

Análise da Diversidade Genética e Estruturação Populacional entre Marcadores SSRs e SNPs em Germoplasma de Feijoeiro Comum (*Phaseolus vulgaris* L.)

Bárbara Müller Salomão de Faria¹, Ivandilson Pessoa Pinto de Menezes², Georgios Joannis Pappas Júnior³, Paula Arielle Mendes Ribeiro Valdisser⁴, Gesimária Ribeiro Costa Coelho⁴, Tereza Cristina de Oliveira Borba⁵, Aluana Gonçalves de Abreu⁵, Claudio Brondani⁵, Everaldo Gonçalves de Barros⁶, Rosana Pereira Vianello⁵

Resumo

O objetivo desse estudo foi avaliar o poder de informação genética dos marcadores SSRs e SNPs para aplicações no melhoramento genético do feijoeiro comum. O DNA genômico foliar de 88 genótipos (55 cultivares e 33 *landraces*) foi extraído para caracterização molecular da diversidade e estruturação genética de 24 SSRs-di, 34 BES-SSRs e 345 SNPs. Os SSRs-di apresentaram maior média de alelos por loco gênico (9,91) e revelaram maior diversidade genética média (0,721), sendo o grupo de marcadores com o maior poder de discriminação genética entre indivíduos. Um conjunto de 14 SSRs (10 SSRs-di e quatro BES-SSRs) com média de alelos privativos de 14,93 por loco e H_E média de 0,84 possibilitou a discriminação de todos os genótipos avaliados. As análises de agrupamento (PCoA, STRUCTURE e dendrogramas) e a análise de variância molecular apresentaram resultados similares, para cada grupo de marcador, mostrando a estruturação dos genótipos em relação à origem do feijoeiro comum nos *pools* gênicos Mesoamericano e Andino, sendo que os SSRs-di ainda possibilitaram a discriminação entre cultivares e *landraces* dentro do *pool* gênico Mesoamericano. Já os SNPs proporcionaram uma resolução maior na discriminação dos genótipos compreendidos nos diferentes *pools* gênicos. Ambas as classes de marcadores avaliadas mostraram-se adequadas para análise genética do feijoeiro comum e, embora a natureza bialélica dos SNPs faça deles menos informativos, o grande número de marcadores simultaneamente genotipados resulta em estimativas de parâmetros genéticos adequadas para uma aplicação operacional no melhoramento do feijoeiro comum.

Introdução

O feijoeiro comum (*Phaseolus vulgaris* L.) é a espécie mais cultivada do gênero *Phaseolus* e mais importante para o consumo direto humano, entre as leguminosas (Broughton et al. 2003). A aplicação de marcadores moleculares no melhoramento de plantas atualmente está em franca expansão. A obtenção de marcadores reprodutíveis, com elevado poder de discriminação e que apresentem um satisfatório padrão de amplificação em *background* genético diverso do feijoeiro é um processo contínuo e dependente de uma adequada seleção dos marcadores que irão compor os sistemas de genotipagem.

Os microssatélites (*Simple Sequence Repeats* - SSRs) 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). Na última década, diversos grupos de pesquisa vêm avançando no desenvolvimento de marcadores SSRs para o feijoeiro comum. Atualmente estão disponíveis mais de 2000 marcadores originados a partir do sequenciamento genômico total ou parcial, bem como de bancos de sequências expressas (Garcia et al. 2011).

Os SNPs (*Single Nucleotide Polymorphisms*) são as mais frequentes formas de variação genética de um organismo, ocorrendo em abundância ao longo de todo o genoma (Ching et al. 2002) e caracterizam-se por serem predominantemente bialélicos. Com os avanços das tecnologias de sequenciamento e da computação 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 o feijoeiro comum (Galeano et al. 2012).

O objetivo deste estudo foi avaliar o potencial dos marcadores SSRs e SNPs em caracterizar a magnitude da diversidade genética de um grupo de acessos de feijoeiro comum e, comparativamente, determinar o

1 Mestranda em Genética e Melhoramento – UFV/Viçosa-MG. E-mail: mullerbsf@gmail.com

2 Professor do Departamento de Biologia – IFG/Urutaí-GO. E-mail: ivan.menezes@ifgoiano.edu.br

3 Professor do Departamento de Biologia Celular – UnB/Brasília-DF. E-mail: gpappas@unb.br

4 Analista da Embrapa Arroz e Feijão – CNPAF/Santo Antônio de Goiás-GO. E-mail: paula.valdisser@embrapa.br

5 Pesquisador da Embrapa Arroz e Feijão – CNPAF/Santo Antônio de Goiás-GO. E-mail: rosana.vianello@embrapa.br

6 Professor Titular do Departamento de Biologia Geral – UFV/Viçosa-MG. E-mail: everaldodebarros@gmail.com

melhor sistema de genotipagem quanto ao poder de informação e facilidade de operacionalização para serem empregados de modo rotineiro por programas de melhoramento genético do feijoeiro comum.

