**Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”**

**Diversidade genética entre acessos cultivados de feijão comum
(*Phaseolus vulgaris* L.): uma abordagem *in silico* a partir dos genes
 α -Phs e FRO1**

Augusto Lima Diniz

Dissertação apresentada para obtenção do título de
Mestre em Ciências. Área de concentração:
Genética e Melhoramento de Plantas

**Piracicaba
2012**

Augusto Lima Diniz
Licenciado e Bacharel em Ciências Biológicas

Diversidade genética entre acessos cultivados de feijão comum (*Phaseolus vulgaris* L.): uma abordagem *in silico* a partir dos genes α -Phs e FRO1

Orientador:
Profa. Dra. **MARIA LUCIA CARNEIRO VIEIRA**

Dissertação apresentada para obtenção do título de
Mestre em Ciências. Área de concentração:
Genética e Melhoramento de Plantas

**Piracicaba
2012**

**Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - ESALQ/USP**

Diniz, Augusto Lima

Diversidade genética entre acessos cultivados de feijão comum (*Phaseolus vulgaris* L.): uma abordagem *in silico* a partir dos genes α -Phs e *FRO1* / Augusto Lima Diniz. - Piracicaba, 2012.
90 p: il.

Dissertação (Mestrado) - - Escola Superior de Agricultura "Luiz de Queiroz", 2012.

1. Absorção 2. Aminoácidos 3. Feijão 4. Ferro 5. Diversidade genética 6. Polimorfismo
7. Sequenciamento genético I. Título

CDD 635.652
D585d

"Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor"

*Aos meus pais Diniz e Azilcia,
Pelo exemplo de vida e constante apoio
Dedico*

*A Deus,
Pelo dom da vida
Ofereço*

AGRADECIMENTOS

À Escola Superior de Agricultura “Luiz de Queiroz” (ESALQ/USP) e ao Departamento de Genética, pela oportunidade de realizar o curso de Mestrado.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela bolsa concedida.

À professora Dra Maria Lucia Carneiro Vieira, pela recepção, apoio, pelos ensinamentos e, principalmente, pela confiança em mim depositada.

Aos professores do programa de pós-graduação em Genética e Melhoramento de plantas, pelos ensinamentos e colaboração.

À doutoranda Luciane Santini pela paciência e pelo auxílio durante o planejamento e desenvolvimento deste trabalho.

Ao Dr Alisson Fernando Chiorato e à Dra Luciana Larsy Benchimol, do Instituto Agronômico de Campinas; e à professora Dra Siu Mui Tsai, do Centro de Energia Nuclear na Agricultura, pelo fornecimento das sementes dos acessos usados.

À Dra Maria Imaculada Zucchi, do Pólo Regional Centro Sul APTA, pelos conhecimentos compartilhados, treinamento e auxílio na análise dos dados.

Aos doutorandos Guilherme da Silva Pereira e Francynês Oliveira Macedo, pela amizade e companheirismo desde a graduação.

Aos amigos do Laboratório de Genética Molecular de Plantas Cultivadas, Alessandra, Anselmo, Bruno, Carla, Carlos, Carmelice, Cássia, Helen, Luciane, Taislene e Zirlane, por toda colaboração, apoio e amizade.

Ao técnico Carlos Alberto de Oliveira, pelo companheirismo e suporte oferecido.

Aos meus familiares e amigos que sempre me apoiaram e incentivaram.

A todos com quem convivi durante todos esses anos que, de alguma forma, contribuíram em minha formação e para a realização deste trabalho.

SUMÁRIO

RESUMO.....	09
ABSTRACT.....	11
LISTA DE FIGURAS.....	13
LISTA DE TABELAS.....	17
1 INTRODUÇÃO.....	19
2 REVISÃO BIBLIOGRÁFICA.....	23
2.1 A cultura do feijão e sua importância para a alimentação humana.....	23
2.2 Origem e domesticação de <i>Phaseolus vulgaris</i>	25
2.3 A faseolina.....	29
2.4 Absorção de Ferro (Fe) em plantas.....	30
2.5 Polimorfismos moleculares e diversidade genética em <i>P. vulgaris</i>	33
2.5.1 SNPs em sequências gênicas de <i>P. vulgaris</i>	34
3 MATERIAL E MÉTODOS.....	37
3.1 Material Vegetal.....	37
3.2 Extração e quantificação do DNA genômico.....	37
3.3 Prospecção <i>in silico</i> das sequências de <i>Phs</i> e <i>FRO1</i> de <i>Phaseolus vulgaris</i> , desenho e síntese de <i>primers</i>	38
3.4 Obtenção dos Fragmentos Internos do Gene (FIG).....	39
3.4.1 Condições de amplificação e detecção dos FIGs.....	39
3.4.2 Purificação e sequenciamento dos FIGs.....	41
3.5 Obtenção dos Fragmentos das Extremidades do Gene (FEG).....	41
3.5.1 Condições de amplificação e detecção dos FEGs.....	43
3.5.2 Seleção, purificação e clonagem dos FEGs.....	44
3.5.3 <i>Miniprep</i> e sequenciamento.....	45
3.6 Análise dos dados.....	46
3.6.1 Processamento e alinhamento das sequências.....	46
3.6.2 Verificação de saturação e escolha do modelo evolutivo.....	47
3.6.3 Parâmetros moleculares e medidas de variabilidade genética.....	47
3.6.4 Caracterização do polimorfismo nos genes α - <i>Phs</i> e <i>FRO1</i>	48
3.6.5 Metodologia de agrupamento.....	48
3.6.6 Testes de neutralidade seletiva de locos e análise da variância molecular (AMOVA).....	48

4 RESULTADOS E DISCUSSÃO.....	51
4.1 Prospecção <i>in silico</i> das sequências de <i>Phs</i> e <i>FRO1</i> de <i>P. vulgaris</i> e desenho de <i>primers</i>	51
4.2 Obtenção dos FIGs.....	52
4.2.1 Validação dos FIGs.....	54
4.3 Obtenção dos FEGs.....	56
4.4 Caracterização do polimorfismo.....	60
4.5 Predição das alterações nas sequências de aminoácidos.....	67
4.6 Verificação de saturação e escolha do modelo evolutivo.....	71
4.7 Diversidade genética entre os grupos gênicos de feijão.....	72
4.8 Diversidade nucleotídica e testes de neutralidade.....	78
5 CONCLUSÕES.....	81
REFERÊNCIAS.....	83

RESUMO

Diversidade genética entre acessos cultivados de feijão comum (*Phaseolus vulgaris* L.): uma abordagem *in silico* a partir dos genes α -Phs e FRO1

Análises de diversidade em feijão comum (*Phaseolus vulgaris* L.), envolvendo caracteres morfológicos e marcadores moleculares, têm mostrado uma estruturação em dois *pools* gênicos principais, um Mesoamericano e outro Andino, os quais diferem, dentre vários aspectos, quanto à morfologia e fisiologia do grão. Uma nova abordagem que vem sendo usada para estimar a variabilidade genética entre acessos de feijoeiro é a prospecção de polimorfismos de base única (SNPs), tanto em sequências arbitrárias do genoma quanto em sequências gênicas de interesse. Também, o estudo de sequências nucleotídicas que codificam proteínas importantes, tais como a faseolina e a ferro redutase, codificadas, respectivamente, pelos genes *Phs* e *FRO1*, deve permitir a geração de novos e informativos marcadores e promover um melhor entendimento da variação existente em *P. vulgaris*. Assim, os objetivos deste trabalho foram obter as sequências dos genes α -*Phs* e *FRO1*, acessar a frequência de SNPs em regiões codantes e não-codantes de ambos os genes, e avaliar a utilidade desses polimorfismos para investigar a diversidade genética em 31 genótipos cultivados de feijão comum e um acesso de *P. lunatus*. Para tanto, *primers* específicos foram desenhados e as sequências obtidas foram alinhadas, permitindo a identificação de vários sítios polimórficos. Parâmetros moleculares foram estimados pelo software Arlequin e MEGA. As análises de agrupamento foram conduzidas através do método *Neighbor-joining*. Foram detectadas 361 bases polimórficas, das quais 260 foram do tipo substituição e 101, indel. A frequência de SNPs em regiões não codantes foi duas vezes maior do que em regiões traduzidas, em ambos os genes; também, a substituições do tipo transição foram mais freqüentes do que as do tipo transversão. Os polimorfismos em regiões codantes levaram à substituição de 17 aminoácidos na proteína faseolina, e 14 na enzima ferro redutase. Tais modificações incluem, na maioria das vezes, alterações de aminoácidos com propriedades similares. Vale destacar que sequência predita de aminoácidos da faseolina exibiu uma diversidade elevada entre os acessos investigados, sendo que alguns desses polimorfismos se prestaram para tipificar alguns acessos. Em contrapartida, a enzima ferro redutase exibiu um padrão de aminoácidos mais homogêneo, principalmente para os acessos andinos. As análises de agrupamento revelaram dois *clusters* bem definidos: um que agrupou os acessos Mesoamericanos e as cultivares brasileiras, e outro que agrupou os Andinos, sugerindo que sequências gênicas são úteis na distinção dos *pools* gênicos de *Phaseolus*. Todavia, os polimorfismos do gene α -*Phs* permitiram uma melhor resolução das relações entre os acessos dentro de cada *pool* gênico, em comparação com os do gene *FRO1*.

Palavras-chave: Polimorfismos; Leguminosa; SNP; Faseolina; Absorção de ferro

ABSTRACT

Genetic diversity in cultivated accessions of common bean (*Phaseolus vulgaris* L.): an *in silico* approach based on the *Phs*- α and *FRO1* genes

Diversity analysis in common bean (*Phaseolus vulgaris* L.), based on morphological traits and molecular markers, has revealed the existence of two major gene pools, the Mesoamerican and Andean pools, which differ in several features, including the grain morphology and physiology. A novel approach to estimating genetic variability is *in silico* mining for single nucleotide polymorphisms (SNP). Arbitrary genome sequences as well as genic sequences were used for this purpose. Additionally, the investigation of nucleotide sequences that code for important proteins, such as phaseolin and iron-reductase, encoded respectively by the *Phs* and *FRO1* genes, should allow new and informative markers to be generated, helping us to gain a better understanding of intraspecific variation in *P. vulgaris*. Therefore, the aims of this study were to sequence the α -*Phs* and *FRO1* genes, to assess SNP frequencies in coding and non-coding regions of both genes, and to assess the possible use of these polymorphisms for investigating the genetic diversity of 31 cultivated genotypes of common bean and one of *P. lunatus*. Specific primers were designed and the sequences obtained were aligned, allowing us to identify several polymorphic sites. Molecular parameters were estimated using the Arlequin and MEGA software packages. Cluster analyses were conducted using the neighbor-joining method. We detected 361 polymorphic sites, consisting of 260 base substitutions and 101 indels. The frequency of SNPs in non-coding regions was twice the frequency in translated regions in both genes. In addition, the occurrence of base transitions was higher than transversions. Polymorphisms in coding regions lead to 17 amino acid substitutions in the phaseolin protein and 14 in the iron reductase enzyme. Most substitutions of this kind include changes that preserve the respective protein functions. The predicted amino acid sequence of phaseolin exhibited high diversity, and some polymorphisms allowed us to typify some accessions. In contrast, the iron reductase enzyme exhibited a more homogeneous pattern of amino acids, especially in the Andean accessions. Cluster analyses revealed two well-defined clusters, one containing the Mesoamerican accessions with the Brazilian cultivars, and another containing the Andean accessions, suggesting that gene sequences are useful for distinguishing the *Phaseolus* gene pools. Interestingly, the polymorphisms detected in the α -*Phs* gene allowed better visualization of the relationships between accessions in the same pool, in comparison to *FRO1*.

Keywords: Polymorphism; Legume; SNP; Phaseolin; Absorption of iron

LISTA DE FIGURAS

- Figura 1 - Diversidade genética entre acessos de *P. vulgaris* a partir de marcadores SSR. A: Distribuição geográfica e genética de acessos de feijoeiro silvestre. B: Relações genéticas entre acessos de feijoeiro. Ramos que terminam com pontos pretos referem-se aos acessos domesticados; os demais, aos acessos silvestres.....28
- Figura 2 - Estratégia de plantas, exceto gramíneas, para a aquisição de ferro do solo. As plantas acidificam a rizosfera em resposta à deficiência de ferro, utilizando uma H⁺ATPase. Isto possibilita a melhor solubilização do íon Fe³⁺, que é reduzido por uma ferro redutase (*FRO1* em ervilha e feijão). O íon ferroso (Fe²⁺) resultante é transportado para dentro da célula por um transportador de raiz (IRT – *Iron Regulated Transporter*).....31
- Figura 3 - Topologia predita da enzima ferro redutase em ervilha. Os cilindros representam os domínios transmembranares I - X, nos quais os sítios Heme de ligação encontram-se destacados (HH). A barra preta denota a localização da sequência correspondente ao sítio de oxiredução.....32
- Figura 4 - Representação esquemática das regiões de desenho de *primers* (setas), e alinhamento dos fragmentos (I, II e III) gerados pelas amplificações ao utilizar cada par de *primers*.....39
- Figura 5 - Esquema de amplificação utilizando uma adaptação do método *PCR-walking* para identificação de sequências alvo. Os fragmentos de DNA, resultantes de uma restrição enzimática, são ligados a adaptadores, cuja extremidade 3' da fita curta encontra-se bloqueada pela presença de um radical amina. A amplificação de sequências alvo envolve duas etapas: no primeiro ciclo da PCR, a extensão ocorre até o final do fragmento que termina com a fita longa do adaptador, o que gera, assim, o sítio de *annealing* do *primer* do adaptador. Em seguida, a partir do segundo ciclo, ocorre a extensão da segunda fita.....42

- Figura 6 - Produto de amplificação (FRO em duplicata) resultante do uso dos *primers* FRO1F e FRO5R com, aproximadamente, 5.5 Kb. O tamanho do amplicon corresponde ao tamanho do putativo transcrito primário do gene *FRO1* em *P. vulgaris*. M1 e M2 são marcadores *Ladder* 1 kb e 100 pb, respectivamente.....53
- Figura 7 - Representação esquemática da formação de *contigs* a partir do alinhamento de FIGs correspondentes aos genes α -*Phs* e *FRO1*. Setas direcionadas para a direita indicam sequenciamento *forward*, e para a esquerda, *reverse*.....55
- Figura 8 - Padrão de amplificação das combinações de enzima de restrição e primers âncora em gel de poliacrilamida (6% p/v) desnaturante, corado com nitrato de prata. Primeira e última canaletas contém *Ladder* de 100+25 pb. As demais canaletas correspondem ao produto de amplificação de cada combinação dos acessos de *P. vulgaris* (Sanilac) em triplicata e de *P. lunatus* (87 JP 12). As setas indicam os fragmentos que foram excisados do gel. Setas pretas correspondem às bandas monomórficas entre *P. vulgaris* e *P. lunatus*, e as vermelhas indicam as bandas monomórficas apenas entre os acessos de *P. vulgaris*.....56
- Figura 9 - Padrão de reamplificação das bandas excisadas, em gel de agarose corado com Sybr Safe. 1: Fragmentos monomórficos entre *P. vulgaris* e *P. lunatus*; 2: fragmentos dos acessos de *P. vulgaris*. A primeira e a última canaletas correspondem ao marcador *Ladder* 100pb. I: a) FROR/*AluI*, b) FROR/*DraI* e c) FROR/*RsaI*; II: PhsF/*RsaI*; III: PhsR/*DraI*; IV: FROF/*DraI*. As setas indicam os fragmentos descartados por apresentarem amplificação inespecífica.....57
- Figura 10 - Representação gráfica do resultado do alinhamento local (BlastN) entre a sequência do FEG FROR/*DraI*_3 e o banco de nucleotídeos do NCBI...59
- Figura 11 - Classes dos polimorfismos observados em sequências de α -*Phs* e *FRO1*. A: A/G e C/T correspondem a transições e as demais classes, a

transversões. B: Número de inserções/deleções (indels) observado classificado de acordo com seu tamanho em pb.....63

Figura 12 - Alinhamento de sequências de *FRO1* entre acessos de *Phaseolus* evidenciando uma região do intron 2. Destaques em vermelho referem-se aos sítios que exibem variação do tipo substituição e indel. As siglas localizadas entre parênteses remetem à classificação dos acessos. M = Mesoamericano, A = Andino, Cr = Crioulo e Cv = Cultivar Embrapa.....64

Figura 13 - Distribuição da frequência de transições (S), transversões (V) e inserções/deleções (Indels) ao longo das sequências dos *contigs* de α -*Phs*. As regiões correspondentes a exons são identificadas pela letra “e” e as correspondentes à introns, pela letra “i”. Número observado de polimorfismos entre as espécies de *Phaseolus* (A) e entre os acessos de *P. vulgaris* (B). As barras localizadas abaixo de cada gráfico correspondem a uma representação esquemática da disposição dos exons ao longo da sequência gênica.....65

Figura 14 - Distribuição da frequência de transições (S), transversões (V) e inserções/deleções (Indels) ao longo das sequências dos *contigs* de *FRO1*. As regiões correspondentes à exons são identificadas pela letra “e” e as correspondentes à introns, pela letra “i”. Número observado de polimorfismos entre as espécies de *Phaseolus* (A) e entre os acessos de *P. vulgaris* (B). As barras localizadas abaixo de cada gráfico correspondem a uma representação esquemática da disposição dos exons ao longo da sequência gênica.....66

Figura 15 - Identificação dos domínios protéicos em uma sequência de nucleotídeos do gene *FRO1* de *P. vulgaris*, por BlastX. FeR: domínio de redução de Fe, conservado na superfamília das ferro redutases; FAD: domínio de ligação de FAD; N: domínio de ligação FAD.....70

Figura 16 - Gráfico de transição (S) e transversão (V) versus distância de Kimura dos genes α -Phs (A), *FRO1* (B) e do conjunto de dados combinado das sequências dos dois genes (C).....72

Figura 17 - Relações entre 32 acessos de feijão a partir das sequências de α -Phs. Os dendrogramas foram construídos pelo método *Neighbor-joining*, usando o modelo de substituição de nucleotídeos Tamura 3 Parâmetros e 500 repetições de *Bootstrap*. As siglas localizadas entre parênteses remetem à classificação dos acessos. M = Mesoamericano, A = Andino, Cr = Crioulo e Cv = Cultivar Embrapa.....74

Figura 18 - Relações entre 32 acessos de feijão a partir das sequências de *FRO1*. Os dendrogramas foram construídos pelo método *Neighbor-joining* usando o modelo de substituição de nucleotídeos Tamura 3 Parâmetros e 500 repetições de *Bootstrap*. As siglas localizadas entre parênteses remetem à classificação dos acessos. M = Mesoamericano, A = Andino, Cr = Crioulo e Cv = Cultivar Embrapa.....75

Figura 19 - Relações entre 32 acessos de feijão a partir das sequências de α -Phs e *FRO1*. Os dendrogramas foram construídos pelo método *Neighbor-joining* usando o modelo de substituição de nucleotídeos Tamura 3 Parâmetros e 500 repetições de *Bootstrap*. As siglas localizadas entre parênteses remetem à classificação dos acessos. M = Mesoamericano, A = Andino, Cr = Crioulo e Cv = Cultivar Embrapa.....76

LISTA DE TABELAS

Tabela 1 - Identificação dos acessos de <i>Phaseolus vulgaris</i> e <i>P. lutanus</i>	37
Tabela 2 - Lista de primers desenhados a partir de sequências de α - <i>Phs</i> e <i>FRO1</i> ...	52
Tabela 3 - Resumo das análises via BlastN dos contigs obtidos para 14 fragmentos das extremidades dos genes α - <i>Phs</i> e <i>FRO1</i>	58
Tabela 4 - Resumo das análises de polimorfismo para sequências de α - <i>Phs</i> e <i>FRO1</i> de acessos de <i>P. vulgaris</i> e <i>P. lunatus</i>	61
Tabela 5 - Resumo das análises de polimorfismo para sequências combinadas de α - <i>Phs</i> e <i>FRO1</i> de acessos de <i>P. vulgaris</i> e <i>P. lunatus</i>	62
Tabela 6 - Posições dos resíduos polimórficos na sequência do gene α - <i>Phs</i>	68
Tabela 7 - Posições dos resíduos polimórficos na sequência do gene <i>FRO1</i>	69
Tabela 8 - Análise Molecular da Variância entre e dentro dos <i>pools</i> gênicos Mesoamericano e Andino de <i>P. vulgaris</i> , a partir das sequências de α - <i>Phs</i> e <i>FRO1</i>	78
Tabela 9 - Diversidade nucleotídica e teste de neutralidade D de Tajima para os <i>pools</i> gênicos Mesoamericano e Andino <i>P. vulgaris</i> , a partir de sequências de α - <i>Phs</i> e <i>FRO1</i>	78

1 INTRODUÇÃO

O feijão comum (*Phaseolus vulgaris* L.) é uma das leguminosas de grão mais importantes para o consumo humano direto. Em 2010 foram plantados aproximadamente 29.882 milhões de hectares em 120 países, sendo que a produção total foi de 23.2 milhões de toneladas. Nove dos dez maiores produtores são considerados países em desenvolvimento, os quais são responsáveis por mais de 70% da produção mundial. Atualmente, o maior produtor é a Índia (20,9%), seguida do Brasil, Mianmar e da China.

O feijão faz parte dos alimentos tradicionais consumidos pela população brasileira, sendo relevante principalmente por seu conteúdo energético e protéico. Além disso, também é considerado um excelente alimento porque fornece nutrientes essenciais como ferro, cálcio, vitaminas, fibra e o aminoácido essencial lisina (RESENDE et al., 2008). Identificar e minimizar os fatores que limitam a produção e o rendimento de feijão é uma preocupação constante de muitos programas de melhoramento, além disso, dada a prevalência do feijão na dieta das populações desses países, a modificação de seu conteúdo nutricional, tornando-o um alimento mais equilibrado, tem também recebido ênfase.

