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

**Mapeamento associativo de locos relacionados à produtividade
de grãos em soja**

Mário Sérgio Sigrist

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

**Piracicaba
2012**

Mário Sérgio Sigrist
Licenciado e Bacharel em Ciências Biológicas

Mapeamento associativo de locos relacionados à produtividade de grãos em soja

Orientador:
Prof. Dr. **JOSÉ BALDIN PINHEIRO**

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

Piracicaba
2012

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

Sigrist, Mário Sérgio
Mapeamento associativo de locos relacionados à produtividade de grãos em soja /
Mário Sérgio Sigrist. - - Piracicaba, 2012.
87 p: il.

Tese (Doutorado) - - Escola Superior de Agricultura "Luiz de Queiroz", 2012.

1. Diversidade genética 2. Genética de populações 3. Germoplasma vegetal 4. Soja
I. Título

CDD 633.34
S578m

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

DEDICO

A minha família, pelo apoio e incentivo em todos os momentos da minha vida.

AGRADECIMENTOS

- Ao Dr. José Baldin Pinheiro por compartilhar sua experiência, pela orientação, amizade e apoio durante todo o curso e pelas estimáveis conversas sobre assuntos diversos.
- À pesquisadora, orientadora do mestrado e amiga Dr^a. Maria Imaculada Zucchi pela contribuição ao longo dos últimos anos para minha formação profissional.
- Ao Departamento de Genética da ESALQ/USP por permitir o uso de suas instalações e materiais para o desenvolvimento deste trabalho.
- À todos os professores da PG por transmitirem a ampla experiência que possuem em genética e melhoramento de plantas.
- Aos funcionários da PG: Léia, Berdan, Domingos de Sálvio Amaral e Márcio Araújo Silva pelo auxílio incondicional e pela convivência.
- À Regina Priolli, Natália Spagnol e Felipe Furlaneto de Souza pelo auxílio em etapas decisivas deste trabalho e pela estimada amizade.
- Aos amigos de longa data do Laboratório de Diversidade Genética e Melhoramento: Miklos Bajay, Marcelo Mattos Cavallari, Fernanda Raquel, Fátima Bosetti, Milene Moller, Michele dos Santos, Carlos Eduardo e Giuliana.
- À todos os colegas do curso de pós-graduação pelos momentos, discussões compartilhadas e prazerosa convivência.
- À CAPES e ao CNPq pela concessão das bolsas de estudo.
- À meus pais, Nilce e Sérgio, aos meus também pais, Luciano e Luiza, aos meus irmãos, Marcus e Alexandre, por serem as pessoas extraordinárias que sempre foram, pelo amor e por todo o incentivo. Sem vocês, não seria possível.
- Ao meu Deus, por não jogar dados.

SUMÁRIO

RESUMO	9
ABSTRACT	11
LISTA DE FIGURAS	13
LISTA DE TABELAS	15
1 INTRODUÇÃO	17
2 REVISÃO BIBLIOGRÁFICA	19
2.1 Soja.....	19
2.2 Aspectos econômicos.....	20
2.3 Base genética e germoplasma.....	21
2.4 Marcadores moleculares.....	22
2.5 Mapas genéticos em soja.....	23
2.6 Mapeamento associativo.....	25
3 MATERIAL E MÉTODOS	31
3.1 Material vegetal.....	31
3.2 Extração de DNA.....	31
3.3 Avaliação fenotípica.....	34
3.4 Avaliação genotípica.....	36
3.5 Análise dos dados.....	41
3.5.1 Dados fenotípicos.....	41
3.5.2 Dados genotípicos.....	43
3.5.3 Desequilíbrio de ligação.....	44
3.5.4 Mapeamento associativo.....	45
4 RESULTADOS	47
4.1 Avaliação agromorfológica.....	47
4.2 Avaliação genético-molecular.....	54
4.3 Divergência genética.....	56
4.4 Desequilíbrio de ligação.....	58
4.5 Mapeamento associativo.....	59
5 DISCUSSÃO	63
6 CONCLUSÕES	73
REFERÊNCIAS	75
ANEXOS	83

RESUMO

Mapeamento associativo de locos relacionados à produtividade de grãos em soja

O mapeamento associativo em soja tem sido recentemente utilizado como alternativa para o mapeamento de locos envolvidos no controle de características quantitativas em plantas. Dentre as vantagens, a técnica possibilita explorar a variabilidade genética disponível em bancos de germoplasma, permitindo maior resolução para a identificação de novos locos e alelos envolvidos no controle de características quantitativas. Um painel associativo composto por 89 genótipos provenientes de diversas regiões do mundo foi caracterizado em relação a 20 características fenotípicas e 114 marcadores moleculares microssatélites. Foram identificados 905 alelos, com média de 7,94 alelos/loco além de ampla variabilidade fenotípica, incluindo materiais mais produtivos em relação às testemunhas comerciais. Para associação entre marcadores e fenótipos, foi utilizada a abordagem de modelo linear misto (MLM), o qual incorpora informações de estrutura populacional e parentesco. Dos 114 marcadores utilizados, 78 foram responsáveis por 285 associações significativas com base nos diferentes métodos de correção para múltiplos testes, sendo 29 marcadores associados a uma única característica. As características produtividade de grãos, massa de 100 sementes, teor de óleo e teor de proteína foram associadas a 36 marcadores, sendo 30% destas associações previamente descritas na literatura. Algumas destas associações foram selecionadas para verificar o efeito alélico sobre o genótipo observado, gerando informações preliminares para a aplicação futura de seleção assistida por marcadores (SAM). O tamanho dos blocos de ligação foram estimados em ~20 cM ($r^2 < 0,05$) e ~2 cM ($r^2 < 0,1$), sugerindo a necessidade de 140 ou 1.300 marcadores para saturação completa do genoma no caso do painel associativo analisado. Os resultados obtidos deverão ser investigados futuramente a fim de confirmar as associações em diferentes populações.

Palavras-chave: *Glycine max*; Desequilíbrio de ligação; Germoplasma; Diversidade genética

ABSTRACT

Association mapping of loci related to soybean yield

Association mapping has recently been used in soybean as an alternative to linkage mapping of loci controlling quantitative traits. Among the advantages, the technique allows to explore the genetic diversity available in germplasm banks, which may result in greater resolution for the identification of new loci and alleles. An association panel comprising 89 genotypes from different regions of the world was characterized with using 20 phenotypic traits and 114 microsatellite markers. A total of 905 alleles were identified, with an average of 7.94 alleles / locus, besides a wide phenotypic variability, including more productive genotypes when compared to the commercial checks. Association between markers and phenotypes was performed using a mixed linear approach (MLM), which incorporates information regarding population structure and kinship. Among the 114 markers used, 78 were responsible for 285 significant associations based on different criteria to correct for multiple tests, and 29 markers were associated with a single trait. Seed yield, 100 seeds weight, oil and protein content were associated with 36 markers, with 30% of these associations previously reported in the literature. Some of these associations were selected to determine the allelic effects over the traits, in order to generate preliminary data for marker assisted selections (MAS). Estimated size of haplotype blocks were ~20 cM ($r^2 < 0.05$) and ~2 cM ($r^2 < 0.1$), suggesting that 140 or 1,300 markers would be necessary to cover the association panel's genome. Future studies should confirm marker-trait associations here found using different populations.

Keywords: *Glycine max*; Linkage disequilibrium; Germplasm; Genetic diversity

LISTA DE FIGURAS

Figura 1 - Variabilidade fenotípica para cor da vagem, semente e hilo dos genótipos avaliados (MULATO, 2009).....	33
Figura 2 - Visão geral do experimento para avaliação fenotípica dos acessos (Departamento de Genética - ESALQ).....	35
Figura 3 - Distribuição genômica dos 114 marcadores microssatélites utilizados no mapeamento associativo. Distância em centiMorgans (cM).....	37
Figura 4 - Perfis genético-moleculares e identificação alélica dos 89 acessos do painel associativo. A) Satt 038, B) Satt 146, C) Satt 174 e D) Satt 577. Diferentes cores representam os alelos presentes em cada loco.....	40
Figura 5 - Valores de Δk para cada valor de k , calculado de acordo com o proposto por Evanno et al. (2005). O maior valor de Δk corresponde ao número ótimo de subgrupos.....	55
Figura 6 - Teste de atribuição para os acessos de soja avaliados ($k=2$). Diferentes genótipos representados por barras verticais, sendo os de mesma cor pertencentes ao mesmo grupo. Diferentes cores em um mesmo indivíduo indicam a porcentagem do genoma compartilhado com cada grupo.	56
Figura 7 - Agrupamento hierárquico pelo método de Ward baseado em dados fenotípicos e de marcadores SSR.....	57
Figura 8 - Decaimento do desequilíbrio de ligação de acordo com a distância genética (cM) no painel associativo.....	58
Figura 9 - Médias das características relativas a cada alelo nos locos associados à produtividade de grãos, massa de 100 sementes, teor de óleo e proteína.....	69

Figura 10 - Decaimento do desequilíbrio de ligação em diferentes populações de soja.....	71
--	----

LISTA DE TABELAS

Tabela 1 - Lista dos materiais avaliados no mapeamento associativo conforme o país de origem.....	32
Tabela 2 - Características de alguns materiais incluídos no painel de mapeamento associativo.....	34
Tabela 3 - Análise de variância dos dados agromorfológicos do painel associativo.....	48
Tabela 4 - Médias dos 20 caracteres avaliados nos 89 genótipos de soja e os grupos estabelecidos pelo teste de comparação de médias de Scott-Knott (5%).....	49
Tabela 5 - Índices de correlação fenotípica (Pearson) entre as características avaliadas.....	52
Tabela 6 - Importância relativa das 20 características avaliadas através da análise por variáveis canônicas.....	53
Tabela 7 - Número de alelos e conteúdo de polimorfismo informativo (PIC) dos 114 locos utilizados no mapeamento.....	54
Tabela 8 - Quantidade total de marcadores associados a cada uma das 20 características.....	60
Tabela 9 - Marcadores associados aos principais características quantitativas avaliadas.....	61
Tabela 10 - Comparação das associações obtidas por mapeamento associativo e por ligação publicados no <i>SoyBase</i> . Em negrito, destaque para os marcadores associados à mesma característica em ambas as abordagens.....	67

1 INTRODUÇÃO

A safra brasileira de soja no ano agrícola 2011/12 foi de 66,3 milhões de toneladas em uma área total de 25 milhões de hectares, consolidando o país como o segundo maior produtor e exportador mundial (COMPANHIA NACIONAL DE ABASTECIMENTO - CONAB, 2012). Mais da metade das exportações do complexo são destinados à China, país cuja demanda pelo grão e seus derivados tende a aumentar nos próximos anos devido às previsões de crescimento populacional e de renda. Embora o Brasil seja um dos poucos países no mundo com possibilidade de aumento da área de produção, diversos setores da sociedade exigem uma maior sustentabilidade da agricultura e menor pressão sobre os ecossistemas naturais. Consequentemente, o país tem como principal desafio para os próximos anos buscar o aumento de produtividade da cultura, atualmente entre 2,5 e 3 t/ha, para atender a demanda mundial pela *commodity*.

Nas últimas décadas, o emprego de novas tecnologias no melhoramento genético vegetal tem proporcionado inúmeros avanços neste sentido, como a disseminação do uso de marcadores moleculares e a identificação dos genes responsáveis por diversas características agrônômicas. De maneira geral, muitas dessas características são controladas por locos de efeito quantitativo (*Quantitative Trait Loci - QTL's*), sendo que o mapeamento dessas regiões é fundamental para que sejam atingidos patamares superiores de produtividade. O mapeamento genético busca estimar quantos e quais são os QTL's responsáveis pela variação nas características, localizar sua posição no genoma, estimar seus efeitos e identificar suas inter-relações. Em última análise, o mapeamento genético possibilita a seleção assistida em programas de seleção (SAM) e a clonagem gênica, reduzindo o tempo e recursos despendidos no desenvolvimento de novas variedades.

O mapeamento de locos que controlam caracteres quantitativos tem sido frequentemente realizado através da análise de ligação, considerando a variação genética presente em progênies de cruzamentos biparentais. Neste caso, somente dois alelos por loco podem ser estudados simultaneamente. Tal abordagem limita o número de recombinações possíveis, o que tende a diminuir a resolução do mapa. Ainda, a necessidade de cruzamentos controlados para obtenção da população de mapeamento pode ser onerosa dependendo da espécie em questão.

Visando superar estes problemas, outros métodos foram propostos para o desenvolvimento de mapas genéticos e análise de regiões genômicas relacionadas a determinado fenótipo. O mapeamento associativo, também conhecido como mapeamento por

desequilíbrio de ligação (DL) baseia-se na detecção da variação fenotípica e genética presente em populações não estruturadas em famílias. Esta metodologia apresenta como vantagens a exploração de uma maior variabilidade genética na população de mapeamento, a qual pode ser composta por indivíduos provenientes de populações naturais ou por acessos de bancos de germoplasma. Nestes casos, espera-se uma maior resolução na identificação de regiões genômicas de interesse, dado que diversas gerações de recombinação são avaliadas.

Para que estudos de associação sejam efetivos, são necessárias informações prévias referentes à população em questão. O conhecimento acerca da estrutura populacional é um dos mais relevantes, de modo a evitar associações espúrias entre marcadores e fenótipo. Ainda, a extensão do desequilíbrio de ligação e a existência e estrutura de blocos de haplótipos exercem papel central em análises de associação. Tais informações determinarão o tipo, o número e a densidade de marcadores a serem utilizados, além dos métodos mais adequados para análise e interpretação dos resultados.

O uso do mapeamento associativo permanece incipiente na cultura da soja, sendo que a maioria dos estudos limita-se a avaliar a estrutura populacional e o grau de desequilíbrio de ligação em diferentes populações. Recentemente, a técnica foi utilizada para a detecção de QTL's responsáveis pelos teores de proteína e óleo, qualidade de grãos e pela resposta à deficiência em micronutrientes. O emprego do mapeamento associativo para identificação de locos responsáveis pela produtividade de grãos carece de estudos mais detalhados, que comparem e validem os resultados obtidos com outras técnicas de mapeamento. Uma vez que a técnica permite identificar alelos úteis para produtividade de grãos em fenótipos desfavoráveis (ex. PI's – *Plant Introductions*), espera-se que o mapeamento associativo promova o uso dos recursos genéticos e amplie a base genética disponível para melhoramento.

Considerando tais aspectos e possibilidades, o presente trabalho tem como objetivo empregar a metodologia para identificação de novos QTL's e alelos relacionados à produtividade em soja, permitindo orientar cruzamentos em programas de hibridação e promover o uso da seleção assistida por marcadores no melhoramento da cultura.

2 REVISÃO BIBLIOGRÁFICA

2.1 Soja

A soja (*Glycine max*, $2n = 40$) é uma planta pertencente à ordem Fabales, família Fabaceae, subfamília Papilionoideae, tribo Phaseoleae, gênero *Glycine*. Tendo como centro de origem a China, a soja foi primariamente domesticada entre as latitudes 30 e 45°N e posteriormente disseminada para a América do Norte, Europa e América do Sul. A espécie é essencialmente autógama, com flores perfeitas e taxa de polinização cruzada menor que 1% (BORÉM; MIRANDA, 2005).

A planta é herbácea, anual e de crescimento determinado, semideterminado ou indeterminado, com altura variando entre 0,3 a 2 metros. Conforme o cultivar e condições ambientais, o ciclo varia de 80 a 200 dias (SEDIYAMA et al., 1999). Uma planta pode chegar a produzir até 400 vagens, sendo que de maneira geral desenvolvem entre 30 a 80 vagens (CÂMARA, 1998). O número de flores produzidas é maior do que o convertido em vagem, devido à variação na taxa de fertilização.

Evidências atuais indicam que o processo de domesticação ocorreu a partir de seu parente silvestre *Glycine soja* na região da China, há aproximadamente 5.000 anos. Como em diversas espécies, os processos de domesticação e seleção promoveram gargalos genéticos ao longo das gerações que, aliado a um processo de domesticação descentralizado, resultou no surgimento de uma infinidade de raças locais de *G. max* na região da Ásia Oriental. Apesar da ampla variação genética, menos de 0,02% do total destas raças locais foram introduzidas para formar a base genética dos programas de melhoramento americanos, durante a primeira metade do século XX (HYTEN et al., 2006). Esta introdução de poucos materiais para formação dos programas de melhoramento ocidentais resultou em um efeito gargalo sobre a variabilidade disponível, aumentando a ação da deriva genética e colaborando com o estreitamento da base genética da cultura.

No Brasil, a primeira referência de plantio experimental de soja data do XIX, quando da introdução do cultivo no Estado da Bahia. Contudo, somente após 1960 a cultura passou a adquirir importância econômica, inicialmente na região Sul devido à melhor adaptação. Com a crescente demanda por matéria-prima protéica nos países desenvolvidos, observou-se o aumento da produção e a rápida expansão da área de cultivo rumo à região Centro-Oeste. Esta disseminação da cultura para regiões de baixas latitudes tornou o Brasil um dos maiores produtores mundiais da leguminosa, fato obtido graças ao melhoramento genético da cultura

visando obter cultivares de período juvenil longo (ALMEIDA et al., 1999). Contudo, semelhante ao caso americano, a base genética do germoplasma cultivado no Brasil é considerada, tendo como origem poucos materiais oriundos da região sul dos Estados Unidos (HIROMOTO; VELLO, 1986).

2.2 Aspectos econômicos

No mundo, o agronegócio da soja movimenta aproximadamente 215 bilhões de dólares. O Brasil é o segundo maior produtor mundial, com área cultivada na safra 2011/12 de 25 milhões de hectares e produção de 66 milhões de toneladas (CONAB, 2012). Os Estados Unidos, maior produtor mundial com uma área cultivada de quase 31 milhões de hectares e 91 milhões de toneladas produzidas, são responsáveis por 38% de toda a produção mundial (UNITED STATES. DEPARTMENT OF AGRICULTURE - USDA, 2010).

No Brasil, o complexo agroindustrial da soja movimenta cerca de 30 bilhões de dólares por ano e tem grande peso na balança comercial, contribuindo com 19% das exportações do agronegócio e 7% das exportações totais do país. A produção nacional, safra 2010/11, foi de 75 milhões de toneladas, sendo a produtividade média do País de 3.115 kg/ha. As maiores produtividades foram verificadas nos Estados do Paraná e Bahia, sendo que a região Centro-Oeste lidera o *ranking* da produção com 34 milhões de toneladas (CONAB, 2012). Na safra 2011/12 houve redução de 12% da produção em relação à safra anterior, principalmente devido à estiagem na região Sul decorrente do fenômeno “La Niña” (CONAB, 2012).

A soja é uma das principais fontes de proteína e óleo vegetal do mundo, sendo cultivada comercialmente e utilizada na alimentação humana e animal. A planta de soja e seus derivados provenientes do processamento industrial dos grãos fornecem ainda matéria-prima para adubação verde e para produção de materiais plásticos e de biodiesel. No Brasil, a soja é predominantemente utilizada para o processamento do grão em óleo e proteína, sendo a proteína processada (torta ou farelo) utilizada na suplementação de ração animal e o óleo como principal fonte para fabricação biodiesel e diversas aplicações na indústria alimentícia.

O óleo de soja é composto principalmente pelos ácidos graxos palmítico, esteárico, oléico, linoléico e linolênico. Diferentes concentrações destes ácidos graxos implicam em diferentes aplicações do óleo, sendo que composição do óleo é controlada geneticamente e passível de ser selecionada (LI et al., 2002).

2.3 Base genética e germoplasma

Estudos sobre a variabilidade genética da cultura destacam que o germoplasma brasileiro possui uma base estreita, originada a partir de poucas linhagens ancestrais (HIROMOTO; VELLO, 1986). O estreitamento da base genética representa uma limitação para o desenvolvimento de programas de melhoramento, como indisponibilidade de alelos para enfrentar estresses bióticos e abióticos, menor adaptabilidade e limitação do potencial produtivo. A uniformidade genética das cultivares é também corroborada por estudos de variabilidade e parentesco empregando marcadores moleculares. (PRIOLLI et al., 2002; HYTEN et al., 2006).

Nos EUA, estima-se que 14 genótipos ancestrais contribuíram com mais de 80% da composição dos cultivares comerciais, situação semelhante ao verificado no germoplasma brasileiro. Em estudo recente, identificou-se a participação de 60 genótipos ancestrais no desenvolvimento de 444 cultivares brasileiras, sendo que os quatro principais ancestrais (CNS, S-100, Nanking e Tokyo) contribuíram com 55,26% da base genética do conjunto de materiais (WYSMIERSKI, 2010).

A maior parte da variabilidade genética da soja é mantida em bancos de germoplasma existentes em diversos países, sendo as principais coleções localizadas nos Estados Unidos, China e Japão. No Brasil, em 2007, a coleção de germoplasma de soja contava com aproximadamente 5.000 acessos caracterizados e disponíveis. A partir do referido ano, iniciou-se o processo de introdução de materiais provenientes da coleção do Banco de Germoplasma dos Estados Unidos (USDA), composta por 22 mil acessos, visando enriquecer a coleção brasileira. Contudo, apesar dos esforços, a variabilidade genética para caracteres fisiológicos, morfológicos e agrônômicos, permanece pouco utilizada pela maioria dos programas de melhoramento. Este fato decorre, principalmente, da preferência dos melhoristas por materiais mais produtivos e adaptados para uso em cruzamentos, colaborando com o estreitamento da base genética (NASS, 2001; NASS; SIGRIST, 2009). Para superar estas limitações, tem sido proposto o desenvolvimento de programas de pré-melhoramento, fundamentados na caracterização do germoplasma exótico e posterior introdução de genes de interesse em materiais adaptados. O processo de caracterização pode empregar desde técnicas multivariadas até o uso de cruzamentos com linhagens testadoras para avaliação de desempenho.

Considerando que o comportamento fenotípico de materiais exóticos pode não ser o mais adequado para avaliar seu potencial, novas abordagens para uso destes materiais tem

sido propostas. De acordo com Tanksley e Maccouch (1997), embora os materiais exóticos sejam menos produtivos, ainda assim podem apresentar alelos úteis em determinados QTL's. O uso de marcadores moleculares e de mapas genéticos tem auxiliado neste sentido, permitindo uma avaliação mais precisa da contribuição de alelos exóticos em programas de melhoramento (KABELKA et al. 2004). Após ampla caracterização, a introdução de alelos úteis pode ser então realizada por meio de retrocruzamentos ou pelo uso de seleção recorrente para posterior extração de linhagens (HIROMOTO; VELLO, 1986).

Devido ao potencial de contribuição dos materiais exóticos como fonte de novos alelos para o melhoramento genético, novas abordagens e técnicas para estudo e utilização destes materiais são essenciais para que novos patamares de produtividade sejam alcançados.

2.4 Marcadores moleculares

O uso de marcadores moleculares no melhoramento genético tem sido amplamente realizado devido a diversas características desejáveis, dentre as quais se destacam: alta taxa de polimorfismo detectada, ausência de influência ambiental, são herdáveis, apresentam baixo custo por dado gerado, são distribuídos em todo o genoma e o processo de análise pode ser, em alguns casos, automatizado (FERREIRA et al., 2007). A automatização e o desenvolvimento de plataformas de genotipagem tem sido uma tendência atual, uma vez que permitem diminuir o custo por dado gerado, além de permitir uma ampla cobertura genômica em menor tempo.

Os marcadores moleculares têm sido utilizados em análise genética com as mais diversas finalidades, tais como: identificação de clones, linhagens, híbridos, cultivares, estudos de fluxo gênico, estimativas de taxas de cruzamento e parentesco, análise da diversidade genética entre indivíduos, construção de mapas genéticos e na seleção assistida por marcadores (SOUZA, 2001). Além disso, os marcadores moleculares se tornaram ferramentas que permitem a clonagem de genes úteis no melhoramento transgênico (FERREIRA; GRATTAPAGLIA, 1998; BUSO et al., 2003). Existe atualmente uma infinidade de tipos de marcadores moleculares, cuja escolha dependerá de uma série de fatores, relacionados ao objetivo da pesquisa, recursos financeiros e infraestrutura disponível (CIAMPI et al., 2007).

Dentre os marcadores moleculares empregados no mapeamento genético, uma das técnicas mais disseminadas é a utilização de locos microssatélites, também denominados *Simple Sequence Repeats* (SSR) (LITT; LUTY, 1989). Os SSRs apresentam características

altamente desejáveis: são marcadores codominantes, encontram-se ampla e uniformemente distribuídos nos genomas eucariotos, são multialélicos e amplificados via PCR. Uma vez desenvolvidos os *primers* que amplificam tais regiões do genoma, estes podem ser facilmente compartilhados entre laboratórios. Ainda, *primers* marcados com fluorescência possibilitam sistemas multiplex semiautomatizados, o que permite avaliar rapidamente um grande número de indivíduos para um grande número de locos em pouco tempo. Em soja, uma infinidade de locos SSR já foram desenvolvidos e mapeados ao longo do genoma (CREGAN et al., 1999; SONG et al., 2004, HWANG et al., 2009). As informações referentes a estes marcadores são de domínio público e encontram-se disponíveis para comunidade científica na base de dados SoyBase (<http://www.soybase.org>).

Na última década, novas classes de marcadores foram desenvolvidas, principalmente visando obter uma maior cobertura genômica, em menor tempo e com menor custo. Dentre estes, destacam-se os SNP's (*Single Nucleotide Polymorphism*), encontrados abundantemente ao longo dos genomas. Os marcadores SNP's tem como base as mutações em base única na cadeia de DNA, trocas ou inserções / deleções, sendo as mais comuns as trocas entre purinas (transições). Os SNP's são marcadores codominantes e bialélicos, podendo ocorrer em regiões codificadoras ou de regulação, porém na maior parte das vezes são encontrados em espaços intergênicos. A principal vantagem destes marcadores em relação ao demais reside no fato de serem a forma mais abundante de variação genômica e apresentarem ampla capacidade de automação (FAN et al., 2003, 2006). Estima-se que o número de SNP's na soja cultivada seja da ordem de 4-5 milhões, considerando uma taxa de 280 SNP's verificados em 76,3 kpb (ZHU et al., 2003). Uma fase inicial de descobrimento de SNP's em soja resultou na identificação de 5.551 marcadores (CHOI et al., 2007) e posteriormente em outros 12.000 adicionais (HYTEN et al., 2008). Esta quantidade de marcadores possibilita uma maior resolução para o mapeamento de QTL's, potencializando a ligação com polimorfismos responsáveis pela variação fenotípica.

