acids Res.* **21**: 1111–1115. doi:10.1093/nar/21.5.1111. PMID:8464696.
- Li, Y.-C., Korol, A.B., Fahima, T., and Nevo, E. 2004. Microsatellites within genes: structure, function, and evolution. *Mol. Biol. Evol.* **21**: 991–1007. PMID:14963101.
- Melo, L.C., Santos, J.B., and Ferreira, D.F. 2002. Mapeamento de QTLs para reação ao oídio e mancha-angular do feijoeiro-comum em diferentes locais. *Pesqui. Agropecu. Bras.* **37**: 1115–1126.
- Melotto, M., Monteiro-Vitorello, C.B., Bruschi, A.G., and Camargo, L.E.A. 2005. Comparative bioinformatic analysis of genes expressed in common bean (*Phaseolus vulgaris* L.) seedlings. *Genome*, **48**: 562–570. doi:10.1139/g05-010. PMID:16121253.
- Miklas, P.N., Coyne, D.P., Grafton, K.F., Mutlu, N., Reiser, J., Lindgren, D.T., and Singh, S.P. 2003. A major QTL for common bacterial blight resistance derives from the common bean great northern landrace cultivar Montana No. 5. *Euphytica*, **131**: 137–146. doi:10.1023/A:1023064814531.
- Miller, M.P. 1997. Tools for population genetic analysis (TFPGA), version 1.3: a Windows program for the analysis of allozyme and molecular population genetic data. Computer software distributed by the author. Available from <http://herb.bio.nau.edu/~miller> [accessed January 2006].
- Morgante, M., Hanafey, M., and Powell, W. 2002. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nat. Genet.* **30**: 194–200. doi:10.1038/ng822. PMID:11799393.
- Morgante, M., and Olivieri, A.M. 1993. PCR-amplified microsatellites as markers in plant genetics. *Plant J.* **3**: 175–182. doi:10.1046/j.1365-3113.1993.t01-9-00999.x. PMID:8401603.
- Nodari, R.O., Tsai, S.M., Yu, K.F., Park, S.J., Zhang, B.L., Haffner, M., et al. 1993. Towards an integrated linkage map of common bean. 3. Mapping genetic factors controlling host-bacteria interactions. *Genetics*, **134**: 341–350. PMID:8514141.
- Oliveira, E.J., Pádua, J.G., Zucchi, M.I., Vencovsky, R., and Vieira, M.L.C. 2006. Origin, evolution and genome distribution of microsatellites. *Genet. Mol. Biol.* **29**: 294–307.
- Peng, J.H., and Lapitan, N.L.V. 2005. Characterization of EST-derived microsatellites in the wheat genome and development of eSSR markers. *Funct. Integr. Genomics*, **5**: 80–96. doi:10.1007/s10142-004-0128-8. PMID:15650880.
- Powell, W., Machray, G.C., and Provan, J. 1996. Polymorphism revealed by simple sequence repeats. *Trends Plant Sci.* **1**: 215–222.
- Schlotterer, C., and Wiehe, T. 1999. Microsatellite, a neutral marker to infer selective sweeps. In *Microsatellites: evolution and applications*. Edited by D.B. Goldstein and C. Schlotterer. Oxford University Press, Oxford. pp. 238–247.
- Scott, K.D. 2001. Microsatellites derived from ESTs, and their comparison with those derived by other methods. In *Plant genotyping: the DNA fingerprinting of plants*. Edited by R.J. Henry. CAB International, Wallingford. pp. 255–237.
- Singh, S.P. 2001. Broadening the genetic base of common bean cultivars: a review. *Crop Sci.* **41**: 1659–1675.
- Tarlan, B., Michaels, T.E., and Pauls, K.P. 2001. Mapping genetic factors affecting the reaction to *Xanthomonas axonopodis* pv. *phaseoli* in *Phaseolus vulgaris* L. under field conditions. *Genome*, **44**: 1046–1056. PMID:11768208.
- Temnykh, S., DeClerck, G., Lukashova, A., Lipovich, L., Cartin-hour, S., and McCouch, S. 2001. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. *Genome Res.* **11**: 1441–1452. doi:10.1101/gr.184001. PMID:11483586.
- Tóth, G., Gaspari, Z., and Jurka, J. 2000. Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res.* **10**: 967–981. doi:10.1101/gr.10.7.967. PMID:10899146.
- Varshney, R.K., Thiel, T., Stein, N., Langridge, P., and Graner, A. 2002. In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species. *Cell. Mol. Biol. Lett.* **7**: 537–546. PMID:12378259.
- Varshney, R.K., Graner, A., and Sorrells, M.E. 2005. Genic microsatellite markers in plants: features and applications. *Trends Biotechnol.* **23**: 48–55. doi:10.1016/j.tibtech.2004.11.005. PMID:15629858.
- Yaish, M.W.F., and Pérez de la Vega, M. 2003. Isolation of (GA)n microsatellite sequences and description of a predicted MADS-box sequence isolated from common bean (*Phaseolus vulgaris* L.). *Genet. Mol. Biol.* **26**: 337–342.
- Yu, Z.H., Stall, R.E., and Vallejos, C.E. 1998. Detection of genes for resistance to common bacterial blight of beans. *Crop Sci.* **38**: 1290–1296.
- Yu, K., Park, S.J., and Poysa, V. 1999. Abundance and variation of microsatellite DNA sequences in beans (*Phaseolus* and *Vigna*). *Genome*, **42**: 27–34. doi:10.1139/gen-42-1-27.
- Yu, K., Park, S.J., Poysa, V., and Gepts, P. 2000. Integration of simple sequence repeat (SSR) markers into a molecular linkage map of common bean (*Phaseolus vulgaris* L.). *J. Hered.* **91**: 429–434. doi:10.1093/jhered/91.6.429. PMID:11218079.
- Wang, Z., Weber, J.L., Zhong, G., and Tanksley, S.D. 1994. Survey of plant short tandem repeats. *Theor. Appl. Genet.* **88**: 1–6.
- Wright, S. 1978. *Evolution and the genetics of populations*, Vol. 4. Variability within and among natural populations. University of Chicago Press, Chicago.

6. Artigo III

“Microsatellites for genetic studies and breeding programs in common bean”

Tatiana de Campos, Luciana Lasry Benchimol, Alisson Fernando Chioratto, Sérgio Augusto Morais Carbonell, Eduardo Fernando Formighieri, e Anete Pereira de Souza.

Publicado na revista
Pesquisa Agropecuária Brasileira (42:589-592; 2007).

Notas Científicas

Microsatellites for genetic studies and breeding programs in common bean

Tatiana de Campos⁽¹⁾, Luciana Lasry Benchimol⁽²⁾, Sérgio Augusto Moraes Carbonell⁽²⁾, Alisson Fernando Chioratto⁽²⁾, Eduardo Fernandes Formighieri⁽³⁾ and Anete Pereira de Souza⁽¹⁾

⁽¹⁾Universidade Estadual de Campinas (Unicamp), Centro de Biologia Molecular e Engenharia Genética, Caixa Postal 6.010, CEP 13083-970 Campinas, SP, Brazil. E-mail: tatyuni@unicamp.br, anete@unicamp.br ⁽²⁾Instituto Agronômico, Centro de Pesquisa e Desenvolvimento de Recursos Genéticos Vegetais, Fazenda Santa Elisa, Caixa Postal 28, CEP 13001-970 Campinas, SP, Brazil. E-mail: llasry@iac.sp.gov.br, carbonel@iac.sp.gov.br, afchioratto@iac.sp.gov.br ⁽³⁾Unicamp, Dep. de Genética e Evolução, Caixa Postal 6.109, CEP 13083-970 Campinas, SP, Brazil. E-mail: eduformi@lge.ibi.unicamp.br

Abstract – Twenty microsatellite loci were identified and characterized in common bean. Microsatellites were tested in 14 genotypes. The allele number ranged from 1 to 3, and the polymorphism information content (PIC) was between 0.14 and 0.65. These polymorphic markers are available to be used for breeding programs.

Index terms: *Phaseolus vulgaris*, SSR, molecular markers.

Microssatélites para estudos genéticos e programas de melhoramento em feijoeiro

Resumo – Vinte locos de marcadores microssatélites foram identificados e caracterizados em feijoeiro. Os microssatélites foram testados em 14 genótipos. O número de alelos variou entre 1 e 3, e o conteúdo informativo de polimorfismo (PIC) entre 0,14 e 0,65. Esses marcadores polimórficos estão disponíveis para serem usados em programas de melhoramento.

Termos para indexação: *Phaseolus vulgaris*, SSR, marcadores moleculares.

Common bean (*Phaseolus vulgaris* L.) is the principal leguminosae used for human nourishment, mainly in South America and Africa, where it represents an important source of protein. Brazil detains the most important productivity and consumption of common bean in the world. However, this crop has not reached high productivity (Silva, 2000). Common bean occurred in two centers of origin in South and Central America, comprising two major gene pools, Andean or large-seeded bean type, and Mesoamerican or small-seeded type (Gepts & Debouck, 1991).

Microsatellites or simple sequence repeats (SSR) are polymerase chain reaction (PCR) based markers developed, for a wide range of plant species, around short segments of DNA, in which a specific motif of one to six nucleotides is repeated in tandem and distributed over the euchromatic part of the genome (Morgante & Olivieri, 1993; Powell et al., 1996). In common bean there are about 200 available SSR markers (Yu et al., 2000; Gaitán-Solís et al., 2002;

Métais et al., 2002; Yaish & Pérez de la Vega, 2003; Buso et al., 2006), a small number when compared with other crops, like soybean (Cregan et al., 1999).

Microsatellite markers are powerful tools but their development is expensive and labour intensive. Consequently, many researchers have tried to use primer pairs developed for one species in another (Cipriani et al., 1999), attributing knowledge as transferability or cross-species amplification.

The variety IAC-UNA was used to construct a microsatellite enriched library for two dinucleotide repeat sequences (CT and GT). This enrichment was based on the procedure described by Billotte et al. (1999). IAC-UNA is a black-seeded variety, developed by Agronomic Institute (IAC, Campinas, SP, Brazil), resistant to anthracnose and susceptible to bean rust, *Fusarium* and angular leaf spot.

The genomic DNA was extracted from leaf tissue using a CTAB extraction method, as described by Hoisington et al. (1994). The extracted DNA was

digested with *RsaI* restriction enzyme, and the digested fragments were linked to *RsaI* adapters. The library was enriched for dinucleotide sequences using (CT)₈ and (GT)₈ biotinylated microsatellite primers with labelled probes that were bound to Streptavidine MagneSphere Paramagnetic Particles as described by manufacturer.

Selected fragments were amplified by PCR using primer sequences complementary to the adapters, and then, attached to the vector pGEM-T (Promega). Plasmids were introduced into XL-1 Blue cells. Transformed cells were cultivated on agar plates containing 100 µg mL⁻¹ of ampicilin and 50 µg mL⁻¹ of X-galactosidase. Single white colonies were transferred onto microplates for long-term storage at 80°C. Sequencing of the inserts was performed using the ABI 377 Big Dye Terminator and 20 sequences containing microsatellites were selected to have primer pairs designed.

Primer designing was performed using Primer Select with the following conditions: amplification DNA size from 150 pb to 350 pb; GC content between 40–60%; temperature annealing (*Ta*) between 45 and 60°C; primer length between 18 and 22 pb; no hairpins or dimmers.

Twenty primers were designed for microsatellite loci, and were selected for characterization using 14 accessions from the IAC Germplasm Bank, including Andean (A) and Mesoamerican (M) gene pools. Total DNA was extracted from the following accessions: 'Sanilac' (M), 'Baetão' (M), 'Red Kidney' (A), 'Jamapa' (M), 'Flor de Mayo' (M), 'Tu' (M), 'Carioca Comum' (M), 'Jabola [CB]' (A), 'Fradinho Cruzeiro' (*Vigna* spp.), '87-JP-12' (A), 'BAT-93' (M), 'Jalo EEP-558' (A), 'IAC-UNA' (M), 'CAL-143' (A).

PCR reactions were carried out in a total volume of 25 µL containing 50 ng of template DNA, 0,8 µM of forward and reverse primer, 100 µM of each dNTP, 1,5 mM MgCl₂, 10 mM Tris-HCl, 50 mM KCl, and 0.5 U *Taq* DNA Polymerase (Invitrogen). Reactions were performed using the following conditions: 1 min at 94°C; then, 30 cycles of [1 min 94°C, 1 min at specific *Ta*, 1 min at 72°C], followed by 5 min at 72°C. Amplification products were checked by electrophoresis on 3% agarose gels, and then loaded on 6% w/v denatured polyacrylamide gels using a 10 bp ladder (Invitrogen) as a size standard. After each run, gels were silver stained according to Creste et al. (2001).

The polymorphism information content (PIC) value was calculated by the following formula:

$$PIC = 1 - \sum_{i=1}^n f_i^2 - \sum_{i=1}^{n-1} \sum_{j=i+1}^n 2 f_i^2 f_j^2$$

in which *n* is the number of alleles; *f_i* and *f_j* are the frequencies of the *i*th and *j*th alleles, respectively (Botstein et al., 1980). Polymorphisms were observed among the amplified alleles.

Fifteen loci were polymorphic, four were monomorphic and one was unable to amplify among the twenty analyzed microsatellites (Table 1). Thirteen primer pairs were perfect dinucleotide motifs, one was imperfect (*FJ 20*), and six presented compound motifs. Two loci could separate Andean from Mesoamerican accessions; showing two different alleles, each one associated to a genetic pool (*FJ 05* and *FJ 17*). These SSRs would be tested in other genotypes, considering that they can be a tool to separate in domestication centers.

The number of alleles ranged from one to three, with an average of 2.07 alleles per locus. Yu et al. (2000) found an allele range that varied from 2 to 10, for a 24 polymorphic SSR evaluation in 12 genotypes. PIC values ranged from 0.14 to 0.65. The highest PIC (0.65) was found in *FJ 14*, that presented the greatest repetition motif, (GA)₁₀, and the most elevated number of alleles (3). Métais et al. (2002) published a range of 0.12 to 0.72 for PIC values, with an average of 0.44, when evaluating 15 polymorphic SSRs in 45 different bean lines of nine different quality types.

The primer pairs were also used to amplify one genotype of *Vigna* spp., accession 'Fradinho Cruzeiro', and one genotype of *Phaseolus lunatus*, accession 87-JP-12, using the same conditions optimized for *Phaseolus vulgaris* accessions. The efficiency of heterologous amplification was 100%. However, it would be necessary a resequencing of the amplification products to check if they were really related to the original sequence from which primer pairs were developed. Even so, this shows a considerable level of sequence conservation within the primer regions flanking microsatellite loci. These results suggest that the new microsatellites reported in this paper could be used for synteny studies, establishing the conservation of genes between species such as *Phaseolus lunatus* and *Vigna* spp.