Material e Métodos

O DNA genômico de 88 genótipos de feijoeiro comum, compreendendo 55 linhagens comerciais (cultivares) lançadas por instituições de pesquisa do Brasil e do exterior e 33 variedades tradicionais (*landraces*) foi extraído a partir de amostras do tecido foliar de acordo com o método do CTAB, conforme descrito por Ferreira and Grattapaglia (1998).

Um conjunto de 58 marcadores SSRs desenvolvidos para o feijoeiro comum, compreendendo 24 dinucleotídeos e 34 BES-SSRs (*BAC-end sequences-SSRs*) contendo repetições do tipo tri-; tetra-; penta-; hexanucleotídeo ou compostos foram analisados através de sistemas de genotipagem multiloco reunindo de três a nove SSRs. Após a amplificação dos locos SSRs por PCR, os fragmentos foram separados via eletroforese capilar conduzida na plataforma ABI3100 *Genetic Analyzer* (Applied Biosystems). A análise dos fragmentos foi realizada utilizando os programas *DataCollection v2.0* e *GeneMapper v4.1* (Applied Biosystems).

Um conjunto de 384 SNPs, identificados a partir do alinhamento do genoma total da linhagem BAT93 (Mesoamericano) com os dados de re-sequenciamento do genótipo Jalo EEP558 (Andino), foi utilizado neste estudo. A análise de marcadores SNPs foi realizada em plataforma Illumina *BeadxPress* (Kim and Misra 2007) utilizando a tecnologia *VeraCode™* (Illumina, 2012). A genotipagem dos SNPs foi realizada por meio do programa *GenomeStudio v1.8.4* (Illumina), de acordo com os seguintes parâmetros: *Gen Call Threshold* de 0,25; *Call Rate* variando de 0,6463 a 1,00 e *GenTrain* $\geq 0,2535$.

A diversidade genética total da amostra foi analisada através das estimativas da frequência alélica, número de alelos por loco, heterozigosidade esperada (H_E) e observada (H_O) usando o programa *PowerMarker v2.25* (Liu and Muse 2005). A identificação dos alelos privados e a análise de desequilíbrio de ligação foram realizadas através do programa *GenAlex v6.5* (Peakall and Smouse 2012). A matriz de distância genética de Roger's modificada (MRD) foi determinada para cada conjunto de marcadores. O índice de fixação (F_{IS}) através da amostra total foi estimado de acordo com a estatística de Wright (Weir and Cockerham 1984) com intervalo de confiança de 95% para 10.000 *bootstraps* usando o *PowerMarker v2.25* (Liu and Muse, 2005).

A estrutura genética dos genótipos de feijoeiro comum foi analisada através da *Principal Coordinates Analysis* (PCoA), com base na MRD, pelo programa *GenAlex v6.5* (Peakall and Smouse, 2012). O programa *STRUCTURE v2.2.4* (Pritchard et al. 2000) foi utilizado para inferir sobre a estruturação populacional dos genótipos estudados, através de agrupamento baseado no modelo Bayesiano. Para determinar o número de grupos genéticos utilizou-se o critério proposto por Evanno et al. (2005), através do programa *Structure Harvester v0.6.93* (Earl and VonHold, 2012). A Análise de Variância Molecular (AMOVA) foi realizada para testar a estrutura da diversidade genética dos genótipos, através do programa *Alerquin v3.5* (Excoffier and Lischer, 2010), com 10.000 permutações. 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).

Resultados e Discussão

Os 24 SSRs-di (SSRs dinucleotídeos), os 34 BES-SSRs (SSRs tri-, tetra-, penta-, hexa- e compostos) e os 345 SNPs polimórficos satisfizeram a premissa de independência de acordo com teste de desequilíbrio de ligação. A maior estimativa de número médio de alelos detectada por loco gênico foi verificada para os SSRs baseados em repetições do tipo dinucleotídeo (9,91 alelos), seguida pelos BES-SSRs (5,58) e os SNPs com, no máximo, dois alelos por loco gênico.