2.5 Mapas genéticos em soja

O desenvolvimento de mapas genéticos é considerado uma das aplicações de maior impacto da tecnologia de marcadores moleculares. O mapeamento genético tem por base a correlação entre a distância entre dois genes e a frequência de recombinação entre eles. Quanto menor a distância entre dois locos (ou marcadores), menor será a probabilidade de

ocorrer *crossing-over*. A distância é usualmente expressa em centiMorgans (cM), sendo calculadas através das frequências de recombinação e funções de mapeamento (LIU, 1998).

No melhoramento de plantas, os mapas genéticos possibilitam a análise completa dos genomas, a decomposição de características genéticas complexas nos seus componentes mendelianos, a localização das regiões genômicas que controlam caracteres de importância, a quantificação do efeito destas regiões na característica estudada e a canalização de toda esta informação para uso em programas de melhoramento. Três requisitos básicos são necessários para o seu desenvolvimento: reprodução sexuada, produção de descendentes e uma fonte de marcadores moleculares com comportamento mendeliano (FERREIRA; GRATTAPAGLIA, 1998). Dentre os métodos de mapeamento clássicos, destacam-se: a análise de marcas simples, o mapeamento por intervalo, o mapeamento por intervalo composto e o mapeamento de múltiplos intervalos (COLLARD et al., 2005). O método de análise de marcas simples normalmente utiliza modelos lineares, enquanto que os demais empregam a máxima verossimilhança, com base na informação da distribuição conjunta dos marcadores e QTL's (LIU, 1998).

O primeiro mapa genético da soja empregando marcadores moleculares foi desenvolvido por Keim et al. (1990). Empregando 150 marcadores RFLP (*Restriction Fragment Length Polymorphism*), foram detectados 26 grupos de ligação em uma população F₂ oriunda do cruzamento entre *G. max* e *G. soja*. Posteriormente, outros estudos foram publicados utilizando, além de RFLP's, marcadores morfológicos e izoenzimáticos (LARK et al., 1993; SHOEMAKER; SPECHT, 1995). Apesar do avanço proporcionado por estes estudos, marcadores do tipo RFLP não são os mais indicados para o mapeamento devido ao baixo polimorfismo e a complexidade em avaliar múltiplas bandas nas sondas, o que pode ter gerado a falta de correlação entre o número de grupos de ligação e o número cromossômico da espécie. Consequentemente, os marcadores SSR foram considerados como os mais apropriados para o mapeamento (AKKAYA et al., 1992)

Em um trabalho de referência no mapeamento genético em soja, Cregan et al. (1999) desenvolveram e mapearam 606 locos SSR, além de outros marcadores, em três diferentes populações de cruzamentos biparentais. O consenso entre estas populações resultou, pela primeira vez, em um mapa contendo os esperados 20 grupos de ligação e comprimento de 2.400 cM. Este mesmo grupo de pesquisa, posteriormente, desenvolveu novos marcadores SSR a partir de bibliotecas BAC (*Bacterial Artificial Chromosome*) e de sequências expressas armazenadas em bancos de dados (EST – *Expressed Sequence Tags*), visando uma

maior saturação dos grupos de ligação e mapeamento de genes de interesse (SONG et al., 2004).

Atualmente, mais de 90 características já foram mapeadas em soja, incluindo caracteres relacionados ao desenvolvimento e reprodução, resistência a insetos-praga e doenças, qualidade da semente e aspectos nutricionais (YAMANAKA et al., 2001; ZHANG et al., 2004; CHOI et al., 2007; HYTEN et al., 2007a; LI et al., 2007). A identificação de polimorfismo e genotipagem em soja, assim como nas grandes culturas agrônomicas, têm focado novas técnicas passíveis de automação e obtenção de dados em larga escala (GUPTA et al., 2008). O uso de SNP's, aliado a plataformas de microarranjo (*microarrays*), permitem uma ampla cobertura do genoma para identificação QTL's. Deste modo, a resolução da nova geração de mapas genéticos tenderá ao nível de QTN's (*Quantitative Trait Nucleotide*), onde alterações de apenas uma base podem ser responsáveis por alterações fenotípicas (HYTEN et al., 2010). Mais recentemente, novas fronteiras poderão ser exploradas com o sequenciamento do genoma da cultura, permitindo integrar os densos mapas genéticos disponíveis com os polimorfismos causadores do fenótipo (SCHMUTZ et al., 2010).

Conforme mencionado, o mapeamento genético via análise de ligação é o método rotineiramente utilizado para identificação de QTL's em plantas. Contudo, este procedimento apresenta como limitações o custo e tempo necessários para síntese de populações de mapeamento, a baixa resolução (10 a 20 cM) e o fato de somente dois alelos por loco poderem ser estudados em progênies de cruzamentos biparentais, limitando a exploração da diversidade alélica (FLINT-GARCIA et al., 2003).

2.6 Mapeamento associativo

O mapeamento associativo avalia o desequilíbrio de ligação entre polimorfismos no DNA e os genes responsáveis por caracteres de interesse. Além da utilidade em mapear QTL's, o mapeamento associativo pode também identificar mutações em uma única base, responsáveis por fenótipos específicos (PALAISA et al., 2004). Uma vez que o mapeamento por associação considera todas as gerações de recombinação prévias à origem da mutação, as regiões alvo no genoma tendem a ser menores em relação ao mapeamento convencional, aumentando a resolução dos mapas (GUPTA et al., 2005). Havendo desequilíbrio de ligação entre o marcador e um loco alvo, então alelos específicos ou haplótipos (combinação de genótipos em grupos de marcadores ligados) podem ser associados a valores fenotípicos com alto nível de significância (CARDON; BELL, 2001).

O mapeamento associativo foi inicialmente proposto para mapear regiões genômicas relacionadas com importantes doenças humanas, sendo posteriormente empregado para o mapeamento de genes de interesse em culturas agrônomicas (PRITCHARD; PRZEWORSKI, 2001; GUPTA et al., 2005). Dois delineamentos principais são utilizados nestes estudos: casos controle e baseados em famílias. Em casos controle, as frequências de marcadores são determinadas em um grupo de indivíduos afetados para determinada característica e comparados com um grupo não-afetado. O delineamento baseado em famílias, também conhecido como teste de transmissão de desequilíbrio, avalia trios familiares compostos por um parental heterozigoto e um filho afetado, buscando correlações na transmissão desigual de alelos (ORIGUZIE et al., 2007). Em plantas, a metodologia usual envolve o uso de populações não estruturadas (sem progênes), compostas de indivíduos não aparentados provenientes de diferentes *backgrounds* genéticos, histórias de seleção e origens, apresentando ampla variação para a característica alvo (ABDURAKHMANOV; ABDUKARIMOV, 2008).

Considerado como ferramenta para maior refinamento dos mapas genéticos, obtidos através do mapeamento convencional de QTL's, o mapeamento associativo inicialmente restringiu-se à análise de regiões específicas do genoma (genes candidatos). Melhorias das técnicas de sequenciamento e automação da genotipagem empregando plataformas específicas permitiram uma maior cobertura genômica e, conseqüentemente, os estudos associativos passaram a avaliar todo o genoma da espécie em questão (*whole genome scanning*). Os primeiros estudos em plantas buscaram avaliar espécies modelo como *Arabidopsis* e milho, porém atualmente a metodologia tem se expandido rapidamente para outras espécies (ORIGUZIE et al., 2007; GUPTA et al., 2005).

Em soja, Wang et al. (2008) utilizaram o mapeamento associativo para identificar locos responsáveis pela resposta à deficiência de ferro. Foram utilizados 84 marcadores SSR em duas populações. Foram encontrados três marcadores associados ao controle da característica, sendo que dois deles foram confirmados em ambas as populações. Além da confirmação dos QTL's encontrados comparando com resultados da literatura, o uso de duas ou mais populações é importante para confirmação dos QTL's encontrados.

Outro foco de estudos associativos em soja é o teor de proteína e óleo nos grãos (JUN et al. 2007, SHI et al., 2010) e aspectos relacionados a qualidade nutricional em variedades de soja destinadas à alimentação humana (HOU et al., 2011). Nestes estudos foram encontrados novos locos e alelos para a característica em questão, fornecendo importantes subsídios para programas de melhoramento.

Os principais objetivos do melhoramento genético em soja, a produtividade de grãos e ciclo de maturação, foram estudados por meio de mapeamento clássico de QTL's (SMALLEY et al., 2004; GUZMAN et al., 2007). Contudo, considerando os resultados dos estudos anteriores, espera-se que o mapeamento por desequilíbrio de ligação (DL) seja uma alternativa promissora para identificação de novos locos e alelos para estas características. A abordagem permite verificar a existência de alelos úteis em acessos de germoplasma pouco adaptados, nos quais o fenótipo pode ser desinteressante para os melhoristas em um primeiro momento. Para que estudos futuros envolvendo estas características sejam efetivos, alguns fatores devem ser levados em conta.

De modo geral, o primeiro aspecto a ser avaliado é a extensão e distribuição do DL, uma vez que é necessária para a associação entre marcador e fenótipo e porque a extensão destas regiões determinará a resolução do mapeamento. O DL é definido como a associação não-aleatória de alelos em diferentes locos, o que termina por alterar a frequência de haplótipos na população teste (FLINT-GRACIA et al., 2003). Diversos fatores influenciam o DL verificado em uma população ou espécie: taxa de mutação e recombinação, sistema de cruzamento, deriva genética, estrutura populacional, grau de parentesco, seleção, epistasia e rearranjos cromossômicos (ABDURAKHMANOV; ABDUKARIMOV, 2008). Estes fatores podem gerar falsas associações entre marcadores e fenótipo no mapeamento associativo, de modo que devem ser consideradas nas análises.

Em situações onde o DL é alto, o mapeamento amplo do genoma é facilitado, pois exige uma menor quantidade de marcadores, apesar da limitada resolução. Nos casos onde a extensão do desequilíbrio é baixa, a melhor opção seria a análise de genes candidatos, pois o mapeamento amplo do genoma exigirá um grande número de marcadores, ampliando os custos. Assim, torna-se necessário entender o padrão do DL para identificar a melhor estratégia a ser adotada. Uma estratégia alternativa é o desenvolvimento de populações com níveis variáveis de DL, permitindo abordagens complementares para uma melhor resolução dos mapas genéticos (RAFALSKI; MORGANTE, 2004).

Existem diversos índices propostos para mensuração do DL. Essencialmente, estas medidas quantificam o desequilíbrio como sendo a diferença entre a frequência observada do haplótipo e a frequência esperada no equilíbrio, assumindo independência (BENNET; BINET, 1956). Dentre os índices mais empregados em estudos de DL encontram-se o D' e r^2 , sendo que o primeiro consiste em uma tentativa de parametrizar o DL e o segundo pode ser interpretado como um coeficiente de determinação (GUPTA et al., 2005). De acordo com Origuzie et al. (2007), D' é útil para comparações entre locos com diferentes frequências, ao

passo que r^2 é uma medida mais confiável em casos onde existem alelos raros ou pouco frequentes na população. Em um primeiro momento, a medida r^2 é a mais indicada como indicação geral do desequilíbrio, sendo D' mais restrito ao mapeamento em alta resolução, após associações iniciais terem sido detectadas.

O cálculo do DL entre pares de marcadores é usualmente realizado por meio de processos iterativos, visando obter estimativas de máxima verossimilhança por meio dos algoritmos de Clark ou algoritmo EM (*expectation maximization*), embora outros já tenham sido propostos (EXCOFFIER; SLATKIN, 1995; STEPHENS; DONNELLY, 2003). Recentemente, abordagens considerando a informação de diversos locos simultaneamente têm sido propostas, uma vez que serão importantes futuramente na elaboração de mapas de DL de genomas inteiros (GUPTA et al., 2005).

A visualização dos padrões de DL pode então ser feita graficamente, demonstrando valores de r^2 (ou D') significativos (p-valor) entre pares de marcadores. Para visualização do decaimento do desequilíbrio e estimativa do tamanho dos blocos de ligação, os valores de r^2 podem ser relacionados com a distância genética entre os marcadores (cM ou pares de base). Neste caso, a distância na qual o valor de r^2 for menor que 0,1 (ou $D' < 0,5$) é considerada como estimativa do tamanho do bloco de ligação (FLINT-GRACIA et al., 2003).

Remington et al. (2001) mediram o DL no genoma de milho (*Zea mays* L.) empregando 47 marcadores SSR, indicando um rápido decaimento do desequilíbrio (12 kb) nas proximidades do loco *su1*. Este estudo também sugeriu que marcadores SSR foram mais eficientes em relação aos SNPs para determinar a estrutura populacional, uma vez que maiores níveis de LD foram detectados entre os SSR, os quais são considerados evolucionariamente mais antigos. Em *Arabidopsis thaliana*, o DL começa a decair dentro de 50 kb, embora persista por mais de 250 kb em determinadas regiões genômicas (NORDBORG et al., 2002). Em cevada, amplos DL se estendem a distâncias de até 50 cM com $r^2 > 0,05$ ou até 10 cM com $r^2 > 0,2$ (MALYSHEVA-OTTO et al., 2006). Em soja, Hyten et al. (2007b) reportaram DL variando entre 90 kb até 574 kb em regiões genômicas de três populações, sendo o menor DL encontrado em acessos de *G. soja*. Estes resultados reforçam a variação do DL em diferentes populações, de modo que podem ser indicadas diferentes abordagens para mapeamento associativo em uma mesma espécie.

Uma vez avaliado o grau de DL presente na população, o próximo passo consiste na associação propriamente dita entre marcadores e fenótipo por meio de testes de associação. O delineamento clássico utilizado no mapeamento associativo consiste em estudos de caso-controle, onde dois grupos de indivíduos discrepantes para a característica em questão são

comparados. Contudo, estes delineamentos têm como principal limitação a ocorrência de estrutura populacional, o que pode levar a falsas associações entre marcador e fenótipo. Visando eliminar este problema, foram propostos diversos delineamentos baseados em testes de transmissão de desequilíbrio (TDT, da tradução em inglês), o qual compara a transmissão ou não-transmissão de alelos para a progênie. Embora contorne o efeito da estrutura populacional, o TDT é um método menos robusto quando comparado ao caso-controle (MACKAY; POWELL, 2007; ABDURAKHMANOV; ABDUKARIMOV, 2008).

Consequentemente, diversas abordagens foram propostas para controle da estrutura populacional em delineamentos caso-controle ou de amostras aleatórias: controle genômico (DEVLIN; ROEDER, 1999), associação estruturada (PRITCHARD et al., 1999, 2000), componentes principais (PRICE et al., 2006) e modelos mistos (YU et al., 2006). A abordagem mais comum para testar a presença de estrutura populacional em testes de associação tem sido a associação estruturada, embora não leve em conta o possível parentesco entre indivíduos e o número correto de subpopulações seja complexo de estimar. Neste sentido, uma saída promissora parece ser o uso de componentes principais e informações de parentesco incorporadas em modelos lineares mistos (ZHAO et al., 2007; ZHU et al., 2008).

Finalmente, o poder estatístico do mapeamento associativo (probabilidade de encontrar associações verdadeiras) deve ser estimado, de modo a eliminar erros do tipo I. Dentre os métodos estatísticos rotineiramente utilizados encontram-se os testes de permutação, *false discovery rate* (FDR) e a determinação do fator *Bayes* (BENJAMINI; HOCHBERG, 1995; BALL, 2005; BALDING, 2006).

3 MATERIAL E MÉTODOS

3.1 Material vegetal

Um conjunto de 81 acessos de soja (PI – *Plant Introduction*) foi selecionado para composição do painel de associação com base em sua origem geográfica, representando diversas regiões dos diferentes continentes. Além destes, foram avaliadas oito testemunhas comerciais, totalizando 89 genótipos (Tabela 1). As sementes foram germinadas em vasos e as plântulas conduzidas em casa de vegetação para obtenção de material suficiente para genotipagem e fenotipagem.

A escolha de acessos provenientes de diferentes regiões geográficas e de diferentes grupos de maturação teve por intuito diminuir o grau de estrutura genética, diminuindo assim a quantidade de falsos positivos no mapeamento associativo. Mais além, a ampla origem dos materiais tende a maximizar a variabilidade genética do painel, o que em última análise possibilita a identificação de novos alelos envolvidos no controle das características avaliadas (Figura 1 e Tabela 2).

3.2 Extração de DNA

Aproximadamente 20 dias após a germinação, folhas de cinco plantas de cada genótipo foram colhidas em *bulk*, e o DNA extraído pelo protocolo CTAB (DOYLE; DOYLE, 1990). A qualidade e concentração do DNA extraído foram avaliadas por eletroforese em gel de agarose (1%) corados com SYBR Safe (Invitrogen) e suas concentrações estimadas pela intensidade de fluorescência emitida, sendo a quantificação do DNA extraído realizada comparando-se esta intensidade com a de concentrações pré-estabelecidas do DNA do fago λ . A extração de todos os materiais foi realizada em duplicata, visando garantir quantidade suficiente de DNA para o trabalho. Uma vez quantificado, o DNA de todos os acessos foi diluído a 10 ng/ μ l para padronização das reações de amplificação (PCR).

Tabela 1 - Lista dos materiais avaliados no mapeamento associativo conforme o país de origem

Id	Genótipo	Origem	Id	Genótipo	Origem
1	PI36906	Manchúria	46	PI315701	EUA
2	PI79861	China	47	PI322695	Angola
3	PI84910	Coréia do Norte	48	PI331793	Vietnã
4	PI90251	Coréia do Sul	49	PI331795	Vietnã
5	PI133226	Indonésia	50	PI341254	Sudão
6	PI145079	Zimbábue	51	PI341264	Libéria
7	PI148259	Indonésia	52	PI360851	Japão
8	PI148260	África do Sul	53	PI377573	China
9	PI153681	São Salvador	54	PI381660	Uganda
10	PI159097	África do Sul	55	PI381680	Uganda
11	PI159927	Peru	56	PI398493	Coréia do Sul
12	PI164885	Guatemala	57	PI407744	China Centro
13	PI165524	Índia	58	PI407764	China Sul
14	PI166141	Nepal	59	PI416828	Japão
15	PI170889	África do Sul	60	PI417563	Vietnã
16	PI171437	China	61	PI417581	EUA Sul
17	PI172902	Turquia	62	PI417582	EUA Sul
18	PI189402	Guatemala	63	PI427276	China Sul
19	PI200832	Burma Mianmar	64	PI438301	Coréia do Norte
20	PI203400	França	65	PI90577	China Nordeste
21	PI203404	Japão	66	PI159922	Peru
22	PI204333	Suriname	67	PI209839	Nepal
23	PI204340	Suriname	68	PI222546	Argentina
24	PI205384	Paquistão	69	PI240665	Filipinas
25	PI205912	Tailândia	70	PI281898	Malásia
26	PI210178	Taiwan	71	PI281911	Filipinas
27	PI210352	Moçambique	72	PI284816	Malásia
28	PI212606	Afeganistão	73	PI306712	Tanzânia
29	PI215692	Israel	74	PI438503-A	EUA
30	PI222397	Paquistão	75	PI281907	Malásia
31	PI222550	Argentina	76	IAC100	Brasil
32	PI229358	Japão	77	Paranagoiana	Brasil
33	PI239237	Tailândia	78	A7002	Brasil
34	PI253664	China	79	CD215	Brasil
35	PI259540	Nigéria	80	Conquista	Brasil
36	PI265491	Peru	81	Pintado	Brasil
37	PI265497	Colômbia	82	Sambaiba	Brasil
38	PI274454-A	Japão	83	Dowling	EUA
39	PI274454-B	Japão	84	PI200526	Japão
40	PI274507	China	85	PI200487	Japão
41	PI283327	Taiwan	86	PI471904	Indonésia
42	PI285095	Venezuela	87	PI459025	Japão
43	PI295952	Rússia	88	PI227687	Japão
44	PI297550	Rússia	89	PI171451	Japão
45	PI306702	Tanzânia

Figura 1 - Variabilidade fenotípica para cor da vagem, semente e hilo dos genótipos avaliados (MULATO, 2009)

Tabela 2 - Características de alguns materiais incluídos no painel de mapeamento associativo (GRIN, 2012)

Acesso	Principais características
Conquista	Cultivar (ampla adaptação)
Sambaíba	Cultivar (Centro-Norte)
CD215	Cultivar (Sul)
Pintado	Cultivar (Centro-Oeste) - Resistência a <i>Heterodera glycines</i>
A7002	Cultivar (Centro-Oeste) - Alto teor de óleo
IAC100	Cultivar (Centro-Sul) - Resistência a insetos
Paranagoiana	Cultivar - Período juvenil longo
Dowling	Cultivar - Resistência a <i>Aphis glycines</i>
PI 133226	Ausência de lipoxigenases
PI 171451	Resistência a insetos
PI 227687	Resistência a insetos
PI 200487	Resistência a <i>Phakopsora pachyrhizi</i>
PI 200526	Resistência a <i>Phakopsora pachyrhizi</i>
PI 471904	Resistência a <i>Phakopsora pachyrhizi</i>
PI 133226	Resistência a <i>Phytophthora sojae</i>
PI 145079	Resistência a <i>Phytophthora sojae</i>
PI 165524	Resistência a <i>Phytophthora sojae</i>
PI 171437	Resistência a <i>Phytophthora sojae</i>
PI 204333	Resistência a <i>Phytophthora sojae</i>
PI 212606	Resistência a <i>Phytophthora sojae</i>
PI 253664	Resistência a <i>Phytophthora sojae</i>
PI 259540	Resistência a <i>Phytophthora sojae</i>
PI 283327	Resistência a <i>Phytophthora sojae</i>
PI 331793	Resistência a <i>Phytophthora sojae</i>
PI 341264	Resistência a <i>Phytophthora sojae</i>
PI 377573	Resistência a <i>Phytophthora sojae</i>
PI 398493	Resistência a <i>Phytophthora sojae</i>
PI 407764	Resistência a <i>Phytophthora sojae</i>
PI 153681	Tolerância a salinidade
PI 164885	Tolerância a salinidade
PI 165524	Tolerância a salinidade
PI 171437	Tolerância a salinidade
PI 203400	Tolerância a salinidade
PI 204333	Tolerância a salinidade
PI 205912	Tolerância a salinidade
PI 210178	Tolerância a salinidade
PI 222550	Tolerância a salinidade

3.3 Avaliação fenotípica

A avaliação agromorfológica dos acessos foi realizada em área experimental do Departamento de Genética da Escola Superior de Agricultura ‘Luiz de Queiroz’ (ESALQ / USP), localizado no município de Piracicaba - SP (22°45' S / 47°38' O). Sementes dos acessos foram multiplicadas em casa de vegetação no primeiro semestre de 2010, visando obter quantidade suficiente de sementes para instalação do experimento à campo (Figura 2) em 09/12/2010.

Figura 2 - Visão geral do experimento para avaliação fenotípica dos acessos (Departamento de Genética - ESALQ)

Foi utilizado o delineamento de blocos completos casualizados com três repetições e 89 tratamentos. A parcela experimental foi composta por duas linhas de três metros, com espaçamento de 0,5 m entre linhas. A semeadura do ensaio foi feita manualmente, tendo por base 15 sementes por metro. Os tratos culturais foram realizados conforme o padrão para a cultura e o lote irrigado sempre que necessário. Foram coletados dados referentes às seguintes características:

- **Número de dias para o florescimento (NDF)** – Número de dias da semeadura até o florescimento de 50% da parcela.
- **Altura da planta no florescimento (APF)** - Média da altura de quatro plantas centrais na data do florescimento, em centímetros.
- **Início da granação (NDR5)** – Número de dias da semeadura até o início do enchimento de grãos em 50% da parcela.
- **Altura da inserção da primeira vagem (AIPV)** – Média da altura da inserção de quatro plantas centrais, em centímetros.
- **Término da granação (NDR7)** – Número de dias da semeadura até a maturação fisiológica em 50% da parcela.
- **Período de granação (PEG)** – Número de dias compreendidos entre os estágios R5 e R7.
- **Número de dias para a maturidade (NDM)** – Número de dias do plantio até 95% das vagens apresentarem coloração madura em 50% da parcela.

- **Altura da planta na maturidade (APM)** – Média das quatro plantas centrais da parcela no momento da maturação, em centímetros.
- **Estande (EST)** – escala de notas (1 a 5) relativas a germinação da parcela, sendo 5 para parcelas com aproximadamente 100% de germinação.
- **Valor agronômico (VA)** – Escala de notas (1 a 5) referente ao aspecto geral das plantas na parcela, sendo 5 o valor ideal.
- **Acamamento (AC)** – Escala de notas, sendo 5 para planta totalmente acamada e 1 para planta totalmente ereta.
- **Retenção foliar (RF)** – Escala de notas (1 a 5) referente a retenção de folhas verdes após a maturação das vagens. Característica frequentemente associada à suscetibilidade a percevejos sugadores.
- **Produtividade de grãos (PG)** – Massa total das sementes produzidas na parcela, convertida em kg/ha.
- **Massa de 100 sementes (MCS)** – Massa de 100 sementes, em gramas.
- **Teor de proteína e óleo** – Foi avaliado o teor de proteína e óleo total, além da porcentagem de ácido palmítico, esteárico, oleico e linoleico. Medido por espectroscopia infravermelha utilizando curva de calibração e equipamento fornecidos pela PERTEN (DA7200).

3.4 Avaliação genotípica

Para genotipagem dos materiais foram utilizados 114 locos SSR previamente descritos na literatura (SONG et al., 2004), selecionados com base nos valores de PIC (*Polymorphism Information Content*) e em sua distribuição ao longo de todos os 20 grupos de ligação do genoma da soja (Figura 3). Na síntese dos *primers* foi adicionada uma sequência específica para pareamento com fluoróforo, permitindo a genotipagem dos materiais em sequenciador automático.

