

BÁRBARA MÜLLER SALOMÃO DE FARIA

**CARACTERIZAÇÃO DE RECURSOS GENÔMICOS DO FEIJOEIRO COMUM  
(*Phaseolus vulgaris* L.) PARA O DESENVOLVIMENTO DE SSRs E SNPs  
ÚTEIS AO ESTUDO GENÉTICO E MELHORAMENTO DA CULTURA**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Magister Scientiae*.

VIÇOSA  
MINAS GERAIS – BRASIL  
2013

Ficha catalográfica preparada pela Seção de Catalogação e  
Classificação da Biblioteca Central da UFV

T

F224c  
2013 Faria, Bárbara Müller Salomão de, 1989-  
Caracterização de recursos genômicos do feijoeiro comum (*Phaseolus vulgaris* L.) para o desenvolvimento de SSRs e SNPs úteis ao estudo genético e melhoramento da cultura / Bárbara Müller Salomão de Faria. – Viçosa, MG, 2013.  
x, 121 f. : il. (algumas color.) ; 29 cm.

Inclui apêndices.

Orientador: Everaldo Gonçalves de Barros.

Dissertação (mestrado) - Universidade Federal de Viçosa.

Inclui bibliografia.

1. Feijão - Melhoramento genético. 2. Marcadores genéticos. 3. Genoma. 4. Leguminosa. I. Universidade Federal de Viçosa. Departamento de Biologia Geral. Programa de Pós-Graduação em Genética e Melhoramento. II. Título.

CDD 22. ed. 576.53

BÁRBARA MÜLLER SALOMÃO DE FARIA

**CARACTERIZAÇÃO DE RECURSOS GENÔMICOS DO FEIJOEIRO COMUM  
(*Phaseolus vulgaris* L.) PARA O DESENVOLVIMENTO DE SSRs E SNPs  
ÚTEIS AO ESTUDO GENÉTICO E MELHORAMENTO DA CULTURA**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Magister Scientiae*.

APROVADA: 29 de julho de 2013.

---

Evandro Novaes

---

Thiago Lívio P. Oliveira de Souza

---

Rosana Pereira Vianello  
(Coorientadora)

---

Everaldo Gonçalves de Barros  
(Orientador)

## AGRADECIMENTOS

Aos meus pais, à minha avó e à minha bisavó, por todo carinho, dedicação, apoio e amor incondicional. Especialmente, aos meus pais, por todos os esforços destinados a minha educação, pelos valores morais e de disciplina, para meu desenvolvimento pessoal e profissional.

Aos meus orientadores, Dr. Everaldo Gonçalves de Barros e Dra. Rosana Pereira Vianello, pela orientação constante e pela confiança. Além disso, agradeço o exemplo de ética e comprometimento profissional, e suas colaborações em minha formação humana e profissional.

Agradeço ao Dr. Maurilio Alves Moreira, ao Dr. Cláudio Brondani, ao Dr. Wellington Martins, à Dra. Tereza Cristina, à Dra. Aluana de Abreu, ao Dr. Georgios Pappas, ao Dr. Helton Pereira e ao Dr. Thiago de Souza, que colaboraram de diferentes formas com o desenvolvimento e conclusão deste trabalho.

Ao Msc. Tetsu Sakomoto pelas análises de bioinformática, pela atenção e dedicação nas análises dos resultados deste trabalho.

Ao Msc. Ivandilson Menezes pelo apoio, ensinamentos e análises de diversidade genética, fundamentais para a realização deste trabalho.

À Paula Arielle, à Dra. Gesimária Coelho, ao Guilherme Prado e ao Ueric de Souza pela enorme contribuição no desenvolvimento deste trabalho. Agradeço pelo apoio e por sempre me despertar a necessidade de aprender e compartilhar novos conhecimentos.

À Universidade Federal de Viçosa (UFV) e, particularmente, ao Instituto de Biotecnologia Aplicada à Agropecuária (BIOAGRO) e ao Programa de Pós-Graduação em Genética e Melhoramento (PPGGM), pela oportunidade de realização deste trabalho.

À Embrapa Arroz e Feijão, à Universidade Federal de Goiás (UFG), ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pelo espaço, apoio técnico e financeiro.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo auxílio financeiro, indispensável, pela concessão da bolsa de mestrado.

Aos estudantes e funcionários do Laboratório de Genética Molecular de Plantas do BIOAGRO/UFV e do Laboratório de Biotecnologia da Embrapa Arroz e Feijão, pelo ótimo ambiente de trabalho.

Aos professores do Programa de Pós-Graduação em Genética e Melhoramento da UFV, e aos colegas de curso, pelos importantes ensinamentos.

Aos meus amigos e familiares, pela dedicação, incentivo, amizade e companheirismo. Agradeço pelas cobranças, conselhos e pela segurança que me transmitiram nos momentos de maior necessidade.

Sinceros agradecimentos a todos que torceram, colaboraram, incentivaram e estimularam a realização deste trabalho.

## SUMÁRIO

RESUMO .....	vi
ABSTRACT .....	ix
INTRODUÇÃO GERAL .....	1
REFERÊNCIAS BIBLIOGRÁFICAS .....	6
CAPÍTULO 1: ANÁLISE DAS BAC-END SEQUENCES DE FEIJOEIRO COMUM PARA O DESENVOLVIMENTO E CARACTERIZAÇÃO DE BES- SSRs E PARA ELUCIDAR A COMPOSIÇÃO DO GENOMA .....	11
RESUMO .....	12
INTRODUÇÃO .....	13
MATERIAL E MÉTODOS .....	17
Sequenciamento e processamento das BAC-end sequences (BESs) .....	17
Identificação das BES-SSRs e desenho dos primers.....	17
Caracterização genética dos BES-SSRs.....	18
Análise estatística dos BES-SSRs .....	19
Anotação Funcional das BESs .....	19
Análise de Genômica Comparativa .....	20
RESULTADOS .....	21
Sequenciamento das BESs.....	21
Identificação e caracterização de BES-SSRs.....	21
Caracterização molecular dos BES-SSRs.....	22
Anotação das BESs .....	24
Comparação de genomas de leguminosas .....	26
DISCUSSÃO .....	28
Sequenciamento das BESs.....	28
Identificação e caracterização de BES-SSRs.....	28
Caracterização molecular dos BES-SSRs.....	32
Anotação das BESs .....	34
Genômica de leguminosas .....	36
CONSIDERAÇÕES FINAIS.....	38
TABELAS E FIGURAS .....	39
REFERÊNCIAS .....	46
APÊNDICES.....	54

CAPÍTULO 2: ANÁLISE DE DIVERSIDADE GENÉTICA E ESTRUTURAÇÃO POPULACIONAL EM GERMOPLASMA DE FEIJOEIRO COMUM UTILIZANDO SSRs e SNPs .....	59
RESUMO .....	60
INTRODUÇÃO .....	61
MATERIAL E MÉTODOS .....	65
Material Vegetal e Extração de DNA .....	65
Genotipagem de SSRs .....	65
Identificação dos SNPs para tecnologia <i>VeraCode</i> <sup>TM</sup> .....	66
Genotipagem de SNPs .....	67
Análise de Diversidade Genética .....	67
Análise de Estruturação Populacional .....	68
Análise de Desequilíbrio de Ligação .....	69
Alinhamento dos SSRs e SNPs no genoma de feijoeiro comum .....	69
RESULTADOS .....	70
Diversidade Genética .....	70
Estruturação Populacional .....	75
Desequilíbrio de Ligação .....	77
Distribuição dos SSRs e SNPs no genoma do feijoeiro comum .....	78
DISCUSSÃO .....	79
Diversidade Genética .....	79
Estruturação Genética .....	83
CONSIDERAÇÕES FINAIS .....	87
TABELAS E FIGURAS .....	88
REFERÊNCIAS .....	99
APÊNDICES .....	106
CONCLUSÕES GERAIS .....	120

## RESUMO

FARIA, Bárbara Müller Salomão de, M.Sc., Universidade Federal de Viçosa, Julho de 2013. **Caracterização de recursos genômicos do feijoeiro comum (*Phaseolus vulgaris* L.) para o desenvolvimento de SSRs e SNPs úteis ao estudo genético e melhoramento da cultura.** Orientador: Everaldo Gonçalves de Barros. Coorientadores: Rosana Pereira Vianello e Maurilio Alves Moreira.

Este trabalho apresenta resultados de pesquisas desenvolvidas no âmbito do Projeto “Consórcio internacional para o sequenciamento do genoma e do transcriptoma do feijoeiro comum (*Phaseolus vulgaris*)” (Prosul/CNPq e CYTED) conduzido na Embrapa Arroz e Feijão (CNPAP) e na Universidade Federal de Viçosa (UFV), sendo este composto por dois subprojetos distintos: “Sequenciamento e caracterização das pontas de BACs” e “Identificação e desenvolvimento de marcadores SNPs em feijoeiro comum”. A dissertação foi estruturada em dois capítulos, cada um derivado de um subprojeto, contendo os objetivos e os principais resultados apresentados a seguir: Capítulo 1- Com o intuito de ampliar e aprofundar o conhecimento sobre a estrutura e composição do genoma de *P. vulgaris* este estudo teve como objetivo sequenciar uma biblioteca BAC (*Bacterial Artificial Chromosome*) a partir de suas extremidades (BESs – *BAC-end sequences*) e analisar as sequências quanto à presença de SSRs e a anotação genômica. Ao todo, 52.270 BESs foram geradas e processadas equivalendo a 32 Mpb (6,5%) do genoma de *P. vulgaris*, com conteúdo de GC estimado em 39%. Foram encontrados um total de 3.789 BES-SSRs, um a cada 8,36 kb. Destes, 2.000 foram adequados para o desenvolvimento de marcadores moleculares, dos quais 194 foram avaliados e 40 deles foram caracterizados quanto a aspectos genéticos operacionais. Das 52.270 BESs, aproximadamente 2% continham sequências que codificavam fatores de transcrição e 3% continham elementos transponíveis. Através da comparação com o banco de dados não-redundante de proteínas do NCBI, foi possível identificar funções putativas para 24.321 BESs, contabilizando 46,53% do total de sequências analisadas. Com base nos termos do *Gene Ontology* (GO), foram atribuídas categorias funcionais a 19.363 BESs inseridas em processos biológicos (52%), função molecular (65%) e componente celular (22%). Este estudo permitiu identificar com sucesso

BES-SSRs altamente polimórficos, buscando motivos SSRs de tri- a hexanucleotídeos, agregando características genéticas adequadas e com potencial de serem utilizados no programa de melhoramento genético de feijoeiro comum. Adicionalmente, as BESs geradas e processadas foram integradas ao projeto internacional de sequenciamento do genoma de feijoeiro comum auxiliando na composição e montagem do mesmo. Capítulo 2- O objetivo desse estudo foi avaliar e comparar o poder de informação genética de um grupo de 58 marcadores SSRs (24 SSRs-di – microssatélites dinucleotídeo e 34 BES-SSRs – microssatélites tri- a hexanucleotídeo) e 345 SNPs para aplicações no melhoramento genético de *P. vulgaris*. Foram avaliados 88 genótipos representativos de 55 acessos melhorados e 33 variedades tradicionais, incluindo oito combinações biparentais composta por cruzamentos inter- e intra-*pool* gênico. Os SSRs-di apresentaram maior média de alelos por loco gênico (9,916) e revelaram maior diversidade genética média (72,1%), sendo o grupo de marcadores com o maior poder de discriminação genética entre indivíduos. Entre os 58 SSRs, 14 com uma elevada média de alelos privativos (14,93/loco) e diversidade genética (84%) possibilitaram a discriminação individual dos 88 genótipos. Os SNPs polimórficos entre cruzamentos biparentais, intra- (17,7%) e inter-*pool* gênico (78,2%), foram avaliados quanto à aplicação prática no mapeamento genético. As abordagens utilizadas para inferir a estrutura genética populacional apresentaram resultados similares entre os grupos de marcadores, discriminando os genótipos nos *pools* gênicos Mesoamericano e Andino, com a maior diferenciação genética ( $K = 2$ ,  $F_{ST} = 0,895$ ) apresentada pelos SNPs. Os SSRs-di possibilitaram discriminar dentro do *pool* gênico Mesoamericano o germoplasma melhorado e tradicional ( $K = 3$ ). O desequilíbrio de ligação testado para todos os marcadores foi elevado, sendo maior para os SNPs (84,92%), reduzindo significativamente quando avaliado separadamente para os genótipos Andinos e Mesoamericanos. A distribuição dos SSRs e SNPs foi ampla e aleatória em todo o genoma de *P. vulgaris*. Em termos de resultados práticos para o programa de melhoramento, neste estudo foram derivados painéis de genotipagem operacionais baseado em SSRs e SNPs com alto e eficiente poder de discriminação do germoplasma por origem. Adicionalmente, um conjunto de SNPs polimórficos entre diversas populações biparentais representa uma aplicação adicional dessa tecnologia para geração de mapas

genéticos de alta densidade compreendendo uma proporção substancial de marcadores compartilhados entre cruzamentos. Esse estudo contribui para a integração crescente da genômica nos programas de melhoramento, aliando metodologias de genotipagem acessíveis e a baixos custos que permitem amostrar de modo eficiente e/ou amplo o genoma de feijoeiro comum.

## ABSTRACT

FARIA, Bárbara Müller Salomão de, M.Sc., Universidade Federal de Viçosa, July, 2013. **Genomic resources characterization of the common bean (*Phaseolus vulgaris* L.) for the development of useful SSRs and SNPs to the genetic and breeding study of the crop.** Adviser: Everaldo Gonçalves de Barros. Co-advisers: Rosana Pereira Vianello and Maurilio Alves Moreira.

The present work reports results from the research developed at the "International consortium for the sequencing of the genome and transcriptome of the common bean (*Phaseolus vulgaris*)" (Prosul/CNPq and CYTED) project conducted at Embrapa Rice and Beans (CNPq) and the Federal University of Viçosa (UFV), which is composed of two distinct subprojects: "Sequencing and characterization of the BACs-ends" and "Identification and development of SNPs markers in common bean". The dissert were structured in two chapters, each derived from their respective subprojects, containing the objectives and the main results presented below: Chapter 1- With the intention of amplifying and deepening the understanding about the *P. vulgaris* genome structure and composition, the goal of this study was to generate BAC-end sequences (BESs) from one BAC (Bacterial Artificial Chromosome) library and analyze the sequences for the presence of SSRs and genomic annotation. In total, 52,270 BESs were generated and processed equaling 32 Mbp (6.5%) of the *P. vulgaris* genome, with 39% of GC content. A total of 3,789 BES-SSRs were found, one for every 8.36 kb. Of these, 2,000 SSRs were appropriated for the development of molecular markers, of these 194 were evaluated and 40 were characterized for genetic and operational aspects. Of the 52,270 BESs, approximately, 2% contained sequences that encode transcription factors and 3% contained transposable elements. Through the comparison with the NCBI non-redundant protein database, it was possible to identify putative functions for 24,321 BESs, accounting for 46.53% of the total sequences analyzed. According to Gene Ontology (GO) terms functional categories were assigned to 19,363 BESs involved in biological process (52%), molecular function (65%) and cellular component (22%). This study allowed us to successfully identify BES-SSRs, highly polymorphic, searching for SSRs motifs of tri- to hexanucleotides, adding appropriate genetic features and with the potential to be used in the common

bean genetic breeding programs. Additionally, the generated and processed BESs were integrated into the international project of common bean genome sequencing assisting in the composition and assembly of the genome. Chapter 2- The objectives of this study were to evaluate and compare the informativeness of 58 SSRs (24 SSRs-di – dinucleotide microsatellites and 34 BES-SSRs – tri- to hexanucleotide microsatellites) and 345 SNPs, as suited tools to *P. vulgaris* breeding programs. A germplasm set of 88 genotypes compounded of 55 breeding material and 33 landraces was evaluated, including eight biparental combinations derived from inter and intra-gene pool crosses. The SSRs-di displayed higher average of alleles/locus (9.916) and gene diversity (72.1%), exhibiting a superior capacity to distinguish the whole group of genotypes. Fourteen SSRs, out of the 58, with a considerable number of private alleles (over 14.93/locus) and high levels of gene diversity (84%), discriminated all 88 genotypes. The polymorphic SNPs, among inter (78.2%) and intra-gene pool (17.7%) combinations, were evaluated concerning their employment in genetic mapping. The approaches utilized to infer the genetic population structure presented correspondent results among all classes of markers, differentiating the genotypes within Andean and Mesoamerican gene pools. The SNPs, however, evidenced a superior ( $K = 2$ ,  $F_{ST} = 0.895$ ) competence to differentiate between the two gene pools. Furthermore, the SSRs-di distinguished, within Mesoamerican gene pool, the breeding material and the landraces germplasm ( $K = 3$ ). Despite the high levels of linkage disequilibrium for both SSRs and SNPs, the latter yielded higher levels (84.92%). The linkage disequilibrium levels dropped when Andean and Mesoamerican genotypes were analyzed separately. The SSRs and SNPs distribution in *P. vulgaris* genome was abundant and random. Concerning the breeding program, SSRs and SNPs genotyping panels are now available, allowing the germplasm origin to be disclosed. In addition, a group of polymorphic SNPs, between several biparental populations, presents itself as an extension of such technology in the development of high-density genetic maps, with a substantial marker overlapping through crosses. This study contributes for the integration of genomic tools within breeding programs, connecting feasible and low cost techniques with efficient/wide genome sampling of the common bean.

## INTRODUÇÃO GERAL

O feijoeiro comum (*Phaseolus vulgaris* L.) é a espécie mais cultivada do gênero *Phaseolus* e a mais importante para o consumo direto por humanos, entre as leguminosas (Broughton *et al.* 2003; Ramirez-Vallejo & Kelly, 1998). É um dos produtos agrícolas de maior importância econômico-social no Brasil, sendo este o maior produtor mundial dessa leguminosa. A espécie *P. vulgaris* destaca-se como uma cultura de grande valor nutricional e como alimento acessível às populações de baixa renda (Broughton *et al.*, 2003). Os centros de diversidade do feijoeiro comum estão localizados no Equador e no norte do Peru, a partir dos quais o feijão silvestre foi disperso para o sul (América do Sul) e para o norte (América Central). A domesticação levou à sua separação e a formação de dois conjuntos gênicos distintos, um Andino (feijões de sementes médias e grandes) e outro Mesoamericano (feijões de sementes pequenas), com importantes centros secundários de diversidade formados na África, Brasil, Europa, Oriente Médio e América do Norte (Beebe *et al.*, 2000; Broughton *et al.*, 2003; Doyle & Luckow, 2003; Gepts, 1998; Gepts *et al.*, 2005; Kami *et al.*, 1995). A espécie *P. vulgaris* é diplóide ( $2n = 2x = 22$ ) e possui um genoma estimado entre 450 a 650 Mb (Bennett & Leitch, 1995), sendo considerado pequeno quando comparado com os de outras leguminosas. Segundo Melotto *et al.* (2005), há indicação que o genoma do feijão apresenta um baixo nível de duplicação gênica e poucas famílias gênicas. Nos últimos anos, grandes esforços têm sido realizados para desenvolver marcadores moleculares e sequenciar o genoma do feijoeiro comum, visando ampliar a exploração dos recursos genômicos dessa espécie. Dentre os marcadores moleculares, os mais utilizados nos estudos de genética molecular de plantas tem sido os marcadores microssatélites e os SNPs (*Single Nucleotide Polymorphisms*).

Os marcadores microssatélites são sequências simples repetidas, denominadas *Simple Sequence Repeats* (SSRs) (Tautz, 1989); ou *Short Tandem Repeats* (STRs) (Edwards *et al.*, 1991) e foram descritos como pequenas sequências contendo de um a seis nucleotídeos repetidos em *tandem* (Litt & Luty 1989; Tautz, 1989). Os microssatélites destacam-se dos demais marcadores por apresentarem um padrão de herança mendeliana do tipo codominante e detectarem múltiplas formas alélicas por loco gênico (Li *et*

al. 2004). Esses marcadores podem ser analisados via PCR (*Polymerase Chain Reaction*), utilizando pares de *primers* que flanqueiam as sequências repetidas, de uma maneira relativamente simples, envolvendo um baixo custo, e pequena quantidade de DNA. Além disso, os produtos da amplificação podem ser diretamente observados, após eletroforese em uma matriz de agarose ou poliacrilamida, via genotipagem manual ou semi-automatizada em analisador automático de DNA (Kalia *et al.*, 2011). Os microssatélites apresentam diversas classificações dependendo de seu tamanho e tipo de unidade repetitiva. Com relação ao número de nucleotídeos por unidade repetitiva, SSR são classificados em mono-, di-, tri-, tetra-, penta- ou hexanucleotídeos. Com base no arranjo de nucleotídeos nos motivos de repetição, Weber (1990) classificou os microssatélites como perfeitos, imperfeitos e compostos. Repetições perfeitas são arranjos em *tandem* de repetições de um único motivo; enquanto que nas repetições imperfeitas as repetições são interrompidas por motivos não repetidos ao longo de sua extensão. Os microssatélites compostos apresentam dois ou mais motivos básicos de repetição e estão presentes em conjunto em várias configurações (Kalia *et al.*, 2011).

Os microssatélites são abundantes e amplamente distribuídos ao longo de todo genoma (Tautz & Renz, 1984). Atualmente, devido ao crescente volume de dados oriundos do sequenciamento parcial ou total dos genomas de plantas, existem evidências que os SSRs constituem uma grande parte do DNA não codificante. No entanto, muitos relatos verificam que um grande número de SSRs está localizado nas regiões transcritas dos genomas, embora, o número de repetições dos SSRs nestas regiões seja relativamente baixo (Morgante *et al.*, 2002; Li *et al.*, 2004). Quando os SSRs estão associados às regiões transcritas do genoma estão presentes especialmente nas regiões não traduzidas (3' e 5'-UTRs) (Morgante *et al.*, 2002). Nos cereais (milho, trigo, cevada, sorgo e arroz), apenas 1,5-7,5% dos SSRs estão localizados em ESTs (*Expressed Sequence Tags*) (Kantety *et al.*, 2002; Thiel *et al.*, 2003). Marcadores EST-SSRs têm sido considerados menos polimórficos que os SSRs derivados de bibliotecas genômicas, devido à alta conservação de sequência em regiões transcritas (Cho *et al.*, 2000). Em plantas, observa-se uma maior proporção de dinucleotídeos com repetição AT (Morgante *et al.*,

2002; Powell *et al.*, 1996) e o trinucleotídeo mais frequente é AAG (Li *et al.*, 2004), embora, em cereais, o trinucleotídeo mais comum seja o CCG (Wang *et al.*, 2012; Varshney *et al.*, 2002). Os microssátelites tetra-, penta- e hexanucleotídeos são menos frequentes nos genomas de plantas, geralmente quanto maior o número de nucleotídeos na unidade repetitiva, menor é a sua frequência dentro do genoma (Kalia *et al.*, 2011). Na última década, diversos grupos de pesquisa vêm avançando no desenvolvimento de marcadores SSRs para o feijoeiro comum. Atualmente estão disponíveis mais de 2.000 marcadores originados a partir do sequenciamento genômico total ou parcial, bem como de bancos de sequências expressas (Garcia *et al.*, 2011).

Os SNPs são as mais frequentes formas de variação genética de um organismo, ocorrendo em abundância ao longo de todo o genoma (Ching *et al.*, 2002). Os SNPs se baseiam na detecção de polimorfismos resultantes da alteração de uma única base e constituem a base genética da maioria das variações alélicas. Devido à sua abundância no genoma, sua lenta taxa de mutação, e pela possibilidade de detecção automatizada, os SNPs vêm sendo cada vez mais utilizados como marcadores moleculares de grande aplicação em genética (McCouch *et al.*, 2010; Yu *et al.*, 2011). Os SNPs caracterizam-se por serem predominantemente bialélicos, ao contrário dos SSRs que são multialélicos. Por esse motivo, os SNPs fornecem menor informação de polimorfismo por loco gênico, que é compensado pela elevada abundância desses marcadores polimórficos distribuídos ao longo dos genomas (McCouch *et al.*, 2010; Jehan & Lakhanpaul, 2006). Os marcadores SNPs estão em estudos iniciais na maioria das espécies vegetais, incluindo *P. vulgaris*, enquanto as tecnologias de marcadores SSRs estão bem estabelecidas para uso generalizado de melhoramento molecular (Bohra *et al.*, 2011). Com os avanços das tecnologias de sequenciamento e da bioinformática, um grande número de marcadores SNPs está sendo desenvolvido para o feijoeiro comum (Souza *et al.*, 2012). Até o momento, a sua aplicação mais concreta tem sido no desenvolvimento de mapas genéticos com ampla cobertura para *P. vulgaris* (Galeano *et al.*, 2012).

Atualmente, bibliotecas genômicas derivadas de BAC (*Bacterial Artificial Chromosome*) têm sido construídas para muitas espécies de plantas, por serem ferramentas úteis para o mapeamento físico, clonagem baseada em

mapeamento e por auxiliar no sequenciamento, na análise funcional e estrutural do genoma (Kim *et al.*, 2007; Wang *et al.*, 2012). As sequências das extremidades de BAC ou BAC-*end sequences* (BESs) são componentes valiosos para auxiliar na montagem e determinação da sequência de grandes genomas ou na ancoragem de dados do sequenciamento *shotgun* de genomas inteiros (Goff *et al.*, 2002). Além disso, as BESs podem fornecer informações sobre a composição da sequência de genomas não sequenciados e serem utilizadas para gerar marcadores moleculares (Tomkins *et al.*, 2004; Shultz *et al.*, 2007). Nos últimos anos, grandes esforços têm sido realizados para o desenvolvimento de marcadores SSRs derivados de BESs (BES-SSRs) em plantas, como nas espécies *Glycine max* (Shultz *et al.*, 2007; Shoemaker *et al.*, 2008; Saini *et al.*, 2008), *P. vulgaris* (Córdoba *et al.*, 2010), *Cajanus cajan* (Bohra *et al.*, 2011), *Quercus robur* (Rampant *et al.*, 2011), *Cicer arietinum* (Thudi *et al.*, 2011), *Triticum aestivum* (Sehgal *et al.*, 2012), *Arachis hypogaea* (Wang *et al.*, 2012), *Juglans regia* (Wu *et al.*, 2012) e nos gêneros *Citrus* (Biswas *et al.*, 2012) e *Saccharum* (Kim *et al.*, 2013).

Diversas espécies de plantas apresentam elevada sintonia entre seus genomas, o que pode facilitar a transferência de informações entre espécies de plantas relacionadas. Isso tem sido bem caracterizado, especialmente em gramíneas, cujos genomas apresentam micro e macrossintonia (Devos & Gale, 2000; Paterson *et al.*, 2009). Contudo, estes estudos também revelam muitas exceções à conservação da sintonia, com frequentes rearranjos gênicos locais, incluindo inversão, duplicação, translocação e inserção/deleção. A conservação da sequência gênica em microregiões do genoma (microsintonia) em leguminosas tem se mostrado útil para a clonagem de genes envolvidos na simbiose com *Rhizobium* spp. em *Medicago sativa*, *Medicago truncatula*, *Lotus japonicus* e *Pisum sativum* (Endre *et al.*, 2002).

Apesar da grande importância econômica e social das leguminosas, o mapa genético integrado desta família apresenta-se em estudos iniciais. Segundo Choi *et al.* (2004), o número de pontos de comparação (marcadores âncoras) para mapas genéticos das duas leguminosas de maior importância econômica no mundo, soja e feijoeiro comum, foram somente 15 e 22, respectivamente. O sequenciamento completo do genoma da soja (Schmutz *et al.*, 2010) foi de extrema importância para estabelecer esta espécie como

referência para mais de 20.000 leguminosas. Este genoma, que tem um ancestral comum com outras espécies de feijão domesticado há 20 milhões de anos, permite integrar o conhecimento das características fenotípicas e mapeadas (Schmutz *et al.*, 2010). O conhecimento do nível de sintonia da soja com outras leguminosas pode facilitar a identificação de variações alélicas em genes de espécies relacionadas, que possam ser transferidos via engenharia genética, e introduzir características desejáveis na espécie, tais como tolerância a estresses bióticos e abióticos. Devido a sua importância no cenário mundial, a comunidade científica internacional que trabalha com leguminosas recomendou que a soja deve ser considerada uma planta modelo para estudos genéticos e moleculares, representando o grupo Phaseoloid, que contém a maioria das leguminosas economicamente importantes, como a própria soja, feijoeiro comum, feijão caupi e guandu (Gepts *et al.*, 2005). Atualmente, as leguminosas *L. japonicus* e *M. truncatula* possuem seus genomas completamente sequenciados e são também consideradas plantas modelo (Sato *et al.*, 2008; Young *et al.*, 2011).

Neste contexto, a integração do conhecimento já disponível sobre o genoma da soja e outras leguminosas utilizadas como plantas modelo representam um potencial real para auxiliar no avanço do conhecimento de *P. vulgaris* e o desenvolvimento de ferramentas moleculares úteis para o melhoramento genético dessa espécie. Assim, o estudo realizado neste trabalho visa à geração e caracterização de um banco de sequências genômicas de *P. vulgaris* derivadas da estratégia de sequenciamento das extremidades de clones de BACs, com ênfase na investigação de aspectos da estrutura e organização do genoma. Além de avaliar o potencial dos marcadores SSRs e SNPs em caracterizar a magnitude da diversidade e estruturação genética de um grupo de acessos de feijoeiro comum e, comparativamente, determinar o melhor sistema de genotipagem quanto ao poder de informação e facilidade de operacionalização para serem empregados de modo rotineiro por programas de melhoramento genético da espécie.

## REFERÊNCIAS BIBLIOGRÁFICAS

- BEEBE, S.; SKROCH, P. W.; TOHME, J. DUQUE, M. C.; PEDRAZA, F. & NIENHUIS, J. Structure of genetic diversity among common bean landraces of Middle American origin based on correspondence analysis of RAPD. **Crop Sci** v. 40, p. 264–273, 2000.
- BENNETT, M.D. & LEITCH, I.J. Nuclear DNA amounts in angiosperms. **Ann Bot.** v. 76, p. 113–176, 1995.
- BISWAS, M.K; CHAI, L; MAYER, C; XU, Q; GUO, W; DENG, X. Exploiting BAC-end sequences for the mining, characterization and utility of new short sequences repeat (SSR) markers in Citrus. **Mol Biol Rep.** v. 39, p. 5373–5386, 2012.
- BOHRA, A; DUBEY, A; SAXENA, RK; PENMETS, POORNIMA, RV; KUMAR, N; FARMER, AD; SRIVANI, G; UPADHYAYA, HD; GOTHALWAL, R; RAMESH, S; SINGH, D; SAXENA, K; KISHOR, PBK; SINGH, NK; TOWN, CD; MAY, GD; COOK, DR & VARSHNEY, RK. Analysis of BAC-end sequences (BESs) and development of BES-SSR markers for genetic mapping and hybrid purity assessment in pigeonpea (*Cajanus* spp.). **BMC Plant Biol.** v. 11:56. p. 11-15, 2011.
- BROUGHTON, W.J.; HERNÁNDEZ, G.; BLAIR, M.; BEEBE, S.; GEPTS, P., VANDERLEYDEN, J. Beans (*Phaseolus* spp.) – model food legumes. **Plant Soil.** v. 252, p. 55-128, 2003.
- CHING, A.; CALDWELL, KS; JUNG, M.; DOLAN, M; SMITH, O.S.H; TINGEY, S; MORGANTE, M. & RAFALSKI, A. J. SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. **BMC Genetics.** v. 3:19, 2002.
- CHO, Y.G.; ISHII, T.; TEMNYKH, S.; CHEN, X.; LIPOVICH, L.; MCCOUCH, S.R.; PARK, W.D.; AYRES, N.; CARTINHO, S. Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). **Theor Appl Genet.** v. 100, p. 713–722, 2000.
- CHOI, H.K.; MUN, J.H.; KIM, D.J.; ZHU, H.; BAEK, J.M.; MUDGE, J.; ROE, B.; ELLIS, N.; DOYLE, J.; KISS, G.B.; YOUNG, N.D.; COOK, D.R. Estimating genome conservation between crop and model legume species. **Proc Natl Acad Sci.** v. 101, p. 15289–15294, 2004.
- CÓRDOBA, JM; CHAVARRO, C; SCHLUETER, JA; JACKSON, AS; BLAIR, MW. Integration of physical and genetic maps of common bean through BAC-derived microsatellite markers. **BMC Genomics.** v. 11:436, 2010.
- DEVOS, K. M & GALE, M. D. Genome relationships: the grass model in current research. **Plant Cell.** v.12:5, p. 637-646, 2000.
- DOYLE, J. J. & LUCKOW, M. A. The rest of the iceberg. Legume diversity and evolution in a phylogenetic context. **Plant Physiology.** v. 131, p. 900–910, 2003.

EDWARDS, A.; CIVITELLO, A.; HAMMOND, H. A.; CASKEY, C. T. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. **Am J Hum Genet.** v. 49, p. 746–756, 1991.

ENDRE, G.; KERESZT, A.; KEVEI, Z.; MIHACEA, S.; KALO, P.; KISS, G. A receptor kinase gene regulating symbiotic nodule development. **Nature.** v. 417, p. 962–966, 2002.

GEPTS, P. Origin and evolution of common bean: past events and recent trends. **Hort Sci.** v. 33, p. 1124-1130, 1998.

GEPTS, P.; BEAVIS, W. D.; BRUMMER, E. C.; SHOEMAKER, R. C.; STALKER, H. T., WEEDEN, N. F. & YOUNG, N. D. Legumes as a Model Plant Family. Genomics for Food and Feed Report of the Cross-Legume Advances through Genomics Conference. **Plant Physiol.** v. 137, p. 1228–1235, 2005.

GOFF, S.A.; RICKE, D.; LAN, T.H.; PRESTING, G.; WANG, R.; DUNN, M.; GLAZEBROOK, J.; SESSIONS, A.; OELLER, P.; VARMA, H.; HADLEY, D.; HUTCHISON, D.; MARTIN, C.; KATAGIRI, F.; LANGE, B.M.; MOUGHAMER, T.; XIA, X.; BUDWORTH, P.; ZHONG, J.; MIGUEL, T.; PASZKOWSKI, U.; ZHANG, S.; COLBERT, M.; SUN, W.; CHEN, L.; COOPER, B.; PARK, S.; WOOD, T. C.; MAO, L.; QUAIL, P.; WING, R.; DEAN, R.; YU, Y.; ZHARKIKH, A.; SHEN, R.; SAHASRABUDHE, S.; THOMAS, A.; CANNINGS, R.; GUTIN, A.; PRUSS, D.; REID, J.; TAVTIGIAN, S.; MITCHELL, J.; ELDREDGE, G.; SCHOLL, T.; MILLER, R. M.; BHATNAGAR, S.; ADEY, N.; RUBANO, T.; TUSNEEM, N.; ROBINSON, R.; FELDHaus, J.; MACALMA, T.; OLIPHANT, A. & BRIGGS, S. A draft sequence of the rice genome (*Oryza sativa* L. ssp. japonica). **Science.** v. 296, p. 92–100, 2002.

JEHAN, T. & LAKHANPAUL, S. Single nucleotide polymorphism (SNP)-methods and applications in plant genetics: a review. **Indian J Biotechnol.** v. 5; p. 435–459, 2006.

KALIA, R. K.; RAI, M. K.; KALIA, S.; SINGH, R.; DHAWAN, A. K. Microsatellite markers: an overview of the recent progress in plants. **Euphytica.** v. 177, p. 309–334, 2011.

KAMI, J.; VELÁSQUEZ, V. B.; DEBOUCK, D. G. & GEPTS, P. Identification of presumed ancestral DNA sequences of phaseolin in *Phaseolus vulgaris*. **Proc Natl Acad Sci. USA.** v. 92, p. 1101–1104, 1995.

KANTETY, R.V.; ROTA, M.L.; MATTHEWS, D.E.; SORRELLS, M.E. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. **Plant Mol Biol.** v. 48:501–510, 2002.

KIM C; LEE TH; COMPTON RO; ROBERTSON JS; PIERCE GJ; PATERSON AH. A genome-wide BAC end-sequence survey of sugarcane elucidates genome composition, and identifies BACs covering much of the euchromatin. **Plant Mol Biol.** v.81:139–147, 2013.

KIM, H.; MIGUEL, P. S.; NELSON, W.; COLLURA, K.; WISSOTSKI, M.; WALLING, J. G.; KIM, J. P.; JACKSON, S. A.; SODERLUND, C. & WING, R. A.

Comparative Physical Mapping Between *Oryza sativa* (AA Genome Type) and *O. punctata* (BB Genome Type). **Genetics**. v.176, p 379–390, 2007.

LI, Y.C.; KOROL, A.B.; FAHIMA, T. & NEVO, E. Microsatellites within genes: structure, function, and evolution. **Mol Biol Evol**. v. 21, p. 991–1007, 2004.

LITT, M. & LUTY, J. A. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. **Am J Hum Genet**. v. 44, p. 397–401, 1989.

MCCOUCH SR, ZHAO K, WRIGHT M, TUNG C-W, EBANA K, THOMSON M, REYNOLDS A, WANG D, DECLERCK G, ALI MDL, MCCLUNG A, EIZENGA G & BUSTAMANTE C. Development of genome-wide SNP assays for rice. **Breeding Sci**. v. 60, p. 524–535, 2010.

MELOTTO, M.; MONTEIRO-VITORELLO, C. B.; BRUSCHI, A. G. & CAMARGO, L. E. A. Comparative bioinformatic analysis of genes expressed in common bean (*Phaseolus vulgaris* L.) seedlings. **Genome**. v. 48, p. 562–570, 2005.

MORGANTE, M.; HANAFEY, M.; POWELL, W. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. **Nat Genet**. v. 30, p. 194–200, 2002.

PATERSON, A.H.; BOWERS, J.E.; FELTUS, F.A.; TANG, H.; LIN, L.; WANG, X. Comparative genomics of grasses promises a bountiful harvest. **Plant Physiol**. v. 149, p. 125-131, 2009.

POWELL, W.; MACHRAY, G.C.; PROVAN, J. Polymorphism revealed by simple sequence repeats. **Trends Plant Sci**. v. 1, p. 215–222, 1996.

RAMIREZ-VALLEJO, P. & KELLY, J. D. Traits related to drought resistance in common bean. **Euphytica**. v. 99, p. 127–136, 1998.

RAMPANT, P.F.; LESUR, I.; BOUSSARDON, C.; BITTON, F.; MARTIN-MAGNIETTE, M.L.; BODÉNÈS, C.; PROVOST, G.L.; BERGÈS, H.; FLUCH, S.; KREMER, A.; & PLOMION, C. Analysis of BAC end sequences in oak, a keystone forest tree species, providing insight into the composition of its genome. **BMC Genomics**. v. 12:292, 2011.

SAINI, N.; SHULTZ, J.; & LIGHTFOOT, D. A. Re-annotation of the physical map of Glycine max for polyploid-like regions by BAC end sequence driven whole genome shotgun read assembly. **BMC Genomics**. v. 9:323, 2008.

SATO, S.; NAKAMURA, Y.; KANEKO, T.; ASAMIZU, E.; KATO, T. *et al.* Genome Structure of the Legume, *Lotus japonicus*. **DNA Research**. v. 15, p. 227–239, 2008.

SCHMUTZ, J.; CANNON, S.B.; SCHLUETER, J.; MA J.; MITROS, T.; NELSON, W.; HYTEN, D.L.; SONG, Q.; THELEN, JJ.; CHENG, J.; XU, D.; HELLSTEN, U.; MAY, GD.; YU, Y.; SAKURAI, T.; UMEZAWA, T.; BHATTACHARYYA, M.K.; SANDHU, D.; VALLIYODAN, B.; LINDQUIST, E.; PETO, M.; GRANT, D.; SHU, S.; GOODSTEIN, D.; BARRY, K.; FUTRELL-GRIGGS, M.; ABERNATHY, B.; DU, J.; TIAN, Z.; ZHU, L.; GILL, N.; JOSHI, T.;

LIBAULT, M.; SETHURAMAN, A.; ZHANG, X.C.; SHINOZAKI, K.; NGUYEN, H.T.; WING, R.A.; CREGAN, P.; SPECHT, J.; GRIMWOOD, J.; ROKHSAR, D.; STACEY, G.; SHOEMAKER, R.C.; JACKSON, S.A. Genome sequence of the palaeopolyploid soybean. **Nature**. v. 463, p.178–183, 2010.

SEHGAL, SK; LI, W; RABINOWICZ, PD; CHAN, A; ŠIMKOVÁ, H; DOLEŽEL, J; & GILL, BS. Chromosome arm-specific BAC end sequences permit comparative analysis of homoeologous chromosomes and genomes of polyploid wheat. **BMC Plant Biol**. v. 12:64, 2012.

SHOEMAKER, RC; GRANT, D; OLSON, T; WARREN, WC; WING, R; YU, Y; KIM, H-R; CREGAN, P; JOSEPH, B; FUTRELL-GRIGGS, M; NELSON, W; DAVITO, J; WALKER, J; WALLIS, J; KREMITSKI, C; SCHEER, D; CLIFTON, SW; GRAVES, T; NGUYEN, H; WU, X; LUO, M; DVORAK, J; NELSON, R; CANNON, S; TOMKINS, J; SCHMUTZ, J; STACEY, G; JACKSON, S. Microsatellite discovery from BAC end sequences and genetic mapping to anchor the soybean physical and genetic maps. *Genome*. v. 51, p. 294–302, 2008.

SHULTZ, J.L; KAZI, S.; BASHIR, R.; AFZAL J.A.; LIGHTFOOT D.A. The development of BAC-end sequence-based microsatellite markers and placement in the physical and genetic maps of soybean. **Theor Appl Genet**. v. 114, p.1081–1090, 2007.

TAUTZ, D. & RENZ, M. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. **Nucleic Acids Res**. v. 12, p. 4127–4138, 1984.

TAUTZ, D. Hypervariability of simple sequences as a general source for polymorphic DNA markers. **Nucleic Acids Res**. v. 17, p.6463–6471, 1989.

THIEL T, MICHALEK W, VARSHNEY RK, GRANER A. Exploiting EST databases for the development of cDNA derived microsatellite markers in barley (*Hordeum vulgare* L.). **Theor Appl Genet**. v.106, p. 411–422, 2003.

THUDI, M.; BOHRA, A.; NAYAK, S.N.; VARGHESE, N.; SHAH, T.M.; PENMETSA, R.V.; THIRUNAVUKKARASU, N.; GUDIPATI, S.; GAUR, P.M.; KULWAL, P.L.; UPADHYAYA, H.D.; KAVIKISHOR, P.B.; WINTER, P.; KAHL, G.; TOWN, C.D.; KILIAN, A.; COOK, D.R.; & VARSHNEY, R.K. Novel SSR Markers from BAC-End Sequences, DArT Arrays and a Comprehensive Genetic Map with 1,291 Marker Loci for Chickpea (*Cicer arietinum* L.). **PLoS ONE**. v.6:11, 2011.

TOMKINS, J.; FREGENE, M.; MAIN, D.; KIM, H.; WING, R.A.; TOHME, J. Bacterial artificial chromosome (BAC) library resource for positional cloning of pest and disease resistance genes in cassava (*Manihot esculenta* Crantz). **Plant Mol Biol**. v. 56, p. 555–561, 2004.

VARSHNEY, R.K.; THIEL, T.; STEIN, N.; LANGRIDGE, P.; GRANER, A. In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species. **Cell Mol Biol Lett**. v. 7, p. 537–546, 2002.

WANG, H.; PENMETSA, R.V.; YUAN, M.; GONG, L.; ZHAO, Y.; GUO, B.; FARMER, A.D.; ROSEN, B.D.; GAO, J.; ISOBE, S.; BERTIOLI, D.J.; VARSHNEY, R.K.; COOK, D.R.; HE, G. Development and characterization of BAC-end sequence derived SSRs, and their incorporation into a new higher density genetic map for cultivated peanut (*Arachis hypogaea* L.). **BMC Plant Biol.** v. 12:10, 2012.

