

DANIELLE FERNANDES DA SILVA

**IDENTIFICAÇÃO E VALIDAÇÃO DE MARCADORES
MOLECULARES INDEL E SNP PARA LIPOXIGENASES DE
SEMENTES DE SOJA**

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

**VIÇOSA
MINAS GERAIS - BRASIL
2009**

DANIELLE FERNANDES DA SILVA

**Identificação e validação de marcadores moleculares indel e
SNP para lipoxigenases de sementes da soja**

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

APROVADA: 06 de Julho de 2009.

Prof. Everaldo Gonçalves de Barros
(Co-orientador)

Prof. Fábio Demolinari de Miranda

Dr. Ivan Schuster

Dra. Rita Maria Alves de Moraes

Prof. Maurilio Alves Moreira
Orientador

A Deus,

Aos meus pais, Maria Gorete e Manoel (in memoriam),

A minha avó Maria,

Ao meu namorado Guilherme.

DEDICO

AGRADECIMENTOS

A Deus pelo dom da vida, por permitir que eu sempre convivesse com pessoas excepcionalmente admiráveis, que tanto contribuíram e contribuem para o meu profissionalismo e humanidade. Por sempre iluminar os meus passos, e me dar forças para superar as dificuldades.

Ao meu pai, Manoel que sempre viverá em mim. Desde sempre e para sempre meu exemplo de alegria, honestidade e amor ao próximo.

À minha mãe Maria Gorete, minha primeira professora, por sempre torcer por mim, me incentivar e preparar tudo sempre para minha felicidade e bem-estar.

À minha avó Maria pelo exemplo de simplicidade, doçura e fé.

À minha madrinha Lili por sempre me tratar com carinho, e torcer pelo meu sucesso.

Ao meu namorado Guilherme pelo amor, paciência, e dedicação em todos os momentos, por me ajudar sempre a cumprir com minhas responsabilidades e por fazer sempre tudo para me deixar feliz.

Às minhas irmãs Vanessa e Andressa pela paciência, compreensão e pela torcida.

A todos meus familiares, especialmente meus tios maternos por sempre terem se empenhado pra fazer da minha pequena família uma grande família.

À família do Guilherme por sempre torcerem por mim.

À Universidade Federal de Viçosa e ao Programa de Pós-Graduação em Bioquímica Agrícola, pela oportunidade de realização deste curso.

Ao Instituto de Biotecnologia Aplicada à Agropecuária (BIOAGRO) por fornecer a estrutura e ambiente para a realização deste trabalho.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa de estudo.

Ao Professor Maurílio Alves Moreira, pela orientação, amizade, por sempre ter confiado em mim.

Ao Professor Everaldo G. de Barros pela co-orientação, pelas sábias sugestões e amizade.

Ao Newton pela ajuda com os materiais para os experimentos, pelas sugestões e pela amizade.

Ao Professor Fábio D. de Miranda pelas sugestões no início deste trabalho e por ter aceitado participar da banca examinadora

À Doutora Rita Maria A. de Moraes por tudo que fez por mim, ainda no início da graduação e por ter aceitado participar da minha banca examinadora.

Ao Dr. Ivan Schuster por ter aceitado participar da minha banca.

Aos Estudantes que muito me ajudaram: Dodora, Eduardo e Ana Cristina, pela amizade, paciência e presteza na execução dos experimentos.

À Mariana por ter me ajudado no desenho dos *primers* e pela amizade e torcida.

Às doçuras da minha república e aos respectivos agregados: Sustanis e Aníbal, Sabina e Fernando, Cíntia, Gisele e André, Luíra, Gigi, Carol e Camila; que fizeram desses dois anos os melhores da minha vida.

Aos companheiros de curso e amigos da UFV, em especial: Pricila Larissa, Jerusa, Maíra e Orlando, pelo companheirismo e pela força empenhada nos momentos de maior dificuldade.

Aos amigos e colegas dos Laboratórios de Proteínas e Biomol pela agradável convivência, por ajudarem as vezes que precisei, em especial: Márcia, Leandro, Renato, Josiane, Giselle, Fernanda, Demerson, Leonardo, Marta, Suelen, Janaína e Luiz. Sentirei saudades de vocês!

Aos amigos do Laboratório de Sequenciamento de DNA pela excelente convivência e ajuda.

Aos funcionários Naldo, Cássio, Gláucia e Zé Carlos pela grande ajuda que me deram sempre com boa vontade.

À todos os cidadãos brasileiros que ao pagarem corretamente seus impostos contribuíram em parte para a bolsa de estudos sem a qual esse trabalho não se realizaria.

BIOGRAFIA

Danielle Fernandes da Silva, filha de Manoel Durso da Silva e Maria Gorete Fernandes da Silva, nasceu em 16 de junho de 1986, em Ubá, MG.

Estudou o ensino médio no Colégio Sagrado Coração de Maria em Ubá, concluindo este em dezembro de 2002.

Em março de 2003, ingressou no bacharelado em Bioquímica da Universidade Federal de Viçosa, MG, graduando-se em agosto de 2007. Durante a graduação, desenvolveu atividades de estágio e Iniciação Científica nas linhas de pesquisa de Métodos bioquímicos de análise, Enzimologia, Genética Molecular de Plantas e Biologia Molecular todas no BIOAGRO/UFV.

Em Agosto de 2007, ingressou no Programa de Pós-Graduação em Bioquímica Agrícola, nível de mestrado, na Universidade Federal de Viçosa, MG, submetendo-se à defesa da dissertação em julho de 2009.

SUMÁRIO

LISTA DE ABREVIACÕES.....	VIII
RESUMO	IX
ABSTRACT	XI
1. INTRODUÇÃO.....	1
2. REVISÃO DE LITERATURA	3
2.1. A SOJA	3
2.2. AS LIPOXIGENASES.....	4
2.3. MARCADORES MOLECULARES	7
2.4. POLIMORFISMO INDEL E SNP	10
2.4.1. Diversidade nucleotídica	10
2.4.2. Marcadores indel e SNP	11
2.4.3. Identificação dos polimorfismos e genotipagem.....	14
3. OBJETIVOS	19
3.1. OBJETIVO GERAL.....	19
3.2. OBJETIVOS ESPECÍFICOS	19
4. MATERIAIS E MÉTODOS.....	20
4.1. SELEÇÃO E PLANTIO DOS MATERIAIS GENÉTICOS.....	20
4.2. POPULAÇÃO PARA A GENOTIPAGEM.....	21
4.3. SEQÜÊNCIAS DE DNA ESTUDADAS.....	21
4.3.1. Lipoxigenase 2	21
4.3.2. Lipoxigenase 3	21
4.4. DESENHO DOS PRIMERS ESPECÍFICOS.....	22
4.4.1. Lipoxigenase 2	22
4.4.2. Lipoxigenase 3	23
4.5. EXTRAÇÃO DE DNA TOTAL DE FOLHAS DE SOJA	233
4.6. AMPLIFICAÇÃO E PURIFICAÇÃO DOS FRAGMENTOS.....	23
4.6.1. Lipoxigenase 2	24
4.6.2. Lipoxigenase 3	24
4.7. SEQUENCIAMENTO AUTOMÁTICO.....	25

4.8.	ANÁLISE, COMPARAÇÃO DAS SEQÜÊNCIAS E IDENTIFICAÇÃO DE POLIMORFISMOS.....	25
4.9.	AVALIAÇÃO FENOTÍPICA DA POPULAÇÃO PARA A GENOTIPAGEM	26
4.10.	GENOTIPAGEM DA POPULAÇÃO CONTRASTANTE	27
4.10.1.	Lipoxigenase 2	27
4.10.2.	Lipoxigenase 3	27
5.	RESULTADOS E DISCUSSÃO.....	28
5.1	AMPLIFICAÇÃO DE FRAGMENTOS DOS GENES <i>LOX</i> POR PCR.....	28
5.1.1.	Lipoxigenase 2	28
5.1.2.	Lipoxigenase 3	29
5.2.	IDENTIFICAÇÃO DE POLIMORFISMOS	322
5.2.1.	Lipoxigenase 2	333
5.2.2.	Lipoxigenase 3	34
5.2.2.1.	Polimorfismos Totais.....	34
5.2.2.2.	<i>Polimorfismos entre variedades Normais.....</i>	<i>377</i>
5.2.2.3.	Polimorfismos entre variedades Normais X Triplo Nula	42
5.3	GENOTIPAGEM DA POPULAÇÃO DE RILS CONTRASTANTES PARA AS LIPOXIGENASES	43
5.3.1.	Lipoxigenase 2	43
5.3.2.	Lipoxigenase 3	46
6.	CONCLUSÕES.....	48
7.	PERSPECTIVAS	49
8.	REFERÊNCIAS BIBLIOGRÁFICAS	50
	ANEXO	63

LISTA DE ABREVIações

A - Adenina

BLAST- *Basic Local Alignment Search Tool*

C - Citosina

CAPS- *Cleaved Amplified Polymorphics Sequence*

DL - Desequilíbrio de ligação

DNA - Ácido desoxirribonucléico

dNTP - desoxirribonucleotídeos

EC - *Enzyme Commission*

EST- *Expressed sequence tags*

G – Guanina

ha - Hectare

Kb - Quilobase

LOX - - Fenótipo negativo para lipoxigenase

LOX + - Fenótipo positivo para lipoxigenase

LOX 1 - Lipoxigenase 1

LOX 2 - Lipoxigenase 2

LOX 3 - Lipoxigenase 3

M - Molar

mL - Mililitro

mM - Milimolar

mRNA - RNA mensageiro

NCBI - *National Center for Biotechnology Information*

ng- Nanograma

pb - Pares de bases

PDB- *Protein Data Bank*

PCR - *Polymerase Chain Reaction*

mRNA - RNA mensageiro

RILs – *Recombinant Inbred Lines*

SNP - *Single Nucleotide Polymorphism*

t - Tonelada

T-Timina

µL - Microlitros

µM - Micromolar

RESUMO

SILVA, Danielle Fernandes da, M.Sc. Universidade Federal de Viçosa, Julho de 2009. **Identificação e validação de marcadores moleculares indel e SNP para lipoxigenases de sementes de soja.** Orientador: Maurílio Alves Moreira. Co-orientadores: Everaldo Gonçalves de Barros e Luiz Orlando Oliveira.

A degradação oxidativa dos ácidos graxos da fração óleo de soja, que ocorre durante o processamento dos grãos, desenvolve o sabor de feijão cru que altera a palatabilidade e conseqüentemente aceitabilidade dos produtos à base de soja, resultando em problemas para a indústria de alimentos. A enzima lipoxigenase (linoleato: O₂ oxirredutase, EC 1.13.11.12), catalizadora dessa reação oxidativa, está presente na semente de soja sob a forma de três isoenzimas codificadas por três genes dominantes, *Lox1*, *Lox2* e *Lox3*. A seleção assistida por marcadores moleculares é uma área que possui grande impacto dentro dos programas de melhoramento. Recentemente, uma nova classe de marcadores moleculares, denominados indel (Inserção/deleção) e SNP (*Single Nucleotide Polymorphism*), vem sendo utilizada. Estes se baseiam na detecção de polimorfismos resultantes da alteração de uma única base no genoma. A identificação de marcadores moleculares dentro dos genes que codificam as lipoxigenases poderá auxiliar na seleção de alelos recessivos que determinam a ausência de lipoxigenases na semente, acelerando assim o processo de melhoramento da soja. Os objetivos desse trabalho foram a confirmação e identificação de polimorfismos presentes nas seqüências dos genes que codificam as lipoxigenases 2 e 3, respectivamente; entre variedades, de origem genética distintas, utilizadas no Programa de Melhoramento da Qualidade da Soja-BIOAGRO/UFV e validação de alguns desses polimorfismos como marcadores moleculares. Um par de *primers* para a LOX2 e 14 pares de *primers* desenhados para amplificar todo o gene da LOX3, foram utilizados para a obtenção dos fragmentos dos genes. As seqüências foram obtidas a partir do seqüenciamento direto dos produtos de PCR de cada par de *primer*. Posteriormente, uma população de RILs contrastantes para presença/ausência das lipoxigenases das sementes, foi

utilizada para a validação de dois SNPs como marcadores dos genes LOX 2 e um indel como marcador para LOX 3 . O método de genotipagem usado foi o seqüenciamento direto dos fragmentos que continham os polimorfismos em estudo. Dois SNPs para LOX 2 e o indel contido no primeiro éxon do gene LOX 3 co-segregaram com o caráter presença/ausência de LOX 2 e 3 respectivamente nos indivíduos da população de RILs. Esses polimorfismos têm potencial para serem utilizados como ferramentas no melhoramento da soja assistido por marcadores moleculares.

ABSTRACT

SILVA, Danielle Fernandes da, M.Sc. Universidade Federal de Viçosa, July, 2009. **Identification and validation of indel and SNP molecular markers for soybean seed lipoxygenases.** Adviser: Maurílio Alves Moreira. Co-advisers: Everaldo Gonçalves de Barros and Luiz Orlando Oliveira.

The oxidative degradation of fatty acids from soybean oil fraction, that occurs during the grain processing, develop a flavor of “unripe bean” that modify the palatability and therefore the acceptability of soybean products, resulting in problems for the food industry. The enzyme lipoxygenases (linoleate: O₂ oxidoreductase, EC 1.13.11.12), that catalyses this oxidation reaction, are present in soybean seed in the three isozyme forms encoded by three dominant genes, *Lox1*, *Lox2* and *Lox3*. The selection assisted by molecular markers is an area that has great impact within breeding programs. Recently, a new class of molecular markers, called indel (insertion / deletion) and SNP (Single Nucleotide Polymorphism) have been used. These are based on detection of polymorphisms by the change of a single base in the genome. The identification of molecular markers within the genes encoding lipoxygenase may help in the selection of recessive alleles that determine the absence of lipoxygenase in the seed, accelerating the process of soybean breeding. The objective of this study was verify and identify the polymorphisms in the sequences of genes *Lox2* and *Lox3*, respectively, between varieties of different genetic origin used in the Program of Improvement and quality of Soybean at BIOAGRO/UFV, and to validate some of these polymorphisms as molecular markers. A pair of primers for LOX2 and 14 pairs of primers designed to amplify the entire gene LOX3 were used to obtain genes fragments. The sequences were obtained from the direct sequencing of PCR products from each primer pair. Subsequently, a population of RILs contrasting for the presence / absence of seed lipoxygenases, were used for the validation of two SNPs as markers for genes *Lox 2* and a indel marker for LOX 3. The genotyping method used was the direct sequencing of fragments that contained the polymorphisms studied. Two SNPs for LOX 2 and the indel contained in the first exon of the gene LOX 3 were co-segregated with the character presence/absence of LOX 2 and 3

respectively in the individuals of the population of RILs. These polymorphisms have the potential to be used as tools in soybean breeding assisted by molecular markers.

1. INTRODUÇÃO

A soja (*Glycine max* (L.) Merrill) constitui o principal produto agrícola de exportação brasileiro, sendo cultivada praticamente em todo território nacional. O grande interesse no cultivo da soja deve-se a composição típica do grão, que apresenta em torno de 40% de teor protéico e 20% de teor de óleo como também a alta produtividade de seus cultivares.

As características nutricionais dos grãos de soja têm sido amplamente estudadas, uma vez que a soja possui alto teor de proteína de boa qualidade, bom teor de fibras e quantidade reduzida de gordura saturada, presença de isoflavonas, entre outras características, que conferem uma relação benéfica entre o consumo de soja e a saúde humana. Estudos relatam uma relação entre o consumo de soja e a redução dos riscos de doenças crônicas não-infecciosas como as doenças cardiovasculares, alguns tipos de cânceres e osteoporose.

A utilização dos produtos derivados de soja no consumo humano é ainda limitada, em consequência das características indesejáveis encontradas no grão tais como textura, odores desagradáveis e sabor adstringente os quais são proporcionados em grande parte pelas enzimas lipoxigenases presentes.

A degradação oxidativa dos ácidos graxos de reserva da soja, ocorrida durante o processamento dos grãos, é responsável pelo sabor de feijão cru também denominado “beany flavor”. A enzima lipoxigenase (linoleato: O₂ oxirredutase, EC 1.13.11.12), catalizadora dessa reação, está presente nas sementes de soja sob a forma de três isoenzimas codificadas por três genes dominantes, *Lox1*, *Lox2* e *Lox3*. Os produtos da ação dessas enzimas são hidroperóxidos, que se degradam formando compostos carbonílicos responsáveis pela palatabilidade indesejada dos produtos à base de soja.

A seleção assistida por marcadores moleculares é uma área que possui grande impacto dentro dos programas de melhoramento de plantas na atualidade. Constitui-se em um processo rápido, eficiente e seguro, sendo de fundamental importância para o desenvolvimento de novos cultivares. Dentre os marcadores moleculares se destacam os marcadores indel e SNP que são muito convenientes aos melhoristas de plantas, por combinarem muitas

vantagens, entre elas sua natureza codominante, elevado grau de polimorfismos, abundância e presença em todo o genoma de plantas, além dos varios métodos disponíveis para acessar esses polimorfismos.

A identificação de marcadores moleculares indel e SNP ligados aos genes que codificam as lipoxigenases de semente de soja pode auxiliar no monitoramento na introdução da característica ausência de LOX em cultivares comerciais, acelerando assim o processo de melhoramento.

Dentro deste contexto, objetivou-se com este trabalho, a confirmação dos polimorfismos descritos por KIM *et al.* (2004) no gene *Lox2* e identificação de novos polimorfismos no gene *Lox3* de soja nos genótipos do programa de melhoramento da qualidade da soja da UFV (PMQS-BIOAGRO/UFV), e posterior validação desses polimorfismos como possíveis marcadores para essas características.

2. REVISÃO DE LITERATURA

2.1. A soja

A soja é uma leguminosa originária da China, região de clima temperado, mas apresenta boa adaptação às condições climáticas encontradas no Brasil, devido principalmente aos trabalhos de melhoramento genético (GONDIM, 2001). Constitui atualmente um dos produtos agrícolas de maior importância no Brasil (Tabela 1), ocupando lugar de destaque na pauta de exportação, com uma produção na safra 2008/2009 de cerca de 57 milhões de toneladas (CONAB 2009).

Tabela 1 – Comparação da produção das principais culturas cultivadas no Brasil

Culturas	Total Área Plantada (1000 ha)	Produção Total (1000 t)	Exportação grãos¹ (t)
Soja	21.730,5	57.137,4	12.601.504
Milho	14.122,3	49.882,3	2.951.907
Trigo	2.423,3	6.015,6	284.196
Feijão	4.159,9	3.524,2	763
Arroz	2.924,7	12.744,3	205.063
Algodão	833,2	1.857,8	216.523

¹Jan/Abr 2009 – CONAB (2009)

Esta leguminosa tem sido a principal fonte de proteína na dieta de milhões de pessoas na Ásia, o que não pode ser dito para o resto do mundo, onde o seu consumo direto e de derivados é ainda pequeno, quando comparado ao uso do óleo (MOREIRA, 1999).

Desta forma, a soja constitui um bom alvo para o melhoramento genético, por representar uma cultura de grande interesse econômico e constituir uma importante fonte de óleo e proteínas (YADAV, 1996). O desenvolvimento de programas de melhoramento de soja no Brasil possibilitou o lançamento de novas cultivares, contribuindo para a grande expansão do cultivo da soja em diversas regiões do país (SEDIYAMA *et al.*, 1999).

As características nutricionais dos grãos de soja têm sido amplamente estudadas, uma vez que a soja possui alto teor de proteína de boa qualidade, bom teor de fibras e quantidade reduzida de gordura saturada conferem uma relação benéfica entre o consumo de soja e a saúde humana. Estudos relatam uma relação entre o consumo de soja e a redução dos riscos de doenças crônicas não-infecciosas como as doenças cardiovasculares, alguns tipos de cânceres e osteoporose. Entretanto a utilização do produto no consumo humano é limitada pela presença de fatores antinutricionais e das lipoxigenases (SILVA *et al.*, 2006).

A soja integral pode ser considerada um alimento de ótima qualidade, porém os produtos derivados da soja não têm boa aceitação no mercado em consequência das características indesejáveis encontradas no grão tais como textura, odores desagradáveis e sabor adstringente, os quais são proporcionados em grande parte pela presença de enzimas lipoxigenases (VANDERGRIFT, 1983, MARTINS *et al.* 2002).

2.2. As lipoxigenases

As lipoxigenases são enzimas que catalisam a adição de oxigênio molecular aos ácidos graxos poliinsaturados, ácido linolênico e linoléico (Figura 1).