In common bean, most of the molecular markers used for genetic studies and breeding programs are RAPD

Table 1. Primer sequences and characteristics of 20 common bean (*Phaseolus vulgaris*) microsatellite markers, obtained from variety IAC-UNA, tested on 14 accessions, comprising Mesoamerican and Andean gene pools.

Locus/ GeneBank accession	Primers (5'→3')	Repeat motif	Na ⁽¹⁾	Ta ⁽²⁾ (°C)	Size of cloned allele (size range) in bp	PIC ⁽³⁾
FJ 01/ DQ469376	GTCGCCGCTACTTCTTTGTT TTTAAATGTTGTGGGAGTGATG	(AC) ₇	2	60	270 (265-270)	0.51
FJ 02/ DQ469377	GGTCCACAATCAAGCAGTCA TATGGAACCTGATAGCAAGTG	(AC) ₁₀ (AT) ₅	2	51,4	251 (258-260)	0.52
FJ 03/ DQ469378	TTCGCGAGCAAGCAACTA TGAATGTTTTAAATGCGTTGAA	(GT) ₆	1	45	178 (177)	0.00
FJ 04/ DQ469379	ATAGATGAAGGATTGGGAGAG GGGAAATTGAAGAGGAGATAC	(AG) ₈	1	45	216 (218)	0.00
FJ 05/ DQ469380	AAGAAACAGAAACAATAAAAC TTTCCATTATTTCAGTCACA	(CT) ₂ (CA) ₆	2	60	212 (220-222)	0.48
FJ 06/ DQ469381	TTGGAACACCGTGGAATGGA GAGGCTTTAGACGTTGGAGACA	(GT) ₇	0	60	152 (-) ⁽⁴⁾	0.00
FJ 07/ DQ469382	GAAAACGCGAAACAACCGA ATGTCTCCAAATCCCAAGTG	(CA) ₈	2	60	290 (283-297)	0.52
FJ 08/ DQ469383	ATGGTCATGGTATCAGTTCA TCTTTTCCATAGTATTCTCTG	(CA) ₆ (TA) ₃	1	60	195 (195)	0.00
FJ 09/ DQ469384	ACCTTAGATAGTGTGTTAGAG CATGACACCTAGGGCAAA	(TG) ₆	1	45	155 (153)	0.00
FJ 10/ DQ469385	AGGGGAGTTGTGTTCTTAC ATACGTACGAGTGACTGGAGA	(TG) ₆	2	45	209 (207-209)	0.14
FJ 11/ DQ469386	AAAAGGATCAAAGAGGAGAAAT GGGCAAGTAAAGCTAAACGAG	(CA) ₅	2	60	297 (308-310)	0.25
FJ 12/ DQ469387	TATCAGCCTAGTTATTTCAAG CATACTTTCTTATTTCTGGA	(CA) ₇	2	60	256 (255-258)	0.25
FJ 13/ DQ469388	TCGATGCAGGATTGGATT CAGGTTGATTGTGATAGGTTAC	(AC) ₉	2	60	266 (267-269)	0.48
FJ 14/ DQ469389	TTCATGGCAAGGTAAGTAAATA TGAATGAACACAACACAACAA	(AG) ₁₀	3	60	148 (145-155)	0.65
FJ 15/ DQ469390	AGAATGGAGGGAAAAGCAAAAG CCGAAGTCCAAGATTAGAAGCC	(ATGAG) ₄ (GT) ₃	2	60	191 (200-205)	0.14
FJ 16/ DQ469391	TGGTGCTACAACAAAAGAGAAT TAGGCATGTGGGTAGGTCAG	(TA) ₆ (TG) ₇	2	60	284 (280-300)	0.20
FJ 17/ DQ469392	TCCCGATTATAGTTCTCATTT AGGGACCTCCTTCATCTC	(TG) ₈ (TA) ₃	2	60	222 (220-230)	0.48
FJ 18/ DQ469393	CATTGAGATTGAGGTTTCGTT AGGTATTTCATCGTGCTTTTC	(TG) ₅	2	60	224 (224-230)	0.48
FJ 19/ DQ469394	ATGTTAGTGCCTTATTTCTCT AAGGTAGGGTTGGGATTGT	(CA) ₇	2	60	205 (210-222)	0.51
FJ 20/ DQ469395	TTGGAACACCGTGGAATGGA GAGGCTTAGACGTTGGAGACA	(AG) ₃ AA (AG) ₃	2	60	251 (250-263)	0.51

⁽¹⁾Number of alleles. ⁽²⁾Temperature annealing. ⁽³⁾Polymorphism information content. ⁽⁴⁾No amplification.

markers, which are dominant, and not reproducible. SSR are codominant, more polymorphic and stable. These new informative microsatellites are an available tool in common bean research. They consist in important source of polymorphism which can be used in breeding programs or in genetic studies, as genetic and QTL mapping, marker-assisted selection and germplasm characterization in common bean.

Acknowledgements

To Fapesp, for financial support and fellowships; to CNPq, for fellowships.

References

BILLOTTE, N.; LAGODA, P.J.L.; RISTERUCCI, A.M.; BAURENS, C. Microsatellite-enriched libraries: applied

- methodology for the development of SSR markers in tropical crops. **Fruits**, v.54, p.277-288, 1999.
- BOTSTEIN, D.; WHITE, R.L.; SKOLNICK, M.; DAVIS, R.W. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. **American Journal of Human Genetics**, v. 32, p.314-331, 1980.
- BUSO, G.S.C.; AMARAL, Z.P.S.; BRONDANI, R.P.V.; FERREIRA, M.E. Microsatellite markers for the common bean *Phaseolus vulgaris*. **Molecular Ecology Notes**, v.6, p.252-254, 2006.
- CIPRIANI, G.; LOT, G.; HUANG, W.G.; MATARAZZO, M.T.; PETERLUNGER, E.; TESTOLIN, R. AC/GT and AG/CT microsatellite repeats in peach [*Prunus persica* (L.) Batsch]: isolation, characterisation and cross-species amplification in *Prunus*. **Theoretical and Applied Genetics**, v.99, p.65-72, 1999.
- CREGAN, P.B.; JARVIK, T.; BUSH, A.L.; SHOEMAKER, R.C.; LARK, K.G.; KAHLER, A.L.; KAYA, N.; VAN TOAI, T.T.; LOHNES, D.G.; CHUNG, J.; SPECHT, J.E. An integrated genetic linkage map of the soybean genome. **Crop Science**, v.39, p.1464-1490, 1999.
- CRESTE, S.; TULMANN-NETO, A.; FIGUEIRA, A. Detection of single sequence repeat polymorphisms in denaturing polyacrylamide sequencing gels by silver staining. **Plant Molecular Biology Reporter**, v.19, p.299-306, 2001.
- GAITÁN-SOLÍS, E.; DUQUE, M.C.; EDWARDS, K.J.; TOHME, J. Microsatellite repeats in common bean (*Phaseolus vulgaris*): isolation, characterization and cross-species amplification in *Phaseolus* ssp. **Crop Science**, v.42, p.2128-2136, 2002.
- GEPTS, P.; DEBOUCK, D. Origin, domestication, and evolution of the common bean (*Phaseolus vulgaris* L.). In: SCHOONHOVEN, A. van; VOYSEST, O. (Ed.). **Common beans: research for crop improvement**. Cali: CIAT, 1991. p.7-53.
- HOISINGTON, D.; KHAIRALLAH, M.; GONZÁLEZ-DE-LEÓN, D. **Laboratory protocols: CIMMYT applied molecular genetics laboratory**. 2nd ed. Mexico: CIMMYT, 1994. 88p.
- MÉTAIS, I.; HAMON, B.; JALOUZOT, R.; PELTIER, D. Structure and level of genetic diversity in various bean types evidenced with microsatellite markers isolated from a genomic enriched library. **Theoretical and Applied Genetics**, v.104, p.1346-1352, 2002.
- MORGANTE, M.; OLIVIERI, A.M. PCR-amplified microsatellites as markers in plant genetics. **Plant Journal**, v.3, p.175-182, 1993.
- POWELL, W.; MACHRAY, G.C.; PROVAN, J. Polymorphism revealed by simple sequence repeats. **Trends in Plant Science**, v.1, p.215-222, 1996.
- SILVA, M.V. da. **Identificação de marcador RAPD ligado ao alelo co-7 de resistência do feijão ao agente causal da antracnose**. 2000. 41p. Dissertação (Mestrado) - Universidade Federal de Lavras, Lavras.
- YAISH, M.W.F.; PÉREZ DE LA VEGA, M. Isolation of (GA)_n microsatellite sequences and description of a predicted MADS-box sequence isolated from common bean (*Phaseolus vulgaris* L.). **Genetics and Molecular Biology**, v.26, p.337-342, 2003.
- YU, K.; PARK, S.J.; POYSA, V.; GEPTS, P. Integration of simple sequence repeat (SSR) markers into a molecular linkage map of common bean (*Phaseolus vulgaris*). **Journal of Heredity**, v.91, p.429-434, 2000.

Received on October 19, 2006 and accepted on February 14, 2007

7. Artigo IV

“A linkage map for common bean (*Phaseolus vulgaris* L.) based on genic and genomic microsatellite markers”

Tatiana de Campos, Paula Rodrigues Oblessuc, Danilo Augusto Sforça, Juliana Morini Kupper Cardoso, Renata Moro Baroni, Adna Cristina Barbosa de Sousa, Luciana Lasry Benchimol, Alisson Fernando Chioratto, Sérgio Augusto Morais Carbonell, Antônio Augusto Franco Garcia, e Anete Pereira de Souza.

Artigo a ser submetido à revista

Theoretical and Applied Genetics (2009).

A linkage map for common bean (*Phaseolus vulgaris* L.) based on genic and genomic microsatellite markers

Tatiana de Campos^{1,2}, Paula Rodrigues Oblessuc^{1,2,3}, Danilo Augusto Sforça^{1,2}, Juliana Morini Kupper Cardoso³, Renata Moro Baroni³, Adna Cristina Barbosa de Sousa^{1,2}, Luciana Lasry Benchimol³, Sérgio Augusto Moraes Carbonell³, Alisson Fernando Chioratto³, Antônio Augusto Franco Garcia⁴, Anete Pereira de Souza^{1,2}.

¹Centro de Biologia Molecular e Engenharia Genética (CBMEG), Universidade Estadual de Campinas (UNICAMP), Cidade Universitária Zeferino Vaz, CP 6010, CEP 13083-970, Campinas-SP, Brazil.

²Departamento de Genética e Evolução – Instituto de Biologia (IB), UNICAMP, Campinas-SP, Brazil.

³Centro de Pesquisa e Desenvolvimento de Recursos Genéticos Vegetais, Fazenda Santa Elisa, IAC, Campinas-SP, Brazil.

⁴Departamento de Genética, Escola Superior de Agricultura Luiz de Queiroz (ESALQ), Universidade de São Paulo (USP), CP 83, CEP 13400-970, Piracicaba-SP, Brazil.

*To whom correspondence should be addressed: Centro de Biologia Molecular e Engenharia Genética (CBMEG), Universidade Estadual de Campinas (UNICAMP), Cidade Universitária Zeferino Vaz, 13083-085, C.P. 6010, Campinas/SP, Brasil.

E-mail: anete@unicamp.br

Fax: (55-19) 3521-1089

Phone: (55-19) 3521-1132 or 3521-1156

Abstract

A genetic linkage map for common bean (*Phaseolus vulgaris* L.) is a valuable tool to breeding programs. The aim of this study was to develop a new genetic linkage map using genic and genomic microsatellite markers and some phenotypic characteristics. A large based mapping population was generated from IAC-UNA x CAL143, consisting of 380 recombinant F₁₀ lines. A total of 871 microsatellites were screened for polymorphism among the parents, and 247 polymorphic markers were found. Morphological traits were also used to extend the map cover. A linkage map was obtained with 198 mapped microsatellites in 11 linkage groups. Three phenotypic markers were mapped: the *V*, *Ct* and *fin* loci. The map has length of 1865.9 cM and an average of 9.4 cM distance between markers.

Keywords: Microsatellites, SSRs, Common bean, *Phaseolus vulgaris* L., Genetic Maps, Phenotypic markers.

Introduction

Common bean (*Phaseolus vulgaris* L.) is a major staple food crop in Africa and Latin America. It occupies the first position on commodities production rank in Brazil, together with coffee, mate, orange, papaya and sugar cane (FAO, 2005). Beans are valued as one of the primary protein source in the diet of the low-income population in developing countries. Moreover, it is a legume favorable to genetics studies because of some characteristics of model plants, such as a small genome content of 637 to 675 Mb per haploid genome (Arumuganathan & Earle, 1991; Guo *et al.*, 2007), to be a true diploid ($2n = 2x = 22$), and to have a self-pollinating reproductive system.

The cultivated bean is morphologically diverse, with variation in many traits, like growth habit, pigmentation, pod, seed and phenology (Singh *et al.*, 1991). The morphological diversity contradicts to the narrow based genetic background in domesticated beans. Less than 5% of the genetic diversity available has been used globally, despite nearly a century of organized bean improvement (Broughton *et al.*, 2003). Several studies have suggested that wild bean populations present greater genetic variability than cultivated beans, due to the founder effect during the domestication process. The domestication syndrome has modulated the genetic variability to exclude the diversity for adaptative conditions and to fix traits selected to consumer's preferences (Romero and Bliss, 1985; Koinange, 1996; Gepts *et al.*, 1999).

The seed coat pattern and color in common bean are controlled by genes that exhibit epistatic interactions (McClean *et al.*, 2002). The flower standards are traits usually used for differentiate superior plants and they can be used like

morphological markers. Morphological traits are used as markers when present qualitative variation and discrete phenotypic segregation classes.

Molecular markers have been used in many crops. The advance of genetic improvement is sustained by the development of new markers which are tools for performing marker-assisted selection (MAS). One example in markers technology is the construct of genetic maps. Genetic maps are an essential tool for several studies. The localization of genome regions that express phenotypic characteristics, sinteny studies, cloning genes based on genetic maps are complements important to be considered in breeding programs. Linkage maps can be used not only to investigated genetic control of simple traits, but also of quantitative traits and their linkage relationships. An important prerequisite for the use of genetic resources is the knowledge of genetic variability accessed by adequate molecular markers for the study.