(continua)

Figura 3 - Distribuição genômica dos 114 marcadores microssatélites utilizados no mapeamento associativo. Distância em centiMorgans (cM)

(conclusão)

Figura 3 - Distribuição genômica dos 114 marcadores microssatélites utilizados no mapeamento associativo. Distância em centiMorgans (cM)

Para reação de amplificação foram utilizados 20 ng do DNA, 0,16 μ M do iniciador *Forward*, 0,2 μ M do iniciador *Reverse*, 0,3 μ M do primer fluorescente (M13), 250 μ M de cada dNTP, solução tampão de PCR, BSA (*Bovine Serum Albumine*), 2,0 mM de $MgCl_2$ e uma unidade da enzima Taq DNA-polimerase em um volume total de 15 μ L. A amplificação das regiões microssatélites (PCR) foi realizada conforme o protocolo: 5 minutos a 94 °C, seguidos por 10 ciclos de 1 minuto de desnaturação a 94 °C, 1 minuto de pareamento a 50 °C e 1 minuto de extensão a 72 °C, sendo que após cada ciclo a temperatura de pareamento foi

diminuída em 1 °C. Em seguida, foram realizados mais 30 ciclos de 40 segundos a 94 °C, 40 segundos a 40 °C e 1 minuto a 72 °C, com extensão final de 10 minutos a 72 °C.

Os produtos gerados pela PCR foram separados por eletroforese em gel de poliacrilamida sob condições desnaturantes utilizando sequenciador automático 4300S DNA Analysis System (LI-COR Corporate). O tamanho dos alelos foi inferido com base em marcadores comerciais padrão (*ladder*) e as imagens obtidas foram arquivadas para posterior marcação dos alelos e genotipagem com auxílio dos softwares SAGA Lite e SAGA MX Generation (LI-COR Corporate). Os resultados foram conferidos e, quando necessário, a identificação dos alelos foi corrigida manualmente (Figura 4).

Figura 4 - Perfis genético-moleculares e identificação alélica dos 89 acessos do painel associativo. **A)** Satt 038, **B)** Satt 146, **C)** Satt 174 e **D)** Satt 577. Diferentes cores representam os alelos presentes em cada loco

3.5 Análise dos dados

3.5.1 Dados fenotípicos

Os dados referentes aos experimentos de campo foram inicialmente testados quanto à normalidade e homogeneidade de variância pelos testes de Barlett e Lilliefors, seguindo as recomendações de Little e Hills (1978). Foi realizada a conferência em relação a presença de *outliers* e procedeu-se a transformação de dados quando necessário. A tabela resultante foi submetida à análise de variância (ANAVA), considerando-se fixo o efeito de tratamentos (acessos) conforme o modelo:

$$Y_{ij} = m_i + t_i + b_j + e_{ij}$$

sendo:

Y_{ij} : é o valor do tratamento i na repetição j;

m_i : a média geral; t_i : o efeito do tratamento i, $i = 1, 2, \dots, I$;

b_j : o efeito da repetição j, $j = 1, 2, \dots, J$;

e_{ij} : o erro experimental associado à parcela ij.

Sendo detectadas diferenças significativas entre os tratamentos (teste F), procedeu-se a comparação de médias pelo teste de Scott-Knott (1974) à 5% de probabilidade de erro. Este procedimento utiliza a razão de verossimilhança para atestar a significância de que os tratamentos podem ser divididos em grupos que maximizem a soma de quadrados entre grupos (RAMALHO et al., 2000).

Com o objetivo de avaliar o grau de divergência genética entre os genótipos, foram calculadas as distâncias generalizadas de Mahalanobis (D^2) entre pares de acessos a partir das médias e da matriz de variâncias e covariâncias residuais. As distâncias generalizadas de Mahalanobis (D^2) foram estimadas a partir da transformação das variáveis por condensação pivotal e padronizadas utilizando a expressão:

$$D_{ii'}^2 = \Sigma (X_{ij} - X_{i'j})^2$$

sendo:

$D_{ii'}^2$: é a distância generalizada de Mahalanobis entre os acessos i e i' ;

$X_{ij} - X_{i'j}$: é a diferença entre os genótipos i e i' em relação à j -ésima variável;

X_{ij} : é a variável transformada padronizada (não-correlacionada), ou seja, a média do i -ésimo genótipo em relação à j -ésima variável com variância igual a 1.

A matriz de distâncias de Mahalanobis foi posteriormente somada a matriz de distâncias de Rogers-W, obtida a partir dos dados de marcadores moleculares, para posterior agrupamento dos acessos. Para análise de divergência genética, foram considerados apenas os dados referentes aos caracteres quantitativos, sendo excluídos os dados de notas (EST, VA, AC e RF).

Para avaliar a contribuição das características agromorfológicas na discriminação dos acessos e a possibilidade de descarte de descritores, procedeu-se a análise por variáveis canônicas (CRUZ; REGAZZI, 2001). A importância relativa dos caracteres foi realizada pela análise dos coeficientes das transformações lineares utilizados para a obtenção das variáveis canônicas (CRUZ, 1990). Quando se utiliza este procedimento é comum a transformação das variáveis originais em variáveis padronizadas e não correlacionadas, pelo processo de condensação pivotal. As variáveis canônicas foram obtidas através da resolução dos sistemas:

$$\text{I)} \quad \left(F - \lambda_t \hat{\Sigma} \right) \alpha_{\sim t} = \phi$$

$$\text{II)} \quad \left| F - \lambda_t \hat{\Sigma} \right| = 0$$

sendo:

F : é a matriz de covariâncias entre médias dos genótipos

$\hat{\Sigma}$: é a matriz de covariâncias residuais;

λ_t : são as raízes características (autovalores);

$\alpha_{\sim t}$: é o vetor característico (autovetor) associado a cada estimativa das raízes características, e os elementos deste vetor são os coeficientes de ponderação.

Realizou-se também a análise da relação entre pares de características pelo índice de correlação de Pearson:

$$\rho = \frac{COV(X,Y)}{\sqrt{VAR(X).VAR(Y)}}$$

sendo:

X e Y referem-se ao par de características comparadas.

Todas as análises referentes aos dados agromorfológicos foram realizadas com auxílio dos softwares SAS (SAS INSTITUTE, 1999) e GENES (CRUZ, 2001).

3.5.2 Dados genotípicos

Os arquivos de genotipagem gerados pelos softwares SAGA Lite e SAGA MX Generation (LI-COR Corporate) foram exportados para a análise dos dados de marcadores SSR. Foi calculado o número de alelos por loco e as respectivas frequências alélicas, as quais foram utilizadas para cálculo do PIC (*Polimorphism Information Content*) no aplicativo PIC Calculator Extra (<http://www.genomics.liv.ac.uk/animal/pic.html>).

Para análise da divergência entre acessos, utilizou-se a distância genética de Rogers modificada (Rogers-W), a qual é estimada com base nas frequências alélicas nos diferentes locos, por meio do software NTSys (ROLF, 1989). A matriz de distâncias obtidas com dados de marcadores foi então somada à matriz de distâncias resultante dos dados agromorfológicos utilizando a soma matricial do software GENES (CRUZ, 2001). A matriz resultante foi utilizada para obtenção de um dendrograma gerado pelo método de agrupamento hierárquico de Ward, o qual busca minimizar o desvio padrão entre os elementos de cada grupo. O agrupamento por esta metodologia tem se mostrado o mais eficaz para recuperação de grupos quando utilizados diferentes conjuntos de variáveis, além de proporcionar resultados similares aos agrupamentos obtidos por métodos bayesianos (MOHAMMADI; PRASANNA, 2003; ODONG et al., 2011).

As matrizes de distância molecular e agromorfológica foram comparadas por correlação matricial utilizando a estatística Z de Mantel (1967), sendo a significância estatística testada com 1.000 permutações aleatórias (*bootstrap*).

O conhecimento acerca do grau de estrutura populacional do painel associativo é imperativo para que o mapeamento de QTL's seja efetivo, evitando associações espúrias entre marcador e fenótipo. Neste sentido, a abordagem de modelos mistos para o mapeamento associativo deve incluir, além dos dados genotípicos e fenotípicos, informações a respeito da estrutura populacional (matriz Q) e do grau de parentesco (matriz K ou *Kinship*) entre os genótipos (YU et al., 2006).

Para estimativa da estrutura genética e obtenção da matriz Q foi empregado o método bayesiano proposto por Pritchard et al. (2000) utilizando o software STRUCTURE. Para cálculo do número de subpopulações mais provável (k) foram utilizados os seguintes parâmetros: 10 corridas independentes com k variando entre 2 e 10, 80.000 *burn-ins*, 300.000 simulações de Monte Carlo Cadeias de Markov (MCMC), modelo de não-mistura (*no admixture*) e frequências alélicas correlacionadas. A seleção do número k mais provável em relação aos propostos foi realizada por meio dos valores de Δk (EVANNO et al., 2005). Por sua vez, a matriz de coancestralidade (K) entre os genótipos foi obtida pelo método de Ritland (1996) utilizando o software SPAGeDI (HARDY; VEKEMANS, 2002).

3.5.3 Desequilíbrio de ligação

O desequilíbrio de ligação (DL) entre pares de marcadores foi calculado pelo coeficiente de determinação r^2 , utilizando o teste de permutação rápida no pacote computacional TASSEL (BRADBURY et al., 2007). A análise do DL foi realizada considerando cada grupo de ligação separadamente, evitando que marcadores em grupos de ligação distintos pudessem apresentar algum grau de DL. Alelos pouco frequentes (< 5%) foram excluídos a priori utilizando a opção disponível no software. Pares de locos foram considerados em DL significativo caso $p\text{-valor} < 0,01$.

Para visualização do grau de decaimento do DL ao longo do genoma, foram considerados os valores significativos de r^2 e as respectivas distâncias genéticas entre pares de marcadores (cM), obtidas com base no mapa público da soja (SONG et al., 2004). As informações foram sumarizadas em um gráfico, sendo utilizada regressão logarítmica para inferir o tamanho médio dos blocos de ligação (haplótipos).

3.5.4 Mapeamento associativo

As análises de associação entre marcador e fenótipo foram realizadas no programa TASSEL (BRADBURY et al., 2007), conforme a metodologia proposta por Yu et al. (2006). Com base na informação dos marcadores SSR, foram obtidas as matrizes Q e K e seus efeitos foram incluídos como covariáveis para ajuste do modelo de associação. A equação utilizada no modelo misto é a seguinte:

$$y = X\beta + S\alpha + Qv + Zu + e$$

sendo:

y: o vetor de observações fenotípicas;

β : um vetor de efeitos fixos (outros efeitos diferentes dos marcadores e estrutura populacional);

α : um vetor de efeitos fixos dos marcadores;

v: um vetor de efeitos fixos de estrutura da população;

u: um vetor de efeitos poligênicos aleatórios desconhecidos;

e: o vetor aleatório de efeitos residuais;

Q: a matriz do STRUCTURE relacionando y a v;

X, S e Z: matrizes de incidência (compostas por zeros e uns), relacionando y a β , α e u, respectivamente.

Todas as associações com p-valor > 0,01 foram desconsideradas das análises posteriores. As associações restantes foram submetidas a correção para múltiplos testes e controle de possíveis Erro Tipo I, considerando assim a proporção de hipóteses nulas indevidamente rejeitadas. Primeiramente, foi utilizado o método do *False Discovery Rate* – FDR (BENJAMINI; HOCHBERG, 1995), com base nos q-valores obtidos através do pacote “qvalue” (STOREY, 2003) implementado no programa R (R DEVELOPMENT CORE TEAM, 2011). Outra abordagem utilizada consistiu no cálculo do menor fator bayesiano (*Minimum Bayes Factor* - minBF) (GOODMAN, 2001). O minBF é calculado com base nos p-valores, conforme a expressão:

$$\min BF = -e * p * \ln(p)$$

sendo:

e: o algoritmo neperiano;

p: o p-valor da associação entre marcador e característica.

Para discussão dos resultados obtidos nos testes de associação, foram selecionadas as características produtividade de grãos (PG), massa de 100 sementes (MCS), teor de óleo total e teor de proteína. Os resultados dos locos associados a estas características foram comparados com os dados públicos do mapa da soja, disponíveis no SoyBase (<http://www.soybase.org/>).

4 RESULTADOS

4.1 Avaliação agromorfológica

Os resultados das análises univariadas do experimento, referentes a cada uma das características avaliadas, encontram-se sumarizados na Tabela 3. O valor do quadrado médio obtido para todas as características foi significativo 5% de probabilidade de erro de acordo com o teste F. Verifica-se ainda, uma grande amplitude dos dados para todas as características, com ampla variabilidade para maturação, produtividade e teores de ácidos graxos e proteínas.

Os coeficientes de variação foram, em sua maioria, considerados entre baixos e medianos, com exceção das características altura de inserção da primeira vagem e produtividade de grãos. Tais características são altamente influenciadas pelo ambiente e, considerando a natureza não adaptada dos materiais testados, estes coeficientes podem ser considerados aceitáveis. Havendo diferença entre os materiais, as médias dos tratamentos para cada característica foram comparadas pelo teste de Scott-Knott (5%) (Tabela 4).

Considerando os resultados do teste de comparação entre médias, fica evidente a existência de germoplasma exótico com características superiores aos materiais comerciais (Id 76 à 82). Este resultado sugere uma primeira evidência da possibilidade de se encontrar alelos favoráveis em materiais atualmente pouco explorados em programas de melhoramento.

A maior produtividade foi verificada para o genótipo PI203400 (Id 20), o qual alia alta produtividade de grãos e massa de 100 sementes como principais características, embora tenha apresentado um ciclo de maturação sete dias mais tardio que a média do painel associativo (126 e 119, respectivamente). Interessante salientar que este material encontra-se caracterizado como tolerante à estresse salino (GRIN, 2012).

Foram também verificados no conjunto de genótipos, materiais produtivos e de ciclo precoce (PI159097 – Id 10) ou com elevado teor de proteína (PI285095 - Id 42). Por sua vez, podem-se citar como exemplos de materiais precoces e com elevados teores de proteína e óleo as PI's 212606 (Id 28), 438503-A (Id 74) e 295952 (Id 43). Dentre os materiais comerciais avaliados destacam-se os cultivares 'CD215' (Id 79), com alto teor de óleo total, e 'Conquista' (Id 80) com elevada produtividade de grãos.

Tabela 3 - Análise de variância dos dados agromorfológicos do painel associativo

Fonte de variação	GL	Quadrado Médio (QM)																			
		NDF (dias)	APF (cm)	NDR5 (dias)	AIPV (cm)	NDR7 (dias)	PEG (dias)	NDM (dias)	APM (cm)	PG (kg/ha)	MCS (g)	Proteína (%)	Óleo (%)	Palmitico (%)	Estearico (%)	Oleico (%)	Linoleico (%)	EST **	VA **	AC **	RF **
Blocos	2																				
Genótipos	88	289,41*	1.533,52*	730,35*	80,21*	549,54*	106,96*	244,58*	2.600,13*	2.181.910,89*	100,54*	35,82*	12,51*	7,94*	0,32*	29,05*	38,48*	0,13*	0,14*	0,45*	0,34*
Resíduo	176	0,99	41,08	4,13	15,86	8,53	11,33	12,78	96,17	262.206,85	2,62	1,20	0,43	0,42	0,03	4,49	1,98	0,05	0,02	0,04	0,05
Total	266																				
CV (%)		1,89	10,70	2,84	23,60	2,84	10,79	2,98	11,03	28,16	10,66	2,29	3,48	7,16	5,18	9,22	2,38	11,87	9,67	12,58	16,84
Média		52,65	59,93	71,52	16,87	102,66	31,21	119,85	88,90	1.818,43	15,20	47,91	18,85	9,09	3,31	22,98	59,14	1,81	1,58	1,65	1,38
Máximo		75,00	122,50	101,00	37,50	145,00	60,00	156,00	177,50	4.166,67	27,67	62,89	24,63	11,52	4,32	36,44	71,41	1,00	1,00	1,00	1,00
Mínimo		29,00	17,50	39,00	5,50	69,00	18,00	89,00	28,75	66,67	4,18	39,68	14,39	0,30	0,73	12,18	51,65	2,24	2,24	2,24	2,24

* Significativo à 5% de probabilidade (teste F)

** Dados transformados em $\sqrt{x}$

NDF – Número de dias para o florescimento, APF – Altura da planta no florescimento, NDR5 – Início da granação, AIPV – Altura da inserção da primeira vagem, NDR7 – Término da granação, PEG – Período de granação, NDM – Número de dias para a maturidade, APM – Altura da planta na maturidade, PG – Produtividade de grãos, MCS – Massa de 100 sementes, EST – Estande, VA – Valor agrônomo, AC – Acamamento, RF – Retenção foliar.

Tabela 4 - Médias dos 20 caracteres avaliados nos 89 genótipos de soja e os grupos estabelecidos pelo teste de comparação de médias de Scott-Knott (5%)

(continua)

NDF (dias)		APF (cm)		NDR5 (dias)		AIPV (cm)		NDR7 (dias)		PEG (dias)		NDM (dias)		APM (cm)		PG (kg/ha)		MCS (g)	
Id	$\bar{x}$	Id	$\bar{x}$	Id	$\bar{x}$	Id	$\bar{x}$	Id	$\bar{x}$	Id	$\bar{x}$	Id	$\bar{x}$	Id	$\bar{x}$	Id	$\bar{x}$	Id	$\bar{x}$
75	72,95 a	68	109,58 a	73	99,33 a	88	27,83 a	35	144,89 a	35	57,79 a	67	139,00 a	88	152,50 a	20	3577,78 a	17	26,17 a
73	72,67 a	75	103,75 a	67	98,67 a	68	26,50 a	73	125,33 b	77	52,00 a	82	138,30 a	66	148,75 a	60	3495,55 a	58	25,94 a
71	72,33 a	73	103,75 a	71	96,00 a	71	26,25 a	68	123,00 b	85	43,05 b	73	138,00 a	35	143,75 a	85	3448,89 a	20	25,85 a
68	71,98 a	72	103,33 a	72	95,33 b	42	25,50 a	67	122,00 b	83	42,67 b	68	136,67 a	33	140,00 a	42	3435,55 a	87	25,66 a
67	71,67 a	33	96,25 b	75	94,67 b	54	25,33 a	29	121,00 b	30	41,00 b	45	134,67 a	67	137,50 a	84	3191,11 a	9	25,34 a
33	68,33 b	71	95,00 b	29	94,00 b	86	24,75 a	39	119,00 c	20	39,00 b	85	132,67 b	12	136,25 a	80	3168,89 a	63	25,01 a
29	67,67 b	66	91,67 b	68	93,00 b	2	24,75 a	75	118,67 c	3	39,00 b	77	132,33 b	73	133,75 a	51	3151,11 a	62	24,22 a
39	67,00 b	29	91,25 b	7	92,33 b	55	24,17 a	72	117,67 c	57	38,67 b	72	132,00 b	68	128,75 b	86	3140,00 a	6	23,75 a
7	67,00 b	26	90,00 b	50	92,00 b	87	24,00 a	71	117,67 c	15	37,67 b	29	132,00 b	23	125,83 b	15	3002,22 a	44	23,02 b
72	66,95 b	55	88,33 b	36	91,67 b	72	23,83 a	85	117,05 c	84	37,33 b	75	130,67 b	45	124,17 b	1	2944,44 a	64	22,27 b
88	66,67 b	50	87,92 b	26	90,67 c	78	23,33 a	78	117,05 c	64	37,33 b	70	130,00 b	31	123,75 b	53	2942,22 a	27	21,89 b
55	66,67 b	82	86,03 b	13	90,67 c	5	23,25 a	26	117,00 c	78	37,05 b	39	130,00 b	13	123,33 b	6	2886,66 a	19	21,78 b
50	66,67 b	5	85,83 b	12	90,67 c	40	22,33 a	66	116,33 c	21	37,00 b	71	129,00 b	86	122,08 b	24	2841,57 a	57	21,78 b
26	66,00 b	25	85,00 b	31	90,33 c	20	21,33 b	23	116,00 c	89	36,67 b	78	128,67 b	18	121,25 b	10	2773,33 a	3	21,74 b
36	64,00 c	54	84,58 b	23	90,00 c	80	21,25 b	82	115,06 c	87	36,67 b	87	128,00 c	59	120,83 b	2	2697,78 b	84	21,68 b
12	62,67 d	22	84,58 b	82	89,41 c	73	21,25 b	12	115,06 c	63	36,33 b	11	127,00 c	75	119,58 b	52	2666,67 b	8	21,30 b
40	62,00 d	7	81,67 c	70	89,33 c	84	21,00 b	40	115,00 c	86	35,33 c	50	126,67 c	60	119,58 b	76	2662,22 b	37	21,08 b
54	61,67 d	88	81,03 c	39	88,33 c	63	20,50 b	38	115,00 c	24	35,33 c	26	126,33 c	82	118,72 c	54	2606,67 b	52	20,55 c
13	61,67 d	40	80,42 c	33	88,33 c	49	20,50 b	7	115,00 c	58	35,00 c	66	126,00 c	22	114,17 c	11	2575,56 b	54	20,26 c
23	61,33 d	12	80,42 c	55	87,67 d	11	20,50 b	13	114,67 c	32	35,00 c	20	126,00 c	26	112,92 c	8	2564,44 b	11	20,25 c
11	61,33 d	11	80,00 c	40	86,67 d	77	20,42 b	87	114,33 c	9	35,00 c	7	126,00 c	20	112,08 c	4	2562,22 b	24	20,14 c
22	60,00 e	39	78,75 c	35	86,67 d	12	20,33 b	50	114,33 c	4	35,00 c	86	125,00 c	36	111,67 c	66	2515,56 b	56	19,99 c
35	59,00 e	18	78,75 c	38	86,33 d	50	20,25 b	31	113,33 d	27	34,67 c	84	125,00 c	53	111,25 c	41	2455,55 b	85	19,97 c
31	59,00 e	67	78,47 c	18	84,33 d	85	20,17 b	36	113,00 d	19	34,33 c	81	125,00 c	7	109,58 c	27	2388,89 b	10	19,65 c
82	58,98 e	81	77,92 c	22	84,00 d	67	20,17 b	20	113,00 d	17	34,33 c	55	125,00 c	50	109,17 c	47	2377,78 b	51	19,60 c
38	58,33 f	13	75,00 c	66	82,67 e	36	20,08 b	18	112,67 d	61	34,16 c	40	125,00 c	42	108,33 c	32	2364,45 b	4	19,48 c
70	58,00 f	35	74,17 c	25	82,33 e	45	19,83 b	33	112,55 d	80	34,00 c	38	125,00 c	54	107,50 c	45	2360,00 b	61	19,46 c
87	57,33 f	31	74,17 c	88	82,00 c	48	19,75 b	70	112,33 d	54	34,00 c	23	125,00 c	72	105,83 c	38	2293,33 b	42	19,44 c
66	57,33 f	23	74,17 c	78	79,33 f	1	19,67 b	25	112,33 d	34	34,00 c	21	125,00 c	71	105,83 c	82	2279,65 b	15	19,27 c
25	57,33 f	48	72,92 c	45	79,00 f	69	19,42 b	54	112,00 d	66	33,67 c	80	124,00 c	29	105,42 c	16	2224,44 b	1	18,72 c
5	57,33 f	86	72,50 c	11	79,00 f	25	19,42 b	21	112,00 d	6	33,67 c	54	124,00 c	55	103,33 c	25	2204,44 b	45	18,29 c
86	57,00 f	41	72,50 c	59	78,67 f	53	19,33 b	77	111,67 d	42	33,33 c	36	124,00 c	40	103,33 c	5	2195,56 b	47	18,18 c
81	57,00 f	60	72,08 c	54	78,00 f	8	19,17 b	86	111,00 d	79	33,00 c	35	124,00 c	25	101,67 c	87	2191,11 b	83	17,78 d
78	57,00 f	53	70,83 c	87	77,67 f	66	19,00 b	11	111,00 d	10	33,00 c	33	124,00 c	78	100,83 d	37	2162,22 b	21	17,74 d
48	57,00 f	69	67,50 d	5	77,67 f	9	18,92 b	84	110,00 d	62	32,67 c	31	124,00 c	5	100,83 d	9	2151,11 b	14	17,60 d
18	57,00 f	38	67,08 d	81	76,67 g	51	18,67 b	55	110,00 d	52	32,67 c	25	124,00 c	65	99,58 d	17	2117,78 b	80	17,32 d
69	56,33 f	36	66,25 d	48	76,67 g	38	18,58 b	45	110,00 d	46	32,67 c	18	124,00 c	48	99,58 d	46	2099,65 b	65	17,15 d
49	55,33 g	80	65,83 d	86	75,67 g	14	18,58 b	22	110,00 d	14	32,67 c	13	124,00 c	11	99,17 d	65	2080,00 b	43	17,03 d
20	55,33 g	87	65,00 d	14	75,33 g	29	18,33 b	88	108,00 e	76	32,33 c	12	124,00 c	49	97,92 d	14	2057,78 b	79	16,68 d
21	54,48 g	51	65,00 d	85	75,00 g	7	18,17 b	83	108,00 e	37	32,33 c	14	123,33 c	69	97,50 d	55	2002,22 b	53	16,26 d
59	54,33 g	85	64,58 d	21	75,00 g	81	18,08 b	81	108,00 e	74	32,29 c	89	122,67 c	38	97,08 d	61	1948,89 b	5	16,05 d
80	53,33 g	20	64,58 d	20	74,00 g	39	18,08 b	14	108,00 e	44	32,00 c	83	122,67 c	1	94,17 d	49	1931,11 b	49	16,01 d
42	53,33 g	49	62,08 d	69	73,67 g	18	18,08 b	80	107,00 e	11	32,00 c	22	122,67 c	21	93,33 d	33	1908,89 b	48	15,76 d
85	52,33 h	59	61,25 d	42	73,67 g	31	17,83 b	42	107,00 e	65	31,67 c	42	122,00 d	2	93,33 d	40	1835,56 c	2	15,76 d
76	52,33 h	2	60,42 d	49	73,33 g	58	17,33 b	5	106,00 e	81	31,33 c	69	121,33 d	46	89,58 d	21	1813,33 c	60	15,71 d
53	51,00 i	65	58,17 e	80	73,00 g	47	17,33 b	89	105,67 e	45	31,00 d	5	120,33 d	80	89,17 d	69	1782,22 c	34	15,14 d
14	51,00 i	78	57,50 e	84	72,67 g	60	17,17 b	59	105,67 e	47	30,67 d	76	119,67 d	77	88,75 d	88	1771,11 c	66	15,10 d
45	50,67 i	45	57,50 e	32	70,33 h	70	16,83 b	32	105,33 e	39	30,67 d	59	119,67 d	81	87,08 d	89	1735,55 c	77	15,06 d
41	50,67 i	70	56,67 e	76	69,67 h	22	16,83 b	15	104,67 e	2	30,67 d	49	119,67 d	14	87,08 d	19	1606,67 c	86	14,18 e
60	50,33 i	1	56,25 e	53	69,67 h	21	16,83 b	69	104,00 e	69	30,33 d	32	119,67 d	51	84,17 e	48	1575,56 c	81	13,26 e
89	50,00 i	47	54,58 e	89	69,00 h	75	16,67 b	49	103,00 e	68	30,00 d	48	119,00 d	85	82,08 d	77	1542,22 c	32	12,89 e
84	49,67 i	21	53,33 g	60	67,00 i	33	16,58 b	48	103,00 e	60	30,00 d	24	118,67 d	47	80,42 d	70	1533,33 c	69	12,76 e
8	49,33 i	42	52,92 e	15	67,00 i	35	16,42 b	76	102,00 e	25	30,00 d	19	118,67 d	87	78,33 e	36	1508,89 c	78	12,64 e
79	48,67 j	32	52,92 e	41	66,67 i	59	15,83 c	27	100,67 e	49	29,67 d	9	118,67 d	41	78,33 e	7	1482,22 c	76	12,58 e
51	48,00 j	77	51,67 e	27	66,00 i	79	15,50 c	60	97,00 f	1	29,33 d	57	117,00 e	76	76,67 e	13	1440,00 c	82	11,98 e
32	48,00 j	6	50,42 e	83	65,33 i	13	15,25 c	53	96,67 f	56	29,00 d	27	117,00 e	39	75,83 e	12	1395,56 c	22	11,54 e
15	48,00 j	84	50,00 e	51	63,33 j	4	15,00 c	79	96,00 f	38	28,67 d	65	116,67 e	6	75,83 e	26	1371,11 c	40	11,45 e
1	47,33 j	79	47,50 f	79	63,00 j	89	14,75 c	30	95,67 f	40	28,33 d	58	116,67 e	70	75,00 e	58	1344,44 c	50	11,10 e
52	47,00 j	8	47,50 f	8	63,00 j	76	14,67 c	6	95,00 f	18	28,33 d	63	116,33 e	84	71,67 e	39	1326,67 c	39	10,94 e
27	47,00 j	52	46,67 f	37	62,00 j	34	14,58 c	37	94,33 f	8	28,33 d	3	116,33 e	74	70,83 e	62	1324,45 c	38	10,65 e
77	46,67 j	15	46,25 f	1	62,00 j	27	14,50 c	52	94,00 f	5	28,33 d	15	115,67 e	8	69,17 e	64	1302,22 c	59	10,47 e
2	46,67 j	37	44,58 f	52	61,33 j	26	14,50 c	8	91,33 g	43	28,00 d	6	115,67 e	4	67,92 e	79	1300,00 c	70	10,26 e
83	45,33 j	14	43,75 f	16	61,33 j	41	14,33 c												

