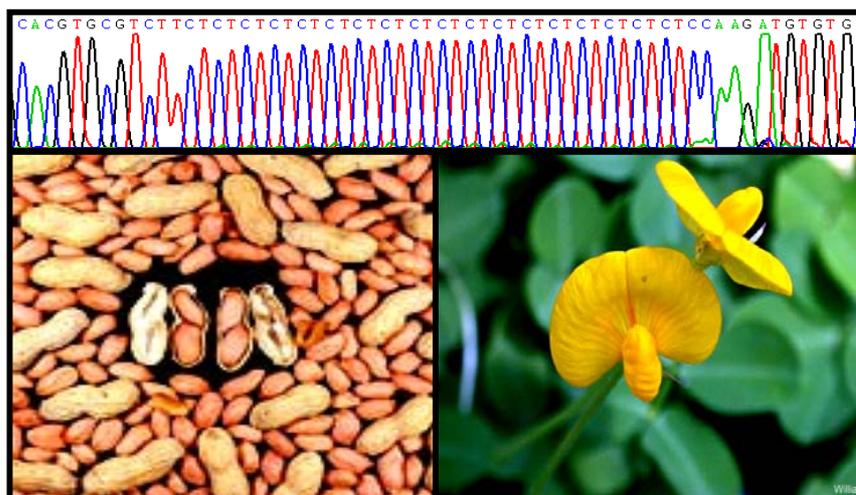


Universidade Católica de Brasília
Pós-Graduação em Ciências Genômicas e Biotecnologia

**Desenvolvimento de Marcadores Microssatélites (SSR)
em
*Arachis hypogaea***



Dissertação de Mestrado

Lélia Cristina Tenório Leoi

Brasília
2003



Universidade Católica de Brasília
Pós-Graduação em Ciências Genômicas e Biotecnologia



Recursos Genéticos e Biotecnologia

Desenvolvimento de Marcadores Microssatélites (SSR) em *Arachis hypogaea*

Lélia Cristina Tenório Leoi

“Dissertação apresentada no Programa de Pós-Graduação em Ciências Genômicas e Biotecnologia, da Universidade Católica de Brasília como requisito parcial para obtenção do grau de Mestre”.

Orientador: David John Bertioli

Brasília
Julho de 2003

BANCA EXAMINADORA

Dr. Cláudio Brondani
(examinador externo)

Dr. Dário Grattapaglia
(examinador interno)

Dra. Rosane Garcia Collevatti
(examinador interno)

Dr. David John Bertioli
(Orientador)

DEDICATÓRIA

AGRADECIMENTOS

Índice

Lista de Figuras e Tabelas.....	VII
Abreviaturas.....	XII
Resumo.....	XIII
Abstract.....	XIV
1 - Introdução	1
1.1 - Origem e Diversidade Genética em Plantas.....	2
1.2 - Caracterização e Histórico do Gênero <i>Arachis</i>	3
1.3 - Citogenética e Genoma do Gênero <i>Arachis</i>	5
1.3.1 - Citogenética e Genoma da Secção <i>Arachis</i>	6
1.4 - <i>Arachis hypogaea</i>	8
1.5 - Pragas e Doenças.....	10
1.6 - Marcadores Moleculares para Análise Genética.....	13
1.6.1 - Marcadores Microssatélites.....	18
1.7 - Aplicações de Marcadores no Melhoramento de Plantas.....	23
1.8 - Marcadores Moleculares em <i>Arachis hypogaea</i>	25
1.9 - A Bioinformática como Ferramenta.....	27
1.9.1 - Processamento de Cromatogramas de Sequências Definitivas.....	28
1.9.2 - Software Staden Package.....	29
2 - Justificativa.....	35
3 - Objetivos.....	36
4 - Materiais.....	37
4.1 - Produtos Químicos, Enzimas e Kits.....	38
4.2 - Plasmídeo, Oligonucleotídeos e Bactérias.....	40
4.2.1 - Vetor de Clonagem	40
4.2.2 - Oligonucleotídeos.....	40
4.2.3 - Isolado Bacteriano	41
4.2.3.1 - Meios de Cultura	41
4.2.3.2- Antibiótico- Meio LB e LB-agar	42
4.3 - Condições de Cultivo para Bactéria	43
4.3.1 - <i>E. coli</i> XL1-Blue	43
4.4 - Soluções	43
5 - Metodologia	46
5.1 - Extração e Quantificação do DNA Genômico Total de <i>Arachis hypogaea</i>	47
5.2 - Seleção da Enzima para Digestão Total do DNA Genômico	48
5.3 - Gel de Agarose para Eletroforese Horizontal	48
5.3.1 - Eluição dos Fragmentos de DNA em Gel de Agarose	49
5.4 - Construção de Biblioteca Genômica Enriquecida/ Ligação aos Adaptadores ..	50
5.4.1 - Controle de Enriquecimento da Biblioteca Genômica por PCR	52
5.4.2 - Análise da hibridização do DNA/Sonda pela Técnica <i>Southern Blot</i>	53
5.5 - Seleção de Clones Positivos para SSR	54
5.5.1 - Ligação do Inseto ao Plasmídeo	54

Índice

5.5.2 - Preparação de Bactérias Eletrocompetentes	54
5.5.3 - Transformação por Eletroporação	55
5.5.4 - Seleção das Colônias	55
5.5.5 - Clonagem e Isolamento do DNA Plasmidial por Lise Alcalina	56
5.6 - Seqüenciamento dos Clones Positivos para SSR	58
5.7 - Construção dos Primers-SSR	58
6 - Resultados e Discussão.....	60
6.1 - Obtenção e Isolamento de SSRs	61
6.2 - Obtenção das Seqüências	64
6.3 - Teste Comparativo com Hibridização e PCR Ancorado	65
6.4 - Análise e Agrupamento das Seqüências para Montagem dos Contigs	70
6.5 - Desenhos dos Primers de SSR	71
6.6 - Análise de Similaridade	75
7 - Conclusões	78
8 - Perspectivas futuras	80
9 - Referências Bibliográficas	82
Anexos	92
Anexo 1	91
Anexo 2 - <i>Primer Notes</i>	93
Anexo 3 - Primers SSR.....	101
Anexo 4 - Blastx- Seqüências com Microsatélites.....	119

Listas de Figuras e Tabelas

	Página
Figura 1 Ilustração didática demonstrando o gargalo genético gerado pela seleção de características, favorecendo a diminuição da variedade alélica e conseqüente perda da heteroziguidade.....	2
Figura 2 Aumento da produção de soja (planta x acre) nos Estados Unidos nos últimos 30 anos através do melhoramento genético.....	3
Figura 3 Ilustração da anatomia da planta do amendoim.....	4
Figura 4 Morfologia do cromossomo SAT (Fernández & Krapovickas, 1994).....	5
Figura 5 Ilustrações da espécie <i>Arachis hypogaea</i> : (A) Fotografia da base da planta. (B) Detalhe da florescência. (C) Desenho feito por <i>Linnaeus</i> em 1753 em uma de suas viagens aos países tropicais.....	8
Figura 6 Forma de utilização das plantas pertencentes à secção <i>Arachis</i> . A e C) A espécie <i>Arachis pintoi</i> sendo usada como cobertura vegetal de solo contra a erosão; B) Planta forrageira utilizada como fonte protéica na alimentação animal e D) Amendoim <i>in natura</i> (vagem c/ sementes), destinado à alimentação humana.	9
Figura 7 Mapa mundial indicando as regiões de origem e cultivo da espécie <i>A. hypogaea</i> na atualidade.....	7
Figura 8 Tipos de pragas que comumente atacam o amendoim. A) Folha de <i>A. hypogaea</i> contaminada por uma doença foliar causada pelo fungo <i>Cercospora arachidicola</i> . B) Mosca-branca do prateamento da folha (ordem Homóptera; <i>Bemisia argentifolii</i>) atacando a folha do amendoim. C) Vagens do amendoim atacadas pelo nematóide da espécie <i>Meloidogyne arenaria</i> e D) Semente de amendoim infectada por <i>Aspergillus flavus</i>	11
Figura 9 Microsatélite com repetições em tandem de dinucleotídeos CT/GA.....	18
Figura 10 Modelo do processo de mutação do loco microsatélite. Na figura, o DNA é representado pela linha azul, repetições microsatélites são os blocos azuis claros e as setas indicam a replicação do DNA molde.....	20
Figura 11 Crescimento exponencial do número de seqüências contidas no <i>GenBank</i> ao longo das duas últimas décadas.....	26
Figura 12 Pregap 4, interativo e configurável. Programas como Phred (análise de qualidade) e Cross-Match (identifica regiões contaminantes de vetor) trabalham na mesma interface..	28
Figura 13 (A) Janela principal do gap4, contem informações sobre as seqüências depositadas, por exemplo, os números de <i>contigs</i> , <i>templates</i> e vetores. No topo estão organizados vários ícones que quando abertos mostram os vários módulos de trabalho. (B) – <i>Contig selector</i> é um dos vários módulos interativos existentes dentro do gap4, permitindo uma rápida visualização de todos os <i>contigs</i> existentes no banco de dados, além de mostrar as posições <i>tags</i> com cores específicas. (C) – O <i>contig editor</i> , para edição das seqüências e (D) Eletroferograma das seqüências apresentadas no <i>contig editor</i>	29
Figura 14 (A) Esta interface permite verificar a redundância das seqüências depositadas através do cruzamento dos dados. O resultado é exposto em forma de gráfico (X;Y), sendo as seqüências consideradas homólogas (total ou parcial) representadas por um traço diagonal. (B) Emite informações sobre a provável homologia selecionada em (A). (C) Sobreposição das seqüências consideradas homólogas. (D) Verificação dos cromatogramas das seqüências.....	31
Figura 15 Representação do vetor de clonagem. pGEM-T. Ori = origem de replicação; Lac Z = α -peptídeo - região codificadora da enzima β -galactosidase; T7 e SP6 = promotores da RNA polymerase que flanqueiam uma região múltipla de clonagem; T = local de inserção da molécula de DNA exógena; fl ori = região fl fago; Amp ^r = gene que confere resistência a ampicilina.....	38

Figura 16	Captura e eluição do DNA, clivado por enzimas de restrição, em membrana de nitrocelulose.....	48
Figura 17	Protocolo de enriquecimento da biblioteca genômica. A primeira linha representa o DNA genômico contendo muitos sítios de clivagem da enzima <i>Sau</i> 3AI. GATC e CTAG representam as extremidades coesivas geradas pela enzima <i>Sau</i> AI no DNA genômico. As repetições (TC/AG) _n presentes nas seqüências são representações de microssatélites.....	50
Figura 18	Etapa 1 - Seleção de colônias transformadas e o replique em placa de petri mapeado com o formato de placa de 96 poços. Etapa 2- Inóculo das 96 colônias transformadas com o auxílio de um replicador em uma placa tipo <i>Deep Well</i> contendo meio líquido e ampicilina.....	54
Figura 19	Eletroforeses em gel de agarose 1% corados com brometo de etídeo. A) Teste de qualidade de DNA em <i>Arachis hypogaea</i> para 4 tecidos diferentes: poços: 1 e 2 - fago λ 100 ng e 200 ng; 3-caule; 4-raiz, 5-flor e 6-folha. B) Quantificação do DNA genômico total extraído do tecido foliar de indivíduos da espécie <i>A. hypogaea</i> ; Fago λ 100 ng e 200 ng.....	61
Figura 20	Eluição do DNA no gel de agarose 1% corado com brometo de etídeo.....	62
Figura 21	(A) Resultado, via eletroforese, da amplificação dos fragmentos da PCR controle de enriquecimento (TC/AG). B) Análise por Southern blotting do filme de raios-X do gel representado em "A". A sonda marcada utilizada foi a (TC) ₁₃ . Poços: 1- Controle positivo DNA (1:1000); 2- Solução de hibridização; 3- Solução 1ª lavagem; 4- Sol. 2ª lav. ; 5-Sol. 3ª lav. ; 6- Sol lav. Final e 7- Solução final (DNA hibridizado).....	62
Figura 22	Ilustração representando o programa <i>Sort Chromats for Microsats</i> para seleção de SSR desenvolvido pelo Dr. David Bertoli.....	64
Figura 23	(A) Auto-radiografia dos clones da placa Ah1TC4 que hibridizaram com a sonda (AG) _n . (B) Ilustração comparativa dos dados obtidos, confrontando o resultado do seqüenciamento das amostras com a hibridização da sonda (mostrada em "A").....	65
Figura 24	(A)- Perfil dos produtos de PCR oriundos do DNA genômico da placa Ah1TC4 entre os poços A01 e H06. O marcador de peso molecular utilizado foi o 1Kb plus. Nesta eletroforese podem-se observar falhas de amplificações em muitas amostras, como por exemplo, os poços referentes às amostras A05 e B05 (círculo azul). Em outros casos, na sua maioria, ocorreram amplificações inespecíficas, comprometendo a análise do comprimento do fragmento amplificado, amostra A02 (círculo vermelho).	67
Figura 24	(B). Perfil dos produtos de PCR oriundos do DNA genômico da placa Ah1TC4 entre os poços A07 e H12. O marcador de peso molecular utilizado foi o 1Kb plus. Amostras que não apresentaram SSR quando seqüenciadas, também foram amplificadas erroneamente, por exemplo, amostra A08 (círculo verde).	
Figura 25	Ilustração mostrando a seleção de clones que contém SSR por PCR-ancorado: Para cada clone são feitas 5 reações de PCR: (1) – par de primers que amplifica o vetor inteiro; (2 a 5) – um par que amplifica do início do vetor até o SSR. Os clones 1 e 2 possuem SSR devido às amplificações dos poços 2 e 3 ou 4 e 5; o clone 3 não tem SSR (Buso <i>et al.</i> , 2000).....	68
Figura 26	Desenhos de primers. (1) Amostra de DNA seqüenciado com região flanqueadora do SSR suficiente para desenhar primers <i>forward</i> e <i>reverse</i> . (2 e 3) Amostra de DNA seqüenciado com região flanqueadora do tamanho insuficiente para desenhar o primer <i>forward</i> ou <i>reverse</i> . Linha pontilhada azul = fim vetor/adaptador e início da seqüência; linha pontilhada vermelha = fim da seqüência e início do adaptador/vetor; linha pontilhada preta = início ou fim da região flanqueadora do SSR.....	71
Figura 27	Fluxograma das etapas de análise computacional das seqüências de <i>A. hypogaea</i> no projeto.....	75
Figura 28	Formato da aplicação dos clones e das respectivas combinações dos primer na placa de PCR	91
Tabela 1	Classificação Biológica (Fernández & Krapovickas, 1994).....	4

Listas de Figuras e Tabelas

Tabela 2	Diferenciações do cromossomo SAT e a variação do tipo em cada secção do gênero <i>Arachis</i> (Fernández & Krapovickas, 1994).....	6
Tabela 3	Características agronômicas, industriais e fitossanitárias encontradas em algumas espécies selvagens de <i>Arachis</i>	12
Tabela 4	Fatos importantes que marcaram as décadas de 1910 a 1980 (Borém, 1998).....	13
Tabela 5	Análise comparativa entre os marcadores moleculares (Gepts, 1993; Ferreira & Grattapaglia, 1995).....	22
Tabela 6	Descrição sucinta dos demais marcadores existentes na literatura (Souza, 2001).....	23
Tabela 7	Principais aplicações de metodologias baseadas em marcadores microssatélites no melhoramento de plantas (Beckmann, 1991; Ferreira & Grattapaglia, 1998).....	24
Tabela 8	Seqüências de oligonucleotídeos usados nos protocolos de construção e seqüenciamento de bibliotecas genômica.....	39
Tabela 9	Composição do meio LB para crescimento de bactéria (Miller, 1972).....	39
Tabela 10	Composição do meio SOC para crescimento de bactéria (Hanahan, 1983).....	40
Tabela 11	Composição do meio TB para crescimento de bactéria (Tartof & Hobbs, 1987).....	40
Tabela 12	Composição do gel de agarose para eletroforese horizontal.....	46
Tabela 13	Parâmetros específicos selecionados para os desenhos dos primers de SSR no programa Primer 3 (McCouch <i>et al.</i> , 1997).....	57
Tabela 14	Comparação dos custos gastos pelas 3 principais técnicas (Hibridização; PCR ancorado e Seqüenciamento Direto) na detecção de SSR em uma placa contendo 96 amostras. Ao total foram desenhado 25 pares de primers para esta placa.....	69
Tabela 15	Caracterização dos novos marcadores SSR desenvolvidos para análise genômica de amendoim.....	72
Tabela 16	Resultado do Blast-X das seqüências presentes na Pasta NO-HITS para busca de similaridade contra um banco de dados local de gene de <i>Arabidopsis thaliana</i>	75
Tabela 17	Resultado do Blast-X das seqüências presentes na Pasta HITS para busca de similaridade contra um banco de dados local de gene de <i>Arabidopsis thaliana</i>	76
Tabela 18	<i>Master Mix</i>	91

Resumo

Abreviaturas

%	Porcentagem	mM	MiliMolar
°C	Grau Celsius	<i>Mse I</i>	<i>Micrococcus species</i>
μF	Micro-faraday	N	Normal
μg	Micrograma	Nm	nanômetro
μL	MicroLitro	OD	Densidade Óptica
μM	MicroMolar	p/v	peso/volume
λ	<i>Lambda</i>	pB	Pares de Base
Ω	Ohms	PBS	<i>Phosphate-Buffered Saline</i>
AFLP	<i>Amplified Fragment Length Polymorphism</i>	PCR	<i>Polymerase Chain Reaction</i>
B&W	<i>Binding & Washing Buffer</i>	pH	Potencial Hidrogeniônico
BSA	Albumina de Soro Bovino	QTLs	<i>Quantitative Trait Loci</i>
CAP3	<i>Sequence Assembly Program</i>	RAPD	<i>Randomly Amplified Polymorphic</i>
CIA	Clorofórmio- Álcool Isoamílico	RFLP	<i>Restriction Fragment Length Polimorphism</i>
Cm	Centímetros	RNAse	Ribonuclease
CTAB	<i>Cetyltrimethylammonium Bromide</i>	rpm	Rotações Por Minuto
DDJB	<i>DNA Data Bank of Japan</i>	SA	<i>Short Adaptor</i>
DMSO	<i>Dimethylsulfoxide</i>	SAT	Satélite
DNA	Ácido Desoxirribonucléico	<i>Sau 3A</i>	<i>Staphylococcus aureus 3A</i>
dNTPs	<i>Deoxyribonucleoside Triphosphates</i>	SDS	<i>Sodium Dodecyl Sulfate</i>
EDTA	Ácido Etileno Diamono Tetrecético	SSC	<i>Sodium, Salt Citrate</i>
EMBL	<i>European Molecular Biology Laboratory</i>	SSPE	<i>Sodium, Salt, Phosphate e EDTA</i>
g	Grama	SSR	<i>Simple Senquence Repeats</i>
GAP	<i>Genome Assembly Program</i>	<i>Taq</i>	<i>Thermus aquaticus</i>
GET	Glicose, EDTA, Tris-Cl	TB	<i>Terrific Broth</i>
H ₂ O _{bidest}	Água Bidestilada	TBE	Tris-Base
HS-NET	<i>High Salt-NET</i>	TdT	Terminal Transferase
IPTG	<i>Isopropyl-beta-D-thiogalactopyranoside</i>	TE	Tris-EDTA
Kb	Kilobase	Tris	Tris-(hidroximetil)-aminometano
KV	KiloVolts	Tris-Cl	Tris HCl
LA	<i>Long adaptor</i>	<i>Tsp</i>	<i>Thermus species</i>
M	Molar	U	Unidade
Meio LB	Meio Luria-Bertani	UV	Radiação Ultravioleta
mg	Miligrama	VNTR	<i>Variable Number of Tandem Repeats</i>
		X-Gal	<i>5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside</i>

Resumo

Entre as plantas cultivadas com base genética estreita, encontra-se o amendoim da espécie *Arachis hypogaea* L. (Leguminosae). Esta espécie é um alotetraplóide, provavelmente oriunda por domesticação da espécie também tetraplóide *A. monticola* e esta por sua vez do cruzamento entre duas espécies silvestres distintas do gênero *Arachis*.

O maior interesse pela prospecção, resgate e caracterização de germoplasma das espécies do gênero *Arachis* reside em seu potencial de fornecimento de genes úteis para o melhoramento do *A. hypogaea*, bem como na preservação da variabilidade genética populacional localizada em áreas de risco de devastação. Marcadores microssatélites são ideais para aplicações no melhoramento de plantas por sua natureza co-dominante, alto polimorfismo e conteúdo informativo.

Neste trabalho foram isolados microssatélites através de metodologias de enriquecimento de bibliotecas genômicas para seqüências repetitivas contendo dinucleotídeos TC. As colônias da biblioteca foram seqüenciadas aleatoriamente não ocorrendo seleção por hibridização ou PCR-ancorado. A partir de uma biblioteca genômica enriquecida foram seqüenciados 598 clones. Essas seqüências foram processadas com o auxílio de um *script Perl* identificando ao total 144 seqüências com microssatélites, após análise de redundância o número de *contigs* abaixou para 138 e destes, apenas 77 (55%) apresentaram condições de desenhar marcadores por apresentarem tamanho suficiente de repetição e suas posições na seqüência serem adequadas para os desenhos dos pares de primers flanqueadores. Todas os cromatogramas que continham as seqüências repetidas foram previamente analisados e visualizadas com o auxílio do programa interativo Staden, cuja função foi a de organizar e verificar as sobreposições de seqüências, além de realizar a inspeção e edição manual das seqüências. Os pares de primers flanqueadores das regiões repetidas foram desenhados com o auxílio do programa Primer3. Todas as seqüências obtidas através do seqüenciamento foram analisadas posteriormente através do programa Blast-X para se verificar homologia com seqüências depositadas no banco de proteínas de *Arabidopsis thaliana*. Entre as 576 seqüências analisadas, 42 apresentaram homologia com algum gene, sendo destas, 13 oriundas de seqüências com regiões ricas em microssatélites. Entre as seqüências sem SSR foram encontradas homologias freqüentes com retrotransposons, sugerindo abundância destes elementos no genoma de *A. hypogaea*.

Abstract

Among the cultivated plants which have suffered a genetic "bottle-neck" is included the peanut, or groundnut *Arachis hypogaea* L. (Leguminosae). This species is an allotetraploid possibly derived from the wild species *A. monticola* which, in turn, is derived from a cross between two distinct wild species of the genus *Arachis*.

The greatest interest in the search, conservation and characterisation of wild species of *Arachis* resides in their potential to provide genes for the improvement of *A. hypogaea* through breeding. Microsatellite markers are ideal markers to be used to aid in this breeding process, because of their co-dominance, high polymorphism and high number of alleles.

In this work, a genomic library enriched for TC repetitive sequences was constructed and used to obtain microsatellite containing sequences. From this library were sequenced 598 clones. The sequences were processed by the use of a Perl script to select 144 sequences with microsatellites. These 144 sequences represented 138 non-redundant sequences. Of these, only 77 (55%) were suitable for primer design and these were completed by further sequencing of the selected clones. Contigs were processed and edited using the software "Staden". Primers were designed using the program "Primer 3". Finally all 576 sequences obtained in the study (including sequences which did not contain microsatellites) were analysed by Blast-X against *Arabidopsis thaliana* proteins. In total 42 significant homologies were found, 13 of these being from microsatellite containing regions. Among the sequences without microsatellites significant homologies with retrotransposons were frequent, suggesting a high frequency of these elements in the *A. hypogaea* genome.

Introdução

1 - INTRODUÇÃO

1.1 - Origem e Diversidade Genética em Plantas.

Embora as plantas com flores (Angiospermae) tenham sido originadas provavelmente no período Jurássico (145 a 206 milhões de anos atrás), a maioria da diversificação e dominância destas na Terra tem sido atribuída ao período Cretáceo (90 a 140 milhões de anos atrás) (Crane *et al.*, 1995).

Por volta de 10 mil anos atrás, com a mudança do homem nômade-caçador para a sociedade agrária, surgiram as primeiras atividades agrícolas, dando início às plantas cultivadas, transição hoje conhecida como Revolução Agrícola (Flannery, 1973). Embora sejam desconhecidas as etapas exatas de como as plantas foram domesticadas pelos humanos, certamente as características presentes nas sementes, como a germinação, e na planta, como a altura e produtividade, teriam sido selecionadas pelos primeiros agricultores (Harlan, 1987). A propagação seletiva de linhagens que continham tais características selecionadas resultou em uma progressiva restrição da base genética na população subsequente, conhecida como gargalo genético (*genetic bottleneck*) (Figura 1).

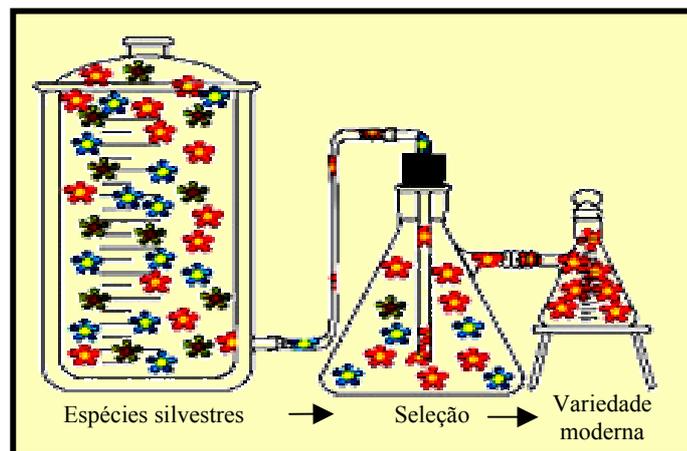


Figura 1: Ilustração didática demonstrando o gargalo genético gerado pela seleção de características, favorecendo a diminuição da variedade alélica e conseqüente perda da heterozigidade.

Na atualidade, a variação genética em plantas cultivadas tem sido continuamente reduzida através da seleção gênica de indivíduos com altos rendimentos na colheita. A reprodução contínua de tais linhagens seja por autofecundação, cruzamento entre variedades modernas geneticamente relacionadas ou por apomixia geram conforme o ilustrado na Figura 2, plantas mais produtivas, porém com uma menor variabilidade alélica. Este procedimento pode levar a exclusão de genes importantes existentes nos ancestrais primitivos e em plantas silvestres, ameaçando e diminuindo a base genética das quais dependeriam as novas variantes.

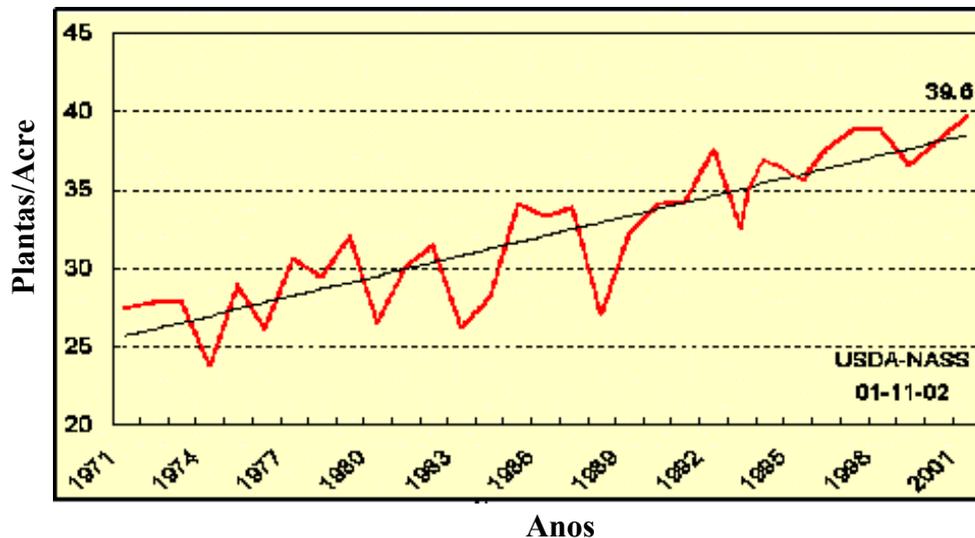


Figura 2. Aumento da produção (Plantas/Acre X Anos) de soja nos Estados Unidos nos últimos 30 anos através do melhoramento genético.

Adaptado da fonte eletrônica: *National Agricultural Statistics Service (USDA)*.

1.2 - Caracterização e Histórico do Gênero *Arachis*

Entre as plantas cultivadas base genética restrita, encontra-se o amendoim da espécie *Arachis hypogaea* L. (Leguminosae). O gênero *Arachis* se distingue da maioria das outras plantas devido ao florescimento sobre o terreno, mas com produção de frutos abaixo da superfície do solo. Esta característica particular de frutificação é devido à presença do ginóforo (mais comumente conhecido como “peg”) que após a fertilização da flor, inicia um alongamento em direção ao solo (geotropismo positivo), transpondo-o (Zamski & Ziv, 1976). O peg então cessa o alongamento iniciando o desenvolvimento geocárpico (vagem)

Introdução

na sua extremidade, resultando na formação da semente popularmente conhecida por amendoim. Suas flores são hermafroditas e geralmente apresentam autopolinização (Godoy *et al.*, 1989) (Figura 3).

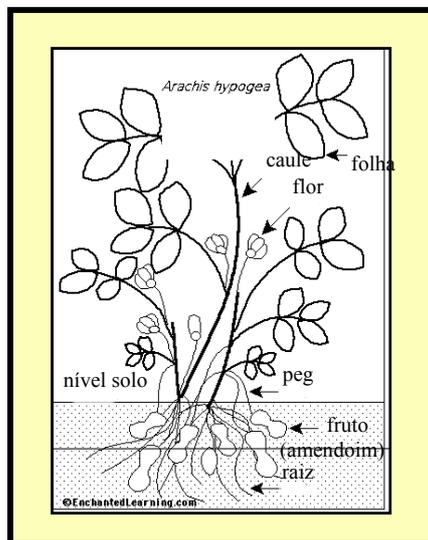


Figura 3. Ilustração da anatomia da planta do amendoim.

Adaptado da fonte eletrônica *Enchanted Learning*.

O gênero *Arachis* é composto por cerca de 80 espécies, todas nativas da América do Sul (Valls & Simpson, 1994). Krapovickas & Gregory (1994) dividiram o gênero em nove seções, baseando-se em suas características morfológicas, capacidade de cruzamento e fertilidade dos híbridos (Tabela 1).

Tabela 1. Classificação Biológica.

Reino			Ordem	Família	Gênero	Secções
						<i>Arachis</i>
						<i>Caulorrhizae</i>
						<i>Erectoides</i>
Plantae	Tracheophyta	Angiospermae	Fabaceae	Leguminosae	<i>Arachis</i>	<i>Extranervosae</i>
	(Spermatophytina)	(Dicotyledoneae)		(Papilionidae)		<i>Heteranthae</i>
						<i>Procumbentes</i>
						<i>Rhizomatosae</i>
						<i>Trirectoides</i>
						<i>Triseminatae</i>

Adaptado de Krapovickas & Gregory, 1994; Fernández & Gregory, 1994.

Dentre os países do Cone Sul e Andinos, há um relevante destaque para o Brasil, país que reúne todas as secções e a maior parte das espécies. Adicionalmente, cinco das secções avançam sobre o Paraguai e apenas duas alcançam a Bolívia, a Argentina e o Uruguai. Do conjunto de secções apresentadas, *Arachis* é a que apresenta as características mais recentes ou derivadas, sendo composta de 27 espécies (Valls & Simpson, 1997).

1.3 - Citogenética e Genoma do Gênero *Arachis*.

De acordo com os aspectos citológicos e genéticos, grupos genômicos têm sido propostos para o gênero *Arachis*. Em geral os cromossomos apresentam-se bastante semelhantes entre si, com exceção de dois pares que podem ser perfeitamente diferenciados: (i) um par pequeno com denominação “A” e (ii) um par com constrição secundária com denominação “SAT” - (satélite) (Husted, 1936). Nos cromossomos “SAT”, um dos braços é dividido formando um segmento proximal e outro distal (satélite) em relação ao centrômero (Figura 4).

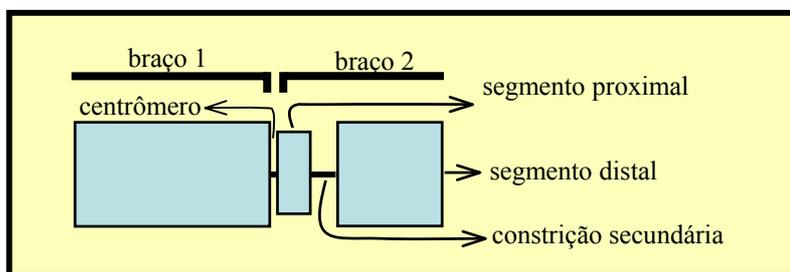


Figura 4. Morfologia do cromossomo SAT.

Adaptado de Fernández & Krapovickas, 1994.

A presença ou ausência do cromossomo “A” e a presença de diferentes tipos de cromossomo SAT caracterizam os diferentes genomas (Fernández & Krapovickas, 1994).

Com base no tamanho relativo do satélite e a posição do centrômero, os pares SAT podem ser classificados em dez tipos (Tabela 2). Todas as espécies estudadas possuem um par de cromossomos SAT, exceto um acesso de *A. valida* onde se encontram dois pares de cromossomos SAT.

Tabela 2. Diferenciações do cromossomo SAT e a variação do tipo em cada secção do gênero *Arachis*.

Tipo SAT	Descrição	Secções
1 	O satélite é menor que o braço 1, mas é maior que o segmento proximal.	<i>Extranervosae</i> .
2 	O satélite é menor que o braço 1 e muito maior que o segmento proximal (0,2µm).	<i>Caulorrhizae; Heteranthae; Erectoides; Extranervosae; Trierectoides</i>
3 	O satélite é mais ou menos igual ao tamanho do braço 1 mais o segmento proximal (0,2µm).	<i>Erectoide; Extranervosae Arachis; Caulorrhizae; Rhizomatosae</i> .
4 	Semelhante ao tipo 3, tanto na forma como no tamanho, porém o braço 1 apresenta em prófase e em prometáfase uma coloração menos intensa que o satélite.	<i>Erectoide</i>
5 	O satélite tem mais ou menos o mesmo tamanho do braço 1 mais o segmento proximal e este é menor que o braço 1.	<i>Arachis</i>
6 	O satélite é maior que o braço 1 e este é maior que o segmento proximal.	<i>Arachis</i>
7 	O satélite terminal mais o intercalar possuem quase o mesmo tamanho do braço 1 com o segmento proximal. Este cromossomo apresenta constrições secundárias, uma entre o segmento proximal e o satélite intercalar e outra entre o intercalar e o satélite terminal.	<i>Arachis</i>
8 	O satélite é a metade em tamanho do braço 1. O braço 1 é maior que o segmento proximal.	<i>Arachis</i>
9 	O satélite é puntiforme e seu tamanho é menor que o segmento proximal. O braço 1 é maior que o segmento proximal.	<i>Arachis, Procumbentes</i>
10 	O satélite é puntiforme, porém menor que o tipo anterior. O braço 1 é mais ou menos igual em tamanho ao segmento proximal.	<i>Heteranthae</i>

Adaptado de Fernández & Krapovickas, 1994.

A maioria das espécies pertencentes ao gênero *Arachis* é diplóide ($2n=2x=20$), e somente quatro são tetraplóides ($2n=4x=40$), duas pertencentes à secção *Arachis*, entre elas o amendoim cultivado (*Arachis hypogaea*), e duas à secção *Rhizomatosae*.

1.3.1 - Citogenética e Genoma da Secção *Arachis*.

Dados oriundos dos estudos citogenéticos e reprodutivos da secção *Arachis* revelaram a presença de três genomas distintos denominados “A”, “B” e “D” (Stalker, 1991).

O genoma “A” está relacionado à presença de um par de cromossomos “A” (Husted, 1933; Husted, 1936). Este par diferencia-se dos demais cromossomos devido à sua condensação na pro-metáfase, bem como por possuir a metade, em comprimento, do maior cromossomo presente no interior do mesmo núcleo. O genoma “B” é caracterizado pela ausência do par de cromossomos “A”, enquanto o genoma “D” está relacionado à presença de seis pares de cromossomos subteloentrícos (Stalker, 1991).

A maioria das espécies da secção *Arachis* apresenta o genoma “A”, enquanto que *A. batizocoi*, *A. ipaënsis*, *A. magna*, *A. valida*, *A. benensis* e *A. hoehnei* possuem genoma “B” e somente *A. glandulifera* o genoma “D”.

As espécies tetraplóides apresentam simultaneamente dois genomas. Uma dessas espécies, como já citada, é a *A. hypogaea*, que reúne dois genomas (A e B) ocorrentes em espécies diplóides distintas da secção *Arachis* (Fernández & Krapovickas, 1994). Os resultados apresentados por Singh (1994) sugerem que por domesticação, *A. hypogaea* (AABB) teria sido originada da espécie *A. monticola* (AABB) e esta pelo cruzamento entre duas espécies silvestre da secção *Arachis*, uma com o genoma “A”, e a outra de genoma “B” (Gregory & Gregory, 1976; Gregory & Gregory, 1979).

Por sua vez, Kirti e colaboradores (1982) estudaram o pareamento cromossômico de híbridos F1 ($P = A. monticola \times A. hypogaea$) e relataram, pelo alto potencial de recombinação e fertilidade dos polens, que as espécies parentais são altamente relacionadas, podendo até mesmo, segundo os autores, a espécie *A. monticola* ser descrita como *A. hypogaea* var. *monticola*. Segundo Wynne & Coffelt (1982), as características de herança qualitativa em *A. hypogaea* parecem ser controladas por locos duplicados, o que suporta a hipótese de que esta é uma espécie alotetraplóide.

Embora várias espécies tenham sido propostas como progenitores de *A. monticola*, diversos trabalhos foram realizados na tentativa de identificar os prováveis doadores dos genomas A e B em *Arachis hypogaea*, cujos dados baseados em evidências citogenéticas, aliadas ou não a dados moleculares e de cruzamentos interespecíficos, sugerem que *A. duranensis* teria sido a espécie doadora do genoma "A" (Fernández & Krapovickas, 1994; Kockert *et al.*, 1991; Kockert *et al.*, 1996; Singh, 1986) e *A. batizocoi* estaria relacionada ao genoma B (Cai *et al.*, 1987; Smartt *et al.*, 1978; Singh, 1986).

1.4 - *Arachis hypogaea*

O primeiro registro do amendoim foi descrito na literatura no início do século XVI por Linnaeus em 1753, como *Arachis* (do grego “*arachos*” significando erva daninha) e *hypogaea* (aposeno subterrâneo) (Figura 5 A a C).



Figura 5. Ilustrações da espécie *Arachis hypogaea*: (A) Fotografia da base da planta. (B) Detalhe da florescência. (C) Desenho feito por *Linnaeus* em 1753 em uma de suas viagens aos países tropicais. Adaptado das fontes eletrônicas: A e B) Clarke Brunt's Home Page- Peanut (*Arachis hypogaea*); C) Missouri Botanical Garden Library.

As suposições sobre a possível origem africana ou asiática foram descartadas através de numerosos estudos. Um desses estudos está relacionado a registros de achados arqueológicos em 1875 em tumbas pré-colombianas situadas nas proximidades de Ancon e Pachamac, na região da costa árida do Peru, datados por volta de 3.900-3.750 anos atrás (Hammons, 1994). Nos últimos 25 anos, a análise de diversas coleções de germoplasma de amendoins silvestres e cultivados obtidos do Noroeste e Nordeste da Argentina, Paraguai, Brasil, Bolívia, Uruguai, Peru e Equador confirmaram definitivamente a origem sul-americana desta espécie (Gregory & Gregory, 1976; INTA, 1986).

Embora *A. hypogaea* seja de longe a espécie que produz a semente mais cultivada e economicamente a mais importante do gênero (Knauff & Ozias-Akins, 1995; Freire *et al.*,

Introdução

1996; Freire, 1997), outras espécies também são utilizadas como fonte de alimentação em pequenas comunidades, por exemplo, indígenas (Stalker & Simpson, 1995). Observa-se ainda que outras espécies são largamente empregadas na cobertura de solos, nos controles da erosão e propagação de ervas daninhas ou como plantas forrageiras servindo de fonte de proteínas para a alimentação animal (Otero, 1941; Kerridge & Hardy, 1994) (Figura 6A à 6D).

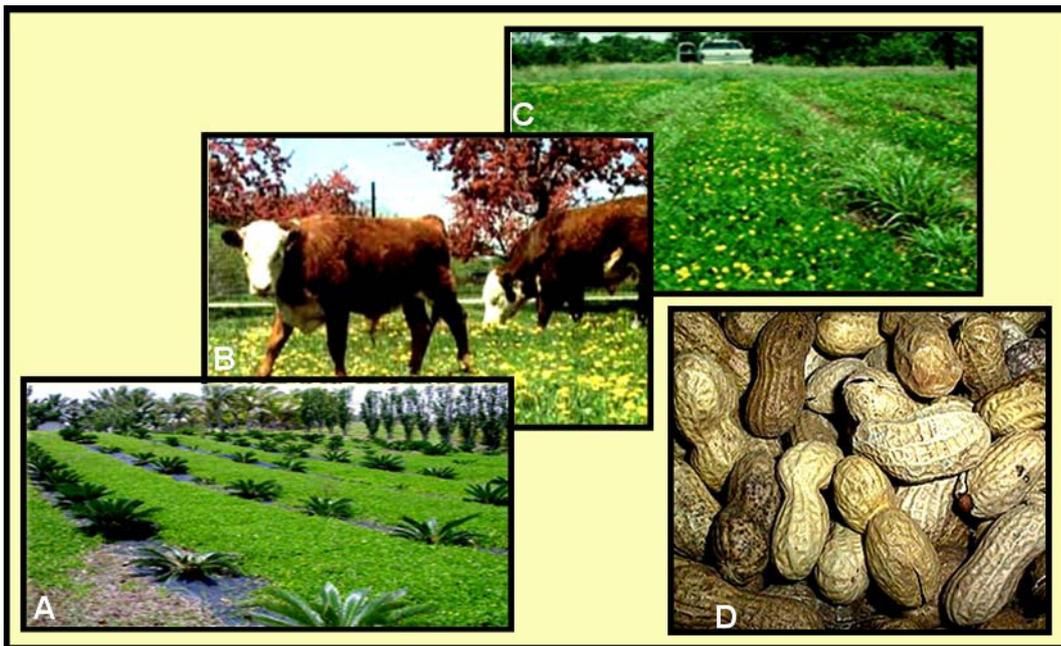


Figura 6. Forma de utilização das plantas pertencentes à secção *Arachis*. A e C) A espécie *Arachis pintoii* sendo usada como cobertura vegetal de solo contra a erosão; B) Planta forrageira utilizada como fonte protéica na alimentação animal e D) Amendoim *in natura* (vagem c/ sementes), destinado à alimentação humana.