WEBER, J.L. Informativeness of human (dC-dA) n (dG-dT) n polymorphisms. **Genomics.** v. 7, p. 524–530, 1990.

WU, J.; GU, Y.Q.; HU, Y.; YOU, F.M.; DANDEKAR, A.M.; LESLIE, C.A.; ARADHYA, M.; DVORAK, J.; & LUO, M-C. Characterizing the walnut genome through analyses of BAC end sequences. **Plant Mol Biol.** v. 78, p. 95–107, 2012.

YOUNG, N.D.; DEBELLÉ, F.; OLDROYD, G.E.; GEURTS, R.; CANNON, SB.; UDVARDI MK.; BENEDITO VA.; MAYER KF.; GOUZY J.; SCHOOF H.; VAN DE PEER Y.; PROOST S.; COOK DR.; MEYERS BC.; SPANNAGL M.; CHEUNG F.; DE MITA S.; KRISHNAKUMAR V.; GUNDLACH H.; ZHOU S.; MUDGE J.; BHARTI AK.; MURRAY JD.; NAOUMKINA MA.; ROSEN B.; SILVERSTEIN KA.; TANG H.; ROMBAUTS S.; ZHAO PX.; ZHOU P.; BARBE V.; BARDOU P.; BECHNER M.; BELLEC A.; BERGER A.; BERGÈS H.; BIDWELL S.; BISSELING T.; CHOISNE N.; COULOUX A.; DENNY R.; DESHPANDE S.; DAI X.; DOYLE JJ.; DUDEZ AM.; FARMER AD.; FOUTEAU S.; FRANKEN C.; GIBELIN C.; GISH J.; GOLDSTEIN S.; GONZÁLEZ AJ.; GREEN PJ.; HALLAB A.; HARTOG M.; HUA A.; HUMPHRAY SJ.; JEONG DH.; JING Y.; JÖCKER A.; KENTON SM.; KIM DJ.; KLEE K.; LAI H.; LANG C.; LIN S.; MACMIL SL.; MAGDELENAT G.; MATTHEWS L.; MCCORRISON J.; MONAGHAN EL.; MUN JH.; NAJAR FZ.; NICHOLSON C.; NOIROT C.; O'BLENESS M.; PAULE CR.; POULAIN J.; PRION F.; QIN B.; QU C.; RETZEL EF.; RIDDLE C.; SALLET E.; SAMAIN S.; SAMSON N.; SANDERS I.; SAURAT O.; SCARPELLI C.; SCHIEX T.; SEGURENS B.; SEVERIN AJ.; SHERRIER DJ.; SHI R.; SIMS S.; SINGER SR.; SINHAROY S.; STERCK L.; VIOLLET A.; WANG BB.; WANG K.; WANG M.; WANG X.; WARFSMANN J.; WEISSENBACH J.; WHITE DD.; WHITE JD.; WILEY GB.; WINCKER P.; XING Y.; YANG L.; YAO Z.; YING F.; ZHAI J.; ZHOU L.; ZUBER A.; DÉNARIÉ J.; DIXON RA.; MAY GD.; SCHWARTZ DC.; ROGERS J.; QUÉTIÉ F.; TOWN CD.; ROE BA. The *Medicago* genome provides insight into the evolution of rhizobial symbioses. **Nature.** v. 480, p. 520–524, 2011.

YU, H.; XIE, W.; WANG, J.; XING, Y.; XU, C. Gains in QTL Detection Using an Ultra-High Density SNP Map Based on Population Sequencing Relative to Traditional RFLP/SSR Markers. **PLoS ONE.** v. 6:3: e17595, 2011.

## **CAPÍTULO 1**

**ANÁLISE DAS BAC-END SEQUENCES DE FEIJOEIRO COMUM PARA O  
DESENVOLVIMENTO E CARACTERIZAÇÃO DE BES-SSRs E PARA  
ELUCIDAR A COMPOSIÇÃO DO GENOMA**

## RESUMO

Com o intuito de ampliar e aprofundar no conhecimento sobre a estrutura e composição do genoma do feijoeiro comum este estudo teve como objetivo sequenciar uma biblioteca BAC (*Bacterial Artificial Chromosome*) a partir de suas extremidades (*BAC-end sequences* – BES), realizar a sua anotação genômica, identificar e caracterizar BES-SSRs. Ao todo, 52.270 BESs foram geradas e processadas equivalendo a 32 Mpb (6,5%) do genoma de *P. vulgaris*, com conteúdo de GC estimado em 39%. Foram encontrados um total de 3.789 BES-SSRs, um a cada 8,36 Kpb. Destes, 2.000 foram adequados para o desenvolvimento de marcadores moleculares, dos quais 194 foram avaliados e 40 deles, baseados em motivos do tipo tri-, tetra-, penta-, hexanucleotídeos e SSRs compostos, foram caracterizados quanto a aspectos genéticos e operacionais. A caracterização molecular dos BES-SSRs foram avaliados em 88 genótipos de feijoeiro comum (55 acessos melhorados e 33 variedades tradicionais) e as genotipagens foram conduzidas através de 10 sistemas semi-automatizados tetraplex. Entre os 40 BES-SSRs, 34 foram polimórficos com diversidade genética média ( $H_E$ ) de 0,487 e 5,6 alelos por loco. Os seis BES-SSRs com  $H_E \geq 0,7$  apresentaram comprimento do motivo maior que 40pb. Das 52.270 BESs, aproximadamente 2% continham seqüências que codificavam fatores de transcrição e 3% continham elementos transponíveis. No total, 47.194 (90,3%) das BESs apresentaram similaridade com GSSs (*Genome Survey Sequences*) e 24.321 (46,5%) BESs com proteínas putativas do banco de dados não-redundantes do *GenBank*. Com base nos termos do *Gene Ontology* (GO), foram atribuídas categorias funcionais a 19.363 BESs inseridas em processos biológicos (52%), função molecular (65%) e componente celular (22%). Este estudo permitiu identificar com sucesso BES-SSRs altamente polimórficos, buscando motivos SSRs de tri- a hexanucleotídeos, agregando características genéticas adequadas e com potencial de serem utilizados no programa de melhoramento genético de feijoeiro comum. Adicionalmente, as BESs geradas e processadas foram integradas ao projeto internacional de sequenciamento do genoma de feijoeiro comum auxiliando na composição e montagem do mesmo.

## INTRODUÇÃO

Os legumes, membros da família Leguminosae, fazem parte de uma das três maiores famílias do Reino Plantae (Doyle and Luckow 2003) compreendendo, aproximadamente, 18.000 espécies divididas nas subfamílias Mimosoideae, Caesalpinoideae e Papilionoideae. Esta última é a mais numerosa e inclui espécies de elevado valor econômico, como o amendoim (*Arachis hypogaea*), a soja (*Glycine max*), o feijoeiro comum (*Phaseolus vulgaris*) e o feijão-caupi (*Vigna unguiculata*) (Choi et al. 2004). Entre as leguminosas, a espécie *P. vulgaris* é a mais importante para o consumo direto humano em quase todos os países de clima tropical e subtropical, assumindo grande importância na alimentação humana por apresentar uma composição química que torna o seu consumo benéfico sob o ponto de vista nutricional. O feijão é usado como uma importante fonte de proteínas, que contém níveis adequados do aminoácido lisina, que exerce efeito complementar às proteínas dos cereais em dietas que combinam leguminosas e cereais (Broughton et al. 2003). O consumo per capita mundial desse legume (Kg/ano) varia consideravelmente entre países, com estimativas de 2 – 3,5 Kg nos EUA a, aproximados, 16 Kg no Brasil (FAO 2009). Os programas de melhoramento genético do feijoeiro comum, essencialmente públicos e de caráter multidisciplinar, atualmente contam com o apoio de ferramentas biotecnológicas, tornando mais eficiente e competitivo o processo de desenvolvimento de novas cultivares.

O tamanho do genoma haploide do feijoeiro comum é estimado em 550 milhões de pares de bases (Mpb), distribuídas em 11 cromossomos (Arumuganatham and Earle 1991). Diversos foram os atributos que viabilizaram o feijoeiro comum como um organismo experimental adequado para estudos moleculares e sequenciamento, como o tamanho de genoma comparável ao das espécies de leguminosas modelo *Medicago truncatula* e *Lotus japonicus* (Cannon et al. 2009) e o reduzido índice de duplicação do genoma (Melotto et al. 2005). Do ponto de vista do genoma estrutural, a espécie *P. vulgaris* apresenta uma posição privilegiada dentro do grupo dos legumes de clima tropical e, com base no estabelecimento de correlação entre genomas, um estudo recente revelou elevado grau de conservação de regiões transcritas

entre os genomas do feijoeiro comum e a soja (Schmutz et al. 2010). Decorrente de sua crescente importância no cenário mundial, em Março de 2012, dois genótipos representantes dos *pools* gênicos Mesoamericano e Andino de feijoeiro comum tiveram seus genomas sequenciados. O sequenciamento do genoma de uma linhagem Mesoamericana (BAT 93), resultante de uma iniciativa integrada por quatro países Iberoamericanos, incluindo Brasil, México, Espanha e Argentina, está disponível através de endereço eletrônico [<http://www.cytod.org/>]. O genoma da variedade Andina (G19833), cujo sequenciamento foi liderado pelos EUA, apoiado pelos institutos DOE-JGI e ARRA, encontra-se publicamente disponível através do acesso [<http://www.phytozome.net/commonbean.php>]. Os sequenciamentos e resultados gerados com livre acesso contribuem para o aumento da importância do feijoeiro comum como uma espécie genética-modelo, impactando positivamente também a pesquisa da soja e outros legumes. Atualmente a anotação das sequências encontra-se em andamento e o grande acervo de informação de genoma tem criado muitas possibilidades de pesquisa, como a mineração de sequências simples repetitivas (SSRs) (Blair et al. 2012a) e polimorfismo de nucleotídeo único (SNPs) (Galeano et al. 2012).

Do ponto de vista genômico, a espécie *P. vulgaris* tem sido objeto de intensivo estudo nos últimos 10 anos. Neste período foram desenvolvidos mapas genéticos (Galeano et al. 2011), foram mapeados QTLs (*Quantitative Trait Loci*) (Asfaw et al. 2012), foram gerados bancos de ESTs (*Expressed Sequence Tags*) com ênfase em estresses bióticos e abióticos (Kalavacharla et al. 2011) e foram construídas bibliotecas de clones BACs (*Bacterial Artificial Chromosomes*) para a manipulação e auxílio na montagem do genoma estrutural (Kami et al. 2006; Schlueter et al. 2008). As bibliotecas de grandes insertos de DNA genômico são consideradas ferramentas extremamente úteis para o estabelecimento de mapas citogenéticos e físicos de grande importância para o sequenciamento genômico, clonagem posicional, caracterização do genoma quanto à estrutura genética, análise funcional e, mais recentemente, em abordagem de transgenia (Schlueter et al. 2008; Ragupathy et al. 2011; Kang and Hérbert 2012). Os BACs têm sido amplamente utilizados como vetores para a clonagem de grandes fragmentos genômicos (100-150 Kpb), conforme demonstrado para várias espécies de importância econômica na

revisão de Yu (2012). Abordagens de mineração de sequências simples repetidas (SSRs), ou microssatélites (Tautz 1989), em bancos de dados derivados de sequenciamento de extremidades de BACs (BES-SSRs) têm sido aplicadas com sucesso para diversas espécies vegetais, incluindo leguminosas como a soja (Shultz et al. 2007), o feijoeiro comum (Córdoba et al. 2010), o feijão guandu (Bohra et al. 2011) e o amendoim (Wang et al. 2012).

Um número crescente de marcadores microssatélites (atualmente  $\geq$  2.000 SSRs) derivados de sequências genômicas e gênicas vem sendo disponibilizado para o feijoeiro comum (Yu et al. 2000; Blair et al. 2003; Hanai et al. 2007; Garcia et al. 2011). As principais aplicações dos marcadores SSRs em análise genômica de feijoeiro comum compreendem estudos de estimativas de diversidade genética e estrutura populacional em germoplasmas cultivado (Vianello-Brondani et al. 2013), de variedades tradicionais (Burle et al. 2010) e silvestre (Kwak and Gepts 2009; Blair et al. 2012b), em adição aos experimentos de mapeamento genético (Galeano et al. 2011) com ênfase na identificação de QTLs (Hanai et al. 2010; Blair et al. 2011a). A quase totalidade dos SSRs descritos na literatura para *P. vulgaris* é derivada de repetições do tipo dinucleotídeos, seguidos em menor quantidade por trinucleotídeos derivados de sequências gênicas. Com os recentes esforços de sequenciamento genômico para diversas espécies de plantas, bem como a disponibilização de bancos de ESTs, os SSRs baseados em tetra-, penta- e hexanucleotídeos tem sido relatados (Sonah et al. 2011), apesar de se mostrarem efetivamente menos abundantes no genoma. Na leguminosa *Medicago sativa*, a proporção conjunta estimada de tetra-, penta- e hexanucleotídeos em 12.371 ESTs foi de aproximados 25,1%, contra 48,8% de tri- e 26,1% de dinucleotídeos (Wang et al. 2013). O desenvolvimento de SSRs derivados de repetições maiores é particularmente interessante devido a maior facilidade e repetibilidade quanto à detecção de polimorfismo e devem ser investigados quanto à qualidade analítica e hipervariabilidade em comparação com as categorias de di- e trinucleotídeos. Em *P. vulgaris*, os poucos SSRs baseados em repetições maiores que tetranucleotídeos (somente 12 marcadores) avaliados foram caracterizados como monomórficos (Blair et al. 2011b). Entretanto, o grande volume de informação do genoma atualmente disponível cria oportunidade para o desenvolvimento de número abundante de

SSR, ampliando as possibilidades de caracterizar em maior escala novas classes de repetições quanto à robustez analítica e potencial de detecção de polimorfismo.

Neste estudo foram sequenciadas as extremidades dos clones de uma biblioteca genômica de BAC derivada da espécie *P. vulgaris* (Kami et al. 2006), variedade BAT 93, gerando informações de sequências que foram caracterizadas quanto à presença de fatores de transcrição, elementos transponíveis, anotação gênica, similaridade com genomas de leguminosas e, por fim, a mineração por sequências simples repetidas baseadas em motivos longos ( $\geq$  tri- e  $\leq$  hexanucleotídeos), com o desenvolvimento de marcadores operacionais úteis ao estudo genético e melhoramento da espécie.

## MATERIAL E MÉTODOS

### Sequenciamento e processamento das *BAC-end sequences* (BESs)

O genoma total da espécie *P. vulgaris*, variedade BAT 93, clonado e mantido em uma biblioteca genômica de fragmentos longos (BAC) desenvolvida por Kami et al. (2006) foi utilizada para o sequenciamento das *BAC-end sequences* (BESs). Essa biblioteca denominada de PVMBBa ou BAT 93 *HindIII* com tamanho médio de insertos de 125 Kpb e abrangendo aproximadamente 20X de cobertura do genoma [<http://www.genome.arizona.edu/orders/direct.html?library=PVMBBa>] foi comercialmente disponibilizada para fins de sequenciamento das BESs pelo *Arizona Genome Institute* (AGI, Tucson, AZ, USA). Cerca de 30.000 clones de BACs tomados aleatoriamente foram sequenciados a partir das extremidades *Forward* e *Reverse* em plataformas de sequenciamento automático ABI 3730xl (*Applied Biosystems*, Foster City, CA, USA), conforme descrito por Kim et al. (2007). As BESs foram inicialmente processadas utilizando o programa *Phred* (Ewing et al. 1998) e os nucleotídeos correspondentes às sequências de vetor foram excluídos pelo programa LUCY (Chou and Holmes 2001). O programa CAP3 (Huang and Madan 1999) foi utilizado para determinar o consenso (*contigs*) das sequências.

### Identificação das BES-SSRs e desenho dos *primers*

As BESs foram processadas de acordo com a metodologia descrita por Martins et al. (2006) que se baseia na busca por microssatélites usando o módulo TROLL (Castelo et al. 2002), juntamente com o pacote Staden (Staden et al. 1999), seguido pelo desenho dos pares de *primers* que flanqueiam os SSRs identificados utilizando o programa Primer3 (Rozen and Skaletsky 2000). Os parâmetros do módulo TROLL foram estabelecidos para a identificação de sequências contendo SSRs do tipo perfeito e compostos, baseados nos motivos di-, tri-, tetra-, penta- e hexanucleotídeos e com no mínimo quatro repetições. Os parâmetros para o desenho dos *primers* foram a temperatura de anelamento (*T<sub>a</sub>*) variando entre 45 a 68°C, com média de 60°C; o tamanho do *amplicon* de 100-400 pb; a estabilidade máxima 3' de 250 pb; o comprimento do *primer* de 18-27 pb – com média de 22 pb; conteúdo GC de 50% ± 5% e

outros parâmetros fixados com valores padrão. A nomenclatura “BES” foi acrescida à nomenclatura “SSRs” para indicar sua origem (BES-SSRs).

### **Caracterização genética dos BES-SSRs**

Visando ao desenvolvimento de marcadores SSRs, na primeira etapa foram avaliados 194 marcadores SSRs privilegiando as repetições do tipo tetra-, penta- e hexanucleotídeo, assim como os compostos com maior número de unidades repetitivas. Esses marcadores foram avaliados quanto ao padrão de amplificação e sugestão de polimorfismo em um painel de quatro genótipos de feijoeiro comum, sendo um representativo do *pool* gênico Andino (BRS Executivo) e três do Mesoamericano (Rudá, BRS Esplendor e BRS Agreste). Os marcadores foram amplificados e submetidos à eletroforese em géis desnaturantes de poliacrilamida 6% conforme descrito em Creste et al. (2001).

Com base no perfil de fácil interpretação e robustez do produto amplificado, assim como a indicação de polimorfismo, foram selecionados 40 BES-SSRs a partir dos 194 avaliados para a síntese dos iniciadores marcados com fluorescência. Considerando o tamanho esperado do fragmento amplificado em pares de base (pb) e a existência de complementaridade entre os *primers*, determinada através do programa *AutoDimer* (Vallone and Butler 2004), os microssatélites foram reunidos em sistemas de genotipagem multiloco, denominados de sistemas *multiplex*. Para aumentar a eficiência dos sistemas, os microssatélites foram marcados com fluorescência azul (6-FAM™), verde (HEX™), amarela (NED™) ou vermelha (PET™), especificadas no Material Suplementar 1, e a laranja (LIZ®) como marcador interno de massa molecular.

Para o conjunto dos 40 BES-SSRs selecionados foram avaliados 88 genótipos de feijoeiro comum, compreendendo 55 linhagens comerciais lançadas por instituições de pesquisa do Brasil e do exterior e 33 variedades tradicionais (*landraces*) brasileiras (Material Suplementar 2). O DNA genômico foi extraído a partir de amostras de tecido foliar de acordo com o método do CTAB, conforme descrito por Grattapaglia and Sederoff (1994).

As reações de amplificação com os SSRs, reunidos em sistemas *tetraplex*, foram conduzidas utilizando o produto comercial *Qiagen Multiplex*

PCR 2X (Qiagen, Hilden, Germany), seguindo a recomendação do fabricante. As reações de PCR foram conduzidas em termociclador *GeneAmp Thermal Cycler 9700 (Applied Biosystems)* com uma etapa inicial de desnaturação a 95°C por 15 min; seguida de 30 ciclos com etapas de desnaturação (94°C por 30 s), anelamento (variando de 56°C a 60°C por 90 s) e extensão (72°C por 1,5 min); uma etapa de extensão final de 72°C por 10 min. A eletroforese foi conduzida na plataforma ABI3500 XL (*Applied Biosystems*). A análise dos fragmentos foi feita utilizando-se o programa *GeneScan Analysis v2.1 (Applied Biosystems)* e a determinação do tamanho dos alelos pelo programa *GeneMapper v4.1 (Applied Biosystems)*.

### **Análise estatística dos BES-SSRs**

A diversidade genética da amostra total foi analisada através das estimativas da frequência alélica, número de alelos por loco, heterozigosidade esperada ( $H_E$ ), heterozigosidade observada ( $H_O$ ) e o índice de fixação ( $F_{IS}$ ) (estimado de acordo com a estatística de Wright com intervalo de confiança de 95% para 10.000 *bootstraps*) usando o programa *PowerMarker v2.25 (Liu and Muse 2005)*. A identificação dos alelos privados foi obtida através do programa *GenAlex v6.5 (Peakall and Smouse, 2012)*. A probabilidade de identidade (PI) genética para cada loco e multilocos foram calculados usando o programa *Cervus v3.0 (Kalinowski et al. 2007)*.

### **Anotação Funcional das BESs**

As BESs foram anotadas utilizando a estratégia de inferência de função baseando-se em similaridades significativas através de comparação com sequências disponíveis na base de dados do NCBI. As BESs foram anotadas com a ajuda do algoritmo BLASTN, sendo comparadas ao banco de dados de GSSs (*Genome Survey Sequences*) do *GenBank* utilizando ponto de corte  $E\text{-value} \leq 1,0E-06$  e o BLASTX confrontadas à base de dados de proteínas não redundantes com as configurações padrões. A fim de atribuir função putativa ao total das BESs e dos BES-SSR, foi utilizada a ferramenta BLAST2GO, para fornecer anotação automática usando termos *Gene Ontology (GO)*, a partir dos BLASTX *hits*. (Conesa et al. 2005). As análises de mapeamento e anotação das BESs de acordo com os termos do GO foram baseadas na similaridade de

sequência e, portanto, sequências sem BLAST *hits* não foram anotadas. O GO fornece um sistema de classificação dos produtos gênicos de acordo com três categorias: processo biológico, função molecular e componente celular; sendo que cada sequência pode ter mais de um termo GO, ou a partir de categorias diferentes, ou a partir da mesma categoria (Ashburner et al. 2000). Os parâmetros adotados na anotação foram às configurações padrões do BLAST2GO ( $E\text{-value} \leq 1,0E-06$  e ponto de corte de 55 da anotação).

As BESs foram comparadas por similaridade com o auxílio do algoritmo BLASTN aos transcritos mapeados no genoma de referência Andino de *P. vulgaris* (Versão 1.0), disponível no Phytozome v8.0 (<http://www.phytozome.net/>), utilizando ponto de corte  $E\text{-value} \leq 1,0E-05$ . As BESs, com correspondência aos transcritos do genoma identificados, foram confrontados com o banco de dados de fatores de transcrição para planta, PlnTFDB – *Plant Transcription Factor Database* (<http://plntfdb.bio.uni-potsdam.de/v3.0/>) (Pérez-Rodríguez et al. 2010), selecionando apenas os fatores de transcrição descritos para feijoeiro comum. Através do BLASTN, as BESs foram também confrontadas com o banco de dados de elementos transponíveis de soja, SoyTEdb – *SoyBase and the Soybean Breeder's Toolbox* (<http://soybase.org/soytedb/>) (Du et al. 2010) com ponto de corte  $E\text{-value} \leq 1,0E-05$ .

### **Análise de Genômica Comparativa**

O total de BES, os BES-SSRs, os fatores de transcrição (TF) e os elementos transponíveis (TE), foram alinhados, separadamente, aos genomas das leguminosas *M. truncatula* (Young et al. 2011) (<http://www.phytozome.net/medicago.php>), *G. max* (Schmutz et al. 2010) (<http://www.phytozome.net/soybean.php>) e de *P. vulgaris* variedade Andina (G19833) (<http://www.phytozome.com/commonbean.php>). Os alinhamentos foram realizados através do algoritmo BLASTN utilizando como corte um valor de  $E\text{-value} \leq 1,0E-05$  e um comprimento mínimo do alinhamento de 50 pb. O melhor *hit*, determinado a partir de cada BES, foi utilizado para desenvolver os mapas no programa Circos para visualização da sintonia (Krzywinski et al. 2009).

## RESULTADOS

### Sequenciamento das BESs

Através do sequenciamento de 30 mil clones de BAC foi possível obter uma visão geral sobre o genoma de feijoeiro comum e derivar informações úteis com potencial de serem empregadas na ampliação do conhecimento e no melhoramento genético dessa espécie. Ao todo, 60 mil sequências genômicas de clones de BACs foram obtidas, compreendendo aproximadamente 36 Mpb que correspondem a uma estimativa de 6,5% do tamanho do genoma de *P. vulgaris*. A partir do sequenciamento das extremidades dos BACs pelo método de *Sanger*, foram gerados 52.270 BAC-*end sequences* (BESs) com *reads* de alta qualidade (87,12%) e tamanho médio de 690 pb. O comprimento total analisado foi de 31.702.945 pb de BES e com conteúdo GC estimado em 38,83%.

### Identificação e caracterização de BES-SSRs

A busca por marcadores microssatélites para o feijoeiro comum a partir das 52.270 BESs resultou na identificação de 3.789 SSRs, na frequência de um SSR a cada 8,36 Kpb. Já a análise que objetivou a identificação de regiões de SSRs completas e passíveis de serem utilizados para o desenvolvimento de marcadores moleculares foi mais restritiva e as BESs foram submetidas a um processo de remoção das sequências curtas e/ou redundantes resultando em 47.698 sequências úteis que foram reunidas em 1.028 *contigs* e 42.082 *singlets* (43.110 total). Do total de *contigs* e *singlets*, foram identificados 2.472 SSRs, com algumas BES contendo mais de um SSRs. Considerando apenas um SSR por BES foram identificados 2.000 SSRs denominados de BES-SSRs.

Os BES-SSRs mais frequentes em feijoeiro comum foram aqueles contendo repetições perfeitas do tipo trinucleotídeos (45,75%) e dinucleotídeos (27,95%), seguidos pelos compostos (19,4%) que são formados por repetições contíguas ou adjacentes de diferentes motivos. Os SSRs baseados em repetições do tipo tetra- (3,85%), penta- (0,95%) e hexanucleotídeos (2,1%) foram menos abundantes (Tabela 1). Os trinucleotídeos compostos por motivos AAG ocorrem com maior frequência (17,12%) entre o total de BES-SSRs, seguidos por AAT (9,83%) e ATG (6,04%). Em relação aos

dinucleotídeos, os motivos AT (16,57%), AG (7,09%) e AC (4,29%) foram os mais frequentes nesta ordem, sendo o motivo GC representado por apenas uma única sequência entre as 559 BES-SSRs de dinucleotídeos. Os motivos mais frequentes de tetra-, penta- e hexanucleotídeos foram AAAT (2,4%), AAAAT (0,8%) e GGGCTT (0,65%), respectivamente (Tabela 2).

Com base no comprimento em pares de base (pb) das sequências de DNA contendo as repetições, os SSRs foram categorizados em Classe I (SSRs  $\geq$  20 pb) e Classe II (SSRs  $\leq$  19 pb), conforme proposto por Temnykh et al. (2001). Entre os 2.000 BES-SSRs, 644 (32,2%) foram representativos da Classe I, com predomínio dos SSRs compostos (56%); e 1.356 (67,8%) da Classe II, com predomínio dos trinucleotídeos (65%) e seguidos pelos dinucleotídeos (29%). Quanto aos penta- e hexanucleotídeos foram identificadas repetições representativas apenas da Classe I, conforme detalhado na Tabela 1.

### **Caracterização molecular dos BES-SSRs**

A partir dos 2.000 BES-SSRs identificados via mineração das sequências genômicas de clones de BACs foram desenhados *primers* para 194 SSRs. Considerando que, em geral, os SSRs mais longos tendem a ser mais hipervariáveis, a seleção dos marcadores para síntese levou em consideração esse critério, além de priorizar as repetições do tipo tetra-, penta- e hexanucleotídeos, até o momento, pouco explorados quanto ao padrão de polimorfismo e qualidade de amplificação em feijoeiro comum. A seleção resultou em um conjunto 164 SSRs com repetições do tipo perfeito, sendo 18 di-, 8 tri-, 77 tetra-, 19 penta-, 42 hexanucleotídeos e 30 SSRs do tipo composto. Desses, 31% são representativos da Classe II, que corresponde aos 60 tetranucleotídeos; e 69% da Classe I que inclui todos os motivos de SSRs. Na primeira etapa de triagem foi verificado que 69 (36%) pares de *primers* não amplificaram e foram descartados. Ao todo, 64% dos locos amplificaram produtos com sucesso em pelo menos um indivíduo testado, com as maiores taxas de êxito na amplificação dos penta- (79%) e tetranucleotídeos (78%), enquanto para os hexanucleotídeos a taxa foi de apenas 48%. Dentre os SSRs que amplificaram, em 24% foi possível detectar polimorfismos com resolução em agarose. Do total de marcadores, foram selecionados 40 BES-SSRs para

síntese com fluorescência, os quais reuniram as características desejadas para o desenvolvimento de sistemas de genotipagem semi-automatizados.

Os 40 BES-SSRs foram representados por 33 SSRs do tipo perfeito, sendo quatro tri-, 10 tetra-, 10 penta-, nove hexanucleotídeos, e sete SSRs do tipo compostos. Os BES-SSRs foram avaliados através de 10 sistemas *tetraplex* nos quais os *primers* foram marcados com quatro tipos distintos de fluorocromos. Dos 10 sistemas avaliados, em oito deles todos os locos amplificaram e nos demais houve falha de amplificação de um a dois locos. Os locos que falharam foram individualmente amplificados com êxito (PvPenta 5, PvComp8, PvPenta8 e PvPenta10). A partir da análise em eletroforese de alta resolução com detecção semi-automatizada e amostragem ampliada para 88 genótipos foram identificados 34 (85%) BES-SSRs polimórficos e seis monomórficos (PvTetra7, PvPenta18, PvTri4, PvPenta9, PvHexa26 e PvHexa28).

Considerando os 34 BES-SSRs polimórficos, o número médio de alelos por loco foi de 5,59; variando de dois (PvHexa12, PvTetra32, PvPenta8, PvPenta14 e PvTetra76) a 19 (PvComp10). A diversidade genética média ( $H_E$ ) encontrada foi de 0,493; variando de 0,135 no loco PvPenta5 a 0,930 no loco PvComp10. A heterozigosidade observada ( $H_o$ ) média estimada foi de 0,004; corroborando com o elevado índice de endogamia ( $F_{IS}=0,933$ ) e reforçando a tendência de homozigose para os locos analisados. Considerando os grupos de germoplasma avaliados, o número médio de alelos por loco gênico foi maior para o material cultivado (5) e menor para as variedades tradicionais (3,44), enquanto as estimativas médias de  $H_E$  foram próximas, com valores de 0,498 e 0,463; respectivamente. Ao todo foram identificados 93 alelos privados em 28 BES-SSRs analisados (73 alelos no material cultivado e 20 nas variedades tradicionais), sendo que os locos PvHexa15, PvComp10 e PvComp4 destacaram-se dos demais com oito, nove e 10 alelos privados, respectivamente (Tabela 3).

Entre as classes de BES-SSRs avaliadas, os trinucleotídeos, representados por três marcadores, apresentaram a maior estimativa média de  $H_E$  (0,73), seguido pelos SSR do tipo compostos (0,66), representados por sete marcadores. Considerando os tetranucleotídeos essa estimativa foi de 0,51;

enquanto para os penta- e hexanucleotídeos os valores foram menores e similares, com índices de 0,35 e 0,33; respectivamente. Dos 34 marcadores, seis (com comprimento da repetição maior que 40 pb) destacaram-se com  $H_E \geq 0,7$ ; apresentando em ordem decrescente as estimativas de  $H_E$  de 0,924 (PvComp10); 0,826 (PvTri8); 0,821 (PvComp4); 0,747 (PvTri6); 0,745 (PvTetra25) e 0,707 (PvComp9). Considerando um  $H_E \geq 0,5$ ; o número de locos aumentou para 16, o que corresponde a 47% dos SSRs caracterizados. Um elevado padrão de qualidade dos perfis de amplificação foi observado com raras ocorrências de produtos inespecíficos e *stuttering* conforme ilustrado no Material Suplementar 3.

### **Anotação das BESs**

A anotação das 52.270 sequências (BLASTN) resultou no alinhamento de 47.194 (90,3%) contra as GSSs do *GenBank*, das quais 31.564 (60,4%) alinharam-se com elevada similaridade ( $E\text{-value} = 0$ ) às sequências genômicas de *P. vulgaris*. Considerando as 2.000 BES-SSRs, 1.611 (80,5%) tiveram uma similaridade significativa com as GSSs. Através da comparação das sequências com o banco de dados não-redundante de proteínas do *GenBank* (BLASTX) foi possível identificar funções putativas para 24.321 (46,53%) das BESs (52.270), das quais 609 (30,45%) são BES-SSRs (2.000). Dentre as 40 BES-SSRs selecionados para genotipagem em sistemas de alta resolução, 13 (32,5%) foram associadas a produtos gênicos putativos (Material Suplementar 1). Com o auxílio da ferramenta BLAST2GO as sequências totais foram submetidas ao processo de mapeamento e anotação, nos quais os valores estimados ( $E\text{-values}$ ) variaram de  $1,0E-06$  a  $1,0E-175$ . As BESs apresentaram maior similaridade (BLAST "*Top-Hits*") com sequências de *P. vulgaris* (17%), seguido por *G. max* (13,5%), *Vitis vinifera* (6%), *M. truncatula* (4%) e *Oryza sativa* (1%). Considerando apenas as BES-SSRs observou-se um predomínio de alinhamento com sequências de *G. max* (12%), enquanto para *P. vulgaris* essa estimativa foi de 8,65%.

Das 24.321 sequências associadas a genes, a classificação representativa das categorias funcionais utilizando os termos do GO foi atribuída a 19.363 BESs (80%) (Figura 1). Dentro de processos biológicos (10.130 produtos gênicos ou sequências), os mais representativos foram

*metabolic process* (9.521), *cellular process* (9.204) e DNA *metabolic process* (6.837). Referente à categoria função molecular (13.632), as atividades dos genes mais representadas foram *binding* (12.555), *organic cyclic compound binding* (8.692) e *nucleic acid binding* (7.478). Na categoria componente celular (4.225), as mais abundantes foram *cell part* (3.936), *intracellular part* (3.500) e *organelle* (3.285). Dentre as 609 BES-SSRs correspondentes a genes, 416 (68%) foram atribuídos a termos do GO, com representatividade das três categorias funcionais e predomínio de *metabolic process* (196) em processos biológicos, *binding* (256) para categoria função molecular e *cell part* (119) para componente celular (Figura 1).

A análise quanto à presença de TFs revelou que em 2% das BESs estão representadas 65 famílias de TFs, das 84 famílias previamente descritas (Pérez-Rodríguez et al. 2010). Entre as 65 famílias, as mais representativas estão destacadas na Figura 2. Para a maior parte das sequências (37%) observou-se alinhamento com o TF mTERF, seguido por DDT (9%), Tify (6%) e ARR-B-MYB (5%) que, em conjunto, perfazem 57% das BESs que contêm TFs. Nas demais BESs (43%) foram identificados 61 famílias diferentes de TFs. Para as BES-SSRs, 58 foram associadas à TFs representativos de 30 famílias diferentes, sendo as famílias bHLH e NAC mais abundantes, com cinco BES-SSRs cada.

O alinhamento das sequências com o objetivo de identificar TEs, utilizando a base de dados de *G. max* (Du et al. 2010), revelou 3% das BESs associadas a esses elementos, com predomínio das sequências representativas da Classe I de *retroelements* (98%) e apenas 2% da Classe II composta por DNA *transposons*. Dentre os *retroelements*, os LTR *retrotransposons* que correspondem a repetições terminais longas foi predominantemente representados pelas famílias *Gypsy* (81%) e *Copia* (17%). Para os TEs da Classe II, apesar de bem menos abundante, foram identificadas as famílias *Mutator* (1,7%), seguida por *PIF-Harbinge* (0,25%) e *Helitron* (0,05%) (Figura 2). Dentre as 2.000 BES-SSRs de *P. vulgaris* apenas 38 (2%) foram identificadas como associadas aos TEs, das quais 33 são da família *Gypsy* (87%), quatro da *Copia* (10%) e uma da *Mutator* (3%).

## Comparação de genomas de leguminosas

O alinhamento das 52.270 BESs derivadas da variedade Mesoamericana BAT 93 contra os genomas de *P. vulgaris*, variedade andina (Pv), *M. truncatula* (Mt) e *G. max* (Gm) está representado na Figura 3-A. Do total de 52.270 BES, 99,4% apresentaram similaridades com a sequência genômica de Pv, seguido por 14,4% de Gm e 2,7% de Mt. Quanto as 2.000 BES-SSRs, 99,4% alinharam ao genoma de Pv, seguido por 16,7% ao de Gm e somente 2,7% ao de Mt. Poucas BES-SSRs foram similares simultaneamente entre os genomas de Mt e Pv (53 BES-SSRs), já entre os genomas Gm e Pv foi 6,2 vezes superior (331 BES-SSRs). Considerando os cromossomos de *P. vulgaris*, onde 10 ou mais BES-SSRs foram comuns ao genoma de *G. max*, foi possível observar um nítido alinhamento entre eles, evidenciando que as sequências mapeadas no cromossomo um de feijoeiro comum (Pv1) foram alinhadas principalmente no cromossomo Gm19 da soja, enquanto as dos cromossomos Pv3, Pv5, Pv7 e Pv11 alinharam-se aos cromossomos Gm17, Gm12, Gm10/Gm20 e Gm6/Gm12 da soja, respectivamente (Figura 3-A).

Considerando os 1.039 fatores de transcrição identificados em *P. vulgaris*, todas as sequências foram alinhadas ao genoma da variedade Andina (100%), enquanto 41% foram comuns aos de *G. max* e 8.5% de *M. truncatula*. Para o alinhamento entre os cromossomos de Pv e Gm contendo mais de 10 TFs comuns, observou uma distribuição variável das sequências ao longo do genoma de Gm quando comparado ao Pv. Alguns TFs associados a um cromossomo específico de feijoeiro tiveram agrupamento de sequências com correspondência para um único cromossomo de soja, como os TFs posicionados no cromossomo um de *P. vulgaris* (Pv1) e muitos localizados no cromossomo 19 da soja (Gm19). O mesmo foi observado para as sequências posicionadas no Pv2 e correspondentes ao Gm1; as do Pv3 no Gm7; as do Pv4 no Gm16; as do Pv9 no Gm4; e por fim, as do Pv11 no cromossomo Gm12. Já para outros TFs, observou-se uma distribuição mais ampla e dispersa no genoma da soja quando comparada à do feijoeiro comum, com os TFs localizados no Pv5 sendo distribuídos nos cromossomos Gm1, Gm12 e Gm13 e os do Pv7 nos Gm10 e Gm20 (Figura 3-B).

Quanto aos 1.643 elementos transponíveis identificados, conforme o esperado, 99,9% alinharam-se ao genoma da variedade Andina de *P. vulgaris* (Pv) e, em menores proporções, aos genomas de Gm (72,5%) e Mt (7,4%). Para o alinhamento entre os cromossomos de Pv e Gm contendo 10 ou mais TEs comuns, observou-se uma ampla dispersão das sequências em *G. max* quando comparadas às de *P. vulgaris*. Como exemplos podem-se destacar os TEs posicionados no cromossomo Pv5 de feijoeiro e dispersos em cinco cromossomos de soja (Gm1/Gm6/Gm15/Gm18/Gm19); do Pv8 em oito Gms (Gm1/Gm2/Gm4/Gm5/Gm6/Gm15/Gm18/Gm19) e do Pv11 em seis Gms (Gm1/Gm5/Gm10/Gm15/Gm18/Gm19). Para regiões comuns aos genomas de *P. vulgaris* e *M. truncatula*, destaca-se os 35 TEs posicionados no Pv11 e igualmente correspondentes no Mt3 (Figura 3-C).

## **DISCUSSÃO**

### **Sequenciamento das BESs**

A análise de *BAC-end sequences* tem se mostrado um recurso valioso para o estudo de genomas de plantas. Em estudos com o feijoeiro comum, bibliotecas de BACs têm sido desenvolvidas e exploradas para diversas aplicações, como a caracterização de uma família de proteínas de sementes (*locus APA*) (Kami et al. 2006), a construção de mapa físico (Schlueter et al. 2008), o mapeamento de QTL para resistência a cretamento bacteriano comum (CBB) (Liu et al. 2010) e, mais recentemente, para o isolamento e caracterização de uma nova família de DNA repetitivo em genótipos do *pool* gênico Mesoamericano (Ribeiro et al. 2011). Entretanto, apenas no estudo de Córdoba et al. (2010) foi relatado o desenvolvimento de uma biblioteca de BAC com ênfase na busca por marcadores SSRs, os quais foram empregados para integrar o mapa genético ao mapa físico de feijoeiro comum. No presente estudo, as informações derivadas do sequenciamento das BESs foram, de imediato, integradas ao projeto internacional de sequenciamento do genoma de feijoeiro comum, variedade Mesoamericana, auxiliando na determinação e montagem do mesmo. Paralelamente, a disponibilidade dessas sequências revelou com sucesso a possibilidade de explorar a composição do genoma e identificar SSRs baseados em repetições longas, até então pouco exploradas para feijoeiro comum, agregando características genéticas adequadas e com potencial de se tornarem operacionais junto ao programa de melhoramento genético dessa cultura. A estimativa na composição de sequências ricas em AT (61%) no genoma de feijoeiro comum através das BESs revelou estar próxima ao observado para outras leguminosas, como a soja, na qual foi descrita uma proporção de 64% (Shultz et al. 2007).

### **Identificação e caracterização de BES-SSRs**

Os BES-SSRs desenvolvidos neste estudo são importantes recursos moleculares para a análise genética da espécie *P. vulgaris*. A mineração de sequências genômicas, derivadas de pontas de clones BACs, buscando SSRs tem-se mostrado bastante útil para o desenvolvimento desses marcadores em diversas leguminosas, como os 3.290 SSRs identificados em soja (Shoemaker

et al. 2008), 18.149 em feijão guandu (Bohra et al. 2011), 1.424 em amendoim (Wang et al. 2012), 6.845 em grão de bico (Thudi et al. 2011) e 875 SSRs em feijoeiro comum (Córdoba et al. 2010). O número de regiões microssatélites identificadas em *P. vulgaris* através das análises aqui descritas mostrou que este tipo de sequência é bem mais numerosa do que qualquer previsão empírica esperada. No estudo atual, a partir das 52.270 BESs foi identificado um número de SSRs quatro vezes superior (3.789) ao relatado por Córdoba et al. (2010), trabalho no qual foram analisadas 40% a mais de sequências (89.017 BESs). Esta diferença se dá, primariamente, devido aos diferentes parâmetros para qualificar uma sequência como sendo microssatélite e aos diferentes programas de busca por SSRs. O estudo de Blair and Hurtado (2013), usando um mesmo conjunto de dados (21.000 ESTs) de feijoeiro comum, apresentou resultados divergentes para a detecção do número e tipos diferentes de SSRs pela comparação de três programas de bioinformática.

Em plantas superiores existe a hipótese de que a frequência de SSRs em regiões codificadoras é superior a regiões não codificadoras (Morgante et al. 2002; Tóth et al. 2000), conforme demonstrado em *Arabidopsis* (Morgante et al. 2002) e em *Citrus* (Biswas et al. 2012). Por outro lado, Victoria et al. (2011) relata que os SSRs seguem um padrão de distribuição aleatória no genoma de plantas, estando de acordo com o encontrado para o feijoeiro comum na comparação entre este estudo e o de Garcia et al. (2011), onde a frequência de SSRs por sequência analisada foi praticamente a mesma. Estimativas próximas com relação à frequência de SSRs nas regiões codificadoras (EST-SSRs: 0,021%) identificados por Garcia et al. (2011) foram observada no estudo atual em regiões não codificadoras (BES-SSRs: 0,026%), enquanto a distribuição das frequências dos motivos diferiu entre EST-SSRs e BES-SSRs, o mesmo foi relatado para café (Dereeper et al. 2013). Do ponto de vista prático a distribuição mais homogênea dos SSRs ao longo de todo o genoma é bastante interessante, pois permitirá o mapeamento direto destes SSRs em *P. vulgaris* isolados a partir de diferentes fontes de sequências, assegurando uma ampla distribuição e amostragem do genoma.