Tabela 4 - Médias dos 20 caracteres avaliados nos 89 genótipos de soja e grupos estabelecidos pelo teste de comparação de médias de Scott-Knott (5%)

(conclusão)

Proteína (%)	Oleo (%)	Palmitico (%)	Estearico (%)	Oleico (%)	Linoleico (%)	EST (notas)	VA (notas)	AC (notas)	RF (notas)
Id $\bar{x}$	Id $\bar{x}$	Id $\bar{x}$	Id $\bar{x}$	Id $\bar{x}$	Id $\bar{x}$	Id $\bar{x}$	Id $\bar{x}$	Id $\bar{x}$	Id $\bar{x}$
74 60,30 a	83 23,43 a	4 10,94 a	52 4,19 a	9 32,19 a	18 68,03 a	82 2,25 a	80 2,24 a	88 2,24 a	63 2,24 a
42 54,86 b	44 22,70 a	53 10,87 a	2 4,00 a	2 31,07 a	13 66,97 a	89 2,24 a	77 2,00 a	67 2,24 a	58 2,16 a
28 54,36 b	79 22,60 a	52 10,63 a	32 3,92 a	57 30,75 a	23 66,87 a	84 2,24 a	70 2,00 a	46 2,24 a	44 2,08 a
67 54,21 b	19 22,52 a	42 10,58 a	68 3,86 b	52 30,67 a	31 66,80 a	73 2,16 a	11 2,00 a	35 2,24 a	9 2,08 a
31 53,64 b	43 22,50 a	47 10,52 a	79 3,78 b	63 30,49 a	35 66,73 a	70 2,16 a	82 1,99 a	30 2,24 a	57 2,07 a
72 52,96 b	52 22,11 a	32 10,47 a	65 3,78 b	58 29,48 a	36 66,43 a	86 2,08 a	78 1,99 a	23 2,24 a	3 2,07 a
73 52,43 c	62 21,77 b	89 10,42 a	41 3,77 b	62 28,72 a	12 66,41 a	85 2,08 a	87 1,91 a	66 2,16 a	4 2,00 a
64 52,23 c	47 21,68 b	45 10,42 a	4 3,72 b	69 28,54 a	59 65,95 a	80 2,08 a	84 1,91 a	59 2,16 a	34 1,99 a
57 52,07 c	65 21,65 b	41 10,36 a	89 3,71 b	44 27,76 a	14 65,86 a	72 2,08 a	53 1,91 a	39 2,16 a	28 1,98 a
88 52,06 c	80 21,18 b	79 10,35 a	48 3,69 b	3 27,55 a	38 65,58 a	60 2,08 a	38 1,91 a	13 2,16 a	73 1,90 a
75 51,96 c	24 21,11 b	87 10,32 a	9 3,63 c	65 27,37 a	26 65,45 a	54 2,08 a	60 1,82 a	12 2,16 a	74 1,82 a
12 51,59 c	77 21,02 b	19 10,32 a	6 3,62 c	4 27,05 a	33 65,26 a	53 2,08 a	40 1,82 a	86 2,08 a	43 1,82 a
18 51,51 c	21 20,88 c	62 10,29 a	8 3,61 c	59 25,02 b	46 64,61 b	32 2,08 a	32 1,82 a	75 2,08 a	19 1,82 a
29 51,29 c	49 20,80 c	48 10,23 a	56 3,60 c	34 24,86 b	30 64,42 b	25 2,08 a	25 1,82 a	45 2,08 a	2 1,82 a
13 51,17 c	82 20,75 c	7 10,20 a	82 3,59 c	64 24,81 b	69 63,89 b	9 2,08 a	20 1,82 a	36 2,08 a	64 1,81 a
50 51,14 c	37 20,73 c	11 10,19 a	11 3,58 c	87 24,64 b	74 63,62 b	1 2,08 a	85 1,73 b	26 2,08 a	62 1,72 b
34 50,96 c	15 20,73 c	76 10,16 a	22 3,54 c	26 24,59 b	16 62,56 b	88 2,00 a	79 1,73 b	21 2,08 a	24 1,67 b
7 50,90 c	78 20,72 c	60 10,16 a	1 3,53 c	54 24,45 b	43 61,61 b	29 2,00 a	68 1,73 b	71 2,00 a	85 1,63 b
36 50,78 c	61 20,64 c	54 10,16 a	54 3,51 c	89 24,38 b	76 61,29 c	6 2,00 a	51 1,73 b	33 2,00 a	68 1,63 b
39 50,61 c	27 20,53 c	16 10,14 a	7 3,48 c	38 24,30 b	49 60,51 c	16 1,99 a	50 1,73 b	31 2,00 a	52 1,63 b
71 50,54 c	86 20,47 c	6 10,05 a	77 3,47 c	41 24,21 b	50 60,24 c	15 1,99 a	49 1,73 b	22 2,00 a	47 1,63 b
23 50,26 d	56 20,43 c	55 10,04 a	70 3,47 c	24 24,16 b	75 60,22 c	4 1,99 a	41 1,73 b	72 1,99 a	39 1,63 b
63 50,14 d	81 20,42 c	1 10,04 a	53 3,47 c	46 24,15 b	10 60,21 c	74 1,91 a	15 1,73 b	69 1,99 a	26 1,63 b
58 50,06 d	3 20,33 c	25 10,03 a	69 3,46 c	20 24,07 b	88 60,19 c	68 1,91 a	14 1,73 b	18 1,99 a	69 1,61 b
33 50,06 d	54 20,20 c	5 9,97 a	47 3,46 c	42 23,98 b	71 59,68 c	67 1,91 a	10 1,73 b	68 1,91 b	65 1,61 b
38 49,95 d	45 20,18 c	61 9,93 a	26 3,46 c	12 23,76 b	67 59,52 c	55 1,91 a	7 1,73 b	55 1,91 b	30 1,61 b
56 49,92 d	10 20,08 c	39 9,93 a	87 3,44 c	36 23,70 b	78 59,49 c	51 1,91 a	76 1,72 b	40 1,91 b	54 1,52 b
35 49,70 d	60 20,05 c	80 9,92 a	86 3,44 c	31 23,67 b	15 59,43 c	42 1,91 a	27 1,72 b	29 1,91 b	29 1,52 b
26 49,69 d	51 20,05 c	40 9,92 a	66 3,44 c	1 23,66 b	7 59,42 c	41 1,91 a	1 1,72 b	7 1,91 b	7 1,52 b
55 49,57 d	11 19,84 d	65 9,91 a	81 3,42 c	6 23,61 b	84 59,36 c	39 1,91 a	89 1,63 b	5 1,91 b	1 1,52 b
69 49,50 d	68 19,58 d	29 9,91 a	80 3,41 c	74 23,57 b	83 59,14 c	38 1,91 a	86 1,63 b	2 1,91 b	37 1,49 b
87 49,40 d	20 19,56 d	50 9,90 a	60 3,40 c	56 23,46 b	85 59,09 c	33 1,91 a	58 1,63 b	73 1,82 b	67 1,47 c
22 49,36 d	6 19,56 d	72 9,89 a	25 3,39 c	11 23,43 b	39 59,06 c	31 1,91 a	55 1,63 b	51 1,82 b	87 1,41 c
9 49,22 d	84 19,50 d	84 9,88 a	57 3,38 c	68 23,37 b	55 59,03 c	11 1,91 a	54 1,63 b	50 1,82 b	72 1,41 c
84 48,89 d	74 19,49 d	51 9,85 a	21 3,38 c	70 23,27 b	82 58,89 c	2 1,91 a	52 1,63 b	49 1,82 b	71 1,41 c
44 48,67 d	4 19,48 d	27 9,82 a	71 3,37 c	18 23,27 b	80 58,82 c	61 1,90 a	48 1,63 b	48 1,82 b	55 1,41 c
43 48,56 d	85 19,41 d	17 9,82 a	62 3,37 c	17 23,21 b	28 58,74 c	46 1,90 a	42 1,63 b	42 1,82 b	51 1,41 c
20 48,38 d	32 19,37 d	34 9,81 a	5 3,37 c	84 23,14 b	20 58,71 c	81 1,88 a	37 1,63 b	41 1,82 b	38 1,41 c
17 48,34 d	53 19,35 d	22 9,80 a	61 3,36 c	33 23,04 b	40 58,65 c	3 1,88 a	16 1,63 b	38 1,82 b	11 1,41 c
16 48,33 d	17 19,30 d	88 9,76 a	40 3,36 c	35 22,98 b	70 58,57 c	77 1,82 a	8 1,63 b	28 1,82 b	61 1,38 c
3 47,98 e	64 19,28 d	86 9,76 a	76 3,35 c	72 22,95 b	66 58,55 c	76 1,82 a	6 1,63 b	8 1,82 b	60 1,38 c
2 47,88 e	41 19,28 d	2 9,72 a	63 3,34 c	67 22,93 b	25 58,52 c	66 1,82 a	5 1,63 b	20 1,81 b	33 1,38 c
89 47,77 e	66 19,14 d	83 9,71 a	51 3,34 c	51 22,82 b	21 58,51 c	64 1,82 a	2 1,63 b	34 1,73 b	84 1,28 c
37 47,76 e	1 19,13 d	3 9,65 b	24 3,34 c	37 22,81 b	5 58,50 c	63 1,82 a	83 1,61 b	81 1,72 b	77 1,28 c
85 47,64 e	87 18,98 d	37 9,63 b	46 3,32 c	45 22,74 b	53 58,33 c	43 1,82 a	81 1,61 b	14 1,72 b	53 1,28 c
41 47,62 e	14 18,95 d	10 9,63 b	78 3,31 c	14 22,66 b	81 58,32 c	40 1,82 a	71 1,52 c	47 1,67 b	46 1,28 c
53 47,45 e	69 18,78 d	49 9,62 b	75 3,30 c	47 22,65 b	56 58,30 c	36 1,82 a	61 1,52 c	85 1,63 b	41 1,28 c
40 47,40 e	40 18,65 e	15 9,62 b	58 3,30 c	79 22,41 b	60 58,18 c	19 1,82 a	46 1,52 c	54 1,63 b	23 1,28 c
25 47,39 e	34 18,61 e	20 9,60 b	43 3,29 c	13 22,37 b	37 58,13 c	12 1,82 a	45 1,52 c	53 1,63 b	13 1,28 c
59 47,25 e	59 18,58 e	24 9,59 b	45 3,28 d	77 22,09 b	27 58,12 c	10 1,82 a	39 1,52 c	25 1,63 b	12 1,28 c
45 47,24 e	76 18,57 e	82 9,58 b	84 3,27 d	23 22,09 b	17 58,02 c	8 1,82 a	36 1,52 c	11 1,63 b	5 1,28 c
62 47,21 e	8 18,51 e	77 9,57 b	3 3,27 d	29 22,03 b	61 58,01 c	7 1,82 a	31 1,52 c	4 1,63 b	56 1,24 c
70 47,07 e	57 18,48 e	58 9,53 b	59 3,26 d	5 21,91 b	45 58,01 c	5 1,82 a	26 1,52 c	1 1,61 b	82 1,23 c
5 47,03 e	28 18,45 e	71 9,51 b	55 3,26 d	66 21,85 b	19 57,95 c	18 1,81 a	24 1,52 c	82 1,55 c	66 1,19 c
46 47,02 e	9 18,43 e	70 9,47 b	39 3,24 d	32 21,85 b	42 57,84 c	75 1,73 b	21 1,52 c	74 1,52 c	89 1,14 c
4 46,84 e	30 18,34 e	21 9,45 b	73 3,22 d	60 21,78 b	29 57,84 c	65 1,73 b	19 1,52 c	62 1,52 c	86 1,14 c
48 46,69 e	48 18,33 e	66 9,41 b	27 3,22 d	81 21,68 c	51 57,83 c	52 1,73 b	73 1,41 c	63 1,49 c	81 1,14 c
24 46,43 f	89 18,23 e	9 9,39 b	88 3,21 d	53 21,57 c	86 57,65 c	49 1,73 b	69 1,41 c	52 1,49 c	79 1,14 c
51 46,41 f	2 18,21 e	81 9,37 b	83 3,20 d	28 21,54 c	34 57,58 c	30 1,73 b	67 1,41 c	78 1,47 c	70 1,14 c
1 46,36 f	16 18,12 e	8 9,31 b	67 3,20 d	22 21,53 c	87 57,46 c	27 1,73 b	66 1,41 c	65 1,41 c	50 1,14 c
60 46,31 f	70 17,85 e	85 9,30 b	30 3,20 d	48 21,42 c	47 57,46 c	20 1,73 b	65 1,41 c	60 1,41 c	48 1,14 c
68 46,19 f	5 17,80 e	64 9,21 b	17 3,17 d	8 21,34 c	48 57,36 c	58 1,72 b	64 1,41 c	57 1,41 c	45 1,14 c
14 46,06 f	25 17,77 e	63 9,18 b	72 3,16 d	55 21,05 c	22 57,25 c	50 1,72 b	63 1,41 c	27 1,38 c	42 1,14 c
32 45,95 f	58 17,59 e	78 9,14 b	29 3,16 d	85 21,03 c	64 57,21 c	45 1,72 b	62 1,41 c	10 1,38 c	40 1,14 c
47 45,75 f	55 17,56 e	68 9,09 b	85 3,15 d	88 21,01 c	89 56,75 d	37 1,72 b	59 1,41 c	3 1,38 c	32 1,14 c
54 45,67 f	63 17,52 e	67 9,02 b	50 3,14 d	7 20,94 c	72 56,66 d	23 1,72 b	47 1,41 c	79 1,28 c	20 1,14 c
66 45,54 f	38 17,40 e	44 8,98 b	19 3,14 d	40 20,88 c	6 56,56 d	17 1,72 b	43 1,41 c	61 1,28 c	17 1,14 c
76 45,51 f	33 17,31 e	75 8,79 b	36 3,13 d	73 20,83 c	1 56,48 d	24 1,67 b	35 1,41 c	58 1,28 c	16 1,14 c
6 45,29 f	73 17,26 e	56 8,67 b	10 3,13 d	78 20,81 c	73 56,42 d	87 1,63 b	34 1,41 c	15 1,28 c	14 1,14 c
8 45,15 f	42 16,74 f	57 8,60 b	34 3,11 d	75 20,81 c	32 56,38 d	71 1,63 b	33 1,41 c	9 1,28 c	10 1,14 c
11 45,11 f	23 16,66 f	43 8,56 b	33 3,10 d	25 20,75 c	8 56,37 d	57 1,63 b	29 1,41 c	6 1,28 c	8 1,14 c
49 44,82 g	22 16,37 f	73 8,41 b	49 3,03 e	16 20,71 c	77 56,29 d	44 1,63 b	28 1,41 c	24 1,24 c	6 1,14 c
15 44,27 g	36 16,36 f	46 7,39 c	44 3,03 e	80 20,58 c	11 56,20 d	35 1,63 b	23 1,41 c	87 1,14 d	88 1,00 c
54 44,16 g	39 16,31 f	14 7,31 c	16 3,03 e	50 20,21 c					

Diversos outros materiais destacam-se para determinadas características em particular, sendo raros os materiais que contemplam mais de duas características desejáveis. Isto se deve, em parte, a baixa pressão de seleção exercida sobre estes materiais ao longo de sua história genealógica, somado ao fato de que algumas características como o teor de óleo e proteína apresentam forte correlação negativa (Tabela 5). Este fato dificulta a seleção e obtenção de materiais favoráveis que contemplem ambos os caracteres. A produtividade de grãos mostrou correlação positiva e significativa com a massa de 100 sementes, teor de óleo total e teor de ácido graxo palmítico. Contudo, a correlação foi fortemente negativa em relação ao teor total de proteínas.

Os dados fenotípicos foram submetidos à análise multivariada para auxiliar na verificação da estrutura genética do painel associativo. Este procedimento resultou em uma matriz de distâncias de Mahalanobis, a qual foi posteriormente somada a matriz de distâncias moleculares para inferência a cerca da divergência genética entre os acessos. Uma questão que surge nesta abordagem consiste em verificar qual das características menos contribui para discriminação dos materiais. O resultado da análise por variáveis canônicas demonstra que o descritor período de granação (PEG), pode ser descartado em futuras análises uma vez que apresentou o menor valor absoluto nas duas primeiras variáveis canônicas, responsáveis por explicar 81.53% da variação entre os materiais (Tabela 6).

Tabela 5 - Índices de correlação fenotípica (Pearson) entre as características avaliadas

	NDF	APF	NDR5	AIPV	NDR7	PEG	NDM	APM	PG	MCS	Proteína	Óleo	Palmítico	Estearico	Oleico	Linoleico	EST	VA	AC	RF
NDF	-																			
APF	0,905 **	-																		
NDR5	0,959 **	0,873 **	-																	
AIPV	0,679 **	0,682 **	0,627 **	-																
NDR7	0,864 **	0,765 **	0,916 **	0,615 **	-															
PEG	-0,462 **	-0,468 **	-0,449 **	-0,189	-0,052	-														
NDM	0,755 **	0,681 **	0,805 **	0,570 **	0,871 **	-0,059	-													
APM	0,783 **	0,825 **	0,793 **	0,634 **	0,735 **	-0,339 **	0,588 **	-												
PG	-0,173	-0,110	-0,131	0,194	-0,066	0,184	-0,106	-0,048	-											
MCS	-0,613 **	-0,541 **	-0,646 **	-0,162	-0,516 **	0,462 **	-0,332 **	-0,575 **	0,427 **	-										
Proteína	0,227 *	0,226 *	0,201	0,069	0,062	-0,359 **	0,034	0,247 *	-0,409 **	-0,327 **	-									
Óleo	-0,623 **	-0,499 **	-0,610 **	-0,360 **	-0,485 **	0,437 **	-0,321 **	-0,550 **	0,291 *	0,617 **	-0,670 **	-								
Palmítico	-0,052	-0,040	-0,103	0,226 *	-0,065	0,113	0,052	-0,231 *	0,429 **	0,438 **	-0,498 **	0,321 **	-							
Estearico	0,041	0,120	-0,008	0,218 *	0,003	0,025	0,071	-0,039	0,239 *	0,158	-0,467 **	0,230 *	0,537 **	-						
Oleico	-0,278 **	-0,203	-0,329 **	-0,044	-0,313 **	0,119	-0,159	-0,155	-0,043	0,341 **	0,248 *	-0,011	-0,067	0,255 *	-					
Linoleico	0,280 **	0,174	0,393 **	-0,055	0,353 **	-0,193	0,110	0,407 **	-0,205	-0,568 **	0,271 **	-0,454 **	-0,730 **	-0,481 **	-0,287 **	-				
EST	0,132	0,194	0,163	0,233 *	0,131	-0,109	0,182	0,135	0,395 **	-0,107	0,097	-0,164	0,180	0,127	-0,062	-0,156	-			
VA	0,140	0,136	0,201	0,354 **	0,264 *	0,087	0,286 **	0,021	0,549 **	0,146	-0,473 **	0,243 *	0,443 **	0,256 *	-0,268 *	-0,165	0,281 **	-		
AC	0,629 **	0,605 **	0,606 **	0,436 **	0,516 **	-0,358 **	0,339 **	0,767 **	-0,190	-0,598 **	0,292 **	-0,624 **	-0,365 **	-0,145	-0,075	0,456 **	-0,090	-0,294 **	-	
RF	-0,371 **	-0,278 **	-0,459 **	-0,235 *	-0,469 **	0,100	-0,222 *	-0,337 **	-0,231 *	0,292 **	0,278 **	0,114	-0,057	-0,026	0,600 **	-0,382 **	0,029	-0,380 **	-0,159	-

* Significativo à 1% (teste t)

** Significativo à 5% (teste t)

NDF – Número de dias para o florescimento, APF – Altura da planta no florescimento, NDR5 – Início da granação, AIPV – Altura da inserção da primeira vagem, NDR7 – Término da granação, PEG – Período de granação, NDM – Número de dias para a maturidade, APM – Altura da planta na maturidade, PG – Produtividade de grãos, MCS – Massa de 100 sementes, EST – Estande, VA – Valor agrônômico, AC – Acamamento, RF – Retenção foliar.