A estrutura das populações do feijoeiro comum é baseada em estudos que utilizaram vários caracteres morfológicos e marcadores moleculares. Baseando-se nos resultados desses trabalhos, uma divisão em dois principais *pools* gênicos está sendo documentada repetidamente, sendo um localizado na América do Sul (Andino) e outro na América Central (Mesoamericano). Todavia, na maioria destes estudos, o baixo nível de polimorfismo e o reduzido número de marcas têm impedido uma caracterização detalhada da estrutura populacional e das relações genéticas dos acessos dentro dos grupos gênicos (KWAK; GEPTS, 2009).

A proteína faseolina, codificada pelo gene *Phs*, tem sido um dos principais alvos das pesquisas moleculares. Esta proteína faz parte do material de reserva das sementes do feijão e apresenta uma relativa diversidade (GEPTS, 1990). Existe uma alta associação entre o tipo de faseolina e o local de origem do material, tanto para acessos silvestres como para domesticados (GEPTS; DEBOUCK, 1991). Os dados sugerem que as cultivares de feijão atuais são resultado de múltiplos eventos de domesticação, possuindo dois centros primários, localizados na América Central e no Sul dos Andes. Ainda, um terceiro centro, localizado na região da Equador, tem

recebido destaque nos estudos sobre os processos de domesticação do feijão (DEBOUCK, 1986; GEPTS, 1990; GEPTS; DEBOUCK, 1991; DEBOUCK et al., 1993).

Dentre vários outros aspectos, os *pools* gênicos também diferem quanto a algumas propriedades fisiológicas básicas do grão, bem como pelas características nutricionais. Uma série de resultados tem mostrado que os feijões andinos tendem a ter maior concentração de ferro na semente do que os feijões mesoamericanos (ISLAM, et al., 2002; ISLAM, et al., 2004; BLAIR, et al., 2010a). A quantidade de ferro acumulada em sementes de feijão está diretamente relacionada com a atividade de uma enzima de membrana plasmática, localizada especialmente em células epidérmicas de raiz, denominada ferro redutase (BLAIR et al., 2010b). Sua função nas raízes é converter o Fe^{3+} presente no solo em Fe^{2+} , o qual, nesta condição, pode ser imediatamente absorvido (GRUSAK, 1995).

Estudos da diversidade no nível das sequências de DNA são necessários em qualquer análise molecular moderna em que os genótipos necessitam ser claramente distinguidos e classificados. A redução no custo e a atual facilidade de sequenciamento direto têm permitido uma série de abordagens para a comparação direta de sequências gênicas, permitindo a detecção de regiões de inserção/deleção (indels) de nucleotídeo(s), bem como de polimorfismos de uma única base (*Single Nucleotide Polymorphisms* - SNPs).

Vários projetos de prospecção de SNPs em larga escala foram e vêm sendo conduzidos para uma ampla gama de organismos (KARCHIN et al., 2005; BALASUBRAMANIAN et al., 2002), inclusive em bancos de dados que abrigam sequências expressas de várias plantas cultivadas (RAFALSKI, 2002; SOLEIMANI et al., 2003; BARBAZUK et al., 2007). Uma alternativa à análise de SNPs em sequências arbitrárias é a análise cuidadosa de variabilidade em genes de interesse, que pode ser útil para a associação das variantes alélicas com diferenças fenotípicas, e para a subsequente seleção do caráter assistida por marcadores ancorados nas respectivas regiões gênicas (CHEN, et al., 2010).

Polimorfismos de nucleotídeos em regiões codificadoras de proteínas podem estar diretamente relacionados com alterações fenotípicas, caso a mudança na sequência de aminoácidos resulte numa alteração funcional significativa (GAITÁN-SOLÍS, 2008). Sequências gênicas sofrem uma pressão seletiva mais intensa do que regiões do genoma ricas em sequências não transcritas. Entretanto, sequências

intrônicas encontram-se sob menor pressão de seleção do que as codantes e, conseqüentemente, acumulam mais alterações.

Em 2004, McClean, Lee e Miklas avaliaram a diversidade entre acessos silvestres e domesticados de *P. vulgaris*, a partir da análise de uma sequência intrônica do gene que codifica a dihidroflavonol 4-redutase. Tal abordagem se mostrou eficaz no estudo da diversidade entre acessos pertencentes aos dois *pools* gênicos, Mesoamericano e no Andino. Porém, os autores sugeriram que a investigação de polimorfismos não só em genes com poucas cópias, tal como estudado, mas também em famílias gênicas e genes com outras funções de interesse, como certos metabólicos e reguladores, pode ser útil.

A análise de genes codificadores de proteínas tais como a faseolina e a ferro redutase, as quais já vêm sendo amplamente investigadas por estudiosos da cultura do feijoeiro, pode promover um melhor entendimento da variação existente na espécie. Ainda, o uso de abordagens similares para investigar polimorfismo em sequências do feijoeiro comum relativamente a espécies afins, deve fornecer informações valiosas no entendimento dos eventos moleculares associados à especiação de *P. vulgaris*.

Neste contexto, os objetivos deste trabalho foram: (i) obter as sequências dos genes α -Phs e *FRO1* para 31 genótipos cultivados de *P. vulgaris* e um acesso de *P. lunatus*; (ii) acessar a frequência de SNPs em regiões codantes e não-codantes de ambos os genes; e (iii) avaliar a utilidade desses polimorfismos para investigar a diversidade genética entre estes genótipos.

2 REVISÃO BIBLIOGRÁFICA

2.1 A cultura do feijão e sua importância para a alimentação humana

A família Leguminosae é muito grande, com 643 gêneros (cerca de 18.000 espécies) agrupados em 40 tribos que são encontradas tanto em ambientes tropicais como temperados (MABBERLEY, 1998). A tribo Phaseoleae, da qual fazem parte *P. vulgaris* (feijão comum), *Vigna unguiculata* (feijão grande) e *Glycine max* (soja), é o mais importante grupo econômico e contém 75% de todas as leguminosas comercializadas no mundo.

Atualmente, são conhecidas cerca de 55 espécies de feijões do gênero *Phaseolus*; contudo, apenas quatro delas apresentam importância econômica: Feijão Escarlata (*P. coccineus*); Feijão Tepari (*P. acutifolius*); Feijão-lima (*P. lunatus*) muito importante no Perú, Bolívia e Equador; e Feijão comum (*P. vulgaris*), o tipo de feijão mais difundido hoje no mundo e principalmente no Brasil. Todas as espécies do gênero são diplóides, sendo a maioria $2n=2x=22$.

O feijão comum está presente em todos os continentes, exceto Antártica. As formas silvestres de *P. vulgaris* são consideradas ancestrais das atuais formas cultivadas, sendo encontradas apenas nas Américas. Ocorrem da Província de San Luis, no norte da Argentina, até o Estado de Chihuahua, no norte do México, cobrindo uma extensão quase contínua de aproximadamente 7.000 km de zonas montanhosas, desde 500 a 2.500 m de altitude. Não apresentam nenhum mecanismo de isolamento reprodutivo, sendo que formas silvestres e cultivadas podem ser cruzadas facilmente, produzindo gerações F_1 e F_2 férteis. Segundo Gepts e Debouck (1991), as diferenças genéticas entre as duas formas afetam apenas poucos locus.

Existem duas principais classes comerciais de feijão comum: feijão vagem e feijão seco. Em 2010 foram plantados 29.882 milhões de hectares em 120 países, sendo que a produção total foi de 23.2 milhões de toneladas. Nove dos dez maiores produtores são considerados países em desenvolvimento, os quais são responsáveis por mais de 70% da produção mundial de feijão. Atualmente, o maior produtor é a Índia (20,9%), seguida do Brasil, Mianmar e da China.

O feijão faz parte dos alimentos tradicionais consumidos pela população brasileira, tendo importância essencialmente por seu conteúdo energético e protéico.

Além disso, também é considerado um excelente alimento porque fornece nutrientes essenciais como ferro, cálcio, vitaminas, fibra e o aminoácido essencial lisina (RESENDE et al., 2008).

Além de se destacar na produção mundial, o Brasil é o maior consumidor de feijão, com consumo estimado em 16 kg/ano por habitante. Todavia, Wander, Ramalho e Andrade (2005) constataram que o consumo de feijão vem diminuindo nos últimos 40 anos. Essa redução poderia estar associada ao maior acesso da população a produtos de origem animal, principalmente a carne bovina. Como forma de complementação da dieta, essa leguminosa poderia estar sendo substituída por outras fontes protéicas.

O consumo diário de feijão está entre 50 a 100 g por dia/pessoa, contribuindo com 28% de proteínas e 12% de calorias ingeridas (DANTAS et al., 2007). Portanto, como alimento básico e sob o ponto de vista quantitativo, o feijão é considerado alimento protéico, embora, seu conteúdo calórico, mineral e vitamínico não possa ser desprezado.

Apesar de serem deficientes em aminoácidos sulfurados, as sementes de feijão são ricas em lisina. Combinações das sementes cozidas com cereais, tal como o arroz, são consideradas complementares, constituindo alimentos básicos em certas regiões da América Latina (VIEIRA, 1992).

Devido a sua composição, o feijão proporciona vários benefícios à saúde, sendo indicado na prevenção e no tratamento de várias doenças, tais como: distúrbios cardíacos, diabetes mellitus, obesidade e câncer, preenchendo as principais recomendações dietéticas para a boa saúde: aumento no consumo de fibras, amido e outros carboidratos complexos, e diminuição no consumo de lipídeos e sódio (SATHE et al., 1984; GEIL; ANDERSON, 1994).

Como mencionado anteriormente, além da alta quantidade de proteína, o feijão contribui como uma das melhores fontes vegetais de ferro, sendo valiosa a sua contribuição em casos de deficiências deste mineral, tal como a anemia ferropriva (IDA – *Iron Deficiency Anemia*). O ferro é um elemento essencial para o crescimento e desenvolvimento humano e está presente em todas as células do organismo. Ainda, atua como constituinte da hemoglobina e mioglobina, e de muitas enzimas, como, por exemplo, aquelas envolvidas em reações de oxi-redução (WICK; PINGGERA; LEHMANN, 1996).

IDA é o tipo mais comum de disfunção nutricional humana e é caracterizada pela indisponibilidade de ferro para o organismo. Várias estratégias podem ser adotadas no combate à IDA, como, por exemplo, a administração oral ou injetável do mineral, porém o sabor do medicamento líquido, a pigmentação dos dentes e necessidade de uma logística apropriada para a entrega dos medicamentos apresentam-se como algumas desvantagens (FROSSARD et al., 2000).

Uma alternativa no incremento da concentração de ferro na dieta é a biofortificação de alimentos básicos, e o feijão tem sido uma cultura estratégica e alvo de ações de biofortificação para aumentar a quantidade de ferro na dieta de seres humanos. Os programas de melhoramento, além de dedicar esforços voltados para identificar e minimizar os fatores que limitam a produção e o rendimento, têm dado ênfase, também, na modificação do conteúdo nutricional do feijão, na tentativa de torná-lo um alimento mais equilibrado. Nesse contexto, a identificação de genes e o melhor entendimento das etapas de aquisição e acúmulo de ferro nas sementes podem auxiliar na seleção e melhoria de genótipos para o caráter.

2.2 Origem e domesticação de *Phaseolus vulgaris*

Evidências oriundas do Novo Mundo, segundo Heiser Jr. (1977), indicam que o feijão foi uma das primeiras plantas cultivadas pelo homem nas Américas. A sua domesticação data de 5.000 anos AC no México e no Peru, data de 3.000 AC. Vavilov (1950), com base em um método fitogeográfico, mostrara que o centro de diversidade genética de inúmeras formas silvestres, localizava-se no México e América Central, pois foi nestas áreas que ele encontrou maior variedade de formas da espécie.

Smartt (1969) estudou a evolução do gênero *Phaseolus* na América e mostrou que os registros arqueológicos encontrados na América Central e do Sul são impressionantes, comparados ao encontrado para outras espécies cultivadas. Datações realizadas pelo método do carbono 14 mostraram que os restos de feijão (aparentemente já domesticado) mais antigos até agora descritos foram encontrados em Huachichona, no norte da Argentina, com 6.700 a 9.600 anos de idade, e na caverna de Guitarrero, no Peru, com 4.300 a 7.000 anos de idade (GEPTS; DEBOUCK, 1991). Restos milenares foram descritos em outras partes da América Latina, mas nunca em outro continente.

A ampla distribuição geográfica do feijão acompanhada da diversidade genética presente em formas cultivadas e silvestres levou ao questionamento sobre quais eventos de domesticação ocorreram na espécie. A domesticação do feijoeiro trouxe-lhe diversas modificações morfológicas e fisiológicas em relação às formas silvestres, as quais foram documentadas por Schawanz (1966). A dispersão dentro das regiões de domesticação provocou uma profunda redução da diversidade genética em decorrência dos efeitos da deriva genética e pressão para adaptação às novas condições ambientais propícias ao cultivo e consumo humano (GEPTS, 1998). As principais mudanças que ocorreram foram: (i) gigantismo, caracterizado pelo aumento no tamanho e peso das sementes, vagens maiores e com menor número de sementes, flores maiores e caule mais robusto; (ii) perda do mecanismo de dispersão das sementes; (iii) mudança do hábito de crescimento, pois as plantas cultivadas apresentam um crescimento mais restrito e forma mais compactada, tornando-as de menor altura e mais eretas; (iv) perda da dormência das sementes. Além disso, houve mudanças na arquitetura do sistema radicular (SCHAWANITZ, 1966).

Segundo Gentry (1969), a domesticação do feijoeiro ocorreu no México de onde foi levado para a região sul da América. Contrariamente, Berglund-Brucher (1976) acredita que a domesticação ocorreu no sul da América e não exclui a possibilidade de haver um segundo centro de domesticação independente na América Central. Registros arqueológicos de feijão, com idade similar, foram encontrados na América Central (Tehuacán Valley, México: 5.000 AC) e nos Andes (Callejón de Huaylas, Peru: 5.680 AC). Baseado nestes dados, vários autores tem sugerido que a domesticação do feijão tenha ocorrido em dois centros independentes: América Central, onde as formas cultivadas apresentam sementes pequenas e outro na região dos Andes, onde as sementes são maiores (HARLAN, 1971; HEISER, 1977; EVANS, 1976).

Análises de amostras atuais e arqueológicas de *P. vulgaris* mostram que, antes da domesticação o conjunto gênico das formas silvestres já divergia em dois grupos localizados em regiões geográficas diferentes. Um deles é encontrado na Mesoamérica, e compreende a metade sul do México e o norte da América Central. O outro se localiza na região dos Andes, abrangendo o Peru, a Bolívia e a Argentina. Esses dois *pools* gênicos são distinguidos tanto morfológicamente (GEPTS; DEBOUCK, 1991; CHIORATO et al., 2005) quanto molecularmente (GEPTS; BLISS,

1985; GEPTS et al., 1986; KOENIG; SINGH; GEPTS, et al. 1990; SINGH; NODARI; GEPTS, 1991; BECERRA-VELÁSQUEZ; GEPTS, 1994; FREYRE et al., 1996; THOME et al., 1996; GEPTS, 2003; CHACÓN; PICKERSGILL; DEBOUCK, 2005; BLAIR et al., 2006; DÍAZ; BLAIR, 2006; HANAI et al., 2007; BENCHIMOL et al., 2007; KWAK; GEPTS, 2009; GILL-LANGARICA et al., 2011).

Em adição a esses dois maiores *pools* gênicos, um terceiro grupo vem sendo relatado com importância na história evolutiva do feijoeiro (DEBOUCK et al., 1993). Mais recentemente, este grupo tem sido considerado um núcleo de diversidade, localizado no Equador e Norte do Peru (KAMI et al., 1995), de onde feijões silvestres se dispersaram tanto na direção norte como na direção sul, para formar dois *pools* gênicos geograficamente distintos: o Mesoamericano e o Andino (GEPTS, 1998). Por sua vez, a divergência pós-domesticação deu origem a três raças domesticadas em cada um destes dois *pools* gênicos (SINGH et al., 1991).

Kwak e Gepts (2009) avaliaram a estruturação populacional de *P. vulgaris*, utilizando vários genótipos (silvestres e cultivados, oriundos de diferentes *pools* gênicos) em relação à putativa população silvestre ancestral, citada no parágrafo anterior. Os autores observaram que esta população encontra-se posicionada, geneticamente, entre os *pools* Andino e Mesoamericano, porém com um viés para o Mesoamericano provocado, provavelmente, pela melhor estruturação dos acessos dentro deste grupo (Figura 1). As relações genéticas entre os acessos silvestres refletem sua distribuição geográfica, o que corrobora a existência de um terceiro grupo. Entretanto, os autores sugerem que um estudo mais aprofundado, com mais acessos silvestres ou mais locos moleculares, deva ser realizado.

Figura 1 - Diversidade genética entre acessos de *P. vulgaris* a partir de marcadores SSR. A: Distribuição geográfica e genética de acessos de feijoeiro silvestre. B: Relações genéticas entre acessos de feijoeiro. Ramos que terminam com pontos pretos referem-se aos acessos domesticados; os demais, aos acessos silvestres

Fonte: Kwak e Gepts, 2009

2.3 A faseolina

A faseolina, glicoproteína codificada pelo gene *Phs*, tem sido um dos principais alvos de pesquisas moleculares em *P. vulgaris*. Esta proteína faz parte do material de reserva das sementes e compõe a fração protéica predominante nas sementes de feijão. Foram identificados pelo menos 10 tipos de alelos desta proteína em diferentes acessos cultivados de feijão e de seus parentes silvestres (GEPTS, 1990).

Por meio de análise eletroforética das proteínas da semente, Gepts (1986) determinou os tipos de faseolina que ocorrem nas formas silvestres e cultivadas de feijão. Essas análises revelaram três tipos predominantes entre os acessos: "S" (Sanilac), "T" (Tendergreen) e "C" (Contender). Segundo Brown (1981) os genes que codificam os diferentes polipeptídeos de cada tipo de faseolina estão fortemente ligados apresentando herança Mendeliana simples, com alelos de ação codominante.

Ainda no trabalho de 1986, Gepts observou que, na América, as formas silvestres de feijão apresentavam faseolina tipo "S" e "M" enquanto nos Andes (Peru e Argentina) exibiam exclusivamente o tipo "T". Verificou, também, uma correspondência entre a distribuição geográfica dos tipos de faseolina entre os feijões silvestres e cultivados. A maioria dos acessos cultivados na América Central apresenta tipo "S", sendo que nos Andes o tipo predominante é o "T". Além disso, alguns acessos andinos apresentaram o tipo "C" e outros dois tipos "H" e "A". Em ambas as regiões, acessos com faseolina tipo "T" tinham sementes maiores que aqueles com tipo "S". Nos Andes, nenhuma diferença foi observada entre os tipos de faseolina "T", "C", "H" e "A" entre acessos cultivados.

Com base nesses resultados, duas hipóteses surgiram para explicar a domesticação do feijão. A primeira propõe que a domesticação tenha ocorrido em uma das duas regiões (América Central ou Andes), resultando em genótipos com sementes pequenas e faseolina do tipo "S", e genótipos com sementes maiores e faseolina tipo "T", "C", "H" e "A". Posteriormente esses genótipos dispersaram para outras regiões. A segunda hipótese sugere dois centros independentes: América Central, com genótipos de sementes pequenas e faseolina do tipo "S"; e os Andes com genótipos de sementes maiores e faseolina tipo "T" (GEPTS et al., 1986).

Os eventos que conduziram ao aparecimento de acessos com tipos de faseolina "C", "H" e "A" ainda não foram determinados. Eles poderiam ter sido domesticados na América Central e, posteriormente, levados para os Andes. Entretanto, esses tipos de faseolina não foram encontrados entre as formas silvestres ou cultivadas da espécie presentes na América Central. Alternativamente, esses tipos de faseolina poderiam ter surgido nos Andes, antes ou depois da domesticação (GEPTS et al., 1986).

Posteriormente, Koenig, Singh e Gepts (1990) identificaram novos tipos de faseolina: "J", em formas silvestres da Argentina, e "I", num acesso silvestre do norte do Peru. Em feijões cultivados, encontraram formas variantes da faseolina "S": "Sb", em alguns cultivares de sementes pequenas do Brasil, como Rio Tibagi e Mulatinho Vagem Roxa; e "Sd", predominando nos cultivares do altiplano central do México.

O perfil eletroforético unidimensional mostra que todos esses tipos de faseolina descritos correspondem a 2 até 6 polipeptídios que diferem quanto ao peso molecular (SALMANOWICZ, 2001; MONTROYA et al., 2008) e ao ponto isoelétrico (BROWN et al., 1981). Tais diferenças refletem, além de modificações pré e pós-traducionais, às variações presentes nas sequências de DNA que codificam esses polipeptídios, os quais podem ser categorizados em duas subfamílias: α -faseolina e β -faseolina (SLIGNTOM et al., 1985), ambas derivadas de um mesmo ancestral comum.

2.4 Absorção de Ferro (Fe) em plantas

O ferro é um micronutriente essencial, importante na biossíntese da clorofila e está presente nos sítios ativos de enzimas e outras proteínas, como aquelas responsáveis pela transferência de elétrons (MESSERSCHIMIDT et al., 2001). Apesar de ser um micronutriente essencial, mecanismos celulares e moleculares regulam a absorção, o transporte e o acúmulo de ferro, até a sua utilização nos mais variados processos metabólicos (HELL; STEPHAN, 2003).