Adaptado das fontes eletrônicas: A e C) DeFrank; B) S. G. Reynolds; C) <<http://pas.byu.edu/AgHrt100/peanut.htm>>.

Hoje o cultivo do amendoim está difundido nas áreas tropicais e subtropicais. Os maiores países produtores são a Índia, China, USA, África do Sul e Central/Oeste e o Brasil (Figura 7). Os principais estados brasileiros produtores de amendoim são respectivamente São Paulo, Rio Grande do Sul e Paraná (Ministério da Agricultura e do Abastecimento, 2000).

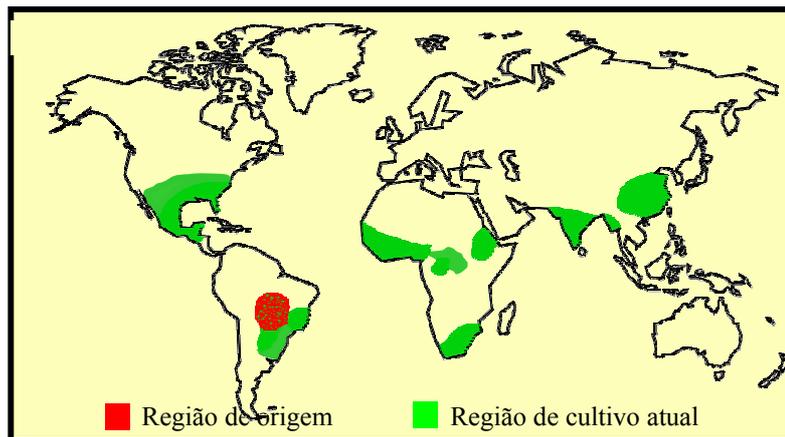


Figura 7. Mapa mundial indicando as regiões de origem e cultivo da espécie *A. hypogaea* na atualidade. Adaptado da fonte eletrônica: <<http://www.mpiz-koeln.mpg.de/pr/garten/schau/ArachishypogaeaL/Peanut.html>>

1.5 - Pragas e Doenças

Devido ao estreitamento da base genética do amendoim domesticado e a perda de genes importantes, principalmente àqueles ligados à resistência contra doenças, têm-se aumentado as preocupações por parte dos pesquisadores em relação a fungos, insetos e a patógenos presentes no solo. Em escala global, as doenças foliares causadas pelos fungos são consideradas como limitantes para a cultura, devido aos seus variados graus de severidade: *Cercospora arachidicola* (mancha castanha) (Figura 8 A); *Cercosporidium personatum* (mancha preta); *Phoma arachidicola* (mancha barrenta); *Puccinia arachidis* (ferrugem); *Sphaceloma arachidis* (verrugose) (Godoy *et al.*, 1997). Dentre os insetos, os pulgões (Homoptera) (Figura 8 B) e as lagartas (Lepidoptera) são os que lideram a lista de pragas. Com relação aos nematóides, estes são responsáveis por sérias perdas nas regiões basais da planta (Figura 8 C) (Isleib *et al.*, 1994). Juntas, estas pragas podem causar perdas superiores a 70% da colheita no cultivado (Subrahmanyam *et al.*, 1985a). Embora não seja diretamente a causa para a diminuição da produção no campo, as sementes infectadas por *Aspergillus flavus* (Figura 8 D) e *A. parasiticus* são comuns na maioria das regiões produtoras de amendoim, aumentando os custos de processamento pós-colheita além da obrigatoriedade dos testes de detecção de aflatoxinas (Santos *et al.*, 2001). A utilização indiscriminada de inseticidas, nematicidas e fungicidas pelos agricultores têm aumentado o

Introdução

custo de produção agrícola. Adicionalmente, face ao aumento da severidade das doenças, tem-se evidenciado uma diminuição da eficácia destes agentes agroquímicos no controle de pragas, principalmente daquelas relacionadas a patógenos do solo (Sholar *et al.*, 1995).

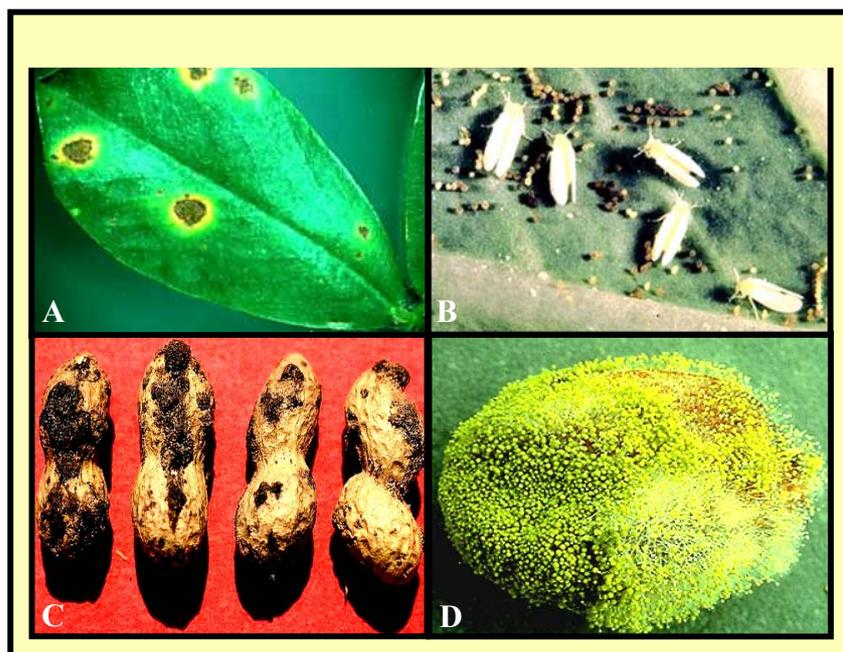


Figura 8. Tipos de pragas que comumente atacam o amendoim. A) Folha de *A. hypogaea* apresentando uma doença foliar causada pelo fungo *Cercospora arachidicola*. B) Mosca-branca do prateamento da folha (ordem Homoptera; *Bemisia argentifolii*) atacando a folha do amendoim. C) Vagens de amendoim atacadas pelo nematóide da espécie *Meloidogyne arenaria* e D) Semente de amendoim infectada por *Aspergillus flavus*. (A) Adaptado: Nancy Paiva; (B e C) Adaptado: Jim Castner, University of Florida e (D) Adaptado: B. Horn em *Compendium of Peanut Diseases*, 2ª ed.

O maior interesse pela prospecção, resgate e caracterização de germoplasma das espécies da secção *Arachis* reside em seu potencial de fornecimento de genes úteis para o melhoramento do amendoim cultivado (Stalker & Moss, 1987; Stalker, 1992), principalmente no que diz respeito à resistência a doenças, em que a base genética de *A. hypogaea* é muito estreita (Subrahmanyam *et al.*, 1985a). A caracterização do germoplasma obtido ao longo de coletas vem trazendo perspectivas interessantes, como a descoberta de novas fontes de resistência a doenças (Subrahmanyam *et al.*, 1985b); a rica qualidade da proteína encontrada nas sementes de algumas espécies; plantas com ciclo extremamente curto e alta resistência ao estresse hídrico, que são de grande utilidade para adaptação de variedades cultivadas do amendoim em áreas com maior carência de chuvas, como na África Ocidental e no Nordeste Brasileiro (Tabela 3).

Tabela 3. Características agronômicas, industriais e fitossanitárias encontradas em algumas espécies silvestres de *Arachis*.

Espécies	Características
<i>A. chacoense</i>	Resistência á cercosporiose, imunidade à ferrugem, trips a afídeos.
<i>A. cardenasii</i>	Resistência à <i>Cercosporidium personata</i> e a ferrugem.
<i>A. duranensis</i>	Resistência à ferrugem.
<i>A. correntina</i>	Imunidade à ferrugem.
<i>A. sternosperma</i>	Resistência à cercosporiose.
<i>A. batizocoi</i>	Resistência à cercosporiose, trips, jassídeos e afídeos.
<i>A. villosa</i>	Imunidade à ferrugem, sistema radicular profundo; maior regeneração na poda e número de sementes; alto conteúdo de óleo na semente; e resistência à seca.
<i>A. monticola</i>	Resistência à ferrugem e à cercosporiose.
<i>A. pusilla</i>	Resistência à ferrugem, cercosporiose e à virose do tomate.
<i>A. villosulicarpa</i>	Florescimento abundante e formação de 4-6 peg/axila foliar.
<i>A. glabrata</i>	Sistema radicular profundo e resistência à cercosporiose, ferrugem e trips; alto conteúdo de proteínas, sais minerais e óleo na semente.
<i>A. hagenbeckii</i>	Folíolo pouco denso e resistência à seca, alto conteúdo de proteína, sais minerais e óleo na semente.
<i>A. marginata</i>	Alto conteúdo de proteínas, sais minerais e óleo nas sementes.
<i>A. prostrata</i>	Resistência à seca.
<i>A. diogoi</i>	Alto conteúdo de sais minerais, proteína e óleo.

Adaptado: Murty *et al.*, 1981; Bajaj, 1984.

A espécies da seção *Arachis* têm sido avaliadas para características como porcentagem de óleos e ácidos graxos, capacidade de fixação de nitrogênio e pelo potencial de forrageamento (Knauff & Ozias-Akins, 1995; Freire *et al.*, 1996; Freire, 1997). Como algumas espécies são caracterizadas por apresentarem resistências múltiplas a determinadas doenças, elas têm como maior potencial agrônômico e de grande impacto econômico a possibilidade de aumentar a resistência do amendoim cultivado (*A. hypogaea*) (Stalker, 1992).

1.6 - Marcadores Moleculares para Análise Genética

A composição genética atual das diversas culturas de plantas é o resultado da domesticação e melhoramento a que estas foram submetidas durante os séculos. O melhoramento executado pelo homem primitivo resultava da simples procura e propagação de plantas com fenótipos mais adequados para as suas necessidades.

Após a redescoberta da lei de Mendel em 1900, o melhoramento de plantas, antes considerado como uma arte primitiva, passou a ser controlado por pesquisadores que, baseados na lei da segregação mendeliana, passaram a produzir progênies segregantes provenientes de linhagens combinadas por hibridização. As décadas seguintes foram marcadas por descobertas importantes, ampliando cada vez mais os conhecimentos da genética (Tabela 4).

Tabela 4. Fatos importantes que marcaram as décadas de 1910 a 1980.

Data	Fatos
1910	Descoberta da heterose; Fagos são descobertos.
1920	Desenvolvimento dos métodos clássicos de melhoramento.
1930	Descoberta da mutagênese.
1940	Grandes avanços na genética quantitativa.
1950	Watson e Crick revelam a estrutura tridimensional do DNA; Os estudos na área Fisiologia avançaram; Técnicas de cultivo de células são desenvolvidas.
1960	Utilização de marcadores bioquímicos; RNA mensageiro foi descoberto.
1970	Enzimas de restrição são identificadas; Hibridização de colônias e <i>Southern blotting</i> foram desenvolvidos para detectar seqüências de DNA específicas; Os primeiros anticorpos monoclonais foram produzidos; Genes de levedura são expressos em bactérias.
1980	O desenvolvimento da biologia molecular; A técnica de PCR (<i>Polymerase Chain Reaction</i>) foi desenvolvida; A primeira transformação de plantas por <i>Agrobacterium</i> (plasmídeos Ti) foi realizada; O primeiro cromossoma artificial foi sintetizado; Os primeiros marcadores moleculares para doenças hereditárias foram encontrados. A técnica de DNA fingerprinting foi desenvolvida. O projeto de seqüenciamento genoma humano foi aprovado.
1990	O primeiro tratamento de terapia gênica foi realizado; O primeiro câncer de mama foi descoberto; Cientistas relatam a clonagem de carneiros, usando DNA de um carneiro adulto.
2000	Cientistas do Estado de São Paulo revelam o código genético da primeira bactéria fitopatogênica, a <i>Xylella fastidiosa</i> ; Obtenção do arroz geneticamente modificado que produz beta-caroteno, precursor da vitamina A.

Adaptado: Borém, 1998;

Fonte eletrônica: Entendendo a biotecnologia, 2000.

Novas ferramentas são necessárias para assegurar que o crescimento da produtividade siga o mesmo ritmo do crescimento populacional. O surgimento de técnicas usando marcadores moleculares ajudou acelerar os métodos convencionais de melhoramento, como seleção assistida por marcadores e o mapeamento genético, diminuindo o tempo de liberação de materiais, o tamanho da população analisada e os custos para avaliação (Borém, 1998).

Os marcadores são definidos como elementos capazes de prever, mapear e caracterizar um fenótipo. Inicialmente, até meados da década de 60, os estudos genéticos eram realizados utilizando-se marcadores morfológicos de fácil identificação no organismo e geralmente regidos por um único gene. Apesar dos marcadores morfológicos terem contribuído significativamente para o estabelecimento dos princípios teóricos do mapeamento genético e das análises de ligação gênica (Knapp, 1991), o pequeno número de marcadores morfológicos diferentes em uma mesma linhagem diminui a probabilidade de se encontrar associações significativas entre esses marcadores e caracteres de importância para programas econômicos ou para estudos teóricos (Borém, 1998). Além disso, esses marcadores freqüentemente são afetados pela ação gênica de dominância, efeito ambiental, pleiotropia e epistasia (Paterson *et al.*, 1991).

O primeiro grande passo para resolver estes problemas foi o desenvolvimento de marcadores isoenzimáticos, que são baseados na análise do produto da expressão de genes (Moss, 1982). Eles revelaram-se como uma nova fonte de marcadores genéticos com inúmeras vantagens sobre os marcadores morfológicos como co-dominância, insensibilidade a pleiotropia e a epistasia (Paterson *et al.*, 1991). Desde a sua resolução pelos métodos histoquímicos, as isoenzimas têm sido extremamente importantes para as investigações sobre variação intra-específica, genética de populações, evolução e mapeamento genético. Em plantas, as isoenzimas têm sido utilizadas principalmente em genética de populações, evolução, e caracterização de germoplasma (Gottlieb, 1981; Tanksley & Orton, 1983; Soltis & Soltis, 1989; Pasteur *et al.*, 1988; Hillis *et al.*, 1996).

No entanto, ainda se trata de um modo indireto de se estudar os genes. As técnicas de biologia molecular foram aprimorando-se e o DNA passou a ser o ponto focal no

desenvolvimento de novos marcadores (Borém, 1998). Os distintos tipos de marcadores moleculares hoje disponíveis diferenciam-se pela tecnologia utilizada para revelar variabilidade no DNA, pelo polimorfismo molecular, expressão genética, distribuição no genoma, custo de implementação e operação (Griffiths *et al.*, 2000).

Inicialmente, a utilização de enzimas de restrição no estudo direto do DNA permitiu a análise e a comparação do comprimento de fragmentos gerados pela clivagem do material genético através da detecção por hibridização da seqüência clonada (Griffiths *et al.*, 2000). As variações nos nucleotídeos do DNA devido à mutação, deleção, inserção e inversão, podem ser detectadas se ocorrerem num sítio de corte das enzimas de restrição, gerando fragmentos de diferentes tamanhos, portanto polimórficos (Lander *et al.*, 1989). A técnica que se baseia nestes fragmentos é denominada de RFLP (*Restriction Fragment Length Polimorphism* - polimorfismo de restrição de DNA) e foi desenvolvida por Botstein e colaboradores (1980). A grande vantagem dos marcadores RFLPs é a sua ampla distribuição no genoma, permitindo uma cobertura adequada e proporcionando a construção de densos mapas genéticos de ligação, que possibilitam a realização de análises genéticas e moleculares e várias aplicações no melhoramento de plantas, como clonagem de genes e mapeamento de QTLs (*Quantitative trait loci*- locos controladores de características quantitativas), além de ter expressão co-dominante (Nodari *et al.*, 1993). Entretanto o elevado custo, instalações apropriadas ao manuseio e descarte de material radioativo e o tempo necessário na geração destes marcadores restringem drasticamente seu uso de forma freqüente (Ferreira & Grattapaglia, 1998).

Com os avanços nos estudos sobre o genoma dos organismos, surgiram marcadores baseados nas seqüências repetitivas do DNA. Jeffreys e colaboradores (1985) mostraram que muitos RFLPs eram causados por minissatélites ou VNTR (*Variable Number of Tandem Repeats*, Wyman & White, 1980), que são regiões dispersas no genoma que contém um número variável de seqüências repetidas com tamanhos intermediários entre 0,5 a 30 Kb e enfileiradas em tandem de DNA com comprimento variável entre 10 a 100 pb, altamente variáveis (Armour *et al.*, 1999). O processo de detecção de VNTR é bastante semelhante ao RFLP, contudo o polimorfismo depende da freqüência de pequenas regiões repetitivas entre os sítios de restrição no genoma (Ferreira & Grattapaglia, 1998).

A vantagem deste tipo de marcador é a grande quantidade de informação que pode ser obtida em um único teste, devido o cruzamento de informações produzidas simultaneamente pelo RFLP com as sondas VNTR, gerando um complexo padrão de bandas abrangendo todos os locos hipervariáveis do genoma concomitantemente (Armour *et al.*, 1999). Outra vantagem desta técnica é a possibilidade de utilização de uma mesma sonda para detecção de locos hipervariáveis em uma larga variedade de espécies (Ferreira & Grattapaglia, 1998). Em contrapartida, devido à complexidade do perfil gerado de bandas, nem todos os alelos correspondentes a cada loco podem ser identificados. Adicionalmente, por ser uma técnica baseada em RFLP, compartilham as mesmas limitações (Ferreira & Grattapaglia, 1998). Uso destes marcadores está concentrado no melhoramento de plantas para a identificação de variedades, cultivares e clones, na análise de diversidade genética e determinação de paternidade (Dallas, 1988; Nybom & Hall, 1991; Rogstad *et al.*, 1988; Broun *et al.*, 1992).

Mais recentemente, o advento de técnicas baseadas em PCR (*Polymerase Chain Reaction* - Reação em Cadeia pela Polimerase) apresentou uma nova opção ao uso de marcadores moleculares. Esta técnica (PCR) foi concebida em 1983 por Kary Mullis (Prêmio Nobel em 1993), publicada em 1987, sendo utilizada de forma rotineira em diversas áreas da biologia, tanto em pesquisa básica como na aplicada a partir de 1988 (Saiki *et al.*, 1988).

PCR é uma técnica baseada na síntese enzimática “*in vitro*” de milhões de cópias de um fragmento específico de DNA na presença da enzima DNA polimerase. O processo inicia-se com a hibridação e extensão enzimática de um par de iniciadores (*primers*) que delimitam a seqüência de DNA de fita dupla, alvo da amplificação.

Com a PCR e as novas técnicas de genética molecular, os marcadores moleculares de DNA (marcadores genéticos baseados na variabilidade de DNA) têm contribuído como ferramenta para incrementar a eficiência do melhoramento de plantas, por meio de metodologias como caracterização de germoplasma, seleção assistida e mapeamento genético (Ferreira & Grattapaglia, 1998).

Na década de 80, surgiu um novo tipo de marcador molecular denominado de RAPD (*Randomly Amplified Polymorphic DNA*; DNA Polimórfico Amplificado ao Acaso, Welsh & McClelland, 1990; Williams *et al.*, 1990). O RAPD utiliza como iniciadores oligonucleotídeos curtos, comumente com 10 nucleotídeos, de seqüência nucleotídica arbitrária. A reação de amplificação procede em condições de baixa estringência, permitindo que ocorra o pareamento entre os iniciadores e o DNA-alvo, mesmo que não haja complementaridade total entre as duas seqüências. Os produtos resultantes da amplificação podem ser visualizados como bandas em géis de agarose ou poliacrilamida. Diferenças ao nível de DNA são inferidas pela presença ou ausência de um determinado fragmento amplificado que produzem um padrão de bandas específico conhecido como impressão digital genômica (*DNA fingerprint*). Os *fingerprints* permitem distinguir diferentes espécies e até mesmo populações distintas da mesma espécie. As principais vantagens deste marcador estão na simplicidade e rapidez na obtenção dos resultados, além da mínima quantidade de DNA necessária para a análise genotípica de um indivíduo. Sendo o RAPD um marcador com característica dominante, sua principal limitação está justamente no baixo conteúdo informativo por loco (Ferreira & Grattapaglia, 1998).

Em 1993, os dois pesquisadores Zabeau e Vos, desenvolveram e divulgaram outro marcador molecular denominado AFLP (*Amplified Fragment Length Polymorphism* - Polimorfismo de Comprimento de Fragmentos Amplificados). Esta técnica combina os alvos para estudo do polimorfismo pelos marcadores RFLP e RAPD: a distribuição aleatória de sítios de restrição entre genomas e a amplificação aleatória de fragmentos empregando-se primers de seqüências arbitrárias. A característica fundamental desta técnica é a capacidade de revelar simultaneamente muitas regiões diferentes distribuídas aleatoriamente ao longo do genoma (Mueller & Wolfenbarger, 1999). A análise de AFLP é baseada na amplificação seletiva via PCR de um subconjunto de fragmentos genômicos gerados após digestão com uma enzima de corte raro combinada com uma enzima de corte freqüente. A principal vantagem desta tecnologia é o grande número de fragmentos que são gerados simultaneamente, aumentando o poder de detecção de variabilidade genética. Por esta razão, ela tem sido utilizada principalmente no estudo da diversidade genética entre indivíduos (Muluvi *et al.*, 1999; Loh *et al.*, 2000) e na construção de mapas genéticos e mapeamento de genes de interesse em diferentes espécies (Virk *et al.*, 1998; Hartl *et al.*,

Os microssatélites têm sido detectados dentro dos genomas de todos os organismos até agora analisados (Hancock, 1999), principalmente no interior de regiões não codificantes do genoma. Pode também ser encontrado numa taxa menor em seqüências de regiões promotoras codificantes, como a região promotora do gene *Drosophila Ubx* (Biggin & Tjian, 1988), sítios de ligação para várias proteínas regulatórias (Lue *et al.*, 1989; Csink & Henikoff, 1998), como por exemplo, ligação de poly (GA)-poly(GT)-proteína em fibroblastos humanos (Aharoni *et al.*, 1993). Morgante e colaboradores (2002) citaram que com exceção dos trinucleotídeos, todos os tipos de microssatélites são significativamente menos freqüentes em seqüências preditas como codificantes de proteínas quando comparadas com a fração não codificante. Neste estudo, os autores analisaram 25.762 seqüências em cinco espécies de plantas (*Arabidopsis thaliana*, arroz, soja, milho e trigo).

Em genomas de eucariotos estas seqüências são muito mais freqüentes e melhor distribuídas ao acaso no genoma. Apresentam tamanhos mais variantes, formando locos genéticos muito mais polimórficos (Hamada *et al.*, 1992). Nos procariotos as quantidades de seqüências repetidas em tandem são muito baixas (Hancock, 1995; Field & Wills, 1998).

A maioria dos microssatélites (48-67%) encontrados em muitas espécies estudadas são dinucleotídeos (Wang *et al.*, 1994; Schug *et al.*, 1998). No genoma humano, dentre todas as seqüências repetitivas em tandem (mono a hexanucleotídeos), o poly A/T é o mais comum (Beckmann & Weber, 1992; Stallings 1992; Tóth *et al.*, 2000; Wren *et al.*, 2000). Das repetições dinucleotídeos, a mais freqüente é AC/TG, que ocorre duas vezes mais que repetições do tipo AT/TA e três vezes mais que repetições AG/TC (Beckmann & Weber, 1992). Em plantas, repetições AT/TA e GA/CT são mais comuns do que repetições CA/GT (Stallings, 1992; Lagercrantz *et al.*, 1993). Por outro lado às repetições do tipo CA/GT são as mais comuns em *Drosophila melanogaster* (Schug *et al.*, 1998). Os procariotos têm poucos microssatélites quando comparados com o genoma de eucariotos e possuem repetições do tipo poly A/T em sua maioria, seguidas das repetições AT/TA (Belkun *et al.*, 1998; Gur-Arie *et al.*, 2000).

Os microssatélites são utilizados como marcadores moleculares devido ao alto nível de polimorfismo encontrado em seus locos (Litt & Luty, 1989; Weber & May, 1989; Tautz, 1989), o que proporciona sua utilização em diversos tipos de estudos populacionais, permitindo analisar desde indivíduos até espécies proximamente relacionadas (Brown *et al.*, 1996). E isto é possível devido ao fato que as regiões flanqueadoras complementares aos primers são frequentemente conservadas dentro da mesma espécie ou entre espécies de gêneros correlatos, ainda que as regiões microssatélites estejam sujeitas à alta taxa de mutação (Hopkins *et al.*, 1999).

O “deslizamento” da DNA polimerase durante a replicação do DNA é tido como a principal causa da variação no número de repetições nesses locos, como foi demonstrado *in vitro* por Sclötterer e Tautz (1992), caracterizando a alta taxa de mutação (10^{-2} a 10^{-6} eventos por locos por geração) destas regiões repetidas quando comparadas com as regiões codificantes. Alguns destes erros são corrigidos pela atividade das exonucleases que revisam e reparam o DNA, mas outros escapam de serem corrigidos, tornando-se mutações, por exemplo, as repetições CTG/CAG ou CGG/CCG podem formar estruturas secundárias denominadas de *hairpin*, impedindo o reparo do DNA em leveduras (Moore *et al.*, 1999). Os diferentes números de repetições caracterizam os diferentes alelos que podem ser detectados em gel de eletroforese utilizando-se poliacrilamida ou agarose especial de alta resolução (Figura 10).

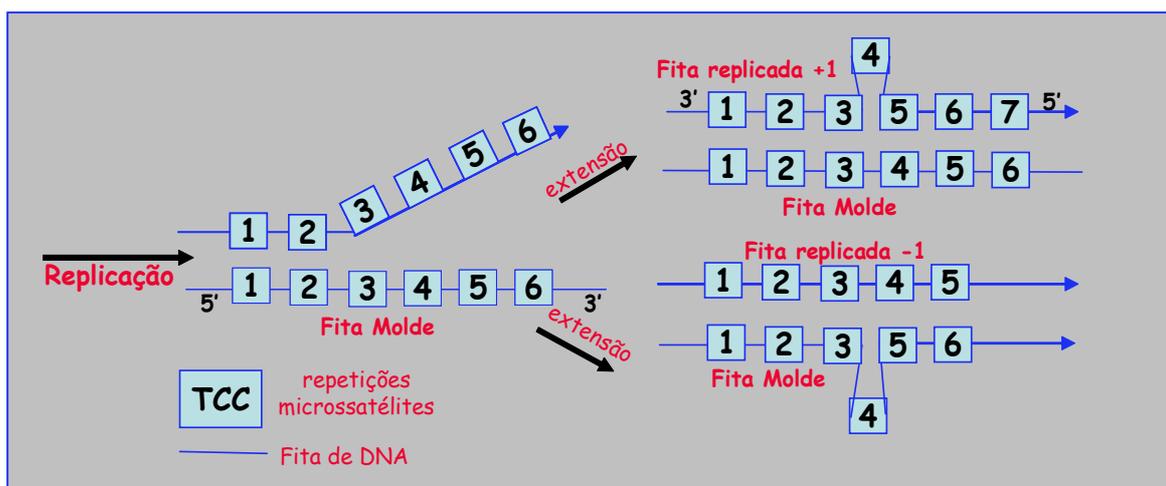


Figura 10. Modelo do processo de mutação do loco microssatélite. Na figura, o DNA é representado pela linha azul, repetições microssatélites são os blocos azuis claros e as setas indicam a replicação do DNA da fita molde.

Adaptado: Eisen, J.1999

Marcadores microssatélites são tipicamente de natureza co-dominante, possuem alto polimorfismo genético, riqueza de alelos por loco, alta heterozigidade e o mais elevado conteúdo informativo dentre os marcadores moleculares. Adicionalmente, cada loco de microssatélite pode ser amplificado por PCR e analisado individualmente ou por *multiplex* onde vários locos podem ser amplificados e analisados de uma só vez em sistemas semi-automatizados ou automatizados, características estas altamente desejadas em um marcador molecular (Ferreira & Grattapaglia, 1998).

Os marcadores microssatélites são mais simples de usar do que os marcadores RFLPs e diferentemente destes, requerem pouca quantidade (nanogramas) e qualidade de DNA. Possuem ainda a vantagem de serem trocados entre pesquisadores ou centros de pesquisa, pois cada loco é definido por um par de seqüências de primer, mantendo o mesmo resultado, independente de onde o pesquisador esteja trabalhando. Ensaio com microssatélites são mais robustos do que RAPDs, e mais transferíveis do que AFLPs. Quando comparados com os outros marcadores, RFLP; RAPD e AFLP, os microssatélites apresentaram o mais alto conteúdo informativo, revelando a média mínima de 0,60 de heterozigidade dos locos das espécies estudadas (Powell *et al.* 1996a; Powell *et al.* 1996b). A grande limitação do uso em larga escala de marcadores microssatélites é a obtenção dos primers específicos, pois requer trabalho intenso no desenvolvimento e o custo inicial é elevado (Rafalski & Tingey, 1993; Brown *et al.*, 1996).

Os marcadores microssatélites podem ser usados para toda e qualquer população segregante, em estudos de ligação e mapeamento genético, principalmente quando co-utilizados com AFLPs, produzindo mapas genéticos muito detalhados (Wenburg *et al.*, 1996; Stephenson *et al.*, 1998). O marcador microssatélite devido à habilidade de revelar alta diversidade de alelos por loco é usado em genotipagem de indivíduos (Cregan *et al.*, 1994). Devido à natureza co-dominante estes marcadores podem ser usados em estudos de parentesco e de proteção e identificação de variedades de plantas em bancos de germoplasma (Lopes *et al.*, 1999; Diwan & Cregan, 1997; Marklund *et al.*, 1994).

Os cinco tipos de marcadores citados anteriormente (Isoenzimas, RFLP, RAPD, AFLP e SSR) são os mais utilizados atualmente no estudo do genoma de plantas, animais e humanos. Para facilitar a diferenciação desses marcadores, a Tabela 5 traz uma breve análise comparativa entre eles. Entretanto, na literatura existem outros marcadores, que

Introdução

muitas vezes apresentam apenas uma pequena modificação em relação a alguns desses já comentados (Tabela 6).

Tabela 5. Análise comparativa entre os marcadores moleculares.

Atributos	Isoenzimas	RFLPs	RAPDs	AFLPs	Microsatélites
Nível de Polimorfismo	Baixo	baixo-alto	baixo-alto	muito alto	muito alto
Estabilidade ambiental	moderada	alta	alta	alta	alta
Número de locos	Moderado<50	alto	alto	alto	alto
Expressão genética	co-dominante	co-dominante	dominante	dominante	co-dominante
Número de alelos por loco	2 – 5	multialélico	2	2	multialélico
Distribuição no genoma	regiões de cópia única	várias	ao acaso	ao acaso	ao acaso
Acessibilidade tecnológica	Muito alta	média	muito alta	média	muito baixa
Aplicabilidade melhoramento	rápido, custo baixo	lento, custo médio	rápido, custo baixo	rápido, custo baixo	lento, custo alto
Identificação de genótipos	Baixa	alta	muito alta	muito alta	muito alta
Avaliação de germoplasma	Média	alta	alta	muito alta	alta
Mapeamento genético	Baixa	alta	alta	alta	muito alta
Mapeamento de regiões específicas	Baixa	média	muito alta	muito alta	média
Mapeamento comparativo	Baixa	muito alta	baixa	baixa	alta
Genética de Autógamas	Baixa	média	alta	muito alta	muito alta
Genética de Alógamas	Média	média	alta	muito alta	muito alta
Análise Filogenética	Média	muito alta	média	Alta	Média

Adaptado de Gepts, 1993; Ferreira & Grattapaglia, 1995.

Tabela 6. Descrição sucinta dos demais marcadores existentes na literatura.

TIPO	DESCRIÇÃO
AP-PCR <i>Arbitrary Primer PCR</i>	Semelhante ao RAPD, constitui-se por ampliações ao acaso utilizando primer com 20 bases, empregando-se temperaturas de anelamento inferiores à TM exigida pelo primer. Característica Dominante, detecção em gel de agarose (Welsh & McClelland, 1990).
CAPS <i>Cleaved Amplified Polymorphic Sequence</i>	Constitui-se da digestão com enzimas de restrição de fragmentos polimórficos amplificados com um primer específico. Função: revelar o polimorfismo presente nos sítios de restrição do fragmento. Co-dominante e detecção em gel de agarose ou prata. (Konieczny & Ausubel, 1993).
DAF <i>DNA Amplified Fingerprinting</i>	Mesmo princípio que o RAPD, utilizando primer mais curtos (entre 5 a 10 bases). Os perfis eletroforéticos dos produtos amplificados são extremamente complexos, apresentando um grande número de bandas. (Caetano-Anóles <i>et al.</i> , 1991).
DGGE <i>Differential Gradient Gel Electrophoresis</i>	Visa à detecção de fragmentos com polimorfismo na seqüência de bases, após eletroforese em gel de acrilamida contendo um gradiente do produto desnaturante. (Myers <i>et al.</i> , 1987)
IMA <i>Inter Microsatellite Amplification</i>	Amplificação do DNA genômico com primers cuja seqüência de bases é constituída de um motivo repetitivo (geralmente dinucleotídeos) seguido de 2 a 6 bases definidas ao acaso. (Zietkiewicz <i>et al.</i> , 1994). Este marcador também é conhecido por Inter-SSR PCR; ISA ou IRA.
SCAR <i>Sequence Characterized Amplified Region</i>	Fragmento de DNA genômico amplificado por PCR com primers específicos (14 a 20 bases). Estes primers são definidos após o conhecimento da seqüência de bases do fragmento. Esta técnica originou-se a partir do isolamento seguido de seqüenciamento de fragmentos amplificados por RAPD (Paran & Michelmore, 1993).
SSCP <i>Single Strand Conformation Profile</i>	Análise da conformação de fragmentos especificamente amplificados, desnaturados e migrados em gel de acrilamida. As fitas simples assumirão uma conformação estrutural diferente conforma a seqüência de bases que elas apresentam, a qual modificará a mobilidade eletroforética da mesma. Através da comparação da migração da fitas simples de um mesmo fragmento amplificados em diferentes indivíduos, é possível determinar aqueles que apresentam polimorfismo na seqüência de bases (Orita <i>et al.</i> , 1989).
STS <i>Site Tagged Sequence</i>	Corresponde a qualquer seqüência genômica amplificada com primers específicos. (Olso <i>et al.</i> , 1989).

Adaptado: Souza, 2001

1.7 - Aplicações de Marcadores no Melhoramento de Plantas.

Os marcadores moleculares facilitam a realização de estudos de genética, taxonomia e evolução de plantas proporcionando um substancial avanço no conhecimento científico. As principais implicações deste avanço no conhecimento se refletem no poder, precisão e rapidez na manipulação da variabilidade genética. De um modo geral, diversas aplicações de marcadores microssatélites em melhoramento genético podem ser distribuídas em aplicações cujos resultados apresentam expectativas de curto, médio e longo prazo.

As aplicações de curto prazo envolvem, basicamente, a identificação e discriminação de genótipos. Nas aplicações analíticas de médio-longo prazo, os marcadores permitem quantificar a variabilidade genética existente ao nível de sequência de DNA e correlacioná-la com a expressão fenotípica em procedimento de mapeamento genético de QTLs (Tabela 7).

Tabela 7. Principais aplicações de metodologias baseadas em marcadores microssatélites no melhoramento de plantas.

Melhoramento Clássico

- **Aplicações de curto prazo:**
 - Identificação de origem parental.
 - Identificação e proteção de variedades
 - Monitoramento de fecundação cruzada e autofecundação em plantas
 - Avaliação de germoplasma e populações de melhoramento.
 - **Aplicações de médio a longo prazo.**
 - Construção de mapas genéticos.
 - Mapeamento genético de QTL (*Quantitative Trait Loci*).
 - Introgessão de características via cruzamento assistido por marcadores microssatélites.
 - Seleção durante o desenvolvimento de linhagens endógamas.
 - Seleção indireta para características de difícil avaliação
 - Seleção precoce em culturas perenes
-

Adaptado: Beckmann, 1991 e Ferreira & Grattapaglia, 1998.

No melhoramento de plantas, o desenvolvimento de mapas genéticos é considerado uma das aplicações de maior impacto, pois possibilita a cobertura completa de genomas; a decomposição de características genéticas complexas nos seus componentes Mendelianos (padrões de distorções de segregação Mendeliana de segmentos cromossômicos, presença de inversões, translocações e duplicações de segmentos de DNA); a localização das regiões que controlam caracteres de importância; e a quantificação do efeito destas regiões na característica.

Os marcadores microssatélites também são utilizados na seleção assistida em programas de retrocruzamento, permitindo com isso a redução do tempo necessário para a transferência de um gene de interesse de um genótipo doador, com menor adaptação, para um genótipo elite e adaptado. Tanksley e colaboradores (1989) afirmam que o uso de marcadores moleculares para acessar e transferir genes de germoplasma exótico para variedades cultivadas poderá vir a ser a contribuição mais significativa desta tecnologia no melhoramento de plantas.

Uma outra aplicação para estes marcadores e a seleção assistida, que permite precocemente a seleção indireta de indivíduos portadores de genes de interesses fortemente ligados a estes marcadores, possibilita ainda a seleção de alelos com efeitos positivos provenientes dos dois ou mais progenitores envolvidos na geração da população segregante (Lande & Thompson, 1990).

1.8 - Marcadores Moleculares em *Arachis hypogaea*.

O amendoim é uma das poucas espécies que apresentam baixa variabilidade genética quando detectados por marcadores moleculares como RFLPs, Isoenzimas e RAPDs (Stalker & Mozingo, 2001). Somente poucos marcadores moleculares têm sido identificados e ligados a genes de resistência (Garcia *et al.*, 1996; Burow *et al.*, 1996).

He e Prakash (1997) foram os primeiros pesquisadores que descreveram as aplicações da tecnologia do AFLP em amendoins. Eles usaram 28 pares de primers para gerar 111 marcadores AFLP em *A. hypogaea*. Embora um considerável número de marcadores AFLP tem sido identificado, substancialmente mais marcadores de DNA são necessários para saturar o mapa de ligação do amendoim e iniciar os estudos genéticos desta planta (He & Prakash, 1997). Mila e Stalker (dados não publicados) relataram que embora a técnica AFLP tenha se mostrado bastante polimórfica para a grande maioria das espécies de plantas estudadas, em amendoins a mesma variação não foi detectada.

Os microssatélites são descritos como mais variáveis do que RFLPs ou RAPDs, e tem sido largamente utilizados em estudos de genoma de plantas (Tanaka *et al.*, 1999). Hopkins e colaboradores (1999) relataram seis microssatélites polimórficos em *A. hypogaea* com o número de alelos por loco variando entre 5 a 14. Em trabalho recente, He e colaboradores (2003) utilizando a metodologia de biblioteca genômica enriquecida para três diferentes dinucleotídeos, AT, GT e GA, além das enzimas de restrição *HindIII* e *MseI* para a fragmentação do DNA genômico total, isolaram e desenharam pares de primers para 56 diferentes microssatélites, sendo que 19 mostraram-se polimórficos dentro dos genótipos estudados. Embora os microssatélites estejam presentes no amendoim, estas repetições não tem sido totalmente identificadas, desenvolvidas ou utilizadas, e as informações detalhadas com relação à abundância de microssatélites ainda permanecem em branco (Hopkins *et al.* 1999; Stalker & Mozingo, 2001).

O desenvolvimento de marcadores SSRs envolve o conhecimento das seqüências únicas adjacentes à região contendo as repetições em tandem, para que seja possível designar um loco específico. A identificação de novos locos SSR pode ser iniciada com a procura de seqüências conhecidas em bancos de dados públicos (*GenBank*) ou através da identificação de clones que contenham locos microssatélites gerados a partir de bibliotecas genômicas. Brown e colaboradores (1996) confrontaram as três metodologias utilizadas para a identificação de marcadores SSR em plantas: (i) bancos de dados públicos, (ii) amplificação com *primers* de espécies relacionadas e (iii) seleção por biblioteca genômica. O banco de dados foi avaliado como o de menor custo em termos de tempo e recursos para obtenção de novos locos microssatélites, mas apresenta uma fonte limitada de dados.

Com relação a transferibilidade de primers SSR entre espécies relacionadas é difícil de prever. A distância taxonômica das espécies de interesse e a conservação das respectivas seqüências flanqueadoras são determinantes para amplificar uma determinada região (Brown *et al.*, 1996). Além disso, as condições de reação precisam ser otimizadas e os produtos seqüenciados ou hibridizados com sondas específicas para verificar quanto à presença da região de seqüência repetida (Westman & Kresovich, 1998). A seleção a partir de uma biblioteca genômica, apesar do esforço e do custo inicial, pode ser a melhor opção para o isolamento de novos locos microssatélites (Maguire, 2001).

A busca ao acaso por regiões repetidas oriundas de uma biblioteca genômica total sem enriquecimento, provou ser um método ineficiente quando comparadas bibliotecas enriquecidas ou com outras técnicas de seleção (Brown *et al.*, 1996). Aproximadamente apenas 0,2% dos clones de biblioteca hibridizaram com uma sonda de microssatélite. Após o seqüenciamento, 70% dos clones seqüenciados não puderam ser utilizados para desenhar os primers por apresentarem problemas específicos, tais como: ausência de microssatélites devido à ocorrência de hibridização inespecífica; falta de espaço para o desenho dos primers nas regiões flanqueadoras; ou apresentavam regiões com poucas repetições. Dos 30% restantes selecionados, 65% deles não apresentaram produtos polimórficos na espécie da planta usada no teste.

1.9 - A Bioinformática como Ferramenta.

A bioinformática é ainda considerada uma recente subdivisão da biotecnologia e representa o “casamento” da biotecnologia com a informática. De modo simples, a bioinformática consiste no depósito e análise de seqüências genéticas em bancos de dados e conseqüente manipulação e análise destas seqüências com a utilização de softwares específicos (Baxevanis, 2001).

A primeira base de dados de biologia molecular parece ter surgido por volta de 1960, quando Dayhoff e colaboradores construíram um catálogo contendo todas as seqüências de proteínas conhecidas até a data. Essas seqüências foram publicadas num livro chamado “*Atlas of Protein Sequences and Structure*”, de 1965 (Baxevanis, 2001).

Com o advento do seqüenciamento do DNA e, principalmente, a partir da década de 1990, do seqüenciamento em larga escala, como os projetos genomas, foi necessária à construção de bancos de dados mais robustos para abrigar a explosão no número de seqüências obtidas pelos pesquisadores (Bairoch & Apweiler, 2002). O estabelecimento de bancos de dados públicos possibilitou que os cientistas tenham acesso à informação proveniente de outros laboratórios e que troquem e compartilhem seqüências genéticas (Wheeller *et al.*, 2002). Todos os dias novas seqüências são depositadas no banco de dados públicos do NCBI (*National Center for Biotechnology Information*), o chamado GenBank (Figura 11).