Com relação aos valores de diversidade genética média (H_E), as estimativas foram de 0,721 para os SSRs-di, com destaque para o loco BM154 ($H_E = 0,926$ e 22 alelos), seguida pelos BES-SSRs ($H_E = 0,487$) e SNPs ($H_E = 0,387$). A elevada heterozigosidade esperada na maioria dos locos SSRs-di analisados indica que, em conjunto, eles possuem um alto poder de discriminação individual para o grupo de genótipos avaliados. As estimativas de heterozigosidade observadas (H_O) foram de 0,0043; 0,022 e 0,034 para BES-SSRs, SNPs e SSRs-di, respectivamente, corroborando os elevados índices de endogamia (F_{IS} ou f) de 0,991; 0,943 e 0,953, reforçando a tendência de homozigose para os locos analisados, conforme o esperado para espécies autógamas. O maior número de alelos privados identificados foi de 221 para os SNPs, seguido de 118 para

os SSRs-di, sendo ambos identificados entre os genótipos Mesoamericanos (Tabela 1).

A Probabilidade de Identidade (PI) combinada para os 24 locos dinucleotídeos ($3,37 \times 10^{-26}$), para os 34 BES-SSRs ($3,61 \times 10^{-20}$) e para os SNPs ($4,65 \times 10^{-119}$) correspondem à probabilidade de que duas amostras tomadas ao acaso tenham genótipos idênticos (Tabela 1). O valor da PI combinada para os SNPs foi muito menor, evidenciado pelo grande número de marcadores utilizados, proporcionando um poder discriminatório bastante superior quando comparado aos SSRs. Considerando todos os microssatélites avaliados neste estudo, a PI combinada para os locos que apresentavam 10 ou mais alelos privados e H_E acima de 0,7 foi estimada em $1,08 \times 10^{-20}$. Um total de 14 locos SSRs foi selecionado para esta estimativa de PI combinada, sendo 10 locos representados pelos SSRs-di e quatro locos por BES-SSRs. Estes 14 SSRs, com média de alelos privados de 14,93 por loco e H_E média de 0,84; possibilitam a discriminação de todos os 88 genótipos de feijoeiro avaliados.

Tabela 1: Descritores genéticos dos locos de microssatélites e SNPs de feijoeiro comum analisados com suas respectivas estimativas obtidas considerando a origem e o tipo de germoplasma, incluindo a frequência alélica máxima ($F_{m\acute{a}x}$), o número de alelo por loco (A), o número de alelos privados (A_p) diversidade gênica (H_e); heterozigosidade observada (H_o), índice de fixação (f) e probabilidade de identidade (PI).

	Grupos	$F_{m\acute{a}x}$	A	A_p	H_E	H_o	f	PI
SSRs-di	Andino	0,590	5,000	44	0,522	0,045	0,916	4,93E-17
	Mesoamericano	0,501	8,083	118	0,619	0,031	0,951	2,60E-21
	Cultivar	0,449	8,375	82	0,679	0,022	0,968	8,60E-25
	Landrace	0,448	6,500	37	0,647	0,054	0,921	3,80E-22
	Total	0,405	9,916	-	0,721	0,034	0,953	3,37E-26
BES-SSRs	Andino	0,805	2,735	28	0,277	0,0015	0,995	5,87E-11
	Mesoamericano	0,730	4,765	97	0,393	0,0051	0,987	1,06E-15
	Cultivar	0,647	5,000	73	0,478	0,003	0,994	7,05E-20
	Landrace	0,641	3,440	20	0,433	0,007	0,986	8,87E-18
	Total	0,641	5,588	-	0,487	0,0043	0,991	3,61E-20
SNPs	Andino	0,951	1,360	13	0,071	0,025	0,665	3,24E-22
	Mesoamericano	0,908	1,962	221	0,141	0,022	0,847	1,07E-43
	Cultivar	0,755	2,000	9	0,357	0,023	0,937	2,10E-109
	Landrace	0,673	1,970	0	0,427	0,022	0,950	2,40E-127
	Total	0,721	2,000	-	0,387	0,022	0,943	4,65E-119

A distribuição dos indivíduos por meio da análise de PCoA foi consistente entre os grupos de marcadores SSRs e SNPs revelando a estruturação dos genótipos de feijoeiro comum em relação aos seus centros de origem, Mesoamericano e Andino. A parcela da variação explicada pelos primeiros eixos variou de 46,1 a 89,6% para marcadores de *loci* SSRs e SNPs, respectivamente, indicando claramente que apesar do menor conteúdo de informação genética por loco gênico, os SNPs genotipados em grande número possibilitaram uma melhor determinação da divergência genética entre grupos.