Em condições aeróbicas e pH neutro ou básico, o ferro é extremamente indisponível para absorção pelas plantas, formando complexos insolúveis no solo (BRIAT, 2006). As plantas lidam com a insolubilidade do ferro utilizando sistemas moleculares para a sua redução ($\text{Fe}^{3+} \rightarrow \text{Fe}^{2+}$) ou secretam moléculas para quelar o ferro não reduzido. Para tanto, desenvolveram duas estratégias para assegurar a

absorção de ferro, a partir das raízes, quando há baixa disponibilidade deste mineral (RÖMHELD; MARSCHNER, 1986). As plantas dicotiledôneas e monocotiledôneas, com exceção das gramíneas, desenvolveram a estratégia I (Figura 2), que apresenta os seguintes passos: (i) excreção de prótons via ATPase tipo – P para acidificar o meio e aumentar a solubilidade do Fe^{3+} ; (ii) redução do Fe^{3+} por uma quelato redutase para a forma de Fe^{2+} , que é mais solúvel; (iii) transporte do Fe^{2+} por transportadores para o interior da célula. A estratégia II evoluiu em gramíneas, como o arroz (*Oryza sativa*) e o milho (*Zea mays*), e encontra-se sumarizada em Mori (1999).

Figura 2 - Estratégia de plantas, exceto gramíneas, para a aquisição de ferro do solo. As plantas acidificam a rizosfera em resposta à deficiência de ferro, utilizando uma H⁺ATPase. Isto possibilita a melhor solubilização do íon Fe^{3+} , que é reduzido por uma ferro redutase (*FRO1* em ervilha e feijão). O íon ferroso (Fe^{2+}) resultante é transportado para dentro da célula por um transportador de raiz (IRT – Iron Regulated Transporter)

Fonte: Modificado de Briat (2006)

Ferro redutases são membros da superfamília dos flavocitocromos, são expressas em células epidérmicas de raiz e encontram-se localizadas na membrana plasmática. Sua atividade varia de acordo com as condições fisiológicas e ambientais como, por exemplo, a disponibilidade de ferro e o pH do solo. Mutantes de ervilha se mostraram bastante úteis para avaliar o papel da redução do ferro e sua aquisição pelas raízes (GRUSAK, 2000).

Análises comparativas entre as sequências de aminoácidos de ferro redutases de vários organismos mostraram que há regiões correspondentes a domínios transmembranares bastante conservadas e análises *in silico* da sequência de aminoácidos de uma ferro redutase (*FRO1*) de ervilha permitiu a predição de sua conformação estrutural (Figura 3). O modelo, proposto por Waters, Blevins e Edie

(2002), exibe dez domínios transmembranares. As regiões N e C-terminal encontram-se na superfície extracelular da membrana e um *loop* intracelular, localizado entre as hélices transmembranares VII e VIII, contém sítios de ligação NADH e FAD. Os sítios de ligação das moléculas que contém ferro (sítios Heme de ligação) estão localizados nas hélices transmembranares V e VII, ao passo que a sequência correspondente ao sítio de oxiredução está localizado no domínio transmembranar VIII.

Figura 3 - Topologia predita da enzima ferro redutase em ervilha. Os cilindros representam os domínios transmembranares I - X, nos quais os sítios Heme de ligação encontram-se destacados (HH). A barra preta denota a localização da sequência correspondente ao sítio de oxiredução

Fonte: Modificado de Waters et al. (2002)

Dentre os vários aspectos supracitados, os grupos gênicos de feijão diferem, também, quanto a algumas propriedades fisiológicas, bem como características nutricionais como a concentração de ferro na semente. Uma série de estudos tem demonstrado que feijões andinos tendem a ter maior concentração de ferro nas sementes do que os mesoamericanos (ISLAM, et al., 2002; ISLAM, et al., 2004; BLAIR, et al., 2010a). Blair et al. (2010b) relataram que a atividade da enzima ferro redutase está diretamente relacionada com o acúmulo de ferro em sementes de feijão; entretanto, não foram encontradas diferenças significativas na atividade enzimática da ferro redutase entre os dois grupos gênicos. Porém, análises mais detalhadas dos genes envolvidos nos processos de aquisição, como a identificação dos polimorfismos na sequência de nucleotídeos do gene *FRO1*, podem ser úteis na tentativa de entender melhor a fonte deste polimorfismo.

2.5 Polimorfismos moleculares e diversidade genética em *P. vulgaris*

Marcadores genéticos são usados para diferenciar genótipos e são herdados segundo as leis Mendelianas. Os polimorfismos podem ser morfológicos, de natureza bioquímica ou molecular; neste caso, deve haver diferença na sequência de nucleotídeos do DNA, levando a um fenótipo molecular distinto (VIEIRA; VELLO; SILVA-FILHO, 2004).

Os marcadores moleculares podem ter ação: (i) dominante: RAPD (*Random Amplified Polymorphic DNA*, WILLIAMS et al., 1990) e AFLP (*Amplified Fragment Length Polymorphism*, VOS et al., 1995) ou (ii) codominante: RFLP (*Restriction Fragment Length Polymorphism*, BOSTSTEIN et al., 1980), SSR (*Simple Sequence Repeat*, TAUTZ, 1989) e SNP (*Single Nucleotide Polymorphism*, JORDE, 1995).

Diversos tipos de marcadores de DNA, desenvolvidos para estudar a diversidade genética e os processos evolutivos e de domesticação em plantas, são considerados melhores comparativamente aos métodos mais antigos, como marcadores morfológicos e bioquímicos (CHARCOSSET; GALLAIS, 2002; GAITÁN-SOLÍS et al., 2002; BLAIR et al., 2009; KWAK; GEPTS, 2009; BURLE et al., 2010). O conhecimento atual da estrutura populacional e da origem da domesticação do feijoeiro é baseado em estudos que utilizaram diversos marcadores moleculares, tais como: RFLPs (BECERRA-VELÁSQUEZ; GEPTS, 1994); RAPDs (FREYRE et al., 1996; BEEBE et al., 2000); AFLPs (THOME et al., 1996; BEEBE et al., 2001; PAPA; GEPTS, 2003); SSRs (BLAIR et al., 2006; DÍAZ; BLAIR, 2006; HANAI et al., 2007; BENCHIMOL et al., 2007; GILL-LANGARICA et al., 2011); DNA mitocondrial (KHAIRALLAH; SEARS; ADAMS, 1992) e cloroplastidial (CHACÓN et al., 2005); SNPs (MCCLEAN; LEE; MIKLAS, 2004; CHEN et al., 2010). Todavia, na maioria desses estudos, o baixo nível de polimorfismo e o reduzido número de marcas usado tem impedido uma caracterização detalhada da estrutura populacional e das relações genéticas entre os acessos dentro dos grupos gênicos (KWAK; GEPTS, 2009).

Vários autores têm confirmado que o germoplasma mesoamericano de *P. vulgaris* é bem mais complexo do que se imagina e que contém uma diversidade importante, que necessita ser explorada, principalmente para características de atributos agrônômicos. Em contrapartida, estudos sobre a diversidade de feijoeiros andinos detectaram que este grupo apresenta uma estreita base genética, a qual,

segundo Thome et al. (1996), pode estar associada com a topografia dos Andes, e que esta diversidade não foi menor devido a introgressões de feijoeiros mesoamericanos. Isto implica que raças locais andinas foram derivadas de uma ou poucas populações selvagens similares (BEEBE et al., 2001).

Apesar de haver um grande número de trabalhos, todos os autores mencionados ressaltam que análises adicionais ainda são necessárias, incluindo a estratificação da diversidade genética, por exemplo, entre acessos dentro de classes comerciais, regiões de origem, raças genéticas e *pools* gênicos. Para tanto, a geração de novas marcas, bem como a inclusão de novos acessos nas análises podem permitir a melhor compreensão da relação entre os genótipos de *P. vulgaris*.

2.5.1 SNPs em sequências gênicas de *P. vulgaris*

Recentemente, uma nova classe de marcadores de DNA tem recebido destaque. Denominados polimorfismos de base única (*Single Nucleotide Polymorphism* – SNP), esses marcadores são a mais frequente forma de polimorfismo no genoma. SNPs se devem tanto à variação de bases (por transição ou transversão) quanto à presença de pequenas inserções ou deleções (indels) quando se comparam fragmentos homólogos de DNA. Polimorfismos de nucleotídeos em regiões codificadoras de proteínas podem estar diretamente relacionados com alterações fenotípicas, caso a mudança na sequência de aminoácidos resulte numa alteração funcional significativa (GAITÁN-SOLÍS, 2008).

Em contraste com a característica multialélica de certos marcadores, como os SSR, SNPs são de natureza bi-alélica. Todavia, haplótipos múltiplos podem ser distinguidos pela combinação de dois ou mais alelos de SNPs, desde que estes estejam ligados, aumentando assim a informatividade dos marcadores (BHATTRAMAKKI; RAFALSKI, 2001).

Em análises de diversidade a partir de sequências de DNA, os parâmetros de diversidade molecular, utilizados para estimar a distância genética, costumam incluir uma contagem de ocorrências de substituições e de indels. Ao caracterizar a frequência de SNPs classificando as substituições em transições ou transversões, por exemplo, é possível eleger o modelo evolutivo de substituição de nucleotídeos que melhor se adapta à sequência estudada. A escolha do modelo evolutivo

influenciará diretamente na construção da árvore filogenética, portanto, a utilização de um modelo inadequado pode implicar em uma interpretação equivocada.

Diferentes modelos foram propostos para estimar a probabilidade da ocorrência de substituição de nucleotídeos. Um dos mais simples, a distância p (*p-distance*), baseia-se apenas na porcentagem de bases discrepantes entre duas sequências devidamente alinhadas. Além deste, dois modelos são bastante utilizados em análises de diversidade: JC (JUKES; CANTOR, 1969), o qual considera que as taxas de substituição ocorrem em frequências iguais, e Kimura 2 Parâmetros (KIMURA, 1980), que considera diferentes frequências de transições e transversões.

A análise de agrupamento trata da identificação de grupos de indivíduos similares após a estimação de uma matriz de similaridade. Há vários métodos de agrupamento que se diferenciam pelo tipo de resultado e pelas diferentes formas de definir a proximidade entre indivíduos ou grupos formados. Em todos os casos, não se conhece à priori, o número de grupos a ser estabelecido e diferentes métodos proporcionam resultados distintos (CRUZ; CARNEIRO, 2003). Os métodos de agrupamento baseiam-se principalmente em métodos hierárquicos e de otimização. Dentro dos hierárquicos, destaca-se o método do vizinho mais próximo (*Neighbor-joining*), no qual os grupos são identificados na forma de dendrogramas e tem sido o mais utilizado nos estudos de diversidade genética de *P. vulgaris* (KWAK.; GEPTS, 2009; CHEN et al., 2010).

Vários projetos de prospecção de SNPs em larga escala foram e vêm sendo conduzidos para uma ampla gama de organismos (BALASUBRAMANIAN et al., 2002; KARCHIN et al., 2005), inclusive em dados depositados em bancos de sequências expressas de várias plantas cultivadas (RAFALSKI, 2002; SOLEIMANI et al., 2003; BARBAZUK et al., 2007). Uma alternativa à análise de SNPs em sequências arbitrárias é a análise cuidadosa da variabilidade de genes de interesse, que pode ser útil para a associação das variantes alélicas com diferenças fenotípicas, e para a subsequente seleção assistida por marcadores do fenótipo associado (CHEN, 2010).

Em 2004, McClean, Lee e Miklas avaliaram a diversidade entre acessos silvestres e domesticados de *P. vulgaris*, baseados no estudo de uma sequência intrônica do gene que codifica a dihidroflavonol-4-redutase. Tal abordagem se mostrou eficaz no estudo da diversidade entre acessos de feijão pertencentes aos

dois *pools* gênicos, Mesoamericano e no Andino. Porém, os autores sugeriram que a investigação de polimorfismos não só em genes com poucas cópias, tal como estudado, mas também em famílias gênicas e genes com outras funções de interesse, como certos metabólicos e reguladores, pode ser útil.

Já em 2010, Chen et al. avaliaram a diversidade nucleotídica de um novo gene *P5CS*, também de *P. vulgaris*, e identificaram 244 sítios polimórficos incluindo substituições e indels. Tais polimorfismos revelaram parte da história evolutiva do gene, bem como permitiram acessar a diversidade entre diferentes classes de feijões oriundos dos dois grupos gênicos.

A análise de genes de interesse vem se mostrando bastante útil tanto em abordagens filogenéticas quanto para acessar a diversidade intraespecífica. Desta forma, o estudo de sequências gênicas de proteínas tais como a faseolina e a ferro redutase, as quais já vêm sendo amplamente investigadas por estudiosos da cultura do feijoeiro, pode permitir a geração de marcas informativas adicionais e promover um melhor entendimento da variação existente em *P. vulgaris*. Ainda, o uso de abordagens similares em estudos sobre o polimorfismo em sequências do feijoeiro relativamente às espécies afins, deve fornecer informações valiosas no entendimento dos eventos moleculares associados à especiação de *P. vulgaris*.

3 MATERIAL E MÉTODOS

3.1 Material Vegetal

O material vegetal utilizado neste trabalho consistiu de 31 acessos de *Phaseolus vulgaris* e um acesso de *P. lunatus* utilizado como “outgroup” (Tabela 1). De acordo com Hanai et al. (2007), esta coleção de acessos compreende uma amostra representativa de feijões cultivados, e inclui as linhagens genitoras de três populações de mapeamento: ‘Bat 93’ X ‘Jalo EEP 558’ (NODARI, 1993), ‘Carioca’ X ‘Flor de Mayo’ (MELO; SANTOS; FERREIRA, 2002) e ‘CAL 143’ X ‘IAC UNA’ (Instituto Agrônomo, Campinas, Brasil).

Tabela 1 - Identificação dos acessos de *Phaseolus vulgaris* e *P. lunatus*

Denominação	Origem	Denominação	Origem
Brasil 2	Mesoamérica	Antioquia 8	Andes
Porriilo	Mesoamérica	Jalo EEP 558	Andes
Rio Tibagi	Mesoamérica	Jabola	Andes
Mulatinho	Mesoamérica	Cal 143	Andes
Carioca comum	Mesoamérica	Red Kidney	Andes
Puebla 152	Mesoamérica	Pompadour	Andes
Great Northern	Mesoamérica	Serro Azul Fosco	Crioulo
Flor de mayo	Mesoamérica	Serro Azul Brilhante	Crioulo
Garbancillo	Mesoamérica	Jalo Precoces	Cultivar Embrapa
Apetito Blanco	Mesoamérica	BRS Pontal	Cultivar Embrapa
Jamapa	Mesoamérica	Aporé	Cultivar Embrapa
Barbunya	Mesoamérica	BRS Pitanga	Cultivar Embrapa
IAC-UNA	Mesoamérica	BRS Vereda	Cultivar Embrapa
Baetão	Mesoamérica	87 JP 12	<i>Phaseolus lunatus</i>
Sanilac	Mesoamérica		
Tu	Mesoamérica		
Bat 93	Mesoamérica		
Red Mexican	Mesoamérica		

3.2 Extração e quantificação do DNA genômico

Amostras foliares jovens de cada acesso foram coletadas, congeladas em nitrogênio líquido e armazenadas em freezer -80 °C, até serem submetidas à extração de DNA genômico usando o método desenvolvido por Doyle e Doyle (1987), com base em CTAB. A quantificação do DNA extraído foi realizada por espectrofotometria com auxílio do equipamento NanoDrop 2000 - Thermo Scientific®.

Foram feitas diluições a 10 ng/μl e 50 ng/μl, as quais foram utilizadas nas reações de amplificação (PCR) e de hidrólise com enzimas de restrição, respectivamente.

3.3 Prospecção *in silico* das sequências de *Phs* e *FRO1* de *P. vulgaris*, desenho e síntese de *primers*

Uma busca por sequências correspondentes aos genes *Phs* e *FRO1* de *P. vulgaris* foi realizada no site público do NCBI (ALTSCHUL et al., 1997; disponível em <<http://www.ncbi.nlm.nih.gov/Genbank/>>). Esta pesquisa foi executada nos bancos de nucleotídeos e de *Expressed Sequence Tags* (ESTs), e foram selecionadas preferencialmente aquelas sequências de genes completos, anotadas com um alto nível de detalhamento.

As sequências selecionadas foram utilizadas para o desenho dos *primers* (*forward* e *reverse*) que foi efetuado usando-se o aplicativo web Primer3 (ROZEN; SKALETSKY, 2000; disponível em: <<http://frodo.wi.mit.edu/primer3/>>). A fim de padronizar as condições de PCR para vários pares de *primers*, os seguintes critérios foram utilizados: (i) temperatura de *annealing* (TA) variando de 55 a 60 °C; (ii) diferença máxima de 3 °C na TA entre os *primers* de cada par; (iii) conteúdo de GC variando de 40 a 60%. Ainda, cada par de *primers* foi avaliado quanto às possibilidades de formação de estruturas secundárias com auxílio do *software* Gene Runner (SPRUYT; BUQUICCHIO, 1994). Considerando as limitações da técnica de sequenciamento idealizada por Sanger, Nicken e Coulson (1997), a qual envolve terminadores irreversíveis, os *primers* foram desenhados de modo que os fragmentos amplificados não ultrapassassem 700 pares bases (pb), na tentativa de garantir que a sequência completa pudesse ser obtida em uma única corrida. Ainda, os *primers* foram desenhados visando à sobreposição de regiões terminais entre dois fragmentos, e assim possibilitar a sobreposição das sequências dos fragmentos gerados (Figura 4). Os *primers* foram sintetizados pela empresa Sigma-Aldrich na escala de 25 nmol.

Figura 4 - Representação esquemática das regiões de desenho de primers (setas), e alinhamento dos fragmentos (I, II e III) gerados pelas amplificações ao utilizar cada par de primers

Ao observar a representação esquemática na figura 4, percebe-se que os fragmentos a serem amplificados não abrangem as bases presentes nos extremos da sequência. Isto ocorre porque tais regiões não permitem o desenho de *primers* segundo os critérios citados anteriormente. Com o objetivo de se obter essas sequências, *primers* adicionais complementares às regiões terminais de cada gene (um *reverse* na extremidade 5' e um *forward* na extremidade 3') foram desenhados e denominados *primers* âncora; estes foram utilizados numa adaptação da metodologia *PCR-walking*, descrita por Cottage et al. (2001).

Como foram adotadas duas metodologias distintas para a obtenção dos fragmentos de DNA correspondentes aos genes *Phs* e *FRO1*, estes foram classificados em duas categorias, de acordo com a sua localização no gene. Aqueles obtidos pela metodologia *PCR-walking* foram denominados Fragmentos das Extremidades do Gene (FEG), ao passo que os demais foram denominados Fragmentos Internos do Gene (FIG).

3.4 Obtenção dos Fragmentos Internos do Gene (FIG)

3.4.1 Condições de amplificação e detecção dos FIGs

Os pares de *primers* sintetizados foram submetidos a testes de amplificação usando o DNA *template* de quatro acessos. As reações de amplificação foram feitas para um volume de 16 µL, contendo 20 ng de DNA, tampão PCR 1X (GoTaq® Flexi Buffer, Promega), 0,3 µM de primers *forward* e *reverse*, 1,5 mM de MgCl₂ (Promega), 200 µM de cada dNTP (GE Healthcare Life Sciences), 0,4 U de Taq DNA polimerase

(GoTaq[®] Flexi DNA, Promega), e água ultra pura (Direct-Q[®] UV, Millipore) e esterilizada para completar o volume. As reações de amplificação em termociclador (GeneAmp[®] PCR System 9.700, Applied Biosystems) foram assim realizadas: desnaturação inicial por 5 min. a 94 °C, seguida por 30 ciclos de amplificação com desnaturação por 40 s a 94 °C, *annealing* de 1 min. com temperatura variando de 55 a 60 °C (a depender do par de *primers*), e extensão de 1 min a 72 °C. Ao final, ocorreu uma extensão adicional por 5 min. a 72 °C. Os produtos resultantes (amplicons) foram separados por eletroforese em gel de agarose (0,8% p/v) e visualizados após coloração com SYBR SAFE[®] (Invitrogen). O marcador de massa molecular (*Ladder* de 100 pb, Invitrogen) foi utilizado para se determinar, comparativamente, o tamanho dos amplicons. Os géis foram fotodocumentados em um sistema de imagem digital MultiDoc-It - UVP[®].

Depois de confirmada a eficiência das condições de amplificação adotadas, todos os demais genótipos foram submetidos à PCR, sendo que as reações foram feitas para um volume final de 50 µl, contendo 100 ng de DNA, e demais componentes com concentrações iguais àquela usada nos testes de amplificação.

No caso em que foi utilizada uma sequência de cDNA para o desenho de *primers*, ou seja, quando não foi possível prever o tamanho dos fragmentos amplificados a partir do DNA genômico, uma PCR adicional foi efetuada para antever o tamanho aproximado do transcrito primário do gene. Nesse caso, os *primers forward* e *reverse* localizados mais próximos a extremidade da sequência, 5' e 3' respectivamente, foram utilizados para amplificar um fragmento que incluísse todos os exons e introns do gene. Para tanto, utilizou-se o kit GoTaq[®] Long PCR Master Mix, o qual possibilitou a amplificação eficiente do fragmento. A amplificação foi realizada em dois estágios. Após uma desnaturação inicial de 1 min. a 95 °C, o primeiro estágio foi composto por sete ciclos, com desnaturação por 15 s a 95 °C, *annealing* de 50 s com temperatura iniciando em 62 °C e tendo uma redução de 1 °C a cada ciclo, e extensão de 3 min. a 68 °C. O segundo estágio foi programado para conter 25 ciclos, com desnaturação de 15 s a 94 °C, *annealing* de 50 s a 55 °C e extensão de 3 min. a 68 °C, ao final dos quais ocorreu uma incubação adicional de 5 min. a 68 °C.