Dessa forma foi criada uma instituição com colaboração internacional para montar um banco de dados primários de seqüências de nucleotídeos chamada de INSDC (*International Nucleotide Sequence Database Collaboration*). Essa instituição contém o NCBI, o EMBL (*European Molecular Biology Laboratory*) e o DDJB (*DNA Data Bank of Japan*) (Tateno *et al.*, 2002). Cada um desses centros possibilita a submissão individual de seqüências de DNA e trocam informações entre si diariamente, sendo que todos os três possuem informações atualizadas de todas as seqüências disponíveis para os pesquisadores (Stoesser *et al.*, 2002).

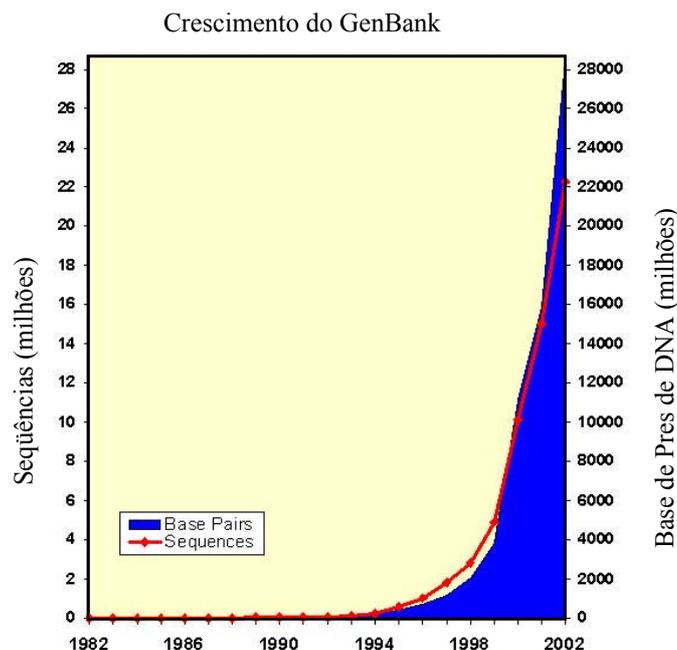


Figura 11. Crescimento exponencial do número de seqüências contidas no GenBank ao longo das duas últimas décadas.

Adaptado da fonte eletrônica: < <http://www.ncbi.nlm.nih.gov/Genbank/genbankstats.html> >

Softwares para análises de seqüências primárias foram desenvolvidos para solucionar problemas comuns ocasionados pela incorreta leitura dos dados gerados pelos equipamentos utilizado nos seqüenciamentos durante a geração de dados. Este processo de análise de dados é chamado *base calling*, que consiste na interpretação dos dados brutos associados às bases A, C, G, T, em cada posição, e a designação de uma pontuação correspondente à qualidade, de forma a refletir uma probabilidade estatística que indica se a *base call* está correta. As pontuações resultantes da análise fornecem de maneira segura a possibilidade de monitorar a qualidade dos dados brutos e de detectar a sobreposição entre as seqüências lidas (Gibas & Jambeck, 2001; Green, 2001). Em adição a este processo, existem outros programas que inspecionam e conferem um valor numérico qualitativo após o *base call*, o mais conhecido é o programa Phred (Ewing *et al.*, 1998) com exceção dos trinucleotídeos

1.9.1 - Processamento de Cromatogramas de Seqüências Definitivas.

A verificação das seqüências é uma das etapas fundamentais no processamento de dados, pois neste momento são verificados não só redundância das seqüências depositadas, como também as corretas sobreposições, facilitando a etapa de edição de nucleotídeos

duvidosos seja devido à má qualidade por contaminantes ou por erros de amplificação na PCR. As sobreposições auxiliam não somente na visualização como facilitam a compreensão das seqüências, principalmente quando se há necessidade de ressequenciá-las várias vezes em ambos os sentidos (*forward e reverse*).

Existem vários programas disponíveis que comparam as seqüências entre si de forma a encontrar quais delas são idênticas ou contêm regiões parecidas o suficiente para que sejam reunidas em uma só. As seqüências que foram sobrepostas pelo programa, ou seja, tiveram similaridade entre si, recebem o nome de *contigs*, e a seqüência resultante destas sobreposições é chamada de *consensus*. Os vários programas disponíveis desempenham funções similares, os mais populares são o Phrap (Green, 1999), o GAP – (*Genome Assembly Program*) (Bonfield *et al.*, 1995) e o CAP3 (*Sequence Assembly Program*) (Huang & Madan, 1999). Estes programas são dotados de padrões de segurança calculados de forma a fornecer critérios objetivos para a subsequente finalização de seqüências (Green, 2001).

Vários pacotes de softwares existem hoje disponíveis no mercado da bioinformática para os pesquisadores, disponibilizando dentro deles não somente os programas de processamento de dados *eg.* Phrap e CAP3, mas outros com funções distintas e de similar importância. Estes pacotes apresentam quase todas as mesmas funções, diferenciando na presença ou ausência de alguns programas específicos, na interface com o usuário ou no preço do produto; SequencherTM (Genes Codes Corporation), Consed (*freeware*) e Staden Package (*freeware*) são exemplos de alguns dos produtos disponíveis.

1.9.2 - Software Staden Package

Com a finalidade de reunir em um único pacote vários programas interativos, Staden e colaboradores desenvolveram o software denominado Staden. O foco principal deste programa foi o de melhorar o processamento de dados de projetos em larga escala, como os genomas, adicionando em um único pacote, componentes de análise avançadas e interativas como: Pregap 4; Gap 4, Prefinish e Spin.

O pregap 4 tem como função processar os dados “cru” do seqüenciamentos antes de serem analisados pelo gap 4. A seqüências passam por vários módulos, como a análise de qualidade; identificação de regiões contaminantes de vetor; cortes de vetores (*cloning vector clipping*), e conversão de formatos de dados (*trace format conversion*). O pregap 4 provê uma interface que permite montar e configurar cada módulo individualmente, processando-os todos juntos automaticamente ao final do programa (Figura 12)

A interface gap 4 tem como funções principais a de gerar *contigs*, realizar a inspeção visual e editar as seqüências de bases duvidosas (Figura 13 A a D). Módulos específicos como *assembly independently* (permite a sobreposição de seqüências novas com as já existentes no banco de dados, porém sem uni-las), *find internal join* (compara todos os contigs que estão presentes na base de dados e as prováveis sobreposições são visualizados em um gráfico, onde será permitindo juntar, editar os *contigs* redundantes ou rejeitá-los) (Figura 14 A a D), *tags* (permite acrescentar comentários e cores específicas a determinadas bases ou a toda uma seqüência), *Disassemble Readings and Break Contig* (erros de sobreposições podem ser corrigidos através da retirada do *contig* do banco de dados ou da região de leitura) entre outros, fazem parte desta interface que podem auxiliar de uma forma mais rápida a completar o projeto do seqüenciamento.

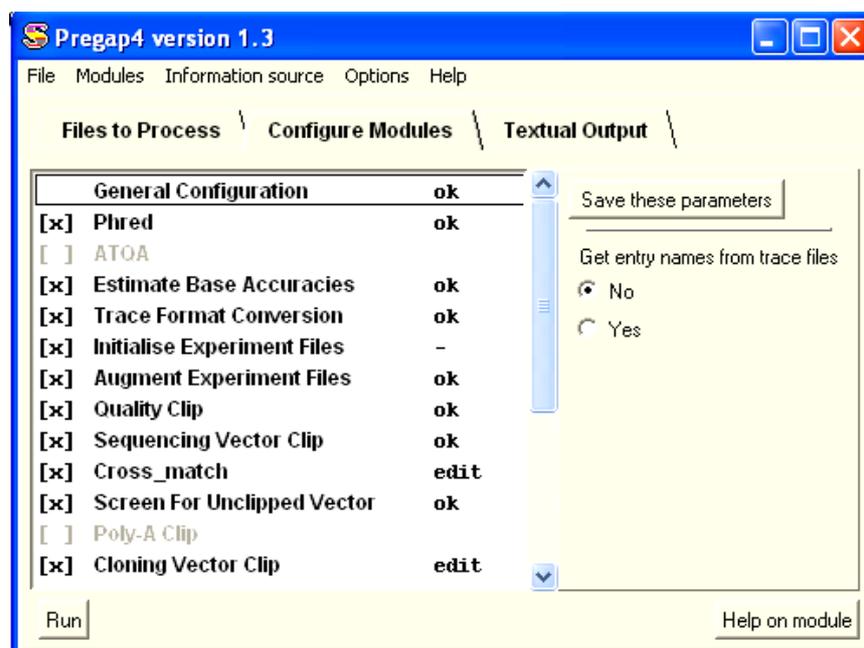


Figura 12. Pregap 4, interativo e configurável. Programas como Phred (análise de qualidade) e Cross-Match (identifica regiões contaminantes de vetor) trabalham na mesma interface.

Introdução

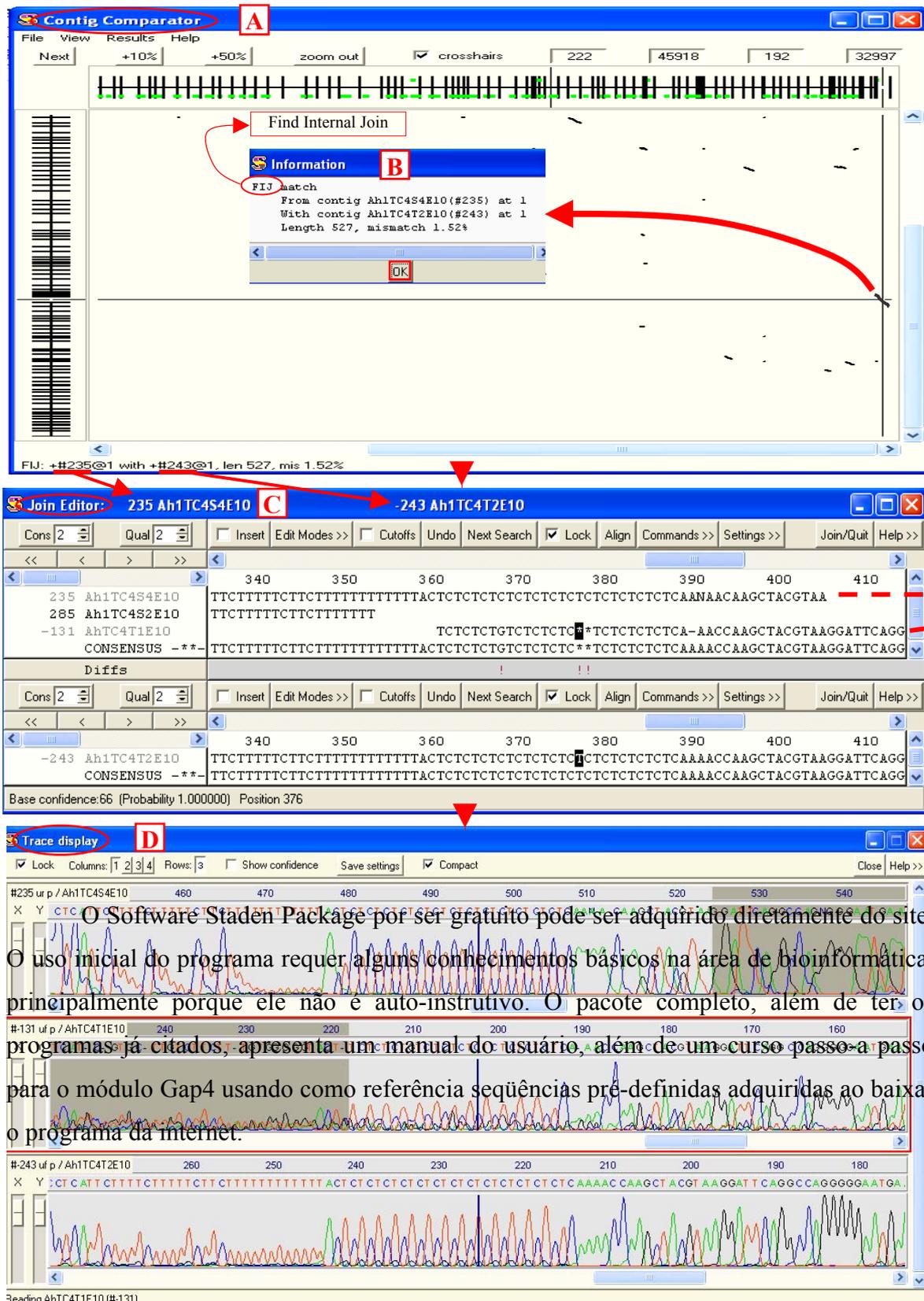


Figura 14. (A) Esta interface permite verificar a redundância das seqüências depositadas através do cruzamento dos dados. O resultado é exposto em forma de gráfico (X;Y), sendo as seqüências consideradas homólogas (total ou parcial) representadas por um traço diagonal. (B) Emite informações sobre a provável homologia selecionada em (A). (C) Sobreposição das seqüências consideradas homólogas. (D) Verificação dos cromatogramas das seqüências.

Introdução

O Software Staden Package por ser gratuito pode ser adquirido diretamente do site. O uso inicial do programa requer alguns conhecimentos básicos na área de bioinformática, principalmente porque ele não é auto-instrutivo. O pacote completo, além de ter os programas já citados, apresenta um manual do usuário, além de um curso passo-a passo para o módulo Gap4 usando como referência seqüências pré-definidas adquiridas ao baixar o programa da internet.

Justificativa e Objetivo

2 - Justificativa

Este trabalho é parte de um projeto geral de busca de genes de resistência e produção de um mapa genético em *Arachis*, que envolve a Universidade Católica de Brasília (UCB); Embrapa Recursos Genéticos e Biotecnologia (CENARGEN); Instituto de Botânica Del Nordeste (Argentina); John Innes Centre (Inglaterra); Universidade de Aarhus (Dinamarca) e Unesp-Botucatu. Para deste trabalho foi proposto o desenvolvimento de marcadores microssatélites a partir de uma biblioteca genômica enriquecida para o dinucleotídeos AG/TC do genoma de *Arachis hypogaea*, espécie cultivada caracterizada por sua susceptibilidade a uma grande variedade de pragas e doenças. Após a conclusão de todas as etapas do trabalho, os primers marcadores microssatélites contribuirão para o desenvolvimento de um mapa genético informativo, onde genes de interesse poderão ser identificados e mapeados.

O trabalho Desenvolvimento de Marcadores Microssatélites (SSR) em *Arachis hypogaea* foi financiado pela Universidade Católica de Brasília, Comunidade Européia e por uma parceria entre a Embrapa e o Banco Mundial (Prodetab).

3 - Objetivo

- O presente trabalho visa o desenvolvimento de marcadores microssatélites a partir de bibliotecas genômicas enriquecidas para dinucleotídeos AG/TC em *Arachis hypogaea* para futuro mapeamento genético desta espécie.

Materials

4 - Materiais

4.1 - Produtos Químicos, Enzimas e Kits.

Produtos	Fornecedor
Agar	Merck, Darmstadt (Germany)
Agarose	Gibco BRL, Gaithersburg (USA)
Álcool Isoamílico	Merck, Darmstadt (Germany)
Ampicilina	USB, Ohio (USA)
Azul de Bromofenol	Gibco BRL, Gaithersburg (USA)
Biotina 16ddUTP	Gibco BRL, Gaithersburg (USA)
Brometo de Etídio	USB, Ohio (USA)
BSA V (<i>Bovine Serum Albumin Fraction V</i>)	Sigma
β -Mercaptoethanol	Merck, Darmstadt (Germany)
Contas Magnéticas - Dynabeads [®] Streptavidin	Dynal, Oslo, Norway
Clorofórmio	Merck, Darmstadt (Germany)
CTAB	USB, Ohio (USA)
DNA-Oligonucleotídeos	Gibco BRL, Gaithersburg (USA)
dNTPs	USB, Ohio (USA)
DYEnamic ET Terminator Cycle Sequencing Kit	Amersham Pharmacia Biotech (UK)
EDTA	USB, Ohio (USA)
Enzimas de Restrição:	
<i>Mse</i> I	Amersham Pharmacia Biotech (UK)
Sau 3AI	Amersham Pharmacia Biotech (UK)
<i>Tsp</i> 509 I	New England, Biolabs
Ethanol	Merck, Darmstadt (Germany)
Extrato de levedura	Merck, Darmstadt (Germany)
Ficoll Tp 400	Gibco BRL, Gaithersburg (USA)
Glicerol	Gibco BRL, Gaithersburg (USA)
IPTG	USB, Ohio (USA)
Isopropanol	Merck, Darmstadt (Germany)
kit de hibridização–RPN 3540/RPN 3541 <i>Genes Images Randon Primer Labelling Module</i>	Amersham Pharmacia Biotech

Substâncias	Fonte / Referência
kit de detecção-RPN 3510/RPN 3511 <i>Genes Images CDP-Star Detection</i>	Amersham Pharmacia Biotech
Meio LB	USB, Ohio (USA)
Membrana de náilon -Hybond N ⁺	USB, Ohio (USA)
Padrão para gel de eletroforese:	
1 Kb Plus	USB, Ohio (USA)
100 bp <i>Ladder</i>	Gibco BRL, Gaithersburg (USA)
DNA Fago λ	Invitrogen
Papel filtro	<i>Watman</i> TM 3MM
Proteinase K	USB, Ohio (USA)
QIAquick Gel Extraction Kit	QIAGEN, Chatsworth (USA)
QIAquick Spin PCR Purification Kit	QIAGEN, Chatsworth (USA)
RNase	USB, Ohio (USA)
SDS	Gibco BRL, Gaithersburg (USA)
T4-DNA-Ligase	Amersham Pharmacia Biotech (UK)
<i>Taq</i> -DNA-Polymerase	Gibco BRL, Gaithersburg (USA)
Terminal Transferase-TdT	Amersham Pharmacia Biotech (UK)
Tetraciclina	USB, Ohio (USA)
Tris	USB, Ohio (USA)
Tryptona	Q-Biogene
TWEEN [®] 20	USB, Ohio (USA)
X-Gal	USB, Ohio (USA)
Xileno Cianol	Gibco BRL, Gaithersburg (USA)

4.2 - Plasmídio, Oligonucleotídeos e Bactérias.

4.2.1 - Vetor de Clonagem

Plasmídio	Tamanho	Qualidades	Fornecedor
pGEM [®] -T	3.000 pb	<i>Cloning PCR Products</i> , Amp ^r , T7/SP6/M13 promoters	Promega (USA)

→ O vetor pGEM-T é a forma commercial do pGem5zfm cortado com EcoRV, com *single Ts* adicionado no 3' ends. (Figura 15)

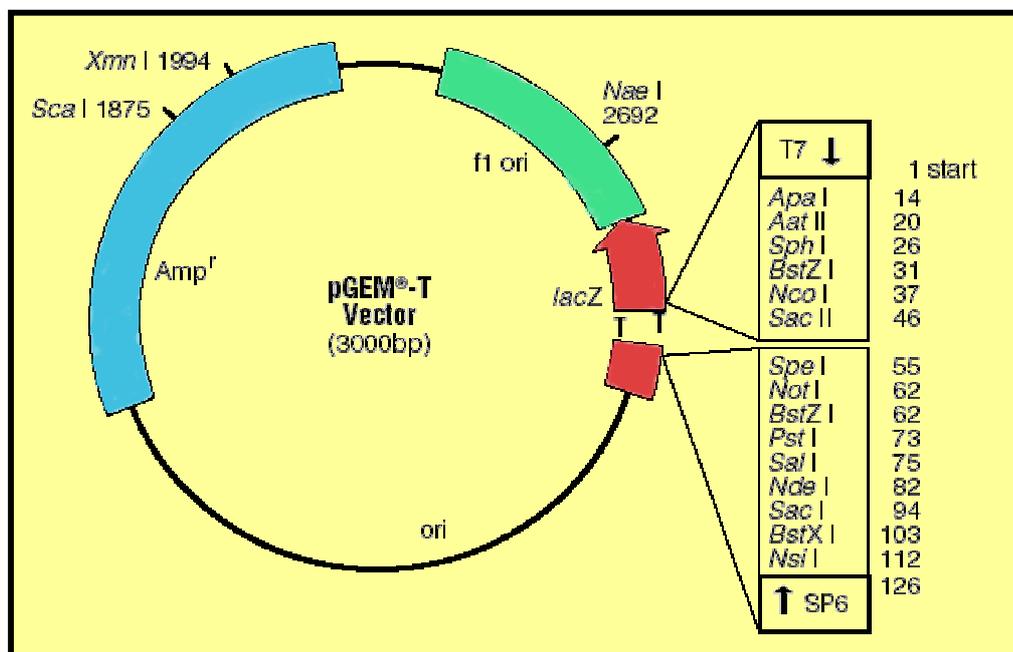


Figura 15. Representação do vetor de clonagem. pGEM-T. Ori = origem de replicação; Lac Z = α -peptídeo - região codificadora da enzima β -galactosidase; T7 e SP6 = promotores da RNA polimerase que flanqueiam uma região múltipla de clonagem; T = local de inserção da molécula de DNA exógena; f1 ori = região f1 fago; Amp^r= gene que confere resistência a ampicilina.

Adaptado:Manual técnico- pGEM-T e pGEM-T Easy vector systems- Promega.

4.2.2 - Oligonucleotídeos

Os oligonucleotídeos foram sintetizados pela Invitrogen (USA). As seqüências de oligonucleotídeos utilizadas nos protocolos de Construção de Biblioteca Genômica Enriquecida –SSR e Seleção e Seqüenciamento dos Clones Positivos - SSR estão descritas na tabela 8.

Tabela 8. Sequências de oligonucleotídeos usados nos protocolos de construção e seqüenciamento de bibliotecas genômicas.

Oligonucleotídeos	Mer	Seqüências
SA <i>Sau</i>	21	5' CAG CCT AGA GCC GAA TTC ACC 3'
LA <i>Sau</i>	25	5' pGAT CGG TGA ATT CGG CTC TAG GCT G 3'
AG ₁₃	26	5' AGA GAG AGA GAG AGA GAG AGA GAG AG 3'
T7 promoter/primer	23	5' TAA TAC GAC TCA CTA TAG GGA GA3'
Sp6 promoter/primer	24	5' CAT ACG ATT TAG GTG ACA CTA TAG 3'

4.2.3 - Isolado Bacteriano

Escherichia coli

Linhagem

XL1-Blue *supE44 hsdR17 rec A1 end A1 gyrA46 thi relA1 lac [F' proAB⁺ lac^q lacZ ΔM15 Tn10 (tet^r)]* (Stratagene, UK)

4.2.3.1 - Meios de Cultura

- **Meio Luria Bertani (LB) -Tabela 9**

Tabela 9. Composição do meio LB para crescimento de bactéria.

Componentes	Concentrações
Triptona	1,0 %
Extrato de levedura	0,5 %
NaCl	1,0 %

Fonte: Miller, 1972.

O pH do meio foi ajustado para 7,5 com 0,1N NaOH e autoclavado por 20 minutos a 121°C. Para o meio sólido, foi acrescentado 1,8% agar (p/v) ao meio LB líquido antes da autoclavagem.

- **Meio SOC e Meio SOB – Tabela 10**

Tabela 10. Composição do meio SOC para crescimento de bactéria.

Componentes	Concentrações
Triptona	2,0 %
Extrato de levedura	0,5 %
NaCl	10,0 mM
KCl	2,5 mM
MgCl ₂	10,0 mM
MgSO ₄	10,0 mM
Glicose	20,0 mM

Fonte: Hanahan, 1983.

Os três primeiros componentes foram dissolvidos e o pH ajustado para 7.0 . O meio foi autoclavado durante 20 minutos a 121°C, e no momento de sua utilização, foram adicionadas a ele as demais soluções previamente filtradas.

O meio SOB é idêntico ao meio SOC, exceto pela ausência da glicose em sua composição.

- **Meio *Terrific* (TB) – Tabela 11**

Tabela 11. Composição do meio TB para crescimento de bactéria.

Componentes	Concentrações
Triptona	1,2%
Extrato de levedura	2,4%
Glicerol	0,4%
KH ₂ PO ₄	0,17 mM
K ₂ HPO ₄	0,17 mM

Fonte: Tartof & Hobbs, 1987

Os três primeiros componentes foram dissolvidos e autoclavados durante 20 minutos a 121°C, sendo as demais soluções filtradas adicionadas no momento do uso.

4.2.3.2- Antibiótico- Meio LB e LB-agar.

Ampicilina	10,0 µgmL ⁻¹
Tetraciclina	12,5 µgmL ⁻¹

Os antibióticos foram adicionados ao meio após ser autoclavado e resfriado entre 55 e 60°C (Sambrook *et. al.*, 2001).

4.3 - Condições de Cultivo para Bactéria.

4.3.1 - *E. coli* XL1-Blue.

As bactérias foram cultivadas em meio líquido SOB (2.3.2), para a etapa de preparação de bactérias eletrocompetentes (2.3.3.2), contendo o antibiótico tetraciclina, com incubação a 37°C sob agitação constante de 250 rpm por 16 horas.

Para o crescimento bacteriano, as estirpes de *E. coli* foram cultivadas em meio LB líquido ou sólido, contendo ampicilina a 37°C com ou sem agitação respectivamente, por cerca de 16 horas.

4.4 - Soluções

- 2X CTAB

2,0% CTAB; 1,4M NaCl; 20mM EDTA; 100mM Tris-Cl pH 8,0; 1,0% PVP

- Proteinase-K

2,0% Proteinase K

- Tampão de Extração de DNA - 2X CTAB

0,2% β-mercaptoetanol; 0,05% Proteinase-K em 2X CTAB.

- 2M NaCl

11,69% NaCl

- Tampão TE, pH 8.0

10mM Tris-HCl, 1mM EDTA

- Solução de ressuspensão do DNA (3.1)

Tampão TE, pH 8.0 com 10µgmL⁻¹ de RNase

- 10X TBE

10,8% Tris Base; 5,5% Ácido Bórico; 0,01M EDTA pH 8.0

- Tampão de carregamento 6X

0,25% Azul de Bromofenol, 0,25% Xileno Cianol, 30% Ficoll Tp 400

- 5M EDTA pH 8.0
18,61% EDTA , ajustar pH
- Tampão NET (2.4.3.1)
0,15M NaCl; 0,1mM EDTA; 20mM Tris-Cl pH 8.0
- Tampão HS-NET (2.4.3.1)
1M NaCl, 0,1mM EDTA, 20mM Tris-CL pH 8.0
- B&W
10mM Tris-Cl pH 7.5; 1mM EDTA, 2mM NaCl
- 20X SSPE
3M NaCl; 0,2M NaH₂PO₄.H₂O; 0,02M EDTA; ajustar pH 7.4
- Solução Desnaturante
0,5M NaOH; 1,5M NaCl
- Solução Neutralizante
0,5M Tris-Cl pH 8.0; 1,5M NaCl
- 20X SSC
3M NaCl; 0,3M Citrato de Sódio; ajustar pH 7.0
- Solução de Hibridização
5X SSC; 0,1% SDS; 5% Dextran Sulfato; 1:20 *liquid Blocking* (Kit de Hibridização, provido por *Genes Images Random Primer Labeling Module*, Amersham Pharmacia)
- Solução de Lavagem 1
1X SSC; 0,1% SDS
- Solução de Lavagem 2
0,5X SSC; 0,1% SDS

- Tampão A

100mM Tris-Cl; 300mM NaCl; 1:10 *liquid blocking* (Kit de Detecção, provido por *Genes Images CDP-Star Detection Module*, Amersham Pharmacia)

- Solução de Lavagem 3

0,3% TweenTM20 em Tampão A

- GET

0,92% glicose (filtrada); 10 mM EDTA pH 8.0; 26 mM Tris-Cl pH 7.4

- Solução 2 (Extração do DNA plasmidial)

0,2M NaOH; 1% SDS

Metodologias

5 - Metodologia

5.1 - Extração e Quantificação do DNA Genômico Total de *Arachis hypogaea*.

Para a extração do DNA total, foi utilizado tecido foliar *A. hypogaea* do banco de germoplasma da Embrapa-Recursos Genéticos e Biotecnologia (CENARGEN). A extração foi feita de acordo com o protocolo de Doyle & Doyle (1987), utilizando 75 mg de amostra por tubo (1,5 mL) para a extração do DNA genômico total. Para o rompimento das paredes e membranas celulares do tecido, foi utilizado o nitrogênio líquido e um bastão de vidro pontiagudo. Ao tecido macerado foram adicionados 700 µL de Tampão de Extração de DNA previamente aquecido a 65°C. Após a completa ressuspensão das amostras na solução, iniciou-se a etapa da incubação por 1 hora a 65°C em banho-maria, com leve agitação a cada 10 minutos. As amostras foram então levadas para a capela de exaustão para a retirada da fase orgânica das amostras. Dois sucessivos processos de extração orgânica foram realizados com a adição de 600 µL de CIA (clorofórmio-álcool isoamílico 24:1). Os tubos foram agitados durante 5 minutos e centrifugados a 12.000 rpm por 10 minutos. Em seguida, os sobrenadantes foram transferidos para tubos novos. Os ácidos nucleicos foram precipitados com a adição de 500 µL de isopropanol absoluto e incubado a -20°C por 1 hora. Os tubos foram centrifugados a 12.000 rpm por 8 minutos e o sobrenadante foi descartado. O pellet foi lavado duas vezes com etanol 70% (± 1 mL) por 5 minutos, descartando-se o sobrenadante. Em seguida foram adicionados 500 µL de 2 M NaCl e incubados a 65°C por 1 hora. Após a completa ressuspensão do *pellet*, os tubos foram gentilmente agitados e incubados a 4°C por 45 minutos. As amostras foram centrifugadas a 2.000 rpm por 10 minutos e os sobrenadantes foram transferidos para novos tubos. Foram acrescentados ao sobrenadante 350 µL de isopropanol absoluto, precipitando o DNA novamente. Os tubos foram incubados a -20°C por 1 hora e centrifugados a 12.000 rpm por 8 minutos. Foram feitas duas lavagens com etanol 70% por 10 minutos e uma lavagem com etanol absoluto (gelado) por 3 minutos. Após o descarte dos sobrenadantes, os *pellets* foram colocados para secar no fluxo laminar por duas horas e então ressuspensos em 30 µL de solução de ressuspensão do DNA. A quantificação do

DNA extraído foi feita em gel de agarose 1% corado com brometo de etídio (2.4.3) sendo utilizado o DNA de fago *lambda* de concentração conhecida como padrão de comparação.

5.2 - Seleção da Enzima para Digestão Total do DNA Genômico.

Para a seleção da enzima de restrição, 250 ng de DNA foram digeridas por 12 horas com três enzimas diferentes: *Mse* I, *Sau* 3AI e *Tsp* 509 I, de acordo com o protocolo de cada fabricante. Os fragmentos da digestão foram visualizados por eletroforese em gel de agarose 1,0 % corado com brometo de etídio ($1,5 \mu\text{g.mL}^{-1}$).

Para a digestão total, foram utilizados 30 μg de DNA genômico; 7 unidades da enzima *Sau* 3AI e 1X *buffer* H (10 mM Tris-HCl pH 7.5; 7 mM MgCl_2 e 100 mM NaCl), sendo em seguida a solução incubada por 12 horas a 37°C.

5.3 - Gel de Agarose para Eletroforese Horizontal.

A visualização dos fragmentos de DNA foram feitos em eletroforese com gel de agarose (1 e 1,5%).

Tabela 12. Composição do gel de agarose para eletroforese horizontal.

Componentes	Concentrações
Agarose	1% ou 1,5%
TBE	1X
Brometo de etídio	$1,5 \mu\text{g.mL}^{-1}$

Foi adicionado a agarose (1% ou 1,5%) em 1X TBE e fundidos em microondas. Deixou-se esfriar até aproximadamente 60°C, adicionando-se em seguida o brometo de etídio. Foi vertido ainda quente em uma cuba de mini-gel (Gibco BRL, Gaithersburg, USA) com o pente posicionado e deixado solidificar por aproximadamente 20 minutos, à temperatura ambiente. Após este período, as amostras de DNA com tampão de carregamento foram aplicados. Foi aplicado no mesmo gel um marcador de peso molecular (1 Kb Plus - ou 100 ng λ fago), em seguida foi feito o DNA migrar por eletroforese. O DNA foi visualizado com trans-iluminador com luz ultravioleta/366 nm.

5.3.1 - Eluição dos Fragmentos de DNA em Gel de Agarose.

O DNA digerido foi aplicado no gel de agarose 1,5%, para que os fragmentos entre 200 a 950 pb fossem recuperados em uma membrana de nitrocelulose (DEAE–Celulose NA-45) previamente cortada (5,0 cm x 1,0 cm) e hidratada: 1ª etapa foi a de submersão da membrana já cortada em uma solução 10 mM EDTA, pH 7.6 por 10 minutos e em seguida, em uma solução 0,5 M EDTA por mais 10 minutos. Na 2ª etapa, a membrana foi lavada oito vezes com H₂O_{bidest} estéril e estocada até o momento do seu uso em H₂O_{bidest} estéril a 4°C. Após estes procedimentos, a eletroforese foi iniciada nas seguintes condições: amperagem 50 mA e voltagem 70 KV. Após 25 minutos, a corrida foi interrompida e o gel rapidamente visualizado no trans-iluminador. Um corte no gel na amplitude dos fragmentos desejados (200 e 950 pb) foi feito com uma lâmina de bisturi, a membrana hidratada foi ajustada no corte do gel e a eletroforese reiniciada. Após 30 minutos de eletroforese, verificou-se que os fragmentos desejados haviam sido absorvidos pela membrana, iniciando a etapa de eluição do DNA.

A membrana foi retirada do gel e lavada com o tampão NET, sendo em seguida cortada em pedaços pequenos e acondicionada em um tubo (1,5mL). Foi adicionado ao tubo 500µL de tampão HS-NET, sendo então incubado a 65°C por 1 hora. Em seguida, a solução foi recuperada em um novo tubo (1,5 mL) e adicionado 1,5 mL de etanol absoluto para precipitar o DNA a -20°C por 12 horas. O tubo foi centrifugado a 12.000 rpm por 15 minutos, o sobrenadante descartado e o *pellet* lavado com etanol 70% e em seguida deixado secar por 30 minutos no fluxo laminar. O *pellet* foi ressuspensão em 30µL de água e em seguida quantificado (Figura 16).

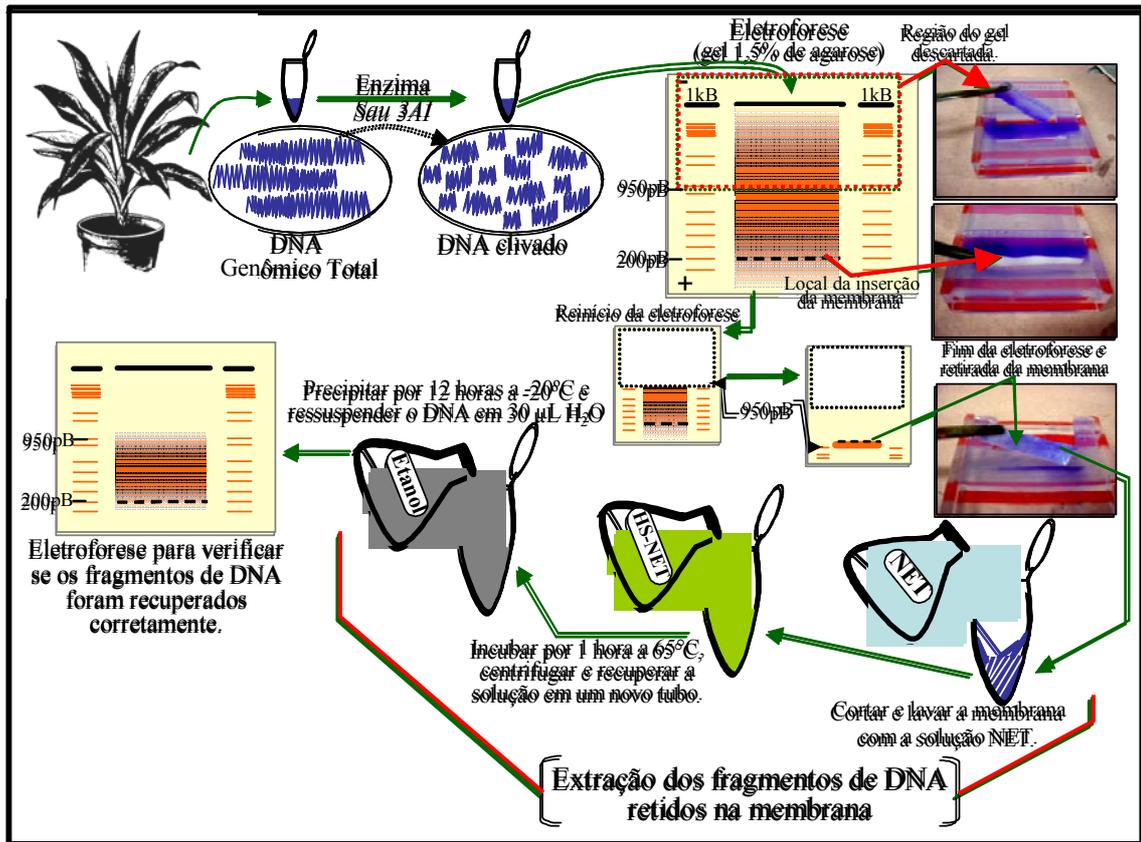


Figura 16. Captura e eluição do DNA, clivado por enzimas de restrição, em membrana de nitrocelulose.

5.4 - Construção de Biblioteca Genômica Enriquecida/ Ligação aos Adaptadores.

A biblioteca genômica foi construída seguindo os protocolos de Rafalski *et al.* (1996); Taramino & Tingey, (1996); White & Powell, (1997); Brondani *et al.*, (1998).

Dos 30 μg do DNA digeridos aplicados no gel de agarose, cerca de 16,6% (5 μg) continham o tamanho dos fragmentos desejados (200 a 950 pb) e foram recuperados pela membrana. Estes fragmentos foram então precipitados com a adição de 50 μL de 3 M acetato de sódio e 950 μL de etanol absoluto para cada 500 μL de amostra, em seguida lavados com etanol 70% e ressuspensos 50 μL de $\text{H}_2\text{O}_{\text{bidest}}$. Os fragmentos foram ligados com T4 DNA ligase (12°C, 16 horas) aos adaptadores (previamente preparados: 80 μL SA *Sau*-500 μM ; 80 μL LA *Sau*-500 μM ; 40 μL TE; 0,1 M NaCl, aquecidos por 15 minutos a 65°C, resfriados por 2 horas em temperatura ambiente e estocados a 4°C.) com extremidades coesivas (CATG) às geradas pela enzima *Sau 3AI* (os adaptadores consistiam de SA *Sau* 21 mer: 5'-CAG CCT AGA GCC GAA TTC ACC 3' e LA *Sau* 25

mer: 5'-pGAT CGG TGA ATT CGG CTC TAG GCT G 3'). Na reação de marcação da sonda de oligonucleotídeos (AG)₁₃ com biotina 16ddUTP, foi utilizado 200 μM de sonda; 1X de Tampão Terminal Transferase (20 mM Tris-Acetato, 50 mM acetato de potássio pH 7.9); 96 unidades de Terminal Transferase-TdT (24 UμL⁻¹); 2 μL de biotina 16ddUTP; 28 μL de H₂O_{bidest} estéril em 40 μL de volume total. A reação foi incubada a 37°C por 30 minutos, posteriormente neutralizada com 4 μL de 0,5 M EDTA, e precipitada com etanol absoluto por 16 horas a -20°C. A reação foi então centrifugada por 30 minutos a 12.000 rpm, lavada uma vez com etanol 70% e ressuspensa em 30 μL de H₂O_{bidest}. Os 200 μM de sonda biotinizada (AG₁₃) da reação, foram ligados a 1 mg de contas magnéticas (*magnetic beads*)/streptavidina em 400 μL de 2X B&W à temperatura ambiente, sob agitação por 1 hora. O excesso de sonda foi removido pela lavagem com 400 μL de 1 X B&W e 400 μL de 5X SSPE com 0,1% SDS. O DNA ligado foi hibridizado ao complexo sonda-conta magnéticas à 65°C por 90 minutos, com volume total de 300 μL (150 μL de 10X SSPE; 0,2% SDS e 150 μL DNA/Adaptador-desnaturado previamente a 95°C), como controle positivo da hibridização, foi utilizada uma diluição de 1:1000 do DNA/adaptador a 65°C por 90 minutos. As contas magnéticas foram sedimentadas por aplicação de um campo magnético e lavadas duas vezes por 5 minutos em 2X SSPE; 0,1% SDS a temperatura ambiente, uma vez por 15 minutos em 2 X SSPE; 0,1% SDS a 65°C. As contas magnéticas foram ressuspensas em 200 μL de H₂O_{bidest} estéril, todas as 5 soluções de lavagens das contas magnéticas após a hibridização foram guardadas para verificação do controle de enriquecimento via PCR (Figura 17).

