



UNIVERSIDADE DE BRASÍLIA
Instituto de Ciências Biológicas
Departamento de Biologia Celular

**DESENVOLVIMENTO E MAPEAMENTO DE MARCADORES
MICROSSATÉLITES E IDENTIFICAÇÃO DE QTLs LIGADOS À
PRODUTIVIDADE E À RESISTÊNCIA À MANCHA PRETA EM *Arachis* spp.**

Tese de Doutorado

MÁRCIO DE CARVALHO MORETZSOHN

Brasília - DF

2006

Universidade de Brasília
Departamento de Biologia Celular
Curso de Pós-Graduação em Biologia Molecular

**DESENVOLVIMENTO E MAPEAMENTO DE MARCADORES MICROSSATÉLITES
E IDENTIFICAÇÃO DE QTLs LIGADOS À PRODUTIVIDADE E À RESISTÊNCIA À
MANCHA PRETA EM *Arachis* spp.**

Márcio de Carvalho Moretzsohn

Orientador: Dr. David John Bertioli

Tese apresentada ao Departamento de Biologia Celular do Instituto de Biologia, da Universidade de Brasília, como requisito parcial para obtenção do grau de Doutor em Ciências Biológicas, Área de concentração: Biologia Molecular.

Brasília-DF, 2006

Termo de Aprovação

Tese apresentada ao Departamento de Biologia Celular da Universidade de Brasília, como requisito parcial para obtenção do grau de Doutor em Ciências Biológicas, área de concentração Biologia Molecular.

Tese defendida e aprovada em 29/11/2006 por:

Dr. David John Bertoli
Universidade Católica de Brasília / UnB

Dr. Dario Grattapaglia
Embrapa Recursos Genéticos e Biotecnologia / UnB

Dr. Márcio Elias Ferreira
Embrapa Recursos Genéticos e Biotecnologia / UnB

Dr. Alexandre Siqueira Guedes Coelho
Universidade Federal de Goiás

Dr. Marcos Aparecido Gimenes
Instituto Agronômico de Campinas

Membro suplente:

Dra. Gláucia Salles Cortopassi Buso (Embrapa Recursos Genéticos e Biotecnologia)

À minha esposa, Anna,
e aos meus filhos, Júlia e Henrique,
DEDICO.

À memória de meu pai,
OFEREÇO.

AGRADECIMENTOS

A realização dessa tese somente foi possível pelo apoio recebido de diversas pessoas e instituições, mas gostaria de agradecer de maneira especial:

À Embrapa, pela oportunidade do treinamento e pela concessão da bolsa de estudo;

À Embrapa Recursos Genéticos e Biotecnologia, pela infra-estrutura de trabalho e pelo apoio recebido;

Ao Departamento de Biologia Celular da UnB, pela concessão desse treinamento e aos professores da UnB, pelos ensinamentos;

Ao Dr. David Bertoli, pela amizade, orientação, sugestões e apoio durante a realização desse trabalho;

Ao Dr. Dario Grattapaglia, pela co-orientação, sugestões e também pela amizade e por todo o apoio recebido;

À Dra. Patrícia Guimarães, pelo acompanhamento desse trabalho, na condição de Conselheira Acadêmica junto à Embrapa;

Às Dras. Soraya Leal-Bertoli e Alessandra Fávero, pela condução dos bioensaios;

Ao BIRD/PRODETAB, à União Européia e ao Generation Challenge Program pelo suporte financeiro e aos Drs. David, Patrícia e Soraya pelo empenho na gestão desses projetos, cuidando do fornecimento dos meios necessários para a realização dessa tese;

Aos estagiários do Laboratório de Genética Vegetal da Embrapa Recursos Genéticos e Biotecnologia, Raphael, Aryanne, Saulo, Ronaldo e Ediene, pela ajuda nas atividades de laboratório em diferentes fases desse trabalho;

Ao Leandro, pelo cuidado com as plantas nas casas de vegetação;

À Lélia Leoi, pela ajuda na construção das bibliotecas genômicas;

À Carol José, pela análise dos marcadores RGA;

Ao Dr. Valls pela concessão dos acessos de *Arachis* utilizados;

À toda a equipe envolvida nos projetos de amendoim, pela colaboração;

Aos meus amigos do Laboratório de Genética Vegetal da Embrapa Recursos Genéticos e Biotecnologia, em especial, Gláucia, Ana Ciampi, Marco Antônio, Zilneide, Marília e Vânia.

Índice

Lista de figuras	viii
Lista de tabelas	x
Resumo	1
Abstract	3
1. INTRODUÇÃO GERAL	5
1.1. O amendoim cultivado (<i>Arachis hypogaea</i> L.)	5
1.2. O gênero <i>Arachis</i>	7
1.3. Uso de espécies silvestres no melhoramento do amendoim	8
1.4. Mapeamento genético	10
1.4.1. Construção de mapas genéticos	10
1.4.2. Mapas genéticos em <i>Arachis</i>	14
1.5. Marcadores moleculares	16
1.5.1. Marcadores microssatélites	16
1.5.2. Marcadores âncoras	19
1.5.3. Análogos a genes de resistência	20
1.6. Mapeamento de QTLs e seleção assistida por marcadores	22
1.7. Mapeamento de QTLs em <i>Arachis</i>	25
2. OBJETIVOS GERAIS	27
 <u>Capítulo 1</u> - Desenvolvimento e caracterização de marcadores microssatélites para o amendoim (<i>Arachis hypogaea</i> L.)	
1. INTRODUÇÃO	29
2. MATERIAL E MÉTODOS	31
2.1. Material vegetal	31
2.2. Extração e quantificação de DNA	31
2.3. Construção de bibliotecas genômicas	33
2.4. Seqüenciamento de DNA e desenho de primers	34
2.5. Condições de PCR e eletroforese dos marcadores microssatélites	35
2.6. Caracterização dos novos marcadores microssatélites	35
3. RESULTADOS E DISCUSSÃO	37
3.1. Desenvolvimento de marcadores microssatélites	37
3.2. Características das regiões repetitivas em <i>Arachis</i>	39
3.3. Polimorfismo e transferibilidade dos novos marcadores microssatélites	40
3.4. Caracterização do conteúdo gênico das bibliotecas usadas para desenvolvimento dos marcadores	46
3.5. Relações genéticas entre acessos de <i>A. hypogaea</i>	48
4. CONCLUSÕES	51
 <u>Capítulo 2</u> - Construção de um mapa de ligação e mapeamento de QTLs ligados à produtividade e à resistência à mancha preta em espécies diplóides de <i>Arachis</i> com genoma AA	
1. INTRODUÇÃO	53
2. MATERIAL E MÉTODOS	55
2.1. Material vegetal	55
2.2. Extração de DNA e PCR dos locos microssatélites	55

2.3. Construção do mapa.....	55
2.4. Avaliação fenotípica.....	58
2.5. Análise de QTLs	59
3. RESULTADOS E DISCUSSÃO	61
3.1. Construção do mapa.....	61
3.2. Análise de QTLs	67
3.2.1. QTLs para peso de sementes	69
3.2.2. QTLs para número de sementes por planta	75
3.2.3. QTLs para resistência à mancha preta (<i>C. personatum</i>).....	78
3.2.4. Co-localização de QTLs	82
3.2.5. Genes candidatos.....	82
4. CONCLUSÕES	84
CONSIDERAÇÕES FINAIS.....	86
REFERÊNCIAS BIBLIOGRÁFICAS	88
ANEXOS	
ANEXO 1 - Moretzsohn <i>et al.</i> , 2005.....	104
ANEXO 2 - Martins <i>et al.</i> , 2006.....	117
ANEXO 3 - Planilha com as informações sobre os marcadores	122

Lista de figuras

INTRODUÇÃO GERAL

Figura 1 – Base genética dos marcadores microssatélites: (A) Seqüência de DNA contendo as regiões repetitivas, (B) genótipos para esse loco de 3 plantas hipotéticas e (C) migração dos diferentes alelos em gel de eletroforese. 17

Capítulo 1 - Desenvolvimento e caracterização de marcadores microssatélites para o amendoim (*Arachis hypogaea* L.)

Figura 1 – Número de marcadores detectados por classe de tamanho de repetições (gráfico A) e relações entre número de repetições dos microssatélites e o polimorfismo detectado nas análises de *A. duranensis* e *A. stenosperma* e em seis acessos de *A. hypogaea* (gráfico B). Nessa análise, foram considerados 225 marcadores que amplificaram produtos com boas resoluções em *A. duranensis* e *A. stenosperma* e 236 marcadores na análise de *A. hypogaea*. 41

Figura 2 - Dendrograma, obtido pelo método UPGMA, baseado nas distâncias de alelos comuns (“shared allele distance”), entre 16 acessos de *A. hypogaea* (Tabela 1). As letras após cada acesso correspondem às subespécies e variedades: FF-*fastigiata/fastigiata*; FV-*fastigiata/vulgaris*; FA-*fastigiata/aequatoriana*; FP-*fastigiata/peruviana*; HH-*hypogaea/hypogaea*; HHi-*hypogaea/hirsuta*. Xingu corresponde ao tipo Xingu, que parece pertencer à subespécie *hypogaea*. 50

Capítulo 2 - Construção de um mapa de ligação e mapeamento de QTLs ligados à produtividade e à resistência à mancha preta em espécies diplóides de *Arachis* com genoma AA

Figura 1 – Padrões de amplificação obtidos com o primer RN8C09, por eletroforese em gel de poliacrilamida 5% (foto A) e com os primers RN10F09 (foto B) e Leg69, que é um marcador âncora (foto C), em géis de agarose 3,5%. Na foto A, a primeira e a última linhas são padrões de tamanhos conhecidos (10-bp DNA ladder). Nos três géis, da esquerda para a direita: *A. duranensis* (acesso K7988), usado como progenitor feminino; *A. stenosperma* (acesso V10309), progenitor masculino; híbrido F₁ e 45 indivíduos F₂ na foto A e 21 indivíduos F₂ nas fotos B e C. A seta, na foto A, indica o fragmento de 250 pares de bases. 62

Figura 2 – Mapa genético obtido pela análise de 93 plantas F₂, resultantes do cruzamento entre *A. duranensis* e *A. stenosperma*. Marcadores que não segregaram de acordo com as proporções esperadas apresentam o símbolo # após o nome, sendo # para distorções a 5%; ## a 1% e ### para marcadores altamente distorcidos, com valores de X² maiores do que 30. Marcadores que compõem o “framework” map estão sublinhados e os possíveis marcadores gênicos estão em itálico. À esquerda de cada grupo aparecem as distâncias de Kosambi. Os marcadores dominantes apresentam Ad ou As após os nomes, dependendo da origem do fragmento amplificado (Ad - *A. duranensis* e As - *A. stenosperma*). Marcadores que amplificaram dois locos apresentam os números _1 e _2 após os nomes. Locos mapeados com LOD<11 estão indicados no mapa por uma seta, seguida do valor de LOD máximo com que foram mapeados. Todos os marcadores âncoras começam por Leg, enquanto os

marcadores RGA têm RGA após o nome da sonda utilizada (RFLP) ou da espécie de origem da banda amplificada (AFLP modificado). Os demais marcadores são microssatélites..... 64

Figura 3 – Distribuição de frequências fenotípicas das três características avaliadas na população F_2 : (A) peso de sementes, (B) número de sementes por planta e (C) resistência à mancha-preta. Os fenótipos dos progenitores feminino (P_1) e masculino (P_2) e do híbrido F_1 são indicados para as características peso de sementes e resistência à mancha-preta. Para número de sementes é apresentada somente a média fenotípica. 68

Figura 4 – Localização de QTLs para as três características fenotípicas avaliadas: peso de sementes (*psem*), número de sementes por planta (*nsem*) e resistência ao fungo *Cercosporidium personatum* (*Cp*). Os marcadores adjacentes aos QTLs aparecem em negrito. As demais informações referentes à nomenclatura dos marcadores e ao mapa de ligação encontram-se na legenda da figura 2 (página 67)..... 74

Figura 5 – QTLs identificados para a característica número de sementes por planta, pelo método de mapeamento por intervalo composto. São apresentados apenas os dois grupos de ligação nos quais foram mapeados QTLs com valores de LOD score maiores do que 4,1. Cada pico acima da linha horizontal, que indica o ponto de corte (LOD score=4,1), representa um possível QTL.. 76

Figura 6 – Mapeamento de QTLs para resistência ao fungo *Cercosporidium personatum*, causador da mancha-preta em *Arachis* spp, pelo método de mapeamento por intervalo composto..... 80

Lista de tabelas

Capítulo 1: Desenvolvimento e caracterização de marcadores microssatélites para o amendoim (*Arachis hypogaea* L.)

Tabela 1 – Acessos de *Arachis hypogaea* incluídos na caracterização dos novos marcadores microssatélites. Os acessos utilizados na triagem inicial dos 271 novos marcadores são identificados com um asterisco após o nome. São apresentados também a classificação taxonômica e o local de coleta de cada acesso..... 32

Tabela 2 – Número de marcadores polimórficos detectados na análise de dois acessos de espécies silvestres de *Arachis* - *A. duranensis* (K 7988) e *A. stenosperma* (V 10309) - e de seis acessos de *A. hypogaea*, incluindo as seis variedades botânicas. Os seis marcadores desenvolvidos a partir de genes de *A. hypogaea*, depositados no GenBank, não foram incluídos..... 42

Tabela 3 – Pares de primers, nota referente à qualidade dos produtos amplificados, amplitude de comprimento dos fragmentos, número de indivíduos analisados (n), número total de alelos amplificados (A) e diversidade gênica (GD) para os 62 novos marcadores microssatélites polimórficos para *A. hypogaea*, identificados no presente estudo 43

Tabela 4 – Conteúdo gênico de diferentes bibliotecas estimado por BlastX (valores de $E < 10^{-8}$) contra *Arabidopsis* e proteínas de leguminosas. Os dados de ESTs são apresentados apenas para comparação com as bibliotecas genômicas. 47

Tabela 5 – Matriz de distâncias genéticas entre 16 acessos de *A. hypogaea*, estimadas pelo método de alelos comuns (“shared allele distance”). As letras após cada acesso correspondem às subespécies, variedades e tipos: FF-*fastigiata/fastigiata*; FV-*fastigiata/vulgaris*; FA-*fastigiata/aequatoriana*; FP-*fastigiata/peruviana*; HH-*hypogaea/hypogaea*; HHi-*hypogaea/hirsuta*; Xingu - *hypogaea* tipo Xingu 49

Capítulo 2: Construção de um mapa de ligação para espécies diplóides de *Arachis* com genoma AA e mapeamento de QTLs de interesse agrônomico para o amendoim

Tabela 1 - Médias e desvios-padrões (s) para três características fenotípicas estimados para os acessos progenitores e plantas F₁ e F₂. Para a população F₂ são apresentados também os valores mínimos e máximos..... 67

Tabela 2 – Coeficientes de correlação de Pearson, obtidos para as três características avaliadas na população F₂..... 68

Tabela 3 – Marcadores significativamente associados à característica peso de sementes pela análise de marcas simples. Todos os marcadores são codominantes..... 69

Tabela 4 – QTLs detectados para a característica peso de sementes, pelo método de mapeamento por intervalo múltiplo. Todos os marcadores são codominantes..	71
Tabela 5 – Marcadores significativamente associados à característica número de sementes por planta, pela análise de marcas simples. Todos os marcadores são codominantes.....	75
Tabela 6 – QTLs detectados para a característica número de sementes por planta, pelo método de mapeamento por intervalo múltiplo. Os marcadores S1A36RGA_B e S1A36RGA_D são dominantes (RGA); os demais são codominantes..	77
Tabela 7 – Marcadores significativamente associados à resistência ao fungo <i>Cercosporidium personatum</i> , causador da mancha preta, pela análise de marcas simples. Todos os marcadores são codominantes..	79
Tabela 8 – QTLs detectados para a característica resistência ao fungo <i>Cercosporidium personatum</i> , causador da mancha preta em <i>Arachis</i> spp, pelo método de mapeamento por intervalo múltiplo	81

Resumo

O amendoim cultivado (*Arachis hypogaea*) é uma cultura importante, amplamente cultivada nas regiões tropicais e subtropicais do mundo. É bastante susceptível a diversos estresses bióticos e abióticos, para os quais muitas das espécies silvestres são resistentes. Como um primeiro passo visando a introgressão desses genes de resistência no amendoim cultivado, um mapa de ligação foi construído, usando uma população F₂, obtida do cruzamento entre duas espécies silvestres diplóides de genoma AA (*A. duranensis* e *A. stenosperma*). Esse mapa foi baseado em marcadores microssatélites e marcadores âncoras, ambos codominantes e transferíveis entre espécies aparentadas. Um total de 271 novos marcadores microssatélites foi desenvolvido no presente trabalho, a partir de bibliotecas genômicas enriquecidas para microssatélites, de bibliotecas de etiquetas de seqüências expressas (ESTs ou “expressed sequence tags”) e por mineração (“data mining”) de seqüências disponíveis no GenBank. Para facilitar o desenvolvimento dos marcadores a partir de um grande número de seqüências, foi utilizado um módulo de programas desenvolvido em parceria com a Universidade Católica de Goiás e com a Universidade Católica de Brasília. Esse módulo integra um programa para busca de microssatélites (TROLL), um conjunto de programas para agrupamento e processamento das seqüências (Staden package) e um terceiro programa, para desenho de primers (Primer3). Com isso, o desenvolvimento dos marcadores pôde ser automatizado, tornando o processo bem mais rápido e eficiente. Dos 271 novos marcadores microssatélites, 66 mostraram-se polimórficos para o amendoim cultivado e 62 deles foram utilizados para análise das relações genéticas entre acessos representando as seis variedades botânicas descritas de *A. hypogaea*. Os resultados obtidos foram consistentes com a classificação taxonômica atual, exceto para as variedades *fastigiata/peruviana* e *fastigiata/aequatoriana*, que formaram um grupo separado e distante das outras duas variedades da subespécie *fastigiata*. Os 271 novos marcadores microssatélites, além de outros 218 publicados ou em publicação para o amendoim foram testados, usando-se os dois progenitores. Desses 489 marcadores microssatélites, 215 (44,0%) mostraram-se polimórficos, sendo 181 codominantes e 34 dominantes (amplificaram alelos nulos em um dos progenitores). Os 99 marcadores microssatélites codominantes que não apresentaram distorções da segregação 1:2:1 esperada ($p < 0,05$), junto aos 40

marcadores âncoras não distorcidos (de um total de 72), foram utilizados para o estabelecimento dos grupos de ligação. Em seguida, os marcadores distorcidos e os dominantes foram incluídos no mapa. Usando um LOD score mínimo de 11, 182 marcadores microssatélites, 66 marcadores âncoras e 12 RGAs foram mapeados em 10 grupos de ligação, como esperado, uma vez que *Arachis duranensis* e *A. stenosperma* são espécies diplóides com $2n=20$ cromossomos. Um comprimento de mapa total igual a 1645,1 cM foi obtido, com uma distância média de 6,32 cM entre marcadores. Esse mapa foi utilizado para o mapeamento de QTLs que controlam três características de interesse para o melhoramento do amendoim. Dez QTLs que controlam o peso de sementes, oito QTLs que controlam o número de sementes produzidas por planta e dois QTLs para resistência ao fungo causador da mancha preta (*Cercosporidium personatum*) foram significativamente identificados pelo método de mapeamento por intervalo múltiplo. Esse é o primeiro mapa baseado em marcadores microssatélites desenvolvido para *Arachis*. Além disso, os primeiros QTLs para componentes da produtividade e para resistência à mancha preta foram mapeados em *Arachis* spp. Uma vez que a maioria dos marcadores microssatélites utilizados foi desenvolvida a partir de ESTs e de bibliotecas genômicas digeridas com enzimas de restrição sensíveis a metilação, mais da metade dos marcadores mapeados, incluindo os marcadores âncoras, são provavelmente gênicos. Os grupos de ligação identificados nesse trabalho, as ordens dos marcadores obtidas e os QTLs mapeados estão sendo validados em outras populações de mapeamento, visando a construção de um mapa de referência para *Arachis* e a obtenção de estimativas mais robustas sobre os QTLs identificados, para uso dessas informações nos programas de melhoramento do amendoim.

Abstract

Cultivated peanut (*Arachis hypogaea*) is an important crop, widely grown in tropical and subtropical regions of the world. It is highly susceptible to several biotic and abiotic stresses to which wild species are resistant. As a first step towards the introgression of these resistance genes into cultivated peanut, a linkage map was constructed, using an F₂ population obtained from a cross between two diploid wild species with AA genome (*A. duranensis* and *A. stenosperma*). The map was based on microsatellite and anchor markers that are both codominant and transferable among related species. A total of 271 new microsatellite markers were developed in the present study from SSR-enriched genomic libraries, ESTs, and by “data-mining” sequences available in GenBank. To aid the development of microsatellite markers from a great number of sequences, a module, developed in collaboration with the Catholic University of Goiás and the Catholic University of Brasília, was used. This module integrates a program for the detection of microsatellites (TROLL), the Staden Package for the assembly and processing of sequences, and a software for primer design (Primer3). By using this module, the SSR marker development was automated, and the process proved to be fast and efficient. Of the 271 new SSR markers, 66 were polymorphic for cultivated peanut and 62 were used for the genetic relationship analysis of accessions representing the six described botanical varieties of *A. hypogaea*. Results were consistent with the actual taxonomic classification, except for the *fastigiata/peruviana* and *fastigiata/aequatoriana* varieties, which formed a separated group, distant from the other two *fastigiata* varieties. The 271 new markers plus another 218 published or being published for peanut were screened against both progenitors. A total of 215 (44.0%) were polymorphic, with 181 codominant and 34 dominant markers (amplified null alleles in one of the progenitors). The 99 codominant markers segregating 1:2:1 ($p < 0.05$) plus the 40 non-distorted anchor markers (of a total of 72 markers) were initially used to establish the linkage groups. Distorted and dominant markers were subsequently included in the map. By using a minimum LOD score of 11 and a maximum recombination fraction of 0.35, 182 microsatellite markers, 66 anchor markers and 12 RGAs mapped in 10 linkage groups, as expected, since *A. duranensis* and *A. stenosperma* are diploid species with $2n=20$ chromosomes. A total map distance of 1645.1 cM was obtained, with an average distance of 6.32 cM between markers. This map was used

for the mapping of QTL for three traits of interest for the peanut breeding programs. Ten QTL controlling seed weight, eight QTLs for seed number and two QTLs for resistance against the fungus *C. personatum* were significantly mapped by using the multiple interval mapping method. This is the first microsatellite-based map published for *Arachis*. Moreover, the first QTLs for yield components and for resistance to late leafspot were mapped in *Arachis* spp. Because most markers used were derived from ESTs and genomic libraries made using methylation sensitive restriction enzymes, more than a half of the markers, including the anchor markers, are genic. Linkage group ordering and the mapped QTL are being validated in other mapping populations, with the aim of constructing a transferable reference map for *Arachis* and to obtain more robust estimates about the mapped QTL for use in the peanut breeding programs.

1. INTRODUÇÃO GERAL

1.1. O amendoim cultivado (*Arachis hypogaea* L.)

O amendoim é uma cultura importante internacionalmente, tanto para produção de óleo, como para consumo direto, constituindo uma das principais fontes de proteínas e vitaminas em diversos países em desenvolvimento da Ásia, África e América do Sul (Savage & Keenan, 1994). Seus grãos contêm cerca de 45% de óleos, 25% de proteínas e 20% de carboidratos (Godoy *et al.*, 1999; Holbrook & Stalker, 2003). Sua folhagem é utilizada como forragem e a torta resultante da extração do óleo é usada como ração animal. Além disso, a capacidade de fixar nitrogênio da raiz do amendoim, mantendo a fertilidade do solo, é especialmente importante para os pequenos produtores que, em geral, não têm condições de utilizar fertilizantes químicos. Por isso, o amendoim é também bastante cultivado para subsistência por pequenos agricultores, principalmente na Ásia, África e no nordeste do Brasil (Singh & Singh, 1991; Hammons, 1994). Outras espécies do gênero *Arachis* possuem grande valor como plantas forrageiras (*A. glabrata*, *A. pintoi*) e alto potencial como cobertura para controle de erosões e como plantas ornamentais (*A. repens*, *A. helodes*, *A. kempff-mercadoi*) (Valls, 2000). Além disso, *A. villosulicarpa* e *A. stenosperma* permanecem sendo cultivadas para produção de grãos alimentícios por indígenas brasileiros (Stalker & Simpson, 1995; Freitas *et al.*, 2003).

O amendoim é cultivado, sob diversos sistemas de produção, em mais de 80 países nas Américas, na Ásia e na África (Singh & Singh, 1991; Holbrook & Isleib, 2001). A produção mundial encontra-se estável nos últimos seis anos, tendo atingido 36,4 milhões de toneladas de amendoim com casca e 5,9 milhões de toneladas de óleo em 2004, o que coloca o amendoim em quinto lugar entre as plantas oleaginosas, atrás da soja (30 milhões de toneladas), dendê (28 Mt), colza (12,6 Mt) e girassol (9,9 Mt) (FAO, 2006). Mais de 90% da produção de amendoim provém da Ásia e da África, sendo a China (34,1% da produção mundial), Índia (28,5%) e Nigéria (5,2%) os maiores produtores mundiais (FAO, 2006). No Brasil, o cultivo de amendoim é pequeno, representando menos de 1% da produção mundial. Apesar disso, a produção nacional aumentou cerca de 34% nos últimos quatro anos, tendo atingido 291.966 toneladas de amendoim com casca e 45.000 toneladas de óleo na

safrá 2004/2005, colhidos em 125.490 hectares (FAO, 2006). Cerca de 85% da área plantada no Brasil encontra-se no estado de São Paulo (CONAB, 2006).

Arachis hypogaea possui duas subespécies – *fastigiata* e *hypogaea*, classificadas em seis variedades botânicas, com base em características morfológicas e nos hábitos de crescimento da planta (Krapovickas & Gregory, 1994). A subespécie *hypogaea* apresenta ciclo longo, hábito rasteiro, sementes com dormência, flores no eixo central e possui duas variedades, *hypogaea* e *hirsuta*. A variedade *hypogaea* apresenta vagens com 2 ou 3 sementes, enquanto a variedade *hirsuta* apresenta vagens com nervuras bem marcadas, com uma saliência em uma das extremidades em forma de bico de papagaio e 3 ou 4 sementes. A subespécie *fastigiata*, de ciclo mais curto, porte ereto, sem dormência e sem flores no eixo central, é dividida nas variedades *fastigiata*, *vulgaris*, *aequatoriana* e *peruviana*. A variedade *fastigiata* apresenta 3 ou 4 sementes por vagem e a variedade *vulgaris*, apenas duas. As variedades *aequatoriana* e *peruviana* apresentam distribuição bastante restrita. A variedade *peruviana* apresenta vagens com reticulação acentuada e nervuras longitudinais salientes. A variedade *aequatoriana* é caracterizada por folíolos pilosos e vagens rugosas, contendo de 3 a 5 sementes coloridas. Agronomicamente, o amendoim tem sido classificado em diferentes grupos, sendo os principais Valência, Spanish e Virgínia. Embora cultivares de cada um desses grupos possam ter em sua origem acessos de diferentes subespécies e variedades, morfológicamente, o grupo Valência é classificado como subsp. *fastigiata* var. *fastigiata*; o grupo Spanish como *fastigiata* var. *vulgaris* e o grupo Virgínia como pertencente à subsp. *hypogaea* var. *hypogaea*.

Arachis hypogaea é considerada uma espécie autógama, com estrutura floral que facilita a autofecundação: anteras e estigma na mesma altura e envoltas por uma quilha (Santos & Godoy, 1999). A peculiaridade do amendoim é que, embora as flores sejam aéreas, os frutos e as sementes são produzidos abaixo da superfície do solo. Após a fertilização, a região basal do ovário alonga-se, formando uma estrutura conhecida como pendão (“peg”) ou ginóforo, que é dotado de geotropismo positivo. Devido à produção subterrânea da semente, a dispersão natural do amendoim torna-se limitada, resultando em populações com lenta expansão de perímetro e grande isolamento geográfico (Valls, 2000).

1.2. O gênero *Arachis*

O gênero *Arachis* pertence à família Leguminosae ou Fabaceae e é nativo da América do Sul, provavelmente, da região de divisa entre o Brasil Central e o Paraguai (Gregory *et al.*, 1980). Sua distribuição natural é restrita ao Brasil, Bolívia, Paraguai, Argentina e Uruguai (Valls & Simpson, 1994). Possui 80 espécies descritas, reunidas em nove seções taxonômicas, de acordo com a morfologia, distribuição geográfica e viabilidade de cruzamentos (Krapovickas & Gregory, 1994; Valls & Simpson, 2005). As nove seções são: *Trirectoides*, *Erectoides*, *Extranervosae*, *Triseminatae*, *Heteranthae*, *Caulorrhizae*, *Procumbentes*, *Rhizomatosae* e *Arachis*. A seção *Arachis*, à qual pertence o amendoim, apresenta outras 30 espécies descritas, que são classificadas nessa seção por cruzarem com *A. hypogaea*.

A grande maioria das espécies da seção *Arachis* é diplóide, mas o amendoim cultivado (*A. hypogaea*) é alotetraplóide ($2n=4x=40$), com genoma AABB. O genoma AA é caracterizado pela presença de um par de cromossomos, chamado A (Husted, 1936), de tamanho bastante reduzido e que apresenta um menor nível de condensação da eucromatina em relação aos demais cromossomos (Seijo *et al.*, 2004). As espécies diplóides de genoma AA podem ser perenes ou anuais, mas cruzam-se facilmente entre si, produzindo híbridos com viabilidade de pólen de média a alta (Gregory & Gregory, 1979; Krapovickas & Gregory, 1994; Fávero, 2004). As demais espécies diplóides com $2n=20$, todas anuais, formam um grupo mais heterogêneo, com aparentes subgrupos (Valls, 2000; Moretzsohn *et al.*, 2004). Essas espécies, que não apresentam o par A, são consideradas como possuidoras do genoma BB de *A. hypogaea*, com exceção de *A. glandulifera*, que possui o genoma chamado DD, caracterizado pela presença de seis pares de cromossomos subtelocêntricos (Stalker, 1991). Além disso, três espécies possuem $2n=18$ cromossomos (*A. decora*, *A. praecox* e *A. palustris*) (Lavia, 1996; 1998; Peñaloza & Valls, 1997) e uma outra espécie é também tetraplóide, *A. monticola*, considerada por vários autores como uma espécie progenitora do amendoim cultivado (Gregory & Gregory, 1976; Raina & Mukai, 1999; Raina *et al.*, 2001; Moretzsohn *et al.*, 2004; Milla *et al.*, 2005).

A origem de *A. hypogaea* tem sido objeto de bastante debate. Por ser um alotetraplóide, acredita-se que tenha surgido da hibridização entre duas espécies

diplóides. Diversas espécies têm sido sugeridas como as possíveis doadoras dos genomas AA e BB (revisto por Fávero *et al.*, 2006). Com base em dados de marcadores moleculares, de distribuição geográfica e de FISH (“Fluorescent In Situ Hybridization”), acredita-se que o amendoim seja resultante de um cruzamento entre *A. ipaënsis*, doadora do genoma BB, e *A. duranensis*, doadora do genoma AA (Kochert *et al.*, 1991; 1996; Seijo *et al.*, 2004). Recentemente, o amendoim pôde ser “refeito” a partir do cruzamento entre essas duas espécies (Fávero *et al.*, 2006). Assim, *A. ipaënsis*, usada como parental feminina foi cruzada com *A. duranensis*, usada como doadora de pólen. A planta F₁ resultante, diplóide, foi duplicada com colchicina, obtendo-se uma planta tetraplóide AABB. Essa planta foi cruzada com *A. hypogaea*, gerando híbridos férteis e dando suporte à hipótese de que *A. duranensis* e *A. ipaënsis* são os prováveis doadores dos genomas AA e BB, respectivamente, para o amendoim.

O amendoim, portanto, deve ter-se originado da hibridização de duas espécies diplóides, seguida de um evento de duplicação espontânea de cromossomos (Halward *et al.*, 1991; Young *et al.*, 1996). A planta resultante devia apresentar maior vigor híbrido, mas, por ser tetraplóide, tornou-se isolada reprodutivamente de seus parentes silvestres. Como consequência, todas as raças locais de amendoim são, provavelmente, derivadas de uma ou poucas plantas. Esse fato levou a uma variabilidade genética reduzida, o que tem dificultado os avanços no melhoramento molecular do amendoim. Por outro lado, espécies silvestres de *Arachis* apresentam maior variabilidade genética (Hilu & Stalker, 1995; Stalker *et al.*, 1995; Burow *et al.*, 2001; Moretzsohn *et al.*, 2004; Milla *et al.*, 2005), além de possuírem resistências a diferentes fatores bióticos e abióticos, que são de interesse para o melhoramento do amendoim (Stalker & Simpson, 1995; Rao *et al.*, 2003b; Dwivedi *et al.*, 2003; 2006).

1.3. Uso de espécies silvestres no melhoramento do amendoim

Por constituírem importante fonte de genes úteis, espécies silvestres de *Arachis* vêm sendo introduzidas em programas de melhoramento do amendoim. A transferência de genes dessas espécies para *A. hypogaea* tem tido algum sucesso, sempre a partir de espécies da seção *Arachis*. *A. monticola*, espécie também tetraplóide, foi usada em cruzamentos interespecíficos com *A. hypogaea*, resultando

nas primeiras cultivares de amendoim com genes de espécies silvestres, Spancross e Tamnut 74 (Isleib & Wynne, 1992). Apesar disso, essas cultivares não apresentam nenhuma característica fenotípica que possa ser identificada como derivada de espécies silvestres.

Simpson & Starr (2001) registraram a primeira cultivar de amendoim com genes identificáveis que foram transferidos de espécies silvestres de *Arachis*. Essa cultivar, chamada COAN, foi obtida a partir de retrocruzamentos (RC₅) de *A. hypogaea* cv. Florunner, usada como parental recorrente, com um complexo híbrido interespecífico, chamado TxAG-6, resultante do cruzamento entre {*A. batizocoi* x (*A. cardenasii* x *A. diogeni*)} e posterior duplicação de cromossomos por colchicina. COAN é resistente ao nematóide-das-galhas (*Meloidogyne arenaria* e *M. javanica*), mas extremamente susceptível ao “tomato spotted wilt virus”- TSWV, que é uma das principais doenças do amendoim, especialmente nos Estados Unidos (Holbrook & Stalker, 2003). A seleção de plantas na geração RC₇ desse mesmo programa de retrocruzamentos resultou na cultivar NemaTAM, resistente a nematóides, mais produtiva do que a COAN, mas também bastante susceptível ao TSWV (Starr *et al.*, 2002; Simpson *et al.*, 2003).

Outras tentativas de introgressão de genes oriundos de espécies silvestres de *Arachis* para o amendoim vêm sendo feitas em diferentes programas de melhoramento (Dwivedi *et al.*, 2003; Holbrook & Stalker, 2003). No entanto, essas iniciativas são ainda bastante reduzidas, se considerado o grande potencial que essas espécies representam. A introgressão de genes úteis em *A. hypogaea* apresenta uma série de dificuldades que tem limitado o uso dessas espécies no melhoramento do amendoim. Mesmo quando espécies compatíveis são utilizadas, barreiras de esterilidade estão presentes, devido a diferentes níveis de ploidia, incompatibilidades genômicas e diferenças genéticas (Holbrook & Stalker, 2003). O método mais utilizado tem sido o retrocruzamento, usando-se as espécies silvestres como parental doador e *A. hypogaea* como parental recorrente. Dessa forma, junto com o(s) gene(s) de interesse, partes do genoma do parental doador são transferidas para o parental recorrente (“linkage drag”). Frequentemente, essas regiões do genoma possuem genes que controlam características indesejáveis, como por exemplo, baixa produtividade e qualidade inferior. O processo de introgressão apenas dos genes úteis pode ser otimizado e acelerado pelo uso de seleção assistida por marcadores. Para isso, a construção de mapas genéticos e o

mapeamento de genes que controlam as características de interesse são de fundamental importância e essas técnicas começam a ser introduzidas nos principais programas de melhoramento da cultura (Dwivedi *et al.*, 2006).

1.4. Mapeamento genético

1.4.1. Construção de mapas genéticos

Os fundamentos teóricos sobre ligação gênica, base para a construção de mapas genéticos, surgiram com os trabalhos de Morgan e colaboradores (1910) e Sturtevant, que em 1913 sugeriu a utilização de frequência de recombinantes como medida da distância entre dois genes (Coelho & Silva, 2002). Mas foi com o surgimento dos marcadores moleculares, na década de 80, que a construção de mapas genéticos saturados tornou-se possível. Marcadores moleculares são disponíveis em grande número, são polimórficos, seletivamente neutros e não sofrem influência ambiental. Mapas genéticos de marcadores moleculares permitem identificar, mapear e medir a magnitude do efeito dos principais genes envolvidos no controle de características monogênicas ou quantitativas (Ferreira & Grattapaglia, 1998). Uma vez mapeados, marcadores associados a esses genes podem ser utilizados para seleção indireta, o que torna mais rápido e eficiente o melhoramento tradicional (seleção assistida por marcadores).

Mapas genéticos também servem de base para o isolamento de genes, visando sua utilização em biotecnologia. Essa estratégia é conhecida como clonagem baseada em mapeamento ou clonagem posicional. Para isso, é necessária a construção de uma biblioteca, contendo grandes fragmentos de DNA genômico (por exemplo, BACs – “bacterial artificial chromosomes”). Nos últimos anos, bibliotecas de BACs têm sido construídas para diversas espécies de plantas, incluindo o amendoim cultivado (Yuksel & Paterson, 2005) e espécies silvestres de *Arachis* com genomas AA (*A. duranensis*) e BB (*A. ipaënsis*) (Bertioli e colaboradores, ainda não publicado).

A construção de mapas genéticos envolve três etapas principais: (1) produção da população de mapeamento; (2) identificação de polimorfismos e (3) análise de ligação entre os marcadores.

Populações de mapeamento

A construção de mapas genéticos baseia-se na existência de desequilíbrio de ligação, que é definido como desvios das frequências alélicas, em relação às frequências esperadas sob independência (Coelho & Silva, 2002). Fatores como migração e seleção podem causar desequilíbrio de ligação e, por isso, para a construção de mapas genéticos é necessário o desenvolvimento de populações de mapeamento. Nesse caso, o desequilíbrio de ligação é causado, basicamente, pela ligação física entre os locos. Os indivíduos utilizados como progenitores na formação dessas populações devem ser contrastantes para características de interesse, além de possuírem uma distância genética suficiente para a identificação de marcadores polimórficos, possibilitando a construção do mapa. O tamanho das populações de mapeamento em plantas varia, geralmente, de 50 a 250 indivíduos, mas populações maiores são necessárias para mapeamento de alta resolução (Mohan *et al.*, 1997; Collard *et al.*, 2005).

Diferentes tipos de população podem ser utilizados para a construção de mapas, sendo as mais comuns as populações obtidas por retrocruzamentos, populações F_2 , linhagens puras recombinantes (RILs ou “Recombinant Inbred Lines”), linhagens de duplo-haplóides e, no caso de espécies de fecundação cruzada, cruzamentos entre indivíduos heterozigotos (Ferreira & Grattapaglia, 1998; Collard *et al.*, 2005). Cada um desses delineamentos possui vantagens e desvantagens. Populações F_2 e de retrocruzamentos são as mais fáceis e rápidas de serem obtidas, mas apresentam a desvantagem de que os indivíduos são únicos, o que dificulta a avaliação de uma determinada característica em diferentes ambientes. Uma vantagem da população F_2 sobre a população de retrocruzamentos, é que os indivíduos da geração F_2 são formados por dois gametas com locos em desequilíbrio de ligação, e não por um único, como no caso dos indivíduos obtidos por retrocruzamentos (Coelho & Silva, 2002). Linhagens puras recombinantes são obtidas por autofecundações sucessivas, a partir de F_2 , gerando uma série de linhagens homozigotas, cada uma contendo uma combinação única de segmentos cromossômicos provenientes dos progenitores. A principal desvantagem é o longo tempo necessário para desenvolvimento dessas populações, que requer, geralmente, de seis a oito gerações (Collard *et al.*, 2005). Populações de duplo-haplóides são produzidas pela indução de duplicação cromossômica a partir de

grãos-de-pólen, regenerando plantas inteiras. Esses dois últimos delineamentos apresentam a vantagem de produzir populações imortalizadas, que podem ser multiplicadas e reproduzidas, sem alterações genéticas. Isso possibilita a condução de ensaios em diferentes ambientes e anos. Para espécies alógamas, de ciclo longo ou que apresentam depressão por endogamia, a obtenção de linhagens endogâmicas é impraticável. Nesse caso, o cruzamento entre indivíduos heterozigotos é geralmente utilizado para a construção de mapas. A escolha do tipo de população mais apropriada ao mapeamento está relacionada à espécie em questão, mas também ao tipo de marcador empregado (Tanksley *et al.*, 1988), considerando que o máximo de informação genética é alcançado com uma população F_2 e marcadores codominantes (Staub *et al.*, 1996; Carneiro & Vieira, 2002).

Identificação de polimorfismos

O segundo passo na construção de mapas genéticos é a identificação de marcadores polimórficos. Nesse caso, algumas centenas de marcadores são testadas nas plantas progenitoras, dependendo da distância genética entre elas. Para escolha do tipo de marcador molecular a ser utilizado devem ser consideradas as características de cada marcador, com suas vantagens e desvantagens, além de outros fatores, como a disponibilidade de marcadores para a espécie em questão, a infra-estrutura laboratorial disponível e o treinamento do pessoal envolvido. Uma vez identificados, os marcadores polimórficos são genotipados na população de mapeamento e os padrões de segregação analisados (1:2:1, 3:1 ou 1:1, dependendo do tipo de população e marcador utilizados). Para isso, são utilizados geralmente testes de X^2 .

Análise de ligação entre os marcadores

As etapas seguintes envolvem a detecção de ligação entre os marcadores, a determinação das distâncias genéticas e o ordenamento dos marcadores nos grupos de ligação.

A ligação entre cada par de marcadores é verificada por comparações entre as frequências alélicas observadas e as frequências esperadas sob independência. Para isso, geralmente são utilizados os testes de significância de X^2 ou pela razão

de verossimilhança (LR). Segundo Lynch & Walsh (1998), LR é calculada pela razão da função de verossimilhança, sob a hipótese de que há ligação e sob a hipótese nula de que não há ligação. Esse teste estatístico tem distribuição aproximada de X^2 . Para facilitar a interpretação desse teste, tem sido utilizada a estatística do LOD score (“Log of the odds”), que consiste em utilizar a estatística da razão de verossimilhança convertida para o logaritmo na base 10. Em geral, consideram-se ligados locos que apresentam valores de LOD superiores ou iguais a 3, uma vez que esse valor indica que a hipótese alternativa é $10^3 = 1000$ vezes mais verossímil do que a hipótese nula de segregação independente.

As distâncias genéticas entre os marcadores ligados são calculadas, pelas frequências de recombinantes, em porcentagens. No entanto, essa medida não é aditiva, uma vez que permutas duplas em segmentos adjacentes não são detectadas. Com isso, a inclusão de novos marcadores no mapa requer que as distâncias já mapeadas sejam ajustadas. Para se evitar esse problema, são utilizadas funções de mapeamento. Um grande número de funções de mapeamento foi desenvolvido, mas as mais utilizadas em plantas são as funções de Haldane (1919) e de Kosambi (1944) (Liu, 1998; Coelho & Silva, 2002). A função de Haldane assume que as ocorrências de permutas gênicas em segmentos adjacentes são independentes (sem interferência), enquanto a de Kosambi leva em consideração a interferência, sendo, provavelmente, mais próxima da situação real. As funções de mapeamento são, em geral, expressas em cM (centiMorgan) que é igual a 0,01 Morgan. Um Morgan equivale à distância entre dois genes ao longo da qual se espera que ocorra uma permuta (“crossing-over”) por gameta por geração (Weir, 1996).