Modificações específicas foram necessárias para amplificação do FIG NFRO2, as quais incluíram as seguintes condições de amplificação: uma desnaturação inicial por 2 min. a 94 °C, seguidos por 30 ciclos compostos por

desnaturação a 94 °C durante 15 s e *annealing*/extensão a 65 °C durante 3 min., ao final dos quais ocorreu uma incubação adicional de 10 min. a 72 °C.

3.4.2 Purificação e sequenciamento dos FIGs

A purificação dos produtos de PCR foi feita usando-se o Illustra GFX PCR Purification Kit[®] (GE Healthcare), conforme as especificações do fabricante. As concentrações dos produtos purificados foram estimadas por espectrofotometria, usando-se o equipamento NanoDrop 2000 - Thermo Scientific[®], e o sucesso do procedimento de purificação foi verificado a partir da eletroforese de alíquotas de 5 µl desses produtos em gel de agarose (0,8% p/v), corado com SYBR SAFE[®] (Invitrogen). Neste caso, a quantificação das amostras foi estimada a partir da comparação com alíquotas de DNA do fago λ de concentrações conhecidas. Os géis foram visualizados sob luz UV e fotodocumentados pelo sistema de imagem digital MultiDoc-It - UVP[®].

Posteriormente, foi efetuada uma normalização da concentração dos produtos da purificação, de acordo com instruções do fabricante do *kit* de reação de sequenciamento (BigDye[®] Terminator v3.1 Cycle Sequencing Kit), a fim de garantir uma relação ótima entre o tamanho do fragmento e a quantidade a ser usada nas reações. As reações de sequenciamento foram feitas usando-se diretamente 30 a 40 ng do produto de PCR purificado para a incorporação de dNTPs marcados, ou seja, não foi necessária a clonagem dos mesmos. As reações foram analisadas por eletroforese no equipamento ABI 3730 DNA Analyser.

3.5 Obtenção dos Fragmentos das Extremidades do Gene (FEG)

Para a obtenção de sequências complementares às regiões terminais dos genes, foi empregada uma adaptação da técnica de *PCR-walking*, descrita por Cottage et al., (2001). Nesta técnica, fragmentos de DNA, resultantes da hidrólise por restrição enzimática, são ligados a adaptadores especificamente desenhados e então, amplificados seletivamente. O adaptador é composto por dois oligonucleotídeos: um que serve como fita longa do adaptador (5'-ACTCGATTCTCAACCCGAAAGTATAGATCCCA-3') e outro como fita curta do adaptador (3'-H₂N TTCATATCTAGGGT-5'). O primeiro oligonucleotídeo contém uma

região complementar ao sítio de *annealing* do *primer* do adaptador (5'-ACTCGATTCTCAACCCGAAAG-3'), ao passo que o segundo oligonucleotídeo possui a extremidade 3' bloqueada por um radical amina. Assim, no primeiro ciclo de amplificação, o sítio de *annealing* do *primer* do adaptador não existe, o que garante a extensão da primeira fita apenas a partir do *primer* âncora (fase linear da PCR). Esta primeira extensão ocorre até o final do fragmento, que termina com a fita longa do adaptador, o que gera, assim, o sítio de *annealing* do *primer* do adaptador. Então, a partir do segundo ciclo, ocorre a extensão da segunda fita (fase exponencial da PCR). Um esquema ilustrativo é apresentado na Figura 5.

Figura 5 - Esquema de amplificação utilizando uma adaptação do método *PCR-walking* para identificação de sequências alvo. Os fragmentos de DNA, resultantes de uma restrição enzimática, são ligados a adaptadores, cuja extremidade 3' da fita curta encontra-se bloqueada pela presença de um radical amina. A amplificação de sequências alvo envolve duas etapas: no primeiro ciclo da PCR, a extensão ocorre até o final do fragmento que termina com a fita longa do adaptador, o que gera, assim, o sítio de *annealing* do *primer* do adaptador. Em seguida, a partir do segundo ciclo, ocorre a extensão da segunda fita

Fonte: Adaptado de Lunardi et al. (2009)

Inicialmente, o DNA genômico dos 32 acessos foi digerido com três enzimas de restrição. Optou-se por usar aquelas enzimas que geram fragmentos abruptos: *AluI* (New England BioLabs®) e *RsaI* (New England BioLabs®), que reconhecem 4 pb e *DraI* (Fermentas™) que reconhece 6 pb. Um total de 200 ng de DNA foi digerido com 6 U de enzima, em um volume de 20 µl, seguindo as condições do fabricante. A hidrólise ocorreu durante 4 h a 37 °C, com posterior inativação das enzimas a 80 °C durante 20 min.

O adaptador foi preparado pela mistura equimolar dos dois oligonucleotídeos, para uma concentração final de 40 µM. A mistura foi aquecida por 10 min. a 65 °C e resfriada, lentamente, a temperatura ambiente.

A seguir, o adaptador foi ligado às extremidades dos fragmentos gerados pela restrição. O volume final da reação de ligação totalizou 40 µl, contendo 20 µl do DNA digerido, 1,25 µl de adaptador (40 µM), 2 µl de tampão de ligação (10X), 1,2 U de ligação T4-DNA ligase (New England BioLabs®) e água ultra pura (Direct-Q® UV, Millipore) e esterilizada para completar o volume. A ligação se deu em 14 horas a 16 °C, com posterior inativação da enzima a 65 °C durante 20 min.

3.5.1 Condições de amplificação e detecção dos FEGs

As reações de amplificação foram realizadas em um volume de 20 µl, contendo 3 µl de DNA digerido e ligado aos adaptadores, 1,5 mM de MgCl₂, 0,2 mM de dNTP, 0,25 µM do *primer* âncora, 0,25 µM do *primer* do adaptador, 10X do tampão de PCR, 1 U da enzima Taq DNA polimerase (Promega®) e água ultra pura (Direct-Q® UV, Millipore) e esterilizada para completar o volume.

A amplificação foi realizada em dois estágios. Após uma desnaturação inicial de 5 min. a 94 °C, seguiu-se sete ciclos, com desnaturação por 45 s a 94 °C, *annealling* de 50 s com temperatura iniciando em 62 °C e tendo uma redução de 1 °C a cada ciclo, e extensão de 1 min. a 72 °C. No segundo estágio adotou-se 25 ciclos, com desnaturação de 45 s a 94 °C, *annealling* de 50 s a 55 °C e extensão de 1 min. a 72 °C, ao final dos quais ocorreu uma incubação adicional de 7 min. a 72 °C.

Os produtos da PCR foram desnaturados pela adição de igual volume de tampão de corrida desnaturante (0,2% de azul de bromofenol, 0,2% de xileno cianol,

10 mM EDTA, pH 8,0, e 95% de formamida) e aquecimento a 95 °C por 5 min. Uma alíquota de 3 µl de cada amostra desnaturada foi então aplicada em gel de poliacrilamida [acrilamida/bisacrilamida (19:1) 6%, uréia 7,5 M, tampão TBE 1X] de 0,4 mm de espessura, montado em aparato de eletroforese vertical de 38 x 50 cm (Sequi-Gen® GT Sequencing Cell, Bio-Rad), o qual foi submetido a uma potência constante de 80 W e temperatura de 45 °C por cerca de 4 h. *Ladders* de 25 e 100 pb foram usados como marcadores de massa molecular.

A coloração foi realizada com nitrato de prata (AgNO₃), de acordo com Creste, Tulmann Neto e Figueira (2001), com algumas modificações descritas por Santini (2009). Os géis foram analisados sob fonte de luz branca e digitalizados em *scanner* (Epson Expression 10000 XL).

3.5.2 Seleção, purificação e clonagem dos FEGs

Após a separação dos fragmentos por eletroforese, as marcas monomórficas de todos os acessos, e que apresentaram um tamanho entre 200 e 700 pb, foram excisadas do gel de poliacrilamida para posterior clonagem e sequenciamento. Os fragmentos excisados foram eluídos em 50 µl de TE 10:1 (10 mM Tris-HCl, pH 8,0; 1mM EDTA, pH 8,0) e, posteriormente, transferidos para colunas de filtragem e submetidos à centrifugação a 6.000 rpm durante 5 min., em centrífuga de mesa (USA Scientific®). O DNA recuperado foi utilizado em uma PCR com o intuito de reamplificar cada fragmento e confirmar os seus tamanhos.

As reações de amplificação totalizaram um volume de 50 µl, contendo 10 µl de DNA, 0,25 µM do *primer* “âncora”, 0,25 µM do *primer* do adaptador, 25 µl de *MasterMix Go Taq Flexi* (Promega®) e água ultra pura (Direct-Q® UV, Millipore) e esterilizada para completar o volume. As condições de amplificação foram as mesmas descritas no item 3.5.1. Uma alíquota de 5 µl das reações foi submetida à eletroforese em gel de agarose (1,0% p/v), adicionado com SYBR SAFE® (Invitrogen). Foi utilizado *Ladder* de 100 pb como marcador de massa molecular. Os géis foram visualizados sob luz UV e fotodocumentados pelo sistema de imagem digital MultiDoc-It-UVP®. O volume restante das reações foi submetido à purificação, utilizando-se o Illustra GFX PCR Purification Kit® (GE Healthcare), conforme especificações do fabricante. As concentrações dos fragmentos purificados foram estimadas por espectrofotometria, usando-se o equipamento NanoDrop 2000 -

Thermo Scientific[®], e o sucesso do procedimento de purificação foi verificado a partir de eletroforese de alíquotas de 5 µl desses produtos em gel de agarose (0,8% p/v), corado com SYBR SAFE[®] (Invitrogen).

Os produtos de purificação foram, então, ligados ao vetor de clonagem *pGem-T Easy Vector System* (Promega[®]). A ligação se deu em volume de 20 µl, contendo 10 µl de tampão de ligação (2x), 25 ng do vetor, 6 µl dos fragmentos de DNA, 3 U de enzima T4 DNA ligase (Promega[®]) e água ultra pura (Direct-Q[®] UV, Millipore) e esterilizada para completar o volume. A concentração do DNA utilizada na ligação foi normalizada para garantir a relação ótima entre o tamanho do inserto e a quantidade de vetor, conforme as especificações do fabricante do vetor. A ligação se deu *overnight* a 4 °C.

O produto da ligação foi usado para transformar células eletrocompetentes de *Escherichia coli* DH5α (HANAHAN, 1983). A transformação foi realizada em volume de 72 µl, contendo 70 µl de células eletrocompetentes e 2 µl do produto de ligação, seguida de incubação em meio SOC [20 g/l de triptona, 5 g/l de extrato de levedura, 0,5 g/l de NaCl, pH 7,5] por 1 h a 37 °C em agitação de 200 rpm. Em seguida, 60 µl do produto da transformação, 60 µl de IPTG 20% (0,2 g.ml⁻¹ de IPTG) e 60 µl XGAL 2% (20 mg.ml⁻¹ de X-GAL; 5 ml de formamida), foram plaqueados em meio LB sólido (SAMBROOK; RUSSELL, 2001) acrescido de 1 µl/ml de ampicilina.

As placas foram incubadas em BOD a 37 °C por cerca de 18 h para o crescimento das colônias e, então, armazenadas a 4 °C até o dia seguinte, o que possibilitou a diferenciação entre colônias transformadas (coloração branca) e não transformadas (coloração azul). Foram selecionadas quatro colônias brancas de cada placa, as quais foram transferidas para placas do tipo *deep well*, contendo 200 µl de meio *circle grow*, acrescido de 1 µl/ml de ampicilina. A incubação se deu *overnight* a 37 °C.

3.5.3 Miniprep e sequenciamento

As *minipreps* para a purificação dos insertos foram realizadas por lise alcalina (SAMBROOK; RUSSELL, 2001). Uma PCR usando os *primers* pUC/M13 -40F (5'-GTTTTCCCAGTCACGAC-3') e pUC/M13 -46R (5'-GAGCGGATAACAATTTTCACAGG-3'), complementares às regiões do vetor de

clonagem (*pGem-T Easy Vector System* - Promega®) foi realizada para a verificação dos insertos.

Essas reações se deram em um volume de 25 µl, contendo 2 µl de DNA, 1,5 mM de MgCl₂ (Promega), 0,2 mM de cada dNTP (GE Healthcare Life Sciences), 2 µM de cada *primer*, 1X do tampão de PCR (GoTaq® Flexi Buffer, Promega), 1U da enzima Taq DNA polimerase (*Go Taq Flexi* - Promega®) e água ultra pura (Direct-Q® UV, Millipore) e esterilizada para completar o volume.

Os 35 ciclos de amplificação ocorreram após uma desnaturação inicial de 5 min. a 94 °C, com desnaturação de 30 s a 94 °C, *annealing* de 60 s a 55 °C, e extensão de 90 s a 72 °C; ao final dos ciclos ocorreu uma incubação adicional de 10 min. a 72 °C. A separação e verificação do tamanho dos fragmentos foram realizadas por eletroforese em gel de agarose (1,0% p/v), utilizando como padrão de massa molecular o marcador *Ladder* 100 pb (Invitrogen).

Alíquotas do produto de *miniprep* foram submetidas a reações de sequenciamento, usando-se o BigDye® Terminator v3.1 Cycle Sequencing Kit para a incorporação de dNTPs marcados e continham cerca de 100 ng de fragmento como *template*. As reações foram analisadas por eletroforese no equipamento ABI 3730 DNA Analyser.

3.6 Análise dos dados

3.6.1 Processamento e alinhamento das sequências

Após a obtenção das sequências, o *software* CodonCode Aligner v. 3.5.7 (CodonCode Corporation) foi utilizado para a visualização dos eletroferogramas e limpeza das sequências, ou seja, eliminação das extremidades e de regiões de baixa qualidade. A qualidade dos picos foi verificada através do algoritmo Phred/Phrap/Consed (EWING et al., 1998), implementado no *software* citado, que fornece um índice para cada base. Os índices com valores acima de 20 são considerados aceitáveis, pois representam no mínimo 99% de acurácia na determinação das bases, qualidade suficiente para as análises subsequentes. No caso das sequências dos fragmentos que foram submetidos à clonagem, as regiões que correspondiam ao vetor usado foram identificadas e eliminadas.

Ainda usando o CodonCode Aligner v. 3.5.7 (CodonCode Corporation), foram formados os *contigs* com as sequências correspondentes a cada acesso. Posteriormente, os *contigs* obtidos para cada acesso foram alinhados usando o algoritmo de alinhamento global múltiplo CLUSTAL W (THOMPSON et al., 1994), o qual leva em consideração os valores de qualidade das bases e um valor limite de 70% de identidade. Este alinhamento permitiu que os SNPs e indels candidatos fossem identificados, preliminarmente, por meio de inspeção visual, corrigindo-se, manualmente, eventuais equívocos de leitura. As sequências processadas foram novamente submetidas ao alinhamento global múltiplo CLUSTAL W (THOMPSON et al., 1994), desta vez utilizando o *software* MEGA 5.1 (TAMURA et al., 2011), uma vez que este permite exportar o arquivo de saída, gerado pelo alinhamento, em formatos requeridos pelo *software* usado para calcular certos parâmetros moleculares (item 3.6.3).

A validação das sequências obtidas para cada acesso foi realizada por algoritmo de alinhamento local em base de dados de nucleotídeos (BlastN) implementado no site do NCBI (ALTSCHUL et al., 1997).

3.6.2 Verificação de saturação e escolha do modelo evolutivo

Antes de realizar a análise de agrupamento, a qualidade da informação filogenética contida nos dados obtidos foi verificada quanto ao nível de saturação de substituições, usando o *software* DAMBE (XIA; XIE, 2001). O modelo de substituição de bases empregado foi o K80 (KIMURA, 1980) contra a taxa de transversão (V) e transição (S). Gráficos foram gerados a partir da sequência de cada gene, bem como para a combinação de ambos os genes, e foram analisados visualmente.

A escolha do modelo evolutivo que fundamentou a construção das árvores de distância genética a partir das sequências de DNA foi realizada pelo *software* MEGA 5.1 (TAMURA et al., 2011).

3.6.3 Parâmetros moleculares e medidas de variabilidade genética

O número total de locos polimórficos, bem como o número de substituições, seja do tipo transversão ou transição, o número de inserções/deleções, a diversidade nucleotídica, incluindo π (NEI, 1987) e Θ (WATTERSON, 1975) foram

determinados pelo *software* MEGA 5.1 (TAMURA et al., 2011), o qual também foi usado para estimar as distâncias genéticas entre os acessos.

3.6.4 Caracterização do polimorfismo nos genes α -Phs e *FRO1*

A delimitação das regiões correspondentes aos exons e introns foi determinada *in silico* através de alinhamento local entre as sequências do gene completo e do respectivo cDNA. As sequências correspondentes ao gene *FRO1* foram submetidas ao alinhamento pelo algoritmo BlastN em base de dados de nucleotídeos implementado no site do NCBI (ALTSCHUL et al., 1997), uma vez que apenas uma sequência do cDNA deste gene está depositada. Já as sequências referentes ao gene α -Phs e ao seu cDNA foram submetidas ao alinhamento pelo algoritmo de Smith-Waterman, utilizando o aplicativo web EMBOSS Water (RICE; BLEASBY; ISON, 2000; disponível em: <[HTTP://http://www.ebi.ac.uk/Tools/psa/](http://www.ebi.ac.uk/Tools/psa/)>). Em seguida, as sequências de cada exon e intron foram submetidas separadamente aos cálculos dos parâmetros moleculares citados no item 3.6.3, através do *software* Arlequin v.3.5 (EXCOFFIER; LISCHER, 2010).

3.6.5 Metodologia de agrupamento

Os dendrogramas das relações entre os acessos foram obtidos usando o *software* MEGA 5.1 (TAMURA et al., 2011), empregando-se o critério de *Neighbor-joining* (NJ) e o teste não-paramétrico de *bootstrap* para estimar a confiabilidade dos ramos, com 500 repetições. Foram construídas árvores de NJ para ambos os genes e para o conjunto de sequências combinados para avaliar a contribuição de cada gene aos resultados finais.

3.6.6 Testes de neutralidade seletiva de locos e análise da variância molecular (AMOVA)

O teste estatístico “D” de neutralidade seletiva de locos, proposto por Tajima (1989) foi calculado pelo programa Arlequin v.3.5 (EXCOFFIER; LISCHER, 2010). Adicionalmente, uma análise de variância molecular foi realizada para cada gene, separadamente, bem como para o conjunto de sequências combinadas, visando

avaliar a estrutura genética dos *pools* gênicos Andino e Mesoamericano, por meio do mesmo *software*. Para tanto, foram conduzidos dois tipos de análises para testar a existência de variação molecular atribuída à estrutura genética: (i) entre os *pools* gênicos e (ii) entre acessos dentro de cada *pool* gênico.

4 RESULTADOS E DISCUSSÃO

4.1 Prospecção *in silico* das sequências de *Phs* e *FRO1* de *P. vulgaris* e desenho de *primers*

Utilizando o termo “*Phaseolus vulgaris*”, foram realizadas buscas nos bancos de dados de nucleotídeos no NCBI, tendo sido identificadas várias sequências depositadas referentes ao gene *Phs*. Entretanto, foram selecionadas apenas duas sequências, referentes ao gene α -*Phs*, as quais exibiam uma anotação bastante detalhada. Uma correspondia ao mRNA (código de acesso X02980.1) e a outra, ao gene completo (código de acesso X52626.1), a qual inclui sua região regulatória, o promotor. Tal sequência serviu para o desenho de *primers*, e permitiu incluir a região do promotor nas análises de diversidade nucleotídica.

O gene α -*Phs* possui 4.763 pb, dos quais 1.535 (32,2%) são transcritos. O alinhamento local entre as duas sequências permitiu identificar que α -*Phs* é composto por seis exons, sendo que o primeiro deles corresponde a uma região que não faz parte da proteína.

Por outro lado, foi encontrada apenas uma sequência referente ao gene *FRO1* de *P. vulgaris* com 2.115 pb (código de acesso HM440564.1). Esta sequência é o cDNA do gene, e foi obtida e depositada por Blair et al. (2010), que também realizaram comparações com genes homólogos de ervilha (*Pisum sativum*) e alfafa (*Medicago sativa*) sugerindo que *FRO1* em *P. vulgaris* está localizado no cromossomo b07 e contém 8 exons e 7 introns, em ambas as comparações.

Ao todo, 17 pares de *primers* foram desenhados, sintetizados e utilizados para amplificar o DNA genômico dos acessos de *P. vulgaris*, sendo identificados da seguinte forma: os *primers* *Phs* correspondem àqueles desenhados a partir da sequência de α -*Phs*, e os *primers* *FRO*, aos desenhados a partir da sequência de *FRO1* (Tabela 2).