O número de regiões microssatélites identificadas em *P. vulgaris* através do programa TROLL mostrou que este tipo de sequência é bastante numerosa sendo equivalente a um SSR a cada 8,36 Kpb, apresentando uma maior

densidade de SSRs quando comparado com amendoim, onde foram identificados 27,32 Kpb/SSR derivados de BES (Wang et al. 2012) e cujo genoma está estimado em 2.800 Mpb. Essa proporção parece estar de acordo com a tendência observada por Morgante et al. (2002), onde genomas menores possuem maior porcentagem de seu genoma composto por regiões microssatélites do que genomas maiores, em plantas. A densidade de SSRs encontrada neste trabalho é semelhante com a descrita para outras leguminosas como soja (6,82 Kpb/SSR), feijão guandu (5,65 Kpb/SSR) e grão de bico (4,85 Kpb/SSR) (Saini et al. 2008; Bohra et al. 2011; Thudi et al. 2011). Certamente, a densidade de SSRs identificada ao longo dos genomas varia de acordo com a espécie e a característica do genoma em questão, bem como o seu nível de ploidia (Biswas et al. 2012; Tóth et al. 2000; Morgante et al. 2002); com os parâmetros dos programas adotados para a busca por microssatélites (Leclercq et al. 2007); com os métodos empregados para a construção da biblioteca de BAC, como o uso de diferentes enzimas de restrição, aumentando a probabilidade de cobertura de todas as regiões presentes no genoma e, conseqüentemente, amplia as chances de mais SSRs serem identificados (Wang et al. 2012).

Nesse estudo as sequências contendo repetições mononucleotídicas de SSR não foram analisadas devido à abundância de motivos poli-A encontradas em outras espécies (Thudi et al. 2011). Adicionalmente, para fins operacionais essas sequências contendo repetições em *tandem* de um único nucleotídeo não seriam utilizadas para o desenvolvimento de ferramentas moleculares. Embora tenha sido descrito que o número de locos no genoma e o número de repetições total no microssatélite tende a mostrar uma relação decrescente com o tipo de motivo, sendo mais comum encontrarmos microssatélites dinucleotídicos do que trinucleotídicos, e assim por diante ao longo de todo o genoma (Agüero et al. 2000). Neste estudo os SSRs mais abundantes foram os trinucleotídeos (46%), e não os dinucleotídeos (26%) como descrito para outras leguminosas (Wang et al. 2012; Thudi et al. 2011; Bohra et al. 2011; Córdoba et al. 2010). O trabalho de Victoria et al. (2011) relata que os trinucleotídeos são os SSRs predominantes em plantas superiores, como o observado para feijoeiro comum neste estudo, para soja (Shoemaker et al. 2008) e *Arabidopsis thaliana* (Lawson and Zhang 2006), não existindo consenso entre as espécies

vegetais com relação a essa lógica. Em regiões codificantes, espera-se que os trinucleotídeos sejam mais frequentes decorrente da teoria de que exista um processo de seleção negativa contra mutações que alterem a fase de leitura dos códons (Morgante et al. 2002), embora os SSRs apresentem um padrão distinto de distribuição entre as diferentes regiões gênicas (5' UTR, 3'UTR, introns e exons) (Lawson and Zhang 2006). No estudo atual, o predomínio de tri- em relação aos dinucleotídeos quando comparado com os demais estudos conduzidos para o feijoeiro comum (Córdoba et al. 2010) pode ser justificado devido, também, a diferenças nos parâmetros para qualificar ou não uma sequência como sendo microssatélite. Enquanto os parâmetros do módulo TROLL foram estabelecidos para a identificação de SSRs do tipo perfeito e compostos com no mínimo quatro repetições, os SSRs identificados por Córdoba et al. (2010) utilizando o programa *BatchPrimer3* (You et al. 2008) foram definidos com no mínimo cinco repetições para di-, quatro para tri- e três para tetra- e para pentanucleotídeos. Independente dessas diferenças nas análises distintas de fragmentos genômicos de DNA em busca por regiões SSRs observa-se que há um elevado número destas no genoma de *P. vulgaris*, tendo sido possível identificar tipos e motivos abundantes.

Em relação aos trinucleotídeos observou-se um predomínio de motivos AAG (17,12%), os quais também foram relatados como mais abundantes em amendoim (Wang et al. 2012). Entre os dinucleotídeos, os motivos AT (16,57%) foram mais abundantes corroborando estudos prévios para feijoeiro comum (Córdoba et al. 2010) e com estudos de BES-SSRs em outras leguminosas como amendoim, feijão guandu e soja (Wang et al. 2012; Bohra et al. 2011; Shultz et al. 2007); demonstrando que isolar esses motivos torna-se viável por meio dessa estratégia, uma vez que esses motivos não podem ser isolados pela construção de bibliotecas enriquecidas, pois formam estruturas secundárias, i.e. (AT) $n$ /(TA) $n$ . O dinucleotídeo GC é uma repetição extremamente rara em genomas de eucariotos (Katti et al. 2001), a identificação de apenas um único motivo GC em 559 SSRs dinucleotídeos detectados neste trabalho exemplifica bem a escassez no genoma de feijoeiro comum. A dificuldade em construir bibliotecas para tetra-, penta- e hexanucleotídeos, pelo fato de serem menos abundantes no genoma, restringiu a disponibilidade dessas sequências para o desenvolvimento de pares de

*primers* e uso na caracterização molecular do feijoeiro comum. Sendo assim, o isolamento dessas regiões repetitivas a partir de fragmentos genômicos aleatórios representa uma estratégia atrativa, podendo resultar na geração de resultados bastante expressivos e importantes no que tange o desenvolvimento e uso de marcadores SSRs operacionais para o feijoeiro comum.

### **Caracterização molecular dos BES-SSRs**

Bibliotecas de BACs de *P. vulgaris* certamente são excelentes fontes de SSRs e, apesar de que somente parte desses marcadores identificados terem sido testados nesse estudo, não parece haver qualquer tipo de limitação em relação ao seu nível de polimorfismo e emprego dessa metodologia para genotipagem em larga escala. Nesse estudo foram explorados uma parte dos tri-, tetra-, hexanucleotídeos e SSRs do tipo compostos, bem como todos os pentanucleotídeos que foram identificados a partir das BESs. A triagem inicial realizada através da análise em agarose permitiu selecionar os marcadores com perfil de amplificação consistente e com potencial de detecção de polimorfismo a partir do grupo de genótipos inicialmente avaliados, apresentando uma taxa de sucesso de amplificação de 79%. Resultados similares de eficiência de amplificação foram relatados para o feijoeiro comum (76,5%) e grão de bico (79%), cujos SSRs foram explorados também a partir de BESs (Córdoba et al. 2010; Thudi et al. 2011). Com relação à proporção de SSRs de acordo com o tamanho do motivo, com os critérios estabelecidos notou-se um predomínio dos SSRs do tipo compostos (56%) categorizados na Classe I que compreendem as repetições com sequências de DNA mais longas. Estudos anteriores relatam proporções acima de 30% dessa classe de repetições em feijoeiro comum (Hanai et al. 2007; Garcia et al. 2011) e, juntamente com o estudo atual, disponibilizam e possibilitam explorar o potencial dessas repetições para fins de análise genética vegetal.

Conforme relatado por diversos autores, parece existir uma correlação entre o número de repetições dos motivos de um loco SSR com o seu conteúdo de informatividade genética, sugerindo que SSRs representativos da Classe I seriam mais úteis para análise genética (Byrne et al. 1996; Cho et al. 2000). Dentre os SSR avaliados nesse estudo, todos os seis com  $H_E \geq 0,7$  foram da Classe I e representados por três SSRs do tipo compostos e três

perfeito, sendo dois tri- e um tetranucleotídeo, seguindo padrão de maior polimorfismo por apresentarem mais de um tipo de motivo e/ou uma região microssatélite maior. O número médio de alelos por loco variou de 9,7 para os SSR compostos a 3,25 para os pentanucleotídeos, enquanto no trabalho de Vianello-Brondani et al. (2013) o número médio de alelos observados foi de 7,79 em dinucleotídeos. Isto demonstra que os SSRs compostos formados por repetições contíguas ou adjacentes variáveis podem mostrar maior polimorfismo por apresentarem mais de um tipo de motivo e uma região microssatélite maior. Nos locos do tipo tetra-, penta- e hexanucleotídeos o número médio de alelos foi de 4,60, 3,25 e 3,85, respectivamente, e significativamente menor do que os valores relatados para os dinucleotídeos, o que está de acordo com o esperado em função das maiores taxas de mutação que atuam nas repetições di- em função do escorregamento da enzima durante a replicação e alinhamento desigual do DNA durante a recombinação (Renwick et al. 2001).

Nesse estudo destacaram-se sete SSRs, sendo um hexa-, dois penta- e quatro tetranucleotídeos, com média da  $H_E$  de 0,61 e 5,14 alelos por loco gênico, com destaque para o loco PvTetra25 com 12 alelos e  $H_E$  de 0,74. Esses resultados sugerem que existe a possibilidade de se identificar um número bem maior de SSRs baseados em repetições  $\geq$  trinucleotídeos com elevado poder de informatividade genética, comparáveis ou superiores aos dinucleotídeos atualmente disponíveis. Em humanos, por questões de precisão experimental e intercâmbio de resultados, o estabelecimento de um sistema de tipagem de DNA forense envolve um esforço organizado de validação entre diversos laboratórios do mundo (Kimpton et al. 1996), no qual foram estabelecidos um conjunto de 13 SSRs internacionalmente recomendados baseados em repetições do tipo penta- e tetranucleotídeos (Budowle et al. 1999). Para o feijoeiro comum não existem relatos, até onde se conhece, para o desenvolvimento de SSRs baseados em motivos maiores, sendo que os tetra-, penta- e hexanucleotídeos identificados por Blair et al. (2011b) não foram polimórficos. Isto sugere que os SSRs desenvolvidos neste trabalho sejam os primeiros deste tipo para *P. vulgaris*. Além disso, os microssatélites tri-, tetra- e pentanucleotídicos apresentam menor “*stutter*”, e por consequência disto seus alelos parecem ter uma melhor definição do que os de

dinucleotídeos, facilitando a análise, aumentando a eficiência na distinção de indivíduos e reduzindo erros de genotipagem.

### **Anotação das BESs**

A grande similaridade das sequências BESs (60,4%) com as sequências GSSs de *P. vulgaris* foi devido ao grande volume de dados depositados oriundo do sequenciamento total do genoma da espécie. Entretanto, a similaridade das BESs (17%) com sequências proteicas de feijoeiro comum foram bastante reduzidas, devido as BESs representarem sequências genômicas e também ao limitado conteúdo da anotação atualmente disponível, o que mostra a necessidade crescente de mais esforços na anotação do genoma de feijoeiro comum. A anotação funcional das sequências foi realizada através do BLAST2GO. A maior proporção das sequências, 63% das BES e 70% das BES-SSR, permaneceram sem anotação, e podem ser consideradas como novas sequências a serem exploradas como recurso genômico para *P. vulgaris*. Os resultados encontrados neste trabalho sugerem que as BESs e BES-SSRs foram amplamente distribuídas nas três categorias funcionais principais do *Gene Ontology*. A caracterização funcional das sequências de BES-SSRs de feijoeiro comum contribuiu para agregar informações sobre o domínio funcional das proteínas associadas e sua anotação, permitindo desenvolver marcadores microssatélites com informações funcionais. Os 13 BES-SSRs, dos 40 BES-SSR sintetizados, que tiveram resultados do BLASTX foram associados a produtos gênicos com diversas funções putativas, entre elas o auxílio no desenvolvimento e defesa da planta (*calcium-transporting ATPase*) (Frei dit Frey et al. 2012) e na resposta a estresses bióticos e abióticos (*constans-like zinc finger protein*) (Noguero et al. 2013). Assim, as regiões microssatélites, detectadas e avaliadas neste estudo, estão próximas a genes associados a características de interesse agrônomo, podendo posteriormente ser utilizadas no melhoramento genético do feijoeiro comum.

Os fatores de transcrição são importantes no controle e manutenção da transcrição de genes em todo o genoma. Estes genes devem ser expressos em padrões espaço-temporal específicos e determinados por complexos de proteínas compostos por fatores de transcrições gerais (Pérez-Rodríguez et al. 2010). A identificação de TFs (2%) em feijoeiro comum representa um passo

importante no sentido de compreender as redes reguladoras dos genes, que podem estar associados a características de interesse no melhoramento de plantas. A família mTERF (37%) foi a mais representativa de todos os TFs identificados para *P. vulgaris*. A família mTERF é uma grande família de proteínas, identificadas em metazoários e plantas. As proteínas pertencentes a esta família estão localizadas nas mitocôndrias, onde atuam como reguladores da transcrição mitocondrial e no controle da replicação do mtDNA (Roberti et al. 2009). O estudo de Kleine (2012) em *A. thaliana* relata que a expressão de genes organelares é crucial para o desenvolvimento das plantas, fotossíntese e respiração, mas a nossa compreensão dos mecanismos que a controlam ainda é relativamente escasso. O sequenciamento do genoma de *Arabidopsis* levou à identificação de uma diversificada família de genes de mTERF, ainda pouco compreendida, porém revelando que as duplicações em *tandem* e uma duplicação em bloco contribuiu com a origem de um grande número de genes mTERF em *A. thaliana* (Kleine 2012).

Os elementos transponíveis são os componentes repetitivos mais abundantes de todos os genomas caracterizados de eucariotos superiores e desempenham papéis importantes na regulação da expressão gênica e também no que tange à composição e estrutura de genomas. Assim, a identificação e caracterização TEs no genoma de feijoeiro comum é de relevância para a anotação precisa dos genes e para a investigação da interação funcional dos genes e os TEs (Du et al. 2010). Neste estudo foram identificados 3% de TEs associados às BESs de feijoeiro comum, um resultado bem abaixo do estimado para grão de bico (22,16%) e mais próximo do obtido para amendoim (5,86%) (Bohra et al. 2011; Wang et al. 2012). Outras espécies de plantas, diferentes das leguminosas, apresentam *BAC-ends sequences* associadas aos TEs com porcentagens um pouco superiores ao encontrado neste trabalho, como o encontrado para cana-de-açúcar (6,6%), noqueira comum (10,7%) e café (11,9%) (Kim et al. 2013; Wu et al. 2012; Dereeper et al. 2013). A estimativa de TEs (2%) obtida nos BES-SSRs de feijoeiro foi próxima à proporção relatada em *Citrus* (1,35%) (Biswas et al. 2012). Conforme encontrado em outros estudos de plantas (Bohra et al. 2011; Wang et al. 2012; Biswas et al. 2012; Wu et al. 2012; Kim et al. 2013; Dereeper et al. 2013), há um maior número de elementos transponíveis da Classe I (*retroelements*) do

que os da Classe II (DNA *transposons*), corroborando com o encontrado para feijoeiro comum. Dentre os *retroelements*, os LTR *retrotransposons* de feijoeiro foi predominantemente representados pelas famílias *Gypsy* (81%) e *Copia* (17%), as quais diferem apenas na ordem em que seus genes estão dispostos e em algumas partes de suas sequências, sendo que as mesmas proporções foram observadas em *Citrus*, nogueira comum, cana-de-açúcar, café (Biswas et al. 2012; Wu et al. 2012; Kim et al. 2013; Dereeper et al. 2013) e também com o identificado no genoma total da soja (Schmutz et al. 2010). Do ponto de vista operacional, além de influenciar bastante no tamanho do genoma, *retrotransposons* também exercem função no rearranjo genômico, podendo gerar pontos de recombinação, deleções, duplicações, inversões e translocações. Nesse contexto, a caracterização desses elementos mostra-se importante para futuros estudos de filogenia, aplicação de elementos genéticos repetitivos como marcadores moleculares, assim como na imediata aplicação auxiliando a montagem do genoma de *P. vulgaris*.

### **Genômica de leguminosas**

O alinhamento das sequências BESs, da variedade Mesoamericana BAT 93, utilizando os genomas de *P. vulgaris* variedade Andina (Pv), *M. truncatula* (Mt) e *G. max* (Gm), os quais em conjunto representam os genomas de referência para leguminosas, é uma valiosa fonte de informação para estabelecimento de correlação entre essas espécies. As análises realizadas através da ferramenta BLASTN foram graficamente representadas utilizando o programa *Circos* (Figura 3) que auxiliou na interpretação dos resultados da comparação genômica entre essas várias sequências. Em uma primeira avaliação, essa análise possibilitou a incorporação e localização das BESs com TFs e TEs, bem como dos BES-SSRs avaliados no genoma de *P. vulgaris*, indicando uma distribuição ampla das sequências com representatividade de todos os cromossomos. Observou-se que o alinhamento dos TEs com o genoma da soja (72,5%) foi relativamente alto em relação aos BES-SSRs (14,4%) e TFs (41%), o que pode ser justificado pelo uso da base de dados de elementos transponíveis de soja (SoyTEdb) (Du et al. 2010). Apesar de *M. truncatula* ser considerado um modelo bem estabelecido no estudo das leguminosas (Young et al. 2011), os alinhamentos entre as BES-SSRs (2,7%), TFs (8,5%) e TEs (7,4%) de feijoeiro comum foram significativamente menores

que os de soja. Neste estudo, o alto nível de similaridade entre o genoma do feijoeiro comum e da soja é devido à relação filogenética mais próxima destas leguminosas que pertencem ao mesmo clado Phaseoloid, enquanto que há uma limitada sintonia com *M. truncatula*, que pertence ao clado Hologalegina (Gepts et al. 2005).

O estudo de Galeano et al. (2011), através de mapeamento fino e análise de sintonia entre a soja e o feijoeiro comum, demonstrou que o cromossomo Pv7 está extremamente relacionado com Gm10 e Gm20, assim como os dados de alinhamento das BES-SSRs e TFs encontrados neste trabalho. O sequenciamento do genoma completo da soja permitiu identificar que o Gm20 é altamente homólogo ao braço longo do Gm10, com algumas partes distribuídas em diversos outros lugares do genoma (Schmutz et al. 2010). Assim, os cromossomos Gm10 e Gm20 da soja, juntamente com o Pv7 de feijoeiro são fortes candidatos na identificação das duplicações cromossômicas ancestrais dos genomas de leguminosas (Galeano et al. 2011). Entretanto, segundo evidências do estudo de McClean et al. (2010), o genoma da soja foi fracionado e remontado em algum ponto após o evento de duplicação. A ampla distribuição da maioria das BES-SSRs, TFs e TEs de feijoeiro distribuídas por todo o genoma da soja não segue a proporção um para dois (alinhamento em um cromossomo de feijoeiro com dois cromossomos de soja), que é suportado pela quebra e união cromossômica possivelmente associada com evento de tetraploidização durante a evolução do genoma da soja (Galeano et al. 2011).

## CONSIDERAÇÕES FINAIS

Além de todas estas descrições e contribuições para o desenvolvimento de marcadores moleculares SSRs, deve-se destacar a importância das sequências geradas neste trabalho no âmbito do projeto internacional de sequenciamento do feijoeiro comum. Na prática, tanto as sequências geradas, como os resultados extraídos de suas análises, serão aplicados em outros projetos em larga escala. Desta forma, as análises das BESs de *P. vulgaris* mostraram-se bastante informativa, possibilitando acessar e descrever aspectos nunca ou pouco estudados no genoma do gênero *Phaseolus*. Além disso, pode funcionar como uma excelente estratégia para se obter ferramentas moleculares para os genomas de feijoeiro comum, uma cultura de grande interesse agrônomo e socioeconômico mundial.

Um dos grandes avanços desse estudo foi a geração de uma bateria de SSRs em adição aos SSRs baseados em di- e trinucleotídeos atualmente disponíveis com uma qualidade analítica satisfatória e facilidade de interpretação dos resultados gerados com a clara declaração dos alelos. Uma vez empregados por outros grupos de pesquisa, a integração dos resultados gerados poderão criar uma real possibilidade de compartilhamento e comparação dos perfis moleculares a partir da caracterização de bancos de germoplasma internacionais de feijoeiro comum, nos quais os genótipos de referência internacionais BAT 93, Jalo EEP 558 e G19833 poderão ser utilizados como uma conexão para a integração dos dados. Dessa forma, do ponto de vista prático, no mínimo quatro sistemas tetraplex compreendendo SSRs com elevado poder informativo ( $H_E$  combinada de 0,672) e discriminatório (PI combinada de  $1,309E-14$ ) constituem-se ferramentas moleculares para uso em âmbito internacional. Adicionalmente, novos marcadores SSRs estão disponíveis para serem avaliados em estudos futuros, assim como a possibilidade de desenhos alternativos de novos iniciadores visando ampliar a compatibilidade das co-amplificações reduzindo custos e atribuindo agilidade ao processo de genotipagem em larga escala.

## TABELAS E FIGURAS

**Tabela 1:** Caracterização dos BES-SSRs identificados incluindo o número total de marcadores por tipo de motivo, estimativa das frequências dos motivos em relação ao número total de SSRs identificados e classificação com base no comprimento em pares de base da repetição em Classe I (SSRs  $\geq$  20 pb) e Classe II (SSRs  $\leq$  19 pb).

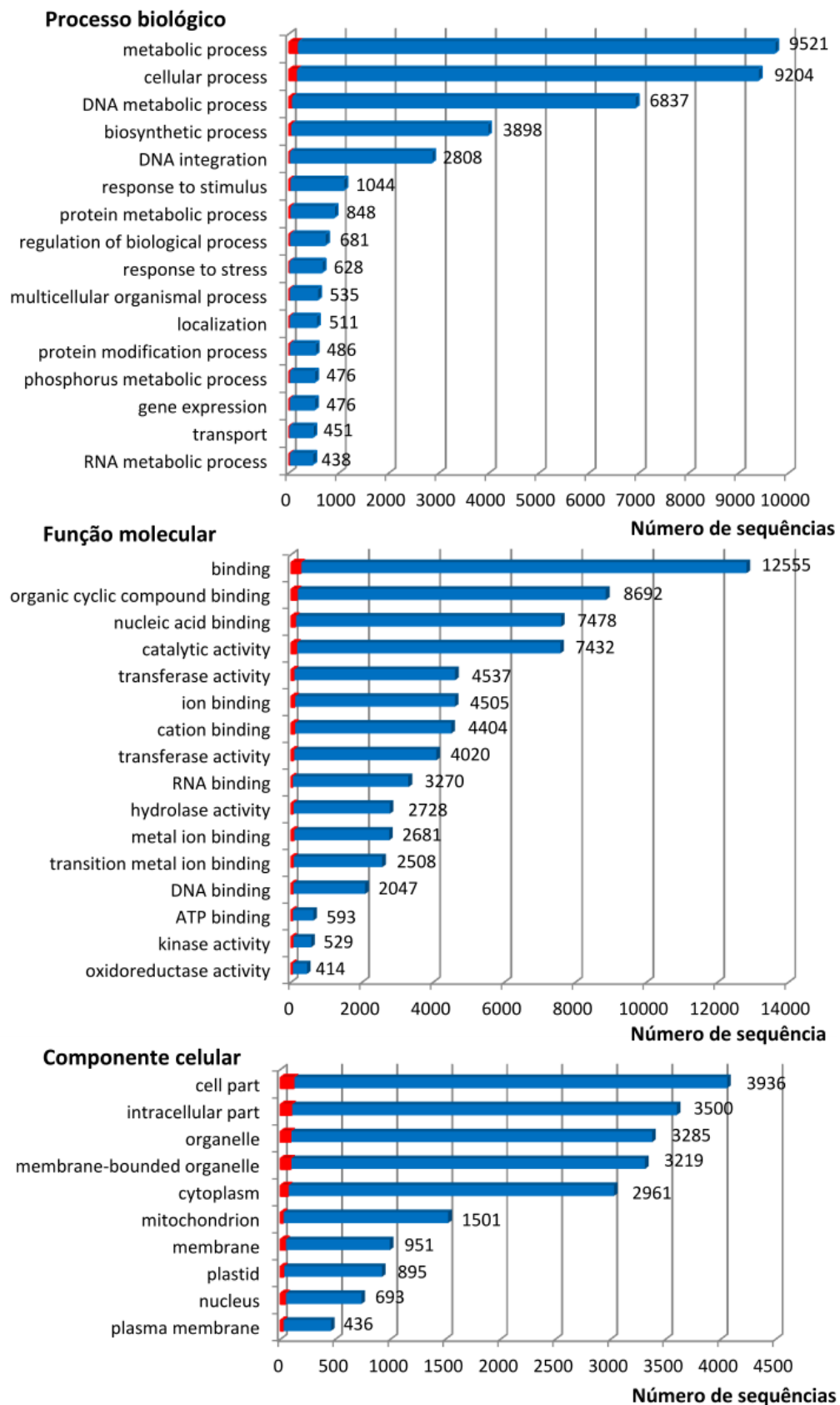
Motivo de repetição	Número de SSRs	Frequência (%)	Classe I	Classe II
Dinucleotídeo	559	27,95	167	392
Trinucleotídeo	915	45,75	37	878
Tetranucleotídeo	77	3,85	17	60
Pentanucleotídeo	19	0,95	19	0
Hexanucleotídeo	42	2,10	42	0
Composto	388	19,40	362	26
Total de SSRs	2000	100%	644	1356

**Tabela 2:** Frequência individual dos tipos de motivos BES-SSRs, mais representativos dos 6,5% do genoma do feijoeiro comum, categorizados de acordo com o número de número de repetições que compõem o motivo.

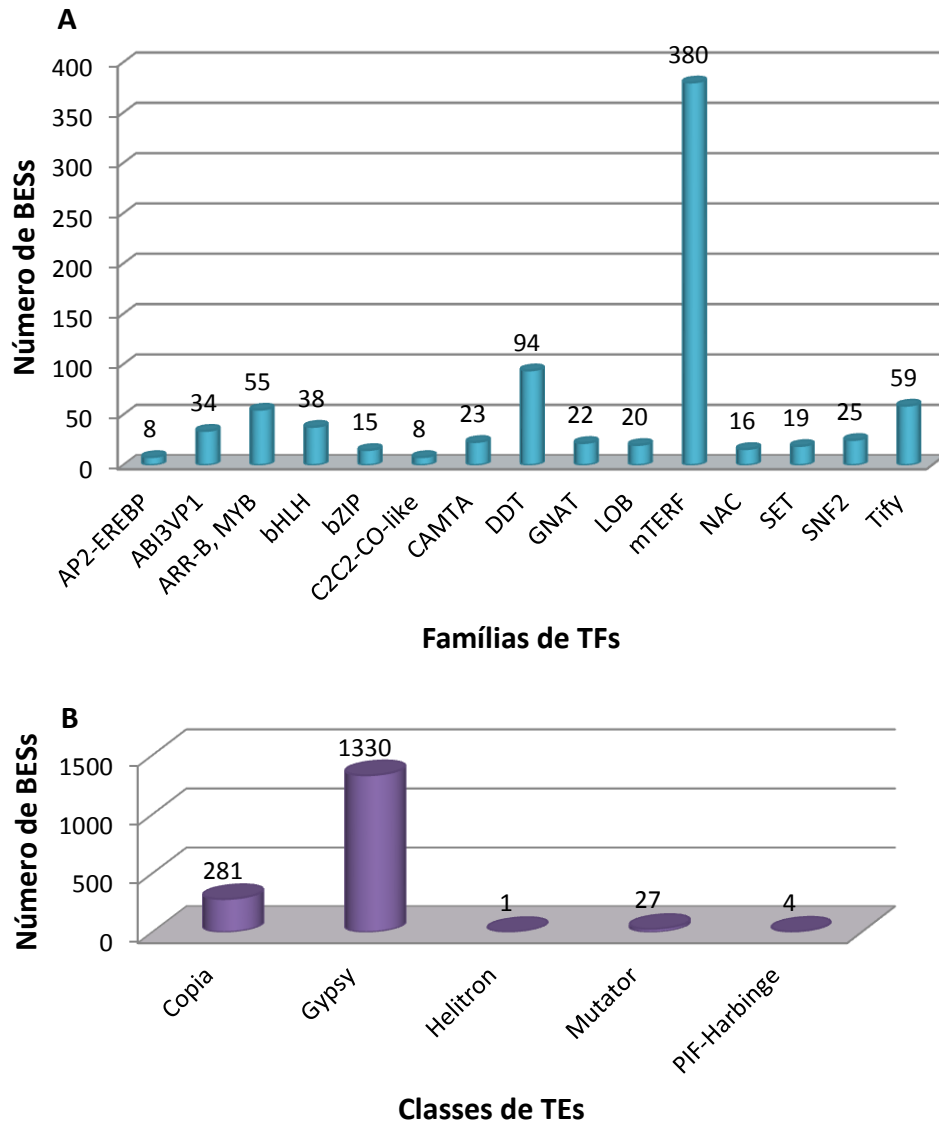
Motivo de repetição	Número de SSRs	Frequência (%)
<b>Dinucleotídeo</b>		
AT TA	332	16,57
AG GA CT TC	141	7,09
AC CA TG GT	85	4,29
GC CG	1	0,05
<b>Trinucleotídeo</b>		
AAG AGA GAA CTT TTC TCT	343	17,12
AAT ATA TAA ATT TTA TAT	197	9,83
ATG TGA GAT CAT ATC TCA	121	6,04
AAC ACA CAA GTT TTG TGT	95	4,74
AGG GGA GAG CCT CTC TCC	58	2,89
AGC GCA CAG GCT CTG TGC	55	2,74
ACC CCA CAC GGT GTG TGG	35	1,75
GGC GCG CGG GCC CCG CGC	5	0,25
<b>Tetranucleotídeo</b>		
AAAT AATA ATAA TAAA TTTA TTAT TATT ATTT	48	2,40
<b>Pentanucleotídeo</b>		
AAAAT AAATA AATAA ATAAA TAAAA TTTTA TTTAT TTATT TATTT ATTTT	16	0,80
<b>Hexanucleotídeo</b>		
GGCTTG GCTTGG CTTGGG TTGGGC TGGGCT GGGCTT CCGAAC CGAACC GAACCC AACCCG ACCCGA CCCGAA	13	0,65

**Tabela 3:** Descritores genéticos referentes aos 34 BES-SSRs de feijoeiro comum analisados, incluindo o número de alelo por loco ( $nA$ ), heterozigosidade observada ( $H_o$ ), diversidade gênica ( $H_E$ ), probabilidade de identidade ( $PI$ ) e estimativa da maior frequência alélica ( $Fmáx$ ).

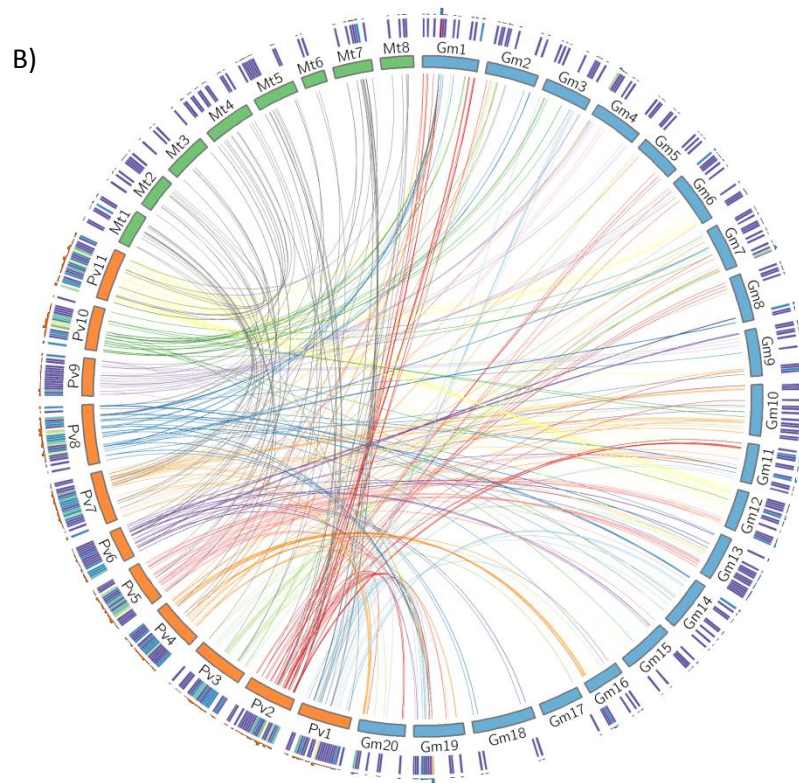
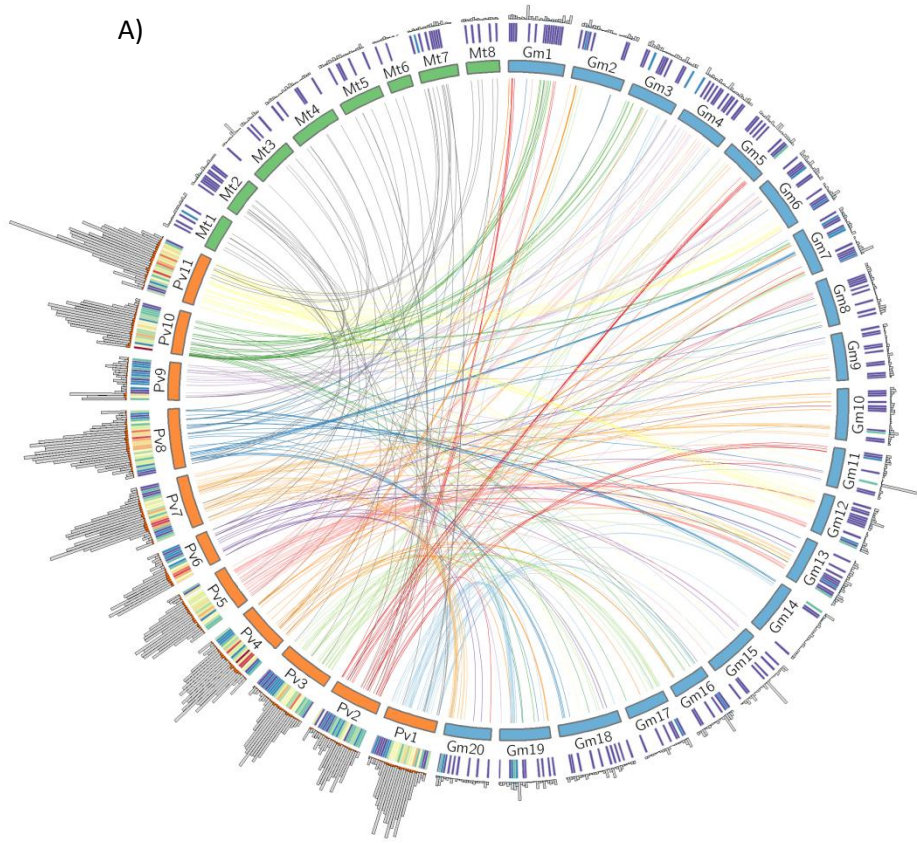
<b>Marcador</b>	<b><math>nA</math></b>	<b><math>H_o</math></b>	<b><math>H_E</math></b>	<b><math>PI</math></b>	<b><math>Fmáx</math></b>
<b>Composto</b>					
PvComp10	19	0,013	0,924	0,013	0,149
PvComp4	17	0	0,821	0,047	0,376
PvComp9	9	0,023	0,707	0,125	0,465
PvComp2	5	0,024	0,678	0,16	0,464
PvComp8	9	0,012	0,609	0,195	0,583
PvComp21	5	0	0,511	0,314	0,635
PvComp27	3	0	0,391	0,44	0,743
<b>Média</b>	<b>9,571</b>	<b>0,010</b>	<b>0,663</b>	<b>3,29E-07</b>	<b>0,488</b>
<b>Hexanucleotídeo</b>					
PvHexa20	3	0	0,506	0,362	0,558
PvHexa10	3	0,012	0,456	0,387	0,671
PvHexa19	3	0	0,418	0,419	0,712
PvHexa12	2	0	0,336	0,499	0,788
PvHexa36	4	0	0,231	0,606	0,873
PvHexa39	3	0	0,209	0,642	0,885
PvHexa15	9	0	0,186	0,667	0,902
<b>Média</b>	<b>3,857</b>	<b>0,001</b>	<b>0,334</b>	<b>7,60E-03</b>	<b>0,770</b>
<b>Pentanucleotídeo</b>					
PvPenta13	3	0	0,614	0,226	0,506
PvPenta4	3	0	0,559	0,275	0,575
PvPenta10	6	0,012	0,437	0,342	0,738
PvPenta14	2	0	0,336	0,499	0,788
PvPenta19	4	0	0,314	0,488	0,822
PvPenta8	2	0	0,265	0,578	0,844
PvPenta16	3	0	0,168	0,7	0,910
PvPenta5	3	0	0,134	0,755	0,929
<b>Média</b>	<b>3,250</b>	<b>0,001</b>	<b>0,353</b>	<b>1,58E-03</b>	<b>0,764</b>
<b>Tetranucleotídeo</b>					
PvTetra25	12	0,012	0,745	0,098	0,424
PvTetra65	5	0,012	0,657	0,182	0,476
PvTetra73	4	0	0,622	0,2	0,544
PvTetra57	6	0	0,592	0,211	0,597
PvTetra49	3	0,014	0,478	0,354	0,659
PvTetra32	2	0,012	0,458	0,4	0,648
PvTetra50	4	0	0,453	0,331	0,723
PvTetra47	4	0	0,43	0,393	0,712
PvTetra76	2	0	0,199	0,663	0,889
<b>Média</b>	<b>4,667</b>	<b>0,005</b>	<b>0,515</b>	<b>9,19E-06</b>	<b>0,631</b>
<b>Trinucleotídeo</b>					
PvTri8	11	0	0,826	0,052	0,325
PvTri6	7	0	0,747	0,111	0,337
PvTri5	10	0	0,639	0,179	0,535
<b>Média</b>	<b>9,333</b>	<b>0</b>	<b>0,737</b>	<b>1,03E-03</b>	<b>0,399</b>

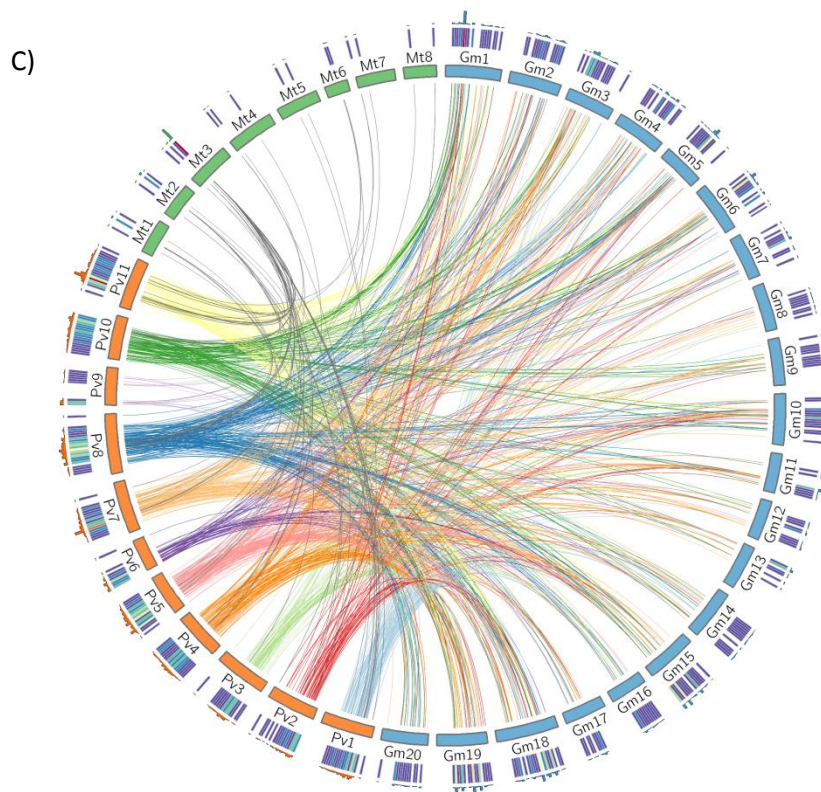


**Figura 1:** Classificação funcional dos 52.270 BES (azul) e dos 2.000 BES com microssatélites (BES-SSRs) (vermelho) derivados de *P. vulgaris* a partir da anotação de termos GO, incluindo as atribuições baseadas em processo biológico, função molecular e componente celular.



**Figura 2:** Associação (A) das famílias de fatores de transcrição (TFs) e (B) das classes de elementos transponíveis (TEs) com as BESs. Somente as maiores famílias de TFs foram representadas.





**Figura 3: Análise comparativa entre os genomas de *Phaseolus vulgaris*, *Medicago truncatula* e *Glycine max*.** Análise comparativa de alta resolução das BAC-end sequences versus as seqüências dos genomas das leguminosas *P. vulgaris* Andino (barras alaranjadas), *M. truncatula* (barras verdes) e *G. max* (barras azuis). (A) Análise comparativa das BAC-end sequences com microssatélites (BES-SSRs) versus seqüências dos genomas de leguminosas: O esquema (A) inclui três círculos concêntricos. O círculo mais externo representa o alinhamento do total de 52.270 BES (cinza) contra os três genomas de leguminosas, independentemente. O círculo intermediário representa o *heat map*, barras coloridas indicando a densidade do BLAST hits por dois Mb de seqüência (2 Mb). O círculo interno está representando os cromossomos haploides dos genomas de *P. vulgaris* (Pv1-11), *M. truncatula* (Mt1-8) e *G. max* (Gm1-20). As linhas coloridas representam as BES-SSRs com ortologia na relação comparativa entre os três genomas de leguminosas. Os esquemas (B) e (C) apresentam dois círculos concêntricos, o primeiro mais externo representa o *heat map* e o segundo círculo interno representa os cromossomos haploides dos genomas. As linhas coloridas representam os (B): Fatores de Transcrições (TF) e (C): Elementos Transponíveis (TE) com ortologia na relação entre os três genomas de leguminosas.