Microsatellites or SSR ("Simple Sequence Repeats", Tautz & Renz, 1984) are one of the most indicated techniques to study polymorphism between DNA sequences. This molecular marker is based on PCR reaction to detect loci variations of repetitive sequences. They present high levels of polymorphism, co-dominance, multiallelism, and mendelian pattern. Microsatellites have been used in several genetic analysis in mammals (Moore *et al.*, 1998), insects (Harper *et al.*, 2003), birds (Primmer *et al.*, 1997), fishes (Rico *et al.*, 1993) and plants (Condit & Hubbell, 1991; Yang *et al.*, 1994; Yu *et al.*, 2000). Microsatellites are recently available for common bean (Yu *et al.*, 1999; Gaitan-Sólis *et al.*, 2002; Métais *et al.*, 2002; Blair *et al.*, 2003; 2008; Buso *et al.*, 2006; Hanai *et al.*, 2007; Benchimol *et al.*, 2007; Campos *et al.*, 2007; Cardoso *et al.*, 2008), results of researches efforts to develop genetic knowing about the crop. When microsatellites are derived from

ESTs, they can be associated with genes of known function and used as functional markers, tagging genes of interest in a more efficient manner (Hackauf & Wehling, 2002).

Genetic mapping in common bean has started based on morphologic markers and isozymes (Basset, 1991). The maps developed by Vallejos *et al.* (1992), Gepts *et al.* (1993), Nodari *et al.* (1993), and Adam-Blondon *et al.* (1994) were first comprised of restriction length polymorphism (RFLPs) markers. A core map was established in common bean based on a recombinant inbred population BAT 93 X Jalo EEP558 (Freyre *et al.*, 1998). Random amplified polymorphic DNA (RAPDs) (Adam-Blondon *et al.*, 1994; Bai *et al.*, 1997, Rodríguez-Suárez *et al.* 2006) and amplified fragment length polymorphisms (AFLPs) (Tar'an *et al.*, 2002) were also used to construct linkage maps for the crop.

Some of the SSRs described above have already been used in diversity analysis (Zhang *et al.* 2008) and in genetic maps (Yu *et al.*, 2000; Blair *et al.*, 2003, 2006; Grisi *et al.*, 2007). There are still linkage groups with low saturation and gaps remain considerable. A central effort of common bean genome research is the generation of integrated genetic and physical map. A draft physical map was developed and covered approximately 9,54% of the genome (Schlueter *et al.*, 2008). The objectives of the present study were: to validate new microsatellites, to develop a linkage map based on genic and genomic microsatellites markers and phenotypic markers and to compare this map with the physical map available.

Material and Methods

Parental genotypes

A mapping population was obtained from the cross of CAL143 with IAC-UNA. CAL143 is an Andean line with red stringed seeds (cream coat color) and calima type. IAC-UNA is a black seeded variety developed by the Agronomic Institute (IAC), and it is from the Mesoamerican gene pool. The two parents are divergent for many phenotypic characteristics and have contrasting responses to many common bean pathogens such as anthracnose, angular leaf spot and rust. This mapping population was developed at IAC, and it was advanced by inbreeding to F₁₀ generation and 380 lines were selected to be mapped.

Microsatellites amplifications and analysis

Total genomic DNA for each of 380 recombinant inbred lines was isolated from bulked young leaves of ten plants per line, using CTAB extraction method as described in Hoisington *et al.* (1994).

Eight hundred and seventy one microsatellites were tested for their polymorphism among CAL143 and IAC-UNA. From these, 567 have been already published. There were 44 markers of PV series (Yu *et al.*, 2000; Caixeta *et al.*, 2005), 46 markers from BM, AG and GAT series (Gaitán-Solis *et al.*, 2002), 2 markers from PVbng series (Murray *et al.*, 2002), 49 markers from Bmd series (Blair *et al.*, 2003), 15 markers from M series (Yaish & Perez de la Vega, 2003), 18 markers from Guerra-Sanz (2004), 71 markers from PVBR series (Buso *et al.*, 2006; Grisi *et al.*, 2007), 80 from PvM and FJ series (Hanai *et al.*, 2007), 222 markers from SSR-IAC series (Benchimol *et al.*, 2007; Cardoso *et al.*, 2008), and 20 markers from FJ series (Campos *et al.*, 2007). The series PvM (developed from

ESTs), Bmd and PV from Yu *et al.* (2000) and the markers from Guerra-Sanz (2004) (developed from GenBank gene searches) were gene based microsatellites. The remaining 304 markers tested were developed in this work according to Benchimol *et al.* (2007) and the polymorphic markers were further described (**Table 1**).

The microsatellite fragments were amplified by PCR containing: 40 ng of template DNA, 0.8 μ M of forward and reverse primer, 150 μ M of each dNTP, 1.5 mM MgCl₂, 10 mM Tris-HCl, 50 mM KCl, and 0.5 U Taq DNA Polymerase. Reactions were performed using the following conditions: 1 min at 94°C, followed by 30 cycles of 94°C for 1 min, specific annealing temperature for 1 min and 72°C for 1 min, finishing by 72°C for 5 min. Some microsatellites were mapping using cycles of touchdown (TD) profile (**Table 1**) described in Hanai *et al.* (2007). PCR products were separated by electrophoresis in denaturing polyacrylamide gels (6% w/v) and silver stained (Creste *et al.* 2001). Molecular sizes of the fragments were estimated using a 10bp ladder (Invitrogen). For 39 loci, it was utilized fluorescent primers according to Schuelke (2000), and they were genotyped at automatic sequencer ABI 3730.

Table 1. Description of the polymorphic microsatellites for CAL143 and IAC-UNA developed in this work. Including primer sequences, repeat motif, allele size and annealing temperature (Ta).

	Locus/ Genbank accession	Primer Sequence (5' - 3')	Motif	Size (pb)	Ta (°C)
1	SSR-IAC223/ FJ529449	GTGGCACCTGGAATAAGACAAC ACAAATGCTCACTCACCAAATC	(AG)15	214-220	TD
2	SSR-IAC224/ FJ529450	ATTATGGATTTAGGGTAGGTGA GCTTGACAATTGGAAGTGG	(TC)13	116-120	TD
3	SSR-IAC226/ FJ529452	TTTTTCTCGTTTTACTTTTATCTG TACTGATTTATTTCTTCCACTTC	(TG)8	150-155	60
4	SSR-IAC227/ FJ529453	GGATCTGTGCCTTCTCTGTG TTCCATATCCCCAAAACCTT	(AG)18	135-148	45
5	SSR-IAC228/ FJ529454	CAAATTTTGTCTGAACTGAT TTCTCTCCTCAAATGTAACAACT	(CA)6	203-	60
6	SSR-IAC229/ FJ529455	CCTAAGGATGAACCACTCTAATA ACCCTTGCATGTGTTGTTT	(CA)7	198-	60
7	SSR-IAC230/ FJ529456	GGATTCTGGCATTGATAGAC AGAGAAGAAATTAGGGAGAACAG	(CT)6	188-190	60
8	SSR-IAC231/ FJ529457	TGCATTTTGAAAGGACAGAT ACCAGGCATAGGTAGGACA	(TA)7 A (AG)16	246	60
9	SSR-IAC232/ FJ529458	GGTTTGCATTCTTCATTTATTT ACCCATTCCCAGACTTCC	(TG)6(TA)5	241-245	60
10	SSR-IAC233/ FJ529459	GGCAACTCTAAGGCAATCC CGAAGAATGTCAGAGAAATAAAC	(GT)9(TA)7	271	60
11	SSR-IAC234/ FJ529460	GACCGTTAAATGGAATCAGTT TGCCTATGTTTATGTGCTTGTT	(TC)5	258	60
12	SSR-IAC235/ FJ529461	TTCAGCATGCATATTCAAGTGT GGCTCCCAGGCATAGTTCT	(TA)10	253	60
13	SSR-IAC236/ FJ529462	TCTTCCTCTTCCTCTTCCTCTA TCTGATGTGGCAACGATAAAAT	(CT)12	170	60
14	SSR-IAC237/ FJ529463	TGAAAGGCCACAACGACAAGT GCACCCAATCTCCACAAA	(GT)2 AT (TG)5	155-160	60
15	SSR-IAC238/ FJ529464	GAGATAGATCATAGACCCCAA CTTTCAATTTTAACTCATCCTC	(AG)18	233	60
16	SSR-IAC239/ FJ529465	AAAGAAAAAGTGGGGAAAAGA GCTCATGCAAAAATAGACTCA	(AG)15	247	60
17	SSR-IAC240/ FJ529466	TGAATGTATGAGATTCGTAGTTA ATTTTATTAGTGGCATTGTCCT	(CT)10	213-228	60
18	SSR-IAC241/ FJ529467	AATGTGGCTATGATGGAAGAGG AAGAGAATGGGGGAAAAGGAG	(TC)8	151	60
19	SSR-IAC242/ FJ529468	ATGAGATGCGAGGATTTTAT GGATGGGTAAGAAGGCTGAA	(CA)5	242-280	60
20	SSR-IAC243/ FJ529469	GCTGGTGAGGCTTGTTGA	(AT)2 (GT)3	145	60

	FJ529469	AATTGACCTTGGATCTTCTTAG			
21	SSR-IAC244/ FJ529470	CGCGTTGTATTGCAGTAAAGAA AACTCCGTGGAACCCCTGTGAT	(TC)9	192	60
22	SSR-IAC245/ FJ5294471	TAGGCCATCATCATTTCAGTTTG GGTTGCCCTCATAGTCATCAC	(AT)2 (GT)3	185-192	60
23	SSR-IAC246/ FJ529472	TTCCCTTCATGGTCAACA CAGTAACTACAGATCAGCACAC	(AG)10	225-228	60
24	SSR-IAC247/ FJ529473	TGGATGAGACGGCTTTT CTAGATACATTCTTGTCTTCAGT	(AT)3 (GT)3	210-214	TD
25	SSR-IAC248/ FJ529474	TTATAAGTTTACATCAGTTTGGTT AGAATCAGGCCTTGCTTGT	(GT)2 GC (GT)4	286-288	TD
26	SSR-IAC249/ FJ529475	TTTATGAGTGGTCGAAGTGAAC AAGGGGAGGGTATTAGACAGT	(AG)9	227-231	60
27	SSR-IAC250/ FJ529476	TATCTAGATCTTTGCCCTTGAA CATTGCCTGTATTGTTTGAA	(GT)4 AT (GT)3	238	60
28	SSR-IAC251/ FJ529477	AAGTCTTGCCCTTTTTGTTTAT TTTTTGCTAAGTTGGTTCC	(AC)11 (AT)12	278	45
29	SSR-IAC252/ FJ529478	AATGAATTCGCAACTGTC GTGGCTAAGATGAAGAACAT	(GA)5	303-305	60
30	SSR-IAC253/ FJ529479	GTTTCGGTCGGGGATGGA TTAGACGGCTACTGGCAAAGAG	(GA)10	218-220	60
31	SSR-IAC254/ FJ529480	TTGGGTTTTATGTTTTCTGTGA CCATATCTTGTTGTGCCTTGAC	(TG)6	245-249	60
32	SSR-IAC255/ FJ529481	TTACAATTCACAACAACAGAGA TTAAGAGAAAATGAAGGATGAG	(AG)23	148-168	60
33	SSR-IAC256/ FJ529482	TATTTTATTAGTGGCATTGTCC TGAGAGATCCTGTAGAAGTAGTAGT	(GA)9	156-160	60
34	SSR-IAC257/ FJ529483	GCAACTGAAAGGCTAAGATT TATTGGAAAATATGGGAGAA	(GT)8	300-302	60
35	SSR-IAC258/ FJ529484	CATTGTTCGGTGTTCGAGAAGTC CCCACGCTCTTGTTGCTGTC	(GT)7	175-178	60
36	SSR-IAC259/ FJ529485	TATATGCCTGCACCACTGTAAC GAAAATCCGGAAACTCAAGAA	(TC)11	241-247	45
37	SSR-IAC260/ FJ529486	TGAACAGTGCAGCAGTAACAA CACCAGACACCAATCATCAA	(AG)9	119-129	60
38	SSR-IAC261/ FJ529487	TTCCCAAACACCACACCTAAGT TCACCGCGCACGAGATAA	(AC)8	260-270	60
39	SSR-IAC262/ FJ529488	ATATCGTTTGATATCCTTACACA CAAACACTGGTTCACATCTCAC	(GT)9	243-245	60
40	SSR-IAC263/ FJ529489	TGCAATTGACAAAAAGTTCGTA TGTATGATAGGCCTCCACCA	(AC)9	156-160	60
41	SSR-IAC264/ FJ529490	TGGGATCTGTGCCTTCTC TTCCATATCCCCAAAACCTT	(GA)15	125-135	45
42	SSR-IAC265/ FJ529491	GTAGGTTTGTGTGCGTGC GGAAAGAAAGTTAAGATTGAGT	(TG)5	245-247	60
43	SSR-IAC266/ FJ529492	TTGAGGATGTAGATTATTTTGTT	(TA)5 (TG)8	269-273	60