Uma vez conhecidas as frações de recombinação entre cada par de marcadores, os grupos de ligação são identificados. Do ponto de vista biológico, grupos de ligação são definidos como grupos de genes (marcadores), cujos locos estão localizados no mesmo cromossomo (Carneiro & Vieira, 2002). Programas como Mapmaker (Lander *et al.*, 1987) e QMOL (Cruz, 2004) permitem a detecção de ligação em um conjunto de centenas de marcadores, como é comum nos experimentos atuais, e o ordenamento dos marcadores dentro dos grupos de ligação. O estabelecimento dos grupos de ligação, usando-se esses programas, é feito pela função de verossimilhança ou LOD score. Dessa forma, um valor de LOD mínimo é utilizado (geralmente 3) para identificação dos grupos de ligação. A

etapa seguinte consiste na determinação da melhor ordem dos marcadores dentro de cada grupo. Historicamente, o ordenamento de locos tem sido um processo que visa minimizar o número de “crossing-overs” ou permutas genéticas (Liu, 1998). O ordenamento de grupos de três marcadores pode ser feito, de maneira simples, pela identificação dos duplo-recombinantes, ou seja, por comparações das classes paternas (com maior frequência) e as classes originadas de permutas duplas, com menor frequência, obtidas a partir do cruzamento entre o triplo heterozigoto e um homozigoto recessivo (Schuster & Cruz, 2004). Para ordenamento de marcadores múltiplos, vários métodos têm sido desenvolvidos, incluindo mínima soma de frações de recombinação adjacentes (SARF), mínimo produto de frações de recombinação adjacentes (PARF), máxima soma de LOD escores adjacentes (SALOD), mínima soma de probabilidades de duplas recombinações (PDR) e máxima verossimilhança (ML) (Liu, 1998; Carneiro & Vieira, 2002; Coelho & Silva, 2002). Esses métodos são impraticáveis quando muitos locos estão envolvidos, uma vez que o número de ordens possível para m locos é igual a $m!/2$. Por isso, alguns algoritmos têm sido utilizados para determinação da ordem dos marcadores nos grupos de ligação, tais como seriação, anelamento simulado, ramos e conexões ou combinações de mais de um desses métodos (Liu, 1998). Seriação e anelamento simulado são algoritmos de aproximação e, portanto, não garantem que a melhor ordem seja obtida. Por outro lado, o método de ramos e conexões garante a melhor ordem, mas a análise pode demorar bastante tempo, no caso de grupos com muitos marcadores. Simulações de ordenamento de locos usando cinco desses métodos (SARF, PARF, SALOD, ML e seriação) em uma população hipotética de duplo-haplóides resultaram em coeficientes de correlação maiores do que 0,99, indicando que todos os métodos produziram ordens de marcadores bastante similares (Wu *et al.*, 2003). Deve-se ressaltar, no entanto, que a acurácia da determinação da melhor ordem de marcadores em um grupo de ligação está diretamente relacionada ao número de indivíduos da população de mapeamento e, quanto maior o número de indivíduos, maior a acurácia.

1.4.2. Mapas genéticos em *Arachis*

Mapas de ligação baseados em diferentes tipos de marcadores moleculares têm sido construídos para as mais diversas espécies de plantas (revisado por Ferreira

& Grattapaglia, 1998; Collard *et al.*, 2005). No entanto, poucos mapas genéticos foram publicados para *Arachis*. Halward e colaboradores (1993) desenvolveram um mapa baseado em marcadores RFLP (“Restriction Fragment Length Polymorphism”), utilizando uma população de 87 indivíduos F₂, derivada de um cruzamento entre duas espécies silvestres diplóides com genoma AA (*A. stenosperma* x *A. cardenasii*). Foram mapeados 117 locos em 11 grupos de ligação, com uma distância de mapa total de 1063 cM e uma distância média de 9,08 cM. Com base nesse mapa, embora pouco saturado, Garcia e colaboradores (1995) utilizaram marcadores RFLP dispersos pelos 11 grupos de ligação, junto a marcadores RAPD (“Random Amplified Polymorphic DNA”) não mapeados, para a análise de linhas de introgressão tetraplóides, resultantes de um cruzamento entre *A. hypogaea* (AABB) e *A. cardenasii* (genoma AA, duplicado com colchicina). Com isso, os autores detectaram a introgressão de regiões cromossômicas da espécie silvestre diplóide no amendoim cultivado. Essa análise mostrou que recombinação, e não substituição cromossômica era o mecanismo de introgressão, o que é um resultado importante para o uso de espécies silvestres de *Arachis* nos programas de melhoramento do amendoim.

Devido à baixa variabilidade genética do amendoim cultivado, populações anfidiplóides sintéticas têm sido usadas para a construção de mapas genéticos (Dwivedi *et al.*, 2006). Dessa forma, uma planta anfidiplóide sintética, obtida pelo cruzamento de {*A. batizocoi* x (*A. cardenasii* x *A. diogeni*)} e posterior duplicação de cromossomos com colchicina, foi utilizada como parental doador em um cruzamento com *A. hypogaea*, gerando uma população RC₁ de 78 indivíduos, que foi utilizada para a construção de um mapa de ligação (Burow *et al.*, 2001). *A. cardenasii* e *A. diogeni* possuem genoma AA, enquanto *A. batizocoi* possui genoma BB. Esse mapa consistiu de 370 marcadores RFLP, mapeados em 23 grupos de ligação, com uma distância total de 2210 cM e uma densidade média de 5,97 cM. Esse foi o primeiro mapa tetraplóide publicado para *Arachis*. Recentemente, o primeiro mapa para o amendoim cultivado, embora bastante preliminar, foi publicado, a partir de uma população F₂ (Herselman *et al.*, 2004). Cinco grupos de ligação com apenas 11 marcadores AFLP (“Amplified Fragment Length Polymorphism”) foram obtidos, cobrindo 139,4 cM do genoma.

1.5. Marcadores moleculares

1.5.1. Marcadores microssatélites

A análise com marcadores RFLP requer pessoal técnico habilitado, instalações adequadas para a manipulação de DNA recombinante e de material radioativo e é uma técnica laboriosa e de alto custo, enquanto marcadores RAPD e AFLP são dominantes e, portanto, pouco informativos. Conseqüentemente, esses três tipos de marcadores moleculares não são ferramentas ideais para uso em seleção assistida. Microssatélites ou SSR (“simple sequence repeats” ou seqüências repetitivas simples) são seqüências de 1 a 6 nucleotídeos que se repetem lado a lado na fita de DNA (Figura 1), no genoma de eucariotos (Tautz & Renz, 1984; Litt & Luty, 1989). Regiões contendo essas seqüências podem ser amplificadas por PCR (Reação em Cadeia da Polimerase), utilizando um par de primers específicos (cerca de 20-25 bases) complementares a seqüências conservadas que flanqueiam o microssatélite. Essa técnica revela polimorfismo em um loco, devido a diferenças no número de vezes que o microssatélite (por exemplo, CA, AG, TTG, ATGC) repete-se naquele loco (Tautz, 1989; Weber & May 1989; Hancock, 1995). Cada segmento amplificado de tamanho diferente representa um alelo diferente do mesmo loco. A taxa de mutação nestas seqüências repetitivas é em geral alta. Isso ocorre devido à alta incidência de “slippage” ou escorregamento da enzima polimerase, causando um pareamento incorreto dos microssatélites durante a replicação, resultando em cópias de trechos de DNA com tamanhos diferentes (Levinson & Gutman, 1987; Schlötterer & Tautz, 1992). Marcadores SSR caracterizam-se por serem codominantes, abundantes, aparentemente distribuídos por todo o genoma, multialélicos e dependentes de pequena quantidade de DNA dos indivíduos analisados, por serem baseados em PCR e em eletroforese em gel.

o uso dos marcadores microssatélites em estudos de genômica comparativa e para validação de mapas de ligação obtidos em diferentes “backgrounds” genéticos. A principal desvantagem desses marcadores é o alto custo inicial para seu desenvolvimento, que geralmente requer a construção de bibliotecas genômicas enriquecidas para seqüências contendo SSRs. No entanto, com a crescente disponibilidade de etiquetas de seqüências expressas (“expressed sequence tags” ou ESTs) e de outros dados de seqüências de DNA, o desenvolvimento de marcadores SSR, por meio de mineração de dados (“data mining”), tornou-se uma opção rápida, eficiente e de baixo custo para muitas espécies de plantas.

Um grande número de marcadores microssatélites tem sido desenvolvido para o amendoim nos últimos anos. Hopkins e colaboradores (1999) publicaram os primeiros marcadores microssatélites para o amendoim, desenvolvidos a partir de bibliotecas genômicas enriquecidas para repetições GT/CA e CT/GA. De um total de 26 pares de primers desenhados, seis mostraram-se polimórficos em uma análise de 19 acessos de *A. hypogaea*, incluindo quatro das seis variedades botânicas conhecidas. Nesse trabalho, apenas as seqüências dos seis primers polimórficos foram publicadas. Posteriormente, 56 novos marcadores foram desenvolvidos a partir de uma biblioteca genômica enriquecida para repetições GA/CT (He *et al.*, 2003). Desses, 19 mostraram-se polimórficos na análise de 24 acessos de amendoim, incluindo as seis variedades conhecidas. Esses mesmos autores publicaram outros 19 marcadores SSR polimórficos para o amendoim, sendo 10 a partir da biblioteca mencionada acima e nove, a partir de uma biblioteca de cDNAs (He *et al.*, 2005). O número de marcadores microssatélites disponível para *Arachis* aumentou consideravelmente com a publicação de Ferguson e colaboradores (2004b). Nesse trabalho, foram desenvolvidos 192 novos marcadores, a partir de bibliotecas enriquecidas para diferentes repetições de di, tri e tetranucleotídeos. A análise de 24 acessos das seis variedades de *A. hypogaea* resultou na identificação de 110 marcadores polimórficos para o amendoim. Além disso, 67 marcadores microssatélites foram desenvolvidos, a partir de uma biblioteca genômica enriquecida para repetições TTG/AAC, dos quais apenas três mostraram-se polimórficos na análise de cinco acessos de *A. hypogaea* representando cinco variedades botânicas (Moretzsohn *et al.*, 2004). Finalmente, 36 marcadores microssatélites foram desenvolvidos para *A. hypogaea*, *A. pintoii* e *A. glabrata* pelo mesmo grupo de pesquisa, a partir de bibliotecas genômicas enriquecidas para

AC/TG, AG/TC, TA/AT e TTC/AAG (Palmieri *et al.*, 2002; 2005; Bravo *et al.*, 2006; Gimenes *et al.*, 2006). Devido à estreita base genética de *A. hypogaea*, esses 396 marcadores disponíveis são ainda insuficientes para a construção de mapas genéticos intra-específicos para o amendoim e, mesmo, para a construção de mapas saturados para as espécies silvestres de *Arachis*. Por isso, neste trabalho, uma bateria de marcadores microssatélites foi desenvolvida, a partir de bibliotecas genômicas enriquecidas, de bibliotecas de cDNAs e por busca de seqüências no GenBank (Capítulo 1).

1.5.2. Marcadores âncoras

Diversas iniciativas de seqüenciamento de ESTs ou do genoma completo encontram-se em andamento para diferentes espécies de plantas, incluindo leguminosas. Com isso, tornou-se possível o desenvolvimento de marcadores universais (âncoras), que são marcadores gênicos e de cópias únicas, transferíveis entre espécies de uma mesma família, o que possibilita seu uso em estudos de genômica comparativa. Mapeamento genético comparativo permite a integração de informações entre diferentes “taxa” e possibilita que espécies menos estudadas, como *Arachis*, se beneficiem da grande quantidade de dados genômicos que estão sendo gerados para espécies-modelos de leguminosas, como *Lotus japonicus* e *Medicago truncatula*, a partir dos projetos de seqüenciamento em larga escala que se encontram atualmente em andamento para estas duas espécies (Young *et al.*, 2005; Cannon *et al.*, 2006).

Em leguminosas, o desenvolvimento desses marcadores tem sido feito seguindo a metodologia descrita por Lyons e colaboradores (1997), mas usando programas de computador recentemente desenvolvidos para aumentar a eficiência do processo (Fredslund *et al.*, 2005; Fredslund *et al.*, 2006). Por comparações de ESTs de *Medicago truncatula*, *Glycine max* e *Lotus japonicus*, são identificadas seqüências conservadas evolutivamente (ECSs ou “evolutionary conserved sequences”). As ECSs são alinhadas e introns são identificados, uma vez que essas regiões são mais sujeitas a mutações e, portanto, tendem a ser mais polimórficas do que as regiões codificadoras. O comprimento dos introns é importante porque a PCR final é limitada a um “amplicon” de 2,5 kb, aproximadamente, usando-se polimerases padrões e também porque introns muito pequenos têm menor probabilidade de ser

polimórficos. Em seguida, seqüências com similaridades únicas no proteoma de *Arabidopsis* são selecionadas para desenho de primers, geralmente com bases degeneradas, complementares às seqüências conservadas que flanqueiam os introns. Esses primers são então usados em PCR com os progenitores das populações de mapeamento e os fragmentos obtidos são seqüenciados para identificação de polimorfismos em introns e desenho de novos primers, específicos para *Arachis* (ou outra espécie em questão). Dessa forma, é possível amplificar regiões ortólogas, com primers específicos para cada espécie analisada.

Alguns desses marcadores revelam polimorfismo de comprimento. No entanto, para a grande maioria, são identificados SNPs (“single nucleotide polymorphisms” ou polimorfismos de nucleotídeo único) e os produtos de PCR são clivados com enzimas de restrição específicas para cada marcador, gerando marcadores dos tipos CAPS – “Cleaved Amplified Polymorphic Sequence” (Konieczny & Ausubel, 1993) ou dCAPS – “derived CAPS” (Neff *et al.*, 1998). A técnica de CAPS consiste do uso de primers específicos para amplificar uma determinada região no genoma, na qual polimorfismos são detectados pela perda ou ganho de um sítio de reconhecimento de uma determinada enzima de restrição. Na técnica dCAPS, um sítio de reconhecimento de uma enzima de restrição que inclui um SNP é introduzido no produto de PCR por um primer que contém uma ou mais bases não complementares ao DNA molde. Essas duas técnicas apresentam as vantagens de produzirem marcadores codominantes e de serem baseadas em PCR e eletroforese em gel de agarose ou poliacrilamida.

1.5.3. Análogos a genes de resistência

Ao longo da evolução, as plantas têm desenvolvido vários mecanismos de defesa para sua proteção contra doenças e parasitas. Um desses mecanismos envolve o reconhecimento específico do patógeno. Entre os eventos celulares que caracterizam este tipo de resistência estão a indução da expressão de genes de defesa, freqüentemente manifestada como uma resposta de hipersensibilidade envolvendo morte celular programada no local da infecção (De Wit, 1995). Estes genes são chamados genes de resistência (R-genes).

Nos últimos anos, um grande número de genes de resistência para diferentes patógenos tem sido isolado em plantas (Dixon *et al.*, 1998; Yoshimura *et al.*, 1998;

Wang *et al.*, 1999; Wang *et al.*, 2002). Esses genes de resistência podem ser agrupados pela similaridade em sua seqüência de aminoácidos e apresentam regiões comuns. O maior grupo de genes de resistência, em plantas, codifica para proteínas com o que se supõe ser uma região para ligação de nucleotídeos (NBS – “nucleotide binding sites”) no N-terminal, assim como uma região com repetições de leucina (LRR – “leucine-rich repeat”) no C-terminal. O domínio LRR parece estar ligado ao reconhecimento do patógeno (Jones & Jones, 1997). Em plantas, a única função associada ao domínio NBS é em resistência a doenças (Tameling *et al.*, 2002). Enquanto a região LRR é muito variável, a região NBS apresenta motivos conservados (Kobe & Deisenhofer, 1994). Esses motivos (por exemplo, P-loop, Kinase-2, Kinase-3a) têm sido bastante utilizados para o desenho de primers degenerados. Os fragmentos amplificados usando-se esses primers são seqüenciados e o produto de tradução putativo desses fragmentos tem indicado similaridade com genes de resistência conhecidos. Portanto, estes fragmentos são chamados análogos a genes de resistência (RGAs). Esses trabalhos têm mostrado que RGAs são freqüentemente ligados a genes de resistência ou, mesmo, partes de genes de resistência funcionais (Kanazin *et al.* 1996; Aarts *et al.*, 1998; Collins *et al.*, 2001; Donald *et al.*, 2002; McIntyre *et al.*, 2005; Mammadov *et al.*, 2006).

Essa metodologia foi utilizada, com sucesso, em *Arachis* (Bertioli *et al.*, 2003; Yuksel *et al.*, 2005). No primeiro trabalho, primers degenerados para a região NBS foram desenhados e usados para amplificação em *A. hypogaea* var. Tatu e quatro espécies silvestres (*A. duranensis*, *A. cardenasii*, *A. stenosperma* e *A. simpsonii*) que apresentam resistências a várias doenças que atacam o amendoim. Setenta e oito RGAs foram identificados, sendo que 63 deles apresentavam ORFs (“Open Reading Frames”) não interrompidos. Análises filogenéticas dessas 63 seqüências e de RGAs de outras espécies, como *Glycine max*, *Lotus japonicus*, *Medicago truncatula*, *Phaseolus vulgaris* e *Arabidopsis thaliana* mostraram a formação de clados, específicos para as leguminosas, de acordo com o tipo de RGA.

Representantes de cada um desses clados foram utilizados como sondas na análise de RFLPs e primers complementares ao motivo kinase-2 foram utilizados em uma técnica de AFLP modificada (Hayes & Saghai Maroof, 2000; van der Linden *et al.*, 2004), para desenvolvimento e mapeamento de RGAs em *Arachis* spp. (José, 2006). Essa técnica é similar ao AFLP (Vos *et al.*, 1995), mas, na segunda PCR é utilizado um primer, específico ou degenerado, complementar a regiões conservadas

do motivo NBS, além de um primer complementar ao adaptador ligado à extremidade do fragmento cortada com a enzima de corte freqüente (nesse caso, *Mse I*).

1.6. Mapeamento de QTLs e seleção assistida por marcadores

Tradicionalmente, nos programas de melhoramento de plantas em que se utilizam espécies silvestres, são comuns os retrocruzamentos para a introdução de genes que conferem características de interesse, como por exemplo, genes de resistência a pragas ou doenças. Em cada geração de retrocruzamento são selecionadas as plantas que apresentam o fenótipo do alelo a ser introgridido e o fenótipo do parental recorrente para as outras características. Este processo resulta na manutenção de parte do genoma doador associada ao gene a ser introgridido (“linkage drag”), ou seja, genes não desejáveis ou deletérios do parental doador podem ser transferidos junto aos genes de interesse. Nesse esquema, várias gerações de retrocruzamento (8 a 10 gerações) são, em geral, necessárias para a eliminação das características indesejáveis.

O processo de seleção após a introgressão pode ser otimizado por meio da construção de mapas genéticos e da identificação de marcadores moleculares associados aos genes de interesse. Desta forma, é possível, em cada fase de seleção, a escolha dos indivíduos que contêm o(s) gene(s) de interesse, mas também, o máximo do genótipo do parental recorrente, diminuindo assim o tempo do processo de melhoramento. Isso é particularmente importante em programas que visam a introgressão de genes de resistência a doenças. Neste caso, a seleção indireta através de marcadores moleculares oferece, além da redução de tempo, a possibilidade de identificação e seleção de genótipos resistentes, sem que haja necessidade da presença do patógeno e ainda possibilita a transferência e a manutenção de mais de um gene de resistência principal durante a seleção (piramidização de genes de resistência).

Para implementação de um esquema efetivo de seleção assistida por marcadores em programas de melhoramento genético, três requerimentos são essenciais: (1) os marcadores devem estar fortemente ligados aos genes que controlam as características de interesse; (2) os marcadores moleculares utilizados devem estar facilmente disponíveis para a análise de grandes populações (portanto,

baseados em PCR), devem ser codominantes e dispersos por todo o genoma e (3) as técnicas utilizadas devem ser de baixo custo e reproduzíveis entre laboratórios (Mohan *et al.*, 1997).

A maioria das características importantes agronomicamente, como produtividade, precocidade, qualidade do grão e resistências a alguns estresses bióticos ou abióticos em plantas é controlada por vários genes (características poligênicas, quantitativas ou de herança complexa). Nesse caso, os fenótipos resultantes apresentam uma variação contínua, em vez de classes fenotípicas discretas. Os genes ou locos cromossômicos que controlam uma determinada característica quantitativa são denominados QTLs (“Quantitative Trait Loci”).

O mapeamento de QTLs consiste em se detectar uma associação entre o fenótipo em questão e o genótipo de marcadores. Marcadores são utilizados para dividir a população de mapeamento em classes genotípicas diferentes de acordo com os genótipos do loco marcador. Em seguida, métodos estatísticos são utilizados para determinar se os indivíduos das diferentes classes genotípicas diferem significativamente, com respeito à característica sendo avaliada (Tanksley, 1993). Se há diferenças significativas, considera-se que os genes que controlam a característica em questão estão ligados ao loco marcador. Os testes de significância desses parâmetros são, geralmente, feitos pela razão de verossimilhança (LR) ou por LOD score. Os valores de LR ou de LOD são estimados para cada posição do cromossomo e então plotados em gráficos, para facilitar a visualização.

Três métodos têm sido mais utilizados para mapeamento de QTLs: análises de marcas simples, mapeamento por intervalo (Lander & Botstein, 1989) e mapeamento por intervalo composto (Jansen, 1993; Zeng, 1993; 1994). O método de marcas simples baseia-se na comparação entre as médias fenotípicas obtidas para as diferentes classes genotípicas de um determinado marcador individualmente. Os métodos estatísticos geralmente utilizados para verificar se as diferenças observadas são significativas incluem o teste de t, análises de variância (ANOVA), regressão linear simples e o método de máxima verossimilhança (Ferreira & Grattapaglia, 1998; Coelho & Silva, 2002). Esse método é o mais simples e não requer a construção de mapas. No entanto, de acordo com Liu (1998), a posição do QTL não pode ser precisamente determinada e não é possível distinguir um QTL de pequeno efeito situado muito próximo ao marcador de um QTL de grande efeito situado mais longe.

O mapeamento por intervalo é feito com base nas frequências conjuntas de um par de marcadores adjacentes e um QTL localizado entre esses dois marcadores e, portanto, requer a existência de um mapa de ligação. De acordo com Schuster & Cruz (2004), esse método consiste resumidamente das seguintes etapas: (a) como a distância entre cada par de marcadores é conhecida, o método investiga o intervalo em incrementos predefinidos, usando a função de mapeamento e estimando os parâmetros do modelo a cada ponto de análise; (b) a função de verossimilhança (ou a análise de regressão, Haley & Knott, 1992) é então usada para os cálculos, utilizando os valores estimados e também sob a hipótese nula de ausência de QTL, sendo calculada na forma de LOD scores para cada ponto ao longo do genoma; (c) o máximo LOD score entre todas as análises indica a presença de um QTL, caso seja maior do que um mínimo estabelecido estatisticamente. Esse método apresenta vantagens em relação ao método de marcas simples, pois permite que sejam estimados os parâmetros envolvidos na análise, como a média, a variância, o efeito do QTL e a distância entre o QTL e os marcadores adjacentes (Lander & Botstein, 1989; Liu, 1998). No entanto, esse método apresenta problemas quando há QTLs ligados ou interações entre QTLs (Liu, 1998). Por isolar o intervalo analisado do restante do genoma, não considera os efeitos de outros QTLs próximos ao QTL contido no intervalo, levando à identificação de “QTLs fantasmas” (Lynch & Walsh, 1998; Coelho & Silva, 2002).

Esses problemas são quase totalmente resolvidos com o método de mapeamento por intervalo composto (Schuster & Cruz, 2004). Esse método combina o mapeamento por intervalo com a análise de regressão linear múltipla, com o intuito de se eliminar a interferência de QTLs adjacentes ao intervalo que está sendo mapeado, reduzindo a identificação de QTLs inexistentes (Coelho & Silva, 2002). Atualmente, esse tem sido o método mais utilizado, por permitir um aumento na precisão e no poder de detecção de QTLs, especialmente quando QTLs ligados estão envolvidos. No método de mapeamento por intervalo composto, marcadores, selecionados por regressão, são utilizados como cofatores, permitindo controlar os efeitos de outros QTLs, enquanto o QTL do intervalo de interesse é investigado. No entanto, esse método considera um único QTL de cada vez. Recentemente, um quarto método foi desenvolvido, chamado mapeamento por intervalo múltiplo (Kao *et al.*, 1999). Esse método utiliza intervalos múltiplos de marcadores simultaneamente para a identificação de múltiplos possíveis QTLs. Com isso, podem ser estimados os

efeitos de interação entre QTLs (epistasia), os valores genotípicos dos indivíduos e as herdabilidades das características quantitativas (Kao *et al.*, 1999). Comparado aos métodos de marcas simples, de mapeamento por intervalo e por intervalo composto, esse método permite um aumento na precisão e no poder de detecção de QTLs (Kao *et al.*, 1999; Coelho & Silva, 2002).

1.7. Mapeamento de QTLs em *Arachis*

Poucos estudos de mapeamento de genes individuais ou de QTLs têm sido publicados em *Arachis*. Garcia e colaboradores (1996) identificaram dois genes dominantes, ligados, que conferiam resistência ao nematóide-das-galhas (*Meloidogyne arenaria* raça 1), usando uma população F₂ tetraplóide, derivada de um cruzamento entre *A. hypogaea* e *A. cardenasii* (tratada com colchicina). Esses genes foram designados *Mae*, gene que reduz a capacidade de reprodução do patógeno (produção de ovos) e *Mag*, gene que inibe a formação de galhas. Marcadores RAPD, SCAR e RFLP, junto a BSA ou “Bulked Segregant Analysis” (Michelmore *et al.*, 1991), foram utilizados para análise dessa população F₂, sendo detectada a ligação de um marcador RAPD a 10 cM e a 14 cM de *Mag* e *Mae*, respectivamente. Também para mapeamento de genes de resistência a nematóides, uma população de retrocruzamento (BC₄F₂) entre *A. hypogaea* e uma planta anfidiplóide sintética, obtida por duplicação de cromossomos com colchicina da planta resultante de cruzamentos entre três espécies silvestres diplóides {*A. batizocoi* x (*A. cardenasii* x *A. diogeni*)}^c foi utilizada, junto à técnica de BSA (Burow *et al.*, 1996). Três marcadores RAPD foram identificados ligados a um único gene dominante (*RKN* – para “Root-knot nematode”), que confere resistência a *M. arenaria*, estando o mais próximo a 5,4 cM desse gene. Análises de seis populações BC₅F₂ resultantes desses cruzamentos, possibilitaram a identificação de três marcadores RFLP distantes a 4,2; 5,2 e 11,2 cM desse gene (Choi *et al.*, 1999). Os dois marcadores RFLP mais próximos a esse gene de resistência foram utilizados, posteriormente, na seleção assistida do amendoim (Church *et al.*, 2000). Com isso, indivíduos homocigotos para resistência a *M. arenaria* puderam ser identificados, ainda no viveiro, resultando em grande economia de tempo e espaço. No entanto, esses marcadores encontram-se ainda distantes dos genes de resistência, resultando em erros de identificação das plantas selecionadas (estimados em 4%).

Stalker & Mozingo (2001) relataram a identificação de marcadores RAPD que explicavam cerca de 35% da variação para resistência à mancha castanha (*Cercospora arachidicola*) em uma população de amendoim que tinha *A. cardenasii* em seu “pedigree” e 10% da variação em um cruzamento de *A. hypogaea* x *A. hypogaea*. Esse último foi o primeiro relato de marcadores moleculares associados a genes de resistência em um cruzamento intraespecífico de *A. hypogaea*. Recentemente, marcadores AFLP, em conjunto com BSA, foram utilizados para o mapeamento de genes de resistência a um afídeo, vetor do vírus “rosette”, que é a principal doença do amendoim causada por vírus na África (Herselman *et al.*, 2004). Foi utilizada uma população segregante F₂, resultante do cruzamento entre dois acessos de *A. hypogaea*. Um marcador foi mapeado a 3,9 cM de um único gene recessivo, que explicava 76,1% da variação para resistência ao afídeo.

2. OBJETIVOS GERAIS

- 1 - Desenvolver uma bateria de marcadores microssatélites para o amendoim, a partir de bibliotecas genômicas enriquecidas para dinucleotídeos, de bibliotecas de ESTs e por mineração (“data mining”) de seqüências depositadas no GenBank para *Arachis* spp. (Capítulo 1);
- 2 - Caracterizar os marcadores obtidos em um grupo de acessos de *A. hypogaea*, representando as seis variedades botânicas descritas para o amendoim (Capítulo 1);
- 3 - Construir um mapa genético para espécies diplóides de *Arachis*, de genoma AA, usando marcadores microssatélites, marcadores âncoras e RGAs (Capítulo 2);
- 4 - Mapear QTLs que controlam características de interesse agrônômico para o melhoramento genético do amendoim, sendo dois componentes da produtividade (peso de sementes e número de sementes por planta) e resistência ao fungo causador da mancha preta (*Cercosporidium personatum*) no amendoim (Capítulo 2).

Capítulo 1

Desenvolvimento e caracterização de marcadores microssatélites para o amendoim (*Arachis hypogaea* L.)*

*Este capítulo é a tradução de parte do artigo publicado na Theoretical and Applied Genetics (Moretzsohn *et al.*, 2005 – ANEXO 1), acrescida da análise das relações genéticas entre acessos de *Arachis hypogaea*.

1. INTRODUÇÃO

Arachis hypogaea possui considerável variabilidade para diversas características morfológicas, fisiológicas e agrônômicas. No entanto, pouca variação tem sido detectada por marcadores moleculares, tais como RAPD, AFLP, RFLP e microssatélites (Halward *et al.*, 1991; Kochert *et al.*, 1991, 1996; Lanham *et al.*, 1992; Paik-Ro *et al.*, 1992; Hilu & Stalker, 1995; He & Prakash, 1997; Subramanian *et al.*, 2000; Gimenes *et al.*, 2002; Herselman, 2003; Moretzsohn *et al.*, 2004; He *et al.*, 2005; Milla *et al.*, 2005). Acredita-se que essa baixa variabilidade genética seja decorrente de três fatores: (1) isolamento reprodutivo do amendoim em relação às espécies silvestres diplóides, devido à poliploidização (Young *et al.*, 1996); (2) evento de poliploidização recente, a partir de um ou poucos indivíduos de cada espécie ancestral, associado ao modo de reprodução autógamo do amendoim (Halward *et al.*, 1991) e (3) uso de poucas linhagens-elite e pouco germoplasma exótico nos programas de melhoramento da cultura, resultando em uma base genética estreita (Knauff & Gorbet, 1989; Isleib & Wynne, 1992; Isleib *et al.*, 2001).

Essa pouca variabilidade genética tem dificultado os avanços no melhoramento molecular do amendoim. A identificação de marcadores moleculares polimórficos seria, portanto, de grande utilidade. Marcadores microssatélites constituem a ferramenta ideal para a construção de mapas genéticos, para o mapeamento de QTLs e para o uso em seleção assistida por marcadores, por serem multialélicos, codominantes e baseados em PCR. Microssatélites têm sido utilizados em diversas espécies de plantas com esse fim (revisto por Gupta & Varshney, 2000, Ellegren, 2004). No entanto, pouco tem sido feito com o amendoim e, até o momento, nenhum mapa baseado em marcadores microssatélites foi publicado para *Arachis*.

Até recentemente, marcadores microssatélites eram desenvolvidos basicamente a partir de bibliotecas genômicas enriquecidas para essas seqüências. Esse processo apresenta custo elevado, é lento e bastante intensivo em mão-de-obra, considerando-se todas as etapas. Nos últimos anos, diversos projetos de seqüenciamento de ESTs têm surgido para as mais diversas espécies de animais e plantas. Esses projetos têm mostrado que SSRs estão localizados também em regiões transcritas do genoma (Kantety *et al.*, 2002; Morgante *et al.*, 2002; Gao *et al.*, 2003; Eujayl *et al.*, 2004; Li *et al.*, 2004; Varshney *et al.*, 2005). Muitas dessas

seqüências são depositadas em bases de dados públicas e constituem uma nova fonte de desenvolvimento de marcadores microssatélites. Com isso, esses marcadores podem ser gerados rapidamente e quase sem custo.

Embora um grande número de marcadores microssatélites tenha sido publicado para o amendoim nos últimos anos (Hopkins *et al.*, 1999; He *et al.*, 2003; 2005; Ferguson *et al.*, 2004b; Moretzsohn *et al.*, 2004), os marcadores disponíveis são ainda insuficientes para a construção de mapas saturados para o amendoim. Além disso, o uso de marcadores desenvolvidos de diferentes formas, amostrando diferentes regiões do genoma, com diferentes modos de evolução, deverá possibilitar uma melhor cobertura do genoma, na construção de mapas genéticos.

Portanto, os objetivos do presente trabalho foram: (1) desenvolver um grande número de marcadores microssatélites para o amendoim, a partir de bibliotecas genômicas enriquecidas para dinucleotídeos TC/AG e AC/TG, de bibliotecas de ESTs e por mineração (“data mining”) de seqüências publicadas no GenBank para *Arachis* spp.; (2) caracterizar os novos marcadores em um grupo de acessos de *A. hypogaea*; e (3) analisar as relações genéticas entre acessos de *A. hypogaea*, representando as seis variedades conhecidas de amendoim.

2. MATERIAL E MÉTODOS

2.1. Material vegetal

Uma única planta de *A. hypogaea* cv. Tatu foi utilizada para a construção das duas bibliotecas genômicas. Para a caracterização dos novos marcadores e análise das relações genéticas, foram incluídos 16 acessos de *A. hypogaea* representando as duas subespécies (*hypogaea* e *fastigiata*) e as seis variedades descritas por Krapovickas & Gregory, 1994 (Tabela 1). Essas plantas foram obtidas no Banco Ativo de Germoplasma de *Arachis* (BAG-*Arachis*), mantido na Embrapa Recursos Genéticos e Biotecnologia, Brasília, DF.

2.2. Extração e quantificação de DNA

DNA genômico total foi extraído de folíolos jovens, de acordo com o método descrito por Ferreira & Grattapaglia (1998), com algumas modificações. Após a extração com clorofórmio-álcool isoamílico (24:1), a fase aquosa foi dividida em dois tubos e acrescentados 3 volumes de tampão de precipitação (CTAB 1%, Tris-HCl 50 mM, EDTA 20 mM). Após centrifugação, o sobrenadante foi descartado e o “pellet” ressuspensão em NaCl 1,2 M. Em seguida, o DNA foi precipitado com etanol 100% e lavado duas vezes com etanol 70%.

A concentração de DNA de cada tubo foi estimada por eletroforese em gel de agarose a 1,5% de concentração, comparando-se as intensidades de fluorescência de cada amostra corada com brometo de etídio com diferentes padrões de DNA Lambda.

Tabela 1 – Acessos de *Arachis hypogaea* incluídos na caracterização dos novos marcadores microsatélites. Os acessos utilizados na triagem inicial dos 271 novos marcadores são identificados com um asterisco após o nome. São apresentados também a classificação taxonômica e o local de coleta de cada acesso.

Número	Acesso	Espécie / Subespécie / Variedade (Tipo)	Local de coleta
1	cv. BR-1*	<i>A. hypogaea</i> ssp. <i>fastigiata</i> var. <i>fastigiata</i> (Valência)	Paraíba, Brasil
2	cv. IAC Tatu-ST	<i>A. hypogaea</i> ssp. <i>fastigiata</i> var. <i>fastigiata</i> (Valência)	São Paulo, Brasil
3	V 6265	<i>A. hypogaea</i> ssp. <i>fastigiata</i> var. <i>fastigiata</i> (Valência)	Bahia, Brasil
4	Mf 3207*	<i>A. hypogaea</i> ssp. <i>fastigiata</i> var. <i>vulgaris</i> (Spanish)	Desconhecido
5	V 13115	<i>A. hypogaea</i> ssp. <i>fastigiata</i> var. <i>vulgaris</i> (Spanish)	Minas Gerais, Brasil
6	Mf 1560*	<i>A. hypogaea</i> ssp. <i>fastigiata</i> var. <i>peruviana</i>	Pastaza, Ecuador
7	Sv 429	<i>A. hypogaea</i> ssp. <i>fastigiata</i> var. <i>peruviana</i>	Yurimáguas, Peru
8	Mf 1678*	<i>A. hypogaea</i> ssp. <i>fastigiata</i> var. <i>aequatoriana</i>	Sucurubios, Ecuador
9	Mf 2389	<i>A. hypogaea</i> ssp. <i>fastigiata</i> var. <i>aequatoriana</i>	Desconhecido
10	cv. IAC Runner 886	<i>A. hypogaea</i> ssp. <i>hypogaea</i> var. <i>hypogaea</i> (Virgínia)	São Paulo, Brasil
11	V 10067	<i>A. hypogaea</i> ssp. <i>hypogaea</i> var. <i>hypogaea</i> (Virgínia)	Santa Catarina, Brasil
12	V 13009*	<i>A. hypogaea</i> ssp. <i>hypogaea</i> var. <i>hypogaea</i> (Virgínia)	Tocantins, Brasil
13	Of 109	<i>A. hypogaea</i> ssp. <i>hypogaea</i> var. <i>hypogaea</i> (Xingu)	Mato Grosso, Brasil
14	Of 128	<i>A. hypogaea</i> ssp. <i>hypogaea</i> var. <i>hypogaea</i> (Xingu)	Mato Grosso, Brasil
15	Mf 3618	<i>A. hypogaea</i> ssp. <i>hypogaea</i> var. <i>hirsuta</i>	Desconhecido
16	Mf 1538*	<i>A. hypogaea</i> ssp. <i>hypogaea</i> var. <i>hirsuta</i>	Pichincha, Ecuador

2.3. Construção de bibliotecas genômicas

Duas bibliotecas de DNA genômico total, enriquecidas para dinucleotídeos TC/AG e AC/TG foram construídas, seguindo o protocolo descrito por Rafalski e colaboradores (1996), com modificações. DNA genômico total foi extraído de uma única planta de amendoim (*A. hypogaea* cv. Tatu) e alíquotas foram digeridas separadamente com três diferentes enzimas (*Mse*I, *Sau*3AI e *Tsp*509I), a fim de se identificar qual enzima produzia o perfil de digestão mais adequado para a construção de bibliotecas, ou seja, com fragmentos de DNA de diferentes tamanhos, distribuídos uniformemente no gel, sem bandas. Os fragmentos de 200 a 800 pares de bases (pb) foram transferidos para membranas de celulose (DEAE-cellulose membranes NA 45, Schleicher & Schuell BioScience, NH, EUA), por eletroforese em gel de agarose a 1,5% (“low melting agarose”). Os fragmentos foram precipitados das membranas (Sambrook *et al.*, 1989) e ligados a adaptadores *Sau* 3AI (Adaptador curto: 5'-CAGCCTAGAGCCGAATTCACC-3' e adaptador longo: 5'-pGATCGGTGAATTCGGCTCTAGGCTG-3'). Fragmentos contendo seqüências microssatélites foram selecionados por hibridização com dinucleotídeos ((TC)₁₃ e (AC)₁₃) ligados a biotina e recuperados por contas magnéticas ligadas a estreptavidina (Dynabeads[®] Streptavidin, Dynal Biotech., Noruega). A fração enriquecida foi amplificada por PCR, usando-se o adaptador curto de *Sau* 3AI como primer. Os produtos de PCR foram purificados com o sistema de purificação QIAquick PCR Purification System (Qiagen, CA, EUA), clonados em plasmídeo pGEM-T[™] (Promega, WI, EUA) e, em seguida, transformados em *E. coli*, cepa XL1-Blue (Promega) por eletroporação. As células transformadas foram multiplicadas em placas com o meio de crescimento LB-Amp (50 µg de ampicilina/ml), contendo X-Gal e IPTG a 37° C. DNA plasmidial das colônias brancas foi extraído por lise alcalina (“mini-preps”) e preparado para seqüenciamento. As “mini-preps” foram feitas de acordo com o protocolo descrito por Sambrook e colaboradores (1989), com pequenas adaptações para extração em placas de 96 poços.

2.4. Seqüenciamento de DNA e desenho de primers

As reações de seqüenciamento foram realizadas com os primers T7 e SP6 e *Big-Dye Terminator Cycle Sequencing Kit*, versão 3.1 (Applied Biosystems, CA, EUA) ou *DYEnamic ET Terminator Cycle Sequencing Kit* (Amersham Pharmacia Biotech, PA, EUA) usando-se os seqüenciadores automáticos de DNA ABI 3100 e ABI 377. Para facilitar a seleção das seqüências que continham microssatélites, o agrupamento de seqüências redundantes e os produtos seqüenciados nos dois sentidos, um módulo de programas foi desenvolvido, em parceria com a Universidade Católica de Goiás e com a Universidade Católica de Brasília, sob coordenação do Dr. Wellington Martins (ANEXO 2). Com isso, além das bibliotecas genômicas, o módulo foi utilizado para buscar (“data mining”) microssatélites em seqüências de bibliotecas de ESTs produzidas pela equipe do projeto (Guimarães *et al.*, 2005) e em seqüências depositadas no GenBank. Todas as novas seqüências identificadas no presente estudo foram submetidas à base de dados do National Center for Biotechnology Information (NCBI, EUA; GenBank, acessos números DQ099037-DQ099242).

Seqüências consensos foram geradas, no caso de grupos de seqüências similares (“contigs”), e primers complementares às regiões flaqueadoras dos microssatélites foram desenhados, utilizando o programa Primer 3 (Whitehead Institute of Biomedical Research - http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi). Para reduzir problemas de amplificação de bandas inespecíficas, alguns critérios restritivos foram utilizados: (1) temperatura de fusão (T_m – “melting temperature”) entre 57°C e 65°C; (2) diferença de T_m entre pares de primers menor que 1°C; (3) conteúdo de GC entre 40% e 60% e (4) ausência de complementaridade entre primers. Para os demais parâmetros, foram utilizados os valores pré-estabelecidos pelo programa (“default values”), exceto para o comprimento total da seqüência amplificada, que foi entre 100 e 400 pares de bases.

2.5. Condições de PCR e eletroforese dos marcadores microssatélites

As PCRs foram realizadas em volumes de 13 μ L, contendo tampão de PCR 1X (Tris-HCl 10mM pH8,3, KCl 50mM), $MgCl_2$ 2,0 mM, 0,2 mM de cada dNTP, 1 unidade de *Taq* DNA polimerase, 1 μ g/ μ L de BSA purificado (New England Biolabs, MA, EUA), 5 pmol de cada primer e 10 ng de DNA genômico. As amplificações foram realizadas em termocicladores ABI 9700 (Applied Biosystems) ou PTC-200 (MJ Research, MA, EUA), com as seguintes condições: 94°C por 5 minutos (1 ciclo), 94°C por 1 min, 48-64°C por 1 min, 72°C por 1 min (30 ciclos); e uma extensão final a 72°C por 7 minutos (1 ciclo). As temperaturas de anelamento foram otimizadas para cada par de primers, visando à obtenção de melhor resolução e maior especificidade dos produtos de PCR. Essa otimização de temperatura foi feita em géis de agarose a 3,5%. Para genotipagem, os produtos de PCR foram separados em géis de poli-acrilamida a 5% de concentração, em condições desnaturantes e corados com nitrato de prata (Creste *et al.*, 2001). Os comprimentos dos fragmentos foram estimados nos géis, por comparações com padrões de bandas de tamanhos conhecidos (10-pb DNA ladder).

2.6. Caracterização dos novos marcadores microssatélites

Os 271 novos marcadores microssatélites foram inicialmente testados em um subgrupo de seis acessos de *A. hypogaea*, sendo um de cada uma das seis variedades de amendoim descritas (Tabela 1 - acessos 1, 4, 6, 8, 12 e 16). Os marcadores desenvolvidos a partir de bibliotecas genômicas foram denominados de TC (para bibliotecas de *A. hypogaea* enriquecidas para repetições TC) ou AC (para repetições AC), seguido do número da placa com os clones em meio de cultura (TC1 a TC11 ou AC1 a AC3) e a identificação do clone (A01 a H12, de acordo com a localização na placa de 96 poços). Marcadores desenvolvidos a partir de bibliotecas de cDNA foram denominados de ML, para folha madura ("mature leaf"), RN para raiz não inoculada, RI para raiz inoculada com o nematóide *Ditylenchus africanus* ou RM para raiz inoculada com o nematóide *Meloidogyne*. Após as duas letras iniciais foram incluídos o número da placa (de 1 a 22) e a identificação do clone (A01 a H12). Marcadores obtidos por mineração de seqüências no GenBank foram chamados pelo nome com o qual a seqüência foi submetida. Os nomes desses

marcadores começam com gi-, pPGP ou pPGS. Uma nota subjetiva foi dada para cada marcador, referente à qualidade das bandas no gel. Assim, nota 3 foi dada aos marcadores que resultaram em bandas com resoluções muito boas; nota 2, para resoluções medianas e nota 1, para géis de qualidade inferior, mas passíveis de análise. O número de alelos por loco, a amplitude de comprimento dos fragmentos e a diversidade gênica (GD) foram estimados para os marcadores polimórficos usando um grupo de 16 acessos de *A. hypogaea* (Tabela 1). A diversidade gênica foi calculada de acordo com a fórmula: $GD = 1 - (\sum p_i^2)$, sendo p_i a frequência do alelo i nos 16 acessos (Nei, 1973). Para caracterização das seqüências usadas no desenvolvimento dos marcadores, bases de dados BLAST foram formatadas para todas as possíveis proteínas codificadas pelos genomas de *Arabidopsis* e de todas as leguminosas com seqüências depositadas no GenBank. Buscas por similaridades foram realizadas usando BlastX (Altschul *et al.*, 1990) e as concordâncias foram consideradas significativas para valores de E (E-values) inferiores a 1×10^{-8} .

As distâncias genéticas entre acessos individuais foram calculadas a partir da planilha com os genótipos (comprimento dos alelos em pares de bases) de cada acesso, pelo método dos alelos comuns ou “shared allele distance” (Bowcock *et al.*, 1994). Essas estimativas foram obtidas pela fórmula $\{1 - (\text{alelos comuns} / \text{total de alelos})\}$, usando o programa “Genetic Distance Calculator”, disponível em <http://www2.biology.ualberta.ca/jbrzusto/sharedst.php>. Distâncias baseadas em alelos comuns tendem a dar uma representação mais exata das relações genéticas, no caso de indivíduos geneticamente próximos e de marcadores SSR, quando comparadas a outras estimativas de distância (Takezaki & Nei, 1996; Bertin *et al.*, 2001). A matriz de distâncias resultante foi submetida à análise de agrupamento usando UPGMA (“unweighted pair-group method analysis”) e o programa NTSYS, versão 2.1 (Rohlf, 2000).