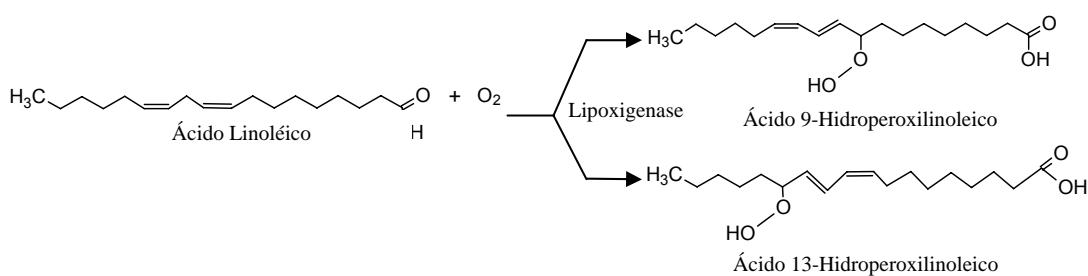


Figura 1 – Reação primária catalisada pelas lipoxigenases usando o ácido linoléico como substrato indicando os dois possíveis produtos, ácido 9- e 13-hidroperoxilinoico.

Especificamente, as lipoxigenases são dioxigenases que contêm um átomo de ferro associado às suas estruturas tridimensionais, responsáveis por catalisar a oxigenação estereoespecífica e a hidroxidação de lipídios,

contendo uma dupla ligação do tipo cis, cis-1,4-pentadieno em suas estruturas. Nas sementes de plantas estas enzimas utilizam os ácidos linoléico e linolênico como substratos.

Todas as lipoxigenases são proteínas que possuem apenas uma subunidade polipeptídica e podem ser encontradas nas sementes nas três formas isoenzimáticas: lipoxigenase 1 (LOX 1), lipoxigenase 2 (LOX 2) e lipoxigenase 3 (LOX 3). A herança que condiciona a ausência/presença das lipoxigenases nos grãos de soja é mendeliana simples, sendo que os alelos que determinam a ausência dessas isoenzimas são recessivos. Os locos que controlam a LOX 1 e 2 estão ligados e o loco que controla LOX3 apresenta segregação independente (KITAMURA *et al.*, 1983).

AXELROD *et al.* (1981) purificaram quatro isoenzimas distintas de lipoxigenase de semente de soja, que foram designadas LOX 1, LOX 2, LOX3a e LOX3b. Contudo, pela grande similaridade, tanto no comportamento quanto na composição das isoenzimas LOX3a e LOX3b esses autores consideraram, para efeito analítico, a existência de somente uma isoenzima LOX 3. As três isoenzimas são globulares, solúveis em soluções salinas diluídas, sendo cada uma formada por uma única cadeia polipeptídica (SIEDOW, 1991). As lipoxigenases de sementes de soja têm sido bastante estudadas, devido a sua importância na determinação da qualidade dos produtos de soja (HILDEBRAND, 1988).

A LOX 1 é a mais efetiva na hidroperoxidação de ácidos graxos livres, com pH ótimo em torno de 9; enquanto que a LOX 2 e 3 são mais efetivas na hidroperoxidação de ácidos graxos neutros, como metil ésteres de ácidos graxos (GARDNER *et al.*, 1995). Os valores de pH ótimo para cada enzima LOX 2 e 3 são respectivamente em torno de 6,5 e 7 (SIEDOW, 1991).

Dentre as três lipoxigenases, a LOX 1 é a que apresenta maior atividade “*in vitro*” e maior estabilidade térmica. As três lipoxigenases de sementes são proteínas monoméricas similares, com massas moleculares entre 94 e 97 kD, apresentando pontos isoelétricos distintos (HILDEBRAND, 1988).

Estas enzimas estão amplamente distribuídas na natureza. A atividade de lipoxigenases em soja não varia somente entre os diferentes órgãos e tecidos, mas também entre os diferentes estádios do desenvolvimento da planta

(MULLER, 1997). As lipoxigenases de plantas estão envolvidas em uma série de diferentes aspectos da fisiologia da planta, incluindo o crescimento e desenvolvimento, resistência a pragas, senescência e respostas a ferimentos (PDB).

O sabor e odor desagradáveis presentes em produtos derivados da soja conhecidos como “beany flavor” são originados da associação de compostos carbonílicos de cadeia curta com a fração protéica. Estes compostos são produtos finais de uma série de reações que se inicia com a hidroperoxidação de ácidos graxos poliinsaturados, catalisada pelas lipoxigenases que constituem aproximadamente 1% do total das proteínas do grão de soja (MONTEIRO *et al.*, 2004). A eliminação genética das lipoxigenases é considerada a forma mais eficiente de reduzir o “beany flavor” (sabor de feijão cru), porque o tratamento térmico utilizado para a inativação dessas isoenzimas, apesar de ter eficiência razoável, prejudica sensivelmente as propriedades funcionais das proteínas presentes no grão (KINSELLA, 1979).

A seleção de genótipos superiores por meio do uso de marcadores moleculares de DNA em programas que objetivam a introgressão de genes por meio de retrocruzamentos é um exemplo largamente utilizado de seleção assistida que busca principalmente introduzir características de herança simples, selecionar genótipos mais próximos ao genótipo recorrente com menor número de ciclos, acelerando assim a obtenção de novos cultivares melhorados (CARPENTIERI-PÍPOLO *et al.*, 2006).

No Programa de Melhoramento da Qualidade da Soja para a Agroindústria desenvolvido no Instituto de Biotecnologia Aplicada à Agropecuária da Universidade Federal de Viçosa (PMQS-BIOAGRO/UFV) foram desenvolvidos cultivares comerciais que apresentam ausência de LOX na semente, por meio de retrocruzamentos, visando à diminuição do “beany flavor”.

A combinação de métodos clássicos de melhoramento genético com tecnologias de genética molecular de análises de DNA têm contribuído significativamente para a ampliação do conhecimento genético e acelerar os programas de melhoramento (FERREIRA & GRATTAPAGLIA, 1998).

2.3. Marcadores Moleculares

Segundo SAKIYAMA (1993) o termo marcador tem sido utilizado para designar fatores morfológicos, fisiológicos, bioquímicos ou genéticos que por co-segregarem com os genes de interesse, permitem o estudo comparativo de genótipos e de suas progênies. Entre esses, os marcadores de DNA são os mais utilizados na atualidade.

Nos estudos de genética e melhoramento, nos anos 60, eram principalmente utilizados marcadores morfológicos, ou seja, fenótipos de fácil identificação visual, como cor dos animais, das pétalas, morfologia foliar, entre outros. No entanto, devido ao reduzido número de marcadores morfológicos polimórficos, a probabilidade de encontrar associações entre esses marcadores e características importantes é reduzida (FERREIRA & GRATTAPAGLIA, 1995).

Este quadro começou a mudar com a utilização dos primeiros marcadores moleculares, o termo aloenzimas introduzido por MARKET & MOLLER (1959) que define variantes de uma mesma proteína, detectados por eletroforese. Este tipo de marcador tem a vantagem de ser relativamente barato, no entanto apresenta um baixo número de variantes nas proteínas o que também limitou o desenvolvimento de mapas genéticos saturados. Em meados dos anos 80, quando os métodos de avaliação de variações do DNA tornaram-se amplamente disponíveis, marcadores baseados no DNA substituíram amplamente as aloenzimas em estudos de mapeamento (BORÉM & CAIXETA, 2006).

Considerando que o DNA é o material genético dos organismos e, que variações fenotípicas entre esses indivíduos são reflexo direto de diferenças nas seqüências de nucleotídeos do DNA; uma série de técnicas pode ser utilizada para acessar essas variações no DNA. De acordo com GUPTA (1999), os diversos marcadores moleculares podem ser divididos em três classes de acordo com o método utilizado para a sua detecção: marcadores baseados na hibridização do DNA, como os polimorfismos de comprimento de fragmentos de restrição do DNA o RFLP (*Restriction Fragment Length Polymorphism*); os baseados na reação em cadeia da polimerase PCR (*Polymerase Chain Reaction*), tais como RAPD, SCAR, AFLP e SSR; e por

último os baseados na variação entre seqüências e na utilização de *chips* de DNA, como exemplo os marcadores indel e SNP- *Single Nucleotide Polymorphism*.

Os polimorfismos de comprimento de fragmentos de restrição do DNA (RFLP) foram os primeiros marcadores moleculares desenvolvidos. A metodologia RFLP consiste na digestão do DNA com uma variedade de enzimas de restrição e análise eletroforética do DNA após sofrerem clivagem específica após o uso de enzimas de restrição. Após a separação dos fragmentos, bandas individuais são detectadas pela utilização de sondas de DNA marcadas, que tem bases complementares à determinada região do genoma. Marcadores RFLP possuem expressão codominante, assim, possibilitando a identificação de genótipos homozigotos e heterozigotos (BOLTSTEIN, 1989). Os polimorfismos detectados pelo RFLP podem ter se resultado de mudanças no DNA de uma única base tais como inserções e deleções causando uma perda ou ganho de sítios de restrição (McCOUCH *et al.*, 1988; EDWARDS *et al.*, 2004). Esse marcador foi utilizado na construção de mapas genéticos de diversas espécies vegetais, tais como soja (NJITI *et al.*, 2002); milho (GROH *et al.*, 1998; SÉNE *et al.*, 2000; CARDINAL *et al.* 2001), arroz (HE *et al.*, 2001; CHEN *et al.*,2003; UGA *et al.*, 2003) e trigo (NACHIT *et al.*, 2001).

A amplificação do DNA por PCR baseia-se no pareamento de um par de *primers*, que são pequenas moléculas de DNA fita simples, utilizadas como iniciadores e que delimitam a seqüência de DNA de fita dupla, alvo da amplificação (MULLIS & FALLONA,1987). Marcadores baseados no PCR requerem muito menos DNA que os marcadores RFLP, além de serem mais compatíveis com a genotipagem automatizada (GUIMARÃES *et al.*, 2007).

Na técnica de RAPD, fragmentos de DNA são amplificados no genoma com a utilização de *primers* curtos, aproximadamente 10 nucleotídeos, de seqüência arbitrária (WILLIAMS *et al.*, 1990; WELSH & McCLELLAND, 1990). Os marcadores RAPD têm a vantagem de que um simples *primer* poder revelar muitos locos de uma só vez em que cada loco corresponde a uma região diferente do genoma analisado. Contudo, não é possível a distinção do genótipo homozigoto dominante do heterozigoto, pois estes marcadores são

dominantes. Os marcadores RAPD foram utilizados no mapeamento genético de vários organismos, como soja (HNETKOVSKY *et al.*, 1996; CHANG *et al.*, 1997; FERREIRA *et al.*, 2000; NJITI *et al.*, 2002) ; arroz (CAI & MORISHIMA, 2002; UGA *et al.*, 2003) e eucalipto (FERREIRA & GRATTAPAGLIA, 1995).

Os AFLPs são marcadores derivados da amplificação seletiva de fragmentos de restrição (VOS *et al.*, 1995) . Permite a obtenção de grande número de polimorfismo e isso tem levado a sua utilização na obtenção de mapas altamente saturados (BALLVORA *et al.*, 1995, MEKSEM *et al.*, 1995, GEBHARDT, 1995). Os marcadores AFLP têm muitas das vantagens dos marcadores RAPD, e ainda permitem uma maior reprodutibilidade dos dados (GUIMARÃES *et al.*, 2007).

Os microssatélites, também conhecidos como SSR (*Simple Sequence Repeats*), são marcadores moleculares com herança mendeliana e consistem de regiões de DNA de unidades mono-, di-, tri-, tetra- ou pentanucleotídica repetidas em tandem; estão dispersas ao acaso por todo o genoma (GUPTA & VARSHNEY, 2000). Os marcadores microssatélites são obtidos pela amplificação destas seqüências de nucleotídeos repetidas por meio do PCR, utilizando para isto um par de *primers* específicos contendo de 20 a 30 bases de comprimento, complementares a seqüências únicas conservadas que flanqueiam o microssatélite (WEBER & MAY, 1989). Diferenças no número de repetições em tandem é que são responsáveis pelos polimorfismos de tamanho entre os fragmentos de PCR. Devido à expressão codominante destes marcadores e ao seu multialelismo, eles podem ser aplicados a todos os tipos de populações segregantes empregadas no mapeamento genético e estudos de ligação (FERREIRA & GRATTAPAGLIA, 1995). Foram utilizados marcadores microssatélites no mapeamento da soja : CHANG *et al.*(1997), CREGAN *et al.* (1999), MEKSEM *et al.* (2001), IQBAL *et al.* (2001), YUAN *et al.* (2002) e NJITI *et al.* (2002); em trigo NACHIT *et al.* (2001) e em milho CARDINAL *et al.* (2001).

O seqüenciamento do DNA permitiu a detecção de alterações de bases únicas no DNA e possibilitaram o desenvolvimento de marcadores baseados em polimorfismos em um único par de bases (*Single Nucleotide Polymorphism* – SNP). Existe uma divergência entre os conceitos de SNP presentes na

literatura. Segundo KWOK (1996), KRUGLYAK (1997) e COLLINS (1998) SNPs são variações de nucleotídeos ou pequenas inserções/deleções nas seqüências de bases em fragmentos homólogos de DNA, sendo estes as mais freqüentes formas de polimorfismo observado no genoma humano. No entanto, BROOKES (1999), RAFALSKI (2002) e VIGNAL *et al.* (2002) separam o conceito de SNPs do de indels, sendo o primeiro ocasionado por substituições de base e o segundo por inserções e deleções porque muito provavelmente eles são originados por mecanismos genéticos distintos.

2.4. Polimorfismo indel e SNP

2.4.1. Diversidade nucleotídica

As variabilidades presentes no genoma, caracterizadas como mutações são alterações no material genético que podem ser herdáveis. Quando estas variabilidades são localizadas, ou seja, alteram apenas poucas bases, são chamadas de mutações pontuais (ALBERTS *et al.*, 2002). Essas mutações pontuais podem ser substituições ou inserções/deleções de bases (indels). As substituições podem ser de duas naturezas: as que envolvem trocas de bases de mesma natureza química, ou seja, entre duas purinas (A/G ou G/A) ou duas pirimidinas (C/T ou T/C) e são denominadas transições ou as transversões que são substituições que acarretam numa incorporação de base de natureza química distinta, ou seja, uma purina por uma pirimidina ou vice-versa (BROOKES, 1999). Essas alterações, algumas vezes, têm origem em erros de incorporação de bases durante a replicação do DNA.

No entanto, devido à redundância do código genético, nem todas as alterações de pares de base acarretarão alterações dos aminoácidos da proteína, principalmente quando ocorrem na terceira base do códon (ALBERTS *et al.*, 2002). Estas alterações de bases que não acarretam modificação do aminoácido são chamadas de silenciosas. Como este tipo de mutação não tem nenhum efeito positivo ou negativo na descendência do indivíduo, acaba por ser considerada uma mutação neutra (LEWIN *et al.*, 2000).

Os polimorfismos de um único nucleotídeo (SNPs – *single nucleotide polymorphism*) que ocorrem em éxons e que são responsáveis por uma

variação em um aminoácido da proteína são chamados de não sinônimos. Os que não acarretam variação da seqüência de aminoácidos são chamados de sinônimos, podendo a substituição ser conservativa ou não conservativa em função das características dos aminoácidos envolvidos na troca (Figura 2). Mesmo não alterando a proteína, os SNPs sinônimos podem alterar a estrutura do RNA mensageiro e desestabilizá-lo, afetando a concentração final da proteína (GRIFFITHS *et al.*, 2001).



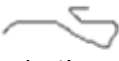
	Indivíduo 1	GCA	AGA	GAT	AAT	TGT	...	Proteína
		Ala	Arg	Asp	Asn	Cys		
	Indivíduo 2	GCG	AGA	GAT	AAT	TGT	...	Ativa
Mudança no códon, sem alterar aminoácido		Ala	Arg	Asp	Asn	Cys		
Indivíduo 3	GCA	AAA	GAT	AAT	TGT	...	Inativa	
Mudança no códon e Mudança no aminoácido		Ala	Lys	Asp	Asn	Cys		

Figura 2. Ilustração de possíveis variações na seqüência de um gene. No indivíduo 2 a substituição não alterou o aminoácido incorporado, caracterizando o SNP sinônimo. No indivíduo 3 é mostrado um SNP do tipo não-sinônimo onde a substituição alterou o aminoácido incorporado, alterando também a estrutura da proteína.

Existem também as substituições que acabam por gerar códons de terminação. Estas podem inviabilizar a funcionalidade da proteína por interromper a seqüência da mesma, fatalmente impossibilitando a sua atividade (VOET & VOET, 2006).

Quando ocorre uma inserção ou deleção (indel) dentro da região codificadora de um gene, o quadro de leitura (*frameshift*) pode ser alterado, acarretando na mudança dos aminoácidos do restante da proteína. Esse tipo de mutação é o mais prejudicial, pois acarreta mudança drástica da estrutura final do polipeptídeo (LEWIN *et al.*, 2000).

2.4.2. Marcadores indel e SNP

A disponibilidade crescente de seqüências parciais de genes, nos bancos de dados, promoveu a descoberta de uma grande quantidade de polimorfismos de uma base entre indivíduos de uma mesma espécie. Tais polimorfismos de

uma única base são denominados SNPs - *Single Nucleotide Polymorphism* (GUIMARÃES & COSTA, 2002). Para que tal posição seja considerada como um SNP, assume-se para um dado loco, que o alelo menos freqüente na população tenha uma abundância acima de 1%. Essa freqüência mínima impede que polimorfismos verdadeiros sejam confundidos com raras mutações pontuais e erros de incorporação de bases durante a manipulação *in vitro* do DNA. A princípio, os SNPs podem ser polimorfismos bi-, tri-, ou tetralélicos, entretanto na prática são em geral bialélicos (BROOKES, 1999). Segundo VIGNAL *et al.* (2002) esse fato é justificado pela ocorrência desigual (1,5 - 2,5 vezes mais) de transições (A↔G, T↔C) em comparação com as transversões (A↔C, A↔T, G↔C e G↔T) isso porque a taxa de desaminação espontânea da 5-metilcitosina para timina é alta em CG dinucleotídeos (COULONDRE *et al.*, 1978).

Indels e SNPs ocorrem com grande freqüência em todos os genomas, e provavelmente são responsáveis pela maioria das contribuições do genótipo na variação do fenótipo (BOTSTEIN & RISCH, 2003). Esses polimorfismos ocorrem tanto em regiões codificadoras (éxons), não codificadoras (íntrons), regiões transcritas e não traduzidas quanto em regiões promotoras dos genomas (RAFALSKI, 2002).

SNPs podem também afetar o *splicing* alternativo, quando se encontram nas junções íntron-éxon; gerar alterações no padrão de expressão de genes quando ocorrem em regiões promotoras; gerar ou suprimir códons de terminação ou poliadenilação na molécula de RNA mensageiro; e alterar códons de iniciação de tradução e, conseqüentemente afetar a quantidade de proteína produzida (KWOK & GU, 1999)

SNPs em regiões cis-regulatórias podem ter um impacto considerável em um fenótipo, a partir do momento que podem alterar a afinidade de fatores de transcrição por estas regiões, alterando o nível de transcrição e, podendo levar a uma mudança na disponibilidade de uma determinada proteína. Acredita-se que esses SNPs tenham um papel importante na evolução das espécies por levarem a uma modulação leve de determinados traços fenotípicos (WRAY, 2007).

Grande parte dos estudos tem relatado a associação entre um único SNP e uma determinada patologia, ou mesmo definido um SNP como um marcador para pré-disposição a uma determinada doença. Visando à compreensão das bases moleculares de doenças de características genéticas mais complexas, pesquisas têm sido intensificadas no sentido de tentar associar ao fenótipo alterado, conjuntos de determinados SNPs, denominados haplótipos (BROOKES, 1999).

Os haplótipos são constatados quando dois ou mais SNPs segregam conjuntamente (RAFALSKY, 2002). Desta maneira, juntos estarão então caracterizando um desequilíbrio de ligação (DL) que é simplificada a falta de segregação independente entre os alelos em dois ou mais locos. O DL é geralmente dependente da história de recombinações entre os polimorfismos. Fatores tais como deriva genética, seleção entre populações, migração (miscigenação) e a redução do tamanho populacional (*bottleneck*), podem modificar o DL entre os marcadores e as características relacionadas. Fatores que aumentam o DL são o endocruzamento, tamanho pequeno da população, isolamento genético entre linhagens, subdivisão entre populações, baixa taxa de recombinação, seleção natural e artificial. Por outro lado, alguns fatores que diminuem o DL são a fecundação cruzada, altas taxas de recombinação e mutação (REMINGTON, 2001).