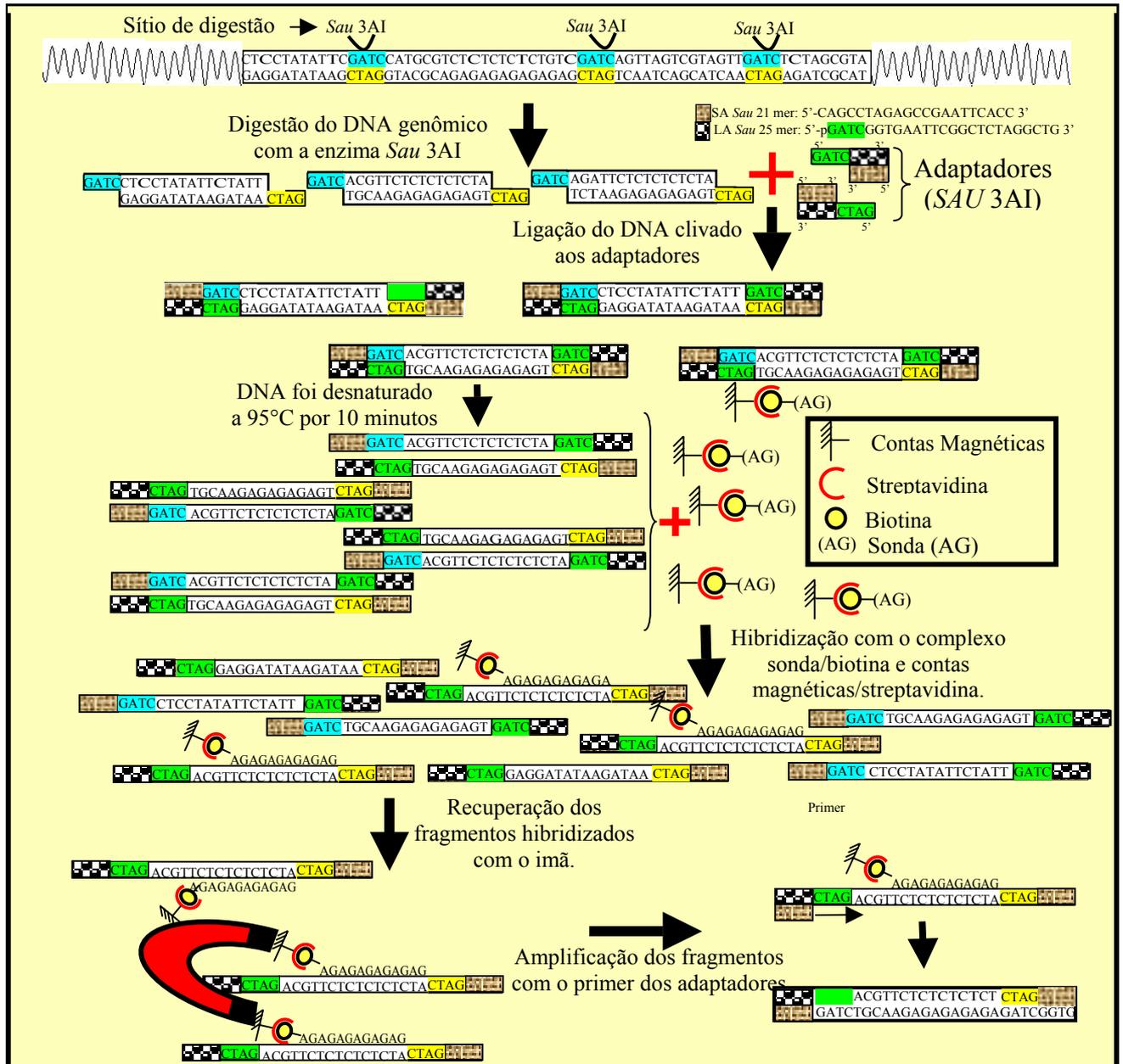


Figura 17. Protocolo de enriquecimento da biblioteca genômica. A primeira linha representa o DNA genômico contendo muitos sítios de clivagem da enzima *Sau 3AI*. **GATC** e **CTAG** representam as extremidades coesivas geradas pela enzima *Sau 3AI* no DNA genômico. As repetições (TC/AG)_n presentes nas seqüências são representações de microssatélites.

5.4.1 - Controle de Enriquecimento da Biblioteca Genômica por PCR.

O controle de enriquecimento da sonda (AG)₁₃ via PCR foi feito com 7 alíquotas diferentes: 1-DNA controle (1:1000); 2-Solução de hibridização; 3-Solução da 1ª lavagem; 4-Solução de 2ª lavagem; 5-Solução da 3ª lavagem; 6-Solução da 4ª lavagem e 7-Solução com as contas magnéticas. Para a reação de PCR foi utilizado 1X Tampão de PCR (10 mM

Tris-Cl pH 8,3; 5 mM KCl); 2,5 mM MgCl₂; 10 μM primer *SA Sau* (adaptador, primer complementar ao sítio alvo da enzima de restrição *Sau* 3AI); 2 mM dNTP; 1,0 unidade da enzima *Taq* polimerase e 2 μL do DNA em 25 μL de reação final. A amplificação dos fragmentos por PCR foi realizada utilizando o seguinte programa: 95°C por 3 minutos, 94°C por 2 minutos, 25 ciclos com etapas de desnaturação (45 segundos a 94°C), ligação (45 segundos a 56°C) e extensão (2 minutos a 72°C); e uma etapa final de extensão de 7 minutos a 72°C. Os produtos de PCR foram visualizados em gel de agarose 2% corado com brometo de etídio.

5.4.2 - Análise da hibridização do DNA/Sonda pela Técnica *Southern Blot*.

A recuperação preferencial dos fragmentos contendo regiões SSR foi confirmada por *Southern Blot*. A técnica iniciou-se com a eletroforese da reação de PCR das 7 alíquotas descritas no item anterior em gel de agarose 2% (2.4.4) sendo desnaturado (solução desnaturante) por 30 minutos e neutralizado (solução neutralizante) por mais 30 minutos e em seguida lavado com água destilada. A montagem do *Southern transfer* iniciou-se com a colocação do gel sobre o papel filtro embebido com a solução 10X SSPE, sobre este uma membrana de náilon e em cima da membrana várias camadas de papel filtro, para que por capilaridade ocorresse transferência dos fragmentos de DNA presentes no gel para a membrana (Sambrook *et al.*, 2001). Após 16 horas de transferência, os fragmentos de DNA foram fixados à membrana por luz ultravioleta, utilizando o aparelho cross-linker. Na etapa seguinte, o DNA fixado à membrana foi hibridizado com a solução de hibridização contendo 2 ng.μL⁻¹ da sonda de oligonucleotídeos (AG)_n marcada com biotina a 60°C por 16 horas. Após a hibridização foram feitas 2 lavagens (solução de lavagem “1” a 60° C por 15 minutos e solução de lavagem “2” a 60° C por 15 minutos) para remoção de moléculas de sondas ligadas inespecificamente à membrana e ao DNA. A detecção da sonda marcada hibridizada com a seqüência de interesse no DNA iniciou-se com a incubação da membrana no tampão “A” por 1 hora em temperatura ambiente com suave agitação. Em seguida, iniciou-se a incubação com uma solução contendo 4 μL anticorpo *anti-florescein AP* (kit de detecção), 20 μL tampão A e 0,10 g BSA V (Albumina de Soro Bovino, fração V) por 1 hora em temperatura ambiente e com suave agitação. Após este período, a membrana foi lavada 3 vezes por 10 minutos com a solução de lavagem “3” em temperatura ambiente com suave agitação sendo por fim, o excesso de

tampão drenado da membrana. O reagente de detecção foi então colocado sobre a membrana ($30 \mu\text{L}\cdot\text{cm}^{-2}$), deixando-o agir por 5 minutos. O excesso da solução foi retirado e a membrana ainda ligeiramente úmida foi envolvida com o plástico *Saran Wrap* e exposta ao filme de raios-X por 1 hora. Após esse período, o filme de raios-X foi revelado (solução de revelação, Kodak) gerando uma auto-radiografia que mostrou o pareamento (específico) DNA-Sonda $(AG)_n$, confirmando a presença de microssatélites dinucleotídeos $(TC)_n$.

5.5 - Seleção de Clones Positivos para SSR.

5.5.1 - Ligação do Inseto ao Plasmídio.

Após a confirmação da presença de SSR, o DNA ligado às contas magnéticas foi amplificado em 25 alíquotas por PCR utilizando-se 1X de Tampão de PCR (10 mM Tris-HCl pH 8.3; 5 mM KCl); 2,5 mM MgCl_2 ; 10 μM primer SA *Sau*; 2 mM dNTP; 1,0 unidade da enzima Taq polimerase e 2 μL do DNA em 25 μL de reação final por alíquota. A PCR foi realizada utilizando o seguinte programa: 95°C por 3 minutos, 94°C por 2 minutos; 25 ciclos com etapas de desnaturação (45 segundos a 94°C), ligação (45 segundos a 56°C) e extensão (2 minutos a 72°C); etapa de final extensão por 7 minutos a 72°C. O produto amplificado foi purificado utilizando-se o QIAquick PCR *Purification* (Qiagen®) e quantificado em gel de agarose 2% corado em brometo de etídio sendo utilizado um DNA de concentração conhecida como padrão de comparação.

Na ligação do inserto ao plasmídio pGEM-t, foi utilizado 15 ng de DNA genômico enriquecido para microssatélites; 50 ng de plasmídio; 1X *Rapid Ligation Buffer*-T4 DNA ligase; 3,0 unidades de T4 ligase para 10 μL de volume final, a reação foi incubada por 16 horas a 4°C. Após a incubação, a reação foi precipitada com a adição de $\text{H}_2\text{O}_{\text{bidest}}$ e estéril, 10 μL de 7,5 M de acetado de amônio e 60 μL de etanol absoluto, a solução foi centrifugada a 14.000 rpm por 20 minutos e o sobrenadante descartado. Em seguida a ligação foi então lavada com etanol 70% e ressuspensa em 5 μL de $\text{H}_2\text{O}_{\text{bidest}}$ estéril.

5.5.2 - Preparação de Bactérias Eletrocompetentes.

Uma única colônia de *E.coli*-XL1Blue foi pré-inoculada em 4 mL de meio SOB e sob agitação constante de 250 rpm, por 16 horas, a 37°C. Após o período de crescimento,

os 4 mL (pré-inóculo) de cultura foram inoculados em 400 mL de meio SOB e deixados crescer a 37°C e 250 rpm até atingir $O.D._{600} = 0.2 - 0.25$. Em seguida, as células bacterianas passaram por 5 sucessivas lavagens com glicerol 10% gelado estéril, e em seguida centrifugados a 5000 rpm, por 20 minutos a 4°C. Por fim, o pellet foi ressuspensão em 500 µL de glicerol 10% gelado e estéril, sendo então dividido em alíquotas de 40 µL por tubo (1,5 mL), congeladas imediatamente em nitrogênio líquido e estocadas a -80°C até a hora do uso. A eficiência das células eletrocompetentes foi testada usando o protocolo descrito na etapa de transformação (2.4.6.2), sendo usado como controle um vetor plasmidial conhecido e fechado.

5.5.3 - Transformação por Eletroporação (Chassy and Flickinger, 1987; Chassy *et al.* 1988; Dower *et al.*, 1988).

Para a etapa de transformação foi adicionado 1 µL (15ng.µL⁻¹) de ligação precipitada e 40 µL de células para eletroporação *Escherichia coli* (2.3.3.2) (eficiência de $3,2.10^7$ células em 40 µL de 10% glicerol) em uma cuveta de eletroporação previamente resfriada no gelo. O eletroporador Gene Pulser® (Biorad) foi ajustado às seguintes condições: capacitância - 25 µF; resistência - 200 Ω e 800 Ω; voltagem - 1,80 KV. Imediatamente após a eletroporação, foi adicionado 1 mL de meio SOC na cuveta, transferindo cuidadosamente todo o preparado para um novo tubo (1,5 mL), incubando-o a 37°C, durante 1 hora. Em seguida, o preparado foi centrifugado a 5.000 rpm por 3 minutos, tendo 800 µL do meio descartado. As células foram novamente ressuspensas nos 200 µL de meio restante no tubo e plaqueadas em duas placas de petri com 20 mL meio LB-agar suplementado com 100 mgL⁻¹ de ampicilina, 10 µL de 0,1M IPTG (isopropyl-beta-D-thiogalactopyranoside) e 40 µL de 2% X-Gal (5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside) e incubadas durante 16 horas a 37°C.

5.5.4 - Seleção das Colônias - Branco/Azul

Após as 16 horas de incubação, observou-se a presença de colônias brancas e azuis, sendo as colônias brancas, oriundas da ausência de degradação do substrato cromogênico, selecionadas e novamente replicadas para uma nova placa mapeada com numeração, contendo meio LB-agar com ampicilina, IPTG e X-Gal e incubadas a 37°C por 16 horas.

5.5.5 - Clonagem e Isolamento do DNA Plasmidial por Lise Alcalina.

O procedimento utilizado para a clonagem e isolamento do DNA plasmidial foi inteiramente realizado em placa de 96 poços. As etapas deste procedimento incluíram o crescimento bacteriano em meio de cultura líquido *Terrific*, lise alcalina e purificação com sistema de filtragem Millipore em placas de 96 poços mediante centrifugação.

As colônias que permaneceram brancas na segunda seleção foram replicadas para uma nova placa de petri, mapeada com o formato de placa 96 poços contendo meio LB (agar e ampicilina), e incubadas a 37°C por 15 horas. Depois de crescidas, as bactérias foram inoculadas na placa tipo *deep well* (microplaca) com um único toque do replicador (com 96 dentes) em 1,5 mL de meio de cultura *Terrific*, acrescido de ampicilina, que é seletivo para o plasmídeo pGEM-T, e incubada por cerca de 16 horas a 37 °C sob gitação constante a 250 rpm (Figura 18).

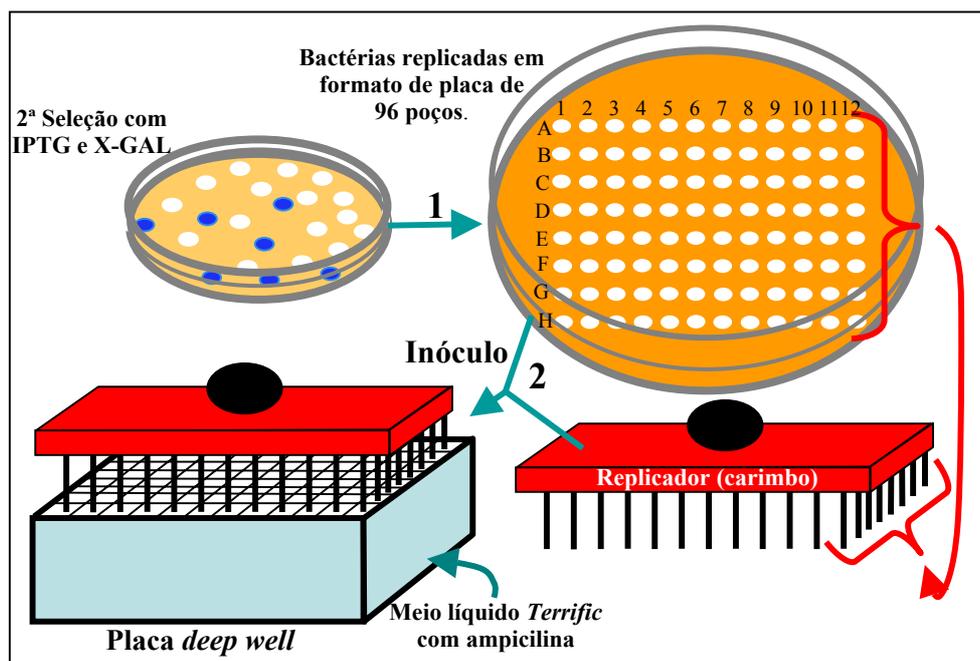


Figura 18. Etapa 1 - Seleção de colônias transformadas e o replique em placa de petri mapeado com o formato de placa de 96 poços. Etapa 2- Inóculo das 96 colônias transformadas com o auxílio de um replicador (carimbo) em uma placa tipo *Deep Well* contendo meio líquido e ampicilina.

Após o crescimento bacteriano, a microplaca foi centrifugada por 6 minutos a 4000 rpm, para sedimentar as células. O sobrenadante foi descartado e a microplaca invertida em papel absorvente por 2 minutos, e em seguida acrescentados 240µL de GET por poço na placa, a qual foi selada, agitada por 2 minutos para homogeneização das soluções e

centrifugada por 9 minutos, a 4000 rpm. O sobrenadante foi descartado, sendo a microplaca invertida em papel absorvente por 2 minutos. Foram acrescentados mais 80 μL de GET por poço, e em seguida a microplaca foi agitada pra homogeneizar a soluções.

A suspensão foi transferida para outra microplaca de 250 μL de polipropileno de fundo redondo (tipo Elisa) já contendo 2,5 μL de 10mgmL^{-1} de RNase por poço. A cada poço foram adicionados 80 μL da solução 2. Com um adesivo novo a microplaca foi selada e invertida 30 vezes para homogeneizar as soluções, sendo em seguida incubada por 10 minutos em temperatura ambiente e centrifugada por alguns segundos para retirar toda a suspensão do adesivo. Foram adicionados à suspensão mais 80 μL de 3M KOAc por poço da microplaca, que foi selada e homogeneizada 30 vezes por inversão. A microplaca ficou incubada em temperatura ambiente por 10 minutos e depois foi centrifugada até 4000 rpm. Sem o adesivo, a microplaca foi incubada a 90°C por 30 minutos. Com a utilização de um novo adesivo a microplaca foi selada e resfriada em gelo por 10 minutos e centrifugada a 20°C por 6 minutos a 4000 rpm.

Com o auxílio de ligas elásticas, uma placa *Millipore* (MAGV N22) foi fixada no topo de outra microplaca de fundo “V” de 250 μL de polipropileno. Todo o volume contido na microplaca de fundo redondo foi transferido para a placa *Millipore*. O aparato foi centrifugado a 20°C a 4000 rpm por 6 minutos. A placa *Millipore* foi removida e ao filtrado resultante na placa fundo “V” acrescentados 100 μL de isopropanol. Com um novo adesivo a placa foi selada e o material homogeneizado por inversão (30 vezes). A placa foi centrifugada a 20°C , a 4000 rpm por 45 minutos. Descartou-se o sobrenadante invertendo a placa, e aos poços, foram adicionados 200 μL de etanol 70% gelado. Centrifugou-se a placa a 20°C , a 4000 rpm por 5 minutos e ao final o sobrenadante foi descartado. Em seguida, a microplaca foi colocada no fluxo laminar por duas horas com a finalidade de secar o DNA plasmidial. O DNA foi ressuspensão em 50 μL de H_2O destilada e autoclavada, permanecendo por 12 horas à temperatura ambiente e em seguida armazenado a -20°C .

A quantificação do DNA plasmidial foi feita em gel de agarose 1% corado com brometo de etídio, sendo utilizado o DNA de fago *lambda* de concentração conhecida como padrão de comparação.

5.6 - Seqüenciamento dos Clones Positivos para SSR.

As reações de seqüenciamento foram feitas utilizando o Kit de seqüenciamento DYEnamic™ ET Terminator Cycle Sequencing para ABI 377, seguindo as seguintes concentrações para a amplificação dos fragmentos por PCR: 0,3µM de primers complementares as seqüências dos vetores (promotores T7 ou SP6); 150 ng de DNA; e 2µL de *Sequencing reagent premix* (marcação fluorescente do Kit de seqüenciamento) em 10 µL de reação final por amostra. As amplificações foram feitas utilizando um termociclador nas seguintes condições: 30 ciclos com etapa de desnaturação (96°C por 20 segundos); ligação (50°C por 15 segundos) e de extensão (60°C por 60 segundos). Os produtos das amplificações foram purificados seguindo o protocolo do Kit de seqüenciamento e seqüenciados na plataforma ABI 377 da *Applied Biosystems*.

5.7 - Construção dos Primers-SSR

As seqüências foram selecionadas e analisadas por programas específicos. O primeiro programa utilizado neste projeto, *Sort chromat for microsats (Script Perl)* desenvolvido pelo Prof. Dr. David Bertoli, tinha o propósito de identificar e selecionar os cromatogramas com seqüências repetidas em tandem (com 1 a 5 nucleotídeos), salvando-os em uma nova pasta denominadas *HITS*. Para as seqüências onde os SSRs não foram encontrados ou ocorreu alguma falha de seqüenciamento, o programa automaticamente enviou os cromatogramas destas seqüências uma outra pasta denominada *NO-HITS*.

O segundo passo foi o de identificar na pasta *HITS* as qualidades das amostras seqüenciadas e analisar os microssatélites presentes. Esta verificação foi feita com o auxílio do programa Phred, que qualifica a amostra fornecendo um valor de confiança para cada base identificada pelo seqüenciador automatizado, e em seguida pelo programa Staden Package (Staden, 1996, <http://www.mrc-lmb.cam.ac.uk/pubseq>) que executou o corte do vetor presente em cada seqüência analisada pelo Phred, além da verificação das sobreposições dos cromatogramas, realizando alinhamentos quando necessários, e montando os *contigs* para a construção das seqüências *consensus*.

O próximo programa utilizado foi o Primer 3 (Rozen & Skaletsky, 1996, http://www-genome.wi.mit.edu/genome_software/other/primer3.html), cuja função foi desenhar os pares de primers complementares às seqüências de DNA que flanquearam os microssatélites presentes na amostra. Foram especificados alguns parâmetros para o desenho dos pares de primers no programa Primer 3 (Tabela 13).

Tabela 13. Parâmetros específicos selecionados para os desenhos dos primers SSR no programa Primer 3.

Classe	Parâmetro		
Tamanho do Primer	mínimo:17mer	ótimo:20mer	máximo:25mer
Temperatura de Ligação	mínimo:57°C	ótimo:60°C	máximo:63°C
Quantidade de GC no Primer	mínimo:30%		máximo:70%
Tamanho do Produto	mínimo:100 pb		máximo:300 pb

Fonte: McCouch et al., 1997

Por último foi realizada uma busca de similaridade das seqüências presentes tanto na pasta *No-Hits* como na pasta *Hits* contra as proteínas putativas de *Arabidopsis thaliana*, feito a partir do programa Blastx (<http://www.ncbi.nlm.nih.gov>).

Resultados e Discussão

6 - Resultados e Discussão

Neste trabalho, a metodologia utilizada foi a da construção da biblioteca genômica enriquecida para a repetição do dinucleotídeo AG/TC. Esta opção foi a escolhida devido a procura de seqüências conhecidas no banco de dados público ter sido ineficiente. Com relação à utilização de pares de primers entre espécies relacionadas, não foi abordada, uma vez que poucos são os primers disponíveis e em sua maioria são para seqüências repetidas de trinucleotídeos, que são menos polimórficos e informativos do que as repetições de dinucleotídeos (Cho *et al.*, 2000). Porém, em estudos recentes de polimorfismo utilizando primers flanqueadores de regiões com repetições de trinucleotídeos e dinucleotídeos em *A. hypogaea* desenvolvidos pelo MSc. Márcio Moretsohn (comunicação pessoal), observou-se que dos 38 pares de primers para repetições de trinucleotídeos, 36 % eram polimórficos, e dos 29 pares de primers para repetições dinucleotídeos, 45 % eram polimórficos.

Uma busca em bancos de dados públicos de DNA revelou que a mais freqüente repetição de dinucleotídeo no genoma de plantas é o AT (Morgante & Olivieri, 1993). Porém, repetições AG foram às escolhidas na construção da biblioteca genômica enriquecida em *A. hypogaea*, pois apesar de apresentarem a segunda maior freqüência de dinucleotídeos nos genomas de plantas até agora estudados (Stallings, 1992; Lagercrantz *et al.*, 1993), não são autocomplementares como as repetições AT, viabilizando o processo da seleção, principalmente nas análises por hibridização. Adicionalmente, as repetições AG também foram as selecionadas por serem mais comumente utilizadas, juntamente com as repetições AC, na construção de mapas genéticos em plantas (Smith *et al.*, 1997; Stepherson *et al.*, 1998).

6.1 - Obtenção e Isolamento de SSRs

O amendoim, em face de sua grande produção de carboidrato, pode apresentar grande quantidade de polissacarídeos no DNA extraído (Ahmed & Young, 1982), sendo necessário antes de iniciar a extração a realização de alguns testes para verificar a qualidade do DNA da amostra a ser trabalhada.

O resultado do teste de extração de DNA considerou o tecido foliar novo o melhor dentre todas as amostras extraídas de *A. hypogaea* (raiz, caule, folha e flor), pois apresentou menor viscosidade do que as demais amostras, por exemplo, impurezas ou polissacarídeos (Figura 19).

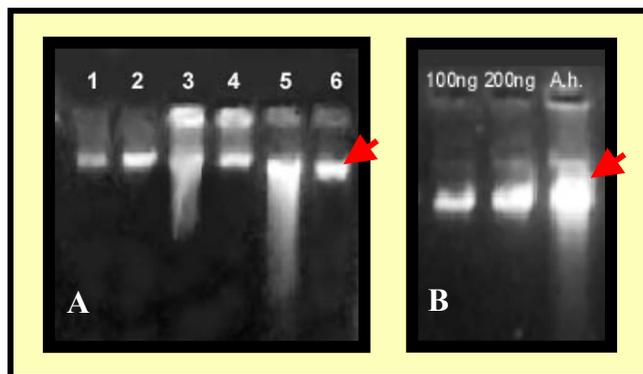


Figura 19. Eletroforeses em gel de agarose 1% corados com brometo de etídio. A) Teste de qualidade de DNA em *Arachis hypogaea* para 4 tecidos diferentes: poços: 1 e 2 - fago λ 100ng e 200ng; 3-caule; 4-raiz, 5-flor e 6-folha. B) Quantificação do DNA genômico total extraído do tecido foliar de indivíduos da espécie *Arachis hypogaea* ; Fago λ 100ng e 200 ng.

Extrações de DNA genômico feitos com kit de extração DNeasy[®] (*Plant Mini Kit*, Qiagen), não demonstraram maior eficiência na eliminação dos polissacarídeos, embora tenha apresentado rendimento superior a 33% de DNA extraído em relação ao método de extração CTAB – Doyle & Doyle (1997) (dados não apresentados). O Kit não foi empregado como método padrão de extração de DNA neste projeto devido ao alto custo de sua utilização quando comparado ao método CTAB.

A digestão de *A. hypogaea* com as três diferentes enzimas de restrição (*Mse* I, *Sau* 3AI e *Tsp* 509 I) revelou que a enzima *Sau* 3AI foi a que produziu o perfil de clivagem do DNA mais adequado para a construção da biblioteca genômica. Dois fatores foram decisivos na escolha desta enzima. O primeiro fator foi a compatibilidade da enzima *Sau* 3AI com os adaptadores já existentes no laboratório, barateando o custo do projeto. O segundo fator considerado foi o tamanho médio dos fragmentos gerados com a clivagem do DNA por esta enzima, sendo a média ótima em torno do comprimento de 500pb. O intuito foi de obter uma larga e variada proporção do genoma de *A. hypogaea* com fragmentos entre 200 e 950 pb, além desta variação de tamanho ser satisfatória para o seqüenciamento inteiro do inserto (Figura 20).

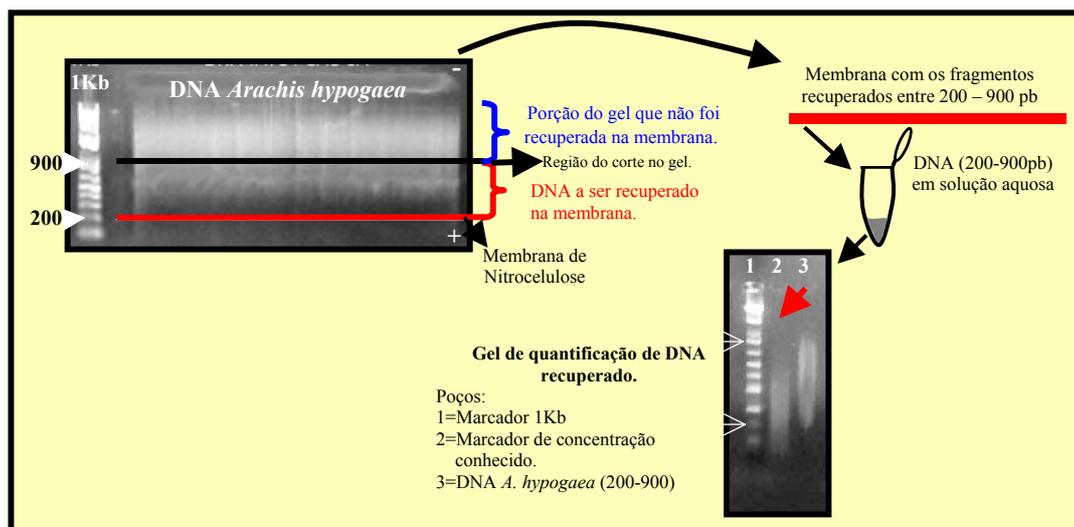


Figura 20: Eluição do DNA no gel de agarose 1% corado com brometo de etídio.

A biblioteca genômica foi enriquecida com a sonda $(AG)_n$, pois repetições de dinucleotídeos TC/AG são comuns em genomas de plantas (Stallings, 1992; Lagercrantz *et al.*, 1993). Após a confirmação da presença de microssatélites na etapa da construção da biblioteca genômica com a técnica *Southern Blotting* (Figura 21), o DNA amplificado foi quantificado, ligado ao plasmídeo pGEM-T, transformado e então plaqueado em placas de petri.

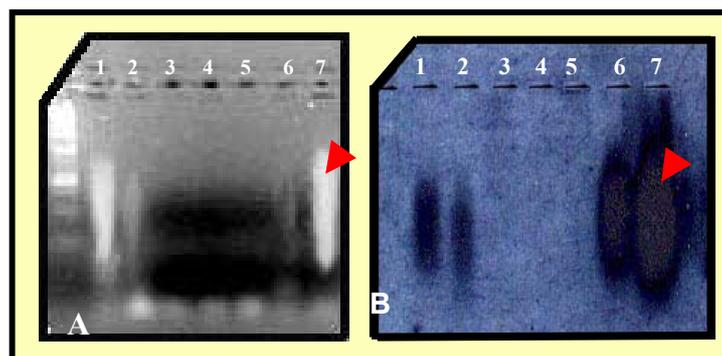


Figura 21. A) Resultado, via eletroforese, da amplificação dos fragmentos da PCR controle de enriquecimento (TC/AG). B) Análise por Southern blotting do filme de raios-X do gel representado em "A". A sonda marcada utilizada foi a $(TC)_{13}$. Poços: 1- Controle positivo DNA (1:1000); 2- Solução de hibridização; 3- Solução 1ª lavagem; 4- Sol. 2ª lav. ; 5- Sol. 3ª lav. ; 6- Sol lav. Final e 7- Solução final (DNA hibridizado).

A confirmação da inserção dos fragmentos de DNA genômico de *Arachis* no interior das bactérias foi feita por duas etapas distintas: A primeira foi baseada na presença do gene de resistência a ampicilina do plasmídeo pGEM-T, cuja função foi degradar o antibiótico presente no meio sólido, preservando a bactéria. E a segunda etapa, baseou-se na seleção pela coloração branco-azul (IPTG/X-Gal) das colônias de bactérias.

Das colônias resistentes ao antibiótico presente na placa de petri, 74% eram brancas e 26% eram colônias azuis. Uma nova seleção com IPTG e X-gal foi realizada nas colônias selecionadas como brancas e o resultado obtido foi de 93 % de colônias brancas contra 7 % de falsos positivos da seleção anterior.

As colônias que permaneceram brancas foram replicadas para uma nova placa no formato de placa de 96 poços (tipo placa de PCR): Cada nova placa continha 96 colônias brancas, dispostas em 12 colunas (numeradas de 01 a 12) por 8 fileiras (classificadas de A a H). Estas placas foram codificadas respeitando alguns critérios, como por exemplo: Ah1TC1A01 - onde respectivamente Ah = *Arachis hypogaea*; 1 = número da biblioteca; TC = tipo de repetição selecionada; 1 = número da placa; A01 = posição da amostra de SSR na placa.

6.2 - Obtenção das Seqüências.

O seqüenciamento foi feito utilizando a quantidade média de 200 ng de DNA plasmideal por amostra, não sendo observado qualquer diferença na qualidade da seqüência quando colocado concentrações entre 100 e 400 ng de DNA plasmideal. As leituras obtidas continham uma certa quantidade de seqüências contaminantes do vetor e algumas delas apresentavam má qualidade, provavelmente devido a alguma contaminação do DNA plasmideal da amostra com o de outra ou mesmo com o DNA bacteriano durante o processo de lise alcalina em placa. Porém não se pode descartar a hipótese de compressões, que compreendem a presença de bases complementares na extremidade final do fragmento amplificado provocando a formação de um *hairpin* (alça em grampo), alterando o resultado final do seqüenciamento. As compressões são mais comuns em seqüências ricas em GC (Watson *et al.*, 1987).

Ao todo foram seqüenciadas seis placas contendo cada uma 96 amostras da biblioteca Ah1 no seqüenciador ABI 377, gerando 576 seqüências, sendo algumas

posteriormente ressequenciadas para confirmação de dados. De todos os 576 cromatogramas analisados no programa seletor de SSR- (*Sort Chromats for Microsats-Script Perl*), em 144 existiam prováveis microssatélites, sendo estes enviados automaticamente para a pasta HITS. Os demais 432 cromatogramas foram depositados na pasta NO-HITS, pois não continham SSR ou apresentavam má qualidade de informação (Figura 22).

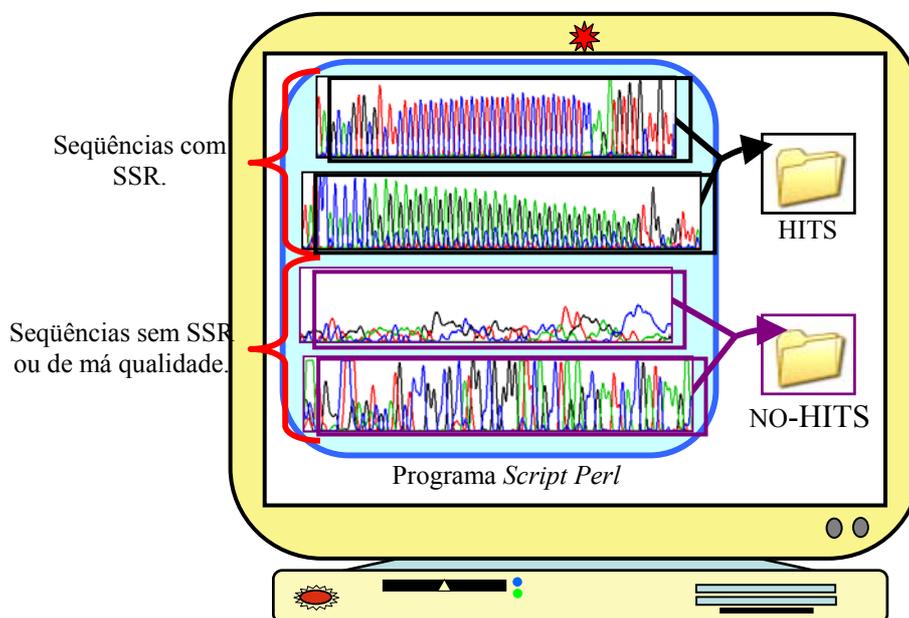


Figura 22. Ilustração representando o programa *Sort Chromats for Microsats* para seleção de SSR desenvolvido pelo Dr. David Bertoli.

6.3 - Teste Comparativo com Hibridização e PCR Ancorado.

No intuito de comparar a eficiência da metodologia utilizada, a placa de número 4 (Ah1TC4), escolhida aleatoriamente entre as seis já sequenciadas, foi submetida ao teste de seleção de clones positivos para SSR por hibridização, sendo também analisada pela metodologia de PCR ancorado (Anexo 1). É importante salientar que de todas as amostras da placa (Ah1TC4) que foram sequenciadas, apenas três não funcionaram, sendo excluídas-as do teste comparativo.

No resultado da seleção por hibridização, dos 93 clones presentes na placa Ah1TC4 (três não funcionaram), 54 hibridizaram com a sonda (AG)_n, como mostrado na autoradiografia da figura 23. Porém quando os dados foram confrontados com o cromatograma das amostras anteriormente seqüenciadas, revelou-se que 48,15% das amostras eram falsos positivos; 51,85% eram positivos verdadeiros; e 1 amostra era falso negativo.

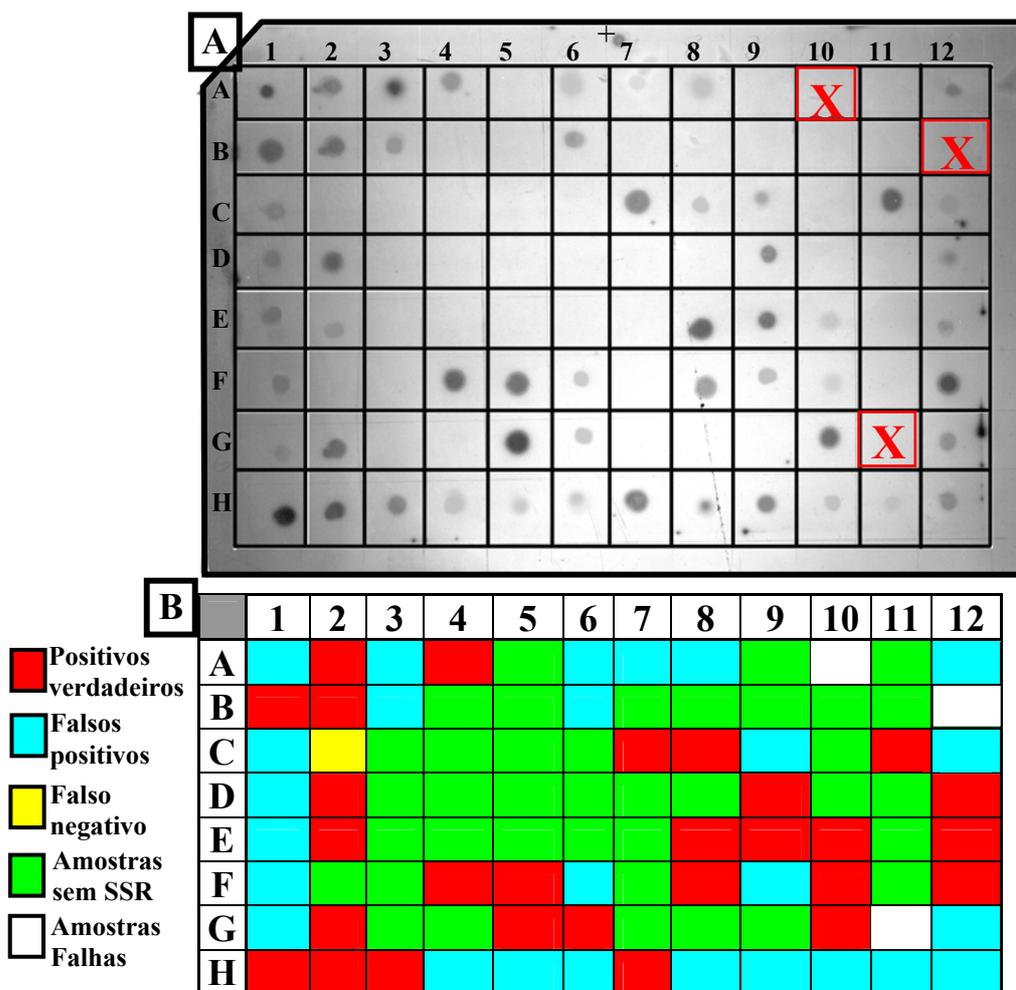


Figura 23. (A) Autoradiografia dos clones da placa Ah1TC4 que hibridizaram com a sonda (AG)_n. (B) Ilustração comparativa dos dados obtidos, confrontando o resultado do seqüenciamento das amostras com a hibridização da sonda (mostrada em "A").

A inexatidão do resultado pode ter sido gerada por uma hibridização inespecífica da sonda (AG) em partes da seqüência ricas nos nucleotídeos C ou T. A contaminação entre os próprios clones foi descartada devido à ausência da etapa inicial de crescimento bacteriano sobre a membrana, como prevê os protocolos de hibridização (Sambrook *et al.*, 2001). Esta etapa foi alterada devido à dificuldade de se trabalhar com a membrana sem

que os clones bacterianos contaminassem-se entre si. Para a obtenção dos resultados, foi adicionado na etapa inicial 0,5 μ L do DNA plasmideal de cada amostra diretamente sobre a membrana, efetuando em seguida o uso do *crosslinker*, evitando com isso a contaminação entre os clones, melhorando significativamente a qualidade dos resultados obtidos.

No PCR ancorado o resultado não foi conclusivo, pois muitas amostras revelaram problemas com amplificações inespecíficas, bem como falhas no processo da PCR, não gerando bandas (Figura 24 A e B).

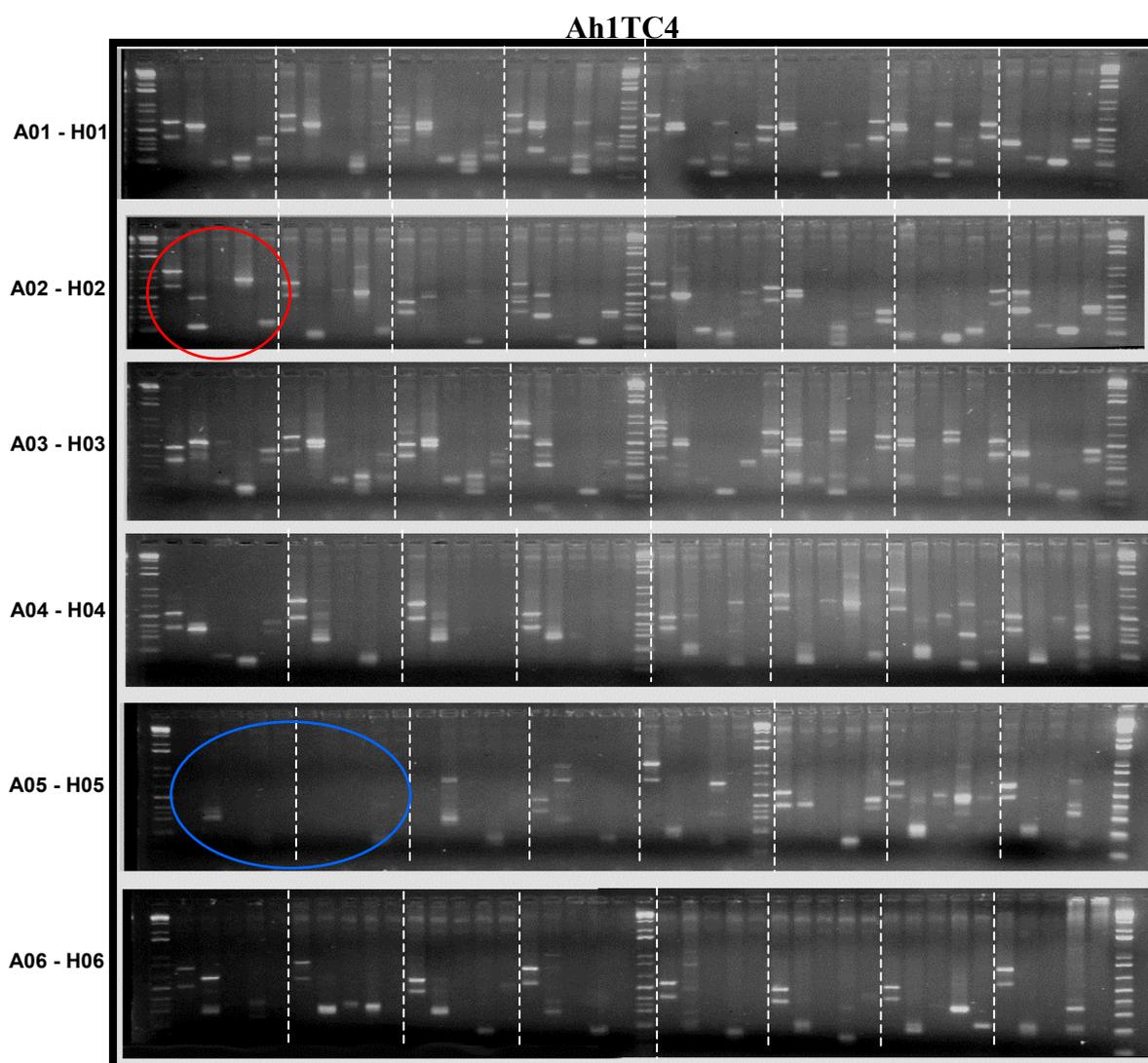


Figura 24 (A)- Perfil dos produtos de PCR oriundos do DNA genômico da placa Ah1TC4 entre os poços A01 e H06. O marcador de peso molecular utilizado foi o 1Kb plus. Nesta eletroforese podem-se observar falhas de amplificações em muitas amostras, como por exemplo, os poços referentes às amostras A05 e B05 (círculo azul). Em outros casos, na sua maioria, ocorreram amplificações inespecíficas, comprometendo a análise do comprimento do fragmento amplificado, amostra A02 (círculo vermelho).