3. RESULTADOS E DISCUSSÃO

3.1. Desenvolvimento de marcadores microssatélites

Digestões do DNA genômico de *A. hypogaea* com três diferentes enzimas (*Mse*I, *Sau*3AI e *Tsp*509I) mostraram que *Sau*3AI produziu o perfil mais adequado para a construção da biblioteca, por apresentar fragmentos de 200 a 800 pb uniformemente distribuídos no gel, sem bandas. *Sau*3AI é sensível a metilação na base citosina e, portanto, corta preferencialmente em regiões ricas em genes, que, em geral, não são metiladas. Com isso, as bibliotecas produzidas são enriquecidas para seqüências de baixo número de cópias, que são geralmente ricas em genes. Duas bibliotecas foram construídas, com base no tipo de dinucleotídeo usado (TC e AC). Essas repetições foram escolhidas por estarem entre os microssatélites mais abundantes em plantas (Wang *et al.*, 1994; Gupta & Varshney, 2000). De um total de 1152 clones positivos (colônias brancas) seqüenciados, 224 (19,4%) continham SSRs, sendo 210 não redundantes. Alguns desses clones foram seqüenciados também na fita oposta, para possibilitar a identificação de grupos de seqüências homólogas (“contigs”) com alta qualidade. As seqüências dos 210 clones positivos únicos foram submetidas ao programa Primer3, sendo obtidos 121 (10,5% dos clones seqüenciados) novos pares de primers (ANEXO 3). O desenho de primers não foi possível para as demais 89 seqüências únicas identificadas, devido à ocorrência de pequenos números de repetições (menos de 5 unidades), baixo conteúdo de GC nas regiões flangeadoras ao SSR ou à proximidade do SSR ao final do inserto de DNA clonado. A porcentagem de primers desenhados, em relação ao número de clones seqüenciados (10,5%), é significativamente menor do que a proporção obtida (41,4%) para o amendoim em um estudo anterior que incluiu uma seleção adicional de clones positivos por PCR-ancorado (Moretzsohn *et al.*, 2004). Esse resultado mostrou que essa seleção adicional é eficiente por aumentar significativamente o rendimento de clones úteis seqüenciados, portanto, reduzindo os custos. No entanto, essa etapa é lenta e sua omissão acelerou sobremaneira o processo de desenvolvimento dos marcadores microssatélites.

Uma busca por seqüências microssatélites foi realizada em um total de 2740 seqüências de ESTs. Dessas, 98 (3,6%) apresentaram repetições de di, tri, tetra ou pentanucleotídeos e 81 (3,0%) possibilitaram o desenho de primers (ANEXO 3).

Esses números são similares às porcentagens obtidas para outras espécies de plantas, que têm variado de 1,7% em trigo a 7,5% em cevada: (Cardle *et al.*, 2000; Cordeiro *et al.*, 2001; Varshney *et al.*, 2002; Thiel *et al.*, 2003; Eujayl *et al.*, 2004; Gao *et al.*, 2004; Khlestkina *et al.*, 2004; Jung *et al.*, 2005).

Finalmente, seqüências disponíveis no GenBank, para as quais não se tinham desenhado primers, foram analisadas quanto à presença de repetições (>5 unidades de repetição). Um total de 69 novos pares de primers microssatélites foram desenhados (ANEXO 3), sendo 38 de seqüências genômicas, 25 de ESTs e seis de genes de *A. hypogaea*, tais como proteína de resposta a frio, desaturase e lectina.

A presença e a localização de microssatélites em todas as seqüências foi feita com o módulo Staden-TROLL. Isso permitiu que o processo fosse feito de maneira rápida, com parâmetros facilmente ajustáveis quanto ao tipo e comprimento dos microssatélites de interesse, além de considerar a qualidade da região seqüenciada para desenho dos primers. A partir de um grande número de seqüências, na forma de cromatogramas ou texto, uma base de dados foi construída, contendo todas e somente as seqüências que continham microssatélites, incluindo aquelas para as quais primers já tinham sido publicados. Centenas de “contigs” foram estabelecidas, eliminando as seqüências redundantes. Dessa forma, foi possível assegurar que todos os 271 marcadores recentemente desenvolvidos eram únicos. Além disso, as análises de BlastN dessas seqüências contra bases de dados de todas seqüências publicadas de *Arachis* que continham microssatélites forneceram evidências de que todos os marcadores eram novos.

3.2. Características das regiões repetitivas em *Arachis*

Dentre as 271 seqüências para as quais primers foram desenhados, 159 (58,7%) apresentaram repetições perfeitas, 82 (30,2%), repetições compostas e 30 (11,1%), repetições imperfeitas. Seqüências obtidas das bibliotecas enriquecidas para TC/AG e AC/TG foram constituídas, como esperado, quase inteiramente de repetições TC/AG e AC/TG, respectivamente. Para as bibliotecas de ESTs, 153 repetições foram encontradas nas 106 seqüências incluídas, uma vez que mais de um microsatélite foi encontrado em algumas seqüências. Repetições de trinucleotídeos foram o tipo de repetição mais abundante, com 112 (73,2%) microsatélites, seguido de di (22,2%), tetra (3,3%) e pentanucleotídeos (1,3%). Essas proporções têm variado bastante nos poucos trabalhos sobre microsatélites derivados de ESTs publicados até o momento. Para trinucleotídeos, as porcentagens têm variado de 30,2% em milho a 90,3% em cana-de-açúcar; para dinucleotídeos, de 9,7% em cana-de-açúcar a 39,0% em soja; para tetranucleotídeos, de 2,2% em soja e *Arabidopsis* a 10,3% em *Medicago*; e de 0,6% em cevada a 9,0% em *Medicago* no caso de pentanucleotídeos (Cordeiro *et al.*, 2001; Varshney *et al.*, 2002; Gao *et al.*, 2003; Thiel *et al.*, 2003; Eujayl *et al.*, 2004; Tian *et al.*, 2004; Gonzalo *et al.*, 2005; Chen *et al.*, 2006; Han *et al.*, 2006). No entanto, a maioria desses estudos tem mostrado que SSRs triméricos são mais abundantes em ESTs de plantas do que os outros tipos de SSRs. Esses resultados não são inesperados e também têm sido relatados para outros organismos eucariotos superiores, incluindo *Drosophila*, *Caenorhabditis elegans*, mamíferos, humanos e leveduras (revisto por Li *et al.*, 2004). Isso tem sido atribuído à seleção negativa contra mutações que alterem a fase de leitura em regiões codificadoras (Metzgar *et al.*, 2000) ou à seleção positiva para manutenção de cadeias específicas de aminoácidos (Morgante *et al.*, 2002).

Repetições AAG/TTC (27/153), AAT/TTA (21/153) e AAC/TTG (14/153) foram os trinucleotídeos mais comuns em ESTs de *A. stenosperma*. Consideráveis diferenças nas freqüências de tipos de repetições têm também sido detectadas em plantas, mas essas três classes estão geralmente entre os trinucleotídeos mais comuns em outras espécies de dicotiledôneas (Cardle *et al.*, 2000; Gao *et al.*, 2003; Eujayl *et al.*, 2004; Tian *et al.*, 2004; Han *et al.*, 2006). Para muitas monocotiledôneas, CCG/GGC têm-se mostrado as repetições mais comuns

(Cordeiro *et al.*, 2001; Morgante *et al.*, 2002; Varshney *et al.*, 2002; Gao *et al.*, 2003; La Rota *et al.*, 2005), mas foram detectadas em apenas duas seqüências no presente estudo. Os motivos AAT/TTA são os mais raros (<1%) em animais e em monocotiledôneas (Li *et al.*, 2004), mas estão entre os mais freqüentes em ESTs de *Arachis* e foram detectados com altas freqüências também em soja e *Medicago* (Gao *et al.* 2003; Eujayl *et al.* 2004; Tian *et al.* 2004).

Repetições AG/TC e AT/TA (16 de 153, cada) apresentaram as maiores freqüências entre os dinucleotídeos, seguidos de AC/TG (6/153). Os motivos CG/GC não foram detectados em nenhuma seqüência. Resultados similares têm sido descritos para diversas espécies de plantas, tendo AG/TC apresentado as freqüências mais altas e CG/GC, as mais baixas (Kantety *et al.*, 2002; Eujayl *et al.*, 2004; Tian *et al.*, 2004). Embora as freqüências de AT/TA sejam geralmente menores do que as encontradas no presente estudo, altas freqüências têm também sido detectadas em outras espécies de leguminosas, como *Medicago* e soja (Morgante *et al.*, 2002; Gao *et al.*, 2003; Eujayl *et al.*, 2004). As possíveis funções de microssatélites em regiões codificadoras ainda não são claras, bem como as razões pelas quais os tipos de repetições são tão variáveis em diferentes espécies de plantas. No entanto, esses resultados mostram que a predominância de um tipo específico de microssatélite em uma espécie não é um indicativo da representação desse microssatélite em outra espécie de planta.

3.3. Polimorfismo e transferibilidade dos novos marcadores microssatélites

Os 271 novos marcadores microssatélites (ANEXO 3) foram avaliados para detecção de polimorfismo em seis amostras de *A. hypogaea*, sendo uma de cada uma das seis variedades descritas (Tabela 1), um acesso de *A. stenosperma* (V 10309) e um de *A. duranensis* (K 7988). Como esperado, o polimorfismo entre os dois acessos de espécies diferentes foi maior do que o encontrado para as seis amostras de amendoim (Figura 1 e Tabela 2).

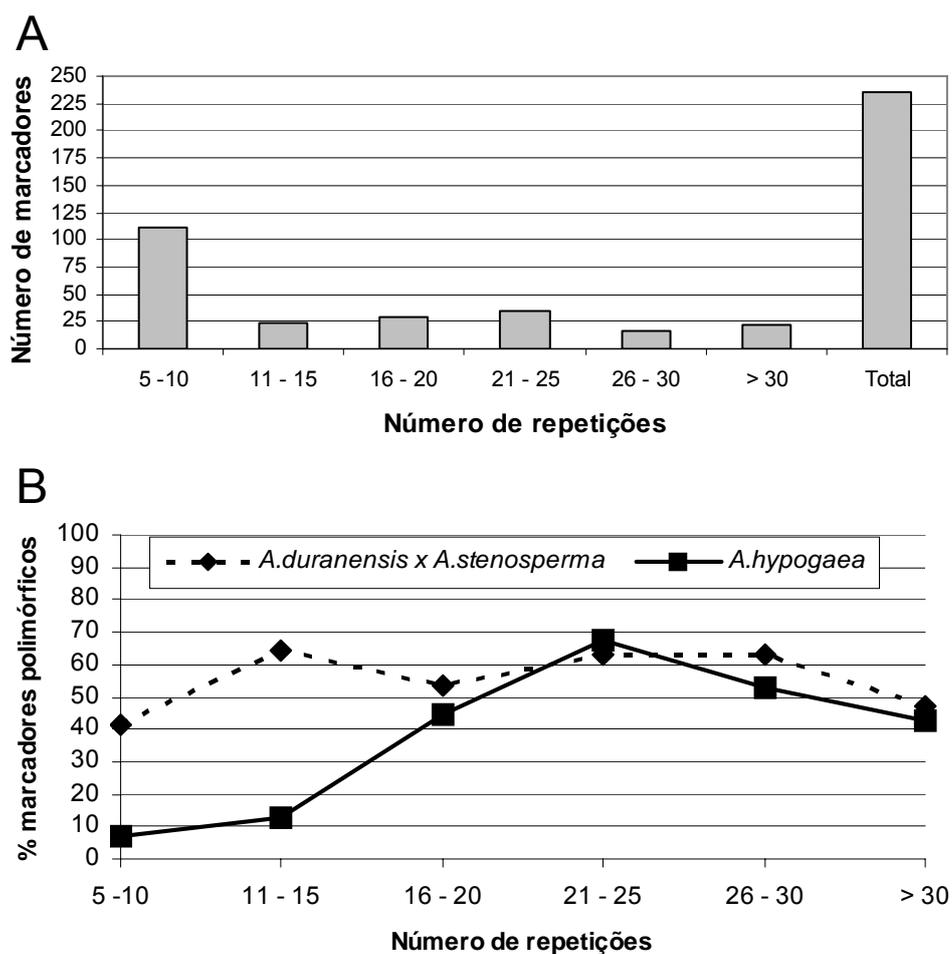


Figura 1 – Número de marcadores detectados por classe de tamanho de repetições (gráfico A) e relações entre número de repetições dos microssatélites e o polimorfismo detectado nas análises de *A. duranensis* e *A. stenosperma* e em seis acessos de *A. hypogaea* (gráfico B). Nessa análise, foram considerados 225 marcadores que amplificaram produtos com boas resoluções em *A. duranensis* e *A. stenosperma* e 236 marcadores na análise de *A. hypogaea*.

Para as espécies silvestres, 27 (10,0%) pares de primers não amplificaram nenhum fragmento passível de análise, 11 (4,1%) amplificaram fragmentos apenas em *A. stenosperma* e 4 (1,5%) apenas em *A. duranensis*. Portanto, 229 (84,5%) pares de primers foram totalmente transferíveis entre as duas espécies. Desses 229 marcadores, 113 (49,3%) foram polimórficos.

Tabela 2 – Número de marcadores polimórficos detectados na análise de dois acessos de espécies silvestres de *Arachis* - *A. duranensis* (K 7988) e *A. stenosperma* (V 10309) - e de seis acessos de *A. hypogaea*, incluindo as seis variedades botânicas. Os seis marcadores desenvolvidos a partir de genes de *A. hypogaea*, depositados no GenBank, não foram incluídos.

Tipo de repetição	Bibliotecas genômicas				Bibliotecas de cDNA
	TC/AG	AC/TG	Diversos	TOTAL	Diversos
K7988 x V10309					
Marcadores polimórficos					
Codominantes	48 (48,5%)	5 (22,7%)	13 (34,2%)	66 (41,5%)	44 (41,5%)
Dominantes	13 (13,1%)	-	-	13 (8,2%)	2 (1,9%)
Marcadores monomórficos	32 (32,3%)	15 (68,2%)	17 (44,7%)	64 (40,3%)	49 (46,2%)
Sem amplificação	6 (6,1%)	2 (9,1%)	8 (21,1%)	16 (10,0%)	11 (10,4%)
TOTAL	99	22	38	159	106
<i>A. hypogaea</i>					
Marcadores polimórficos	48 (48,5%)	5 (22,7%)	4 (10,5%)	57 (35,8%)	8 (7,5%)
Marcadores monomórficos	42 (42,4%)	13 (59,1%)	27 (71,1%)	82 (51,6%)	82 (77,4%)
Sem amplificação	9 (9,1%)	4 (18,2%)	7 (18,4%)	20 (12,6%)	16 (15,1%)
TOTAL	99	22	38	159	106

Para *A. hypogaea*, 37 pares de primers (13,7%) não amplificaram fragmentos com boa qualidade. Para os demais 234 pares de primers, 66 (28,2%) foram polimórficos para pelo menos uma das seis amostras. Desses, 62 foram utilizados para análise da diversidade alélica, usando 16 acessos de *A. hypogaea* (Tabela 3). Os demais quatro marcadores polimórficos (AC2C02, RN0-602, Seq14G03, e TC1E06) não foram incluídos nessa análise, por apresentarem resoluções de bandas insatisfatórias. O número de alelos detectado variou de 2 a 12 para cada um dos 62 locos polimórficos analisados, com uma média de 5,87 alelos por loco. Os valores da diversidade gênica (GD) variaram de 0,121 (para os locos TC11B11 e gi-385) a 0,910 (para o loco TC2D06), com uma média de 0,690. GD é uma estimativa da probabilidade de que dois genes escolhidos ao acaso em uma população sejam diferentes (Nei, 1987). Os valores obtidos no presente estudo estão provavelmente superestimados, uma vez que alguns dos marcadores podem ter amplificado alelos de locos duplicados ou de genomas diferentes. Dessa forma, esses dados não podem ser comparados com os dados obtidos para espécies diplóides, mas o valor médio obtido (0,690) é similar aos valores detectados com marcadores

microsatélites para diversas espécies poliplóides, incluindo o amendoim (Provan *et al.*, 1996; Chavarriga-Aguirre *et al.*, 1998; Prasad *et al.*, 2000; He *et al.*, 2003; Ferguson *et al.*, 2004b; Moretzsohn *et al.*, 2004). Por outro lado, valores médios de GD, estimados por marcadores RFLP, foram consideravelmente menores para cinco espécies de *Arachis*, incluindo *A. hypogaea* (Paik-Ro *et al.*, 1992). Esses resultados demonstram a grande utilidade dos marcadores microsatélites para análise da variabilidade genética do amendoim cultivado.

Tabela 3 – Pares de primers, nota referente à qualidade dos produtos amplificados, amplitude de comprimento dos fragmentos, número de indivíduos analisados (n), número total de alelos amplificados (A) e diversidade gênica (GD) para os 62 novos marcadores microsatélites polimórficos para *A. hypogaea*, identificados no presente estudo.

Loco	Nota	Amplitude de comprimento	n	A	GD
AC2B3	3	296-308	14	6	0,720
AC2B5	3	200-212	15	5	0,680
AC2C8	2	146-220	15	7	0,828
AC2C12	3	204-212	16	5	0,685
TC1A1	2	202-222	13	6	0,735
TC1A2	3	240-276	16	11	0,815
TC1D2	3	242-278	16	4	0,458
TC1E1	3	154-248	14	11	0,892
TC2A2	3	194-212	15	7	0,800
TC2B1	2	190-215	11	4	0,610
TC2B9	3	190-200	16	5	0,645
TC2C7	2	182-210	16	3	0,492
TC2D6	3	196-224	15	10	0,910
TC2E5	3	102-122	15	8	0,864
TC3B5	3	248-270	12	4	0,710
TC3E2	3	270-310	16	9	0,708
TC3E5	3	358-370	16	5	0,798
TC3G1	2	250-256	15	4	0,745
TC3G5	3	118-130	14	2	0,423
TC3H2	3	280-300	15	4	0,671
TC3H7	2	246-270	16	9	0,863
TC4A2	3	280-284	16	3	0,411
TC4D2	2	202-240	12	5	0,786
TC4D9	2	220-240	14	2	0,389
TC4E10	3	300-344	16	8	0,841
TC4F12	3	220-232	15	6	0,763
TC4G2	2	130-166	11	6	0,784
TC4G10	2	205-220	15	3	0,543

Loco	Nota	Amplitude de comprimento	n	A	GD
TC4H2	2	184-192	12	4	0,725
TC4H7	2	166-220	13	8	0,742
TC5A6	2	164-200	14	7	0,825
TC5D6	2	174-184	13	5	0,782
TC6E1	3	154-186	15	9	0,802
TC6G9	3	132-146	12	6	0,746
TC6H3	3	210-228	14	8	0,754
TC7A2	2	308-320	15	6	0,706
TC7C6	3	148-176	16	9	0,827
TC7D3	3	142-148	15	4	0,579
TC7E4	3	290-300	16	4	0,694
TC7G10	3	110-142	15	4	0,451
TC7H11	2	340-360	14	7	0,847
TC9B8	3	98-108	15	8	0,777
TC9C12	3	256-300	15	7	0,855
TC9F4	3	122-142	15	6	0,800
TC9F10	3	286-320	15	10	0,871
TC11A2	3	284-292	14	6	0,825
TC11A4	3	172-204	15	12	0,906
TC11B11	2	144-146	16	2	0,121
TC11E4	2	110-136	16	12	0,909
TC11F12	2	182-188	14	2	0,254
TC11H6	3	190-214	16	8	0,847
gi-385	3	330-340	16	2	0,121
gi-832	3	200-210	15	2	0,480
gi-1107	3	360-384	14	5	0,680
RN0-615	2	390-400	15	3	0,625
RN0-681	2	310-350	16	6	0,520
RI1F06	3	312-372	11	8	0,823
RN3E10	3	252-255	11	2	0,416
RN17F12	2	260-278	12	5	0,812
Seq4E8	3	274-286	10	6	0,842
Seq4F10	2	180-188	9	4	0,706
Seq16C7	3	242-258	14	5	0,571
Média		-	14,41	5,87	0,690

O número de unidades de repetição encontrado nas 271 seqüências variou de 5 a 54 (ANEXO 3). Alguns estudos têm sugerido que locos com repetições maiores tendem a ser mais polimórficos (Weber, 1990; Thomas & Scott, 1993; Udupa *et al.*, 1999; Burstin *et al.*, 2001). Para os seis acessos de *A. hypogaea* houve uma certa tendência nesse sentido, com máximo polimorfismo sendo detectado por

microssatélites com 21-25 unidades de repetição (Figura 1). O polimorfismo detectado por microssatélites com 5-10 repetições foi bastante baixo (7,2%). Entre as duas espécies silvestres, o polimorfismo variou menos com o tamanho das repetições e 40% dos marcadores SSR com apenas 5-10 repetições mostraram-se polimórficos. Marcadores com repetições maiores detectaram níveis similares de polimorfismo entre as duas espécies silvestres e nos acessos de *A. hypogaea*. Os números de repetições foram obtidos somente para *A. hypogaea*, no caso das bibliotecas genômicas e para *A. stenosperma*, nas bibliotecas de cDNA. Embora fragmentos de tamanhos similares tenham sido obtidos nas três espécies analisadas, variações consideráveis no número de repetições têm sido descritas mesmo entre espécies aparentadas (Peakall *et al.*, 1998; Liewlaksaneeyanawin *et al.*, 2004; Guo *et al.*, 2006). Os resultados obtidos são importantes e poderão direcionar futuros esforços de desenvolvimento de marcadores microssatélites em *Arachis*. No entanto, o seqüenciamento de alguns fragmentos nas três espécies analisadas é necessário para confirmação desses resultados.

Tanto para as espécies silvestres como para o amendoim cultivado, repetições AG/TC mostraram-se bastante mais polimórficas do que as repetições AC/TG (Tabela 2). Efeitos semelhantes do tipo de repetição nos níveis de polimorfismo detectados para o amendoim foram também observados por Ferguson e colaboradores (2004b).

Marcadores SSR foram desenvolvidos, no presente estudo, usando clones provenientes de bibliotecas genômicas enriquecidas e de ESTs (159 e 106 marcadores, respectivamente). Por isso, foi possível comparar os níveis de polimorfismo detectados por marcadores obtidos dessas duas fontes (Tabela 2). Em geral, o número de unidades de repetição encontrado em seqüências de ESTs foi muito menor do que o de seqüências genômicas e, correspondentemente, o polimorfismo de ESTs foi muito menor em *A. hypogaea* do que entre as duas espécies silvestres (7,5% e 35,8%, respectivamente) (Tabela 2). Alguns trabalhos têm demonstrado que marcadores SSR derivados de ESTs são menos polimórficos do que os derivados de bibliotecas genômicas, para outras espécies de plantas, tais como: arroz (Cho *et al.*, 2000), cana-de-açúcar (Cordeiro *et al.*, 2001), trigo (Eujayil *et al.*, 2002; Gupta *et al.*, 2003), cevada (Thiel *et al.*, 2003; Chabane *et al.*, 2005) e soja (Song *et al.*, 2004). Isso tem sido atribuído à possível seleção contra alterações nessas seqüências conservadas (Scott *et al.*, 2000). No entanto, para os acessos de

A. duranensis e *A. stenosperma*, as porcentagens foram idênticas: 41,5% para ESTs e para bibliotecas genômicas, considerando apenas os marcadores codominantes. As razões para essas diferenças não estão claras e necessitam de mais estudos para serem elucidadas.

3.4. Caracterização do conteúdo gênico das bibliotecas usadas para desenvolvimento dos marcadores

Arachis hypogaea possui um genoma bastante extenso ($2,8 \times 10^9$ pb ou 2,7 pg / 1C, Bennet & Leitch, 2005) e pouco caracterizado. O mapeamento genético comparativo com outras espécies de leguminosas permitirá, portanto, que as pesquisas com o amendoim possam se beneficiar do conhecimento adquirido para outros genomas, especialmente das espécies-modelo *Lotus japonicus* e *Medicago truncatula*. Para tais comparações, o posicionamento de genes em um mapa de *Arachis* é essencial. Os marcadores ESTs gerados no presente trabalho foram derivados de seqüências transcritas que são quase inteiramente gênicas. No entanto, nem todas as similaridades desses marcadores são facilmente identificáveis e de um total de 106 marcadores, apenas 67 (63,2%) deram similaridades significativas usando BlastX.

O conteúdo gênico de bibliotecas genômicas depende da maneira como elas foram produzidas. Os conteúdos gênicos de diferentes bibliotecas usadas para desenvolvimento de marcadores SSR, incluindo as desenvolvidas por outros autores, foram estimados usando BlastX (Tabela 4). Nesta tabela, “sítios de metilação clivados / fragmento clonado” representa o número médio de sítios de metilação (CG e CNG, em que metilação na citosina protege contra restrição) clivados por fragmento clonado. Por exemplo, *Pst*I cliva CTGCAG e, dessa forma, cliva dois sítios CNG. Metilação em qualquer das duas citosinas protege contra restrição e, uma vez que as duas extremidades dos fragmentos clonados são cortadas por *Pst*I, quatro sítios de metilação CNG são cortados por fragmento clonado. Para *Sau*3AI que cliva GATC, a restrição dos sítios de metilação depende da seqüência flanqueadora. Portanto, em média, um sítio CG e um sítio CNG serão clivados por fragmento digerido com *Sau*3AI. Quando usada em conjunto com *Bam*HI, que não corta nenhum sítio CG ou CNG, em média, 0,5 sítio CG e 0,5 sítio CNG será clivado por fragmento clonado. Dados de Ferguson e colaboradores

(2004b) incluem todas as seqüências com SSRs disponíveis como dados suplementares e não apenas as utilizadas para desenvolvimento dos marcadores. Os dados de Moretzsohn e colaboradores (2004) também incluem algumas seqüências que não foram utilizadas para desenvolvimento de marcadores. A proporção de seqüências com similaridades significativas (valor de $E < 10^{-8}$) variou de mais de 20% a 0%. Houve uma nítida tendência, com bibliotecas desenvolvidas com o uso de enzimas sensíveis a metilação em sítios CG ou CNG sendo mais ricas em genes. A única exceção foi o relativamente alto conteúdo gênico das seqüências publicadas por Moretzsohn e colaboradores (2004), o que pode ser explicado pelo fato de que essas bibliotecas foram enriquecidas exclusivamente para trinucleotídeos. As similaridades encontradas por BlastX contra o genoma de *Arabidopsis* e proteínas de leguminosas para todos os marcadores SSR são apresentadas no ANEXO 3.

Tabela 4 – Conteúdo gênico de diferentes bibliotecas estimado por BlastX (valores de $E < 10^{-8}$) contra *Arabidopsis* e proteínas de leguminosas. Os dados de ESTs são apresentados apenas para comparação com as bibliotecas genômicas.

Biblioteca	Enzima(s)	Sítios de metilação clivados / fragmento clonado CG:CNG	Seleção / enriquecimento	Similaridades / total de seqüências (%)
EST	NA	NA	di, tri, tetra, e penta SSRs	67/106 (63,2%)
pPGP (Ferguson <i>et al.</i> , 2004b)	<i>Pst</i> I	0:4	GT GA AAC ATC AGT ATT CAC CTT CTG ATCC GATA AAAT	31/164 (20,7%)
TC	<i>Sau</i> 3AI	1:1	TC	17/102 (16,7%)
Genômica (Moretzsohn <i>et al.</i> , 2004)	<i>Tsp</i> 509I	0:0	TTG	8/87 (9,2%)
AC	<i>Sau</i> 3AI	1:1	AC	2/22 (9,1%)
pPGS (Ferguson <i>et al.</i> , 2004b)	<i>Sau</i> 3AI/ <i>Bam</i> HI	0,5:0,5	GT GA AAC ATC AGT ATT CAC CTT CTG	7/288 (2,4%)
Genômica (He <i>et al.</i> , 2003)	<i>Hind</i> III/ <i>Mse</i> I	0:0	AT GT GA	0/56 (0,0%)

3.5. Relações genéticas entre acessos de *A. hypogaea*

As distâncias genéticas entre os 16 acessos de *A. hypogaea* foram estimadas pelo método de alelos comuns, usando 62 marcadores microssatélites. Uma distância média de 0,72 foi obtida e todos os acessos puderam ser diferenciados, ou seja, nenhum par de acessos apresentou todos os alelos em comum (Tabela 5). Um dendrograma, baseado em UPGMA, foi construído para os 16 acessos (Figura 2). Três grupos principais foram formados, um contendo os acessos da subespécie *fastigiata* variedade *fastigiata* (FF) e de *fastigiata/vulgaris* (FV); outro contendo os acessos de *hypogaea/hypogaea* (HH), *hypogaea/hirsuta* (HHi) e *hypogaea*, tipo Xingu e um terceiro grupo, com os acessos de *fastigiata/aequatoriana* (FA) e *fastigiata/peruviana* (FP). Acessos das quatro variedades de *fastigiata* formaram subgrupos separados, enquanto os da subsp. *hypogaea* agruparam entre si e com os dois acessos do tipo Xingu, que também parecem pertencer a essa subespécie. Esse agrupamento de acessos, de acordo com as variedades, sugere que pouca introgressão entre variedades tem ocorrido e mostra que a classificação dos diferentes *taxa*, com base em características morfológicas é corroborada pelos dados moleculares.

Apesar disso, acessos das variedades *peruviana* e *aequatoriana* mostraram-se mais similares a acessos da subsp. *hypogaea* do que com os da subsp. *fastigiata*, à qual pertencem atualmente (Krapovickas & Gregory, 1994). Resultados similares foram obtidos em estudos recentes, baseados em marcadores microssatélites (He *et al.*, 2005) e AFLP (He & Prakash, 2001; Tallury *et al.*, 2005). Além disso, estudos baseados em 12 locos SSR (Ferguson *et al.*, 2004a) e em 94 bandas RAPD e 67 bandas ISSR (Raina *et al.*, 2001) mostraram que acessos da variedade *peruviana* formavam um grupo separado dos demais. Os autores sugeriram, portanto, que essa variedade fosse classificada como uma terceira subespécie. No presente trabalho, foram incluídos 62 locos microssatélites e 364 alelos, o que confere bastante consistência aos resultados obtidos e levanta dúvidas quanto à classificação taxonômica atual. A análise de um grande número de acessos de cada variedade, provenientes de diferentes localidades, encontra-se em andamento na Embrapa Recursos Genéticos e Biotecnologia e deverá elucidar essa questão.

Tabela 5 – Matriz de distâncias genéticas entre 16 acessos de *A. hypogaea*, estimadas pelo método de alelos comuns (“shared allele distance”). As letras após cada acesso correspondem às subespécies, variedades e tipos: FF-*fastigiata/fastigiata*; FV-*fastigiata/vulgaris*; FA-*fastigiata/aequatoriana*; FP-*fastigiata/peruviana*; HH-*hypogaea/hypogaea*; HHi-*hypogaea/hirsuta*; Xingu - *hypogaea* tipo Xingu.

	Tatu FF	Runner HH	Of128 Xingu	Of109 Xingu	Mf1678 FA	Mf1560 FP	V13115 FV	V6265 FF	Mf1538 HHi	V13009 HH	V10067 HH	Mf2389 FA	Mf3207 FV	Mf3618 HHi	BR1 FF	Sv429 FP
Tatu-FF	0,000															
Runner-HH	0,709	0,000														
Of128-Xingu	0,727	0,627	0,000													
Of109-Xingu	0,736	0,645	0,572	0,000												
Mf1678-FA	0,727	0,745	0,700	0,790	0,000											
Mf1560-FP	0,718	0,663	0,700	0,781	0,627	0,000										
V13115-FV	0,718	0,700	0,772	0,772	0,790	0,745	0,000									
V6265-FF	0,590	0,700	0,809	0,854	0,854	0,800	0,618	0,000								
Mf1538-HHi	0,763	0,727	0,781	0,809	0,809	0,745	0,754	0,718	0,000							
V13009-HH	0,709	0,718	0,618	0,754	0,754	0,700	0,709	0,690	0,645	0,000						
V10067-HH	0,690	0,681	0,600	0,709	0,772	0,69	0,718	0,681	0,636	0,109	0,000					
Mf2389-FA	0,763	0,781	0,672	0,800	0,545	0,727	0,818	0,772	0,772	0,772	0,736	0,000				
Mf3207-FV	0,718	0,800	0,836	0,881	0,790	0,827	0,581	0,636	0,763	0,800	0,809	0,745	0,000			
Mf3618-HHi	0,745	0,681	0,636	0,700	0,772	0,736	0,790	0,790	0,736	0,736	0,736	0,745	0,745	0,000		
BR1-FF	0,500	0,700	0,772	0,845	0,718	0,745	0,745	0,609	0,772	0,745	0,745	0,736	0,590	0,700	0,000	
Sv429-FP	0,754	0,709	0,754	0,845	0,754	0,509	0,800	0,763	0,754	0,700	0,718	0,763	0,672	0,681	0,663	0,000

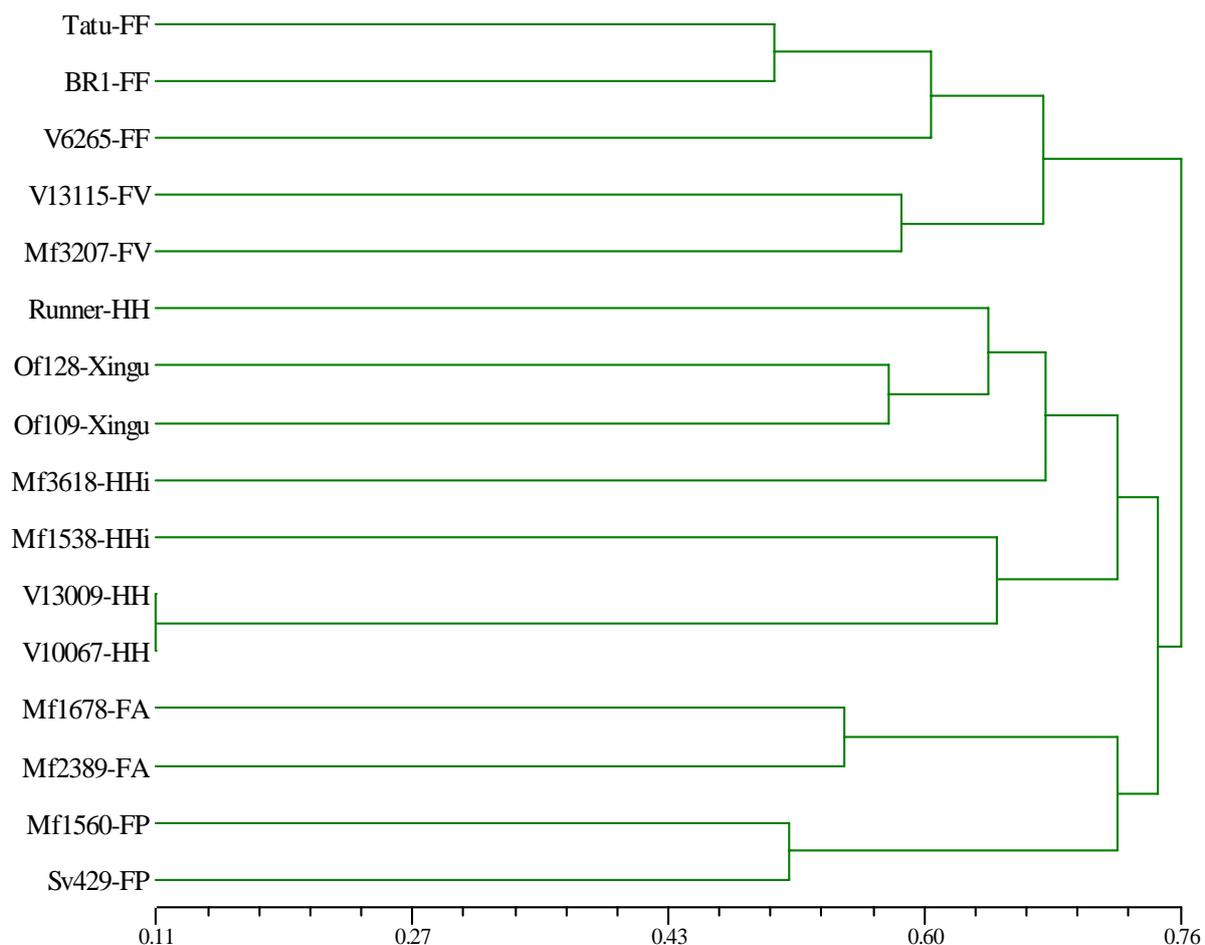


Figura 2 - Dendrograma, obtido pelo método UPGMA, baseado nas distâncias de alelos comuns (“shared allele distance”), entre 16 acessos de *A. hypogaea* (Tabela 1). As letras após cada acesso correspondem às subespécies e variedades: FF-*fastigiata/fastigiata*; FV-*fastigiata/vulgaris*; FA-*fastigiata/aequatoriana*; FP-*fastigiata/peruviana*; HH-*hypogaea/hypogaea*; HHi-*hypogaea/hirsuta*. Xingu corresponde ao tipo Xingu, que parece pertencer à subespécie *hypogaea*.

4. CONCLUSÕES

Marcadores microssatélites foram desenvolvidos a partir de bibliotecas genômicas, de ESTs e por busca em seqüências publicadas no GenBank. Marcadores derivados de bibliotecas de cDNA mostraram-se menos polimórficos para *A. hypogaea* do que os derivados de bibliotecas genômicas, embora, entre *A. duranensis* e *A. stenosperma* porcentagens idênticas de polimorfismo tenham sido detectados por marcadores derivados das duas bibliotecas.

Os resultados mostraram ainda que novas iniciativas de desenvolvimento de marcadores microssatélites para *A. hypogaea* devem ser baseadas em bibliotecas genômicas enriquecidas, preferencialmente de dinucleotídeos, em especial, de TC/AG, com números de repetições superiores a 15. Isso aumenta a probabilidade de identificação de marcadores polimórficos.

Apesar da estreita base genética do amendoim cultivado, o presente estudo mostrou a utilidade dos marcadores microssatélites na detecção de polimorfismo em acessos de *A. hypogaea*. Os marcadores polimórficos identificados abrem novas perspectivas para estudos em amendoim, tais como uma adequada análise da variabilidade genética. A análise preliminar realizada nesse estudo mostrou a formação de três grupos de similaridade e levantou dúvidas quanto à classificação taxonômica das variedades *peruviana* e *aequatoriana*. Além disso, uma vez que um grande número de marcadores foi desenvolvido, torna-se possível a construção de mapas genéticos em *Arachis*, baseados em marcadores microssatélites, o que significa um grande avanço para a futura implementação de seleção assistida por marcadores nos programas de melhoramento genético do amendoim.

Capítulo 2

Construção de um mapa de ligação e mapeamento de QTLs ligados à produtividade e à resistência à mancha preta em espécies diplóides de *Arachis* com genoma AA*

*Este capítulo é a tradução da segunda parte do artigo publicado na *Theoretical and Applied Genetics* (Moretzsohn *et al.*, 2005 – ANEXO 1), com mais marcadores mapeados e acrescida do mapeamento de QTLs.

1. INTRODUÇÃO

O amendoim (*A. hypogaea*) é um alotetraplóide dissômico, com genoma AABB. Os genomas diplóides AA e BB são encontrados em diversas espécies silvestres da seção *Arachis*, que podem ser cruzadas com o amendoim. Embora estudos recentes tenham revelado variabilidade genética no amendoim cultivado, os níveis detectados são ainda insuficientes para a construção de mapas genéticos saturados. Por outro lado, as espécies silvestres de *Arachis* apresentam uma maior variabilidade (Hilu & Stalker, 1995; Stalker *et al.*, 1995; Burow *et al.*, 2001; Moretzsohn *et al.*, 2004; Milla *et al.*, 2005) e constituem uma importante fonte de genes úteis, tais como genes de resistência a estresses bióticos e abióticos (Stalker & Simpson, 1995; Pande & Rao, 2001; Rao *et al.*, 2003b; Dwivedi *et al.*, 2003). Além disso, diversos trabalhos têm mostrado a grande utilidade de espécies silvestres, de diferentes culturas, para o aumento da produtividade, devido à ampliação da base genética e a combinações gênicas superiores (Tanskley & McCouch, 1997; Brondani *et al.*, 2002; Fridman *et al.*, 2004; Talame *et al.*, 2004; Singh *et al.*, 2005).

No entanto, a introgressão desses genes úteis no amendoim não é uma tarefa fácil, devido, principalmente, a barreiras de recombinação e à transferência de genes não desejados junto aos genes de interesse (“linkage drag”). Dessa forma, a construção de mapas genéticos para espécies diplóides e tetraplóides de *Arachis*, baseados em marcadores transferíveis, como os microssatélites e marcadores âncoras possibilitará a construção de um mapa de referência para *Arachis* e a determinação do número e da posição genômica dos genes que controlam características de interesse. Com isso, será dado o primeiro passo para implementação de seleção assistida por marcadores nos programas de melhoramento, visando a introgressão mais rápida e eficiente desses genes no amendoim cultivado.

Arachis hypogaea é bastante susceptível a uma grande variedade de doenças e pragas. Dentre elas, as doenças fúngicas de parte aérea são consideradas o principal problema da cultura no Brasil e um dos principais problemas nos diversos países produtores, especialmente a mancha preta (*Cercosporidium personatum*), mancha parda (*Cercospora arachidicola*) e ferrugem (*Puccinia arachidicola*) (Pezzopane *et al.*, 1998; Almeida, 1999; Dwivedi *et al.*, 2003). As cercosporioses (mancha parda e mancha preta) têm sido responsáveis por reduções que chegam a

50% da produção comercial do amendoim, em várias regiões do mundo (Backman & Crawford, 1984; Almeida, 1999). Embora possam ser controladas quimicamente, os fungicidas são caros e tóxicos, podendo causar danos à saúde humana e ao meio ambiente. A utilização de variedades resistentes é, portanto, a maneira mais eficiente de controle dessas doenças. Poucas fontes de resistência às doenças fúngicas de parte aérea têm sido encontradas em *A. hypogaea*, mas altos níveis de resistência têm sido descritos para algumas espécies silvestres de *Arachis* (Company *et al.*, 1982; Gardner & Stalker, 1983; Stalker, 1984; Subrahmanyam *et al.*, 1985; Fávero, 2004).

Produtividade é uma característica complexa, determinada por vários genes e bastante influenciada pelo meio ambiente. É geralmente medida em termos de vários componentes, que, em leguminosas, incluem o número de vagens por planta, o número de sementes por vagem e o peso médio de sementes (Moot & McNeil, 1995; Timmerman-Vaughan *et al.*, 2005). QTLs que controlam componentes da produtividade têm sido mapeados para diversas espécies de plantas (Doganlar *et al.*, 2000; Zhang *et al.*, 2004; Humphry *et al.*, 2005; Timmerman-Vaughan *et al.*, 2005; Blair *et al.*, 2006; Yoon *et al.*, 2006). No entanto, até o momento nenhum QTL para produtividade ou para resistência à mancha preta foi mapeado em espécies de *Arachis*.

Os objetivos do presente estudo foram, portanto, o desenvolvimento de um mapa genético de ligação para espécies silvestres diplóides de *Arachis*, de genoma AA, baseado em marcadores microssatélites e em marcadores âncoras, além de alguns RGAs e, a partir desse mapa, identificar QTLs que controlam três características de interesse para o melhoramento do amendoim: o número de sementes por planta, o peso de sementes e a resistência ao fungo causador da mancha preta (*Cercosporidium personatum*).

2. MATERIAL E MÉTODOS

2.1. Material vegetal

Uma população F_2 foi construída por autopolinização de uma única planta F_1 , derivada do cruzamento entre *A. duranensis*, acesso K7988, usado como progenitor feminino e *A. stenosperma*, acesso V10309, usado como progenitor masculino. Para se obter o número desejado de plantas F_2 , a planta F_1 foi clonada (por mudas). Estas espécies foram escolhidas por possuírem genomas relacionados ao do amendoim (Kochert *et al.*, 1996), apresentarem alta capacidade de cruzamento (Krapovickas & Gregory, 1994), serem molecularmente contrastantes (Moretzsohn *et al.*, 2004) e, principalmente, por serem segregantes para resistência a nematóides e fungos de parte aérea (Leal-Bertioli *et al.*, 2000; Fávero, 2004). Estas plantas foram obtidas da coleção de germoplasma mantida na Embrapa Recursos Genéticos e Biotecnologia (Brasília-DF, Brasil).

2.2. Extração de DNA e PCR dos locos microssatélites

DNA genômico total foi extraído de folíolos jovens, geralmente ainda não abertos, de acordo com o método descrito por Ferreira & Grattapaglia (1998), com as modificações descritas no capítulo 1 desta tese. As quantificações do DNA extraído e as PCRs também foram realizadas como descrito no capítulo 1.

Os produtos de PCR foram visualizados em géis de poliacrilamida a 5% de concentração, corados com nitrato de prata (Creste *et al.*, 2001). Alguns marcadores, altamente contrastantes entre os dois progenitores, foram visualizados em géis de agarose (3,5%), contendo brometo de etídio. Os comprimentos dos fragmentos foram estimados nos géis, por comparações com padrões de bandas de tamanhos conhecidos (10-pb DNA ladder).