Tabela 2 - Lista de primers desenhados a partir de sequências de α -Phs e FRO1

Nome	TE	TO	Sequência do <i>primer</i> (5'-3')	TA
Phs1F	655	650	CCCCAACCAAGATGAACAC	63
Phs1R			TTGTCATGTGTTGACCCTTG	62
Phs2F	516	510	CACCCAACCAAAATAGCTTC	61
Phs2R			CCTTTTTTCCTGTGTTCTTACC	59
Phs3F	522	500	GGAACAAAAACGGAACGAAC	63
Phs3R			CAAAGTGTCCAACACCTCG	62
Phs4F	564	600	GTGAAAACCATCACCGTCC	63
Phs4R			AACGACTACAGAGTGCCATG	60
Phs5F	536	550	ATGCCATTCAAACCCAC	61
Phs5R			GGTGGCGGAAGAGAATG	62
Phs6F	544	550	ACCATCTCAACCCACACAC	61
Phs6R			TGGACCTGAACTCCACAAG	61
Phs7F	513	500	GAGGTTGACCAACAATCC	62
Phs7R			GTTGGGCTTCTGTGCTAG	58
Phs8F	619	650	AACAACCCTCAGATTCATG	57
Phs8R			TTTTCTAACCAAGACGACAG	57
Phs9F	613	600	GACCGATAACTCCTTGAATG	59
Phs9R			AGAGAACGTAAGCCCCAAC	61
Phs10F	615	600	GAAGGTAAGACGGACAATG	57
Phs10R			CACATAAAACACAAGCGTC	57
FRO1F	-	950	GAACCCCAAACCTCGAAAAC	61
FRO1R			TGATGCTTGACTCTGATGTG	60
FRO2F	-	**	GTGAAAAAGTGTGGCAAGC	61
FRO2R			ATGCGAGCTGATAGCAATG	62
FRO3F	-	700	CCTCGTTGACAGACACCTC	61
FRO3R			TGAGCCAGAAACAGAAAGC	61
FRO4F	-	550	GAGCTTGTCTTGGTGAGTG	58
FRO4R			TCCACCAGAGAAAACTGC	57
FRO5F	-	500	GGGCGAAGTTTATCATTGG	62
FRO5R			CATGCAAAGCTGATGGATTG	63
PhsR*	-	-	GCGGTATGTGGTGATTTTCAG	63
PhsF*	-	-	TGCACCTTCACTAGTAATAC	52
FROR*	-	-	GGGAATGTGAAAAGCAAGAG	61
FROF*	-	-	CCAAGGAACATGAGACATG	59

* *Primers* âncoras utilizados na amplificação dos FEGs.

** Não houve amplificação.

TE: Tamanho esperado do fragmento (pb); TO: Tamanho observado do fragmento (pb), determinado a partir de comparação com marcador *Ladder* 100 pb; TA: Temperatura de *annealing* (°C)

4.2 Obtenção dos FIGs

Todos os *primers* desenhados a partir da sequência dos genes α -Phs e FRO1 amplificaram fragmentos únicos, cujos tamanhos corresponderam ao predito, com exceção do fragmento FRO2, do qual não se obteve amplificação.

Uma vez que a sequência molde de *FRO1* correspondia ao mRNA, esta não incluía as bases correspondentes aos introns (regiões não traduzidas). A região escolhida pelo *software* Primer 3, como sendo ótima para o desenho do *primer* FRO2F, compreendia a junção de dois éxons, o que impossibilitou, portanto, o *annealing* do *primer* no DNA genômico, prejudicando a amplificação. Após realizar um alinhamento local entre a sequência do FIG FRO1 e a sequência de cDNA do *FRO1*, foi constatado que as primeiras 13 bases do *primer* FRO2F encontram-se a 78 pb *downstream* das últimas seis bases.

Uma vez que os *primers* foram desenhados para permitir a sobreposição de regiões terminais, a extremidade 5' da sequência do FIG FRO1 foi usada para o desenho de um novo *primer* denominado NFRO2F (5'-AAGCAAAGTTGAGGAGTGTG-3'), que substituiu o *primer* FRO2F nas reações de amplificação. Todavia, mesmo garantindo que o *primer* NFRO2F seria capaz de se *annealing* completo, foram encontradas dificuldades para amplificar o FIG FRO2.

O resultado da amplificação com os *primers* FRO1F e FRO5R, o qual pode ser observado na figura 6, permitiu prever o tamanho da região transcrita do gene *FRO1*, a qual corresponde a, aproximadamente, 5 kb. Desse modo, pôde-se inferir que o FIG FRO2 correspondia a um fragmento de aproximadamente 2,3 kb e que as condições de PCR aplicadas, principalmente a pouca eficiência da enzima Taq DNA polimerase e o tempo dedicado à extensão do fragmento durante os ciclos da PCR, não permitiam sua amplificação. Posteriormente, o tamanho real do FIG FRO2 pôde ser confirmado a partir da utilização do kit GoTaq® Long PCR Master Mix.

Figura 6 - Produto de amplificação (FRO em duplicata) resultante do uso dos *primers* FRO1F e FRO5R com, aproximadamente, 5.5 Kb. O tamanho do amplicon corresponde ao tamanho do putativo transcrito primário do gene *FRO1* em *P. vulgaris*. M1 e M2 são marcadores *Ladder* 1 kb e 100 pb, respectivamente

4.2.1 Validação dos FIGs

Todas as sequências correspondentes aos FIGs foram submetidas à checagem de qualidade de seus picos no *software* CodonCode Aligner, e aquelas que exibiram índices de *Phred* inferior a 20 foram eliminadas. Alinhamentos foram executados para as sequências dos dois genes em cada um dos 32 acessos separadamente, o que permitiu a obtenção de *contigs* individuais. Como pode ser observado na figura 7, para cada um dos acessos foi obtido um *contig* de 4.466 pb correspondente ao α -*Phs* e dois *contigs* correspondentes ao *FRO1* que, somados, contem 2.736 pb. Apesar de não se ter alcançado toda a sequência do FIG FRO2, esses *contigs* contem toda a região codante de ambos os genes.

A validação dos FEGs foi realizada através da comparação dos *contigs* obtidos com as sequências presentes na base de dados do NCBI, por BlastN. Os resultados dos alinhamentos indicaram que todas as sequências obtidas com os *primers* Phs (1-10) correspondiam ao gene α -*Phs*, bem como aquelas obtidas com os *primers* FRO (1-5) eram similares às do gene *FRO1*.

Figura 7 - Representação esquemática da formação de contigs a partir do alinhamento de FIGs correspondentes aos genes α -Phs e FRO1. Setas direcionadas para a direita indicam sequenciamento *forward*, e para a esquerda, *reverse*

4.3 Obtenção dos FEGs

Ao todo, foram testadas 12 combinações de três enzimas de restrição com quatro primers âncoras, visando à obtenção de fragmentos correspondentes às extremidades 5' e 3' dos genes *α -Phs* e *FRO1*. A seleção das melhores combinações se baseou no padrão de amplificação e no número de bandas bem resolvidas entre os 32 acessos de *Phaseolus*. O padrão de amplificação das combinações escolhidas, bem como os fragmentos selecionados para serem sequenciados, encontra-se exposto na figura 8.

Figura 8 - Padrão de amplificação das combinações de enzima de restrição e primers âncora em gel de poliacrilamida (6% p/v) desnaturante, corado com nitrato de prata. Primeira e última canaletas contém *Ladder* de 100+25 pb. As demais canaletas correspondem ao produto de amplificação de cada combinação dos acessos de *P. vulgaris* (Sanilac) em triplicata e de *P. lunatus* (87 JP 12). As setas indicam os fragmentos que foram excisados do gel. Setas pretas correspondem às bandas monomórficas entre *P. vulgaris* e *P. lunatus*, e as vermelhas indicam as bandas monomórficas apenas entre os acessos de *P. vulgaris*

Um total de 41 fragmentos foi excisado do gel de poliacrilamida, eluído em TE, purificado em coluna e reamplificado de acordo com os procedimentos descritos no item 3.5.1. Deste total, 21 fragmentos apresentaram-se monomórficos entre *P. vulgaris* e *P. lunatus*, e 20 fragmentos exibiram monomorfismo apenas entre os acessos de *P. vulgaris*. Após a reamplificação, 27 fragmentos exibiram uma única banda, a qual correspondia ao tamanho do fragmento utilizado como *template* na reação (Figura 9). Estes foram, portanto, selecionados para serem sequenciados.

Figura 9 - Padrão de reamplificação das bandas excisadas, em gel de agarose corado com Sybr Safe. 1: Fragmentos monomórficos entre *P. vulgaris* e *P. lunatus*; 2: fragmentos dos acessos de *P. vulgaris*. A primeira e a última canaletas correspondem ao marcador *Ladder* 100pb. I: a) FROR/*AluI*, b) FROR/*DraI* e c) FROR/*RsaI*; II: PhsF/*RsaI*; III: PhsR/*DraI*; IV: FROF/*DraI*. As setas indicam os fragmentos descartados por apresentarem amplificação inespecífica

Após a purificação, clonagem e sequenciamento dos FEGs selecionados, foram originadas duas leituras de sequenciamento para cada FEG, ou seja, foram analisadas 52 sequências. Depois da verificação de qualidade das bases de cada sequência, aproveitaram-se 36 para o alinhamento, as quais correspondiam a 18 FEGs, sendo duas leituras para cada. O tamanho final das sequências, em todos os casos, foi igual ao tamanho do fragmento excisado do gel de poliacrilamida. Os *contigs* obtidos a partir de cada par de sequências foram utilizados para a validação.

Os resultados dos alinhamentos, resumidos na tabela 3, apontaram que nenhum FEG obtido a partir de amplificações com os *primers* PhsF, PhsR e FROF apresentou similaridade com sequências depositadas de α -Phs e *FRO1*.

Tabela 3 - Resumo das análises via BlastN dos *contigs* obtidos para 14 fragmentos das extremidades dos genes α -Phs e FRO1

FEG ID	Acesso GenBank	Descrição	Identidade (%)	Score	E-values
FROR/ <i>Alu</i> _1	GU215964.1	<i>P. vulgaris</i> clone	85	185	1 ⁻⁴³
FROR/ <i>Alu</i> _2	GU215964.1	<i>P. vulgaris</i> clone	82	298	1 ⁻⁷⁷
FROR/ <i>Alu</i> _3	GU215964.1	<i>P. vulgaris</i> clone	83	306	9 ⁻⁸⁰
FROR/ <i>Alu</i> _4	AF325187.1	<i>P. coccineus</i> putative reverse transcriptase	89	66	2 ⁻⁷
FROR/ <i>Dra</i> _1	HQ993069.1	<i>Juglans nigra</i> clone B unknown	97	52	2 ⁻³
FROR/ <i>Dra</i> _2	GU215964.1	<i>P. vulgaris</i> clone	83	183	4 ⁻⁴³
FROR/ <i>Dra</i> _3	HM440564.1	<i>P. vulgaris</i> FRO1mRNA	100	268	4 ⁻⁶⁸
FROR/ <i>Rsa</i> _1	JN375330.1	<i>Phoenix dactylifera</i> (mitocôndria)	94	360	6 ⁻⁹⁶
FROF/ <i>Dra</i> _1	EU196765.1	<i>P. vulgaris</i> (cloroplasto)	100	509	4 ⁻¹⁴¹
FROF/ <i>Dra</i> _2	EU196765.1	<i>P. vulgaris</i> (cloroplasto)	100	511	1 ⁻¹⁴¹
FROF/ <i>Dra</i> _3	AF325187.1	<i>P. coccineus</i> putative reverse transcriptase	90	536	4 ⁻¹⁴⁹
FROF/ <i>Dra</i> _4	XM_003551623.1	<i>Glycine Max</i> proteína putativa	91	187	9 ⁻⁴⁴
FROF/ <i>Dra</i> _5	GQ279301.1	<i>P. vulgaris</i> clone	77	226	1 ⁻⁵⁵
PhsF/ <i>Rsa</i> _1	EU196765.1	<i>P. vulgaris</i> (cloroplasto)	99	459	6 ⁻¹²⁶
PhsF/ <i>Rsa</i> _2	EU196765.1	<i>P. vulgaris</i> (cloroplasto)	99	616	5 ⁻¹⁷³
PhsF/ <i>Rsa</i> _3	AP009647.1	<i>Lotus japonicus</i> clone	76	48.2	3.9 ⁻²
PhsR/ <i>Dra</i> _1	EU196765.1	<i>P. vulgaris</i> (cloroplasto)	99	605	8 ⁻¹⁷⁰
PhsR/ <i>Dra</i> _2	HQ840770.1	<i>P. vulgaris</i> clone	83	214	4-52

Em contrapartida, dos oito FEGs obtidos a partir do *primer* FROR, apenas um, denominado FROR/DraI_3, apresentou 100% de identidade com a sequência de cDNA do gene *FRO1*. Esse fragmento revelou 202 bases localizadas à montante da primeira adenina da sequência depositada no banco de dados, a qual remete, provavelmente, ao final da UTR 5' do gene *FRO1*, e permitiu confirmar o tamanho em pb do intron 1 (Figura 10). Uma vez que apenas uma das quatro regiões buscadas foi obtida, os polimorfismos relativos à extremidade 5' do gene *FRO1* ainda não foram investigados para todos os acessos estudados no presente trabalho.

Figura 10 - Representação gráfica do resultado do alinhamento local (BlastN) entre a sequência do FEG FROR/DraI_3 e o banco de nucleotídeos do NCBI

Todavia, outros estudos que se utilizaram de adaptações da técnica de PCR *walking*, na ocasião da validação dos fragmentos obtidos, constataram que algumas sequências não apresentavam similaridade com as regiões alvo e/ou com sequências depositadas no banco de dados utilizado para a validação. Por conseguinte, vários autores têm demonstrado que, apesar do sucesso na obtenção de fragmentos cujas sequências correspondem àquelas regiões de interesse, a técnica também amplifica fragmentos indesejados.

Adaptações da técnica proposta por Cottage et al. (2001) já foram empregadas para a obtenção de sequências que flanqueiam regiões específicas em fungos (LUNARDI et al., 2009). Outra técnica similar, denominada *NBS-profiling*, permite a obtenção de sequências correspondentes a regiões análogas a genes de resistência em plantas, as quais podem ser utilizadas como marcadores RGA (VAN DER LINDEN et al., 2004; CALENCE et al., 2005; BRUGMANS et al., 2008; HANAI et al., 2010). Também há registros de adaptações da técnica *NBS-profiling* para amplificarsequências ancoradas aregiões de elementos transponíveis de cana-de-açúcar (PALHARES et al., 2012).

É importante ressaltar que tanto as regiões análogas a genes de resistência quanto de elementos transponíveis, por exemplo, são múltiplas no genoma. Regiões conservadas entre as sequências alvo possibilitam o desenho de *primers* degenerados, os quais ampliam a quantidade de sítios de *annealing* e, consequentemente, permitem a amplificação de mais de uma região alvo numa mesma PCR.

Uma provável explicação para o baixo rendimento na obtenção de sequências alvo no presente caso seria um desbalanceamento na proporção de cópias da sequência alvo no genoma em relação à quantidade de sequências que são amplificadas aleatoriamente pela técnica, associado à provável falta de especificidade dos *primers* utilizados. Deste modo, uma alternativa para obter sucesso na obtenção de sequências alvo com poucas cópias no genoma seria ampliar a amostragem de fragmentos a serem sequenciados ou associar as técnicas PCR-*walking* e *nested*-PCR numa tentativa de se obter amplicons mais específicos. No último caso, o desenho de outros *primers* se faz necessário.

4.4 Caracterização do polimorfismo

As sequências correspondentes aos contigs de *α -Phs* e *FRO1*, os quais continham 4.466 pb e 2.736 pb, respectivamente, exibiram um total de 361 bases polimórficas que incluíram 260 substituições e 101 indels. É importante ressaltar que estes polimorfismos foram detectados entre todos os acessos de *Phaseolus* aqui estudados. Ao investigar o polimorfismo intraespecífico para *P. vulgaris*, ou seja, excluindo o acesso de *P. lunatus* da análise, verificou-se a existência de 213 bases polimórficas, das quais 150 eram substituições e 63, indels.

O número de ocorrências e as frequências de polimorfismos em cada um dos genes estudados, bem como para o conjunto combinado de dados encontram-se detalhados nas tabelas 4 e 5, respectivamente.

Tabela 4 - Resumo das análises de polimorfismo para sequências de α -Phs e FRO1 de acessos de *P. vulgaris* e *P. lunatus*

Parâmetro	<i>α-Phs</i>						<i>FRO1</i>					
	Cod	Cod*	n-Cod	n-Cod*	Total	Total*	Cod	Cod*	n-Cod	n-Cod*	Total	Total*
Transições	20	19	69	24	89	43	27	13	30	23	57	36
Transversões	27	27	44	19	71	46	16	9	27	16	43	25
Substituições	47	46	113	43	160	89	43	22	57	39	100	61
Indels	17	17	31	21	48	38	5	3	48	22	53	25
Sítios polimórficos	64	63	144	64	208	127	48	25	105	61	153	86
Frequência de transições	0,014	0,013	0,023	0,008	0,020	0,010	0,015	0,007	0,034	0,027	0,021	0,014
Frequência de transversões	0,018	0,018	0,015	0,006	0,016	0,010	0,009	0,005	0,031	0,018	0,016	0,009
Frequência de substituições	0,032	0,031	0,038	0,014	0,036	0,020	0,024	0,012	0,065	0,045	0,037	0,023
Frequência de indels	0,012	0,012	0,010	0,007	0,011	0,009	0,003	0,002	0,055	0,025	0,019	0,009
Frequência de sítios polimórficos	0,044	0,043	0,048	0,021	0,047	0,028	0,027	0,013	0,119	0,070	0,056	0,032
Total de bases avaliadas	1462	1462	3004	2997	4466	4459	1775	1870	879	866	2736	2654

* Valores obtidos a partir da análise de polimorfismo intraespecífico, desconsiderando *P. lunatus*

Cod: Região codante da sequência

n-Cod: Região não codante da sequência

Tabela 5 - Resumo das análises de polimorfismo para sequências combinadas de α -Phs e FRO1 de acessos de *P. vulgaris* e *P. lunatus*

Parâmetro	<i>FRO1 + α-Phs</i>					
	Cod	Cod*	n-Cod	n-Cod*	Total	Total*
Transições	47	32	99	47	146	79
Transversões	43	36	71	35	114	71
Substituições	90	68	170	82	260	150
Indels	22	20	79	43	101	63
Sítios polimórficos	112	88	249	125	361	213
Frequência de transições	0,03	0,02	0,057	0,0346	0,04	0,023
Frequência de transversões	0,03	0,02	0,045	0,0248	0,03	0,02
Frequência de substituições	0,06	0,04	0,102	0,0594	0,07	0,043
Frequência de indels	0,01	0,01	0,065	0,0324	0,03	0,018
Frequência de sítios polimórficos	0,07	0,06	0,167	0,0918	0,1	0,061
Total de bases avaliadas	3237	3332	3883	3863	7120	7195

* Valores obtidos a partir da análise de polimorfismo intraespecífico, desconsiderando *P. lunatus*

Cod: Região codante da sequência

n-Cod: Região não codante da sequência

Para ambos os genes, a frequência de substituições foi superior a de indels, assim como a frequência de locos polimórficos foi maior quando os acessos de *P. vulgaris* foram contrastados com *P. lunatus*, de acordo com o que teoricamente se esperava.

Todas as substituições encontradas foram bi-alélicas, ou seja, havia apenas duas alternativas de nucleotídeos para cada posição. Foi encontrada uma frequência duas vezes maior de SNPs em regiões não codantes do que em regiões traduzidas. Isto implica que a pressão de seleção nestas regiões é bem mais marcante.

Além disso, alguns tipos de substituições de bases ocorrem com maior frequência do que outros e também estão sujeitas à frequência daquela base no segmento estudado. Ao avaliar a composição nucleotídica das sequências de α -Phs e *FRO1* e, detectou-se que a frequência de C e G foi de 35,2% e 37,7%, respectivamente. Também, verificou-se que as substituições detectadas nas sequências de ambos os genes mostram maior frequência de transições (54.6%), ou seja, a substituição de uma pirimidina por outra pirimidina (C/T), ou de uma purina por outra purina (A/G). No geral, essas mutações por substituição seguiram a normalidade, pois as substituições por transição costumam ocorrer com maior

frequência no genoma de diversos organismos (LEWIN, 2009). Destas, foi observado que ocorre uma maior frequência de transições do tipo A/G em ambos os genes.

As classes de polimorfismo que foram identificadas, incluindo os tipos de substituição e os tamanhos dos indels, encontram-se melhor detalhadas na figura 11.

Figura 11 - Classes dos polimorfismos observados em sequências de α -Phs e FRO1. A: A/G e C/T correspondem a transições e as demais classes, a transversões. B: Número de inserções/deleções (indels) observado classificado de acordo com seu tamanho em pb

Observou-se uma prevalência de indels que se caracterizavam pela inserção/deleção de um ou dois nucleotídeos, principalmente em regiões não codantes das sequências. Também, foram detectados na região não codante de

FRO1, quatro indels que correspondem à inserção/deleção de várias bases consecutivas (7, 13, 14 e 43 pb). Entretanto, estes foram tratados pelos *softwares* utilizados como sendo 77 eventos independentes.

Estes blocos de indels representaram, na maioria dos casos, inserções/deleções de sequências quando as duas espécies eram comparadas. Todavia, também foi observado um indel de 14 pb entre os acessos de *P. vulgaris*, o qual está presente em todos os acessos do grupo gênico Andino, em três acessos do grupo Mesoamericano e em um acesso classificado como “crioulo”. O resultado do alinhamento entre as sequências do intron 2 de *FRO1* pode ser observado na figura 12, na qual se encontram destacados alguns SNPs e o indel mencionado.