## REFERÊNCIAS

Agüero F, Verdún RE, Frasc AC, Sánchez DO (2000) A random sequencing approach for the analysis of the *Trypanosoma cruzi* genome: general structure, large gene and repetitive DNA families, and gene discovery. *Genome Res* 10(12):1996–2005

Arumuganatham K, Early ED (1991) Nuclear DNA content of some important plant species. *Plant Mol Biol Rep* 9(3):208–218

Asfaw A, Blair MW, Struik PC (2012) Multienvironment quantitative trait loci analysis for photosynthate acquisition, accumulation, and remobilization traits in common bean under drought stress. *G3 (Bethesda)* 2(5):579–595. doi: 10.1534/g3.112.002303

Ashburner M, Ball CA, Blake JA et al (2000) Gene Ontology : tool for the unification of biology. *Nat Genet* 25(1):25–29. doi: 10.1038/75556

Biswas MK, Chai L, Mayer C, Xu Q, Guo W, Deng X (2012) Exploiting BAC-end sequences for the mining, characterization and utility of new short sequence repeat (SSR) markers in *Citrus*. *Mol Biol Rep* 39(5):5373–5386. doi: 10.1007/s11033-011-1338-5

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. doi: 10.1007/s00122-003-1398-6

Blair MW, Astudillo C, Rengifo J, Beebe SE, Graham R (2011a) QTL analyses for seed iron and zinc concentrations in an intra-genepool population of Andean common beans (*Phaseolus vulgaris* L.). *Theor Appl Genet* 122(3):511–521. doi: 10.1007/s00122-010-1465-8

Blair MW, Hurtado N, Chavarro CM, Muñoz-Torres MC, Giraldo MC, Pedraza F, Tomkins J, Wing R (2011b) Gene-based SSR markers for common bean (*Phaseolus vulgaris* L.) derived from root and leaf tissue ESTs: an integration of the BMC series. *BMC Plant Biol* 11:50. doi: 10.1186/1471-2229-11-50

Blair MW, Hurtado N, Sharma P (2012a) New gene-derived simple sequence repeat markers for common bean (*Phaseolus vulgaris* L.). *Mol Ecol Resour* 12:661–668. doi: 10.1111/j.1755-0998.2012.03136.x

Blair MW, Soler A, Cortés AJ (2012b) Diversification and population structure in common beans (*Phaseolus vulgaris* L.). *PLoS ONE* 7(11):e49488. doi: 10.1371/journal.pone.0049488

Blair MW, Hurtado N (2013) EST-SSR markers from five sequenced cDNA libraries of common bean (*Phaseolus vulgaris* L.) comparing three bioinformatic algorithms. *Mol Ecol Resour* 13(4):688–695. doi: 10.1111/1755-0998.12099

Bohra A, Dubey A, Saxena RK et al (2011) Analysis of BAC-end sequences (BESs) and development of BES-SSR markers for genetic mapping and hybrid purity assessment in pigeonpea (*Cajanus* spp.). *BMC Plant Biol* 11:56. doi: 10.1186/1471-2229-11-56

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

Budowle B, Moretti TR, Baumstark AL, Defenbaugh DA, Keys KM (1999) Population data on the thirteen CODIS core short tandem repeat loci in African Americans, U.S. Caucasians, Hispanics, Bahamians, Jamaicans, and Trinidadians. *J Forensic Sci* 44(6):1277–1286

Burle ML, Fonseca JR, Kami JA, Gepts P (2010) Microsatellite diversity and genetic structure among common bean (*Phaseolus vulgaris* L.) landraces in Brazil, a secondary center of diversity. *Theor Appl Genet* 121:801–813. doi: 10.1007/s00122-010-1350-5

Byrne N, Marquez-Garcia M, U, Smith DS, Moran GF (1996) Conservation and genetic diversity of microsatellite loci in the genus *Eucalyptus*. *Aust J Bot* 44(3):331–341. doi:10.1071/BT9960331

Cannon SB, May GD, Jackson SA (2009) Three sequenced legume genomes and many crop species: rich opportunities for translational genomics. *Plant Physiol* 151(3):970–977

Castelo AT, Martins W, Gao GR (2002) TROLL--tandem repeat occurrence locator. *Bioinformatics* 18(4):634–636

Cho YG, Ishii T, Temnykh S, Chen, X, Lipovich L, McCouch SR, Park WD, Ayres N, Cartinhour S (2000) Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). *Theor Appl Genet* 100(5):713–722

Choi H, Kim D, Uhm T, Limpens E, Lim H, Mun JH, Kalo P, Penmetsa RV, Seres A, Kulikova O, Roe BA, Bisseling T, Kiss GB, Cook DR (2004) A sequence-based genetic map of *Medicago truncatula* and comparison of marker colinearity with *M. sativa*. *Genetics* 166(3):1463–1502

Chou, H, Holmes, MH (2001) DNA sequence quality trimming and vector removal. *Bioinformatics* 17(12):1093–1104

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21(18):3674–3676. doi: 10.1093/bioinformatics/bti610

Córdoba JM, Chavarro C, Schlueter JA, Jackson SA, Blair MW (2010) Integration of physical and genetic maps of common bean through BAC-derived microsatellite markers. *BMC Genomics* 11:436. doi: 10.1186/1471-2164-11-436

Creste S, Tulmann Neto A, Figueira A (2001) Detection of single sequence repeat polymorphisms in denaturing polyacrylamide sequencing gels by silver staining. *Plant Mol Biol Rep* 19(4): 299–306

Dereeper A, Guyot R, Tranchant-Dubreuil C et al (2013) BAC-end sequences analysis provides first insights into coffee (*Coffea canephora* P.) genome composition and evolution. *Plant Mol Biol* 83(3):177–189. doi: 10.1007/s11103-013-0077-5

Doyle JJ, Luckow MA (2003) The rest of the iceberg. Legume diversity and evolution in a phylogenetic context. *Plant Physiol* 131(13):900–910. doi: 10.1104/pp.102.018150.groups

Du J, Grant D, Tian Z, Nelson RT, Zhu L, Shoemaker RC, Ma J (2010) SoyTEdb: a comprehensive database of transposable elements in the soybean genome. *BMC Genomics* 11:113. doi: 10.1186/1471-2164-11-113

Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 8(3):175–185. doi: 10.1101/gr.8.3.175

FAO - Food and Agriculture Organization of the United Nations - Statistical Database (2009). Available at <http://www.fao.org/faostat>. Accessed 10 Jan 2013

Frei dit Frey N, Mbengue M, Kwaaitaal M et al (2012) Plasma membrane calcium ATPases are important components of receptor-mediated signaling in plant immune responses and development *Plant Physiol* 159(2):798–809. doi: 10.1104/pp.111.192575

Galeano CH, Fernandez AC, Franco-Herrera N, Cichy KA, McClean PE, Vanderleyden J, Blair MW (2011) Saturation of an intra-gene pool linkage map: towards a unified consensus linkage map for fine mapping and synteny analysis in common bean. *PLoS ONE* 6(12):e28135. doi: 10.1371/journal.pone.0028135

Galeano CH, Cortés AJ, Fernández AC, Soler Á, Franco-Herrera N, Makunde G, Vanderleyden J, Blair MW (2012) Gene-based single nucleotide polymorphism markers for genetic and association mapping in common bean. *BMC Genet* 13:48. doi: 10.1186/1471-2156-13-48

Garcia RAV, Rangel PN, Brondani C, Martins WS, Melo LC, Carneiro MS, Borba TC, Brondani RP (2011) The characterization of a new set of EST-derived simple sequence repeat (SSR) markers as a resource for the genetic analysis of *Phaseolus vulgaris*. *BMC Genet* 12:41. doi: 10.1186/1471-2156-12-41

Gepts P, Beavis WD, Brummer EC, Shoemaker RC, Stalker HT, Weeden NF, Young ND.(2005) Legumes as a model plant family. genomics for food and feed report of the cross-legume advances through genomics conference. *Plant Physiol* 137(4):1228–1235. doi: <http://dx.doi.org/10.1104/pp.105.060871>

Grattapaglia D, Sederoff R (1994) Genetic linkage maps of *Eucalyptus grandis* and *E. urophylla* using a pseudo-testcross strategy and RAPD markers. *Genetics* 137(4):1121–1137

Hanai LR, de Campos T, Camargo LE et al. (2007) Development, characterization, and comparative analysis of polymorphism at common bean SSR loci isolated from genic and genomic sources. *Genome* 50(3):266–277. doi: 10.1139/G07-007

Hanai LR, Santini L, Camargo LE, Fungaro MH, Gepts P, Tsai SM, Vieira ML (2010) Extension of the core map of common bean with EST-SSR, RGA, AFLP, and putative functional markers. *Mol Breed* 25(1):25–45. doi: 10.1007/s11032-009-9306-7

Huang X, Madan A (1999) CAP3: A DNA Sequence Assembly Program. *Genome Res* 9(9):868–877. doi: 10.1101/gr.9.9.868

Kalavacharla V, Liu Z, Meyers BC, Thimmapuram J, Melmaiee K (2011) Identification and analysis of common bean (*Phaseolus vulgaris* L.) transcriptomes by massively parallel pyrosequencing. *BMC Plant Biol* 11:135. doi: 10.1186/1471-2229-11-135

Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol Ecol* 16(5):1099–1106. doi: 10.1111/j.1365-294X.2007.03089.x

Kami J, Poncet V, Geffroy V, Gepts P (2006) Development of four phylogenetically-arrayed BAC libraries and sequence of the APA locus in *Phaseolus vulgaris*. *Theor Appl Genet* 112(6):987–998. doi: 10.1007/s00122-005-0201-2

Kang W, Hérbert JM (2012) A Sox2 BAC transgenic approach for targeting adult neural stem cells. *PLoS ONE* 7(11):e49038. doi: 10.1371/journal.pone.0049038

Katti M V, Ranjekar PK, Gupta VS (2001) Differential distribution of simple sequence repeats in eukaryotic genome sequences. *Mol Biol Evol* 18(7):1161–1167

Kim H, San Miguel P, Nelson W, Collura K, Wissotski M, Walling JG, Kim JP, Jackson SA, Soderlund C, Wing RA (2007) Comparative physical mapping between *Oryza sativa* (AA genome type) and *O. punctata* (BB genome type). *Genetics* 176(1):379–390. doi: 10.1534/genetics.106.068783

Kim C, Lee TH, Compton RO, Robertson JS, Pierce GJ, Paterson AH (2013) A genome-wide BAC end-sequence survey of sugarcane elucidates genome composition, and identifies BACs covering much of the euchromatin. *Plant Mol Biol* 81(1-2):139–147. doi: 10.1007/s11103-012-9987-x

Kimpton CP, Oldroyd NJ, Watson SK, Frazier RR, Johnson PE, Millican ES, Urquhart A, Sparkes BL, Gill P (1996) Validation of highly

discriminating multiplex short tandem repeat amplification systems for individual identification. *Electrophoresis* 17(8):1283–1293

Kleine T (2012) *Arabidopsis thaliana* mTERF proteins: evolution and functional classification. *Front Plant Sci* 3:233. doi: 10.3389/fpls.2012.00233

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA (2009) Circos: an information aesthetic for comparative genomics. *Genome Res* 19(9):1639–1645. doi: 10.1101/gr.092759.109

Kwak M, Gepts P (2009) Structure of genetic diversity in the two major gene pools of common bean (*Phaseolus vulgaris* L., Fabaceae). *Theor Appl Genet* 118(5): 979–992. doi: 10.1007/s00122-008-0955-4

Lawson MJ, Zhang L (2006) Distinct patterns of SSR distribution in the *Arabidopsis thaliana* and rice genomes. *Genome Biol* 7(2):R14. doi: 10.1186/gb-2006-7-2-r14

Leclercq S, Rivals E, Jarne P (2007) Detecting microsatellites within genomes : significant variation among algorithms. *BMC Bioinformatics* 8:125. doi: 10.1186/1471-2105-8-125

Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21(9):2128–2129. doi: 10.1093/bioinformatics/bti282

Liu SY, Yu K, Huffner M, Park SJ, Banik M, Pauls KP, Crosby W (2010) Construction of a BAC library and a physical map of a major QTL for CBB resistance of common bean (*Phaseolus vulgaris* L.). *Genetica* 138(7):709–716. doi: 10.1007/s10709-010-9450-3

Martins W, de Sousa D, Proite K, Guimarães P, Moretzsohn M, Bertoli D (2006) New softwares for automated microsatellite marker development. *Nucleic Acids Res* 34(4):e31. doi: 10.1093/nar/gnj030

McClellan PE, Mamidi S, McConnell M, Chikara S, Lee R (2010) Synteny mapping between common bean and soybean reveals extensive blocks of shared loci. *BMC Genomics* 11:184. doi: 10.1186/1471-2164-11-184

Melotto M, Monteiro-vitorello CB, Bruschi AG, Camargo LEA (2005) Comparative bioinformatic analysis of genes expressed in common bean (*Phaseolus vulgaris* L.) seedlings. *Genome* 48(3): 562–570. doi: 10.1139/G05-010

Morgante M, Hanafey M, Powell W (2002) Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nat Genet* 30(2):194–200. doi: 10.1038/ng822

Noguero M, Atif RM, Ochatt S, Thompson RD (2013) The role of the DNA-binding One Zinc Finger (DOF) transcription factor family in plants. *Plant Sci* 209:32–45. doi: 10.1016/j.plantsci.2013.03.016

Peakall R, Smouse PE (2012) GenAEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. *Bioinformatics* 28(19):2537–2539. doi: 10.1093/bioinformatics/bts460

Pérez-Rodríguez P, Riaño-Pachón DM, Corrêa LG, Rensing SA, Kersten B, Mueller-Roeber B (2010) PlnTFDB: updated content and new features of the plant transcription factor database. *Nucleic Acids Res* 38(1):D822–827. doi: 10.1093/nar/gkp805

Ragupathy R, Rathinavelu R, Cloutier S (2011) Physical mapping and BAC-end sequence analysis provide initial insights into the flax (*Linum usitatissimum* L.) genome. *BMC Genomics* 12:217. doi: 10.1186/1471-2164-12-217

Renwick A, Davison L, Spratt H, King JP, Kimmel M (2001) DNA dinucleotide evolution in humans: fitting theory to facts. *Genetics* 159(2): 737–747

Ribeiro T, dos Santos KGB, Fonsêca A, Pedrosa-Harand A (2011) Isolation and characterization of a new repetitive DNA family recently amplified in the Mesoamerican gene pool of the common bean (*Phaseolus vulgaris* L., Fabaceae). *Genetica* 139(9):1135–1142. doi: 10.1007/s10709-011-9615-8

Roberti M, Polosa PL, Bruni F, Manzari C, Deceglie S, Gadaleta MN, Cantatore P (2009) The MTERF family proteins: mitochondrial transcription regulators and beyond. *Biochim Biophys Acta* 1787(5):303–311. doi: 10.1016/j.bbabi.2009.01.013

Rozen S, Skaletsky H (2000) Primer3 for general users and for biologist programmers. In *Bioinformatics Methods and Protocols*. Edited by: Krawetz S, Misener S. The Humana Press Inc, New Jersey. *Methods Mol Biol* 132:365–386

Saini N, Shultz J, Lightfoot DA (2008) Re-annotation of the physical map of *Glycine max* for polyploid-like regions by BAC end sequence driven whole genome shotgun read assembly. *BMC Genomics* 9:323. doi: 10.1186/1471-2164-9-323

Schlueter JA, Goicoechea JL, Collura K, et al (2008) BAC-end sequence analysis and a draft physical map of the common bean (*Phaseolus vulgaris* L.) genome. *Trop Plant Biol* 1(1):40–48. doi: 10.1007/s12042-007-9003-9

Schmutz J, Cannon SB, Schlueter J et al (2010) Genome sequence of the palaeopolyploid soybean. *Nature* 463(7278):178–183. doi: 10.1038/nature08670

Shoemaker RC, Grant D, Olson T et al (2008) Microsatellite discovery from BAC end sequences and genetic mapping to anchor the soybean physical and genetic maps. *Genome* 51(4):294–302. doi: 10.1139/G08-010

Shultz JL, Kazi S, Bashir R, Afzal JA, Lightfoot DA (2007) The development of BAC-end sequence-based microsatellite markers and placement in the physical and genetic maps of soybean. *Theor Appl Genet* 114(6):1081–1090. doi: 10.1007/s00122-007-0501-9

Sonah H, Deshmukh RK, Sharma A, Singh VP, Gupta DK, Gacche RN, Rana JC, Singh NK, Sharma TR (2011) Genome-wide distribution and organization of microsatellites in plants: an insight into marker development in *Brachypodium*. *PLoS ONE* 6(6):e21298. doi: 10.1371/journal.pone.0021298

Staden R, Beal KF, Bonfield JK (1999) The staden package, 1998. In *Bioinformatics Methods and Protocols*. Edited by: Misener S, Krawetz SA. The Humana Press Inc, New Jersey. *Methods Mol Biol* 132:115–130

Tautz D (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Res* 17(16):6463-6471. doi: 10.1093/nar/17.16.6463

Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, 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(8):1441–1452. doi: 10.1101/gr.184001

Thudi M, Bohra A, Nayak SN et al (2011) Novel SSR markers from BAC-end sequences, DArT arrays and a comprehensive genetic map with 1,291 marker loci for chickpea (*Cicer arietinum* L.). *PLoS ONE* 6(11):e27275. doi: 10.1371/journal.pone.0027275

Tóth G, Gáspári Z, Jurka J (2000) Microsatellites in Different Eukaryotic Genomes: Survey and Analysis. *Genome Res* 10(7):967-981. doi: 10.1101/gr.10.7.967

Vallone PM, Butler JM (2004) AutoDimer: a screening tool for primer-dimer and hairpin structures. *Biotechniques* 37(2):226–231

Vianello-Brondani RP, Cardoso PCB, Menezes IPP, Valdisser PAMR, Borba TCO, Peloso MJD, Brondani C (2013) Discrimination of common bean commercial cultivars using panels of highly informative microsatellite markers. *Genet Mol Res: Ahead of Print*

Victoria FC, Maia LC, de Oliveira AC (2011) In silico comparative analysis of SSR markers in plants. *BMC Plant Biol* 11:15. doi: 10.1186/1471-2229-11-15

Wang H, Penmetsa RV, Yuan M et al (2012) Development and characterization of BAC-end sequence derived SSRs, and their incorporation into a new higher density genetic map for cultivated peanut (*Arachis hypogaea* L.). *BMC Plant Biol* 12:10. doi: 10.1186/1471-2229-12-10

Wang Z, Yan H, Fu X, Li X, Gao H (2013) Development of simple sequence repeat markers and diversity analysis in alfalfa (*Medicago sativa* L.). *Mol Biol Rep* 40(4):3291–3298. doi: 10.1007/s11033-012-2404-3

Wu J, Gu YQ, Hu Y, You FM, Dandekar AM, Leslie CA, Aradhya M, Dvorak J, Luo MC (2012) Characterizing the walnut genome through analyses of BAC end sequences. *Plant Mol Biol* 78(1-2):95–107. doi: 10.1007/s11103-011-9849-y

You FM, Huo N, Gu YQ, Luo MC, Ma Y, Hane D, Lazo GR, Dvorak J, Anderson OD (2008) BatchPrimer3: a high throughput web application for PCR and sequencing primer design. *BMC Bioinformatics* 9:253. doi: 10.1186/1471-2105-9-253

Young ND, Debelle F, Oldroyd GED et al (2011) The *Medicago* genome provides insight into the evolution of rhizobial symbioses. *Nature* 480(7378):520–524. doi: 10.1038/nature10625

Yu K (2012) Bacterial artificial chromosome libraries of pulse crops: characteristics and applications. *J Biomed Biotechnol* 2012:493186. doi: 10.1155/2012/493186

Yu K, Park SJ, Poysa V, Gepts P (2000) Integration of simple sequence repeat (SSR) markers into a molecular linkage map of common bean (*Phaseolus vulgaris* L.). *J Hered* 91(6): 429–434. doi: 10.1093/jhered/91.6.429

## APÊNDICES

**Material Suplementar 1:** Descrição dos 40 BES-SSRs utilizados no desenvolvimento de sistemas de genotipagem semi-automatizados, com as respectivas informações: marcação fluorescente, sequências dos primers *Foward (FWD)* e *Reverse (REV)*, tamanho esperado do fragmento amplificado em pares de base (pb), tipo de motivo, classificação quanto ao tamanho da repetição, informação da sequência de origem e resultado da anotação com base no BLASTX:

<i>Primers</i>	<i>Cor</i>	<i>FWD primer</i>	<i>REV primer</i>	<b>Tamanho (pb)</b>	<b>Motivo</b>	<b>Classes SSRs</b>	<b>Nome Seq.</b>	<b>Descrição Seq. Blastx</b>
PvTetra7	6-FAM	CAGTGTGGCCAAAATTCTAACA	CTGTCCAATATCCAAACTGCAA	130	(AAAT)4	Classe II	PVMBBa0022E01.r.scf	---NA---*
PvComp4	6-FAM	TCTTCACCACTTTGAAAACACG	TCAGATAAATGTTGGTATTGGCA	217	(AATATC)5 (AATATA)7 (AT)9	Classe I	PVMBBa0069L17.r.scf	---NA---*
PvPenta4	6-FAM	AGCAACTTTTCGGTCTGGTAAAGT	TGAATCATTGCTCCTAACCCCTT	239	(ATTTT)4	Classe I	PVMBBa0040J18.r.scf	constans-like zinc finger protein
PvPenta5	6-FAM	AACTGTTCTTGTGGGTTCAAT	GCCCAAGGAGTAATTAACAATAAAA	254	(TTTAT)4	Classe I	PVMBBa0002H18.f.scf	---NA---*
PvComp9	6-FAM	TGAATGCAACATCTAATACTAACTCAC	TGTTCTAGGTCTAAAGGCCACA	267	(AT)10 (TATG)4 (TATG)4 (GTAT)4 (AT)7	Classe I	PVMBBa0036O16.f.scf	---NA---*
PvPenta8	6-FAM	AAGAGCATGTTTACTTCACTCATT	TGGGGATGGTGTGGTGGT	290	(TTTAT)4	Classe I	PVMBBa0075I14.f.scf	---NA---*
PvTetra47	6-FAM	TGGAATGGAGAAGAGACATCCT	CAAACCATGTTCCAGCATCTA	321	(TTAG)4	Classe II	PVMBBa0019C23.f.scf	---NA---*
PvPenta14	6-FAM	GAAACTATTCACGGAACAAGCC	AGGAGTGGTGGAGGCAGTATAA	336	(AAAAT)4	Classe I	PVMBBa0037O10.f.scf	---NA---*
PvTetra65	6-FAM	TCTGAATCAAACATGTCTCAGA	GGCCAAACTGTTTAAGGTGAG	366	(AATA)4	Classe II	PVMBBa0055J14.r.scf	---NA---*
PvPenta18	6-FAM	GTTTTCTTGAATAGAGTTTGAATCG	CCTACAATGTGGTGGTTGAAT	393	(AATTT)4	Classe I	PVMBBa0025C01.f.scf	---NA---*
PvComp2	NED	CTCCGAATCAGAAACCCTATTG	GTGGATGAAGAGAAAGGCAAAG	201	(TCTTTC)6 (TTCTCC)5	Classe I	PVMBBa0040L02.r.scf	pat15_arath ame: full=patellin-5
PvTetra25	NED	CACTTTCCTTATGCCTTCACAG	TAATTTGACCAATGCCAAACAC	226	(ACTA)22	Classe I	PVMBBa0032C05.f.scf	---NA---*
PvTri6	NED	CGAATGGGAGAAGAAGGTTATG	GACATTCTGTGTCGTTTCCAA	242	(AGA)16	Classe I	PVMBBa0068I16.f.scf	copia-type polyprotein
PvHexa20	NED	GTGTCTTCTATAGGTGTCCCG	GAGTATTTCAAAGCTTGGCCTT	257	(CCCAAG)4	Classe I	PVMBBa0012D14.f.scf	mutant gag-pol polyprotein
PvTri5	NED	AGAAAATCATGCAGGTTGAGGT	TCCAGCTAAATAGATGATACGTAATTG	269	(TAA)10	Classe I	PVMBBa0067C05.f.scf	---NA---*
PvPenta10	NED	GTTCTTCTATTTCCATCTATC	AATATACATAAGAGTCACTTCT	297	(ATAAA)4	Classe I	PVMBBa0021C21.r.scf	---NA---*
PvTetra50	NED	TAACATGGTTAGGCCTTTTGAA	TCGTACGGATCCAAGTATTAATTT	324	(AAAT)4	Classe II	PVMBBa0015F10.f.scf	---NA---*
PvComp21	NED	ACGAGTTATTGTTCCAGATGGG	TTATCCTTCTTATGCGGAGACC	353	(AGATGA)4 (TG)5	Classe I	PVMBBa0086N03.f.scf	catalase-1 2
PvHexa39	NED	TTAATGCCTCCACTTGTGTTG	CATGAGGCCCAAGTCAAAA	380	(TTTTTC)5	Classe I	Cluster_727_2_Contig	gag-pol polyprotein
PvPenta19	NED	TTAGGTCTTCAAAGAGATTTGG	TGTGGTAGTAGATGTTAAAGTCATTT	399	(AATTT)4	Classe I	PVMBBa0035L01.r.scf	---NA---*
PvHexa10	PET	AACTGTTTATCGCATCCAGAA	ATGCAATCAAGGAATGCTCATA	184	(TGCACA)4	Classe I	PVMBBa0059N20.f.scf	---NA---*
PvTri4	PET	GTACACACGGTCTCCAAAGA	CATGTCTGCTTGTGAGTCT	225	(GAA)10	Classe I	PVMBBa0052P13.r.scf	hypothetical protein MTR_1g007770

**Material Suplementar 1:** Continuação.

<i>Primers</i>	<i>Cor</i>	<i>FWD primer</i>	<i>REV primer</i>	<b>Tamanho (pb)</b>	<b>Motivo</b>	<b>Classes SSRs</b>	<b>Nome Seq.</b>	<b>Descrição Seq. Blastx</b>
PvTri8	PET	CAATGTGGAACAACTGAGGTG	AAGCAAAGTGCTGAATTGCTG	240	(ATA)10	Classe	PVMBBa0057118	NA
PvTetra32	PET	ATTCCTGCCACTAACGAAGTGT	CTAAAGGCCTAGCAGATTGCAT	255	(AGAA)4	Classe II	PVMBBa0088D12.r.scf	calcium-transporting atpase
PvComp10	PET	AAATTCTATGATCAACCCGTGG	TGATCCCTTGTAGAGGAATCTCA	268	(TA)13 (ATTT)4	Classe I	PVMBBa0086J12.r.scf	---NA---*
PvPenta9	PET	GACCAAACGAAGCACCTAAAAT	CACAGTCATGTAAATTGTTGAAATC	293	(TAAAA)4	Classe I	PVMBBa0067J13.r.scf	---NA---*
PvTetra49	PET	TGGGTAGAGCTTGGTCTTCATT	AGTTTGTGAGTGATGTGATGGG	324	(ATAA)4	Classe II	PVMBBa0048O16.f.scf	---NA---*
PvHexa36	PET	TCACTTTGGCACCTCCTTATTT	AGGATTGTTTGCCTAAACCAGA	352	(CTTGGG)4	Classe I	PVMBBa0037K12.r.scf	---NA---*
PvPenta16	PET	ACATTTGGTTTTGGTTTTGGTC	TCTAAAATGGTCTCGAATTTATTCAC	379	(AATAA)4	Classe I	PVMBBa0035B15.r.scf	---NA---*
PvComp27	PET	CAAGATGAACATCACCATTCT	AAAAGAACATTTGTACGTCCA	395	(TATT)4 (ATA)6 (TTA)5	Classe I	PVMBBa0088P08.f.scf	---NA---*
PvHexa12	VIC	TGCTAAATAGCCAAAGCAACAA	CATCACACAGCACCAAGTATT	203	(AACTCA)5	Classe I	PVMBBa0036N02.r.scf	rna-directed dna polymerase (reverse transcriptase) chromo zinc cchc-type
PvHexa15	VIC	CCAAACGAACCGACTATTTCTC	TTTGACCTCCCTTTATGTTCT	235	(AAGCCC)4	Classe I	PVMBBa0090E12.r.scf	retrotransposon unclassified
PvHexa19	VIC	TGCTTCTCTGCTGACTTGGAA	AATTACAAGCCTGAAGCTGCTC	252	(TATGCT)4	Classe I	PVMBBa0062F11.r.scf	dual specificity protein kinase pyk1-like
PvComp8	VIC	TCACTATGTGAAATTGAACCCA	TTCCTACCTAACTTACTTGTACCACTT	262	(AT)7 (TAAC)4	Classe I	PVMBBa0064O08.f.scf	---NA---*
PvHexa26	VIC	TCACTTTGGCACCTCCTTATTT	ATCATTGCGCCTCACACTTAAA	289	(CTTGGG)4	Classe I	PVMBBa0050D17.r.scf	---NA---*
PvHexa28	VIC	AAACAAAGCATGAGGAAGGTGT	AAGATCATTGCGCTTCACATTT	314	(CTTGGG)4	Classe I	PVMBBa0061E11.r.scf	mutant gag-pol polyprotein
PvPenta13	VIC	ACTGAAGAAAGTACTAGAAACCTTAC A	CCCCTTTTAATCAGAGAATTTTA	331	(TAAAA)4	Classe I	PVMBBa0054A19.r.scf	---NA---*
PvTetra57	VIC	ATCAACTATGGCGGATTGACTT	GAAAACAAATCCTTTTGACCCA	359	(TTTA)5	Classe I	PVMBBa0084K21.f.scf	g-type lectin s-receptor-like serine threonine-protein kinase at4g27290-like
PvTetra73	VIC	TGGTATCGAAGCATTAGGTTCA	GTA CTGGGTAAACGGGTGTGAAT	389	(TTTG)4	Classe II	PVMBBa0084E01.f.scf	---NA---*
PvTetra76	VIC	TACTCAAGCTTCTTCTGCAC	TGAAATATATGTTGCGGAAT	399	(TTTA)4	Classe II	PVMBBa0069M02.r.scf	---NA---*

\*---NA--- Não anotada.

**Material Suplementar 2:** Informação dos 88 genótipos de feijoeiro comum, utilizados na análise de caracterização molecular dos marcadores SSRs e SNPs, quanto ao tipo de grão, o *pool* gênico de origem e o tipo de germoplasma melhorado ou tradicional.

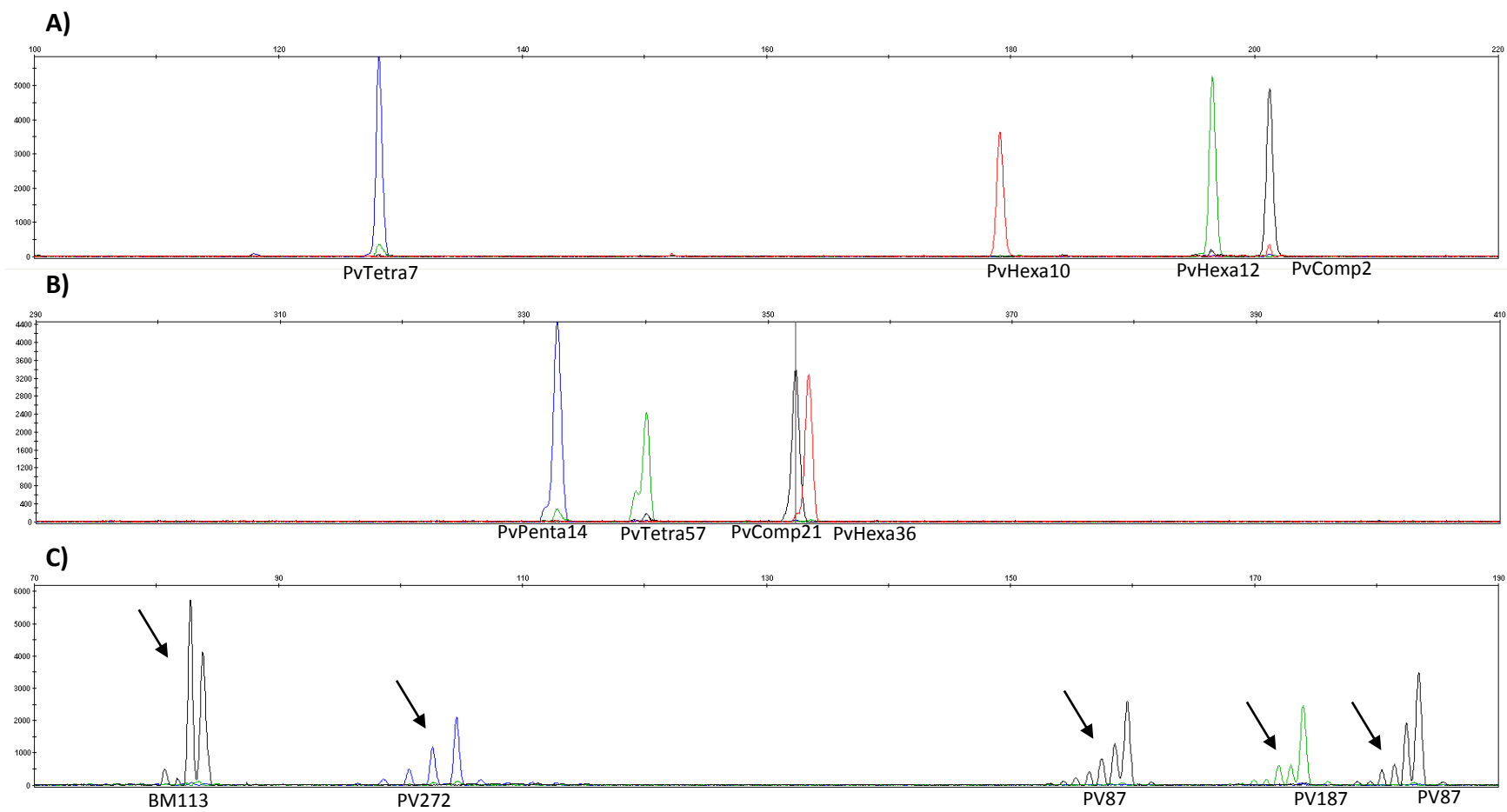
Número*	Genótipo	Tipo de grão	Origem	Germoplasma
1	AND 277	Vermelho	Andino	Melhorado
2	BRS Radiante	Rajado	Andino	Melhorado
3	WAF 75	Branco	Andino	Melhorado
4	Jalo Precoce	Jalo	Andino	Melhorado
5	Ouro Branco	Branco	Andino	Melhorado
6	BRS Executivo	<i>Cranberry</i>	Andino	Melhorado
7	BRSMG Realce	Rajado	Andino	Melhorado
8	Jalo EEP 558	Jalo	Andino	Melhorado
9	AND	Vermelho	Andino	Melhorado
10	BRS Embaixador	<i>Dark Red Kidney</i>	Andino	Melhorado
11	Cavalo de Rama - SC	Outros	Andino	Variedade tradicional
12	Cariocão - MG	Outros	Andino	Variedade tradicional
13	Jalão - MG	Jalo	Andino	Variedade tradicional
14	Crioulo Pintado - SC	Outros	Andino	Variedade tradicional
15	Jaulinho - MG	Jalo	Andino	Variedade tradicional
16	Massa Fina - MG	Outros	Andino	Variedade tradicional
17	Aspargo Redondo - PR	Outros	Andino	Variedade tradicional
18	Ameixa RS	Outros	Andino	Variedade tradicional
19	Preto Bonito/ Bagem Rosa - MG	Preto	Andino	Variedade tradicional
20	Amendoim SC	Roxo	Andino	Variedade tradicional
21	Jalo - GO	Jalo	Andino	Variedade tradicional
22	IPR 88 - Uirapuru	Preto	Mesoamericano	Melhorado
23	BRS Agreste	Mulatinho	Mesoamericano	Melhorado
24	Rudá	Carioca	Mesoamericano	Melhorado
25	BRS 7762 Supremo	Preto	Mesoamericano	Melhorado
26	Corrente	Mulatinho	Mesoamericano	Melhorado
27	CNFP10132	Preto	Mesoamericano	Melhorado
28	Pérola	Carioca	Mesoamericano	Melhorado
29	BRS Aurora	Branco	Mesoamericano	Melhorado
30	BRS Expedito	Preto	Mesoamericano	Melhorado
31	US Pinto 111	Pinto	Mesoamericano	Melhorado
32	Red Mexican	Vermelho	Mesoamericano	Melhorado
33	PI181996	Preto	Mesoamericano	Melhorado
34	BAT 477	Mulatinho	Mesoamericano	Melhorado
35	BRS Grafite	Preto	Mesoamericano	Melhorado
36	BRS Esplendor	Preto	Mesoamericano	Melhorado
37	Diamante Negro	Preto	Mesoamericano	Melhorado
38	Ouro Negro	Preto	Mesoamericano	Melhorado
39	Bola Cheia	Carioca	Mesoamericano	Melhorado
40	BRS Timbó	Roxo	Mesoamericano	Melhorado
41	BRS 9435 Cometa	Carioca	Mesoamericano	Melhorado
42	BRS Campeiro	Preto	Mesoamericano	Melhorado
43	Cuco	Carioca	Mesoamericano	Melhorado
44	BRSMG Majestoso	Carioca	Mesoamericano	Melhorado
45	BRS Estilo	Carioca	Mesoamericano	Melhorado

## Material Suplementar 2: Continuação.

Número*	Genótipo	Tipo de grão	Origem	Germoplasma
46	BRS Ametista	Carioca	Mesoamericano	Melhorado
47	BAT 93	Mulatinho	Mesoamericano	Melhorado
48	BRSMG Tesouro	Roxo	Mesoamericano	Melhorado
49	BRS Esteio	Preto	Mesoamericano	Melhorado
50	SEA 5	Mulatinho	Mesoamericano	Melhorado
51	BRS Valente	Preto	Mesoamericano	Melhorado
52	BRS Pitanga	Roxo	Mesoamericano	Melhorado
53	IAPAR 57	Carioca	Mesoamericano	Melhorado
54	BRS Pontal	Carioca	Mesoamericano	Melhorado
55	BRS Vereda	Rosinha	Mesoamericano	Melhorado
56	BRS 10408 Notável	Carioca	Mesoamericano	Melhorado
57	BRS Requite	Carioca	Mesoamericano	Melhorado
58	BRS Marfim	Mulatinho	Mesoamericano	Melhorado
59	IPR Juriti	Carioca	Mesoamericano	Melhorado
60	Magnífico	Carioca	Mesoamericano	Melhorado
61	IPR Colibri	Carioca	Mesoamericano	Melhorado
62	BRSMG Talismã	Carioca	Mesoamericano	Melhorado
63	CNFC 10729	Carioca	Mesoamericano	Melhorado
64	BRS Horizonte	Carioca	Mesoamericano	Melhorado
65	G23554	Mulatinho	Mesoamericano	Melhorado
66	PHAS 8328	Outros	Mesoamericano	Melhorado
67	Preto 60 Dias - GO	Preto	Mesoamericano	Variedade tradicional
68	Branco -ES	Mulatinho	Mesoamericano	Variedade tradicional
69	Chumbinho / Paranazinho - MG	Outros	Mesoamericano	Variedade tradicional
70	Mulatinho desconhecido - BA	Preto	Mesoamericano	Variedade tradicional
71	Mulatinho Vagem Roxa - BA	Mulatinho	Mesoamericano	Variedade tradicional
72	Rosinha - MG	Rosinha	Mesoamericano	Variedade tradicional
73	Pardinho 2 - MG	Roxo	Mesoamericano	Variedade tradicional
74	Vermelhinho Crioulo Marrom - SC	Rosinha	Mesoamericano	Variedade tradicional
75	Barriga Verde - PR	Preto	Mesoamericano	Variedade tradicional
76	Vermelho 1 Epamig - MG	Roxo	Mesoamericano	Variedade tradicional
77	Bico Ouro Carioca - MG	Carioca	Mesoamericano	Variedade tradicional
78	Carioca Determinado - DF	Carioca	Mesoamericano	Variedade tradicional
79	Mulatinho Vagem Roxa e Branca - MG	Mulatinho	Mesoamericano	Variedade tradicional
80	Mulatinho 3 - MG	Mulatinho	Mesoamericano	Variedade tradicional
81	Roxão MG	Roxo	Mesoamericano	Variedade tradicional
82	Cachinho - MG	Mulatinho	Mesoamericano	Variedade tradicional
83	Mulatinho Vagem Lisa - MG	Mulatinho	Mesoamericano	Variedade tradicional
84	Mulatinho 3 - BA	Mulatinho	Mesoamericano	Variedade tradicional
85	Bico Ouro Mulatinho- MG	Mulatinho	Mesoamericano	Variedade tradicional
86	Bico Ouro Redondo -MG	Mulatinho	Mesoamericano	Variedade tradicional
87	Triete 1 - MG	Roxo	Mesoamericano	Variedade tradicional
88	Banha Carioca - MG	Carioca	Mesoamericano	Variedade tradicional

\*Correspondente ao número dos genótipos dos dendrogramas *neighbour-joining* (Figura 5) e da do *STRUCTURE* (Figura 7) no CAPÍTULO 2.

**Material Suplementar 3:** Perfil de amplificação simultânea observado para os locos BES-SSRs baseados em repetições do tipo di-, tetra-, penta-, hexanucleotídeos e compostos. Enquanto nos painéis (A) e (B) derivados de SSRs com repetições  $\geq$  tetranucleotídeos observa-se um padrão de amplificação alélico específico e de fácil chamada, no painel (C), formado apenas por dinucleotídeos, observa-se a presença de “*stutter*” indicado pelas setas.



## **CAPÍTULO 2**

### **ANÁLISE DE DIVERSIDADE GENÉTICA E ESTRUTURAÇÃO POPULACIONAL EM GERMOPLASMA DE FEIJOEIRO COMUM UTILIZANDO SSRs e SNPs**

## RESUMO

O objetivo desse estudo foi avaliar e comparar o poder de informação genética de um grupo de 58 marcadores SSRs (24 SSRs-di e 34 BES-SSRs) e 345 SNPs para aplicações no melhoramento genético de *Phaseolus vulgaris*. Foram avaliados 88 genótipos representativos de 55 acessos melhorados e 33 variedades tradicionais, incluindo oito combinações biparentais composta por cruzamentos inter- e intra- *pool* gênico. Os SSRs-di apresentaram maior média de alelos por loco gênico (9,916) e revelaram maior diversidade genética média (72,1%), sendo o grupo de marcadores com o maior poder de discriminação genética entre indivíduos. Entre os 58 SSRs, 14 SSRs com uma elevada média de alelos privativos (14,93/loco) e diversidade genética (84%) possibilitaram a discriminação individual dos 88 genótipos avaliados. Os SNPs polimórficos entre cruzamentos biparentais, intra- (17,7%) e inter-*pool* gênico (78,2%), foram avaliados quanto à aplicação prática no mapeamento genético. As análises de agrupamento e estruturação apresentaram resultados similares entre os grupos de marcadores, discriminando os genótipos nos *pools* gênicos Mesoamericano e Andino, com a maior diferenciação genética ( $K = 2$ ,  $F_{ST} = 0,895$ ) apresentada pelos SNPs. Os SSRs-di possibilitaram discriminar dentro do *pool* gênico Mesoamericano o germoplasma melhorado e tradicional ( $K = 3$ ). O desequilíbrio de ligação testado para todos os marcadores foi elevado, sendo maior para os SNPs (84,92%), reduzindo significativamente quando avaliado separadamente para os genótipos Andinos e Mesoamericanos. A distribuição dos SSRs e SNPs foi ampla e aleatória em todo o genoma de *P. vulgaris*. Em termos de resultados práticos para o programa de melhoramento, neste estudo foram derivados painéis de genotipagem operacionais baseado em SSRs e SNPs com alto e eficiente poder de discriminação de acessos do germoplasma por origem. Adicionalmente, um conjunto de SNPs polimórficos entre diversas populações biparentais representa uma aplicação adicional dessa tecnologia para geração de mapas genéticos de alta densidade compreendendo uma proporção substancial de marcadores compartilhados entre cruzamentos. Esse estudo contribui para a integração crescente da genômica nos programas de melhoramento, aliando metodologias de genotipagem acessíveis e a baixos custos que permitem amostrar de modo eficiente e/ou amplo o genoma de feijoeiro comum.

## INTRODUÇÃO

Em relação às leguminosas, a espécie *Phaseolus vulgaris* (feijoeiro comum) é a mais cultivada dentro do gênero *Phaseolus* e utilizada para o consumo direto humano em todos os continentes. Destaca-se como uma cultura de grande valor nutricional por apresentar componentes e características que tornam seu consumo vantajoso e um alimento acessível às populações de baixa renda (Broughton et al. 2003). O continente Americano responde por 43% do consumo mundial, seguido da Ásia (34,5%), África (18,5%), Europa (3,7%) e Oceania (0,1%). Os países em desenvolvimento são responsáveis por 86,7% do consumo mundial, sendo o Brasil considerado o maior produtor com produção oscilando entre 3,2 a 3,3 milhões de toneladas na última década e o consumo estimado em 18,4 Kg/habitante/ano em 2013. As importações líquidas brasileiras, que eram em torno de 100 mil toneladas/ano até 2006/2007, aumentaram para 200 mil toneladas/ano (2012/2013), apesar do enorme potencial para expansão da produção nacional (Silva and Wander 2013). Por meio dos programas de melhoramento genético várias cultivares de feijoeiro comum tem sido desenvolvidas, ampliando o mercado e as formas de uso pelo produtor. Processos que envolvem o uso de ferramentas biotecnológicas podem contribuir para reduzir o tempo na obtenção de genótipos superiores, assim como auxiliar no desenvolvimento de materiais mais competitivos em termos de qualidade e custos de produção.