	FJ529492	CATCATTTGTGCAGTTACCAG			
44	SSR-IAC267/ FJ529493	TGAGTGAACCAGCATAATCTAA CACCCGGTTGAAAATACA	(CT)15	139-162	51,4
45	SSR-IAC268/ FJ529494	ATGTAGATCTGGTAAGGAGTGAA AATTAGATTAGTCAAGAAACAAAAC	(TC)9	190-193	60
46	SSR-IAC269/ FJ529495	TGCGCCACTGTTTCGTATT ACCCAACCCAGGACTTCAC	(AC)6	145-147	60
47	SSR-IAC270/ FJ529496	ATGTGTTTAGCAAGGATGGTCT GCTGATTTGCTGTAACGAAC	(TG)6	214-216	60
48	SSR-IAC271/ FJ529497	GATTTTCTTTTCTTCCCTCAA TGCTTCCATTCCATCAA	(CA)7	227-228	49
49	SSR-IAC272/ FJ529498	TAACTGAACAGGCTAAAAGAAC TGATGAGACTAAAAATGACACTT	(CA)6	210-215	60
50	SSR-IAC274/ FJ529500	TCTACACATAACTGCCCTGAAC AACACTGCCACAATAAAGAAA	(TG)7	200-205	60
51	SSR-IAC275/ FJ809021	AGCAACTCATAGCACTTCTG TGAGTAATTCATTGTCTGTCCT	(TA)3G(AC)6	255	48.6
52	SSR-IAC276/ FJ809022	AAGCCCAAGCCCAAACCAG CACCACCAAGACCCGATTCAG	(GAA)4(GT)7	249	52.8
53	SSR-IAC277/ FJ809023	ATGGAAGGCTGCAATACATCAG CAGCGACGGTGCTTACTGG	(AC)6	286	52.1
54	SSR-IAC278/ FJ809024	TCTGTCAGTTTAGCTTCGTCAC CCGTTGGAGGGTCGTTAC	(CT)10	191	51.9
55	SSR-IAC279/ FJ809025	CAATACCACAAAACGCAA CATTATGATTAAATTCTGGTGT	(TG)3GG(TG)2	220	45.4
56	SSR-IAC280/ FJ809026	ACTCCTGGCAAAAATCTCG GCAAAACCTCCATGAAGACAG	(TTGA)3	141	49.0
57	SSR-IAC281/ FJ809027	AGTCCAGGCAGAGCAAGCAAG CGCATTTCATCTCATCCATTATCC	(GT)9	189	52.7
58	SSR-IAC282/ FJ809028	CTCTTGGTTACGCGTGGACTAC CGCCTATGTTGGGTTGAGATG	(CAA)2T(CA)3	234	54.8
59	SSR-IAC283/ FJ809029	GAACTCATTCCCCTCTCC GCAGCTCCTCCATTCTAC	(TTA)7	208	47.9
60	SSR-IAC284/ FJ809030	AAGGAGAAAATAAAATCACAGTC GTTTCGAGGTTATTGGGGA	(CT)13	240	50.7
61	SSR-IAC285/ FJ809031	ACCTAGATGGATTTGTGACC GTCTAGTTTTACGCATTCATTC	(AT)11(TG)11(TA)6(AG)7	204	48.1
62	SSR-IAC286/ FJ809032	TTGTTCTGAGGACTGGATGTAT TCGGTTGATTTTTCGTTTTTA	(TC)18	181	49.1
63	SSR-IAC287/ FJ809033	ACAACGTTAGAAAGGGAAGG GAAAAAAGCGAGAAAAAGAGT	(GA)13	179	51.4
64	SSR-IAC288/ FJ809034	TCCCGTGAGTTACAAATAG ATCACCCTCCTCTCCAG	(ATTT)4	182	46.6

Phenotypic markers

Three phenotypic characteristics were evaluated on F₁₁ mapping population: flower color, growth habit, and curved pod tip shape. Flower color is controlled at the *V* locus (Lamprech, 1939) and segregates to purple and pink in mapping population. Growth habit varies from determinate and indeterminate habit, and it is described like qualitative characteristic controlled by *fin* locus (Lamprecht, 1935; Rudorf, 1958). The tip shape can be curved pod tip shape (*Ct*) or straight pod tip (*ct*) (Al-Mukhtar and Coyne 1981).

Segregation and linkage analyses

Tests of chi-square (χ^2) for 1:1 segregation ratios were performed on a subset of polymorphic markers. The segregation ratios were analyzed by *p*-valor and Bonferroni correction (Lynch & Walsh, 1998), performed in “R” Program version 2.5.1 (*The R Foundation for Statistical Computing*, 2007). The linkage map was constructed using the MAPMAKER 3.0b (Lander *et al.*, 1987). It was considered LOD values superiors to 3 and a maximum genetic distance of 37.5 cM. Initially, markers with known position in others maps were anchored (‘ANCHOR’ command) in the correspondent group. The linkage groups were established by two point analyses (‘TWO POINT’ command). Markers in conflict were observed through ‘LIST STATUS’ command and they were positioned by the greatest LOD value observed (‘ATTACH’ command).

Ordering analyses was conducted with priority for the markers with amplification fails no more than 100 samples in genotyping process (Informativeness Criteria: min Distance 2.0, min #Individuals 280) and without segregation distortion, utilizing ‘SUGGEST SUBSET’ command. For each group, nine

markers with the greatest statistical precision described above were placed first ('COMPARE' command). Additional markers were added using the 'TRY' command. First, the more informative loci were added, and after that, the less informative and with segregation distortion markers. The final ordering was verified by 'RIPPLE' command, using a window with 6 markers. The mapping function used for map distances were Kosambi (1944).

The distance between each marker and the linkage group were obtained by 'MAP' command. The expression graphic of linkage map was developed at MapChart 2.2 (Voorrips, 2002), and Adobe Photoshop 7.0 (Microsoft).

Results

Microsatellite polymorphisms tests

The 871 microsatellites were optimized for amplification conditions and the polymorphism was tested for the parents of the CAL143 X IAC-UNA population. It was obtained 265 (30,4%) polymorphic markers, and 247 (28,4%) presented adequate visualizing pattern without stutter bands or non-specific bands. It was optimized multiple applications in polyacrilamide gel for PCR reactions of the same locus (**Figure 1**). The automated process with fluorescent primers was a fast method to obtain mapping data and it allowed that two different loci could be genotyped simultaneously (**Figure 2**).

There were 8 (38,1%) markers polymorphic of serie PV of Caixeta *et al.* (2005), 19 (41,3%) markers of series BM, AG and GAT, 1 (50,0%) marker of serie PVbng, 38 (53,5%) markers of serie PVBR, 11 (26,3%) of serie FJ of Hanai *et al.* (2007), 69 (31,1%) markers of serie SSR-IAC; 5 (25,0%) markers of serie FJ (Campos *et al.*, 2007). All 15 markers of serie M were monomorphics. The

polymorphism in microsatellites gene based were 5 (21,8%) markers of Yu *et al.* (2000), 15 (32,7%) markers of serie Bmd, 3 (16,7%) markers of Guerra-Sanz (2004), and 10 (25%) markers of serie PvM. In total, there were 33 (25,4 %) polymorphic markers from 130 gene based markers tested and 232 (31,3%) from 741 markers of genomic source.

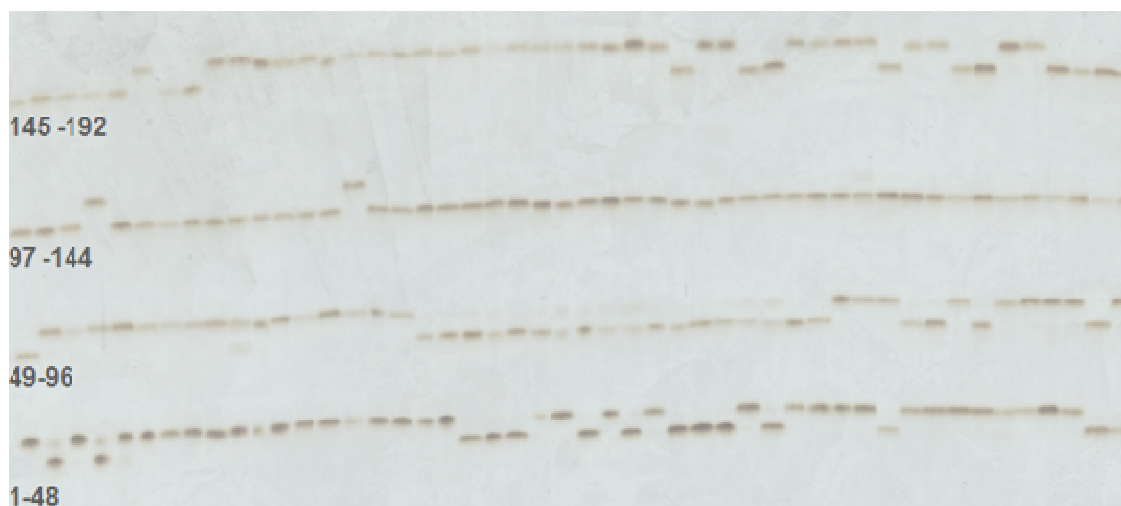


Figure 1. Locus FJ54 genotyped at polyacrilamide gel (6%) with 4 applications of 48 samples for time.

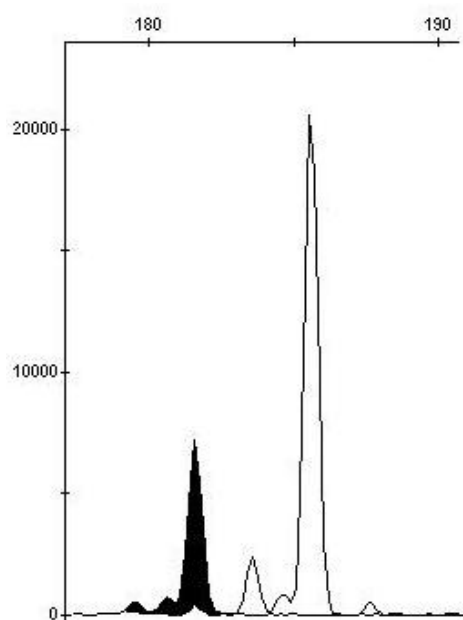


Figure 2. Multiplex genotyping of locus SSR-IAC288 and SSR-IAC261 at automatic sequencer with two fluorescent primers: FAM (in black) and HEX.

Genetic map

A total of 198 microsatellite markers and 3 phenotypic markers were mapped distributed in all 11 linkage groups (**Figure 3**). The total map length was 1865.9 cM with an average linkage group length of 170.5 cM. The average distance between each marker was 9.4 cM, with an average of 18 markers for each group. The number of markers varied from 5 (group B11) to 37 (group B2) for each linkage group.

The linkage groups B2, B3, B4, B5, B7, B8, and B9 presented gaps that suggested sub-groups. Some gaps could be resulted of markers firstly anchored to form the group. These anchored markers were supposed to have known position by others maps developed. In the attempt to solve the gaps, the anchored markers were removed and the two-point analyses were remade with remain markers of each group. Just after all remain markers were positioned, the anchored markers were integrated to the formed the group by TRY command. It solved the gaps that existed in groups B3, B5 and B7. The groups B2, B4, B8 and B9 still presented markers statistical linked ($LOD > 3$) but with genetic distance superior to 37.5 cM. In this cases, the initial group resulted in two sub-groups discriminated by segmented lines (**Figure 3**). The group B8 was separated in A and B because it was not possible to determine the linkage order between them. The subgroup 8A is formed by ten markers with unknown position mapping until then.

The marker PVBR251 had statistical evidence to be linked at group B1, but the distance of it from the adjacent marker (IAC235) was superior to 37.5 cM. Because of this, the marker PVBR251 remained linked to the group B1 and this position was assumed correct, but the genetic distance is not described in the map. The same fact occurred in group B9 with the marker IAC242.

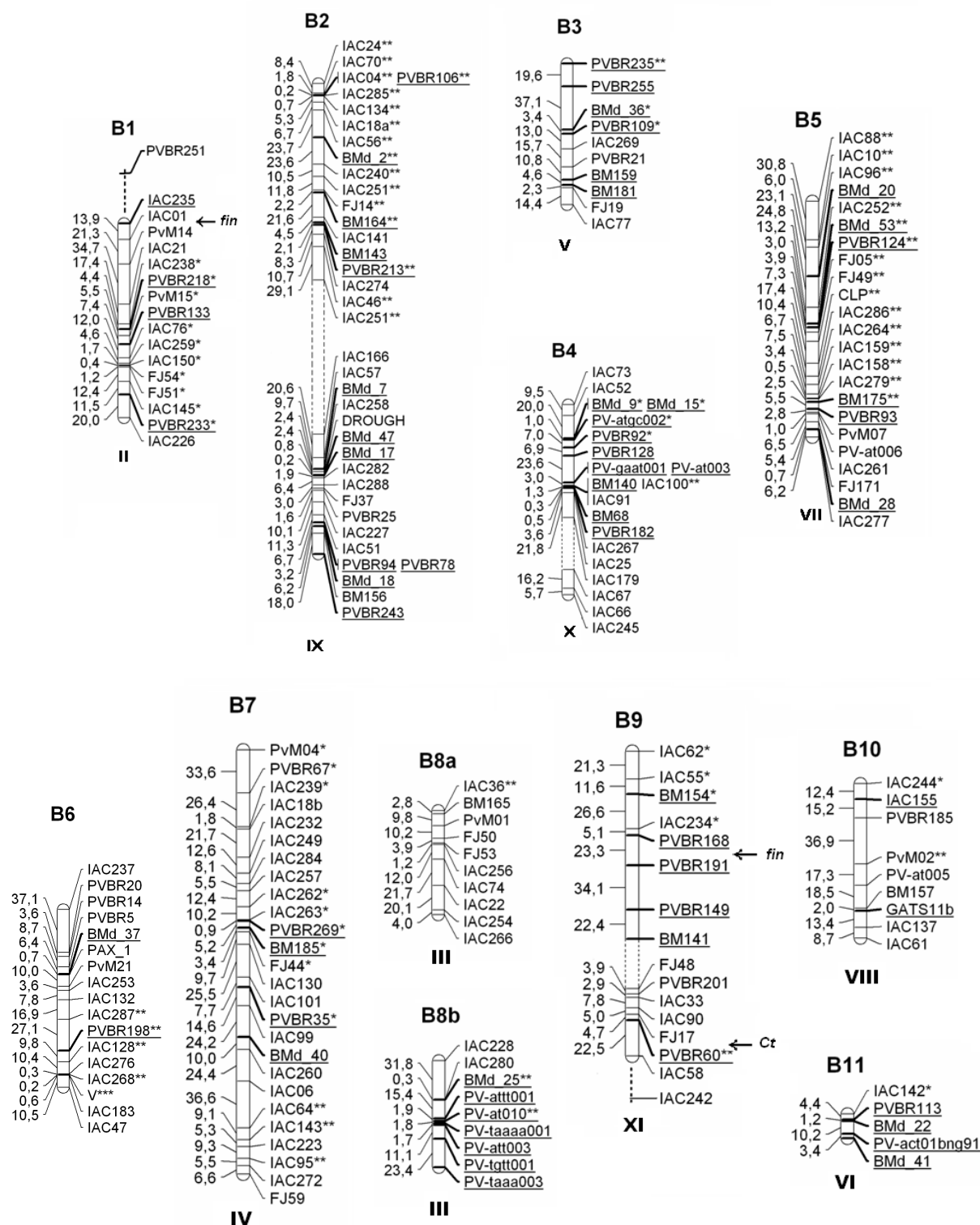


Figure 3. Genetic Map for IAC-UNA X CAL143 (UC). B1 to B11 are referent to linkage groups and, I to XI are the chromosome correspondence according to Pedrosa *et al.* (2003). Microsatellites underlined were anchored to the linkage groups. (*)Markers with segregation distortion to CAL143 and (**) to IAC-UNA.

(***) Phenotypic marker, *V* locus, linked at group B6. Arrows indicate possible position of *fin* and *Ct* loci.

Markers with significant segregation distortion were used to linkage analysis. This phenomenon was observed in 92 (37,3%) microsatellite markers. It was mapped 76 markers in the linkage map with segregation distortion, indicated by asterisks in **Figure 3**. These markers tended to be located together in determinate chromosome regions.