Através da análise de SNPs, pode-se constatar a presença de blocos de DNA incluindo diferentes locos que tendem a ser herdados em conjunto, os quais são denominados haplótipos. Por terem esta característica e por estarem sendo descritos em larga escala (aproximadamente 32.000.000 de submissões no banco de dados de SNPs atualmente [GenBank]), os haplótipos tornaram-se alvos para estudos como marcadores em estudos de associação (HIRSCHHORN, 2005) de genes relacionados a doenças, estudos de demografia genética e de evolução cromossomal. Estudos têm mostrado que, no genoma humano, cerca de 11 milhões de SNPs têm grupos de vizinhos estreitamente relacionados entre si (THE INTERNATIONAL HAPMAP CONSORTIUM, 2003; HIRSCHHORN, 2005).

Os SNPs também vêm ganhando muito espaço em estudos de análises genéticas em plantas. Como exemplo, têm-se os trabalhos realizados com soja

(ZHU, *et al.*, 2003; CHOI, *et al.*, 2007), milho (CHING *et al.*, 2002; THORNSBERRY *et al.*, 2001), *Arabidopsis* (CHO *et al.*, 1999; NORDBORG *et al.*, 2002), cana-de-açúcar (GRIVET *et al.*, 2003) e trigo (SOMERS *et al.*, 2003).

Em variedades elite de milho (CHING *et al.*, 2002) e em *Populus* (INGVARSSON, 2004) ocorre um SNP a cada 60 pb. No gene *CCR* (cinamoil CoA redutase) de *Eucalyptus nitens*, THUMMA *et al.* (2005) observou um SNP a cada 94 pb. Sabendo então do padrão de desequilíbrio de ligação para determinada região, um conjunto de marcadores pode ser desenvolvido de forma a se tentar associar a presença dos mesmos a possíveis fenótipos de interesse.

Considerando o grande aumento do número de entradas de seqüências nos bancos de dados (JANSSEN, 2003; SAIER, 1998), SNPs e indels tornam-se fontes inesgotáveis de possíveis marcadores polimórficos para uso no mapeamento genético de alta resolução, assim como para estudos de associação baseados em genes candidatos ou, possivelmente, em todo o genoma (RAFALSKI, 2002). REMINGTON (2001) sugere que polimorfismos fortemente associados a um loco e posteriormente correlacionado com um fenótipo em uma população aleatória, são fontes de dados muito importantes para a saturação do mapa de ligação em torno dessa região do genoma, aumentando a resolução do mesmo.

2.4.3. Identificação dos polimorfismos e genotipagem

Com o desenvolvimento tecnológico de programas e de equipamentos cada vez mais robustos e sofisticados, vem sendo possível gerar e analisar um grande número de dados de seqüências de DNA e dessa forma identificar SNPs e indels, utilizando as ferramentas da bioinformática. (WANG *et al.*, 2005). Essas ferramentas são utilizadas em várias etapas do estudo de polimorfismos, desde a identificação das variações até a predição do efeito das mesmas. Além disso, esses recursos também são úteis na construção e manutenção de bancos de dados que podem ser acessados pela Internet e que atuam como sedes de referência para a deposição de SNPs (GUIMARÃES & COSTA, 2002). O maior banco de dados público de polimorfismos é mantido pelo *National Center for Biotechnology Information (NCBI)*, o dbSNP

(<http://www.ncbi.nlm.nih.gov/SNP>), que é banco de dados público, o qual contém seqüências de mais de cem mil espécies.

Até o presente momento mais de 250 milhões de seqüências de SNPs já foram catalogados no genoma humano, sendo estas informações, bem como para outras espécies, disponibilizadas gratuitamente no *site* do NCBI. O estudo e a identificação de tais marcadores no genoma da maioria das espécies vegetais ainda ocorrem em menor escala. No entanto, para algumas plantas, considerável progresso tem sido observado, como exemplo: *Arabidopsis thaliana* (JANDER, 2002; SCHMID, 2003; NORDBORG, 2005), milho (*Zea mays* L.) (TENAILLON, 2001; CHING, 2002), arroz (*Oryza sativa* L.) (NASU, 2002; FELTUS, 2004), entre outras espécies.

No genoma de soja, a freqüência de ocorrência de SNPs tem sido o objetivo de estudo de diversos grupos de pesquisa. Nos trabalhos realizados por SCALLOM *et al.* (1987) com genótipos de soja, 3543pb (pares de bases) do gene *Gy₄*, que codifica uma das proteínas de reserva no grão foram analisados, possibilitando a identificação de três SNPs. ZAKHAROVA *et al.* (1989) compararam 789pb de seqüências de cDNAs do gene que codifica para a subunidade A₃B₄ da proteína glicinina em três cultivares e encontraram dois SNPs.

Isso demonstra que seqüências depositadas nos bancos de dados como bibliotecas genômicas e de cDNA, podem ser utilizadas para a identificação de polimorfismos *in silico*. No entanto, erros contidos nas seqüências depositadas têm se mostrado um empecilho para estudos dessa natureza (SOMERS *et al.*, 2003).

Segundo CHOI *et al.* (2007), dois importantes fatores a respeito do genoma de uma espécie devem ser considerados na identificação de SNPs, a freqüência de variabilidade genética e a presença de regiões duplicadas. Para a cultura da soja, considerando os sucessivos processos de seleção em trabalhos de melhoramento e a estreita base genética entre as cultivares utilizadas para a obtenção das linhagens comerciais, é esperada uma baixa freqüência de seqüências variantes no genoma. No estudo realizado por ZHU *et al.* (2003) para detecção de SNPs em soja, vinte e cinco genótipos, representando a base do germoplasma das principais variedades cultivadas

nos Estados Unidos da América, foram avaliados. Os resultados deste trabalho mostraram que a frequência de ocorrência de SNPs, mensurada pela diversidade de nucleotídeos, foi de aproximadamente 0,00053 em 28,7kpb (kilo pares de bases) de seqüências codificadoras e de 0,00111 para regiões não codificadoras. Estes valores são similares aos encontrados por SACHIDANANDAM *et al.* (2001), que analisaram 1,2 bilhões de pares de bases em regiões do genoma humano, indicando a possibilidade de sucesso na identificação de marcadores SNPs no genoma da soja.

Pesquisas recentes realizadas por CHOI *et al.* (2007) possibilitaram obtenção de um mapa do genoma transcrito da soja (disponível para consulta no endereço <http://bfgl.anri.barc.usda.gov/soybean/>), no qual foram identificados e mapeados SNPs em 1141 genes de três populações de RILs. Segundo os autores deste trabalho, a utilização de SNPs aumenta a eficiência de análises genéticas que incluem clonagem posicional, análises de associação, mapeamento e detecção de QTLs e determinação de relação genética entre indivíduos. O genoma seqüenciado da soja se encontra disponível no site: <http://www.phytozome.net/soybean>, sendo essa uma ferramenta muito importante para consulta e desenvolvimento de pesquisas envolvidas com o melhoramento genético da soja.

Muitos métodos de genotipagem se baseiam na reação de PCR e/ou clivagem dos fragmentos com enzimas de restrição. Os mais comuns são: CAPS (*Cleaved amplified polymorphic sequence*) que se baseia na amplificação de fragmentos de DNA cujo padrão de bandas após a clivagem com enzima de restrição (específica para clivagem alelo específica), apresenta polimorfismos que podem ser correlacionado com características de interesse (VARSHNEY *et al.*, 2007). GAUDET *et al.* (2007) descrevem a utilização da metodologia PAMSA (*PCR amplification of multiple specific alleles*) que consiste na utilização de um primer reverso R(*Reverse*) e vários primers diretos - F(*Forward*) específicos para cada um dos alelos SNP. Cada um desses *primers* F terminam nos diferentes alelos do SNP e tem um mau pareamento a quatro pares de base do término 3' para aumentar a especificidade da amplificação alelo específica. Essas técnicas apesar de serem baratas, requerem a separação dos fragmentos em agarose, o que

consome bastante tempo. No entanto, alguns métodos não requerem a modificação ou separação dos produtos do PCR (GIBSON, 2006; MAMOTTE, 2006).

Outros métodos de genotipagem vêm sendo utilizados no melhoramento genético assistido por marcadores indel/SNP. Estes ensaios incluem: a TaqMan assay (LIVAK, 1999), análises por curva de desnaturação (BENNETT *et al.* 2003), *molecular beacons* (TYAGI *et al.*, 1998), pirosequenciamento (LANGAEE & RONAGHI, 2005), chips de DNA ou microarranjos, DHPLC - cromatografia líquida de alta pressão desnaturante (WOLFORD *et al.*, 2000) e espectrometria de massa (BLONDAL *et al.*, 2003). Uma descrição mais detalhada de duas das metodologias mais utilizadas será feita.

Metodologias denominadas *Melting curve analysis* – análises de curva de desnaturação, baseiam-se na diferença da estabilidade térmica entre os pareamentos e os mau pareamentos contidos nos duplexes de DNA (BERNARD *et al.* 1998). A temperatura de desnaturação (T_m) reflete a estabilidade térmica de um duplex de DNA e depende do comprimento do DNA dupla fita, do conteúdo de GC, e do grau de complementaridade entre as fitas (WETMUR, 1991). Assim, uma sonda contendo um mau pareamento com a seqüência alvo se desnatura, saindo desse alvo, numa temperatura mais baixa que uma sonda perfeitamente correspondente.

Essa técnica requer um termociclador *Real Time* para obtenção das curvas de desnaturação, pois esse aparelho permite a quantificação da fluorescência emitida da amostra durante o decorrer da reação. Dessa forma é possível a diferenciação dos amplicons cada um contendo um polimorfismo e conseqüentemente discriminar os alelos recessivos dos dominantes. Essa técnica pode utilizar a sonda Syber Green da Molecular Probes® para quantificar o T_m das duplas fitas, mas também uma diversidade de outras sondas fluorescentes ou fluoróforos associados aos *primers* alelos específicos (BERNARD *et al.*, 1998). A fluorescência captada pelo termociclador é que dá origem às curvas sigmóides que após sofrerem a primeira derivada geram um pico onde o máximo corresponde a temperatura de “melting” (T_m) do fragmento que foi amplificado. Os resultados são picos em T_m distintos para cada um dos homozigotos, o que caracteriza diferenças na composição de bases pela

presença do SNP, e no heterozigoto uma curva com dois máximos, indicando a presença das duas formas polimórficas. Análises de curva de desnaturação com sondas para hibridização já têm sido aplicadas para a detecção de alelos de resistência em plantas (DUFRESNE *et al.* 2004).

Uma outra metodologia de genotipagem de SNPs importante é baseada na reação de mini-sequenciamento. Dideoxinuclotídeos são utilizados para o término ou o alongamento de uma cadeia de DNA, a partir de um *primer*, o qual é desenhado à montante (*upstream*) ao sítio polimórfico.

A metodologia denominada PINPOINT, descrita originalmente por HAFF *et al.* (1997) para genotipagem de SNPs por MALDI-TOF, é baseada na extensão de um *primer* na presença de todos os quatro dideoxynucleotídeos, sendo o local do sítio polimórfico identificado por diferenças de massas, pela adição de nucleotídeos ao *primer*. Uma modificação da metodologia PINPOINT denominada NUDGE (*Nucleotide Depletion Genotyping*) foi descrita por BLONDAL *et al.* (2003), e tem como característica principal, a utilização para a reação de mini-sequenciamento de uma mistura de dideoxynucleotídeos faltando um dos nucleotídeos do sítio polimórfico. Desta forma durante a extensão do *primer* a enzima DNA polimerase se dissocia do molde de DNA na posição onde o nucleotídeo deletado for requerido, o que possibilita a determinação do haplótipo. O protocolo desta metodologia simplifica e reduz consideravelmente os custos da análise. Várias plataformas de análise podem, então, ser utilizadas para a observação dos resultados obtidos com o mini-sequenciamento, incluindo a espectrometria de massas MALDI-TOF (*Matrix Assisted Laser Desorption Ionization Time of Flight*). Entre as principais vantagens dessa metodologia estão a acurácia, a possibilidade de se avaliar múltiplas amostras simultaneamente e a redução dos custos das análises.

Diante desse grande potencial dos marcadores SNPs para análises genéticas, esforços têm sido feitos no desenvolvimento de protocolos que facilitem a identificação de tais marcadores. Na genotipagem podem ser utilizadas as metodologias acima citadas e outras descritas na literatura. Elas devem ser escolhidas de acordo com custo, precisão, agilidade, disponibilidade de equipamentos, dificuldade de desenvolvimento dos ensaios e potencial para detecções de *multiplex* (RAFALSKY, 2002)

3. OBJETIVOS

3.1. Objetivo geral

Confirmação e identificação de polimorfismos presentes nas seqüências dos genes que codificam lipoxigenases 2 e 3, entre as seqüências de DNA de variedades de soja que apresentam o fenótipo LOX+ e variedades com o fenótipo LOX- e posterior validação desses polimorfismos como marcadores moleculares.

3.2. Objetivos específicos

- Confirmar, por meio de seqüenciamento de DNA das variedades estudadas, se estas continham alguns dos três polimorfismos na seqüência intrônica da LOX 2, anteriormente descritos por KIM *et al.* (2004);
- Desenhar *primers* para amplificação do gene *Lox3* de soja;
- Amplificar e seqüenciar todos os fragmentos referentes às seqüências dos genes *Lox2* e *Lox3* de soja;
- Comparar esses seqüenciamentos para todas as variedades estudadas com a finalidade de identificar polimorfismos;
- Caracterizar fenotipicamente uma população contrastante para lipoxigenases utilizando atividade enzimática;
- Genotipar essa população contrastante, com a finalidade de associar polimorfismos encontrados para *Lox2* e para *Lox3*, com os dados fenotípicos.

4. MATERIAIS E MÉTODOS

4.1. Seleção e plantio dos materiais genéticos

Foram utilizados seis genótipos de soja para o estudo das sequências correspondentes às LOX 2 e 3: Cristalina (FT sementes), CS94731 (COOPADAP), UFV 16 (UFV), CAC 1 (COOPADAP) e Hartwig (americana) todas contendo os alelos dominantes para as três enzimas LOX, sendo por isso denominadas “normais” e a isolinha Cristalina Triplo Nula (derivada da variedade Cristalina por retrocruzamentos) contendo os três alelos recessivos para as três LOX de sementes (Tabela 2).

Tabela 2 - Genótipos utilizadas para o seqüenciamento com os respectivos fenótipos.

Variedade	Fenótipo
Cristalina Triplo Nula	LOX 1-, LOX 2- e LOX 3-
Cristalina Normal	LOX 1+, LOX 2+ e LOX 3+
CS94731	LOX 1+, LOX 2+ e LOX 3+
CAC 1	LOX 1+, LOX 2+ e LOX 3+
UFV 16	LOX 1+, LOX 2+ e LOX 3+
Hartwig	LOX 1+, LOX 2+ e LOX 3+

As cinco variedades “normais” foram selecionadas por terem origem genéticas distintas, ou seja, que não apresentavam progenitores em comum. Uma única variedade “triplo nula” foi usada, pois a origem genética dessa recessividade, no programa de melhoramento genético da qualidade da soja em desenvolvimento no BIOAGRO/UFV, é única. Os fenótipos foram confirmados antes do plantio, por meio de testes colorimétricos descritos por SUDA *et al.* (1995).

Todos os plantios foram realizados em casa de vegetação com sistema de controle de temperatura no setor de estufas da UFV. As sementes foram semeadas em vasos contendo três litros de solo previamente adubado. As plantas foram cultivadas entre o período de maio a agosto de 2007.

4.2. População para a genotipagem

Para validação dos SNPs relacionados à manifestação da característica em estudo, foram utilizadas amostras de DNA de folhas de populações de RILs (F_8), originadas do cruzamento de genótipos contrastantes para LOX que foram: Hartwig (variedade americana com os três alelos dominantes para as lipoxigenases) e a variedade Paranagoiana Triplo Nula (isolinha derivada da variedade Paranagoiana por retrocruzamentos com os três alelos recessivos para as lipoxigenases). Essa última variedade como comentado tem a mesma origem genética da variedade Cristalina Triplo Nula. Para a genotipagem de cada gene foram utilizados grupos diferentes de indivíduos dessa mesma população

4.3. Seqüências de DNA estudadas

4.3.1. Lipoxigenase 2

Uma seqüência de 804 nucleotídeos correspondentes a uma região intrônica do gene da lipoxigenase 2, anteriormente descrita por KIM *et al.* (2004), foi usada como base para o estudo dessa mesma região nas variedades analisadas neste trabalho.

4.3.2. Lipoxigenase 3

A seqüência de 5709 nucleotídeos correspondente ao gene da lipoxigenase 3, disponível no *Genbank* (*National Center for Biotechnology Information* – NCBI) com o número de acesso X06928.1, foi utilizada como seqüência base para o estudo desse loco nas variedades selecionadas. Essa seqüência contém: uma região 3' transcrita e não traduzida (UTR- *untranslated region*), a região promotora, 8 íntrons, 9 éxons, e a região 5' UTR (Figura 3).



Figura 3 – Esquema ilustrativo dos 5709 pb correspondentes ao gene da lipoxigenase 3, contendo em azul a região 5' UTR + região promotora no início e a região 3' UTR no final; em amarelo 9 éxons; e em vermelho 8 íntrons.

4.4. Desenho dos *primers* específicos

4.4.1. Lipoxigenase 2

De acordo com KIM *et al.* (2004), um fragmento de 804 pb correspondentes à uma região intrônica do gene que codifica a lipoxigenase 2, contém três SNPs que podem ser usados como marcadores para a presença/ausência dessa enzima em soja. Dessa forma o par de *primers* descritos por esse autor (Tabela 3) foram também utilizados para amplificar essa mesma região no material genético das variedades selecionadas, com o intuito de confirmar a eventual presença desses polimorfismos e a conseqüente utilização desses como marcadores SNP para essa característica em soja. Esses *primers* serão referidos ao longo do trabalho como primer *Lox2*.

Tabela 3 - Par de *primer* utilizado para a amplificação de um fragmento da lipoxigenase 2 de soja.

<i>Primer</i>	Direção	Seqüência 5'→3'	Tamanho do <i>Amplicon</i> (pb)
lox2	R	TTGCAAACAAAGCGAATGGTT	804
	F	GGTGTCGGGAATCCTGAACA	

¹F, *forward primer*; R, *reverse primer*

4.4.2. Lipoxigenase 3

O desenho dos *primers* para o gene *Lox3* foram feitos com base no acesso *GenBank*: X06928.1. Para isso foi utilizado o *software primer 3* (ROZEN & SKALETSKY, 2000) *versão 0.4.0*.

Parâmetros como tamanho entre 25 e 30 nucleotídeos e uma temperatura média de anelamento de 55°C foram utilizados para todos pares de *primers*. Foram necessários 14 pares de *primers* que amplificam fragmentos entre 500 a 600 pb, para que os 5709 pares de bases do gene *Lox3* da soja, fossem abrangidos. Durante o desenho dos *primers* foram consideradas regiões de sobreposição entre esses fragmentos. Todos os pares de *primers* foram utilizados para amplificar a região alvo de cada um deles no DNA genômico de todos os seis genótipos de soja selecionados.

4.5. Extração de DNA total de folhas de soja

O DNA foi extraído a partir de coletas individuais de tecido foliar de cinco plantas de cada cultivar selecionado. A extração foi feita utilizando o kit comercial Wizard® Genomic DNA Purification com o número de catálogo A1125, de acordo com as especificações desse fabricante. O rendimento das extrações foi de 200 ng/μL de DNA com altíssima pureza, verificados e quantificados utilizando géis de agarose 0.8 % corados com brometo de etídio e marcadores de peso obtidos a partir da clivagem do DNA do fago λ clivados com específicas enzimas de restrição. Após essa quantificação, as amostras de DNA foram diluídas para uma concentração final de 10 ng/μL.

4.6. Amplificação e purificação dos fragmentos

Os fragmentos dos genes tanto para o *primer* que amplifica a região estudada do gene *Lox2* quanto para os 14 *primers* que amplificam as regiões do gene *Lox3* foram isolados via PCR, utilizando o DNA total das variedades de soja anteriormente selecionadas. Todas as reações de amplificação continham 3 μL de DNA; 3 μL do tampão de reação 5X da Phoneutria®; 0,6 μL da mistura

dos 4 dNTPs numa concentração estoque de 20 mM; 0,5 µL de cada *primer* na concentração estoque de 5 µM e 1 U de Taq DNA Polimerase Phoeutria[®], adicionando água para completar os 15 µL finais. Todas as reações de amplificação foram conduzidas no termociclador *Mastercycler ep Gradient*, modelo 5341 (Eppendorf).