O melhoramento genético tem buscado uma integração crescente com a área de genética molecular, tendo em vista os avanços e benefícios proporcionados ao conhecimento científico em diferentes etapas dos processos de caracterização, desenvolvimento e seleção de genótipos. Em etapas iniciais dos programas de pré-melhoramento, o emprego de ferramentas moleculares tem auxiliado na quantificação e determinação da diversidade genética. Isto possibilita traçar estratégias de conservação e uso eficiente desses recursos genéticos com ênfase na escolha de acessos geneticamente divergentes e que possuam, ao mesmo tempo, boas características agronômicas, diminuindo os efeitos do estreitamento da base genética dos programas (Buso et al. 2006). Em programas de melhoramento genético atividades de retrocruzamento assistido por marcadores moleculares têm se mostrado atrativas e eficazes reduzindo o número de gerações, facilitando e acelerando a introgressão de

regiões genômicas de interesse (Oliveira et al. 2006) e a piramidação de alelos de resistência a doenças (Alzate-Marin et al. 2005). Na etapa final dos programas de melhoramento, atividades de quantificação da variabilidade genética de linhagens de VCU, determinação da identidade genética para fins de registro e proteção de cultivares, além dos testes de certificação de lotes de sementes comerciais têm contribuído para o aprimoramento e valorização destes programas (Vianello-Brondani et al. 2013a; Chediak et al. 2007). A incorporação da informação genômica no melhoramento de plantas cultivadas tende a ser ampliada devido aos avanços nas técnicas de DNA e ao desenvolvimento permanente de métodos moleculares mais acessíveis e eficazes.

O feijoeiro comum tem a vantagem de apresentar um genoma pequeno dentre as leguminosas (~ 450-650 Mb), com  $2n = 22$  cromossomos, próximo ao tamanho do genoma do feijão-caupi (*Vigna unguiculata*) e representando menos da metade do tamanho do genoma da soja (*Glycine max*), um sétimo do genoma da lentilha (*Lens culinaris*) e da ervilha (*Pisum sativum*), sendo 2.600 vezes menor que o genoma da fava (*Vicia faba*) (Benko-Iseppon 2001). Segundo Melotto et al. (2005), o genoma do feijoeiro comum apresenta ainda um baixo nível de duplicação e poucas famílias gênicas. Ao longo dos últimos 20 anos, avanços significativos foram alcançados rumo ao conhecimento do genoma do feijoeiro comum. Primeiramente, com o desenvolvimento de um mapa genético baseado em RFLP e RAPD (Freyre et al. 1998), e seguindo um progresso contínuo de integração de um número crescente e diverso de marcadores moleculares utilizando diferentes populações biparentais (Yu et al. 2000; Grisi et al. 2007; Blair et al. 2013). Mais recentemente, um mapa genético integrando mais de 1.000 marcadores foi gerado, proporcionando uma análise ampla de todo o genoma e com elevado potencial para estudos de sintenia e mapeamento de QTLs (*Quantitative Trait Loci*) (Galeano et al. 2011; 2012). A obtenção de marcadores robustos, com elevado poder de discriminação e que apresentem um satisfatório padrão de amplificação em *background* genético diverso, é um processo contínuo e dependente de uma adequada seleção dos marcadores que irão compor os sistemas de genotipagem.

Os marcadores microssatélites (*Simple Sequence Repeats* – SSRs) estão amplamente disponíveis para genotipagem molecular em feijoeiro comum (Gaitán-Solís et al. 2002; Grisi et al. 2007; Garcia et al. 2011). Diversos estudos relatam o desenvolvimento de SSRs a partir de bibliotecas genômica enriquecidas, com ênfase em regiões ricas em dinucleotídeos (Buso et al. 2006). Os dinucleotídeos representam a classe mais abundante de repetições com elevado potencial de explorarem o polimorfismo molecular e dos quais atualmente existe uma bateria de marcadores reunindo características desejadas para fins operacionais. Os recentes avanços de sequenciamento genômico total ou parcial, bem como de bancos de sequências expressas, tem permitindo a identificação de um número crescente de regiões SSRs (Hanai et al. 2007). Conforme demonstrado no Capítulo 1, há uma real possibilidade de identificar SSRs baseados em repetições do tipo tetra-, penta- e hexanucleotídeos polimórficos e com características genéticas adequadas para fins de pesquisa e análise genômica do feijoeiro comum. O uso de marcadores SSRs tem sido crescente com o desenvolvimento de sistemas de amplificação simultânea de vários locos aliado ao sistema de detecção semi-automatizada (Vianello-Brondani et al. 2013b), ampliando o potencial de uso dos mesmos através de sistemas mais robustos de genotipagem.

Os SNPs (*Single Nucleotide Polymorphisms*) ou polimorfismos de nucleotídeos únicos são as mais frequentes formas de variação genéticas de um organismo, ocorrendo em abundância ao longo de todo o genoma (Ching et al. 2002). Os SNPs apresenta natureza polimórfica do tipo bialélica, podendo ocorrer em menor frequência polimorfismos do tipo tri- ou tetralélica (Brown 2002). Estimativas da densidade de SNPs em plantas variam em função do hábito reprodutivo da espécie e da região de ocorrência ao longo do genoma (Kaczorowski et al. 2008), tendo sido estimado para soja na frequência de um a cada 2.038 pb em sequência codante e um a cada 191 pb em região não codante (Van et al. 2005). O número mínimo de SNPs necessário para determinar com sucesso os grupos haplotípicos mais comuns em soja, apesar desse variar em função do grupo populacional e do tamanho do genoma, varia de 9.600 a 13.600 SNPs em material melhorado e tradicional, respectivamente (Hyten et al. 2007). A alta frequência de SNPs ao longo do genoma abre a possibilidade de desenvolvimento de mapas genéticos densos e estudos de

associação mais robustos para a identificação de genes de interesse e abordagens genômicas mais precisas (Galeano et al. 2012). Adicionalmente, em função de uma taxa de mutação reduzida quando comparados com os marcadores SSRs, além da disponibilidade de sistemas de detecção automatizados (Thomson et al. 2012; Yu et al. 2011), os SNPs vêm sendo cada vez mais utilizados como marcadores moleculares de grande aplicação em genética vegetal para leguminosas (Loridon et al. 2012; Saxena et al. 2012; Chang et al. 2013; Song et al. 2013). Um número crescente de SNPs tem sido relatado para o feijoeiro comum (Hyten et al. 2010; Cortés et al. 2011; Souza et al. 2012; Blair et al. 2013) ampliando as possibilidades de uso imediato dessa ferramenta pelos programas de melhoramento, bem como para avançar no conhecimento da genética de caracteres de importância agronômica.

Enquanto as ferramentas moleculares baseadas em marcadores SSRs estão bem estabelecidas para uso amplo e com abordagens genéticas diversas no melhoramento genético do feijoeiro comum, o emprego rotineiro dos marcadores SNPs ainda permanece por ser demonstrado. O objetivo desse estudo foi avaliar comparativamente o poder de informação genética dos marcadores SSRs baseados em vários tipos de repetição com detecção semi-automatizada e marcadores SNPs avaliados por meio da tecnologia *Golden Gate* para as aplicações mais comuns no melhoramento genético do feijoeiro comum.

## MATERIAL E MÉTODOS

### Material Vegetal e Extração de DNA

Um total de 88 genótipos de feijoeiro comum, compreendendo 55 linhagens comerciais (material melhorado) lançadas por instituições de pesquisa do Brasil e do exterior e 33 variedades tradicionais (*landraces*) foram caracterizados a nível molecular utilizando marcadores SSRs e SNPs. O total de genótipos de feijoeiros avaliados representam 12 tipos de grãos diferentes (Branco, Carioca, *Cranberry*, *Dark Red Kidney*, Jalo, Mulatinho, Pinto, Preto, Rajado, Vermelho, Rosinha, Roxo e outros), que pertencem a dois *pools* gênicos distintos, Mesoamericano (67 genótipos) e Andino (21 genótipos) (Material Suplementar 2 – Capítulo 1). O DNA genômico foi isolado a partir de amostras de tecido foliar conforme descrito por Grattapaglia and Sederoff (1994), com adaptações. Em seguida, realizou-se a quantificação do DNA no espectrofotômetro *Qubit* (*Invitrogen*, Carlsbad, CA, USA), seguido pelo ajuste da concentração a 5 e 50ng/μL para análise com SSRs e SNPs, respectivamente.

### Genotipagem de SSRs

As genotipagens foram realizadas utilizando 58 marcadores SSRs subdivididos em dois conjuntos: 24 dinucleotídeos (SSRs-di) (Material Suplementar 1) e 34 baseados em motivos tri- (3) , tetra- (9), penta- (8), hexanucleotídeos (7) e SSRs compostos (7) (Material Suplementar 2). O conjunto de 34 marcadores SSRs foi desenvolvido por bioinformática a partir de *BAC-end sequences* (BESs) (Capítulo 1) e foram denominados de BES-SSRs. Os *primers* SSRs foram marcados com fluorescência e analisados em sistemas de coamplificação *multiplex* composto por três a seis locos baseadas no *Standard Dye Set DS-30* para os 24 SSRs-di e DS-33 para os 34 BES-SSRs (*Applied Biosystems*, Foster City, CA, USA). A amplificação dos SSRs foi realizada utilizando produto comercial *Qiagen Multiplex PCR 2X* (*Qiagen*, Hilden, NRW, Germany) em termociclador *GeneAmp Thermal Cycler 9700* (*Applied Biosystems*). A amplificação compreendeu uma etapa inicial de desnaturação a 95°C por 15 min; seguido de 30 ciclos com etapas de desnaturação (94°C por 30 s), anelamento (variando de 56 a 60°C por 90 s) e

extensão (72°C por 1,5 min); e por fim, uma etapa de extensão final de 72°C por 10 min. Os produtos amplificados foram submetidos à eletroforese em analisador automático de DNA ABI3500 (*Applied Biosystems*) e os dados genotipados utilizando os programas *DataCollection 2.0* e *GeneMapper 4.1* (*Applied Biosystems*).

### **Identificação dos SNPs para tecnologia *VeraCode*<sup>TM</sup>**

A partir de um conjunto de 3.487 SNPs brutos, derivados do contraste das sequências de BAT 93 e Jalo EEP 558 do estudo de Hyten et al. (2010), uma filtragem inicial foi realizada eliminando os SNPs redundantes. Posteriormente, os SNPs foram alinhados contra os genomas de referência de *G. max* (Schmutz et al. 2010) obtido do site do Phytozome disponível em [<http://www.phytozome.net/soybean.php>] e *P. vulgaris* obtido via CYTED e acessado em [<http://www.cyteted.org/>] para a identificação de *repeats*. A segunda etapa consistiu na seleção dos SNPs mais adequados para o desenho do ensaio GGGT - *Golden Gate Genotyping Technology* (GGGT assay design) de acordo com o programa *Polybayes* e a probabilidade *SNP vicinity score* disponibilizado pelo alinhamento. Após a finalização dessas etapas, uma série de parâmetros foi definido para o desenho dos oligonucleotídeos flanqueando os SNPs, começando pela identificação de, no mínimo, 60 pb *upstream* e *downstream* do SNP identificado. Isto possibilitou o desenho de três oligonucleotídeos, sendo dois alelo-específico (ASO) para cada variação do SNP e um loco-específico (LSO) ligando-se à região 3' do fragmento de DNA contendo o SNP alvo. Adicionalmente, conforme detalhadamente descrito em Grattapaglia et al. (2011), foram estimados *in silico* a menor frequência alélica (MAF - *Minor Allele Frequency*) para cada SNP, a redundância do SNP em relação aos genomas de referência e a presença de SNPs adicionais na sequência flanqueando o SNP-alvo. Por fim, cinco etapas de filtragem *in silico* foram implementadas (Grattapaglia et al. 2011) com o aumento gradual das restrições na análise, sendo que o conjunto de SNPs selecionados em cada filtragem (300-600 SNPs) foi sendo submetidos ao *Assay Design Tool* (ADT) disponibilizado pela *Illumina* para receber um *score* de validação equivalente a um preditor de probabilidade de sucesso de genotipagem. Os 384 SNPs com maior *score* foram indicados para síntese do *Oligo Pool Assay* (OPA

VC0013574) específico para uso na plataforma *BeadXpress* (*Illumina*, San Diego, CA, USA). A listagem das sequências de um dos *primers* ASO de SNPs e o posicionamento (BLASTN:  $E\text{-value} \leq 1,0E-10$ ) dos mesmos no genoma de referência de feijoeiro comum, variedade Andina (G19833) obtida pelos institutos DOE-JGI e ARRA [<http://www.phytozome.net/commonbean.php>], encontram-se disponível em Material Suplementar 3.

### **Genotipagem de SNPs**

A genotipagem para os 384 SNPs foi realizada via plataforma *Illumina BeadXpress* (Kim and Misra 2007), baseando-se na tecnologia *VeraCode*<sup>TM</sup> (*Illumina*, 2012) e utilizando protocolo recomendando pela *Illumina*. Os procedimentos foram conduzidos no Laboratório de Biotecnologia da Embrapa Arroz e Feijão (Santo Antônio de Goiás, GO, Brasil). A genotipagem dos SNPs foi realizada por meio do programa *GenomeStudio* versão 1.8.4 (*Illumina*) com *Gen Call Threshold* de 0,25; valores de *Call Rate* variando de 0,6463 a 1,00 para chamada dos SNPs e *GenTrain*  $\geq 0,2535$  para o agrupamento dos SNPs. A clusterização (agrupamento) para a chamada dos alelos referente a cada SNP foi feita, *a priori*, de modo automatizado com base na intensidade dos sinais emitidos a partir dos fluoróforos *Cy3* e *Cy5*, os quais foram agrupados em três classes de genótipos representativas dos grupos de homocigotos para os alelos AA ou BB e heterocigoto AB. Para a análise dos dados, os agrupamentos foram individualmente e manualmente ajustados determinando os melhores *clusters* com base no perfil de BAT 93 e Jalo EEP 558.

### **Análise de Diversidade Genética**

A diversidade genética da amostra total foi analisada através das estimativas da frequência alélica, número de alelos por loco, heterocigosidade esperada ( $H_E$ ) e observada ( $H_O$ ), com auxílio do programa *PowerMarker* v2.25 (Liu and Muse, 2005). A identificação dos alelos privados, a probabilidade de identidade (PI) e probabilidade de exclusão (PE) genética para cada loco e multilocos foram obtidas através do programa *GenAlex* v6.5 (Peakall and Smouse 2012). A detecção de genótipos idênticos, através dos SSRs, foi analisada pelo *software Identity* v4 (Sefc et al. 1997). A matriz da distância genética de Rogers modificada (MRD - *Modified Roger's Distance*) (Wright

1978), par a par foi determinada para cada conjunto de marcador. O índice de fixação ( $F_{IS}$ ) na amostra total foi estimado de acordo com a estatística de Wright com intervalo de confiança de 95% para 10.000 *bootstraps* usando o *PowerMarker* v2.25 (Liu and Muse 2005). As árvores filogenéticas foram construídas, a partir da matriz de MRD, pelo algoritmo *Neighbour-joining* não ponderado (Saitou and Nei 1987) implementado pelo programa *DARwin* v5.0.158 (Perrier and Jacquemoud-Collet 2006).

### **Análise de Estruturação Populacional**

A estrutura genética dos genótipos de feijoeiro comum foi analisada através da *Principal Coordinates Analysis* (PCoA). A PCoA foi calculada baseando-se na MRD pelo programa *GenAlex* v6.5 (Peakall and Smouse 2012). A diferenciação genética ( $F_{ST}$ ) na amostra total foi estimada de acordo com a estatística  $F$  de Wright com intervalo de confiança de 95% para 10.000 *bootstraps* usando o *PowerMarker* v2.25 (Liu and Muse 2005). O programa STRUCTURE v2.2.4 (Pritchard et al. 2000) foi utilizado para inferir a estruturação dos genótipos estudados, através de agrupamento baseado no modelo Bayesiano. A estrutura dos indivíduos foi classificada em " $K$  clusters, de acordo com suas similaridades genéticas. O modelo *admixture* foi aplicado, com frequências alélicas correlacionadas e nenhuma informação anterior sobre a população foi utilizada. O número de clusters ( $K$ ) testado variou de 1 a 8, com 10 interações cada. O número de passos do comprimento de *burn-in* e o de repetições MCMC (Markov *chain* Monte Carlo) de 500.000 foram definidos para cada execução. Para determinar o número de grupos genéticos ( $K$  mais provável) utilizou-se o critério proposto por Evanno et al. (2005), através do programa *Structure Harvester* v0.6.93 (Earl and vonHoldt 2012). O software CLUMPP 1.1.2 (Jakobsson and Rosenberg, 2007) foi utilizado para encontrar o consenso entre as 10 interações do  $K$  mais provável e a saída foi utilizada diretamente como entrada no programa de visualização DISTRUCT 1.1 (Rosenberg 2004). A *Analysis of Molecular Variance* (AMOVA) foi realizada para testar a estrutura da diversidade genética dos genótipos, baseado nas diferentes classes de marcadores, através do programa *Arlequin* v3.5 (Excoffier and Lischer 2010), com 10.000 permutações.

## **Análise de Desequilíbrio de Ligação**

A análise de desequilíbrio de ligação (LD) foi obtida através do programa *Tassel* v.2.1 (Bradbury et al. 2007), com 1.000 permutações. Os genótipos heterozigotos foram considerados como dados faltantes, posteriormente os dados dos marcadores moleculares foram transformados em dados haplotípicos. Para avaliar o desequilíbrio de ligação, o coeficiente padronizado de desequilíbrio  $r^2$  foi calculado para o total de marcadores SSRs e SNPs. Os locos foram considerados em LD significativo, se o  $p < 0,001$  e  $r^2 > 0,10$ . Estes parâmetros de LD foram calculados para o total da amostra e, separadamente, para os *pools* gênicos Andinos e Mesoamericanos. Em uma segunda etapa na análise de LD, os locos de SNPs com MAF  $< 0,10$  foram descartados da análise.

## **Alinhamento dos SSRs e SNPs no genoma de feijoeiro comum**

As regiões genômicas adjacentes às sequências repetitivas dos marcadores SSRs e aos SNPs foram alinhadas contra o genoma de *P. vulgaris*, variedade Andina (G19833), [<https://www.phytozome.com/commonbean.php>] utilizando o BLASTN com *E-value*  $\leq 1,0E-25$ . O comprimento mínimo de alinhamento foi de 70 pb para os SSRs e 120 pb para os SNPs. Um único *hit* foi utilizado para desenvolver o mapa no programa *Circos* para visualização do alinhamento das regiões flanqueadoras (Krzywinski et al. 2009).

## RESULTADOS

Os resultados deste trabalho estão estruturados com base na origem dos genótipos de feijoeiro comum, *pools* gênicos Andino e Mesoamericano, subdivididos em germoplasma melhorado e variedades tradicionais (VTs). Os 88 genótipos de feijoeiro foram caracterizados utilizando três conjuntos diferentes de marcadores moleculares, sendo 24 SSRs-di, 34 BES-SSRs (3 tri-, 9 tetra-, 8 penta-, 7 hexanucleotídeos e 7 SSRs compostos) e 384 SNPs. Em algumas análises os grupos de marcadores SSRs-di e BES-SSRs foram avaliados conjuntamente, representando toda a classe de SSRs. A estatística genética descritiva para o conjunto total de marcadores está apresentada na Tabela 1, considerando os valores médios para o grupo total de genótipos de feijoeiro comum. As estimativas derivadas das análises considerando os agrupamentos Andino e Mesoamericano, bem como suas subdivisões por tipo de germoplasma estão descritas no Material Suplementar 4. A nomenclatura “BARC-PV-000XXXX” dos SNPs derivados do trabalho de Hyten et al. (2010) foi reduzida para “PV-XXXX” neste estudo.

### Diversidade Genética

Os resultados mostraram que os SSRs-di apresentaram estimativas superiores em relação aos BES-SSRs para os parâmetros número médio de alelos ( $nA$ ) e diversidade genética média ( $H_E$ ) conforme detalhadamente apresentado na Tabela 1. Considerando os 24 SSRs-di, em média, foram identificados 9,916 alelos por loco gênico com  $H_E$  de 0,721, sendo que para os parâmetros  $H_O$  e  $F_{IS}$  as estimativas foram de 0,034 e 0,953, respectivamente, confirmando a tendência de homozigose para os locos analisados em feijoeiro comum. Considerando os 34 BES-SSRs o  $nA$  estimado foi 1,7X inferior quando comparado com os SSRs-di, com média de 5,588 alelos, e um  $H_E$  médio de 0,487, enquanto os índices de  $H_O$  (0,004) e  $F_{IS}$  (0,991) foram próximos aos estimados para os SSRs-di. Entretanto, dentro das duas categorias de SSRs, existem locos que se destacam por apresentarem estimativas elevadas, como o dinucleotídeo PV163 com 23 alelos ( $H_E$  de 0,889) e o BES-SSR PvComp10 com 19 alelos ( $H_E$  de 0,930). Com relação ao número de alelos privados, em 24 SSRs-di foram identificados 44 exclusivos do *pool* gênico Andino e 118 do Mesoamericano, sendo detectados 18 e 19 alelos privados nos locos PV163 e

BM154, respectivamente. Considerando a diferenciação dos genótipos por tipo de germoplasma, foi identificada, para 22 SSRs-di, uma estimativa de número de alelos privados duas vezes superiores para o germoplasma melhorado (82) em relação às VTs (37), com destaque dos locos PV163 e BM154 com 14 e 15 alelos privados, respectivamente. Considerando os BES-SSRs, foram identificados alelos privados em 31 locos, mantendo a mesma proporção observada para os SSRs-di, com um número menor de alelos exclusivos dentro do *pool* gênico Andino (28) e maior para o Mesoamericano (97), com destaque para os locos PvComp4 e PvComp10 com 15 alelos privados cada. Entre os grupos de germoplasma, os materiais melhorados (73) também apresentaram um maior número de alelos privados em relação às VTs (20), com ênfase para os locos PvHexa15, PvComp10 e PvComp4 com oito, nove e 10 alelos privados, respectivamente.

Em relação ao total de 384 SNPs avaliados neste estudo, 358 (93,23%) apresentaram um padrão adequado de detecção na plataforma *BeadXpress*, dos quais 345 (96,37%) foram polimórficos e 13 (3,63%) monomórficos para os 88 genótipos de feijoeiro comum. Dos 26 excluídos das análises, sete SNPs apresentaram falha de amplificação em pelo menos um alelo e 19 não apresentaram clusters bem definidos. Conforme o esperado para essa classe de marcador, no material melhorado todos os locos caracterizaram-se como bialélicos, enquanto para as VTs observou-se uma média de 1,97; dos quais 10 SNPs foram monomórficos. A análise dos genótipos agrupados pelo *pool* gênico revelou 124 (35,94%) SNPs polimórficos dentro do grupo Andino e um total de 332 (96,23%) no Mesoamericano. Com a maior frequência alélica estimada em 0,721, a  $H_E$  média calculada para os SNPs (0,387) foi da ordem de 1,8X menor em relação aos SSRs-di (0,721) e próxima aos BES-SSRs (0,487), variando de 0,024 para o loco PV-4825 a 0,506 no loco PV-5330. Quanto aos índices  $H_O$  e  $F_{IS}$ , os valores também foram baixos e altos com estimativas de 0,022 e 0,943, respectivamente (Tabela 1). Somente nove alelos privados em nove SNPs individuais foram identificados e exclusivos do germoplasma melhorado (PV-5528, PV-4479, PV-4077, PV-3048, PV-3383, PV-4636, PV-4934, PV-4825 e PV-3706). Considerando o agrupamento por origem do germoplasma, dos 345 SNPs polimórficos um conjunto de 234

diferenciaram com base no padrão alélico os genótipos Andinos (13 alelos exclusivos) dos de origem Mesoamericana (221 exclusivos).

Todos os dois sistemas de genotipagem possibilitaram derivar estimativas satisfatórias de diversidade genética, porém os três grupos de marcadores avaliados concentraram locos em intervalos de distribuição dos valores de  $H_E$  distintos (Figura 1). Para os SSRs-di um total de 87,5% dos locos concentraram estimativas de  $H_E$  variando de 0,6 a 1,0. Enquanto para os BES-SSRs apenas 50% dos locos apresentaram valores compreendidos no intervalo de 0,4 a 0,7, sendo que os demais se distribuíram em valores menores do que 0,4 e maiores que 0,7. Já para os SNPs, 85,5% dos valores de  $H_E$  ocorreram no curto intervalo de 0,3 a 0,6, sendo que o maior valor de diversidade genética foi de 0,506.

As estimativas de distância genética (MRD) média entre indivíduos para os SSRs-di, BES-SSRs e SNPs foram de 0,826; 0,640 e 0,531, respectivamente. Já a MRD entre grupos foi maior entre Andino e Mesoamericano, do que para os grupos de germoplasmas melhorado e VTs, para todos os conjuntos de marcadores. Os SNPs se destacaram dos demais quanto à estimativa de MRD entre grupos Andino e Mesoamericano (0,846) (Tabela 1). As comparações da distância genética entre cada par de acessos foram maiores ou iguais a 0,5 para 99,27% dos SSRs-di, 82,5% dos BES-SSRs e 40,6% dos SNPs (Figura 2).

O coeficiente de correlação de Pearson ( $r$ ) estimado entre as matrizes de distância genética geradas, a partir dos diferentes grupos de marcadores utilizados, apresentou índices variados. Os valores mais elevados foram obtidos entre os marcadores SSRs-di e SNPs com  $r$  de 0,833 (IC<sub>95%</sub> 0,81 a 0,85;  $P < 0,0001$ ) (Figura 3-A) e, seguindo a mesma tendência, comparável ao valor estimado com todos os SSRs (SSRs-di e BES-SSRs) *versus* SNPs com  $r$  de 0,826 (IC<sub>95%</sub> 0,81 a 0,85;  $P < 0,0001$ ) (Figura 3-C), os quais possibilitaram visualizar claramente a formação de dois grupos caracterizados pela menor e maior divergência dos genótipos aos pares. Verificou-se que os microssatélites em geral apresentaram distâncias genéticas de intermediárias a altas e os SNPs apresentaram MRD baixas e altas, poucos pares de genótipos apresentaram distâncias genéticas intermediárias (Figura 2). Um valor menor

de  $r$  foi observado na correlação entre os BES-SSRs e SNPs ( $r = 0,68$ ;  $IC_{95\%}$  0,65 a 0,71;  $P < 0,0001$ ) (Figura 3-B), porém mantendo a mesma visualização da divergência entre grupos (Andino e Mesoamericano) quando se compara SSRs com SNPs. Já a comparação entre os dois grupos de SSRs utilizados, ou seja, SSRs-di *versus* BES-SSRs, o  $r$  foi estimado em 0,663 ( $IC_{95\%}$  0,63 a 0,7;  $P < 0,0001$ ) e não observou-se a formação desses dois grupos distintos bem definidos, ressaltando a ampla distribuição dos valores de MRD, representando os valores intermediários dos microssatélites de forma geral (Figura 3-D). Quando a comparação por Pearson, par a par, foi realizada separadamente entre os genótipos Andinos não verificou a formação de grupos distintos entre as MRDs entre os marcadores. Já para os genótipos Mesoamericanos o resultado foi similar ao encontrado para a análise geral, mostrando que quando se compara SNPs com qualquer grupo de SSRs há a formação de dois grupos. Constatou-se que o primeiro grupo é formado por todas as correlações dos genótipos Mesoamericanos e o segundo grupo formado pelas correlações entre o BRSMG Talismã e os demais do grupo (Material Suplementar 5).

A Probabilidade de Identidade (PI) calculada com base nos 24 SSRs-di apresentou índices que variaram de 0,013 (BM154) a 0,572 (PV251) para os locos com maior e menor poder de discriminação, respectivamente, e a estimativa combinada foi da ordem de  $3,37E-26$ . Considerando os 34 BES-SSRs, com estimativas de locos individuais variando de 0,013 (PvComp10) a 0,755 (PvPenta5), o valor combinado foi ligeiramente maior quando comparado com os SSRs-di e da ordem de  $3,61E-20$ , mesmo utilizando um conjunto adicional de 10 SSRs. Para os 345 SNPs avaliados a PI combinada foi estimada em  $4,67E-119$ , com índices de loco individual variando de 0,375 (PV-3045, PV-3353, PV-5330, PV-4473, PV-4220, PV-4881, PV-2954, PV-4496) a 0,954 (PV-4825) para os locos com maior e menor poder de discriminação, respectivamente (Tabela 1). Em função do menor poder informativo por loco gênico individual um maior número de marcadores SNPs (bialélicos) são necessários para alcançar os mesmos índices quando estimados com os SSRs (multialélicos). Dessa forma, estimativas de PI calculada com base em seis marcadores SSRs-di seriam obtidas a partir da análise com nove marcadores

BES-SSRs e 23 marcadores SNPs. De maneira geral são necessários três vezes mais SNPs para fornecer a mesma PI derivada dos SSRs (Figura 4).

As estimativas da Probabilidade de Exclusão (PE) indicam que cada grupo de marcador, avaliados de modo isolado, apresenta um elevado poder de exclusão de paternidade. Os valores de PE detectados foram de 0,999 para SSRs e um para SNPs, revelando que estes marcadores são suficientes para a discriminação genética dos indivíduos avaliados. As estimativas de PE para todos os marcadores SSRs não alcançam exatos 100%, devido ao baixo número de locos avaliados (máximo 58), mas atingem valores próximos da totalidade. Já para os SNPs são necessário 210 locos para que o PE seja igual a 100%, não sendo necessário avaliar todos os 345 SNPs deste estudo.

Considerando todos os 58 microssatélites avaliados (24 SSRs-di e 34 BES-SSRs), a PI combinada para os locos que apresentavam 10 ou mais alelos privativos e  $H_E$  acima de 0,7 foi estimada em  $1,08E-20$ . Um total de 14 locos SSRs foi selecionado para esta estimativa de PI combinada, sendo dez locos representados pelos SSRs dinucleotídeos (BM154, PV272, PV87, PV163, PV25, BM143, BM187, BM210, BM114 e PV35) e quatro locos por BES-SSRs (PvComp10, PvComp4, PvTri8 e PvTetra25). Estes 14 SSRs com média de alelos privativos de 14,93 por loco e  $H_E$  média de 0,84 possibilitam a discriminação de todos os 88 genótipos de feijoeiro comum avaliados.

A aplicação prática dos SNPs no mapeamento genético para o programa de melhoramento do feijoeiro comum foi realizada através da validação dos 345 SNPs avaliados, por meio da observação da taxa de polimorfismo em diferentes parentais utilizados em populações de mapeamento. Nesta análise, entre os 88 genótipos avaliados, 11 foram utilizados em oito cruzamentos biparentais, os quais são importantes nos estudos de mapeamento genético para tolerância a estresses bióticos e abióticos. Desses cruzamentos, os cinco envolvendo parentais pertencentes ao mesmo *pool* gênico (*intra-pool* gênico Mesoamericano) (BAT 477 x Pérola, *Red Mexican* x Pérola, CNFP10132 x BRS Cometa, BRS Aurora x PI181996, US Pinto 111 x PI181996) tiveram em média 61 (17,7%) SNPs polimórficos. Em relação aos três cruzamentos *inter-pool* gênico (Andino x Mesoamericano) (AND 277 x Rudá, Ouro Branco x CNFP10132, AND 277 x SEA 5) a média de SNPs polimórficos foi de 270

(78,2%). Observou-se que dos 345 SNPs, o número de marcadores polimórficos compartilhados inter-*pool* gênico foi de 212 (61,4%), intra- *pool* gênico foi de apenas um SNP (0,29%), sendo o mesmo SNP (PV-4505) comum entre todos os oito cruzamentos biparentais (intra- e inter- *pool* gênico) (Tabela 2).

As análises das árvores filogenéticas baseadas nas distâncias genéticas de Rogers modificada por Wright (MRD) e geradas separadamente para os três grupos de marcadores, assim como para os SSRs avaliados em conjunto, ilustram os genótipos reunidos com base na origem do germoplasma (Figura 5). Os ramos superiores e inferiores dos agrupamentos são representados pelos genótipos de origem Andina e Mesoamericana, respectivamente, mesclando em ambos os acessos melhorados e tradicionais. O acesso BRSMG Talismã *a priori* classificado como de origem Mesoamericana foi posicionado *a posteriori* junto ao grupo de acessos Andinos.

### **Estruturação Populacional**

A estrutura genética dos 88 genótipos de feijoeiro comum avaliada através da análise PCoA revelou resultados similares para os grupos de marcadores analisados separadamente (SSRs-di, BES-SSRs e SNPs), bem como para os SSRs analisados em conjunto, revelando uma nítida estruturação dos genótipos com base nas origens Mesoamericana e Andina. O primeiro eixo, que explica a maior parcela de variação dos dados, separou os acessos de forma consistente por origem para cada classe de marcador avaliada (Figura 6), com valores extremos de 46,1 e 89,6% da variação molecular para os marcadores SSRs-di e SNPs, respectivamente (Figuras 6-A e 6-D). O segundo eixo explicou de 3,3 a 18,5% da variação molecular para SNPs e SSRs-di, respectivamente, sendo que somente nesse último foi evidenciado uma separação entre os grupos de germoplasma melhorado e tradicional dentro dos acessos de origem Mesoamericana.

A diferenciação genética ( $F_{ST}$ ) entre o germoplasma melhorado e tradicional foi significativamente diferente de zero ( $P < 0,05$ ) para todos os grupos de marcadores avaliados, assim como entre os genótipos de origem Andina e Mesoamericana. Com exceção do valor de  $F_{ST}$  para os SSRs-di, que

indicou uma moderada diferenciação genética, os BES-SSRs e SNPs indicaram uma pequena diferenciação genética entre os subgrupos de melhorado e VTs de feijoeiro comum. Por outro lado, uma elevada diferenciação genética entre os grupos Andino e Mesoamericano foi detectada para todos os tipos de marcadores avaliados nesse estudo (Tabela 1).

A análise de estruturação foi realizada para cada grupo de marcador separadamente e a representação gráfica das estatísticas de  $\Delta K$  está apresentada no Material Suplementar 6. Através dos resultados de  $\Delta K$ , observou-se que o  $K = 2$  foi o *cluster* mais provável para todos os grupos de marcadores individualmente analisados (SSRs-di, BES-SSRs, SSRs total e SNPs). A Figura 7 ilustra a divisão clara ( $K = 2$ ) entre os genótipos de feijoeiro comum de origem Andina (verde) e Mesoamericana (vermelho), com exceção do acesso BRSMG Talismã (62, número do genótipo) posicionado no agrupamento Mesoamericano, porém marcado em verde e identificado como pertencente ao agrupamento Andino através da análise molecular. Adicionalmente, a análise com os SSRs-di revelou ainda a formação de três clusters ( $K = 3$ ) mostrando a subdivisão dentro do *pool* gênico Mesoamericano do germoplasma melhorado (roxo) e tradicional (vermelho), corroborando os resultados da PCoA. Através da análise de estruturação, observou-se a formação de dois grupos, com o menor formado por 21 genótipos de origem Andina e um de origem mista (62), e outro constituído por 66 de origem Mesoamericana. Nas análises subsequentes o genótipo BRSMG Talismã (62) foi posicionado no grupo Andino.

A AMOVA realizada com os diferentes grupos de marcadores moleculares, para inferir sobre a estruturação do germoplasma com base na origem Andina e Mesoamericana, revelou diferentes estimativas sobre a porcentagem da variação genética. Os marcadores SSRs-di explicaram 36,5% da diferenciação, seguido pelos BES-SSRs com 49,6% e uma estimativa combinada para os SSRs de 43,3%. Os marcadores SNPs, destacaram-se dos SSRs, explicando 87,7% da variação genética entre os grupos de germoplasma por origem (Tabela 3).

## Desequilíbrio de Ligação

No que diz respeito ao pressuposto de segregação independente dos locos foi efetuada a análise de desequilíbrio de ligação ( $p < 0,001$  e  $r^2 > 0,10$ ) para os três grupos de marcadores separadamente, conforme apresentado na Tabela 3. Para os 24 SSRs-di foram avaliadas 276 combinações par a par entre marcadores, sendo que 17,75% foram consideradas em LD com base na análise dos 88 genótipos de feijoeiro comum. A estimativa de locos em LD reduziu significativamente quando a análise foi realizada considerando os acessos agrupado por *pool* gênico Andino (0,72%) e Mesoamericano (1,45%). Entre os 34 BES-SSRs, totalizando 561 combinações par a par de locos avaliadas, 21,4% foram caracterizadas em LD considerando todos os acessos, 0,53% dentro do grupo Andino e 4,28% no Mesoamericano. Por fim, a partir dos 345 SNPs (59.340 pares de locos testados) analisados com todos os acessos foram observados 84,92% dos locos em LD, reduzindo-se para 0,25% dentro do grupo Andino e 4,64% no Mesoamericano. Estabelecendo um limite de  $MAF > 0,10$  para os SNPs, somente oito locos foram excluídos da análise de LD considerando o grupo total de acessos, enquanto dentro do grupo Andino e Mesoamericano foram removidos 290 e 204 SNPs, respectivamente. Com relação ao LD estimado somente para os SNPs restantes, ou seja, com  $MAF > 0,10$ , esse aumentou para 88,45% na amostragem total, 3,23% no grupo Andino e 15,35% no Mesoamericano.

Para o estabelecimento de uma análise rotineira e operacional junto ao banco de germoplasma de feijoeiro comum, com o objetivo de discriminar de modo eficiente acessos de origem Andina e Mesoamericana, as análises de LD auxiliaram na seleção de um grupo de SSRs úteis para esse fim. Dos 40 SSRs em LD (15 SSRs-di e 25 BES-SSRs) foram selecionados 16 (oito SSRs-di e oito BES-SSRs) que apresentaram frequência de um alelo para um dos *pool*s gênicos acima de 85%. Os oito locos SSRs-di estão inseridos e destacados no Material Suplementar 1 e os outros oito locos BES-SSRs no Material Suplementar 2. Estes 16 locos foram selecionados para integrar o painel de genotipagem discriminatória por origem do feijoeiro comum. As análises de estruturação realizadas com esses oito SSRs-di e oito BES-SSRs (dados não apresentados) revelaram um padrão nítido de agrupamento dos acessos de acordo com seus centros de origem.

## Distribuição dos SSRs e SNPs no genoma do feijoeiro comum

O alinhamento dos 58 SSRs e 345 SNPs contra o genoma de *P. vulgaris*, variedade andina (Pv), está representado na Figura 8. Dos 24 SSRs-di, 87,5% (21) alinharam-se contra sequências genômicas de Pv (exceto BM201, BM210 e PV163), distribuídos em oito dos 11 cromossomos de *P. vulgaris* e com maior representatividade dos SSRs (cinco no total) no cromossomo Pv2. Em relação aos 34 BES-SSRs, 94,1% (32) alinharam ao genoma de Pv (exceto PvComp4 e PvPenta8), apresentando uma ampla distribuição genômica com dois ou mais BES-SSRs representados nos 11 cromossomos, e o máximo de cinco marcadores no Pv10. Diferentemente ao observado para os SSRs, 100% dos SNPs alinharam no genoma da linhagem Andina demonstrando uma ampla cobertura genômica, com representatividade por cromossomo variando de 23 (Pv5 e Pv9) a 39 SNPs (Pv1 e Pv3). Um grupo de 16 SNPs (PV-3023, PV-4006, PV-3881, PV-4134, PV-3353, PV-3269, PV-4479, PV-3840, PV-3106, PV-4020, PV-3094, PV-4250, PV-4276, PV-4576, PV-4026 e PV-3566) alinharam com mais de 99% de identidade ao genoma de Pv (*E-value* = 0) e comprimento do alinhamento maior do que 350 pb.

## DISCUSSÃO

### Diversidade Genética

A caracterização da diversidade genética, no programa de melhoramento do feijoeiro comum, é de extrema importância para diagnosticar precocemente o estreitamento da base genética e colocar em prática estratégias alternativas para ampliação desta (Frascaroli et al. 2013). O uso de diferentes conjuntos de marcadores para explorar a variabilidade do material melhorado é um recurso fundamental para aumentar a base genética e possibilita examinar as relações genéticas em diferentes níveis de resolução. Apesar do grande número de marcadores SSRs disponibilizados e caracterizados para *P. vulgaris* nos últimos anos, poucos são os estudos utilizando os SNPs para explorar a diversidade e estruturação genética de feijoeiro comum (Cortés et al. 2011; Blair et al. 2013). Até o momento, a aplicação dos SNPs mais concreta tem sido no desenvolvimento de mapas genéticos com ampla cobertura para o feijoeiro comum (Galeano et al. 2012). O presente trabalho é pioneiro em explorar a diversidade e estruturação genética de *P. vulgaris* a partir da comparação de marcadores SSRs e SNPs, utilizando um mesmo grupo de genótipos. A eficiência dos SSRs em discriminar os genótipos com base no *pool* gênico de origem tem sido demonstrada por diversos autores (Kwak and Gepts 2009; Burle et al. 2010) e os resultados estão de acordo com os apresentados no presente estudo. Para os marcadores SNPs a menor diversidade genética por loco gênico (Cortés et al. 2011; Blair et al. 2013), quando comparado com os SSRs (Buso et al. 2006; Vianello-Brondani et al. 2013b), é compensada pela a combinação de um elevado número de marcadores, indicando o uso crescente dessa metodologia para os mais variados estudos genéticos. Embora, os SNPs se caracterizem por serem a forma mais abundante de polimorfismo ao longo do genoma, apenas recentemente o sua utilização em larga escala tem sido demonstrada no melhoramento de plantas (Mammadov et al. 2012). Apesar de um grande número de marcadores SNPs estar se tornando disponível para o feijoeiro comum, os SSRs continuam sendo a classe de marcadores mais informativa em estudos de diversidade genética (Yang et al. 2011).

Os SSRs destacam-se, de forma geral, por detectarem múltiplas formas alélicas por loco gênico (SSRs-di  $nA = 9,916$  e BES-SSRs  $nA = 5,588$ ); enquanto que para os SNPs predominou o bialelismo e a maior frequência alélica (0,721), resultando em menor poder informativo por loco gênico ( $H_E = 0,387$ ). Enquanto para os 345 SNPs foram identificados 690 alelos, para os 58 SSRs esse número foi de 428, ou seja, um número seis vezes maior de SNPs possui apenas 1,6X mais alelos que os SSRs. Dentre os SSRs, os derivados de repetições do tipo dinucleotídeo destacaram pelo elevado multialelismo por loco gênico e estimativas de diversidade genética ( $H_E = 0,721$ ) superiores, possibilitando a discriminação genética dos 88 indivíduos avaliados. Tal característica pode ser atribuída ao fato da unidade de repetição dos dinucleotídeos se formarem de forma mais rápida, em função da maior taxa de deslizamento da DNA polimerase durante a replicação, do que para as repetições maiores (Schlötterer and Tautz 1992). Embora os 58 SSRs tenham se destacado quanto ao número de alelos identificados, através dos 345 SNPs foram identificados um maior número de alelos privados entre os genótipos do *pool* gênico Andino e Mesoamericano, favorecendo a maior diferenciação dos acessos por origem. O maior número de alelos privados entre o germoplasma melhorado e tradicional foi encontrado para os SSRs-di, detectando a diferenciação entre esses grupos. Do ponto de vista do melhoramento genético, essas informações dão subsídios para seleção de genótipos que tenham alelos mais divergentes com potencial de incorporação em blocos de cruzamentos, contribuindo efetivamente para a ampliação da base genética no *pool* gênico do germoplasma melhorado (Vianello-Brondani et al. 2013a). As reduzidas estimativas de heteroziguidade observadas corroboram com os elevados índices de endogamia, encontrados através dos marcadores SSRs e SNPs. Isto reforça a tendência de homozigose para todos os locos analisados, conforme o esperado para espécies autógamas, como o feijoeiro comum, cuja taxa de fecundação cruzada foi estimada com valores abaixo de 0,74% (Ferreira et al. 2000).