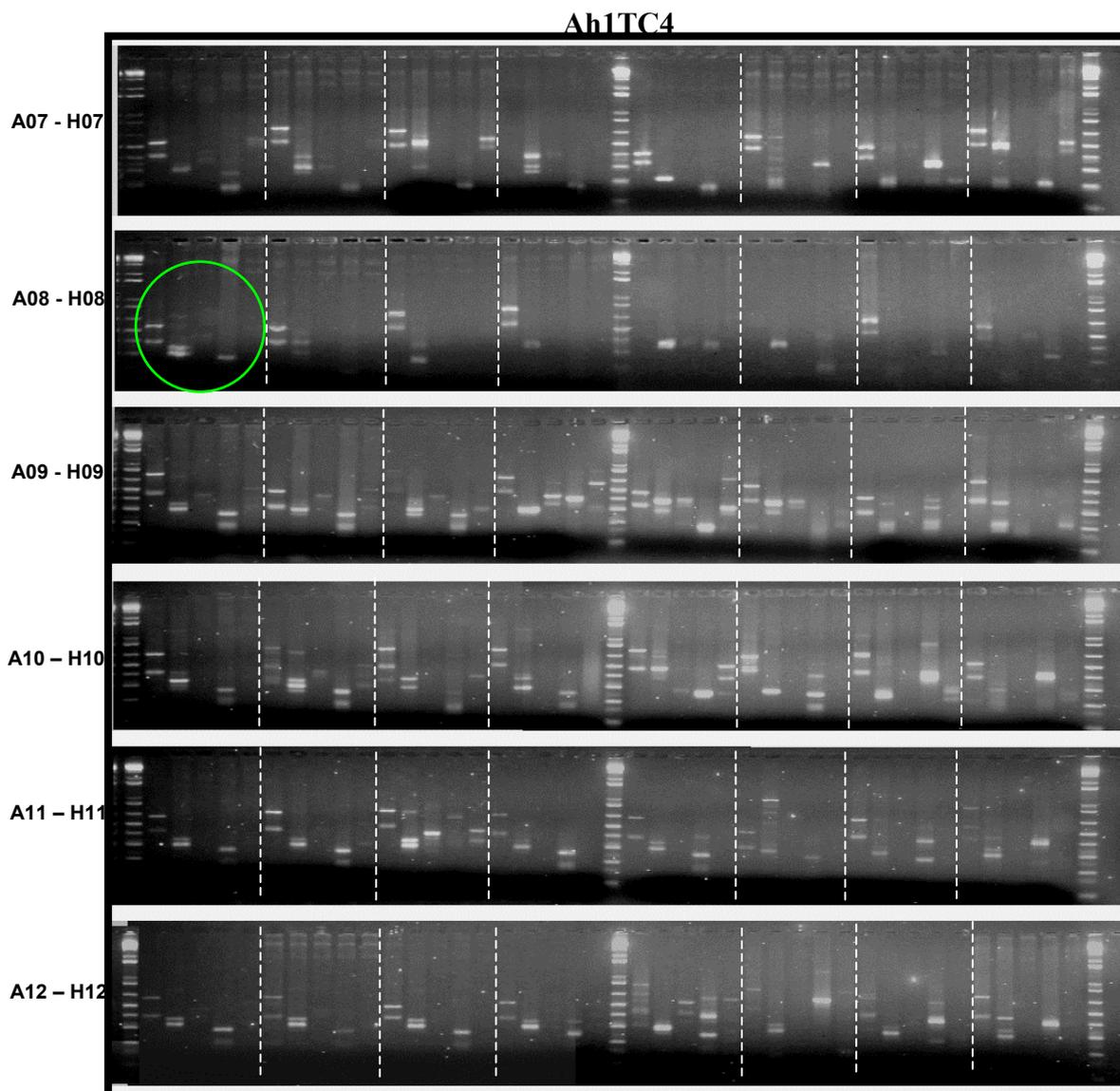


Figura 24 (B). Perfil dos produtos de PCR oriundos do DNA genômico da placa Ah1TC4 entre os poços A07 e H12. O marcador de peso molecular utilizado foi o 1Kb plus. Amostras que não apresentaram SSR quando seqüenciadas, também foram amplificadas erroneamente, por exemplo, amostra A08 (circulo verde).

A hipótese de contaminação dos reagentes ou placa foi descartada após testá-los com controles negativos e positivos para outras reações. O provável motivo para a inexatidão das amplificações pode ser uma ligação inespecífica do primers com uma outra região que contenha uma ou várias seqüências ricas em nucleotídeos A, G, C ou T, podendo estes estar sozinhos ou combinados entre si (Zietkiewicz *et al*, 1994; Fisher *et al.*, 1996).

Para fim de resultado comparativo baseado nos dados apresentados obtivemos a seguintes conclusões: (i) Foi verificado que a técnica do PCR ancorado não foi um bom indicador para seleção destes específicos clones contendo regiões repetidas em tandem, uma vez que apresentou dados duvidosos em quase todas as amostras se tornando ineficaz para o *screening*, porém é necessário salientar que a metodologia do PCR ancorado é amplamente aceita e adotada em vários trabalhos científicos. Outra desvantagem apurada foi o trabalho na elaboração, pois para cada clone dos 96 existentes na placa, foram feitos 5 combinações distintas de primers para a amplificação via PCR, totalizando ao final 6 placas com 480 amostras (Figura 25); (ii) Com relação à técnica de hibridização, o resultado obtido demonstrou um erro de quase 50% dos clones considerados como positivos. A desvantagem desta técnica está basicamente no trabalho intensivo, pois requer não só atenção e cuidado no manuseio com as colônias, como também exige preferencialmente ambiente esterilizado, evitando contaminações; (iii) O seqüenciamento direto, sem nenhuma pré-seleção para SSRs, foi a técnica que demonstrou o resultado mais rápido e verdadeiro dentre todas as comparações. Sua principal vantagem está na eficiência de se poder trabalhar com as 96 amostras quase todas ao mesmo tempo, pois a utilização de pipetas multicanais e a precipitação das amostras feitas em placa reduzem o tempo e o trabalho na execução do serviço. A tabela 14 tem por finalidade comparar as três técnicas levando em consideração os custos para a realização de cada uma delas.

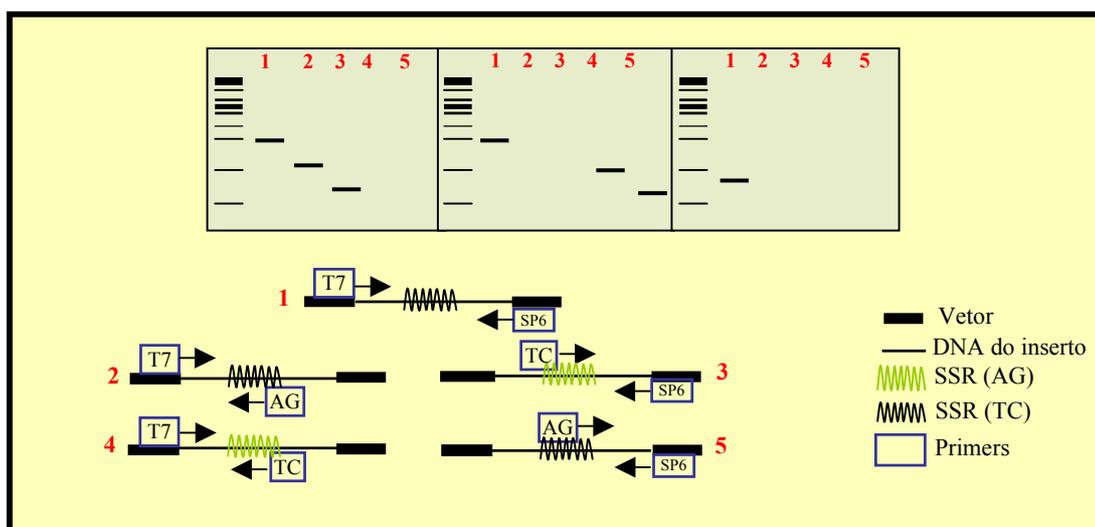


Figura 25. Ilustração mostrando a seleção de clones que contém SSR por PCR-ancorado: Para cada clone são feitas 5 reações de PCR: (1) – par de primers que amplifica o vetor inteiro; (2 a 5) – um par que amplifica do início do vetor até o SSR. Os clones 1 e 2 possuem SSR devido as amplificações do poços 2 e 3 ou 4 e 5; o clone 3 não tem SSR.

Adaptado: Buso *et al.*, 2000.

Tabela 14. Comparação dos custos pelas três principais técnicas (Hibridização; PCR ancorado e Sequenciamento Direto) na detecção de SSR da placa Ah1TC4 contendo 96 clones. (Ao total foram desenhados 25 pares de primers para esta placa)*.

TÉCNICAS	Principais produtos e reagentes utilizados	Preço por Produto (US\$)	Preço por Placa/ 96 amostras (US\$)	Total por placa (valor mínimo) (US\$)	Custo final do seqüenciamento das amostras em relação aos dados reais obtidos nos testes comparativos. \diamond	Custo por par de primers desenhado (\diamond 1/*25).
ibridização						
	Kit de Hibridização. RPN-3540 p/ 30 reações.	300,00	10,00		80,00 dólares. (incluiu a hibridização e o seqüenciamento das 55 amostras selecionadas como clones positivos na etapa de hibridização).	US 3,20
	Kit de Detecção RPN- 3510 p/ 2500 cm ²	340,00	17,00	27,00		
PCR-ancorado	Taq polymerase (250 U) Primers (três pares) (T7 e SP6; Anchor 1 e 2; Anchor 3 e 4).	161,00	161,00	215,00	307,00 dólares (Incluiu a PCR de 480 amostras mais o seqüenciamento das 96 amostras. Os dados da PCR-ancorado não foram conclusivos, impossibilitando a seleção dos prováveis clones positivos).	US 12,28
Seqüenciamento	Kit Seqüenciamento: DYEnamic ET terminator.p/100 reações (4 μ L/amostra)	420,00	91,43 (2 μ L/ amostra)	91,43	91,43 dólares. (Seqüenciamento de 96 amostras)	US 3,65

Fonte eletrônica: Amersham Pharmacia Biotech (preços via Internet/2003).

6.4 - Análise e Agrupamento das Seqüências para Montagem dos Contigs

Os cromatogramas com microssatélites da pasta HITS foram submetidos ao software interativo Staden, cuja função foi organizar e verificar as sobreposições de seqüências, além de realizar a inspeção e edição manual delas (Staden, 1996). Este software pode ser usado tanto no sistema operacional Windows como no Linux.

A primeira etapa de utilização deste software foi feito pelo programa denominado Pregap 4, que entre outras funções executou o controle de qualidade com o Phred, o corte do vetor das seqüências, identificação de microssatélites e compactação dos arquivos dos cromatogramas nos formatos ZTR (Dear & Staden, 1992). O Phred aplicou um valor de confiança (*confidence values*) que definiu a probabilidade de que a base chamada (*base call*) estivesse correta (Ewing & Green, 1998). Neste trabalho foi usado o parâmetro Phred > 15, isto é, a probabilidade de erro em cada nucleotídeo foi de 1 em 31,6. O corte de vetor foi feito por vários módulos diferentes, conseguindo um resultado final com alta precisão.

As seqüências aprovadas pelo Pregap 4 foram enviadas e analisadas em uma segunda etapa chamada Gap 4, cuja função principal foi a formação de contigs, que são sobreposições de cromatogramas. A formação dos contigs permitiu uma análise rápida e clara das seqüências, facilitando a exclusão de artefatos pela simples sobreposição e comparação de cromatogramas, gerando para cada *contig*, editado ou não, uma seqüência consenso (*consensus*). A seqüência consenso possui maior confiabilidade do que qualquer cromatograma individual presente no contig. Por exemplo, dois cromatogramas sobrepostos que possuem o valor de confiabilidade médio igual a 10 pode produzir uma seqüência consenso com valor de Phred igual a 20.

O software Staden é um pacote de programas interativos para montagem de sobreposições de seqüências, podendo utilizar programas como o GAP4, Cap3 e o Phrap. Para as sobreposições de seqüências contendo regiões microssatélites, o programa GAP4 foi o utilizado dentro do pacote Staden, pois diferentemente dos outros dois programas citados, ele considera as regiões de alta complexidade em detrimento das regiões de baixa complexidade, como os microssatélites. Antes de unir os prováveis cromatogramas iguais, o GAP4 realiza uma inspeção de similaridade, priorizando as regiões de alta complexidade e subestimando as regiões de baixa complexidade. Devido a esta peculiar inspeção, não ocorre sobreposição de cromatogramas baseados nas regiões microssatélites, evitando junções com dados errôneos. O Cap3 baseia-se apenas nos valores de confiança, não importando se a região apresenta alta ou baixa complexidade, podendo gerar falsos *contigs* no resultado final. Parecido com o modo operante do GAP4, o Phrap considera preferencialmente as regiões de alta complexidade em detrimento das de baixa complexidade, mas em nossa experiência junções inexatas ocorreram. Estes resultados errôneos são conseqüências diretas das sobreposições baseadas nas regiões repetitivas, que podem apresentar um alto valor de confiabilidade em relação aos das regiões que flaqueiam essas repetições.

6.5 - Desenhos dos Primers de SSR.

Dos 139 *contigs* criados na pasta HITS, apenas 79 apresentaram condições de terem os primers desenhados com o auxílio do programa Primer 3 (Anexo 2). Os demais foram

excluídos por não possuírem tamanho suficiente de repetição; a posição da seqüência simples repetitiva ser inadequada para o desenho dos pares de primers flanqueadores devido a posição do adaptador/vetor; repetições severamente imperfeitas ou não atendiam aos parâmetros configurados no programa Primer 3 (Figura 26).

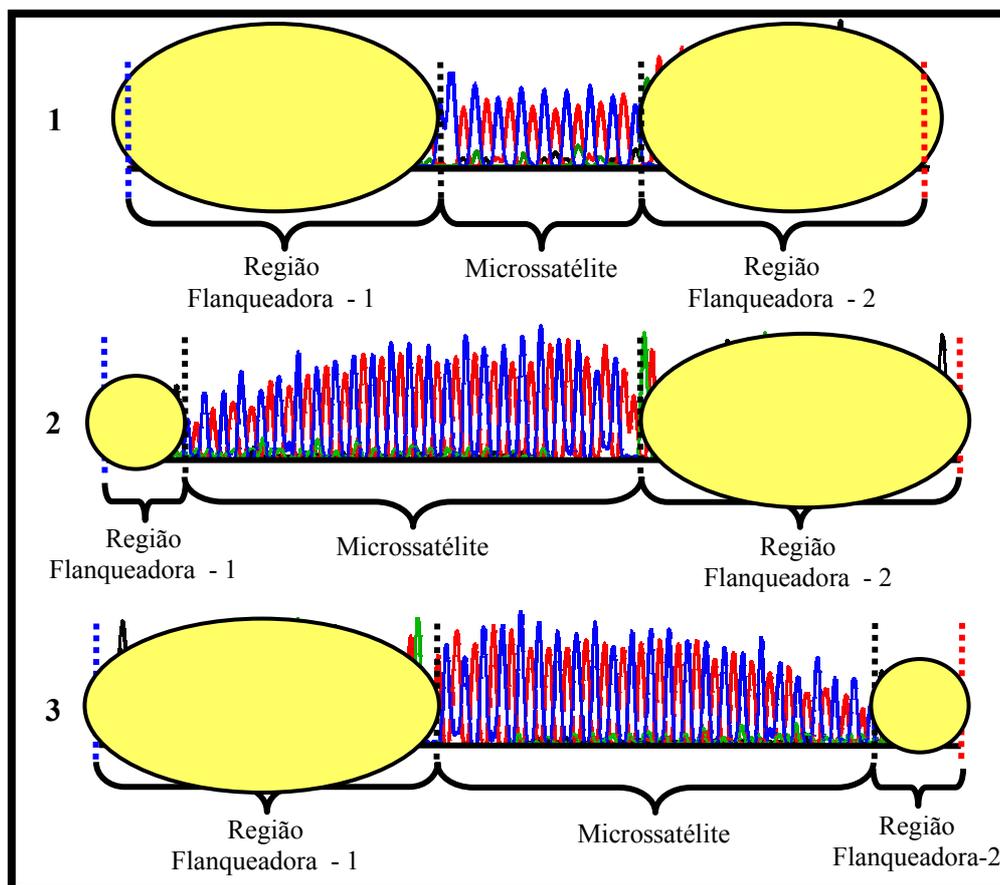


Figura 26. Desenhos de primers. 1) Amostra de DNA seqüenciado com região flanqueadora do SSR suficiente para desenhar primers *forward* e *reverse*. 2 e 3) Amostra de DNA seqüenciado com região flanqueadora do tamanho insuficiente para desenhar o primer *forward* ou *reverse*. Linha pontilhada azul = fim do vetor/adaptador e início da seqüência; linha pontilhada vermelha = fim da seqüência e início do adaptador/vetor; linha pontilhada preta = início ou fim da região flanqueadora do SSR.

As seqüências dos primers flanqueadores das regiões de microssatélites foram desenhadas com 17 a 22 nucleotídeos de comprimento, contendo aproximadamente 50% de GC, a temperatura média de ligação de 60°C e com o tamanho ideal da amplificação entre 100–250 pb (McCouch *et al.*, 1997).

Os microssatélites podem ser classificados como perfeitos, imperfeitos ou compostos (Weber, 1990). A maioria dos fragmentos seqüenciados (62,20%) continha repetições perfeitas, 20,73 % repetições imperfeitas e 17,07% repetições compostas

(Tabela 15). A predominância de repetições perfeitas também foi encontrada em arroz (Chen *et al*, 1997) e em eucalipto (Brondani *et al*, 1998).

Tabela 15. Caracterização dos novos marcadores SSR desenvolvidos para análise genômica de amendoim.

Marcador	Repetição	Tipo
Ah1TC0A01	(AG) ₂₉	PERFEITO
Ah1TC1A01	(TC) ₂₆	PERFEITO
Ah1TC1A02	(TC) ₃₅	PERFEITO
Ah1TC1A08	(GA) ₃₀	PERFEITO
Ah1TC1B01	(CA) ₂ (TC) ₁₁ (CA) ₉ 100-122(TC) ₉ 140-190 (TC) ₂₁	IMPERFEITO
Ah1TC1B02	(CT) ₂₅	PERFEITO
Ah1TC1D01	(AG) ₂₄	PERFEITO
Ah1TC1D02	(TC) ₃₁	PERFEITO
Ah1TC1D12	(TC) ₁₄	PERFEITO
Ah1TC1E01	(CA) ₄ (GA) ₂₉	COMPOSTO
Ah1TC1E05	(TC) ₃₀	PERFEITO
Ah1TC1E06	(TC) ₂₃	PERFEITO
Ah1TC1G04	(CTT) ₄ (CT) ₃₃	COMPOSTO
Ah1TC1H04	(A) ₁₄ 58-344 (TC) ₁₉ 383-577 (CTT) ₈	PERFEITO
Ah1TC2A02	(TC) ₃₂	PERFEITO
Ah1TC2A11	(TC) ₂₈ (GT) ₅	COMPOSTO
Ah1TC2B01	(TC) ₁₃	PERFEITO
Ah1TC2B05	(AG) ₂₄	PERFEITO
Ah1TC2B09	(TC) ₂₇	PERFEITO
Ah1TC2C03	(TC) ₃₉	PERFEITO
Ah1TC2C07	(TC) ₂₃	PERFEITO
Ah1TC2C11	(TC) ₂₄	PERFEITO
Ah1TC2C12	(CA) ₂₇ (TC) ₃	COMPOSTO
Ah1TC2D06	(AG) ₅ (A) ₃ (T) ₄ (AG) ₅ (A) ₅ 59-181(GA) ₃₁	COMPOSTO E PERFEITO
Ah1TC2D08	(TC) ₂₆	PERFEITO
Ah1TC2E05	(TC) ₂₆	PERFEITO
Ah1TC2E11	(TTA) ₆	PERFEITO
Ah1TC2G05	(GA) ₃₅	PERFEITO
Ah1TC3A10	(CT) ₃ GAGTA (CT) ₂₅	IMPERFEITO
Ah1TC3B04	(TC) ₂₄	PERFEITO
Ah1TC3B05	(AG) ₂₆	PERFEITO
Ah1TC3D04	(TC) ₃₄	PERFEITO
Ah1TC3E02	(CT) ₂₆ (CA) ₇ (CG) ₃ CTCGTG(CA) ₅	IMPERFEITO
Ah1TC3E03	(A) ₁₃	PERFEITO
Ah1TC3E05	(GA) ₂₅ (GT) ₃	COMPOSTO
Ah1TC3G01	(TC) ₁₄ (T) ₅ AAC(CT) ₂ GC(CT) ₁₅	IMPERFEITO
Ah1TC3G03	(TCC) ₁₀	PERFEITO
Ah1TC3G05	(CTT) ₁₆	PERFEITO
Ah1TC3H01	(CT) ₃ (T) ₁₄	COMPOSTO
Ah1TC3H02	(CT) ₂₈ GTGCTCTT(CG) ₂ (CT) ₆	IMPERFEITO
Ah1TC3H04	(A) ₁₀	PERFEITO

Marcador	Repetição	Tipo
Ah1TC3H07	(GA) ₂₃ 282-377 (TTC) ₄ 390-398 (TG) ₂ (TC) ₃ (CA) ₃ GT(CA) ₄	PERFEITO E IMPERFEITO
Ah1TC4A02	(GA) ₂₆	PERFEITO
Ah1TC4B01	(TC) ₃₁	PERFEITO
Ah1TC4B02	(TC) ₂₇	PERFEITO
Ah1TC4B07	(A) ₉ G(A) ₄ G(A) ₅ TAGCG(A) ₂₄	IMPERFEITO
Ah1TC4C08	(AC) ₃ (TC) ₃ 154-197 (GA) ₅	COMPOSTO
Ah1TC4C11	(CT) ₃₅	PERFEITO
Ah1TC4D02	(TC) ₂₁	PERFEITO
Ah1TC4D09	(CA) ₂ (CCA) ₂ (CT) ₁₈	COMPOSTO
Ah1TC4D12	(G) ₅ (AG) ₅ TAGTGA(GT) ₃	IMPERFEITO
Ah1TC4E08	(TC) ₂₈	PERFEITO
Ah1TC4E09	(TC) ₂₂	PERFEITO
Ah1TC4E10	(T) ₂ C(T) ₄ C(T) ₅ C(T) ₂ C(T) ₁₂ A(CT) ₁₅	COMPOSTO
Ah1TC4F01	(C) ₄ A(C) ₉	COMPOSTO
Ah1TC4F02	(C) ₁₄	PERFEITO
Ah1TC4F04	(CT) ₂₇ (CA) ₂	COMPOSTO
Ah1TC4F08	(TC) ₃₀	PERFEITO
Ah1TC4F10	(TTC) ₃₃	PERFEITO
Ah1TC4F12	(CT) ₂₃	PERFEITO
Ah1TC4G02	(CT) ₂₇	PERFEITO
Ah1TC4G05	(CT) ₃₄	PERFEITO
Ah1TC4G06	(CT) ₁₄ (GA) ₂	COMPOSTO
Ah1TC4G10	(GA) ₂₂ GCGT(GA) ₆ GT(GA) ₄	IMPERFEITO
Ah1TC4H02	(CT) ₁₇ (A) ₂ (T) ₅ CA(C) ₂ (T) ₂ C(CT) ₅ 164-185(CT) ₅ AACAT(TC) ₂ A(TC) ₇	IMPERFEITO
Ah1TC4H07	(CT) ₃₃ (A) ₂ (T) ₄ G(T) ₉ (G) ₁₀	IMPERFEITO
Ah1TC4H11	(A) ₉ (TC) ₃ (T) ₄ G(T) ₄ CG(A) ₁₆	IMPERFEITO
Ah1TC5A06	(CT) ₂₈	PERFEITO
Ah1TC5A07	(GA) ₁₆ TAT(GA) ₈	IMPERFEITO
Ah1TC5B06	(TG) ₂ (A) ₅ TG(TA) ₃ (TG) ₂ (A) ₄ TGTT(A) ₅ (TA) ₁₀	IMPERFEITO
Ah1TC5C05	(A) ₇ TTG(T) ₃ C(T) ₅ (C) ₇ T(CA) ₂ (CT) ₄ 280-347 (A) ₄ (GA) ₇ 367-441 (CA) ₃ (GA) ₁₃	IMPERFEITO E COMPOSTO
Ah1TC5D01	(CT) ₂₀	PERFEITO
Ah1TC5D06	(GA) ₂₁	PERFEITO
Ah1TC5H11	(TC) ₅ CATT(A) ₃ (CA) ₅	IMPERFEITO
Ah1TC6C07	(A) ₁₃	PERFEITO
Ah1TC6E01	(CT) ₂₄	PERFEITO
Ah1TC6G09	(CT) ₁₉	PERFEITO
Ah1TC6H03	(AG) ₂₂	PERFEITO

Dos 79 pares de primers desenhados, estima-se que 35 deles sejam polimórficos para a principal população de mapeamento (derivado de *A. stenosperma* x *A. duranensis*) do projeto do qual estes primers fazem parte. Estes valores de polimorfismo foram

baseados nos estudos de polimorfismo realizados pelo MSc. Márcio Moretsohn (comunicação pessoal) que demonstraram que 44,8% dos pares de primers para dinucleotídeos testados eram polimórficos. Em outro estudo recente, Guohao e colaboradores (2003) descobriram polimorfismo em 19 dos 56 pares de primers microssatélites desenhados, em variedades de *A. hypogaea*, obtendo a média de 4,25 alelos por loco.

Para um estudo mais detalhado sobre a espécie *A. hypogaea* e o desenvolvimento futuro de um mapa genético, duas outras bibliotecas foram desenvolvidas, porém ainda não seqüenciadas: uma enriquecida para o dinucleotídeo (AC)_n e a segunda sem nenhum enriquecimento. Para aumentar ainda mais a eficiência do número de microssatélites encontrados no genoma total e diminuir a redundância, Chen e colaboradores (1997) sugerem a utilização de outras enzimas de restrição para a construção de novas bibliotecas genômicas, tornando esta classe de marcadores uma fonte quase inesgotável.

6.6 - Análise de Similaridade

Todas as amostras seqüenciadas, pertencentes tanto à pasta HITS como à NO HITS, após a montagem pelo programas GAP4 e Cap3 respectivamente, tiveram suas seqüências consenso submetidas ao programa Blast-X para busca de similaridade contra um banco de dados de proteínas de *Arabidopsis thaliana* (Anexo 3). Uma seqüência foi classificada como homóloga se ele obtivesse o nível de significância *E value* (E) < 10^{-6} . Seqüências com múltiplas homologias e com (E) < 10^{-6} foram classificadas baseadas no melhor *score* de homologia dentro da banco de dados. Entre as 138 seqüências da pasta HITS submetidas à análise, 13 foram significantes quanto à similaridade com genes putativos e 125 não apresentaram similaridade alguma. Quanto às seqüências consensos da pasta NO HITS, das 340 seqüências, 29 foram significativas para algum gene e 311 seqüências não apresentaram similaridade (Figura 27).

As seqüências pertencentes tanto a HITS como a NO HITS que apresentaram homologia com algum gene de *A. thaliana*. Em grande proporção à NO HITS representaram retrotransposons, enquanto que os pertencentes à pasta HITS eram em grande parte homólogos de seqüências *putative proteins* (Tabela 16 e 17). Um total de 436 seqüências de ambas as pastas não apresentaram homologia com nenhum gene.

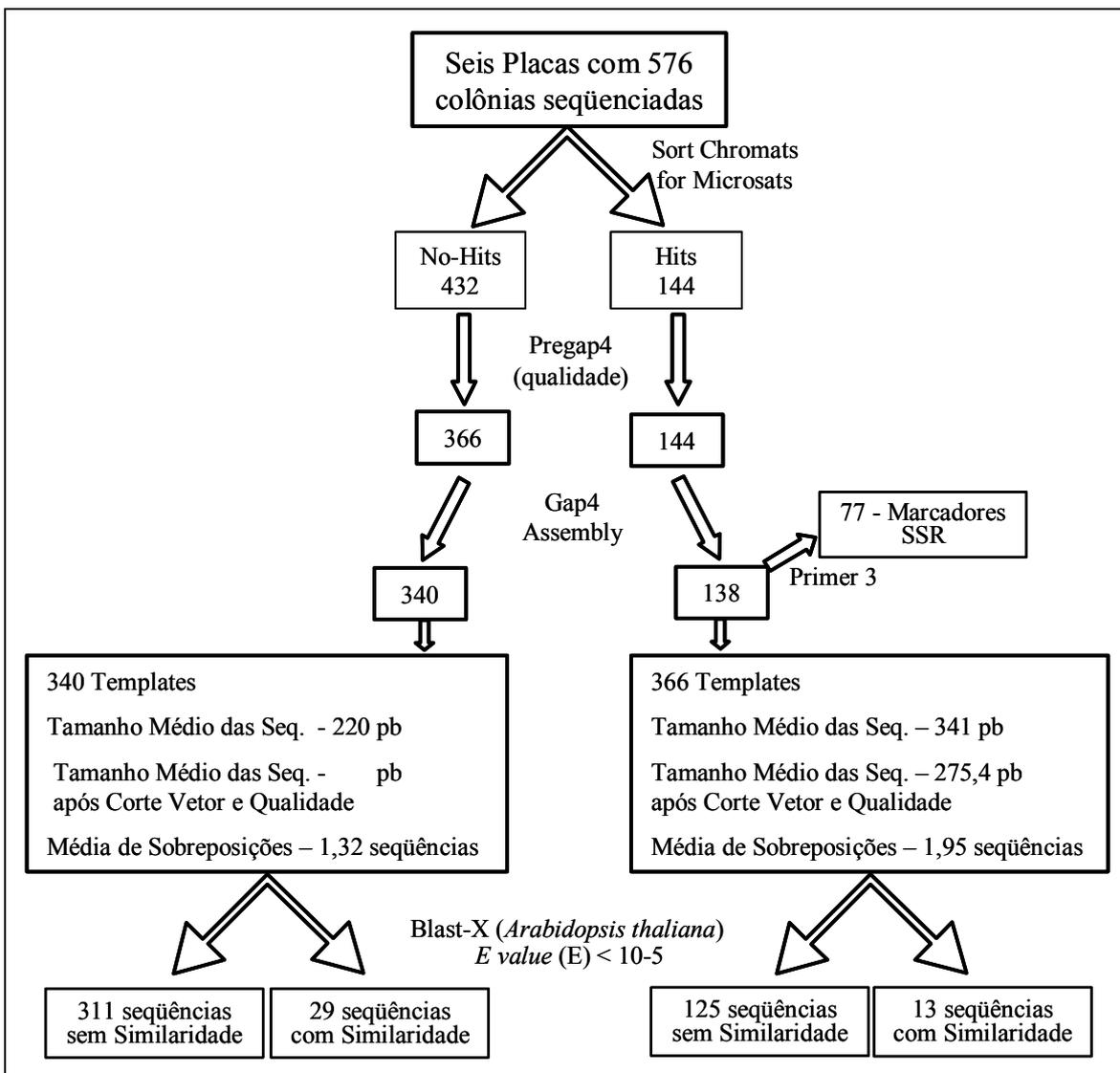


Figura 27. Fluxograma das etapas de análise computacional das seqüências de *A. hypogaea* no projeto.

Tabela 16. Resultado do Blast-X das seqüências presentes na Pasta NO-HITS para busca de similaridade contra um banco de dados local de gene de *Arabidopsis thaliana*.

Homologias das seqüências de <i>A. hypogaea</i> sem SSR	Frequência encontrada (E) < 10 ⁻⁶
<i>retroelement</i>	20
<i>hypothetical protein</i>	4
<i>topoisomerase II</i>	1
<i>disease resistance protein, putative</i>	1
<i>Serine/threonine phosphatase PP7, putative</i>	1
<i>phosphatidylinositol-4-phosphate 5-kinase isolog</i>	1
<i>putative auxin-induced protein</i>	1
Homologias de genes encontrados	29

Tabela 17. Resultado do Blast-X das seqüências presentes na Pasta HITS para busca de similaridade contra um banco de dados local de gene de *Arabidopsis thaliana*.

Homologias das seqüências de <i>A. hypogaea</i> com SSR	Freqüência encontrada (E)<10 ⁻⁶
<i>putative protein</i>	6
<i>retroelement</i>	2
<i>alpha subunit of F-actin capping protein</i>	1
<i>ATP dependent copper transporter</i>	1
<i>auxin-independent growth promoter, putative</i>	1
<i>serine/threonine kinase, putative</i>	1
<i>Tic22, putative</i>	1
Homologias de genes encontrados	13

O seqüenciamento do genoma do organismo modelo *A. thaliana* oferece oportunidade de integração de informação sobre as diversas espécies de plantas. A existência de homologia entre genes de *A. thaliana* e gene de diferentes espécies indica conservação de função, o que permite inferir sobre as funções de genes de outras espécies, como em *A. hypogaea* (Osterlund *et al.*, 2002).

Os microssatélites em geral são encontrados preferencialmente dentro de regiões não codificantes do genoma, mas podem ser encontrados também em regiões codificantes (Tautz *et al.*, 1986; Hancock, 1995). Morgante e colaboradores (2002) em estudos comparativos de microssatélites entre amostras do genoma e amostras transcritas (ESTs-*Expressed Sequence Tags*), verificaram que entre os dinucleotídeos presentes em EST, as repetições do tipo TC/AG foram encontradas em maior quantidade. Isto sugere a similaridade encontrada nas 13 seqüências ricas em repetições TC/AG de *A. hypogaea* com os genes de *A. thaliana* pelo Blast-X. O resultado da homologia das seqüências sem microssatélites (NO HITS) indica talvez uma abundância de retrotransposons no genoma de *A. hypogaea*. Kumar & Bennetzen (1999) citam a alta abundância de retrotransposons no genoma de plantas e em milho estes elementos constituem em quase 50% do seu genoma (San Miguel *et al.*, 1996). Os elementos transponíveis ocasionalmente deslocam-se ou reorganizam as seqüências vizinhas de DNA do genoma. Eles freqüentemente causam deleções e adições de seqüências de nucleotídeos adjacentes. É provável que esses elementos sejam responsáveis por muitas mudanças evolutivas no genoma e sejam importantes na evolução das espécies (Alberts *et al.*, 1997).

Conclusões

7 - Conclusões

Os resultados comprovaram a existência de regiões ricas em repetições TC/AG em *Arachis hypogaea*, possibilitando o desenvolvimento de mais marcadores microssatélites para esta planta cultivada.

O auxílio da informática tanto na busca e qualidade bem como na interpretação dos resultados gerados de microssatélites, facilitou o desenvolvimento dos primers desenhados, influenciando diretamente na qualidade e velocidade final produzida.

A análise de similaridade de regiões do cruzamento de dados com outras espécies, em particular com o genoma de *Arabidopsis thaliana*, permitiu verificar a homologia de algumas regiões do genoma de *A. hypogaea* com genes já identificados em *A. thaliana*. O resultado da homologia indicou que aproximadamente 10% do SSRs do tipo TC/AG encontraram-se em regiões gênicas e indicou também a abundância de retrotransposons no genoma de *A. hypogaea*.

Perspectivas Futuras

8 - Perspectivas Futuras

A identificação de polimorfismo para os primers desenhados e a procura de mais regiões repetidas tanto na biblioteca enriquecida (TC) como para as outras duas já desenvolvidas, porém ainda não seqüenciadas, possibilitarão a construção de um mapa genético informativo, onde genes de interesse (herança qualitativa ou quantitativa) poderão ser identificados e mapeados. Os microssatélites também terão aplicações em análises da diversidade genética entre indivíduos bem como na seleção assistida, através de retrocruzamentos.

Referências Bibliográficas

9- Referências Bibliográficas

- AHARONI, A.; BARAN, N.; MANOR, H. 1993. Characterization of a multisubunit human protein which selectively bind single stranded d(GA)_n and d(GT)_n sequence repeats in DNA. *Nucleic Acids Research*, **21**:5221-5228.
- AHMED, E.H; YOUNG, C.T. 1982. Composition, nutrition, and flavor of peanuts. In H.E. Pattee and C.T. Young (eds), *Peanut Science and Technology*, Am. Peanut Res. Educ. Soc., Inc., Yokum, Tx, p. 655-688.
- ALBERTS, B.; BRAY, D.; LEWIS, J.; RAFF, M.; ROBERTS, K.; WATSON, J.D. 1997. *Biologia Molecular da Célula*. 3ª ed. Editora Artes Médicas, Porto Alegre, RS.
- ARMOUR, J.A.L; ALEGRE, S.A; MILES, S.; WILLIAMS, L.; BADGE, R. 1999. Minisatellites and mutation processes in tandem repetitive DNA. In *Microsatellites: Evolution and Applications*. Edited by D. Goldsteins & Christian Schöterer. Oxford University Press, New York.
- BAIROCH, A.; APWEILER, R. 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acid Reserch*. **28(1)**: 45-48.
- BAXEVANIS, A.D. 2001. Bioinformatics and the internet. In: *Bioinformatics: A Practical Guide to the Analysis of Genes & Proteins*. Edited by Baxevanis, A.D; Ouellette B. F. Wiley-Interscience; 2nd edition.
- BECKMANN, J.S. 1991. Genomic genetic and plant genetic improvement. In: *Gene-mapping techniques and applications*. Schook, L.B.; Lewin, H.A.; MacLaren, D.G. (eds). Marcel Dekker Inc., New York. Pp 201-230.
- BECKMANN, J.S; WEBWE, J.L. 1992. Survey of human and rat microsatellites. *Genomics*. **12**:627-631.
- BELKUM, A.; SCHERER, S.; ALPHEN, L.; VERBRUGH, H. 1998. Short-sequence DNA repeats in prokaryotic genomes. *Microbiology and Molecular Biology Reviews*, **62(2)**:275-293.
- BENNETZEN, J.L. 2000. Comparative sequence analysis of plant nuclear genomes. Microcolinearity and its many exceptions. *Plant Cell* **12**: 1021–1030.
- BIGGIN, M.D.; TJIAN, R. 1988. Transcription factors that active the Ultra-bithorax promoter in developmentally staged extract. *Cell*, **53**:699-711.
- BONFIELD, J.K.; SMITH, K.F.; STADEN, R. 1995. A new DNA sequence assembly program. *Nucleic Acids Res*, **24**:4992-4999.
- BORÉM, A. 1998. *Melhoramento de plantas*. 2º edição, Viçosa: Editora UFV, p.453.
- BOTSTEIN, D.; WHITE, R.L.; SKOLNICK, M.; DAVIS, R.W. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* **32**:314-331.
- BRONDANI, R.P.V.; BRONDANI, C.; TARCHINI, R.; GRATTAPAGLIA, D. 1998. Development, characterization and mapping of microsatellite markers in *Ecalyptus grandis* and *E. urophylla*. *Theor Appl Genet*, **97**:816–827.
- BROUN, P.; GANAL, M.W.; TANKSLEY, S.D. 1992. Telomeric arrays display high levels of heritable polymorphism among closely related plant varieties. *Proc Natl Acad Sci USA* **89**:1354-1357
- BROWN, S.M.; HOPKINS, M.S.; MITCHELL, S.E.; SENIOR, M.L.; WANG, T.Y.; DUNCAN, R.R.; GONZALEZ-CANDELAS, F.; KRESOVICH, S. 1996. Multiple methods for the identification of polymorphic simple sequence repeats (SSRs) in sorghum [*Sorghum bicolor* (L.) Moench]. *Theoretical and Applied Genetics*, **93**:190–198.
- BRÜCHER, H. 1989. *Useful plants of neotropical origin and their wild relatives*. London, Springer-Verlag.
- BUROW, M.D.; STARR, J.L.; SIMPSON, C.E.; PATERSON, A.H. 1996. Identification of RAPD markers in peanut (*Arachis hypogaea*) associated with root-knot nematode resistance derived from *A. cardenasii*. *Molec Breed*, **2**:307-319.