4.6.1. Lipoxigenase 2

Foi utilizada uma etapa inicial de desnaturação a 94°C por 5 min, seguida por 35 ciclos (94°C por 30 s; 55°C por 45 s; e 72°C por 90 s) e um período adicional de polimerização a 72°C, por 10 min. Os produtos de amplificação foram analisados por eletroforese em gel de agarose 0,8%, contendo 0,2 µg/mL de brometo de etídeo, juntamente com um marcador de tamanho (*GeneRuler 100 bp DNA Ladder Plus* - Fermentas[®]), para confirmação da amplificação única e do tamanho esperado para cada fragmento. As bandas do DNA foram visualizadas sob luz ultravioleta e fotodocumentado.

Os fragmentos obtidos foram purificados com o kit *Wizard SV Gel and PCR Clean-Up System* (Promega) de acordo com as recomendações dos fabricantes, para serem posteriormente enviadas para sequenciamento.

4.6.2. Lipoxigenase 3

De forma semelhante à LOX 2, foi realizada para amplificação com os 14 pares de *primers* para *Lox3*, uma etapa inicial de desnaturação a 94°C por 4 min, seguida por 40 ciclos (94°C por 45 s; 55°C por 45 s; e 72°C por 1 min) e um período adicional de polimerização a 72°C, por 4 min.

Todos os produtos de amplificação foram analisados em gel de agarose 0,8% corado com brometo de etídeo juntamente com um marcador de tamanho (*GeneRuler 100 bp DNA Ladder Plus* - Fermentas[®]), para confirmação da amplificação única e do tamanho esperado para cada fragmento.

Após a confirmação da amplificação, os fragmentos de DNA foram purificados com o kit *Wizard SV Gel and PCR Clean-Up System* (Promega), de acordo com as recomendações do fabricante, e em seguida enviadas para o sequenciamento direto.

4.7. Sequenciamento automático

As seqüências de nucleotídeos foram obtidas, através de produtos de PCR, de forma direta numa concentração aproximada de 40ng/ μ l de reação para cada um dos 15 pares de *primers* (um para *Lox2* e 14 para *Lox3*). As reações de seqüenciamentos foram feitas utilizando ambos os *primers* R e F, devido ao caráter investigatório das subseqüentes análises. As reações de seqüenciamento que não geraram bons resultados foram repetidas. Os fragmentos obtidos de cada uma das seis variedades foram seqüenciados cinco vezes (devido a coleta do material para extração de DNA ter sido feita separadamente para cinco plantas de cada variedade).

Foram utilizados nestas reações 100 a 150 ng de DNA, 0,5 μ M de cada *primer* R e F, e 2,0 μ L de *DYEnamic™ ET Terminators Sequencing Premix* em volume final de 5 μ L. As condições de PCR foram 95°C por 10 s, 50°C por 5 s e 60°C por 2 min repetidas 35 vezes. Em seguida, o DNA foi precipitado adicionando-se 27,5 μ L de etanol absoluto e acetato de amônio para uma concentração final de 0,75 M. Após 10 minutos à temperatura ambiente, as reações foram centrifugadas por 45 min a 3.500 g. O DNA então foi lavado com 100 μ L de etanol 70%, centrifugado novamente a 3.500 g por 10 min e secado por 10 min. Após a precipitação, o DNA foi ressuspenso em 5 μ L de *Loading Buffer* (GE Healthcare).

O seqüenciamento dos fragmentos foi realizado em sequenciador automático *MegaBACE™ 500* de 48 capilares (GE Healthcare). As reações de seqüenciamento foram baseadas na técnica de seqüenciamento por terminação de cadeia por didesoxinucleotídeos (ddNTPs), descrita por SANGER *et al.* (1977).

4.8. Análise, comparação das seqüências e identificação de polimorfismos

A detecção de polimorfismos envolve o conhecimento da seqüência de DNA (após o seqüenciamento) para um mesmo loco de vários indivíduos dentro de uma determinada população, permitindo a observação de posições específicas que possam apresentar variações as quais caracterizam os diferentes alelos.

Dessa forma as seqüências obtidas foram analisadas com o auxílio do programa *Sequencher*® versão 4.1.4 (Genes Code Corporation) e a comparação feita por similaridades com as seqüências bases de *Lox2* e *Lox3* para verificação e posicionamento dos SNPs.

A seqüência base para LOX 2 usada para comparação das seqüências foi a mesma descrita por KIM *et al.* (2004) e a seqüência depositada no GenBank para o gene *Lox2* com o número de acesso J03211. Para LOX 3 foi utilizada a seqüência do gene e da seqüência codificante (Cds completa), ambas depositadas no *GenBank* com os respectivos números de acesso X06928.1 e AAB4127 e proteína depositada no PDB, com a finalidade de garantir a correta identificação de cada seqüência. As seqüências foram alinhadas inicialmente par a par (Seqüência R com a F de cada *primer*) analisadas individualmente para checar a qualidade dos dados brutos. Posteriormente as seqüências foram reagrupadas de acordo com cada variedade, em busca de polimorfismos intravarietais e por último as seqüências de todas as variedades foram comparadas em conjunto em busca dos polimorfismos intervarietais. Todos polimorfismos foram identificados utilizando o mesmo programa acima citado, por meio de minuciosa inspeção visual.

4.9. Avaliação fenotípica da população para a genotipagem

A caracterização fenotípica da população RILs foi feita a partir da determinação da atividade enzimática das lipoxigenases 1 e 3. As atividades foram determinadas espectrofotometricamente de acordo com OLIVEIRA *et al.* (1998). Para isto, foram pesados cerca de 60 mg de semente, maceradas com 1800 µL de tampão contendo Tris-HCl 60 mM, pH 8,2, CaCl₂ 15 mM e sacarose 13%, em gral previamente resfriado para manter a atividade das lipoxigenases. O homogenato foi centrifugado a 15,7 g por 20 minutos a 4°C. O sobrenadante foi utilizado para realização da determinação da atividade enzimática. A determinação da atividade dessas enzimas foi realizada medindo-se a absorvância, após 2 minutos de reação, a 234 e 280 nm para LOX 1 e 3, respectivamente. Para LOX 1, as amostras eram constituídas de 1 mL de tampão borato de sódio 0,1 M, pH 9,5, 6 µL de linoleato de sódio 10 mM e 2,5 µL do extrato enzimático. Para LOX 3, as amostras eram constituídas de

1 mL de tampão fosfato de sódio 0,2 M, pH 6,8, 35 µL de linoleato de sódio 10 mM e 15 µL do extrato enzimático. A partir da absorvância obtida em 234 e 280 nm para as amostras, foram determinados os fenótipos presença e ausência de cada enzima nos extratos, por meio da comparação daquelas absorvâncias geradas por cada indivíduo da população de RILs, com os valores a priori determinados das variedades para LOX 1+ e LOX 3 + e para LOX 1- e LOX 3 - para os mesmos comprimentos de onda.

4.10. Genotipagem da população contrastante

Para a validação dos polimorfismos como possíveis marcadores para os genes *Lox2* e *Lox3*, foi utilizada como ferramenta de genotipagem o próprio seqüenciamento direto dos indivíduos da população de RILs, considerando os locos de interesse (que contêm os polimorfismos) para cada lipoxigenase.

A validação dos marcadores foi feita através da correlação direta entre os dados fenotípicos advindo das atividades enzimáticas com os dados do seqüenciamento.

4.10.1. Lipoxigenase 2

Amostras de DNA de 53 RILs foram analisadas utilizando o mesmo *primer* descrito por KIM *et al.* (2004), usado anteriormente para o seqüenciamento das variedades e enviadas para seqüenciamento. O seqüenciamento foi feito apenas na direção direta - F (*Forward*), pois o caráter da análise foi apenas confirmatório.

4.10.2. Lipoxigenase 3

Cinquenta e uma RILs foram amostrados (indivíduos diferentes do tópico anterior), os respectivos materiais genéticos foram amplificados com o *primer* *lox3.2* referente ao primeiro éxon que continha o polimorfismo de interesse. Os produtos de PCR foram enviados para o seqüenciamento apenas na direção direta - F (*Forward*).

Para os dois locos em estudo, o seqüenciamento e a análise das seqüências foram conduzidos da mesma forma descrita para as variedades.

5. RESULTADOS E DISCUSSÃO

5.1. Amplificação de fragmentos dos genes *Lox* por PCR

Para a obtenção dos fragmentos dos genes de interesse, foram feitas reações de amplificação por PCR, utilizando DNA total de folhas de soja e os *primers* específicos correspondentes para cada gene.

5.1.1. Lipoxigenase 2

O par de *primers* usado por KIM *et al.* (2004), que amplifica no genoma da soja um fragmento de 804 pb (correspondente a uma região intrônica do gene *Lox2*), foi utilizado no presente trabalho com a finalidade de confirmar a presença de um ou mais polimorfismos descritos naquele artigo, nas variedades em estudo do Programa de Melhoramento da Qualidade da Soja do BIOAGRO-UFV. O par de *primers* originou um fragmento de tamanho esperado, não havendo amplificação inespecífica (Figura 4), o que é pré-requisito essencial para a qualidade do seqüenciamento.

Uma vez confirmadas as amplificações para todas as variedades, utilizando gel de agarose, os produtos de PCR foram purificados e, em seguida, seqüenciados.

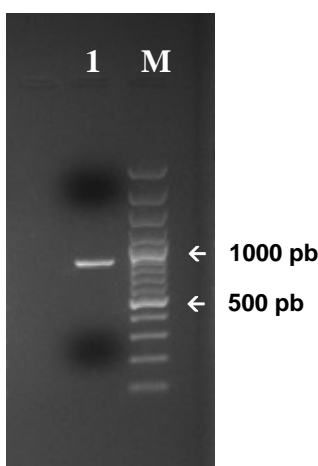


Figura 4. Análise eletroforética em gel de agarose (1,2%) do produto de amplificação do *primer lox 2* utilizando material genético da variedade Cristalina Normal. Canaleta 1 - Fragmento amplificado a partir do par de *primers* para *Lox 2*; Canaleta M – Marcador de tamanho (*GeneRuler 100 bp DNA Ladder Plus* – Fermentas).

5.1.2. Lipoxigenase 3

Os fragmentos de PCR referentes à sequência do gene *Lox3* foram obtidos a partir da amplificação do DNA genômico das variedades de soja estudadas com os 14 pares de *primers* desenhados para abranger a sequência de *Lox3* depositada no banco de dados *GenBank*, com número de acesso X06928.1, cuja sequência nucleotídica completa se encontra no anexo (Figura 1). Após as amplificações os tamanhos dos fragmentos foram analisados por eletroforese em gel de agarose (Figura 5).

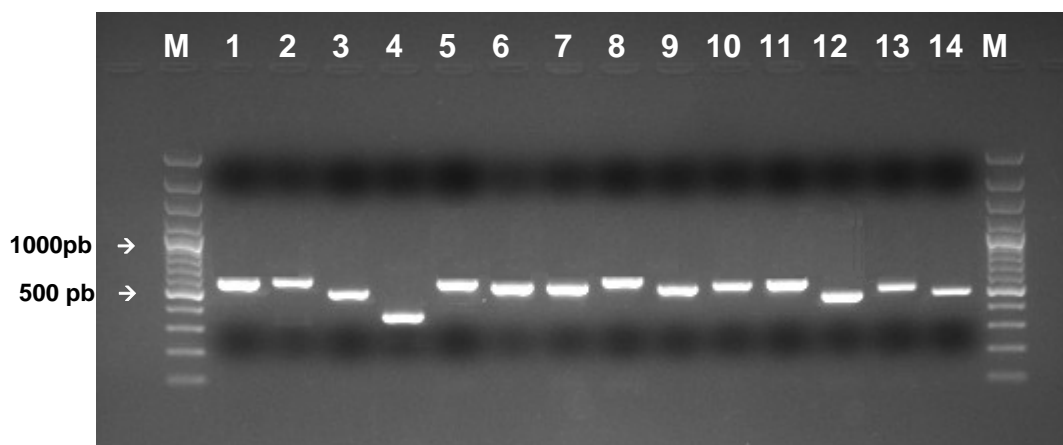


Figura 5. Análise eletroforética em gel de agarose (1,2%) dos produtos de amplificação dos 14 pares de *primers* para *Lox 3*. Canaletas de 1 a 14 contêm os respectivos fragmentos amplificados a partir dos pares de *primers* desenhados para o gene *lox3*, sendo 1 - 566 pb; 2 - 571 pb; 3 - 506 pb; 4 - 365 pb; 5 - 575 pb; 6 - 549 pb; 7 - 544 pb; 8 - 589 pb; 9 - 529 pb; 10 - 550 pb; 11 - 564 pb; 12 - 478 pb; 13 - 530; 14 - 493 pb para as amplificações foram utilizados material genético da variedade Cristalina Normal. Canaleta M – Marcador de tamanho (*GeneRuler 100 bp DNA Ladder Plus* – Fermentas).

O tamanho dos fragmentos amplificados conferiram com o esperado (Tabela 4) e todo o gene foi corretamente abrangido e não houve formação de bandas inespecíficas (Figura 6). Após a amplificação e purificação desses fragmentos para cada variedade estudada, as amostras foram seqüenciadas.

Tabela 4 - Pares de *primers* utilizados para a amplificação de fragmentos do gene da lipoxigenase 3 de soja.

Primer	Direção ¹	Sequencia (5'→3')	Tamanho do <i>amplicon</i> (pb)
lox3.1	R	GTTACTATTGAGTACATTTGTGAATTGGAT	566
	F	ATTATGAATTGAGAATTAACCTTTTAAACG	
lox3.2	R	TGGATGAAGTAATTAGGAAAAATAAGAAAG	571
	F	TTATAAAAAGAGAGTAAGAAAATGTTGTGG	
lox3.3	R	CAATGAAGTAATGATACCTTCCAAAAAG	506
	F	TTCCTTTTCTTCTTTCTTATTTTTCCCTAAT	
lox3.4	R	CTCATATAACAAATAATGAAGACACACAAG	367
	F	AATTATTGGTCAAGGTCTCGACTTAGTT	
lox3.5	R	CAACTCTTCTTCTCTATTTGACTAGTGG	575
	F	GTTTTTACTTTTGTTTTCACTCTTTCTTTT	
lox3.6	R	ATTTTCAAACCTCATAATTGAAAAACAGTTA	549
	F	AAAAATAAAGTTAACAAAACATATTGCAGA	
lox3.7	R	TAGTGGACTAATCTTGCTGATTATATCTGT	544
	F	TACTCTATCACAGACTCAAATCTGTTCTTC	
lox3.8	R	TATATACCTCATCTACAGTGAGCCCTTCTA	589
	F	GTGAGTTTGATAGCTTTGATGAAGTT	
lox3.9	R	ATATTATTTGTTGAGAAAATGTATACCAA	529
	F	GATCATACTAGTCAAATAACCAAAGAACAC	
lox3.10	R	TCACATGACATTAATTATATAACATCAAAG	550
	F	CAGATGAAGGTGTTGAAAGTTCTATTT	
lox3.11	R	CTGAACTTTGGTGTAAATAGTCTTCAAGTA	564
	F	ATATCTTACAATAGGAAACATGAAACTCAA	
lox3.12	R	AAATTCAGATTCCTTCTTCTATCTTTTA	478
	F	TTTAATCTTAAACCGTCCAACCTTAGTAG	
lox3.13	R	TATAATAACGATATTTGGGTAGTTAGTTGC	530
	F	CTTCTAGTAAGGAAGGATTAACCTTTCAGAG	
lox3.14	R	ACTATTCAAGATGTATTGCAAATTCAAG	493
	F	ATAACCCATTGTTCAAATATAATCCGTAAT	

¹F, *forward primer*; R, *reverse primer*

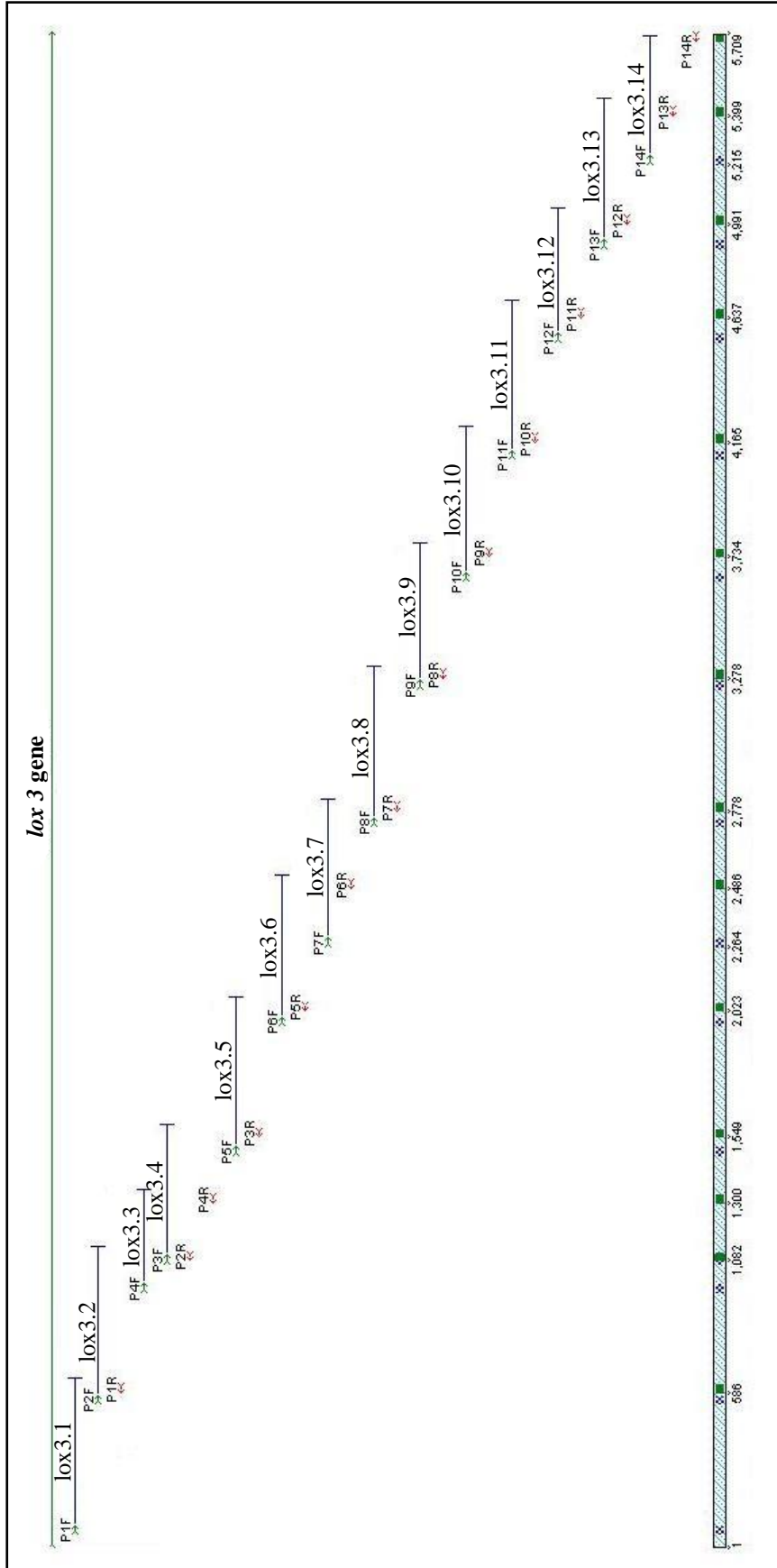


Figura 6 : Esquema de anelamento dos 14 pares de primers utilizados para amplificação do gene *lox3*.

5.2. Identificação de polimorfismos

Como já visto, no total 30 amostras de DNA foram utilizadas e distribuídas da seguinte forma: 5 amostras da variedade cristalina normal; 5 da CAC 1; 5 da UFV 16; 5 da CS94731, 5 da Hartwig e 5 da Cristalina TN. Cada um dos fragmentos de PCR foram seqüenciados 5 vezes para cada variedade, ou seja, 1 vez por indivíduo.

Para cada uma destas amostras foi seqüenciada a região de 804 pb da *Lox 2* e os 14 fragmentos (um para cada par de *primers*) resultantes da amplificação de regiões genômicas correspondentes ao gene *Lox3*.