	460	470	480	490	500	510	520	530	540
Puebla 152 (M)	GCTCTTGGATAATATTATATAATCTTGT	TACAGAAATA	TAAAAGACATGAAAGAAACATATATAGTTTCTGAGATCTATAAC						
IAC-UNA (M)	GCTCTTGGATAATATTATATAATCTTGT	TACAGAAATA	TAAAAGACATGAAAGAAACATATATAGTTTCTGAGATCTATAAC						
Tu (M)	GCTCTTGGATAATATTATATAATCTTGT	TACAGAAATA	TAAAAGACATGAAAGAAACATATATAGTTTCTGAGATCTATAAC						
BRS Pontal (Cv)	GCTCTTGGATAATATTATATAATCTTGT	TACAGAAATA	TAAAAGACATGAAAGAAACATATATAGTTTCTGAGATCTATAAC						
Jalo EEP558 (A)	GCTCTTGAAATAATTATATAATCTTGT	TACAGAAATA	TAAAAGACATGAAAGAAACATATATAGTTTCTGAGATCTATAAC						
Cal 143 (A)	GCTCTTGAAATAATTATATAATCTTGT	TACAGAAATA	TAAAAGACATGAAAGAAACATATATAGTTTCTGAGATCTATAAC						
Pompadour (A)	GCTCTTGAAATAATTATATAATCTTGT	TACAGAAATA	TAAAAGACATGAAAGAAACATATATAGTTTCTGAGATCTATAAC						
Red Kidney (A)	GCTCTTGAAATAATTATATAATCTTGT	TACAGAAATA	TAAAAGACATGAAAGAAACATATATAGTTTCTGAGATCTATAAC						
Jabola (A)	GCTCTTGAAATAATTATATAATCTTGT	TACAGAAATA	TAAAAGACATGAAAGAAACATATATAGTTTCTGAGATCTATAAC						
Antioquia 8 (A)	GCTCTTGAAATAATTATATAATCTTGT	TACAGAAATA	TAAAAGACATGAAAGAAACATATATAGTTTCTGAGATCTATAAC						
Rio Tibagi (M)	GCTCTTGGATAATATTATATAATCTTGT	TACAGAAATA	-----AAAGAAACACGTATAATTACTGAGATCTGTAAC						
Apetito Blanco (M)	GCTCTTGGATAATATTATATAATCTTGT	TACAGAAATA	-----AAAGAAACACGTATAATTACTGAGATCTGTAAC						
Baetão (M)	GCTCTTGGATAATATTATATAATCTTGT	TACAGAAATA	-----AAAGAAACACGTATAATTACTGAGATCTGTAAC						
Porrillo (M)	GCTCTTGGATAATATTATATAATCTTGT	TACAGAAATA	-----AAAGAAACACGTATAATTACTGAGATCTGTAAC						
Sanilac (M)	GCTCTTGGATAATATTATATAATCTTGT	TACAGAAATA	-----AAAGAAACACGTATAATTACTGAGATCTGTAAC						
BRS Pitanga (Cv)	GCTCTTGGATAATATTATATAATCTTGT	TACAGAAATA	-----AAAGAAACACGTATAATTACTGAGATCTGTAAC						
Jamapa (M)	GCTCTTGGATAATATTATATAATCTTGT	TACAGAAATA	-----AAAGAAACACGTATAATTACTGAGATCTGTAAC						
BRS Vereda (Cv)	GCTCTTGGATAATATTATATAATCTTGT	TACAGAAATA	-----AAAGAAACACGTATAATTACTGAGATCTGTAAC						
Garbancillo (M)	GCTCTTGGATAATATTATATAATCTTGT	TACAGAAATA	-----AAAGAAACACGTATAATTACTGAGATCTGTAAC						
Brasil 2 (M)	GCTCTTGGATAATATTATATAATCTTGT	TACAGAAATA	-----AAAGAAACACGTATAATTACTGAGATCTGTAAC						
Jalo Precoce (Cv)	GCTCTTGGATAATATTATATAATCTTGT	TACAGAAATA	-----AAAGAAACACGTATAATTACTGAGATCTGTAAC						
Red Mexican (M)	GCTCTTGGATAATATTATATAATCTTGT	TACAGAAATA	-----AAAGAAACACGTATAATTACTGAGATCTGTAAC						
Barbunia (M)	GCTCTTGGATAATATTATATAATCTTGT	TACAGAAATA	-----AAAGAAACACGTATAATTACTGAGATCTGTAAC						
Mulatinho (M)	GCTCTTGGATAATATTATATAATCTTGT	TACAGAAATA	-----AAAGAAACACGTATAATTACTGAGATCTGTAAC						
Bat 93 (M)	GCTCTTGGATAATATTATATAATCTTGT	TACAGAAATA	-----AAAGAAACACGTATAATTACTGAGATCTGTAAC						
Carioca (M)	GCTCTTGGATAATATTATATAATCTTGT	TACAGAAATA	-----AAAGAAACACGTATAATTACTGAGATCTGTAAC						
SAF (Cr)	GCTCTTGGATAATATTATATAATCTTGT	TACAGAAATA	-----AAAGAAACACGTATAATTACTGAGATCTGTAAC						
Flor de Mayo (M)	GCTCTTGGATAATATTATATAATCTTGT	TACAGAAATA	-----AAAGAAACACGTATAATTACTGAGATCTGTAAC						
Aporé (Cv)	GCTCTTGGATAATATTATATAATCTTGT	TACAGAAATA	-----AAAGAAACACGTATAATTACTGAGATCTGTAAC						
SAB (Cr)	GCTCTTGGATAATATTATATAATCTTGT	TACAGAAATA	-----AAAGAAACACGTATAATTACTGAGATCTGTAAC						
Great Northern (M)	GCTCTTGGATAATATTATATAATCTTGT	TACAGAAATA	-----AAAGAAACACGTATAATTACTGAGATCTGTAAC						
<i>P. lunatus</i>	GCTCTAGAAATAATTATATAATCTTGT	TACAGAAATA	-----AAAGACATGAAAGAAACACATATAATTACTGACATCTATAAC						

Figura 12 - Alinhamento de sequências de *FRO1* entre acessos de *Phaseolus* evidenciando uma região do intron 2. Destaques em vermelho referem-se aos sítios que exibem variação do tipo substituição e indel. As siglas localizadas entre parênteses remetem à classificação dos acessos. M = Mesoamericano, A = Andino, Cr = Crioulo e Cv = Cultivar Embrapa

Ao avaliar a frequência de polimorfismo ao longo da sequência de cada gene (Figuras 13 e 14), observou-se que as porções iniciais de ambos concentram mais da metade de todas as variações detectadas.

Figura 13 - Distribuição da frequência de transições (S), transversões (V) e inserções/deleções (Indels) ao longo das sequências dos *contigs* de *α-Phs*. As regiões correspondentes a exons são identificadas pela letra “e” e as correspondentes à introns, pela letra “i”. Número observado de polimorfismos entre as espécies de *Phaseolus* (A) e entre os acessos de *P. vulgaris* (B). As barras localizadas abaixo de cada gráfico correspondem a uma representação esquemática da disposição dos exons ao longo da sequência gênica

FRO1

Figura 14 - Distribuição da frequência de transições (S), transversões (V) e inserções/deleções (Indels) ao longo das sequências dos *contigs* de *FRO1*. As regiões correspondentes à exons são identificadas pela letra “e” e as correspondentes à introns, pela letra “i”. Número observado de polimorfismos entre as espécies de *Phaseolus* (A) e entre os acessos de *P. vulgaris* (B). As barras localizadas abaixo de cada gráfico correspondem a uma representação esquemática da disposição dos exons ao longo da sequência gênica

De um modo geral, pode-se dizer que foi encontrado um SNP a cada 20 pb. Considerando as regiões codantes, observou-se um SNP a cada 29 pb, enquanto nas regiões não codantes, observou-se um a cada 16 pb. Essas taxas foram superiores às encontradas em milho (TENAILLON et al., 2001), em soja (ZHU et al., 2003), em cevada (BUNDOCK; HENRY, 2004) e até mesmo para o feijão (CHEN et al., 2010).

Polimorfismos de nucleotídeos foram menos abundantes, quando se comparou ESTs de dois genótipos contrastantes de feijoeiro comum, um Andino e outro Mesoamericano, ocorrendo um SNP a cada 387 pb (RAMÍREZ et al., 2005). Em contraste, quando se comparou a frequência de SNPs entre 92 genótipos de feijão a partir de sequências intrônicas do gene que codifica a di-hidroflavonol-4-redutase (*DFR*), foi encontrado um SNP a cada 8 pb (MCCLEAN; LEE; MIKLAS, 2004). Tais resultados enfatizam que a frequência de SNPs em regiões codantes é bem menor do que naquelas em que não há tradução, reforçando a ideia que a pressão de seleção é mais influente naquelas regiões.

A proporção de indels nas sequências dos genes aqui avaliados foi inferior (30%) aos relatados por Chen et al. (2010) para o gene *PvP5CS2* (74,2%), e por McClean, Lee e Miklas. (2004) em um intron do gene *DFR1* (71%), ambos de feijoeiro. Porém, em uma comparação entre sequências de 12 genótipos de *Arabidopsis thaliana*, Schmid, Sorensen e Stracke (2003) determinaram que os indel representaram 14% dos polimorfismos detectados. Em cevada cultivada, a proporção foi de 8% com base numa análise de sequências de cinco cultivares (KANAZIN et al., 2002), enquanto em soja, cerca de 15% de variantes da sequência eram indels (CHOI et al. 2007). Deste modo, apesar da frequência de indels aqui detectada ter sido inferior às obtidas anteriormente para o feijoeiro, é substancialmente mais elevada em sequências gênicas de feijão comum do que nas outras culturas citadas anteriormente. De acordo com Chen et al. (2010), essa diferença pode estar relacionada com o fato de a domesticação do feijoeiro ter ocorrido separadamente em dois grupos gênicos distintos.

4.5 Predição das alterações nas sequências de aminoácidos

Como mencionado, foram detectados SNPs bi-alélicos nas regiões codificadoras de proteína de ambos os genes estudados. Inicialmente efetuou-se a

predição da sequência de aminoácidos das proteínas faseolina e ferro redutase, a partir das respectivas sequências de nucleotídeos, através do *software* GeneRuner (SPRUYT; BUQUICCHIO, 1994).

As traduções *in silico* geraram sequências que corresponderam a 364 aminoácidos para a faseolina e 702 aminoácidos para a ferro redutase. Posteriormente, o alinhamento global múltiplo das sequências preditas evidenciou os polimorfismos de aminoácidos para cada proteína (Tabelas 6 e 7).

Tabela 6 - Posições dos resíduos polimórficos na sequência do gene α -Phs

Acesso	Posição do aminoácido																
	3	10	11	18	25	26	31	34	38	44	50	52	56	112	155	158	340
Brasil 2 (M)	A	L	R	S	A	G	L	P	E	Q	N	L	D	E	I	K	G
Porrillo (M)	I	.	.	F	.	.	.	.	.	.	.	.	.	.	.	.	.
Rio Tibagi (M)	I	.	.	F	.	.	.	.	.	.	.	.	.	.	.	.	.
Mulatinho (M)	I	.	L	.	.	.	.	.	.	.	.	P	.	.	.	.	.
Carioca (M)	I	.	L	.	.	.	.	.	.	.	.	.	.	K	.	.	.
Puebla 152 (M)	I	.	T	.	.	.	W	L	.	.	T	.	E	.	.	.	.
Great Northern (M)	I	S	L	.	.	.	.	.	.	.	.	P	.	.	.	.	.
Flor de Mayo (M)	I	.	P	.	.	.	.	.	.	.	.	P	.	K	.	.	.
Garbancillo (M)	I	.	P	.	.	.	.	.	.	.	.	.	.	K	.	.	.
Apetito Blanco (M)	I	.	L	.	.	.	.	.	Q	.	.	.	.	.	.	.	.
Jamapa (M)	I	.	P	.	.	.	.	.	.	.	.	.	.	K	.	.	.
Barbunia (M)	I	.	L	.	.	.	.	.	.	.	.	.	.	K	.	.	.
IAC-UMA (M)	I	.	L	.	E	R	.	.	.	.	.	P	E	.	.	.	.
Baetão (M)	I	.	L	.	.	.	.	.	.	.	.	P	.	K	.	.	.
Sanilac (M)	I	.	P	.	.	.	.	.	.	.	.	.	.	K	.	.	.
Tu (M)	I	T	.	.	.	.	W	L	.	.	T	.	E	.	.	.	.
Bat 93 (M)	I	.	T	.	.	.	.	.	.	.	.	P	.	K	.	.	.
Red Mexican (M)	I	T	.	.	.	.	W	L	.	.	T	.	.	.	.	.	R
Pompadour (A)	I	.	.	.	.	.	.	.	.	.	.	.	E	.	.	.	R
Antioquia (A)	I	.	.	.	.	.	.	.	.	.	.	P	E	K	.	E	.
Jalo EEP558 (A)	I	.	.	.	.	.	.	.	.	H	.	H	.	.	.	.	.
Jabola (A)	I	.	.	.	.	.	.	.	.	.	.	P	E	K	.	E	.
Cal 143 (A)	I	.	.	.	.	.	.	.	.	.	.	H	.	.	.	.	.
Red Kidney (A)	I	.	.	.	.	.	.	.	.	.	.	P	E	K	.	E	.
SAF (Cr)	I	.	L	F	.	.	.	.	.	.	.	.	.	.	.	.	.
SAB (Cr)	I	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Jalo precoce (Cv)	I	.	.	F	.	.	.	.	.	.	.	.	.	.	.	.	.
BRS Pontal (Cv)	I	.	L	.	.	.	.	.	.	.	.	P	.	K	.	.	.
Aporé (Cv)	I	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
BRS Pitanga (Cv)	I	.	L	.	.	.	.	.	.	.	.	.	.	K	.	.	.
BRS Vereda (Cv)	I	.	L	F	.	.	.	.	.	.	T	.	.	.	.	.	.
<i>P. lunatus</i>	I	.	L	.	.	.	.	.	.	.	.	P	.	.	M	.	.

M: Mesoamericano; A: Andino; Cr: Crioulo; Cv: Cultivar Embrapa

Importante notar que a variação de aminoácidos na faseolina se prestou a distinguir praticamente todos os acessos estudados. Por exemplo, apenas o acesso ‘Brasil 2’ apresentou uma alanina (A) na posição 3 da faseolina, se diferenciando dos demais. Da mesma forma, o acesso ‘Great Northern’ possui uma serina (S) na posição 10 e uma prolina (P) na posição 52; e o acesso ‘IAC-UNA’ se diferencia dos demais por apresentar, e um ácido glutâmico (E) e uma arginina (R) nas posições 25 e 26, respectivamente.

Em contrapartida, a enzima ferro redutase exibiu um padrão de aminoácidos mais homogêneo, principalmente para os acessos andinos, os quais apresentaram uma lisina (K) na posição 444 e uma glutamina (Q) na posição 569.

Tabela 7 - Posições dos resíduos polimórficos na sequência do gene *FRO1*

Acesso	Posição do aminoácido														
	89	327	388	444	457	464	474	498	535	551	569	561	576	686	
Brasil 2 (M)	T	L	S	R	S	L	K	K	H	S	R	A	Y	S	
Porrillo (M)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Rio Tibagi (M)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Mulatinho (M)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Carioca (M)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Puebla 152 (M)	S	P	R	K	L	P	.	.	R	.	Q	.	.	.	
Great Northern (M)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Flor de Mayo (M)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Garbancillo (M)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Apetito Blanco (M)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Jamapa (M)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Barbunia (M)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
IAC-UMA (M)	S	P	R	K	L	P	.	.	R	.	Q	.	.	.	
Baetão (M)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Sanilac (M)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Tu (M)	S	P	R	K	L	P	.	.	R	.	Q	.	.	.	
Bat 93 (M)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Red Mexican (M)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Pompadour (A)	.	.	.	K	.	.	.	.	.	.	Q	.	.	.	
Antioquia (A)	.	.	.	K	.	.	.	.	.	.	Q	.	.	F	
Jalo EEP558 (A)	.	.	.	K	.	.	.	.	.	.	Q	.	.	F	
Jabola (A)	.	.	.	K	.	.	.	.	.	.	Q	.	.	F	
Cal 143 (A)	.	.	.	K	.	.	.	.	.	.	Q	.	.	F	
Red Kidney (A)	.	.	.	K	.	P	Q	E	.	T	Q	T	C	F	
SAF (Cr)	.	.	.	.	.	.	.	.	.	.	Q	.	.	.	
SAB (Cr)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Jalo precoce (Cv)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
BRS Pontal (Cv)	S	P	R	K	L	.	.	.	.	.	.	.	.	.	
Aporé (Cv)	.	.	.	.	L	P	.	.	R	.	Q	.	.	.	
BRS Pitanga (Cv)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
BRS Vereda (Cv)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	

M: Mesoamericano; A: Andino; Cr; Crioulo; Cv: Cultivar Embrapa

Em seguida, as sequências consenso de faseolina e ferro redutase foram utilizadas para o alinhamento contra a base de dados de proteína do NCBI por BlastX. Os resultados dos alinhamentos indicaram o sucesso da predição da sequência de aminoácidos e, para a ferro redutase, foram identificadas as posições dos motivos de redução de ferro, NAD e FAD (Figura 15).

Figura 15 - Identificação dos domínios protéicos em uma sequência de nucleotídeos do gene *FRO1* de *P. vulgaris*, por BlastX. FeR: domínio de redução de Fe, conservado na superfamília das ferro redutases; FAD: domínio de ligação de FAD; N: domínio de ligação FAD

Geptset al. (1986) e Pereira (1990) demonstraram que o processo de domesticação ocasionou uma enorme redução dos tipos de faseolina nos materiais cultivados, os quais representam apenas uma pequena parte da variabilidade dos feijoeiros silvestres. Entretanto, ao avaliar diretamente a sequência do gene que codifica esta proteína, percebe-se a existência de vários polimorfismos que, apesar da natureza degenerada do código genético, acabam se refletindo em alterações na sequência de aminoácidos. A utilização desta abordagem na busca de polimorfismo entre acessos com pouca variabilidade pode ser uma alternativa para o entendimento das relações genéticas dentro de grupos aparentados.

As análises de polimorfismo mostraram que os SNPs identificados causaram a modificação de 17 aminoácidos na faseolina, e 14 na ferro redutase. A menor ocorrência de alterações nos aminoácidos da ferro redutase, em comparação com a faseolina, suporta a idéia que em enzimas, principalmente aquelas que desempenham atividades celulares vitais, as mutações tem maior potencial deletério.

Tais modificações incluem, na maioria das vezes, alterações de aminoácidos por outros com propriedades similares. Todavia, também foram observadas substituições por aminoácidos com propriedades distintas que, principalmente em enzimas, como a ferro redutase, podem implicar em alterações na conformação estrutural da proteína, afetando ou não sua atividade. Destas modificações, pelo menos uma encontra-se no motivo FAD na posição 388 (S → R), e outra no motivo NAD na posição 457 (S → L). Entretanto, faz-se necessária uma abordagem mais

detalhada visando definir a magnitude da influência de cada aminoácido alterado na estrutura tridimensional da proteína e, conseqüentemente, se esta alteração pode afetar a atividade da enzima.

Blair et al. (2010b) relataram que a atividade da enzima ferro redutase está diretamente relacionada com o acúmulo de ferro em sementes de feijão, ou seja, genótipos com uma maior atividade da enzima em suas raízes eram capazes de absorver mais ferro e este, portanto, pode ser direcionado para as sementes através das vias reguladoras de acúmulo do mineral. Entretanto, não foram encontradas diferenças significativas na atividade enzimática da ferro redutase entre os dois grupos gênicos apesar de uma série de estudos ter demonstrado que feijões Andinos tendem a ter maior concentração de ferro nas sementes do que os Mesoamericanos (ISLAM, et al., 2002; ISLAM, et al., 2004; BLAIR, et al., 2010b).

Os domínios de ligação NAD e FAD estão intimamente relacionados com a capacidade de a enzima reduzir o ferro de baixa solubilidade (Fe^{3+}) presente no solo. Uma vez que foram detectadas mutações que resultaram em aminoácidos com propriedades distintas nessas regiões, infere-se que esses polimorfismos possam estar relacionados com a capacidade de conversão de $\text{NADH}^+ + \text{H}^+$ e $\text{FADH}^+ + \text{H}^+$ em NAD^+ e FAD^+ , respectivamente, possivelmente afetando a redução de ferro pela enzima.

Uma vez que as concentrações de ferro em sementes dos acessos utilizados neste estudo não foram estabelecidas, ainda não se podem relacionar, diretamente, tais mutações com a característica fenotípica.

4.6 Verificação de saturação e escolha do modelo evolutivo

A plotagem do número de transições e transversões *versus* o modelo de substituição de bases K80, exposta na figura 16, exibe um padrão de distribuição semelhante a uma reta tanto na avaliação para cada gene separadamente, quanto utilizando as sequências de ambos, combinadas. Esse padrão sugere que não há saturação de transições em relação a transversões e, portanto, todos os dados podem ser utilizados nas análises.

Figura 16 - Gráfico de transição (S) e transversão (V) versus distância de Kimura dos genes α -Phs (A), *FRO1* (B) e do conjunto de dados combinado das sequências dos dois genes (C)

Dentre todos os modelos evolutivos testados pelo *software* MEGA 5.1 (TAMURA et al., 2011), o modelo de substituição de bases Tamura 3 Parâmetros (T92) foi indicado como o mais adequado. Este modelo considera a diferença nas taxas de transições e transversões, bem como a composição da sequência, atribuindo pesos diferenciados para cada parâmetro (TAMURA, 1992). Em sequências codificadoras de proteínas, as bases localizadas na terceira posição do códon estão mais sujeitas à variação do que as correspondentes à primeira e à segunda posição, porque mutações na terceira posição alteram menos frequentemente o aminoácido correspondente devido à natureza degenerada do código genético (GRAUR; LI, 2000; NEI; KUMAR, 2000).

4.7 Diversidade genética entre os grupos gênicos de feijão

A partir dos polimorfismos obtidos neste estudo, foi possível estimar as relações genéticas entre os 31 acessos de feijão comum. As análises de agrupamento *Neighbor-joining* realizadas a partir das sequências dos dois genes,

tanto separadamente quanto em conjunto (Figuras 17, 18 e 19), detectaram em todos os casos a formação de dois *clusters* bem definidos: um que inclui os acessos previamente classificados como Mesoamericanos juntamente com os acessos crioulos e as cultivares produzidas pela Embrapa (*cluster* I) e outro que agrupa todos os acessos classificados como pertencentes ao grupo gênico Andino (*cluster* II). Este padrão de agrupamento concorda em geral com o obtido por Singh (2001), a partir do tamanho das sementes; com os obtidos por Hanai et al. (2007) e Benchimol et al. (2007), a partir de marcadores SSR; também, é consistente com o atual regime hierárquico de divisão de grupos gênicos de *P. vulgaris*.