Tabela 6 - Importância relativa das 20 características avaliadas através da análise por variáveis canônicas

Variável canônica	Raiz (%)	Variância acumulada (%)	Importância relativa dos descritores nas variáveis canônicas																			
			X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉	X ₁₀	X ₁₁	X ₁₂	X ₁₃	X ₁₄	X ₁₅	X ₁₆	X ₁₇	X ₁₈	X ₁₉	X ₂₀
VC1	74,09	74,09	0,779	0,054	-0,265	0,040	1,212	-1,121	0,114	0,175	0,120	-0,167	0,063	-0,182	-0,127	-0,067	-0,030	0,022	0,039	0,364	0,197	-0,057
VC2	7,44	81,53	-0,108	-0,183	-0,961	-0,107	1,142	-1,485	-0,126	0,285	0,294	-0,366	0,192	-0,488	-0,887	0,795	0,544	0,394	-0,170	-0,049	0,023	0,044
VC3	4,05	85,58	0,296	0,114	-0,544	-0,033	0,201	-0,589	-0,210	-0,076	-0,028	-0,034	0,790	0,093	0,276	-0,277	-0,306	-0,630	-0,043	-0,134	0,055	0,037
VC4	2,88	88,46	0,199	0,006	-0,539	0,008	0,426	-0,545	0,260	0,316	0,102	0,552	0,165	0,129	-0,248	0,851	0,738	0,260	-0,202	0,063	-0,013	0,125
VC5	2,42	90,88	-0,152	-0,435	0,889	0,221	-0,571	1,111	-0,020	-0,215	0,383	0,355	0,427	-0,329	0,248	-0,228	0,159	0,188	0,102	-0,011	-0,151	-0,189
VC6	1,68	92,56	-0,512	0,350	0,295	0,010	0,389	0,037	0,226	-0,174	-0,345	-0,235	0,359	0,144	-0,225	0,411	-0,200	-0,461	0,329	0,019	-0,126	0,107
VC7	1,54	94,10	0,027	0,086	-0,509	-0,028	0,649	-0,496	0,130	0,323	-0,080	0,151	0,333	0,641	0,047	-0,523	-0,102	0,155	-0,321	-0,151	0,248	0,019
VC8	1,24	95,35	-0,259	-0,019	-0,435	0,177	0,669	-0,561	0,002	0,624	0,403	-0,340	0,059	-0,116	0,386	-0,130	-0,005	-0,315	-0,149	-0,085	0,258	-0,047
VC9	0,97	96,32	-0,177	0,211	-0,007	0,113	0,334	-0,741	-0,252	-0,021	0,312	-0,001	0,429	0,443	0,025	0,209	0,006	0,167	-0,122	0,208	-0,132	-0,340
VC10	0,88	97,20	-0,099	-0,097	-0,010	0,162	0,076	-0,408	0,206	0,129	-0,442	-0,210	0,112	0,313	1,010	-0,222	0,752	0,770	-0,250	-0,109	-0,101	0,155
VC11	0,61	97,81	0,349	-0,511	-1,352	0,071	1,507	-1,423	-0,010	0,089	0,142	-0,361	0,442	0,495	0,144	0,162	0,305	0,078	-0,053	0,190	-0,105	-0,417
VC12	0,47	98,28	0,130	-0,467	2,009	-0,057	-3,036	3,086	0,429	0,097	0,356	-0,127	0,037	0,197	-0,139	-0,038	-0,052	-0,058	0,202	0,197	0,191	0,405
VC13	0,40	98,69	-0,030	0,079	-0,605	0,182	0,841	-0,941	-0,258	0,254	-0,319	-0,012	-0,211	-0,249	-0,135	-0,282	0,144	-0,120	-0,170	0,779	-0,348	0,333
VC14	0,31	99,00	-0,016	0,449	1,710	0,277	-2,542	3,151	0,040	-0,562	0,309	-0,231	0,092	0,111	0,226	-0,100	0,220	0,325	-0,313	0,099	0,316	0,171
VC15	0,31	99,31	0,147	0,301	1,083	-0,198	-2,080	2,379	0,247	0,158	0,124	-0,054	-0,113	-0,090	0,219	-0,114	0,082	0,448	0,364	-0,289	-0,554	0,041
VC16	0,23	99,54	0,003	-0,054	-1,542	-0,656	2,648	-2,800	-0,464	0,070	0,305	-0,026	-0,050	0,044	-0,002	-0,129	0,212	-0,209	-0,179	-0,060	-0,155	0,393
VC17	0,17	99,71	-0,002	0,229	-1,997	-0,218	2,432	-3,054	0,543	-0,084	0,344	-0,135	-0,087	-0,209	-0,098	-0,189	0,206	-0,172	-0,455	-0,050	-0,095	-0,435
VC18	0,12	99,82	-0,030	0,040	-0,787	0,459	1,250	-1,460	-0,128	-0,138	-0,015	-0,071	-0,187	0,061	-0,049	-0,302	0,401	-0,075	0,326	-0,362	-0,127	-0,186
VC19	0,10	99,93	0,043	0,081	-1,422	-0,388	1,875	-2,133	0,017	-0,125	-0,338	0,115	0,093	0,129	0,164	-0,043	0,240	0,155	0,621	0,350	0,524	-0,294
VC20	0,07	100,00	0,041	-0,052	6,363	-0,151	-8,942	10,364	-0,155	0,125	-0,162	-0,012	-0,047	0,021	0,003	-0,096	0,230	-0,078	0,019	0,099	-0,066	-0,165

X₁ - Número de dias para o florescimento, X₂ - Altura da planta no florescimento, X₃ - Início da granação, X₄ - Altura da inserção da primeira vagem, X₅ - Término da granação, X₆ - Período de granação, X₇ - Número de dias para a maturidade, X₈ - Altura da planta na maturidade, X₉ - Produtividade de grãos, X₁₀ - Massa de 100 sementes, X₁₁ - Teor de proteína, X₁₂ - Teor de óleo, X₁₃ - Teor de ácido palmítico, X₁₄ - Teor de ácido esteárico, X₁₅ - Teor de ácido oleico, X₁₆ - Teor de ácido linoleico, X₁₇ - Estande, X₁₈ - Valor agrônômico, X₁₉ - Acamamento, X₂₀ - Retenção foliar.

4.2 Avaliação genético-molecular

Os 114 locos SSR genotipados nos 89 acessos resultaram em 905 alelos, com média de 7,94 alelos/loco. O número de alelos por loco variou entre dois (AW781285) e 17 (Satt373 e Satt458). O conteúdo de polimorfismo informativo dos marcadores (PIC) é resumido na Tabela 7.

Tabela 7 - Número de alelos e conteúdo de polimorfismo informativo (PIC) dos 114 locos utilizados no mapeamento

Grupo de ligação	Loco	PIC	Número de alelos	Grupo de ligação	Loco	PIC	Número de alelos
A1	Satt545	0,79	10	F	Satt334	0,73	7
A1	Satt236	0,73	8	F	SOYHSP176	0,71	7
A1	Satt174	0,61	5	F	Satt335	0,54	6
A1	Satt599	0,39	3	F	BE806387	0,37	4
A2	Satt329	0,85	11	F	Sct_188	0,37	3
A2	Satt429	0,77	10	G	Satt352	0,80	7
A2	Satt177	0,70	6	G	Satt303	0,78	10
A2	Satt187	0,62	5	G	Satt191	0,76	7
A2	Satt390	0,54	5	G	Satt038	0,75	9
B1	Satt197	0,85	13	G	Satt115	0,74	8
B1	Satt453	0,66	9	G	Satt012	0,74	12
B2	Satt534	0,89	15	G	Satt570	0,63	5
B2	Satt070	0,78	10	G	Satt163	0,59	6
B2	Satt168	0,77	7	G	Satt309	0,49	4
B2	Satt577	0,72	5	G	Satt324	0,39	7
B2	Satt556	0,72	9	H	Satt216	0,86	13
B2	Satt020	0,66	7	H	Satt192	0,80	9
B2	Satt063	0,64	7	H	Satt279	0,76	8
C1	Satt194	0,74	7	H	Satt302	0,50	5
C1	Satt565	0,55	8	H	Satt222	0,48	6
C1	AI794821	0,55	3	I	Satt270	0,80	10
C1	SOYGPART	0,42	3	I	Satt354	0,79	12
C2	Satt371	0,83	10	I	Sct_189	0,75	11
C2	Satt460	0,77	10	I	Satt239	0,70	10
C2	Satt277	0,76	10	I	Satt571	0,56	4
C2	Satt307	0,76	8	I	Satt292	0,37	2
C2	Satt286	0,72	9	J	Satt224	0,89	14
C2	Satt557	0,71	11	J	Satt414	0,80	10
C2	Satt202	0,57	6	J	Satt431	0,76	9
C2	AW734043	0,56	5	J	AW310961	0,55	6
C2	Satt227	0,50	4	J	Satt215	0,38	5
C2	Satt578	0,38	4	K	Satt242	0,81	8
C2	Satt432	0,27	5	K	Satt588	0,73	10
D1a	Satt179	0,79	9	K	Satt001	0,63	9
D1a	Satt184	0,78	9	K	Satt196	0,60	8
D1a	Satt129	0,74	10	K	Satt102	0,55	6
D1a	Satt147	0,72	12	L	Satt373	0,89	17
D1a	Satt071	0,52	3	L	Satt462	0,82	10
D1a	Satt468	0,49	5	L	Satt523	0,74	8
D1a	Satt203	0,38	6	L	Satt156	0,72	9
D1a	AW781285	0,36	2	L	Satt448	0,53	8
D1b	Satt005	0,77	10	L	Satt182	0,47	5
D1b	Satt600	0,77	9	M	Satt463	0,78	9
D1b	Satt274	0,71	7	M	Satt346	0,76	7
D1b	Satt266	0,57	6	M	Satt540	0,74	11
D2	Satt458	0,91	17	N	Satt009	0,85	11
D2	Satt002	0,79	8	N	Satt530	0,79	9
D2	Satt186	0,75	9	N	Satt022	0,76	8
D2	Satt389	0,67	7	N	Satt080	0,57	6
D2	Satt226	0,67	9	N	Satt266	0,57	6
D2	Satt386	0,67	7	N	Satt387	0,42	4
D2	Satt574	0,65	8	O	Satt173	0,87	14
E	Satt230	0,62	7	O	Satt345	0,80	10
F	Satt146	0,87	13	O	Satt487	0,79	10
F	Satt030	0,84	10	O	Satt243	0,73	8
F	Satt586	0,80	8	O	Satt262	0,57	7
F	Satt510	0,74	9	O	BF008905	0,33	3

A distância média entre os marcadores nos grupos de ligação foi de aproximadamente 40 cM, sendo o grupo de ligação C2 o mais saturado (11 marcadores). O grupo de ligação E apresentou somente um marcador, de modo que este grupo não foi incluído em análises posteriores para verificação do grau de decaimento do desequilíbrio de ligação (DL).

Os dados referentes a genotipagem dos acessos utilizando os 114 marcadores informativos foram submetidos a análise de estrutura populacional sem hierarquização *a priori*. Com base na abordagem bayesiana implementada pelo software STRUCTURE, utilizou-se o critério proposto por Evanno et al. (2005) para definição do número de subpopulações mais provável (k). Por esta metodologia, verificou-se um valor ótimo como sendo k=2 (Figura 5). Consequentemente, os acessos foram então submetidos ao teste de atribuição, considerando o painel associativo subdividido em dois grandes grupos de acessos (Figura 6). Esta mesma subdivisão foi utilizada para obtenção da matriz Q, a qual foi posteriormente incorporada no modelo misto utilizado para o teste de associação entre marcadores e fenótipos.

Figura 5 - Valores de Δk para cada valor de k, calculado de acordo com o proposto por Evanno et al. (2005). O maior valor de Δk corresponde ao número ótimo de subgrupos

Figura 6 – Teste de atribuição para os acessos de soja avaliados ($k=2$). Diferentes genótipos representados por barras verticais, sendo os de mesma cor pertencentes ao mesmo grupo. Diferentes cores em um mesmo indivíduo indicam a porcentagem do genoma compartilhado com cada grupo

4.3 Divergência genética

A correlação matricial entre os dados de distância genética baseada em dados fenotípicos (Mahalanobis) e em dados de marcadores (Rogers-W) foi de 0,247. Embora baixa, a correlação mostrou-se significativa a 1% de probabilidade pelo teste de Mantel, baseado em 1000 simulações.

As matrizes de distância foram somadas e a matriz resultante utilizada para agrupamento hierárquico dos acessos pelo método de Ward (Figura 7). Este procedimento foi adotado com intuito de melhor caracterizar o painel associativo em relação à diversidade genética. Esta metodologia proporcionou uma maior separação dos acessos quando comparada com os grupos formados pelo método STRUCTURE. Contudo, verifica-se que os dois principais agrupamentos previamente sugeridos foram mantidos. Com base nas duas metodologias, verifica-se a inexistência de relação entre a divergência genética e a origem geográfica dos materiais.

Figura 7 - Agrupamento hierárquico pelo método de Ward, baseado em dados fenotípicos e de marcadores SSR

4.4 Desequilíbrio de ligação

A inspeção do padrão de decaimento do DL foi realizada graficamente, demonstrando valores de r^2 significativos ($p\text{-valor} < 0,01$) entre pares de marcadores. Para tal, os valores de r^2 foram relacionados com a distância genética entre os marcadores (cM) (Figura 8). Com base na linha de tendência central do gráfico de dispersão, verifica-se um lento decaimento do desequilíbrio de ligação conforme o aumento da distância genética, com marcadores apresentando considerável DL até a distância de 30 cM.

Como estimativa média do tamanho dos blocos de ligação, adotou-se a distância genética onde ocorre o cruzamento da linha de regressão logarítmica com o valor de $r^2 = 0,05$. Considerando este limiar, verifica-se que o tamanho dos blocos haplotípicos para o painel associativo em questão estende-se por uma distância de aproximadamente 20 cM. Com base nestes resultados, considera-se viável o mapeamento de QTL's com menor número de marcadores, uma vez a maior saturação do genoma teria poucas implicações no aumento da resolução do mapeamento (FLINT-GRACIA et al., 2003).

Figura 8 - Decaimento do desequilíbrio de ligação conforme a distância genética (cM) entre pares de marcadores utilizados no mapeamento associativo

4.5 Mapeamento associativo

Os dados de genotipagem foram utilizados para inferência a cerca da estrutura genética populacional e estimativa do grau de parentesco entre os materiais. As matrizes Q (estrutura populacional) e a matriz K (parentesco) obtidas utilizando os softwares STRUCTURE e SPAGeDI foram então inseridas como covariáveis no modelo misto para determinar as associações entre marcadores e fenótipos.

Foram verificadas 505 associações entre os marcadores e as 20 características avaliadas, considerando inicialmente um p-valor $< 0,01$. Com a aplicação da correção dos p-valores para múltiplos testes de hipóteses, verificou-se uma redução para 392 associações com base nos valores significativos de FDR $< 0,01$. Utilizando o método do fator Bayesiano mínimo, a quantidade total de associações que apresentaram forte evidência para rejeição da hipótese de nulidade ($\text{minBF} < 0,02$) foi reduzida para 285 (Anexo A). Este último método tem sido proposto na literatura como alternativa adequada para estudos de mapeamento associativo, de modo que foi a metodologia de correção selecionada para determinar as associações válidas.

Dos 114 marcadores SSR genotipados, 78 (68%) foram responsáveis pelas 285 associações com todas as 20 características avaliadas, sendo que 29 marcadores apresentaram-se associados exclusivamente a uma característica. Os demais locos mostraram-se associados com duas ou mais características, sendo que alguns foram relacionados com até 12 caracteres fenotípicos avaliados (ex: Satt 070 e Satt 371, localizados nos grupos de ligação B2 e C2, respectivamente). A característica com maior quantidade de marcadores associados foi a altura de plantas no florescimento, com 38 marcadores, seguida dos caracteres relacionados com a maturidade (número de dias para a maturidade e para o florescimento, com 36 e 31 marcadores respectivamente) (Tabela 8). Considerando a natureza pleiotrópica de alguns locos, pode-se verificar que diversos locos em comum associados com características correlacionadas, como por exemplo, a alta congruência verificada entre os marcadores associados no controle das características de maturidade e altura.

Considerando as principais características quantitativas avaliadas, foram verificados 18 marcadores associados a produtividade de grãos, 13 a massa de 100 sementes, 5 ao teor de proteínas e 15 relacionados com o teor de óleo total (Tabela 9).

Tabela 8 - Quantidade total de marcadores associados a cada uma das 20 características

Características	Unidade	Quantidade de marcadores associados
Período de granação	dias	1
Estande	notas	1
Retenção foliar	notas	1
Valor agrônômico	notas	2
Acamamento	notas	2
Teor de proteína	%	5
Teor de ácido oleico	%	5
Massa de 100 sementes	g	13
Teor de ácido esteárico	%	13
Teor de ácido linoleico	%	13
Teor de ácido palmítico	%	14
Teor de óleo total	%	15
Altura da planta na maturidade	cm	17
Produtividade de grãos	kg/ha	18
Início da granação	dias	20
Altura da inserção da primeira vagem	cm	20
Término da granação	dias	20
Número de dias para o florescimento	dias	31
Número de dias para a maturidade	dias	36
Altura da planta no florescimento	cm	38
Total de associações		285

Tabela 9 - Marcadores associados às principais características quantitativas avaliadas

Característica	Marcador	GL	$-e^*p*\ln(p)$ ($<0,02$)	Característica	Marcador	GL	$-e^*p*\ln(p)$ ($<0,02$)
MCS	Satt329	A2	2,89E-03	PG	Satt038	G	6,23E-03
MCS	Satt429	A2	5,41E-03	PG	Satt270	I	4,98E-03
MCS	Satt197	B1	1,26E-02	PG	Satt448	L	8,53E-05
MCS	Satt070	B2	1,37E-11	PG	Satt373	L	1,04E-02
MCS	Satt063	B2	4,91E-04	PG	Satt387	N	1,88E-02
MCS	Satt556	B2	1,70E-02	Proteína	Satt168	B2	1,88E-02
MCS	Satt203	D1a	2,48E-06	Proteína	Satt578	C2	2,60E-03
MCS	Satt468	D1a	6,43E-03	Proteína	Satt286	C2	2,90E-03
MCS	Satt386	D2	1,76E-02	Proteína	Satt335	F	8,91E-03
MCS	SOYHSP176	F	6,23E-04	Proteína	Satt224	J	1,57E-09
MCS	Satt191	G	2,84E-06	Óleo	Satt070	B2	2,53E-03
MCS	Satt038	G	1,31E-02	Óleo	Satt371	C2	9,12E-09
MCS	Satt192	H	1,81E-02	Óleo	Satt557	C2	1,81E-05
PG	Satt236	A1	1,38E-04	Óleo	Satt578	C2	4,66E-03
PG	Satt329	A2	1,68E-02	Óleo	Satt307	C2	5,12E-03
PG	Satt070	B2	4,04E-15	Óleo	Satt432	C2	1,13E-02
PG	Satt063	B2	1,46E-02	Óleo	Satt179	D1a	2,57E-06
PG	Satt371	C2	2,92E-06	Óleo	Satt203	D1a	2,19E-04
PG	Satt557	C2	2,03E-04	Óleo	Satt458	D2	4,20E-03
PG	Satt307	C2	5,09E-04	Óleo	Satt186	D2	7,79E-03
PG	Satt203	D1a	6,17E-10	Óleo	Satt146	F	1,40E-02
PG	Satt129	D1a	1,67E-02	Óleo	Satt270	I	9,23E-06
PG	Satt458	D2	2,01E-03	Óleo	Satt001	K	2,14E-03
PG	Satt030	F	5,74E-03	Óleo	Satt448	L	8,49E-06
PG	Satt146	F	1,75E-02	Óleo	Satt540	M	5,89E-03
PG	Satt163	G	2,34E-07				

MCS – Massa de 100 sementes, PG – Produtividade de grãos, Proteína – Teor de proteínas, Óleo – Teor de óleo.

5 DISCUSSÃO

A intensificação no uso dos recursos genéticos vegetais depende, no longo prazo, do conhecimento acerca da biodiversidade armazenada em bancos de germoplasma. O baixo uso da variabilidade genética em programas de melhoramento freia o progresso no desenvolvimento de variedades mais produtivas e de melhor qualidade, bem como tolerantes a estresses bióticos e abióticos (NASS; SIGRIST, 2010).

O processo de endogamia existente em espécies autógamas como a soja, resulta naturalmente em uma diminuição da diversidade genética, devido a eliminação de alelos deletérios juntamente com alelos favoráveis em desequilíbrio de ligação. Somada ao aspecto reprodutivo *per se*, o processo de domesticação, seleção artificial e eventos fundadores (gargalos genéticos) afetam diretamente o nível de variação genética, alterando as frequências alelicas, aumentando o DL e eliminando alelos raros e potencialmente benéficos das populações subsequentes (HOLLAND, 2004; HYTEN et al., 2006).

Considerando a necessidade de variabilidade adequada para futuros programas de melhoramento e visando identificar novas potencialidades em materiais exóticos ou não-adaptados, o mapeamento associativo apresenta-se como importante ferramenta. Uma vez que a técnica não exige o desenvolvimento de populações de mapeamento, existe a possibilidade de realizar a triagem de diversos acessos em busca de genótipos ou características promissoras. Interessante notar que, uma infinidade de alelos favoráveis podem estar presentes em genótipos pouco interessantes, cujo *background* genético seja desfavorável. Deste modo, apenas a seleção fenotípica é insuficiente para melhor aproveitamento da variabilidade genética disponível (TANKSLEY; MACCOUGH, 1997).

O conjunto de genótipos utilizados para compor o painel associativo mostrou ampla variabilidade. Observou-se que os marcadores utilizados exibiram considerável nível de polimorfismo, permitindo a detecção de diversos alelos raros. Em termos fenotípicos, vários materiais mostraram-se superiores em relação às testemunhas para todas as características avaliadas, sugerindo grande potencial para uso em blocos de cruzamento e desenvolvimento de novas populações.

Para estimativa da divergência genética e melhor caracterização dos materiais componentes do painel, as distâncias genéticas foram submetidas ao agrupamento hierárquico utilizando o método de Ward. De acordo com Odong et al. (2011), o método apresenta alta correlação cofenética com outros métodos de agrupamento, com destaque para uma melhor recuperação dos grupos. Consequentemente, trata-se de um método eficaz para caracterizar a

divergência genética em coleções de germoplasma, permitindo estabelecer coleções nucleares e selecionar cruzamentos promissores para programas de introgressão ou para estudos de associação.

O padrão de divergência genética verificado com base nas distâncias fenotípicas e de marcadores sugere pouca correlação entre o grau de divergência genética e a origem dos acessos, resultado observado previamente por Mulato et al. (2010) avaliando conjunto similar de materiais. Uma provável explicação seria que, embora os materiais possuam histórias de seleção distintas, o país de origem do material pode não refletir o local onde a seleção foi praticada, mas apenas o país de origem do banco de germoplasma doador do material.

Por sua vez, os genótipos comerciais brasileiros incluídos no painel para efeito comparativo, apresentaram menor distância genética, formando um grupo facilmente distinguível no dendrograma final. Essa alta similaridade entre os acessos brasileiros era esperada, considerando que a base genética do germoplasma brasileiro é estreita e que aproximadamente metade do conjunto gênico é originada a partir de apenas quatro genótipos ancestrais (PRIOLLI et al., 2002; WYSMIERSKI, 2010).

De modo a superar a limitação de pouca variabilidade, diversos estudos abordaram o emprego de PI's para incremento da base genética de populações de soja (NARVEL et al., 2000; SEBOLT et al., 2000; SMALLEY et al., 2004, FASOULA et al., 2004; KABELKA et al., 2004; GUZMAN et al., 2007) Em geral, foi verificada grande vantagem no emprego de genótipos exóticos e não-adaptados, tanto no aumento da variância genética quanto na média das populações. Em alguns casos, utilizando marcadores microsatélites para mapear QTL's relacionados à produtividade de grãos, teor de óleo e proteína, diversos novos alelos foram identificados, permitindo encontrar alelos de interesse agrônômico em *backgrounds* genéticos desfavoráveis. Mais além, a identificação de marcadores relacionados à QTL's permitiu acelerar a introdução dos novos alelos em materiais elite por meio da seleção assistida por marcadores.

O resultado dos avanços no mapeamento de QTL's nos últimos anos encontram-se disponíveis no SoyBase, onde é possível visualizar os mapas genéticos e os QTL's possivelmente associados. Contudo, conforme ressaltado neste trabalho, a maioria dos estudos empregou a abordagem de mapeamento clássico por ligação, avaliando populações oriundas de cruzamentos biparentais. A principal limitação desta abordagem é a baixa resolução, uma vez que poucas gerações de recombinação podem ser avaliadas.

Em soja, até o momento, existem cinco trabalhos publicados nos quais foi empregado o mapeamento associativo, utilizando diferentes metodologias. Wang et al. (2008) e Jun et al.

(2008) avaliaram a variação nas frequências alélicas em locos para identificar QTL's relacionados com a tolerância a deficiência em ferro e teor de proteínas. Nesta técnica, os diferentes materiais foram agrupados em populações caso e controle, semelhante aos primeiros estudos envolvendo o mapeamento associativo em humanos. Contudo, esta abordagem apresenta como principal limitação a possibilidade de associações espúrias, por desconsiderar a estrutura genética da população.

Os demais estudos associativos buscaram solucionar os problemas relacionados à estrutura populacional. Hou et al. (2011) empregou a abordagem de modelos mistos generalizados ao incorporar a informação da estrutura populacional para avaliar caracteres em soja vegetal (*edamame*). Atualmente, a metodologia mais aceita para mapeamento associativo é o uso de modelos mistos, onde são incorporadas tanto a informação sobre a estrutura populacional quanto a de parentesco entre indivíduos (YU et al., 2006). Esta abordagem foi utilizada em soja no presente trabalho, por Shi et al. (2010) e Li et al. (2011), com intuito de detectar QTL's. Nestes últimos exemplos, buscou-se relacionar marcadores ao teor de óleo, proteína, estresses bióticos e abióticos.

Embora esses estudos de mapeamento associativo em soja existentes na literatura tenham promovido avanços, uma limitação para interpretação dos resultados publicados reside na ausência de correção para múltiplos testes. O uso de p-valores tem sido o método mais utilizado no meio científico para medir o grau de evidência para falsear hipóteses nulas. Contudo, sem o ajuste para múltiplos testes, o limiar usual estabelecido para p-valores $< 0,01$ tende a inflacionar as taxas de erro tipo I (YU et al., 2006).

Considerando isso, diversos métodos foram propostos (HELD, 2010). Para testar a significância das associações entre marcadores e fenótipo utilizando o modelo linear misto (MLM), o número de associações obtidas considerando p-valores $< 0,01$ foi comparado com as obtidas empregando o método do *false discovery rate* ($FDR < 0,01$) e do *minimum bayes factor* ($\min BF < 0,02$) (BENJAMINI; HOCHBERG, 1995; GOODMAN, 2001).

O método FDR é definido como a proporção esperada de hipóteses nulas verdadeiras dentro do conjunto de hipóteses nulas rejeitadas. Por sua vez, o $\min BF$ consiste no reescalonamento dos p-valores, de modo a aumentar a evidência para rejeição da hipótese nula (HELD, 2010). Para interpretação, os valores obtidos são interpretados com base no grau de evidência para rejeitar a hipótese nula de não associação, sendo que valores de $\min BF < 0,02$ indicam fortes evidências para rejeição (GOODMAN, 2001).

Com base na redução do número de associações, julgou-se o método do $\min BF$ como o que melhor representaria as associações verdadeiras, procedimento similar ao adotado por

Abdurakhmanov et al. (2008). Vale ressaltar que, até o momento, não foram propostas soluções definitivas para melhor correção de múltiplos testes oriundos do MLM empregado no mapeamento associativo.

O uso de MLM, incorporando informações sobre a estrutura genética dos acessos e grau de parentesco aliado a correção para múltiplos testes, são procedimentos essenciais para controlar possíveis associações espúrias no mapeamento associativo. Uma vez obtidas associações verdadeiras com maior grau de confiabilidade, o próximo passo consiste em testar as associações encontradas em diferentes locais, anos e em populações distintas, com diferentes *backgrounds* genéticos (HOU et al., 2011). No presente trabalho, o mapeamento associativo foi realizado a partir de dados fenotípicos obtidos em um único local e população, de modo que as associações encontradas deverão ser confirmadas em oportunidades futuras (Anexo A).

Outra forma de verificar a veracidade das associações consiste em estabelecer comparações com QTL's previamente identificados por estudos empregando o mapeamento por ligação, tais como os disponibilizados na base de dados SoyBase. Para tanto, foram selecionados todos os 36 marcadores SSR associados às características produtividade de grãos, massa de 100 sementes e teor de óleo e proteína (Tabela 10). Dentre as 51 associações entre marcador e fenótipo consideradas, 15 (~30%) foram previamente descritas utilizando o mapeamento convencional.