Todas as seqüências obtidas foram analisadas no programa *Sequencher*TM versão 4.1.4 (Genes Code Corporation) e comparadas com as seqüências utilizadas para o desenho dos *primers* ou foram analisadas com o auxílio do BLAST (ALTSHUL *et al.*, 1997), confirmando a identidade e a localização espacial dos fragmentos nas seqüências depositadas. A identificação dos polimorfismos foi feita por meio de cuidadosa análise visual.

Uma das formas de identificação de polimorfismos no DNA requer a obtenção de uma seqüência base (da mesma espécie em estudo) referente ao loco de interesse, e assim realizar o ressequenciamento em outras variedades para assim encontrar pares de bases polimórficos (SOLEIMANI *et al.*, 2003). Essa foi a forma através da qual os polimorfismos contidos entre as variedades do PMQS-BIOAGRO foram identificados.

Marcadores SNPs vêm sendo utilizados para saturar o mapa de ligação de importantes genes da soja, além de auxiliarem na clonagem posicional e no melhoramento assistido por marcadores (HA & BOERMA, 2008).

SNPs e indels são grandes fontes de marcadores polimórficos para serem usados no aumento da resolução dos mapas genéticos, estudos de associação (RAFALSKI, 2002) e, freqüentemente, esses polimorfismos são opções para o desenvolvimento de marcadores, muito próximo ou no interior do gene de interesse (GUIMARÃES *et al.*, 2007).

5.2.1. Lipoxigenase 2

Dos três polimorfismos descritos por KIM *et al.* (2004), dois foram confirmados após seqüenciamento e comparação entre as seqüências das variedades selecionadas neste trabalho, nas mesmas posições, ou seja, T/C (251) e o G/A (367) (Figura 7). Os cromatogramas indicativos desses polimorfismos estão no anexo (Figura 2).

```
GGTGTCTCGGGAATCCTGAAACAGAGGAGGAGGGCATAAGATAAAAAGGGACAGTGGTGTGAT 60
GAGGAAGAATGTGTTGGACTTCAACAGCGTGGCTGATCTTACTAAAGGAAATGTTGGGGG 120
ACTCATAGGCACCGGCTCAACGTTGTTGGCTCAACACTTGATAACCTCACTGCTTTCTT 180
GGGCCGAAGTGTCTGCCCTACAGCTCATTAGTGTACCAAACCTCTTGGTTCATTTCTTCT 240
TCCTTCCACAT-AATCAATAACTTCTATATTCAACAATAAAGTGTTTAATCTCTATACTCTC 300
ATTCATTTCAATCAATGAAAAAAAAAATCATAAGACTTTTAACTAAAATTAACCTATGTAA 360
AGAATCG-CAAACAAAAAACTATATAATATTAAAGTTTATTTACTTTTTTTTATAATGACA 420
AAAAAATTATTGTATATGGTGCACAAATTTTTGTACTCTTTAAAAATATATCACTTTAT 480
ACATAGCCAAACATATTTATTTTGTATAGTATTAACCTTATTTGGGTACGTACCTTAATAA 540
TATTATTATGTGTGTATGTATGGTCTGTTTGTAGCAAATGGAAAAGGAAAAGTTGGAAAG 600
GGATACGTTCTTGAAGGGATTATTGTGTCTGTTACCAACTTTGGGAGCAGGGGAGTCTGC 660
ATTC AATATT CAGTTT GAATGGGACGAAAGCATGGGAATCCCCGGTGCGTTTTACATAAA 720
GAACTACATGCAAGTTGAGTTTTACCTCAAGAGTTTGACTCTTGAAGACGTTCCAAACCA 780
AGGAACACATTTCGCTTTGTTTGCAA 804
```

Figura 7 – Sequência intrônica do gene *Lox2*, com os dois SNPs existentes entre as variedades normais (nucleotídeo acima) e triplo nulas (nucleotídeo baixo), destacados em vermelho. As setas indicam o local de anelamento dos primers *lox2*.

Segundo KIM *et al.* (2004), três polimorfismos de um único nucleotídeo foram identificados após o seqüenciamento de um fragmento de 804 pb correspondentes a uma região intrônica do gene responsável por codificar a lipoxigenase 2 em soja. As duas variedades eram contrastantes para a característica, sendo uma LOX 2+ (Pureunkong) e a outra LOX 2- (Jimpumkong 2). Os polimorfismos presentes nessa seqüência são duas transições e uma transversão, sendo o primeiro alelo da variedade LOX 2+ e o segundo LOX 2- nessa ordem: T/C, G/A e C/A (anexo - Figura 3). No fragmento de 804 pb estudado por este autor, os polimorfismos encontrados se situam nas posições 251, 367 e 420, respectivamente. No estudo desenvolvido por

KIM *et al.* 2004, somente o primeiro polimorfismo foi utilizado para genotipagem de 90 RILs contrastantes, obtendo 100% de concordância entre o fenótipo e o genótipo.

5.2.2. Lipoxigenase 3

Dentre todos os SNPs identificados para quaisquer comparações entre as diferentes seqüências obtidas, foi observada uma freqüência relativamente alta de SNPs do tipo transição na região codificadora. Esse tipo de resultado foi observado também em estudos prévios de identificação de SNPs envolvendo outras culturas (GUPTA *et al.*, 2001; PÓVOA 2005).

5.2.2.1. Polimorfismos Totais

Ao comparar a seqüência base para *Lox3* do banco de dados (Williams 82), usada para o desenho dos *primers*, com todas as regiões seqüenciadas das variedades do PMQS-BIOAGRO, foi possível ter uma visão geral do número de polimorfismos em cada região do gene, bem como a freqüência dos mesmos. As regiões do gene *Lox3* foram descritas por YENOFKY *et al.* (1988) contendo, além das regiões 5' e 3' UTR, 9 éxons e 8 íntrons (Tabela 5).

Tabela 5. Número de polimorfismos para cada região do gene *Lox 3* obtidas nesse trabalho comparadas com a seqüência de *Lox 3* depositada no *GenBank* (número de acesso: X06928) utilizada para o desenho dos *primers*.

Região do gene <i>Lox3</i>	Total de pb compreendidos	Número de polimorfismos	
		SNPs	indels
5' UTR e região promotora	815	11	1
9 Éxons	2767	11	1
8 Íntrons	1519	3	2
3' UTR	608	94	33
Total	5610	118	37

No estudo conjunto das seqüências das seis variedades (Cristalina, CAC 1, UFV 16, CS94731, Hartwig e Cristalina TN), foi detectado um total de 156 polimorfismos identificados nos 5610 pb abrangidos pelo seqüenciamento, caracterizando uma média de 1 polimorfismo a cada 36 pb. Em genes de milho,

esta frequência diminui para cerca de 1 polimorfismo a cada 60,8 pb em linhagens elite (CHING, 2002) e, em trigo, a densidade de polimorfismos é de 1 a cada 370 pb (PROCUNIER *et al.*, 2003).

Se forem considerados apenas os polimorfismos da região codificadora serão 12 polimorfismos em 2767 pb, o que corresponde a uma frequência de polimorfismos de 1 a cada 230,5 pb. RAFALSKI (2002) descreveu que, em um grupo de 502 ESTs de milho, foi encontrada uma frequência de 1 polimorfismo a cada 130 pb para a região codificante de linhagens cultivadas.

De um total de 119 SNPs, 11 ocorrem na 5' UTR, 11 em éxons, 3 em íntrons e 94 na 3'UTR. E de forma semelhante à PÓVOA (2005) que num estudo das bases moleculares da qualidade da madeira de eucalipto, ao sequenciar o gene *cad2* que codifica para a enzima cinamil-álcool desidrogenase observou que o número de transições se concentraram mais nos éxons enquanto que as transversões foram predominantes no 3' UTR.

Dos SNPs, 48 foram transições e 70 foram transversões, constituindo, respectivamente, 41% e 59% das substituições de bases. ZHU *et al.* (2003), em um estudo de ESTs totalizando 9,3 Kpb, identificaram 233 SNPs, sendo 112 transversões (48%) e 121 transições (52%). Essa proporção foi bastante próxima dos resultados obtidos no presente trabalho. Em humanos, diferentemente, a razão entre transições e transversões é 2:1 (WANG *et al.*, 1998). O número de transições normalmente é maior porque uma transversão gera uma distorção de largura na molécula de DNA e a probabilidade do sistema de reparo da molécula atuar neste tipo de mutação é muito mais rara do que no caso de uma transição (LEWIN, 2000).

Os elementos cis-regulatórios de um gene, segundo LEVINE & TIJAN, (2003) se dividem basicamente em três tipos: região promotora mínima, regiões ativadoras ou inibidoras da transcrição. Essas seqüências cis regulatórias, e as proteínas que com elas interagem, chamadas elementos trans-regulatórios (SCHREM *et al.*, 2002), são componentes fundamentais para a transcrição de um gene.

O CAAT *box* e o TATA *box*, componentes do promotor mínimo, contidos no gene *Lox3* estão posicionados entre os nucleotídeos 732 e 786 (YENOFKY *et al.*, 1988) considerando o início da seqüência depositada como

+1 . Os dois últimos SNPs localizados na região 5' UTR são uma transversão (G/C) na posição 514 e uma transição (A/G) na 796. Assim, esses polimorfismos não estão contidos nem nas posições específicas dos cis-elementos, nem entre elas, possivelmente afetando pouco o reconhecimento dos promotores pelos fatores de transcrição semente-específicos. Dessa forma, o fator que determina a recessividade para LOX 3 possivelmente não se encontra no promotor. Por outro lado, estudos como os de HOOGENDOORN *et al.* (2003) e MASOTTI *et al.* (2005), descrevem uma grande quantidade de promotores em genes humanos com SNPs capazes de alterar a atividade desses promotores.

Ao contrário do esperado, o valor bruto das substituições nos éxons foi maior do que aquele observado na região intrônica, mas isto se deve primariamente ao fato deste gene conter mais pares de base em regiões exônicas (2767 pb) do que em regiões intrônicas (1519 pb). Foram analisados praticamente o dobro de nucleotídeos em éxons em relação a íntrons, sendo a quantidade relativa de substituições em regiões intrônicas (1 a cada 304 pb) menor do que em regiões exônicas (1 a cada 230 pb).

Dos 11 polimorfismos que estão na região codificadora, um apenas é um indel localizado na posição 959 do gene *Lox3* e os demais ou são SNPs sinônimos que não acarretam mudança no aminoácido incorporado ou são polimorfismos não sinônimos conservativos, que mesmo levando à incorporação de outro aminoácido, mantém a atividade da enzima. Essa conservação da atividade enzimática é mantida porque ou um aminoácido de mesmas características físicoquímicas foi incorporado, ou a região da proteína onde ocorreram as variações aminoacídicas não são diretamente determinantes para a atividade e estabilidade dessa enzima.

A região 3' UTR demonstrou ser extremamente polimórfica. Resultados similares foram obtidos por PÓVOA (2005) no estudo dos polimorfismos presentes no gene *cad 2*, para o qual observou-se um padrão de polimorfismos mais freqüente na região 3' UTR (1 SNP a cada 15 pb) do que na região exônica (1 a cada 24 pb); no entanto, a região 3' UTR do gene *Lox3* descrita no presente trabalho mostrou-se ainda mais polimórfica do que a relatada no

trabalho de PÓVOA (2005). Todos esses polimorfismos presentes serão melhor abordados nos dois tópicos subseqüentes.

5.2.2.2. Polimorfismos entre variedades Normais

Todas as seqüências obtidas foram uma a uma comparadas com a seqüência de referência da lipoxigenase 3 (da variedade americana Williams 82) com a finalidade de descartar possíveis seqüências incorretas e conferir a qualidade das seqüências.

Este trabalho teve, inicialmente, como prioridade a escolha de variedades de origens genéticas distintas. No entanto, nenhuma variação foram observadas entre as seqüências de *Lox 3* das variedades Cristalina, CAC 1, UFV 16 e Hartwig. Pode ser observado um alto desequilíbrio de ligação entre os polimorfismos. Isso mostra uma alta similaridade genética entre as variedades, o que reforça a idéia de que as variedades comerciais de soja apresentam uma estreita base genética. Segundo HYTEN *et al.* (2006) a baixa freqüência de polimorfismos entre variedades comerciais de soja é devido a baixa diversidade genética contida entre os ancestrais selvagens da soja (*Glycine soja*).

Procedeu-se ao seqüenciamento de todos os 14 fragmentos de PCR para as cinco variedades normais (Cristalina, CAC 1, UFV 16, CS94731 e Hartwig) e para a variedade triplo nula (Cristalina TN). Após a obtenção das seqüências e cuidadosa análise visual dos cromatogramas no programa *Sequencher*[®], não foram observados polimorfismos intravarietais, confirmando a homogeneidade genética entre os indivíduos. De forma inesperada, os polimorfismos contidos nas seqüências da variedade normal CS94731 apresentaram muita semelhança aos contidos na variedade Cristalina TN, se mostrando diferentes apenas nas regiões UTR. No entanto não se entende a razão dessa semelhança uma vez que a origem genética de tais variedades é distinta. Um posterior estudo da diversidade genética entre essas variedades é necessário para esclarecer essa semelhança entre esses genótipos.

A seqüência de 857 aminoácidos correspondentes à lipoxigenase 3 está depositada no PDB com o número de acesso P09186. Essa seqüência aminoacídica também foi utilizada para obtenção da região codificante

denominada (*cds*), através da utilização do tBlastn, que utiliza uma sequência protéica para encontrar a sequência nucleotídica correspondente depositada no banco de dados. A *cds* completa para lipoxigenase 3 está contida no *GenBank* com o número de acesso AAB4127. Essas duas sequências também foram utilizadas para posicionar os polimorfismos.

Seis SNPs foram identificados na região codificadora, após serem comparadas as seqüências das variedades normais com a seqüência base (W82). Dentre as seis substituições nos éxons, as cinco primeiras são SNPs não sinônimos, pois determinam a incorporação de um aminoácido diferente na proteína; e a última é um SNP sinônimo, que determina a incorporação de um aminoácido igual (Tabela 6).

Tabela 6 - Comparação de SNPs contidos entre as seqüências codificadoras do gene *Lox3* das variedades normais (N) e a seqüência base (W82).

Posição na seqüência nucleotídica*	Posição na seqüência codificante*	Mudança nas bases		Mudança no códon		Mudança de Aminoácido**	
		W82	N	W82	N	W82	N
934	73	C	G	CAT	GAT	His	Asp
1031	169	C	T	CCC	TCC	Pro	Ser
1640	335	T	C	CTT	CCT	Leu	Pro
2104	601	G	A	GTC	ATC	Val	Ile
3466	1283	G	A	GGT	GAT	Gly	Asp
4085	1809	C	T	GAC	GAT	Asp	Asp

*Refere-se se a posição a partir da extremidade 5' da seqüência.

** Refere-se à seqüência a partir da extremidade aminoterminal.

De acordo com os dados publicados no PDB, as cinco primeiras variações encontradas na comparação entre as variedades normais Cristalina, CAC 1, UFV 16 e Hartwig com a seqüência base da variedade americana Williams 82, são variações naturais presentes entre as seqüências das lipoxigenases 3, ou seja, são polimorfismos não sinônimos, mas conservativos, pois apesar de levarem a incorporação de um aminoácido diferente, a enzima continua ativa. Em análise mais minuciosa dessa seqüência pôde ser observado que todas as modificações acontecem em regiões de mudança de estrutura secundária e muito provavelmente é por isso que a enzima continua ativa, uma vez que teoricamente esses aminoácidos estão localizados em regiões de transição de estrutura secundária, e que normalmente não estão diretamente associadas

com a estabilidade e/ou atividade da enzima. Para o último SNP, a mudança de base ocorreu na terceira base do códon, levando à substituição de um ácido aspártico por outro, contribuindo, dessa forma, para a manutenção da estrutura e, conseqüentemente, da atividade das enzimas independente dessa variação.

Dentre as 6 substituições em éxons, a primeira é uma transversão (C→G) e as outras 5 são transições (Figura 8). Isso confirma o que já foi comentado, de que existem mais transições do que transversões na região codificante.

As regiões 3' UTR apresentaram total identidade ao comparar as variedades Cristalina, CAC 1, UFV 16, Hartwig e a seqüência base, originada da variedade Williams 82 (Tabela 7). Segundo ALBERTS *et al.* (2002) a região 3' UTR geralmente tem um papel importante na estabilidade do mRNA, influenciando fortemente na expressão de alguns genes. Nesse sentido um estudo posterior de expressão gênica pode confirmar se essa semelhança se estende para o padrão de expressão para *Lox3*.

Tabela 7. Número de polimorfismos para cada região do gene *Lox3* se comparadas somente as seqüências normais obtidas nesse trabalho com a seqüência base *Lox3* utilizada para o desenho dos *primers*.

Região do gene	Número de polimorfismos	
	SNPs	Indels
5' UTR e promotor	5	0
Éxons Totais	8	0
Ítrons Totais	1	0
3' UTR	0	0
Total Geral	14	0

5.2.2.3. Polimorfismos entre variedades Normais X Triplo Nula

Considerando que para identificação de um SNP correlacionado com um dado fenótipo devem-se usar genótipos contrastantes para esse fenótipo, buscou-se identificar, no presente trabalho, os polimorfismos entre as variedades normais e a triplo nula que pudessem ser correlacionados com o fenótipo para LOX 2 e 3. Dessa forma ao comparar separadamente as seqüências do genótipos contrastantes, as variedades normais LOX 3+ (Cristalina, CAC 1, UFV 16, CS594731 e Hartwig) e a triplo nula (Cristalina

TN), foram encontrados 3 polimorfismos em região exônica sendo estes 1 indel e 2 SNPs (Tabela 8)

```

ATGCTTGGGGTCTTCTCCATAGGGGTCATAAGATAAAAAGGGACAGTGGTGTGATGCGC 60
AAGAATGTGTTGCACGTGAATAGCGTAACCAGCGTTGGGGGAATTATTTGGTCAAGGTCTC 120
GACTTAGTTGGGTCAACACTCGATACTCTTACTGCCTTCTTGGGCCGATCCGTGTCTCTC 180
CAGCTTATTAGTGCTACCAAAGCTGATGCCAACGGAAAAGGGAACTTGGAAAAGGCTACC 240
TTTTTGGAAAGGTATCATTACTTCATTGCCAACTTTGGGAGCAGGCCAATCTGCATTCAA 300
ATTAATTTTGAATGGGATGATGGGAGTGGAAATCTTGGAGCATTTTATATCAAGAATTTT 360
ATGCAAACCTGAGTTTTTCTTGTGAGTTTGTACTCTTGAAGACATTCCAAACCATGGAAGC 420
ATCCACTTTGTTTGC AATTTCGTGGATTACAATGCCAACTCTTCAAAAAGTGACCGCATT 480
TCTTTTGCCAACCAGACATATCTTCCAAGTGAGACACCAGCTCCACTAGTCAAATATAGA 540
GAAGAAGAGTTGCATAATTTAAGAGGAGATGGAACCTGGAGAACGCAAAGAGTGGGAAAGG 600
CTCTATGATTATGATGTCTACAATGATTTAGGTGATCCGGATAAAGGTGAAAATCATGCC 660
CGTCCTTGTTCTTGGAGGAAATGACACCTTTTCTTATCCTCGTAGGGGGAGAAACTGGTAGA 720
AAACCAACAAGGAAAGATCCTAATAGTGAGAGTAGGAGCAATGATGTTTATCTTCCAAGA 780
GATGAGGCTTTTGGACACTTGAAGTCATCTGACTTTCTTACTTATGGACTAAAATCCGTA 840
TCTCAAAATGTTCTTCCATTATTGCAATCTGCTTTTGATTTGAATTTACACCCCCTGAG 900
TTTGATAGCTTTGATGAAGTTCATGGACTCTATTAGGCGGAATTAAGCTGCCAACAGAT 960
ATAATCAGCAAGATTAGTCCACTACCCGTGCTTAAGGAAATCTTCCGAACTGATGGTGAA 1020
CAGGCCCTTAAGTTTCTCCTCCTAAAGTAATTC AAGTGAGTAAGTCTGCATGGATGACT 1080
GATGAAGAATTTGCAAGAGAAATGCTTGCTGGTGAAATCCAAACTTGATTCTGTTGCTTT 1140
AAGGATTTCCCTCCACGAAGCAAGCTAGATAGCCAAGTCTATGGTGTATCATACTAGTCAA 1200
ATAACCAAGAACACCTAGAGCCCACTTAGAAGGGCTCACTTGTAGATGAGGCAATTCAA 1260
AACAAGAGATTGTTCTACTAGTTCATTCATTGACCCCAATCATGCCATATTTGAGGCGAATA 1320
AATGCAACCTCCACAAAGGCTTATGCTACCAGAACCATCCTTTTCTGAAAAATGACGGA 1380
ACTTTAAGGCCACTTGCCATAGAGTTGAGTTTGCCACATCCTCAGGGAGATCAATCTGGT 1440
GCTTTTAGTCAAGTTTTTCTGCTGCAGATGAAGGTGTTGAAAGTCTATTTGGCTGCTA 1500
GCAAAGGCTTATGATGTTGTGAATGACTCGTGTCTATCATCAACTTGTGAGCCATTGGTTA 1560
AACACTCATGCAGTTGTTGAGCCATTATCATAGCAACAAACAGGCATCTCAGTGTGTTT 1620
CACCCATTTTATAAACTCCTTACCCCTCACTATCGTGACACCATGAACATAAATGGCCTT 1680
GCTCGGTTATCACTGGTCAACGACGGTGGCGTTATAGAACAAACATTTTGTGGGGAAGG 1740
TATTCGTGGAAATGTCTGCTGTAGTTTACAAGGATTGGGTTTTTACAGATCAAGCATTG 1800
CCTGCTGATCTTATAAAAAGAGGAATGGCAATTGAGGATCCATCGTGCCCTCATGGCATT 1860
CGCCTTGTGATAGAGGACTACCCTTTACTGTTGATGGACTTGAGATATGGGATGCTATC 1920
AAGACATGGGTCATGAATACGTTTTCTTGTACTACAAATCAGATGACACACTTAGAGAA 1980
GATCCTGAACTCCAAGCCTGCTGGAAAGAACCTCGTAGAGGTGGGTCTATGGAGACAAGAAA 2040
AATGAGCCATGGTGGCCTAAGATGCAAACTCGTGAAGAGCTAGTTGAAGCTTGGCTATC 2100
ATCATATGGACTGCTTCAAGCACTTCAATGCAGCTGTTAATTTTGGACAGTATCCCTATGGA 2160
GGTTTTAATCTTAAACCGTCCAACCTCTTAGTAGGCGATTATGCCTGAGAAAGGTTCTGCT 2220
GAGTATGAGGAGCTGAGGAAGAATCCCAGAAGGCTTACTTGAAGACTATTACACCAAAG 2280
TTTCAGACCCCTATTGACCTTTCTGTTATAGAAAATCTTGTCAAGGCATGCATCTGATGAG 2340
GTGTACCTTGGGGAGAGGGACAATCCAAATTGGACATCTGATACAAGAGCATTAGAGGCT 2400
TTTAAAAGGTTTGGAAATAAACTGGCACAAATGAGAATAAACTCTCAGAGAGAAACAAC 2460

```

Figura 8 – Sequência codificante da LOX 3, com os seis SNPs existentes entre todas as seqüências das variedades normais , destacados em vermelho.