2.3. Construção do mapa

Os 271 novos marcadores microssatélites, 162 já publicados (Hopkins *et al.*, 1999; Palmieri *et al.*, 2002; 2005; He *et al.*, 2003; 2005; Moretzsohn *et al.*, 2004; Bravo *et al.*, 2006; Gimenes *et al.*, 2006) e 56 novos marcadores microssatélites desenvolvidos a partir de bibliotecas de cDNA de *A. stenosperma* (Proite *et al.*, em preparação) foram analisados nos dois acessos progenitores, para identificação dos

marcadores polimórficos. Além disso, foram incluídos 72 marcadores âncoras, polimórficos para essa população, desenvolvidos na Universidade de Aarhus, na Dinamarca, pelo grupo do Dr. Jens Stougaard. Os marcadores polimórficos foram analisados na população F_2 de mapeamento, constituída de 93 indivíduos. O teste de qui-quadrado (X^2) foi utilizado para testar a hipótese nula de segregação 1:2:1 ou 3:1 de todos os marcadores. O teste de FDR ou razão de falsas descobertas (Benjamin & Hocheberg, 1995) foi utilizado para determinação do nível de significância, ou seja, para correção do erro tipo I, que é a rejeição de uma hipótese nula verdadeira. Apenas os marcadores que não apresentaram distorções da proporção 1:2:1 esperada, a um limite de significância de 5%, foram utilizados inicialmente para a construção do mapa. As análises de ligação foram feitas usando o programa Mapmaker, versão 2.0 (Lander *et al.*, 1987). Um LOD score mínimo de 11 e uma fração de recombinação máxima (θ) de 0,35 foram utilizados como parâmetros para estabelecimento dos grupos de ligação, usando a função “group”. A ordem mais provável dos marcadores em cada grupo de ligação foi estimada pelo método de máxima verossimilhança, usando a função “first order” ou por análise de três pontos (função “three point”), usando-se sub-grupos, para grupos com mais do que sete marcadores. As ordens dos marcadores foram confirmadas por comparações das verossimilhanças (LOD scores) de todas as possíveis ordens usando o comando “compare” e por permutações de todas as trincas de marcadores adjacentes (função “ripple”). Em uma segunda etapa, o LOD score foi reduzido até 5, para possibilitar a inclusão de novos marcadores nos grupos. O comando “try” foi então usado para determinar a posição exata dos novos marcadores dentro de cada grupo. Em seguida, os marcadores distorcidos e os marcadores dominantes (RGAs e microssatélites segregando 3:1, ou seja, presentes em um dos progenitores e ausentes no outro) foram incluídos, usando a função “group”. As novas ordens obtidas foram novamente confirmadas com os comandos “first order”, “compare” e/ou “ripple”. As frações de recombinação foram convertidas em distâncias de mapa, em centimorgans (cM), usando a função de mapeamento de Kosambi, que leva em consideração a interferência (I). Os locos que puderam ser ordenados com um suporte maior ou igual a 1000:1 foram incluídos no mapa referência (“framework map”). Ou seja, foram aceitas as ordens de marcadores com uma verossimilhança mil vezes maior do que a segunda melhor ordem possível. Apenas os marcadores codominantes foram utilizados nessa análise. O tamanho do genoma foi estimado de

acordo com Hulbert e colaboradores (1988) considerando apenas as comparações entre pares de marcadores que constituíam o mapa referência.

Os marcadores que desviaram das proporções esperadas foram analisados quanto à natureza da distorção, se zigótica ou gamética. Para isso, dois testes de X^2 foram realizados, de acordo com Lorieux e colaboradores (1995). O primeiro testou a hipótese $p=q$, ou seja, se a frequência do alelo de *A. duranensis* era igual à do alelo de *A. stenosperma* em um determinado loco, pela fórmula $X^2=[(2np-n)^2+(2nq-n)^2]/n$, sendo $p(A)=(AA+AB/2)/n$, $q(B)=(BB+AB/2)/n$ e n =número total de indivíduos. O segundo X^2 foi utilizado para testar se as frequências dos genótipos eram iguais às esperadas sob equilíbrio de Hardy-Weinberg (EHW). O teste foi realizado pela equação $X^2=[(AA-np^2)^2]/np^2+[(Aa-2npq)^2]/2npq+[(aa-nq^2)^2]/nq^2$, sendo p^2 =número esperado de indivíduos AA sob EHW (homozigotos para o alelo de *A. duranensis*), q^2 =número esperado de indivíduos BB sob EHW (homozigotos para *A. stenosperma*) e $2pq$ =número esperado de indivíduos AB sob EHW (heterozigotos). Marcadores que deram valores de X^2 não significativos para o primeiro teste apresentaram tipo de seleção zigótica; marcadores com valores de X^2 significativos para este teste apresentaram seleção zigótica ou gamética, dependendo se o segundo teste foi também significativo ou não significativo, respectivamente (Lorieux *et al.*, 1995).

Os marcadores desenvolvidos a partir de bibliotecas genômicas foram denominados de TC ou AC (para bibliotecas enriquecidas para repetições TC ou AC), seguido do número da placa com o meio de cultura (TC1 a TC11 ou AC1 a AC3) e da identificação do clone na placa de 96 poços (A01 a H12). Marcadores desenvolvidos a partir de bibliotecas de cDNA foram denominados de ML, para folha madura ("mature leaf"), RN para raiz não inoculada, RI para raiz inoculada com o nematóide *Ditylenchus africanus* ou RM para raiz inoculada com *Meloidogyne*. Após as duas letras iniciais foram incluídos o número da placa (de 1 a 22) e a identificação do clone (A01 a H12). Marcadores obtidos por mineração de seqüências no GenBank foram chamados pelo nome com o qual a seqüência foi submetida. Os nomes desses marcadores começam com gi-, pPGP ou pPGS. Os nomes dos marcadores já publicados foram mantidos conforme a publicação original. Os publicados por He e colaboradores (2003; 2005) começam por PM seguido de um número de identificação, enquanto os demais já publicados começam por Ag, Ah ou Ap (Hopkins *et al.*, 1999; Palmieri *et al.*, 2002; 2005; Moretzsohn *et al.*, 2004; Bravo

et al., 2006; Gimenes *et al.*, 2006). Todos os marcadores âncoras foram denominados por Leg, seguido de um número ou um nome de identificação. Finalmente, os análogos a genes de resistência apresentam o nome RGA, após o nome da sonda utilizada (S1A36 ou S1A37), no caso de RFLPs, ou das iniciais da espécie da qual o fragmento foi amplificado (AFLP modificado), sendo *Ad* para *A. duranensis* e *As* para *A. stenosperma*.

2.4. Avaliação fenotípica

Os progenitores, a planta F_1 e os 93 indivíduos F_2 foram avaliados para resistências / suscetibilidades ao fungo *Cercosporidium personatum*, causador da mancha preta no amendoim. A cultura de *C. personatum* foi isolada de folhas de amendoim, em Campinas-SP, na safra 2002/2003 e mantida em meio aveia-agar, com repicagens a cada 14 dias (Fávero *et al.*, 2005). As avaliações foram feitas usando-se a técnica da folha destacada, descrita por Moraes & Salgado (1984). As folhas foram coletadas de plantas adultas, mantidas em vasos, em casa de vegetação. Esta técnica consistiu, resumidamente, do acondicionamento das folhas com um chumaço de algodão úmido na ponta do pecíolo e com o lado adaxial para cima em placas de petri, contendo uma camada de algodão no fundo, coberta por uma folha de papel de filtro e sobre esta, uma lâmina de vidro para sustentar os folíolos. Após adição de 20 mL de água destilada e autoclavada por placa de petri, foi feita a inoculação do patógeno, por aspersão de uma solução de água destilada, Tween 20 a 0,5% e esporos, a uma concentração de 50.000 esporos / mL sobre as folhas, dentro das placas. Todas as etapas, desde a coleta das folhas até a inoculação do patógeno foram feitas no mesmo dia. As placas foram mantidas em uma sala com temperatura controlada a 23-25°C, acondicionadas em estantes, sob luz fluorescente. A iluminação foi programada para 10 horas de luz e 14 horas de escuro. As avaliações foram efetuadas 42 dias após a inoculação do patógeno, pela razão entre a área de lesão e a área foliolar, em mm². Foram avaliadas quatro placas de petri por planta, com quatro folíolos cada uma. As médias das 16 avaliações para cada planta foram consideradas para a análise.

Dois componentes da produtividade foram também avaliados: peso de 15 sementes e número de sementes por planta que, nesse caso, é igual ao número de vagens por planta, uma vez que todas as plantas possuíam apenas uma semente

por vagem. O peso de sementes foi obtido pela média de 15 sementes, escolhidas ao acaso. Para 14 plantas F_2 que produziram muito pouco foram consideradas as médias de 3 a 13 sementes. O número de sementes por planta foi obtido pelo total de sementes coletado de cada planta, ao longo de três anos. Como *A. duranensis* e *A. stenosperma* são plantas anuais, ao final de cada ciclo, as plantas eram transplantadas para outro vaso e as sementes colhidas. O número total de sementes colhido de cada planta nas três colheitas foi considerado para análise.

As distribuições fenotípicas das três características foram analisadas pelos testes de assimetria e curtose, cujas significâncias foram avaliadas por testes de Lilliefors (Lilliefors, 1967). Além disso, foram estimadas as correlações de Pearson entre as três características, avaliadas pelo teste de significância de Student. Essas análises foram realizadas com o programa GQMOL (Cruz, 2004).

2.5. Análise de QTLs

Quatro métodos foram utilizados para detecção dos QTLs que controlam as três características avaliadas: (1) análise de marcas simples, com comparação entre as médias das classes genotípicas pelo método de regressão linear simples; (2) mapeamento por intervalo, (3) mapeamento por intervalo composto e (4) mapeamento por intervalo múltiplo. As análises de marcas simples foram realizadas com o programa GQMOL (Cruz, 2004) e as demais análises, com o programa QTL Cartographer, versão 2.5 (Wang *et al.*, 2006). Os pontos de corte para cada característica foram determinados pelo programa, via testes de permutação (Churchill & Doerge, 1994; Doerge & Churchill, 1996), sendo utilizadas 1000 permutações, com três repetições, e um nível de significância de 0,05. Os testes de hipótese para presença do QTL nos três métodos de mapeamento por intervalo foram realizados a cada 1 cM do mapa de ligação (“walk speed” = 1). As apresentações gráficas dos grupos de ligação e dos QTLs foram obtidas usando o programa MapChart, versão 2.1 (Voorrips, 2002).

A análise de mapeamento por intervalo múltiplo foi feita usando-se a estatística de Critério de Informação Bayesiano (“Bayesian Informaton Criterion” - BIC) para criação de um modelo inicial, ou seja, para identificação de QTLs e quantificação de seus efeitos, conforme recomendação de Zeng e colaboradores (1999). A partir do modelo inicial realizaram-se os seguintes passos:

- 1) otimização das posições dos QTLs (“optimizing QTL positions”);
- 2) otimização das posições dos QTLs dentro do intervalo entre os dois marcadores adjacentes mais próximos ao QTL;
- 3) busca de novos QTLs (“search for new QTLs”) e incorporação ao modelo, se significativos (“testing for existing QTLs”);
- 4) busca de epistasias (“search for epistasis”) e incorporação ao modelo, se significativas (“testing for existing epistasis”);
- 5) repetição dos passos 1 a 4 até que novos QTLs ou epistasias não fossem detectados;
- 6) otimização das posições dos QTLs para o modelo MIM final;
- 7) testes dos efeitos aditivos, de dominância e epistáticos, se presentes (“testing all”);
- 8) obtenção do resumo dos dados analisados através da ferramenta “MIM Model Summary”, ou seja, as posições, as razões de verossimilhança, os efeitos epistáticos e as partições da variância explicada pelos QTLs foram estimados para os QTLs significativos (“Model summary file”) e plotados em gráfico (“Graphic result file”).

3. RESULTADOS E DISCUSSÃO

3.1. Construção do mapa

A análise de 489 locos microssatélites em *A. duranensis*, acesso K7988 e *A. stenosperma*, acesso V10309, mostrou que 215 (44,0%) eram polimórficos (Figura 1), sendo 181 codominantes e 34 dominantes, ou seja, amplificaram alelos nulos em um dos progenitores. Dentre os 181 marcadores codominantes, 99 não apresentaram distorções da segregação 1:2:1 esperada ($p < 0,05$) e esses marcadores, junto aos 40 marcadores âncoras (Figura 1) não distorcidos (de um total de 72), foram utilizados para o estabelecimento dos grupos de ligação, usando um LOD score mínimo de 11. Em seguida, os marcadores distorcidos e os dominantes foram incluídos no mapa, usando um LOD score mínimo de 5 e uma fração de recombinação máxima de 0,35. No total, 182 marcadores microssatélites, 66 marcadores âncoras e 12 RGAs foram mapeados em 10 grupos de ligação (Figura 2), como esperado, uma vez que *A. duranensis* e *A. stenosperma* são espécies diplóides com $2n=20$ cromossomos. No entanto, 39 marcadores não puderam ser mapeados, o que pode indicar que partes do genoma não foram cobertas por esse mapa.

Um comprimento total de 1645,1 cM foi obtido e os grupos variaram de 58,7 cM (com 6 marcadores) a 279,3 cM (43 marcadores), com uma distância média de 6,32 cM entre marcadores adjacentes. Os grupos de ligação foram numerados de acordo com os números publicados na primeira versão do mapa (Moretzsohn *et al.*, 2005 - ANEXO 1). Seis pares de primers amplificaram consistentemente dois locos polimórficos (AC2C2, TC1D1, TC4H2, Ah745, Seq15D6 e gi-7295) e os 12 locos foram incluídos no mapa. Esses marcadores foram indicados pelos números -1 e -2 após o nome do marcador (Figura 2). Os marcadores mapeados com LOD scores < 11 (de 5 a 11), que foi o valor utilizado para estabelecimento dos grupos de ligação, estão indicados no mapa por uma seta seguida do LOD score máximo com que esses marcadores foram mapeados.

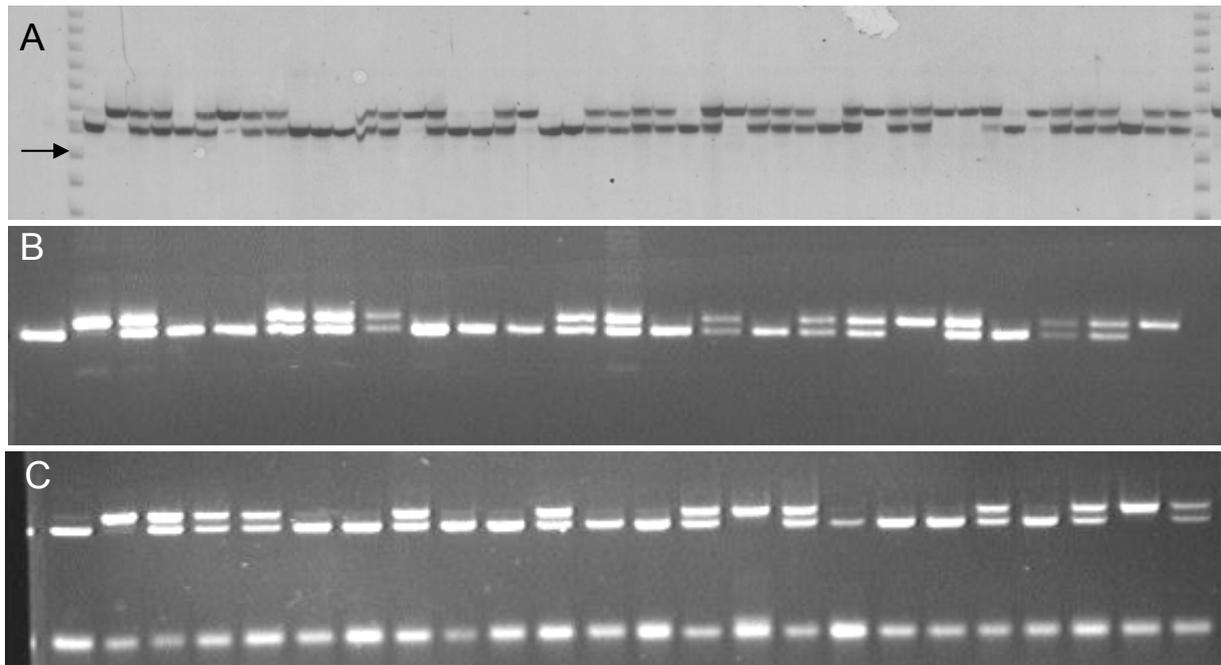


Figura 1 – Padrões de amplificação obtidos para o loco RN8C09, por eletroforese em gel de poliacrilamida 5% (foto A) e para os locos RN10F09 (foto B) e Leg69, que é um marcador âncora (foto C), em géis de agarose 3,5%. Na foto A, a primeira e a última linhas são padrões de tamanhos conhecidos (10bp DNA ladder). Nos três géis, da esquerda para a direita: *A. duranensis* (acesso K7988), usado como progenitor feminino; *A. stenosperma* (acesso V10309), progenitor masculino; híbrido F₁ e 45 indivíduos F₂ na foto A e 21 indivíduos F₂ nas fotos B e C. A seta na foto A indica o fragmento de 250 pares de bases.

O comprimento total obtido com base nessa população (1645,1 cM) é consideravelmente maior do que os 1063 cM obtidos para outra população diplóide, de genoma AA, resultante de um cruzamento entre *A. duranensis* e *A. cardenasii* (Halward *et al.*, 1993) e do que metade dos 2210 cM obtidos para uma população tetraplóide (Burow *et al.*, 2001). A presença de marcadores distorcidos, principalmente em favor de heterozigotos, tem sido apontada como causadora de expansão do comprimento de mapas genéticos (Knox & Ellis, 2002) e, talvez, possa explicar essas diferenças.

O mapa de referência, com os marcadores ordenados com suporte de 1000:1, foi composto de 158 marcadores, cobrindo 1422,8 cM, com uma distância média de 9,0 cM entre marcadores. O número de marcadores por grupo de ligação variou de seis (grupo 10) a 28 (grupo 2). Esses 158 marcadores aparecem sublinhados no mapa (Figura 2).

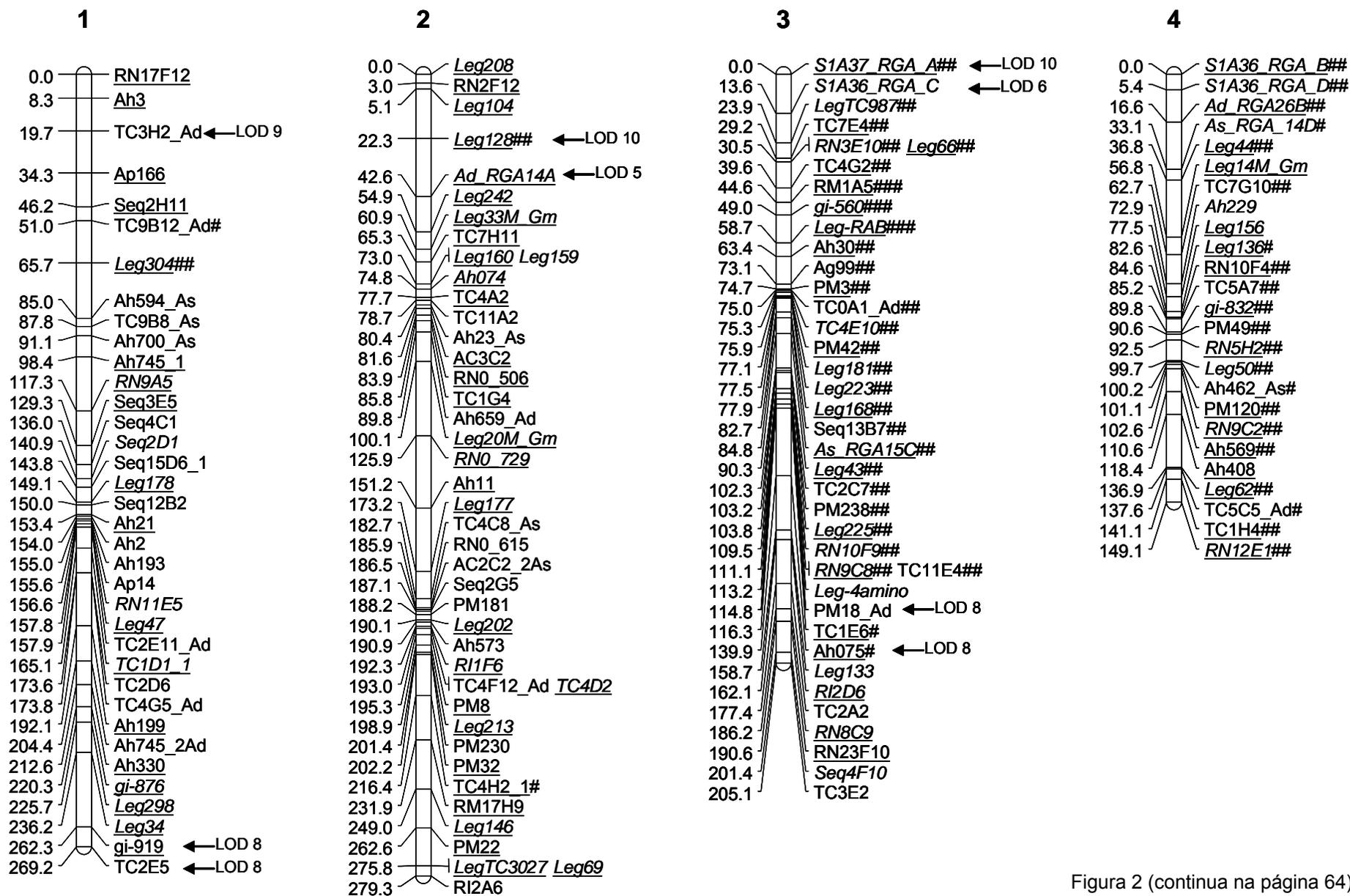


Figura 2 (continua na página 64).

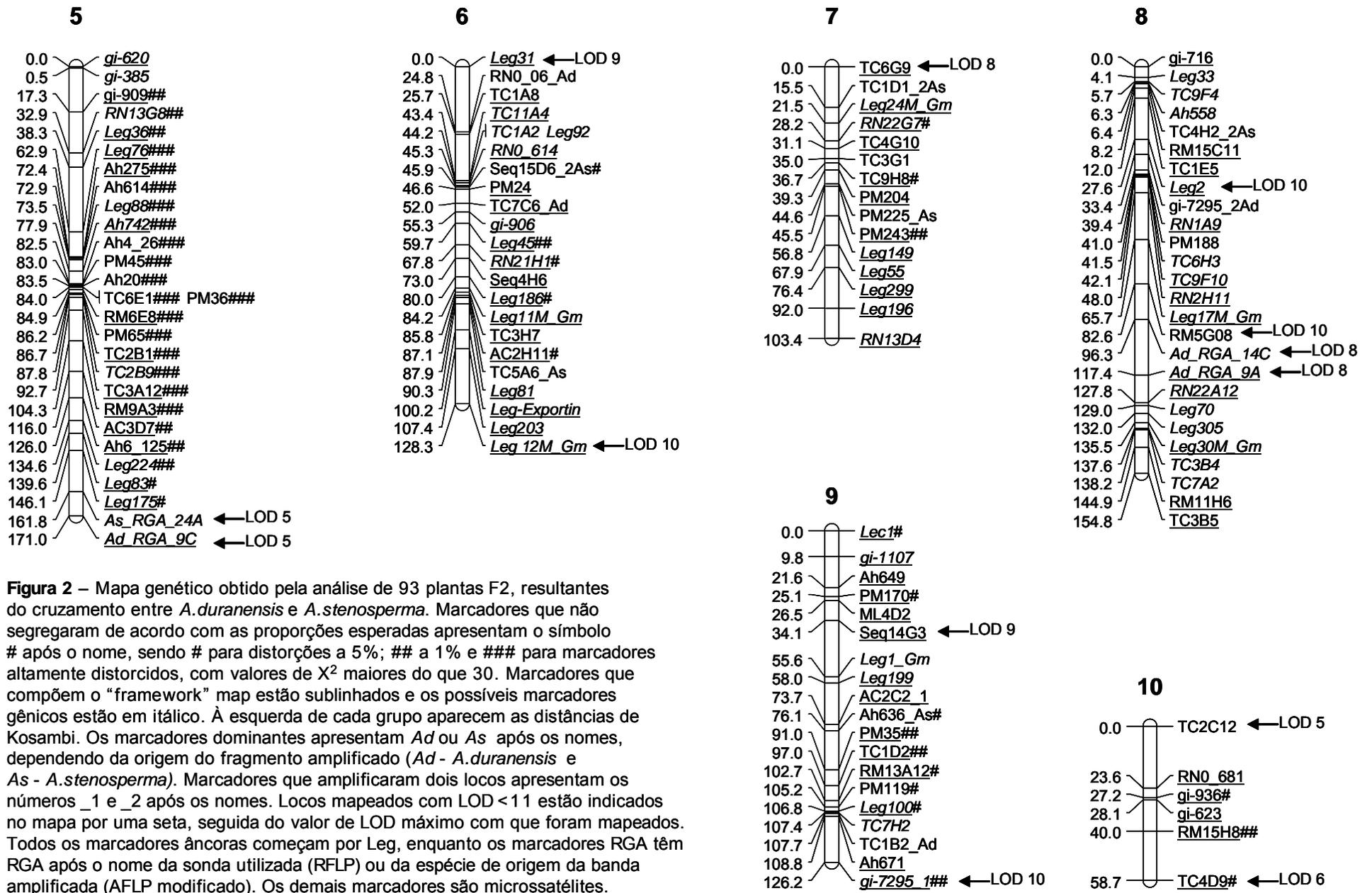


Figura 2 – Mapa genético obtido pela análise de 93 plantas F2, resultantes do cruzamento entre *A. duranensis* e *A. stenosperma*. Marcadores que não segregaram de acordo com as proporções esperadas apresentam o símbolo # após o nome, sendo # para distorções a 5%; ## a 1% e ### para marcadores altamente distorcidos, com valores de X² maiores do que 30. Marcadores que compõem o “framework” map estão sublinhados e os possíveis marcadores gênicos estão em itálico. À esquerda de cada grupo aparecem as distâncias de Kosambi. Os marcadores dominantes apresentam Ad ou As após os nomes, dependendo da origem do fragmento amplificado (Ad - *A. duranensis* e As - *A. stenosperma*). Marcadores que amplificaram dois locos apresentam os números _1 e _2 após os nomes. Locos mapeados com LOD < 11 estão indicados no mapa por uma seta, seguida do valor de LOD máximo com que foram mapeados. Todos os marcadores âncoras começam por Leg, enquanto os marcadores RGA têm RGA após o nome da sonda utilizada (RFLP) ou da espécie de origem da banda amplificada (AFLP modificado). Os demais marcadores são microssatélites.

Do total de 299 marcadores polimórficos (incluindo os 12 RGAs), 133 (44,5%) desviaram das proporções 1:2:1 (108 marcadores) ou 3:1 (25 marcadores) esperadas ($p < 0,05$). Dos 108 marcadores codominantes que desviaram da proporção 1:2:1 esperada, 59 apresentaram um excesso de alelos de *A. duranensis* (progenitor feminino); 25, de alelos de *A. stenosperma* (progenitor masculino) e 24, um excesso de heterozigotos. Para os 25 marcadores dominantes distorcidos, quinze apresentaram um excesso de alelos de *A. stenosperma* e dez, de *A. duranensis*. Marcadores distorcidos foram mapeados em todos os grupos, com exceção do grupo 8. No grupo 1, apenas um marcador codominante apresentou distorção de segregação (Leg304), a favor do alelo de *A. stenosperma*. O único marcador dominante distorcido deste grupo (TC9B12_Ad) apresentou também um excesso de alelos de *A. stenosperma*. O grupo 2 apresentou apenas dois marcadores distorcidos, sendo um a favor de *A. stenosperma* (Leg128) e um, do heterozigoto (TC4H2). O grupo 3 foi composto, basicamente, de marcadores distorcidos (25 de um total de 34 marcadores codominantes e três de seis dominantes), sendo 23 com excesso de alelos de *A. duranensis* e cinco com excesso de heterozigotos. Os 21 marcadores distorcidos do grupo 4, de um total de 25, apresentaram um excesso de alelos de *A. stenosperma*. Por outro lado, os grupos 5 e 9 apresentaram marcadores distorcidos a favor dos alelos de *A. duranensis* ou com um déficit de indivíduos heterozigotos. O grupo 6 apresentou três locos com excesso de heterozigotos e dois com excesso de alelos de *A. stenosperma*, sendo dois dominantes. Finalmente, os grupos 7 e 10 apresentaram todos os marcadores distorcidos com excesso de heterozigotos. A proporção de marcadores distorcidos encontrada nesse estudo é maior do que os 25% descritos para os dois mapas, baseados em marcadores RFLP, publicados para *Arachis* e também envolvendo cruzamentos interespecíficos (Halward *et al.*, 1993; Burow *et al.*, 2001). Distorções de segregação têm sido descritas para a grande maioria dos mapas genéticos publicados e têm sido atribuídas a diferentes causas, tais como alelos recessivos deletérios (Berry *et al.*, 1995), alelos auto-incompatíveis (Barzen *et al.*, 1995), re-arranjos estruturais nos cromossomos (Quillet *et al.*, 1995), seleção pré ou pós zigótica de combinações alélicas ou mesmo ao tamanho diferente entre os genomas dos progenitores, levando a um pareamento atípico dos cromossomos (Jenczewski *et al.*, 1997). Além disso, erros de genotipagem também podem causar desvios das proporções esperadas.

As análises de X^2 dos 108 marcadores codominantes distorcidos mostraram que 88 apresentaram genótipos na F_2 que não estavam distribuídos de acordo com o esperado sob Equilíbrio de Hardy-Weinberg, indicando que a distorção nesses locos foi causada na fase zigótica. Um número considerável desses marcadores apresentou distorção em favor dos heterozigotos. Por isso, uma das causas da distorção zigótica observada pode ter sido o vigor híbrido, como relatado para outras espécies, tais como batata (Bonierbale *et al.*, 1988), alfafa (Diwan *et al.*, 2000) e melão (Ritschel *et al.*, 2004). Os dados dos demais 20 marcadores indicaram que a distorção de segregação foi causada também na fase gamética. Independentemente da causa, a distorção em cromossomos, ou em partes deles, traz algumas limitações para o uso desse mapa em programas de seleção assistida. Novos mapas genéticos e experimentos de citogenética são necessários para um melhor entendimento desses resultados e esses encontram-se em andamento como parte de um projeto que está sendo desenvolvido na Embrapa Recursos Genéticos e Biotecnologia e colaboradores, financiado pelo Generation Challenge Program.

Quarenta e cinco dos 182 locos microssatélites mapeados são ESTs ou genes caracterizados. Dentre os demais 137, que foram desenvolvidos de seqüências genômicas, 25 apresentaram similaridades significativas na análise de BlastX. Além disso, 66 marcadores âncoras, que são gênicos e de cópias únicas, foram mapeados. Portanto, cerca de 136 (54,8%) dos 248 marcadores mapeados provavelmente representam genes. Além disso, foram mapeados 12 RGAs que são, provavelmente, genes ou partes de genes de resistência. Como esperado, as similaridades foram diversas, mas incluíram algumas de possível interesse para qualidade de semente (alergênicos em semente e proteína ligada à maturação de semente); resistência a doenças (lectinas, inibidor de proteases, “cinnamoyl-CoA reductase”, proteína induzida por lesão); resistência a seca (aquaporina e proteína ativada em resposta a desidratação) (ANEXO 3).

O tamanho do genoma foi estimado pelo método de Hulbert e colaboradores (1988). Somente os marcadores do mapa de referência foram incluídos nessa análise, para evitar estimativas errôneas da cobertura do genoma. Uma distância total de 1424,7 cM foi estimada, que é similar aos 1422,8 cM cobertos pelo mapa de referência e inferior ao comprimento total do mapa obtido no presente trabalho (1645,1 cM). Portanto, apesar dos possíveis erros de genotipagem terem sido corrigidos com o uso da ferramenta “error detection” do Mapmaker, alguns erros

devem ter sido mantidos, superestimando algumas distâncias entre marcadores. Além disso, como já comentado, distorções de segregação, principalmente em favor de heterozigotos, que foram comuns no presente estudo, podem levar a uma superestimativa do comprimento de mapas genéticos. O tamanho do genoma haplóide de *A. hypogaea* foi estimado em 2.813.000.000 de pares de bases (Bennet & Leitch, 2005). Se espécies silvestres de *Arachis* possuem genomas similares e dada a distância total do mapa estimada, 1 cM deve corresponder, em média, a cerca de 2.000.000 de pares de bases no genoma de *Arachis*.

3.2. Análise de QTLs

Um total de 93 plantas F₂ foi avaliado para as três características. Os valores médios e os desvios-padrões (s) estimados para os acessos progenitores e plantas F₁ e F₂ são apresentados na tabela 1. Para a população F₂ são apresentados também os valores mínimos e máximos obtidos para cada característica. Os dados do número de sementes por planta para os acessos progenitores e para o híbrido F₁ não puderam ser coletados.

Tabela 1 - Médias e desvios-padrões (s) para três características fenotípicas estimados para os acessos progenitores e plantas F₁ e F₂. Para a população F₂ são apresentados também os valores mínimos e máximos.

Característica	K7988		V10309		F1		População F2				
	Média	s	Média	s	Média	s	Média	s	Mínimo	Máximo	N
Peso de sementes	0,1510	-	0,2430	-	0,1013	-	0,1732	0,0543	0,0133	0,3242	87
Sementes por planta	-	-	-	-	-	-	200,05	145,27	1,00	709,00	82
Resistência a <i>C.personatum</i>	7,8879	0,1061	0,1678	0,0047	0,0000	0,0000	0,3151	1,3717	0,0000	12,7782	90

A distribuição fenotípica da característica peso de sementes, avaliada na população F₂, mostrou uma distribuição normal (Figura 3-A). No entanto, os testes de Lilliefors baseados nos coeficientes de assimetria e curtose foram significativos (p<0,01) para as outras duas características incluídas na análise (Figura 3-B e C). A distribuição da característica número de sementes por planta mostrou um desvio em favor de um menor peso de semente, enquanto a resistência à mancha preta mostrou-se fortemente desviada para resistência. Tentativas de normalizar esses dados foram feitas, mas os resultados foram similares aos dados originais.

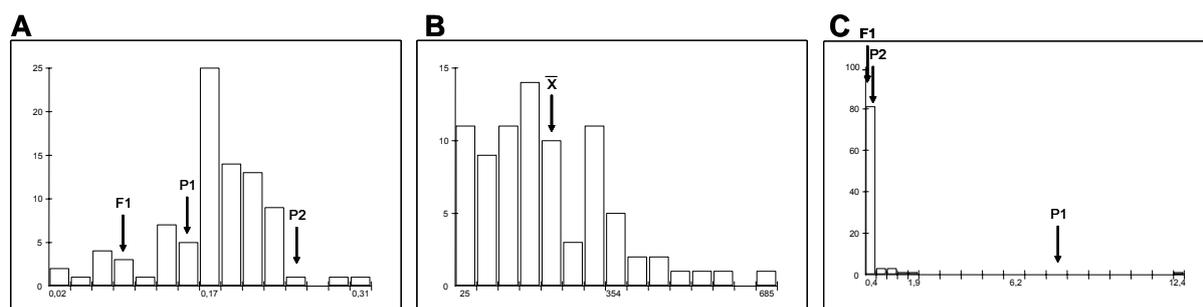


Figura 3 – Distribuição de freqüências fenotípicas das três características avaliadas na população F₂: (A) peso de sementes, (B) número de sementes por planta e (C) resistência à mancha-preta. Os fenótipos dos progenitores feminino (P₁) e masculino (P₂) e do híbrido F₁ são indicados para as características peso de sementes e resistência à mancha-preta. Para número de sementes é apresentada somente a média fenotípica.

As correlações obtidas para as três características são apresentadas na Tabela 2. Uma correlação significativa ($p < 0,05$) e negativa (-0,2199) foi obtida para as características número de sementes por planta e resistência à mancha preta, que foi avaliada dividindo-se a área da lesão pela área do folíolo. Conseqüentemente, quanto menores os valores, mais resistentes as plantas. Portanto, essa correlação significativa e negativa é interessante, uma vez que a seleção de indivíduos com maior número de sementes por planta, levaria, provavelmente, à seleção indireta de indivíduos mais resistentes e vice-versa. A correlação entre peso de sementes e resistência à mancha preta também foi negativa, mas não significativa (Tabela 2). A correlação entre peso e número de sementes foi positiva e não significativa, ao contrário de alguns trabalhos com diferentes espécies de plantas, que têm mostrado haver uma indesejável correlação negativa entre essas duas características (Xiao *et al.*, 1998; Timmerman-Vaughan *et al.*, 2005; Herrmann *et al.*, 2006; Yoon *et al.*, 2006).

Tabela 2 – Coeficientes de correlação de Pearson, obtidos para as três características avaliadas.

Característica	Peso de sementes	Número de sementes
Peso de sementes		
Número de sementes	0,0971	
Resistência a <i>C. personatum</i>	-0,0413	-0,2199*

*significativo a 5% de probabilidade, pelo teste t.

3.2.1. QTLs para peso de sementes

Dez marcadores foram significativamente associados com peso de sementes pela análise de marcas simples, usando o método de regressão linear múltipla (Tabela 3). Esses marcadores encontram-se localizados em quatro grupos de ligação, sendo um no grupo 1, quatro no grupo 2, quatro no grupo 3 e um último marcador no grupo 9. As porcentagens da variação explicada (R^2) por esses 10 marcadores variaram de 3,06 a 9,36 (Tabela 3).

Tabela 3 – Marcadores significativamente associados à característica peso de sementes pela análise de marcas simples. Todos os marcadores são codominantes.

Marcador ^a	Grupo de ligação	Posição (cM)	R^2 (%) ^b
Leg304*	1	65,7	6,63
Leg202*	2	190,1	3,06
Ah573***	2	190,9	6,64
RM17H9*	2	231,9	6,64
Leg146**	2	249,0	8,84
gi-560*	3	49,0	8,49
Ah30*	3	63,4	9,36
Ag99*	3	73,1	8,13
PM3*	3	74,7	6,69
gi-7295*	9	126,2	6,92

^aMarcadores significativamente associados a 5% (*), 1% (**) e 0,1% (***).

^bPorcentagem da variação fenotípica explicada pelo marcador.

O valor mínimo de LOD score para considerar válidos os QTLs para peso de sementes foi estimado em 4,8 pela média de três repetições de 1000 permutações (Churchill & Doerge, 1994; Doerge & Churchill, 1996). Este valor foi utilizado como ponte de corte nas análises de mapeamento por intervalo. Nenhum QTL foi encontrado com LOD scores iguais ou superiores a 4,8 pelos métodos de mapeamento por intervalo e por intervalo composto. No entanto, dez QTLs foram significativamente identificados pelo método de mapeamento por intervalo múltiplo (Tabela 4). Os QTLs localizados nos grupos 2, 3 e 9 coincidem ou estão próximos aos marcadores identificados pelo método de marcas simples. Os dois QTLs identificados no grupo 3, embora bastante próximos (Figura 4), foram

consistentemente identificados como dois picos e apresentaram valores de LOD significativos, ambos superiores a 8 (Tabela 4). Além disso, a ferramenta “testing for existing QTLs” não excluiu nenhum dos dois QTLs. No entanto, uma maior saturação de marcadores nessa região é necessária para a confirmação da existência desses dois QTLs. Nenhum efeito de epistasia foi detectado entre os dez QTLs.

Mapeamento por intervalo múltiplo é, atualmente, um dos métodos mais precisos e com maior poder de detecção de QTLs (Kao *et al.*, 1999; Coelho & Silva, 2002; Mayer, 2005). No entanto, ainda não foi definido qual o melhor critério para estimar o ponto de corte a ser usado por esse método. Kao e colaboradores (1999), que desenvolveram o método, sugerem que sejam usados os mesmos valores estimados, por permutações, para o mapeamento por intervalo e por intervalo composto, até que sejam definidos os critérios para modelos baseados em múltiplos QTLs. Com isso, no presente trabalho são descritos e mapeados os primeiros QTLs para componentes da produtividade em *Arachis* spp.

Como esperado, o presente trabalho mostrou que peso de sementes possui herança quantitativa em *Arachis*. Vários trabalhos têm mostrado que um grande número de locos controla essa característica em diferentes espécies de leguminosas (Fatokun *et al.*, 1992; Maughan *et al.*, 1996; Timmerman-Vaughan *et al.*, 1996; Csanadi *et al.*, 2001; Humphry *et al.*, 2005).

Tabela 4 – QTLs detectados para a característica peso de sementes, pelo método de mapeamento por intervalo múltiplo. Todos os marcadores são codominantes.

QTL	GL ^a	Intervalo ^b	Posição ^c	LOD ^d	a ^e	d ^e	a (%) ^f	d (%) ^f	d/a ^g	Ação gênica ^h
<i>psem2.1</i>	2	PM181 / Leg202	189,2	7,96	-0,0676	0,0167	32,0	1,0	0,25	A
<i>psem2.2</i>	2	TC4H2 / RM17H9	226,4	9,29	0,0287	-0,0242	-9,2	-4,0	0,84	D
<i>psem3.1</i>	3	TC1E6 / Ah075	136,7	8,46	-0,0443	-0,036	14,0	4,6	0,81	D
<i>psem3.2</i>	3	Ah075 / Leg133	148,0	9,12	0,0109	0,0541	-0,9	10,1	4,96	SD
<i>psem4</i>	4	PM49 / RN5H2	92,4	8,02	-0,0067	-0,0098	-0,8	1,9	1,46	SD
<i>psem5</i>	5	AC3D7 / Ah6-125	117,0	7,99	-0,0037	-0,0185	-0,4	1,8	5,00	SD
<i>psem6</i>	6	Leg203 / Leg12M-Gm	114,9	7,94	-0,0043	-0,0215	0,2	2,1	5,00	SD
<i>psem7</i>	7	PM243 / Leg149	47,5	11,19	-0,0137	0,0499	1,0	7,2	3,64	SD
<i>psem9.1</i>	9	Lec1 / gi-1107	3,9	13,52	0,0061	0,0547	-0,7	16,3	8,97	SD
<i>psem9.2</i>	9	RM13A12 / PM119	102,8	10,54	0,0496	0,0162	11,6	2,1	0,33	A

^aGrupo de ligação a que pertence o QTL.

^bMarcadores adjacentes ao QTL.

^cPosição no grupo de ligação, em cM, em que foi observado o maior valor de LOD score.

^dLOD score máximo observado.

^eEfeitos aditivo (a) e de dominância (d). O efeito aditivo refere-se ao efeito da substituição dos alelos de *A. duranensis* pelos alelos de *A. stenosperma*. Valor negativo indica diminuição do valor da característica. Para o efeito de dominância, valor positivo indica que o heterozigoto possui, em média, valor fenotípico superior à média dos dois homozigotos e valor negativo, que o heterozigoto possui, em média, valor fenotípico inferior à média dos dois homozigotos.

^fEfeitos aditivo (a) e de dominância (d), em porcentagem da variação fenotípica total.

^{g,h}Ação gênica, calculada por $[d] / [a]$ e interpretada, de acordo com Edwards *et al.*, 1987, como $d/a < 0,55$, ação aditiva (A) ou de dominância parcial; $0,55 < d/a < 1,25$, dominante (D) e $d/a > 1,25$, ação sobredominante (SD).

Os efeitos fenotípicos, aditivos e/ou dominantes, desses dez QTLs foram estimados pela análise de variância através do método de mapeamento por intervalo múltiplo (Tabela 4). Alelos com efeitos aditivos significativos foram identificados, ou seja, alelos que, quando presentes aumentaram o peso de sementes da planta. No QTL *psem2.1*, por exemplo, o alelo Leg202_172pb, proveniente do progenitor feminino (*A. duranensis*), que produz sementes com menor peso, possui um grande efeito aditivo (0,0676 ou 32,0% da variância aditiva total), ou seja, está ligado à produção de sementes com maior peso. Alelos que conferem um aumento da característica, mesmo sendo provenientes de parentais agronomicamente inferiores, têm sido identificados para diferentes espécies de plantas (Xiao *et al.*, 1998; Rao *et al.*, 2003a; Frary *et al.*, 2004; Wang *et al.*, 2004; Yoon *et al.*, 2006). Outros dois QTLs, *psem3.1* e *psem9.2*, apresentaram alelos com significativos efeitos aditivos: TC1E6_134pb (14,0%) e RM13A12_235pb (11,6%), respectivamente, ambos provenientes do progenitor masculino (*A. stenosperma*). Esses três QTLs, em conjunto, responderam por 57,6% da variância aditiva e possuem pouco efeito de dominância (7,7% na soma dos três QTLs), sendo, portanto de grande potencial para uso em seleção assistida por marcadores. Além disso, dois QTLs apresentaram possíveis efeitos de heterose (*psem7* e *psem 9.1*). Obviamente, essas informações precisam ser validadas em populações maiores, com uma menor proporção de marcadores distorcidos e, principalmente, em populações que incluam o amendoim cultivado e analisadas em experimentos replicados em diferentes ambientes.