Através da análise de estruturação implementada pelo STRUCTURE, a partir dos resultados de ΔK , observou-se que $K = 2$ foi o *cluster* mais representativo para todos os grupos de marcadores, representando o agrupamento dos genótipos conforme aos seus centros de origem. Para os SSRs-di, o $K = 3$ mostra que os indivíduos pertencentes ao grupo Mesoamericano subdivide em cultivares e *landraces*, assim como na análise de PCoA. Esses resultados indicam uma relação bastante estreita entre o germoplasma tradicional e cultivado revelando a ocorrência de hibridização entre os mesmos levando a uma homogeneização genética entre os acessos. A capacidade de detecção de um grande número de alelos por loco gênico, conforme atribuído para os SSRs-di, favoreceu a identificação das diferenças genética possibilitando a determinação de um número maior de grupos. As informações baseadas nas MRD utilizadas na construção de dendrogramas de *Neighbour-joining*, para cada grupo de marcador, mostraram a formação dos dois agrupamentos principais para feijoeiro comum, representando 21 genótipos Andinos e 67 Mesoamericanos, assim como os outros resultados de agrupamento.

A AMOVA referente às estimativas de diferenciação entre os *polls* gênicos revelou índices variando de 36,5% para SSRs-di; 49,6% para BES-SSRs e 87,7% para os SNPs. Apesar das duas classes de marcadores utilizadas terem sido adequadas para acessar a diferenciação genética entre os grupos de gemoplasmata avaliados, os SNPs possibilitaram uma maior diferenciação entre os grupos amostrados quando comparados aos SSRs. O número seis vezes maior de SNPs utilizados, bem como o fato deles serem derivados do contraste entre um genótipo Andino e Mesoamericano explica o elevado poder discriminatório quanto à origem desses acessos.

O grande poder de informação dos SSRs os qualifica para obter uma boa representação da diversidade genética existente em feijoeiro comum. O número crescente de marcadores SNPs decorrentes do sequenciamento do genoma de *P. vulgaris* possibilitará o estabelecimento de painéis de genotipagem contendo centenas de marcadores, mais reprodutíveis e informativos, adequados para a análise genética rotineira do feijoeiro comum a custos acessíveis.

Apoio

Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq e Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES.

Referências

- Broughton WJ, Hern G, Blair M, et al. (2003) Beans (*Phaseolus* spp.) – model food legumes. **Plant and Soil** **252**: 55–128.
- Ching A, Caldwell KS, Jung M et al. (2002) SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. **BMC Genetics** **3(19)**: 1471-2156.
- Earl DA and VonHoldt BM (2011) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. **Conservation Genetics Resources** **4(2)**: 359–361.
- Evanno G, Regnaut S and Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. **Molecular Ecology** **14(8)**: 2611–2620.
- Excoffier L and Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. **Molecular ecology resources** **10(3)**: 564–567.
- Ferreira ME and Grattapaglia D (1998) **Introdução ao uso de marcadores moleculares em análise genética**. 2 ed. Embrapa Cenargen, Brasília, 219p.
- Galeano CH, Cortés AJ, Fernández AC et al (2012) Gene-Based Single Nucleotide Polymorphism Markers for Genetic and Association Mapping in Common Bean. **BMC Genetics** **13**:48
- Garcia RAV, Rangel PN, Brondani C et al. (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 Genetics** **12**:41.
- Kim S and Misra A (2007) SNP genotyping: Technologies and biomedical applications. **Annual Review of Biomedical Engineering** **9**: 289-320.
- Li YC, Korol AB, Fahima T, Nevo E (2004) Microsatellites within genes: structure, function, and evolution. **Molecular Biology and Evolution** **21**:991–1007.
- Liu K and Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. **Bioinformatics** **21(9)**: 2128–2129.
- Peakall R and Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. **Bioinformatics** **28(19)**: 2537–2539.
- Perrier X and Jacquemoud-Collet JP (2006) **DARwin: Dissimilarity Analysis and Representation for Windows**. Version 5.0.158. Available at <http://darwin.cirad.fr/darwin>. Accessed on Apr 4, 2013.
- Pritchard JK, Stephens M and Donnelly P (2000) Inference of population structure using multilocus genotype data. **Genetics** **155(2)**: 945–959.
- Saitou N and Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. **Molecular Biology and Evolution** **4(4)**: 406–425.
- Souza TLPO, Barros EG, Bellato CM et al. (2012) Single nucleotide polymorphism discovery in common bean. **Molecular Breeding** **30**:419–428
- Weir BS and Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. **Evolution** **38**:1358-1370.