Em relação às correlações das MRDs obtidas entre os marcadores observou-se que quando os SNPs são confrontados com os SSRs (SSRs-di, BES-SSRs, SSRs-di/BES-SSRs) há formação de dois grupos, que refletem pares de genótipos com a menor e maior divergência. A menor (0,2-0,4) retrata

a divergência dentro de grupos (*intra-pool* gênico). Já a maior (0,8-1,0) representa a divergência entre os *pools* gênicos Andino e Mesoamericano (*inter-pool* gênico) na análise geral (Figura 3-A, B e C) e entre o BRSMG Talismã e os demais genótipos do grupo Mesoamericano (Material Suplementar 5). A cultivar BRSMG Talismã, desenvolvida pelos programas de melhoramento da Embrapa Arroz e Feijão, da Universidade Federal de Lavras (UFLA), da Universidade Federal de Viçosa (UFV) e da Empresa de Pesquisa Agropecuária de Minas Gerais (EPAMIG), tem origem mista entre os *pools* gênicos Andino e Mesoamericano. Este genótipo é oriundo de seleção recorrente envolvendo os seguintes parentais: BAT 477, IAPAR 14, FT 84-29, Jalo EEP 558, A-252, A-77, Ojo de Liebre, ESAL 645, Pintado, Carioca, P-85, P-103, H-4, AN 910522, ESAL 624 e Carioca MG. Como neste estudo foi analisado um indivíduo por genótipo, os resultados mostram um viés para a cultivar que apresentou composição genética típica do germoplasma Andino, sendo considerada como pertencente ao *pool* gênico Mesoamericano por características fenotípicas, com o grão do tipo carioca. Este viés foi importante para destacar que os SNPs são marcadores eficientes em relatar a divergência entre os grupos Andino e Mesoamericano, o que não foi indicado pelas correlações da MRD entre os marcadores SSRs (Figura 3-D). Esta discrepância de posicionamento do mesmo genótipo em agrupamentos diferentes já foi previamente relatado por Briñez et al. (2012), para o genótipo BAT 477 em feijoeiro comum.

A diversidade genética estimada no *pool* gênico Mesoamericano foi ligeiramente maior do que dentro do *pool* gênico Andino calculada através dos marcadores SSRs e SNPs, em contraste com os resultados obtidos para SSRs e SNPs nos estudos de Blair et al. (2006) e Cortés et al. (2011), respectivamente, e semelhante ao obtido para SNPs no estudo de Blair et al. (2013). A maior estimativa de diversidade genética encontrada para os genótipos Mesoamericanos pode ser resultante de um efeito amostral, uma vez que foram genotipados 3,2X mais indivíduos deste *pool* gênico quando comparado ao Andino, conforme também relatado por Blair et al. (2013). As estimativas de distância genética, bem como os índices de diferenciação genética, entre grupos de acessos que compõem o germoplasma melhorado e VTs foram inferiores aos estimados *inter-pool* gênico. Os índices de MRD e  $F_{ST}$

indicam que ainda há informações adicionais a serem exploradas para ampliar a base genética das cultivares brasileiras, predominantemente de origem Mesoamericana, a partir da introdução da diversidade genética contida nas VTs, minimizando a necessidade de retornar à introdução do material silvestre, como vem ocorrendo para a soja (Li et al. 2010).

Os grupos de marcadores avaliados neste estudo possuem ainda um elevado poder de discriminação individual demonstrado pela alta PE e pelos baixos valores de PI combinada, em concordância com estudos prévios para feijoeiro comum (Vianello-Brondani et al. 2013a, b). O valor da PI combinada para os SNPs foi inferior em função do grande número de marcadores utilizados, proporcionando um poder discriminatório superior quando comparado aos SSRs. Em adição, os marcadores SSRs do tipo tetra-, penta-, hexanucleotídeos (BES-SSRs), que sabidamente são menos informativos com base em dados da literatura (Blair et al. 2011a) foram polimórficos neste estudo, além de fornecerem uma PE combinado ( $> 0,999$ ) semelhante aos resultados dos SSRs-di e SNPs. De modo geral, um número apenas 3X maior de marcadores SNPs foi necessário para alcançar os mesmos índices quando estimados com os SSRs em feijoeiro comum. Em milho, predominante alógamo, Van Inghelandt et al. (2010) propõem de 7 a 11X mais SNPs quando comparado aos SSRs. Estes resultados corroboram a expectativa de uma maior extensão de LD em feijoeiro comum, com hábito preferencialmente autógamo. Esses dados sugerem ainda que será possível adotar uma abordagem de varredura ampla do genoma (“*whole genome scan*”) em *P. vulgaris*, dado ao menor número de SNPs necessários para capturar os blocos haplotípicos representativos do genoma.

Os marcadores SNPs utilizados nesse estudo derivados do polimorfismo entre um genótipo Andino e outro Mesoamericano forneceram um número substancialmente maior de marcadores polimórficos para fins de análise de mapeamento inter-*pool* gênico (78,2%) e intra-*pool* gênico (17,7%) em *P. vulgaris* do que tecnologias anteriores. Estimativas próximas foram obtidas por Cortés et al. (2011), que detectaram com base na análise de 94 marcadores SNPs, um nível de polimorfismo inter-*pool* gênico e intra-*pool* gênico de 71,7% e 15,4%, respectivamente. Embora microssatélites codominantes sejam significativamente mais informativos em termos de loco individual, estimativas

inferiores de polimorfismo para cruzamento envolvendo linhagens pertencentes ao mesmo *pool* gênico, com índices ao redor de 11%, têm sido relatadas em estudos de QTLs (Garcia et al. 2012; Blair et al. 2011b). Considerando cruzamentos com feijoeiro comum envolvendo genitores de centros de origens distintos, os SSRs também foram menos polimórficos quando comparados com os SNPs atuais (78,2%), com estimativas de polimorfismo variando de 7,7% (Galeano et al. 2011) a 24% (Garcia et al. 2011). Do ponto de vista prático, a estratégia de obtenção de um grande número de SNPs a partir de genótipos que compõem a base genética de interesse, ou seja, para a qual se quer maximizar o potencial de polimorfismo do marcador, mostrou ser adequada. Desse modo é possível focar no desenvolvimento de marcadores SNPs que sejam geneticamente informativos para as populações desenvolvidas no âmbito dos programas de melhoramento, e que geralmente pertencem a um mesmo *pool* gênico (Beebe et al. 2013), porém são contrastantes para o fenótipo de interesse. A análise de polimorfismo dos marcadores SNPs identificou, através das oito composições biparentais, um total de 212 marcadores comuns nos cruzamentos inter-*pool* gênico e apenas um SNP intra- *pool* gênico. A disponibilidade desses SNPs em maior escala vai facilitar a geração e integração de mapas genéticos com maior resolução necessários para identificação QTLs, permitindo uma integração e comparação de informações genômicas em estudos futuros.

### **Estruturação Genética**

Os resultados de estruturação genética para os marcadores SSRs e SNPs estão em concordância com outros trabalhos de feijoeiro comum, revelando dois subgrupos pelo *pool* gênico de origem (Kwak and Gepts 2009; Burle et al. 2010; Cortés et al. 2011; Blair et al. 2013). Nesse estudo, o conjunto de SNPs caracterizados mostrou ser o mais adequado para diferenciar de modo eficiente o germoplasma Andino do Mesoamericano ( $F_{ST} = 0,895$ ), quando comparado com os marcadores SSRs (di- SSRs  $F_{ST} = 0,724$  e BES-SSRs  $F_{ST} = 0,733$ ). O maior potencial dos marcadores SNPs em diferenciar os indivíduos com base na origem decorreu da escolha dos genótipos-alvo específicos, as cultivares BAT 93 e Jalo EEP 558 (Hyten et al. 2010), criando um viés de amostragem que resultou no aumento da diferenciação entre os

*pools* gênicos. Essa tendência em agrupar os acessos amostrados, seguindo o mesmo padrão de diferenciação dos genótipos alvos utilizados para derivar os SNPs, tem sido observados em outros estudos, como relatado para a cevada (Moragues et al. 2010), arroz (Thomson et al. 2012) e feijoeiro comum (Blair et al. 2013). Segundo Nagasaki et al. (2010), dependendo da finalidade do estudo e do grupo de germoplasma avaliado, diferentes conjuntos de SNPs devem ser utilizados, sendo que o uso de grupos de SNPs comuns torna-se atrativo, pois possibilita o compartilhamento de resultados e consolidação de um banco de dados integrados. Para isso, a estratégia de aumentar a representatividade genômica dos SNPs, através da busca por polimorfismo em fragmentos amostrados a partir de um número maior de raças e tipos de grãos inter- e intra-*pool* gênico, representará a melhor maneira de explorar a variabilidade genética de *P. vulgaris*.

As análises de estruturação populacional compartilharam resultados similares entre SNPs e SSRs, evidenciando a estruturação dos genótipos de feijoeiro comum por origem. A PCoA explicou a maior variação entre os grupos Andino e Mesoamericano através do primeiro eixo para os SSRs e SNPs, assim como no estudo de Cortés et al. (2011) com SNPs. A maior variação do primeiro eixo foi detectada para os SNPs (89,6%), indicando claramente que apesar do menor conteúdo de informação genética por loco gênico, estes marcadores genotipados em grande número possibilitaram uma melhor determinação da divergência genética entre grupos por origem. A análise de estruturação populacional, através do STRUCTURE, indicou que o  $K = 2$  é o *cluster* mais provável, diferenciando os genótipos Andinos dos Mesoamericanos conforme os dados da literatura para feijoeiro comum (Kwak and Gepts 2009; Burle et al. 2010; Vianello-Brondani et al. 2013a,b; Cortés et al. 2011; Blair et al. 2013). Estes mesmos dados foram corroborados pelas informações baseadas nas MRD utilizadas na construção das árvores filogenética de *Neighbour-joining*. A AMOVA referente às estimativas de diferenciação entre os *pools* gênicos revelou uma variação maior para os SNPs (87,7%), certamente decorrente da origem dos SNPs utilizados, conforme previamente discutido, e também pelo maior número de locos utilizados (6X).

Adicionalmente, as análises de agrupamento realizadas com os SSRs-di evidenciaram a subestruturação dos indivíduos pertencentes ao grupo

Mesoamericano em material melhorado e tradicional ( $K = 3$ ). Esses resultados indicam uma relação bastante estreita entre o germoplasma tradicional e melhorado, revelando a ocorrência de cruzamentos entre os mesmos e levando a uma homogeneização genética entre os acessos. A ocorrência de subestruturação em germoplasma de feijoeiro comum, baseado em análise moleculares, tem sido amplamente relatada, possibilitando a diferenciação entre germoplasma melhorado do silvestre, entre as raças e entre os tipos de grãos em níveis maiores ( $K > 3$ ) de estruturação genética (Kwak and Gepts 2009; Burle et al. 2010; Vianello-Brondani et al. 2013a,b; Cortés et al. 2011; Blair et al. 2013). Apesar deste estudo não avaliar genótipos silvestres, ele foi capaz de diferenciar entre o germoplasma melhorado das VTs para os SSRs-di. Essa informação pode ser extremamente importante para a caracterização do germoplasma de *P. vulgaris*, de forma rotineira, pelo tipo de acesso.

O maior poder dos SSRs-di em estruturar o germoplasma Mesoamericano em melhorado e tradicional deve-se ao fato desses marcadores terem sido caracterizados e selecionados a partir de grupos de acessos prioritariamente do *pool* gênico Mesoamericano (Grisi et al. 2007; Garcia et al. 2011), atendendo à demanda dos programas de melhoramento genético brasileiros. A identificação de um grupo maior de marcadores baseados em repetições de trinucleotídeos ou superiores (BES-SSRs), com o potencial discriminatório dos SSRs-di, depende exclusivamente de um esforço adicional na seleção e caracterização de novos marcadores, derivados das BESs, mais informativos geneticamente para serem incorporados aos sistemas de genotipagem. Já os marcadores SNPs, desenvolvidos entre o polimorfismo das cultivares BAT 93 e Jalo EEP 558, foram importantes para distinção dos genótipos Andinos e Mesoamericanos, mas menos úteis para distinguir dentro de cada *pool* gênico, assim como para Cortés et al. (2011).

Considerando todos os genótipos de feijoeiro comum, Andinos e Mesoamericanos ( $K = 2$ ), foi detectado um alto nível de LD ( $p < 0,001$  e  $r^2 > 0,10$ ), com os SSRs e SNPs (17,75 a 84,92%). No entanto, um menor LD foi observado para as subamostras separadas entre os *pools* gênicos Andino (0,25 a 0,72%) e Mesoamericano (1,45 a 4,64%), indicando que a estruturação é a principal causa de LD em feijoeiro comum. Embora ainda não esteja descrito na literatura análise de LD, aliado ao estudo de diversidade e

estruturação genética do feijoeiro comum, baseados em SNPs, resultados correlatos baseados em AFLP (Rossi et al. 2009) e SSRs (Kwak and Gepts, 2009; Burle et al. 2010) encontram-se relatados. O trabalho de Burle et al. (2010) detectou um LD de 80%, entre as combinações aos pares de SSRs, para todos os acessos, e um menor nível entre Andino (8%) e Mesoamericano (23%). Resultado semelhante a esse foi encontrado neste trabalho, com os marcadores SSRs e SNPs, para LD com  $p$ -valor  $< 0,01$  e  $r^2 > 0,10$  (dados não mostrados). Quando o  $p$ -valor mínimo foi reduzido de 0,01 para 0,001 os resultados de LD diminuíram consideravelmente (Tabela 3), por isso os dados aqui apresentados por uma análise mais estrigente foram inferiores aos da literatura para SSRs em feijoeiro comum, exceto para SNPs cujo LD continuou elevado para toda a amostra (84,92%). Mesmo tentando avaliar os efeitos do LD para os 345 SNPs, utilizando a  $MAF > 0,10$  que descartou os locos com dados faltantes e monomórficos para os subgrupos Andino e Mesoamericano os valores de LD aumentaram, o que não era esperado para o estudo, mas é justificado pela redução proporcional dos locos avaliados. Apesar deste trabalho não apresentar análise de mapeamento, estudos entre marcadores próximos no mesmo cromossomo podem ser uma alternativa para averiguar a existência, principalmente, de LD por ligação física, não apenas pelo efeito da estruturação. Isto poderia ser melhor observado para os SNPs, que apresentam um número maior de marcadores por cromossomo, comparado com os marcadores SSRs mais dispersos e em menor número distribuídos no genoma de feijoeiro comum (Figura 8).

Os resultados deste estudo, também, mostraram um nível de LD ligeiramente maior dentro do *pool* gênico Mesoamericano do que no Andino, contrariamente ao observado por Rossi et al. (2009). Este dado pode ser justificado pela maior amostragem dos genótipos Mesoamericanos utilizados e a seleção dos genótipos deste *pool* gênico durante a domesticação no Brasil. Sendo assim, o LD tem sido um método eficaz para identificar episódios de seleção durante a domesticação e na seleção de genes candidatos ou locos de interesse agrônômico através do mapeamento associativo. Sugerindo ainda que estas análises devem ser realizadas em amostras separadas para germoplasma Andino e Mesoamericano diretamente (Kwak and Gepts 2009).

## CONSIDERAÇÕES FINAIS

Os resultados mostraram nitidamente um maior poder dos marcadores SSRs em caracterizar a diversidade alélica entre os genótipos quando comparados com os SNPs. Entretanto, os dois grupos de marcadores foram consistentes quanto à determinação dos parâmetros de estrutura populacional. Os recentes esforços do sequenciamento genômico de *P. vulgaris*, aliados às recentes abordagens de genotipagem por sequenciamento (GBS), que começaram a ser desenvolvidas para plantas em 2010 (Maughan et al 2010; Elshire et al. 2011), apresentam uma excelente perspectiva para derivar milhares de SNPs em feijoeiro comum.

O desenvolvimento e a validação de marcadores SNPs ainda são etapas demoradas e caras. Com base nos sistemas de genotipagem utilizados nesse estudo, a análise com marcadores SNPs em plataforma *BeadXpress (Illumina)*, além de mais rápida, proporcionou próximo a duas ordens de magnitude mais marcadores polimórficos do que um ensaio com marcadores SSRs em plataforma ABI 3500 (*Applied Biosystems*). Em termos de custo de genotipagem (*data point/amostra*), o que se observou com base no uso dessas tecnologias, foi uma estimativa três vezes menor para os SNPs quando comparado aos SSRs. Do ponto de vista de implementação de tecnologia, os marcadores SSRs ainda são mais acessíveis e podem ser genotipados em géis de alta resolução de poliacrilamida, porém demanda mais tempo e mão de obra.

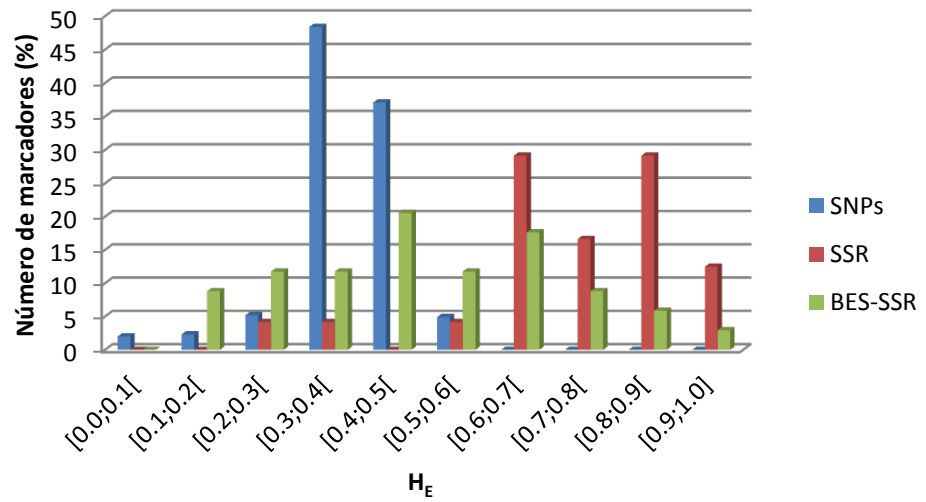
Este estudo revelou com sucesso painéis de genotipagem com características genéticas adequadas para diferenciação de acessos inter-*pool* gênico (SNPs e SSRs) e intra-*pool* gênico (SSRs-di), passíveis de adoção em escala internacional para análise genética do feijoeiro comum. Adicionalmente, a indicação de um conjunto de SNPs polimórficos, em diversas populações biparentais, representa uma aplicação adicional dessa tecnologia para geração de mapas genéticos de alta densidade, compreendendo uma proporção substancial de marcadores compartilhados entre cruzamentos. Os experimentos realizados contribuem para a integração crescente da genômica nos programas de melhoramento genético do feijoeiro comum.

## TABELAS E FIGURAS

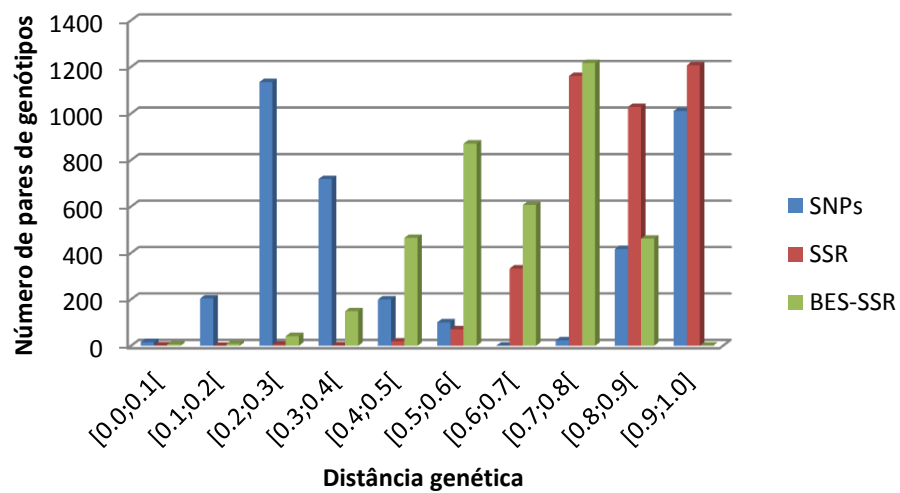
**Tabela 1:** Estatística descritiva dos 24 SSRs-di, 34 BES-SSRs e 345 SNPs caracterizados em 88 genótipos de feijoeiro comum:

	Grupo	Tamanho da amostra	F <sub>máx</sub>	Alelos	nA	Alelos privados	H <sub>E</sub>	H <sub>O</sub>	F <sub>IS</sub>	PI	MRD entre indivíduos	MRD entre grupos	F <sub>ST</sub>
SSRs	Andino	21	0,590	120	5	44	0,522	0,045	0,916	4,93E-17	0,711	0,585	0,724*
	Mesoamericano	67	0,501	194	8,083	118	0,619	0,031	0,951	2,60E-21	0,768		
	Melhorado	55	0,449	201	8,375	82	0,679	0,022	0,968	8,60E-25	0,810	0,287	0,086*
	Tradicional	33	0,448	156	6,5	37	0,647	0,054	0,921	3,80E-22	0,798		
	<b>Total</b>	<b>88</b>	<b>0,405</b>	<b>238</b>	<b>9,916</b>	-	<b>0,721</b>	<b>0,034</b>	<b>0,953</b>	<b>3,37E-26</b>	<b>0,826</b>	-	-
BES-SSRs	Andino	21	0,805	93	2,735	28	0,277	0,0015	0,995	5,87E-11	0,459	0,590	0,733*
	Mesoamericano	67	0,730	162	4,765	97	0,393	0,0051	0,987	1,06E-15	0,582		
	Melhorado	55	0,647	170	5,00	73	0,478	0,003	0,994	7,05E-20	0,643	0,175	0,036*
	Tradicional	33	0,641	117	3,440	20	0,433	0,007	0,986	8,87E-18	0,617		
	<b>Total</b>	<b>88</b>	<b>0,641</b>	<b>190</b>	<b>5,588</b>	-	<b>0,487</b>	<b>0,0043</b>	<b>0,991</b>	<b>3,61E-20</b>	<b>0,640</b>	-	-
SNPs	Andino	21	0,951	469	1,360	13	0,071	0,025	0,665	3,24E-22	0,237	0,846	0,895*
	Mesoamericano	67	0,908	677	1,962	221	0,141	0,022	0,847	1,07E-43	0,327		
	Melhorado	55	0,755	690	2,00	9	0,357	0,023	0,937	2,10E-109	0,507	0,141	0,023*
	Tradicional	33	0,673	680	1,97	0	0,427	0,022	0,950	2,40E-127	0,560		
	<b>Total</b>	<b>88</b>	<b>0,721</b>	<b>690</b>	<b>2</b>	-	<b>0,387</b>	<b>0,022</b>	<b>0,943</b>	<b>4,67E-119</b>	<b>0,531</b>	-	-

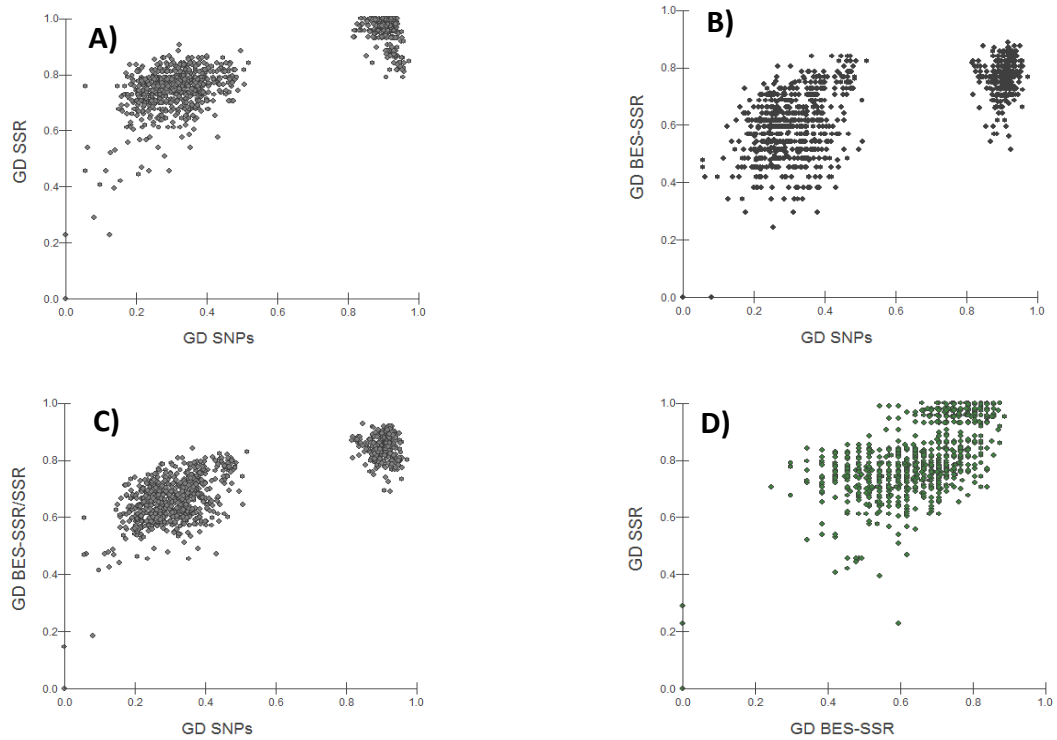
Descritores abreviados: estimativa da maior frequência alélica (F<sub>máx</sub>), número de alelo por loco (nA), diversidade gênica (H<sub>E</sub>), heterozigiosidade observada (H<sub>O</sub>), índice de endogamia (F<sub>IS</sub>), probabilidade de identidade (PI), distância genética de Rogers modificada (MRD) e diferenciação genética (F<sub>ST</sub>).



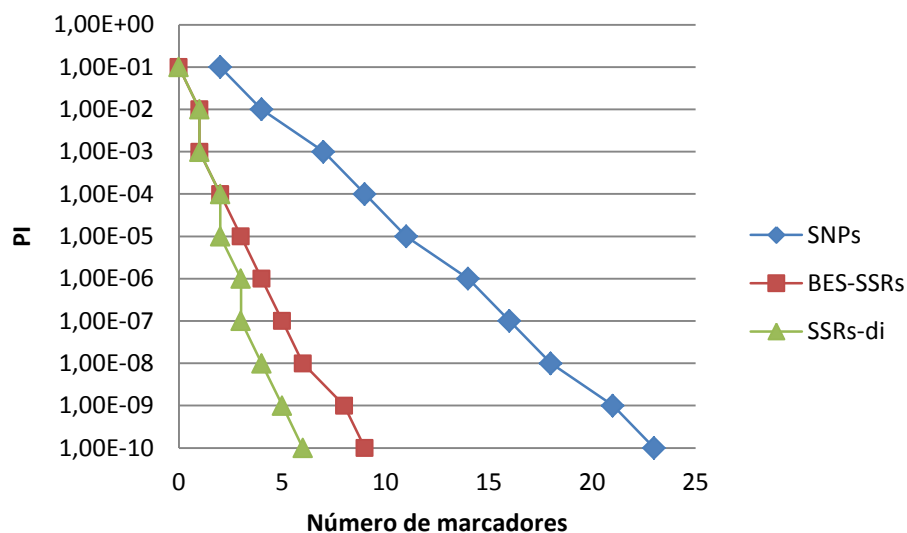
**Figura 1:** Distribuição da diversidade genética ( $H_e$ ) para os 24 SSRs-di, 34 BES-SSRs e 345 SNPs utilizados na genotipagem dos 88 genótipos de feijoeiro comum.



**Figura 2:** Distribuição da distância genética de Rogers modificada (MRD) para os 24 SSRs-di, 34 BES-SSRs e 345 SNPs utilizados na genotipagem dos 88 genótipos de feijoeiro comum.



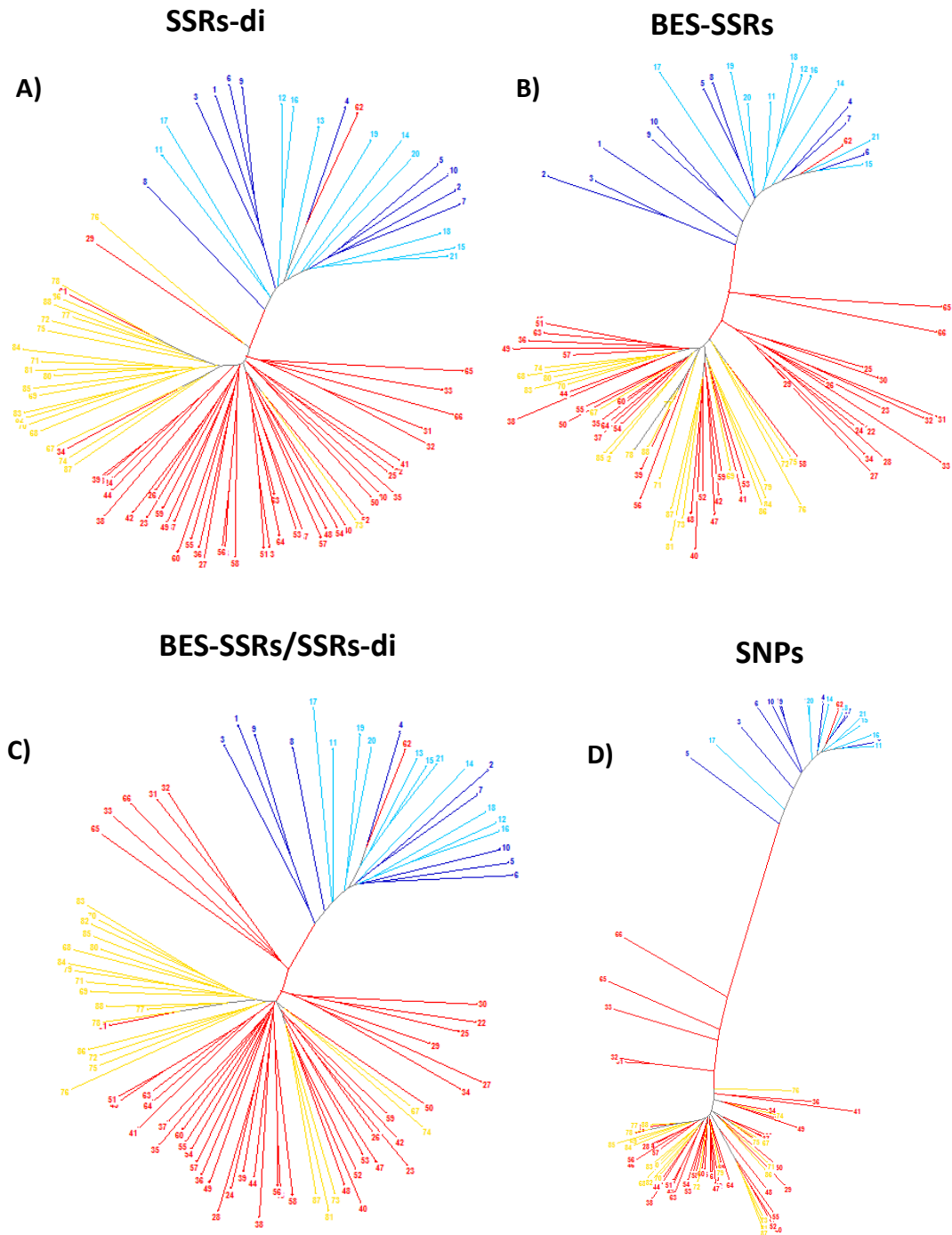
**Figura 3:** Correlação da distância genética, par a par pelo método de Pearson, baseado nas comparações das três abordagens de marcadores moleculares: A- SSRs-di *versus* SNPs, B- BES-SSRs *versus* SNPs, C- BES-SSRs/SSRs-di *versus* SNPs e D- SSRs-di *versus* BES-SSRs.



**Figura 4:** Distribuição dos valores de probabilidade de identidade ( $PI$ ) para os marcadores SSRs-di, BES-SSRs e SNPs; indicando o número de marcadores necessários para atingir um mesmo valor de  $PI$  discriminante entre os 88 genótipos de feijoeiro comum. Os marcadores (SSRs e SNPs) foram sendo incluídos a partir de seus valores individuais de  $PI$  (ordenados do menor para o maior) e, conseqüentemente, maiores valores de  $H_E$ .

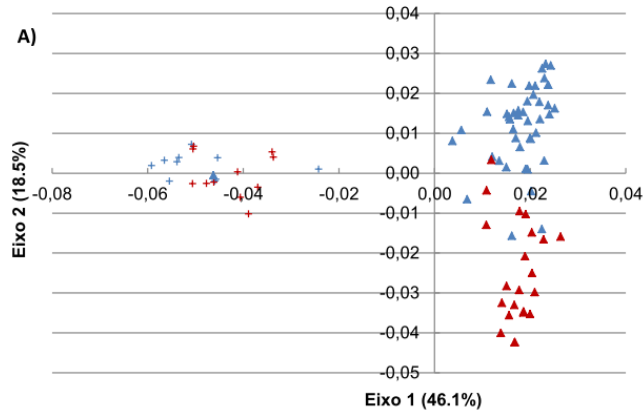
**Tabela 2:** Avaliação do polimorfismo dos cruzamentos biparentais entre os *pools* gênicos Andino (A) e Mesoamericano (M) ou intra-*pool* gênico Mesoamericano (M x M) e análise dos SNPs com polimorfismos comuns intra-, inter-*pool* gênico e entre o total (intra- e inter-*pool* gênico), através dos 345 SNPs de feijoeiro comum:

<b>Combinação Parental</b>		Tipo de cruzamento	<b>Total SNP (345)</b>	
Parental 1	Parental 2		N. Poli.	% Poli.
AND 277	Rudá	A x M	272	78,84
Ouro Branco	CNFP10132	A x M	255	73,91
AND 277	SEA 5	A x M	282	81,74
BAT 477	Pérola	M x M	27	7,82
Pérola	Red Mexican	M x M	58	16,81
CNFP10132	BRS Cometa	M x M	54	15,65
BRS Aurora	PI181996	M x M	105	30,43
PI181996	US Pinto 111	M x M	60	17,39
<b>Polimorfismo comum</b>				
	Inter- <i>pool</i> gênico	3(A x M)	212	61,45
	Intra- <i>pool</i> gênico	5(M x M)	1	0,29
Inter- <i>pool</i> gênico	Intra- <i>pool</i> gênico	3(A x M) + 5(M x M)	1	0,29

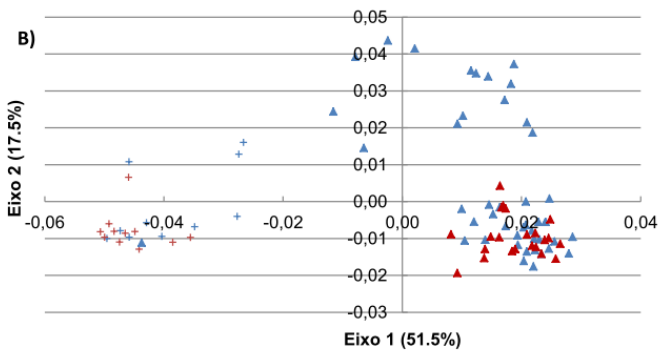


**Figura 5:** Dendrogramas *Neighbour-joining* baseados nas distâncias genéticas de Rogers modificada por Wright representando o agrupamento dos 88 genótipos de feijoeiro para os diferentes grupos de marcadores: **A)** 24 SSRs-di, **B)** 34 BES-SSRs, **C)** 58 microssatélites (34 BES-SSRs + 24 SSRs-di), e **D)** 345 SNPs. Os genótipos Andinos estão representados em azul escuro (material melhorado) e claro (variedades tradicionais) e os Mesoamericanos em vermelho (material melhorado) e amarelo (variedades tradicionais).

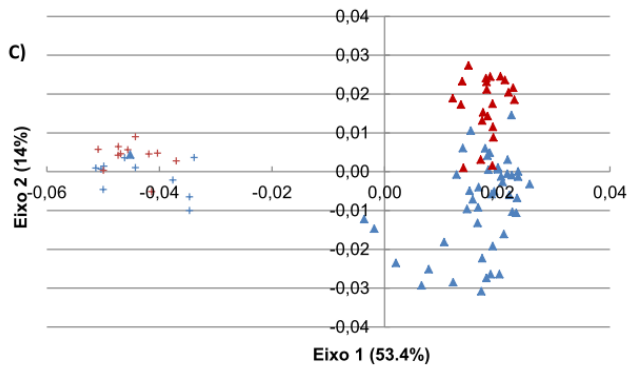
**SSRs-di**



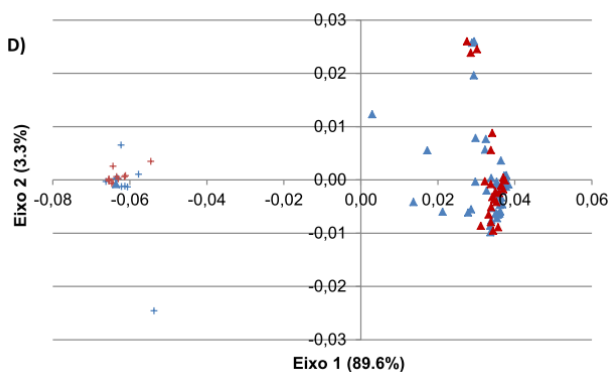
**BES-SSRs**



**BES-SSRs/SSRs-di**



**SNPs**



**Legenda:**

- + And\_C
- + And\_L
- ▲ Mes\_C
- ▲ Mes\_L

**Figura 6:** Análise de coordenadas principais (PCoA) dos 88 genótipos de feijoeiro comum baseado na distância genética de Rogers modificada, calculada para os **A)** 24 SSRs-di, **B)** 34 BES-SSRs, **C)** 58 microssatélites (34 BES-SSRs + 24 SSRs-di), e **D)** 345 SNPs.

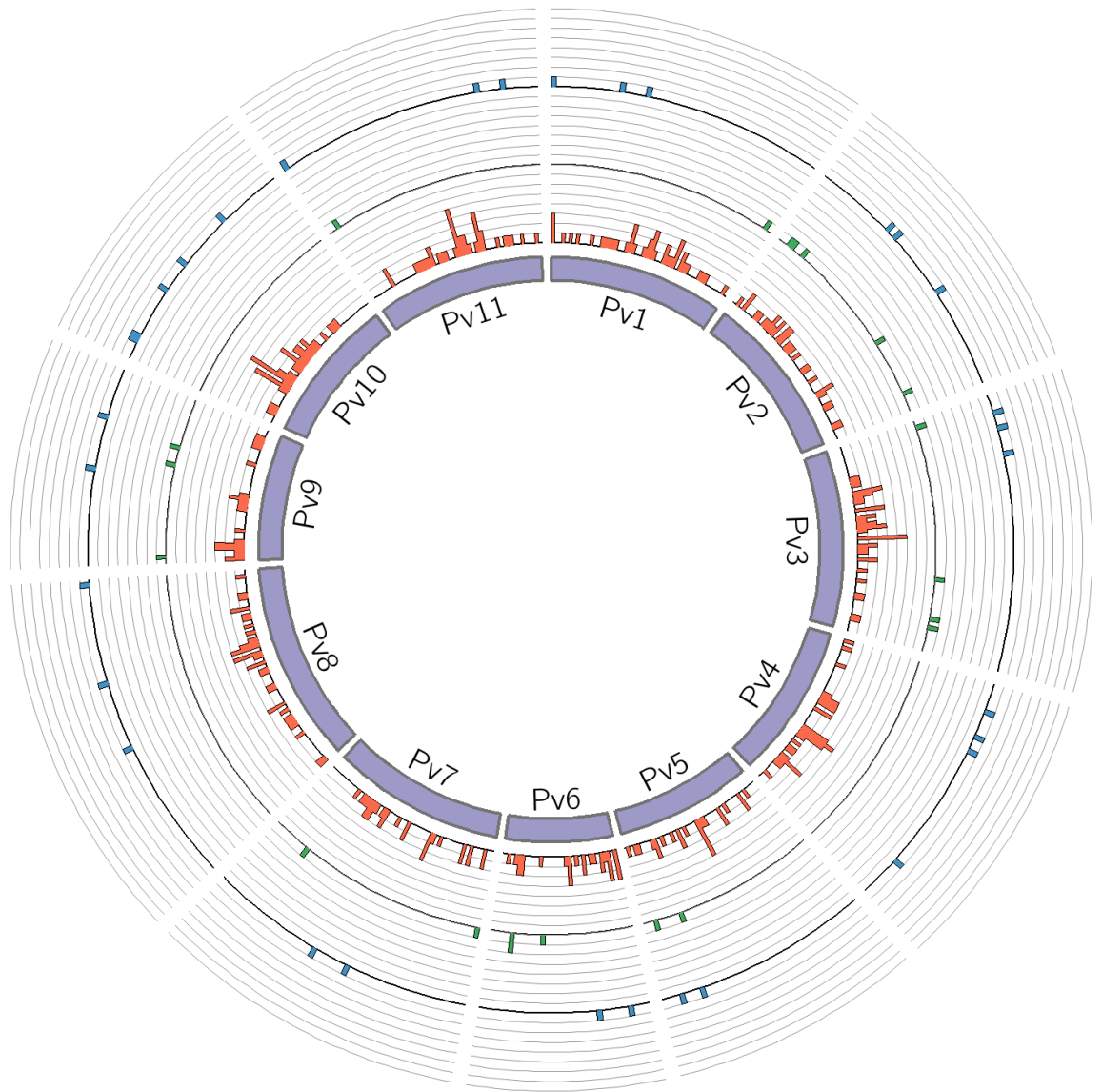


**Figura 7:** Estruturação populacional inferida pela abordagem Bayesiana baseada nos grupos de marcadores SSRs-di, BES-SSRs, SSRs-di/BES-SSRs e SNPs, para  $K = 2$  e 3. Cada indivíduo é representado por uma linha vertical, que é dividida em segmentos coloridos na proporção da divisão prevista para 2-3 subpopulações. As barras em verde indicam o germoplasma de origem Andina, as vermelhas o Mesoamericano e as roxas os acessos melhorados.

**Tabela 3:** Estimativas de desequilíbrio de ligação calculadas aos pares de locos ( $p < 0,001$  e  $r^2 > 0,1$ ) e análise de variância molecular (AMOVA) baseadas em 24 SSRs-di, 34 BES-SSRs e 345 SNPs.

	Pares de locos ligados			Total de pares de locos	AMOVA
	Total (88)*	Andino (22)*	Mesoamericano (66)*		
<b>SSRs-di</b>	49 (17,75%)	2 (0,72%)	4 (1,45%)	276	36,5%
<b>BES-SSRs</b>	120 (21,40%)	3 (0,53%)	24 (4,28%)	561	49,6%
<b>SNPs</b>	50393 (84,92%)	151 (0,25%)	2751 (4,64%)	59340	87,7%

\* Tamanho da amostra.



**Figura 8:** Distribuição das regiões flangeadoras dos SSRs e SNPs através do alinhamento no genoma de *Phaseolus vulgaris*, variedade Andina. A figura inclui quatro círculos concêntricos. O círculo interno representa o genoma de feijoeiro comum (barras em roxo), destacando os cromossomos haploides (Pv1-11). Os histogramas estão representados, do mais interno para o mais externo, o alinhamento de SNPs (vermelho), SSRs-di (verde) e BES-SSRs (azul), respectivamente. Cada barra do histograma equivale a uma região de 1 Mb de sequência.