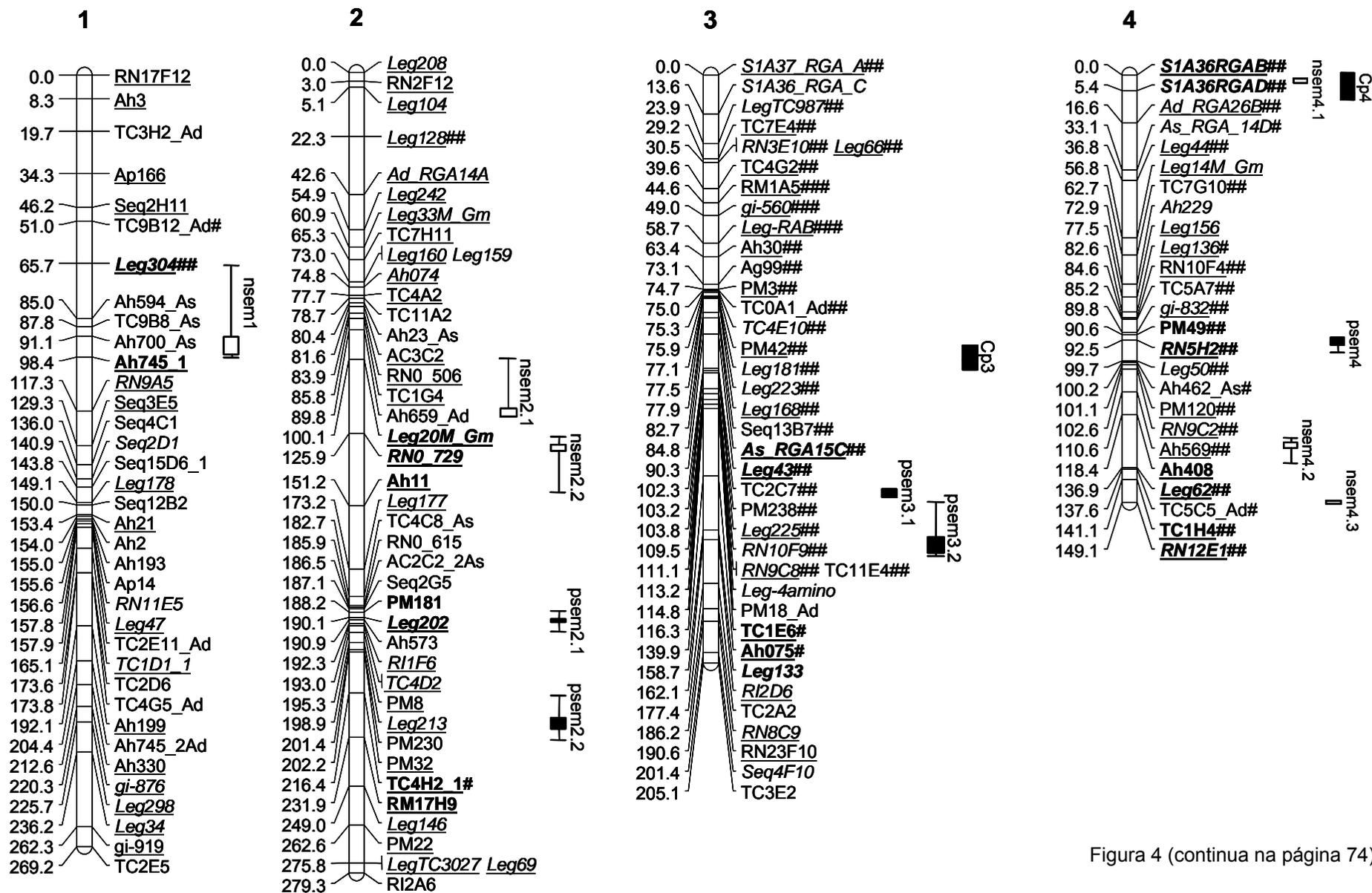


Figura 4 (continua na página 74).

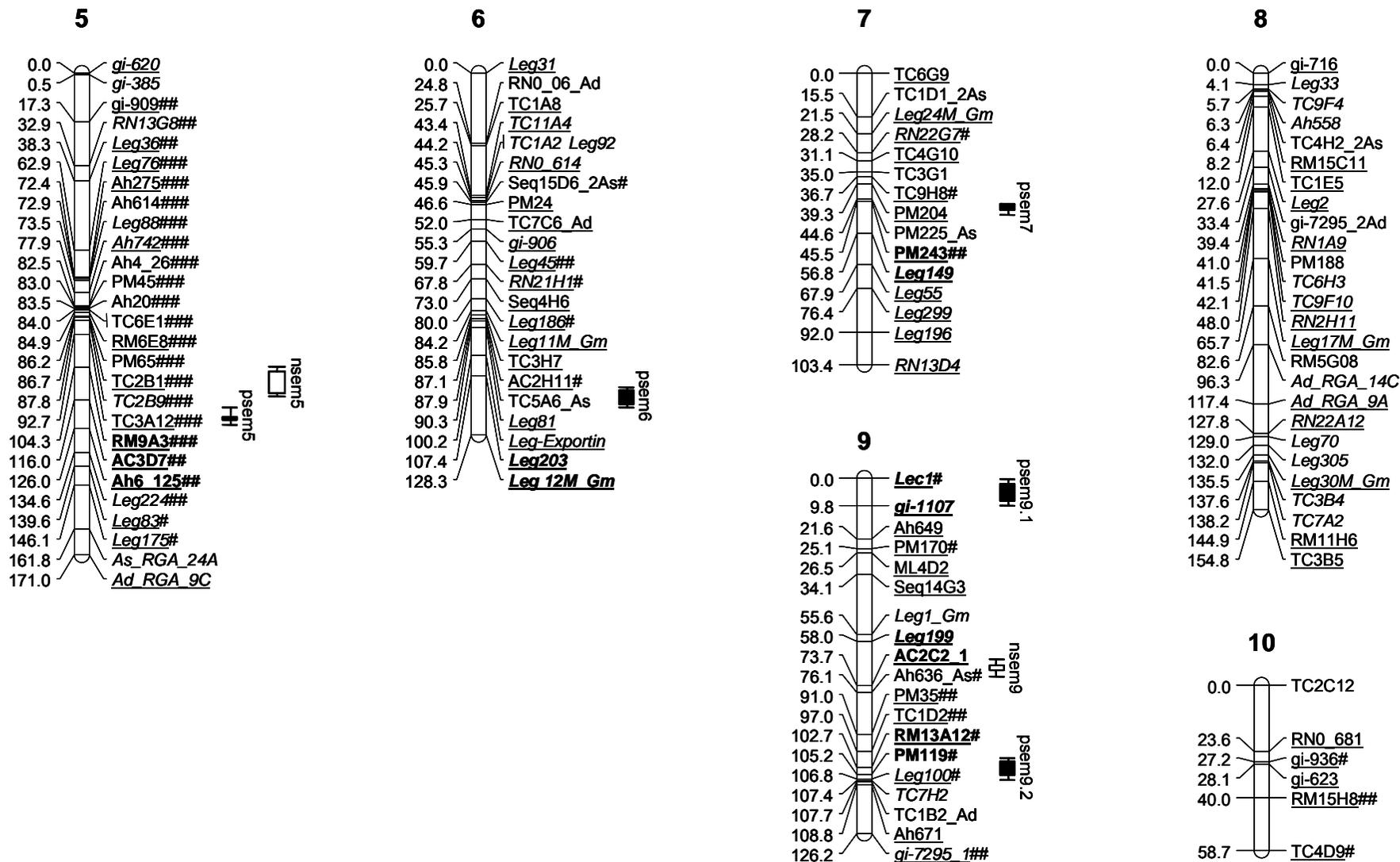


Figura 4 – Localização de QTLs para as três características fenotípicas avaliadas: peso de sementes (*psem*), número de sementes por planta (*nsem*) e resistência ao fungo *Cercosporidium personatum* (*Cp*). Os marcadores adjacentes aos QTLs aparecem em negrito. As demais informações referentes à nomenclatura dos marcadores e ao mapa de ligação encontram-se na legenda da figura 2 (página 67).

3.2.2. QTLs para número de sementes por planta

A análise de marcas simples resultou na identificação de 28 marcadores significativamente associados com a característica número de sementes por planta (Tabela 5). Esses marcadores encontram-se localizados em seis grupos de ligação, dispersos por oito regiões do mapa, sendo uma no grupo 1, uma no grupo 2, duas no grupo 3, duas no grupo 4, uma região no grupo 5 e uma última região no grupo 7. Além disso, quatro marcadores que não mapearam em nenhum dos grupos deram associações significativas (Leg24, PM50, TC11F12 e TC9E8). As porcentagens da variação explicada (R^2) por esses 28 marcadores variaram de 4,01 a 14,34 (Tabela 5).

Tabela 5 – Marcadores significativamente associados à característica número de sementes por planta, pela análise de marcas simples. Todos os marcadores são codominantes.

Marcador ^a	Grupo de ligação	Posição (cM)	R^2 (%) ^b
Ah745*	1	98,4	7,44
Leg20M-Gm*	2	100,1	4,01
Leg_RAB**	3	58,7	7,64
Ah30**	3	63,4	11,84
RN8C9**	3	186,2	9,21
Seq4F10*	3	201,4	7,86
Ah229**	4	72,9	9,28
Leg156*	4	77,5	7,19
gi-832**	4	89,8	10,06
PM49*	4	90,6	8,40
RN5H2*	4	92,5	8,92
Ah4_26**	5	82,5	10,36
PM45*	5	83,0	7,70
Ah20*	5	83,5	8,48
TC6E1*	5	84,0	8,41
PM36*	5	84,0	8,41
PM65*	5	86,2	8,71
TC2B1*	5	86,7	7,89
TC3A12**	5	92,7	14,34
TC3G1_2**	7	35,0	11,88
TC9H8**	7	36,7	9,07
PM204**	7	39,3	8,89
PM243**	7	45,5	12,89
Leg149**	7	56,8	8,27
Leg24**	- ^c	-	7,97
PM50*	- ^c	-	7,25
TC11F12**	- ^c	-	11,27
TC9E8**	- ^c	-	11,23

^aMarcadores significativamente associados a 5% (*) e a 1% (**).

^bPorcentagem da variação fenotípica explicada pelo marcador.

^cMarcadores que não mapearam.

Usando-se o método de permutações, o valor mínimo de LOD score para considerar válidos os QTLs para a característica número de sementes por planta foi estimado em 4,1, pela média de três repetições de 1000 permutações. Nenhum QTL foi encontrado com LOD scores iguais ou superiores a esse valor pelo método de mapeamento por intervalo, mas dois QTLs foram identificados pelo método de mapeamento por intervalo composto (Figura 5). Esses QTLs foram localizados no grupo 2, posição 22,2 cM, próximo ao marcador Leg128, com LOD 4,22 e no grupo 5, posição 92,8 entre TC3A12 e RM9A3, com LOD estimado em 5,13. O marcador TC3A12, do grupo 5, também foi significativamente ($p < 0,01$) associado a essa característica pelo método de marcas simples (Tabela 5).

Considerando-se o valor mínimo de LOD de 4,1, oito QTLs foram identificados pelo método de mapeamento por intervalo múltiplo (Tabela 6, Figura 4). Desses, sete estão próximos ou nos mesmos grupos de ligação de QTLs detectados pelos métodos de marcas simples e de mapeamento por intervalo composto. Além disso, foi identificado um QTL no grupo 9, para o qual nenhum QTL havia sido detectado pelos demais três métodos utilizados.

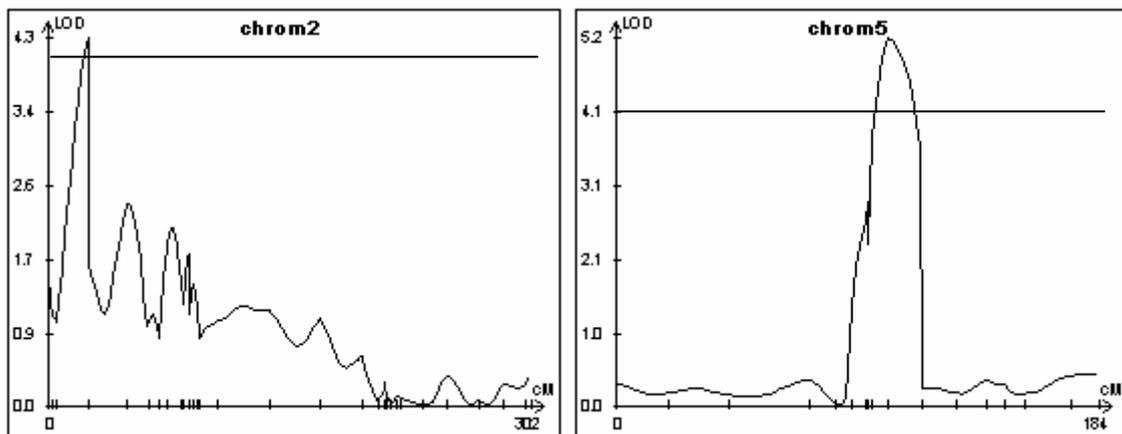


Figura 5 – QTLs identificados para a característica número de sementes por planta, pelo método de mapeamento por intervalo composto. São apresentados apenas os dois grupos de ligação nos quais foram mapeados QTLs com valores de LOD score maiores do que 4,1. Cada pico acima da linha horizontal, que indica o ponto de corte (LOD score=4,1), representa um possível QTL.

Tabela 6 – QTLs detectados para a característica número de sementes por planta, pelo método de mapeamento por intervalo múltiplo. Os marcadores S1A36RGA_B e S1A36RGA_D são dominantes (RGA); os demais são codominantes.

QTL	GL ^a	Intervalo ^b	Posição ^c	LOD ^d	a ^e	d ^e	a (%) ^f	d (%) ^f	d/a ^g	Ação gênica ^h
<i>nsem1</i>	1	Leg304 / Ah745	93,4	9,38	61,55	43,70	4,4	-1,4	0,71	D
<i>nsem2.1</i>	2	Leg20M-Gm / RN0-729	117,8	14,93	180,55	-13,07	29,3	0,0	0,07	A
<i>nsem2.2</i>	2	RN0-729 / Ah11	122,9	8,37	-119,66	49,01	3,7	2,6	0,41	A
<i>nsem4.1</i>	4	S1A36RGA_B / S1A36RGA_D	2,0	7,55	-71,10	145,65	7,9	15,0	2,01	SD
<i>nsem4.2</i>	4	Ah408 / Leg62	130,4	6,01	40,39	-90,63	-1,1	-1,8	2,24	SD
<i>nsem4.3</i>	4	TC1H4 / RN12E1	147,1	6,12	-118,07	30,62	13,1	0,8	0,26	A
<i>nsem5</i>	5	RM9A3 / AC3D7	109,8	9,53	41,34	-158,01	-3,3	26,8	3,82	SD
<i>nsem9</i>	9	Leg199 / AC2C2	67,3	4,24	-44,44	-5,62	3,5	0,0	0,12	A

^aGrupo de ligação a que pertence o QTL.

^bMarcadores adjacentes ao QTL.

^cPosição no grupo de ligação, em cM, em que foi observado o maior valor de LOD score.

^dLOD score máximo observado.

^eEfeitos aditivo (a) e de dominância (d). O efeito aditivo refere-se ao efeito da substituição dos alelos de *A. duranensis* pelos alelos de *A. stenosperma*. Valor negativo indica diminuição do valor da característica. Para o efeito de dominância, valor positivo indica que o heterozigoto possui, em média, valor fenotípico superior à média dos dois homozigotos e valor negativo indica que o heterozigoto possui, em média, valor fenotípico inferior à média dos dois homozigotos.

^fEfeitos aditivo (a) e de dominância (d), em porcentagem da variação fenotípica total.

^{g,h}Ação gênica, calculada por $[a] / [d]$ e interpretada, de acordo com Edwards *et al.*, 1987, como $d/a < 0,55$, ação aditiva (A) ou de dominância parcial; $0,55 < d/a < 1,25$, dominante (D) e $d/a > 1,25$, ação sobredominante (SD).

Alelos com grandes efeitos aditivos foram identificados para esses oito QTLs (Tabela 6). Os alelos Leg20M-Gm_156pb (*psem2.1*), proveniente do progenitor masculino (*A. stenosperma*), e RN12E1_146pb (*psem4.3*), proveniente do progenitor feminino (*A. duranensis*), são particularmente interessantes, pois apresentaram efeitos aditivos significativos (29,3% e 13,1% da variância aditiva, respectivamente) e quase nenhum efeito de dominância. O QTL localizado no grupo 5 foi identificado pelos métodos de marcas simples e de mapeamento por intervalo composto e por intervalo múltiplo. Esse QTL apresentou um efeito aditivo reduzido (-3,3%), mas um efeito de dominância negativo significativo (-158,01 ou 26,8% da variância total), indicando que plantas homozigotas para o alelo proveniente de *A. duranensis* (RM9A3_144pb) deverão produzir mais sementes do que plantas heterozigotas ou homozigotas para o alelo proveniente de *A. stenosperma* (RM9A3_155pb). Novamente nenhum efeito de epistasia foi identificado entre esses oito QTLs.

Como esperado, o presente trabalho mostrou que número de sementes por planta também possui herança quantitativa em *Arachis*. Os números de QTLs que controlam essa característica têm variado de três a nove em diferentes espécies de leguminosas (Zhang *et al.*, 2004; Timmerman-Vaughan *et al.*, 2005; Blair *et al.*, 2006; Herrmann *et al.*, 2006).

3.2.3. QTLs para resistência à mancha preta (*Cercosporidium personatum*)

Oito marcadores foram significativamente associados à resistência ao fungo causador da mancha preta no amendoim (*C. personatum*), pela análise de marcas simples (Tabela 7). Esses marcadores encontram-se localizados em três regiões do mapa, nos grupos 2, 3 e 5. Além disso, quatro marcadores que não mapearam em nenhum dos grupos deram associações significativas (Leg7-Gm, gi-370, Leg24 e PM50). As porcentagens da variação explicada (R^2) por esses oito marcadores variaram de 7,19% a 11,58%.

Tabela 7 – Marcadores significativamente associados à resistência ao fungo *Cercosporidium personatum*, causador da mancha preta, pela análise de marcas simples. Todos os marcadores são codominantes.

Marcador ^a	Grupo de ligação	Posição (cM)	R ² (%) ^b
Leg69*	2	275,8	7,71
RI2A6*	2	279,3	9,50
Leg181*	3	77,5	8,11
RM9A3*	5	104,3	9,48
Leg7-Gm*	- ^c	-	7,32
gi-370*	- ^c	-	7,19
Leg24**	- ^c	-	11,58
PM50*	- ^c	-	9,59

^aMarcadores significativamente associados a 5% (*) e a 1% (**).

^bPorcentagem da variação fenotípica explicada pelo marcador.

^cMarcadores que não mapearam.

Usando-se o método de permutações, o valor mínimo de LOD score para considerar válidos os QTLs associados à resistência à mancha preta foi estimado em 51,8 pela média de três repetições de 1000 permutações. As razões para o alto valor de ponto de corte estimado por permutações estão possivelmente ligadas ao pequeno contraste entre as 93 plantas F2 avaliadas e ao fato de que a distribuição fenotípica dessa característica mostrou-se fortemente distorcida em favor da resistência ao fungo. Além disso, a presença de marcadores distorcidos tem sido apontada como uma das causas da obtenção de altos valores no teste de permutações (Westerbergh & Doebley, 2004).

Considerando-se o limite mínimo de 51,8, nenhum QTL pôde ser significativamente identificado por nenhum dos três métodos de mapeamento por intervalo. Apesar disso, três QTLs foram identificados pelo método de mapeamento por intervalo composto com valores de LOD superiores a 44, sendo dois no grupo 3 (entre os marcadores S1A37RGA_A e S1A36RGA_C, com LOD 45,7 e entre AsRGA_15C e Leg43, com LOD 47,8) e um no grupo 4, entre S1A36RGA_B e S1A36RGA_D, com LOD 44,9 (Figura 6). Os QTLs do grupo 3 entre AsRGA_15C e Leg43 e do grupo 4 também foram identificados pelo método de mapeamento por intervalo múltiplo, com valores de LOD de 36,3 e 12,5, respectivamente (Tabela 8).

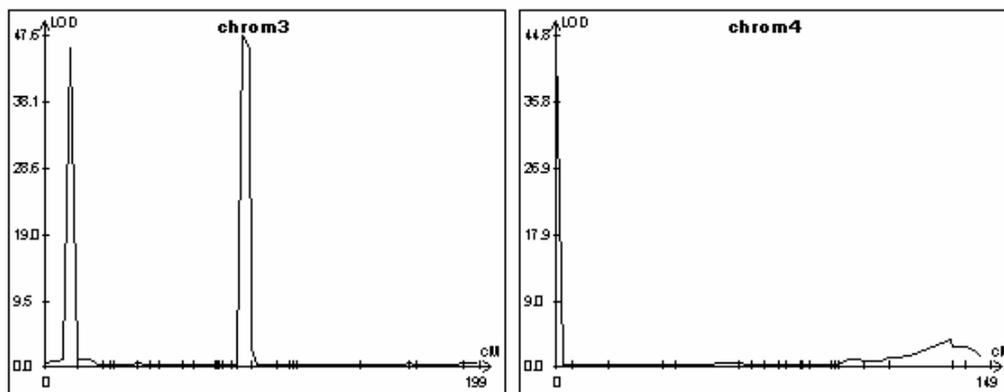


Figura 6 – Mapeamento de QTLs para resistência ao fungo *Cercosporidium personatum*, causador da mancha-preta em *Arachis* spp, pelo método de mapeamento por intervalo composto.

Os efeitos fenotípicos desses dois QTLs foram estimados pela análise de variância através do método de mapeamento por intervalo múltiplo (Tabela 8). Os dois QTLs explicaram, no total, 63,3% da variância aditiva, sendo que o QTL *Cp3* apresentou efeito aditivo de 52,7%. Esse QTL foi localizado próximo a um marcador do tipo RGA (AsRGA_15), enquanto o QTL *Cp4* foi identificado em uma região na qual mapearam quatro marcadores RGA, com um valor de LOD máximo na posição 2,0 cM, entre os marcadores S1A36RGA_B e S1A36RGA_D. O mapeamento de QTLs adjacentes a marcadores RGA confere maior suporte à hipótese de que esses QTLs são verdadeiros, embora não tenham atingido o valor de ponto de corte estimado. Poucos estudos têm sido publicados referentes ao controle genético da resistência a fungos em amendoim e esses relatos apresentam resultados divergentes, variando de dois a cinco genes recessivos, enquanto outros estudos reportam efeito predominante aditivo para essa característica (revisto por Dwivedi *et al.*, 2003). Além disso, nenhum desses genes foi mapeado até o momento. Os resultados obtidos no presente trabalho indicam que há, pelo menos, dois genes controlando essa característica, sendo um de grande efeito e que, nesse caso, explicou 52,7% da variação fenotípica aditiva. Famílias F3 estão sendo geradas para essa população e para uma população tetraplóide que inclui o amendoim cultivado como um dos progenitores, para avaliação da resistência a esse fungo, visando à validação dos resultados aqui obtidos.

Tabela 8 – QTLs detectados para a característica resistência ao fungo *Cercosporidium personatum*, causador da mancha preta em *Arachis* spp, pelo método de mapeamento por intervalo múltiplo.

QTL	GL ^a	Intervalo ^b	Posição ^c	LOD ^d	a ^e	d ^e	a (%) ^f	d (%) ^f	d/a ^g	Ação gênica ^h
<i>Cp3</i>	3	AsRGA_15C / Leg43	90,2	36,3	-5,6640	-5,7053	52,7	32,3	1,01	D
<i>Cp4</i>	4	S1A36RGA_B / S1A36RGA_D	2,0	12,5	0,6952	-0,6508	10,6	3,5	0,94	D

^aGrupo de ligação a que pertence o QTL.

^bMarcadores adjacentes ao QTL.

^cPosição no grupo de ligação, em cM, em que foi observado o maior valor de LOD score.

^dLOD score máximo observado.

^eEfeitos aditivo (a) e de dominância (d). O efeito aditivo refere-se ao efeito da substituição dos alelos de *A. duranensis* pelos alelos de *A. stenosperma*. Valor negativo indica diminuição do valor da característica, que, nesse caso, significa um aumento da resistência à doença. Para o efeito de dominância, valor positivo indica que o heterozigoto possui, em média, valor fenotípico superior à média dos dois homozigotos, sendo mais susceptível, e valor negativo indica que o heterozigoto possui, em média, valor fenotípico inferior à média dos dois homozigotos (mais resistente).

^fEfeitos aditivo (a) e de dominância (d), em porcentagem da variação fenotípica total.

^{g,h}Ação gênica, calculada por $[a] / [d]$ e interpretada, de acordo com EDWARDS *et al.*, 1987, como $d/a < 0,55$, ação aditiva (A) ou de dominância parcial; $0,55 < d/a < 1,25$, dominante (D) e $d/a > 1,25$, ação sobredominante (SD).

3.2.4. Co-localização de QTLs

Considerando-se as três características, 20 QTLs foram detectados em oito dos dez grupos de ligação, pelo método de mapeamento por intervalo múltiplo (Figura 4). Apenas nos grupos 8 e 10 nenhum QTL foi identificado. Em duas regiões do genoma, QTLs para diferentes características foram localizados no mesmo intervalo de marcadores ou em intervalos adjacentes. No grupo 4, os QTLs *nsem4.1* (número de sementes por planta) e *Cp4* (para resistência à mancha preta) foram mapeados no mesmo intervalo de marcadores, entre S1A36RGA_B e S1A36RGA_D (Figura 4, Tabelas 6 e 8). Em ambos os casos, o alelo favorável foi proveniente de *A. stenosperma*, confirmando a correlação significativa e negativa identificada (Tabela 2). Portanto, esses QTLs seriam úteis em um programa de seleção assistida, uma vez que o uso desses marcadores para seleção de plantas mais resistentes levaria também à seleção de plantas que produzem mais sementes. Os dados são insuficientes para sugerir que isso seja devido à presença de genes com efeitos pleiotrópicos.

Os QTLs *nsem5* e *psem5* (peso de sementes) foram identificados em intervalos adjacentes no grupo 5, entre RM9A3 e AC3D7 e entre AC3D7 e Ah6-125, respectivamente (Figura 4). Nesse caso, os alelos favoráveis nos dois QTLs foram provenientes de *A. duranensis*, possibilitando que esses marcadores também fossem utilizados em uma eventual seleção assistida.

3.2.5. Genes candidatos

Dentre os 19 locos associados à característica peso de sementes pelo método de mapeamento por intervalo múltiplo (Tabela 4), 9 deram similaridades significativas na análise de BlastX ($e < 10^{-10}$), sendo quatro a proteínas com funções ainda não conhecidas (ANEXO 3). O QTL *psem9.1* foi localizado entre os marcadores gi-1107 e Lec-1, ambos com similaridades a um precursor de lectina que se liga a galactose (“galactose-binding lectin precursor”). Portanto, esses marcadores podem estar envolvidos na síntese de proteínas em sementes, com a possibilidade de fazerem parte dos genes de interesse e não apenas de estarem ligados a eles (genes candidatos). As demais similaridades não têm, aparentemente, nenhuma relação com vias metabólicas relacionadas ao desenvolvimento de sementes, embora essa seja uma característica bastante complexa.

A característica número de sementes por planta é também bastante complexa. Dos 15 locos associados a QTLs que a controlam (Tabela 6), apenas cinco deram similaridades significativas a genes já descritos (ANEXO 3), mas nenhum possivelmente relacionado à produção de sementes.

Os quatro locos associados à resistência ao fungo *C. personatum* pelo método de mapeamento por intervalo múltiplo (Tabela 8) deram similaridades significativas. Três desses marcadores são RGAs, do tipo NBS-TIR (Bertioli *et al.*, 2003) e, portanto, podem ser os genes de resistência em questão ou parte deles. O marcador Leg43 deu similaridade a uma dioxigenase (“homogentisate 1,2 dioxigenase”) e, provavelmente não está envolvido em respostas à infecção por patógenos.

Os genes candidatos identificados nesse estudo precisam ser validados em outros experimentos. No entanto, o uso de marcadores gênicos, e não apenas de marcadores ligados a esses genes, permitirá uma maior eficiência no controle da região do genoma a ser introgridida.

4. CONCLUSÕES

O mapa desenvolvido no presente estudo é o primeiro mapa baseado em marcadores microssatélites publicado para *Arachis*. Além disso, por incluir marcadores âncoras e microssatélites desenvolvidos a partir de bibliotecas de cDNA e de bibliotecas genômicas digeridas com enzimas sensíveis a metilação, apresenta vários marcadores gênicos mapeados (136 dos 248 marcadores mapeados). Esses marcadores são de particular utilidade para estudos de genômica comparativa e, por estarem localizados na porção expressa do genoma, apresentam maior probabilidade de estarem ligados a genes de interesse ou de fazerem parte desses genes.

Uma vez que *A. hypogaea* possui uma base genética estreita, espécies silvestres de *Arachis* constituem uma importante fonte de genes que controlam características de interesse para o amendoim. A disponibilidade de mapas genéticos baseados em marcadores codominantes e moderadamente saturados irá representar um considerável aumento da capacidade de mapear genes úteis e de implementar a seleção assistida por marcadores nos programas de melhoramento do amendoim.

No presente trabalho foram também mapeados os primeiros QTLs para componentes da produtividade em amendoim e para resistência ao fungo *Cercosporidium personatum*, causador da mancha-preta, que é uma das principais doenças que atacam o amendoim. Alguns desses marcadores apresentaram similaridades significativas com genes envolvidos nessas vias metabólicas e são, portanto, genes candidatos, com potencial uso em mapeamento de associação ou mesmo para clonagem desses genes.

A validação dos grupos de ligação obtidos, tanto em termos da ordem dos marcadores quanto das estimativas das frações de recombinação, encontra-se em andamento, em outras duas populações de mapeamento: uma F₂ diplóide, gerada pelo cruzamento de duas espécies com genoma BB (*A. ipaënsis* x *A. magna*) e uma população F₂ tetraplóide, obtida do cruzamento entre um acesso de *A. hypogaea* (cv. IAC Runner 886) e uma planta anfidiplóide, resultante do cruzamento entre duas espécies diplóides (*A. duranensis* – genoma AA x *A. ipaënsis* – genoma BB) tratadas com colchicina. Essas mesmas populações estão sendo utilizadas para validação dos QTLs identificados. Para isso, famílias F3 estão sendo avaliadas em diferentes

localidades com o objetivo de obtenção de estimativas mais robustas, visando ao uso dessas informações nos programas de melhoramento do amendoim. O desenvolvimento de mapas genéticos para espécies tetraplóides e diplóides de *Arachis* com genomas AA e BB, baseados em microssatélites e marcadores âncoras, que são codominantes e transferíveis entre espécies, possibilitará a construção de um mapa de referência para *Arachis* e o mapeamento de QTLs em diferentes “backgrounds” genéticos. Adicionalmente, o mapa desenvolvido nesse estudo é o único mapa publicado para *Arachis* construído com marcadores derivados de seqüências que são, todas, publicamente disponíveis no GenBank. Portanto, pesquisadores de diferentes laboratórios poderão usar esses mesmos marcadores para o desenvolvimento de mapas, usando outras populações de mapeamento, contribuindo, assim, para a construção de um mapa de referência para *Arachis*.

Finalmente, o mapeamento de marcadores de cópias únicas e transferíveis entre diferentes espécies de leguminosas (âncoras) possibilitará a inclusão de *Arachis* em estudos de genômica comparativa de leguminosas. Estudos de sintenia têm sido descritos para gramíneas e *Brassica* spp., mas, até o momento, muito pouco tem sido feito em leguminosas. Resultados preliminares mostraram que vários grupos de ligação apresentaram marcadores em comum e na mesma ordem nos genomas de *Arachis*, *Lotus* e *Medicago*. *A. hypogaea* possui um genoma cerca de dez vezes maior do que os genomas de *Lotus* e de *Medicago*, cujos genomas estão sendo seqüenciados. Portanto, esses marcadores gênicos conservados entre as espécies irão facilitar sobremaneira a identificação de genes ortólogos no amendoim.

CONSIDERAÇÕES FINAIS

O presente trabalho descreve o desenvolvimento de ferramentas genéticas e genômicas, visando a utilização de espécies silvestres de *Arachis* nos programas de melhoramento genético do amendoim. Com isso, foi dado o primeiro passo para implementação de seleção assistida por marcadores, o que irá facilitar a introgressão de genes de interesse, presentes nessas espécies, para o amendoim cultivado. Nesse sentido, os principais resultados alcançados foram:

- Construção de duas bibliotecas genômicas enriquecidas para seqüências microssatélites (TC/AG e AC/TG) de amendoim;
- Seqüenciamento de 1152 clones positivos resultantes destas duas bibliotecas;
- Desenvolvimento de 271 novos marcadores microssatélites para *Arachis*, a partir de bibliotecas genômicas, de ESTs e por busca de seqüências publicadas no GenBank;
- Identificação de 66 novos marcadores microssatélites polimórficos para *A.hypogaea*, o que constitui um grande aumento na disponibilidade desses marcadores para análises genéticas do amendoim;
- Marcadores derivados de bibliotecas genômicas mostraram-se, em geral, mais polimórficos para *A. hypogaea* do que os derivados de bibliotecas de cDNA, embora, para o cruzamento interespecífico (*A. duranensis* x *A. stenosperma*), porcentagens idênticas de polimorfismo tenham sido detectados por marcadores derivados das duas bibliotecas;
- A análise de relações genéticas entre acessos de *A. hypogaea* mostrou a formação de três grupos de similaridade e levantou dúvidas quanto à classificação taxonômica das variedades *peruviana* e *aequatoriana*.
- O primeiro mapa baseado em marcadores microssatélites foi desenvolvido e publicado para *Arachis*;

- Por incluir marcadores âncoras e microssatélites desenvolvidos a partir de bibliotecas digeridas com enzimas sensíveis a metilação, mais da metade dos marcadores mapeados são gênicos;
- Os primeiros QTLs para componentes da produtividade e para resistência ao fungo *Cercosporidium personatum*, causador da mancha-preta, foram mapeados em *Arachis*;
- O mapeamento de marcadores-âncoras possibilita a inclusão de *Arachis* em estudos de genômica comparativa de leguminosas, o que permitirá que esta cultura se beneficie dos avanços em genética e genômica de outras leguminosas mais estudadas.

REFERÊNCIAS BIBLIOGRÁFICAS

- AARTS, M.G., TE LINTEL HEKKERT, B.; HOLUB, E.B.; BEYNON, J.L.; STIEKEMA W.J.; PEREIRA, A. Identification of R-gene homologous DNA fragments genetically linked to disease resistance loci in *Arabidopsis thaliana*. **Molecular Plant-Microbe Interactions**, v.11, p.251-258, 1998.
- ALMEIDA, A.M. Principais doenças fúngicas do amendoim. In: **Anais da II Reunião Itinerante de Fitossanidade do Instituto Biológico**, Marília, SP. 1999. 115 p.
- ALTSCHUL, S.F.; GISH, W.; MILLER, W.; MYERS, E.W.; LIPMAN, D.J. Basic local alignment search tool. **Journal of Molecular Biology**, v.215, p.403-410, 1990.
- BACKMAN, P.A.; CRAWFORD, M.A. Relationship between yield loss and severity of early and late leafspot diseases of peanut (*Arachis hypogaea* L.). **Phytopathology**, v.74, p.1101-1103, 1984.
- BARZEN, E.; MECHELKE, W.; RITTER, E.; SCHULTE-KAPPERT, E.; SALAMINI, F. An extended map of the sugar beet genome containing RFLP and RFLP loci. **Theoretical and Applied Genetics**, v.90, p.189-193, 1995.
- BENJAMIN, Y.; HOCHBERG, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. **Journal of the Royal Statistical Society**, v.57, p.289-300, 1995.
- BENNET, M.D.; LEITCH, I.J. Nuclear DNA amounts in Angiosperms: Progress, problems and prospects. **Annals of Botany**, v.95, p.45-90, 2005.
- BERRY, S.T.; LEON, A.J.; HANFREY, C.C.; CHALLIS, P.; BURKHOLZ, A.; BARNES, S.R.; RUFENER, G.K.; LEE, M.; CALIGARI, P.D.S. Molecular marker analysis of *Helianthus annuus* L.: 2. Construction of an RFLP linkage map for cultivated sunflower. **Theoretical and Applied Genetics**, v.91, p.195-199, 1995.
- BERTIN, P.; GREGOIRE, D.; MASSART, S.; de FROIDMONT, D. Genetic diversity among European cultivated spelt revealed by microsatellites. **Theoretical and Applied Genetics**, v.102, p.148-156, 2001.
- BERTIOLI, D.J.; LEAL-BERTIOLI, S.C.M.; LION, M.B.; SANTOS, V.L.; PAPPAS Jr, G.; CANNON, S.B.; GUIMARÃES, P.M. A large scale analysis of resistance gene homologues in *Arachis*. **Molecular Genetics and Genomics**, v.270, p.34-45, 2003.
- BLAIR, M.W.; IRIARTE, G.; BEEBE, S. QTL analysis of yield traits in an advanced backcross population derived from a cultivated Andean x wild common bean (*Phaseolus vulgaris* L.) cross. **Theoretical and Applied Genetics**, v.112, p.1149-1163, 2006.
- BONIERBALE, M.W.; PLAISTED, R.L.; TANKSLEY, S.D. RFLP maps based on common set of clones reveal modes of chromosomal evolutions in potato and tomato. **Genetics**, v.120, p.1095-1103, 1988.
- BOWCOCK, A.M.; RUIZ-LINARES, A.; TOMFOHRDE, J.; MINCH, E.; KIDD, J.R.; CAVALLI-SFORZA, L.L. High resolution of human evolutionary trees with polymorphic microsatellites. **Nature**, v.368, p.455-457, 1994.
- BRAVO, J.P.; HOSHINO, A.A.; ANGELICI, C.M.L.C.D.; LOPES, C.R.; GIMENES, M.A. Analysis of genetic variability in germplasm of *Arachis* section *Arachis* by using microsatellites. **Genetics and Molecular Biology**, 2006 (no prelo).
- BRONDANI, C.; RANGEL, P.H.N.; BRONDANI, R.P.V.; FERREIRA, M.E. QTL mapping and introgression of yield-related traits from *Oryza glumaepatula* to cultivated rice (*Oryza sativa*) using microsatellite markers. **Theoretical and Applied Genetics**, v.104, p.1192-1203, 2002.
- BUROW, M.D.; SIMPSON, C.E.; PATERSON, A.H.; STARR, J.L. Identification of peanut (*Arachis hypogaea* L.) RAPD markers diagnostic of root-knot nematode

- (*Meloidogyne arenaria* (Neal) Chitwood) resistance. **Molecular Breeding**, v.2, p.369-379, 1996.
- BUROW, M.D.; SIMPSON, C.E.; STARR, J.L.; PATERSON, A.H. Transmission genetics of chromatin from a synthetic amphidiploid to cultivated peanut (*Arachis hypogaea* L.): Broadening the gene pool of a monophyletic polyploid species. **Genetics**, v.159, p.823-837, 2001.
- BURSTIN, J.; DENIOT, G.; POTIER, J.; WEINACHTER, C.; AUBERT, G.; BARANGER, A. Microsatellite polymorphism in *Pisum sativum*. **Plant Breeding**, v.120, p.311-317, 2001.
- CANNON, S.B.; STERCK, L.; ROMBAUTS, S.; SATO, S.; CHEUNG, F.; GOUZY, J.P.; WANG, X.; MUDGE, J.; VASDEWANI, J.; SCHIEX, T.; SPANNAGL, M.; MONAGHAN, E.; NICHOLSON, C.; HUMPHRAY, S.J.; SCHOOF, H.; MAYER, K.F.X.; ROGERS, J.; QUÉTIER, F.; OLDROYD, G.E.; DEBELLÉ, F.; COOK, D.R.; ROE, B.A.; TOWN, C.D.; TABATA, S.; VAN DE PEER, Y.; YOUNG, N.D. Legume evolution viewed through the *Medicago truncatula* and *Lotus japonicus* genomes. **Proceedings of the National Academy of Sciences of the United States of America**, v.103, p.14959-14964, 2006.
- CARDLE, L.; RAMSAY, L.; MILBOURNE, D.; MACAULAY, M.; MARSHALL, D.; WAUGH, R. Computational and experimental characterization of physically clustered simple sequence repeats in plants. **Genetics**, v.156, p.847-854, 2000.
- CARNEIRO, M.S.; VIEIRA, M.L.C. Mapas genéticos em plantas. **Bragantia**, v.61, p.89-100, 2002.
- CHABANE, K.; ABLETT, G.A.; CORDEIRO, G.M.; VALKOUN, J.; HENRY, R.J. EST versus genomic derived microsatellite markers for genotyping wild and cultivated barley. **Genetic Resources and Crop Evolution**, v.52, p.903-909, 2005.
- CHAVARRIAGA-AGUIRRE, P.; MAYA, M.M.; BONIERBALE, M.W.; KRESOVICH, S.; FREGENE, M.A.; TOHME, J.; KOCHERT, G. Microsatellites in Cassava (*Manihot esculenta* Crantz): discovery, inheritance and variability. **Theoretical and Applied Genetics**, v.97, p.493-501, 1998.
- CHEN, C.; ZHOU, P.; CHOI, Y.A.; HUANG, S.; GMITTER JR, F.G. Mining and characterizing microsatellites from citrus ESTs. **Theoretical and Applied Genetics**, v.112, p.1248-1257, 2006.
- CHO, Y.G.; ISHII, T.; TEMNYKH, S.; CHEN, X.; LIPOVICH, L.; McCOUCH, S.R.; PARK, W.D.; AYRES, N.; CARTINHO, S. Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). **Theoretical and Applied Genetics**, v.100, p.713-722, 2000.
- CHOI, K.; BUROW, M.D.; CHURCH, G.; BUROW, G.; PATERSON, A.H.; SIMPSON, C.E.; STARR, J.L. Genetics and mechanism of resistance to *Meloidogyne arenaria* in peanut germplasm. **Journal of Nematology**, v.31, p.283-290, 1999.
- CHOUMANE, W.; WINTER, P.; BAUM, M.; KAHL, G. Conservation of microsatellite flanking sequences in different taxa of Leguminosae. **Euphytica**, v.138, p.239-245, 2004.
- CHURCH, G.; SIMPSON, C.E.; BUROW, M.D.; PATERSON, A.H.; STARR, J.L. Use of RFLP markers for identification of individuals homozygous for resistance to *Meloidogyne arenaria* in peanut. **Nematology**, v.2, p.575-580, 2000.
- CHURCHILL, G.A.; DOERGE, R.W. Empirical threshold values for quantitative trait mapping. **Genetics**, v.138, p.963-971, 1994.
- COELHO, A.S.G.; SILVA, H.D. Construção de Mapas Genéticos e Mapeamento de QTL's. ESALQ, Piracicaba, 2002. 77 p. (apostila).

- 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**, v.142, p.169-196, 2005.
- COLLINS, N.; PARK, R.; SPIELMEYER, W.; ELLIS, J.; PRYOR, A.J. Resistance gene analogs in barley and their relationship to rust resistance genes. **Genome**, v.44, p.375-381, 2001.
- COMPANY, M., STALKER, H.T.; WYNNE, J.C. Cytology and leafspot resistance in *Arachis hypogaea* X wild species hybrids. **Euphytica**, v.31, p.885-893, 1982.
- CONAB - Companhia Nacional de Abastecimento, 2006. Disponível em <http://www.conab.gov.br/>. Acesso em 10 jul 2006.
- CORDEIRO, G.M.; CASU, R.; McINTYRE, C.L.; MANNERS, J.M.; HENRY, R.J. Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to erianthus and sorghum. **Plant Science**, v.160, p.1115-1123, 2001.
- CRESTE, S.; TULMANN NETO, A.; FIGUEIRA, A. Detection of single sequence repeat polymorphisms in denaturing polyacrylamide sequencing gels by silver staining. **Plant Molecular Biology Reporter**, v.19, p.299-306, 2001.
- CRUZ, C.D. GQMOL versão 2004.2.1. Disponível em <http://www.ufv.br/dbg/gqmol/gqmol.htm>
- CSANÁDI, G.; VOLLMANN, J.; STIFT, G.; LELLEY, T. Seed quality QTLs identified in a molecular map of early maturing soybean. **Theoretical and Applied Genetics**, v.103, p.912-919, 2001.
- De WIT, P.L.G.M. Fungal avirulence genes and plant resistance genes: unraveling the molecular basis of gene-for-gene interactions. **Advances in Botanical Research**, v.21, p.147-185, 1995.
- DIWAN, N.; BOUTON, J.H.; KOCHERT, G.; CREGAN, P.B. Mapping of simple sequence repeat (SSR) DNA markers in diploid and tetraploid alfafa. **Theoretical and Applied Genetics**, v.101, p.165-172, 2000.
- DIXON, M.S.; HATZIXANTHIS, K.; JONES, D.A.; HARRISON, K.; JONES, J.D. The tomato Cf-5 disease resistance gene and six homologs show pronounced allelic variation in leucine-rich repeat copy number. **Plant Cell**, v.10, p.1915-1925, 1998.
- DOERGE, R. W.; CHURCHILL, G.A. Permutation tests for multiple loci affecting a quantitative character. **Genetics**, v.142, p.285-294, 1996.
- DOGANLAR, S.; FRARY, A.; TANKSLEY, S.D. The genetic basis of seed-weight variation: tomato as a model. **Theoretical and Applied Genetics**, v.100, p.1267-1273, 2000.
- DONALD, T.M.; PELLERONE, F.; ADAM-BLONDON, A.F.; BOUQUET, A. Identification of resistance gene analogs linked to a powdery mildew resistance locus in grapevine. **Theoretical and Applied Genetics**, v.104, p.610-618, 2002.
- DWIVEDI, S.L.; CROUCH, J.H.; NIGAM, S.N.; FERGUSON, M.E.; PATERSON, A.H. Molecular breeding of groundnut for enhanced productivity and food security in the semi-arid tropics: Opportunities and challenges. **Advances in Agronomy**, v.80, p.153-221, 2003.
- DWIVEDI, S.L.; BERTIOLI, D.J.; CROUCH, J.H.; VALLS, J.F.M.; UPADHYAYA, H.D.; FÁVERO, A.P.; MORETZSOHN, M.C.; PATERSON, A.H. Peanut Genetics and Genomics: Toward Marker-assisted Genetic Enhancement in Peanut (*Arachis hypogaea* L.). In: KOLE, C. (Ed.). **Oilseeds**. Series: Genome Mapping and Molecular Breeding in Plants, Vol. 2, Springer, 2006. 302p.