Besides *V* locus, two other phenotypic characteristics measured were mapped in the linkage groups, but they were not positioned and integrated to the genetic map, because they formed gaps inside the groups. The supposed positions of these markers are indicated in the genetic map by arrows. The characteristic curved tip shape, *Ct* locus, was significantly linked at group B9. The growth habit, the *fin* locus was a detected conflict marker, possible linked in group B1 (LOD=3,76) and in group B9 (LOD=3,95).

Discussion

Mapping Population

The new mapping population IAC-UNA X CAL143 established by this study represents the largest one developed until nowadays in terms of total number of recombinant inbred lines. Besides reaching F₁₀ generation, parents were chosen from different gene pools (Mesoamerican and Andean) and therefore, they possess phenotypically contrasting features for many traits, like to be resistant in soils that present low phosphorus, low nitrogen, and low pH. It was observed that parents segregate for resistance to diseases like anthracnose (caused by *Colletotrichum*

lindemuthianum), angular leaf spot (caused by *Pseudocercospora griseola*), rust (caused by *Uromyces phaseoli*), and halo blight (caused by *Pseudomonas savastanoi* pv. *phaseolicola*). These characteristics delineate an ideal mapping population with several important traits to be mapped.

A moderate polymorphism rate was found (30.4%) and it was expected for a cross between two cultivated genotypes. This result is similar to what was reported by other studies, such as 45.7% (Blair *et al.*, 2006) and 42.8% (Grisi *et al.*, 2007) using contrasting parents. If the parents were from the same gene pool, they would probably have a narrower genetic base and lesser content of genetic variability.

Probably, one reason for the lower polymorphism rate found was the extensive number of microsatellite with small motif repeats (considered from 5 units) that were used by our group. As we observed (unpublished data), polymorphism tended to be significantly associated (correlation $r=0,84$, $P=0,05$) with the size of the motif repeats in common bean. If we considered just the PVBR series, they presented the most polymorphic microsatellite serie (53.5%). In this work we described 64 new microsatellite markers that presented parental polymorphism (CAL143 X IAC-UNA) and will be useful for other genetic studies in common bean.

Among the microsatellite gene based, in linkage group B4 the loci BMd9, Bmd15 and PVatgc002 were mapped closely and they were derived from different phytohemagglutinin gene sequences. Blair *et al.* (2003) found these markers in the same linkage group and they suggested that it represents a single location of a gene family cluster.

Genetic Map

The final UC genetic map covered all 11 linkage groups of the common bean. The total length (1865.9 cM) was consistent with other genetic mapping studies (Blair *et al.*, 2003; Grisi *et al.* 2007). Vallejos *et al* (1992) had estimated that an ideal common bean map must have about 1200 cM, but that was based on BJ population, which have 75 recombinant lines. The UC (CAL143 X IAC-UNA) population has higher power to detect recombinant events, because of the huge number of recombinant lines (n=380). For this reason, it is expected a higher capacity of mapping saturation and a quite robust map to be generated, with precise estimates associated.

Some linkage groups presented highly dense cover distribution of markers (B5, B7), and others still have gaps (B2, B8), or did not present many markers yet (B11 with 5 markers). In the microsatellite based genetic map presented by Grisi *et al.* (2007), the B11 group (chromosome 6) was also the smaller linkage group detected, with 4 markers.

A total of 199 markers were linked to our UC map, including 131 markers that were not mapped before. The use of anchor markers was efficient to form the final linkage groups and ordinate the markers with unknown position. However, there was the need of remaking the two-point analysis according to our sampling data for some markers that could not be linked to the specific groups they were referenced before.

The segregation distortion observed for 92 markers did not exclude them as potential informative for recombinant detections. For this reason, they were incorporated to the linkage analysis. According to Shizhong (2008), the use of markers with segregation distortion could elevate the level of mapping saturation

and the detection power of QTLs. The phenomenon of markers with segregation distortion in common bean had already been described by many authors (Freyre *et al.*, 1998; Blair *et al.*, 2003; Grisi *et al.*, 2007; Hougaard, *et al.*, 2008). It was observed not only for microsatellite markers, but in Legume anchor markers (Hougaard *et al.*, 2008), and RFLP markers (Freyre *et al.*, 1998). The reasons for this fact were not totally elucidated yet, but it could be associated to a preferential selection of genomic regions that conferred adaptative advantages during a population mapping development.

In UC map, it was detected a concentration of markers with distortion to Andean alleles in the groups B1, and to Mesoamerican alleles in B2 and B5. In the linkage group B2 only the superior region presented markers with segregation distortion. For the group B5, the distortion region is extended to the superior extremity for the middle portion of the group. In group B1, the Andean alleles predominated from middle to inferior extremity (**Figure 3**). Freyre *et al.* (1998) also found the distortion for B1 and B2 for the same parental gene pools. Blair *et al.* (2003) found the same segregation distortion to Mesoamerican alleles on regions of chromosome B5.

Phenotypic Markers

The *V* locus, responsible by color flower, was mapped in group B6. The same result was obtained by Nodari *et al.* (1992) and McClean *et al.* (2002). The *V* locus was detected closely linked to microsatellites IAC268 (0.2 cM) and to IAC183 (0.6 cM) in the extremity of B6, and they could be used for future cloning experiments.

The *fin* locus is associated to a gene that controls determinate versus indeterminate growth habit. It was identified in two probable groups (B1 and B9), with almost the same LOD value. However, it was not positioned on the map because it caused a gap in both groups. It is necessary more markers in this group to saturate the region and to complete the formed gap. Koinange *et al.* (1996) also mapped the *fin* locus on B1. Ta'an *et al.* (2001) mapped the growth habit gene in a linkage group K, equivalent to B9. The existence of only one locus to define this trait was already questioned, and it is probable that multiple genes to determinate growth habit exist in common bean (McClellan *et al.*, 2002). The characteristic tip shape, *Ct* locus, was not mapped because of gap formation described above.

Bibliography

- Adam-Blondom, A., M. Sévignac M., Dron, M. (1994). A genetic map of common bean to localize specific resistance genes against anthracnose. *Genome*, 37, 915-924.
- Al-Mukhtar, F.A., Coyne, D.P. (1981). Inheritance and association of flower, ovule, seed, pod, and maturity characters in dry edible beans (*Phaseolus vulgaris* L.). *Journal of the American Society for Horticultural Science*, 106:713-719.
- Arumuganathan, K. & Earle, E.D. (1991). Nuclear DNA content of some important plant species. *Plant Molecular Biology Reporter*, 9:208-218.
- Bai, Y., Michaels, T.E., Pauls, K.P. (1997). Identification of RAPD markers linked to common bacterial blight resistance genes in *Phaseolus vulgaris* L. *Genome*, 40:544-551.
- Basset, M.J. (1991). A revised linkage map of common bean. *HortScience*, 26:834-836.
- Benchimol, L.L., Campos, T., Carbonell, S.A.M., Colombo, C. A., Chioratto, A.F., Formighieri, E.F., Gouvea, L.R.L., Souza, A.P. (2007). Structure of genetic diversity among
-

common bean (*Phaseolus vulgaris* L.) varieties of Mesoamerican and Andean origins using new developed microsatellite markers. *Genetic Resources and Crop Evolution*, 54:1747-1762.

Blair, M.W., Pedraza, F., Buedia, H.F., Gaitán-Solís, E., Beebe, S.E., Gepts, P., Tohme, J. (2003). Development of a genome-wide anchored microsatellite map for common bean (*Phaseolus vulgaris* L.). *Theoretical and Applied Genetics*, 107:1362-1374.

Blair, M.W., Giraldo, M.C., Buendia, H.F., Tovar, E., Duque, M.C., Beebe, S.E. (2006). Microsatellite marker diversity in common bean (*Phaseolus vulgaris* L.). *Theoretical and Applied Genetics*, 113:100-109.

Blair, M.W., Buendía, H.F., Giraldo, M.C., Métais, I., Peltier, D. (2008). Characterization of AT-rich microsatellites in common bean (*Phaseolus vulgaris* L.). *Theoretical and Applied Genetics*, 118:91-103.

Broughton, W. J., G. Hernández, M. Blair, S. Beebe, P. Gepts., Vanderleyden, J. (2003) Beans (*Phaseolus* spp.): model food legumes. *Plant Soil*, 252: 55–128.

Buso, G.S.C., Amaral, Z.P.S., Brondani, R.P.V., Ferreira, M.E. (2006). Microsatellite markers for the common bean *Phaseolus vulgaris*. *Molecular Ecology Notes*, 6:252-254.

Caixeta, E.T., Borém, A., Kelly, J.D. (2005). Development of microsatellite markers based on BAC common bean clones. *Crop Breeding and Applied Biotechnology*, 5:125-133.

Campos, T., Benchimol, L.L., Carbonell, S.A.M., Chioratto, A.F., Formighieri, E.F., Souza, A. P. (2007). Microsatellites for genetic studies and breeding programs in common bean. *Pesquisa Agropecuaria Brasileira*, 42:589-592.

Cardoso, J.M.K., Oblessuc, P.R., Campos, T., Sforça, D.A., Carbonell, S.A.M., Chioratto, A.F., Formighieri, E.E. Souza, A.P., Benchimol, L.L. (2008). New microsatellite markers developed from an enriched microsatellite common bean library. *Pesquisa Agropecuária Brasileira*, 43:929-936.

- Condit, R., Hubbell, S.P. (1991). Abundance and DNA sequence of two-base regions in tropical tree genomes. *Genome*, 34:66-71.
- Creste, S., Tulmann, A., Figueira, A. (2001). Detection of Single Sequence Repeat Polymorphism in denaturing Polyacrylamide Sequencing Gels by Silver Staining. *Plant Molecular Biology Reporter*, 19:299-306.
- FAO – Food and Agriculture Organization of the United Nations (2005). FAOSTATdatabase:<http://www.fao.org/es/ess/top/topproduction.html?lang=en&country=21&year=2005>. Cited 7 Oct 2008.
- Freyre, C.I., Skroch, P., Geffroy, V., Adam-Blondon, A.F., Shirmohamadali, A., Johnson, W., Llaca, V., Nodari, R., Pereira, P., Tsai, S.M., Tohme, J., Dron, M., Nienhuis, J., Vallejos, C., Gepts, P. (1998). Towards an integrated linkage map of common bean. 4. Development of a core map and alignment of RFLP maps. *Theoretical and Applied Genetics*, 97:847-856.
- Gaitán-Solís, E., Duque, M.C., Edwards, K.J., Tohme, J. (2002). Microsatellite in Common Bean (*Phaseolus vulgaris*): Isolation, Characterization, and Cross-Species Amplification in- *Phaseolus* ssp. *Crop Science*, 42: 2128-2136.
- Gepts, P.R., Nodari, R., Tsai, R., Koinange, E.M.K., Llaca, V., Gilbertson, R., Guzman, P. (1993). Linkage Mapping in Common Bean. *Annual Report Bean Improvement Cooperative*, 36:24-38.
- Gepts, P., Papa, R., González, A., Acosta, J. Delagdo-Salinas, A. 1999. Human effects on *Phaseolus vulgaris* adaption during and after domestication. P. 161-181. In L.W.D. Van Raamsdonk and J.C.M. den Nijs (ed.) Plant evolution in man-made habitats. Hugode Vries Laboratory, Amsterdam, the Netherlands.
- Grisi, M.C.M., Blair, M.W., Gepts, P., Brondani, C., Pereira, P.A.A., Brondani, R.P.V. (2007). Genetic mapping of a new set of microsatellite markers in a reference common

bean (*Phaseolus vulgaris*) population BAT93 x Jalo EEP558. *Genetics and Molecular Research*, 6:691-706.

Guo, X., Castill-Ramirez, S., Gonzalez, V., Bustos, P., Fernandez-Vazquez, J.L., Santamaria, R.I., Arellano, J., Cevallos, M.A., Davila, G. (2007). Rapid evolutionary change of common bean (*Phaseolus vulgaris* L.) plastome, and the genomic diversification of legume chloroplasts. *BMC Genomics*, 8:228.

Guerra-Sanz, J.M. (2004). New SSR markers of *Phaseolus vulgaris* from sequence databases. *Plant Breeding*, 123: 87-89.

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

Hanai, L. R., Campos, T., Camargo, L. E. A., Benchimol, L. L., Souza, A. P., Melloto, M., Carbonell, S. A. M., Chioratto, A. F., Consoli, L., Formighieri, E. F., Siqueira, M., Tsai, S. M., Vieira, M. L. C. (2007). Development, characterization, and comparative analysis of polymorphism at common bean SSR loci isolated from genic and genomic sources. *Genome*, 50:266-277.

Harper, G.L., Maclean, N., Goulson, D. (2003). Microsatellite markers to assess the influence of population size, isolation and demographic change on the genetic structure of the UK butterfly *Polyommatus bellargus*. *Molecular Ecology*, 12:3349-3357.

Hoisington D., M. Khairallah and D. González-de-León, 1994: Laboratory Protocols: CIMMYT Applied Molecular Genetics Laboratory. Second Edition. Mexico, D.F.: CIMMYT.

Hornik, K. 2008. The R FAQ. ISBN 3-900051-08-9. <http://CRAN.R-project.org/doc/FAQ/R-FAQ.html>.

- Hougaard, B.K., Madsen, L.H., Sandal, N., Moretzsohn, M. de C., Fredslund, J., Schauser, L., Nielsen, A.M., Rohde, T., Sato, S., Tabata, S., Bertoli, D.J., Stougaard, J. (2008). Legume Anchor Markers Link Syntenic Regions Between *Phaseolus vulgaris*, *Lotus japonicus*, *Medicago truncatula* and *Arachis*. *Genetics*, 179:2299-2312.
- Hughes, C.R., Queller, D.C. (1993). Detection of highly polymorphic microsatellite loci in a species with little allozyme polymorphism. *Molecular Ecology*, 2:131-137.
- Koinange, E.M.K., Singh, S.P., Gepts, P. (1996). Genetic control of the domestication syndrome in common bean. *Crop Science*, 36:1037-1045.
- Lamprecht, H. 1935. Zur Genetik von *Phaseolus vulgaris*. X. Über Infloreszenztypen und ihre Vererbung. *Hereditas*, 20: 71-93.
- Lamprecht, H. 1939. Zur Genetik von *Phaseolus vulgaris*. XIV. Über die Wirkung der Gene P, C, J, Ins, Can, G, B, V, Vir, Och und Flav. *Hereditas*, 25:255-288.
- Lander, E.S., Green, P., Abrahamson, J., Barlow, A., Daly, M.J., Lincoln, S.E., Newburg, L. (1987). Mapmaker: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics*, 1:174-181.
- Lynch, M., Walsh, B. (1998). Genetics and analysis of quantitative traits. Sunderland: Sinauer Associates. 980p.
- Margarido, G.R.A., Souza, A.P., Garcia, A.A.F. (2007). OneMap: software for genetic mapping in outcrossing species. *Hereditas*, 144:78-79.
- McClearn, P.E., Lee, P.K., Otto, C., Gepts, P., Bassett, M.J. (2002). Molecular and phenotypic mapping of genes controlling seed coat pattern and color in common bean (*Phaseolus vulgaris* L.). *The Journal of Heredity*, 93:148-152.