- BUSO, G. S. C.; BRONDANI, R. V.; AMARAL, Z. P. S.; REIS, A. M.; FERREIRA, M. E. 2000. Desenvolvimento de primers SSR para análise genética de pimentas e pimentões (*Capsicum spp.*) utilizando biblioteca genômica enriquecida. Em Boletim de Pesquisa nº 15 – Embrapa Recursos Genéticos e Biotecnologia. Brasília,DF.
- CAETANO-ANÓLLES, G.; BASSAM, B.J.; GRESSHOF, P.M. 1991. High resolution DNA amplification fingerprint using very short arbitrary oligonucleotide primers. *Bio/technology*, **9**:553-557.
- CAI, Q.; LU, S.; CHINNAPPA, C.C. 1987. Analysis of karyotypes and giemsa C-banding patterns in eight species of *Arachis*. *Genome*, **29**:187-194.
- CARLEY, D.H.; Fletcher, S. M. 1995. In overview of world peanut markets, In: eds H. E. Pattee and H. T. Stalker, *Advances in Peanut Science, Am. Peanut Res. And Educ. Soc.* pp 554-557.
- CHASSY, B.M.; MERCENIER, A.; FLICKINGER, J. 1988. Transformation of bacteria by electroporation. *Tibtech*, **6**:303-309.
- CHEN, X; TEMNYKH, S.; XU, Y., CHO, Y.G.; MCCOUCH, S.R. 1997. Development of a microsatellite framework map providing genome-wide coverage in rice (*Oriza sativa* L.) *Theor Appl. Genet.*, **95**:673-679.
- CRANE, P.R.; FRIISM, E.M.; PEDERSEN, K.J. 1995. The original and early diversification of angiosperms. *Nature*, **374**:27-33.
- CREGAN, P.B; AKKAYA M.S.; BHAGWAT A.A.; LAVI, U.; RONGWEN, J. 1994. Length polymorphisms of simple sequence repeat (SSR) DNA as molecular markers in plants, pp 47-56. In: P.M. Gresshoff (ed.), *Plant Genome Analysis*. CRC .Press, Boca Raton, Florida.;
- CSINK, A.K.; HENIKOFF, S. 1998. Something from nothing: the evolution and utility of satellite repeats. *Trends in Genetics*, **14**:200–204.
- DALLAS, J.F. 1988. Detection of DNA “fingerprints” of cultivated rice by hybridization with a human minisatellite DNA probe. *Proc. Nat. Acad. Sci.* **85**:6831-6835.
- DAVIES, J.; JACIB, F. 1968. Genetic mapping of the regulator and operator genes of the lac operon. *Journal Molecular Biology*, **36**:413-417.
- DAYHOFF, M. O., SCHWARTZ, R. M. AND B. C. ORCUTT. 1978. In "Atlas of Protein Sequences and Structure,"Vol. 5, Suppl. 3, pp. 345-352. Nat. Biomed. Res. Found., Washington, D. C.
- DEAR, S.; STADEN, R. 1992. A standard file format for data from DNA sequencing instruments. *DNA Sequence* **3**:107-110.
- DIWAN, N.; CREGAN, P.B. 1997. Automated sizing of fluorescent labeled simple sequence repeat (SSR) markers to assay genetic variation in soybean. *Theor. Appl. Genet*, **95**: 723–733.
- DOWER, W.J.; MILLER, J.F.; RAGSDALE, C.W. 1988. High efficiency transformation of *E. coli* by high voltage electroporation. *Nucl. Acids Rec*, **16**:6127-6145.
- DOYLE, J.J.; DOYLE, J.L. 1997. Isolation of plant DNA from fresh tissue. *Focus*, **12**:13-15.
- EISEN, J.A. 1999. Mechanistic basis for microsatellite instability. In *Microsatellites: Evolution and Applications*. Edited by. D. Goldsteins & Christian Schötterer. Oxford University Press, New York.
- EWING, B.; GREEN, P. 1998. Base-Calling of Automated Sequencer Traces Using Phred. II. Error Probabilities. *Genome Research*, **8(3)**:186-194.
- FERNANDEZ, A.; KRAPOVICKAS, A. 1994. Cromosomas y evolución em *Arachis* (Leguminosae). *Bonplandia*, **8**: 187-220.
- FERREIRA, M. E.; GRATTAPAGLIA, D. 1998. Introdução ao uso de marcadores RAPD e RFLP em análise genética. Brasília: Embrapa-CENARGEM, 3ªed. 220p.
- FIELD, D; WILLS, C. 1998. Abundant microsatellite polymorphism in *Saccharomyces cerevisiae* and different distributions of microsatellites in eight prokaryotes and *S. cerevisiae*, result from strong mutation pressures and a variety of selective forces. *Proc. Natl. Acad. Sci.*, **95**:1647-1652.

- FISHER, P.J.; GARDNER, R.C; RICHARDSON, T.E. 1996. Single locus microsatellites isolated using 5' anchored PCR. *Nucleic Acids Research*, **24**:4369–4371.
- FLANNERY, K.V. 1973. The origins of agriculture. *Ann. Rev. of Anthropic*, **2**:271-310.
- FREIRE, R.M.M. 1997. Estudo de aminoácidos em genótipos de amendoim (*Arachis hypogaea* L.). João Pessoa: UFPb,. 118p. Tese Mestrado.
- FREIRE, R.M.M.; SANTOS, R.C. dos; BELTRÃO, N. E. de M. 1996. Qualidade nutricional e industrial de algumas oleaginosas herbáceas cultivadas no Brasil. *Óleos e Grãos*, n. 28, p. 49-53.
- GARCIA, G.M.; STALKER, H.T.; SHROEDER, E.; KOCHERT, G.A. 1996. Identification of RAPD, SCAR and RFLP markers tightly linked to nematode resistance genes introgressed from *Arachis cardenasii* to *A. hypogaea*. *Genome*, **39**:836-845.
- GEPTS, P. 1993. The use of molecular and biochemical markers in crop evolution studies. In: *Evolutionary Biology*, V27. Edited by: MK Hecht, et al. Plenum Press, NY, NY.
- GIBAS, C.; JAMBECK. P. 2001. Desenvolvimento Bioinformática: ferramenta de software para aplicações em biologia. Editora Campus. Rio de Janeiro, RJ.
- GODOY, I.J.; MORAIS, S.A.; ZANOTTO, M.D.; SANTOS, R.C. 1989. Melhoramento do amendoim. In Borém (Ed). *Melhoramento de Espécies Cultivadas*. Editora UFV, Viçosa, pp 51-93.
- GODOY, J.I.; MORAES, D.A de; VEIGA, R.F.A. 1997. Avaliação agronômica de germoplasma de amendoim com ênfase para resistência à doença nas condições brasileiras. In: Simpósio Latino-Americano de Recursos Genéticos Vegetais, 1, Campinas. *IAC/CENARGEM*, pg. 18-19.
- GOTTLIEB L.D. 1981. Electrophoretic evidence and plant populations. *Progr Phytochem* **7**:1-46.
- GREEN, E.D. 1999. Documentation for Phrap and Cross_match (version 0.990319).
- GREEN, E.D. 2001. Strategies for systematic sequencing of complex genomes. *Genetics*, **2**:573-583.
- GREGORY, M. P.; GREGORY, W. C. 1979. Exotic germplasm of *Arachis* L. Interspecific hybrids. *J. Hered*, **70**:185-193.
- GREGORY, W C.; GREGORY, M P. 1976. Groundnut. In: *Evolution of crop plants*. Ed. N. W. Simmonds. Longman Group. London. pg. 151-154.
- GRIFFITHS, A.J.F.; MILLER, J.H.; SUZUKI, D.T.; LEWONTIN, R.C; GELBART, W.M. 2000. An introduction to Genetic Analysis. 7th ed. New York: W H Freeman & Co.
- GRODZICKER, T.; WILLIAMS, J.; SHARP, P.; SAMBROOK, J. 1974. Physical mapping of temperature-sensitive mutations of adenoviruses. *Cold Spring Harbor Symp. Quant. Biol*, **39**:439-446.
- GUR-ARIE, R.; COHEN, C.; EITAN, Y.; SHELEF, L.; HALLERMAN, E; KASHI, Y, 2000. Simple sequence repeats in *Escherichia coli*: Abundance, distribution, composition and polymorphism. *Genome Research*, **10**: 62-71.
- HAMADA, H.; PETRINO, M.G.; KAKUNAGA, T. 1982. A novel repeat element with Z DNA forming potential is widely found in evolutionary diverse eukaryotic genomes. *Proc. Nati. Acad. Sci USA*, **79**:6455-6469.
- HAMMONS, R.O. 1994. The origin and history of groundnut. In Smartt, J. *The Groundnut Crop. A scientific basis for improvement*. London, Chapman and Hall, chapter2, p.24-42.
- HANAHAN, D. 1983. Studies on transformation of *Escherichia coli* with plasmid. *Journal of Molecular Biology*, **166**:557-559.
- HANCOCK, J. 1995. The contribution of slippage-like processes to genome evolution. *Journal of Molecular Evolution*, **41**:1038-1047.
- HANCOCK, J. 1996. Simple sequences and the expanding genome. *BioEssays*, **18**:421- 425.
- HANCOCK, J. 1999. Microsatellites and other simple sequences: genomic context and mutational mechanisms. In *Microsatellites: Evolution and Applications*. Edited by. D. Goldsteins & Christian Schötterer. Oxford University Press, New York.

- HARLAN, J. R. 1987. Crops and Man. *American Society of Agronomy*, Madison.
- HARTL, L.; MOHLER, V.; ZELLER, F.J.; HSAM, S.L.K.; SCHWEIZER, G. 1999. Identification of AFLP markers closely linked to the powdery mildew resistance genes Pm1c and Pm4a in common wheat (*Triticum aestivum* L.). *Genome*, **42**:322–329.
- HE, G.H.; PRAKASH, C.S. 1997. Identification of polymorphic DNA markers in cultivated peanut (*Arachis hypogaea* L.). *Euphytica*, **97**:143-149.
- HILLIS, D.M.; MORITZ, C.; MABLE, B.K. 1996. *Molecular Systematics*, 2nd Edition. Sinauer Associates, Sunderland MA.
- HOPKINS, M.S.; CASA, A.M.; WANG, T.; MITCHELL, S.E.; DEAN, R.E.; KOCHERT, G.D.; KRESOVICH, S. 1999. Discovery and characterization of polymorphic simple sequence repeats (SSRs) in cultivated peanut (*Arachis hypogaea* L.). *Crop Sci*, **39**:1243-1247.
- HOPKINS, M.S.; CASA, A.M.; WANG, T.; MITCHELL, S.E.; DEAN, R.E.; KOCHERT, G.D.; KRESOVICH, S. 1999. Discovery and characterisation of polymorphic simple sequence repeats (SSRs) in peanut. *Crop Science*, **39**:1243–1247.
- HUANG, X.; MADAN, A. 1999. CAP 3: Sequence Assembly Program. *Genome Research*, **9**:868-877.
- HUSTED, L. 1933. Cytological studies of the peanut *Arachis*. I. Chromosome number and morphology. *Cytologia*, **5**:109-117.
- HUSTED, L. 1936. Cytological studies of the peanut *Arachis*. II. Chromosome number, morphology and behaviour, and their application to the problem of the origin of the cultivated forms. *Cytologia*, **7**:396-423.
- INSTITUTO NACIONAL DE TECNOLOGIA AGROPECUÁRIA-INTA. 1986. **Mani**: historia, importância, técnica de cultivo, uso y comercialización. Manfredi, Córdoba., 52p. (Cuaderno de actualización técnica,3).
- ISLEIB, T.G.; WYNNE, J.C.; NIGAM, S.N. 1994. Groundnut breeding In. Ed. J. Smartt, The Groundnut Crop: A Scientific Basis of Improvement, Chapman and Hall, London, p.552-623.
- JEFFREYS, A.J.; WILSON, V.; THEIN, S.L. 1985. Hypervariable “minisatellite” regions in human DNA. *Nature*, **316**:76-79.
- KERRIDGE, P.C.; HARDY, B. 1994. Biology and agronomy of forage *Arachis*. Cali, CIAT.
- KIRTI, P.B.; MURTY, U.R.; BHARATHI, M.; RAO, N.G.P. 1982. Chromosome pairing in F1 hybrid *Arachis hypogaea* L. x *A. monticola* Krap. et Rig. *Theoretical Applied Genetic*. **62**:139-144.
- KNAPP, P.E. 1991. Studies of glial lineage and proliferation in vitro using an early marker for committed oligodendrocytes. *J. Neurosci Res*, **30(2)**: 336-45.
- KNAUFT, D.A.; OZIAS-AKINS, P. 1995. Recent methodologies for germplasm enhancement and breeding. in eds. H.E. Pattee and H.T. Stalker, Advances in Peanut Science, *Am. Peanut Res. And Educ. Soc.*, Inc., Stillwater, OK, p.54- 94.
- KOCHERT, G.; HALWARD, T.; BRANCH, W.D.; SIMPSON, C.E. 1991. RFLP variability in peanut (*Arachis hypogaea* L.) cultivars and wild species. *Theor. Appl. Genet.* **81**:565–570.
- KOCHERT, G.; STALKER, H.T.; GINENES, M.; GALGARO, L.; MOORE, K. 1996. RFLP and cytogenetic evidence for the progenitor species of allotetraploid cultivated peanut, *Arachis hypogaea*, (Leguminosae) *Am. J. Bot.* **83**:1282–1291.
- KONIECZNY, A.; AUSUBEL, F.M. 1993. A procedure for mapping Arabidopsis mutations using codominant ecotype-specific PCR-based markers. *Plant Journal*. **4**:403-410.
- KRAPOVICKAS, A.; GREGORY, W.C. 1994. Taxonomía del género *Arachis* (Leguminosae). *Bonplandia*, **8(1-4)**: 1-186.
- KRAPOVICKAS, A.; RIGONI, V.A. 1951. Estudios citológicos en el género *Arachis*. *Rev. Invest. Agr.* **5**:289–294.
- KUMAR, A.; BENNETZEN, J.L. 1999. Plant retrotransposons. *Annu.Rev. Genet.* **33**: 479–532.

- LAGERCRANTZ, U.; ELLEGREN, H.; ANDERSON, L. 1993. The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. *Nucleic Acids Research*, **21**:1111-1115.
- LANDE, R. AND THOMPSON, R. 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics*, **124**: 743-756.
- LANDER E.S.; BOTSTEIN D. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, **121**:185-199.
- LAVIA, G.I. 1996. Estudos cromosômicos em *Arachis L. Bonplandia*, **9**:111-120.
- LAVIA, G.I. 1998. Karyotypes of *Arachis palustris* and *A. praecox* (Section *Arachis*), two species with basic chromosome number $x=9$. *Cytologia*, **63**: 177-181.
- LITT, M.; LUTY, J. 1989. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *American Journal Human Genetic*, **44**:398-401.
- LOH, J.P.; KIEW, R.; HAY, A.; KE, A.; GAN, L.H.; GAN, Y.Y. 2000. Intergeneric and interspecific relationships in Araceae tribe Caladieae and development of molecular markers using amplified fragment length polymorphism (AFLP). *Annals of Botany*, **85**:371-378.
- LOPES, M.S.; SEFC, K.M.; EIRAS DIAS, E.; STEINKELLNER, H. 1999. The use of microsatellites for germplasm management in a Portuguese grapevine collection. *Theoretical and Applied Genetics* **99**:733-739.
- LUE N.L.; BUCHMAN, A.R.; KORNBERG, R.D. 1989. Activation of yeast RNA polymerase II transcription by a thymidine-rich upstream element in vitro. *Proceedings of the National Academy of Sciences USA*, **86**:486-490.
- MAGUIRE, T.L. 2001. Producing and Exploiting Enriched Microsatellite Libraries. In Plant Genotyping: The DNA Fingerprinting of Plants. Edited by: R J Henry. Centre for Plant Conservation Genetics, Southern Cross University, Lismore, Australia.
- MARKLUND, S.; ELLEGREN, H.; ERIKSSON, S.; SANDBERG, K.; ANDERSSON, L. 1994. Parentage testing and linkage analysis in the horse using a set of highly polymorphic microsatellites. *Animal Genetics*, **25**: 19-23.
- MILLER, J. H. 1972. Experiments in molecular genetics. New York: Cold Spring Harbor.
- MOORE, H.; GREENWELL, P.W.; LIU, C.P.; ARNHEIM, N.; PETES T.D. 1999. Triplet repeats form secondary structures that escape DNA repair in yeast. *Proceedings of the National Academy of Sciences USA*, **96**:1504-1509.
- MOSS, D.W. 1982. Isoenzymes. Capman & Hall, London & New York.
- MOXON, E.R.; RAINEY P.B. 1995. Pathogenic bacteria: the wisdom of their genes, p. 255-268. In B. A. M. Van der Zeijst, L. Van Alphen, W. P. M. Hoekstra, and J. D. A. van Embden (ed.), Ecology of pathogenic bacteria. Royal Dutch Academy of Sciences, second series, no. 96. Royal Dutch Academy of Sciences, Amsterdam, The Netherlands.
- MUELLER, U.G.; WOLFENBARGER, L.L. 1999. AFLP genotyping and fingerprinting. *Trends in Ecology and Evolution* **14**:389-394.
- MULLIS, K.; FALOONA, F. 1987. Specific synthesis of DNA in vitro via a polymerase catalyzed chain reaction. *Methods Enzymol*, **55**:335-350.
- MULUVI, G.M.; SPRENT, J.I.; SORANZO, N.; PROVAN, J.; ODEE, D.; FOLKARD, G.; MCNICOL, J.W.; POWELL, W. 1999. Amplified fragment length polymorphism (AFLP) analysis of genetic variation in *Moringa oleifera* Lam. *Molecular Ecology*, **8**:463-470.
- MURTHY, U.R.; RAO, N.G.P.; KIRTI, P.B.; BHARATI, M. 1981. Cytogenetics and groundnut improvement Report (1978 - 1981). Índia, Hyderabad: ICRISAT, 66p.
- MYERS, R.M.; MANIATS, T.; LERMAN, S. 1987. Detection and localization of single base changes by desaturating gradient gel electrophoresis. *Methods in Enzymology*. **155**:501-527.

- NODARI, R.O.; TSAI, S.M.; GILBERTSON, R.L.; GEPTS, P. 1993. Towards an integrated linkage map of common bean. II. Development of an RFLP-based linkage map. *Theor. Appl. Genet.* **85**:513–520.
- NYBOM, H; HALL, H.K. 1991. Minisatellite DNA “fingerprint” can distinguish *Rubus* and estimate their degree of relatedness. *Euphytica*, **53**:107-114.
- OLSO, M.; HOOD, L.; CANTOR, R.; DOSTSTEIN, D.A. 1989. A common language for physical mapping of the human genome. *Science* . **254**:1434-1435.
- ORITA, M.; IWAHANA, H.; KANAZAWA, H.; HAYASHI, K.; SEKIYA, T. 1989. Detection of polymorphism of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proceedings of the National Academy of Sciences of the United States of America*. **86**:2766-2770.
- OSTERLUND, M.T.; PATERSON, A.H. 2002. Applied plant genomics: the secret is integration. *Current Opinion in Plant Biology*. **5**:141-145.
- OTERO, J.R. 1941. Notas de uma viagem aos campos do sul do Mato Grosso. Rio de Janeiro, Ministério da Agricultura.
- PARAN, I; MICHELMORE, R.W. 1993. Development of reliable PCR-based markers linked to downy mildew resistant genes in lettuce. *Theor. Appl. Genet.* **85**:985-993.
- PASTEUR, N.; PASTEUR, G.; BONHOMME, F.; CATALAN, J.; BRITTON-DAVIDIAN J. 1988. Practical isozyme genetics. Ellis Horwood, Chichester, UK.
- PATERSON, A.H.; TANKSLEY, S.D.; SORRELS, M.E. 1991. DNA markers in plant improvement. *Advances in Agronomy*, **46**:39-90.
- POWELL, W.; MACHRAY, G.C.; PROVAN, J. 1996a. Polymorphism revealed by simple sequence repeats. *Trends in Plant Science*, **1**:215–222.
- POWELL, W.; MORGANTE, M.; ANDRE, C.; HANAFEY, M.; VOGEL, J.; TINGEY, S.; RAFALSKI, A. 1996b. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Molecular Breeding*, **2**:225–238.
- RAFALSKI, J.A.; TINGEY, S.V. 1993. Genetic diagnostics in plant breeding: RAPDs, microsatellites and machines. *Trends Genet*, **9**:275-280.
- RAFALSKI, J.A.; VOGEL, J.M.; MORGANTE, M.; POWELL, W.; ANDRE, C.; TINGEY, S.V. 1996. Generating new DNA markers in plants. In *Non-mammalian genomic analysis: A practical guide*. Edited by B. Birren and E. Lai. Academic Press, New York.
- RAINA, S.N.; MUKAY, Y. 1999. Genomic in situ hybridization in *Arachis (Fabaceae)* identifies the diploid wild progenitors of cultivated (*A. hypogaea*) and related wild (*A. monticola*) peanut species. P1. *Syst. And Evol*, **214**:251-262.
- ROGSTAD, S.H.; PATTON, J.C.; SCHAAL, B.A. 1988. M13 repeat probe detects DNA minisatellite-like sequences in gymnosperms and angiosperms. *Proceedings of the National Academy of Science USA* **85**:9176–9178.
- SAIKI, R.K.; GELFAND, D.H.; STOEFFEL, S.; SHCARF, S.J.; HIGUCHI, G.; HORN, G.T.; MULLIS, K.B.; ERLICH, H.A. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, **239**:487-491.
- SAMBROOK, J.; RUSSELL, D.W. 2001. *Molecular Cloning: a laboratory manual*. 3rd ed. New York: Cold Spring Harbor Laboratory Press.
- SAN MIGUEL, P.; GAUT, B.S.; TIKHONOV, A.; NAKAJIMA, Y.; BENNETZEN, J.L. 1998. The paleontology of intergene retrotransposons of maize. *Nature Genet.* **20**: 43–45.
- SANTOS, C.C.M. DOS; LOPES, M. DO R.V. E KOSSEKI, S.Y.; 2001. Ocorrência de aflatoxinas em amendoim e produtos de amendoim comercializados na região de São José do Rio Preto/SP. *Rev. Inst. Adolfo Lutz*. **60(2)**:153-157.
- SCHLÖTTERER, C.; TAUTZ, D. 1992. Slippage synthesis of simple sequence DNA. *Nucleic Acids Research*, **20**:211-215.

- SCHUG, M.D.; WETTERSTRAND, K.A.; GAUDETTE, M.S. 1998. The distribution and frequency of microsatellite loci in *Drosophila melanogaster*. *Molecular Ecology*, **7**:57–70.
- SHOLAR, R.E.; MOZINGO, R.W.; BEASLEY, J.P. 1995. Peanut cultural practices. Pages 354-382 In: H. E. Pattee and T. H. Stalker (eds.) *Advances in Peanut Science. Am. Peanut Res. and Educ. Soc.*, Stillwater, OK.
- SINGH, A.K. 1986. Utilization of wild relatives in genetic improvement of *Arachis hypogaea* L. Part 8. Synthetic amphidiploids and their importance in interspecific breeding. *Theor. Appl. Genet.*, **72**:433-439.
- SINGH, A.K. 1994. Groundnut, *Arachis hypogaea* (Leguminosae-Papilionoideae). In: Sartt, J. Simmonds, N. W.. ed. *Evolutions of crop plants*, 2. ed. London, Longman, chapter 48, p.246-250.
- SMART, J.; GREGORY, W.C.; GREGORY, M.P. 1978. The genomes of *Arachis hypogaea*. 1. Cytogenetic studies of putative genome donors. *Euphítica*, **27**:665- 675.
- SOLTIS, D.E.; SOLTIS, P.S. 1989. *Isozymes in plant biology*. Dioscorides. Portland, OR.
- SOUTHERN, E.M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology*, **98**:503-217.
- SOUZA, A.P. 2001. Biologia molecular aplicada ao melhoramento. Em: Plantas: Recursos Genéticos & Molhoramento. Nass, L.L.; Valois, A.C.C.; Melo, I. S de; Valadares-Ingliš, M.C. Fundação MT-Rondonópolis.940-965.
- STALKER, H. T., T. G. PHILLIPS, J. P. MURPHY, AND T. M. JONES. 1994. Diversity of isozyme patterns in *Arachis* species. *Theor. Appl. Genet.* **87**:746–755.
- STALKER, H.T. 1991. A new species in section *Arachis* with D genome. *Am. J. Bot.* **78**:630-637.
- STALKER, H.T. 1992. Utilizing *Arachis* Germplasm Resources. p. 281–295. In: Groundnut-a global perspective. Proc. Intern. Workshop. November 25–29, 1991, Int. Crops Res. Inst. Semi-Arid Tropics. Patancheru, A. P., India.
- STALKER, H.T.; MOSS, J.P. 1987. Speciation, cytogenetics, and utilization of *Arachis* species. *Adv. Agronomy*. **41**:1–40.
- STALKER, H.T.; MOZINGO, L.G. 2001. Molecular markers of *Arachis* and marker-assisted selection. *Peanut Science*, **28**:117-123.
- STALKER, H.T.; SIMPSON, C.E. 1995. Genetic resources in *Arachis*. p. 14–53. In: H. E. Pattee and H. T. Stalker (eds.), *Advances in peanut science. Am. Peanut Res. Educ. Soc.*, Stillwater, OK.
- STALLINGS, R.S. 1992. CpG suppression in vertebrate genomes does not account for the rarity of (CpG) microsatellite repeats. *Genomics*, **17**:890-891.
- STEPHENSON, P., G. BRYAN, J. KIRBY, A. COLLINS, and K. DEVOS *et al.*, 1998 Fifty new microsatellite loci for the wheat genetic map. *Theor. Appl. Genet.* **97**:946-949.
- STOESSER, G.; BAKER, W.; VAN DEN BROEK, A.; CAMON, E.; GARCIA-PASTOR, M.; KANZ, C.; KULIKOVA, T.; LEINONEN, R.; LIN, Q.; LOMBARD, V.; LOPEZ, R.; REDASCHI, N.; STOEHRE, P.; TULI, M.A.; TZOUVARA, K.; VAUGHAN, R. 2002. The EMBL Nucleotide Sequence Database. *Nucleic Acids Res.*, **30**:21-26
- SUBRAHMANYAM, P.; GHANEKAR, A.M.; NOLT, B.L.; REDDY, D.V.R.; McDONALD, D. 1985a. Resistance to groundnut diseases in wild *Arachis*. In: ICRISAT. *Proceedings of an International Workshop on Cytogenetics of Arachis*. Patancheru., p. 49-55
- SUBRAHMANYAM, P.; SMITH, D.H.; SIMPSON, C.E. 1985b. Resistance to *Didymella arachidicola* in wild *Arachis* sp. *Oléagineux*, **40**:553-556.
- TACHIDA, H.; IIZUKA, M. 1992. Persistence of repeated sequences that evolve by replication slippage. *Genetics*, **131**:471–478.
- TANAKA, K.; TSUMURA, Y.; NAKAMURA, T. 1999. Development and polymorphism of microsatellite markers for *Fagus crenata* and the closely related species, *F. japonica*. *Theor Appl Genet.* **99**:11-15.

- TANKSLEY, S.D.; ORTON, T.J. 1983. Isozymes in plant genetics and breeding. Part A *Elsevier*, New York. 516 pp
- TANKSLEY, S.D; YOUNG, N.D.; PATERSON, A.H.; BONIERBLE, M.W. 1989. RFLP mapping in plant breeding: New tools for an old science. *Bio/Technology*. **7**:257-263
- TARAMINO, G.; TINGEY, S.V. 1996. Simple sequence repeats for germplasm analysis and mapping in maize. *Genome*, **39**:277-287.
- TARTOF, K.D.; HOBBS, C.A. 1987. Improved media for growing plasmid and cosmid clones. *Focus*, **9**:12.
- TATENO, Y.; IMANISHI, T.; MIYAZAKI, S.; FUKAMI-KOBAYASHI, K.; SAITOU, N.; SUGAWARA, H.; GOJOBORI, T. 2002. DNA Data Bank of Japan (DDBJ) for genome scale research in life science. *Nucleic Acids Res.*, **30**:27-30.
- TAUTZ, D. 1989. Hypervariability of simple sequence as a general source for polymorphic DNA markers. *Nucleic Acids Research*, **17**:6463-6471.
- TAUTZ, et al 1986 – livro microsatelites
- TÓTH, G.; GÁSPÁRI, Z.; JURKA, J. 2000. Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Research*, **10**: 967–981.
- VALLS, J.E.M.; SIMPSON, C.E. 1994. Taxonomy, natural distribution, and attributes of *Arachis*. In: Kerridge, P.C. & Hardy, B. ed. *Biology and Agronomy of Forage Arachis*. Cali, CIAT, chapter 1, p.1-18.
- VALLS, J.E.M.; SIMPSON, C.E. 1997. Novas espécies de *Arachis L.* - IN: Simpósio Latino Americano de Recursos Genéticos, 1. Programas e Resumos, Campinas. p.27.
- VAVILOV, N.I. 1940. *The News Systematics*. Oxford: Clarendon, p. 549-566.
- VIRK, P.S.; FORD-LLOYD, B.; NEWBURY, H.J. 1998. Mapping AFLP markers associated with subspecific differentiation of *Oryza sativa* (rice) and an investigation of segregation distortion. *Heredity*, **81**:613-620.
- WANG, Z.; WEBER, J.L.; ZHONG, G.; TANKSLEY, S.D. 1994 Survey of plant short tandem DNA repeats. *Theoretical and Applied Genetics*, **88**:1–6.
- WATSON, J.D.; HOPKINS, N.H.; ROBERTS, J.W.; STEITZ, J.A.; WEINER, A.M. 1987. *Molecular Biology of the Gene*. 4ª edição. Benjamin/ Cummings Pub., Menlo Park, CA (USA).
- WEBER, J.L.; MAY, P.E. 1989. Abundant class of human DNA polymorphism which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.*, **44**:273-283.
- WELSH, J.; McCLELLAND, M. 1990. Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Research*, **18(24)**:7213-7218.
- WENBURG, J.K.; OLSEN, J.B.; BENTZEN, P. 1996. Multiplexed systems of microsatellites for genetic analysis in coastal cutthroat trout (*Oncorhynchus clarki clarki*) and steelhead (*O. Mykiss*). *Mol. Marinebio. Biotech.*, **5**:273-283.
- WESTMAN, A.L.; KRESOVICH, S. 1998. The potential for cross-taxa simple-sequence repeat (SSR) amplification between *Arabidopsis thaliana* L. and crop brassicas. *Theoretical and Applied Genetics*, **96**: 272–281.
- WHEELER, D.L; CHURCH, D.M.; LASH, A.E.; LEIPE, D.D.; MADDEN, T.L.; PONTIUS, J.U.; SCHULER, G.D.; SCHRIML, L.M.; TATUSOVA, T.A.; WAGNER, L.; RAPP, B.A. 2002. Database resources of the National Center for Biotechnology Information: 2002 update. *Nucleic Acids Res.* **30(1)**:13-6.
- WHITE, G.; POWELL, W. 1997. Isolation and characterization of microsatellite loci in *Swietenia humilis* (Meliaceae): an endangered tropical hardwood species. *Molecular Ecology*, **6**:851-860.
- WILLIAMS, J.; KUBILIEK, A.; LIVAK, J.; RAFALSKI, J.; TINGEY, S. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research*, **18(22)**:6531-6535.

- WREN, J.D.; FORGACS, E.; FONDON, J.W. 2000 Repeat polymorphisms within gene regions: Phenotypic and evolutionary implications. *American Journal of Human Genetics*, **67**: 345–356.
- WYMAN, A.R.; WHITE, R. 1980. A highly polymorphic locus in human DNA. *Proc Natl Acad Sci USA*, **77**: 6754-6758.
- WYNNE, J.C.; COFFELT, T.A. 1982. Genetics of *Arachis hypogaea* L. In: H.E. Pattee and C.T. Young (eds), Peanut Science and Technology, *Am. Peanut Res. Educ. Soc., Inc.*, Yokum, Tx, p.50-94.
- ZABEAU, M.; VOS, P. 1993. Selective restriction fragment length amplification: a general method for DNA fingerprinting. In: **European Patent Application EP 534858A1**.
- ZAMSKI, E.; M. ZIV. 1976. Pod formation and its geotropic orientation in the peanut, *Arachis hypogaea* L., in relation to light and mechanical stimulus. *Ann. Bot.* **40**:631–636.
- ZIETKIEWICZ, E.; RAFALSKI, A.; LABUDA, D. 1994. Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. *Genomics* **20**: 176–183.

Documentos e Fontes em Meios Eletrônicos.

- CLARKE BRUNT'S HOME PAGE. **Plants of interest**. Peanut (*Arachis hypogaea*). Disponível em: <<http://www.viridis.net>>. Acesso em: 25 nov. 2002
- BORÉM, A.; FONTES, E.G.; FONTES, E.B.; OTONI, W. **Entendendo a Biotecnologia**. 2000, CD-ROM.
- ENCHANTED LEARNING. PLANTS. **Peanut Plant Anatomy**. Disponível em: <<http://www.enchantedlearning.com/subjects/plants>>. Acesso em: 02 fev. 2003.
- MISSOURI BOTANICAL GARDEN LIBRARY, **Rare Books from MBG Library**. Disponível em: <<http://ridgwaydb.mobot.org/mobot/rarebooks/>>. Acesso em: 08 mar. 2003.
- NATIONAL AGRICULTURAL STATISTICS SERVICE (USDA)2001-2002 **Statistical Highlights of U.S. Agriculture**. Disponível em: <<http://www.usda.gov>>. Acesso em 10 dez. 2002.
- OWENS, S. **Center for Electron Optics**, MSU – Disponível em: <<http://www.commtechlab.msu.edu/sites/dlc-me/zoo/zah0700.html>>. Acesso em 02 fev. 2003.
- ROZEN, S.; SKALETSKY, H.J. 1996. **Primer 3**, Disponível em: <http://www.genome.wi.mit.edu/genome_software/other/primer3.html>. Acesso em 10 nov. 2002.

Anexos

Anexo 1

1 - Protocolo de PCR –Ancorado

Cada placa de PCR comportou 16 clones, que foram amplificados com uma combinação de 4 primers diferentes dando um total de 80 amostras. A placa Ah1TC4 era composta de 96 clones gerando ao final da etapa 480 amostras. Os primers utilizados foram o T7; SP6; Primer Repetição AG e Primer Repetição CT, que quando combinados dois a dois entre si, originaram 5 diferentes combinações, cada uma com concentração de 5µM.(Figura 28 e Tabela 18)

	2	3	4	5	6	7	8	9	10	11	12
Clone 1											
Clone 2											
Clone 3											
Clone 4											
Clone 5											
Clone 6											
Clone 7											
Clone 8											
Clone 9											
Clone 10											
Clone 11											
Clone 12											
Clone 13											
Clone 14											
Clone 15											
Clone 16											
Primer T7 e SP6											
Primer T7 e Primer Repetição AG											
Primer SP6 e Primer Repetição CT											
Primer T7 e Primer Repetição CT											
Primer SP6 e Primer Repetição AG											

Formato da Placa de PCR

Figura 28 . Formato da aplicação dos clones e das respectivas combinações dos primer na placa de PCR.

Tabela 18. Master Mix (quantidade de reagentes necessárias para uma placa de PCR- ancorado.)

Reagentes	Quantidade
H ₂ O	750 µL
Tampão 10X	135 µL
Taq. polymerase	5,5 µL
MgCl ₂ 25mM	27 µL
dNTP (2mM)	135 µL
Total	1053 µL

O *master mix* foi dividido em 5 alíquotas diferentes, cada uma com 210,6 µL, sendo em seguida acrescentado 5,5 µL de cada combinação de primer em uma alíquota diferente. Em cada poço da placa de PCR foi acrescentado 10 µL do *master mix* com o primer específico mais 2µL de DNA de cada clone. As amplificações foram feitas utilizando um termociclador nas seguintes condições: desnaturação inicial de 94°C por 5 minutos; 30

ciclos com etapa de desnaturação (94°C por 1 minuto); ligação (58°C por 1 minuto) e de extensão (72°C por 1 minuto); extensão final de 72°C por 7 minutos.

As amplificações foram visualizadas em gel de agarose 1,5% corado com brometo de etídio (1,5 µg.mL⁻¹).

2 - Protocolo de Hibridização Membrana.

A membrana de náilon foi cortada no mesmo comprimento da placa de PCR, sobre ela foi colocado 0,5 µL do DNA plasmidial de cada clone pertencente à placa Ah1TC4. A disposição de cada clones na membrana permaneceu a mesma da placa da biblioteca. Após alguns minutos, os DNAs plasmidiais dos clones foram fixados na membrana com luz ultravioleta pelo aparelho cross-linker. As etapas seguintes são as mesmas que estão descritas no item 4.4.2 – Análise da hibridização do DNA/Sonda pela Técnica *Southern Blot* da parte de Metodologia desta dissertação.

Development of microsatellite markers in peanut (*Arachis hypogaea* L.).

Lélia C. T. Leoi, David J. Bertioli

Pós Graduação em Ciências Genômicas e Biotecnologia, Universidade Católica de Brasília.

Abstract. Microsatellites are widely used as genetic markers because they are co-dominant, multiallelic, easily scored and highly polymorphic. A major drawback of microsatellite markers is the time and cost required to characterize them. We have isolated and sequenced 107 microsatellite-containing clones from one small insert library of *Arachis hypogaea* enriched for (AG/TC) repeats. Primer pairs were designed for 79 microsatellite loci.

Key Words: Simple sequence repeats (SSR) · Microsatellite · Molecular Markers · Enrichment · *Arachis hypogaea*

Peanut is widely used as an oilseed crop around the world and as a direct source of human food. Several species of peanut have been cultivated for their edible seeds, but only *Arachis hypogaea* L. has been domesticated and widely distributed, mainly in countries as China, USA and India that are the biggest producers in the world. In spite of the existence of substantial diversity among cultivated peanut genotypes for various morphological, physiological and agronomic traits, very little DNA variation had been detected by using protein or DNA-based markers (Kochert *et al.*, 1991; Halward, *et al.*, 1991; Paik-Ro *et al.*, 1992; Stalker *et al.*, 1994). Isozyme analyses of cultivated peanut have shown little variation (Grieshammer and Wynne 1990), analyses of seed storage proteins have shown that variation exists among species of section *Arachis* (Singh *et al.* 1991a; Bianchi-Hall *et al.* 1993, 1994), but proteins are not useful for descriptions at the species level. Restriction fragment length polymorphisms (RFLP) represented the first marker system that has a sufficiently large number of polymorphisms that could be used to create linkage maps and to implement indirect selection strategies. In *A. hypogaea*, little molecular variation has been detected by using RFLP technologies (Kochert *et al.* 1991). Hopkins and collaborators (1999) reported six polymorphic SSRs in *A. hypogaea* with the number of fragments amplified per SSR ranging from 2 to 14. Few microsatellites have been identified in peanut and detailed information about their abundance and distribution in the genome are not available.

Microsatellites or simple sequence repeats (SSR) are tandemly repeated motifs of one to six bases found in every organism investigated so far (Tautz & Rens, 1984; Weber

PRIMER NOTES

& May, 1989; Weber, 1990). They are present in both coding and non-coding regions although they are found more frequently within non-coding regions. The functional significance of such sequences is still unknown although when found in exons in the human genome, they are sometimes associated with diseases (Ross, 1995). Microsatellites show higher levels of polymorphism than other genetic markers (Rafalski *et al.*, 1996). The origin of such polymorphism is still under debate though it appears most likely to be due to slippage events during DNA replication (Schlötterer & Tautz, 1992). The high informative content due to high polymorphism, co-dominance and rapid analysis by PCR make microsatellites well poised for exploitation in plant breeding applications such as linkage analysis, agronomic trait selection, germplasm assessment, varietal identification, as well as population genetic applications in both natural and breeding populations (Brown *et al.*, 1996; Maguire *et al.*, 2000).

There are a number of ways to obtain microsatellite markers. Approaches for the isolation of new microsatellite loci are generally based on searching known sequences in public databases, or the identification of clones containing microsatellite loci by screening genomic libraries (Brown *et al.*, 1996, Billotte *et al.*, 2001). Enrichment can be realized by hybridizing SSR-containing fragments with biotin-labeled probes that are captured by magnetic beads coated with streptavidin. This methodology can increase the efficiency in obtaining of microsatellites from 50% to 90% (Powell *et al.*, 1996; Edward *et al.*, 1996).

The objective of this study was to identify microsatellites from the cultivated peanut genome; these microsatellites will be used for mapping and characterization of the *Arachis* genome.

Genomic DNA from *A. hypogaea* leaf was isolated using a CTAB extraction protocol (Doyle & Doyle). Approximately 30µg of DNA was digested with *Sau* 3AI (Gibco) and size selected (200 – 950 pb) by 1,5% agarose gel electrophoresis. Microsatellite enriched genomic library was carried out following previously developed and optimized protocols (Rafalski *et al.* 1996, Taramino & Tingey, 1996; White & Powell, 1997 e Brondani *et al.*, 1998). Size selected fragments were ligated to adaptors (SA *Sau* 21 mer: 5'-CAGCCTAGAGCCGAATTCACC 3' and LA *Sau* 25 mer: 5'-pGATCGGTGAATTCGGCTCTAGGCTG 3') and hybridized with 5'biotinylated, 3'aminated (AG)₁₂ oligonucleotides bound to streptavidin-coated magnetic beads (Dynabeads®Streptavidin, Dynal). The hybridization was followed by several washing

PRIMER NOTES

steps in 2xSSPE (0.3M NaCl, 20mM NaH₂PO₄H₂O, 2mM EDTA, pH 7.4) and the enriched DNA was then eluted in water. DNA fragments were polymerase chain reaction (PCR) amplified, using adaptor SA *Sau* as a primer, and ligated into pGEMT vector (Promega). The ligation was performed overnight at 4°C. Fifteen ng of the ligation was transformed into 40µL *Escherichia coli* XL1-blue competent cells following Sambrook *et al.* (2001). A total of 576 positives colonies were screened based on the method blue/white (IPTG and X-Gal). They were sequenced on an Applied Biosystems 377 sequencer using DYEnamic ET Terminator Cycle Sequencing Kit (Amersham Pharmacia Biotech) with T7 and SP6 primers. One hundred and thirty and nine positive clones had had simple sequence repeats and seventy seven were adequate for design of primers (Table 1). Primer pairs complementary to sequences flanking the repeat element were designed using the Primer 3 software http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi. All primers were designed for 60°C annealing temperature and a total amplicon size between 100 and 300 bp to simplify multiplex PCR and rapid gel separation.

Of the 77 clones sequenced, 58,44% (45 clones) contained simple or compound microsatellites with 12 or more repeats. The size distributions of all 45 clones which contain repeats are shown up in figure 1. The average clone contained 26 repeats units. The size distribution is not necessarily indicative of the size of all genomic TC repeats because of the selection imposed by the hybridization procedure. The distribution in repeat length was comparable to those isolated in similar study of *Copaifera langsdorffii* (Leguminosae) (Ciampi *et al.*, 1999).

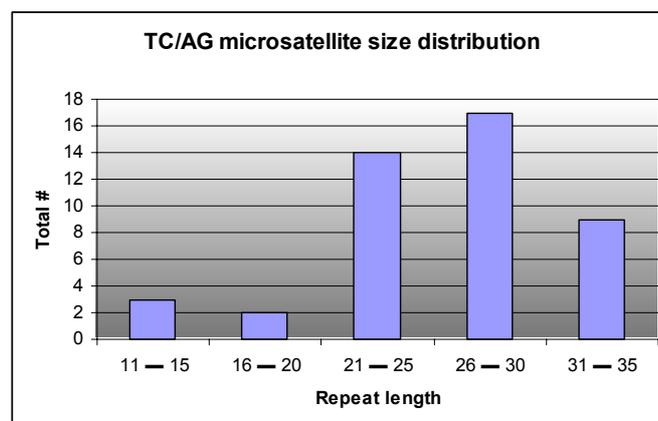


Figure 1. Distribution of microsatellite loci developed by the hybrid capture method. The loci were grouped by repeat length. The average repeat length was 26 bp.

PRIMER NOTES

Table 1. Primer names, repeat motifs and type.