- EDWARDS, M.D.; STUBER, C.W.; WENDEL, J.F. Molecular-Marker-Facilitated Investigations of Quantitative Trait Loci in Maize. I. Numbers, Genomic Distribution and Types of Gene Action. **Genetics**, v.116, p.113-125, 1987.
- ELLEGREN, H. Microsatellites: simple sequences with complex evolution. **Nature reviews**, v.5, p.435-445, 2004.
- EUJAYL, I.; SLEDGE, M.K.; WANG, L.; MAY, G.D.; CHEKHOVSKIY, K.; ZWONITZER, J.C.; MIAN, M.A.R. *Medicago truncatula* EST-SSRs reveal cross-species genetic markers for *Medicago* spp. **Theoretical and Applied Genetics**, v.108, p.414-422, 2004.
- EUJAYL, I.; SORRELLS, M.E.; BAUM, M.; WOLTERS, P.; POWELL, W. Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat. **Theoretical and Applied Genetics**, v.104, p.399-407, 2002.
- FAO. Food and Agriculture Organization of the United Nations (2004). Disponível em <http://faostat.fao.org/faostat/collections?subset=agriculture>. Acesso em 10 jul 2006.
- FATOKUN, C.A.; MENANCIO-HAUTEA, D.I.; DANESH, D.; YOUNG, N.D. Evidence for orthologous seed-weight genes in cowpea and mung bean based on RFLP mapping. **Genetics**, v.132, p.841-846, 1992.
- FÁVERO, A. P. Cruzabilidade entre espécies silvestres de *Arachis* visando à introgressão de genes de resistência a doenças no amendoim cultivado. Tese (Doutorado) - Escola Superior de Agricultura 'Luiz de Queiroz', Universidade de São Paulo, Piracicaba, 2004. 165 p.
- FÁVERO, A.P.; GUIMARÃES, P.M.; BERTIOLI, D.; LEAL-BERTIOLI, S.C.M. **Avaliação para resistência a mancha preta e ferrugem em híbridos entre o amendoim e espécies silvestres de *Arachis***. Boletim de Pesquisa & Desenvolvimento, n.96, Brasília: Embrapa Recursos Genéticos e Biotecnologia, 17p., 2005.
- FÁVERO, A.P.; SIMPSON, C.E.; VALLS, J.F.M.; VELLO, N.A. Study of the evolution of cultivated peanut through crossability studies among *A. ipaënsis*, *A. duranensis*, and *A. hypogaea*. **Crop Science**, v.46, p.1546-1552, 2006.
- FERGUSON, M.E.; BRAMEL, P.J.; CHANDRA, S. Gene Diversity among Botanical Varieties in Peanut (*Arachis hypogaea* L.). **Crop Science**, v.44, p.1847-1854, 2004a.
- FERGUSON, M.E.; BUROW, M.D.; SCHULZE, S.R.; BRAMEL, P.J.; PATERSON, A.H.; KRESOVICH, S.; MITCHELL, S. Microsatellite identification and characterization in peanut (*A. hypogaea* L.). **Theoretical and Applied Genetics**, v.108, p.1064-1070, 2004b.
- FERREIRA, M.E.; GRATTAPAGLIA, D. **Introdução ao uso de marcadores moleculares em análise genética**. 3^a ed. Brasília: EMBRAPA-CENARGEN, 1998. 220 p.
- FRARY, A.; FULTON, T.M.; ZAMIR, D.; TANKSLEY, S.D. Advanced backcross QTL analysis of a *Lycopersicon esculentum* x *L. pennellii* cross and identification of possible orthologs in the Solanaceae. **Theoretical and Applied Genetics**, v.108, p.485-496, 2004.
- FREDSLUND, J.; MADSEN, L.H.; HOUGAARD, B.K.; SANDAL, N.; STOUGAARD, J.; BERTIOLI, D.; SCHAUSER, L. GeMprospector - Online Design of Cross-Species Genetic Marker Candidates in Legumes and Grasses. **Nucleic Acids Research**, v.34, p.W670 - W675, 2006.

- FREDSLUND, J.; SCHAUSER, L.; MADSEN, L.H.; SANDAL, N.; STOUGAARD, J. PriFi: using a multiple alignment of related sequences to find primers for amplification of homologs. **Nucleic Acids Research**, v.33, p.W516–W520, 2005.
- FREITAS, F.O.; PEÑALOZA, A.P.S.; VALLS, J.F.M. O amendoim contador de história. Brasília: Embrapa Recursos Genéticos e Biotecnologia, 2003. **Documentos**, n.107. 2003. 12p.
- FRIDMAN, E.; CARRARI, F.; LIU, Y-S; FERNIE, A.R.; ZAMIR, D. Zooming In on a Quantitative Trait for Tomato Yield Using Interspecific Introgressions. **Science**, v.305, p.1786 – 1789, 2004.
- GAO, L.F.; JING, R.L.; HUO, N.X.; LI, Y.; LI, X.P.; ZHOU, R.H.; CHANG, X.P.; TANG, J.F.; MA, Z.Y.; JIA, J.Z. One hundred and one new microsatellite loci derived from ESTs (EST-SSRs) in bread wheat. **Theoretical and Applied Genetics**, v.108, p.1392-1400, 2004.
- GAO, L.F.; TANG, J.F.; LI, H.W.; JIA, J.Z. Analysis of microsatellites in major crops assessed by computational and experimental approaches. **Molecular Breeding**, v.12, p.245-261, 2003.
- GARCIA, G.M.; STALKER, H.T.; KOCHERT, G. Introgression analysis of an interspecific hybrid population in peanuts (*Arachis hypogaea* L.) using RFLP and RAPD markers. **Genome**, v.38, p.166-176, 1995.
- GARCIA, G.M.; STALKER, H.T.; SHROEDER, E.; KOCHERT, G. Identification of RAPD, SCAR, and RFLP markers tightly linked to nematode resistance genes introgressed from *Arachis cardenasii* into *Arachis hypogaea*. **Genome**, v.39, p.836-845, 1996.
- GARDNER, M.E.B.; STALKER, H.T. Cytology and Leafspot Resistance of Section *Arachis* Amphidiploids and Their Hybrids with *Arachis hypogaea*. **Crop Science**, v.23, p.1069-1074, 1983.
- GIMENES, M.A.; HOSHINO, A.A.; BARBOSA, A.V.G.; PALMIERI, D.A.; BRAVO, J.P.; LOPES, C.R. Characterization and transferability of microsatellite markers from the cultivated peanut (*Arachis hypogaea* L.). **BMC Plant Biology**, 2006 (no prelo).
- GIMENES, M.A.; LOPES, C.R.; VALLS, J.F.M. Genetic relationships among *Arachis* species based on AFLP. **Genetics and Molecular Biology**, v.25, p.349-353, 2002.
- GODOY, I.J.; MORAES, S.A.; ZANOTTO, M.D.; SANTOS, R.C. Melhoramento do amendoim. In: BORÉM, A. (Ed.). **Melhoramento de espécies cultivadas**, Viçosa, UFV; 1999. p.51-94.
- GONZALO, M.J.; OLIVER, M.; GARCIA-MAS, J.; MONFORT, A.; DOLCET-SANJUAN, R.; KATZIR, N.; ARÚS, P.; MONFORTE, A.J. Simple-sequence repeat markers used in merging linkage maps of melon (*Cucumis melo* L.). **Theoretical and Applied Genetics**, v.110, p.802-811, 2005.
- GREGORY, M.P.; GREGORY, W.C. Exotic germplasm of *Arachis* L. interspecific hybrids. **Journal of Heredity**, v.70, p.185-193, 1979.
- GREGORY, W.C.; GREGORY, M.P. Groundnut. In: SIMMONDS, N.W. (Ed.) **Evolution of Crop Plants**. London, Longman; 1976, p.151-154.
- GREGORY, W.C.; KRAPOVICKAS, A.; GREGORY, M.P. Structure, variation, evolution, and classification in *Arachis*. In: SUMMERFIELD, R.J.; BUNTING, A.H. (Ed.) **Advances in Legume Science**. Kew, Royal Botanical Gardens; 1980, p.469-481.
- GUIMARÃES, P.M.; PROITE, K.; LEAL-BERTIOLI, S.C.M.; BERTIOLI, D. **Análise in silico da expressão gênica diferencial de *Arachis stenosperma* inoculado**

- com *Meloidogyne arenaria*. Boletim de Pesquisa & Desenvolvimento, n.85, Brasília: Embrapa Recursos Genéticos e Biotecnologia, 25p., 2005.
- GUO, W.; WANG, W.; ZHOU, B.; ZHANG, T. Cross-species transferability of *G. arboreum*-derived EST-SSRs in the diploid species of *Gossypium*. **Theoretical and Applied Genetics**, v.112, p.1573–1581, 2006.
- GUPTA, P.K.; RUSTGI, S.; SHARMA, S.; SINGH, R.; KUMAR, N.; BALYAN, H.S. Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. **Molecular and General Genomics**, v.270, p.315–323, 2003.
- GUPTA, P.K.; VARSHNEY, R.K. The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. **Euphytica**, v.113, p.163-185, 2000.
- GUTIERREZ, M.V.; VAZ PATTO, M.C.; HUGHET, T.; CUBERO, J.I.; MORENO, M.T.; TORRES, A.M. Cross-species amplification of *Medicago truncatula* microsatellites across three major pulse crops. **Theoretical and Applied Genetics**, v.110, p.1210-1217, 2005.
- HALEY, C.S.; KNOTT, S.A. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. **Heredity**, v.69, p.315-324, 1992.
- HALWARD, T.M.; STALKER, H.T.; KOCHERT, G. Development of an RFLP linkage map in diploid peanut species. **Theoretical and Applied Genetics**, v.87, p.379-384, 1993.
- HALWARD, T.M.; STALKER, H.T.; LARUE, E.A.; KOCHERT, G. Genetic variation detectable with molecular markers among unadapted germ-plasm resources of cultivated peanut and related wild species. **Genome**, v.34, p.1013-1020, 1991.
- HAMMONS, R.O. The origin and history of the groundnut. In: SMARTT, J. (Ed.). **The Groundnut Crop. A scientific basis for improvement**. London, Chapman; Hall, 1994. p.24-42.
- HAN, Z.; WANG, C.; SONG, X.; GUO, W; GOU, J.; LI, C.; CHEN, X.; ZHANG, T. Characteristics, development and mapping of *Gossypium hirsutum* derived EST-SSRs in allotetraploid cotton. **Theoretical and Applied Genetics**, v.112, p.430–439, 2006.
- HANCOCK, J.M. The contribution of slippage-like processes to genome evolution. **Journal of Molecular Evolution**, v.41, p.1038–1047, 1995.
- HAYES, A.J.; SAGHAI MAROOF, M.A. Targeted resistance gene mapping in soybean using modified AFLPs. **Theoretical and Applied Genetics**, v.100, p.1279-1283, 2000.
- HE, G.; MENG, R.; GAO, H.; GUO, B.; GAO, G.; NEWMAN, M.; PITTMAN, R.N.; PRAKASH, C.S. Simple sequence repeat markers for botanical varieties of cultivated peanut (*Arachis hypogaea* L.). **Euphytica**, v.142, p.131–136, 2005.
- HE, G.; MENG, R.; NEWMAN, M.; GAO, G.; PITTMAN, R.N.; PRAKASH, C.S. Microsatellites as DNA markers in cultivated peanut (*A. hypogaea* L.). **BMC Plant Biology**, v.3:3, 2003. [<http://www.biomedcentral.com/1471-2229/3/3>]
- HE, G.; PRAKASH, C.S. Evaluation of genetic relationships among botanical varieties of cultivated peanut (*Arachis hypogaea* L.) using AFLP markers. **Genetic Resources and Crop Evolution**, v.48, p.347-352, 2001.
- HE, G.; PRAKASH, C.S. Identification of polymorphic DNA markers in cultivated peanut (*Arachis hypogaea* L.). **Euphytica**, v.97, p.143-149, 1997.
- HERRMANN, D.; BOLLER, B.; STUDER, B.; WIDMER, F.; KÖLLIKER, R. QTL analysis of seed yield components in red clover (*Trifolium pratense* L.). **Theoretical and Applied Genetics**, v.112, p.536–545, 2006.

- HERSELMAN L. Genetic variation among Southern African cultivated peanut (*A. hypogaea* L.) genotypes as revealed by AFLP analysis. **Euphytica**, v.133, p.319-327, 2003.
- HERSELMAN, L.; THWAITES, R.; KIMMINS, F.M.; COURTOIS, B.; VAN DER MERWE, P.J.A.; SEAL, S.E. Identification and mapping of AFLP markers linked to peanut (*Arachis hypogaea* L.) resistance to the aphid vector of groundnut rosette disease. **Theoretical and Applied Genetics**, v.109, p.1426-1433, 2004.
- HILU, K.W.; STALKER, H.T. Genetic relationships between peanut and wild species of *Arachis* sect. *Arachis* (*Fabaceae*): evidence from RAPDs. **Plant Systematics and Evolution**, v.198, p.167-178, 1995.
- HOLBROOK, C. C.; ISLEIB, T.G. Geographical distribution and genetic diversity in *Arachis hypogaea*. **Peanut Science**, v.28, p.80-84, 2001.
- HOLBROOK, C.C.; STALKER, H.T. Peanut Breeding and Genetic Resources. In: JANICK, J. (Ed.). **Plant Breeding Reviews**, v.22, 2003, pp.297-355.
- HOPKINS, M.S.; CASA, A.M.; WANG, T.; MITCHELL, S.E.; DEAN, R.; KOCHERT, G.D.; KRESOVICH, S. Discovery and characterization of polymorphic Simple Sequence Repeats (SSRs) in peanut. **Crop Science**, v.39, p.1243-1247, 1999.
- HULBERT, S.H.; ILOTT, T.W.; LEGG, E.J.; LINCOLN, S.E.; LANDER, E.S.; MICHELMORE, R.W. Genetic analysis of the fungus, *Bremia lactucae*, using Restriction Fragment Length Polymorphisms. **Genetics**, v.120, p.947-958, 1988.
- HUMPHRY, M.E.; LAMBRIDES, C.J.; CHAPMAN, S.C.; AITKEN, E.A.B.; IMRIE, B.C.; LAWN, R.J.; McINTYRE, C.L.; LIU, C.J. Relationships between hard-seededness and seed weight in mungbean (*Vigna radiata*) assessed by QTL analysis. **Plant Breeding**, v.124, p.292-298, 2005.
- HUSTED, L. Cytological studies on the peanut, *Arachis*. II. Chromosome number, morphology and behavior, and their application to the problem of the origin of the cultivated forms. **Cytologia**, v.7, p.396-422, 1936.
- ISLEIB, T.G.; HOLBROOK, C.C.; GORBET, D.W. Use of plant introductions in peanut cultivar development. **Peanut Science**, v.28, p.96-113, 2001.
- ISLEIB, T.G.; WYNNE, J.C. Use of plant introductions in peanut improvement. In: **Use of Plant Introductions in Cultivar Development**, Part 2. Madison: CCSA, Cap. 4, 1992. p.75-116.
- JANSEN, R.C. Interval mapping of multiple quantitative trait loci. **Genetics**, v.135, p.205-211, 1993.
- JENCZEWSKI, E.; GHÉRARDI, M.; BONNIN, I.; PROSPERI, J.M.; OLIVIERI, I.; HUGUET, T. Insight on segregation distortions in two intraspecific crosses between annual species of *Medicago* (Leguminosae). **Theoretical and Applied Genetics**, v.94, p.682-691, 1997.
- JONES, D.A.; JONES, J.D.G. The roles of leucine-rich repeat proteins in plant defences. **Advances in Botanical Research**, v.24, p.89-167, 1997.
- JOSÉ, A.C.V.F. Desenvolvimento e Mapeamento de Marcadores Moleculares RGA em Espécies Silvestres de *Arachis*. 154p. Dissertação (Mestrado em Ciências Genômicas e Biotecnologia). Universidade Católica de Brasília. Brasília, 2006
- JUNG, S.; ABBOTT, A.; JESUDURAI, C.; TOMKINS, J.; MAIN, D. Frequency, type, distribution and annotation of simple sequence repeats in Rosaceae ESTs. **Functional and Integrative Genomics**, v.5, p.136-143, 2005.
- KANAZIN, V.; MAREK, L.F.; SHOEMAKER, R.C. Resistance gene analogs are conserved and clustered in soybean. **Proceedings of the National Academy of Sciences of the United States of America**, v.93, p.11746-11750, 1996.

- KANTETY, R.V.; LA ROTA, M.; MATTHEWS, D.E.; SORRELLS, M.E. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. **Plant Molecular Biology**, v.48, p.501-510, 2002.
- KAO, C.H.; ZENG, Z.B.; TEASDALE, R.D. Multiple Interval Mapping for Quantitative Trait Loci. **Genetics**, v.152, p.1203-1216, 1999.
- KHLESTKINA, E.K.; THAN, M.H.; PESTSOVA, E.G.; RODER, M.S.; MALYSHEV, S.V.; KORZUN, V.; BORNER, A. Mapping of 99 new microsatellite-derived loci in rye (*Secale cereale* L.) including 39 expressed sequence tags. **Theoretical and Applied Genetics**, v.109, p.725-732, 2004.
- KNAUFT, D.A.; GORBET, D.W. Genetic diversity among peanut cultivars. **Crop Science**, v.29, p.1417-1422, 1989.
- KNOX, M.R.; ELLIS, T.H.N. Excess Heterozygosity Contributes to Genetic Map Expansion in Pea Recombinant Inbred Populations. **Genetics**, v.162, p.861-873, 2002.
- KOBE B.; DEISENHOFER J. The leucine-rich repeat: a versatile binding motif. **Trends in Biochemical Sciences**, v.19, p.425-420, 1994.
- KOCHERT, G.; HALWARD, T.; BRANCH, W.D.; SIMPSON, C.E. RFLP variability in peanut (*Arachis hypogaea*) cultivars and wild species. **Theoretical and Applied Genetics**, v.81, p.565-570, 1991.
- KOCHERT, G.; STALKER, H.T.; GIMENES, M.; GALGARO, L.; LOPES, C.R.; MOORE, K. RFLP and cytogenetic evidence on the origin and evolution of allotetraploid domesticated peanut, *Arachis hypogaea* (Leguminosae). **American Journal of Botany**, v.83, p.1282-1291, 1996.
- KONIECZNY, A.; AUSUBEL, F.M. A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype-specific markers. **The Plant Journal**, v.4, p.403-410, 1993.
- KRAPOVICKAS, A.; GREGORY, W. C. Taxonomía del género *Arachis* (Leguminosae). **Bonplandia**, v.8, p.1-186, 1994.
- La ROTA, M.; KANTETY, R.V.; YU, J.K.; SORRELLS, M.E. Nonrandom distribution and frequencies of genomic and EST-derived microsatellite markers in rice, wheat, and barley. **BMC Genomics**, v.6(1):23, 2005.
- LANDER, E.; BOTSTEIN, D. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. **Genetics**, v.121, p.185-199, 1989.
- LANDER, E.S.; GREEN, P.; ABRAHAMSON, J.; BARLOW, A.; DALY, M.J.; LINCOLN, S.E.; NEWBURG, L. MAPMAKER: an interactive computer package for constructing primary genetic maps of experimental and natural populations. **Genomics**, v.1, p.174-181, 1987.
- LANHAM, P.G.; FENNEL, S.; MOSS, J.P.; POWELL, W. Detection of polymorphic loci in *Arachis* germplasm using random amplified polymorphic DNAs. **Genome**, v.35, p.885-889, 1992.
- LAVIA, G.I. Estudios cromosómicos en *Arachis* (Leguminosae). **Bonplandia**, v.9, p.111-120, 1996.
- LAVIA, G.I. Karyotypes of *Arachis palustris* and *A. praecox* (Section *Arachis*), two species with basic chromosome number $x = 9$. **Cytologia**, v.63, p.177-181, 1998.
- LEAL-BERTIOLI, S.C.M.; GUIMARÃES, P.M.; BRUZZI LION, M.; CARNEIRO, R.G.; VALLS, J.F.M.; BERTIOLI, D.J. **Busca de resistência ao nematóide das galhas, *Meloidogyne* spp., e seqüências análogas a genes de resistência em acessos silvestres de *Arachis***. Boletim de Pesquisa & Desenvolvimento, n.20, Brasília: Embrapa Recursos Genéticos e Biotecnologia, 17p., 2000.

- LEVINSON, G; GUTMAN, G.A. High frequencies of short frameshifts in poly-CA/TG tandem repeats borne by bacteriophage M13 in *Escherichia coli* K-12. **Nucleic Acids Research**, v.15, p.5323-5338, 1987.
- LI, Y-C.; KOROL, A.B.; FAHIMA, T.; NEVO, E. Microsatellites within genes: Structure, function, and evolution. **Molecular Biology and Evolution**, v.21:991-1007, 2004.
- LIEWLAKSANEYANAWIN, C.; RITLAND, C.E.; EL-KASSABY, Y.A.; RITLAND, K. Single-copy, species-transferable microsatellite markers developed from loblolly pine ESTs. **Theoretical and Applied Genetics**, v.109, p.361-369, 2004.
- LILLIEFORS, H. W. On the Kolmogorov-Smirnov test for normality. **Biometrika**, Great Britain, v.62, p.399-402, 1967.
- 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**, v.44, p.397-401, 1989.
- LIU, B.H. **Statistical Genomics: Linkage, Mapping, and QTL Analysis**. CRC Press, Boca Raton, 1998. 611 p.
- LORIEUX, M.; PERRIER, X.; GOFFINET, B.; LANAUD, C.; GONZÁLEZ DE LEÓN, D. Maximum-likelihood models for mapping genetic markers showing segregation distortion. 2. F2 populations. **Theoretical and Applied Genetics**, v.90, p.81-89, 1995.
- LYNCH, M.; WALSH, B. **Genetics and analysis of quantitative traits**. Sunderland: Sinauer Associates, 1998. 980 p.
- LYONS, L.; LAUGHLIN, T.; COPELAND, N.; JENKINS, N.; WOMACK, J.; O'BRIEN, S. Comparative anchor tagged sequences (CATS) for integrative mapping of mammalian sequences. **Nature Genetics**, v.15, p.47-56, 1997.
- MAMMADOV, J.A.; LIU, Z.; BIYASHEV, R.M.; MUEHLBAUER, G.J.; SAGHAI MAROOF, M.A. Cloning, genetic and physical mapping of resistance gene analogs in barley (*Hordeum vulgare* L.). **Plant Breeding**, v.125, p.32-42, 2006.
- MAUGHAN, P.J.; SAGHAI MAROOF, M.A.; BUSS, G.R. Molecular marker analysis of seed weight: genomic locations, gene action, and evidence for orthologous evolution among three legume species. **Theoretical and Applied Genetics**, v.93, p.574-579, 1996.
- MAYER, M. A comparison of regression interval mapping and multiple interval mapping for linked QTL. **Heredity**, v.94, p.599-605, 2005.
- McINTYRE, C.L.; CASU, R.E.; DRENTH, J.; KNIGHT, D.; WHAN, V.A.; CROFT, B.J.; JORDAN, D.R.; MANNERS, J.M. Resistance gene analogues in sugarcane and sorghum and their association with quantitative trait loci for rust resistance. **Genome**, v.48, p.391-400, 2005.
- METZGAR, D.; BYTOF, J.; WILLS, C. Selection against frameshift mutations limits microsatellite expansion in coding DNA. **Genome Research**, v.10, p.72-80, 2000.
- MICHELMORE, R.W.; PARAN, I.; KESSELI, R.V. Identification of markers linked to disease resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. **Proceedings of the National Academy of Sciences of the United States of America**, v.88, p.9828-9832, 1991.
- MILLA, S.R.; ISLEIB, T.G.; STALKER, H.T. Taxonomic relationships among *Arachis* sect. *Arachis* species as revealed by AFLP markers. **Genome**, v.48, p.1-11, 2005.

- MOHAN, M.; NAIR, S.; BHAGWAT, A.; KRISHNA, T.G.; YANO, M.; BHATIA, C.R.; SASAKI, T. Genome mapping, molecular markers and marker-assisted selection in crop plants. **Molecular Breeding**, v.3, p.87-103, 1997.
- MOOT, D.J.; McNEIL, D.L. Yield components, harvest index and plant type in relation to yield differences in field pea genotypes. **Euphytica**, v.86, p.31-40, 1995.
- MORAES, S.A.; SALGADO, C.L. Utilização da técnica de folhas destacadas de amendoim (*Arachis hypogaea* L.) para inoculações com *Cercospora arachidicola* Hori e *Cercospora personata* (Bert. & Curt.) Ell. & Ev. **Summa Phytopathologica**, v.8, p.39-55, 1984.
- MORETZSOHN, M.C.; HOPKINS, M.S.; MITCHELL, S.E.; KRESOVICH, S.; VALLS, J.F.M.; FERREIRA, M.E. Genetic diversity of peanut (*Arachis hypogaea* L.) and its wild relatives based on the analysis of hypervariable regions of the genome. **BMC Plant Biology**, v.4:11, 2004. [<http://www.biomedcentral.com/1471-2229/4/11>].
- MORGANTE, M.; HANAFEY, M.; POWELL, W. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. **Nature Genetics**, v.30, p.194-200, 2002.
- NEFF, M.M.; NEFF, J.D.; CHORY, J.; PEPPER, A.E. dCAPS, a simple technique for the genetic analysis of single nucleotide polymorphisms: experimental applications in *Arabidopsis thaliana* genetics. **The Plant Journal**, v.14, p.387-392, 1998.
- NEI, M. Analysis of gene diversity in subdivided populations. **Proceedings of the National Academy of Sciences of the United States of America**, v.70, p.3321-3323, 1973.
- NEI, M. **Molecular Evolutionary Genetics**. Columbia University Press, New York, 1987. 512 p.
- PAIK-RO, O. G.; SMITH, R.L.; KNAUFT, D.A. Restriction fragment length polymorphism evaluation of six peanut species within the *Arachis* section. **Theoretical and Applied Genetics**, v.84, p.201-208, 1992.
- PALMIERI, D.A.; BECHARA, M.D.; CURI, R.A.; GIMENES, M.A.; LOPES, C.R. Novel polymorphic microsatellite markers in section *Caulorrhizae* (*Arachis*, Fabaceae). **Molecular Ecology Notes**, v.5, p.77-79, 2005.
- PALMIERI, D.A.; HOSHINO, A.A.; BRAVO, J.P.; LOPES, C.R.; GIMENES, M.A. Isolation and characterization of microsatellite loci from the forage species *Arachis pintoii* (Genus *Arachis*). **Molecular Ecology Notes**, v.2, p.551-553, 2002.
- PANDE, S.; RAO, J.N. Resistance of Wild *Arachis* Species to Late Leaf Spot and Rust in Greenhouse Trials. **Plant Disease**, v.85, p.851-855, 2001.
- PEAKALL, R.; GILMORE, S.; KEYS, W.; MORGANTE, M.; RAFALSKI, A. Cross-Species Amplification of Soybean (*Glycine max*) Simple Sequence Repeats (SSRs) Within the Genus and Other Legume Genera: Implications for the Transferability of SSRs in Plants. **Molecular Biology and Evolution**, v.15, p.1275-1287, 1998.
- PEÑALOZA, A.D.P.S.; VALLS, J.F.M. Contagem do número cromossômico em acessos de *Arachis decora* (Leguminosae). In: VEIGA, R.F.A.; BOVI, M.L.A.; BETTI, J.A.; VOLTAN, R.B.Q. (Eds.) SIMPÓSIO LATINO-AMERICANO DE RECURSOS GENÉTICOS VEGETAIS, 1., 1997, Campinas. **Programas e Resumos**. Campinas: IAC/Embrapa-Cenargen, 1997. p.39.
- PEZZOPANE, J.R.M.; PEDRO JÚNIOR, M.J.; MORAES, S.A.; GODOY, J.I.; PEREIRA, J.C.V.N.A.; SILVEIRA, L.C.P. Chuva e previsão de épocas de

- pulverização para controle das manchas foliares do amendoim. **Bragantia**, v.57, p.285:295, 1998.
- PRASAD, M.; VARSHNEY, R.K.; ROY, J.K.; BALYAN, H.S.; GUPTA, P.K. The use of microsatellites for detecting DNA polymorphism, genotype identification and genetic diversity in wheat. **Theoretical and Applied Genetics**, v.100, p.584-592, 2000.
- PROVAN, J.; POWELL, W.; WAUGH, R. Microsatellite analysis of relationships within cultivated potato (*Solanum tuberosum*). **Theoretical and Applied Genetics**, v.92, p.1078-1084, 1996.
- QUILLET, M.C.; MADJIDIAN, N.; GRIVEAU, Y.; SERIEYS, H.; TERSAC, M.; LORIEUX, M.; BERVILLÉ, A. Mapping genetic factors controlling pollen viability in an interspecific cross in *Helianthus* sect. *Helianthus*. **Theoretical and Applied Genetics**, v.91, p.1195-1202, 1995.
- RAFALSKI, J.A.; VOGEL, J.M.; MORGANTE, M.; POWELL, W. ANDRE, C.; TINGEY, S.V. Generating and using DNA markers in plants. In: BIRREN, B.; LAI, E. (Eds). **Analysis of non-mammalian genomes – a practical guide**. Academic Press, New York, 1996. p.75-134.
- RAINA, S.N.; MUKAI, Y. Genomic in situ hybridization in *Arachis* (*Fabaceae*) identifies the diploid wild progenitors of cultivated (*A. hypogaea*) and related wild (*A. monticola*) peanut species. **Plant Systematics and Evolution**, v.214, p.251-262, 1999.
- RAINA, S.N.; RANI, V.; KOJIMA, T.; OGIHARA, Y.; SINGH, K.P.; DEVARUMATH, R.M. RAPD and ISSR fingerprints as useful genetic markers for analysis of genetic diversity, varietal identification, and phylogenetic relationships in peanut (*Arachis hypogaea*) cultivars and wild species. **Genome**, v.44, p.763-772, 2001.
- RAO, G.U.; BENCHAIM, A.; BOROVSKY, Y.; PARAN, I. Mapping of yield-related QTLs in pepper in an interspecific cross of *Capsicum annuum* and *C. frutescens*. **Theoretical and Applied Genetics**, v.106, p.1457-1466, 2003a.
- RAO, N.K.; REDDY, L.J.; BRAMEL, P.J. Potential of wild species for genetic enhancement of some semi-arid food crops. **Genetic Resources and Crop Evolution**, v.50, p.707–721, 2003b.
- RITSCHHEL, P.S.; LINS, T.C.L.; TRISTAN, R.L.; BUSO, G.S.C.; BUSO, J.A.; FERREIRA, M.E. Development of microsatellite markers from an enriched genomic library for genetic analysis of melon (*Cucumis melo* L.). **BMC Plant Biology**, v.4:9, 2004 [<http://www.biomedcentral.com/1471-2229/4/9>]
- ROHLF, F.J. **NTSYS-PC: numerical taxonomy and multivariate analysis system - version 2.1**. Exeter Software, New York, 2000.
- SAMBROOK, J.; FRITSCH, E.F.; E MANIATIS, T. **Molecular Cloning – A laboratory manual**. 2nd ed. Cold Spring Harbor Laboratory Press, 1989.
- SANTOS, R.C.; GODOY, I.J. Hibridação em amendoim. In: BORÉM, A. (Ed.). **Hibridação artificial de plantas**, Viçosa:UFV, 1999, p.83-100.
- SAVAGE, G.P.; KEENAN, J.I. The composition and nutritive value of groundnut kernels. In: SMARTT, J. (Ed.). **The groundnut crop: a scientific basis for improvement**. Chapman and Hall, Londres, 1994. p.173-213.
- SCHLÖTTERER, C.; TAUTZ, D. Slippage synthesis of simple sequence DNA. **Nucleic Acids Research**, v.20, p.211-215, 1992.
- SCHUSTER, I.; CRUZ, C.D. **Estatística Genômica Aplicada a Populações Derivadas de Cruzamentos Controlados**. Editora UFV, Viçosa, 2004. 568 p.

- SCOTT, K.D.; EGGLE, P.; SEATON, G.; ROSETTO, M.; ABLETT, E.M.; LEE, L.S.; HENRY, R.J. Analysis of SSRs derived from grape ESTs. **Theoretical and Applied Genetics**, v.100, p.723–726, 2000.
- SEIJO, J.G.; LAVIA, G.I.; FERNÁNDEZ, A.; KRAPOVICKAS, A.; DUCASSE, D.; MOSCONE, E.A. Physical mapping of the 5S and 18S–25S rRNA genes by FISH as evidence that *Arachis duranensis* and *A. ipaënsis* are the wild diploid progenitors of *A. hypogaea* (Leguminosae). **American Journal of Botany**, v.91, p.1294-1303, 2004.
- SIMPSON, C.E.; STARR, J.L.; CHURCH, G.T.; BUROW, M.D.; PATERSON, A.H. Registration of 'NemaTAM' Peanut. **Crop Science**, v.43, p.1561, 2003.
- SIMPSON, C.E.; STARR, J.L. Registration of 'Coan' peanut. **Crop Science**, v.41, p.918, 2001.
- SINGH, B.; SINGH, U. Peanut as a source of protein for human foods. **Plant Foods for Human Nutrition**, v.41, p.165-177, 1991.
- SINGH, S.; GUMBER, R.K.; JOSHI, N.; SINGH, K. Introgression from wild *Cicer reticulatum* to cultivated chickpea for productivity and disease resistance. **Plant Breeding**, v.124, p.477-480, 2005.
- SONG, Q.J.; MAREK, L.F.; SHOEMAKER, R.C.; LARA, 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**, v.109, p.122-128, 2004.
- STALKER, H.T. A new species in section *Arachis* of peanuts with D genome. **American Journal of Botany**, v.78, p.630-637, 1991.
- STALKER, H.T. Utilizing *Arachis cardenasii* as a source of *Cercospora* leafspot resistance for peanut improvement. **Euphytica**, v. 33, p.529-538, 1984.
- STALKER, H.T.; MOZINGO, L.G. Molecular genetics of *Arachis* and marker-assisted selection. **Peanut Science**, v.28, p.117–123, 2001.
- STALKER, H.T.; SIMPSON, C.E. Germplasm resources in *Arachis*. In: PATTEE, H.E.; STALKER, H.T. (Eds). **Advances in Peanut Science**, APRES, Stillwater, OK, 1995. Cap. 2, pp 14-53.
- STARR, J. L., MORGAN, E. R., SIMPSON, C. E. Management of the peanut root-knot nematode, *Meloidogyne arenaria*, with host resistance. Online. **Plant Health Progress**, 2002. doi:10.1094/PHP-2002-1121-01-HM.
- STAUB, J.E.; SERQUEN, F.C.; GUPTA, M. Genetic markers, map construction, and their application in plant breeding. **HortScience**, v.31, n.5, p.729-740, 1996.
- SUBRAHMANYAM, P.; GHANEKAR, A.M.; NOLT, B.L.; REDDY, D.V.R.; McDONALD, D. Resistance to groundnut diseases in wild *Arachis* species In: **Proceedings of International Workshop on Cytogenetics of Arachis**, ICRISAT, India, Patancheru, 1985.
- SUBRAMANIAN, V.; GURTU, S.; NAGESWARA RAO, R.C.; NIGAM, S.N. Identification of DNA polymorphism in cultivated groundnut using random amplified polymorphic DNA (RAPD) assay. **Genome**, v.43, p.656-660, 2000.
- TAKEZAKI, N.; NEI, M. Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. **Genetics**, v.139, p.457–462, 1996.
- TALAME, V.; SANGUINETI, M.C.; CHIAPPARINO, E.; BAHRI, H.; SALEM, B.M.; FORSTER, B.P.; ELLIS, R.P.; RHOUMA, S.; ZOUMAROU, W.; WAUGH, R.; TUBEROSA, R. Identification of *Hordeum spontaneum* QTL alleles improving field performance of barley grown under rainfed conditions. **Annals of Applied Biology**, v.144, p.309-319, 2004.

- TALLURY, S.P.; HILU, K.W.; MILLA, S.R.; FRIEND, S.A.; ALSAGHIR, M.; STALKER, H.T.; QUANDT, D. Genomic affinities in *Arachis* section *Arachis* (Fabaceae): molecular and cytogenetic evidence. **Theoretical and Applied Genetics**, v.111, p.1229-1237, 2005.
- TAMELING, W.I.; ELZINGA, S.D.; DARMIN, P.S.; VOSSEN, J.H.; TAKKEN, F.L.; HARING, M.A.; CORNELISSEN, B.J. The tomato R gene products I-2 and MI-1 are functional ATP binding proteins with ATPase activity. **Plant Cell**, v.14, p.2929-2939, 2002.
- TANKSLEY, S.D. Mapping polygenes. **Annual Review of Genetics**, v.27, p.205-233, 1993.
- TANKSLEY, S.D.; McCOUCH, S.R. Seed Banks and Molecular Maps: Unlocking Genetic Potential from the Wild. **Science**, v.277, p.1063-1066, 1997.
- TANKSLEY, S.D.; MILLER, J.; PATERSON, A.; BERNATSKY, R. **Molecular mapping of plant chromosomes**. In: GUSTAFSON, J.P.; APPELS, R. (Eds.). New York: Plenum Press, 1988. p.157-173.
- TAUTZ, D. Hypervariability of simple sequences as a general source for polymorphic DNA markers. **Nucleic Acids Research**, v.25, p.6463-6471, 1989.
- TAUTZ, D.; RENZ, M. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. **Nucleic Acids Research**, v.25, p.4127-4138, 1984.
- THIEL, T.; MICHALEK, W.; VARSHNEY, R.K.; GRANER, A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). **Theoretical and Applied Genetics**, v.106, p.411-422, 2003.
- THOMAS, M.R.; SCOTT, N.S. Microsatellite repeats in grapevine reveal DNA polymorphisms when analyzed as sequence-tagged sites (STSs). **Theoretical and Applied Genetics**, v.86, p.985-990, 1993.
- TIAN, A.G.; WANG, J.; CUI, P.; HAN, Y.J.; XU, H.; CONG, L.J.; HUANG, X.G.; WANG, X.L.; JIAO, Y.Z.; WANG, B.J.; WANG, Y.J.; ZHANG, J.S.; CHEN, S.Y. Characterization of soybean genomic features by analysis of its expressed sequence tags. **Theoretical and Applied Genetics**, v.108, p.903-913, 2004.
- TIMMERMAN-VAUGHAN, G.M.; McCALLUM, J.A.; FREW, T.J.; WEEDEN, N.F.; RUSSELL, A.C. Linkage mapping of quantitative trait loci controlling seed weight in pea (*Pisum sativum* L.). **Theoretical and Applied Genetics**, v.93, p.431-439, 1996.
- TIMMERMAN-VAUGHAN, G.M.; MILLS, A.; WHITFIELD, C.; FREW, T.; BUTLER, R.; MURRAY, S.; LAKEMAN, M.; MCCALLUM, J.; RUSSELL, A.; WILSON, D. Linkage Mapping of QTLs for Seed Yield, Yield Components, and Developmental Traits in Pea. **Crop Science**, v.45, p.1336-1344, 2005.
- UDUPA, S.M.; ROBERTSON, L.D.; WEIGAND, F.; BAUM, M.; KAHL, G. Allelic variation at (TAA)_n microsatellite loci in a world collection of chickpea (*Cicer arietinum* L.) germplasm. **Molecular and General Genetics**, v.261, p.354-363, 1999.
- VALLS, J.F.M. Diversidade genética no gênero *Arachis* e a origem do amendoim. In: BANDEL, G.; AGUIAR-PERECIN, M.L.R.; OLIVEIRA, G.C.X. (Ed.). 17º Encontro sobre temas de genética e melhoramento, "genômica: uma abordagem em plantas", v.17. **Anais**. Piracicaba: ESALQ, 2000, p.51.
- VALLS, J.F.M.; SIMPSON, C.E. New species of *Arachis* L. (Leguminosae) from Brazil, Paraguay and Bolivia. **Bonplandia**, v.14, p.35-64, 2005.

- VALLS, J.F.M.; SIMPSON, C.E. Taxonomy, natural distribution, and attributes of *Arachis*. In: KERRIDGE, P.C.; HARDY, B. (Eds.). **Biology and Agronomy of Forage *Arachis***. Cali: CIAT, 1994, p.1-18.
- Van der LINDEN, C.G.; WOUTERS, D.C.; MIHALKA, V.; KOCHIEVA, E.Z.; SMULDERS, M.J.M.; VOSMAN, B. Efficient targeting of plant disease resistance loci using NBS profiling. **Theoretical and Applied Genetics**, v.109, p.384-393, 2004.
- VARSHNEY, R.K.; SIGMUND, R.; BORNER, A.; KORZUN, V.; STEIN, N.; SORRELLS, M.E.; LANGRIDGE, P.; GRANER, A. Interspecific transferability and comparative mapping of barley EST-SSR markers in wheat, rye and rice. **Plant Science**, v.168, p.195-202, 2005.
- VARSHNEY, R.K.; THIEL, T.; STEIN, N.; LANGRIDGE, P.; GRANER, A. In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species. **Cellular and Molecular Biology Letters**, v.7, p.537-546, 2002.
- VOORRIPS, R.E. MapChart: Software for the graphical presentation of linkage maps and QTLs. **The Journal of Heredity**, v.93, p.77-78, 2002.
- VOS, P.; HOGERS, R.; BLEEKER, M.; REIJANS, M.; van de LEE, T.; HORNES, M.; FRIJTERS, A.; POT, J.; PELEMAN, J.; KUIPER, M.; ZABEAU, M. AFLP: a new technique for DNA fingerprinting. **Nucleic Acids Research**, v.23, p.4407-4414, 1995.
- WANG S., BASTEN, C.J.; ZENG, Z.-B. Windows QTL Cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh, NC. 2006. (<http://statgen.ncsu.edu/qtlcart/WQTLCart.htm>)
- WANG, D.; GRAEF, G.L.; PROCOPIUK, A.M.; DIERS, B.W. Identification of putative QTL that underlie yield in interspecific soybean backcross populations. **Theoretical and Applied Genetics**, v.108, p.458-67, 2004.
- WANG, Y.-H.; CHOI, W.; THOMAS, C.E.; DEAN, R.A. Cloning of disease-resistance homologues in end sequences of BAC clones linked to *Fom-2*, a gene conferring resistance to *Fusarium* wilt in melon (*Cucumis melo* L.). **Genome**, v. 45, p.473-480, 2002.
- WANG, Z.; WEBER, J.L.; ZHONG, G.; TANKSLEY, S.D. Survey of plant short tandem DNA repeats. **Theoretical and Applied Genetics**, v.88, p.1-6, 1994.
- WANG, Z.X.; YANO, M.; YAMANOUCHI, U.; IWAMOTO, M.; MONNA, L.; HAYASAKA, H.; KATAYOSE, Y.; SASAKI, T. The *pib* gene for rice blast resistance belongs to the nucleotide binding and leucine-rich repeat class of plant disease resistance genes. **The Plant Journal**, v.19, p.55-64, 1999.
- WEBER, J.L. Informativeness of human (dC-dA)_n (dG-dT)_n polymorphisms. **Genomics**, v.7, p.524-530, 1990.
- WEBER, Z.; MAY, P.E. Abundant class of human DNA polymorphism which can be typed using the polymerase chain reaction. **American Journal of Human Genetics**, v.44, p.388-396, 1989.
- WEIR, B.S. **Genetic Data Analysis II**. Sinauer Associates, Massachusetts, 1996. 445p.
- WESTERBERGH, A.; DOEBLEY, J. Quantitative trait loci controlling phenotypes related to the perennial versus annual habit in wild relatives of maize. **Theoretical and Applied Genetics**, v.109, p.1544-1553, 2004.
- WU, J.; JENKINS, J.; ZHU, J.; McCARTY, J.; WATSON, C. Monte Carlo simulations on marker grouping and ordering. **Theoretical and Applied Genetics**, v.107, p.568-573, 2003.