Tabela 8. Número de polimorfismos para cada região do gene *Lox3* se comparadas somente as sequências obtidas nesse trabalho (Entre as variedades normais e a triplo nula)

Região do gene	Número de polimorfismos	
	SNPs	indels
5' UTR e promotor	6	1
Éxons Totais	2	1
Íntrons Totais	2	2
3' UTR	94	33
Total Geral	105	37

O indel foi identificado na posição 959 e constituiu da ausência de uma guanina nessa posição na variedade recessiva e na presença dessa base nas variedades normais (Figura 9).

Os SNPs identificados foram (nessa ordem: alelo da variedade triplo nula/normal): G/A na posição 3572 (correspondente ao 7º éxon) e A/G na posição 2391 (correspondente ao 9º éxon). Essas modificações de bases acarretaram a incorporação dos mesmos aminoácidos na proteína, sendo esses arginina e alanina, respectivamente. Esses polimorfismos ocorrem na última base do códon, sendo assim conservativos. Isso reforça a idéia de que o indel existente no primeiro éxon seja a razão molecular da recessividade para *Lox3*.

Uma vez que a janela de leitura, também denominada ORF (*Open Reading Frame*), contida no RNA, será mudada a partir dessa deleção no primeiro éxon, possivelmente a tradução da proteína da variedade recessiva será iniciada, mas logo que seu mau dobramento for detectado ela será devidamente sinalizada e direcionada para a degradação.

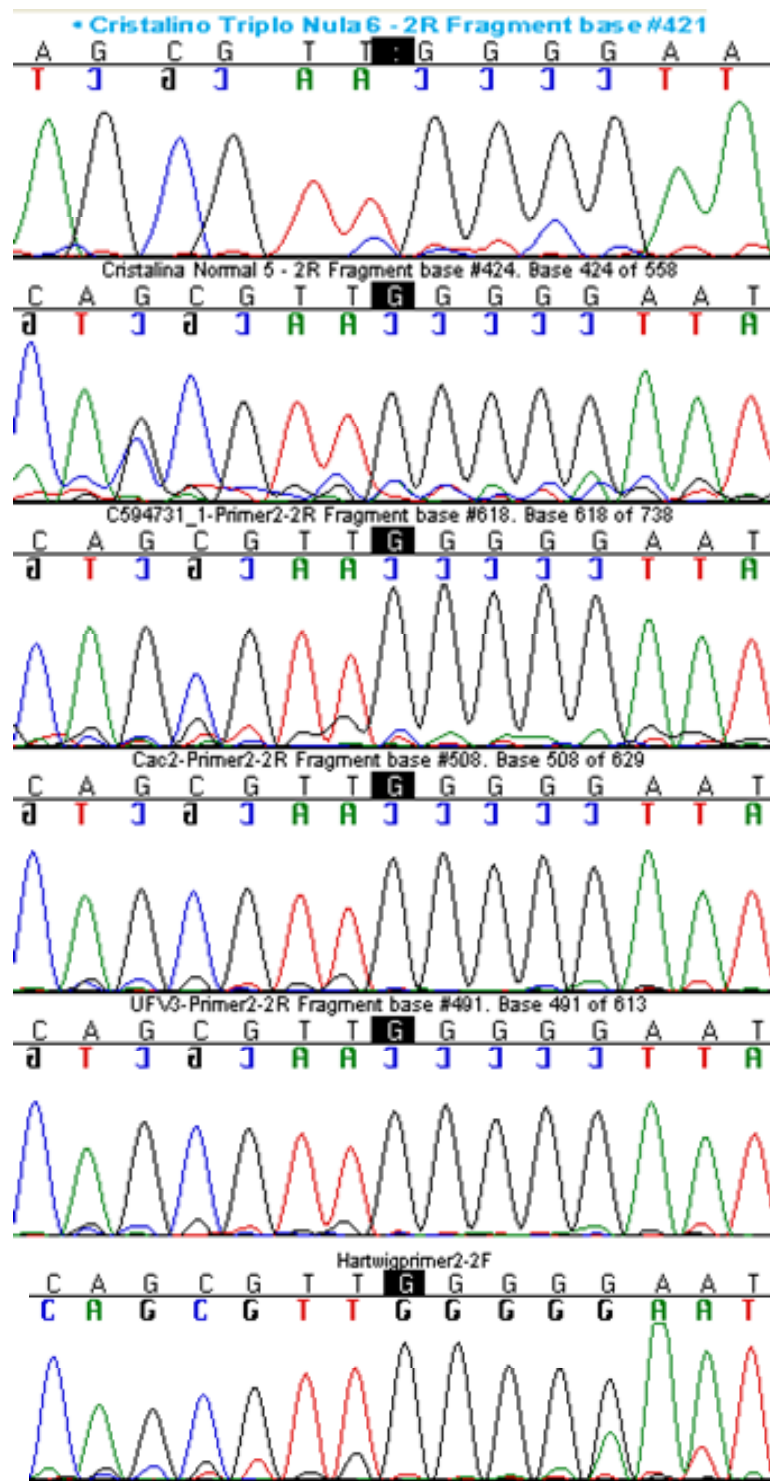


Figura 9 – Região dos cromatogramas correspondentes às seqüências das seguintes variedades (de cima para baixo): Cristalina TN, Cristalina, CS94731, CAC 1, UFV 16 e Hartwig evidenciando o indel contido no primeiro éxon do gene *Lox3*.

Nessa comparação entre as seqüências o número de polimorfismos em éxons foi menor que o contido em íntrons, ao contrário dos números obtidos

nas outras comparações anteriores. Diferentemente da comparação apenas entre as variedades normais, todos os polimorfismos na região 3' UTR ocorreram nessa comparação entre seqüências (Normais x TN). Como já comentado, essa diversidade pode futuramente ser alvo de estudos, principalmente porque essa região pode desempenhar um papel importante na estabilidade do mRNA podendo determinar diferenças na expressão gênica.

Um trabalho de desenvolvimento de marcador SNP para o gene *Lox3* em arroz foi desenvolvido por SHIRASAWA *et al.* (2008). E a razão molecular da recessividade desse gene foi um SNP A/G (alelo recessivo/alelo dominante) posicionado no sétimo éxon. Esse SNP acarretou na substituição de um códon que determina a incorporação do aminoácido triptofano na proteína em um códon de parada (TGA). Dessa forma a proteína se tornou inativa.

5.3. Genotipagem da população de RILs contrastantes para as lipoxigenases

A população de RILs utilizada é constituída de 127 indivíduos a partir do cruzamento das variedades Hartwig x Y23 (Paranagoiana TN). Pôde-se verificar que mesmo a variedade Hartwig sendo americana, ela apresentou todos os polimorfismos iguais às demais variedades seqüenciadas do PMQS-BIOAGRO, e não com a seqüência base de *Lox3* da variedade Williams 82. Sendo que a maioria das variedades produzidas no Brasil, tiveram como ancestrais variedades de origem norte americanas, essa semelhança pode ser devido a ancestralidade compartilhada entre a variedade Hartwig e as demais variedades (Cristalina, CAC 1, UFV 16).

5.3.1. Lipoxigenase 2

O material genético de 53 RILs foram amplificadas utilizando o mesmo *primer* *lox2*. Os produtos da amplificação foram seqüenciados somente na direção direta (*Forward*). Para os mesmos indivíduos foi feita a determinação do fenótipo LOX 2 através da atividade enzimática para a enzima LOX 1. Isso é possível porque os genes *Lox1* e *Lox2* estão ligados; assim, determinar o fenótipo para LOX 1 é o mesmo que para LOX 2. Houve como esperado uma

correspondência de 100% (Quadro 1) entre o fenótipo (atividade para LOX 1) e o genótipo (seqüenciamento).

Quadro 1. Correlação entre o fenótipo e o genótipo para LOX 2.

Fenótipo Alelo	LOX 2 +		LOX 2 -	
	C	D	C	D
Haplótipo TG	28	0	0	0
Haplótipo CA	0	0	25	0

C – N° de indivíduos concordantes; D- N° de indivíduos discordantes

Após o seqüenciamento e o cruzamento de dados da caracterização fenotípica, confirmou-se a associação do haplótipo TG em 28 indivíduos com a presença da enzima LOX 2 e o haplótipo CA em 25 com a ausência de LOX 2.

HA & BOERMA (2008) em estudo do gene *rcs3* que confere às plantas de soja resistência ao fungo *Cercospora sojina* Hara identificaram um SNP (A/G) posicionado no nucleotídeo 209 do gene que está 100% correlacionado com o fenótipo. Nesse trabalho foi utilizando uma população de 96 plantas F2; e sondas fluorescentes alelo específicas e a metodologia Melting Curve assay, como estratégia de genotipagem.

Um marcador SNP foi associado com a característica supernodulação em soja. O polimorfismo A/T na posição 959 do gene *GmNARK*, responsável por codificar o receptor responsável pela autoregulação da nodulação, co-segregou completamente com os indivíduos da população F2 de 40 indivíduos, respectivamente normais e supernodulantes (KIM *et al.*, 2005).

De acordo com RAFALSKI (2002), para distinguir haplótipos em uma seqüência com vários polimorfismos, deve-se considerar o número de SNPs. Os SNPs contidos na seqüência LOX 2 conjuntamente podem gerar quatro haplótipos possíveis: CA, CG, TA e TG. No entanto, se verifica entre os indivíduos a ocorrência de dois: CA e TG separados por 116 pb (Figura 10). Os dois SNPs que separadamente são considerados alelos, conjuntamente são denominados haplótipos. Existem então, dois haplótipos: CA para LOX 2+ e TG para LOX 2-, de acordo com as 53 RILs avaliadas.

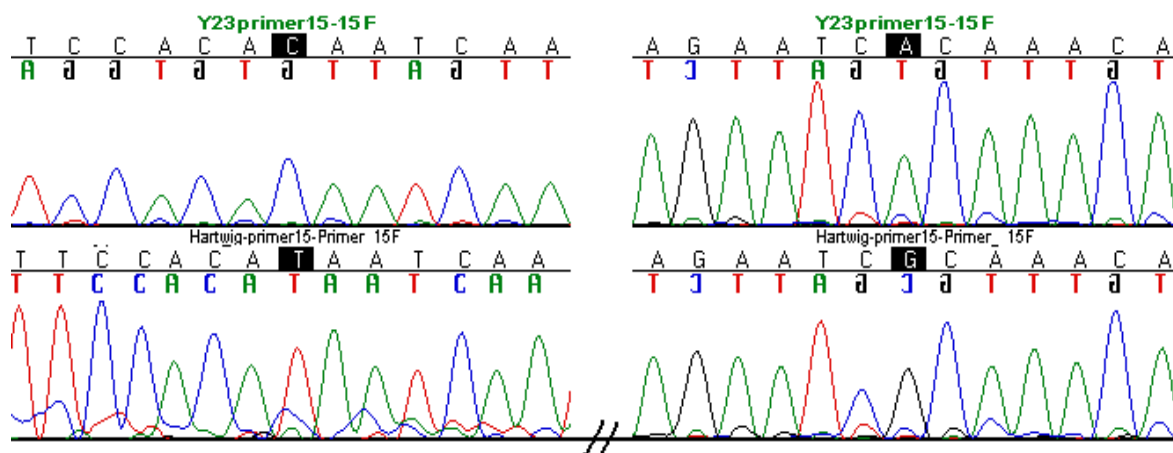


Figura 10 – Região dos cromatogramas correspondentes às sequências das variedades parentais da população de RILs utilizada na genotipagem (de cima para baixo): Y23 (Paranagoiana TN) e Hartwig, evidenciando os haplótipos contidos no gene *Lox2*. As duas barras entre as duas porções da Figura indicam que parte da sequência foi retirada.

O D' (índice de desequilíbrio de ligação) entre esses dois alelos é máximo, pois só existem indivíduos com um haplótipo ou com outro (Tabela 10).

Tabela 10 . Número de indivíduos contendo cada um dos alelos. O índice de desequilíbrio de ligação (DL) $D'=1$

Alelo 2 \ Alelo 1	C	T
A	25	0
G	0	28

Segundo BANSAL (2003), RAFALSKI (2002), REMINGTON (2001) o DL existe entre dois locos ligados quando alelos destes ocorrem no mesmo haplótipo de maneira mais freqüente do que o esperado. Isto caracteriza uma segregação dependente entre dois ou mais locos analisados (LONG, 2004). Segundo ZHU *et al.* (2003), uma consequência do desequilíbrio de ligação é a redução da variabilidade genética que resulta numa limitada diversidade de haplótipos. Em análises de dois locos separados por 55 pb em 570 genótipos de soja, 3 haplótipos dos quatro possíveis considerando o equilíbrio de ligação

foram identificados (CORYELL *et al.*, 1999), apresentando assim um D' menor que o D' para os haplótipos identificados no gene *Lox2*.

5.3.2. Lipoxigenase 3

De forma semelhante, 51 indivíduos foram subamostrados da mesma população de RILS, amplificados com o *primer lox3.2* referente ao primeiro éxon onde se encontra o indel caracterizado pela presença/ausência de uma guanina nessa posição (Figura 11).

Esse polimorfismo poderia estar associado com a presença/ausência da LOX 3 nas sementes. Após a amplificação com o *primer 3.2*, todos os fragmentos referentes aos indivíduos da população foram enviados para o sequenciamento apenas na direção direta (*Forward*). A determinação fenotípica foi feita por meio da atividade enzimática para LOX 3.

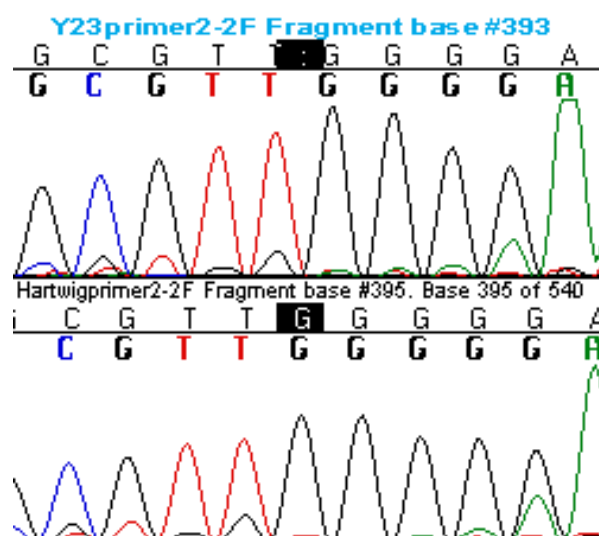


Figura 11 – Região dos cromatogramas correspondentes às sequências das variedades parentais da população de RILS utilizada na genotipagem (de cima para baixo): Y23 (Paranagoiana TN) e Hartwig, evidenciando o indel contido no primeiro éxon do gene *Lox3*.

Houve uma associação direta entre o genótipo e o fenótipo. Vinte e um indivíduos apresentaram a guanina na posição 259 (G259) e foram LOX 3+ e 30 apresentaram o indel (: 959) sendo LOX 3- (Quadro 2). Dessa forma pode-se concluir que o indel presente no primeiro éxon pode ser utilizado como marcador para a característica presença/ausência de LOX 3 nas sementes de

soja, sendo a deleção associada a recessividade e a presença da guanina determinante da dominância.

Quadro 2 . Correlação entre o fenótipo e o genótipo para LOX 3

Fenótipo Alelo	LOX 3 +		LOX 3 -	
	C	D	C	D
G959	21	0	0	0
:959 (deleção)	0	0	30	0

C – N° de indivíduos concordantes; D- N° de indivíduos discordantes

HAYASHI *et al.* (2006) descreveram o uso de indels como marcadores para os seguintes genes: *Piz*, *Piz-t*, *Pik*, *Pik-m*, que conferem resistência ao fitopatógeno do arroz *Magnaporthe grisea*. As populações utilizadas variaram de acordo com o gene, mas todas eram populações F2 ou subgrupo dessas populações. Todos os marcadores indels foram eficientes na seleção assistida por marcadores.

Marcadores indel para o gene *Almt1*, que condiciona tolerância ao alumínio em trigo, foram desenvolvidos por RAMAN *et al.* (2006). Os marcadores indel ALMT1-SSR3a e ALMT1-SSR3b derivaram de marcadores SSR. Ambos co-segregaram completamente com a tolerância ao alumínio, sendo que a validação foi feita utilizando 20 variedades parentais de origens diversas. Segundo esse autor, marcadores Indel podem ser úteis na seleção assistida por marcadores bem como estudos de evolução de populações.

6. CONCLUSÕES

- A presença de dois dos três polimorfismos descritos por KIM *et al.* (2004) foram visualizados a partir do seqüenciamento de parte do gene e foram confirmados os resultados que confirmam que eles podem ser usados como marcadores da ausência/presença da enzima LOX 2 nas variedades de soja estudadas.
- A identificação dos marcadores SNP para lipoxigenase 2, permite a conseqüente determinação do genótipo *Lox1*, uma vez que os genes são ligados.
- Foi possível estimar o número de polimorfismos por Kb, bem como o número de SNPs na região codificadora e não codificadora; por meio do seqüenciamento de grande parte da sequência nucleotídica responsável pela expressão da enzima LOX 3.
- Um indel presente no primeiro éxon da região codificante da lipoxigenase 3 provavelmente é a razão molecular da recessividade apresentada pelas variedades LOX 3 -.

7. PERSPECTIVAS

Para que os polimorfismos estudados nesse trabalho sejam utilizados pelo Programa de Melhoramento da Qualidade da Soja (PMQS) do BIOAGRO/UFV, é necessário que uma estratégia de genotipagem que priorize rapidez e baixo custo seja desenvolvida.

A genotipagem de outras populações também se faz necessário para confirmar a robustez e a eficiência desses marcadores moleculares.