## REFERÊNCIAS

Alzate-Marin AL, Cervigni GDL, Moreira MA, Barros EG (2005) Seleção assistida por marcadores moleculares visando ao desenvolvimento de plantas resistentes a doenças , com ênfase em feijoeiro e soja. *Fitopatol Bras* 30(4):333–342

Beebe SE, Rao IM, Blair MW, Acosta-Gallegos JA (2013) Phenotyping common beans for adaptation to drought. *Front Physiol* 4:35. doi: 10.3389/fphys.2013.00035

Benko-Iseppon AM (2001). Estudos moleculares no caupi e em espécies relacionadas: avanços e perspectivas. *Embrapa Documentos* 56: 327–332

Blair MW, Cortés AJ, Penmetsa RV, Farmer A, Carrasquilla-Garcia N, Cook DR (2013) A high-throughput SNP marker system for parental polymorphism screening, and diversity analysis in common bean (*Phaseolus vulgaris* L.). *Theor Appl Genet* 126(2):535–548. doi: 10.1007/s00122-012-1999-z

Blair MW, Hurtado N, Chavarro CM, Muñoz-Torres MC, Giraldo MC, Pedraza F, Tomkins J, Wing R (2011a) Gene-based SSR markers for common bean (*Phaseolus vulgaris* L.) derived from root and leaf tissue ESTs: an integration of the BMC series. *BMC Plant Biol* 11:50. doi: 10.1186/1471-2229-11-50

Blair MW, Astudillo C, Rengifo J, Beebe SE, Graham R (2011b) QTL analyses for seed iron and zinc concentrations in an intra-genepool population of Andean common beans (*Phaseolus vulgaris* L.). *Theor Appl Genet* 122(3):511–521. doi: 10.1007/s00122-010-1465-8

Blair MW, Giraldo MC, Buendía HF, Tovar E, Duque MC, Beebe SE (2006) Microsatellite marker diversity in common bean (*Phaseolus vulgaris* L.). *Theor Appl Genet* 113(1):100–109. doi: 10.1007/s00122-006-0276-4

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL : software for association mapping of complex traits in diverse samples. *Bioinformatics* 23(19):2633–2635. doi: 10.1093/bioinformatics/btm308

Briñez B, Blair MW, Kilian A, Carbonell SAM, Chiorato AF, Rubiano LB (2012) A whole genome DArT assay to assess germplasm collection diversity in common beans. *Mol Breeding* 30(1):181–193. doi: 10.1007/s11032-011-9609-3

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

Brown TA (2002) *Genomes*. 2nd Edition. Oxford: Wiley-Liss. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK21128/>

Burle ML, Fonseca JR, Kami JA, Gepts P (2010) Microsatellite diversity and genetic structure among common bean (*Phaseolus vulgaris* L.) landraces in Brazil, a secondary center of diversity. *Theor Appl Genet* 121(5):801–813. doi: 10.1007/s00122-010-1350-5

Buso GSC, Amaral ZPS, Brondani RPV, Ferreira ME (2006) Microsatellite markers for the common bean *Phaseolus vulgaris*. *Mol Ecol Notes* 6:252–254. doi: 10.1111/j.1471-8286.2006.01210.x

Chang S, Hartman GL, Singh RJ, Lambert KN, Hobbs HA, Domier LL (2013) Identification of high-quality single-nucleotide polymorphisms in *Glycine latifolia* using a heterologous reference genome sequence. *Theor Appl Genet* 126(6):1627–1638. doi: 10.1007/s00122-013-2079-8

Chediak GLC, Brondani RPV, Del Peloso MJ, Melo LC, Brondani C (2007) Análise de pureza genética de sementes de feijoeiro comum utilizando marcadores microssatélites em sistema de genotipagem multiplex. Santo Antônio de Goiás, Embrapa Arroz e Feijão. *Boletim de Pesquisa e Desenvolvimento* 28, ISSN 1678-9601; 28, 20 p

Ching A, Caldwell KS, Jung M, Dolan M, Smith OS, Tingey S, Morgante M, Rafalski AJ (2002) SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. *BMC Genet* 3:19

Cortés AJ, Chavarro MC, Blair MW (2011) SNP marker diversity in common bean (*Phaseolus vulgaris* L.). *Theor Appl Genet* 123(5):827–845. doi: 10.1007/s00122-011-1630-8

Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genet Resour* 4(2):359–361. doi: 10.1007/s12686-011-9548-7

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6(5):e19379. doi: 10.1371/journal.pone.0019379

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14(8):2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 10(3):564–567. doi: 10.1111/j.1755-0998.2010.02847.x

Ferreira JJ, Alvarez E, Fueyo MA, Roca A, Giraldez R (2000) Determination of the outcrossing rate of *Phaseolus vulgaris* L. using seed protein markers. *Euphytica* 113(3):257–261. doi: 10.1023/A:1003907130234

Frascaroli E, Schrag TA, Melchinger AE (2013) Genetic diversity analysis of elite European maize (*Zea mays* L.) inbred lines using AFLP,

SSR, and SNP markers reveals ascertainment bias for a subset of SNPs. *Theor Appl Genet* 126(1):133–141. doi: 10.1007/s00122-012-1968-6

Freyre R, Skroch PW, Geffroy V, Adam-Blondon AF, Shirmohamadali A, Johnson WC, Llaca V, Nodari RO, Pereira PA, Tsai SM (1998) Towards an integrated linkage map of common bean. 4. Development of a core linkage map and alignment of RFLP maps. *Theor Appl Genet* 97(5-6): 847–856

Gaitán-Solís E, Duque MC, Edwards KJ, Tohme J (2002). Microsatellite repeats in common bean (*Phaseolus vulgaris*): Isolation, characterization and cross-species amplification in *Phaseolus* ssp. *Crop Sci* 42:1228–1236

Galeano CH, Fernandez AC, Franco-Herrera N, Cichy KA, McClean PE, Vanderleyden J, Blair MW (2011) Saturation of an intra-gene pool linkage map: towards a unified consensus linkage map for fine mapping and synteny analysis in common bean. *PLoS ONE* 6(12):e28135. doi: 10.1371/journal.pone.0028135

Galeano CH, Cortés AJ, Fernández AC, Soler Á, Franco-Herrera N, Makunde G, Vanderleyden J, Blair MW (2012) Gene-based single nucleotide polymorphism markers for genetic and association mapping in common bean. *BMC Genet* 13:48. doi: 10.1186/1471-2156-13-48

Garcia RAV, Rangel PN, Brondani C, Martins WS, Melo LC, Carneiro MS, Borba TC, Brondani RP (2011) The characterization of a new set of EST-derived simple sequence repeat (SSR) markers as a resource for the genetic analysis of *Phaseolus vulgaris*. *BMC Genet* 12:41. doi: 10.1186/1471-2156-12-41

Garcia RAV, Rangel PN, Bassinello PZ, Brondani C, Melo LC, Sibov ST, Vianello-Brondani RP (2012) QTL mapping for the cooking time of common beans. *Euphytica* 186(3):779–792. doi: 10.1007/s10681-011-0587-7

Grattapaglia D, Sederoff R (1994) Genetic linkage maps of *Eucalyptus grandis* and *E. urophylla* using a pseudo-testcross strategy and RAPD markers. *Genetics* 137(4):1121–1137

Grattapaglia D, Silva-Junior OB, Kirst M, de Lima BM, Faria DA, Pappas GJ Jr (2011) High-throughput SNP genotyping in the highly heterozygous genome of *Eucalyptus*: assay success, polymorphism and transferability across species. *BMC Plant Biol* 11:65. doi: 10.1186/1471-2229-11-65

Grisi MC, Blair MW, Gepts P, Brondani C, Pereira PA, Brondani RP (2007) Genetic mapping of a new set of microsatellite markers in a reference common bean (*Phaseolus vulgaris*) population BAT 93 x Jalo EEP 558. *Genet Mol Res* 6(3):691–706

Hanai LR, de Campos T, Camargo LE, et al. (2007) Development, characterization, and comparative analysis of polymorphism at common bean

SSR loci isolated from genic and genomic sources. *Genome* 50(3):266–277. doi: 10.1139/G07-007

Hyten DL, Choi IY, Song Q, Shoemaker RC, Nelson RL, Costa JM, Specht JE, Cregan PB (2007) Highly variable patterns of linkage disequilibrium in multiple soybean populations. *Genetics* 175(4):1937–1944. doi: 10.1534/genetics.106.069740

Hyten DL, Song Q, Fickus EW, Quigley CV, Lim JS, Choi IY, Hwang EY, Pastor-Corrales M, Cregan PB (2010) High-throughput SNP discovery and assay development in common bean. *BMC Genomics* 11:475. doi: 10.1186/1471-2164-11-475

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23(14):1801–1806. doi: 10.1093/bioinformatics/btm233

Kaczorowski KA, Kim KS, Diers BW, Hudson ME (2008) Microarray-based genetic mapping using soybean near-isogenic lines and generation of SNP markers in the Rag1 aphid resistance interval. *Plant Genome* 1(2):89–98

Kim S, Misra A (2007) SNP genotyping: technologies and biomedical applications. *Annu Rev Biomed Eng* 9:289–320. doi: 10.1146/annurev.bioeng.9.060906.152037

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA (2009) Circos: an information aesthetic for comparative genomics. *Genome Res* 19(9):1639–1645. doi: 10.1101/gr.092759.109

Kwak M, Gepts P (2009) Structure of genetic diversity in the two major gene pools of common bean (*Phaseolus vulgaris* L., Fabaceae). *Theor Appl Genet* 118(5): 979–992. doi: 10.1007/s00122-008-0955-4

Li YH, Li W, Zhang C, Yang L, Chang RZ, Gaut BS, Qiu LJ (2010) Genetic diversity in domesticated soybean (*Glycine max*) and its wild progenitor (*Glycine soja*) for simple sequence repeat and single-nucleotide polymorphism loci. *New Phytol* 188(1):242–253. doi: 10.1111/j.1469-8137.2010.03344.x

Liu K, Muse S V (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21(9):2128–2129. doi: 10.1093/bioinformatics/bti282

Loridon K, Burgarella C, Chantret N, Martins F, Gouzy J, Prospéri JM, Ronfort J (2012) Single-nucleotide polymorphism discovery and diversity in the model legume *Medicago truncatula*. *Mol Ecol Resour* 13(1):84–95. doi: 10.1111/1755-0998.12021

Mammadov J, Aggarwal R, Buyyarapu R, Kumpatla S (2012) SNP markers and their impact on plant breeding. *Intl J Plant Genomics* 2012:728398. doi: 10.1155/2012/728398

Maughan PJ, Yourstone SM, Byers RL, Smith SM, Udall JA (2010) Single-nucleotide polymorphism genotyping in mapping populations via genomic reduction and next-generation sequencing: proof of concept. *Plant Genome J* 3(3):166–178. doi: 10.3835/plantgenome2010.07.0016

Melotto M, Monteiro-vitorello CB, Bruschi AG, Camargo LEA (2005) Comparative bioinformatic analysis of genes expressed in common bean (*Phaseolus vulgaris* L.) seedlings. *Genome* 48(3): 562–570. doi: 10.1139/G05-010

Moragues M, Comadran J, Waugh R, Milne I, Flavell AJ, Russell JR (2010) Effects of ascertainment bias and marker number on estimations of barley diversity from high-throughput SNP genotype data. *Theor Appl Genet* 120(8):1525–1534. doi: 10.1007/s00122-010-1273-1

Nagasaki H, Ebana K, Shibaya T, Yonemaru J, Yano M (2010) Core single-nucleotide polymorphisms—a tool for genetic analysis of the Japanese rice population. *Breeding Sci* 60(5):648–655

Oliveira GV, Carneiro PCS, Carneiro JES, Cruz CD (2006). Adaptabilidade e estabilidade de linhagens de feijão comum em Minas Gerais. *Pesquisa Agropecuária Brasileira* 41(2):257-265

Peakall R, Smouse PE (2012) GenAIEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28(19):2537–2539. doi: 10.1093/bioinformatics/bts460

Perrier X, Jacquemoud-Collet JP (2006) DARwin software. Available at. <http://darwin.cirad.fr/>. Accessed on Feb 23, 2013

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155(2):945–959

Rosenberg NA (2004) Distruct: a program for the graphical display of population structure. *Mol Ecol Notes* 4:137–138. doi: 10.1046/j.1471-8286.2003.00566.x

Rossi M, Bitocchi E, Bellucci E, Nanni L, Rau D, Attene G, Papa R (2009) Linkage disequilibrium and population structure in wild and domesticated populations of *Phaseolus vulgaris* L. *Evol Appl* 2(4):504–522. doi: 10.1111/j.1752-4571.2009.00082.x

Saitou N, Nei M (1987) The Neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4(4):406–425

Saxena RK, Penmetsa RV, Upadhyaya HD et al (2012) Large-scale development of cost-effective single-nucleotide polymorphism marker assays for genetic mapping in pigeonpea and comparative mapping in legumes. *DNA Res* 19(6):449–461. doi: 10.1093/dnares/dss025

Schlötterer C, Tautz D (1992) Slippage synthesis of simple DNA. *Nucleic Acids Res* 20(2):211–215

Schmutz J, Cannon SB, Schlueter J et al (2010) Genome sequence of the palaeopolyploid soybean. *Nature* 463(7278):178–183. doi: 10.1038/nature08670

Sefc KM, Steinkellner H, Wagner HW, Glössl J, Regner F (1997) Application of microsatellite markers to parentage studies in grapevine. *Vitis* 36(4):179–183

Silva OF, Wander AE (2013) O feijão-comum no Brasil passado, presente e futuro. Santo Antônio de Goiás, Embrapa Arroz e Feijão. Documentos 287, ISSN 1678-9644; 287, 63 p

Song Q, Hyten DL, Jia G, Quigley CV, Fickus EW, Nelson RL, Cregan PB (2013) Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. *PLoS ONE* 8(1):e54985. doi: 10.1371/journal.pone.0054985

Souza TLPO, Barros EG, Bellato CM, Hwang EY, Cregan PB, Pastor-Corrales MA (2012) Single nucleotide polymorphism discovery in common bean. *Mol Breeding* 30(1):419–428. doi: 10.1007/s11032-011-9632-4

Thomson MJ, Zhao K, Wright M et al (2012) High-throughput single nucleotide polymorphism genotyping for breeding applications in rice using the BeadXpress platform. *Mol Breeding* 29(4):875–886. doi: 10.1007/s11032-011-9663-x

Van Inghelandt D, Melchinger AE, Lebreton C, Stich B (2010) Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. *Theor Appl Genet* 120(7):1289–1299. doi: 10.1007/s00122-009-1256-2

Van K, Hwang EY, Kim MY, Park HJ, Lee SH, Cregan PB (2005) Discovery of SNPs in soybean genotypes frequently used as the parents of mapping populations in the United States and Korea. *J Hered* 96(5): 529–535. doi: 10.1093/jhered/esi069

Vianello-Brondani RP, Cardoso PC, Veiga MM, de Menezes IP, Valdisser PA, Borba TC, Melo LC, Del Peloso MJ, Brondani C (2013a) Molecular characterization of high performance inbred lines of Brazilian common beans. *Genet Mol Res* 12(AOP). [Epub ahead of print]

Vianello-Brondani RP, Cardoso PCB, Menezes IPP, Valdisser PAMR, Borba TCO, Peloso MJD, Brondani C (2013b) Discrimination of common bean commercial cultivars using panels of highly informative microsatellite markers. *Genet Mol Res: Ahead of Print*

Wright S (1978) Evolution and the genetics of populations. Variability within and among natural populations, vol 4. University of Chicago Press, Chicago, IL. 511 p

Yang X, Xu Y, Shah T, Li H, Han Z, Li J, Yan J (2011) Comparison of SSRs and SNPs in assessment of genetic relatedness in maize. *Genetica* 139(8):1045–1054. doi: 10.1007/s10709-011-9606-9

Yu H, Xie W, Wang J, Xing Y, Xu C, Li X, Xiao J, Zhang Q (2011) Gains in QTL detection using an ultra-high density SNP map based on population sequencing relative to traditional RFLP/SSR markers. PLoS ONE 6(3):e17595. doi: 10.1371/journal.pone.0017595

Yu K, Park SJ, Poysa V, Gepts P (2000) Integration of simple sequence repeat (SSR) markers into a molecular linkage map of common bean (*Phaseolus vulgaris* L.). J Hered 91(6): 429–434. doi: 10.1093/jhered/91.6.429

## APÊNDICES

**Material Suplementar 1:** Descrição dos 24 marcadores microssatélites do tipo dinucleotídeos (SSRs-di) utilizados na caracterização dos 88 genótipo de feijoeiro comum, incluindo a informação da marcação fluorescente (cor), as sequências dos *primers Forward (FWD)* e *Reverse (REV)*, o tamanho do produto amplificado em pares de base (pb) e o motivo de repetição do respectivo loco.

Marcador	Cor	<i>FWD primer</i>	<i>REV primer</i>	Tamanho do produto (pb)	Motivo
BM143	HEX	GGGAAATGAACAGAGGAAA	ATGTTGGGAACTTTTAGTGTG	100 - 170	(GA)35
PV25	6-FAM	GAGCTTCTCCGTCTGTGT	CGAACTGAATCAGAAAGGAA	150 - 180	(CT)23
BM164**	NED	CCACCACAAGGAGAAGCAAC	ACCATT CAGGCCGATACTCC	130-190	(GT)9(GA)21
BM114	6-FAM	AGCCTGGTAAAATGCTCATAG	CATGCTTGTTCCTAACTCTCT	230 - 260	(TA)8(GT)10
BM138	NED	TGTCCCTAAGAACGAATATGGAATC	GAATCAAGCAACCTTGGATCATAAC	190-210	(GT)13
PV169	HEX	TGGAAAGTCGGAGGAGAAGA	AAAAGGGTCCCAACCAAAAC	195-220	(GA)6
PV5**	HEX	ATTAGACGCTGATGACAGAG	AGCAGAATCCTTTGAGTGTG	160 - 220	(GA)22
PV35	6-FAM	TCTACGCGTTCCTCTGTCT	AGTGGATGTGTGGGAAAAGC	190- 260	(TC)10
BM202	6-FAM	ATGCGAAAGAGGAACAATCG	CCTTTACCCACACGCCTTC	100 - 173	(GA)9GT(GA)4
BM189	NED	CTCCCACTCTACCCTCACT	GCGCCAAGTGAACCTAAGTAGA	80 - 120	(CT)13
BM210**	NED	ACCACTGCAATCCTCATCTTTG	CCCTCATCTCCATTCTTATCG	160 - 220	(CT)15
BM155**	HEX	GTTTATGTTTGTGACAGTTCA	CAGAAGTAGTGTGGTTGATACA	100-130	(CA) 8
BM187	HEX	TTTCTCCAACCTCACTCTTTCC	TGTGTTTGTGTTCCGAATTATGA	150 - 220	(CT)10T(CT)14
PV113	NED	TGCATTCTTCTCCCATCTT	TTGATTTGATTTGATCAGTGGTG	60 - 110	(GA)32
PV87	NED	CTCATTGCGTCTACCAAGTGC	CCTAGGTTCCGCAGCATGT	150 - 201	(GA)16
PV272	6-FAM	CAGAACAGAAGAAGAAACAGAAAATG	GCGTGTTCCTCTGTGTGTGT	70 - 120	(GA)17
BM154	6-FAM	TCTTGCGACCGAGCTTCTCC	CTGAATCTGAGGAACGATGACCAG	205 - 318	(CT)17
PV13	6-FAM	TGAGAAAGTTGATGGGATTG	ACGCTGTTGAAGGCTCTAC	159 - 200	(AG)23
PV11**	HEX	AAACTCAAAGTCGTTGTTC	CCACTGACTCTAGCTCCTCC	175 - 192	(TC)8(GT)4
BM181**	NED	ATGCTGCGAGTTAATGATCG	TGAGGAGCAAACAGATGAGG	170 - 200	(CT)17
BM183	6-FAM	CTCAAATCTATTCACTGGTCAGC	TCTTACAGCCTTGCAGACATC	130 - 160	(TC)14
PV163	6-FAM	TGAGAGTGGAGAAGGAGAGAGA	TGACAACACTGCAAACACCA	180 - 350	(AG)16
BM201**	NED	TGGTGCTACAGACTTGATGG	TGTCACCTCTCTCTCCAAT	90 - 120	(GA)15
PV251**	HEX	TGAAGTTGCAGCTAGGTTGG	GGTTGTGCTTGTGTTGTTGG	190 - 220	(CT)7

\*\*Loco selecionado com elevado poder de discriminação do germoplasma de feijoeiro comum de origem Andina e Mesoamericana.

**Material Suplementar 2:** Descrição dos 34 marcadores microssatélites derivados de BES (BES-SSRs) utilizados na caracterização dos 88 genótipo de feijoeiro comum, incluindo a informação da marcação fluorescente (cor), as sequências dos *primers Forward (FWD)* e *Reverse (REV)*, o tamanho do produto amplificado em pares de base (pb) e o motivo de repetição do respectivo loco.

Marcador	Cor	FWD primer	REV primer	Tamanho do produto (pb)	Motivo
PvComp4	6-FAM	TCTTCACCACTTTGAAAACACG	TCAGATAAATGTTGGTATTGGCA	217	(AATATC)5 (AATATA)7 (AT)9
PvPenta4	6-FAM	AGCAACTTTCGGTCTGGTAAGT	TGAATCATTGCTCCTAACCCCTT	239	(ATTTT)4
PvPenta5	6-FAM	AACTGTTCTTGTGGGTCAAT	GCCCAAGGAGTAATTAACAATAAAA	254	(TTTAT)4
PvComp9	6-FAM	TGAATGCAACATCTAATACTAACTCAC	TGTTCTAGGTCTAAAGGCCACA	267	(AT)10 (TATG)4 (TATG)4 (GTAT)4 (AT)7
PvPenta8	6-FAM	AAGAGCATGTTTACTTCACTCATTTT	TGGGGATGGTGTTTGTTTT	290	(TTTAT)4
PvTetra47**	6-FAM	TGGAATGGAGAAGAGACATCCT	CAAACCATGTTTCCAGCATCTA	321	(TTAG)4
PvPenta14**	6-FAM	GAAACTATTCACGGAACAAGCC	AGGAGTGGTGGAGGCAGTATAA	336	(AAAAT)4
PvTetra65**	6-FAM	TCTGAATCAAACATGTCCTCAGA	GGCCAACTTGTTTAAAGGTGAG	366	(AATA)4
PvComp2	NED	CTCCGAATCAGAAACCTATTG	GTGGATGAAGAGAAAGGCCAAAG	201	(TCTTTC)6 (TTCTCC)5
PvTetra25	NED	CACTTTCCTTATGCCTTCACAG	TAATTTGACCAATGCCAAACAC	226	(ACTA)22
PvTri6**	NED	CGAATGGGAGAAGAAGGTTATG	GACATTCCTGTGTCGTTTCCAA	242	(AGA)16
PvHexa20**	NED	GTGTCTTCTATAGGTGTCCCG	GAGTATTTCAAAGCTTGGCCTT	257	(CCCAAG)4
PvTri5	NED	AGAAAATCATGCAGGTTGAGGT	TCCAGCTAAATAGATGATACGTAATTG	269	(TAA)10
PvPenta10	NED	GTTCTTCTATTTCCATCTATC	AATATACATAAGAGTCACTTCT	297	(ATAAA)4
PvTetra50	NED	TAACATGGTTAGGCCTTTTGAA	TCGTACGGATCCAAGTATTAATTT	324	(AAAT)4
PvComp21**	NED	ACGAGTTATTGTTCCAGATGGG	TTATCCTTCTTATGCGGAGACC	353	(AGATGA)4 (TG)5
PvHexa39	NED	TTAATGCCTCCACTTGTTGTTG	CATGAGGCCCAAGTCAAAA	380	(TTTTTC)5
PvPenta19	NED	TTAGGTCTTCAAAGAGATTTGG	TGTGGTAGTAGATGTTAAAGTCATTT	399	(AATTT)4
PvHexa10	PET	AACTTGTTTATCGCATCCAGAA	ATGCAATCAAGGAATGCTCATA	184	(TGCACA)4
PvTri8	PET	CAATGTGGAACAACTGAGGTG	AAGCAAAGTGTCTGAATTGCTG	240	(ATA)10
PvTetra32	PET	ATTCTGCCACTAACGAAGTGT	CTAAAGGCCTAGCAGATTGCAT	255	(AGAA)4
PvComp10	PET	AAATTCTATGATCAACCCGTGG	TGATCCCTGTAGAGGAATCTCA	268	(TA)13 (ATTT)4
PvTetra49	PET	TGGGTAGAGCTTGGTCTTCATT	AGTTTGTGAGTGTGATGGG	324	(ATAA)4
PvHexa36	PET	TCACTTTGGCACCTCCTATTT	AGGATTGTTTGCCTAAACCAGA	352	(CTTGGG)4
PvPenta16	PET	ACATTTGGTTTTGGTTTTGGTC	TCTAAAATGGTCTCGAATTTATTCAC	379	(AATAA)4
PvComp27**	PET	CAAGATGAACATCACCATTCT	AAAAGAACATTTGTCAGTCCA	395	(TATT)4 (ATA)6 (TTA)5
PvHexa12	VIC	TGCTAAATAGCCAAAGCAACAA	CATCACCACAGCACCAGTATT	203	(AACTCA)5
PvHexa15	VIC	CCAAACGAACCGACTATTTCTC	TTTGACCTCCCTTATGTTCT	235	(AAGCCC)4

**Material Suplementar 2:** Continuação.

<b>Marcador</b>	<b>Cor</b>	<b>FWD primer</b>	<b>REV primer</b>	<b>Tamanho do produto (pb)</b>	<b>Motivo</b>
PvHexa19	VIC	TGCTTCTCTGTCGACTTGGAA	AATTACAAGCCTGAAGCTGCTC	252	(TATGCT)4
PvComp8	VIC	TCACTATGTGAAATTGAACCCA	TTCCTACCTAACTTACTTGTACCACTT	262	(AT)7 (TAAC)4
PvPenta13**	VIC	ACTGAAGAAAAGTACTAGAAAACCTTACA	CCCCTTTTAATCAGAGAATTTTA	331	(TAAAA)4
PvTetra57	VIC	ATCAACTATGGCGGATTGACTT	GAAAACAAATCCTTTTGACCCA	359	(TTTA)5
PvTetra73	VIC	TGGTATCGAAGCATTAGGTTCA	GTAAGGTAACGGGTGTGAAT	389	(TTTG)4
PvTetra76	VIC	TACTCAAGCTTCTCTGCAC	TGAAATATATGTTGCGGAAT	399	(TTTA)4

\*\*Loco selecionado com elevado poder de discriminação do germoplasma de feijoeiro comum de origem Andina e Mesoamericana.

**Material Suplementar 3:** Descrição dos 384 marcadores SNPs utilizados na caracterização dos 88 genótipo de feijoeiro comum, incluindo a informação do nome completo dos SNPs (Hyten et al. 2010), a nomenclatura reduzida (SNP *id*), o cromossomo de ancoragem (Crom.), a variação da base (SNP), o comprimento da sequência do *primer* ASO, o posicionamento inicial e final no genoma, o *E*-value correspondente ao alinhamento e a porcentagem (%) de identidade.

SNP	SNP <i>id</i> *	Crom.	SNP	Tamanho da sequência (pb)**	Posição inicial (pb)	Posição final (pb)	<i>E</i> -value	Identidade (%)
BARC-PV-0002895	PV-2895	Pv04	[C/T]	235	21793951	21794185	4,00E-115	98,30
BARC-PV-0002896	PV-2896	Pv04	[C/A]	171	16059644	16059474	5,00E-83	99,42
BARC-PV-0002897	PV-2897	Pv07	[G/A]	206	2029792	2029588	1,00E-99	98,54
BARC-PV-0002932	PV-2932	Pv07	[C/A]	241	39896079	39896318	7,00E-118	98,33
BARC-PV-0002954	PV-2954	Pv08	[G/A]	241	18845111	18844871	1,00E-120	99,17
BARC-PV-0002964	PV-2964	Pv05	[C/T]	241	25096338	25096098	1,00E-120	99,17
BARC-PV-0003019	PV-3019	Pv08	[T/C]	241	26300577	26300817	2,00E-118	98,76
BARC-PV-0003020	PV-3020	Pv11	[A/T]	202	33278887	33278689	2,00E-97	98,99
BARC-PV-0003023	PV-3023	Pv01	[G/A]	358	37831349	37831706	0,0	99,72
BARC-PV-0003025	PV-3025	Pv02	[T/A]	208	14421666	14421873	2,00E-102	99,04
BARC-PV-0003026	PV-3026	Pv08	[T/C]	241	38903479	38903239	9,00E-122	99,59
BARC-PV-0003027	PV-3027	Pv08	[T/G]	193	31900058	31900249	4,00E-80	94,82
BARC-PV-0003032	PV-3032	Pv10	[T/G]	211	25659837	25660048	6,00E-103	98,58
BARC-PV-0003037	PV-3037	Pv03	[C/T]	241	26733336	26733096	9,00E-122	99,59
BARC-PV-0003042	PV-3042	Pv08	[A/G]	241	45914304	45914066	1,00E-115	97,91
BARC-PV-0003045	PV-3045	Pv07	[A/G]	231	38968650	38968880	3,00E-116	99,57
BARC-PV-0003048	PV-3048	Pv06	[G/C]	268	4207613	4207880	1,00E-136	99,63
BARC-PV-0003052	PV-3052	--	[A/C]	384	--	--	--	--
BARC-PV-0003067	PV-3067	Pv09	[A/G]	282	27301927	27302208	4,00E-141	98,58
BARC-PV-0003077	PV-3077	Pv04	[T/C]	241	19976092	19976332	1,00E-120	99,17
BARC-PV-0003084	PV-3084	Pv01	[G/T]	199	22041972	22042170	2,00E-98	99,50
BARC-PV-0003092	PV-3092	Pv03	[A/C]	241	20494704	20494464	1,00E-120	99,17
BARC-PV-0003094	PV-3094	Pv07	[C/A]	400	40001763	40002162	0,0	99,50
BARC-PV-0003097	PV-3097	Pv03	[T/C]	241	23857839	23857599	4,00E-115	97,10
BARC-PV-0003100	PV-3100	Pv11	[A/G]	224	30951084	30951307	3,00E-111	99,11
BARC-PV-0003106	PV-3106	Pv07	[A/C]	359	37110125	37109767	0,0	99,16
BARC-PV-0003108	PV-3108	Pv11	[A/G]	241	40505469	40505229	9,00E-122	99,59
BARC-PV-0003118	PV-3118	Pv03	[A/G]	241	30615383	30615143	1,00E-120	99,17
BARC-PV-0003122	PV-3122	Pv07	[A/C]	241	8352412	8352652	9,00E-122	99,59
BARC-PV-0003123	PV-3123	Pv02	[G/A]	227	46197362	46197588	5,00E-114	99,56
BARC-PV-0003131	PV-3131	Pv11	[G/A]	241	28546349	28546109	9,00E-122	99,59
BARC-PV-0003140	PV-3140	Pv11	[G/C]	313	29127517	29127829	1,00E-161	99,68
BARC-PV-0003167	PV-3167	Pv03	[T/C]	176	29332158	29331983	1,00E-84	98,86
BARC-PV-0003186	PV-3186	Pv03	[A/G]	272	43875126	43874855	6,00E-134	98,16
BARC-PV-0003194	PV-3194	Pv11	[A/G]	349	38884891	38885239	2,00E-180	99,43
BARC-PV-0003197	PV-3197	Pv04	[T/C]	241	28817325	28817085	9,00E-122	99,59
BARC-PV-0003208	PV-3208	Pv10	[C/A]	343	9555050	9555392	8,00E-174	98,83
BARC-PV-0003215	PV-3215	Pv05	[A/G]	256	18470009	18469754	4,00E-126	98,44
BARC-PV-0003223	PV-3223	Pv08	[C/T]	241	26782138	26782378	4,00E-120	99,17
BARC-PV-0003224	PV-3224	Pv07	[G/T]	241	2506111	2506351	2,00E-119	98,76
BARC-PV-0003237	PV-3237	Pv09	[T/C]	241	4719689	4719929	7,00E-118	98,34
BARC-PV-0003239	PV-3239	Pv03	[T/G]	241	38532431	38532191	3,00E-117	97,93
BARC-PV-0003241	PV-3241	Pv08	[A/G]	237	46736737	46736973	2,00E-119	99,58
BARC-PV-0003259	PV-3259	Pv03	[G/C]	241	22679069	22678829	2,00E-119	98,76
BARC-PV-0003262	PV-3262	Pv03	[C/A]	241	48908987	48908747	1,00E-120	99,17
BARC-PV-0003264	PV-3264	Pv08	[C/A]	241	49091822	49092062	9,00E-122	99,59
BARC-PV-0003269	PV-3269	Pv03	[T/C]	350	26219182	26218833	0,0	99,71

**Material Suplementar 3: Continuação.**

SNP	SNP <i>id</i> *	Crom.	SNP	Tamanho da sequência (pb)**	Posição inicial (pb)	Posição final (pb)	<i>E-value</i>	Identidade (%)
BARC-PV-0003278	PV-3278	Pv04	[T/C]	240	21957304	21957064	4,00E-120	99,17
BARC-PV-0003281	PV-3281	Pv03	[A/G]	240	14089180	14088941	3,00E-121	99,58
BARC-PV-0003282	PV-3282	Pv05	[T/C]	236	23577310	23577075	1,00E-115	98,31
BARC-PV-0003294	PV-3294	Pv11	[C/T]	241	14098511	14098271	9,00E-122	99,59
BARC-PV-0003298	PV-3298	Pv07	[G/A]	241	33856919	33857159	9,00E-122	99,59
BARC-PV-0003319	PV-3319	Pv11	[T/C]	241	34371573	34371813	1,00E-120	99,17
BARC-PV-0003330	PV-3330	Pv08	[G/A]	241	44488692	44488452	4,00E-115	97,93
BARC-PV-0003331	PV-3331	Pv08	[C/A]	241	53912385	53912145	1,00E-120	99,17
BARC-PV-0003337	PV-3337	Pv09	[A/G]	237	15024521	15024757	2,00E-118	99,16
BARC-PV-0003353	PV-3353	Pv03	[T/C]	354	32947666	32947313	0,0	99,72
BARC-PV-0003368	PV-3368	Pv09	[G/A]	205	3182606	3182402	8,00E-102	99,51
BARC-PV-0003369	PV-3369	Pv02	[G/A]	241	39945106	39944866	9,00E-122	99,59
BARC-PV-0003383	PV-3383	Pv06	[C/T]	241	2901506	2901266	9,00E-122	99,59
BARC-PV-0003388	PV-3388	scaffold_14	[C/T]	352	57341	56990	0,0	99,43
BARC-PV-0003399	PV-3399	Pv01	[A/G]	241	7765298	7765538	9,00E-122	99,59
BARC-PV-0003400	PV-3400	Pv08	[A/G]	241	36151836	36152079	3,00E-91	91,80
BARC-PV-0003401	PV-3401	Pv04	[A/T]	216	35303593	35303378	1,00E-104	98,61
BARC-PV-0003403	PV-3403	Pv04	[T/C]	241	19803949	19804189	2,00E-118	98,34
BARC-PV-0003413	PV-3413	Pv07	[G/A]	240	35970659	35970898	4,00E-120	99,17
BARC-PV-0003416	PV-3416	Pv11	[C/T]	241	27616995	27617235	9,00E-122	99,59
BARC-PV-0003419	PV-3419	Pv03	[A/G]	241	28339411	28339171	1,00E-120	99,17
BARC-PV-0003422	PV-3422	Pv02	[C/T]	264	16987369	16987630	4,00E-131	98,85
BARC-PV-0003439	PV-3439	Pv11	[T/C]	198	27445558	27445361	1,00E-95	98,48
BARC-PV-0003449	PV-3449	Pv10	[T/G]	225	20645850	20646074	9,00E-112	99,11
BARC-PV-0003450	PV-3450	Pv08	[G/A]	231	1609675	1609445	3,00E-116	99,57
BARC-PV-0003454	PV-3454	Pv01	[G/A]	241	34868524	34868764	1,00E-120	99,17
BARC-PV-0003463	PV-3463	Pv04	[T/A]	241	27901826	27902066	9,00E-122	99,59
BARC-PV-0003468	PV-3468	Pv08	[T/C]	241	20468756	20468516	3,00E-111	96,28
BARC-PV-0003474	PV-3474	Pv01	[G/A]	241	3640287	3640047	9,00E-122	99,59
BARC-PV-0003476	PV-3476	Pv04	[C/G]	241	27499140	27499380	1,00E-120	99,17
BARC-PV-0003493	PV-3493	Pv04	[C/G]	341	7220558	7220218	1,00E-172	98,83
BARC-PV-0003496	PV-3496	Pv09	[G/A]	238	14368701	14368464	4,00E-120	99,58
BARC-PV-0003501	PV-3501	Pv07	[A/G]	162	36954743	36954904	5,00E-78	99,38
BARC-PV-0003503	PV-3503	Pv01	[C/T]	236	32155689	32155454	2,00E-108	95,76
BARC-PV-0003508	PV-3508	Pv03	[T/G]	241	13607597	13607837	2,00E-118	98,76
BARC-PV-0003541	PV-3541	Pv08	[T/C]	188	33970663	33970476	2,00E-92	99,47
BARC-PV-0003550	PV-3550	Pv10	[T/C]	223	11034873	11034651	9,00E-112	99,55
BARC-PV-0003551	PV-3551	Pv10	[A/G]	241	24564962	24565202	1,00E-120	99,17
BARC-PV-0003561	PV-3561	Pv10	[C/T]	203	4780454	4780252	8,00E-97	98,03
BARC-PV-0003562	PV-3562	Pv10	[C/G]	232	31818841	31818621	7,00E-108	98,65
BARC-PV-0003566	PV-3566	Pv11	[C/T]	394	19806262	19805869	0,0	99,75
BARC-PV-0003574	PV-3574	Pv08	[A/T]	241	40299150	40299387	1,00E-115	98,32
BARC-PV-0003583	PV-3583	Pv01	[C/A]	241	24095198	24095438	9,00E-122	99,59
BARC-PV-0003585	PV-3585	Pv02	[A/G]	237	27452228	27451992	2,00E-119	99,58
BARC-PV-0003586	PV-3586	Pv04	[C/A]	241	40905739	40905979	9,00E-122	99,59
BARC-PV-0003594	PV-3594	Pv11	[T/G]	343	21597261	21597603	3,00E-178	99,71
BARC-PV-0003596	PV-3596	Pv07	[A/G]	241	6232304	6232064	1,00E-120	99,17
BARC-PV-0003600	PV-3600	Pv03	[T/G]	241	20204216	20204456	9,00E-122	99,59
BARC-PV-0003623	PV-3623	Pv05	[C/G]	241	38483476	38483716	9,00E-122	99,59
BARC-PV-0003624	PV-3624	Pv07	[G/A]	326	42414256	42413937	4,00E-161	98,44
BARC-PV-0003629	PV-3629	Pv05	[G/A]	174	10003510	10003683	2,00E-82	98,28
BARC-PV-0003651	PV-3651	Pv01	[T/C]	235	43764566	43764800	7,00E-113	97,45
BARC-PV-0003652	PV-3652	Pv06	[T/C]	351	303104	303454	2,00E-180	99,15

**Material Suplementar 3: Continuação.**

SNP	SNP <i>id</i> *	Crom.	SNP	Tamanho da sequência (pb)**	Posição inicial (pb)	Posição final (pb)	<i>E-value</i>	Identidade (%)
BARC-PV-0003657	PV-3657	Pv08	[A/C]	241	15199525	15199765	1,00E-120	99,17
BARC-PV-0003658	PV-3658	Pv02	[A/G]	241	3864873	3864633	9,00E-122	99,59
BARC-PV-0003667	PV-3667	Pv06	[T/C]	241	13890179	13889939	9,00E-122	99,59
BARC-PV-0003680	PV-3680	Pv05	[T/C]	189	40079781	40079593	6,00E-93	99,47
BARC-PV-0003682	PV-3682	Pv07	[C/T]	235	6308947	6309181	2,00E-117	99,15
BARC-PV-0003697	PV-3697	Pv04	[A/G]	240	29755617	29755379	3,00E-117	98,33
BARC-PV-0003700	PV-3700	Pv11	[A/G]	210	34337639	34337430	1,00E-104	99,52
BARC-PV-0003703	PV-3703	Pv02	[C/T]	234	31875852	31876085	7,00E-118	99,57
BARC-PV-0003706	PV-3706	Pv06	[A/T]	182	3589443	3589624	4,00E-89	99,45
BARC-PV-0003734	PV-3734	Pv09	[G/A]	219	35842580	35842362	1,00E-109	99,54
BARC-PV-0003741	PV-3741	Pv01	[G/A]	241	21863092	21862852	2,00E-119	98,76
BARC-PV-0003756	PV-3756	Pv05	[A/T]	241	19962810	19962570	9,00E-122	99,59
BARC-PV-0003777	PV-3777	Pv11	[G/A]	241	27722646	27722406	9,00E-122	99,59
BARC-PV-0003791	PV-3791	Pv10	[C/A]	235	12084113	12084329	3,00E-71	88,69
BARC-PV-0003800	PV-3800	Pv06	[T/C]	345	26775431	26775738	3,00E-143	95,78
BARC-PV-0003812	PV-3812	Pv02	[T/A]	241	36661642	36661402	9,00E-122	99,59
BARC-PV-0003814	PV-3814	Pv07	[C/A]	311	32495999	32496306	5,00E-150	97,08
BARC-PV-0003821	PV-3821	Pv01	[G/A]	241	16706614	16706854	9,00E-122	99,59
BARC-PV-0003822	PV-3822	Pv07	[T/C]	296	18808719	18809014	4,00E-151	99,32
BARC-PV-0003826	PV-3826	Pv04	[G/T]	372	18457546	18457175	0,0	99,46
BARC-PV-0003839	PV-3839	Pv05	[A/G]	241	12710940	12711180	9,00E-122	99,59
BARC-PV-0003840	PV-3840	Pv07	[G/T]	359	15324865	15325223	0,0	99,44
BARC-PV-0003881	PV-3881	Pv02	[T/C]	352	20764666	20764315	0,0	99,72
BARC-PV-0003883	PV-3883	Pv04	[A/G]	187	30070108	30069922	9,00E-91	98,93
BARC-PV-0003886	PV-3886	Pv01	[T/A]	241	759253	759493	9,00E-122	99,59
BARC-PV-0003889	PV-3889	Pv11	[G/A]	190	15007356	15007545	2,00E-93	99,47
BARC-PV-0003892	PV-3892	Pv03	[G/T]	275	22291199	22290925	2,00E-138	98,91
BARC-PV-0003894	PV-3894	Pv02	[A/G]	241	12053587	12053827	9,00E-122	99,59
BARC-PV-0003905	PV-3905	Pv10	[T/C]	241	3736855	3736615	2,00E-118	98,34
BARC-PV-0003923	PV-3923	Pv11	[G/A]	347	29196023	29196369	2,00E-175	98,56
BARC-PV-0003924	PV-3924	Pv03	[G/A]	241	10229678	10229918	9,00E-122	99,59
BARC-PV-0003932	PV-3932	Pv08	[T/C]	241	39578966	39578726	4,00E-120	99,17
BARC-PV-0003933	PV-3933	Pv01	[A/G]	241	28574903	28575143	9,00E-122	99,59
BARC-PV-0003950	PV-3950	Pv07	[T/G]	239	28835363	28835126	2,00E-118	99,16
BARC-PV-0003967	PV-3967	Pv11	[C/T]	241	33277466	33277706	9,00E-122	99,59
BARC-PV-0003968	PV-3968	Pv05	[T/G]	242	19772575	19772334	1,00E-115	97,52
BARC-PV-0003973	PV-3973	Pv01	[C/T]	270	5654089	5654358	2,00E-134	98,52
BARC-PV-0003993	PV-3993	Pv03	[T/A]	241	28493801	28494041	9,00E-122	99,59
BARC-PV-0003994	PV-3994	Pv04	[T/C]	241	17061646	17061406	4,00E-120	99,17
BARC-PV-0003999	PV-3999	Pv02	[C/T]	241	35794426	35794186	2,00E-119	98,76
BARC-PV-0004006	PV-4006	Pv01	[A/G]	365	37698983	37698619	0,0	99,45
BARC-PV-0004007	PV-4007	Pv05	[T/C]	241	32605700	32605460	9,00E-122	99,59
BARC-PV-0004019	PV-4019	Pv05	[T/C]	241	30123567	30123327	2,00E-118	98,34
BARC-PV-0004020	PV-4020	Pv07	[A/C]	401	25015030	25015430	0,0	98,25
BARC-PV-0004026	PV-4026	Pv11	[A/G]	387	27147586	27147200	0,0	98,71
BARC-PV-0004027	PV-4027	Pv03	[G/C]	240	14284106	14284342	2,00E-119	99,58
BARC-PV-0004028	PV-4028	Pv03	[C/T]	241	23879992	23879752	3,00E-117	97,93
BARC-PV-0004043	PV-4043	Pv07	[C/A]	201	22397327	22397127	3,00E-96	98,01
BARC-PV-0004054	PV-4054	Pv08	[A/T]	241	36805495	36805256	2,00E-113	97,52
BARC-PV-0004057	PV-4057	Pv08	[A/T]	233	58174226	58173994	3,00E-116	99,14
BARC-PV-0004069	PV-4069	Pv10	[C/T]	235	18947842	18947608	2,00E-117	99,15
BARC-PV-0004071	PV-4071	Pv04	[A/G]	241	29876206	29875966	9,00E-122	99,59
BARC-PV-0004077	PV-4077	Pv01	[A/G]	241	17425597	17425357	9,00E-122	99,59