A partir das comparações, podem-se verificar algumas incongruências entre as duas abordagens. Por exemplo, algumas associações obtidas através do mapeamento de populações oriundas de cruzamentos biparentais não foram detectadas pelo mapeamento associativo. Isto decore do fato de que, no mapeamento associativo, determinado loco deve expressar seu efeito em diversos genótipos para ser detectado, ao passo que no mapeamento via análise de ligação este mesmo loco será detectado caso não haja segregação de outros locos (WANG et al., 2008).

Por outro lado, também foram detectadas associações pelo mapeamento associativo que não foram detectadas no mapeamento por análise de ligação. Uma provável explicação para estes casos reside na capacidade do mapeamento associativo em explorar eventos de recombinação ancestrais, permitindo acessar eventos de recombinação e regiões genômicas “ocultas” em populações oriundas de cruzamentos biparentais.

Tabela 10 - Comparação das associações obtidas por mapeamento associativo e por ligação publicados no *SoyBase*. Em negrito, destaque para os marcadores associados à mesma característica em ambas as abordagens

Marcador	CARACTERÍSTICAS	
	Mapeamento associativo	Mapeamento por ligação (<i>Soybase</i>)
Satt373	PG	Produtividade de grãos, florescimento, altura de planta, teor de proteína, teor de óleo
Satt387	PG	Produtividade de grãos, resistência à mofo-branco, altura de plantas, maturidade
Satt557	PG	Produtividade de grãos, maturidade, resistência à fitóftora, altura de plantas
Satt030	PG	N/D
Satt038	PG	Resistência a nematóides, teor de ácido linolênico
Satt063	PG	Eficiência no uso de ferro, caracteres foliares, teor de proteínas, ácido linoleico, ácido linolênico
Satt070	PG	Eficiência no uso de ferro, resistência a mofo-branco, massa de sementes
Satt129	PG	Resistência à mofo-branco
Satt146	PG	Altura de plantas, caracteres foliares
Satt163	PG	Massa de sementes, resistência a nematóides, ácido oleico
Satt203	PG	Teor de óleo
Satt236	PG	Caracteres foliares
Satt270	PG	Teor de óleo, teor de proteína
Satt307	PG	Resistência à fitóftora, maturidade, massa de sementes, comprimento de internós
Satt329	PG	Morfologia de sementes
Satt371	PG	Resistência a nematóide, tolerância à alumínio, altura de plantas
Satt448	PG	N/D
Satt458	PG	Resistência a mofo-branco, teor de óleo, massa de sementes, tolerância a estresse hídrico
Satt070	MCS	Massa de sementes, eficiência no uso de ferro, resistência a mofo-branco, ácido palmítico
Satt197	MCS	Massa de sementes, produtividade de grãos, resistência a mofo-branco, caracteres foliares
Satt329	MCS	Massa de sementes
Satt063	MCS	Eficiência no uso de ferro, caracteres foliares, teor de proteínas, ácido linoleico, ácido linolênico
Satt038	MCS	Resistência a nematóides, teor de ácido linolênico
Satt191	MCS	Resistência a mofo-branco, altura de plantas, produtividade de grãos, teor de óleo, acamamento
Satt192	MCS	N/D
Satt203	MCS	Teor de óleo
Satt386	MCS	N/D
Satt429	MCS	Maturidade
Satt468	MCS	Teor de óleo, morfologia da flor
Satt556	MCS	Uso de nitrogênio, teor de óleo, caracterese foliares, produtividade de grãos, maturidade
SOYHSP176	MCS	Altura de plantas, taxa de aborto de sementes
Satt179	Óleo	Teor de óleo, massa de sementes
Satt270	Óleo	Teor de óleo, teor de proteína
Satt203	Óleo	Teor de óleo
Satt458	Óleo	Teor de óleo, resistência a mofo-branco, massa de sementes, estresse hídrico
Satt540	Óleo	Teor de óleo, produtividade de grãos, teor de proteína
Satt432	Óleo	Teor de óleo
Satt001	Óleo	N/D
Satt070	Óleo	Eficiência no uso de ferro, resistência a mofo-branco, massa de sementes, ácido palmítico
Satt146	Óleo	Altura de plantas, caracteres foliares
Satt186	Óleo	Produtividade de grãos, maturidade
Satt307	Óleo	Resistência à fitóftora, maturidade, massa de sementes, comprimento de internós
Satt371	Óleo	Resistência a nematóide, tolerância à alumínio, altura de plantas
Satt448	Óleo	N/D
Satt557	Óleo	Produtividade de grãos, maturidade, resistência à fitóftora, altura de plantas
Satt578	Óleo	Maturidade, qualitativos da semente, teor de proteína
Satt168	Proteína	Teor de proteína, resistência a nematóide, uso de nitrogênio, produtividade de grãos
Satt335	Proteína	Teor de proteína, acamamento, resistência à mofo-branco, altura de plantas, maturidade
Satt578	Proteína	Teor de proteína, maturidade, qualitativos da semente
Satt224	Proteína	Resistência a nematóides
Satt286	Proteína	Produtividade de grãos, maturidade

N/D – Informação não disponível, PG – Produtividade de grãos, MCS – Massa de 100 sementes, Óleo – Teor de óleo, Proteína – Teor de proteína.

Um ponto interessante verificado foi a existência de locos associados a mais de uma característica. Considerando a natureza complexa de caracteres quantitativos, existe a possibilidade de esses locos apresentarem um efeito pleiotrópico sobre estas características, o que acaba sendo refletido nas correlações fenotípicas entre os caracteres. Por exemplo, as características produtividade grãos, massa de 100 sementes e teor de óleo apresentaram locos comuns associados (ex: Satt070) e também apresentaram correlações fenotípicas positivas e significativas.

Associações em comum também foram verificadas para características que apresentam correlação fenotípica negativa. O loco Satt578 apresentou-se associado ao teor de óleo e teor de proteínas, sendo que estes caracteres apresentaram um alto índice de correlação negativa ($-0,67$), índice este idêntico ao apresentado por Shi et al. (2010). Vale ressaltar que a associação detectada entre marcador e fenótipo não indica qual a direção desta associação. Consequentemente, é provável que os alelos responsáveis pelo aumento do teor de óleo neste loco, sejam os mesmos relacionados com a diminuição de proteína. Embora estas características representem um desafio ao melhoramento genético, marcadores como o Satt578, permitem flexibilidade para uso em programas de seleção assistida de diferentes objetivos (SHI et al., 2010).

Uma vez detectadas as associações com elevado grau de confiança, a próxima informação de interesse para programas de seleção assistida por marcadores é a identificação de quais alelos em determinado loco estão relacionados com as alterações fenotípicas de interesse. Alguns métodos foram propostos para este fim, sendo o mais simples o realizado por Wang et al. (2008). Esta técnica consiste em avaliar as médias fenotípicas de cada alelo do loco associado, de modo a identificar os alelos que contribuem com a variação verificada. A metodologia é facilitada no caso de espécies autógamas, dado que o genótipo é representado por apenas um alelo em linhagens puras.

Para avaliar com maior profundidade alguns locos associados à produtividade de grãos, massa de 100 sementes, teor de óleo e teor de proteína, e exemplificar a aplicação prática dos resultados para a seleção assistida por marcadores (SAM), foi empregado a técnica realizada por Wang et al. (2008) (Figura 9). Os locos utilizados para representação foram selecionados por não estarem associados à outra característica (efeito pleiotrópico), além de serem locos confirmados pelos dados de mapeamento por ligação disponíveis no SoyBase. Os resultados, embora apresentem implicações práticas para a SAM, devem ser considerados com precaução uma vez que alguns dos alelos amostrados são raros no painel associativo. Consequentemente, as médias foram obtidas com base em quantidade de dados diferentes.

Contudo, no aspecto geral, esses resultados fornecem indícios preliminares acerca dos efeitos alélicos destes locos sobre os fenótipos observados.

Figura 9 - Médias das características relativas a cada alelo nos locos associados à produtividade de grãos, massa de 100 sementes, teor de óleo e proteína

Outra abordagem foi proposta por Jun et al. (2008), onde os materiais avaliados foram separados em grupos com maior e menor teor de proteína, para posterior análise da variação da frequência dos alelos entre estes grupos. Breseguelo e Sorrels (2006) propuseram a comparação com alelos nulos (raros e/ou ausentes) para cada loco, a fim de estimar os efeitos dos alelos observados sobre o fenótipo avaliado. Outra técnica interessante consiste no emprego de modelos de regressão passo-à-passo (*stepwise*), com objetivo de estimar o valor teórico máximo de um fenótipo, caso todos os alelos favoráveis estivessem presentes nos diferentes locos (KRAAKMAN et al., 2004). Neste último caso, é possível estabelecer uma combinação ótima de alelos para seleção assistida por marcadores.

Finalmente, a caracterização do desequilíbrio de ligação (DL) é outra importante informação gerada por este trabalho para que as associações sejam testadas em trabalhos futuros. Dentre as implicações, pode-se citar a maior eficiência na detecção e confirmação das associações entre marcadores e fenótipo, indicando a quantidade mínima de marcadores necessária para saturação do genoma.

Em situações onde o DL é alto, o mapeamento amplo do genoma (GWAS) é facilitado devido a exigência de menor quantidade de marcadores para se atingir uma resolução aceitável. Por outro lado, DL pouco extensos favorecem a abordagem associativa através de genes candidatos, uma vez que tal situação exigirá um grande número de marcadores. Usualmente, considera-se em DL significativo pares de marcadores que apresentam $r^2 < 0,01$. Esses valores, quando contrastados graficamente com a distância genética entre os marcadores, fornecem uma estimativa do tamanho dos blocos de ligação presente na população em questão.

Na literatura, é possível encontrar diversas estimativas para o tamanho dos blocos de ligação em soja, com valores entre 2 cM até 50 cM dependendo da população em análise (HYTEN et al., 2006; ABDURARHMONOV; ABDUKARIMOV, 2008). Sabe-se que diversos fatores intrínsecos à população podem moldar o DL verificado, como por exemplo, a estrutura populacional, efeito fundador (deriva genética), grau de parentesco entre os materiais e seleção. Hyten et al. (2006) exemplificaram bem estas situações ao comparar o decaimento do DL em diferentes regiões genômicas de diferentes populações (Figura 10). Com base nos resultados, verifica-se a existência de blocos de haplótipos mais extensos em populações de base genética estreita.

Figura 10 – Decaimento do desequilíbrio de ligação em diferentes populações de soja e em diferentes regiões genômicas (Adaptado de HYTEN et al., 2007)

O DL de ligação encontrado para o painel associativo deste trabalho estendeu-se por uma distância de aproximadamente 20 cM ($r^2 < 0,05$). Considerando que o comprimento estimado do genoma da soja encontra-se estimado em 1.115 mpb (mega pares de base), e que 1 cM corresponde a aproximadamente 400 kpb (kilo pares de base), tem-se que para o painel utilizado no presente estudo seriam necessários aproximadamente 140 marcadores para cobertura genômica (HYTEN et al. 2007; SCHMUTZ et al., 2010). Por outro lado, considerando como o limiar para estimativa do tamanho dos blocos o valor de $r^2 < 0,01$ tem-se que a extensão do DL estaria abaixo de 2 cM. Neste caso, a estimativa dos blocos se assemelha ao obtido por Hyten et al. (2007) para a população composta por raças locais (*landraces*), população cuja natureza genética assemelha-se à utilizada presente estudo.

Com base em um comprimento de 2 cM dos blocos haplotípicos, seriam necessários aproximadamente 1.300 marcadores para saturação completa do genoma e permitir estudos de associação genômica ampla (GWAS). A genotipagem de populações com esta quantidade de marcadores pode ser contemplada atualmente utilizando o *Universal Soy Linkage Panel 1.0* (HYTEN et al., 2010). O painel consiste em um conjunto de 1.536 marcadores polimórficos do tipo SNP, com ampla capacidade de automação para genotipagem de 768 amostras de DNA por semana com custo relativamente baixo.

O mapeamento associativo em soja, embora recente, tem apresentado resultados promissores. Considerando a necessidade crescente de melhor caracterizar e identificar novos

locos e alelos presentes em germoplasma, esta abordagem tende a promover o uso dos recursos genéticos em programas de melhoramento.

Com a crescente automação do processo de genotipagem e consequente disponibilidade de informação genotípica de diversos materiais, estudos associativos poderão futuramente ser realizados com base na meta-análise de dados de marcadores. Com isso, os pesquisadores envolvidos poderão voltar suas atenções para o desenvolvimento de plataformas de fenotipagem mais eficientes, ou então aprimorar as metodologias de análise e validação de marcadores atualmente disponíveis.

Este trabalho evidencia as potencialidades do uso do mapeamento associativo, além de fornecer subsídios para implantação da seleção assistida em bancos de germoplasma e programas de melhoramento. Contudo, os resultados devem ser considerados preliminares, devendo ser comprados com outros estudos abordando diferentes locais, anos e populações. Mais além, sugere-se o uso de outros tipos de marcadores com maior capacidade de automação, promovendo maior saturação do mapa genético.

Nos últimos anos, a identificação de regiões genômicas envolvidas no controle de características complexas tem passado por grandes avanços. Novas tecnologias, possibilitando ampliar exponencialmente a geração de dados genéticos, aliadas ao desenvolvimento da capacidade de análise e processamento de dados, tornaram a o mapeamento associativo e a seleção genômica ampla uma realidade atual. O próximo passo consiste em ampliar a capacidade atual de fenotipagem, visando acompanhar os avanços recentes na genética molecular. Em última análise, somente a completa sinergia entre ambas poderá garantir a produtividade futura.

6 CONCLUSÕES

- As análises de agrupamento utilizando diferentes metodologias não foram equivalentes com a suposta origem geográfica dos genótipos. Tal fato sugere que esta origem pode estar mais relacionada com o país do banco de germoplasma doador, do que com o local onde ocorreu a história seletiva do genótipo.
- A extensão do DL no painel associativo foi estimada entre 2 cM ($r^2 < 0,1$) e 20 cM ($r^2 < 0,05$). Considerando o tamanho do genoma da soja publicado recentemente, avalia-se ser possível obter uma saturação genômica aceitável com as plataformas de genotipagem atualmente disponíveis.
- Algumas associações entre marcador e fenótipo foram corroboradas por metodologias distintas, disponíveis na literatura. Esta convergência de resultados fornece certa evidência de que tais associações sejam verdadeiras. Por outro lado, novas associações foram propostas, devendo ser posteriormente investigadas utilizando dados de diferentes locais, anos e populações.
- O painel associativo apresentou ampla variabilidade genética fenotípica e molecular, indicando o potencial destes materiais para obter novas combinações de alelos favoráveis, além de promover a ampliação da base genética para seleção. Por exemplo, verifica-se que acessos mais produtivos e com alto teor de óleo encontram-se geneticamente distantes na análise de agrupamento, permitindo a combinação destas características em programas de melhoramento. Neste caso, a seleção nas populações segregantes pode ser auxiliada pela variação alélica dos marcadores associados (ex: Satt373 e Satt179), após validação dos mesmos.
- Dentre as metodologias de testes de hipóteses para controle de falsos positivos, o *minimum bayes factor* (minBF) mostrou-se como o mais restritivo, podendo ser uma alternativa útil para controle de erros tipo I em estudos de mapeamento associativo.

REFERÊNCIAS

- ABDURAKHMONOV, Y.I.; ABDUKARIMOV, A. Application of association mapping to understanding the genetic diversity of plant germplasm resources. **International Journal of Plant Genomics**, New York, v. 2008, n. 574927, p. 1-18, 2008.
- ABDURAKHMONOV, Y.I.; SAHA, S.; JENKINS, J.N.; BURIEV, Z.T.; SHERMATOV, S.E.; SCHEFFLER, B.E.; PEPPER, A.E.; YU, J.Z.; KOHEL, R.J.; ABDUKARIMOV, A. Linkage disequilibrium based association mapping of fiber quality traits in *G. hirsutum* L. variety germplasm. **Genetica**, Gravenhage, v. 136, p. 401-417, 2009.
- AKKAYA, M.S.; BHAGWAT, A.A.; CREGAN, P. B. Length polymorphism of simple sequence repeat DNA in soybean. **Genetics**, Pittsburgh, v. 132, p. 1131-1139, 1992.
- ALMEIDA, L.A.; KIIHL, R.A.S.; MIRANDA, M.A.C.; CAMPELO, G.J.A. Melhoramento da soja para regiões de baixas latitudes. In: QUEIROZ, M.A.; GOEDERT, C.O.; RAMOS, S.R.R. (Ed.). **Recursos genéticos e melhoramento de plantas para o nordeste brasileiro**. 1999. Disponível em: <<http://www.cpatas.embrapa.br>>. Acesso em: 10 dez. 2011.
- BALDING, D.J. A tutorial on statistical methods for population association studies. **Nature Reviews in Genetics**, London, v. 7, p. 781-791, 2006.
- BALL, R.D. Experimental designs for reliable detection of linkage disequilibrium in unstructured random population association studies. **Genetics**, Pittsburgh, v. 170, n. 2, p. 859-873, 2005.
- BENJAMINI, Y.; HOCHBERG, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. **Journal of the Royal Statistical Society**, London, v. 57, n. 1, p. 289-300, 1995.
- BENNET, J.H.; BINET, F.E. Association between mendelian factors with mixed selfing and random mating. **Heredity**, London, v. 10, p. 51-55, 1956.
- BAOHUI LIU; TOSHIRO FUJITA; ZE-HONG YAN; SHINICHI SAKAMOTO; DONGHE XU; JUN ABE. QTL Mapping of domestication-related traits in soybean (*Glycine max*). **Annals of Botany**, Oxford, v. 100, p. 1027-1038, 2007.
- BORÉM, A.; MIRANDA, G.V. **Melhoramento de plantas**. Viçosa: UFV, 2005. 525 p.
- BRESEGHELLO, F.; SORRELLS, M.E. Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. **Genetics**, Pittsburgh, v. 172, p. 1165-1177, 2006.
- BUSO, G.S.C.; CIAMPI, A.Y.; MORETZHON, M.C.; AMARAL, Z.P.S.; BRONDANI, R.V. Marcadores microssatélites em espécies vegetais – Desenvolvimento e caracterização de marcadores microssatélites em espécies vegetais tropicais. **Biotecnologia Ciência e Desenvolvimento**, Brasília, v. 30, p. 46-50, 2003.

BRADBURY, P.J.; ZHANG, Z.; KROON, D.E.; CASSTEVEN, T.M.; RAMDOSS, Y.; BUCKLER, E.S. TASSEL: Software for association mapping of complex traits in diverse samples. **Bioinformatics**, Oxford, v. 23, p. 2633–2635, 2007.

CÂMARA, G.M.S. **Soja: tecnologia de produção**. Piracicaba: ESALQ, 1998. 293 p.
CARDON, L.; BELL, J. Association study designs for complex diseases. **Nature Reviews in Genetics**, London, v. 2, p. 91–99, 2001.

CHOI, I.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.S.; HWANG, E.Y.; YI, S.I.; YOUNG, N.D.; SHOEMAKER, R.C.; VAN TASSEL, CP.; SPECHT, J.E.; CREGAN, P.B. A soybean transcript map: gene distribution, haplotype and single nucleotide polymorphism analysis. **Genetics**, Pittsburgh, v.176, p. 685–696, 2007.

CIAMPI, A.Y.; VINSON, C.C.; GAIOTTO, F.A. Estimativa da diversidade genética em arbóreas nativas tropicais utilizando microsatélites. In: NASS, L.L. (Ed.). **Recursos genéticos vegetais**. Brasília: Embrapa Recursos Genéticos e Biotecnologia, 2007. p. 421-440.

COLLARD, B.C.Y.; JAHUFER, M.Z.Z.; BROUWER, J.B.; PANG, E.C.K. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. **Euphytica**, Wageningen, v. 142, p. 169–196, 2005.

COMPANHIA NACIONAL DO ABASTECIMENTO. 2010. Disponível em: <<http://www.conab.gov.br/>>. Acesso em: 01 dez. 2010.

_____. 2012. Disponível em: <<http://www.conab.gov.br/>>. Acesso em: 11 out. 2012.

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, J.; SPECHT, J.E. An integrated genetic linkage map of the soybean genome. **Crop Science**, Madison, v. 39, p. 1464–1490, 1999.

DEVLIN, B.; ROEDER, K. Genomic control for association studies. **Biometrics**, Washington, v. 55, n. 4, p. 997-1004, 1999.

DOYLE, J.J.; DOYLE, J.L. Isolation of plant DNA fresh tissue. **Focus**, Rockville, v. 12, p. 13-15, 1990.

EVANNO, G.; REGNAUT, S.; GOUDET, J. Detecting the number of clusters of individuals using the software structure: a simulation study. **Molecular Ecology**, Oxford, v. 14. p. 2611–2620, 2005.

EXCOFFIER, L.; SLATKIN, M. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. **Molecular Biology and Evolution**, Chicago, v. 12, p. 921–927, 1995.

FAN, J.B.; OLIPHANT, A.; SHEN, R.; KERMANI, B.G.; GARCIA, F.; GUNDERSON, K.L.; HANSEN, M.; STEEMERS, F.; BUTLER, S.L.; DELOUKAS, P.; GALVER, L.; HUNT, S.; MCBRIDE, C.; BIBIKOVA, M.; RUBANO, T.; CHEN, J.; WICKHAM, E.; DOUCET, D.; CHANG, W.; CAMPBELL, D.; ZHANG, B.; KRUGLYAK, S.; BENTLEY, D.; HAAS, J.; RIGAULT, P.; ZHOU, L.; STUELPNAGEL, J.; CHEE, M.S. Highly parallel SNP genotyping. **Cold Spring Harbor Symposia on Quantitative Biology**, Cold Spring Harbor, v. 68, p. 69–78, 2003.

FAN, J.B.; CHEE, M.S.; GUNDERSON, K.L. Highly parallel genomic assays. **Nature Reviews in Genetics**, London, v.7, p. 632-644, 2006.

FASOULA, V.A.; HARRIS, D.K.; BOERMA, H.G. Validation and designation of quantitative trait loci for seed protein, seed oil, and seed weight from two soybean populations. **Crop Science**, Madison, v. 44, p. 1218-1225, 2004.

FERREIRA, M.E.; GRATTAPAGLIA, D. **Introducción al uso de marcadores moleculares en el análisis genético**. Brasília: Embrapa, Cenargen, 1998. 220 p.

FERREIRA, M.E.; MORETZSOHN, M.C.; BUSO, G.S.C. Fundamentos de caracterização molecular de germoplasma vegetal. In: NASS, L.L. (Ed.). **Recursos genéticos vegetais**. Brasília: Embrapa Recursos Genéticos e Biotecnologia, 2007. p. 377-420.

FLINT-GARCIA, S.A.; THORNSBERRY, J.M.; BUCKLER IV, E.S. Structure of linkage disequilibrium in plants. **Annual Review on Plant Biology**, Palo Alto, v. 54, p. 357-374, 2003.

GUPTA, P.K.; RUSTGI, S.; KULWAL, P.L. Linkage disequilibrium and association studies in higher plants: Present status and future prospects. **Plant Molecular Biology**, Dordrecht, v. 57, n.4, p. 461–485, 2005.

GUPTA, P.K., RUSTGI, S.; MIR, R.R. Array-based high-throughput DNA markers for crop improvement. **Heredity**, London, v. 101, p. 5–18, 2008.

GUZMAN, P.S.; DIERS, B.W.; NEECE, D.J.; MARTIN, S.K.; LEROY, A.R.; GRAU, C.R.; HUGHES, T.J.; NELSON, R.L. QTL Associated with yield in three backcross-derived populations of soybean. **Crop Science**, Madison, v. 47, p. 111–122, 2007.

HELD, L. A nomogram for P values. **BMC Medical Research Methodology**, London, v. 10, p. 21, 2010.

HIROMOTO, D.M.; VELLO, N.A. The genetic base of Brazilian soybean (*Glycine max*) cultivars. **Revista Brasileira de Genética**, Ribeirão Preto, v. 9, p. 295-306, 1986.

HOU, J.F.; WANG, C.L.; HONG, X.J.; ZHAO, J.M.; XU, C.C.; GU, E.N.; GA, O.J.Y.; XING, H. Association analysis of vegetable soybean quality traits with SSR markers. **Plant Breeding**, Berlin, v. 130, n. 4, p. 444-449, 2011.