Primer	Motif	Type
Ah1TC0A01	(AG) ₂₉	PERFECT
Ah1TC1A01	(TC) ₂₆	PERFECT
Ah1TC1A02	(TC) ₃₅	PERFECT
Ah1TC1A08	(GA) ₃₀	PERFECT
Ah1TC1B01	(CA) ₂ (TC) ₁₁ (CA) ₉ 100-122 (TC) ₉ 140-190 (TC) ₂₁	IMPERFECT
Ah1TC1B02	(CT) ₂₅	PERFECT
Ah1TC1D01	(AG) ₂₄	PERFECT
Ah1TC1D02	(TC) ₃₁	PERFECT
Ah1TC1D12	(TC) ₁₄	PERFECT
Ah1TC1E01	(CA) ₄ (GA) ₂₉	COMPOUND
Ah1TC1E05	(TC) ₃₀	PERFECT
Ah1TC1E06	(TC) ₂₃	PERFECT
Ah1TC1G04	(CTT) ₄ (CT) ₃₃	COMPOUND
Ah1TC1H04	(A) ₁₄ 58-344 (TC) ₁₉ 383-577 (CTT) ₈	PERFECT
Ah1TC2A02	(TC) ₃₂	PERFECT
Ah1TC2A11	(TC) ₂₈ (GT) ₅	COMPOUND
Ah1TC2B01	(TC) ₁₃	PERFECT
Ah1TC2B05	(AG) ₂₄	PERFECT
Ah1TC2B09	(TC) ₂₇	PERFECT
Ah1TC2C03	(TC) ₃₉	PERFECT
Ah1TC2C07	(TC) ₂₃	PERFECT
Ah1TC2C11	(TC) ₂₄	PERFECT
Ah1TC2C12	(CA) ₂₇ (TC) ₃	COMPOUND
Ah1TC2D06	(AG) ₅ (A) ₃ (T) ₄ (AG) ₅ (A) ₅ 59-181(GA) ₃₁	COMPOUND E PERFECT
Ah1TC2D08	(TC) ₂₆	PERFECT
Ah1TC2E05	(TC) ₂₆	PERFECT
Ah1TC2E11	(TTA) ₆	PERFECT
Ah1TC2G05	(GA) ₃₅	PERFECT
Ah1TC3A10	(CT) ₃ GAGTA (CT) ₂₅	IMPERFECT
Ah1TC3B04	(TC) ₂₄	PERFECT
Ah1TC3B05	(AG) ₂₆	PERFECT
Ah1TC3D04	(TC) ₃₄	PERFECT
Ah1TC3E02	(CT) ₂₆ (CA) ₇ (CG) ₃ CT CGTG(CA) ₅	IMPERFECT
Ah1TC3E03	(A) ₁₃	PERFECT
Ah1TC3E05	(GA) ₂₅ (GT) ₃	COMPOUND
Ah1TC3G01	(TC) ₁₄ (T) ₅ AAC(CT) ₂ GC(CT) ₁₅	IMPERFECT
Ah1TC3G03	(TCC) ₁₀	PERFECT
Ah1TC3G05	(CTT) ₁₆	PERFECT
Ah1TC3H01	(CT) ₃ (T) ₁₄	COMPOUND
Ah1TC3H02	(CT) ₂₈ GTGCTCTT (CG) ₂ (CT) ₆	IMPERFECT
Ah1TC3H04	(A) ₁₀	PERFECT
Ah1TC6H03	(AG) ₂₂	PERFECT

Primer	Motif	Type
Ah1TC3H07	(GA) ₂₃ 282-377 (TTC) ₄ 390-398 (TG) ₂ (TC) ₃ (CA) ₃ GT(CA) ₄	PERFECT E IMPERFECT
Ah1TC4A02	(GA) ₂₆	PERFECT
Ah1TC4B01	(TC) ₃₁	PERFECT
Ah1TC4B02	(TC) ₂₇	PERFECT
Ah1TC4B07	(A) ₉ G(A) ₄ G(A) ₅ TAGCG(A) ₂₄	IMPERFECT
Ah1TC4C08	(AC) ₃ (TC) ₃ 154-197 (GA) ₅	COMPOUND
Ah1TC4C11	(CT) ₃₅	PERFECT
Ah1TC4D02	(TC) ₂₁	PERFECT
Ah1TC4D09	(CA) ₂ (CCA) ₂ (CT) ₁₈	COMPOUND
Ah1TC4D12	(G) ₃ (AG) ₅ TAGTGA(GT) ₃	IMPERFECT
Ah1TC4E08	(TC) ₂₈	PERFECT
Ah1TC4E09	(TC) ₂₂	PERFECT
Ah1TC4E10	(T) ₂ C(T) ₄ C(T) ₅ C(T) ₂ C(T) ₁₂ A(CT) ₁₅	COMPOUND
Ah1TC4F01	(C) ₄ A(C) ₉	COMPOUND
Ah1TC4F02	(C) ₁₄	PERFECT
Ah1TC4F04	(CT) ₂₇ (CA) ₂	COMPOUND
Ah1TC4F08	(TC) ₃₀	PERFECT
Ah1TC4F10	(TTC) ₃₃	PERFECT
Ah1TC4F12	(CT) ₂₃	PERFECT
Ah1TC4G02	(CT) ₂₇	PERFECT
Ah1TC4G05	(CT) ₃₄	PERFECT
Ah1TC4G06	(CT) ₁₄ (GA) ₂	COMPOUND
Ah1TC4G10	(GA) ₂₂ GCGT(GA) ₆ GT(GA) ₄	IMPERFECT
Ah1TC4H02	(CT) ₁₇ (A) ₂ (T) ₅ CA(C) ₂ (T) ₂ C(CT) ₅ 164-185 (CT) ₅ AACAT(TC) ₂ A(TC) ₇	IMPERFECT
Ah1TC4H07	(CT) ₃₃ (A) ₂ (T) ₄ G(T) ₆ (G) ₁₀	IMPERFECT
Ah1TC4H11	(A) ₉ (TC) ₃ (T) ₄ G(T) ₄ CG(A) ₁₆	IMPERFECT
Ah1TC5A06	(CT) ₂₈	PERFECT
Ah1TC5A07	(GA) ₁₆ TAT(GA) ₈	IMPERFECT
Ah1TC5B06	(TG) ₂ (A) ₅ TG(TA) ₃ (TG) ₂ (A) ₄ TGTT(A) ₅ (TA) ₁₀	IMPERFECT
Ah1TC5C05	(A) ₇ TTG(T) ₃ C(T) ₅ (C) ₇ T(CA) ₂ (CT) ₄ 280-347 (A) ₄ (GA) ₇ 367-441 (CA) ₃ (GA) ₁₃	IMPERFECT E COMPOUND
Ah1TC5D01	(CT) ₂₀	PERFECT
Ah1TC5D06	(GA) ₂₁	PERFECT
Ah1TC5H11	(TC) ₃ CATT(A) ₃ (CA) ₅	IMPERFECT
Ah1TC6C07	(A) ₁₃	PERFECT
Ah1TC6E01	(CT) ₂₄	PERFECT
Ah1TC6G09	(CT) ₁₉	PERFECT

PRIMER NOTES

The results provided by the present study highlight a simple and reliable way in obtaining microsatellite markers from the cultivated peanut. A significance of the results is that development of microsatellite markers is important not only for peanut researchers but also for peanut industry as the paucity of DNA markers in cultivated peanut has so far precluded detailed genetic research on this crop.

PRIMER NOTES

Reference:

- BILLOTTE, N.; RISTERUCCI, A.M.; BARCELOS, E.; NOYER, J.L.; AMBLARD, P.; BAURENS, F.C. 2001. Development, characterization, and across-taxa utility of oil palm (*Elaeis guineensis* Jacq.) microsatellite markers. *Genome*, **44**:413-425
- BROWN, S.M.; SZEWC-MCFADDEN, A.K.; KRESOVICH, S. 1996. Development and application of simple sequence repeat (SSR) loci for plant genome analysis *In: Methods of genome analysis in plants (Edited by: Jauhar PP) CRC Press, Inc. Boca Raton New York London Tokyo*, 147-159.
- CIAMPI, A.; BRONDANI, R.; GRATTAPAGLIA, D. 1999. Otimização de sistemas fluorescentes de genotipagem multiloco e desenvolvimento de primers microssatélites para *Copaifera langsdorffii* (Copaíba) – leguminosae-Caesalpinoideae. Boletim de Pesquisa nº 15. Embrapa- Brasília – DF.
- EDWARDS, K.J.; BARKER, J.H.A.; DALY, A.; JONES, C.; KARP, A. 1996. Microsatellite libraries enriched for several microsatellite sequences in plants. *Biotechniques*, **20**:758-760
- HOPKINS, M.S.; CASA, A.M.; WANG, T.; MITCHELL, S.E.; DEAN, R.E.; KOCHERT, G.D.; KRESOVICH, S. 1999. Discovery and characterization of polymorphic simple sequence repeats (SSRs) in cultivated peanut (*Arachis hypogaea* L) *Crop Sci*, **39**:1243-1247
- MAGUIRE, T.L.; EDWARDS, K.J.; SAENGER, P.; HENRY, R. 2000. Characterisation and analysis of microsatellite loci in a mangrove species *Avicennia marina* (Forsk.) Vierh. (Avicenniaceae). *Theoretical and Applied Genetics*. **101**:279–285.
- POWELL, W.; MACHRAY, G.C.; PROVAN, J. 1996. Polymorphism revealed by simple sequence repeats. *Trends in Plant Science* **1**:215–222.
- RAFALSKI, J.A.; VOGEL, J.M.; MORGANTE, M.; POWELL, W.; ANDRE, C.; TINGEY, S.V. 1996. Generating new DNA markers in plants. *In Non-mammalian genomic analysis: A practical guide. Edited by B. Birren and E. Lai. Academic Press, New York.*
- ROSS, C.A. 1995. When more is less: pathogenesis of glutamine repeat neurodegenerative diseases. *Neuron*, **15**:1033-1041.
- SCHLÖTTERER, C.; TAUTZ, D. 1992. Slippage synthesis of simple sequence DNA. *Nucleic Acids Research*, **20**:211-215.
- WEBER, J.L.; MAY, P.E. 1989. Abundant class of human DNA polymorphism which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.*, **44**:273-283. Weber, J.L. (1990) Informativeness of human (dC-dA)n (dG-dT)n polymorphisms. *Genomics* **7**, 524–530.

Anexo 3

Primers SSR

KEYS (in order of precedence):

```
***** target
>>>>> left primer
<<<<<< right primer
```

Ah1TC1A01

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	147	20	60.03	50.00	4.00	2.00	
TCAACGCGACACAAGAAGTC							
RIGHT PRIMER	362	20	59.94	45.00	6.00	0.00	
GTCGGTAAATCCGACGAAAA							
SEQUENCE SIZE: 396							
INCLUDED REGION SIZE: 396							
PRODUCT SIZE: 216							

Ah1TC1A02

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	113	20	59.53	40.00	8.00	3.00	
GCAATTTGCACATTATCCGA							
RIGHT PRIMER	367	22	60.15	40.91	4.00	1.00	
CATGTTTCGGTTTCAAGTCTCAA							
SEQUENCE SIZE: 370							
INCLUDED REGION SIZE: 370							
PRODUCT SIZE: 255							

Ah1TC1B01

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	18	20	60.00	50.00	2.00	2.00	
CACAAACACATCAACCTCCG							
RIGHT PRIMER	307	20	59.88	60.00	2.00	0.00	
AGGAGCAGAGAGCAGAGTGG							
SEQUENCE SIZE: 315							
INCLUDED REGION SIZE: 315							
PRODUCT SIZE: 290							

Ah1TC1B01- 1

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	18	20	60.00	50.00	2.00	2.00	
CACAAACACATCAACCTCCG							
RIGHT PRIMER	171	20	60.73	45.00	5.00	3.00	
GGTTGAAAACCCGATTCACA							
SEQUENCE SIZE: 315							
INCLUDED REGION SIZE: 315							
PRODUCT SIZE: 154							

Anexo 3

Ah1TC1B01- 2

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	161	20	59.97	50.00	5.00	2.00	
GGTTTTCAACCCAGCAGGTA							
RIGHT PRIMER	307	20	59.88	60.00	2.00	0.00	
AGGAGCAGAGAGCAGAGTGG							
SEQUENCE SIZE: 315							
INCLUDED REGION SIZE: 315							
PRODUCT SIZE: 147							

Ah1TC1T1D12

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	68	20	59.87	50.00	2.00	0.00	
CCCTTTTCATTCTCCCTTTCC							
RIGHT PRIMER	236	21	59.86	47.62	4.00	0.00	
TTCTCCTGCACTAGGTTTCCA							
SEQUENCE SIZE: 315							
INCLUDED REGION SIZE: 315							
PRODUCT SIZE: 169							

Ah1TC1E05

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	402	20	57.76	45.00	3.00	1.00	
GAAGGATAAGCAATCGTCCA							
RIGHT PRIMER	628	20	59.99	45.00	3.00	0.00	
GGATGGGATTGAACATTTGG							
SEQUENCE SIZE: 778							
INCLUDED REGION SIZE: 778							
PRODUCT SIZE: 227							

Ah1TC1E06

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	253	20	60.18	50.00	6.00	2.00	
ACCGTTACGAACGCTTTGTC							
RIGHT PRIMER	393	20	59.53	55.00	4.00	0.00	
TCCCTCTCATACGACACCCT							
SEQUENCE SIZE: 403							
INCLUDED REGION SIZE: 403							
PRODUCT SIZE: 141							

Ah1TC1H04

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	23	22	57.28	31.82	8.00	3.00	
AAAATAACGTCGTTTTGGATTC							
RIGHT PRIMER	177	20	59.95	50.00	3.00	0.00	
AAAGGAGAGGCTGATGACGA							
SEQUENCE SIZE: 691							
INCLUDED REGION SIZE: 691							
PRODUCT SIZE: 155							

Anexo 3

Ah1TC1H04 (2)

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	243	20	60.39	40.00	3.00	0.00	
TGCTTACGAATGACGCAAAA							
RIGHT PRIMER	511	20	59.90	50.00	5.00	2.00	
AGTGCAGTGCCATCTTGTTG							
SEQUENCE SIZE: 691							
INCLUDED REGION SIZE: 691							
PRODUCT SIZE: 269							

Ah1TC1H04

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	526	21	59.19	52.38	6.00	1.00	
CTGTGCACCTTCTCACTCTCA							
RIGHT PRIMER	685	20	59.97	50.00	4.00	2.00	
AGTTGGAGTTCCCGAAGGTT							
SEQUENCE SIZE: 691							
INCLUDED REGION SIZE: 691							
PRODUCT SIZE: 160							

Ah1TC2B09

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	40	20	60.04	45.00	4.00	0.00	
GCCCATATCAAGCTCCAAAA							
RIGHT PRIMER	240	20	59.98	50.00	3.00	2.00	
TAGCCAGCGAAGGACTCAAT							
SEQUENCE SIZE: 318							
INCLUDED REGION SIZE: 318							
PRODUCT SIZE: 201							

Ah1TC2C11

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	174	20	59.87	50.00	4.00	2.00	
TCCTAATGCACTTGTGACGC							
RIGHT PRIMER	327	21	59.93	38.10	4.00	2.00	
GAACCTGGAATTTTTCATGA							
SEQUENCE SIZE: 369							
INCLUDED REGION SIZE: 369							
PRODUCT SIZE: 154							

Ah1TC2C12

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	40	24	58.89	33.33	1.00	1.00	
GGAGTTTGTGATTTGAGTTTGA							
RIGHT PRIMER	230	18	59.72	50.00	3.00	2.00	
CCCGCTATCCCCAAAAT							
SEQUENCE SIZE: 245							
INCLUDED REGION SIZE: 245							
PRODUCT SIZE: 191							

Anexo 3

Ah1TC2E11

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	116	20	60.15	45.00	4.00	0.00	
AAACTTGGACGTTGGCTTTG							
RIGHT PRIMER	309	20	60.01	50.00	2.00	1.00	
CAGGTTCGGGTTCACACTTT							
SEQUENCE SIZE: 436							
INCLUDED REGION SIZE: 436							
PRODUCT SIZE: 194							

Ah1TC3B04

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	21	20	60.14	55.00	5.00	3.00	
GAAGAAGAAGTCACTGCGGC							
RIGHT PRIMER	261	25	59.88	36.00	6.00	3.00	
AAGCTAGTTTCTGATTAAAGCACCA							
SEQUENCE SIZE: 288							
INCLUDED REGION SIZE: 288							
PRODUCT SIZE: 241							

Ah1TC3E03

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	40	20	60.20	45.00	6.00	2.00	
TAAGAAGCCGAATTCAACCG							
RIGHT PRIMER	277	22	59.81	45.45	4.00	0.00	
TCGTAGTGTGCAAGGTTTAGGA							
SEQUENCE SIZE: 303							
INCLUDED REGION SIZE: 303							
PRODUCT SIZE: 238							

Ah1TC3G05

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	1	20	59.18	55.00	4.00	3.00	
GATCCCAAGTCTCCAGAGGA							
RIGHT PRIMER	134	20	59.84	50.00	2.00	0.00	
AACACAAGGAGGCAGAGGA							
SEQUENCE SIZE: 287							
INCLUDED REGION SIZE: 287							
PRODUCT SIZE: 134							

Ah1TC3H01

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	48	20	59.52	60.00	5.00	3.00	
CTAGGATGGGTGACCTCCTG							
RIGHT PRIMER	230	20	59.87	50.00	4.00	3.00	
TAACACTATCCGCGAAGCCT							
SEQUENCE SIZE: 239							
INCLUDED REGION SIZE: 239							
PRODUCT SIZE: 183							

Anexo 3

Ah1TC4F04

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	3	20	59.25	60.00	4.00	1.00	
TCCCCACTACCTCGACTCTC							
RIGHT PRIMER	106	20	59.92	50.00	2.00	0.00	
GGTATTTGGGGAAGGTTGTT							
SEQUENCE SIZE: 645							
INCLUDED REGION SIZE: 645							
PRODUCT SIZE: 104							

Ah1TC4F12

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	1	20	58.72	45.00	4.00	0.00	
GATCTTTCCGCCATTTTCTC							
RIGHT PRIMER	220	20	59.64	50.00	3.00	1.00	
GGTGAATGACAGATGCTCCA							
SEQUENCE SIZE: 350							
INCLUDED REGION SIZE: 350							
PRODUCT SIZE: 220							

Ah1TC4G02

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	1	20	60.33	50.00	5.00	2.00	
GATCCAACGTGAATTGGGC							
RIGHT PRIMER	151	20	59.14	50.00	3.00	3.00	
CACACCAGCAACAAGGAATC							
SEQUENCE SIZE: 175							
INCLUDED REGION SIZE: 175							
PRODUCT SIZE: 151							

Ah1TC4G06

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	35	20	60.15	50.00	4.00	0.00	
ATTTACATTCCCTAGCCCC							
RIGHT PRIMER	273	22	59.61	40.91	4.00	0.00	
CATCGACTGACTTGAAAAATGG							
SEQUENCE SIZE: 292							
INCLUDED REGION SIZE: 292							
PRODUCT SIZE: 239							

Ah1TC4H11

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	54	20	60.13	45.00	5.00	3.00	
TTTGGACAACGACGGTTCA							
RIGHT PRIMER	250	20	60.36	50.00	4.00	3.00	
CCAGGCTTCTCATGCTTCAT							
SEQUENCE SIZE: 347							
INCLUDED REGION SIZE: 347							
PRODUCT SIZE: 197							

Anexo 3

Ah1TC5D01

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	20	20	59.39	50.00	3.00	0.00	
CATTGACCACTCACATCCGT							
RIGHT PRIMER	145	20	60.35	55.00	2.00	2.00	
GATGGGAGTGTGTATTCGGC							
SEQUENCE SIZE: 198							
INCLUDED REGION SIZE: 198							
PRODUCT SIZE: 126							

Ah1TC6C07

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	26	21	59.82	42.86	4.00	0.00	
TCCAACCACAAAAGAATTGTCC							
RIGHT PRIMER	95	25	57.54	32.00	5.00	2.00	
GACATTACTTACCCATTCAATTCAT							
SEQUENCE SIZE: 189							
INCLUDED REGION SIZE: 189							
PRODUCT SIZE: 70							

Ah1TC6E01

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	34	20	60.09	55.00	2.00	0.00	
CTCCCTCGCTTCTCTTTCT							
RIGHT PRIMER	204	20	59.89	45.00	4.00	0.00	
ACGCATTAACCACACACCAA							
SEQUENCE SIZE: 208							
INCLUDED REGION SIZE: 208							
PRODUCT SIZE: 171							

Ah1TC1E06

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	253	20	60.18	50.00	6.00	2.00	
ACCGTTACGAACGCTTTGTC							
RIGHT PRIMER	393	20	59.53	55.00	4.00	0.00	
TCCCTCTCATACGACACCCCT							
SEQUENCE SIZE: 403							
INCLUDED REGION SIZE: 403							
PRODUCT SIZE: 141							

Ah1TC1G04

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	16	20	60.14	50.00	3.00	1.00	
TGCTGTGAGAGAAATGGCAG							
RIGHT PRIMER	289	20	59.82	45.00	4.00	2.00	
GCGCATTCTTCGATTAAAGG							
SEQUENCE SIZE: 321							
INCLUDED REGION SIZE: 321							
PRODUCT SIZE: 274							

Anexo 3

Ah1TC1D12

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	49	20	59.82	50.00	4.00	1.00	
GCCAAAATTCTCAGTGCCTC							
RIGHT PRIMER	154	20	59.85	45.00	2.00	0.00	
AGTGAGAAGCGAACCCAAAA							
SEQUENCE SIZE: 315							
INCLUDED REGION SIZE: 315							
PRODUCT SIZE: 106							

Ah1TC2D08

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	58	19	59.65	57.89	3.00	3.00	
ATGTGGGGAGGTCGGTAAC							
RIGHT PRIMER	357	20	60.91	50.00	4.00	2.00	
TCACAGGTTTTGTGTGCTCG							
SEQUENCE SIZE: 439							
INCLUDED REGION SIZE: 439							
PRODUCT SIZE: 300							

Ah1TC2E05

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	96	20	60.06	45.00	6.00	2.00	
GAATTTATAAGCGTGGCGA							
RIGHT PRIMER	195	20	60.04	50.00	2.00	0.00	
CCATCCCTTCTTCCTTCACA							
SEQUENCE SIZE: 220							
INCLUDED REGION SIZE: 220							
PRODUCT SIZE: 100							

Ah1TC2C03

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	60	20	59.47	50.00	4.00	1.00	
AGACGTGAGTGCTTGGTTCA							
RIGHT PRIMER	272	20	59.98	55.00	6.00	2.00	
CAGCCTAGAGCCGAATTCAC							
SEQUENCE SIZE: 323							
INCLUDED REGION SIZE: 323							
PRODUCT SIZE: 213							

Ah1TC2C07

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	32	20	59.86	50.00	3.00	0.00	
CACCACACTCCCAAGGTTTT							
RIGHT PRIMER	196	20	59.99	50.00	4.00	2.00	
TCAAGAACGGCTCCAGAGTT							
SEQUENCE SIZE: 282							
INCLUDED REGION SIZE: 282							
PRODUCT SIZE: 165							

Ah1TC1D02

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	1	20	60.16	45.00	4.00	1.00	
GATCCAAAATCTCGCCTTGA							
RIGHT PRIMER	264	20	60.33	50.00	6.00	0.00	
GCTGCTCTGCACAACAAGAA							
SEQUENCE SIZE: 273							
INCLUDED REGION SIZE: 273							
PRODUCT SIZE: 264							

Anexo 3

Ah1TC3H02

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	4	20	59.92	50.00	4.00	2.00	
CTCTCCGCCATCCATGTAAT							
RIGHT PRIMER	288	20	60.04	55.00	6.00	2.00	
ATGGTGAGCTCGACGCTAGT							
SEQUENCE SIZE: 456							
INCLUDED REGION SIZE: 456							
PRODUCT SIZE: 285							

Ah1TC3A12= Ah1TC2B09

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	40	20	60.04	45.00	4.00	0.00	
GCCCATATCAAGCTCCAAAA							
RIGHT PRIMER	240	20	59.98	50.00	3.00	2.00	
TAGCCAGCGAAGGACTCAAT							
SEQUENCE SIZE: 466							
INCLUDED REGION SIZE: 466							
PRODUCT SIZE: 201							

Ah1TC3G03

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	10	20	60.13	50.00	7.00	3.00	
ATCTGCAGCCTCAAGCTGAT							
RIGHT PRIMER	242	20	59.65	55.00	4.00	0.00	
GCCGGTATGAGAGATTGGAG							
SEQUENCE SIZE: 454							
INCLUDED REGION SIZE: 454							
PRODUCT SIZE: 233							

Ah1TC4D02

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	333	20	60.16	50.00	4.00	0.00	
AAGTTGTTCCCGTTGCACTC							
RIGHT PRIMER	444	23	57.62	30.43	8.00	1.00	
AAAACACCATAAGGTGAATCAA							
SEQUENCE SIZE: 451							
INCLUDED REGION SIZE: 451							
PRODUCT SIZE: 112							

Ah1TC3D04

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	33	20	58.02	45.00	4.00	0.00	
TTTCGTCATTTTCAGCTCCTC							
RIGHT PRIMER	183	21	59.99	47.62	4.00	1.00	
TTCAGCCTAGAGCGGTATTCA							
SEQUENCE SIZE: 190							
INCLUDED REGION SIZE: 190							
PRODUCT SIZE: 151							

Anexo 3

Ah1TC4E08

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	134	20	59.64	55.00	3.00	1.00	
TTCATCCACCCTCCTGACT							
RIGHT PRIMER	344	20	60.14	55.00	3.00	2.00	
GATGCGTGAGTGGTCATACG							
SEQUENCE SIZE: 501							
INCLUDED REGION SIZE: 501							
PRODUCT SIZE: 211							

Ah1TC4F01

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	40	20	59.94	45.00	4.00	2.00	
GAACAACCGGGAGCAATTTA							
RIGHT PRIMER	293	25	59.93	44.00	8.00	3.00	
CGTCCAGTTCCTATAGAACCTATCA							
SEQUENCE SIZE: 296							
INCLUDED REGION SIZE: 296							
PRODUCT SIZE: 254							

Ah1TC4F08

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	46	20	58.35	40.00	6.00	0.00	
TGCAATTTCGACTTCTTGGTT							
RIGHT PRIMER	205	22	58.41	36.36	3.00	3.00	
TCATAAACACCGATTGACATTG							
SEQUENCE SIZE: 207							
INCLUDED REGION SIZE: 207							
PRODUCT SIZE: 160							

Ah1TC5H11

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	500	20	59.62	40.00	4.00	3.00	
AAAAATCGGACCATGATTGG							
RIGHT PRIMER	658	20	59.91	55.00	1.00	0.00	
TCTTCTCCTCCCTCACTCCA							
SEQUENCE SIZE: 690							
INCLUDED REGION SIZE: 690							
PRODUCT SIZE: 159							

Ah1TC1D02

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	1	20	60.16	45.00	4.00	1.00	
GATCCAAAATCTCGCCTTGA							
RIGHT PRIMER	264	20	60.33	50.00	6.00	0.00	
GCTGCTCTGCACAACAAGAA							
SEQUENCE SIZE: 274							
INCLUDED REGION SIZE: 274							
PRODUCT SIZE: 264							

Ah1TC1E01

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	11	20	59.77	55.00	3.00	3.00	
CAGCAAAGAGTCGTCAGTCG							
RIGHT PRIMER	228	23	58.55	34.78	6.00	2.00	
GAAAGTTCCTTGAGCAAATTCA							
SEQUENCE SIZE: 287							
INCLUDED REGION SIZE: 287							
PRODUCT SIZE: 218							

Anexo 3

5as2= Ah1TC0A01

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	52	20	58.85	45.00	4.00	0.00	
CAGCTCATTTTTTCACCTCCA							
RIGHT PRIMER	347	20	60.18	45.00	4.00	2.00	
CCATAACCCCAAAAATGCAG							
SEQUENCE SIZE: 408							
INCLUDED REGION SIZE: 408							
PRODUCT SIZE: 296							

Ah1TC2B01

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	18	20	59.72	45.00	4.00	0.00	
TTGCAGAAAAGGCAGAGACA							
RIGHT PRIMER	171	24	59.34	41.67	4.00	2.00	
GAAAGAAGCTAAGAAGGACCCATA							
SEQUENCE SIZE: 220							
INCLUDED REGION SIZE: 220							
PRODUCT SIZE: 154							

Ah1TC1D01

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	31	20	60.20	50.00	3.00	0.00	
TGCCAATCTCCTCTTCAACC							
RIGHT PRIMER	232	20	60.25	55.00	3.00	1.00	
TCAGGCAAGGGTTCCCTACTG							
SEQUENCE SIZE: 312							
INCLUDED REGION SIZE: 312							
PRODUCT SIZE: 202							

AhTC3B05

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	271	20	59.17	45.00	4.00	1.00	
GGAGAAAACGCATTGGAAC							
RIGHT PRIMER	530	20	59.29	45.00	5.00	2.00	
TTTGTCCCCTTGGGAATAGT							
SEQUENCE SIZE: 539							
INCLUDED REGION SIZE: 539							
PRODUCT SIZE: 260							

AhTC3G01

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	134	20	59.96	50.00	5.00	0.00	
GACGGTAATCGTGCCCTAAA							
RIGHT PRIMER	381	20	60.21	55.00	4.00	1.00	
TGCAGTAGTGGCAGCAGAAC							
SEQUENCE SIZE: 455							
INCLUDED REGION SIZE: 455							
PRODUCT SIZE: 248							

Anexo 3

AhT1C3H07 (1)

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	98	20	59.93	45.00	3.00	1.00	
CAATGGGAGGCAAATCAAGT							
RIGHT PRIMER	344	20	60.05	45.00	4.00	1.00	
GCCAAATGGTTCCTTCTCAA							
SEQUENCE SIZE: 476							
INCLUDED REGION SIZE: 476							
PRODUCT SIZE: 247							

AhT1C3H07 (2)

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	156	20	59.70	45.00	5.00	3.00	
TCATCGAACAAAACGTCGTC							
RIGHT PRIMER	441	20	59.82	45.00	6.00	0.00	
CGAATTCCTTTCTTGCCTTG							
SEQUENCE SIZE: 476							
INCLUDED REGION SIZE: 476							
PRODUCT SIZE: 286							

AhT1C3H07 (3)

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	315	20	60.01	50.00	4.00	0.00	
AGGGGGAGAATTGAGAAGGA							
RIGHT PRIMER	441	20	59.82	45.00	6.00	0.00	
CGAATTCCTTTCTTGCCTTG							
SEQUENCE SIZE: 476							
INCLUDED REGION SIZE: 476							
PRODUCT SIZE: 127							

Ah1TC4B01

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	8	21	59.47	47.62	5.00	1.00	
GAGGCTCCAACATTGCTCTAA							
RIGHT PRIMER	300	25	57.32	48.00	5.00	3.00	
CTAGTAGGTTCTTGTGAGAGAGAGC							
SEQUENCE SIZE: 303							
INCLUDED REGION SIZE: 303							
PRODUCT SIZE: 293							

Ah1TC4B07

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	305	22	60.17	36.36	3.00	0.00	
TGAAGCAAGAAAAAGCAGTGAA							
RIGHT PRIMER	470	21	59.98	42.86	4.00	0.00	
CCTTGCCTTTTGGTTTAAAGG							
SEQUENCE SIZE: 485							
INCLUDED REGION SIZE: 485							
PRODUCT SIZE: 166							

Anexo 3

AhTC4T1G10

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	188	20	60.09	45.00	4.00	2.00	
TTCGGTCATGTTTGTCAG							
RIGHT PRIMER	398	20	60.41	55.00	6.00	2.00	
CTCGAGTGCTCACCCCTCAT							
SEQUENCE SIZE: 448							
INCLUDED REGION SIZE: 448							
PRODUCT SIZE: 211							

Ah1TC5A07

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	188	20	59.53	55.00	2.00	0.00	
GTTTGGTTCTCCCTCCTCCT							
RIGHT PRIMER	333	20	60.04	50.00	3.00	2.00	
AGCCTCTTCATTCCCTCAT							
SEQUENCE SIZE: 414							
INCLUDED REGION SIZE: 414							
PRODUCT SIZE: 146							

Ah1TC5C05 (1)

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	165	20	60.00	55.00	3.00	0.00	
GCTGAGAGAACAAGCCCAAC							
RIGHT PRIMER	404	20	59.99	50.00	6.00	2.00	
CCTTTAACACGTGCCCATCT							
SEQUENCE SIZE: 512							
INCLUDED REGION SIZE: 512							
PRODUCT SIZE: 240							

Ah1TC5C05 (2)

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	385	20	59.99	50.00	6.00	2.00	
AGATGGGCACGTGTTAAAGG							
RIGHT PRIMER	512	20	60.07	50.00	2.00	0.00	
GTCCAACCAAATCCAACAC							
SEQUENCE SIZE: 512							
INCLUDED REGION SIZE: 512							
PRODUCT SIZE: 128							

Ah1TC2B05= AhTC2G03

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	657	20	60.46	55.00	6.00	3.00	
GGCGGAAGAGCTTAAGAAG							
RIGHT PRIMER	885	20	60.05	40.00	6.00	0.00	
TTGCAATGGATCTCACAAA							
SEQUENCE SIZE: 915							
INCLUDED REGION SIZE: 915							
PRODUCT SIZE: 229							

Ah1TC4H02 (1)

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	69	20	60.08	50.00	2.00	0.00	
ACCGCAAATCATCCATCTC							
RIGHT PRIMER	253	20	60.12	60.00	5.00	1.00	
GATAGCGTCAGAGGCAGAGG							
SEQUENCE SIZE: 808							
INCLUDED REGION SIZE: 808							
PRODUCT SIZE: 185							

Anexo 3

Ah1TC1B02

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	71	20	59.94	35.00	8.00	0.00	
AACATGCATGCAAATGGAAA							
RIGHT PRIMER	296	21	59.42	42.86	7.00	0.00	
GCCAAAGTCACTTGTGCTT							
SEQUENCE SIZE: 306							
INCLUDED REGION SIZE: 306							
PRODUCT SIZE: 226							

Ah1TC2A11

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	9	20	60.80	55.00	2.00	2.00	
CAGCAGATTGACGGGTTAGC							
RIGHT PRIMER	231	20	59.77	55.00	3.00	3.00	
CAGCAAAGAGTCGTCAGTCG							
SEQUENCE SIZE: 241							
INCLUDED REGION SIZE: 241							
PRODUCT SIZE: 223							

h1TC4H07

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	298	20	59.99	55.00	5.00	3.00	
CCTCCGTTGCTCTTCTGAAC							
RIGHT PRIMER	476	22	59.74	45.45	4.00	2.00	
GATCAAGCACTTCAGACAATGG							
SEQUENCE SIZE: 476							
INCLUDED REGION SIZE: 476							
PRODUCT SIZE: 179							

Ah1TC5A06

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	93	20	59.70	50.00	3.00	1.00	
TCGGTTTGGGAGACTCTT							
RIGHT PRIMER	263	20	59.87	50.00	3.00	1.00	
TTGTAAGCAGACGCCACATC							
SEQUENCE SIZE: 517							
INCLUDED REGION SIZE: 517							
PRODUCT SIZE: 171							

Ah1TC4E09

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	105	20	59.55	50.00	3.00	0.00	
CTTTCTTCCCCTTGAACCT							
RIGHT PRIMER	255	25	51.50	28.00	4.00	2.00	
GATCAAGTAAAAATGTTAGTATAAG							
SEQUENCE SIZE: 255							
INCLUDED REGION SIZE: 255							
PRODUCT SIZE: 151							

Ah1TC6G09

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	90	21	59.98	47.62	8.00	2.00	
GGAGGTTGCATGCATCATAGT							
RIGHT PRIMER	222	23	58.45	34.78	5.00	2.00	
TCATTGAACGTATTTGAAAGCTC							
SEQUENCE SIZE: 224							
INCLUDED REGION SIZE: 224							
PRODUCT SIZE: 133							

Anexo 3

Ah1TC5D06

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	18	24	59.84	33.33	6.00	0.00	
GAAATTTTAGTTTTTCAGCACAGCA							
RIGHT PRIMER	194	23	59.96	34.78	6.00	2.00	
TTTTCCCTCTTAAATTTTCTCG							
SEQUENCE SIZE: 194							
INCLUDED REGION SIZE: 194							
PRODUCT SIZE: 177							

Ah1TC2D06

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	88	20	60.04	50.00	3.00	0.00	
AGGGGGAGTCAAAGGAAAGA							
RIGHT PRIMER	301	20	60.74	50.00	4.00	1.00	
TCACGATCCCTTCTCCTTCA							
SEQUENCE SIZE: 405							
INCLUDED REGION SIZE: 405							
PRODUCT SIZE: 214							

Ah1TC2G05

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	52	20	59.85	50.00	4.00	0.00	
GCCGAGCTAGTTTGATTTGG							
RIGHT PRIMER	338	21	59.75	38.10	4.00	0.00	
TTGGATTTGAATGGAGGAATG							
SEQUENCE SIZE: 390							
INCLUDED REGION SIZE: 390							
PRODUCT SIZE: 287							

Ah1TC3A10

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	32	20	59.76	45.00	4.00	2.00	
GCATGGGGTAAATCTTCCAA							
RIGHT PRIMER	244	20	59.81	50.00	4.00	0.00	
ATGTGCCTATCAGGGGTTT							
SEQUENCE SIZE: 433							
INCLUDED REGION SIZE: 433							
PRODUCT SIZE: 213							

Ah1TC3E02

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	318	21	59.55	42.86	6.00	0.00	
TGAAAGATAGGTTTCGGTGGA							
RIGHT PRIMER	590	20	59.71	50.00	3.00	1.00	
CAAACCGAAGGAGGAACTTG							
SEQUENCE SIZE: 609							
INCLUDED REGION SIZE: 609							
PRODUCT SIZE: 273							

Ah1TC3E05

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	103	20	59.15	55.00	8.00	0.00	
CACCCTTGAGTTGGTGAGG							
RIGHT PRIMER	397	20	59.95	50.00	3.00	3.00	
CTTCTTCTTCTCCCGCAATG							
SEQUENCE SIZE: 427							
INCLUDED REGION SIZE: 427							
PRODUCT SIZE: 295							

Anexo 3

Ah1TC4F02

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	8	20	61.21	55.00	5.00	2.00	
GCACTGCACCCCAATCTCTA							
RIGHT PRIMER	178	20	60.22	55.00	2.00	1.00	
GATGGGTGGTTTGGTGTCTC							
SEQUENCE SIZE: 230							
INCLUDED REGION SIZE: 230							
PRODUCT SIZE: 171							

Ah1TC1A08

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	2	20	59.83	50.00	4.00	2.00	
AAGGGGTTAAGGCATGACT							
RIGHT PRIMER	215	19	61.75	52.63	5.00	3.00	
CCACAAATGGGTCGTTCGAT							
SEQUENCE SIZE: 228							
INCLUDED REGION SIZE: 228							
PRODUCT SIZE: 214							

Ah1TC4E10

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	186	20	60.07	55.00	4.00	0.00	
ACGTCATCTTCCCTCCTCCT							
RIGHT PRIMER	485	20	60.04	45.00	4.00	0.00	
CCATTTTCTCCTCGAACCAA							
SEQUENCE SIZE: 527							
INCLUDED REGION SIZE: 527							
PRODUCT SIZE: 300							

Ah1TC4C11

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	85	20	60.09	55.00	3.00	1.00	
TCCTGACTGGGTCCCTTGTGTC							
RIGHT PRIMER	348	20	58.91	50.00	4.00	2.00	
CCAAAGGGGAGTACGAACAT							
SEQUENCE SIZE: 487							
INCLUDED REGION SIZE: 487							
PRODUCT SIZE: 264							

Ah1TC4D09

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	68	20	60.18	50.00	2.00	0.00	
TTGTGCTCTGCTCTTGGTTG							
RIGHT PRIMER	277	20	59.87	50.00	4.00	0.00	
CTTGCTGGAGGAAACACACA							
SEQUENCE SIZE: 487							
INCLUDED REGION SIZE: 487							
PRODUCT SIZE: 210							

Ah1TC4F10

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	3	25	58.70	40.00	4.00	0.00	
TCCAAGTACCTTCTTCTTCTTC							
RIGHT PRIMER	214	21	59.54	47.62	6.00	2.00	
CGTTGAAGTCAATCCAGATCC							
SEQUENCE SIZE: 442							
INCLUDED REGION SIZE: 442							
PRODUCT SIZE: 212							

Anexo 3

Ah1TC4G05

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	34	20	60.11	60.00	3.00	3.00	
AACCCACTACGGGACTACCC							
RIGHT PRIMER	323	20	59.99	55.00	5.00	0.00	
ACGACGTGGAGGAGAAGAGA							
SEQUENCE SIZE: 352							
INCLUDED REGION SIZE: 352							
PRODUCT SIZE: 290							

Ah1TC4A02= Ah1TC4H01

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	43	20	59.90	40.00	6.00	1.00	
ATTCAAATCGGAATGGCAAG							
RIGHT PRIMER	319	20	59.81	50.00	3.00	0.00	
GAGCAAAGGGCGAATCTATG							
SEQUENCE SIZE: 363							
INCLUDED REGION SIZE: 363							
PRODUCT SIZE: 277							

Ah1TC2A02

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	5	20	59.70	55.00	3.00	0.00	
CTCCCTTGTGGGTATGTGGT							
RIGHT PRIMER	201	20	60.01	45.00	2.00	1.00	
GGCTCCCATTCATTCTCAAA							
SEQUENCE SIZE: 287							
INCLUDED REGION SIZE: 287							
PRODUCT SIZE: 197							

Ah1TC4C08 (1)

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	50	20	59.81	50.00	6.00	1.00	
CTCCCAAAGCTTCTGAATGG							
RIGHT PRIMER	260	20	60.03	45.00	5.00	2.00	
TGGAGAATTGCCGATTTAGG							
SEQUENCE SIZE: 520							
INCLUDED REGION SIZE: 520							
PRODUCT SIZE: 211							

Ah1TC4C08 (2)

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	334	20	60.32	55.00	4.00	0.00	
ACACTACCTTCAGCGCCAAC							
RIGHT PRIMER	516	21	59.17	38.10	5.00	2.00	
TCTGTGTGAATTTCCGCATTA							
SEQUENCE SIZE: 520							
INCLUDED REGION SIZE: 520							
PRODUCT SIZE: 183							

Ah1TC4D12

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	127	20	59.76	60.00	2.00	0.00	
ATGGAGGAGAGGGAGAGAGG							
RIGHT PRIMER	278	20	60.03	45.00	4.00	0.00	
AAATCTTTCCCAACCAACC							
SEQUENCE SIZE: 306							
INCLUDED REGION SIZE: 306							
PRODUCT SIZE: 152							

Ah1TC6H03

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	3	21	59.43	42.86	6.00	0.00	
TCACAATCAGAGCTCCAACAA							
RIGHT PRIMER	225	20	60.15	55.00	4.00	2.00	
CAGGTCACCAGGAACGAGT							
SEQUENCE SIZE: 264							
INCLUDED REGION SIZE: 264							
PRODUCT SIZE: 223							

Anexo 4

Blastx- Sequências com Microssatélites

This is a cleaned up blastx which only shows those results with significant hits the threshold of significance used was 10⁻⁶.