Assim que completamente corroborada a hipótese de que duas substituições para a enzima lipoxigenase 2 e um indel para lipoxigenase 3 são respectivamente responsáveis pela inatividade das enzimas; e uma boa metodologia de genotipagem seja escolhida, será possível o uso desses marcadores no PMQS do BIOAGRO na UFV.

8. REFERÊNCIAS BIBLIOGRÁFICAS

- ALBERTS, B.; JOHNSON, A.; LEWIS, J.; RAFF, M.; ROBERTS, K. ; WALTER, P. *Molecular Biology of The Cell*. 4th ed. Garland Publishing. NewYork. p.1661. 2002.
- ALTSCHUL, S.F.; MADDEN, T.L.; SCHAFFER, A.A.; ZHANG, J.; ZHANG, Z.; MILLER, W. e LIPMAN, D.J. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research*. v.25, p. 3389-3402. 1997.
- AXELROD, B.; CHEESBROUGH, T.M.; LAASKO, S.L. Lipoxygenase from soybean. *Methods in Enzymology*. v. 71, p. 441-51. 1981.
- BALLVORA, A.; HESSELBACH, J.; NIEWOHNER, J.; LEISTER, D.; SALAMINI, F.; GEBHARDT, C. Marker enrichment and high-resolution map of the segment of potato chromosome VII harbouring the nematode resistance gene. *Molecular and General Genetics*. v.249, p. 82-90. 1995.
- BANSAL, A.; BOYD, P.R.; & MCGINNIS, R. Tools for statistical analysis of genetic data. In: BARNES, M.R. e GRAY, I.C. *Bioinformatics for Geneticists*. John Wiley & Sons, Ltd. p. 217-245. 2003.
- BENNETT, C.D.; CAMPBELL, M.N.; COOK, C.J.; EYRE, D.J.; NAY, L.M.; NIELSEN, D.R.; RASMUSSEN, R.P.; BERNARD, P.S. The LightTyper (TM): High-throughput genotyping using fluorescent melting curve analysis. *BioTechniques* v.34, p. 1288-1295. 2003.
- BERNARD, P.S.; LAY, M.J.; WITTEWER, C.T. Integrated amplification and detection of the C677T point mutation in the methylenetetrahydrofolate reductase gene by fluorescence resonance energy transfer and probe melting curves. *Analytical Biochemistry*. v.255, p. 101-107.1998.
- BORÉM, A. & CAIXETA, E.T. *Marcadores Moleculares*. Editora UFV. Viçosa-MG. Terceira edição. 374p. 2006.
- BLONDAL, T.; WAAGE, B.G.; SMARASON, S.V.; JONSSON, F.; FJALLDAL, S.B.; STEFANSSON, K.; GULCHER, J. e SMITH, A.V. A novel MALDI-TOF based methodology for genotyping single nucleotide polymorphism. *Nucleic Acids Research*. v.31. n.24 e 155. 2003.
- BOTSTEIN, D. & RISCH, N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nature Genetics*, London. v. 33, p. 228-237. 2003.

- BOTSTEIN, D., WHITE, R.L., SKOLNICK, M., DAVIS, R.W. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics*, 32: 314-331, 1980.
- BROOKES, A. J. The essence of SNP. *Gene*, Amsterdam, v.234, p. 177-186, 1999.
- CAI, H.W. & MORISHIMA, H. QTL clusters reflect character associations in wild and cultivated rice. *Theoretical and Applied Genetics*, v. 104, p. 1217-1228, 2002.
- CARDINAL, A.J.; LEE, M.; SHAROPORA, N.; WOODMAN-CLIKEMAN, W.L.; LONG, M.J. Genetic mapping and analysis of quantitative trait loci for resistance to stalk tunneling by the European corn bores in maize. *Crop Science*, v. 41, p. 835-845, 2001.
- CARPENTIERI-PÍPOLO, V.; GARCIA, J.D.; RINALDI, D.A. Desenvolvimento e aplicações de marcadores moleculares. In: *Biotecnologia na Agricultura: aplicações e biossegurança*, p. 372. 2006.
- CHANG, S.J.C.; DOUBLER, T.W.; KILO, V.Y.; ABU-THREDEIH, J.; PRABHU, R.; FREIRE, V.; SUTTNER, R.; KLEIN, J.; SCHMIDT, M.E.; GIBSON, P.T.; LIGHTFOOT, D.A. Association of loci underlying field resistance to soybean sudden death syndrome (SDS) and cyst nematode (SCN) race 3. *Crop Science*. v. 37, p. 965-971, 1997.
- CHEN, H.; WANG, S.; XING, Y.; XU, C.; HAYES, P.M.; ZHANG, O. Comparative analyses of genomic locations and race specificities of loci for quantitative resistance to *Pyricularia grisea* in rice and barley. *Proceedings of the National Academy of Sciences, USA*, v. 100, n. 5, p. 2544-2549, 2003.
- CHING, A.; CALDWELL, K. S.; JUNG, M.; DOLAN, M.; SMITH, O. S. SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. *BMC Genetic*. v.3; p.19, 2002.
- CHO, R. J. *et al.* Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. *Nature Genetics*, v. 23, p. 203 - 207, 1999.
- CHOI, Y.; HYTEN, D.L.; MATUKUMALLI, L.K.; SONG Q.; CHAKY, J.M.; QUIGLEY, C.V.; CHASE, K.; LARK, K.G.; REITER, R.S.; YOON, M.; HWANG, E.; 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*, v.176; p.685–696, 2007.

- COLLINS, F.S.; BROOKS, L.D. e CHARKRAVARTI, A. A DNA polymorphism discovery resource for research on human genetic variation. *Genome Research*, v.8; p.1229–1231, 1998.
- Companhia Nacional de Abastecimento. Acompanhamento de safra brasileira: grãos, nono levantamento, junho/2009 / Companhia Nacional de Abastecimento. – Brasília, P. 39, CONAB. 2009.
- CORYELL, V.H.H.; JESSEN, J.M.; SCHUPP, D.; WEBB e P. KEIM. Allele-specific hybridization markers for soybean. *Theor. Appl. Genet.* V.98. p. 690–696. 1999.
- COULONDRE, C.; MILLER, J.H.; FARABAUGH, P.J.; AND GILBERT, W. Molecular basis of base substitution hot spots in *Escherichia coli*, *Nature*, vol. 274, p.775–780. 1978.
- CREGAN, P. B.; JARVIK, T.; BUSH, A.L.; SHOEMAKER,R.C.; LARK, K.G.; KAHLER, A.L.; KAYA, N.; VANTOAI, T.T.; LOHNES, D.G.; CHUNG, SPECHT,J.E. An Integrated Genetic Linkage Map of the Soybean Genome. *Crop Science*. 39:1464–1490. 1999.
- DUFRESNE, P.J.; JENNI, S.; FORTIN, M.G. FRET hybridization probes for the rapid detection of disease resistance alleles in plants: detection of corky root resistance in lettuce. *Molecular Breeding*. v.13, p. 323-332. 2004.
- EDWARDS, J.D.; LEE, V.M. e McCOUCH, S.R. Sources and predictors of resolvable indel polymorphism assessed using rice as a model. *Mol. Genet. Genomics*, v. 27, p. 298–307. 2004.
- FELTUS, F. A.; WAN, J.; SCHULZE, S. R.; ESTILL, J. C.; JIANG, N. An SNP resource for rice genetics and breeding based on subspecies indica and japonica genome alignments. *Genome Research*. v.14; p.1812-1819. 2004
- FERREIRA, A.R.; FOUTZ, K.R.; KEIM, P. Soybean genetic map of RAPD markers assigned to an existing scaffold RFLP map. *The American Genetic Association*. v. 91, p. 392-396. 2000.
- FERREIRA, M.E. & GRATTAPAGLIA, D. Introdução ao uso de marcadores RAPD e RFLP em análise genética. Brasília: EMBRAPA-CENARGEN, 220p. 1995.
- GARDNER, W.H. Biological roles and biochemistry of a lipoxygenase pathway. *Horticultural Science*. v.30, p.10, 1995.
- GAUDET, M.; FARA, A.; SABATTI, M.; KUZMINSKY, E.; MUGNOZZA, G.S. Single-reaction for SNP Genotyping on Agarose Gel by Allele-specific

- PCR in Black Poplar (*Populus nigra* L.). *Plant Molecular Biology Reporter*. v.25, p.1-9. 2007.
- GEBHARDT, C. A high-resolution map of the vicinity of the R1 locus on chromosome V potato based on RFLP and AFLP markers. *Molecular and General Genetics*. v.249, p. 7481. 1995.
- GIBSON, N.J. The use of real-time PCR methods in DNA sequence variation analysis. *Clinica Chimica Acta*, v.363. p. 32-47. 2006.
- GONDIM, T.C. de O. Qualidade Fisiológica de Sementes de Soja com Presença ou Ausência de Lipoxigenases, em Diferentes Épocas de Colheita. Viçosa, Minas Gerais, UFV, 61p. Dissertação (Mestrado em Fitotecnia) - Universidade Federal de Viçosa, 2001.
- GRIFFITHS, A.J.F.; GELBART, W.M.; MILLER, J.H. e LEWONTIN, R.C. *Genética moderna*. Rio de Janeiro: Guanabara Koogan. p. 589. 2001.
- GRIVET, L.; GLASZMANN, J.C.; VINCENTZ, M.; DA SILVA, F. E ARRUDA, P. ESTs as a source for sequence polymorphism discovery in sugarcane: example of the Adh genes. *Theoretical and Applied Genetics*. v.106. p.190–197. 2003.
- GROH, S.; GONZÁLEZ-DE-LEÓN, D.; KHAIRALLAH, M. M.; JIANG, C.; BERGVINSON, D.; BOHN, M.; HOISINGTON, D. A.; MELCHINGERO, A. E. QTL Mapping in Tropical Maize: III. Genomic Regions for Resistance to *Diatraea* spp. and Associated Traits in Two RIL Populations. *Crop Science*. v.38. p.1062–1072. 1998.
- GUIMARÃES, E.P.; RUANE, J.; SCHERF, B.D.; SONNINO, A.; DARGIE, J.D. Marker-assisted selection. Roma-Itália. Editor: Food and agriculture organization of the united nations. 494 p. 2007.
- GUIMARÃES, E.P.M. & COSTA, M.C.R. SNPs: Sutis diferenças de um código. *Biotecnologia Ciência & Desenvolvimento*, Brasília, v.26, p.24-27. 2002.
- GUPTA, P.K.; VARSHNEY, R.K.; SHARMA, P.C. e RAMESH, B. Molecular Markers and Their Applications in Wheat Breeding, *Plant Breeding*. v. 118, p. 369–390. 1999
- HA, B.K. & BOERMA, H.R. High-throughput SNP genotyping by melting Curve analysis for resistance to southern root-knot nematode and Frogeye leaf spot in soybean. *Journal Crop Science Biotechnology*. V.11 n.2 p.91-100. 2008.

HAFF, L.A. & SMIRNOV, I.P. Single-nucleotide polymorphism identification assays using a thermostable DNA polymerase and delayed extraction MALDI-TOF mass spectrometry. *Genome Research*. v.7, p.378-388. 1997.

HAYASHI, K.; YOSHIDA, H.; ASHIKAWA, I. Development of PCR-based allele-specific and Indel marker sets for nine rice blast resistance genes. *Theoretical and Applied Genetics*. v.113.p.251–260. 2006.

HE, P.; Li, J.Z.; ZHENG, X.W.; SHEN, L.S.; LU, C.F.; CHEN, Y.; ZHU, L.H. Comparison of molecular linkage maps and agronomic trait loci between DH and RIL populations derived from the same rice cross. *Crop Science*. v. 41, p. 1240– 1246. 2001.

HILDEBRAND, D. F.; HAMILTON-KEMP, T. R.; LEGG, C. S.; BOOKJANS, G. Plant lipoxygenases: occurrence, properties and possible functions. *Current Topics in Plant Biochemistry and Physiology*. Columbia, v. 7, p. 201-219, 1988.

HILDEBRAND, D.F. Lipoxygenases. *Physiologia Plantarum*. v.76, p. 249-253.1988.

HNETKOVSKY, N.; CHANG, S.J.C.; DOUBLER, T.W.; GIBSON, P.T.; LIGHTFOOT, D.A. Genetic mapping of loci underlying field resistance to soybean sudden death syndrome (SDS). *Crop Science*. v. 36, p. 393-400. 1996.

HIRSCHHORN, J. N. & DALY, M. J. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet*. 6 (2): 95-108. 2005.

HOOGENDOORN, B.; COLEMAN, S. L, GUY, C.A.; SMITH, K., BOWEN, T.; BUCKLAND , P.L.; O'DONOVAN, M.C. Functional analysis of human promoter polymorphisms. *Human Molecular Genetics*. v.12, n.18, Sep 15, p.2249-54. 2003.

<http://www.expasy.ch/uniprot/P09186>, acessado em: 19/06/09

<http://bfgl.anri.barc.usda.gov/soybean/>, acessado em: 19/06/09

<http://www.phytozome.net/soybean>, acessado em:19/06/09

<http://www.rcsb.org/pdb/home/home.do>, acessado em: 19/06/09

HYTEN, D. L.Q.; SONG, Y.; ZHU, I. Y.; CHOI, R. L.; NELSON, R.L.; COSTA, J.M.; SPECHT, J.E.; SHOEMAKER, R.C.; CREAGAN, P.B., Impacts of

genetic bottlenecks on soybean genome diversity. Proceedings of the National Academy of Sciences. v.103. p. 16666–16671. 2006

INGVARSSON, P. K. Nucleotide polymorphism and linkage disequilibrium within and among natural populations of European aspen (*Populus tremula* L., Salicaceae). *Genetics*. 169 (2): 945-953. 2005.

IQBAL, M.J., MEKSEM, K., NJITI, V.N., KASSEM, M.A., LIGHTFOOT, D.A. Microsatellite markers identify three additional quantitative trait loci for resistance to soybean sudden-death syndrome (SDS) in Essex X Forrest RILs. *Theoretical and Applied Genetics*, v. 102, p. 187-192. 2001.

JANDER, G.; NORRIS, S.R.; ROUNSLEY, S.D.; BUSH, D.F.; LEVIN, I.M. Arabidopsis map-based cloning in the post-genome era. *Plant Physiology*, v.129; p.440-450. 2002.

JANSSEN, P.; AUDIT, B.; CASES, I.; DARZENTAS, N.; GOLDOVSKY, L.; KHLESTKINA, E.K.; SALINA, E.A. SNP Markers: Methods of Analysis, Ways of Development, and Comparison on an Example of Common Wheat. *Russian Journal of Genetics*. v.42. n.6. p.585-594. 2005.

KIM, M.Y.; HA, B.K.; JUN, T.H.; HWANG, E.Y.; VAN, K.; KUK, Y.; LEE, S. Single nucleotide polymorphism discovery and linkage mapping of lipoxygenase-2 gene (Lx 2) in soybean. *Euphytica*, v. 135, n. 2 , p. 169-177. 2004

KIM, M.Y.; VAN, K.; LESTARI, P.; MOON, J.K.; LEE, S.H. SNP identification and SNAP marker development for a GmNARK gene controlling supernodulation in soybean. *Theoretical Applied Genetics*. V. 110. p. 1003-1010. 2005.

KINSELLA, J.E. Functional properties of soy proteins. *Journal of the American Oil Chemists' Society* ., v.56, p.242-258. 1979.

KITAMURA, K.; DAVIES, C.S.; KAIZUMA, N. Genetic analysis of a null-allele for lipoxygenase-3 in soybean seeds. *Crop Science*, v.23, n.5, p.924-927. 1983.

KRUGLYAK, L. The use of a genetic map of biallelic markers in linkage studies. *Natural Genetic*, v.17; p.21–24. 1997.

KWOK, P.Y. & GU, Z. Single nucleotide polymorphism libraries: why and how are we building them. *Molecular Medicine Today*, v. 5, p. 538 – 543. 1999.

- KWOK, P.Y.; DENG, Q.; ZAKERI, H. e NICKERSON, D.A. Increasing the information content of STS-based genome maps: identifying polymorphisms in mapped STSs. *Genomics*, v.31; p.123–126. 1996
- LANGAEE, T. & RONAGHI, M. Genetic variation analyses by Pyrosequencing. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*. v.573. p. 96-102. 2005.
- LEVINE, M. & TJIAN, R. Transcription regulation and animal diversity. *Nature*, v.424, n.6945, Jul 10, p.147-51. 2003.
- LEWIN, B. *Genes VII*. Oxfor: Oxford University Press, p 955. 2000.
- LIVAK, K.J. Allelic discrimination using fluorogenic probes and the 5' nuclease assay. *Genetic Analytical – Biomolecular Engineering*. v.14; p. 143-149. 1999.
- LONG, J.R.; ZHAO, L.J.; LIU, P.Y.; LU, Y.; DVORNYK, V.; SHEN, H.; LIU, Y.J.; ZHANG, Y.Y.; XIONG, D.H.; XIAO, P. e DENG, H.W. Patterns of linkage disequilibrium and haplotype distribution in disease candidate genes. *BMC Genet*. v.5, n.1,p. 11. 2004.
- MAMOTTE, C.D. Genotyping of single nucleotide substitutions. *Clinical Biochemist Reviews*. v.27. p. 63-75. 2006.
- MARKET, C.L. & MOLLER, F. Multiple forms of enzymes tissue, ontogenic and species specific patterns. *Proceeding of the National Academy of Sciences*, v. 45, p. 753-762. 1959
- MARTINS, C.A.O.; SEDIYAMA,C.S; MOREIRA, M.A.; REIS, M.S.; ROCHA, V.S; OLIVEIRA, M.G.A. Efeito da eliminação genética das lipoxigenases sobre as características agrônômicas da soja. *Pesq. Agropec. bras.* , Brasília, v.37, n.10,p.1398. 2002.
- MASOTTI, C., L.M. ARMELIN-CORREA; SPLENDORE, A, LIN, C.J.; BARBOSA, A.; SOGAYAR,M. C.; PASSOS-BUENO; M. R. A funcional SNP in the promoter region of TCOF1 is associated with reduced gene expression and YY1 DNA-protein interaction. *Gene*, v.359, p.44-52. 2005.
- McCOUCH, S.R.; KOCHERT, G.; YU, Z.H.; WANG, Z.Y.; KHUSH, G.S.; COFFMAN, W.R. e TANKSLEY, S.D. Molecular mapping of rice chromosomes. *Theoretical Applied Genetics*. v.76, p. 815–829. 1988.
- MEKSEM, K.; NJITI, V.N.; BANZ, W.J.; IQBAL, M.J.; KASSEM, M.M.; HYTEN, D.L.; YUANG, J.; WINTERS, T.A.; LIGHTFOOT, D.A. Genomic regions

that underlie soybean seed isoflavone content. *Journal of Biomedicine and Biotechnology*, 1:1, p. 38-44. 2001.

MEKSEM, K., LEISTER, D., PELEMAN, J., ZABEAU, M., SALAMINI, F. AND GEBHARDT, C. A high-resolution map of the vicinity of the R1 locus on chromosome V of potato based on RFLP and AFLP markers. *Plant Breeding*, v.249, n. 1, p.74-81. 1995.

MONTEIRO, M.R.P.; COSTA, N. M. B. ; OLIVEIRA, M. G. A.; PIRES, C. V.; MOREIRA, M. A. Protein quality of soy lines lacking Kunitz's Trypsin Inhibitor and Lipoxygenase isozymes. *Annual Review of Nutrition*., Campinas, 17(2):195-205. 2004

MOREIRA M.A. Programa de melhoramento genético da qualidade de óleo e proteína de soja desenvolvido na UFV. In. Congresso Brasileiro de Soja, Londrina-PR. Embrapa-Soja, p 99-104. 1999.

MULLER, M.J. Enzymes involved in jasmonic acid biosynthesis. *Plant Physiology*, v. 100, p. 653-663. 1997.

MULLIS, K. & FALLONA, F. Specific synthesis of DNA in vitro via a polymerase catalyzed chain reaction. *Methods in Enzymology*. v. 55, p. 335-350. 1987.

NACHIT, M.M.; ELOUAFI, I.; PAGNOTTA, M.A.; SALEH, A.E.; IACONO, E.; LABHILILI, M.; ASBATI, A.; AZRAK, M.; HAZZAM, H.; BENSCHER, D.; KHAIRALLAH, M.; RIBAUT, J.M.; TANZARELLA, O.A.; PORCEDDU, E.; SORRELLS, M.E. Molecular linkage map for a intraspecific recombinant inbred population of durum wheat (*Triticum turgidum* L. var. durum). *Theoretical and Applied Genetics*, v. 102, p. 177-186. 2001.