- HYTEN, D.L.; SONG, Q.; ZHU, Y.; CHOI, I.Y.; NELSON, R.L.; COSTA, J.M.; SPECHT, J.E.; SHOEMAKER, R.C.; CREGAN, P.B. Impacts of genetic bottlenecks on soybean genome diversity. **Proceedings of the National Academy of Sciences of the USA**, Washington, v. 103, p. 16666–16671, 2006.
- HYTEN, D.L.; HARTMAN, G.L.; NELSON, R.L.; FREDERICK, R.D.; CONCIBIDO, V.C.; NARVEL, J.M.; CREGAN, P.B. Map location of the *rpp1* locus that confers resistance to soybean rust in soybean. **Crop Science**, Madison, v. 47, n. 2, p. 837-838, 2007a.
- HYTEN, D.L.; CHOI, I.Y.; SONG, Q.; SHOEMAKER, R.C.; NELSON, R.L.; COSTA, J.M.; SPECHT, J.E.; CREGAN, P.B. Highly variable patterns of linkage disequilibrium in multiple soybean populations. **Genetics**, Pittsburgh, v. 175, p. 1937–1944, 2007b.
- HYTEN, D.L.; SONG, Q.; CHOI, I.Y.; YOON, M.S.; SPECHT, J.E.; MATUKUMALLI, L.K.; NELSON, R.L.; SHOEMAKER, R.C.; YOUNG, N.D.; CREGAN, P.B. High-throughput genotyping with the GoldenGate assay in the complex genome of soybean. **Theoretical and Applied Genetics**, New York, v.116, p. 945–952, 2008.
- HYTEN, D.L.; CHOI, I.Y.; SONG, Q.; SPECHT, J.E.; CARTER, T.E.; SHOEMAKER, R.C.; HWANG, E.Y.; MATUKUMALLI, L.K.; CREGAN, P.B. A high density integrated genetic linkage map of soybean and the development of a 1536 universal soy linkage panel for quantitative trait locus mapping. **Crop Science**, Madison, v. 50, p. 960–968, 2010.
- HWANG, T.Y.; HWANG, T.Y.; SAYAMA, T.; TAKAHASHI, M.; TAKADA, Y.; NAKAMOTO, Y. High-density integrated linkage map based on SSR markers in soybean. **DNA Research**, Oxford, v.16, p. 213–225, 2009.
- WANG, J.; MCCLEAN, P.E.; LEE, R.; GOOS, R.J.; HELMS, T. Association mapping of iron deficiency chlorosis loci in soybean (*Glycine max* L. Merr.) advanced breeding lines. **Theoretical and Applied Genetics**, New York, v. 116, p. 777-787, 2008.
- JUN, T.H.; VAN, K.; KIM, M.Y.; LEE, S.H.; WALKER, D.R. Association analysis using SSR markers to find QTL for seed protein content in soybean. **Euphytica**, Wageningen, v. 162, p. 179–191, 2008.
- KABELKA, E.A.; DIERS, B.W.; FEHR, W.R.; LEROY, A.R.; BAIANU, I.C.; YOU, T.; NEECE, D.J.; NELSON, R.L. Putative alleles for increased yield from soybean plant introductions. **Crop Science**, Madison, v. 44, p. 784–791, 2004.
- KEIM, P.; DIERS, B.W.; OLSON, T.C.; SHOEMAKER, R.C. RFLP mapping in soybean: association between marker loci and variation in quantitative traits. **Genetics**, Pittsburgh, v. 126, p. 735–742, 1990.
- KRAAKMAN, A.T.W. Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. **Genetics**, Pittsburgh, v.168, p. 435-446, 2004.
- LARK, K.G.; WEISEMANN, J.M.; MATTHEWS, B.F.; PALMER, R.; CHASE, K.; MACALMA, T. A genetic map of soybean (*Glycine max* L.) using an intraspecific cross of two cultivars: ‘Minsoy’ and ‘Noir 1’. **Theoretical and Applied Genetics**, New York, v. 86, p. 901–906, 1993.

LI, Z.; WILSON, R.F.; RAYFORD, W.E.; BOERMA, H.R. Molecular mapping genes conditioning reduced palmitic acid content in N87-2122-4 soybean. **Crop Science**, Madison, v. 42, p. 373-378, 2002.

LI, Y.; HILL, C.B.; CARLSON, S.R.; DIERS, B.W.; HARTMAN, G.L. Soybean aphid resistance genes in the soybean cultivars Dowling and Jackson map to linkage group M. **Molecular Breeding**, Dordrecht, v. 19, p. 25-34, 2007.

LI, Y.H.; SMULDERS, M.J.M.; CHANG, R.Z.; QIU, L.J. Genetic diversity and association mapping in a collection of selected Chinese soybean accessions based on SSR marker analysis. **Conservation Genetics**, Dordrecht, v. 12, p. 1145-1157, 2011.

LITT, M.; LUTY, J.A. A hypervariable microsatellite revealed by in-vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. **American Journal of Human Genetics**, Houston, v. 44, p. 397-401, 1989.

LIU, B.H. **Statistical genomics**. New York: CRC Press, 1998. 610 p.

MACKAY, I.; POWELL, W. Methods for linkage disequilibrium mapping in crops. **Trends in Plant Science**, Cambridge, v. 12, n.2, p. 57-63, 2007.

MALYSHEVA-OTTO, L.V.; GANAL, M.W.; RÖDER, M.S. Analysis of molecular diversity, population structure and linkage disequilibrium in a worldwide survey of cultivated barley germplasm (*Hordeum vulgare* L.). **BMC Genetics**, London, v. 7, p. 6, 2006.

MOHAMMADI, S.A.; PRASANNA, B.M. Analysis of genetic diversity in crop plants – Salient statistical tools and considerations. **Crop Science**, Madison, v. 43, p. 1235-1248, 2003.

MULATO, B.M. **Diversidade genética em germoplasma de soja identificada por marcadores SSR, EST-SSR e caracteres agromorfológicos**. 2009. 86 p. Dissertação (Mestrado em Genética e Melhoramento de Plantas) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, 2009.

MULATO, B.M.; MOLLER, M.; ZUCCHI, M.I.; QUECINI, V.; PINHEIRO, J.B. Genetic diversity in soybean germplasm identified by SSR and EST-SSR markers. **Pesquisa Agropecuária Brasileira**, Brasília, v. 45, n. 3, p. 276-283, 2010.

NASS, L.L. Utilização de recursos genéticos vegetais no melhoramento. In: NASS, L.L.; VALOIS, A.C.C.; MELO, I.S.; VALADARES-INGLIS, M.C. (Ed.). **Recursos genéticos e melhoramento - plantas**. Rondonópolis: Fundação MT, 2001. p. 29-56.

NASS, L.L.; SIGRIST, M.S. Espécies silvestres: potencial de exploração via pré-melhoramento. In: BORÉM, A.; LOPES, M.T.G.; CLEMENT, C.R. (Ed.). **Domesticação e melhoramento: espécies amazônicas**. Viçosa: Suprema, 2009. p. 101–116.

NARVEL, J.M.; FEHR, W.R.; CHU, W.C.; GRANT, D.; SHOEMAKER, R.C. Simple sequence repeat diversity among soybean plant introductions and elite genotypes. **Crop Science**, Madison, v. 40, n. 5, p. 1452-1458, 2000.

NORDBORG, M.; BOREVITZ, J.O.; BERGELSON, J.; BERRY, C.C.; CHORY, J. The extent of linkage disequilibrium in *Arabidopsis thaliana*. **Nature Genetics**, London, v. 30, p. 190–193, 2002.

ODONG, T.L. Determination of genetic structure of germplasm collections: are traditional hierarchical clustering methods appropriate for molecular marker data? **Theoretical and Applied Genetics**, New York, v. 123, p. 195–205, 2011.

ORAGUZIE, N.C.; RIKKERINK, E.H.A.; GARDINER, S.E.; SILVA, H.N. **Association mapping in plants**. New York: Springer, 2007. 277 p.

PALAISSA, K.; MORGANTE, M.; TINGEY, S.; RAFALSKI, A. Long-range patterns of diversity and linkage disequilibrium surrounding the maize Y1 gene are indicative of an asymmetric selective sweep. **Proceedings of the National Academy of Sciences of the USA**, Washington, v. 101, n. 26, p. 9885–9890, 2004.

PRIOLLI, R.H.G.; MENDES-JUNIOR, C.T.; ARANTES, N.E.; CONTEL, E.P.B. Characterization of Brazilian soybean cultivars using microsatellite markers. **Genetics and Molecular Biology**, Ribeirão Preto, v. 25, p. 185–193, 2002.

PRICE, L.; PATTERSON, N.J.; PLENGE, R.M.; WEINBLATT, M.E.; SHADICK, N.A.; REICH, D. Principal components analysis corrects for stratification in genome-wide association studies. **Nature Genetics**, London, v. 38, p. 904–909, 2006.

PRITCHARD, J.K.; ROSENBERG, N.A. Use of unlinked genetic markers to detect population stratification in association studies. **American Journal of Human Genetics**, Houston, v. 65, p. 220–228, 1999.

PRITCHARD, J.K.; STEPHENS, M.; ROSENBERG, N.A.; DONNELLY, P. Association mapping in structured populations. **American Journal of Human Genetics**, Houston, v. 67, p. 170–181, 2000.

PRITCHARD, J.K.; PRZEWORSKI, M. Linkage disequilibrium in humans: models and data. **American Journal of Human Genetics**, Houston, **69**: 1–14, 2001.

REMYINGTON, D.L.; THORNSBERRY, J.M.; MATSUOKA, Y.; WILSON, L.M.; WHITT, S.R. Structure of linkage disequilibrium and phenotypic associations in the maize genome. **Proceedings of the National Academy of Sciences of the USA**, Washington, v. 98, p. 11479–11484, 2001.

RAFALSKI, A.; MORGANTE, M. Corn and humans: recombination and linkage disequilibrium in two genomes of similar size. **Trends in Genetics**, Cambridge, v. 20, p. 103–111, 2004.

SCHMUTZ, J.; CANNON, S.B.; SCHLUETER, J.; MA, J.; MITROS, T.; NELSON, W.; HYTEN, D.L.; SONG, Q.; THELEN, J.J.; CHENG, J.; XU, D.; HELLSTEN, U.; MAY, G.D.; YU, Y.; SAKURAI, T.; UMEZAWA, T.; BHATTACHARYA, M.K.; SANDHU, D.; VALLIYODAN, B.; LINDQUIST, E.; PETO, M.; GRANT, D.; SHU, S.; GOODSTEIN, D.; BARRY, K.; FUTRELL-GRIGGS, M.; ABERNATHY, B.; DU, J.; TIAN, Z.; ZHU, L.; GILL, N.; JOSHI, T.; LIBAULT, M.; SETHURAMAN, A.; ZHANG, X.C.; SHINOZAKI, K.; NGUYEN, H.T.; WING, R.A. CREGAN, P.; SPECHT, J.; GRIMWOOD, J.; ROKHSAR, D.; STACEY, G.; SHOEMAKER, R.C.; JACKSON, S.A. Genome sequence of the palaeopolyploid soybean. **Nature**, London, v. 463, p. 178-183, 2010.

SEBOLT, A.M.; SHOEMAKER, R.C.; DIERS, B.W. Analysis of a quantitative trait locus allele from wild soybean that increases seed protein concentration in soybean. **Crop Science**, Madison, v. 40, p. 1438-1444, 2000.

SEDIYAMA, T.; TEIXEIRA, R.C.; REIS, M.S. Melhoramento da soja. In: BOREM, A. (Ed.). **Melhoramento de espécies cultivadas**. Editora UFV, Viçosa, 1999. p. 487-533.

SHI, A.; CHEN, P.; ZHANG, B.; HOU, A. Genetic diversity and association analysis of protein and oil content in food-grade soybeans from Asia and the United States. **Plant Breeding**, Berlin, v. 129, p. 250-256, 2010.

SHOEMAKER, R.C.; SPECHT, J.E. Integration of the soybean molecular and classical genetic linkage groups. **Crop Science**, Madison, v. 35, p. 436-446, 1995.

SMALLEY, M.D.; FEHR, W.R.; CIANZIO, S.R.; HAN, F.; SEBASTIAN, S.A.; STREIT, L.G. Quantitative trait loci for soybean seed yield in elite and plant introduction germplasm. **Crop Science**, Madison, v. 44, p. 436-442, 2004.

SONG, Q.J.; MAREK, L.F.; SHOEMAKER, R.C.; LARK, K.G.; CONCIBIDO, V.C.; DELANNAY, X.; SPECHT, J.E.; CREGAN, P.B. A new integrated genetic linkage map of the soybean. **Theoretical and Applied Genetics**, New York, v. 109, p. 122-128, 2004.

SOUZA, A.P. Biologia molecular aplicada ao melhoramento. In: NASS, L.L.; VALOIS, A.C.C.; MELO, I.S.; VALADARES-INGLIS, M.C. (Ed.). **Recursos genéticos e melhoramento - plantas**. Rondonópolis: Fundação MT, 2001. p. 939-965.

STEPHENS, M.; DONNELLY, P. A comparison of bayesian methods for haplotype reconstruction from population genotype data. **American Journal of Human Genetics**, Houston, v. 73, n. 5, p. 1162-1169, 2003.

TANKSLEY, S.D.; MCCOUCH, S.R. Seed banks and molecular maps: unlocking genetic potential from the wild. **Science**, Washington, v. 277, p. 1063-1066, 1997.

UNITED STATES DEPARTMENT OF AGRICULTURE. Disponível em: <<http://www.usda.gov>>. Acesso em: 21 fev. 2010.

WYSMIERSKI, P.T. **Contribuição genética dos ancestrais de soja às cultivares brasileiras**. 2010. 99 p. Dissertação (Mestrado em Genética e Melhoramento de Plantas) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, 2009.

YAMANAKA, N.; NINOMIYA, S.; HOSHI, M.; TSUBOKURA, Y.; YANO M.; M.; NAGAMURA, Y.; SASAKI, T.; HARADA, K. An informative linkage map of soybean reveals QTL's for flowering time, leaflet morphology and regions of segregation distortion. **DNA Research**, Oxford, v. 8, p. 61–72, 2001.

YU, J.; PRESSOIR, G.; BRIGGS, W.H.; BI, I.V.; YAMASAKI, M.; DOEBLEY, J.F.; MCMULLEN, M.D.; GAUT, B.S.; NIELSEN, D.M.; HOLLAND, J.B.; KRESOVICH, S.; BUCKLER, E.S. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. **Nature Genetics**, London, v. 38, n. 2, p. 203–208, 2006.

ZHANG, W.K.; WANG, Y.J., LUO, G.Z., ZHANG, J.S., HE, C.Y., WU, X.L., GAI, J.Y.; CHEN, S.Y. QTL mapping of ten agronomic traits on the soybean (*Glycine max* L. Merr.) genetic map and their association with EST markers. **Theoretical and Applied Genetics**, New York, v. 108, p. 1131–1139, 2004.

ZHAO, K.; ARANZANA, M.J.; KIM, S.; LISTER, C.; SHINDO, C.; TANG, C.; TOOMAJIAN, C.; ZHENG, H.; DEAN, C.; MARJORAM, P.; NORDBORG, M. An Arabidopsis example of association mapping in structured samples. **PLoS Genetics**, San Francisco, v. 3, n. 1, p. e4, 2007.

ZHU, Y.L.; SONG, Q.J.; HYTEN, D.L.; VAN TASSELL, C.P.; MATUKUMALLI, L.K.; GRIMM, D.R.; HYATT, S.M.; FICKUS, E.W.; YOUNG, N.D.; CREGAN, P.B. Single-nucleotide polymorphisms in soybean. **Genetics**, Austin, v. 163, p. 1123–1134, 2003.

ZHU, C.; GORE, M.; BUCKLER, E.S.; YU, J. Status and prospects of association mapping in plants. **Plant Genome**, Madison, v. 1, p. 5–20, 2008.

ANEXOS

Anexo A - Total de associações entre marcadores e características avaliadas, de acordo com diferentes testes de hipótese

(continua)

Associação			Significância			Associação			Significância		
Característica	Marcador	GL	p-valor (<0,01)	q-valor (FDR<0,01)	-e*p*ln(p) (< 0,02)	Característica	Marcador	GL	p-valor (<0,01)	q-valor (FDR<0,01)	-e*p*ln(p) (< 0,02)
AC	SOYHSP176	F	2,55E-06	2,74E-05	8,93E-05	APF	Satt030	F	2,90E-04	1,30E-03	6,42E-03
AC	Satt352	G	8,51E-04	3,06E-03	1,64E-02	APF	Satt022	N	3,05E-04	1,36E-03	6,71E-03
AC	Satt432	C2	2,20E-03	6,27E-03	ns	APF	Satt303	G	3,26E-04	1,44E-03	7,12E-03
AC	Satt468	D1a	3,70E-03	9,32E-03	ns	APF	Satt277	C2	4,13E-04	1,78E-03	8,75E-03
AC	Satt540	M	3,70E-03	9,32E-03	ns	APF	Sct_189	I	4,46E-04	1,88E-03	9,36E-03
AC	Satt242	K	4,30E-03	ns	ns	APF	Satt182	L	5,71E-04	2,31E-03	1,16E-02
AC	Satt600	D1b	4,40E-03	ns	ns	APF	Satt429	A2	7,00E-04	2,70E-03	1,38E-02
AC	Satt545	A1	4,70E-03	ns	ns	APF	Satt373	L	7,08E-04	2,70E-03	1,40E-02
AC	Satt292	I	7,60E-03	ns	ns	APF	Satt266	D1b	7,98E-04	2,94E-03	1,55E-02
AIPV	Satt224	J	1,28E-13	1,14E-11	1,03E-11	APF	Satt266	D1b	7,98E-04	2,94E-03	1,55E-02
AIPV	Satt070	B2	7,05E-09	1,68E-07	3,60E-07	APF	Satt545	A1	9,51E-04	3,33E-03	1,80E-02
AIPV	SOYHSP176	F	3,18E-08	6,91E-07	1,50E-06	APF	Satt226	D2	1,10E-03	3,68E-03	ns
AIPV	Satt335	F	2,89E-07	4,35E-06	1,18E-05	APF	Satt487	O	1,20E-03	3,96E-03	ns
AIPV	Satt202	C2	1,17E-05	9,53E-05	3,61E-04	APF	Satt184	D1a	1,30E-03	4,24E-03	ns
AIPV	Satt080	N	1,17E-05	9,53E-05	3,61E-04	APF	Satt468	D1a	1,60E-03	4,92E-03	ns
AIPV	Satt286	C2	1,67E-05	1,27E-04	4,99E-04	APF	Satt279	H	1,80E-03	5,37E-03	ns
AIPV	Satt431	J	2,00E-05	1,44E-04	5,90E-04	APF	Satt389	D2	3,10E-03	8,15E-03	ns
AIPV	Satt216	H	3,62E-05	2,42E-04	1,01E-03	APF	Satt236	A1	3,40E-03	8,79E-03	ns
AIPV	Satt191	G	5,30E-05	3,32E-04	1,42E-03	APF	AI794821	C1	3,50E-03	8,98E-03	ns
AIPV	Satt352	G	8,06E-05	4,81E-04	2,07E-03	APF	Satt186	D2	3,50E-03	8,98E-03	ns
AIPV	Satt371	C2	9,42E-05	5,42E-04	2,37E-03	APF	BE806387	F	3,60E-03	9,17E-03	ns
AIPV	Satt266	D1b	1,07E-04	5,88E-04	2,66E-03	APF	Satt390	A2	4,50E-03	ns	ns
AIPV	Satt266	D1b	1,07E-04	5,88E-04	2,66E-03	APF	Satt063	B2	6,40E-03	ns	ns
AIPV	Satt557	C2	1,32E-04	6,83E-04	3,21E-03	APF	Satt453	B1	7,80E-03	ns	ns
AIPV	Satt448	L	3,92E-04	1,70E-03	8,37E-03	APF	Satt242	K	8,20E-03	ns	ns
AIPV	Satt179	D1a	7,56E-04	2,83E-03	1,48E-02	APF	Satt005	D1b	8,70E-03	ns	ns
AIPV	Satt324	G	8,44E-04	3,05E-03	1,62E-02	APF	Satt163	G	9,90E-03	ns	ns
AIPV	BF008905	O	9,58E-04	3,33E-03	1,81E-02	EST	Satt196	K	1,61E-05	1,25E-04	4,82E-04
AIPV	Satt329	A2	9,95E-04	3,43E-03	1,87E-02	EST	AW781285	D1a	2,30E-03	6,44E-03	ns
AIPV	Satt203	D1a	1,40E-03	4,44E-03	ns	EST	Satt371	C2	4,80E-03	ns	ns
AIPV	AI794821	C1	1,70E-03	5,15E-03	ns	Estéreoico	Satt390	A2	4,67E-07	6,72E-06	1,85E-05
AIPV	Satt115	G	2,20E-03	6,27E-03	ns	Estéreoico	BE806387	F	6,42E-07	8,64E-06	2,49E-05
AIPV	Satt545	A1	2,80E-03	7,54E-03	ns	Estéreoico	Satt302	H	3,24E-06	3,41E-05	1,12E-04
AIPV	Satt387	N	4,70E-03	ns	ns	Estéreoico	Satt386	D2	4,19E-06	4,10E-05	1,41E-04
AIPV	Satt600	D1b	5,10E-03	ns	ns	Estéreoico	Satt565	C1	4,56E-06	4,41E-05	1,53E-04
AIPV	Satt038	G	5,90E-03	ns	ns	Estéreoico	Satt345	O	9,11E-06	7,81E-05	2,88E-04
AIPV	SOYGPART	C1	7,50E-03	ns	ns	Estéreoico	Satt115	G	1,65E-05	1,26E-04	4,95E-04
AIPV	Sct_189	I	8,60E-03	ns	ns	Estéreoico	Satt510	F	6,70E-05	4,08E-04	1,75E-03
AIPV	Satt030	F	9,90E-03	ns	ns	Estéreoico	Satt373	L	2,31E-04	1,10E-03	5,27E-03
APM	Satt070	B2	1,02E-11	4,34E-10	7,03E-10	Estéreoico	Satt324	G	2,71E-04	1,25E-03	6,05E-03
APM	Satt191	G	5,73E-11	1,87E-09	3,68E-09	Estéreoico	Satt534	B2	4,40E-04	1,87E-03	9,26E-03
APM	Satt224	J	3,00E-09	8,15E-08	1,60E-07	Estéreoico	Sct_189	I	5,50E-04	2,25E-03	1,12E-02
APM	Satt371	C2	1,42E-08	3,23E-07	6,98E-07	Estéreoico	Satt557	C2	6,57E-04	2,57E-03	1,31E-02
APM	SOYHSP176	F	4,51E-08	9,59E-07	2,07E-06	Estéreoico	Satt168	B2	5,80E-03	ns	ns
APM	Satt431	J	1,02E-07	1,81E-06	4,46E-06	Estéreoico	Satt173	O	6,00E-03	ns	ns
APM	Satt203	D1a	1,89E-07	3,13E-06	7,96E-06	Linoleico	Satt534	B2	6,82E-09	1,67E-07	3,49E-07
APM	Satt352	G	5,63E-07	7,75E-06	2,20E-05	Linoleico	Satt224	J	8,50E-08	1,57E-06	3,76E-06
APM	Satt448	L	7,55E-06	6,77E-05	2,42E-04	Linoleico	Satt545	A1	1,39E-05	1,11E-04	4,23E-04
APM	Satt202	C2	1,91E-05	1,39E-04	5,65E-04	Linoleico	Satt352	G	1,45E-05	1,14E-04	4,38E-04
APM	Satt080	N	1,91E-05	1,39E-04	5,65E-04	Linoleico	Satt462	L	2,00E-05	1,44E-04	5,88E-04
APM	Satt129	D1a	4,11E-05	2,66E-04	1,13E-03	Linoleico	Satt020	B2	9,72E-05	5,53E-04	2,44E-03
APM	Satt335	F	4,90E-05	3,13E-04	1,32E-03	Linoleico	Satt174	A1	1,08E-04	5,90E-04	2,67E-03
APM	Satt226	D2	4,39E-04	1,87E-03	9,22E-03	Linoleico	Satt329	A2	1,67E-04	8,37E-04	3,95E-03
APM	Satt329	A2	4,97E-04	2,07E-03	1,03E-02	Linoleico	Satt346	M	2,07E-04	1,01E-03	4,77E-03
APM	Sct_189	I	6,39E-04	2,53E-03	1,28E-02	Linoleico	Satt458	D2	2,09E-04	1,01E-03	4,82E-03
APM	Satt286	C2	8,21E-04	2,98E-03	1,59E-02	Linoleico	Satt177	A2	5,69E-04	2,31E-03	1,16E-02
APM	Satt266	D1b	1,10E-03	3,68E-03	ns	Linoleico	Satt168	B2	5,99E-04	2,39E-03	1,21E-02
APM	Satt266	D1b	1,10E-03	3,68E-03	ns	Linoleico	Satt599	A1	8,11E-04	2,97E-03	1,57E-02
APM	Satt453	B1	1,20E-03	3,96E-03	ns	Linoleico	Satt005	D1b	1,40E-03	4,44E-03	ns
APM	Satt373	L	2,10E-03	6,09E-03	ns	Linoleico	Satt303	G	1,40E-03	4,44E-03	ns
APM	Satt063	B2	2,30E-03	6,44E-03	ns	Linoleico	Satt487	O	1,70E-03	5,15E-03	ns
APM	Satt182	L	2,40E-03	6,69E-03	ns	Linoleico	Satt586	F	3,10E-03	8,15E-03	ns
APM	Satt557	C2	2,60E-03	7,20E-03	ns	Linoleico	Satt197	B1	3,20E-03	8,37E-03	ns
APM	Satt468	D1a	3,10E-03	8,15E-03	ns	Linoleico	Satt335	F	4,00E-03	9,98E-03	ns
APM	Satt600	D1b	3,70E-03	9,32E-03	ns	Linoleico	Satt510	F	4,10E-03	ns	ns
APM	Satt038	G	4,40E-03	ns	ns	Linoleico	Satt600	D1b	4,80E-03	ns	ns
APM	Satt009	N	5,00E-03	ns	ns	MCS	Satt070	B2	1,71E-13	1,39E-11	1,37E-11
APM	Satt540	M	5,20E-03	ns	ns	MCS	Satt203	D1a	5,44E-08	1,13E-06	2,48E-06
APM	BF008905	O	5,60E-03	ns	ns	MCS	Satt191	G	6,30E-08	1,24E-06	2,84E-06
APM	Satt216	H	6,00E-03	ns	ns	MCS	Satt063	B2	1,64E-05	1,26E-04	4,91E-04
APM	Satt186	D2	6,80E-03	ns	ns	MCS	SOYHSP176	F	2,13E-05	1,51E-04	6,23E-04
APM	Satt577	B2	7,30E-03	ns	ns	MCS	Satt329	A2	1,18E-04	6,30E-04	2,89E-03
APM	Satt292	I	8,80E-03	ns	ns	MCS	Satt429	A2	2,38E-04	1,13E-03	5,41E-03
APM	Satt270	I	9,30E-03	ns	ns	MCS	Satt468	D1a	2,90E-04	1,30E-03	6,43E-03
APF	Satt224	J	9,89E-21	9,67E-18	1,24E-18	MCS	Satt197	B1	6,27E-04	2,49E-03	1,26E-02
APF	Satt371	C2	8,26E-14	8,97E-12	6,77E-12	MCS	Satt038	G	6,59E-04	2,57E-03	1,31E-02
APF	Satt191	G	1,09E-13	1,07E-11	8,88E-12	MCS	Satt556	B2	8,91E-04	3,17E-03	1,70E-02
APF	Satt070	B2	2,49E-13	1,74E-11	1,97E-11	MCS	Satt386	D2	9,29E-04	3,27E-03	1,76E-02
APF	SOYHSP176	F	5,02E-13	3,07E-11	3,87E-11	MCS	Satt192	H	9,55E-04	3,33E-03	1,81E-02
APF	Satt335	F	1,79E-11	7,00E-10	1,20E-09	MCS	Satt557	C2	1,30E-03	4,24E-03	ns
APF	Satt448	L	2,56E-11	9,27E-10	1,70E-09	MCS	Satt600	D1b	2,00E-03	5,87E-03	ns
APF	Satt203	D1a	2,52E-09	7,04E-08	1,36E-07	MCS	Satt458	D2	2,50E-03	6,94E-03	ns
APF	Satt352	G	1,12E-07	1,96E-06	4,86E-06	MCS	Satt187	A2	2,80E-03	7,54E-03	ns
APF	Satt115	G	2,59E-07	4,08E-06	1,07E-05	MCS	Satt335	F	2,80E-03	7,54E-03	ns
APF	Satt329	A2	5,27E-07	7,36E-06	2,07E-05	MCS	Satt020	B2	4,60E-03	ns	ns
APF	Satt129	D1a	1,15E-06	1,37E-05	4,28E-05	MCS	Satt586	F	5,30E-03	ns	ns
APF	Satt270	I	1,28E-06	1,47E-05	4,73E-05	MCS	Satt352	G	5,50E-03	ns	ns
APF	Satt202	C2	3,39E-06	3,45E-05	1,16E-04	MCS	Satt202	C2	6,60E-03	ns	ns
APF	Satt080	N	3,39E-06	3,45E-05	1,16E-04	MCS	Satt080	N	6,60E-03	ns	ns
APF	Satt431	J	4,12E-06	4,07E-05	1,39E-04	MCS	Satt371	C2	6,90E-03	ns	ns
APF	BF008905	O	4,94E-06	4,74E-05	1,64E-04	MCS	Satt215	J	7,10E-03	ns	ns
APF	Satt179	D1a	5,11E-06	4,85E-05	1,69E-04	MCS	Satt009	N	7,30E-03	ns	ns
APF	Satt286	C2	6,92E-06	6,27E-05	2,24E-04	MCS	Satt001	K	7,40E-03	ns	ns
APF	Satt038	G	8,16E-06	7,25E-05	2,60E-04	MCS	Satt230	E	7,50E-03	ns	ns
APF	SOYGPART	C1	9,89E-06	8,41E-05	3,10E-04	MCS	Satt334	F	9,20E-03	ns	ns
APF	Satt600	D1b	3,96E-05	2,58E-04	1,09E-03	NDF	SOYHSP176	F	6,07E-14	8,48E-12	5,02E-12
APF	Satt168	B2	6,41E-05	3,97E-04	1,68E-03	NDF	Satt371	C2	7,09E-14	8,67E-12	5,84E-12
APF	Satt577	B2	8,76E-05	5,10E-04	2,23E-03	NDF	Satt070	B2	4,09E-13	2,67E-11	3,18E-11
APF	Satt557	C2	1,28E-04	6,73E-04	3,11E-03	NDF	Satt191	G	1,54E-12	8,86E-11	1,14E-10
APF	Satt216	H	1,37E-04	7,05E-04	3,31E-03	NDF	Satt335	F	2,61E-12	1,39E-10	1,90E-10
APF	Satt386	D2	1,62E-04	8,25E-04	3,85E-03	NDF	Satt448	L	3,16E-10	9,36E-09	