Como mostram os dendrogramas obtidos, os *clusters* gerados a partir dos polimorfismos do gene α -*Phs* permitiram uma melhor resolução das relações entre acessos dentro de cada *cluster*, em comparação com os gerados a partir do gene *FRO1*. Já o dendrograma gerado a partir do conjunto combinado de dados, permitiu o aumento dos valores de *bootstrap* de alguns *clusters*, reforçando a confiabilidade dos subgrupos como, por exemplo, aquele que inclui os acessos ‘Puebla 152’ e ‘Tu’. Entretanto, em geral, o uso combinado de sequências de α -*Phs* e *FRO1* gerou agrupamentos distintos dos obtidos através das análises individuais.

Figura 17 - Relações entre 32 acessos de feijão a partir das sequências de α -Phs. Os dendrogramas foram construídos pelo método *Neighbor-joining*, usando o modelo de substituição de nucleotídeos Tamura 3 Parâmetros e 500 repetições de *Bootstrap*. As siglas localizadas entre parênteses remetem à classificação dos acessos. M = Mesoamericano, A = Andino, Cr = Crioulo e Cv = Cultivar Embrapa

Figura 18 - Relações entre 32 acessos de feijão a partir das sequências de *FRO1*. Os dendrogramas foram construídos pelo método *Neighbor-joining* usando o modelo de substituição de nucleotídeos Tamura 3 Parâmetros e 500 repetições de *Bootstrap*. As siglas localizadas entre parênteses remetem à classificação dos acessos. M = Mesoamericano, A = Andino, Cr = Crioulo e Cv = Cultivar Embrapa

Figura 19 - Relações entre 32 acessos de feijão a partir das sequências de α -Pbs e FRO1. Os dendrogramas foram construídos pelo método *Neighbor-joining* usando o modelo de substituição de nucleotídeos Tamura 3 Parâmetros e 500 repetições de *Bootstrap*. As siglas localizadas entre parênteses remetem à classificação dos acessos. M = Mesoamericano, A = Andino, Cr = Crioulo e Cv = Cultivar Embrapa

Hanai et al. (2007) verificaram que, ao utilizar marcadores SSR provenientes de sequências expressas e genômicas para avaliar as relações entre acessos cultivados de feijão, os dendrogramas obtidos a partir de cada marcador se mostraram levemente distintos. Já Benchimol et al. (2007), ao utilizarem outros marcadores SSR obtidos a partir de uma biblioteca genômica enriquecida com microsatélites e parte dos acessos utilizados por Hanai et al. (2007), verificaram um padrão de agrupamento distinto.

Apesar de ter sido utilizada uma amostra de acessos similar àquela avaliada por Hanai et al. (2007) e conter acessos comuns aos avaliados por Benchimol et al. (2007), o padrão de agrupamento obtido a partir de sequências de *α -Phs* e *FRO1* diferem dos anteriores.

No tocante ao agrupamento dos acessos crioulos e cultivares produzidas pela Embrapa, juntamente com os demais acessos classificados como sendo do *pool* Mesoamericano, sabe-se que no Brasil, feijoeiros desse grupo são preferidos pelos consumidores, e esta preferência se dá pelos tipos de grãos carioca e preto. Cultivares com características andinas, devido ao tamanho da semente, possuem pouca aceitabilidade no comércio. Estas exigências, muitas vezes, reduzem o número de genitores para cruzamentos e conseqüentemente reduzem a base genética das cultivares produzidas (CARBONELL et al., 1999). Voysest, Valencia e Amezquita (1994) analisaram 203 cultivares desenvolvidas na América Latina, e constataram que 64% correspondem à raça Mesomericana, pertencente ao *pool* Mesoamericano e 13% correspondem à raça Nueva Granada, pertencente ao *pool* Andino. Relataram também que novas fontes de germoplasma no Brasil aumentaram desde 1984, mas até 1992, os níveis de diversidade genética entre as cultivares estavam baixos, isto porque, cultivares como Carioca Comum eram sempre usados como genitores na obtenção de genótipos com coloração carioca do tegumento e resistentes à antracnose.

Uma análise de variância molecular foi realizada com o intuito de visualizar quanto os grupos de acessos variam para os genes avaliados. Desta forma, esta análise reforça os resultados obtidos pelos agrupamentos realizados anteriormente, nos quais o gene *α -Phs* permitiu uma melhor resolução das relações entre acessos dentro dos *clusters*. Além disso, ao avaliar a porcentagem de variação obtida para o conjunto de dados, também fica evidente que a utilização dessa abordagem reduz o

potencial informativo para a diversidade dentro de grupos, quando esta é comparada com os valores obtidos a partir da utilização do gene α -Phs isoladamente.

Tabela 8 - Análise Molecular da Variância entre e dentro dos *pools* gênicos Mesoamericano e Andino de *P. vulgaris*, a partir das sequências de α -Phs e *FRO1*

Fonte de variação	(% de variação)		
	α -Phs	<i>FRO1</i>	α -Phs + <i>FRO1</i>
Entre grupos	66,99	75,43	71,15
Dentro de grupos	33,01	24,57	28,85

4.8 Diversidade nucleotídica e testes de neutralidade

Duas estimativas de diversidade nucleotídica foram calculadas (Tabela 9). A diversidade nucleotídica (π), ou seja, a diferença média de bases entre duas sequências aleatórias da amostra foi de 0.006137 para o gene *FRO1* e de 0.003651 para o gene α -Phs. Já a estimativa média de Θ (obtido a partir de S), que é baseado no número observado de locais polimórficos de uma amostra, foi de 0.005456 para o gene *FRO1* e de 0.003794 para o gene α -Phs.

Tabela 9 - Diversidade nucleotídica e teste de neutralidade D de Tajima para os *pools* gênicos Mesoamericano e Andino *P. vulgaris*, a partir de sequências de α -Phs e *FRO1*

Gene	Grupo gênico	m	S	P_s	Θ	π	D
α -Phs	A	6	21	0.004664	0.002042	0.002221	0.546488
	M	25	41	0.009135	0.002419	0.002109	- 0.489758
	Cod	31	22	0.023140	0.005085	0.003749	- 0.918548
	n-Cod	31	12	0.023715	0.005936	0.004565	- 0.748776
	Total	31	68	0,015158	0,003794	0,003651	- 0,141848
<i>FRO1</i>	A	6	31	0.013237	0.005797	0.004412	- 1.514677
	M	25	48	0.020504	0.005430	0.004431	- 0.706950
	Cod	31	22	0,010456	0,002617	0,001873	- 0,995013
	n-Cod	31	29	0.034360	0.008601	0.012297	1.539.608
	Total	31	51	0,021795	0,005456	0,006137	0,463702

m = Número de sequências avaliadas, S= Número de sítio segregantes, $P_s = S/m$, Θ = Diversidade nucleotídica média, π = Diversidade nucleotídica, D = Teste Tajima de neutralidade, M: *pool* Mesoamericano, A: *pool* Andino, Cod: região codante, n-Cod: região não codante

De acordo com ambos os estimadores, a diversidade nucleotídica em *FRO1* é maior do que em α -Phs. Quando se comparam os valores de π e Θ entre regiões codantes e não codantes, percebe-se, para ambos os genes, uma maior diversidade

nucleotídica em regiões não codantes. Entretanto, vale ressaltar que os valores de π e Θ entre essas regiões são mais contrastantes para o gene *FRO1* do que para o gene α -*Phs*. Ao se comparar essa diversidade entre os *pools* gênicos, percebe-se que o valor de Θ para o gene α -*Phs*, indica uma maior variabilidade dentro do *pool* Mesoamericano, ao passo que, para o gene *FRO1*, o *pool* Andino exibe mais variabilidade. No primeiro caso, tal observação concorda com os resultados obtidos por Chen et al. (2010) para o gene *PvP5CS2*. Utilizando marcadores de AFLP, BEEBE et al. (2001) também observaram uma variação reduzida entre genótipos andinos. Mcclean, Lee e Miklas (2004) detectaram que a variação de um intron entre os genótipos andinos também foi reduzida em comparação com a variação entre os genótipos mesoamericanos. No entanto, BLAIR et al. (2006), utilizando marcadores SSR, observaram níveis similares de diversidade entre os *pools* gênicos. Deste modo, estes resultados sugerem que o *pool* andino pode ter uma base genética mais estreita do que o *pool* mesoamericano, embora esta conclusão depende da amostra de acessos utilizada em cada estudo.

A diversidade média de nucleotídeos obtida para os genes *FRO1* ($\Theta = 0.00545$) e α -*Phs* ($\Theta = 0.00365$) foram superiores aos valores obtidos para o gene *PvP5CS2* ($\Theta = 0.00287$) em feijão (CHEN et al., 2010) e para *Arabidopsis* ($\Theta = 0.0033$) (OLSEN et al., 2002), porém inferiores aos valores obtidos para o milho ($\Theta = 0.0181$) (EYRE-WALKER et al., 1998) e a cevada ($\Theta = 0,0081$) (MORRELL et al., 2005).

O teste D de Tajima foi realizado para examinar se a variação nas sequências gênicas apresentavam desvios o modelo de substituição de bases Tamura 3 Parâmetros (TAMURA, 1992). Valores positivos indicam um excesso de polimorfismo de frequência intermédia que pode ser o resultado de seleção estabilizadora ou subdivisão da população (HARTL; CLARK, 1997). Para considerar essas possibilidades, os dados foram subdivididos de acordo com o grupo gênico (Mesoamericano ou Andino) e com as regiões do gene (codante e não codante), e as estatísticas foram recalculadas (Tabela 9).

Os valores de D obtidos para as regiões não codantes do gene *FRO1* é positivo, o que acaba sendo refletido nessa estimativa para o gene completo. Entretanto, para o grupo Andino, Mesoamericano e para a região codante, os valores foram negativos. Tais resultados indicam que, apesar da região não codante do

gene apresentar muita variação, a região codante sofre uma pressão de seleção negativa intensa, ou seja, mutações nessas regiões tendem a ser deletérias.

Para o gene α -*Phs* percebe-se que, aparentemente, os processos de seleção que ocorreram no grupo Andino diferem daqueles que ocorreram no Mesoamericano. Porém, de um modo geral, os valores de D obtidos indicam que a faseolina sofreu uma diminuição da variabilidade, provavelmente durante o processo de domesticação. Tal resultado concorda com Gepts et al. (1986) e Pereira (1990) que demonstraram que o processo de domesticação ocasionou uma enorme redução dos tipos de faseolina nos materiais cultivados. Porém, assim como foi mencionado anteriormente, o gene α -*Phs* exibe um elevado grau de polimorfismo entre os acessos aqui estudados.

5 CONCLUSÕES

- As sequências de *α -Phs* e *FRO1* exibem uma maior frequência de sítios polimórficos em regiões não codantes do que em regiões codantes, e tais polimorfismos se prestam para agrupar os acessos de *Phaseolus vulgaris* em dois *pools* gênicos, o Mesoamericano e o Andino;
- Os polimorfismos do gene *α -Phs* permitem uma melhor resolução das relações entre os acessos dentro de cada *pool* gênico, em comparação com os do gene *FRO1*;
- As substituições nucleotídicas observadas nos genes *α -Phs* e *FRO1* se refletem em alterações de aminoácidos; a faseolina exibe uma diversidade mais elevada entre os acessos investigados, sendo que alguns desses polimorfismos se prestam para tipificar alguns acessos.

REFERÊNCIAS

- ALTSCHUL, S.F.; MADDEN, T.L.; SCHÄFFER, A.A.; ZHANG, J.; ZHANG, Z.; MILLER, W.; LIPMAN, D.J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. **Nucleic Acids Research**, Oxford, v. 25, p. 3389-3402, 1997.
- BALASUBRAMANIAN, S.; HARRISON, P.; HEGYI, H.; BERTONE, P.; LUSCOMBE, N.; ECHOLS, N.; MCGARVEY, P.; ZHANG, Z.; GERSTEIN, M. SNPs on human chromosomes 21 and 22 - Analysis in terms of protein features and pseudogenes. **Pharmacogenomics**, London, v. 3, p. 393-402, 2002.
- BARBAZUK, W.B.; EMRICH, S.J.; CHEN, H.D.; LI, L.; SCHNABLE, P.S. SNP discovery via 454 transcriptome sequencing. **Plant Journal**, Malden, v. 51, p. 910-918, 2007.
- BECERRA-VELÁSQUEZ, V.L.; GEPTS, P. RFLP diversity in common bean (*Phaseolus vulgaris* L.). **Genome**, Ottawa, v. 37, p. 256-263, 1994.
- BEEBE, S.E.; SKROACK, 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 Science**, Madison, v. 40, p. 264-273, 2000.
- BEEBE, S.E.; RENGIFO, J.; GAITAN, E.; DUQUE, M.C.; TOHME, J. Diversity and origin of Andean landraces of common bean. **Crop Science**, Madison, v. 41, p. 854-862, 2001.
- BENCHIMOL, L.L.; CAMPOS, T.; CARBONELL, S.A.M.; CHIORATO, A.F.; COLOMBO, C.A.; FORMIGUIERI, E.F.; GOUVEIA, L.R.L.; SOUZA, A.P. Structure of genetic diversity among common bean (*Phaseolus vulgaris* L.) varieties of Mesoamerican and Andean origins using new developed microsatellite markers. **Genetic Resources and Crop Evolution**, Dordrecht, v. 54, p. 1747-1762, 2007.
- BERGLUND-BRÜCHER, O.; BRÜCHER, H. The south American wild bean (*Phaseolus aborigineus* Burk.) as ancestor of the common bean. **Economic Botany**, Bronx, v. 30, n. 3, p. 257-272, 1976.
- BHATTRAMAKKI, D.; RAFALSKI, A. Discovery and application of single nucleotide polymorphism markers in plants. In: HENRY, R.J. (Ed). **Plant genotyping: the DNA fingerprinting of plants**. Wallingford: CAB International, 2001. p. 179-191.
- BLAIR, M.W.; GIRALDO, M.C.; BUENDIA, H.F.; TOVAR, E.; DUQUE, M.C. Microsatellite marker diversity in common bean (*Phaseolus vulgaris* L.). **Theoretical and Applied Genetics**, Oxford, v. 113, p. 100-109, 2006.
- BLAIR, M.W.; DÍAZ, L.M.; BUENDIA, H.F.; DUQUE, M.C. Genetic diversity, seed size associations and population structure of a core collection of common beans (*Phaseolus vulgaris* L.). **Theoretical and Applied Genetics**, Oxford, v. 119, p. 955-973, 2009.
- BLAIR, M.W.; GONZALES, L.F.; KIMANI, P.; BUTARE, L. Inter-genepool introgression, genetic diversity and nutritional quality of common bean (*Phaseolus vulgaris* L.) landraces from central Africa. **Theoretical and Applied Genetics**, Oxford, v. 121, p. 237-248, 2010a.
- BLAIR, M.W.; KNEWTSON, S.J.B.; ASTUDILLO, C.; LI, C-M.; FERNANDEZ, A.C. Variation and inheritance of iron reductase activity in the roots of common bean (*Phaseolus vulgaris* L.) and association with seed iron accumulation QTL. **BMC Plant Biology**, London, v. 10, p. 215-227, 2010b.
- BOTSTEIN, D.; WHITE, R.L.; SKOLNICK, M.; DAVIS, R.W. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. **The American Journal of Human Genetics**, Houston, v. 32, p. 314-331, 1980.
- BRIAT, J.F. Cellular and whole organism aspects of iron transport and storage in plants. **Topics in Current Genetics**, Göteborg, v. 14, p. 193-213, 2006.

BROWN, J.; MA, Y.; BLISS, F.; HALL, T. Genetic variation in the subunits of globulin-1 storage protein of French bean. **Theoretical and Applied Genetics**, Oxford, v. 59, p. 83-88, 1981.

BRUGMANS, B.; WOUTERS, D.; VAN OS, H.; HUTTEN, R.; LINDEN, G.V.D.; VISSER, G.F.R.; VAN ECK H.J., VAN DER VOSSEN, E.A.G. Genetic mapping and transcription analyses of resistance gene loci in potato using NBS profiling. **Theoretical and Applied Genetics**, Oxford, v. 117, p. 1379-1388, 2008.

BUNDOCK, P.C.; HENRY, R.J. Single nucleotide polymorphism, haplotype diversity and recombination in the *Isa* gene of barley. **Theoretical and Applied Genetics**, Oxford, v. 109, n. 3, p. 543-551, 2004.

BURLE, M.L.; FONSECA, J.R.; KAMI, J.A., GEPTS, P. Microsatellite diversity and genetic structure among common bean (*Phaseolus vulgaris* L.) landraces in Brazil, a secondary center of diversity. **Theoretical and Applied Genetics**, Oxford, v. 121, p. 801-813, 2010.

CALENGE, C.; SPICHIGER, R.; CHESSEL, D.; CHATELAIN, C. Discriminant analysis of the spatial distribution of plant species occurrences: I. **Theoretical aspects**, *Candollea*, v. 60, n. 2, p. 563-575, 2005.

CARBONELL, S.A.M.; ITO, M.F.; POMPEU, A.S.; FRANCISCO, F.G.; RAVAGNANI, S.; ALMEIDA, A.L.L. Raças fisiológicas de *Colletotrichum lindemuthianum* e reação de linhagens e cultivares de feijoeiro no Estado de São Paulo. **Fitopatologia Brasileira**, Brasília, v. 24, p. 60-65, 1999.

CHACÓN, M.I.; PICKERSGILL, B.; DEBOUCK, D.G. Domestication patterns in common bean (*Phaseolus vulgaris* L.) and the origin of the Mesoamerican and Andean cultivated races. **Theoretical and Applied Genetics**, Oxford, v. 110, p. 432-444, 2005.

CHARCOSSET, A.; GALLAIS, A. Application of markers in selection. In: DE VIENNE, D. (Ed). **Molecular markers in plant genetics and biotechnology**. New Hampshire: Science Publishers, 2002. p. 153-176.

CHEN, J.; ZHANG, X.; JING, R.; BLAIR, M. W.; MAO, X.; SHUMIN, W. Cloning and genetic diversity analysis of a new P5CS gene from common bean (*Phaseolus vulgaris* L.). **Theoretical and Applied Genetics**, Oxford, v. 120, p. 1393-1404, 2010.

CHIORATO, A.F.; CARBONELL, S.A.M.; COLOMBO, C.A.; DIAS, L.A.S.; ITO, M.F. Genetic diversity of common bean accessions in the germplasm bank of the Instituto Agronômico - IAC. **Crop Breeding and Applied Biotechnology**, Viçosa, v. 5, p. 1-9, 2005.

CHOI, I.; HYTEN, D.L.; MATUKUMALLI, L.K.; SONG, Q.; CHAKY, J.M.; QUIGLEY, C.V.; CHASE, K.; LARK, K.G.; REITER, R.S.; YOON, M.S.; HWANG, E.Y.; YI, S.I.; YOUNG, N.D.; SHOEMAKER, R.C.; TASSELL, C.P.; SPECHT, J.E.; CREGAN, P.B. A soybean transcript map: gene distribution, haplotype and single-nucleotide polymorphism analysis. **Genetics**, Pittsburgh, v. 176, p. 685-696, 2007.

COTTAGE, A.; YANG, A.; MAUNDERS, H.; LACY, C.; RAMSAY, N. Identification of DNA sequences flanking T-DNA insertions by PCR-walking. **Plant Molecular Biology Reporter**, Athens, v. 19, p. 312-327, 2001.

CRESTE, S. ; TULMANN NETO, A. ; FIGUEIRA, A. Detection of single sequence repeat polymorphisms in denaturing polyacrylamide sequencing gels by silver staining. **Plant Molecular Biology Reporter**, Athens, v. 19, p. 299-306, 2001.

CRUZ, C.D.; CARNEIRO, P.C.C. **Modelos biométricos aplicados ao melhoramento genético**. Viçosa: UFV, 2003. 885 p.

DANTAS, B.F.; RIBEIRO, L.S.; ARAGÃO, C. A. Germination, initial and cotyledon protein content of bean cultivars under salinity stress. **Revista Brasileira de Sementes**, Londrina, v. 26, n. 2, p. 106-110, 2007.

DEBOUCK, D.G.; TORO, O.; PAREDES, O.M.; JOHNSON, W.C.; GEPTS, P. Genetic diversity and ecological distribution of *Phaseolus vulgaris* in northwestern South America. **Economic Botany**, Bronx, v. 47, p. 408–423, 1993.

DEBOUCK, D.G. Primary diversification of *Phaseolus* in the Americas: three centers? **Plant Genetic Resources Newsletter**, Rome, v. 67, p. 2-8, 1986.

DÍAZ, L.M.; BLAIR, M.W. Race structure within the Mesoamerican gene pool of common bean (*Phaseolus vulgaris* L.) as determined by microsatellite markers. **Theoretical and Applied Genetics**, Oxford, v. 114, p. 143-154, 2006.

DOYLE, J.J.; DOYLE, J.L. A rapid DNA isolation procedure to small amounts of fresh leaf tissue. **Phytochemical Bulletin**, Cornell, v. 19, p. 11-15, 1987.

EVANS, A.M. Beans. In: SIMMONDS, N. W. **Evolution of crop plants**. Hong Kong: Longman Scientific and Technical, 1986. p. 168-172.

EWING, B.; HILLIER, L.; WENDL, M.; GREEN, P. Basecalling of automated sequencer traces using phred.1. Accuracy assessment. **Genome Research**, Woodbury, v. 8, p. 175-185, 1998.

EXCOFFIER, L.; LISCHER, E.L.H. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. **Molecular Ecology Resources**, Malden, v. 10, p. 564-567, 2010.

EYRE-WALKER, A.; GAUT, R.L.; HILTON, H.; FELDMAN, D.L.; GAUT, B.S. Investigating the bottleneck leading to the domestication of maize. **Proceedings of the National Academy of Sciences**, Washington, v. 95, p. 4441-4446, 1998.