- Métais, I., Hamon, B., Jalouzot, R., Peltier, D. (2002). Structure and level of genetic diversity in various bean types evidenced with microsatellite markers isolated from a genomic enriched library. *Theoretical and Applied Genetics*, 104:1346-1352.
- Moore, S.S., Hale, P., Byrne, K. (1998). NCAM: a polymorphic microsatellite locus conserved across eutherian mammal species. *Animal Genetics*, 29:33-36.
- Murray, J., Larsen, J., Michaels, T.E., Schaafsma, A., Vallejos, C.E., Pauls, K.P. (2002). Identification of putative genes in bean (*Phaseolus vulgaris*) genomic (Bng) RFLP clones and their conversion to STSs. *Genome*, 45, 1013-1024.
- Nodari, R.O., Tsai, S.M., Gilbertson, R.L., Gepts, P. (1993). Towards an integrated map of common bean-2: development of a RFLP-based linkage map. *Theoretical and Applied Genetics*, 85:513-520.
- Pedrosa, A., Vallejos, C.E., Bachmair, A., Schweizer, D. (2003). Integration of common bean (*Phaseolus vulgaris* L.) linkage and chromosomal maps. *Theoretical and Applied Genetics*, 106:205-212.
- Primmer, C.R., Raudsepp, T., Chowdhary, B.P., Møller, A.P., Ellegren, H. (1997). Low frequency of microsatellites in the Avian Genome. *Genome Research*, 7:471-482.
- Rico, C., Zadworny, D., Kuhnlein, U., Fitzgerald, G.J. (1993). Characterization of hypervariable microsatellite loci in the threespine stickleback *Gasterosteus aculeatus*. *Molecular Ecology*, 2:271-272.
- Rodríguez-Suárez, C., Méndez-Vigo, B., Pañeda, A., Ferreira, J.J., Giraldez, R. (2006). A genetic linkage map of *Phaseolus vulgaris* L. and localization of genes for specific resistance to six races of anthracnose (*Colletotrichum lindemuthianum*). *Theoretical and Applied Genetics*, 114:713-722.

- Romero, A.J., Bliss, F.A. (1985). Heritable variation in the phaseolin protein of nondomesticated common bean, *Phaseolus vulgaris* L. *Theoretical and Applied Genetics*, 71:478-480.
- Rudorf, W. (1958). Genetics of *Phaseolus aborigineus* Burkart. Proc. X Intern. Cong. Genet. 2:243. (Abstr.).
- Schlueter, J.A., Goicoechea, J.L., Collura, K., Gill, N., Lin, J., Yu, Y., Kudrna, D., Zuccolo, A., Vallejos C.E., Muñoz-Torres, M., Blair, M.W., Tohme, J., Tomkins, J., McClean, P., Wing, R.A., Jackson, S.A. (2008). BAC-end Sequence Analysis and a Draft Physical Map of the Common Bean (*Phaseolus vulgaris* L.) Genome. *Tropical Plant Biology*, 1:40–48.
- Schuelke, M. (2000). An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology*, 18:233-234.
- Shizhong, X. (2008). Quantitative Trait Locus Mapping Can Benefit From Segregation Distortion. *Genetics*, 180:2201-2208.
- Singh, S., Nodari, R., Gepts, P. (1991). Genetic diversity in cultivated common bean. II. Marker-based analysis of morphological and agronomic traits. *Crop Science*, 31:19-23.
- Ta'an, B., Michaels, T.E., Pauls, K.P. (2001). Identification of markers associated with quantitative traits in beans. *Annual Report of the Bean Improvement Cooperative*, 44:9-10.
- Ta'an, B., Michaels Thomas, E., Pauls, K.P. (2002). Genetic Mapping of Agronomic traits in Common Bean. *Crop Science*, 42:544-556.
- Vallejos, C. E., Sakiyama, N. S., Chase, C. D. (1992). A molecular marker-based linkage map of *Phaseolus vulgaris* L. *Genetics*, 131: 733-740,
-

- Voorrips, R.E. (2002). MapChart: software for the graphical presentation of linkage maps and QTLs. *Journal of Heredity*, 93:77-78.
- Tautz D. & Renz M. (1984). Simple sequence repeats are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Research*, 12:4127-4137.
- Yaish, M.W.F. & Pérez de la Vega, M. (2003). Isolation of (GA)_n microsatellite sequences and description of a predicted MADS-box sequence isolated from common bean (*Phaseolus vulgaris* L.). *Genetics and Molecular Biology*, 26: 337-342.
- Yang, G.P., Shagai Maroof, M.A., Xu, C.G., Zhang, Q., Biyashev, R.M. (1994). Comparative analysis of microsatellite DNA polymorphism in landraces and cultivars of rice. *Molecular and General Genetics*, 245:187-194
- Yu, K., Park, S.J., Poysa, V. (1999). Abundance and variation of microsatellite DNA sequences in beans (*Phaseolus vulgaris* and *Vigna*). *Genome*, 42:27-34.
- Yu, K., Park, S.J., Poysa, V., Gepts, P. (2000). Integration of simple sequence repeat (SSR) markers into a molecular linkage map of common bean (*Phaseolus vulgaris* L.). *Journal of Heredity*, 91:429-434.
- Zhang, X.; Blair, M.W.; Wang, S. (2008). Genetic diversity of Chinese common bean (*Phaseolus vulgaris* L.) landraces assessed with simple sequence repeat markers. *Theoretical and Applied Genetics*, 117:629-640.

8. Discussão

O uso de marcadores moleculares para acessar o perfil genético em feijoeiro tem sido amplamente empregado (Jarne & Lagoda, 1996): para investigar a origem e a diversidade de cultivares (Gepts, 1998), a domesticação de *Phaseolus vulgaris* (Koinange *et al.*, 1996), o controle genético de importantes doenças (Nodari *et al.*, 1993; Miklas *et al.*, 2003; Blair *et al.*, 2006b), para a construção e a integração dos mapas (Blair *et al.*, 2003; Tarlan *et al.*, 2001; Yu *et al.*, 2000).

O desenvolvimento de microssatélites em feijoeiro permitiu um maior conhecimento genético da cultura. Através da técnica de construção de bibliotecas enriquecidas (Bilotte *et al.*, 1999) foi possível obter um grande número de microssatélites e observar o padrão dos motivos repetidos nos marcadores. O trabalho atual desenvolveu 488 microssatélites e disponibilizou 183 locos na forma de 3 artigos publicados.

O número total de locos variou de 02 a 08 alelos por loco. Yu *et al.* (2000) encontraram uma variação de 02 a 10 alelos quando avaliaram 24 microssatélites polimórficos em 12 genótipos. Buso *et al.* (2006) relataram uma variação de 3 a 10 alelos para 10 locos polimórficos de microssatélites caracterizados em 85 acessos.

O menor PIC encontrado foi de 0,14 e foi obtido para 29 locos, que apresentaram somente um genótipo com 1 alelo, em homozigose, diferente em relação aos demais genótipos. O genótipo de *Vigna* spp, Fradinho Cruzeiro, foi o material responsável pelo polimorfismo do alelo único em 59% destes casos.

Observamos motivos de repetições majoritariamente curtas e interrompidas no genoma do feijoeiro. Isso nos indica a tendência de conservação dos motivos em uma planta autógama e com ciclo de vida curto. Como a taxa de

fecundação cruzada é baixa (cerca de 5%) o material genético é, na maior parte, conservado entre as gerações, as quais avançam por autofecundação num período de ao redor de 3 meses. Acreditamos que o conjunto desses dois fatores levou à diminuição dos motivos microssatélites ao longo da evolução do genoma da planta e na conservação deste padrão observado.

O critério de seleção de dinucleotídeos a partir de 5 repetições pode ter sido a causa do grande número de locos monomórficos observados. Houve alta correlação entre o número repetições e os valores de PIC ($r = 0.84$). Assim, repetições maiores estavam frequentemente associadas a locos mais polimórficos. O maior valor de PIC obtido (0,89) foi obtido pelo microssatélite SSR-IAC267 portador do motivo (CT)₁₅, que apresentou 8 alelos. Apenas 28% dos locos desenvolvidos foram polimórficos para mapeamento. Microssatélites de repetições curtas podem ter sido ineficientes para acessar o polimorfismo entre os genótipos usados. Além disso, os materiais escolhidos no banco de germoplasma do IAC eram genótipos cultivados, de base genética estreita devido à seleção ao longo de muitos anos de características favoráveis ao plantio e colheita, conhecida como a síndrome da domesticação (Koinange *et al.*, 1996). A aparente diversidade morfológica na cor de tegumento dos genótipos é uma característica qualitativa, controlada por poucos locos.

Gaitán-Sólis *et al.* (2002) descreveram que a eficiência de um dado microssatélite não dependeria do número de padrões gerados já que eles verificaram que locos com o mesmo número de alelos possuíam poder discriminatórios diferenciados. Porém, estes autores utilizaram um menor número de locos (68) que nosso estudo e com outro critério de seleção de motivos. Além

disso, o único microssatélite com dinucleotídeo de 5 repetições apresentou-se monomórfico (Gaitán-Sólis *et al.* 2002).

Os locos forneceram dados que corroboram para a separação dos genótipos Andinos dos Mesoamericanos. De acordo com as distâncias genéticas obtidas pelo método de Rogers Modificado (DMR), foi feito um agrupamento UPGMA que separou os dois grupos (Benchimol *et al.*, 2007). Em uma análise bayesiana desenvolvida no programa Structure 2.2 também houve a separação dos genótipos de acordo com o centro de origem (dados ainda não publicados). No trabalho de Blair *et al.*, (2006a) os microssatélites também foram eficientes na separação de 44 materiais quanto ao centro de origem pertencente. Estes resultados mostram a eficiência dos marcadores microssatélites para separar *pools* gênicos distintos. O trabalho de desenvolvimento de microssatélites é laborioso, porém uma vez estabelecidos marcadores robustos, estes poderão ser usados amplamente pela comunidade do feijoeiro sem problemas de transferibilidade entre laboratórios.

De acordo com as distâncias genéticas obtidas, os genitores da população de mapeamento Cal143 e IAC-UNA apresentam distâncias genética superiores (0,848) que aquelas observadas em relação aos genótipos BAT93 e Jalo EEP558 (0,784) da população referência do mapa *core* (Freyre *et al.*, 1998).

É importante ressaltar que a população de mapeamento desenvolvida neste trabalho é única no mundo nas vantagens que apresenta. Não temos conhecimento de outra população RIL de tamanho efetivo tão grande (380 linhagens na geração F₁₀) para feijoeiro. Além disso, os genitores pertencem a centros de origem distintos e são altamente contrastantes para diferentes características, tais como hábito de crescimento, produção de sementes, resistência à antracnose, mancha angular e ferrugem. As características divergentes entre os genitores possibilitam

o mapeamento de diferentes características quantitativas e qualitativas. Portanto, esta população permitirá um grande avanço nas pesquisas de mapeamento em feijão. Desta forma, pretendemos distribuí-la a diferentes pesquisadores da área, a fim de permitir que os avanços sejam rápidos.

O mapa desenvolvido (Artigo IV) representa o mapa genético com maior número de microssatélites até o momento. O mapeamento das características qualitativas (cor de flor, hábito de crescimento e formato do ápice da vagem) foi efetuado, e também foi possível propor um novo modelo de herança para uma das características.

De acordo com os dados de segregação e de mapeamento em duas regiões, estimamos que a característica hábito de crescimento seja governada por dois locos, que precisam se apresentar em recessividade para a expressão fenotípica de hábito determinado. Koinange *et al.* (1996) também haviam mapeado a mesma característica no grupo B1. Ta'an *et al.* (2001) mapearam hábito de crescimento no grupo B9. Nossos resultados indicaram o mapeamento nestes dois grupos B1 e B9, resultado possivelmente do uso da população de mapeamento maior, que aumenta a capacidade de detecção de recombinantes.

9. Conclusões

Desenvolvimento de Marcadores Microsatélites

- (1) O desenvolvimento de um banco enriquecido em microsatélites para a variedade IAC-UNA revelou 714 microsatélites. Este número representa um sucesso na técnica de enriquecimento de 59%, e está de acordo com os números encontrados em trabalhos semelhantes.
- (2) Dentre os microsatélites encontrados, foi possível o desenho de **488** iniciadores para amplificação dos locos. Eles foram otimizados quanto a suas condições de anelamento e genotipados.
- (3) A avaliação do polimorfismo dos microsatélites genômicos desenvolvidos e a comparação destes com locos derivados de ESTs revelaram níveis de polimorfismos mais baixos. No entanto, o número de microsatélites-ESTs disponíveis ainda é inferior aos genômicos, o que pode ter influenciado as análises. Além disso, o critério adotado para seleção das repetições de dinucleotídeos ($n > 5$) pode ter sido a causa do polimorfismo inferior.
- (4) Foi encontrada correlação significativa ($r=0,84$) entre tamanho do motivo microsatélite e o índice de polimorfismo (PIC). Assim, geralmente os locos monomórficos apresentavam poucas repetições no motivo microsatélite, tal fato foi relatado em outros estudos também.
- (5) Acredita-se que a distribuição dos motivos em pequenas repetições seja característico da espécie *Phaseolus vulgaris*, um possível resultado dos fatores: ciclo de vida curto e natureza autógama de reprodução, que levaram à diminuição das repetições ao longo das gerações.

Mapeamento Genético dos Microssatélites

- (1) Os 247 locos microssatélites polimórficos entre os genitores da população de mapeamento promoveram a elaboração de um mapa genético da ordem de 1865 cM, representando o mapa com o maior número de microssatélites até o momento.
- (2) A população de mapeamento utilizada apresenta o maior número de linhagens recombinantes (380 linhas puras) já desenvolvida para a cultura, e o maior nível de autofecundação alcançado (geração F₁₀).
- (3) O uso de marcadores já utilizados em outros mapas genéticos foi eficiente para ancorar os grupos de ligação, possibilitando o posicionamento das novas marcas, e assim, foi possível a obtenção dos 11 grupos esperados.
- (4) A utilização de marcadores com desvio da segregação esperada de 1:1 revelou regiões cromossômicas de alguns grupos de ligação que apresentam seleção para um dos genitores. Tal fato também foi relatado em outros estudos e a causa é ainda desconhecida e comum entre cruzamentos entre *pools* gênicos distintos.