NASU, S.; SUZUKI, J.; OHTA, R.; HASEGAWA, K.; YUI, R. Search for and analysis of single nucleotide polymorphisms (SNPs) in rice (*Oryza sativa*, *Oryza rufipogon*) and establishment of SNP markers. *DNA Research*, v.9; p.163-171. 2002.

NJITI, V.N.; MEKSEM, K.; IQBAL, M.J.; JOHNSON, J.E.; KASSEM, M.A.; ZOBRIK, K.F.; KILO, V.Y.; LIGHFOOT, D.A. Common loci underlie field resistance to soybean sudden death syndrome in Forrest, Pyramid, Essex, and Douglas. *Theoretical and Applied Genetics*, v. 104, p. 294-300. 2002.

NORDBORG, M.; HU, T.T.; ISHINO, Y.; JHAVERI, J.; TOOMAJIAN, C. The pattern of polymorphism in *Arabidopsis thaliana*. *Public Library of Science Biology*, v.3: p196. 2005.

- OLIVEIRA, D.A.; PIOVESAN, N.D.; MORAES, R.M.A.; ROCHEBOIS, G.B.; OLIVEIRA, M.G.A.; BARROS, E.G.; MOREIRA, M.A. Identification of the tree 56 genotypic classes for soybean lipoxygenases 1 and 3 based on enzymatic activity. *Biotechnology Techniques*, 12:71-74. 1998.
- PÓVOA, A.M. Frequência de snps, estrutura de haplótipos e Desequilíbrio de ligação para os genes *CAD2* e *COMT2* da via de lignificação em eucalyptus. 126fl. Dissertação de mestrado. Universidade Católica De Brasília. Brasília, defendida em 26/08/05.
- PROCUNIER, J.D.; GRAY, M.; LIAKAT, A.M. Single Nucleotide Polymorphisms (SNPs) in Hexaploid Wheat and High Throughput SNP Detection by Invader Operation System. *Proc. Plant and Animal Genomes*. 11th conference., San-Diego. P.251. 2003.
- RAFALSKI, A. Applications of single nucleotide polymorphisms in crop genetics. *Current Opinion in Plant Biology*. 5 (2): 94-100. 2002.
- RAMAN, H.; RAMAN, R.; WOOD, R. e MARTIN, P. Repetitive indel Markers within the ALMT1 Gene Conditioning Aluminium Tolerance in Wheat (*Triticum aestivum* L.). *Molecular Breeding*. v.18.p.171–183. 2006.
- REMINGTON, D. L., THORNSBERRY, J. M., MATSUOKA, Y., WILSON, L.M., WHITT, S. R., DOEBLEY, J., KRESOVICH, S., GOODMAN, M. M. e Buckler, E.S. T. Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proceedings of the National Academy of Science*. v. 20, p.11479-11484. 2001.
- ROZEN, S. & SKALETSKY, H. Primer3 on the WWW for General Users and for Biologist Programmers. *Methods in Molecular Biology*, vol. 132: *Bioinformatics Methods and Protocols*, p. 365-386. 2000.
- SACHIDANANDAM, R.; WEISSMAN, D.; SCHMIDT, S.C.; KAKOL, J.M.; STEIN, L.D. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*, v.409; p.928-933. 2001
- SAIER, M.H. Genome sequencing and informatics: new tools for biochemical discoveries. *Plant Physiol*. 117 (4): 1129-1133. SIEDOW, J.N. Plant lipoxygenase, structure and function. *Annual Review of Plant Physiology and Plant Molecular Biology* . v. 42, p.145-88. 1998.
- SAKIYAMA, N.S. Marcadores moleculares e as hortaliças. *Horticultura brasileira*. v.11, n.2, p.204-206. 1993

- SANGER, F., NICKLEN, S. e COULSON, A.R. DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Science USA 74:5463-5467. 1977.
- SCALLON, B.J.; DICKINSON, C.D.; NIELSEN, N.C.; Characterization of a null-allele for the Gy4 glycinin gene from soybean. Molecular Gene and Genetic, v.208; p.107–113. 1987.
- H. SCHREM, J. KLEMPNAUER, J. BORLAK. Liver-enriched transcription factors in liver function and development. Part I and Part II. Pharmacol Rev, v.54, n.1, Mar,p.129-330. 2002.
- SCHMID, K J.; SORENSEN, T.R.; STRACKE, R.O.; TORJEK, T. Large-scale identification and analysis of genome-wide single-nucleotide polymorphisms for mapping in Arabidopsis thaliana. Genome Research, v.13; p.1250-1257. 2003.
- SEDIYAMA, T.; TEIXEIRA, R.C.; REIS, M.S. Melhoramento de soja. In: BORÉM, A. (Ed) Melhoramento de espécies cultivadas. Viçosa – MG, p.487-534. 1999.
- SÉNE, M.; CAUSSE, M.; DAMERVAI, C.; THÉVENOT, C.; PRIOUL, J.L. Quantitative trait loci affecting amylose, amylopectin and starch content in maize recombinant inbred lines. Plant Physiology Biochemistry, v. 38 (6), p. 459-472. 2000.
- SHIRASAWA, K.; TAKEUCHI, Y.; EBITANI, T.; SUZUKI, Y. Identification of gene for rice (*Oryza sativa*) seed lipoxygenase-3 involved in the generation of stale flavor and development of SNP markers for lipoxygenase-3 deficiency. Breeding Science. v.58. p. 169-176. 2008.
- SIEDOW, J.N. Plant lipoxygenase: Structure and Function. Annual Reviews. v. 42. p. 145-188. 1991.
- SILVA, M.S; NAVES, M.M.; OLIVEIRA, R.B.; LEITE, O.S.M. Composição Química e Valor Protéico do Resíduo de Soja em Relação ao Grão de Soja. Revista Ciência e Tecnologia de Alimentos., Campinas, 26(3): 571-576. 2006.
- 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, v. 21, p. 281 – 288. 2003.
- SOMERS, D.J.; KIRKPATRICK, R.; MONIWA, M. e WALSH, A. Mining Single-Nucleotide Polymorphisms from Hexaploid Wheat ESTs, Genome, v. 49, p. 431–437. 2003.

- SOUTHERN, E.M.. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal. Molecular. Biology.* v.98, p. 503–510. 1975.
- SUDA, I.; HAJIKA, M.; NISHIBA, Y.; FURUTA, S.; IGITA, K. Simple and rapid method for the selective detection of individual lipoxygenase isozymes in soybean seeds. *Journal of Agricultural and Food. Chemmistry, Easton,* v.43, p.742-747. 1995.
- TENAILLON, M.I.; SAWKINS, M.C.; LONG, A.D.; GAUT, R.L.; DOEBLEY, J.F. Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays ssp.(L.)*). *Proceedings of the National Academy of Sciences U.S.A,* v.98; p.9161-9166. 2001.
- THE INTERNATIONAL HAPMAP PROJECT. The International HapMap Project. *Nature.* 426 (6968): 789-796. 2003.
- THORNSBERRY, J.M.; GOODMAN, M.M.; DOEBLEY, J.; KRESOVICH, S.; NIELSEN, D. e BUCKLER, E.S.T. Dwarf polymorphisms associate with variation in flowering time. *Nature Genetics.* 28 (3): 286-289. 2001.
- THUMMA, B. R., NOLAN, M. F., EVANS, R. e MORAN, G. F. Polymorphisms in Cinnamoyl CoA Reductase (CCR) are Associated with Variation in Microfibril Angle in *Eucalyptus spp.* *Genetics.* 2005.
- TYAGI, S.; BRATU, D.P.; KRAMER, F.R. Multicolor molecular beacons for allele discrimination. *Nature Biotechnology.* 16: 49-53. 1998.
- UGA, Y.; FUKUTA, Y.; CAI, H.W.; IWATA, H.; OHSAWA, R.; MORISHIMA, H.; FUJIMURA, T. Mapping QTLs influencing rice floral morphology using recombinat inbred lines derived from a cross between *Oryza sativa L.* and *Oryza rufipogon Griff.* *Theoretical and Applied Genetics,* v. 107, p. 218-226. 2003.
- VANDERGRIFT, W.L. Value of soybean meals in swine feeding. *GEORGIA NUTRITION CONFERENCE FOR THE FEED INDUSTRY,* Atlanta. *Procedings.* University of Georgia, p. 75-81. 1983.
- VARSHNEY, R.K.; NAYAK, S.; JAYASHREE, B.; ESHWAR, K.; UPADHYAYA, H.D. e HOISINGTON, D.A. Development of cost-effective SNP assays for chickpea genome analysis and breeding. *An Open Access Journal,* v.3, n.1. 2007.
- VIGNAL, A.; MILAN, D.; SANCRISTOBAL, M.; EGGEN, A. A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution, Paris,* v.34, p.275-305. 2002.

- VOET, D.; VOET, J.G. e PRATT, C.W. Fundamentos de Bioquímica, 3ª ed., Artmed, Porto Alegre. 2006.
- VOS, P.; HOGERS, R.; BLEEKER, M.; REIJANS, M.; VAN DE LEE, T.; HORNES, M.; FRIJTERS, A.; POT, J.; PELEMAN, J. e KUIPER, M. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research*. 23: 4407–4414. 1995.
- WANG, D.G.; FAN, J.B.; SIAO, C.J.; BERNO, A.; YOUNG, P. Large-scale identification, mapping, and genotyping of single nucleotide polymorphisms in the human genome. *Science*. V 280. P. 1077–1082. 1998.
- WEBER, J.L. & MAY, P.E. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *American Journal of Human Genetics*. 44: 388–396. 1989.
- WELSH, J. & McCLELLAND, M. Finger-printing genomes using PCR with arbitrary primers. *Nucleic Acids Research*. 18: 7213–7218. 1990.
- WETMUR, J.G. DNA probes - Applications of the principles of nucleic-acid hybridization. *Critical Reviews in Biochemistry and Molecular Biology*. v.26. p. 227-259. 1991.
- WILLIAMS, J.G., KUBELIK, A.R., LIVAK, K.J. RAFALSKI, J.A., TINGEY, S.V. DNA polymorphism amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research*, v. 18, p. 6531-6535. 1990.
- WOLFORD, J.K.; BLUNT, D.; BALLECER, C.; PROCHAZKA, M. High-throughput SNP detection by using DNA pooling and denaturing high performance liquid chromatography (DHPLC). *Human Genetic*. v.107. p.483-487. 2000.
- WRAY, G.A. The evolutionary significance of cis-regulatory mutations. *Nature Reviews Genetics*, v.8, n.3, Mar, p.206-16. 2007.
- YADAV, N.S. Genetic modification of soybean oil quality. In: VERMA, D. P. S., SHOEMAKER, R. C. (Eds) *Soybean genetics, molecular biology and biotechnology*. USA: CAB INTERNATIONAL, p. 127-188. 1996.
- YENOFISKY, R.L.; FINE, M.; LIN, C. Isolation and characterization of a soybean (*Glycine max*) lipoxygenase-3 gene. *Molecular and General Genetics*. V.211. p. 215-222. 1988.
- YUAN, J., NJITI, V. N., MEKSEM, K., IQBAL, M. J., TRIWITAYAKORN, K., KASSEM, My. A., DAVIS, G. T., SCHMIDT, M. E., LIGHTFOOT, D. A. Quantitative trait loci in two soybean recombinant inbred line

populations segregating for yield and disease resistance. *Crop Science*, v. 42, p. 271–277. 2002.

ZAKHAROVA, E.S.; EPISHIN, S.M.; VINETSKI, Y.P. An attempt to elucidate the origin of cultivated soybean via comparison of nucleotide sequences encoding glycinin B4 polypeptide of cultivated soybean, *Glycine max*, and its presumed wild progenitor, *Glycine soja*. *Theoretical and Applied Genetics*, v.78; p.852–856. 1989.

ZHU, Y.L.; SONG, Q.J.; HYTEN, D.L.; VAN TASSELL, C.P.; MATUKUMALLI, L.K. Single-nucleotide polymorphisms in soybean. *Genetics*, v.163; p.1123-1134. 2003.

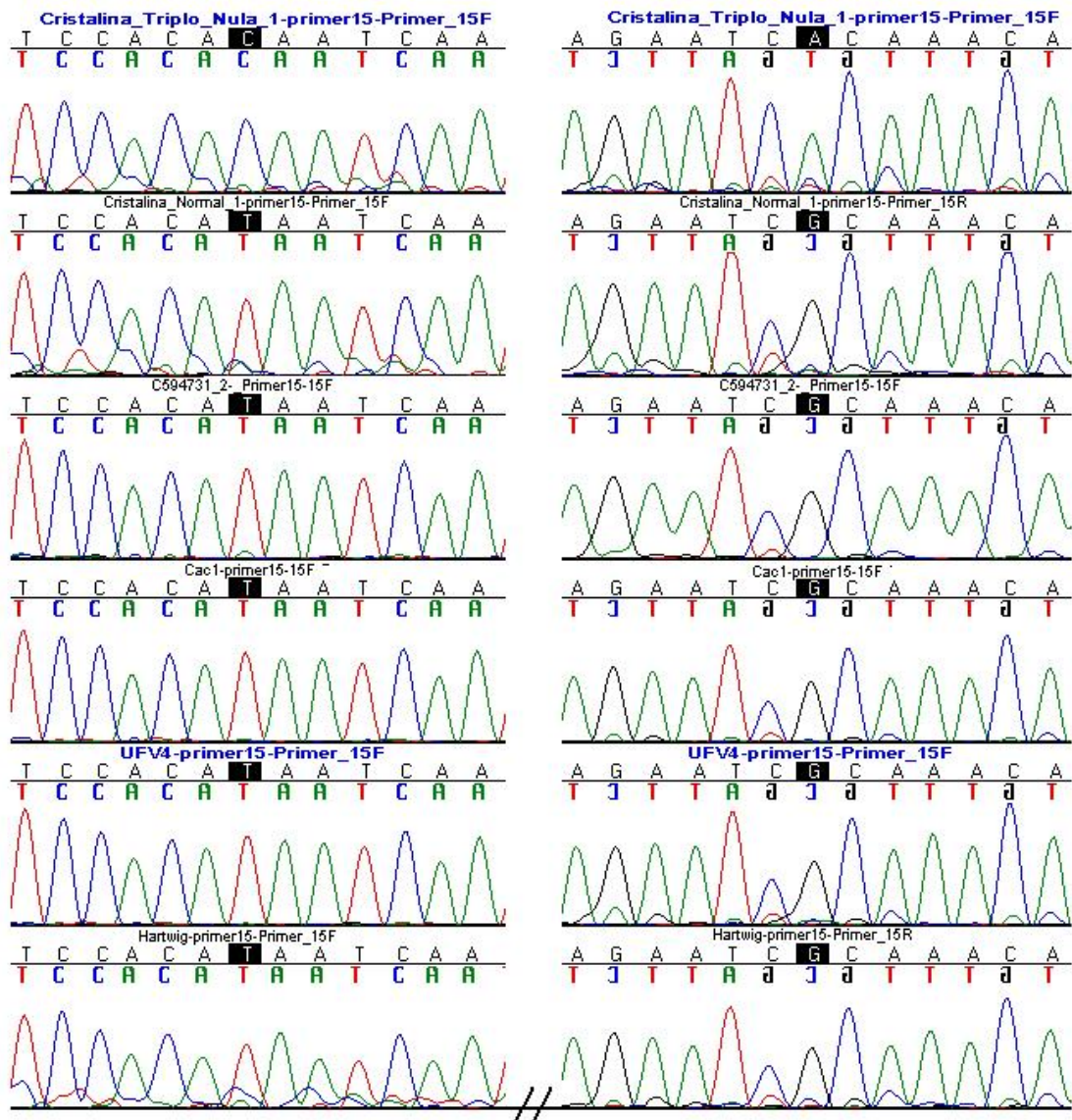


Figura 2 - Região dos cromatogramas correspondente às sequências das variedades (de cima para baixo): Cristalina TN, Cristalina, CS94731, CAC 1, UFV 16 e Hartwig evidenciando os haplótipos contidos no gene da *Lox2*. As duas barras entre as duas porções da figura, indica que parte da sequência foi retirada.

Pureunkong	1	5'	GGTGTCTGGGAATCCTGAACAGAGGAGGAGGGCATAAGATAAAAAGGGACAGTGGTGTGAT
Jinpumkong 2	1	5'	GGTGTCTGGGAATCCTGAACAGAGGAGGAGGGCATAAGATAAAAAGGGACAGTGGTGTGAT
Pureunkong	61		GAGGAAGAATGTGTTGGACTTCAACAGCGTGGCTGATCTTACTAAAGGAAATGTTGGGGG
Jinpumkong 2	61		GAGGAAGAATGTGTTGGACTTCAACAGCGTGGCTGATCTTACTAAAGGAAATGTTGGGGG
Pureunkong	121		ACTCATAGGCACCGGCCTCAACGTTGTTGGCTCAACACTTGATAACCTCACTGCTTTCTT
Jinpumkong 2	121		ACTCATAGGCACCGGCCTCAACGTTGTTGGCTCAACACTTGATAACCTCACTGCTTTCTT
Pureunkong	181		GGGCCGAAGTGTGCGCCCTACAGCTCATTAGTGCTACCAAACCTCTTGGTTCATTTCTTCT
Jinpumkong 2	181		GGGCCGAAGTGTGCGCCCTACAGCTCATTAGTGCTACCAAACCTCTTGGTTCATTTCTTCT
Pureunkong	241		TCCTTCCACATTAATCAATAACTTCTATATTCAAAATTAAGTGTTAATCTCTATACTCTC
Jinpumkong 2	241		TCCTTCCACATTAATCAATAACTTCTATATTCAAAATTAAGTGTTAATCTCTATACTCTC
			*
Pureunkong	301		ATTCATTTTCATTCAATGAAAAAATCATAAGACTTTTAACTAAAATTAACCTATGTAA
Jinpumkong 2	301		ATTCATTTTCATTCAATGAAAAAATCATAAGACTTTTAACTAAAATTAACCTATGTAA
Pureunkong	361		AGAATCGCAAACAAAAAATATATAATATTAAGTTTATTTACTTTTTTTTATAATGAC
Jinpumkong 2	361		AGAATCACAAACAAAAAATATATAATATTAAGTTTATTTACTTTTTTTTATAATGAC
			*
Pureunkong	421		AAAAAAATTATGTATATGGTGCACAAATTTTGTACTCTTTAAAAATATATCACTTTAT
Jinpumkong 2	421		AAAAAAATTATGTATATGGTGCACAAATTTTGTACTCTTTAAAAATATATCACTTTAT
Pureunkong	481		ACATAGCCAAACATATTTATTTTGTATAGTATTAACCTATTTGGGTACGTACCTTAATAA
Jinpumkong 2	481		ACATAGCCAAACATATTTATTTTGTATAGTATTAACCTATTTGGGTACGTACCTTAATAA
Pureunkong	541		TATTATTATGTGTGTATGTATGGTCTGTTTGTAGCAAATGGAAAAGGAAAAGTTGGAAAG
Jinpumkong 2	541		TATTATTATGTGTGTATGTATGGTCTGTTTGTAGCAAATGGAAAAGGAAAAGTTGGAAAG
Pureunkong	601		GGATACGTTCTTGAAGGGATTATGTGTGCTTACCAACTTTGGGAGCAGGGGAGTCTGC
Jinpumkong 2	601		GGATACGTTCTTGAAGGGATTATGTGTGCTTACCAACTTTGGGAGCAGGGGAGTCTGC
Pureunkong	661		ATTCAATATTCAGTTTGAATGGGACGAAAGCATGGGAATCCCCGGTGCCTTTACATAAA
Jinpumkong 2	661		ATTCAATATTCAGTTTGAATGGGACGAAAGCATGGGAATCCCCGGTGCCTTTACATAAA
Pureunkong	721		GAACATACATGCAAGTTGAGTTTACCTCAAGAGTTTGACTCTTGAAGACGTTCCAAACCA
Jinpumkong 2	721		GAACATACATGCAAGTTGAGTTTACCTCAAGAGTTTGACTCTTGAAGACGTTCCAAACCA
Pureunkong	781		AGGAACCATTGCTTTGTTTGCAA 3'
Jinpumkong 2	781		AGGAACCATTGCTTTGTTTGCAA 3'

Figura 3 - Alinhamento entre as sequências de cada fragmento de PCR derivado do gene *Lox2* das variedades Pureunkong and Jinpumkong 2. Os sítios de SNPs estão marcados com asterisco.