- XIAO, J.; LI, J.; GRANDILLO, S.; AHN, S.N.; YUAN, L.; TANKSLEY, S.D.; McCOUCH, S.R. Identification of Trait-Improving Quantitative Trait Loci Alleles From a Wild Rice Relative, *Oryza rufipogon*. **Genetics**, v.150, p.899–909, 1998.
- YOON, D.-B.; KANG, K.-H.; KIM, H.-J.; JU, H.-G.; KWON, S.-J.; SUH, J.-P.; JEONG, O.-Y.; AHN, S.-N. Mapping quantitative trait loci for yield components and morphological traits in an advanced backcross population between *Oryza grandiglumis* and the *O. sativa japonica* cultivar Hwaseongbyeo. **Theoretical and Applied Genetics**, v.112, p.1052–1062, 2006.
- YOSHIMURA, S.; YAMANOUCHI, U.; KATAYOSE, Y.; TOKI, S.; WANG, Z.X.; KONO, I.; KURATA, N.; YANO, M.; IWATA, N.; SASAKI, T. Expression of Xa1, a bacterial blight-resistance gene in rice, is induced by bacterial inoculation. **Proceedings of the National Academy of Sciences of the United States of America**, v.95, p.1663-1668, 1998.
- YOUNG, N.D.; CANNON, S.B.; SATO, S.; KIM, D.; COOK, D.R.; TOWN, C.D.; ROE, B.A.; TABATA, S. Sequencing the Genespaces of *Medicago truncatula* and *Lotus japonicus*. **Plant Physiology**, v.137, p.1174-1181, 2005.
- YOUNG, N.D.; WEEDEN, N.F.; KOCHERT, G. Genome mapping in legumes (Family Fabaceae). In: PATERSON, A.H. (Ed.). **Genome mapping in plants**. R.G. Landes, Austin, TX, 1996. pp.212-227.
- YUKSEL, B.; ESTILL, J.C.; SCHULZE, S.R.; PATERSON, A.H. Organization and evolution of resistance gene analogs in peanut. **Molecular and General Genetics**, v. 274, p.248-263, 2005.
- YUKSEL, B.; PATERSON, A.H. Construction and characterization of a peanut HindIII BAC library. **Theoretical and Applied Genetics**, v.111, p.630-639, 2005.
- ZENG, Z.B. Precision mapping of quantitative trait loci. **Genetics**, v.136, p.1457-1466, 1994.
- ZENG, Z.B. Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. **Proceedings of the National Academy of Sciences of the United States of America**, v.90, p.10972-10976, 1993.
- ZENG, Z.B.; KAO, C.H.; BASTEN, C.J. Estimating the genetic architecture of quantitative traits. **Genetical Research**, v.74, p.279-289, 1999.
- 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**, v.108, p.1131-1139, 2004.

ANEXOS

ANEXO 1

Artigo publicado na Theoretical and Applied Genetics:

MORETZSOHN, M.C.; LEOI, L.; PROITE, K.; GUIMARÃES, P.M.; LEAL-BERTIOLI, S.C.M.; GIMENES, M.A.; MARTINS, W.S.; VALLS, J.F.M.; GRATTAPAGLIA, D.; BERTIOLI, D.J. A microsatellite based, gene-rich linkage map for the AA genome of *Arachis* (Fabaceae). **Theoretical and Applied Genetics**, v.111, p.1060-1071, 2005.

ANEXO 2

Artigo publicado na Nucleic Acids Reserch:

MARTINS, W.; SOUSA, D.; PROITE, K.; GUIMARÃES, P.M.; MORETZSOHN, M.;
BERTIOLI, D. New softwares for automated microsatellite marker development.
Nucleic Acids Research, v.34, e31, 2006.

ANEXO 3

Arquivo suplementar com as informações sobre os marcadores (Moretzsohn *et al.*, 2005 - ESM 1):

Para os 271 novos marcadores microssatélites: nome do loco; nome do primer (nome do loco reduzido); “primers forward e reverse” (5' - 3'); repetição; tipo de repetição; temperatura de anelamento (Ta); polimorfismo para o cruzamento *A. duranensis* (K7988) x *A. stenosperma* (V10309)^a; grupo de ligação (GL); polimorfismo para os seis acessos de *A. hypogaea* (*A. hyp*)^a; número de locos amplificados para *A. hypogaea* (# locos); nota subjetiva referente à qualidade dos produtos amplificados, sendo 3 a melhor (nota); resultados da análise de BlastX; significância das similaridades encontradas por BlastX (E-value) e uns breves comentários.

Para os 162 marcadores já publicados: nome do loco, polimorfismo para o cruzamento *A. duranensis* (K7988) x *A. stenosperma* (V10309)^a, grupo de ligação (GL), referência do artigo em que os marcadores foram publicados, resultados da análise de BlastX e significância das similaridades encontradas por BlastX (E-value).

Os números dos acessos do GenBank para as 207 novas seqüências e para as 67 seqüências de Moretzsohn *et al.*, 2004, que não haviam sido submetidas ao GenBank, são também incluídas.

^a P = polimórfico, M = monomórfico, Hífen (-) = ausência de amplificação.

Nome do arquivo: Tese Moretzsohn AC.doc
Pasta: G:
Modelo: C:\Documents and Settings\cliente\Dados de aplicativos\Microsoft\Modelos\Normal.dot
Título: Desenvolvimento de um mapa de ligação para espécies silvestres de Arachis com genoma AA e mapeamento de QTL de interesse agron
Assunto:
Autor: Márcio Moretzsohn
Palavras-chave:
Comentários:
Data de criação: 28/2/2007 20:15:00
Número de alterações:21
Última gravação: 28/2/2007 21:27:00
Salvo por: EMBRAPA
Tempo total de edição: 68 Minutos
Última impressão: 28/2/2007 21:28:00
Como a última impressão
Número de páginas: 117
Número de palavras: 34.887 (aprox.)
Número de caracteres: 188.396 (aprox.)

M. C. Moretzsohn · L. Leoi · K. Proite
P. M. Guimarães · S. C. M. Leal-Bertioli
M. A. Gimenes · W. S. Martins
J. F. M. Valls · D. Grattapaglia · D. J. Bertioli

A microsatellite-based, gene-rich linkage map for the AA genome of *Arachis* (Fabaceae)

Received: 23 March 2005 / Accepted: 23 June 2005 / Published online: 9 August 2005
© Springer-Verlag 2005

Abstract Cultivated peanut (*Arachis hypogaea*) is an important crop, widely grown in tropical and subtropical regions of the world. It is highly susceptible to several biotic and abiotic stresses to which wild species are resistant. As a first step towards the introgression of these resistance genes into cultivated peanut, a linkage map based on microsatellite markers was constructed, using an F₂ population obtained from a cross between two diploid wild species with AA genome (*A. duranensis* and *A. stenosperma*). A total of 271 new microsatellite markers were developed in the present study from SSR-enriched genomic libraries, expressed sequence tags (ESTs), and by “data-mining” sequences available in GenBank. Of these, 66 were polymorphic for cultivated peanut. The 271 new markers plus another 162 published for peanut were screened against both progenitors and 204 of these (47.1%) were polymorphic, with 170

codominant and 34 dominant markers. The 80 codominant markers segregating 1:2:1 ($P < 0.05$) were initially used to establish the linkage groups. Distorted and dominant markers were subsequently included in the map. The resulting linkage map consists of 11 linkage groups covering 1,230.89 cM of total map distance, with an average distance of 7.24 cM between markers. This is the first microsatellite-based map published for *Arachis*, and the first map based on sequences that are all currently publicly available. Because most markers used were derived from ESTs and genomic libraries made using methylation-sensitive restriction enzymes, about one-third of the mapped markers are genic. Linkage group ordering is being validated in other mapping populations, with the aim of constructing a transferable reference map for *Arachis*.

Communicated by S. J. Knapp

Electronic supplementary material is available for this at <http://dx.doi.org/10.1007/s00122-005-0028-x>.

M. C. Moretzsohn (✉) · K. Proite · P. M. Guimarães
S. C. M. Leal-Bertioli · J. F. M. Valls · D. Grattapaglia
Embrapa Recursos Genéticos e Biotecnologia, C.P. 02372,
CEP 70.770-900 Brasília, DF, Brazil
E-mail: marciocm@cenargen.embrapa.br
Tel.: +55-61-4484615
Fax: +55-61-3403624

M. C. Moretzsohn · K. Proite · D. Grattapaglia
Departamento de Biologia Celular,
IB - Universidade de Brasília (UnB), Campus Universitário,
CEP 70.910-900 Brasília, DF, Brazil

L. Leoi · D. J. Bertioli · D. Grattapaglia
Universidade Católica de Brasília, Campus II, SGAN 916,
CEP 70.790-160 Brasília, DF, Brazil

M. A. Gimenes
Departamento de Genética, IB-UNESP, Rubião Jr.,
CEP 18618-000 Botucatu, SP, Brazil

W. S. Martins
Departamento de Ciência da Computação, Universidade Católica
de Goiás, Goiânia, GO, Brazil

Introduction

The genus *Arachis*, family Fabaceae, is native to South America, and contains 69 described species assembled into nine sections, according to their morphology, geographic distribution, and crossability (Krapovickas and Gregory 1994). The cultivated species, peanut (*Arachis hypogaea*), is a key crop for small farmers throughout the tropics, but especially in Africa and Asia, it is a valuable source of dietary protein and oil, and of fodder for livestock. In addition, the peanut's ability to fix nitrogen is important in the maintenance of soil fertility especially where farmers are too poor to afford chemical fertilizers. Peanut world production is increasing and has reached 35.6 million tons (in the shell) and 5.8 million tons of oil per year (FAO 2003). Currently peanut is the fifth most important oilseed in the world in terms of volume of oil production.

Cultivated peanut is an allotetraploid (genome AABB). It is a member of the section *Arachis* that also includes about 25 diploid and one other tetraploid wild species (*A. monticola*) (Krapovickas and Gregory 1994).

Its origin was via hybridization of two diploid wild species followed by a rare spontaneous duplication of chromosomes (Halward et al. 1991; Young et al. 1996). The resulting tetraploid plant would have had hybrid vigor, but have been reproductively isolated from its wild relatives. Therefore, all land races of peanuts are probably derived from one, or a few plants. This led to a very limited genetic diversity, which has constrained advances in genetics that are necessary for efficient modern breeding, and led to low diversity for some traits of agricultural interest. In contrast, wild diploid *Arachis* species are genetically very diverse and have been selected during evolution by a range of environments and biotic stresses, providing a rich source of variation in agronomically important traits.

In *Arachis*, few linkage maps have been published. Halward et al. (1993) developed an RFLP-based map using an F₂ population derived from a cross between two diploid wild species with AA genome (*A. stenosperma* and *A. cardenasii*). A total of 117 loci were mapped in 11 linkage groups. Another RFLP-based map was published for a tetraploid population (Burow et al. 2001). A synthetic amphidiploid *A. batizocoi* × (*A. cardenasii* × *A. diogoi*)^{4x}, used as a donor parent, was crossed with *A. hypogaea* to generate a BC₁ population of 78 progeny. Three hundred and seventy RFLP loci were mapped onto 23 linkage groups. A BC₄F₂ population obtained through this crossing schedule was used for the identification of RAPD markers linked to root-knot nematode resistance genes (Burow et al. 1996). RFLP and RAPD markers have also been used to follow the introgression of *A. cardenasii* (AA genome) chromosome segments into *A. hypogaea* genome (Garcia et al. 1995). This analysis has shown that chromosome recombination, and not chromosome substitution, was the mechanism of introgression, which is an important result for the use of wild *Arachis* species in peanut-breeding programs. RAPD, SCAR, and RFLP markers have proved to be useful for the identification of markers tightly linked to nematode resistance genes introgressed from *A. cardenasii* into *A. hypogaea* (Garcia et al. 1996). Recently, AFLP markers were used to map genes conferring resistance to the aphid vector of groundnut rosette disease, which is the most destructive viral disease of peanut in Africa (Herselman et al. 2004).

RFLP analysis is time consuming and labor intensive, while RAPD and AFLP are dominant markers of low information content. Consequently, these are not the ideal tools for high throughput analysis and for the implementation of marker assisted selection (MAS) in peanut-breeding programs. In contrast, microsatellites are multiallelic, highly polymorphic, typically codominant genetic markers, and require small amounts of DNA, since they are based on PCR and gel electrophoresis. Their main disadvantage is the high initial cost for the development of markers using genomic DNA libraries enriched for simple sequence repeat (SSR) sequences. With the availability of expressed sequence tags (ESTs) and other DNA sequence data, development of SSR markers

through data mining has become a fast, efficient, and low-cost option for many plant species. Furthermore, the use of markers developed from different sources should enable a better coverage of the whole genome.

Although the number of microsatellite markers published for peanut has increased considerably in the last 5 years (Hopkins et al. 1999; He et al. 2003; Ferguson et al. 2004; Moretzsohn et al. 2004), it is still not sufficient for the construction of saturated linkage maps. Thus, the objectives of the current study were to (1) develop a large set of microsatellite markers from SSR-enriched genomic DNA libraries, from ESTs, and by data mining published *Arachis* spp. sequences; (2) characterize the newly developed markers in a set of *A. hypogaea* accessions, and (3) construct a microsatellite-based, gene-rich map using a diploid AA genome F₂ population.

Materials and methods

Plant material

A single *A. hypogaea* cv. Tatu plant was used for the construction of the two genomic DNA libraries. EST libraries were made from leaf and root tissues from *A. stenosperma* (accession V10309). Sixteen *A. hypogaea* samples (Table 1) were used for the characterization of the newly developed SSR markers. For map construction, an F₂ population was produced by selfing a unique F₁ plant derived from a cross between *A. duranensis* (accession K7988), used as the female parent and *A. stenosperma* (accession V10309), used as the male. In order to obtain enough seeds for the generation of the F₂ population, the F₁ plant was cloned by cuttings. These species were chosen because they are easily crossed and their progeny are fertile, they are contrasting at the DNA level (Moretzsohn et al. 2004), related to the cultivated peanut, and they segregate for several traits of interest. All plants were obtained from the Brazilian *Arachis* germplasm collection, maintained at Embrapa Genetic Resources and Biotechnology — CENARGEN (Brasília-DF, Brazil).

DNA extraction

Total genomic DNA was extracted from young leaves using the protocol of Grattapaglia and Sederoff (1994) modified by the inclusion of an additional precipitation step with 1.2 M NaCl. DNA concentration was estimated by agarose gel electrophoresis comparing the fluorescence intensities of the ethidium bromide stained samples to those of λ DNA standards.

Genomic libraries construction

Total genomic DNA libraries enriched for dinucleotide repeats were constructed following a standard protocol

Table 1 *Arachis hypogaea* accessions analyzed in this study

Number	Accession	Species/subspecies/variety (Type)	Collection site
1	cv. BR-1	<i>A. hypogaea</i> ssp. <i>fastigiata</i> var. <i>fastigiata</i> (Valencia)	Paraíba, Brazil
2	cv. Runner886	<i>A. hypogaea</i> ssp. <i>hypogaea</i> var. <i>hypogaea</i> (Virginia)	São Paulo, Brazil
3	cv. Tatu	<i>A. hypogaea</i> ssp. <i>fastigiata</i> var. <i>fastigiata</i> (Valencia)	São Paulo, Brazil
4	Mf 1538	<i>A. hypogaea</i> ssp. <i>hypogaea</i> var. <i>hirsuta</i>	Pichincha, Ecuador
5	Mf 1560	<i>A. hypogaea</i> ssp. <i>fastigiata</i> var. <i>peruviana</i>	Pastaza, Ecuador
6	Mf 1678	<i>A. hypogaea</i> ssp. <i>fastigiata</i> var. <i>aequatoriana</i>	Sucurubios, Ecuador
7	Mf 2389	<i>A. hypogaea</i> ssp. <i>fastigiata</i> var. <i>aequatoriana</i>	Unknown
8	Mf 3207	<i>A. hypogaea</i> ssp. <i>fastigiata</i> var. <i>vulgaris</i> (Spanish)	Unknown
9	Mf 3618	<i>A. hypogaea</i> ssp. <i>hypogaea</i> var. <i>hirsuta</i>	Unknown
10	Of 109	<i>A. hypogaea</i> ssp. <i>hypogaea</i> var. <i>hypogaea</i> (Xingu type)	Mato Grosso, Brazil
11	Of 128	<i>A. hypogaea</i> ssp. <i>hypogaea</i> var. <i>hypogaea</i> (Xingu type)	Mato Grosso, Brazil
12	Sv 429	<i>A. hypogaea</i> ssp. <i>fastigiata</i> var. <i>peruviana</i>	Yurimaguas, Peru
13	V 10067	<i>A. hypogaea</i> ssp. <i>hypogaea</i> var. <i>hypogaea</i> (Virginia)	Santa Catarina, Brazil
14	V 13009	<i>A. hypogaea</i> ssp. <i>hypogaea</i> var. <i>hypogaea</i> (Virginia)	Tocantins, Brazil
15	V 13115	<i>A. hypogaea</i> ssp. <i>fastigiata</i> var. <i>vulgaris</i> (Spanish)	Minas Gerais, Brazil
16	V 6265	<i>A. hypogaea</i> ssp. <i>fastigiata</i> var. <i>fastigiata</i> (Valencia)	Bahia, Brazil

(Rafalski et al. 1996) with minor modifications. Peanut total genomic DNA, was isolated from a single *A. hypogaea* plant (cv. Tatu) and aliquots were digested separately with three different enzymes (*Mse*I, *Sau*3AI, and *Tsp*509I) to identify which one produced the most adequate fragment profile for library development. Fragments ranging in size from 200 to 800 bp were transferred to NA 45 DEAE cellulose membranes (Schleicher & Schuell BioScience, NH, USA), after electrophoresis in 1.5% low-melting agarose gel. DNA was resuspended in TE buffer and ligated to *Sau*3AI adaptors (Short adaptor: 5'-CAGCCTAGAGCCGA-ATTCACC-3' and Long adaptor: 5'-pGATCGGTGAATTCGGCTCTAGGCTG-3'). Two genomic libraries enriched for dinucleotide repeats were constructed. Fragments were selected by hybridization with biotinylated dinucleotide repeats ((TC)₁₃ and (AC)₁₃) and recovered by magnetic beads linked to streptavidin (Dynabeads Streptavidin, Dynal Biotech., Norway). Enriched fractions were amplified by PCR, using the Short *Sau* adaptor as a primer. Amplification products were purified using QIAquick PCR purification system (Qiagen, CA, USA), cloned into plasmid pGEM-T (Promega, WI, USA), and then transformed into *E. coli* strain XL1-Blue (Promega) by electroporation. Transformed cells were grown on LB-Amp plates (50 µg/ml ampicillin) containing X-Gal and IPTG at 37°C. Plasmid DNA of the white colonies was extracted by alkaline lysis (Mini-preps) and prepared for sequencing. Mini-prep followed the protocol described by Sambrook et al. (1989), with some minor adaptations for extraction in 96-well plates.

EST libraries construction

Two cDNA libraries were made, each with either young leaves or roots of *A. stenosperma*. Total RNA was extracted using Trizol (Invitrogen, CA, USA) and mRNA purified using the Poly A⁺ RNA-Oligotex (Qiagen, CA,

USA) according to the manufacturer's protocol. Full-length cDNA libraries were constructed using the SMART cDNA synthesis kit in λTriplEx2 (Clontech, CA, USA). The resulting cDNA was packed into λ phages using the Gigapack III Gold packaging kit (Stratagene, CA, USA). From the initial plating, the libraries were estimated to contain 10⁷ (roots) and 10⁸ (leaves) recombinant clones. The pTriplEx2 phagemid clones in *E. coli* were obtained using the mass in-vivo excision protocol according to the manufacturer's instructions (Clontech).

DNA sequencing and primer design

Sequencing reactions were performed with T7 and SP6 primers and the Big-Dye Terminator Cycle Sequencing Kit, version 3.1 (Applied Biosystems, CA, USA) or the DYEnamic ET Terminator Cycle Sequencing Kit (Amersham Pharmacia Biotech) using the Applied Biosystems automated DNA 3100 and 377 sequencers. For the ESTs, the cDNA inserts were sequenced from the 5' end using specific PCR primers for the λTriplEx2. To facilitate the selection of SSR-containing sequences from all libraries and the efficient assembly of redundant sequences and forward and reverse sequencing runs, a dedicated module (available from WSM — wsmartins@wsmartins.net — on request) has been developed which integrates the Staden package interface for sequence assembly (Staden et al. 2003) and Troll for finding SSRs (Castelo et al. 2002). In addition to the ESTs and the enriched genomic libraries, the module was used to “data-mine” SSR markers from data available in GenBank. The new sequences identified in this study were submitted to the National Center for Biotechnology Information database (NCBI, USA; GenBank accession #s DQ099037-DQ099242). Primers complementary to unique DNA sequences flanking the SSRs were designed using the computer program Primer 3 (Whitehead Institute of Biomedical Research —

genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi). To reduce problems with spurious banding patterns, some stringent criteria in primer sequence design were applied: (1) primer T_m ranging from 57 to 65°C; (2) 1°C difference in T_m between primer pairs; (3) GC content ranging from 40 to 60%; (4) absence of complementarity between primers. Default values were used for the other parameters, except the “product size ranges” that was between 100 and 400 bp.

PCR amplification of SSR marker loci

PCR reactions were performed in 13 μ l volumes, containing 1X PCR buffer (10 mM Tris-HCl pH 8.3, 50 mM KCl), 2.0 mM MgCl₂, 0.2 mM of each dNTP, 1 U of *Taq* DNA polymerase, 1 μ g/ μ l of purified BSA (New England Biolabs, MA, USA), 5 pmol of each primer and 10 ng of genomic DNA. Amplifications were performed using either a 9700 System (Applied Biosystems, CA, USA) or a PTC-100 (MJ Research, MA, USA) thermal cycler, with the following conditions: 94°C for 5 min (1 cycle), 94°C for 1 min, 48–64°C for 1 min, 72°C for 1 min (30 cycles); and 72°C for 7 min (1 cycle). The annealing temperature was optimized for each primer pair to produce clear DNA band amplification, without spurious fragments. For genotype determination, the amplified products were separated on 5% denaturing polyacrylamide gels stained with silver nitrate (Creste et al. 2001). Some SSR markers highly contrasting between the progenitors of the mapping population were run on 3.5% agarose gels stained with ethidium bromide. Fragment sizes were estimated by comparing with a 10-bp DNA ladder standard (Gibco/BRL, MD, USA).

Characterization of the newly developed SSR markers

The 271 newly developed SSR markers were first tested on a set of six *A. hypogaea* accessions, one from each of the six described peanut varieties (Table 1 — accessions 1, 4, 5, 6, 8, and 14). The expected size and the quality of the amplified fragments were observed on 3.5% agarose gels. When necessary, the annealing temperatures were adjusted to obtain better specificity of PCR amplification. Primers developed from genomic libraries are named by Ah1TC or Ah2TC (for *A. hypogaea* library enriched for TC repeats) or Ah2AC (for AC repeats) followed by the plate number (TC1–TC11 or AC1–AC3) and the clone identification (A01–H12). Primers developed from cDNA libraries are named by AS1 (for *A. stenosperma* libraries), followed by two letters (ML, for mature leaf, RN or RI for root) and the plate number (from 1 to 22) and the clone identification (A01–H12). Primers obtained by data-mining GenBank were denominated by the name of the submitted sequence. These marker names start with GI, pPGP or pPGS. A subjective score was given to each marker from 1 to 3 denoting the fragment amplification

quality, 3 being the best. The number of alleles per locus, the product size ranges, and gene diversity (GD) were estimated for the polymorphic markers using a set of 16 peanut accessions (Table 1). Gene diversity was calculated according to the formula: $GD = 1 - (\sum p_i^2)$, where p_i is the frequency of the i th allele in the 16 samples (Nei 1973). For the characterization of the sequences used for marker development, BLAST databases were formatted from all putative proteins encoded by the *Arabidopsis* genome and all Fabaceae proteins in GenBank. Similarity searches were done using BlastX (Altschul et al. 1990), and hits were considered significant if they had an E-value of less than 1×10^{-8} .

Map construction

The 271 newly developed SSR markers plus another 162 already published SSR markers (Hopkins et al. 1999; Palmieri et al. 2002; He et al. 2003; Moretzsohn et al. 2004; Palmieri et al. 2005; Bravo et al. 2005; Gimenes et al. 2005) were screened against the two progenitors. Polymorphic markers were analyzed on the mapping population consisting of 93 F₂ individuals. A χ^2 test was performed to test the null hypothesis of 1:2:1 or 3:1 segregations on all scored markers. Only marker loci that did not show segregation distortion ($P > 0.05$) from the expected 1:2:1 were used for the initial map construction, in order to eliminate spurious linkages. The linkage analysis was done using Mapmaker Macintosh version 2.0 (Lander et al. 1987). A minimum LOD score of 12.0 and maximum recombination fraction (θ) of 0.3 were set as thresholds for linkage groups (LG) determination with the “group” command. The most likely marker order within each LG was estimated by the matrix correlation method using the “first order” command or by the three-point analysis (“three point” command) for groups containing more than seven markers. Marker orders were confirmed by comparing the log-likelihood of the possible orders using multipoint analysis (“compare” command) and by permuting all adjacent triple orders (“ripple” command). Loci that could be ordered with a likelihood ratio support $\geq 1,000:1$ were used to establish a framework map. Codominant marker loci that did not show segregation distortion at $P > 0.01$ were used in this analysis. In a second step, the LOD score was set to 6.0 in order to include new markers in the groups. The “try” command was then used to determine the exact position of the new markers within each group. In a following step, distorted markers and markers segregating 3:1 (present in one parent and absent in the other) were included using the “group” command. The new marker orders were again confirmed with the “first order”, “compare”, and/or “ripple” commands. Recombination fractions were converted into map distances in centimorgans (cM) using the Kosambi’s mapping function. Genome map sizes were estimated according to Hulbert et al. (1988) taking into consideration only pairwise comparisons between markers placed on the framework map.

Results and discussion

Development of microsatellite markers

Digestion of the *A. hypogaea* genomic DNA with three different enzymes (*Mse*I, *Sau*3AI, and *Tsp*509I) revealed that *Sau*3AI produced the most adequate profile for library development, with fragments ranging from 200 to 800 bp in size. *Sau*3AI is sensitive to cytosine methylation and therefore is expected to preferentially cut in hypomethylated gene-rich DNA, producing libraries enriched for low copy sequences that are usually gene-rich. Two libraries were constructed based on dinucleotide repeat motifs (TC and AC). These repeats were chosen because they are among the most abundant SSR repeats in plant species (Wang et al. 1994; Gupta and Varshney 2000). Out of the 1,152 positive clones (white colonies) sequenced, 224 (19.4%) had SSRs, 210 being non-redundant. When necessary, these clones were sequenced on the reverse strand to enable the construction of high-quality contigs. Of the 210 unique SSR positive clones, 121 (Electronic supplementary material — ESM 1) were suitable for primer design (10.5% of the clones sequenced). The design of primers for the other 89 unique SSRs identified was not possible due to the low GC content of the regions flanking the SSR, or the close proximity of the SSR to the end of the cloned insert DNA. The percentage of primers designed, in relation to the number of clones sequenced (10.5%), is significantly lower than the proportion found (41.4%) for peanut in a previous study that included an additional selection of positive clones by anchored-PCR (Moretzsohn et al. 2004). This shows that this step is effective in significantly improving the yield of useful clones sequenced and reducing costs. However, it is time consuming and its omission did accelerate the process of SSR marker development.

A total of 2,740 EST sequences were screened for SSRs. Ninety-eight showed di-, tri-, tetra- or pentanucleotide repeats (3.6%) and 81 (3.0%) were suitable for primer design (ESM 1). These numbers are similar to the percentages reported for other plant species, ranging from 1.7 in wheat to 7.5 in barley (Varshney et al. 2002; Thiel et al. 2003; Eujayl et al. 2004; Gao et al. 2004; Khlestkina et al. 2004).

Finally, sequences available in GenBank, for which SSR primers had not been designed, were screened for the presence of repeats (>5 core units). A total of 69 new microsatellite primer pairs were designed (ESM 1), 38 from genomic sequences, 25 from ESTs, 1 from a patented sequence, and 5 from *A. hypogaea* genes, such as cold stress responsive protein, desaturase, and seed galactose-binding lectin.

Screening for SSR sequences was done using the Staden-TROLL module, and was quick and convenient with complete integration into the Staden Package. All SSR containing sequences, including those for which primers had been previously published, were used to construct a peanut SSR database in the Staden Gap4

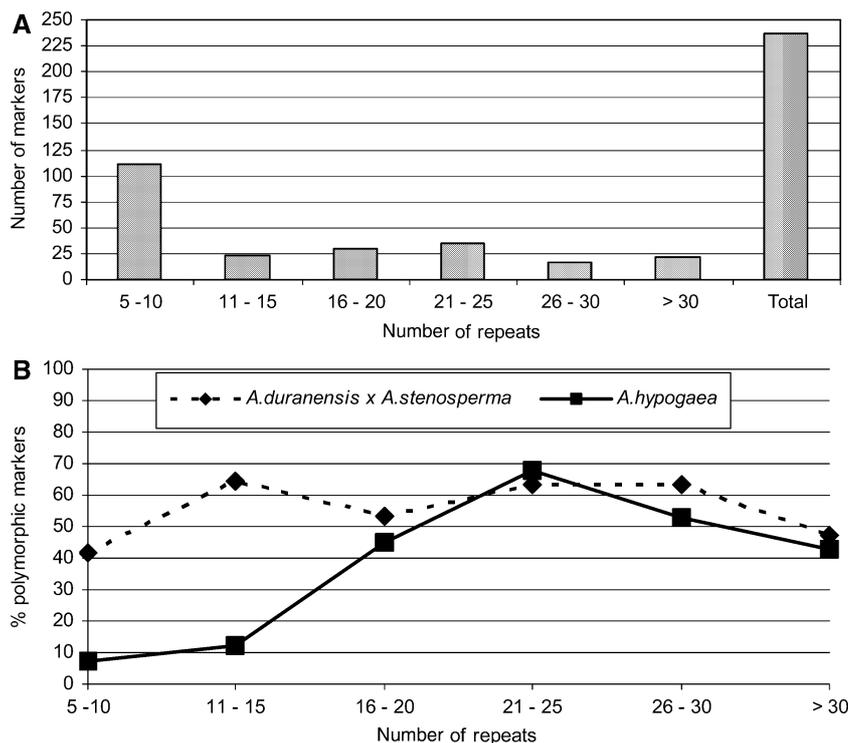
interface. Hundreds of contigs were established thus eliminating redundant sequences. By doing this, we could be sure that all the 271 newly developed SSR markers were unique. Additionally a BlastN search of these sequences against a database of all published *Arachis* SSR containing sequences further ensured that all markers were new.

Characteristics of *Arachis* SSR repeats

Of the 271 sequences for which primers were designed, 159 (58.7%) had perfect repeats, 82 (30.2%) had compound repeats, and 30 (11.1%) had imperfect repeats. Repeats in the TC/AG- and AC/TG-enriched genomic libraries were, as expected, almost entirely TC/AG and AC/TG respectively. For the ESTs, in the 106 markers developed, a total of 153 repeats were found, since some sequences had more than one repeat motif. The trinucleotide motifs were the most abundant type of SSRs detected, with 112 (73.2%) SSRs, followed by di- (22.2%), tetra- (3.3%) and pentanucleotide repeats (1.3%). These proportions are variable in the few published studies of EST derived microsatellites in plants. For trinucleotides, percentages have ranged from 30.2% in maize to 90.3% in sugarcane; dinucleotide motifs from 9.7% in sugarcane to 39.0% in soybean; tetranucleotide motifs from 2.2% in soybean and *Arabidopsis* to 10.3% in *Medicago*; and pentanucleotide repeats from 0.6% in barley to 9.0% in *Medicago* (Cordeiro et al. 2001; Varshney et al. 2002; Gao et al. 2003; Thiel et al. 2003; Eujayl et al. 2004; Tian et al. 2004). However, all these studies have shown that the trimeric SSRs are more abundant in plant EST sequences than the other SSR unit-size classes. These findings are not unexpected and have also been reported for other higher eukaryotes including *Drosophila*, *Caenorhabditis elegans*, mammals, humans, and yeast (reviewed by Li et al. 2004). This has been attributed to negative selection against frameshift mutations in coding regions (Metzgar et al. 2000) and possibly by positive selection for specific single amino-acid stretches (Morgante et al. 2002).

AAG/TTC (27 out of 153), AAT/TTA (21/153), and AAC/TTG (14/153) motifs were the most abundant trinucleotides. Considerable differences in the frequencies of trinucleotide repeat types have also been reported, but these three classes were usually between the most common tri-motifs in other dicot species (Cardle et al. 2000; Gao et al. 2003; Eujayl et al. 2004; Tian et al. 2004). In several monocot species, CCG/GGC repeats have been shown to be the most common motif (Cordeiro et al. 2001; Morgante et al. 2002; Varshney et al. 2002; Gao et al. 2003), but were detected in only two sequences in the present study. The AAT/TTA motifs are the least common (<1%) in animals and in monocot species (Li et al. 2004). In contrast, these motifs were the second most frequent motif found in *Arachis* ESTs and had also high frequencies in soybean and *Medicago* (Gao et al. 2003; Eujayl et al. 2004; Tian et al. 2004).

Fig. 1 Number of markers detected per repeat size class (a) and relationships between repeat numbers of the SSRs and the polymorphism observed when markers were screened against *A. duranensis* and *A. stenosperma*, and in six accessions of *A. hypogaea* (b). A total of 225 markers that amplified products with good resolution were considered in the analysis of *A. duranensis* and *A. stenosperma*, and 236 markers in the analysis of *A. hypogaea*



AG/TC and AT/TA (16 out of 153 each) had the highest frequencies among the dinucleotide repeats, followed by AC/TG (6/153). CG/GC repeats were not found in any of the sequences. Similar results have been described for several plant species, with AG/TC repeats showing the highest frequencies and CG/GC the lowest (Kantety et al. 2002; Eujayl et al. 2004; Tian et al. 2004). Although frequencies of AT/TA repeats have been usually lower than those found in the present study, high frequencies have also been described for other legume species, such as *Medicago* and soybean (Morgante et al. 2002; Gao et al. 2003; Eujayl et al. 2004). The putative functions of SSRs in ESTs are still not clear as are the reasons why repeat motif types are so variable in the different plant species. However, these findings showed that the predominance of a particular repeat motif in one species is not always an indication of representation in another plant species.

Polymorphism and transferability of the new SSR markers

The 271 new SSR markers (ESM 1) were screened for polymorphism against six *A. hypogaea* samples, being one from each of the six described varieties and against the progenitors of the diploid mapping population. As expected, polymorphism between the two accessions of different species was greater than for the six accessions within the cultivated species (Fig. 1, Table 2).

For the wild species progenitors, 27 (10.0%) primer pairs did not amplify any interpretable fragment, 11 (4.1%) amplified fragments only in *A. stenosperma* and 4

(1.5%) in *A. duranensis*. Thus, 229 (84.5%) primer pairs were fully transferable between the two wild species. For these 229 markers, 113 (49.3%) were polymorphic.

For *A. hypogaea*, 37 primer pairs (13.7%) did not amplify any clearly interpretable fragment. Of the remaining 234 primer pairs, 66 (28.2%) were polymorphic for at least one of the six samples. Sixty-two polymorphic markers were used for the analysis of allelic diversity using a sample of 16 *A. hypogaea* accessions (Table 3). The other four polymorphic markers (AC2C2, RN0-602, Seq14G3, and TC1E6) were not included in this analysis, because they resulted in very poor amplification products. The number of alleles detected ranged from 2 to 12 at each of the 62 polymorphic loci analyzed, with an average of 5.87 alleles per locus. GD values ranged from 0.121 (for loci TC11B11 and GI-385) to 0.910 (for locus TC2D6), with an average value of 0.690. GD is an estimate of the probability that two randomly chosen genes from a population are different. The values obtained in the present study are probably overestimated, since some of the markers amplified alleles at duplicated loci. Thus, they cannot be compared to those obtained for diploid species, but the average value (0.690) is similar to values found for several polyploid species, including peanut (Provan et al. 1996; Chavarriaga-Aguirre et al. 1998; Prasad et al. 2000; He et al. 2003; Ferguson et al. 2004; Moretzsohn et al. 2004). In contrast, average GD values estimated by RFLP markers were considerably lower for five *Arachis* species, including *A. hypogaea* (Paik-Ro et al. 1992). These results demonstrate the usefulness of microsatellite markers for genetic diversity analysis of cultivated peanut.

Table 2 Number of polymorphic markers detected by a screening against accessions K7988 (*A. duranensis*) and V10309 (*A. stenosperma*), used as progenitors of the F₂ mapping population, and against six *A. hypogaea* accessions belonging to the six described varieties

Repeat type	Genomic library				cDNA library
	TC/AG	AC/TG	Various	Total	Various
K7988 × V10309					
Polymorphic markers					
Codominant	48 (48.5%)	5 (22.7%)	13 (34.2%)	66 (41.5%)	44 (41.5%)
Dominant	13 (13.1%)	–	–	13 (8.2%)	2 (1.9%)
Monomorphic markers	32 (32.3%)	15 (68.2%)	17 (44.7%)	64 (40.3%)	49 (46.2%)
No amplification	6 (6.1%)	2 (9.1%)	8 (21.1%)	16 (10.0%)	11 (10.4%)
Total	99	22	38	159	106
<i>A. hypogaea</i>					
Polymorphic markers					
Monomorphic markers	48 (48.5%)	5 (22.7%)	4 (10.5%)	57 (35.8%)	8 (7.5%)
No amplification	42 (42.4%)	13 (59.1%)	27 (71.1%)	82 (51.6%)	82 (77.4%)
No amplification	9 (9.1%)	4 (18.2%)	7 (18.4%)	20 (12.6%)	16 (15.1%)
Total	99	22	38	159	106

The number of repeat units found in the 271 sequences ranged from 5 to 54 (ESM 1). Several studies have suggested that loci with longer repeats are much more likely to be variable (Weber 1990; Thomas and Scott 1993; Udupa et al. 1999; Burstin et al. 2001). For the six accessions of *A. hypogaea* there was a trend, with maximum polymorphism of SSRs with 21–25 repeat units (Fig. 1). Polymorphism for SSRs with 5–10 repeats was low at only 7.2%. For the progenitors of the mapping population, polymorphism varied less with repeat size, and even small SSRs with 5–10 repeats showed more than 40% polymorphism. In contrast, the polymorphism of apparently genic markers derived from libraries that had longer repeats, was similar between the two wild and six cultivated accessions.

For both wild and cultivated germplasm, AG/TC repeats are more polymorphic than AC/TG repeats (Table 2), similar effects of repeat type on polymorphism were also observed by Ferguson et al. (2004).

Primer pairs were developed in the present study by using clones from SSR-enriched genomic libraries and ESTs (159 and 106, respectively). Thus, it was possible to compare polymorphism levels detected by SSR markers from these two sources (Table 2). Generally the number of repetitions was much smaller in ESTs than in genomic sequences and correspondingly the polymorphism of ESTs was much lower in *A. hypogaea* than between the two wild species (7.5 and 35.8% respectively, see Table 2). It has been reported that EST-derived SSRs are less polymorphic than those derived from genomic libraries for other plant species, such as rice (Cho et al. 2000), sugarcane (Cordeiro et al. 2001), bread wheat (Eujayl et al. 2002; Gupta et al. 2003), barley (Thiel et al. 2003), and soybean (Song et al. 2004). This has been attributed to possible selection against alterations in these conserved sequences (Scott et al. 2000). However, for *A. duranensis* and *A. stenosperma* accessions, the percentages were identical: 41.5% for ESTs and for genomic libraries, considering only the codominant markers. The reasons for these differences are not clear, and need further investigation to be elucidated.

Characterization of genic content of libraries used for marker development

A. hypogaea has a large and poorly characterized genome. Comparative genetic mapping with other legumes will allow peanut research to begin benefitting from the knowledge of other genomes, especially those of the model legumes *Lotus japonicus* and *Medicago truncatula*. For such comparisons the placement of genic markers on the *Arachis* map is essential. This map includes EST and genome-derived markers from a number of sources. EST markers are derived from transcribed sequences that are almost entirely genic. Nevertheless, not all the homologies of these markers are easily identified, and from 106 only 67 (63.2%) gave significant BlastX hits.

The gene content of genomic libraries depends upon the way that they are made. Using BlastX we estimated the genic content of different libraries that was used to develop SSR markers for peanut (Table 4). The proportion of sequences giving BlastX hits varied from more than 20% to 0%. There was a clear trend, libraries made using restriction enzymes susceptible to methylation at CG or CNG sites are more gene-rich. The single exception being the relatively high gene content of the sequences from Moretzsohn et al. (2004). There the anomaly can be explained by the library being enriched exclusively for tri-repeat SSRs. All eight BlastX hits encode poly-amino acid tracts in their open reading frames. The top BlastX hits for *Arabidopsis* and legume proteins of all markers is in ESM 1.

Map construction

Out of the 433 SSR markers screened against *A. duranensis*, accession K7988 and *A. stenosperma*, accession V10309, 204 (46.8%) were polymorphic, with 170 segregating in a codominant fashion and 34 in a dominant 3:1 ratio (Fig. 2). Of the 170 codominant markers, 80 did not show distortions from the expected 1:2:1 segregation ratio ($P < 0.05$) and were initially used to establish the LG. Distorted and dominant markers were

Table 3 Primer pairs, a quality score for the amplification products, fragment size range, number of individuals analyzed (*n*), total number of alleles amplified (*A*), and gene diversity (GD) for 62 new polymorphic SSR markers identified for *A. hypogaea* in this study

Loci	Score	Size range	<i>n</i>	<i>A</i>	GD
AC2B3	3	296–308	14	6	0.720
AC2B5	3	200–212	15	5	0.680
AC2C8	2	146–220	15	7	0.828
AC2C12	3	204–212	16	5	0.685
TC1A1	2	202–222	13	6	0.735
TC1A2	3	240–276	16	11	0.815
TC1D2	3	242–278	16	4	0.458
TC1E1	3	154–248	14	11	0.892
TC2A2	3	194–212	15	7	0.800
TC2B1	2	190–215	11	4	0.610
TC2B9	3	190–200	16	5	0.645
TC2C7	2	182–210	16	3	0.492
TC2D6	3	196–224	15	10	0.910
TC2E5	3	102–122	15	8	0.864
TC3B5	3	248–270	12	4	0.710
TC3E2	3	270–310	16	9	0.708
TC3E5	3	358–370	16	5	0.798
TC3G1	2	250–256	15	4	0.745
TC3G5	3	118–130	14	2	0.423
TC3H2	3	280–300	15	4	0.671
TC3H7	2	246–270	16	9	0.863
TC4A2	3	280–284	16	3	0.411
TC4D2	2	202–240	12	5	0.786
TC4D9	2	220–240	14	2	0.389
TC4E10	3	300–344	16	8	0.841
TC4F12	3	220–232	15	6	0.763
TC4G2	2	130–166	11	6	0.784
TC4G10	2	205–220	15	3	0.543
TC4H2	2	184–192	12	4	0.725
TC4H7	2	166–220	13	8	0.742
TC5A6	2	164–200	14	7	0.825
TC5D6	2	174–184	13	5	0.782
TC6E1	3	154–186	15	9	0.802
TC6G9	3	132–146	12	6	0.746
TC6H3	3	210–228	14	8	0.754
TC7A2	2	308–320	15	6	0.706
TC7C6	3	148–176	16	9	0.827
TC7D3	3	142–148	15	4	0.579
TC7E4	3	290–300	16	4	0.694
TC7G10	3	110–142	15	4	0.451
TC7H11	2	340–360	14	7	0.847
TC9B8	3	98–108	15	8	0.777
TC9C12	3	256–300	15	7	0.855
TC9F4	3	122–142	15	6	0.800
TC9F10	3	286–320	15	10	0.871
TC11A2	3	284–292	14	6	0.825
TC11A4	3	172–204	15	12	0.906
TC11B11	2	144–146	16	2	0.121
TC11E4	2	110–136	16	12	0.909
TC11F12	2	182–188	14	2	0.254
TC11H6	3	190–214	16	8	0.847
GI-385	3	330–340	16	2	0.121
GI-832	3	200–210	15	2	0.480
GI-1107	3	360–384	14	5	0.680
RN0-615	2	390–400	15	3	0.625
RN0-681	2	310–350	16	6	0.520
RI1F06	3	312–372	11	8	0.823
RN3E10	3	252–255	11	2	0.416
RN17F12	2	260–278	12	5	0.812
Seq4E8	3	274–286	10	6	0.842
Seq4F10	2	180–188	9	4	0.706
Seq16C7	3	242–258	14	5	0.571
Average	–	–	14.41	5.87	0.690

Table 4 Genic content of different libraries as estimated by BlastX hits (E-values of $<10^{-8}$) against *Arabidopsis* and legume putative proteins

Library	Enzyme(s)	Cleaved methylation sites/cloned fragment	Selection/enrichment	Blast hits/total sequences (%)
EST	NA	NA	di-, tri-, tetra-, and penta-SSRs	67/106 (63.2%)
pPGP (Ferguson et al. 2004)	<i>Pst</i> I	0:4	GT GA AAC ATC AGT ATT CAC CTT CTG ATCC GATA AAAT	31/164 (20.7%)
TC	<i>Sau</i> 3AI	1:1	TC	17/102 (16.7%)
Genomic (Moretzsohn et al. 2004)	<i>Tsp</i> 509I	0:0	TTG	8/87 (9.2%)
AC	<i>Sau</i> 3AI	1:1	AC	2/22 (9.1%)
PPGS (Ferguson et al. 2004)	<i>Sau</i> 3AI/ <i>Bam</i> HI	0.5:0.5	GT GA AAC ATC AGT ATT CAC CTT CTG	7/288 (2.4%)
Genomic (He et al. 2003)	<i>Hind</i> III/ <i>Mse</i> I	0:0	AT GT GA	0/56 (0.0%)

Data for ESTs are presented only for comparison to the genomic libraries. Cleaved methylation sites/cloned fragment presents the average number of methylation sites (CG and CNG sites, where cytosine methylation protects against restriction) cleaved per fragment cloned. For instance, *Pst*I cleaves CTGCAG, and so cleaves two CNG sites. Methylation at either of the cytosines protects *Pst*I sites against restriction, and since both ends of the cloned fragments are cut by *Pst*I, four CNG methylation sites are cut per cloned fragment. For *Sau*3AI that cleaves GATC, the cleavage of methylation sites depends upon the flanking DNA sequence, on average 0.5 CG and 0.5 CNG sites will be cleaved per *Sau*3AI fragment terminal. Sequence data from Ferguson et al. (2004) are the assembled SSR containing sequences made available in their supplementary data, and contains sequences that have not been used for marker development. Data from Moretzsohn et al. (2004) also includes some sequences that were not used for marker development. The data shows a clear trend, libraries made from DNA cleaved by restriction enzymes sensitive to methylation at CG and CNG sites are more gene-rich

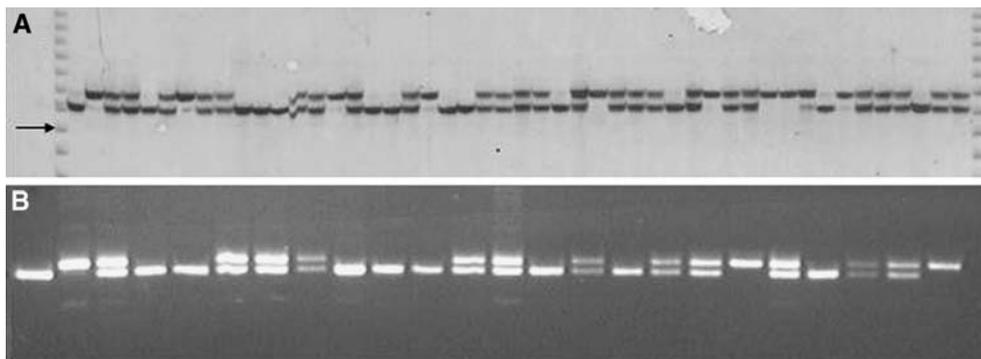


Fig. 2 Amplification patterns obtained with primer RN8C09 in 5% polyacrylamide gel electrophoresis (a) and with primer RN10F09 in 3.5% agarose gel (b). In panel A, the first and the last lanes are 10 bp DNA ladder. In both gels, from left to right: *A. duranensis* (accession K7988) used as the female progenitor, *A. stenosperma* (accession V10309) used as male, the F₁ hybrid, and 45 and 21 F₂ individuals in panels A and B, respectively. Arrow indicates a ladder fragment of 250 bp

further included in the map. Using a minimum LOD score of 12.0 and a maximum recombination fraction (θ) of 0.30, 170 markers mapped into 11 LG covering 1,230.89 cM of total map distance (Fig. 3). This total length is comparable to the 1,063 cM found in a map previously reported from a cross between two AA genome diploid species, *A. duranensis* and *A. cardenasii* (Halward et al. 1993) and to half of the 2,210 cM found for a published tetraploid map (Burow et al. 2001). A few markers were grouped with LOD scores <12.0 (from 6.0 to 12.0) and these are indicated in the map. Groups ranged from 20.07 cM (4 markers) to 280.10 cM (32 markers), with an average distance of 7.24 cM between adjacent markers. Linkage groups were numbered sequentially from the longest to the shortest. Eight SSR primer pairs amplified consistently two loci (AC2C2, TC1B2, TC1D1, TC4H2, TC7H9, Ah573, Ah745, and Seq15D6) and the 16 loci were included in the map. These markers were identified by the numbers -1 and -2 after the marker names (Fig. 3). The framework map, consisting of markers positioned with very high confidence, was composed of 74 markers covering 1,006.20 cM with an average distance of 13.59 cM between markers. The number of markers per linkage group in the framework map ranged from 3 (group 11) to 15 (group 2). Marker names that constitute the framework map were underlined in the map (Fig. 3). *Arachis duranensis* and *A. stenosperma* are diploid species with $2n=20$ chromosomes and, thus, 10 LG were expected instead of the 11 obtained in the present study. By reducing the stringency of the LOD score to 10.0, groups 1 and 2 merged in a single group, but linked with a very high recombination fraction (>40.0 cM). Thus, they were kept as separate groups. In addition, the upper portion of group 3 was joined with markers that constitute the framework map by decreasing the LOD score to 8.0. Thirty-four segregating markers remained un-

linked, suggesting that at least parts of the genome have not been covered by this map. The inclusion of new markers would saturate this map, and probably result in the expected 10 LG.