10. Perspectivas

- (1) O mapa genético será saturado com outros marcadores microssatélites recentemente publicados e, também espera-se utilizar outros tipos de marcadores.
- (2) Outras características qualitativas, responsáveis pela cor e padrão de desenho no tegumento da semente, poderão ser mapeadas com a realização de cruzamentos testes para a determinação dos alelos específicos dos genitores e a referente genotipagem na população de mapeamento.
- (3) Pretende-se realizar outros ensaios de mensurações de características quantitativas de produção, em outros locais e em diferentes épocas de plantio, para a detecção de possíveis QTLs e sua interação com o ambiente.
- (4) O mapeamento de locos que conferem resistência a doenças também é um dos objetivos do grupo. O mapeamento de QTLs para resistência a mancha angular já foi desenvolvido, e pretende-se estender os estudos para antracnose e mofo branco, além de outras características contrastantes para a população de mapeamento CAL143 e IAC-UNA.

11. Bibliografia

- Adam-Blondon, A.; Sévignac, M.; Dron, M. (1994). A Genetic Map of Common Bean to Localize Specific Resistance Genes Against Anthracnose. *Genome*, 37:915-924.
- Arumuganathan, K. & Earle, E.D. (1991). Nuclear DNA content of some important plant species. *Plant Molecular Biology Reporter*, 9:208-218.
- Asíns, M.J. (2002). Present and future of quantitative trait locus analysis in plant breeding. *Plant breeding*, 121:281-291.
- Bai, Y.; Michaels, T.E.; Pauls, K.P. (1997). Identification of RAPD markers linked to common bacterial blight resistance genes in *Phaseolus vulgaris* L. *Genome*, 40:544-551.
- Basset, M.J. (1991). A revised linkage map of common bean. *HortScience*, 26:834-836.
- Bassett, M.J. (1994). The griseoalbus (gray-white) seed coat color is controlled by an allele (pgri) at the P locus in common bean. *HortScience*, 29:1178-1179.
- Bassett, M.J. (1997). Tight linkage between the fin locus for plant habit and the Z locus for partly colored seed coat patterns in common bean. *Journal of the American Society for Horticultural Science*, 122:656-658.
- Bassett MJ and McClean PE, 2000. A brief review of the genetics of partly colored seed coats in common bean. *Annual Report of the Bean Improvement Cooperative*, 43:99-101.
- Becerra-Velazquez, L.; Gepts, P. (1994). RFLP diversity of common bean (*Phaseolus vulgaris* L.) in its centres of origin. *Genome*, 37:256-263
- Beckmann, J.S.; Weber, J.L. (1992). Survey of human and rat microsatellites. *Genomics*, 12: 627-631.
- Benchimol, L.L.; Campos, T.; Carbonell, S.A.M.; Colombo, C. A.; Chioratto, A.F.; Formighieri, E.F.; Gouvea, L.R.L.; Souza, A.P. (2007). Structure of genetic diversity among common bean (*Phaseolus vulgaris* L.) varieties of Mesoamerican and Andean origins using new developed microsatellite markers. *Genetic Resources and Crop Evolution*, 54:1747-1762.

- Beninger, C.W.; Hosfield, G.L.; Bassett, M.J. (1999). Flavonoid composition of three genotypes of dry bean (*Phaseolus vulgaris* L.) differing in seedcoat color. *Journal of the American Society for Horticultural Science*, 124:514–518.
- Beninger, C.W.; Hosfield, G.L.; Bassett, M.J.; Owens, S. (2000). Chemical and morphological expression of the B and Asp seedcoat genes in *Phaseolus vulgaris* L. *Journal of the American Society for Horticultural Science*, 125:52–58.
- Bhering, L.L.; Cruz, C.D. (2008). Tamanho de população ideal para mapeamento genético em famílias de irmãos completos. *Pesquisa Agropecuária Brasileira*, 43:379-385.
- Bianchini, A.; Menezes, J.R.; Maringone, A.C. (1989). Doenças e seu controle. In: Instituto Agrônômico do Paraná. O feijão no Paraná. Londrina, 189-216p. (IAPAR. Circular, 63).
- Billotte, N.; Lagoda, P.J.L.; Risterucci, A.M.; Baurens, F.C. (1999). Microsatellite-enriched libraries: applied methodology for the development of SSR markers in tropical crops. *Fruits*, 54, 277-288.
- Blair, M.W.; Pedraza, F.; Buedia, H.F.; Gaitán-Sólis, E.; Beebe, S.E.; Gepts, P.; Tohme, J. (2003). Development of a genome-wide anchored microsatellite map for common bean (*Phaseolus vulgaris* L.). *Theoretical and Applied Genetics*, 107: 1362-1374.
- Blair, M.W.; Giraldo, M.C.; Buendia, H.F.; Tovar, E.; Duque, M.C.; Beebe, S.E. (2006a). Microsatellite marker diversity in common bean (*Phaseolus vulgaris* L.). *Theoretical and Applied Genetics*, 113:100-109.
- Blair, M.W.; Iriarte, G.; Beebe, S. (2006b). QTL analysis of yield traits in an advanced backcross population derived from a cultivated Andean X wild common bean (*Phaseolus vulgaris* L.) cross. *Theoretical and Applied Genetics*, 112:1149-1163.
- Blair, M.W.; Buendía, H.F.; Giraldo, M.C.; Métais, I.; Peltier, D. (2008). Characterization of AT-rich microsatellites in common bean (*Phaseolus vulgaris* L.). *Theoretical and Applied Genetics*, 118(1):91-103.
- Borém, A. (1998). Melhoramento de plantas. 2.ed. Viçosa: UFV, 453p.
- Brookes, A.J. (1999). The essence of SNPs. *Gene*, 234:177-186.
- Broughton, W.J., Hernández, G.; Blair, M.; Beebe, S.; Gepts, P.; Vanderleyden, J. (2003). Beans (*Phaseolus* spp.): model food legumes. *Plant Soil*, 252: 55–128.

-
- Brown, A.H.D.; Clegg, M.T.; Kahler, A.L.; Weir, B.S. (1990). Plant population genetics, breeding, and genetic resources. UC Davis: International Symposium on Population Genetics and Germoplasm Resources in Crop Improvement. 449p.
- Botstein, D.; White, R.L.; Skolnick, M.; Davis, R.W. (1980). Construction of a genetic map in man using restriction fragment length polymorphism. *American Journal Human Genetics*, 32:314-331.
- Buso, G.S.C.; Amaral, Z.P.S.; Brondani, R.P.V.; Ferreira, M.E. (2006). Microsatellite markers for the common bean *Phaseolus vulgaris*. *Molecular Ecology Notes*, 6:252-254.
- Caixeta, E.T.; Borém, A.; Kelly, J.D. (2005). Development of microsatellite markers based on BAC common bean clones. *Crop Breeding and Applied Biotechnology*, 5:125-133.
- Carneiro, M. S.; Vieira, M. L. C. (2002). Mapas genéticos em plantas. *Bragantia*, 61:89-100.
- Chacón, M.I.; Pickersgill, S.B.; Debouck, D.G. (2005). Domestication patterns in common bean (*Phaseolus vulgaris* L.) and the origin of the Mesoamerican and Andean cultivated races. *Theoretical and Applied Genetics*, 110:432-444.
- Cober, E.R.; Tanner, J.W. (1995). Performance of related indeterminate and tall determinate soybean lines in short-season areas. *Crop Science*, 35:361-364.
- Condit, R.; Hubbell, S.P. (1991). Abundance and DNA sequence of two-base regions in tropical tree genomes. *Genome*, 34:66-71.
- Emerson, R.A. (1909). Factors for mottling in beans. *American Breeders Association*, 5:368-376.
- Falconer, D.S.; Mackay, T.F.C. (1996). Introduction to Quantitative Genetics. 4.ed. London: Longman Scientific e Technical, 464p.
- Faleiro, F.G.; Ragagnin, V.A.; Moreira, M.A.; Barros, E.G. (2004). Use of molecular markers to accelerate the breeding of common bean lines resistant to rust and anthracnose. *Euphytica*, 138:213-218.
- Ferreira, M.E.; Grattapaglia, D. (1995). Introdução ao Uso de Marcadores RAPD, RFLP em Análise Genética. Brasília: EMPRAPA-CENARGEN, 220p.
- Freyre, C.I.; Skroch, P.; Geffroy, V.; Adam-Blondon, A.F.; Shirmohamadali, A.; Johnson, W.; Llaca, V.; Nodari, R.; Pereira, P.; Tsai, S.M.; Tohme, J.; Dron, M.; Nienhuis, J.; Vallejos, C.; Gepts, P. (1998). Towards an integrated linkage map of common bean.
-

4. Development of a core map and alignment of RFLP maps. *Theoretical and Applied Genetics*, 97:847-856.
- Gaitán-Solís, E.; Duque, M.C.; Edwards, K.J.; Tohme, J. (2002). Microsatellite in Common Bean (*Phaseolus vulgaris*): Isolation, Characterization, and Cross-Species Amplification in- *Phaseolus* ssp. *Crop Science*, 42: 2128-2136.
- Garcia, A.A.F.; Benchimol, L.L.; Barbosa, A.M.M.; Geraldi, I.O.; Souza Jr., C.L.; Souza, A.P. (2004). Comparison of RAPD, RFLP, AFLP and SSR markers for diversity studies in tropical maize inbred lines. *Genetics and Molecular Biology*, 27:579-588.
- Gepts, P.; Bliss, F.A. (1986) Phaseolin variability among wild and cultivated common beans (*Phaseolus vulgaris*) from Colombia. *Economic Botany*, 40:469-478
- Gepts, P.R.; Nodari, R.; Tsai, R.; Koinange, E.M.K.; Llaca, V.; Gilbertson, R.; Guzman, P. (1993). Linkage Mapping in Common Bean. *Annual Report Bean Improvement Cooperative*, 36:24-38.
- Gepts, P. (1998). Origin and evolution of common bean: past events and recent trends. *Hortscience*, 33:1124-1130.
- Giordano, M., Oefner, P.J., Underhill, P.A., Sforza, L.L.C., Tosi, R., Richiardi, P.M. (1999). Identification by denaturing high-performance liquid chromatography of numerous polymorphisms in a candidate region for multiple sclerosis susceptibility. *Genomics*, 56:1069-1076.
- Grisi, M.C.M.; Blair, M.W.; Gepts, P.; Brondani, C.; Pereira, P.A.A.; Brondani, R.P.V. (2007). Genetic mapping of a new set of microsatellite markers in a reference common bean (*Phaseolus vulgaris*) population BAT93 x Jalo EEP558. *Genetics and Molecular Research*, 6:691-706.
- Guerra-Sanz, J.M. (2004). New SSR markers of *Phaseolus vulgaris* from sequence databases. *Plant Breeding*, 123: 87-89.
- Guidolin, A.F. (2003). Regeneração de plantas de *Phaseolus vulgaris* L. a partir de calos e transformação genética via *Agrobacterium*. Piracicaba: CENA, 100p. Tese doutorado.
- Griffiths, A.J.F.; Miller, J.H.; Suzuki, D.T.; Lewontin, R.C.; Gelbart, W.M. (2000). An introduction to genetic analysis. New York: W. H. Freeman, 1-103 p.
- Haldane, J.B.S. (1919). The combination of linkage values, and the calculation of distance between the loci of linked factors. *Journal of Genetics*, 8:299-309.

- Hanai, L. R.; Campos, T.; Camargo, L. E. A.; Benchimol, L. L.; Souza, A. P.; Melloto, M.; Carbonell, S. A. M.; Chioratto, A. F.; Consoli, L.; Formighieri, E. F.; Siqueira, M.; Tsai, S. M.; Vieira, M. L. C. (2007). Development, characterization, and comparative analysis of polymorphism at common bean SSR loci isolated from genic and genomic sources. *Genome*, 50:266-277.
- Harper, G.L.; Maclean, N.; Goulson, D. (2003). Microsatellite markers to assess the influence of population size, isolation and demographic change on the genetic structure of the UK butterfly *Polyommatus bellargus*. *Molecular Ecology*, 12:3349-3357.
- Hougaard, B.K.; Madsen, L.H.; Sandal, N.; Moretzsohn, M. de C.; Fredslund, J.; Schauser, L.; Nielsen, A.M.; Rohde, T.; Sato, S.; Tabata, S.; Bertoli, D.J.; Stougaard, J. (2008). Legume Anchor Markers Link Syntenic Regions Between *Phaseolus vulgaris*, *Lotus japonicus*, *Medicago truncatula* and *Arachis*. *Genetics*, 179:2299-2312.
- Jarne, P.; & Lagoda, P.J.L. (1996). Microsatellites, from molecules to populations and back. *Trends in Ecology and Evolution*, 11:424-429.
- Johannsen, W. (1909) Elemente der exakten Erblchkeitslehre. Gustav Fischer, Jena
- Johnson, W.C.; Gepts, P. (2002). The role of epistasis in controlling seed yield and other agronomic traits in an Andean · Mesoamerican cross of common bean (*Phaseolus vulgaris* L.). *Euphytica*, 125:69–72.
- Jung, G.; Coyne, D.P.; Skroch, P.W.; Nienhuis, J.; Arnaud-Santana, E.; Bokosi, J.; Ariyaratne, H.M.; Steadman, J.R.; Beaver, J.S.; Kaeppler, S.M. (1996). Molecular markers associated with plant architecture and resistance to common blight, web blight, and rust in common beans. *Journal of the American Society for Horticultural Science*, 121: 794–803.
- Kepler, E.F. (2000). Melhoramento genético animal no Brasil. Embrapa Gado de Corte, 88p.
- Khairallah, M.; Sears, B.; Adams, M. (1992). Mitochondrial restriction fragment length polymorphisms in wild *Phaseolus vulgaris* L.: insights on the domestication of the common bean. *Theoretical and Applied Genetics*, 84:915–922
- Kijas, J.M.H.; Fowler, J.C.S.; Garbett, C.A.; Thomas, M.R. (1994). Enrichment of microsatellites from the citrus genome using biotinylated oligonucleotide sequences bound to streptavidin-coated magnetic particles. *BioTechniques*, 1:657-662.