Anexo A - Total de associações entre marcadores e características avaliadas, de acordo com diferentes testes de hipótese

(continuação)

Associação			Significância			Associação			Significância		
Característica	Marcador	GL	p-valor (<0,01)	q-valor (FDR<0,01)	-e*p*ln(p) (< 0,02)	Característica	Marcador	GL	p-valor (<0,01)	q-valor (FDR<0,01)	-e*p*ln(p) (< 0,02)
NDF	Satt352	G	9,80E-09	2,28E-07	4,92E-07	NDR5	Satt448	L	1,89E-06	2,12E-05	6,76E-05
NDF	Satt600	D1b	6,45E-07	8,64E-06	2,50E-05	NDR5	Satt557	C2	3,37E-06	3,45E-05	1,16E-04
NDF	Satt115	G	7,39E-07	9,51E-06	2,84E-05	NDR5	Satt352	G	5,70E-06	5,36E-05	1,87E-04
NDF	Satt286	C2	8,49E-07	1,06E-05	3,23E-05	NDR5	BF008905	O	1,80E-05	1,33E-04	5,34E-04
NDF	Satt557	C2	1,06E-06	1,30E-05	3,95E-05	NDR5	Satt286	C2	7,69E-05	4,64E-04	1,98E-03
NDF	Satt129	D1a	5,89E-06	5,48E-05	1,93E-04	NDR5	Satt202	C2	8,57E-05	5,02E-04	2,18E-03
NDF	Satt203	D1a	1,32E-05	1,07E-04	4,04E-04	NDR5	Satt080	N	8,57E-05	5,02E-04	2,18E-03
NDF	Satt577	B2	1,73E-05	1,29E-04	5,16E-04	NDR5	Satt577	B2	1,09E-04	5,92E-04	2,71E-03
NDF	Satt270	I	2,47E-05	1,73E-04	7,12E-04	NDR5	Satt115	G	1,91E-04	9,43E-04	4,45E-03
NDF	BF008905	O	2,50E-05	1,73E-04	7,21E-04	NDR5	Satt226	D2	4,44E-04	1,88E-03	9,32E-03
NDF	Satt431	J	2,91E-05	2,00E-04	8,27E-04	NDR5	Satt600	D1b	4,86E-04	2,03E-03	1,01E-02
NDF	Satt202	C2	3,81E-05	2,50E-04	1,05E-03	NDR5	Satt431	J	7,10E-04	2,70E-03	1,40E-02
NDF	Satt080	N	3,81E-05	2,50E-04	1,05E-03	NDR5	Satt386	D2	7,48E-04	2,81E-03	1,48E-02
NDF	Satt303	G	1,17E-04	6,30E-04	2,88E-03	NDR5	Satt279	H	8,15E-04	2,97E-03	1,58E-02
NDF	Satt329	A2	1,24E-04	6,55E-04	3,04E-03	NDR5	Satt129	D1a	9,98E-04	3,43E-03	1,88E-02
NDF	Satt030	F	1,24E-04	6,55E-04	3,04E-03	NDR5	Satt390	A2	1,40E-03	4,44E-03	ns
NDF	SOYGPART	C1	2,72E-04	1,25E-03	6,08E-03	NDR5	Satt453	B1	1,50E-03	4,72E-03	ns
NDF	Satt386	D2	2,86E-04	1,29E-03	6,35E-03	NDR5	Satt156	L	1,60E-03	4,92E-03	ns
NDF	Satt179	D1a	3,28E-04	1,44E-03	7,15E-03	NDR5	Satt030	F	1,70E-03	5,15E-03	ns
NDF	Satt038	G	3,64E-04	1,68E-03	7,85E-03	NDR5	Sct_189	I	2,30E-03	6,44E-03	ns
NDF	Satt022	N	6,50E-04	2,56E-03	1,30E-02	NDR5	Satt270	I	2,70E-03	7,42E-03	ns
NDF	Satt429	A2	7,07E-04	2,70E-03	1,40E-02	NDR5	Satt588	K	3,80E-03	9,53E-03	ns
NDF	Satt266	D1b	7,99E-04	2,94E-03	1,55E-02	NDR5	Satt303	G	4,40E-03	ns	ns
NDF	Satt266	D1b	7,99E-04	2,94E-03	1,55E-02	NDR5	Satt586	F	4,50E-03	ns	ns
NDF	Satt390	A2	8,98E-04	3,18E-03	1,71E-02	NDR5	Satt329	A2	4,80E-03	ns	ns
NDF	Satt226	D2	1,10E-03	3,68E-03	ns	NDR5	Satt022	N	4,90E-03	ns	ns
NDF	Sct_189	I	1,30E-03	4,24E-03	ns	NDR5	Satt277	C2	5,30E-03	ns	ns
NDF	Satt182	L	1,30E-03	4,24E-03	ns	NDR5	SOYGPART	C1	5,40E-03	ns	ns
NDF	Satt588	K	2,80E-03	7,54E-03	ns	NDR5	Satt468	D1a	7,60E-03	ns	ns
NDF	Satt373	L	3,60E-03	9,17E-03	ns	NDR5	Satt334	F	9,20E-03	ns	ns
NDF	Satt586	F	4,20E-03	ns	ns	NDR7	Satt070	B2	4,77E-12	2,22E-10	3,38E-10
NDF	Satt453	B1	4,30E-03	ns	ns	NDR7	Satt191	G	3,06E-11	1,07E-09	2,01E-09
NDF	Satt487	O	4,40E-03	ns	ns	NDR7	Satt371	C2	5,58E-09	1,44E-07	2,88E-07
NDF	Satt216	H	5,30E-03	ns	ns	NDR7	Satt335	F	6,46E-08	1,24E-06	2,91E-06
NDF	Satt277	C2	6,00E-03	ns	ns	NDR7	Satt448	L	1,12E-06	1,35E-05	4,19E-05
NDF	Satt389	D2	8,60E-03	ns	ns	NDR7	SOYHSP176	F	2,30E-06	2,53E-05	8,14E-05
NDF	Sct_188	F	8,90E-03	ns	ns	NDR7	Satt202	C2	8,65E-06	7,55E-05	2,74E-04
NDF	Satt184	D1a	9,60E-03	ns	ns	NDR7	Satt080	N	8,65E-06	7,55E-05	2,74E-04
NDF	Satt468	D1a	1,00E-02	ns	ns	NDR7	Satt352	G	9,02E-06	7,80E-05	2,85E-04
NDM	Satt070	B2	3,23E-18	1,58E-15	3,54E-16	NDR7	Satt600	D1b	3,09E-05	2,11E-04	8,73E-04
NDM	Satt371	C2	5,40E-18	1,76E-15	5,84E-16	NDR7	Satt030	F	6,72E-05	4,08E-04	1,76E-03
NDM	Satt191	G	4,78E-16	9,35E-14	4,58E-14	NDR7	Satt577	B2	9,35E-05	5,41E-04	2,36E-03
NDM	Satt335	F	4,62E-14	7,53E-12	3,86E-12	NDR7	Satt431	J	1,02E-04	5,73E-04	2,55E-03
NDM	Satt448	L	2,46E-13	1,74E-11	1,94E-11	NDR7	Satt286	C2	1,38E-04	7,06E-04	3,34E-03
NDM	Satt286	C2	1,40E-11	5,70E-10	9,53E-10	NDR7	Satt115	G	1,92E-04	9,43E-04	4,48E-03
NDM	Satt115	G	6,00E-11	1,89E-09	3,84E-09	NDR7	Satt557	C2	2,40E-04	1,13E-03	5,44E-03
NDM	Satt270	I	5,74E-09	1,44E-07	2,96E-07	NDR7	Satt279	H	2,71E-04	1,25E-03	6,06E-03
NDM	Satt545	A1	2,39E-08	5,31E-07	1,14E-06	NDR7	BF008905	O	3,24E-04	1,44E-03	7,07E-03
NDM	Satt303	G	2,75E-07	4,27E-06	1,13E-05	NDR7	Satt390	A2	5,29E-04	2,17E-03	1,09E-02
NDM	Satt600	D1b	4,45E-07	6,59E-06	1,77E-05	NDR7	Satt270	I	8,33E-04	3,02E-03	1,61E-02
NDM	Satt129	D1a	7,35E-07	9,51E-06	2,82E-05	NDR7	BE806387	F	1,40E-03	4,44E-03	ns
NDM	Satt431	J	8,74E-07	1,08E-05	3,32E-05	NDR7	Satt226	D2	1,50E-03	4,72E-03	ns
NDM	Satt202	C2	1,22E-06	1,42E-05	4,52E-05	NDR7	Satt156	L	1,70E-03	5,15E-03	ns
NDM	Satt080	N	1,22E-06	1,42E-05	4,52E-05	NDR7	Satt586	F	2,20E-03	6,27E-03	ns
NDM	Satt352	G	1,35E-06	1,53E-05	4,97E-05	NDR7	Satt386	D2	2,30E-03	6,44E-03	ns
NDM	Satt179	D1a	3,75E-06	3,78E-05	1,27E-04	NDR7	Satt588	K	2,70E-03	7,42E-03	ns
NDM	Satt577	B2	1,07E-05	9,02E-05	3,33E-04	NDR7	Satt129	D1a	3,20E-03	8,37E-03	ns
NDM	Satt203	D1a	1,08E-05	9,03E-05	3,35E-04	NDR7	Satt303	G	3,40E-03	8,79E-03	ns
NDM	Satt038	G	1,16E-05	9,53E-05	3,57E-04	NDR7	Satt453	B1	4,10E-03	ns	ns
NDM	Satt030	F	1,40E-05	1,11E-04	4,26E-04	NDR7	Satt570	G	4,20E-03	ns	ns
NDM	Satt557	C2	1,59E-05	1,24E-04	4,79E-04	NDR7	Satt387	N	4,80E-03	ns	ns
NDM	Sct_188	F	3,69E-05	2,45E-04	1,02E-03	NDR7	Satt510	F	5,90E-03	ns	ns
NDM	SOYHSP176	F	4,32E-05	2,78E-04	1,18E-03	NDR7	Sct_189	I	7,70E-03	ns	ns
NDM	Satt266	D1b	5,19E-05	3,27E-04	1,39E-03	NDR7	Satt354	I	8,10E-03	ns	ns
NDM	Satt266	D1b	5,19E-05	3,27E-04	1,39E-03	NDR7	Satt329	A2	9,90E-03	ns	ns
NDM	Satt329	A2	5,78E-05	3,60E-04	1,53E-03	Oleico	Satt070	B2	8,91E-10	2,56E-08	5,05E-08
NDM	BF008905	O	6,52E-05	4,01E-04	1,71E-03	Oleico	Satt487	O	5,29E-04	2,17E-03	1,09E-02
NDM	Satt390	A2	9,53E-05	5,45E-04	2,40E-03	Oleico	Satt371	C2	6,84E-04	2,65E-03	1,36E-02
NDM	Satt279	H	1,06E-04	5,88E-04	2,65E-03	Oleico	Satt570	G	7,08E-04	2,70E-03	1,40E-02
NDM	Satt184	D1a	1,64E-04	8,31E-04	3,90E-03	Oleico	Satt179	D1a	7,72E-04	2,88E-03	1,50E-02
NDM	Satt156	L	1,65E-04	8,32E-04	3,90E-03	Oleico	Satt030	F	1,60E-03	4,92E-03	ns
NDM	Satt510	F	2,78E-04	1,27E-03	6,19E-03	Oleico	Satt586	F	3,70E-03	9,32E-03	ns
NDM	Satt570	G	3,61E-04	1,58E-03	7,78E-03	Oleico	Satt191	G	7,10E-03	ns	ns
NDM	SOYGPART	C1	4,56E-04	1,91E-03	9,55E-03	Oleico	Satt005	D1b	8,10E-03	ns	ns
NDM	AI794821	C1	5,74E-04	2,31E-03	1,16E-02	Oleo	Satt371	C2	1,48E-10	4,52E-09	9,12E-09
NDM	Satt540	M	1,10E-03	3,68E-03	ns	Oleo	Satt179	D1a	5,65E-08	1,15E-06	2,57E-06
NDM	Satt386	D2	1,60E-03	4,92E-03	ns	Oleo	Satt448	L	2,03E-07	3,31E-06	8,49E-06
NDM	BE806387	F	1,60E-03	4,92E-03	ns	Oleo	Satt270	I	2,21E-07	3,54E-06	9,23E-06
NDM	Satt346	M	1,80E-03	5,37E-03	ns	Oleo	Satt557	C2	4,55E-07	6,64E-06	1,81E-05
NDM	Satt324	G	1,90E-03	5,61E-03	ns	Oleo	Satt203	D1a	6,77E-06	6,19E-05	2,19E-04
NDM	Satt429	A2	2,00E-03	5,87E-03	ns	Oleo	Satt001	K	8,38E-05	4,97E-04	2,14E-03
NDM	Satt177	A2	2,10E-03	6,09E-03	ns	Oleo	Satt070	B2	1,01E-04	5,71E-04	2,53E-03
NDM	Satt588	K	2,40E-03	6,69E-03	ns	Oleo	Satt458	D2	1,79E-04	8,93E-04	4,20E-03
NDM	Satt194	C1	2,90E-03	7,77E-03	ns	Oleo	Satt578	C2	2,01E-04	9,83E-04	4,66E-03
NDM	Satt216	H	3,00E-03	7,95E-03	ns	Oleo	Satt307	C2	2,24E-04	1,07E-03	5,12E-03
NDM	Satt354	I	3,60E-03	9,17E-03	ns	Oleo	Satt540	M	2,63E-04	1,23E-03	5,89E-03
NDM	Satt242	K	4,40E-03	ns	ns	Oleo	Satt186	D2	3,61E-04	1,58E-03	7,79E-03
NDM	Satt345	O	4,80E-03	ns	ns	Oleo	Satt432	C2	5,56E-04	2,27E-03	1,13E-02
NDM	Satt389	D2	5,00E-03	ns	ns	Oleo	Satt146	F	7,12E-04	2,70E-03	1,40E-02
NDM	Satt277	C2	5,90E-03	ns	ns	Oleo	Satt556	B2	1,60E-03	4,92E-03	ns
NDM	Satt002	D2	6,00E-03	ns	ns	Oleo	Satt174	A1	1,80E-03	5,37E-03	ns
NDM	Satt236	A1	7,10E-03	ns	ns	Oleo	Satt239	I	2,70E-03	7,42E-03	ns
NDM	Satt009	N	7,10E-03	ns	ns	Oleo	Satt020	B2	3,50E-03	8,98E-03	ns
NDM	Satt222	H	7,70E-03	ns	ns	Oleo	Satt462	L	3,80E-03	9,53E-03	ns
NDM	Satt586	F	8,20E-03	ns	ns	Oleo	Satt030	F	3,90E-03	9,75E-03	ns
NDM	Satt022	N	8,40E-03	ns	ns	Oleo	AW781285	D1a	4,20E-03	ns	ns
NDR5	Satt371	C2	2,71E-12	1,39E-10	1,96E-10	Oleo	Satt236	A1	5,00E-03	ns	ns
NDR5	Satt191	G	3,73E-12	1,82E-10	2,67E-10	Oleo	Satt600	D1b	5,50E-03	ns	ns
NDR5	Satt070	B2	9,43E-08	1,71E-06	4,15E-06	Oleo	Satt277	C2	7,70E-03	ns	ns
NDR5	Satt335	F	1,31E-07	2,25E-06	5,65E-06	Oleo	Satt335	F	8,10E-03	ns	ns
NDR5	SOYHSP176	F	2,87E-07	4,35E-06	1,17E-05	Oleo	Satt012	G	9,20E-03	ns	ns

Anexo A - Total de associações entre marcadores e características avaliadas, de acordo com diferentes testes de hipótese

(conclusão)

Associação			Significância		
Característica	Marcador	GL	p-valor (<0,01)	q-valor (FDR<0,01)	-e*p*ln(p) (< 0,02)
Palmitico	Satt324	G	7,19E-08	1,35E-06	3,22E-06
Palmitico	Satt335	F	1,63E-07	2,75E-06	6,92E-06
Palmitico	Satt578	C2	6,78E-07	8,96E-06	2,62E-05
Palmitico	Satt173	O	7,60E-07	9,65E-06	2,91E-05
Palmitico	Satt545	A1	2,13E-06	2,37E-05	7,56E-05
Palmitico	Satt371	C2	2,60E-06	2,76E-05	9,10E-05
Palmitico	Satt352	G	2,11E-05	1,51E-04	6,19E-04
Palmitico	Satt540	M	2,19E-05	1,54E-04	6,38E-04
Palmitico	Satt462	L	3,11E-05	2,11E-04	8,79E-04
Palmitico	Satt070	B2	3,52E-05	2,37E-04	9,82E-04
Palmitico	Satt458	D2	1,19E-04	6,32E-04	2,93E-03
Palmitico	Satt224	J	1,30E-04	6,80E-04	3,16E-03
Palmitico	Satt557	C2	2,71E-04	1,25E-03	6,05E-03
Palmitico	Satt586	F	5,78E-04	2,32E-03	1,17E-02
Palmitico	Satt303	G	1,20E-03	3,96E-03	ns
Palmitico	Satt329	A2	1,50E-03	4,72E-03	ns
Palmitico	Satt389	D2	2,20E-03	6,27E-03	ns
Palmitico	Satt177	A2	2,80E-03	7,54E-03	ns
Palmitico	Satt071	D1a	3,00E-03	7,95E-03	ns
Palmitico	Satt262	O	4,20E-03	ns	ns
Palmitico	Satt022	N	5,20E-03	ns	ns
Palmitico	Satt279	H	5,60E-03	ns	ns
Palmitico	Satt510	F	5,70E-03	ns	ns
Palmitico	Satt373	L	7,70E-03	ns	ns
Palmitico	Satt102	K	9,50E-03	ns	ns
PEG	Satt224	J	4,82E-07	6,83E-06	1,91E-05
PEG	Satt191	G	1,90E-03	5,61E-03	ns
PEG	Satt523	L	3,00E-03	7,95E-03	ns
PEG	Satt309	G	7,30E-03	ns	ns
PEG	Satt386	D2	7,80E-03	ns	ns
PG	Satt070	B2	3,94E-17	9,63E-15	4,04E-15
PG	Satt203	D1a	8,92E-12	3,96E-10	6,17E-10
PG	Satt163	G	4,48E-09	1,18E-07	2,34E-07
PG	Satt371	C2	6,49E-08	1,24E-06	2,92E-06
PG	Satt448	L	2,42E-06	2,63E-05	8,53E-05
PG	Satt236	A1	4,09E-06	4,07E-05	1,38E-04
PG	Satt557	C2	6,24E-06	5,76E-05	2,03E-04
PG	Satt307	C2	1,70E-05	1,28E-04	5,09E-04
PG	Satt458	D2	7,82E-05	4,69E-04	2,01E-03
PG	Satt270	I	2,17E-04	1,05E-03	4,98E-03
PG	Satt030	F	2,55E-04	1,20E-03	5,74E-03
PG	Satt038	G	2,80E-04	1,27E-03	6,23E-03
PG	Satt373	L	5,03E-04	2,08E-03	1,04E-02
PG	Satt063	B2	7,45E-04	2,81E-03	1,46E-02
PG	Satt129	D1a	8,70E-04	3,12E-03	1,67E-02
PG	Satt329	A2	8,76E-04	3,13E-03	1,68E-02
PG	Satt146	F	9,19E-04	3,24E-03	1,75E-02
PG	Satt387	N	9,99E-04	3,43E-03	1,88E-02
PG	Satt022	N	1,10E-03	3,68E-03	ns
PG	Satt286	C2	1,80E-03	5,37E-03	ns
PG	Satt184	D1a	2,10E-03	6,09E-03	ns
PG	Satt168	B2	2,20E-03	6,27E-03	ns
PG	Satt192	H	2,30E-03	6,44E-03	ns
PG	Satt115	G	2,80E-03	7,54E-03	ns
PG	Satt177	A2	4,30E-03	ns	ns
PG	Satt182	L	4,30E-03	ns	ns
PG	Satt345	O	4,40E-03	ns	ns
PG	Satt001	K	4,90E-03	ns	ns
PG	SOYGPART	C1	5,90E-03	ns	ns
PG	Satt545	A1	6,40E-03	ns	ns
PG	Satt540	M	6,40E-03	ns	ns
PG	Satt277	C2	7,20E-03	ns	ns
PG	Satt600	D1b	7,40E-03	ns	ns
PG	SOYHSP176	F	7,80E-03	ns	ns
PG	Satt463	M	8,70E-03	ns	ns
PG	Satt191	G	9,20E-03	ns	ns
Proteína	Satt224	J	2,36E-11	8,88E-10	1,57E-09
Proteína	Satt578	C2	1,04E-04	5,81E-04	2,60E-03
Proteína	Satt286	C2	1,18E-04	6,30E-04	2,90E-03
Proteína	Satt335	F	4,22E-04	1,81E-03	8,91E-03
Proteína	Satt168	B2	1,00E-03	3,43E-03	1,88E-02
Proteína	Satt277	C2	1,60E-03	4,92E-03	ns
Proteína	Satt324	G	1,70E-03	5,15E-03	ns
Proteína	Satt557	C2	1,90E-03	5,61E-03	ns
Proteína	Satt540	M	2,20E-03	6,27E-03	ns
Proteína	Satt173	O	2,90E-03	7,77E-03	ns
Proteína	Satt530	N	3,00E-03	7,95E-03	ns
Proteína	Satt274	D1b	3,30E-03	8,58E-03	ns
Proteína	Satt022	N	5,80E-03	ns	ns
Proteína	Satt020	B2	6,60E-03	ns	ns
Proteína	Satt279	H	7,30E-03	ns	ns
Proteína	Satt545	A1	7,70E-03	ns	ns
Proteína	Satt071	D1a	9,20E-03	ns	ns
Proteína	Satt307	C2	9,60E-03	ns	ns
RF	Satt371	C2	1,92E-04	9,43E-04	4,47E-03
RF	Satt191	G	1,20E-03	3,96E-03	ns
RF	Satt577	B2	1,40E-03	4,44E-03	ns
RF	Satt202	C2	1,40E-03	4,44E-03	ns
RF	Satt080	N	1,40E-03	4,44E-03	ns
RF	Satt286	C2	2,10E-03	6,09E-03	ns
RF	Satt277	C2	2,30E-03	6,44E-03	ns
RF	Satt530	N	2,80E-03	7,54E-03	ns
RF	Satt334	F	3,30E-03	8,58E-03	ns
RF	Satt387	N	5,30E-03	ns	ns
RF	Satt156	L	7,20E-03	ns	ns
VA	Satt224	J	5,75E-11	1,87E-09	3,69E-09
VA	Satt557	C2	6,53E-04	2,56E-03	1,30E-02
VA	Satt371	C2	1,10E-03	3,68E-03	ns
VA	Satt286	C2	1,80E-03	5,37E-03	ns
VA	Satt191	G	6,90E-03	ns	ns
VA	AI794821	C1	7,50E-03	ns	ns