A total of 104 (51.0%) out of the 204 markers deviated from the expected F₂ ratio of 1:2:1 (90 markers) or 3:1 (14 markers) at $P > 0.05$ level. Of the 90 markers that showed distortion from the expected 1:2:1 segregation ratio, 41 markers were skewed toward the *A. duranensis* allele, 44 markers toward the *A. stenosperma* allele, and 5 markers toward the heterozygote. For the 14 dominant distorted markers, 5 showed an excess of *A. duranensis* alleles and 9 of *A. stenosperma* alleles. The proportion of distorted markers found in this study is higher than the 25% found for the two RFLP-based maps published for *Arachis* (Halward et al. 1993; Burow et al. 2001). Distorted markers were scattered on LG 1, 3, 4, 5, and 6 (Fig. 3). Groups 1, 3, and 6 had markers with an excess of both *A. duranensis* and *A. stenosperma* alleles, while all the distorted markers of group 4 and group 5 were skewed toward the *A. duranensis* and *A. stenosperma* alleles, respectively (Fig. 3). Segregation distortion has been described for most of the published molecular marker-based linkage maps in plants, and has been attributed to different causes, such as: deleterious recessive alleles (Berry et al. 1995), self-incompatibility alleles (Barzen et al. 1995), structural rearrangements (Quillet et al. 1995) or differences in DNA content (Jenczewski et al. 1997). Whatever the cause is, segregation distortion in some chromosomes, or in parts of chromosomes, must have important implications for using this map information in a MAS program. New genetic maps and cytogenetic experiments that are in progress will enable a better understanding and use of these findings.

Thirty-six of the 170 loci placed on the map are ESTs or characterized genes. Of the other 134 that were developed from genomic sequences, 25 gave BlastX hits. Therefore, about 61 of the mapped microsatellites most likely represent genes. As expected, homologies are diverse, but they include ones of possible interest for: seed quality (seed allergen and seed maturation protein); disease resistance (lectin, protease inhibitor, cinnamoyl-CoA reductase, and wound-induced protein); and drought resistance (dehydration responsive element binding protein, and aquaporin) (ESM 1).

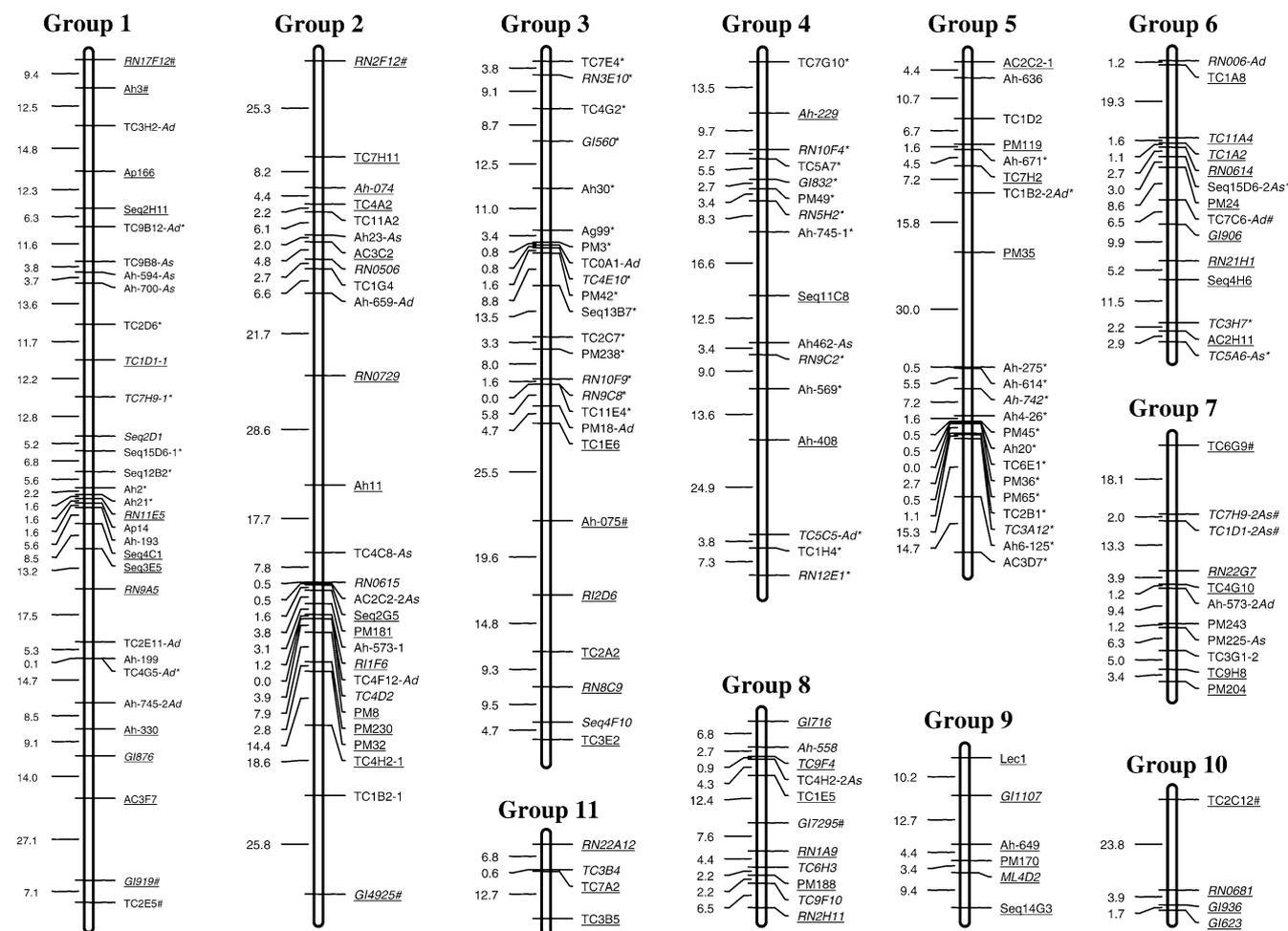
Total genome size was estimated using the method of Hulbert et al. (1988). Only framework markers were used in this procedure to avoid an overestimate of genome coverage. A total map distance of 1,424.70 cM was estimated, of which 1,006.20 cM (70.6%) and 1,230.89 cM (86.4%) were covered by the framework map and by the total map constructed in this study, respectively. The haploid genome size of *A. hypogaea* was estimated as being 2,813 Mbp (Bennet and Leitch 2005). If wild species have similar genome sizes and given the estimated total map distance, the average physical equivalent of 1 cM would correspond to 1.97 Mbp in the *Arachis* genome.

Conclusions

This is the first microsatellite based and gene-rich linkage map published for *Arachis*. Since *A. hypogaea* has a relatively narrow genetic base, wild species of *Arachis* constitute an important source of genes for traits of interest to peanut breeding programs. The availability of moderate coverage, only microsatellite map will certainly represent a considerable improvement in our ability to map useful genes and to implement MAS. The validation of the LG obtained in the present study, both

in terms of locus ordering and recombination fraction estimates, is now in progress using two other mapping populations: a diploid F₂ generated by crossing two BB genome species (*A. ipaënsis* × *A. magna*) and a tetraploid F₂ obtained by crossing an *A. hypogaea* accession (cv. IAC Runner 886) and an amphidiploid plant resulting from a cross between two wild diploid species (*A. duranensis* — AA genome × *A. ipaënsis* — BB genome) treated with colchicine. The development of genetic

Fig. 3 Peanut genetic linkage map, based on SSR markers, obtained through the analysis of 93 F₂ plants, generated from a cross between two diploid wild species with AA genome, *A. duranensis* × *A. stenosperma*. Using a minimum LOD score of 12.0 and a maximum recombination fraction (θ) of 0.30, 170 markers mapped into 11 LG. Marker names that grouped with LOD scores < 12.0 have a symbol # after the marker name. Linkage groups were numbered sequentially from the longest to the shortest. Probable genic markers are shown in *italics*. Framework map was composed by 74 markers that could be ordered with a likelihood support > 1,000:1. Marker names that constitute the framework map were underlined in the map. Numbers on the left of each group are Kosambi map distances. Markers that showed significant distortions from the 1:2:1 segregation ratio are indicated by an asterisk. Dominant markers were identified by *Ad* or *As* after the marker name if they amplified alleles from *A. duranensis* or *A. stenosperma*, respectively. Markers that amplified two loci have numbers 1 and 2 after the marker name



maps for *Arachis* diploid wild species with AA and BB genomes will facilitate detailed inheritance studies of co-segregating microsatellites and traits of interest in the tetraploid population. Transferability of SSR loci to different species will allow the construction of a reference map for *Arachis* and the mapping of QTLs in different genetic backgrounds. Additionally, the present SSR-based map is the only published map for *Arachis* constructed with markers derived from sequences that are all currently available in GenBank. Thus, researchers from different laboratories can use these markers for map development using other mapping populations and contributing to the construction of a consensus map for *Arachis*. Markers linked to resistance genes are being identified to potentially implement MAS strategies in peanut-breeding programs, therefore accelerating the introgression of wild genes into cultivated peanut.

Acknowledgments The authors thank Raphael FTM Oliveira, Aryanne G Amaral, and Juliana Dias for their technical assistance. This research is part of the Ph.D. Thesis of Marcio de Carvalho Moretzsohn and has been supported by the European Union: INCO-DEV, Contract ICA4-CT-2001-10072 Project "ARAMAP" and by the World Bank and EMBRAPA "The Agricultural Technology Development Project for Brazil" (PRODETAB), project 004/01/01.

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410
- Barzen E, Mechelke W, Ritter E, Schulte-Kappert E, Salamini F (1995) An extended map of the sugar beet genome containing RFLP and RFLP loci. *Theor Appl Genet* 90:189–193
- Bennet MD, Leitch IJ (2005) Nuclear DNA amounts in angiosperms: progress, problems and prospects. *Ann Bot* 95:45–90
- Berry ST, Leon AJ, Hanfrey CC, Challis P, Burkholz A, Barnes SJ, Rufener GK, Lee M, Caligari PDS (1995) Molecular marker analysis of *Helianthus annuus* L. 2. Construction of an RFLP linkage map for cultivated sunflower. *Theor Appl Genet* 91:195–199
- Bravo JP, Hoshino AA, Angelici CMLCD, Lopes CR, Gimenes MA (2005) Analysis of genetic variability in germplasm of *Arachis* section *Arachis* by using microsatellites. *Gen Mol Biol* (in press)
- Burow MD, Simpson CE, Paterson AH, Starr JL (1996) Identification of peanut (*Arachis hypogaea* L.) RAPD markers diagnostic of root-knot nematode (*Meloidogyne arenaria* (Neal) Chitwood) resistance. *Mol Breed* 2:369–379
- Burow MD, Simpson CE, Starr JL, Paterson AH (2001) Transmission genetics of chromatin from a synthetic amphidiploid to cultivated peanut (*Arachis hypogaea* L.): broadening the gene pool of a monophyletic polyploid species. *Genetics* 159:823–837
- Burstin J, Deniot G, Potier J, Weinachter C, Aubert G, Baranger A (2001) Microsatellite polymorphism in *Pisum sativum*. *Plant Breed* 120:311–317
- Cardle L, Ramsay L, Milbourne D, Macaulay M, Marshall D, Waugh R (2000) Computational and experimental characterization of physically clustered simple sequence repeats in plants. *Genetics* 156:847–854
- Castelo A, Martins WS, Gao G (2002) TROLL: tandem repeat occurrence locator. *Bioinformatics J* 18:634–636
- Chavarriaga-Aguirre P, Maya MM, Bonierbale MW, Kresovich S, Fregene MA, Tohme J, Kochert G (1998) Microsatellites in Cassava (*Manihot esculenta* Crantz): discovery, inheritance and variability. *Theor Appl Genet* 97:493–501
- Cho YG, Ishii T, Temnykh S, Chen X, Lipovich L, McCouch SR, Park WD, Ayres N, Cartinhour S (2000) Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). *Theor Appl Genet* 100:713–722
- Cordeiro GM, Casu R, McIntyre CL, Manners JM, Henry RJ (2001) Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to erianthus and sorghum. *Plant Sci* 160:1115–1123
- Creste S, Tulmann Neto A, Figueira A (2001) Detection of single sequence repeat polymorphisms in denaturing polyacrylamide sequencing gels by silver staining. *Plant Mol Biol Rep* 19:299–306
- Eujayl I, Sledge MK, Wang L, May GD, Chekhovskiy K, Zwonitzer JC, Mian MAR (2004) *Medicago truncatula* EST-SSRs reveal cross-species genetic markers for *Medicago* spp. *Theor Appl Genet* 108:414–422
- Eujayl I, Sorrells ME, Baum M, Wolters P, Powell W (2002) Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat. *Theor Appl Genet* 104:399–407
- FAO (2003) <http://faostat.fao.org/faostat/collections?subset=agriculture>
- Ferguson ME, Burow MD, Schulze SR, Bramel PJ, Paterson AH, Kresovich S, Mitchell S (2004) Microsatellite identification and characterization in peanut (*A. hypogaea* L.). *Theor Appl Genet* 108:1064–1070
- Gao LF, Jing RL, Huo NX, Li Y, Li XP, Zhou RH, Chang XP, Tang JF, Ma ZY, Jia JZ (2004) One hundred and one new microsatellite loci derived from ESTs (EST-SSRs) in bread wheat. *Theor Appl Genet* 108:1392–1400
- Gao LF, Tang JF, Li HW, Jia JZ (2003) Analysis of microsatellites in major crops assessed by computational and experimental approaches. *Mol Breed* 12:245–261
- Garcia GM, Stalker HT, Kochert G (1995) Introgression analysis of an interspecific hybrid population in peanuts (*Arachis hypogaea* L.) using RFLP and RAPD markers. *Genome* 38:166–176
- Garcia GM, Stalker HT, Shroeder E, Kochert G (1996) Identification of RAPD, SCAR, and RFLP markers tightly linked to nematode resistance genes introgressed from *Arachis cardenasii* into *Arachis hypogaea*. *Genome* 39:836–845
- Gimenes MA, Hoshino AA, Barbosa AVG, Palmieri DA, Bravo JP, Lopes CR (2005) Characterization and transferability of microsatellite markers from the cultivated peanut (*Arachis hypogaea* L.). *Bot J Linn Soc* (in press)
- Grattapaglia D, Sederoff R (1994) Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-test-cross: Mapping strategy and RAPD markers. *Genetics* 137:1121–1137
- Gupta PK, Rustgi S, Sharma S, Singh R, Kumar N, Balyan HS (2003) Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. *Mol Gen Genomics* 270:315–323
- Gupta PK, Varshney RK (2000) The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. *Euphytica* 113:163–185
- Halward TM, Stalker HT, Kochert G (1993) Development of an RFLP linkage map in diploid peanut species. *Theor Appl Genet* 87:379–384
- Halward TM, Stalker HT, Larue EA, Kochert G (1991) Genetic variation detectable with molecular markers among unadapted germ-plasm resources of cultivated peanut and related wild species. *Genome* 34:1013–1020
- He G, Meng R, Newman M, Gao G, Pittman RN, Prakash CS (2003) Microsatellites as DNA markers in cultivated peanut (*A. hypogaea* L.). *BMC Plant Biol* 3:3. <http://www.biomedcentral.com/1471-2229/3/3>
- Herselman L, Thwaites R, Kimmins FM, Courtois B, van der Merwe PJA, Seal SE (2004) Identification and mapping of AFLP markers linked to peanut (*Arachis hypogaea* L.) resistance to the aphid vector of groundnut rosette disease. *Theor Appl Genet* 109:1426–1433

- Hopkins MS, Casa AM, Wang T, Mitchell SE, Dean R, Kochert GD, Kresovich S (1999) Discovery and characterization of polymorphic Simple Sequence Repeats (SSRs) in peanut. *Crop Sci* 39:1243–1247
- Hulbert SH, Ilott TW, Legg EJ, Lincoln SE, Lander ES, Michelmore RW (1988) Genetic analysis of the fungus, *Bremia lactucae*, using Restriction Fragment Length Polymorphisms. *Genetics* 120:947–958
- Jenczewski E, Gherardi M, Bonnin I, Prosperi JM, Olivieri I, Huguet T (1997) Insight on segregation distortions in two intraspecific crosses between annual species of *Medicago* (Leguminosae). *Theor Appl Genet* 94:682–691
- Kantety RV, La Rota M, Matthews DE, Sorrells ME (2002) Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. *Plant Mol Biol* 48:501–510
- Khlestkina EK, Than MH, Pestsova EG, Roder MS, Malyshev SV, Korzun V, Borner A (2004) Mapping of 99 new microsatellite-derived loci in rye (*Secale cereale* L.) including 39 expressed sequence tags. *Theor Appl Genet* 109:725–732
- Krapovickas A, Gregory WC (1994) Taxonomía del género *Arachis* (Leguminosae) Bonplandia 8:1–186
- Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L (1987) MAPMAKER: an interactive computer package for constructing primary genetic maps of experimental and natural populations. *Genomics* 1:174–181
- Li YC, Korol AB, Fahima T, Nevo E (2004) Microsatellites within genes: structure, function, and evolution. *Mol Biol Evol* 21:991–1007
- Metzgar D, Bytof J, Wills C (2000) Selection against frameshift mutations limits microsatellite expansion in coding DNA. *Genome Res* 10:72–80
- Moretzsohn MC, Hopkins MS, Mitchell SE, Kresovich S, Valls JFM, Ferreira ME (2004) Genetic diversity of peanut (*Arachis hypogaea* L.) and its wild relatives based on the analysis of hypervariable regions of the genome. *BMC Plant Biol* 4:11. <http://www.biomedcentral.com/1471-2229/4/11>
- Morgante M, Hanafey M, Powell W (2002) Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nat Genet* 30:194–200
- Nei M (1973) Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci USA* 70:3321–3323
- Paik-Ro OG, Smith RL, Knauff DA (1992) Restriction fragment length polymorphism evaluation of six peanut species within the *Arachis* section. *Theor Appl Genet* 84:201–208
- Palmieri DA, Hoshino AA, Bravo JP, Lopes CR, Gimenes MA (2002) Isolation and characterization of microsatellite loci from the forage species *Arachis pintoi* (Genus *Arachis*). *Mol Ecol Notes* 2:551–553
- Palmieri DA, Bechara MD, Curi RA, Gimenes MA, Lopes CR (2005) Novel polymorphic microsatellite markers in section *Caulorrhizae* (*Arachis*, Fabaceae). *Mol Ecol Notes* 5:77–79
- Prasad M, Varshney RK, Roy JK, Balyan HS, Gupta PK (2000) The use of microsatellites for detecting DNA polymorphism genotype identification and genetic diversity in wheat. *Theor Appl Genet* 100:584–592
- Provan J, Powell W, Waugh R (1996) Microsatellite analysis of relationships within cultivated potato (*Solanum tuberosum*). *Theor Appl Genet* 92:1078–1084
- Quillet MC, Madjidian N, Griveau Y, Serieys H, Tersac M, Lorieux M, Bervillé A (1995) Mapping genetic factors controlling pollen viability in an interspecific cross in *Helianthus* sect. *Helianthus*. *Theor Appl Genet* 91:1195–1202
- Rafalski JA, Vogel JM, Morgante M, Powell W, Andre C, Tingey SV (1996) Generating and using DNA markers in plants. In: Birren B, Lai E (eds) *Analysis of non-mammalian genomes — a practical guide*. Academic, New York, pp 75–134
- Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning — A Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press, New York
- Scott KD, Eggler P, Seaton G, Rosetto M, Ablett EM, Lee LS, Henry RJ (2000) Analysis of SSRs derived from grape ESTs. *Theor Appl Genet* 100:723–726
- Song QJ, Marek LF, Shoemaker RC, Lara KG, Concibido VC, Delannay X, Specht JE, Cregan PB (2004) A new integrated genetic linkage map of the soybean. *Theor Appl Genet* 109:122–128
- Staden R, Judge DP, Bonfield JK (2003) Managing sequencing projects in the GAP4 environment. In: Krawetz SA, Womble DD (eds) *Introduction to Bioinformatics. A Theoretical and practical approach*. Human Press Inc, Totawa
- Tian AG, Wang J, Cui P, Han YJ, Xu H, Cong LJ, Huang XG, Wang XL, Jiao YZ, Wang BJ, Wang YJ, Zhang JS, Chen SY (2004) Characterization of soybean genomic features by analysis of its expressed sequence tags. *Theor Appl Genet* 108:903–913
- Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor Appl Genet* 106:411–422
- Thomas MR, Scott NS (1993) Microsatellite repeats in grapevine reveal DNA polymorphisms when analyzed as sequence-tagged sites (STSs). *Theor Appl Genet* 86:985–990
- Udupa SM, Robertson LD, Weigand F, Baum M, Kahl G (1999) Allelic variation at (TAA)_n microsatellite loci in a world collection of chickpea (*Cicer arietinum* L.) germplasm. *Mol Genet* 261:354–363
- Varshney RK, Thiel T, Stein N, Langridge P, Graner A (2002) In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species. *Cell Mol Biol Lett* 7:537–546
- Wang Z, Weber JL, Zhong G, Tanksley SD (1994) Survey of plant short tandem DNA repeats. *Theor Appl Genet* 88:1–6
- Weber JL (1990) Informativeness of human (dC-dA)_n (dG-dT)_n polymorphisms. *Genomics* 7:524–530
- Young ND, Weeden NF, Kochert G (1996) Genome mapping in legumes (Family Fabaceae). In: Paterson AH (ed) *Genome mapping in plants*. RG Landes, Austin, TX

New softwares for automated microsatellite marker development

Wellington Martins, Daniel de Sousa¹, Karina Proite², Patrícia Guimarães², Marcio Moretzsohn² and David Bertoli³

Department of Computer Science, Catholic University of Goiás, Brazil, ¹Department of Computer Science, Catholic University of Rio de Janeiro, Brazil, ²Embrapa Genetic Resources and Biotechnology, Brasília, Brazil and ³Genomic Sciences and Biotechnology, Catholic University of Brasília, Brazil

Received November 23, 2005; Revised and Accepted January 31, 2006

ABSTRACT

Microsatellites are repeated small sequence motifs that are highly polymorphic and abundant in the genomes of eukaryotes. Often they are the molecular markers of choice. To aid the development of microsatellite markers we have developed a module that integrates a program for the detection of microsatellites (TROLL), with the sequence assembly and analysis software, the Staden Package. The module has easily adjustable parameters for microsatellite lengths and base pair quality control. Starting with large datasets of unassembled sequence data in the form of chromatograms and/or text data, it enables the creation of a compact database consisting of the processed and assembled microsatellite containing sequences. For the final phase of primer design, we developed a program that accepts the multi-sequence 'experiment file' format as input and produces a list of primer pairs for amplification of microsatellite markers. The program can take into account the quality values of consensus bases, improving success rate of primer pairs in PCR. The software is freely available and simple to install in both Windows and Unix-based operating systems. Here we demonstrate the software by developing primer pairs for 427 new candidate markers for peanut.

INTRODUCTION

A molecular marker is an identifiable DNA region whose inheritance can be followed along generations. Molecular markers have become a powerful tool in many areas of biology. Microsatellites, also known as simple sequence repeats (SSR) or short tandem repeats (STR), are repeated sequence motifs,

whose unit of repetition is between 1 and 6 bp. They are highly abundant in the genomes of eukaryotes, polymorphic and usually co-dominant and transferable between different mapping populations. Microsatellite markers can also be used in automated genotyping techniques. Thus, they have become one of the most useful molecular markers for a large number of organisms.

Researchers working on the development of microsatellite markers need an efficient way to go from usually hundreds or thousands of trace and/or text sequence files to the identification of new potential markers. However, the softwares currently available for data-mining microsatellites are either restricted to small datasets, not available for Windows, or are not targeted towards marker development. In addition, they do not offer base quality control and with one exception (Repeat-Masker when used with Staden), they are not integrated into a sequence analysis software, and as such require already assembled data as input (1–3), <http://www.maizemap.org/bioinformatics/SSRFINDER/README.ssrfinder> and <http://www.repeatmasker.org/>.

We have developed a Staden module that integrates the program TROLL (4) to the Staden package. TROLL is an open source program based on the Aho-Corasick algorithm for the detection of perfect microsatellites. The Staden Package is a free, open source, software tool used for sequence assembly and analysis (5) (<http://staden.sourceforge.net/>). The Staden package, working with the module and TROLL, enables efficient processing of a large number of sequences in chromatogram and/or text form to produce a small database with only microsatellite containing sequences, processed, assembled in contigs and that have not previously been used for marker development. For the primer design we have developed a program that uses Primer3 (6) to design primers flanking the SSR located by the module. It is available as a stand-alone program and through a Web service. This program accepts multiple sequences in the 'experiment file' format of Staden as input and produces a list of primer pairs

*To whom correspondence should be addressed. Tel: +55 62 32271381; Fax: +55 62 32271073; Email: wsmartins@ucg.br
Correspondence may also be addressed to David Bertoli. Tel: +55 61 34484735; Fax: +55 61 33403658; Email: david@pos.ueb.br

for amplification of microsatellite markers. The program can take into account the quality values of consensus bases, improving success rate of primer pairs in PCR. The Staden Package, TROLL and the module are freely available for both Linux and Windows platforms.

MATERIALS AND METHODS

General details of the software

The TROLL module was developed in the Tcl/Tk scripting language, following the Staden Package module development specifications. It works through Pregap4 reading the Staden Experiment Files one by one. These files contain, among other information, the DNA sequences, the corresponding quality values and any vector contamination masking information. All sequences are then concatenated and a special character (#) is used as a separator. TROLL then locates the SSRs, these are then filtered, disregarding those in low quality or vector contaminated regions and the result is written back on the Experiment Files. A file containing the names of experiment files with microsatellites is generated (pregap.SSR.passed), and this can be used for entering into the Gap4 interface of Staden to create a database of entirely microsatellite containing sequences. The tags added to the experiment files may be used by Gap4 to aid in assembly by masking and to visualize the microsatellites. Additionally the tags are used by the primer design program to define the target region around which the Primer3 designs primers (see below).

The following module configuration options are available:

- (i) Tag type: used to control Tag type used to mark microsatellite repeats.
- (ii) Motif length: used to set the motif lengths (mono, di, tri, tetra, penta and/or hexa) that will be used in the search for SSRs.
- (iii) Repeat min: used to set the minimum number of motif repetitions for SSRs of each motif length.
- (iv) Motif file: the location of the file containing the motif list required by the TROLL program.
- (v) Create SSR.passed file?: for the creation of a file with the names of all sequences containing SSRs. This file is used to enter into Gap4 to create a database of microsatellite containing sequences.
- (vi) Create no_SSR_passed file?: used to create a file with the name of all sequences that do not contain SSRs.
- (vii) Filter quality: used to set the base quality threshold which will be used when analyzing SSR regions.

The program for primer design is responsible for parsing the experiment format file exported from Gap4 and calling the Primer3 program repeatedly to produce primer pairs flanking the SSRs. For each sequence present in the input file, the program analyses the tags added by the TROLL module (default is 'REPT') and produces a list of the repeats found. Since the input file may contain assembled readings, overlapping repeats are discarded. Based on the list of repeats produced, and on the input parameters chosen by the user (e.g.: GC content, TM, primer length etc.), input Primer3 files are created and processed. The program then analyses Primer3 output to cluster repeats being flanked by the same primer pairs. When all sequences have been analysed, the program

outputs a tab spaced file (primers.passed) that includes the contig name, primer names, microsatellite sequence(s), primer sequences, annealing temperatures and the quality of the contig sequences with the primers annealed. In addition, the program provides three other output files, marker_seqs.fasta containing the sequences for which the primer design was successful, not_marker_seqs.fasta with sequences for which primer design was not successful, and primers.failed, which also contains the sequences for which primer design was not successful, but in the experiment file format of Staden. The primers.failed file can be used for another round of analysis using the primer design program with less stringent conditions for the primer design.

Downloading and using the programs

Download and install:

- (i) The Staden package from <http://staden.sourceforge.net/>.
- (ii) The TROLL module (TROLL.p4m) from <http://finder.sourceforge.net/> (follow the link to Resources). Troll_p4m is distributed as a zip file, containing the module itself, troll binaries (Linux and Windows), a DLL Windows library and installation instructions.
- (iii) The primer design program from <http://finder.sourceforge.net/> (Follow the link to Resources). The primer design program is available as a Web service, requiring no program installation. Alternatively, the primer design program can be downloaded as a Perl script and installed locally. For the Perl script to work you will need to install Primer3, which is available from http://frodo.wi.mit.edu/primer3/primer3_code.html.
- (iv) The Step-by-step protocol for use of the module from <http://finder.sourceforge.net/> (Follow the link to Resources). The Step-by-step protocol for use of the module is written for novice users of the Staden Package. However, to use the TROLL module to its full potential, the user needs to be familiar with the Staden. The course documentation that is downloaded with the software is highly recommended (for Windows users in C:\Program files\Staden package\course\).

Examples of marker development from three datasets

TC-repeat enriched genomic library. Peanut (*Arachis hypogaea*) genomic DNA library enriched for TC repeats was made using genomic DNA cut with the restriction enzyme Sau3AI, oligonucleotides and magnetic beads according to a standard protocol (7). Two hundred and eighty genomic clones were sequenced using the forward primer, and plasmids that contained microsatellites were identified using the module. These plasmids were sequenced using the reverse primer, and the resulting 338 chromatograms were processed using the module, as described in the Step-by-Step protocol (Supplementary Data), but also using a naming scheme allowing Staden to identify forward and reverse read pairs from individual clones. After assembly in Gap4, where necessary, forward and reverse reads were joined manually using the 'Find read pairs', 'Find internal joins' and 'Join Contigs' commands. Consensus sequences were exported in experiment file format from the Gap4 interface, and used as input file for the Web interface of the primer design program.

Expressed sequence tags (ESTs). Eight thousand eight hundred ESTs from *Arachis stenosperma* in the form of traces were processed through the Staden Pregap4 interface. The BLAST module was run to eliminate published molecular markers for peanut using a BLAST database (8) formatted from 707 known *Arachis* SSR markers [(9–12); M. Moretzsohn and D. Bertioli, unpublished data] (<http://www.biomedcentral.com/1471-2229/3/3>, <http://www.biomedcentral.com/1471-2229/4/11>); and the TROLL module was run to identify and select di-, tri-, tetra- and penta-microsatellites with six or more motif repetitions. The resultant pregap.SSR.passed file was entered into the Staden Gap4 interface, using the Gap4 assembler with masking, to create a database of microsatellite containing sequences that have not yet been used for marker development. Consensus sequences were exported in experiment file format from the Gap4 interface, and used as input file for the primer design program.

Text data from GenBank. Three thousand eight hundred and ninety *Arachis* sequences from GenBank in simple text format were processed essentially as described in the Step-by-step protocol for text data with the addition of the use of Cross-Match using the UniVec database to remove contaminating vector sequences, and the use of the BLAST module for the elimination of already known *Arachis* markers (as described for the processing of ESTs).

RESULTS AND DISCUSSION

The number of molecular markers available for *Arachis*, and many other species, is still a limiting factor for genetic mapping. When we had begun working with microsatellite marker development we could not find a software that dealt in an integrated way with the problems that marker development presented, most importantly: the identification of microsatellite containing sequences, the removal of vector sequences and known marker sequences, the processing of sequences into non-redundant contigs, and a large-scale interface with Primer3. To overcome this we developed the softwares published here. Recently we published new microsatellite markers developed using this software in various stages of development, together with the first microsatellite-based genetic map

for *Arachis*. In total 271 new microsatellite markers were developed, 69 data-mined from GenBank, 121 from microsatellite enriched genomic libraries and 81 from ESTs (12). The software in the final stage of development described here solves the problems we encountered and allows the efficient data-mining of new markers.

The TROLL module integrates simply and completely into the Staden Package and performed its task extremely fast. By concatenating the input sequences and calling the TROLL program only once, the module finds all SSRs in a single pass, thus maintaining TROLL's linear processing time. TROLL's simplicity and specialized function is targeted exactly at the types of repeats that are most useful for microsatellite marker development: repeats with di-, tri-, tetra- and penta-nucleotide motif sizes. Although TROLL does not find imperfect repeats, we did not find this to be a problem. Long imperfect repeats largely consist of shorter perfect ones, and isolated short imperfect repeats are of very limited use for marker development. The primer design program worked well and, when used with chromatogram data and quality control, effectively eliminated primers being designed to low quality regions of the consensus sequences (for an example output see Table 1). Here we illustrate the use of the module and primer design program with three different datasets, each of which presenting slightly different difficulties for marker development. Staden working with the module proved effective in all cases.

TC-repeat enriched genomic library

Processing the 338 chromatograms produced 43 microsatellite containing contigs. This type of data where every sequence contains a microsatellite, and many of the sequence reads 'die-out' at the end of the microsatellite is difficult to assemble correctly. Incorrect joining of reads that end in microsatellites can easily be a problem. The use of masking in the assembly avoided this problem. The primer design program produced primer pairs for 30 microsatellite markers. The processing of these sequences is included just for demonstration, markers from these sequences are already published (12). Only one of the primer pairs (3.3%) failed to amplify an identifiable PCR product.

Table 1. Example output from the primer design program formatted and simplified

Contig name	Microsatellite	Product size	FWD primer	T_m	REV primer	T_m
AS1RM8P1E08	(CGG)6	218	GAACGCGTCAAAGAACATAACA	60.173	TTAACAACAGCAACAGCTTCGT	59.984
AS2RM11P1H01	(AG)7	352	AAGAAACGACGGTGAATCTGTT	60.040	GCCTTTTCTCTCTCCCTCT	59.621
AS1RM15P1H08	(TC)6	266	GTAGTGGCAGAGCCGTCTTCT	60.081	GTGGCCAATCTGTAAATCCAAT	60.082
AS1RM15P1H08	(CAG)6	309	TCTTCGCCTTGGTATATTGGAT	59.829	CAGGGGTGTAAGTGCAAGTGTA	60.088
AS1RM14P1B10	(CAA)8	316	ATCGAATCCCAACTCACTACT	60.001	GCTTGGGTTCTTGTGGAGTAAC	60.040
AS1RN11P1C04	(TA)8	372	CGGGGAAATTTTATATTTC	58.501	ATTTTGTAAGGAAATCGCTCGT	59.188
AS1RM11P1A04	(TTC)7	181	GAGACCTTATGGTGGATCTTGG	59.827	AACGCTCCGTACAAGAGAAGAG	60.075
AS2RM7P1E10	(AATA)6	170	CACGTGAAAGTTCATATCGTGTC	59.553	ATTCATGATTCCTTACGCGACT	59.993
AS1RM9P1B07	(ATC)8	322	TGGGGAAATCTTAACACAAGG	60.210	GCAAGAAGGAGTTGAAGAAGGA	59.998
AS2RM7P1A09	(AT)8	357	AATCAACTTTGGACTTGTA	54.102	AAGCATAGGGAAAATGAAAT	53.159
AS2RM11P1H07	(GAA)6	220	CGATAAAGTCTCAGGTGAGCAA	59.517	AGAGGGAGTGGTGAATGAATGT	59.862
AS2RM6P1F03	(CT)7	135	TCCTTCAAGCTCGTCTTCTACC	60.019	GAAATCTGACGCAATGTTCAAG	59.752
AS1RN26P1G09	(TTA)20	234	GTTATTGCTGTGGTGAAGAACG	59.677	TTGACTTTCATATGCACCTCTCA	59.761
AS2RM9P1B06	(AT)8	195	TAAAGTTGATCTTGAATCTCC	52.967	CGCGATCATTTATTTAACTT	52.482
AS2RM1P1A05	(CTT)8	209	TCTTCCAATCTCTCTGCTTC	59.962	ATGGAGTGACCAACATAAACC	59.983

Raw output from the program is a tab-spaced file with, in addition to the information above, primer names and base qualities of the contig regions that bind to the primers.

ESTs

Of the 8800 ESTs, 282 were above the specified threshold of similarity to known *Arachis* SSR markers and were eliminated by the BLAST module. The TROLL module detected 213 other chromatograms with microsatellites. Using the Gap4 interface these were assembled into 142 contigs. Although the original dataset is large, the resultant database is small, and is operated efficiently on a standard personal computer. The primer design program produced primer pairs for 75 new microsatellite markers. These primer pairs have not been fully tested yet, but in our previous work (12) only 6 of 81 EST primer pairs (7.4%) failed to amplify an identifiable PCR product.

Text data from GenBank

A total of 563 of the 3890 sequences from GenBank were above the specified threshold of similarity to known *Arachis* markers and were eliminated by the BLAST module, TROLL detected 482 other sequences with microsatellites and these were assembled into 463 contigs. Although counter-intuitive, it is possible to use quality control even with text data. Staden assigns a nominal quality value of two for bases of text data, but when text data overlaps within a contig, a higher value of confidence is given. This can be used, for instance, to design primers only using consensus sequences supported by at least two reads. However, since the GenBank data we processed has low redundancy, primers were designed without quality control and primer pairs for 352 markers were designed. It is to be expected that data from GenBank may give a lower success rate of primers compared to chromatogram data. Some sequences may have been rejected for marker development by the original authors because the repeat type has relatively low polymorphism, or because of low sequence quality. Nevertheless, in our previous work, data-mining from GenBank was a productive and low cost option for marker development. Of the 69 markers data-mined, only 9 (13%) did not amplify any scorable fragments (12).

Example databases and the new primers designed are in Supplementary Data. In total, 427 pairs of primers for new candidate markers for *Arachis* are presented.

In summary, advances in genetic mapping are reliant upon the development of sufficient numbers of molecular markers, and large scale marker development depends upon efficient processing of the data. The Staden software with the TROLL module and primer design program provides a unified and efficient interface for this.

Availability

TROLL, the module and primer design program, and instructions for installation and use are available free and can be downloaded from the Resources section of <http://finder.sourceforge.net/>. A Web version of the primer design program is also available from the same link.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank the Conselho Nacional de Pesquisa e Desenvolvimento (Reference: 400617/2003-0), the Pró-Reitoria de Pós-Graduação e Pesquisa—Universidade Católica de Goiás, the Universidade Católica de Brasília, the European Union INCO Programme (ARAMAP reference: ICA4-2001-10072), and the Generation Challenge Program (Project number 31) for financial support. The authors also thank anonymous referees for their comments. Funding to pay the Open Access publication charges for this article was provided by Generation Challenge Programme.

Conflict of interest statement. None declared.

REFERENCES

- Benson,G. (1999) Tandem repeats finder: a program to analyse DNA sequences. *Nucleic Acids Res.*, **27**, 573–580.
- Kolpakov,R., Bana,G. and Kucherov,G. (2003) mreps: efficient and flexible detection of tandem repeats in DNA. *Nucleic Acids Res.*, **31**, 3672–3678.
- Kurtz,S., Choudhuri,J., Ohlebusch,E., Schleiermacher,C., Stoye,J. and Giegerich,R. (2001) REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.*, **29**, 4633–4642.
- Castelo,A.T., Martins,W.S. and Gao,G.R. (2002) Tandem repeat occurrence locator. *Bioinformatics*, **8**, 634–636.
- Staden,R., Beal,K.F. and Bonfield,J.K. (1998) The Staden Package. In Misener,S. and Krawetz,S.A. (eds), *Computer Methods in Molecular Biology, Bioinformatics Methods and Protocols*. The Humana Press Inc., Totowa, NJ, Vol.132, pp.115–130.
- Rozen,S. and Skaletsky,H. (2000) Primer3 on the WWW for general users and for biologist programmers. In Krawetz,S. and Misener,S. (eds), *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. The Humana Press Inc., Totowa, NJ, pp. 365–386.
- Rafalski,J.A., Vogel,J.M., Morgante,M., Powel,W., Andre,C. and Tingey,S.V. (1996) Generating and using DNA markers in plants. In Birren,B. and Lai,E. (eds), *Analysis of non-mammalian genomes: a practical guide*. Academic, NY, pp. 75–134.
- Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
- Ferguson,M.E., Burow,M.D., Schulze,S.R., Bramel,P.J., Paterson,A.H., Kresovich,S. and Mitchell,S. (2004) Microsatellite identification and characterization in peanut (*A.hypogaea* L.). *Theor. Appl. Genet.*, **108**, 1064–1070.
- He,G., Meng,R., Newman,M., Gao,G., Pittman,R.N. and Prakash,C.S. (2003) Microsatellites as DNA markers in cultivated peanut (*A.hypogaea* L.). *BMC Plant Biol.*, **3**, 3.
- Moretzsohn,M.C., Hopkins,M.S., Mitchell,S.E., Kresovich,S., Valls,J.F.M. and Ferreira,M.E. (2004) Genetic diversity of peanut (*Arachis hypogaea* L.) and its wild relatives based on the analysis of hypervariable regions of the genome. *BMC Plant Biol.*, **4**, 11.
- Moretzsohn,M.C., Leoi,L., Proite,K., Guimaraes,P.M., Leal-Bertioli,S.C.M., Gimenes,M.A., Martins,W.S., Valls,J.F.M., Grattapaglia,D. and Bertioli,D.J. (2005) A microsatellite based, gene-rich linkage map for the AA genome of *Arachis* (Fabaceae). *Theor. Appl. Genet.*, **111**, 1060–1071.