-
- Koenig, R.L.; Gepts, P. (1989). Allozyme diversity in wild *Phaseolus vulgaris*: further evidence for two major centers of genetic diversity. *Theoretical and Applied Genetics*, 78:809–817.
- Koinange, E.M.K.; Singh, S.P.; Gepts, P. (1996). Genetic control of the domestication syndrome in common bean. *Crop Science*, 36:1037–1045.
- Kosambi, D.D. (1944). The estimation of map distances from recombinant values. *Annals of Eugenics, Cambridge, Inglaterra*, 12:172-175.
- Kumar L.S. (1999). DNA markers in plant improvement: an overview. *Biotechnology Advances*, 17:143-182.
- Kwak, M.; Velasco, D.; Gepts, P. (2008). Mapping Homologous Sequences for Determinacy and Photoperiod Sensitivity in Common Bean (*Phaseolus vulgaris*). *Journal of Heredity*, 99:283-291.
- Kwok, P.Y., Gu, Z. Single nucleotide polymorphism libraries: why and how are we building them? *Molecular Medicine Today*, 12:538-543.
- Lamprecht, H. (1935). Zur Genetik von *Phaseolus vulgaris*. XII. ber die Vererbung der Blüten- und Stammfarbe. *Hereditas*, 21:129–166.
- Lander, E.S.; Green, P.; Abrahamson, J.; Barlow, A.; Daly, M.J.; Lincoln, S.E.; Newburg, L. (1987). Mapmaker: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics*, 1:174-181.
- Lewis, G., Schrire, B.; Mackinder, B.; Lock, M. (2005). Legumes of the World. The Royal Botanic Gardens, Kew, Richmond, UK.
- Levinson, G. & Gutman, G.A. (1987). Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Molecular Biology and Evolution*, 4:203-21.
- Lincoln, S.E.; Daly, M.J.; Lander, E.S. (1992). Constructing genetic maps with Mapmaker Exp 3.0. 3.ed. Cambridge: Whitehead Institute for Biometrical Research, 230p.
- Lynch, M.; Walsh, B. (1998). Genetics and analysis of quantitative traits. Sunderland: Sinauer Associates. 980p
- Maras, M.; Susnik, S.; Sustar-Vozlic, J.; Meglic, V. (2006). Temporal Changes in Genetic Diversity of Common Bean (*Phaseolus vulgaris* L.) Accessions Cultivated between 1800 and 2000. *Russian Journal of Genetics*, 42:775-782.
-

- Martínez-Castillo, J.; Zizumbo-Villareal, D.; Gepts, P.; Colunga-García Marín, P. (2007). Gene Flow and Genetic Structure in the Wild-Weedy-Domesticated Complex of *Phaseolus lunatus* L. in its Mesoamerican Center of Domestication and Diversity. *Crop Science*, 47:58-66.
- McClellan, P.E.; Lee, P.K.; Otto, C.; Gepts, P.; Bassett, M.J. (2002). Molecular and phenotypic mapping of genes controlling seed coat pattern and color in common bean (*Phaseolus vulgaris* L.). *The Journal of Heredity*, 93:148-152.
- Melotto, M.; Monteiro-Vitorello, C.B.; Bruschi A.G.; Camargo, L.E. (2005). Comparative bioinformatics analysis of genes expressed in common bean (*Phaseolus vulgaris* L.) seedlings. *Genome*, 48:562-570.
- Métais, I.; Hamon, B.; Jalouzot, R.; Peltier, D. (2002). Structure and level of genetic diversity in various bean types evidenced with microsatellite markers isolated from a genomic enriched library. *Theoretical and Applied Genetics*, 104:1346-1352.
- Miklas, P.N.; Coyne, D.P.; Grafton, K.F.; Mutlu, N.; Reiser, J.; Lindgren, D.T.; Singh, S.P. (2003). A major QTL for common bacterial blight resistance derives from the common bean great northern landrace cultivar Montana No.5. *Euphytica*, 131:137-146.
- Moore, G.; Devos, K.M.; Wang Z.; Gale, M.D. (1995). Cereal Genome Evolution: Grasses, line up and form a circle. *Current Biology*, 5:737-9.
- Moore, S.S.; Hale, P.; Byrne, K. (1998). NCAM: a polymorphic microsatellite locus conserved across eutherian mammal species. *Animal Genetics*, 29:33-36.
- Mullis, K.; Fallona, F. (1987). Specific synthesis of DNA in vitro via polymerase catalysed chain reaction. *Methods Enzymology*, 55:335-350.
- Murphy, R.W.; Sites, J.W.Jr.; Buth, D.G.; Haufler, C.H. (1990). Proteins I: isozyme electrophoresis. In: Hillis, D.M.; Moritz, C. (Ed.) *Molecular Systematics*. Sunderland: Sinauer Associates, 45-126p.
- Nodari, R.O.; Tsai, S.M.; Gilbertson, R.L.; Gepts, P. (1993). Towards an integrated map of common bean-2: development of a RFLP-based linkage map. *Theoretical and Applied Genetics*, 85:513-520.
- Norton, J.B. (1915). Inheritance of habit in the common bean. *The American Naturalist*, 49:547-561.

- Oetting, W.S.; Lee, H.K.; Flanders, D.J.; Wiesner, G.L.; Sellers, T.A.; King, R.A. (1995). Linkage analysis with multiplexed short tandem repeat polymorphisms using infrared fluorescence and M13 tailed primers. *Genomics* 30: 450-458.
- Oliveira, L.K.; Melo, L.C.; Brondani, C.; Peloso, M.J.D.; Brondani, R.P.V.. (2008). Backcross assisted by microsatellite markers in common bean. *Genetics and Molecular Resources*, 7: 1000-1010.
- Pedrosa, A.; Vallejos, C.E.; Bachmair, A.; Schweizer, D. (2003). Integration of common bean (*Phaseolus vulgaris* L.) linkage and chromosomal maps. *Theoretical and Applied Genetics*, 106:205-212.
- Prakken, R. (1972). Inheritance of colours in *Phaseolus vulgaris* L. III. On genes for red seed coat colour and a general synthesis. *Meded Landbouwhogeschool Wageningen*, 29:1-82.
- Powell, W., Morgante, M.; Andre, C.; Hanafey, M.; Vogel, J.; Tingey, S.; Rafalski, A. (1996a). The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Molecular Breeding*, 2:225-238.
- Powell, W.; Machray, G.C.; Provan, J. (1996b). Polymorphism revealed by simple sequence repeats. *Trends in Plant Science*, 1:215-222.
- Primmer, C.R.; Raudsepp, T.; Chowdhary, B.P. Møller, A.P.; Ellegren, H. 1997. Low frequency of microsatellites in the Avian Genome. *Genome Research*, 7:471-482.
- Ramirez, M.; Graham, M.A.; Blanco-López, L.; Silvente, S.; Medrano-Soto, A.; Blair, M.W.; Hernández, G.; Vance, C.P.; Lara, M. (2005). Sequencing and Analysis of Common Bean ESTs. Building a Foundation for Functional Genomics. *Plant Physiology*, 137:1211-1227.
- Refseth, U.H.; Fangan, B.M.; Jakobsen, K.S. (1997). Hybridization capture of microsatellites directly from genomic DNA. *Electroforesis*, 18: 1519-1523.
- Richards, A.J. (1997). Plant Breeding Systems. London: Stanley Thornes Pub, 529p.
- Rico, C.; Zadworny, D.; Kuhnlein, U.; Fitzgerald, G.J. (1993). Characterization of hypervariable microsatellite loci in the threespine stickleback *Gasterosteus aculeatus*. *Molecular Ecology*, 2:271-272.
- Rodríguez-Suárez, C.; Méndez-Vigo, B.; Pañeda, A.; Ferreira, J.J.; Giraldez, R. (2006). A genetic linkage map of *Phaseolus vulgaris* L. and localization of genes for specific resistance to six races of anthracnose (*Colletotrichum lindemuthianum*). *Theoretical and Applied Genetics*, 114:713-722.
-

- Schlötterer, C.; Tautz, D. (1992). Slippage synthesis of sample sequence DNA. *Nucleic Acids Research*, 20:211-215.
- Schlötterer, C. (2000). Evolutionary dynamics of microsatellite DNA. *Chromosoma*, 109:365-371.
- Schlueter, J.A.; Goicoechea, J.L.; Collura, K.; Gill, N.; Lin, J.; Yu, Y.; Kudrna, D.; Zuccolo, A.; Vallejos C.E.; Muñoz-Torres, M.; Blair, M.W.; Tohme, J.; Tomkins, J.; McClean, P.; Wing, R.A.; Jackson, S.A. (2008). BAC-end Sequence Analysis and a Draft Physical Map of the Common Bean (*Phaseolus vulgaris* L.) Genome. *Tropical Plant Biology*, 1:40-48
- Schuster, I.; Cruz, C.D. (2004). Estatística genômica - aplicada a populações derivadas de cruzamentos controlados. Viçosa: UFV, 568p.
- Silva, M.V. da (2000). Identificação de marcador RAPD ligado ao alelo co-7 de resistência do feijão ao agente causal da antracnose. Lavras: UFLA, 41p. Tese Mestrado.
- Singh, S.; Nodari, R.; Gepts, P. (1991a). Genetic diversity in cultivated common bean. II. Marker-based analysis of morphological and agronomic traits. *Crop Science*, 31:19-23.
- Singh, S.P.; Gepts, P.; DeBouck, D. (1991b). Races of common bean (*Phaseolus vulgaris*, *Fabaceae*). *Economic Botany*, 45: 379-396.
- Singh, S.P. (1999). Production and Utilization. In: Singh, S.P.(ed), Common Bean Improvement in the Twenty-First Century. Kluwer, Dordrecht, Netherlands, 1-24p.
- Southern, E.M. (1975). Detection of specific sequences among DNA fragments separated by electrophoresis. *Journal of Molecular Biology*, 98:503-517.
- Stam, P. (1993). Construction of integrated genetic maps by means of a new computer package: JoinMap. *Plant Journal*, 3:739-744.
- Surface, F.M. (1916). A note on the inheritance of eye pattern in beans and its relation to type of vine. *American Naturalist*, 50:577-586.
- Tanksley, S.D. (1993). Mapping polygenes. *Annual Review of Genetics*, 27:205-233.
- Ta'an, B.; Michaels, T.E.; Pauls, K.P. (2001). Identification of markers associated with quantitative traits in beans. *Annual Report of the Bean Improvement Cooperative*, 44:9-10.
- Tar'an, B.; Michaels Thomas, E.; Pauls, K.P. (2002). Genetic Mapping of Agronomic traits in Common Bean. *Crop Science*, 42:544-556.

- Tarlan, B.; Michaels, T.E.; Pauls, K.P. (2001). Mapping genetic factors affecting the reaction to *Xanthomonas axonopodis* pv. phaseoli in *Phaseolus vulgaris* L. under field conditions. *Genome*, 44:1046-1056.
- Tautz D. & Renz M. (1984). Simple sequence repeats are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Research*, 12:4127-4137.
- Tschermak, E. von. (1912). Bastardierungsversuche an Levkojen, Erbsen und Bohnen mit Rücksicht auf die Faktorenlehre. *Molecular and General Genetics*, 7:81-234.
- Theissen, G.; Becker, A.; Di Rosa, A.; Kanno, A.; Kim, J.T.; Münster, T.; Winter, K.; Saeldler, H. (2000). A short history of MADSbox genes in plants. *Plant Molecular Biology*, 42:115-149.
- Vallejos, C. E.; Sakiyama, N. S.; Chase, C. D. (1992). A molecular marker-based linkage map of *Phaseolus vulgaris* L. *Genetics*, 131: 733-740.
- Vieira, C. (1988). Doenças e pragas do feijoeiro. Universidade Federal de Viçosa, Viçosa, MG, Brasil.
- Vieira, C. (2004). Memórias de meio século de estudo sobre a cultura do feijão. Universidade Federal de Viçosa, Viçosa, MG, Brasil.
- Wallace, D.H.; Yourstone, K.S.; Masaya, P.N.; Zobel, R.W. (1993). Photoperiod gene control over partitioning between reproductive and vegetative growth. *Theoretical and Applied Genetics*, 86:6-16.
- Weber, R.D.; May, P.E. (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *American Journal Human Genetics*, 44:388-396.
- Weeden, N.F.; Muehlbauer, F.J.; Ladizinsk, G. (1992). Extensive conservation of linkage relationships between pea and lentil genetic maps. *Journal of Heredity*, 83:123-129.
- Williams, J.K.F.; Kubelik, A.R.; Livak, K.G.; Rafalki, J.A.; Tingey S.V. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research*, 18:6531-6535.
- Winter, K.U.; Becker, A.; Münster, T.; Kim, J.T.; Saeldler, H.; Theissen, G. (1999). MADS-box genes reveal that gymnosperms are more closely related to conifers than to flowering plants. *Proceedings of the National Academy of Sciences of the United States of America*, 96:7342-7347.
-

- Yaish, M.W.F. & Pérez de la Vega, M. (2003). Isolation of (GA)_n microsatellite sequences and description of a predicted MADS-box sequence isolated from common bean (*Phaseolus vulgaris* L.). *Genetics and Molecular Biology*, 26 (3): 337-342.
- Yang, G.P.; Shagai Maroof, M.A.; Xu, C.G.; Zhang, Q.; Biyashev, R.M. (1994). Comparative analysis of microsatellite DNA polymorphism in landraces and cultivars of rice. *Molecular and General Genetics*, 245:187-194.
- Yu, K.; Park, S.J.; Poysa, V. (1999). Abundance and variation of microsatellite DNA sequences in beans (*Phaseolus vulgaris* and *Vigna*). *Genome*, 42:27-34.
- Yu, K.; Park, S.J.; Poysa, V.; Gepts, P. (2000). Integration of simple sequence repeat (SSR) markers into a molecular linkage map of common bean (*Phaseolus vulgaris* L.). *The Journal of Heredity*, 91:429-434.
- Zabeau, M and P. Vos. (1993). Selective restriction fragment amplification: a general method for DNA fingerprinting. European Patent Office, nº0 534 858 A1.
- Zhang, X.; Blair, M.W.; Wang, S. (2008). Genetic diversity of Chinese common bean (*Phaseolus vulgaris* L.) landraces assessed with simple sequence repeat markers. *Theoretical and Applied Genetics*, 117:629-640.
- Zhao, X.; Kochert, G. (1992). Characterization and genetic mapping of a short, highly repeat, interspersed DNA sequence from rice (*Oryza sativa* L.). *Molecular and General Genetics*, 231:353-359.
- Zhao, X.; Kochert, G. (1993). Phylogenetic distribution and genetic mapping of a (GGC)_n microsatellite from rice (*Oryza sativa* L.). *Plant Molecular Biology*, 21:607-614.