FREYRE, R.; RÍOS, R.; GUZMÁN, L.; DEBOUCK, D.; GEPTS, P. Ecogeographic distribution of *Phaseolus* ssp. (Fabaceae) in Bolivia. **Economic Botany**, Bronx, v. 50, n. 2, p.195-215, 1996.

FROSSARD, E.; CONDRON, L.M.; OBERSON, A.; SINAJ, S.; FARDEAU, J.C. Processes governing phosphorus availability in temperate soils. **Journal of Environmental Quality**, Madison, v. 29, p. 12–53, 2000.

GAITÁN-SOLÍS, S. E.; DUQUE, M. C.; EDWARDS, K. J.; TOHME, J. Microsatellite repeats in common bean (*Phaseolus vulgaris*): Isolation, characterization, and cross-species amplification in *Phaseolus* ssp. **Crop Science**, Madison, v. 42, p. 2128–2136, 2002.

GAITÁN-SOLÍS, E.; CHOI, I.-Y.; QUIGLEY, C.; CREGAN, P.; TOHME, J. Single nucleotide polymorphisms in common bean: their discovery and genotyping using a multiplex detection system. **The Plant Genome**, Madison, v. 1, p. 125-134, 2008.

GEIL, P.B.; ANDERSON, J.W. Nutrition and health implications of dry beans: a review. **Journal of the American College of Nutrition**, Clearwater, v. 13, n. 6, p. 549-558, 1994.

GENTRY, H.S. Origin of the common bean, *Phaseolus vulgaris*. **Economic Botany**, Bronx, v. 23, p. 55-69, 1969.

GEPTS, P.; BLISS, F.A. F1 hybrid weakness in the common bean: differential geographic origin suggests two gene pools in cultivated bean germplasm. **Journal of Heredity**, Washington, v. 76, p. 447-450, 1985.

GEPTS, P.; OSBORN, T.C.; RASHKA, K.; BLISS, F.A. Phaseolin-protein variability in wild forms and landraces of the common bean (*Phaseolus vulgaris*): evidence for multiple centers of domestication. **Economic Botany**, Bronx, v. 40, p. 451-468, 1986.

GEPTS, P.; DEBOUCK, D. Origin, domestication, and evolution of the common bean (*Phaseolus vulgaris* L.). In: VAN SCHOONHOVEN, A.; VOYSEST, O. **Common beans: research for crop improvement**. Wallingford: Common wealth Agricultural Bureaux International, 1991. p. 7–53.

GEPTS, P. Biochemical evidence bearing on the domestication of *Phaseolus* beans. **Economic Botany**, Bronx, v. 44, p. 28–38, 1990.

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

GEPTS P. Ten thousand years of crop evolution. In: CHRISPEELS, M.; SADAVA, D. (Ed.). **Plants, Genes, and Crop Biotechnology**. Sudbury: Bartlett and Jones, 2003. p. 328-359.

GILL-LANGARICA, H.R.; MURUAGA-MARTÍNEZ, J.S.; VARGAS-VÁZQUEZ, P.; ROSALES-SERNA, R.; MAYEK-PÉREZ, N. Genetic diversity analysis of common beans based on molecular markers. **Genetics and Molecular Biology**, Ribeirão Preto, v. 34, p. 595-605, 2011.

GRAUR D.; LI, W-H. **Fundamentals of Molecular Evolution**. 2nd. ed. Sunderland: Sinauer Associates, 2000. 481 p.

GRUSAK, M.A. Whole-root iron (III)-reductase activity through the life cycle of iron-grown. *Pisum sativum* L. (Fabaceae): relevance to the iron nutrition of developing seeds. **Planta**, Heidelberg, v. 197, p. 111-117, 1995.

GRUSAK, M.A. Strategies for improving the iron nutritional quality of seed crops: lessons learned from the study of unique iron-hyperaccumulating pea mutants. **Pisum Genetics**, Novosibirsk, v. 32, p. 1–5, 2000.

HANAHAN, D. Studies on the transformation of *Escherichia coli* with plasmids. **Encyclopedia of life sciences**, Hertfordshire, 2007. DOI: 10.1002/9780470015902.a0020119.

HANAI, L.R.; CAMPOS, T.; CAMARGO, L.R.A.; BENCHIMOL, L.L.; SOUZA, A.P.; MELOTTO, M.; CARBONELL, S.A.M.; CHIORATTO, A.R.; CONSOLI, L.; FORMIGHIERI, E.F.; SIQUEIRA, M.V.B.M.; TSAI, S.M.; VIEIRA, M.L.C. Development, characterization, and comparative analysis of polymorphism at common bean SSR loci isolated from genic and genomic sources. **Genome**, Ottawa, v. 50, p. 266-277, 2007.

HANAI, L.R.; SANTINI, L.; CAMARGO, L.E.A.; FUNGARO, M.H.P.; GEPTS, P.; TSAI, S.M.; VIEIRA, M.L.C. Extension of the core map of common bean with EST-SSR, RGA, AFLP and putative functional markers. **Molecular Breeding**, Dordrecht, v. 25, p. 24-45, 2010.

HARLAN, J.R. Agricultural origins: centers and no centers. Washington. **Science**, Washington, v. 174, p. 468-474, 1971.

HARTL, D.L.; CLARK, A.G. **Principles of population genetics**. 3rd. ed. Sunderland: Sinauer Associates, 1997. 481p.

HEISER Jr., C.B. **Sementes para a civilização: a história da alimentação humana**. São Paulo: Companhia Editora Nacional/EDUSP, 1977. 253p.

HELL, R.; STEPHAN, U.W. Iron uptake, trafficking and homeostasis in plants. **Planta**, Heidelberg, v. 216, p. 541–551, 2003.

ISLAM, F.M.A.; BASFORD, K.E.; JARA, C.; REDDEN, R.J.; BEEBE, S.E. Seed compositional and disease resistance differences among gene pools in cultivated common bean. **Genetic Resources and Crop Evolution**, Dordrecht, v. 49, p. 285-293, 2002.

ISLAM, F.M.A.; BEEBE, S.E.; MUÑOZ, M. THOME, J.; RADDEN R.J.; BASFORD, K.E. Using molecular markers to assess the effect of introgression on quantitative attributes of common bean in the Andean gene pool. **Theoretical and Applied Genetics**, Oxford, v. 121, p. 237-248, 2004.

JORDE, L.B. Linkage disequilibrium as a gene-mapping tool. **The American Journal of Human Genetics**, Houston, v. 56, p. 11-14, 1995.

JUKES, T.H.; CANTOR, C.R. Evolution of Protein Molecules. In MUNRO H. N. (Ed), **Mammalian Protein Metabolism**. New York: Academic Press, 1969. p. 21–132.

KAMI, J.A.; VELASQUEZ, V.B.; DEBOUCK, D.G.; GEPTS, P. Identification of presumed ancestral DNA sequences of phaseolin in *Phaseolus vulgaris*. **Proceedings of National Academy of Sciences**, Washington, v. 92, p. 1101-1104, 1995.

KANAZIN, V.; TALBERT, H.; SEE, D.; DECAMP, F.; NEVO E.; BLAKE, T. Discovery and assay of single-nucleotide polymorphisms in barley (*Hordeum vulgare*). **Plant Molecular Biology**, Dordrecht, v. 48, p. 529-537, 2002.

KARCHIN, R.; DIEKHANS, M.; KELLY, L.; THOMAS, D.J.; PIEPER, U.; ESWAR, N.; HAUSSLER, D.; SALI, A. LS-SNP: Large-scale annotation of coding non-synonymous SNPs based on multiple information sources. **Bioinformatics**, Oxford, v. 21, n. 12, p. 2814–2820, 2005.

KHAIRALLAH, M.M.; SEARS, B.B.; ADAMS, M.W. Mitochondrial restriction fragment length polymorphism in wild *Phaseolus vulgaris* L.: insights on the domestication of the common bean. **Theoretical and Applied Genetics**, Oxford, v. 84, p. 915-922, 1992.

KIMURA, M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. **Journal of Molecular Evolution**, Chicago, v. 16, p. 111–120, 1980.

KOENIG, R.L.; SINGH, S.P.; GEPTS, P. Novel phaseolin types in wild and cultivated common bean (*Phaseolus vulgaris*, Fabaceae). **Economic Botany**, New York, v. 44, p. 50-60, 1990.

KWAK, M.; GEPTS, P. Structure of genetic diversity in the two major gene pools of common bean (*Phaseolus vulgaris* L., Fabaceae). **Theoretical and Applied Genetics**, Oxford, v. 118, p. 979-992, 2009.

LUNARDI, L.U.; GUEMBAROVSKI, R.L.; HANAI, L.R.; CRISTIANO, V.; VIEIRA, M.L.C.; SARTORI, D.; FUNGARO, M.H.P. Identification of a splicing coactivator gene that affects the production of ochratoxin A in *Aspergillus carbonarius*. **Brazilian Archives of Biology and Technology**, Curitiba, v. 52, p. 134-141, 2009.

MABBERLEY, D. **The Plant-Book**. 2nd. ed. Cambridge: University Press, 1998. 1040 p.

MCCLEAN P.E.; LEE R.K.; MIKLAS, P.N. Sequence diversity analysis of dihydroflavonol 4-reductase intron 1 in common bean. **Genome**, Ottawa, v. 47, p. 266-280, 2004.

MELO, L.C.; SANTOS, J.B.; FERREIRA, D.F. Mapeamento de QTLs para reação ao oídio e mancha-angular do feijoeiro-comum em diferentes locais. **Pesquisa Agropecuária Brasileira**, Brasília, v. 37, p. 1115-1126, 2002.

MESSERSCHIMDT, A.; HUBER, R.; POULOS, T.; WIEGHART, K. **Handbook of Metalloproteins**. New York: John Wiley, 2001. 1248 p.

MONTOYA, C.A.; LETERME, P.; VICTORIA, N.F.; TORO, O.; SOUFFRANT, W.B.; BEEBE, S.; LALLÈS J.P. The susceptibility of phaseolin to in vitro proteolysis is highly variable across *Phaseolus vulgaris* bean varieties. **Journal of Agricultural and Food Chemistry**, California, v. 56, p. 2183-2191, 2008.

MORI, S. Iron acquisition by plants. **Current Opinion in Plant Biology**, Saint Louis, v. 2, p. 250-253, 1999.

MORRELL, P.L.; TOLENO, D.M.; LUNDY, K.E.; CLEGG, M.T. Low levels of linkage disequilibrium in wild barley (*Hordeum vulgare* ssp. *spontaneum*) despite high rates of self-fertilization. **Proceedings of the National Academy of Sciences**, Stanford, v. 102, p. 2442–2447, 2005.

NEI, M. **Molecular Evolutionary Genetics**. New York: Columbia University Press. 1987. 512 p.

NEI, M.; KUMAR, S. **Molecular Evolution and Phylogenetics**. New York: Oxford University Press, 2000. 333 p.

NODARI, R.O.; TSAI, S.M.; YU, K.F.; PARK, S.J. SHANG, B.L.; HAFFNER, M. Towards an integrated linkage map of common bean.3. Mapping genetic factors controlling host-bacteria interactions. **Genetics**, Pittsburgh, v. 134, p. 341-350, 1993.

OLSEN, K.M.; WOMACK, A.; GARRETT, A.R.; SUDDITH, J.I.; PURUGGANAN, M.D. Contrasting evolutionary forces in the *Arabidopsis thaliana* floral developmental pathway. **Genetics**, Pittsburgh, v. 160, p. 1641–1650, 2002.

PALHARES, A.C. ; RODRIGUES-MORAIS, T.B. ; VAN SLUYS, M.-A. ; DOMINGUES, D.S. ; MACCHERONI Jr., W. ; JORDÃO Jr., H. ; MARCONI, T.G. ; SOUZA, A.P. ; MOLLINARI, M. ; GAZAFFI, R. ; GARCIA, A. A. F. ; VIEIRA, M.L.C. . "A novel linkage map of sugarcane with evidence for clustering of retrotransposon-based markers". **BMC Genetics**, 2012. In press.

PAPA, R.; GEPTS, P. Asymmetry of gene flow and differential geographical structure of molecular diversity in wild and domesticated common bean (*Phaseolus vulgaris* L.) from Mesoamerica. **Theoretical and Applied Genetics**, Oxford, v. 106, p. 239-250, 2003.

PEREIRA, P.A.A. Evidências de domesticação e disseminação do feijoeiro comum e conseqüências para o melhoramento genético da espécie. **Pesquisa Agropecuária Brasileira**, Brasília, v. 25, n.1, p. 19-23, 1990.

RAFALSKI, A. Applications of single nucleotide polymorphisms in crop genetics. **Current Opinion in Plant Biology**, London, v. 5, p. 94-100, 2002.

RAMIREZ, M.; GRAHAM, M.A.; BLANCO-LOPEZ, L.; SILVENTE, S.; MEDRANO-SOTO, A.; BLAIR, M.W.; HERNANDEZ, G.; VANCE, C.P.; LARA, M. Sequencing and Analysis of Common Bean ESTs. Building a Foundation for Functional Genomics. **Plant Physiology**, Washington, v. 137, p. 1211-1227, 2005.

RESENDE, O.; CORRÊA, P.C.; FARONI, L.R.D.A.; CECON, P.R. Avaliação da qualidade tecnológica do feijão durante armazenamento. **Ciência e agrotecnologia**, Lavras, v. 32, n. 2, p. 517-524, 2008.

RICE P.; LONGDEN I.; BLEASBY, A. EMBOS: The European Molecular Biology Open Software Suite. **Trends in Genetics**, Amsterdam, v. 16, p. 276-277, 2000.

RÖMHELD, V.; MARSCHNER, H. Evidence for a specific uptake system for iron phytosiderophores in roots of grasses. **Plant Physiology**, Washington, v. 80, p.175–180, 1986.

ROZEN, S.; SKALETISKY, H.J. **Primer3**: pick primers from a DNA sequence. 2000. Disponível em: <<http://frodo.wi.mit.edu/primer3/input.htm>>. Acesso em: 02 fev. 2011.

SALMANOWICZ, B.P. Phaseolin seed variability in common bean (*Phaseolus vulgaris*) by capillary gel electrophoresis. **Journal of Applied Genetics**, Poznan, v. 42, p. 269-281, 2001.

SAMBROOK, J.; RUSSEL, D.W. **Molecular cloning**: a laboratory manual. 3rd ed. New York: cold Sprig HARBOR Laboratory Press, 2001. 999 p.

SANGER, F.; NICKEN, S.; COULSON, A.R. DNA sequencing with chain terminating inhibitors. **Proceedings of the National Academy of Sciences**, Washington, v. 74, p. 5463-5467, 1997.

SANTINI, L. **Mapeamento de genes análogos de resistência a patógenos em feijão** (*Phaseolus vulgaris*). 2009. 95p. Dissertação (Mestrado em Genética e Melhoramento de Plantas) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2009.

SATHE, S.K.; DESHPANDE, S.S.; SALUNKHE, D.K. Dry beans of *Phaseolus*. A review. Part 2. Chemical composition: carbohydrates, fiber, minerals, vitamins and lipids. **Critical Reviews in Food Science and Nutrition**, Cleveland, v. 21, n. 2, p. 41-91, 1984.

SCHAWANITZ, F. **The origin of cultivated plants**. Cambridge: Harvard University Press, 1967. 488 p.

SCHMID K.J.; SORENSEN T.R.; STRACKE R. Large-scale identification and analysis of genomewide single-nucleotide polymorphisms for mapping in *Arabidopsis thaliana*. **Genome**, Woodbury, v. 13, p. 1250-1257, 2003.

SINGH, S.P.; NODARI, R.; GEPTS, P. Genetic diversity in cultivated common bean.I. Allozymes. **Crop Science**, Madison, v. 31, p.19-23, 1991.

SINGH, S.P. Broadening the genetic base of common bean cultivars: a review. **Crop Science**, Madison, v. 41, p. 1659-1675, 2001.

SLIGHTOM, J.L.; DRONG, R.F.; KLASSY, R.C.; HOFFMAN, L.M. Nucleotide sequence from phaseolin cDNA clones: the major storage proteins from *Phaseolus vulgaris* are encoded by two unique gene families. **Nucleic Acids Research**, Oxford, v. 13, p. 6483-6498, 1985.

SMARTT, J. Evolution of American Phaseolus beans under domestication. In UCKO, P. J.; DIMBLEBY, G. W. (Eds). **The Domestication and Exploitation of Plants and Animals**. London: Duckworth, 1969. p. 451-454.

SOLEIMANI, V.D.; BAUM, B.R.; JOHNSON, D.A. Efficient validation of single nucleotide polymorphisms in plants by allele-specific PCR, with an example from barley. **Plant Molecular Biology Reporter**, Athens, v. 21, p. 281-288, 2003.

SPRUYT, M.; BUQUICCHIO, F. **Gene Runner for Windows**, 1994. Disponível em: <http://www.generunner.net/> Acesso em : 05 fev, 2011.

TAJIMA, F. Statistical methods to test for nucleotide mutation hypothesis by DNA polymorphism. **Genetics**, Pittsburgh, v. 123, p. 585-595, 1989.

TAMURA, K.; DDUDLEY, J.; NEI, M.; KUMAR, S. MEGA 5.1: Molecular evolutionary genetics analysis (MEGA) software version 5.0. **Molecular Biology and Evolution**, Chicago, v. 24, p. 1596-1599, 2011.

TAMURA, K. Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G + C-content biases. **Molecular Biology and Evolution**, Chicago, v. 9, p. 678-687, 1992.

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

TENAILLON, M.I.; SAWKINS, M.C.; LONG, A.D.; GAUT, R.L.; DOEBLEY, J.F.; GAUT, B.S. Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays ssp. mays* L.). **Proceedings of the National Academy of Sciences**, Washington, v. 98, p. 9161-9166, 2001.

THOMPSON, J.D.; HIGGINS, D.G.; GIBSON, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. **Nucleic Acids Research**, Oxford, v. 22, p. 4673-4680, 1994.

TOHME, J.; ORLANDO, G.; BEEBE, S.; DUQUE, M. C. AFLP analysis of gene pools of a wild bean core collection. **Crop Science**, Madison, v. 36, p. 1375-1384, 1996.

VAN DER LINDEN, C.G.; WOUNTERS, D.C.A.E.; MIHALKA, V.; KOCHIEVA, E.Z.; SMULDERS, M.J. M.; VOSMAN, B. Efficient targeting of plant disease resistance loci using NBS profiling. **Theoretical and Applied Genetics**, Oxford, v. 109, p. 384-393, 2004.

VAVILOV, N.I. **The origin, variation, immunity and breeding of cultivated plants**. Massachusetts: Chronica Botanica, 1950. 366 p.

VIEIRA, C. Leguminosas de grãos: importância na agricultura e na alimentação humana. **Informe agropecuário**, Belo Horizonte, v. 16, n. 174, p. 5-11, 1992.

VIEIRA, M.L.C.; VELLO, N.A.; SILVA FILHO, M.C. Genética e Melhoramento Vegetal. In: MIR, L. (Org.). **Genômica**. São Paulo: Atheneu, 2004. p. 679-699.

VOS, P.; HOGERS, R.; BLEEKER, M.; REIJANS, M.; VANDELEE, T.; HORNES, M.; FRIJTERS, A.; POT, J.; PELEMAN, J.; KUIPER, M.; ZABEAU, M. AFLP - A new technique for DNA-fingerprinting. **Nucleic Acids Research**, Oxford, v. 23, p. 4407-4414, 1995.

VOYSEST, O.; VALENCIA, M.C.; AMEZQUITA, M.C. Genetic diversity among Latin American Andean and Mesoamerican common bean cultivars. **Crop Science**, Madison, v. 34, p. 1100-1110, 1994.

WANDER, A.E.; RAMALHO, M.A.P.; ANDRADE M.J.B. de. **Cultivo do feijão da primeira e segunda safras na região Sul de Minas Gerais**: Introdução e importância econômica, 2005. 2 p. Disponível em:
<[HTTP://sistemasdeproducao.cnptia.embrapa.br/FontesHTML/Feijao/FeijaoPrimSegSafrasulMG/index.htm](http://sistemasdeproducao.cnptia.embrapa.br/FontesHTML/Feijao/FeijaoPrimSegSafrasulMG/index.htm)>. Acesso em: 10 fev. 2012.

WATERS, B.M.; BLEVINS, D.G.; EDIE, D.J. Characterization of FRO1, a pea ferric-chelate reductase involved in root iron acquisition. **Plant Physiology**, Washington, v. 129, p. 85-94, 2002.

WATTERSON, G.A. On the number of segregating sites in genetical models without recombination. **Theoretical Population Biology**, Amsterdam, v. 7, p. 256-276, 1975.

WICK, M.; PINGGERA, W.; LEHMANN, P. **Iron Metabolism**: Diagnosis and Therapy of Anemias. Austria: Springer Verlag GmbH, 1996. 129p.

WILLIAMS, J.G.K.; KUBELIK, A.R.; LIAK, K.J.; RAFALSKI, J.A.; TINGEY, S.V. DNA polymorphisms amplified by arbitrary primers are useful as genetic-markers. **Nucleic Acids Research**, Oxford, v. 18, p. 6531-6535, 1990.

XIA, X.; XIE, Z. Dambe: Data analysis in molecular biology and evolution. **Journal of Molecular Evolution**, Chicago, v. 9, p. 315-329, 2001.

ZHU, H.; KIM, D.J.; BAEK, J.M.; CHOI, H.K.; ELLIS, L.C.; KUESTER, H.; MCCOMBIE, W.R.; PENG, H.M.; COOK, D.R. Syntenic relationships between *Medicago truncatula* and *Arabidopsis* reveal extensive divergence of genome organization. **Plant Physiology**, Washington, v.131, p.1018-1026, 2003.