**Material Suplementar 3: Continuação.**

SNP	SNP <i>id</i> *	Crom.	SNP	Tamanho da sequência (pb)**	Posição inicial (pb)	Posição final (pb)	<i>E-value</i>	Identidade (%)
BARC-PV-0004081	PV-4081	Pv10	[C/T]	241	20987084	20987324	9,00E-122	99,59
BARC-PV-0004089	PV-4089	Pv06	[A/T]	241	30081534	30081774	1,00E-120	99,17
BARC-PV-0004090	PV-4090	Pv01	[C/G]	241	22006688	22006928	4,00E-120	99,17
BARC-PV-0004095	PV-4095	Pv10	[A/C]	240	10740046	10739807	2,00E-119	99,17
BARC-PV-0004096	PV-4096	Pv08	[T/C]	241	30418332	30418092	7,00E-118	98,34
BARC-PV-0004097	PV-4097	Pv07	[C/T]	235	32013306	32013072	2,00E-118	99,57
BARC-PV-0004112	PV-4112	Pv08	[A/C]	241	35345832	35346072	6,00E-119	98,76
BARC-PV-0004115	PV-4115	Pv09	[G/A]	204	17742582	17742382	6,00E-93	96,52
BARC-PV-0004129	PV-4129	Pv08	[T/C]	206	38383965	38383760	2,00E-102	99,51
BARC-PV-0004130	PV-4130	Pv01	[C/T]	198	15376035	15376232	8,00E-97	98,99
BARC-PV-0004134	PV-4134	Pv02	[T/C]	354	23186946	23186593	0,0	99,72
BARC-PV-0004140	PV-4140	Pv04	[C/T]	179	30989219	30989041	2,00E-86	98,88
BARC-PV-0004144	PV-4144	Pv07	[C/T]	449	25942183	25942550	2,00E-136	90,24
BARC-PV-0004147	PV-4147	Pv06	[G/A]	347	9350543	9350197	8,00E-174	97,98
BARC-PV-0004174	PV-4174	Pv04	[C/A]	241	13976865	13977104	2,00E-114	97,92
BARC-PV-0004175	PV-4175	Pv10	[C/T]	241	16739128	16739368	9,00E-122	99,59
BARC-PV-0004181	PV-4181	Pv04	[G/A]	241	37289747	37289986	4,00E-120	99,17
BARC-PV-0004183	PV-4183	Pv01	[T/C]	234	38921455	38921688	7,00E-118	99,57
BARC-PV-0004184	PV-4184	Pv11	[T/G]	225	17738698	17738474	7,00E-113	99,56
BARC-PV-0004206	PV-4206	Pv01	[C/T]	241	35290554	35290314	9,00E-122	99,59
BARC-PV-0004216	PV-4216	Pv04	[G/C]	163	28269707	28269548	1,00E-73	97,50
BARC-PV-0004220	PV-4220	Pv10	[A/C]	241	21443732	21443492	9,00E-122	99,59
BARC-PV-0004230	PV-4230	Pv09	[T/C]	241	16079406	16079166	9,00E-122	99,59
BARC-PV-0004233	PV-4233	Pv01	[A/G]	241	26840676	26840916	9,00E-122	99,59
BARC-PV-0004238	PV-4238	Pv02	[G/A]	241	45830746	45830506	7,00E-118	98,34
BARC-PV-0004239	PV-4239	Pv02	[C/T]	241	7726742	7726982	9,00E-122	99,59
BARC-PV-0004243	PV-4243	Pv02	[G/A]	121	11287495	11287375	2,00E-55	99,17
BARC-PV-0004250	PV-4250	Pv08	[A/G]	350	49354925	49354576	0,0	99,71
BARC-PV-0004260	PV-4260	Pv03	[A/G]	241	14713402	14713162	3,00E-117	98,34
BARC-PV-0004265	PV-4265	Pv01	[A/C]	354	37913241	37912905	2,00E-115	89,09
BARC-PV-0004266	PV-4266	Pv11	[G/A]	241	40418593	40418833	7,00E-118	98,34
BARC-PV-0004268	PV-4268	Pv10	[G/A]	236	21389315	21389550	3,00E-117	99,15
BARC-PV-0004270	PV-4270	Pv09	[T/G]	220	4622624	4622405	2,00E-108	99,09
BARC-PV-0004274	PV-4274	Pv02	[C/A]	342	40080996	40081337	3,00E-172	98,25
BARC-PV-0004276	PV-4276	Pv09	[T/C]	378	583137	583514	0,0	99,74
BARC-PV-0004281	PV-4281	Pv03	[A/C]	241	9310173	9309933	4,00E-120	99,17
BARC-PV-0004292	PV-4292	Pv04	[G/T]	228	34845024	34845251	1,00E-104	96,93
BARC-PV-0004295	PV-4295	Pv04	[A/G]	241	19659029	19659269	9,00E-122	99,59
BARC-PV-0004303	PV-4303	Pv11	[T/C]	241	18001892	18002132	1,00E-120	99,17
BARC-PV-0004309	PV-4309	Pv01	[A/G]	241	51961774	51962014	9,00E-122	99,59
BARC-PV-0004312	PV-4312	Pv10	[G/A]	371	2651237	2650986	2,00E-106	94,47
BARC-PV-0004332	PV-4332	Pv10	[T/A]	241	32724197	32723957	1,00E-120	99,17
BARC-PV-0004335	PV-4335	Pv04	[G/A]	241	27301618	27301858	9,00E-122	99,59
BARC-PV-0004339	PV-4339	Pv11	[A/G]	241	6326053	6325813	9,00E-122	99,59
BARC-PV-0004351	PV-4351	Pv05	[T/C]	240	27553191	27552951	1,00E-115	97,51
BARC-PV-0004354	PV-4354	Pv10	[A/G]	241	10960142	10960382	2,00E-118	98,34
BARC-PV-0004374	PV-4374	Pv10	[A/G]	241	23304134	23304374	9,00E-122	99,59
BARC-PV-0004385	PV-4385	Pv08	[T/C]	241	13392017	13392257	1,00E-120	99,17
BARC-PV-0004417	PV-4417	Pv07	[G/C]	229	8173048	8173276	4,00E-115	99,56
BARC-PV-0004423	PV-4423	Pv01	[A/G]	234	44909893	44910122	3,00E-111	98,26
BARC-PV-0004427	PV-4427	Pv09	[C/T]	195	8614249	8614054	2,00E-87	95,41
BARC-PV-0004435	PV-4435	Pv06	[C/T]	241	687375	687615	2,00E-118	98,34
BARC-PV-0004466	PV-4466	Pv03	[A/C]	176	13608178	13608003	9,00E-86	99,43

**Material Suplementar 3: Continuação.**

SNP	SNP <i>id</i> *	Crom.	SNP	Tamanho da sequência (pb)**	Posição inicial (pb)	Posição final (pb)	<i>E-value</i>	Identidade (%)
BARC-PV-0004473	PV-4473	Pv04	[C/G]	241	2442480	2442240	9,00E-122	99,59
BARC-PV-0004475	PV-4475	Pv06	[T/C]	183	13889372	13889554	5,00E-88	98,91
BARC-PV-0004479	PV-4479	Pv04	[G/T]	367	340663	341029	0,0	99,18
BARC-PV-0004483	PV-4483	Pv11	[G/C]	163	33126977	33126815	2,00E-77	98,77
BARC-PV-0004489	PV-4489	Pv08	[A/G]	241	52275335	52275095	1,00E-120	99,17
BARC-PV-0004496	PV-4496	Pv03	[T/G]	241	32946081	32946320	4,00E-120	99,17
BARC-PV-0004505	PV-4505	Pv03	[G/A]	241	35867120	35867360	9,00E-122	99,59
BARC-PV-0004506	PV-4506	Pv11	[G/T]	123	22452159	22452281	2,00E-56	99,19
BARC-PV-0004523	PV-4523	Pv09	[C/T]	241	34878827	34879067	4,00E-120	99,17
BARC-PV-0004532	PV-4532	Pv05	[A/C]	241	32155447	32155210	1,00E-96	93,72
BARC-PV-0004536	PV-4536	Pv01	[C/A]	322	27138385	27138706	1,00E-166	99,69
BARC-PV-0004539	PV-4539	Pv11	[A/G]	241	24295033	24294794	3,00E-101	94,61
BARC-PV-0004548	PV-4548	Pv08	[C/G]	241	351331	351091	4,00E-120	99,17
BARC-PV-0004561	PV-4561	Pv09	[A/G]	237	898299	898534	9,00E-117	98,73
BARC-PV-0004576	PV-4576	Pv10	[T/C]	358	13203846	13203489	0,0	99,44
BARC-PV-0004580	PV-4580	Pv07	[C/A]	241	36892747	36892987	1,00E-120	99,17
BARC-PV-0004605	PV-4605	Pv08	[A/G]	241	20468413	20468176	4,00E-115	98,32
BARC-PV-0004606	PV-4606	Pv10	[G/C]	241	22213483	22213723	9,00E-122	99,59
BARC-PV-0004621	PV-4621	Pv06	[A/G]	241	2027276	2027036	9,00E-122	99,59
BARC-PV-0004634	PV-4634	Pv04	[C/T]	241	38172659	38172419	4,00E-115	97,51
BARC-PV-0004636	PV-4636	Pv06	[G/A]	229	4161993	4161765	4,00E-115	99,56
BARC-PV-0004639	PV-4639	Pv11	[T/C]	241	26923697	26923937	9,00E-122	99,59
BARC-PV-0004651	PV-4651	Pv08	[T/C]	241	10788748	10788508	9,00E-122	99,59
BARC-PV-0004667	PV-4667	Pv01	[T/A]	241	27670439	27670199	9,00E-122	99,59
BARC-PV-0004670	PV-4670	Pv01	[C/T]	189	306622	306810	1,00E-89	97,88
BARC-PV-0004677	PV-4677	Pv05	[T/C]	236	37115819	37116054	2,00E-114	97,88
BARC-PV-0004678	PV-4678	Pv06	[C/T]	342	27854444	27854785	2,00E-175	99,12
BARC-PV-0004686	PV-4686	Pv08	[A/G]	223	16257792	16258014	9,00E-112	99,55
BARC-PV-0004709	PV-4709	Pv02	[A/G]	241	12964252	12964492	9,00E-122	99,59
BARC-PV-0004715	PV-4715	Pv02	[G/C]	241	35339465	35339225	9,00E-122	99,59
BARC-PV-0004718	PV-4718	Pv04	[G/A]	191	28664251	28664439	7,00E-92	98,94
BARC-PV-0004725	PV-4725	Pv09	[G/A]	241	18286045	18285805	9,00E-122	99,59
BARC-PV-0004776	PV-4776	Pv05	[A/G]	219	3380429	3380647	2,00E-107	98,63
BARC-PV-0004791	PV-4791	Pv06	[T/C]	241	27055369	27055129	9,00E-122	99,59
BARC-PV-0004797	PV-4797	Pv06	[T/C]	263	14707463	14707725	3,00E-132	99,24
BARC-PV-0004798	PV-4798	Pv10	[G/A]	241	25767409	25767169	9,00E-122	99,59
BARC-PV-0004800	PV-4800	Pv03	[A/G]	241	18972805	18973045	9,00E-122	99,59
BARC-PV-0004801	PV-4801	Pv06	[C/T]	241	20647643	20647883	9,00E-122	99,59
BARC-PV-0004803	PV-4803	Pv02	[A/G]	342	19278389	19278730	4,00E-176	99,42
BARC-PV-0004808	PV-4808	Pv02	[C/G]	240	1983422	1983661	2,00E-119	99,17
BARC-PV-0004813	PV-4813	Pv01	[T/C]	238	35803120	35803357	2,00E-114	97,48
BARC-PV-0004815	PV-4815	Pv02	[G/T]	241	3634177	3633937	9,00E-122	99,59
BARC-PV-0004818	PV-4818	Pv01	[C/T]	241	166274	166514	2,00E-119	98,76
BARC-PV-0004823	PV-4823	Pv01	[A/G]	241	45650641	45650401	1,00E-115	97,93
BARC-PV-0004825	PV-4825	Pv02	[G/A]	223	16868077	16867855	9,00E-112	99,55
BARC-PV-0004829	PV-4829	Pv09	[T/C]	242	1014815	1015054	3,00E-121	99,58
BARC-PV-0004857	PV-4857	Pv10	[G/A]	231	10084154	10084384	4,00E-115	99,13
BARC-PV-0004861	PV-4861	Pv03	[T/C]	241	26480117	26480357	9,00E-122	99,59
BARC-PV-0004876	PV-4876	Pv11	[C/T]	241	33393171	33392931	9,00E-122	99,59
BARC-PV-0004878	PV-4878	Pv08	[G/A]	241	39792980	39792740	9,00E-122	99,59
BARC-PV-0004881	PV-4881	Pv04	[A/C]	241	39276332	39276092	9,00E-122	99,59
BARC-PV-0004882	PV-4882	Pv04	[C/T]	241	25071415	25071175	2,00E-119	98,76
BARC-PV-0004895	PV-4895	Pv06	[C/T]	241	9855716	9855956	1,00E-120	99,17

**Material Suplementar 3: Continuação.**

SNP	SNP <i>id</i> *	Crom.	SNP	Tamanho da sequência (pb)**	Posição inicial (pb)	Posição final (pb)	<i>E-value</i>	Identidade (%)
BARC-PV-0004897	PV-4897	Pv09	[G/A]	231	5130589	5130819	3,00E-116	99,57
BARC-PV-0004899	PV-4899	Pv08	[A/C]	241	34044702	34044462	9,00E-122	99,59
BARC-PV-0004934	PV-4934	Pv06	[A/G]	241	2942870	2943110	9,00E-122	99,59
BARC-PV-0004938	PV-4938	Pv11	[G/A]	238	29274209	29274446	4,00E-120	99,58
BARC-PV-0004951	PV-4951	Pv03	[T/C]	156	18982179	18982024	1,00E-74	99,36
BARC-PV-0004953	PV-4953	Pv08	[A/G]	241	13542857	13542617	4,00E-120	99,17
BARC-PV-0004976	PV-4976	Pv10	[A/G]	241	13608332	13608092	9,00E-122	99,59
BARC-PV-0004979	PV-4979	Pv03	[T/C]	229	34518524	34518296	7,00E-113	98,69
BARC-PV-0004981	PV-4981	Pv03	[T/C]	241	16054857	16054617	2,00E-118	98,34
BARC-PV-0004989	PV-4989	Pv03	[C/T]	241	15979269	15979509	2,00E-119	98,76
BARC-PV-0004996	PV-4996	Pv02	[A/G]	241	14421556	14421316	9,00E-122	99,59
BARC-PV-0004998	PV-4998	Pv08	[C/T]	241	25421606	25421846	4,00E-110	96,68
BARC-PV-0005014	PV-5014	Pv11	[T/C]	241	27950939	27950699	9,00E-122	99,59
BARC-PV-0005017	PV-5017	Pv08	[G/A]	142	42204557	42204416	3,00E-59	95,77
BARC-PV-0005019	PV-5019	Pv10	[C/G]	151	28676657	28676807	6,00E-72	99,34
BARC-PV-0005028	PV-5028	Pv04	[G/A]	241	28700052	28699812	9,00E-122	99,59
BARC-PV-0005040	PV-5040	Pv06	[G/T]	241	10803796	10804036	6,00E-119	98,76
BARC-PV-0005047	PV-5047	Pv05	[T/C]	241	19657897	19657657	9,00E-122	99,59
BARC-PV-0005049	PV-5049	Pv05	[A/T]	212	1741289	1741078	1,00E-99	97,64
BARC-PV-0005054	PV-5054	Pv11	[G/A]	235	45099093	45098859	2,00E-118	99,57
BARC-PV-0005055	PV-5055	Pv03	[C/A]	241	26072145	26072385	9,00E-122	99,59
BARC-PV-0005065	PV-5065	Pv02	[A/G]	241	28993554	28993794	9,00E-122	99,59
BARC-PV-0005091	PV-5091	Pv05	[G/A]	176	5256081	5256256	4,00E-84	98,86
BARC-PV-0005097	PV-5097	Pv03	[G/A]	199	27628543	27628739	1,00E-95	98,98
BARC-PV-0005100	PV-5100	Pv03	[C/T]	268	26195899	26195632	6,00E-129	97,01
BARC-PV-0005136	PV-5136	Pv10	[C/A]	273	23564112	23563840	2,00E-139	99,63
BARC-PV-0005141	PV-5141	Pv03	[C/T]	159	11143930	11144088	3,00E-75	98,74
BARC-PV-0005180	PV-5180	Pv10	[T/C]	281	14562432	14562152	6,00E-144	99,64
BARC-PV-0005188	PV-5188	Pv02	[G/C]	250	8688723	8688474	2,00E-124	98,80
BARC-PV-0005192	PV-5192	Pv03	[G/C]	245	42866828	42867072	6,00E-124	99,59
BARC-PV-0005202	PV-5202	Pv05	[T/C]	269	19427903	19428170	2,00E-128	97,41
BARC-PV-0005238	PV-5238	Pv10	[T/C]	259	12329117	12329373	6,00E-129	99,22
BARC-PV-0005249	PV-5249	Pv05	[T/A]	244	33827832	33827589	2,00E-123	99,59
BARC-PV-0005253	PV-5253	Pv04	[T/C]	266	18035157	18035422	1,00E-135	99,62
BARC-PV-0005261	PV-5261	Pv11	[A/C]	268	42496144	42496411	2,00E-133	98,51
BARC-PV-0005262	PV-5262	Pv11	[T/C]	271	19468783	19468513	3,00E-137	99,26
BARC-PV-0005264	PV-5264	Pv05	[G/T]	237	5429230	5428994	2,00E-119	99,58
BARC-PV-0005268	PV-5268	Pv11	[G/C]	260	16847626	16847885	3,00E-132	99,62
BARC-PV-0005288	PV-5288	Pv07	[T/C]	277	37094976	37095241	1,00E-135	99,62
BARC-PV-0005299	PV-5299	Pv10	[G/A]	263	12533998	12533736	6,00E-134	99,62
BARC-PV-0005304	PV-5304	Pv03	[C/G]	278	49313427	49313704	3,00E-142	99,64
BARC-PV-0005330	PV-5330	Pv11	[T/C]	241	32309234	32309474	9,00E-122	99,59
BARC-PV-0005343	PV-5343	Pv04	[A/C]	255	18382001	18382255	2,00E-129	99,61
BARC-PV-0005392	PV-5392	--	[G/C]	236	--	--	--	--
BARC-PV-0005393	PV-5393	Pv01	[T/G]	252	28847622	28847371	7,00E-128	99,60
BARC-PV-0005421	PV-5421	Pv04	[A/C]	256	32968874	32968619	5,00E-130	99,61
BARC-PV-0005437	PV-5437	Pv10	[C/A]	264	12374825	12374562	2,00E-134	99,62
BARC-PV-0005450	PV-5450	Pv02	[C/G]	237	17023385	17023149	2,00E-119	99,58
BARC-PV-0005461	PV-5461	Pv05	[C/T]	265	19835344	19835608	1,00E-131	98,87
BARC-PV-0005470	PV-5470	Pv05	[C/G]	243	12528797	12528555	7,00E-123	99,59
BARC-PV-0005477	PV-5477	Pv09	[G/A]	254	3345461	3345208	6,00E-129	99,61
BARC-PV-0005480	PV-5480	Pv01	[G/A]	265	14913446	14913182	5,00E-135	99,62
BARC-PV-0005490	PV-5490	Pv03	[C/T]	261	24415788	24416048	8,00E-133	99,62

**Material Suplementar 3: Continuação.**

SNP	SNP <i>id</i> *	Crom.	SNP	Tamanho da sequência (pb)**	Posição inicial (pb)	Posição final (pb)	<i>E-value</i>	Identidade (%)
BARC-PV-0005500	PV-5500	Pv04	[C/G]	280	44466047	44466327	3,00E-142	99,29
BARC-PV-0005519	PV-5519	Pv11	[G/A]	268	34266169	34266436	1,00E-136	99,63
BARC-PV-0005524	PV-5524	Pv01	[G/A]	269	18198379	18198111	3,00E-137	99,63
BARC-PV-0005527	PV-5527	Pv09	[G/T]	256	3958312	3958567	5,00E-130	99,61
BARC-PV-0005528	PV-5528	Pv09	[G/A]	277	33823183	33822907	2,00E-138	98,56
BARC-PV-0005542	PV-5542	Pv03	[G/A]	241	8942293	8942053	9,00E-122	99,59
BARC-PV-0005546	PV-5546	Pv11	[G/A]	258	6369118	6369375	4,00E-131	99,61
BARC-PV-0005552	PV-5552	Pv06	[G/A]	272	7333065	7333333	8,00E-98	90,74
BARC-PV-0005555	PV-5555	Pv03	[T/A]	259	50421229	50420971	1,00E-131	99,61
BARC-PV-0005556	PV-5556	Pv02	[C/T]	284	13375643	13375926	1,00E-145	99,65
BARC-PV-0005569	PV-5569	Pv08	[A/G]	263	36885297	36885560	4,00E-131	98,86
BARC-PV-0005575	PV-5575	Pv07	[C/G]	245	17197111	17197355	6,00E-124	99,59
BARC-PV-0005606	PV-5606	Pv04	[C/T]	289	38209483	38209195	2,00E-143	98,62
BARC-PV-0005610	PV-5610	Pv11	[G/A]	263	28419464	28419202	1,00E-131	98,86
BARC-PV-0005645	PV-5645	Pv01	[C/T]	272	22620806	22621077	6,00E-139	99,63
BARC-PV-0005667	PV-5667	Pv04	[T/C]	275	38102036	38102286	8,00E-123	98,41
BARC-PV-0005702	PV-5702	Pv04	[G/A]	274	18592092	18592360	3,00E-137	99,63
BARC-PV-0005704	PV-5704	Pv08	[A/G]	265	33908717	33908456	2,00E-133	99,62
BARC-PV-0005710	PV-5710	Pv02	[T/G]	257	17584026	17583770	1,00E-130	99,61
BARC-PV-0005712	PV-5712	Pv05	[C/A]	261	17038310	17038050	8,00E-133	99,62
BARC-PV-0005717	PV-5717	Pv01	[A/C]	263	32458008	32457749	3,00E-132	99,62
BARC-PV-0005722	PV-5722	Pv03	[C/T]	241	13857933	13858173	9,00E-122	99,59
BARC-PV-0005724	PV-5724	Pv10	[G/A]	259	19391268	19391012	1,00E-130	99,61
BARC-PV-0005729	PV-5729	Pv11	[T/C]	282	27294862	27294581	2,00E-144	99,65
BARC-PV-0005732	PV-5732	Pv04	[G/A]	254	29723542	29723289	6,00E-129	99,61
BARC-PV-0005749	PV-5749	Pv10	[A/T]	265	12555057	12555321	5,00E-135	99,62
BARC-PV-0005755	PV-5755	Pv10	[G/C]	249	12848208	12848456	3,00E-126	99,60
BARC-PV-0005758	PV-5758	Pv03	[T/C]	240	21358704	21358943	3,00E-121	99,58
BARC-PV-0005764	PV-5764	Pv11	[A/T]	257	41450864	41451120	1,00E-130	99,61
BARC-PV-0005771	PV-5771	Pv10	[G/A]	257	10826301	10826557	1,00E-130	99,61
BARC-PV-0005778	PV-5778	Pv01	[C/G]	271	28274075	28274345	1,00E-136	99,26
BARC-PV-0005784	PV-5784	Pv01	[C/T]	264	23593926	23594189	2,00E-134	99,62
BARC-PV-0005793	PV-5793	Pv11	[T/C]	191	49341414	49341224	4,00E-94	99,48
BARC-PV-0005801	PV-5801	Pv10	[G/A]	251	12421800	12422050	3,00E-127	99,60
BARC-PV-0005856	PV-5856	Pv04	[A/C]	249	17898413	17898166	1,00E-125	99,60
BARC-PV-0005865	PV-5865	Pv08	[T/C]	266	47809804	47810062	1,00E-131	99,61
BARC-PV-0005874	PV-5874	Pv01	[A/G]	288	29271320	29271033	1,00E-146	99,31
BARC-PV-0005875	PV-5875	Pv01	[G/A]	260	11710436	11710695	1,00E-130	99,23
BARC-PV-0005878	PV-5878	Pv05	[C/A]	258	25557842	25558099	4,00E-131	99,61
BARC-PV-0005885	PV-5885	Pv06	[T/C]	309	21085874	21086182	2,00E-159	99,68
BARC-PV-0005886	PV-5886	Pv10	[C/T]	276	21354676	21354951	2,00E-109	92,75
BARC-PV-0005888	PV-5888	Pv02	[T/G]	252	10044878	10045129	7,00E-128	99,60
BARC-PV-0005890	PV-5890	Pv06	[T/A]	260	11794791	11795050	3,00E-132	99,62
BARC-PV-0005895	PV-5895	Pv01	[A/G]	272	33511571	33511300	6,00E-139	99,63
BARC-PV-0005909	PV-5909	Pv06	[G/T]	240	28122852	28123091	3,00E-121	99,58
BARC-PV-0005924	PV-5924	Pv09	[G/A]	269	4180925	4180657	3,00E-137	99,63
BARC-PV-0005942	PV-5942	Pv10	[A/C]	252	17545909	17545659	3,00E-127	99,60
BARC-PV-0005944	PV-5944	Pv07	[C/T]	266	18519506	18519241	1,00E-135	99,62
BARC-PV-0005992	PV-5992	Pv06	[T/C]	243	14745441	14745683	7,00E-123	99,59
BARC-PV-0006051	PV-6051	Pv09	[C/A]	262	2293751	2293490	2,00E-133	99,62
BARC-PV-0006061	PV-6061	Pv10	[A/G]	266	18587835	18587570	1,00E-135	99,62
BARC-PV-0006077	PV-6077	Pv06	[C/G]	269	297535	297267	3,00E-137	99,63
BARC-PV-0006098	PV-6098	Pv08	[G/A]	257	14021144	14020889	6,00E-129	99,22

**Material Suplementar 3: Continuação.**

SNP	SNP <i>id</i> *	Crom.	SNP	Tamanho da sequência (pb)**	Posição inicial (pb)	Posição final (pb)	<i>E-value</i>	Identidade (%)
BARC-PV-0006107	PV-6107	Pv10	[T/C]	284	12082318	12082601	3,00E-142	98,59
BARC-PV-0006116	PV-6116	Pv10	[A/C]	264	23759470	23759207	2,00E-134	99,62
BARC-PV-0006124	PV-6124	Pv01	[G/C]	229	35727729	35727501	2,00E-113	99,13
BARC-PV-0006129	PV-6129	Pv07	[C/T]	269	18243501	18243769	3,00E-137	99,63
BARC-PV-0006148	PV-6148	Pv08	[A/G]	259	32112722	32112467	2,00E-123	98,05
BARC-PV-0006196	PV-6196	Pv02	[T/G]	260	21913477	21913218	3,00E-132	99,62
BARC-PV-0006205	PV-6205	Pv06	[G/A]	260	26968242	26968500	1,00E-130	99,23
BARC-PV-0006218	PV-6218	Pv03	[A/G]	277	23640774	23640498	5,00E-140	99,28
BARC-PV-0006235	PV-6235	Pv10	[T/C]	253	18586244	18586496	2,00E-128	99,60
BARC-PV-0006238	PV-6238	Pv08	[T/C]	266	38694759	38695021	1,00E-130	98,86
BARC-PV-0006245	PV-6245	Pv01	[C/T]	246	39654920	39655165	2,00E-124	99,59
BARC-PV-0006268	PV-6268	Pv06	[C/T]	270	13971871	13971608	8,00E-133	99,24
BARC-PV-0006276	PV-6276	Pv03	[C/G]	243	18433246	18433004	7,00E-123	99,59
BARC-PV-0006292	PV-6292	Pv08	[A/G]	187	22736133	22736319	3,00E-90	98,93
BARC-PV-0006295	PV-6295	Pv09	[A/G]	266	17233042	17232777	1,00E-135	99,62
BARC-PV-0006304	PV-6304	Pv03	[G/A]	228	26411626	26411399	7,00E-113	99,12
BARC-PV-0006352	PV-6352	Pv11	[C/G]	247	27200954	27200708	4,00E-125	99,60
BARC-PV-0006356	PV-6356	Pv11	[G/A]	271	20515685	20515955	1,00E-136	99,26
BARC-PV-0006364	PV-6364	Pv06	[C/T]	234	6675806	6676039	7,00E-118	99,57

\* A nomenclatura "BARC-PV-000XXXX" dos SNPs derivados do trabalho de Hyten et al. (2010) foi reduzida para "PV-XXXX" neste estudo (SNP id).

\*\* As sequências dos *primers* ASO estão omitidas devido ao longo tamanho (média de 250 pb).

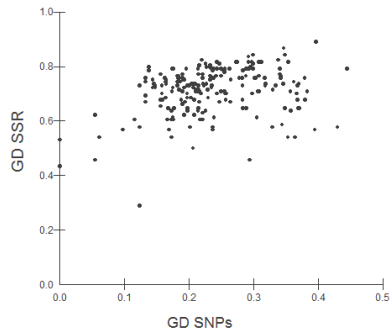
**Material Suplementar 4:** Estatística descritiva detalhada dos 24 SSRs-di, 34 BES-SSRs e 345 SNPs caracterizados em 88 genótipos de feijoeiro comum:

		Grupo	Tamanho da amostra	$F_{máx}$	Alelos	$nA$	Alelos privados	$H_E$	$H_O$	$F_{IS}$	$PI$
SSRs	Andino	Melhorado	10	0,653	78	3,25	13	0,408	0,000	1,000	1,30E-13
		Tradicional	11	0,629	93	3,88	21	0,433	0,087	0,831	1,06E-14
	Mesoamericano	Melhorado	45	0,536	174	7,25	61	0,581	0,027	0,955	2,18E-20
		Tradicional	22	0,596	100	4,17	14	0,474	0,038	0,927	3,95E-15
	<b>Total</b>		<b>88</b>	<b>0,405</b>	<b>238</b>	<b>9,92</b>	<b>106</b>	<b>0,712</b>	<b>0,034</b>	<b>0,953</b>	<b>3,33E-26</b>
BES-SSRs	Andino	Melhorado	10	0,778	76	2,24	10	0,331	0,003	0,991	1,44E-11
		Tradicional	11	0,862	64	1,88	8	0,193	0,000	1,000	1,10E-7
	Mesoamericano	Melhorado	45	0,708	148	4,35	54	0,423	0,003	0,993	1,97E-16
		Tradicional	22	0,783	88	2,59	11	0,297	0,010	0,968	2,10E-11
	<b>Total</b>		<b>88</b>	<b>0,641</b>	<b>190</b>	<b>5,59</b>	<b>83</b>	<b>0,481</b>	<b>0,004</b>	<b>0,991</b>	<b>3,75E-20</b>
SNPs	Andino	Melhorado	10	0,945	442	1,28	0	0,072	0,027	0,685	1,20E-24
		Tradicional	11	0,959	403	1,17	0	0,049	0,023	0,609	7,60E-17
	Mesoamericano	Melhorado	45	0,896	677	1,96	3	0,155	0,022	0,867	1,89E-49
		Tradicional	22	0,940	453	1,31	0	0,082	0,022	0,756	7,21E-27
	<b>Total</b>		<b>88</b>	<b>0,721</b>	<b>690</b>	<b>2,00</b>	<b>3</b>	<b>0,382</b>	<b>0,022</b>	<b>0,943</b>	<b>4,67E-119</b>

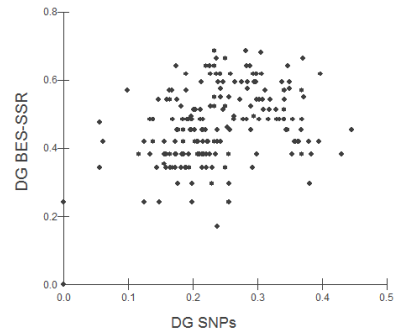
Descritores abreviados: estimativa da maior frequência alélica ( $F_{máx}$ ), número de alelo por loco ( $nA$ ), diversidade gênica ( $H_E$ ), heterozigiosidade observada ( $H_O$ ), índice de endogamia ( $F_{IS}$ ) e probabilidade de identidade ( $PI$ ).

**Material Suplementar 5:** Correlação da distância genética pelo método de Pearson baseado nas comparações entre os marcadores moleculares par a par: 1- SSRs-di *versus* SNPs, 2- BES-SSRs *versus* SNPs, 3- BES-SSRs/SSRs-di *versus* SNPs e 4- SSRs-di *versus* BES-SSRs. A1-A4 para os genótipos Andinos e B1-B4 para genótipos Mesoamericanos.

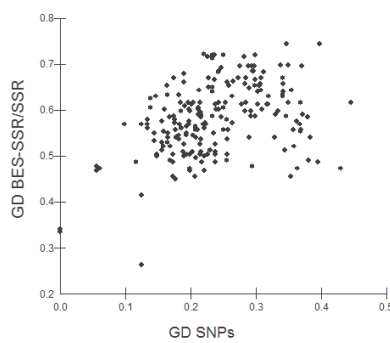
**A1)**  $r = 0,29$  (IC95% 0,16 a 0,40)  $p < 0,0001$



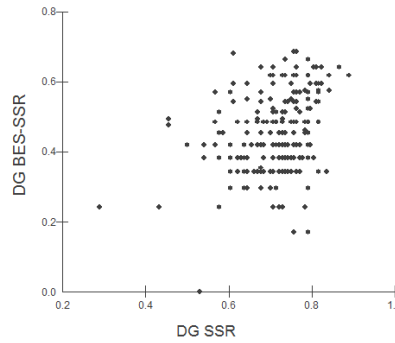
**A2)**  $r = 0,37$  (IC95% 0,25 a 0,48)  $p < 0,0001$



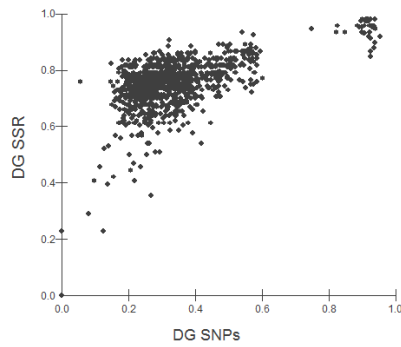
**A3)**  $r = 0,40$  (IC95% 0,28 a 0,51)  $p < 0,0001$



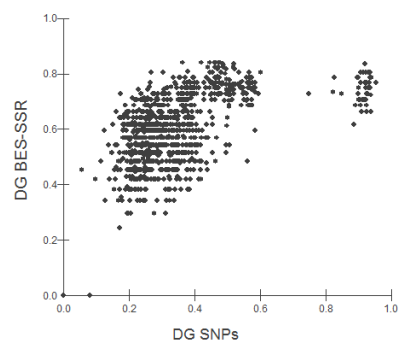
**A4)**  $r = 0,32$  (IC95% 0,19 a 0,43)  $p < 0,0001$



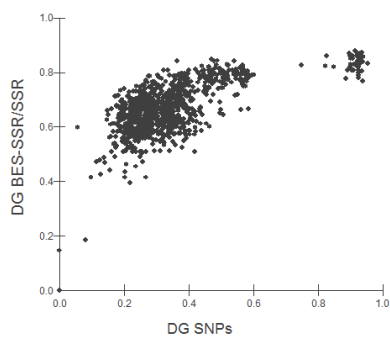
**B1)**  $r = 0,62$  (IC95% 0,58 a 0,65)  $p < 0,0001$



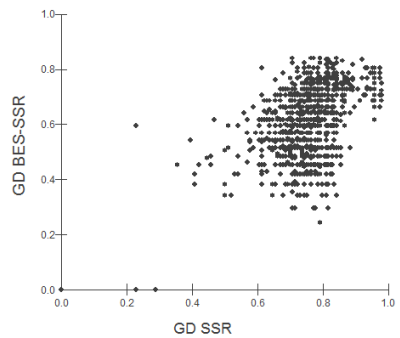
**B2)**  $r = 0,52$  (IC95% 0,48 a 0,57)  $p < 0,0001$



**B3)**  $r = 0,66$  (IC95% 0,62 a 0,69)  $p < 0,0001$

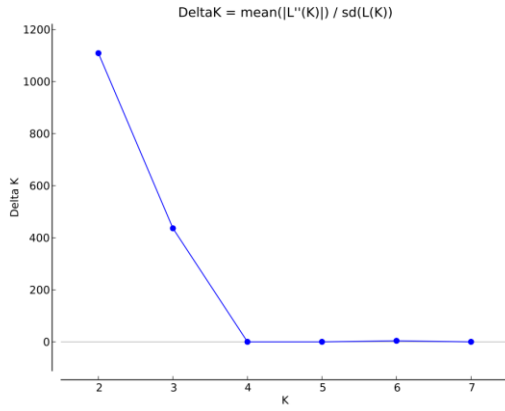


**B4)**  $r = 0,46$  (IC95% 0,41 a 0,51)  $p < 0,0001$

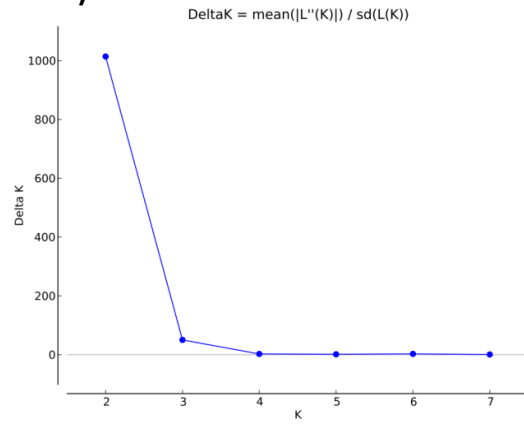


**Material Suplementar 6:** Representação gráfica das estatísticas de  $\Delta K$  através da estimativa da média  $L(K)$  para determinar o  $K$  mais provável entre os *clusters* obtidos pelo STRUCTURE para os marcadores : A- SSRs-di, B- BES-SSRs, C- BES-SSRs/SSRs-di e D- SNPs.

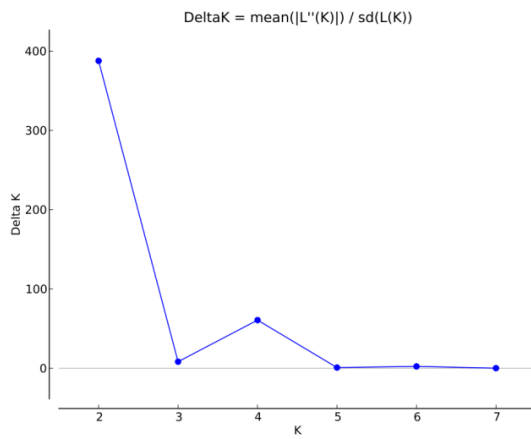
**A)**



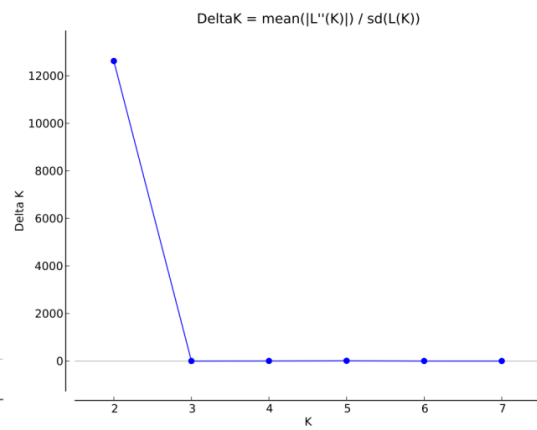
**B)**



**C)**



**D)**



## CONCLUSÕES GERAIS

- A análise de BAC-*end sequences* de *Phaseolus vulgaris*, variedade BAT 93 (Mesoamericana), possibilitou acessar e descrever aspectos pouco estudados do genoma da espécie, além de ter mostrado ser uma estratégia eficiente para se derivar informações moleculares úteis para aplicações nas áreas de genética e melhoramento.
- Através de sequências genômicas representativas de 6% do genoma de *P. vulgaris* foi estimada a presença de 2% de fatores de transcrição, 3% de elementos transponíveis e 46,53% de sequências associadas a genes.
- As sequências de BES-SSRs, fatores de transcrição e elementos transponíveis apresentaram notável similaridade entre os genomas de outras leguminosas modelo (*Medicago truncatula*) e/ou de interesse agrônomico (*Glycine max* e *P. vulgaris*, variedade Andina).
- A mineração por BES-SSRs baseados em motivos longos ( $\geq$  trinucleotídeos) e compostos permitiu quantificar a riqueza e variedade de 2.000 marcadores, os quais nunca haviam sido desenvolvidos e caracterizados quanto ao potencial de uso em feijoeiro comum.
- Foram identificados SSRs tetra-, penta- e hexanucleotídeos com eficiência semelhante aos dinucleotídeos, além do diferencial quanto à qualidade analítica dos dados gerados, conferindo maior confiabilidade na declaração dos alelos identificados.
- Marcadores SSRs adicionais baseados em repetições  $\geq$  trinucleotídeos encontram-se disponíveis no banco de SSRs, gerado a partir deste estudo, possibilitando a triagem quanto ao polimorfismo e seleção de locos adicionais para otimizar novos sistemas de genotipagem em feijoeiro comum.
- As sequências, geradas neste estudo, foram de grande contribuição para o desenvolvimento de marcadores SSRs e para o maior conhecimento do genoma, sendo integradas ao projeto internacional de sequenciamento total do genoma de feijoeiro comum.
- A caracterização molecular de 58 SSRs e 345 SNPs, utilizando uma mesma amostra de genótipos, permitiu a comparação dos parâmetros genéticos

gerados entre essas classes de marcadores e, também, permitiu inferir sobre o melhor potencial de uso em estudos genéticos com o feijoeiro comum.

- Os marcadores SSRs-di destacaram-se dos BES-SSRs e SNPs pela hipervariabilidade alélica e mostraram-se mais adequados para capturar a diversidade genética existente entre acessos de germoplasma melhorado e tradicional, que compartilhem um maior grau de similaridade genética.

- Apesar dos SSRs e SNPs apresentarem o mesmo potencial de estruturação dos acessos de feijoeiro comum, com base em seus centros de origem Andino e Mesoamericano, os SNPs foram mais eficientes quanto ao poder de diferenciação genética entre esses grupos de acessos.

- Foram selecionados 16 marcadores SSRs, com elevado poder de discriminação do germoplasma de feijoeiro comum de origem Andina e Mesoamericana, que serão incorporados às rotinas de genotipagem molecular do Banco Ativo de Germoplasma (BAG) da Embrapa Arroz e Feijão para fins de triagem dos acessos quanto às suas origens.

- Embora os marcadores SSRs sejam consideravelmente mais informativos em termos de loco individual, eles apresentam um custo por dado gerado (*data point*) mais elevado da ordem de 3X, além de não apresentarem potencial de genotipagem simultânea com um grande número de marcadores (*capacidade multiplex*).

- O conjunto de SNPs polimórficos, em oito composições biparentais, indica que o ensaio OPA caracterizado neste estudo fornece centenas de marcadores polimórficos para a geração de mapas genéticos com uma proporção considerável de SNPs compartilhados entre cruzamentos.