This is a cleaned up version of a blast consensus_microsats.aa_blastx, with non-hits and non significant hits removed.

Total number of records inspected is 139

Number of non-hits is 37

Number of significant hits is 13

Number of non-significant hits is 89

Reference:

ALTSCHUL, S.F.; MADDEN, T.; SCHAFFER, A.; ZHANG, J.; ZHANG, Z.; MILLER, W.; LIPMAN, D. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**:3389-3402.

Ah1TC1A02

Sequences producing significant alignments:

At1g12930 unknown protein

Score	E
(bits)	Value
61	2e-10

>At1g12930 unknown protein

Length = 938

Score = 60.8 bits (146), Expect = 2e-10
Identities = 25/35 (71%), Positives = 31/35 (88%)
Frame = +2

Query: 77 VHKCATILQQLAAICTLSERTTLKSILCWQTLHGW 181
VHKC+TILQQLAAIC+L ERT+ K +LCW++LGW
Sbjct: 822 VHKCSTILQQLAAICSLCERTSWKGMCLWKS LQGW 856

Ah1TC2A09

Sequences producing significant alignments:

AT4g07640 putative athila transposon protein

Score	E
(bits)	Value
59	9e-10

>AT4g07640 putative athila transposon protein

Length = 866

Score = 58.5 bits (140), Expect = 9e-10
Identities = 38/102 (37%), Positives = 57/102 (55%), Gaps = 2/102 (1%)
Frame = -2

Query: 301 PSATKEKEKE-VLKPYPTRAPYPQRLMKSEKDGQFSRFLFIFKKLQINIPFAEAIEQMPL 125
P K KEK V PY P P+P R K+ D + F + K++++ IP +A+ +P
Sbjct: 437 PITVKNKEKVFVPPYKPELPPFGRHKKALADKYRAMFAKNIKEVELRIPLVDALALIPD 496

Query: 124 YAKFLKELMTKKRSWRNEETVLLTEECSAIIQHK-LPQXLKD 2

KFLK+L+ +R + V+L+ CSAIIQ K +P+ L D
Sbjct: 497 SHKFLKDLIV-ERIQEVQGMVVL SHGCSAIIQKKIIPKKLSD 537

Ah1TC1D01

Sequences producing significant alignments:

AT4g11980 putative protein

Score	E
(bits)	Value
60	7e-14

>AT4g11980 putative protein

Length = 310

Score = 60.5 bits (145), Expect (2) = 7e-14
Identities = 26/32 (81%), Positives = 29/32 (90%)
Frame = -1

Query: 231 QARVPTGRIILELPTGMLDDDKGDFIGTGVRE 136
Q RVPTG+I+LELP GMLDDDKGDF+GT VRE
Sbjct: 164 QVRVPTGKIVLELPAGMLDDDKGDFVGTAVRE 195

Query: 52 QVEEEIGIKFKLEDMVD 2
+VEEEIGIK K EDMVD

Sbjct: 195 EVEEEIGIKLKKEDMVD 211

Anexo 4

Ah1TC3A12

Sequences producing significant alignments:
AT4g11260 putative protein

Score E
(bits) Value
54 4e-19

>AT4g11260 putative protein

Length = 358

Score = 53.9 bits (128), Expect(3) = 4e-19
Identities = 22/37 (59%), Positives = 30/37 (80%)
Frame = +1

Query: 355 TACIKLEEYQTAKVALQGTGASFPDDSRFTKLVQDCD 465
TAC+KLEEY TAK AL+ GAS AP++ +F K++ +CD
Sbjct: 78 TACMKLEEYSTAKAALEK GASVAPNEPKFKMIDECD 114

Score = 45.4 bits (106), Expect(3) = 4e-19
Identities = 22/29 (75%), Positives = 25/29 (85%)
Frame = +3

Query: 186 TEAVSDANKAIQLSPSLAKAYLRKGYAVL 272
TEAV DANKAI+L P+LAKAYLRKG A +
Sbjct: 53 TEAVVDANKAIELEPTLAKAYLRKGTACM 81

Score = 31.2 bits (69), Expect(3) = 4e-19
Identities = 13/22 (59%), Positives = 17/22 (77%)
Frame = +1

Query: 1 DPSNPLLYADRAQAHIKLNFT 66
DP+ +ADRAQA+IK+ NFT
Sbjct: 32 DPNCAAFFADRAQANIKIDNFT 53

Ah1TC3H07

Sequences producing significant alignments:
AT3g05520 alpha subunit of F-actin capping protein

Score E
(bits) Value
59 5e-19

>AT3g05520 alpha subunit of F-actin capping protein

Length = 308

Score = 58.5 bits (140), Expect(2) = 5e-19
Identities = 23/40 (57%), Positives = 36/40 (89%)
Frame = -2

Query: 196 DVKSILNDDVLFDEAASEAFPVYKNSHLICLPLPNRSGDV 77
D+K++L+D+ +++EAA EAFPVYK+H+ICL +P+ +GDV
Sbjct: 40 DLKAVLSDEEVYNEAAMEAFPVYKTHMICLEMPGAGDV 79

Score = 52.0 bits (123), Expect(2) = 5e-19
Identities = 27/39 (69%), Positives = 30/39 (76%), Gaps = 3/39 (7%)
Frame = -1

Query: 398 MADEEEEE---SELSDKQKVEIAKWFLNLSPPGEIQYVAK 291
MADEE+E +ELS QK EIAKWF LN+P GEI YVAK
Sbjct: 1 MADEEDELLETELSYDQKKEIAKWFFLNAPAGEINYVAK 39

Ah1TC4D02

Sequences producing significant alignments:
AT3g23710 Tic22, putative

Score E
(bits) Value
55 1e-08

>AT3g23710 Tic22, putative

Length = 313

Score = 55.5 bits (132), Expect = 1e-08
Identities = 28/62 (45%), Positives = 46/62 (74%), Gaps = 1/62 (1%)
Frame = +2

Query: 179 VYALNPNWEEFLLIAGASTSKNISIFCFNEDDAQALLCQVFTVHPRLR-DSSKVVVVALN 355
VYAL+N EEF+L++G S+ K++ + E+DA+ LL ++ ++ PR+R + SKVV +AL+
Sbjct: 124 VYALSNSNEEFVLVSGTSSGKSLGLLFCKEEDAETLLKEMKSMDFRMRKEGSKVVALALS 183

Query: 356 KV 361
KV
Sbjct: 184 KV 185

Anexo 4

Ah1TC4E10

Sequences producing significant alignments:
At1g30270 serine/threonine kinase, putative

Score E
(bits) Value
90 8e-19

>**At1g30270 serine/threonine kinase, putative**
Length = 480

Score = 89.7 bits (221), Expect = 8e-19
Identities = 42/54 (77%), Positives = 49/54 (89%)
Frame = -3

Query: 162 ATNSGRTRVGGKYELGRTLGEENFAKVKFAKNTVTGDNVAIKILDKEKVLKHKMI 1
+++ GRTRVGGKYELGRTLGEENFAKVKFA+N GDNVAIK++DKEKVLK+KMI
Sbjct: 20 SSSGGRTRVGGKYELGRTLGEETFAKVKFARNVENGDNVAIKVIDKEKVLKHKMI 73

Ah1TC4H07

Sequences producing significant alignments:
AT5g44790 ATP dependent copper transporter

Score E
(bits) Value
67 4e-12

>**AT5g44790 ATP dependent copper transporter**
Length = 1001

Score = 67.0 bits (162), Expect = 4e-12
Identities = 46/84 (54%), Positives = 55/84 (64%), Gaps = 12/84 (14%)
Frame = +3

Query: 138 LEEVRRLXSYDSRNGIDD-----GR----MRWIQVRVTGMSCAACSTSIESALYAVD 284
+EEV LL SY + DD GR +R IQV VTGM+CAACS S+E+AL V+
Sbjct: 24 MEEVG-LLDSYHNEANADDILTKIEEGRDVSGLRRIQVGVGTGMTCAACSNVVEAALMNVN 82

Query: 285 SVITASVALL-NKADVVPALVK 353
V ASVALL N+ADVVP+P LVK
Sbjct: 83 GVFKASVALLQNRADVVPALVK 106

Score = 30.8 bits (68), Expect = 0.35
Identities = 17/47 (36%), Positives = 27/47 (57%), Gaps = 1/47 (2%)
Frame = +3

Query: 213 QVRVTGMSCAACSTSIESALYAVDSVITASVAL-LNKADVVPALV 350
Q + GM+CAAC S+E L + V A VAL + +V ++P ++
Sbjct: 136 QFTIGGMTCAACVNSVEGILRDLPGVKRAVVALSTSLGEVEYDPNVI 182

Ah1TC5C05

Sequences producing significant alignments:
AT3g30300 auxin-independent growth promoter, putative

Score E
(bits) Value
83 7e-17

>**AT3g30300 auxin-independent growth promoter, putative**
Length = 651

Score = 83.2 bits (204), Expect = 7e-17
Identities = 42/68 (61%), Positives = 52/68 (75%), Gaps = 6/68 (8%)
Frame = -2

Query: 232 MKGEVK--MKSKMKWVGLVGLVLSAFSIFTHFLARFTEMGIAEYQSSVTIFSWRPIFD- 62
MKGE K +KS+MKW+GL+GLVLSAFS+ HFLLA FT+ I++Y VTIFSWRPIFD
Sbjct: 1 MKGEGKVFLLKSRMKWIGLLGLVLSAFSLLVHFLLAGFTDDSIDYSIPVTIFSWRPIFDN 60

Query: 61 ---LQHSP 47
+H+P
Sbjct: 61 PRFARHTP 68

Anexo 4

Ah1TC5C10

Sequences producing significant alignments: (bits) Value
At1g36560 hypothetical protein 96 1e-20

>At1g36560 hypothetical protein

Length = 524

Score = 95.9 bits (237), Expect = 1e-20
 Identities = 47/104 (45%), Positives = 71/104 (68%)
 Frame = -3

Query: 338 KALCDLGLGINLMPLSVIEKMGYGVQADKISLEMVDSSRKQAYGQVEDVLVKVEGLYTP 159
 K LCDLG ++LMPLSV +++G + ISL + D S +Q +G +ED+ +K+ + P
 Sbjct: 138 KCLCDLGASVSLMPLSVAKRLGFNKYKYNISLILADGSVRQPHGVLEDLPIKIGKVEVP 197
 Query: 158 ADFIVLDTVKEEDESIIILGRPFLLATARAVIDVDRGELVLQLNKD 27
 DFI+L+ +E + +ILGRPFLLATA A+IDV +G++ L + KD
 Sbjct: 198 TDFIILNMDEEPKDFLILGRPFLLATAGAIIDVKQKIDLNMGMK 241

Ah1TC5F05

Sequences producing significant alignments: (bits) Value
At2g10680 pseudogene 57 4e-13

>At2g10680 pseudogene

Length = 929

Score = 57.4 bits (137), Expect(2) = 4e-13
 Identities = 32/93 (34%), Positives = 55/93 (58%), Gaps = 1/93 (1%)
 Frame = -3

Query: 486 GVIEDMIVKVRPFAPFTDFVVLIDIEGHKSAFLILGRPFVLVIGRTLIDVEKGEVTLKVNEE 307
 G++ED+ V + PTDFVVL+++G LILGRPFL +IDV++G++ L + E+
 Sbjct: 634 GLLDVDVPMINGVEVPTDFVVLMDGESKDFLILGRPFLLASVGAVIDVKQKINLNLG 693

Query: 306 -KFVLNAVKAVQHLDIPEECLSIDFIDSLVDEV 211
 K + A++ I + ++ +D L +E+
 Sbjct: 694 VKMKFDIRDAMKKPTIEGQTFVLEEMDQLGNEL 726

Score = 33.9 bits (76), Expect(2) = 4e-13
 Identities = 18/44 (40%), Positives = 28/44 (62%)
 Frame = -2
 Query: 616 ITDLGTNINLIPASLIKKLCLTNEVKPIRMYLQLADGSIKIPSG 485
 + DLG +++L+P S++KKL + KP + L LAD + K P G
 Sbjct: 592 LCDLGASVSLMPLSVVKKLGFVH-YKPCDLTLILADRTSKTPFG 634

Ah1TC5H11

Sequences producing significant alignments: (bits) Value E

AT4g16910 retrotransposon like protein 101 4e-22

>AT4g16910 retrotransposon like protein

Length = 687

Score = 101 bits (251), Expect = 4e-22
 Identities = 48/105 (45%), Positives = 68/105 (64%)
 Frame = +2

Query: 2 ILERIGPVAYRVALPPHLSNLHDVFHVSQRLRKYTPDASHVLEPESVQLREDLTLVPVPR 181
 ++ER+G VAY++ LPP L H+VFHVSQRLR + +E L+E++T+ PVR
 Sbjct: 562 VIERVGAVAYKLDLPPKLDLAFHNVFHVSQRLRCLSEQEESMEDVPPGLKENMTVEAWPVR 621

Query: 182 IDDTSIKRLREKEVSLVKVAWNRGGVEEHTWELESKMRTDYPHLF 316
 I D K R K + L+K+ WN GG EE+TWE E+KM+ ++P F
 Sbjct: 622 IMDQMKGTRGKSMDDLKILWNCGGREEYTWETETKMKANFPPEWF 666

Anexo 4

Ah1TC6H03

Sequences producing significant alignments:
At1g16920 hypothetical protein

Score E
(bits) Value
105 3e-24

>**At1g16920 hypothetical protein**
Length = 216

Score = 105 bits (263), Expect = 3e-24
Identities = 51/56 (91%), Positives = 54/56 (96%)
Frame = -2

Query: 263 IGDSGVGKSNLLSRFTRNEFNLESKSTIGVEFATKTLNIDSKVVKAQIWDTAGQER 96
IGDSGVGKSNLLSRFT+NEFNLESKSTIGVEFAT+TL +D KVVKAQIWDTAGQER
Sbjct: 19 IGDSGVGKSNLLSRFTKNEFNLESKSTIGVEFATRLKVDGKVVKAQIWDTAGQER 74

Blastx- Sequências sem Microssatélites**05-6F-**

Sequences producing significant alignments:
At1g20390 hypothetical protein

Score E
(bits) Value
77 3e-15

>**At1g20390 hypothetical protein**
Length = 1791

Score = 77.4 bits (189), Expect = 3e-15
Identities = 44/135 (32%), Positives = 65/135 (47%), Gaps = 12/135 (8%)
Frame = +1

Query: 97 ADVLSKLATRQAENTSALSQILDKPSFEQDTILSITQV-----PDWRTPFL 240
AD L+ LA T ++ + +DKPS + + I DWR
Sbjct: 1326 ADALAALALTSDDLRRRIIPVESIDKPSIDSTDAVEIVNTIRSSNAPDPADPTDWRVEIR 1385
Query: 241 DYINTGTMPNDESNLPPFRKASFYTVLGNTLYRRGHSRPLLKCSWEEAEDVMAETHEG 420
DY++ GT+P+D+ R KA+ YT++ L + +L C+ E ++M ETHEG
Sbjct: 1386 DYLSDGTLP SDKWTARRLRKAAYTLMKEHLLKVSFAFGAMLNCLHGTEINEIMKETHEG 1445
Query: 421 VCGNHIGDRALAAKI 465
GNH G RALA K+
Sbjct: 1446 AAGNHSGGRALALKL 1460

11-12F-

Sequences producing significant alignments:
At2g06320 putative retroelement pol polyprotein

Score E
(bits) Value
144 3e-35

>**At2g06320 putative retroelement pol polyprotein**
Length = 466

Score = 144 bits (362), Expect = 3e-35
Identities = 72/110 (65%), Positives = 83/110 (75%), Gaps = 1/110 (0%)
Frame = -3

Query: 332 IHDLWGIDYMGPPPSYTFKYILVAVEYVSKWVEAIATTTCDTNVVLQFLKKNIFTRFGV 153
I D+WGID+MGPPF SY KYILV V+YVSKWVEAIA+ T D VVL+ K IF RFGV
Sbjct: 167 IFDVWIDFMGPPSSYGNKYILVVVDYVSKWVEAIASPTNDKVVLLKLFKTIIFPRFGV 226
Query: 152 PKGLISDGGSHFCNKQINSLHXYGVTHKVATPYHPQTMGKL-TSNRELK 6
+ISDGG HF NK +LL K+GV HKVATPYHPQT G++ SNRE+K
Sbjct: 227 SWVVISDGGKHFINKVFENLLKHKGVKHKVATPYHPQTSQGVEISNREIK 276

Anexo 4

Ah1TC1S1A06

	Score	E
Sequences producing significant alignments:	(bits)	Value
Atlg37060 Athila retroelment ORF 1, putative	64	5e-11

>Atlg37060 Athila retroelment ORF 1, putative

Length = 1734

Score = 63.5 bits (153), Expect = 5e-11
 Identities = 36/101 (35%), Positives = 50/101 (48%), Gaps = 5/101 (4%)
 Frame = -1

Query: 468 DHLRIEVPVAGASLPSTEISETFPDEQLFAIQEA-----PWFADITNYKAVRFIPQEYSR 304
 DHLSR+ I ++ P+EQL AIQ+ FW+AD NY P S
 Sbjct: 1300 DHLSRMRE-----DEVLIIDSMPEEQLMAIQQLNEKKLPWYADHVNYLVSGEPPNLS 1354
 Query: 303 VQRKKLISDAKYLL*DEPYLFKRCADGMIRRCVPREEAQRI 181
 ++KK D ++ DEPYL+ C D + R CV +E + I
 Sbjct: 1355 YEKKKFFKDINHFWDEPYLYTLCKDKIYRTCVSEDEIEGI 1395

Ah1TC1S1B04

	Score	E
Sequences producing significant alignments:	(bits)	Value
AT3g31340 Athila ORF 1, putative	122	2e-28

>AT3g31340 Athila ORF 1, putative

Length = 781

Score = 122 bits (305), Expect = 2e-28
 Identities = 66/163 (40%), Positives = 96/163 (58%)
 Frame = +2

Query: 47 LTEECSAVILKSFPEKLDKRGSLFIPCILKGNCTKTAFCDLGASINLIPASTIRKLGTE 226
 + CSA+ + PEKL D GSF++PC + + + CDLGA +NL+P S ++LG+T
 Sbjct: 516 MVSTCSAIPPATIEPEKLDGPGSFLVPCIRIGKSAFERCLCDLGA+VNL+PMSKRLGITN 575
 Query: 227 EVKPTRIYLQLADDSTKYPSGVIEDMIVRVGPFAPPTDFVQEMEEHKSATLILGRPFLA 406
 KP+RI L LAD S ++P G+ E++ VRVG F PTDFVV E+++ L LGRPFL
 Sbjct: 576 -FKPSRISLILADRSVRFVGLAENVHVRVGDYIPTDFVVLELDKEPHDPLTLGRPFLN 634
 Query: 407 TRRYLIDVQ*GEITLRVDDDEFKLNNAVAMQHPDTSKDCMKVD 535
 T +IDV+ I L++ D + + ++P VD
 Sbjct: 635 TVGAIIDVRRSTINLQIGDFALEFDMKGRKNPTIEGHAFSVD 677

Ah1TC1S1C04

	Score	E
Sequences producing significant alignments:	(bits)	Value
Atlg37060 Athila retroelment ORF 1, putative	202	1e-52

>Atlg37060 Athila retroelment ORF 1, putative

Length = 1734

Score = 202 bits (515), Expect = 1e-52
 Identities = 94/173 (54%), Positives = 125/173 (71%)
 Frame = +2

Query: 2 IISSALEPQEEKALIQVLRTHKTALGWSISDLKGISPARCTHKKILLEGDAKPVVQRRL 181
 I++ L + LI L ++ A+G+S+ D+KGISP CTH+I LE ++ ++ QRRL
 Sbjct: 848 IVNDELTAQVNNLLITELMKYRKAIGYSLDDIKGISPTLCTHRIHLENESYSSIEPQRRL 907
 Query: 182 NPAMKEVVQKEVTKLLEAGIYPIISDPSWVSPVQVVPKGGMTVVHNEKNELVPTRTVTG 361
 NP +KEVV+KE+ KLL+AG+IYPIISDS WVSPV VPKGGMTVV N K+EL+PTRT+TG
 Sbjct: 908 NPNLKEVVKKEILKLLDAGVIYPIISDSTWVSPVHCVPKGGMTVVKNSKDELIPTRTITG 967
 Query: 362 WRMCIDYRRLNATATRKDHLPSPLIDQMLHPLVPHASHCSLPAYPPYHHLALHP 520
 RMC+YR+LN A+RK+H P P ID ML L H +C L +Y + + +HP
 Sbjct: 968 HRMCIEYRKLNVASRKEHFPLPFIDHMLERLANHPYCYFLDSYSGFFQIPIHP 1020

Anexo 4

Ah1TC1S1C07

Sequences producing significant alignments:
At2g12040 T10J7.2

Score E
(bits) Value
209 1e-54

>At2g12040 T10J7.2

Length = 976

Score = 209 bits (531), Expect = 1e-54
Identities = 99/149 (66%), Positives = 120/149 (80%)
Frame = -1

Query: 563 DQAKIEVIEKLPANVKAIRSFLGHAGFYRRFIKDFS KIAKPLSNLLAADTPFVFDTEC 384
D+AKIEV+ L N+KA+RSFLGHAGFYRRFIKDFS KIA+PL++LL + F F EC
Sbjct: 156 DRAKIEVMTSLQALDNLKAVRSFLGHAGFYRRFIKDFS KIA RPLTSLCKEVKFEFTQEC 215

Query: 383 LQAFEMLKAKLV TAPVISTPDWTLPFELMCDASDHAIGAVLGQRHDKLLHVIIYASRVLN 204
AF+ +K L++AP++ PDW LPFE+MCDASD A+GAVLGQR DK LH IYYASR L+
Sbjct: 216 HDAFQQIKQALISAPIVQPPDWLDPFEVMCDASDFAVGAVLGQRDKKLLHAIYYASRTLD 275

Query: 203 DAQKNYTTTEKELLAVVY AIDKFRSYLVG 117
DA++NY TTEKE LAVV+A +KFRSYLVG
Sbjct: 276 DAKRNYATTEKEFLAVVFAFEKFRSYLVG 304

Ah1TC1S1C08

Sequences producing significant alignments:
AT3g28340 unknown protein

Score E
(bits) Value
58 1e-09

>AT3g28340 unknown protein

Length = 364

Score = 58.2 bits (139), Expect = 1e-09
Identities = 38/118 (32%), Positives = 51/118 (43%)
Frame = +2

Query: 5 YMQKYDSYAPGASESPTIHKQKVDCSYERNLKNRHNQSAVHIATTL DLE*LIGSLAAV 184
+ +D + E+P ++ +DCS N + SAVHIA TLD L G+++AV
Sbjct: 30 FPDSFDDASSDLMEAPA-YQNGLDLDCSVLAKNRLL LACDPSAVHIAMTLDPAYLRGTVSAV 88

Query: 185 HSVVKHTSCPEXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXKVYVSNVRLVNALISPSI 358
HS++KHTSCPE KVV V LIS SI
Sbjct: 89 HSILKHTSCPENIFFHFIASGTSQGLAKTLSSVFP SLSFKVYTFEETT VKNLISSSI 146

Ah1TC1S1D11

Sequences producing significant alignments:
AT4g08030 putative Athila-like protein

Score E
(bits) Value
75 4e-14

>AT4g08030 putative Athila-like protein

Length = 285

Score = 75.5 bits (184), Expect = 4e-14
Identities = 42/88 (47%), Positives = 56/88 (62%)
Frame = -2

Query: 439 SRVNFRLPSV*EKL GICGVQATRV SLEMVDSSRKQGYGQVEDVLVKVEGLYTPADFIVLD 260
S +N P V E+LG+ + TR++L VD S++ G +EDV VKV PADF+VLD
Sbjct: 15 SSINLMPKFVDERLGMTNYRPT RITLLFVDRSKRIPEGILEDVSVKVGNLIPADFVVDL 74

Query: 259 TVKEEDESII LGRPF LATA RAVIDVDRG 176
KE + +ILGR F+ATA A IDV +G
Sbjct: 75 YDKPKDPLILGRAFIATAGARIDVKKG 102

Anexo 4

Ah1TC1S1G06

Sequences producing significant alignments:
AT3g23890 topoisomerase II

Score E
(bits) Value
81 3e-16

>AT3g23890 topoisomerase II

Length = 1473

Score = 81.3 bits (199), Expect = 3e-16
Identities = 50/124 (40%), Positives = 68/124 (54%), Gaps = 3/124 (2%)
Frame = -1

Query: 598 HAIARYIYTELNPITRCLFHEDDDKLEYLEYLNEDGRSIEPSW*LIMLFGQSCLLFQQSDM 419
A ARYI+T+L+P+TR LF +DDD LL+YLNEDG+ IEP+W
Sbjct: 788 SASARYIFTKLSPVTRILFPKDDDLLLDYLNEDGQRIEPTW----- 829

Query: 418 RTYSNSLFCQHRYPPIIPLVLVNGSEGY---WNRLDFLHSQL*S*GIPI*RILTGEAM 248
Y+PIIP VLVNG+EG W+ F+ + + + R+L GE+M
Sbjct: 830 -----YMPIIPTVLVNGAEGIGTGWS--TFIPNYPREIVANVRLLNGESM 874
Query: 247 LPMD 236
+PMD
Sbjct: 875 VPMD 878

Score = 36.6 bits (83), Expect = 0.010
Identities = 13/17 (76%), Positives = 16/17 (93%)
Frame = -2

Query: 333 IGTGWISYIPNYPREL 283
IGTGW ++IPNYPRE+
Sbjct: 845 IGTGWSTFIPNYPREI 861

Ah1TC2S1A09

Sequences producing significant alignments:
AT4g07640 putative athila transposon protein

Score E
(bits) Value
59 1e-09

>AT4g07640 putative athila transposon protein

Length = 866

Score = 58.5 bits (140), Expect = 1e-09
Identities = 38/102 (37%), Positives = 57/102 (55%), Gaps = 2/102 (1%)
Frame = -2

Query: 360 PSATKEKEKE-VLKPYPTRAPYPQRLMKSEKDGQFSRFLFIFKKLQINIPFAEAIEQMPL 184
P K KEK V PY P P+P R K+ D + F + K++++ IP +A+ +P
Sbjct: 437 PITVKNKEKVFVPPYPKPELPPFGRHKALADKYRAMFAKNIKEVELRIPLVDALALIPD 496

Query: 183 YAKFLKELMTKKRSWRNEETVLLTEECSAIQHK-LPQXLKD 61
KFLK+L+ +R + V+L+ CSAIIQ K +P+ L D
Sbjct: 497 SHKFLKDLIV-ERIQEVQGMVVLSHGCSAIIQKKIIPKKLSD 537

Ah1TC2S1D11

Sequences producing significant alignments:
At2g07660 putative retroelement pol polyprotein

Score E
(bits) Value
196 5e-51

>At2g07660 putative retroelement pol polyprotein

Length = 949

Score = 196 bits (498), Expect = 5e-51
Identities = 90/153 (58%), Positives = 120/153 (77%)
Frame = -1

Query: 462 FWGAFQEAFATQLSLSTAYHPQMDGQSERTIQTLLEDMLRACLLDQFESWDRYMPLEVFAY 283
FW AFQ+A T+++LSTAYHPQ DGQSERTIQTLLEDMLRAC+LD +W+++Y+ L+EFAY
Sbjct: 701 FWNAFQKALGTRVNLSTAYHPQTDGQSERTIQTLLEDMLRACVLDWGGNWEKYLRLIEFAY 760
Query: 282 NKNYHASIGMAPYEALYGRKCQSPWCWYEVGEKSLTGPEMISSETTEQIKKIRSRILVAQS 103
N ++ ASIGM+PYEALYGR C++PLCW VGE+ L GP ++ ETTE++K ++ ++ AQ
Sbjct: 761 NNSFQASIGMSPYEALYGRACRTPLCWTPVGERRLFGPTIVDETTERRMKFLKIKLKEAQD 820

Query: 102 RQKSYADQMWKLEFEEGEHVFMKVTLTTGIGR 4
RQKSYA++ K LEF+ G+ V++K G GR
Sbjct: 821 RQKSYANKRRKELEFQVGDVLYLKAMTYKGAGR 853

Anexo 4

Ah1TC3SE06

	Score	E
Sequences producing significant alignments:	(bits)	Value
AT3g14470 disease resistance protein, putative	70	3e-13

>AT3g14470 disease resistance protein, putative

Length = 1054

Score = 70.5 bits (171), Expect = 3e-13
 Identities = 39/95 (41%), Positives = 58/95 (61%), Gaps = 3/95 (3%)
 Frame = -1

Query: 334 SLYIDLLSMEASRASKLKCVRVLSFRERDVL---PDSIGKLIHLRYLNLSRTDVKTLPE 164
 S +D ++S E L +RVLS + PD + H R+L+LSRT+++ LP+S
 Sbjct: 562 SCCLDQMVS-EKLLPTLTRLRVLSLSHYKIARLPPDFFKNISHARFLDLSRTELEKLPKS 620

Query: 163 LCNLFNQLTMLYACYKLSMLPNDMHKLVNVRHLD 59
 LC ++NLQTL+L C L LP D+ L+N+R+LD
 Sbjct: 621 LCYMYNLQTLTLLSYCSSLKELPTDISNLINLRYLD 655

Score = 33.9 bits (76), Expect = 0.029
 Identities = 17/40 (42%), Positives = 24/40 (59%)
 Frame = -1

Query: 244 LPDSIGKLIHLRYLNLSRTDVKTLPELCLNLFNQLTMLY 125
 LP I LI+LRYL+L T ++ +P L +LQTL +
 Sbjct: 641 LPTDISNLINLRYLDLIGTKLRQMPRRFGRKSLQTLTTF 680

Ah1TC3SE11

	Score	E
Sequences producing significant alignments:	(bits)	Value
At1g37060 Athila retroelment ORF 1, putative	170	4e-43

>At1g37060 Athila retroelment ORF 1, putative

Length = 1734

Score = 170 bits (431), Expect = 4e-43
 Identities = 82/137 (59%), Positives = 108/137 (77%)
 Frame = -2

Query: 587 IINSSLSQDQENELLQVLRHKAIGWTPADLKGISSAICIKILLEDADAKPSIQSQRRL 408
 I+N L+ DQ N L+ L +++ AIG++ D+KGIS +C H+I LE+++ SI+ QRRL
 Sbjct: 848 IVNDELTAQVNLITELMKYRKAIGYSLDDIKGISPTLCTHRIHLENESYSSIEPQRRL 907

Query: 407 NPIMKEVVQKEVMKLVQGGVIYPISDSPWVSPVHVVLKGGITVVPNERNELIPTRTVTG 228
 NP +KEVV+KE++KL GVIYPISDS WVSPVH V KGGG+TVV N ++ELIPTRT+TG
 Sbjct: 908 NPNLKEVVKKEILKLLDAGVIYPISDSTWVSPVHCVPKKGGMTVVKNSKDELIPTRTITG 967

Query: 227 WRMCIDYRKLNEATRKD 177
 RMC I+YRKL N A+RK+
 Sbjct: 968 HRMCIEYRKLNVASRKE 984

Ah1TC3SF11

	Score	E
Sequences producing significant alignments:	(bits)	Value
AT3g45040 putative protein	83	2e-17

>AT3g45040 putative protein

Length = 598

Score = 83.2 bits (204), Expect = 2e-17
 Identities = 38/56 (67%), Positives = 49/56 (86%)
 Frame = -1

Query: 295 DPFKRLSLCIYWVGVICLSVLYVYDISKNSRFERILLRKYHLLAVLMFVPALILQ 128
 +P KRLSLCIYW+ +I +SV Y+IS++S+ ERILLRKYHLL+AVLMF+PAL+LQ
 Sbjct: 336 EPLKRLSLCIYWILLIVSVSRFYNISRSKVERILLRKYHLLMAVLMFLPALVLQ 391

Anexo 4

Ah1TC3SG02

Sequences producing significant alignments:
AT3g61330 copia-type polyprotein

Score E
(bits) Value
156 3e-39

>**AT3g61330 copia-type polyprotein**
Length = 1352

Score = 156 bits (395), Expect = 3e-39
Identities = 77/128 (60%), Positives = 88/128 (68%)
Frame = -1

Query: 386 DHPHQLCEACLLGKHSRKSFPKQSKSRAIKPLQLIHADVCGPIKPLSLGKSAYFLLFIDD 207
+HP+Q+CE CLLGK + SFPK+S SRA KPL+LIH DVCPIKP SLGKS YFLLFIDD
Sbjct: 496 NHPNQVCEGCLLGKQFKMSFPKESSSRAQKPLELIHTDVCPIKPKSLGKSNYFLLFIDD 555
Query: 206 YSRXTWVYFLXXXXXXXXXXXXXXXXXXXXXXXXXEGSYEIKALRTDRGGEFTSNEFNMFCEDYGI 27
+SR TWVYFL E IK +R+DRGGEFTS EF +CED GI
Sbjct: 556 FSRKTWVYFLKEKSEVFEIFKFKFAHVEKESGLVIKTMRSRDRGGEFTSKEFLKYCEDNGI 615
Query: 26 *RPLTIPR 3
R LT+PR
Sbjct: 616 RRQLTVPR 623

Ah1TC4SA10

Sequences producing significant alignments:
At1g37060 Athila retroelment ORF 1, putative

Score E
(bits) Value
150 9e-38

>**At1g37060 Athila retroelment ORF 1, putative**
Length = 1734

Score = 150 bits (380), Expect = 9e-38
Identities = 69/90 (76%), Positives = 76/90 (83%)
Frame = -1

Query: 272 P*DQEKIAFTCLSGVFAYRRMPFGLCNAPVTFQRCMLSIFSDIVEKFLEVMDDFSVYGD 93
P DQ K FTC G FAY+RMPFGLCNAP TFQRCM SIFSD++E+ +EVFMDDFSVYG
Sbjct: 1020 PNDQGKTFFTCPYGTFAFKRMPFGLCNAPATFQRCMTSIFSDLIEEMVEVFMDDFSVYGS 1079
Query: 92 SFSSCLDHLAFVLKRCQETNLVLNWEKCHF 3
SFSSCL +L VLKRC+ETNLVLNWEKCHF
Sbjct: 1080 SFSSCLLNLRCVLRKCEETNLVLNWEKCHF 1109

Ah1TC4SA06

Sequences producing significant alignments:
At1g48120 serine/threonine phosphatase PP7, putative

Score E
(bits) Value
74 2e-14

>**At1g48120 serine/threonine phosphatase PP7, putative**
Length = 1338

Score = 74.3 bits (181), Expect = 2e-14
Identities = 43/125 (34%), Positives = 65/125 (51%)
Frame = +3

Query: 30 LLSTFVER*RPETHIFHISWGEITITLQDVAYHLGLHTDGNINSINGCTRDFQQFHGHPTWQ 209
L++ VER RPETH FH+ GEIT+TLQDV LGL DG ++ G T+ W
Sbjct: 84 LITALVERWRPETHTFHLPAGEITVTLQDVNILLGLRVDGPAVTGSTK-----YNWA 135
Query: 210 *VANSLSDKPPPQSEDRKSNFRLKMLWLRNKVAHIPRADEATFR*YRICYLLMLIGTYL 389
+ L P + S+ + + WLR ++P DE T + + ++L L+ +L
Sbjct: 136 DLCEDLLGHRPGPKDLHGSH--VSLAWLRENFRNLPADPDEVTLKCHTRAFVLALMSGFL 193
Query: 390 FTDES 404
+ D+S
Sbjct: 194 YGDKS 198

Anexo 4

Ah1TC4SA09

Sequences producing significant alignments:	Score	E
AT3g31340 Athila ORF 1, putative	(bits)	Value
	141	2e-34

>AT3g31340 Athila ORF 1, putative

Length = 781

Score = 141 bits (355), Expect = 2e-34
 Identities = 68/156 (43%), Positives = 102/156 (64%)
 Frame = -3

Query: 499 EMTKECSAILQRELPEKKDDPRSFYIPCTIGHITIEKSFCDLGASINLMPLSLMRKLQIS 320
 EM CSAI +PEK DP SF +PC IG E+ CDLGA +NLMPLS+ ++L I+
 Sbjct: 515 EMVSTCSAIPPATIPKELGDPGSFVLPKRIGKSAFERCLCDLGAGVNLMPMSKRLGIT 574

Query: 319 ELKSTRIVLQMDKSIKQALGVVENVLIKVGKFFLPADFVISDIEEDPNTPIILGRPFLA 140
 K +RI L +AD+S++ +G+ ENV ++VG F++P DFV+ +++++P+ P+ LGRPFL
 Sbjct: 575 NFKPSRISLILADRSVRFPVGLAENVHVRVGDYIPTDFVLELDKEPHDPLTLGRPFLN 634

Query: 139 TGRALIDAEKE*LLLRVHNEHLAFHVFKTMHEPTQE 32
 T A+ID + + L++ + L F + T PT E
 Sbjct: 635 TVGAIIDVRRSTINLQIGDFALEFDMKGRKNPTIE 670

Ah1TC4SB03

Sequences producing significant alignments:	Score	E
AT4g07600 NO TITLE	(bits)	Value
	166	3e-42

>AT4g07600 NO TITLE

Length = 630

Score = 166 bits (419), Expect = 3e-42
 Identities = 74/104 (71%), Positives = 85/104 (81%)
 Frame = +1

Query: 1 DQMLDRLSGKSHYCFLDGYTG YFQIHIAPEDQEKTFFTCPFGTYAYKRMFPGLCNVPATF 180
 DQML+RL+ +YCFLDGY+G+FQI I P D EKTFTTCP+GT+AY+RMPFGLCN PATF
 Sbjct: 379 DQMLERLANHPYYCFLDGYSGFFQIPIHPNDHEKTFFTCPYGTFAYERMPFGLCNAPATF 438

Query: 181 QRCMMSLFSDFIEDCIEVFMDDFSVMYGD SFILCLYGSSTVLTRC 312
 QRCM S+FSD IE+ +EVFMDDFSVMYG SF CL VLTRC
 Sbjct: 439 QRCMTSIFSDIIEEMVEVFMDDFSVMYGPSFSSCLLNLGRVLTRC 482

Ah1TC4SB11

Sequences producing significant alignments:	Score	E
At2g07730 putative non-LTR retroelement reverse transcriptase	(bits)	Value
	122	1e-30

>At2g07730 putative non-LTR retroelement reverse transcriptase

Length = 970

Score = 122 bits (306), Expect(2) = 1e-30
 Identities = 61/130 (46%), Positives = 81/130 (61%)
 Frame = -2

Query: 391 PASRRIYSGRGTPDNIIIAQEVLFHFMKNTKSKKGTLAFKIDLEKAYDRVYWRFLAHTLKS 212
 PA GR + DNI++ QE +H M+ K +KG + K+DLEKAYDR+ W FL TL +
 Sbjct: 249 PAQANFIPGRSLSDNIVLVQEAHVSMRRKKGRKGWMLLKLDEKAYDRIRWNFLQETLVA 308

Query: 211 FGFPISTINLIMNYVTASSLSILWNGSYLNGFTPSRGLRQ*DPMSPYLFLVLC TEQLACFI 32
 G + IM VT S+S+LWNG + F P+RGLRQ DP+SPYLFVLC E+L I
 Sbjct: 309 TGLSEVWTHRIMAGVTDPSMSVLWNGGKTD SFVPARGLRQGDPLSPYLFVLCLELRLCHLI 368

Query: 31 SHQVNLALWD 2
 + ++ WD
 Sbjct: 369 --EASVGTWD 376

Score = 27.3 bits (59), Expect(2) = 1e-30
 Identities = 11/26 (42%), Positives = 18/26 (68%)
 Frame = -3

Query: 447 IIRKVLGNRLRPHLAEIVGPLQGGFI 370
 II K++ NRL+ + +++GP Q FI
 Sbjct: 230 IITKMIVNRLKKVILKLGPAQANFI 255

Anexo 4

Ah1TC4SD01

Sequences producing significant alignments:
AT4g16910 retrotransposon like protein

Score E
(bits) Value
79 4e-17

>AT4g16910 retrotransposon like protein
Length = 687

Score = 79.3 bits (194), Expect(2) = 4e-17
Identities = 39/101 (38%), Positives = 60/101 (58%)
Frame = -2

Query: 372 LGRWGMRVALTTPHLSNLHDVHFVSQLQKYTPDASHVLEPESVQLREDLTLFPVAPVIIDDT 193
+G ++ L P L H+VFHVSQL+K + +E L+E++T+ PV I D
Sbjct: 566 VGAVAYKLDLPPKLDFAHNVFHVSQLRKCLSEQEESMEDVPPGLKENMTVEAWPVRIMDQ 625
Query: 192 SIKRLRGKEVSLVKVAWSRGGVEEHTWELESEMRTDYPHLF 70
K RGK + L+K+ W+ GG EE+TWE E++M+ ++P F
Sbjct: 626 MKKGTTRGKSMDLLKILWNCGGREEYTWETETKMKANFPPEWF 666

Score = 24.3 bits (51), Expect(2) = 4e-17
Identities = 7/14 (50%), Positives = 12/14 (85%)
Frame = -1

Query: 397 GPFQILERIGPVG 356
GP++++ER+G V Y
Sbjct: 558 GPKVIERVGAVAY 571

Ah1TC4SH06

Sequences producing significant alignments:
At2g14400 putative retroelement pol polyprotein

Score E
(bits) Value
143 2e-34

>At2g14400 putative retroelement pol polyprotein
Length = 1466

Score = 143 bits (360), Expect = 2e-34
Identities = 68/128 (53%), Positives = 97/128 (75%)
Frame = -3

Query: 401 C*GVMPFGLKNAGATYQRLMNKVFTDHIRKVMVEYVDDMLVKTQSEESLLSDLTQVFDTI 222
C VMPFGL+NAGATY RL+NK+F++H+ K MEVY+DDML+K+ +E + L + F +
Sbjct: 583 CYKVMFPGLRNAGATYPRVLNKMFSHVGVKTMVEYIDDMLIKSLKKEDHVKHLEECFAIL 642

Query: 221 RRHDM*LNPACYTFAVEAEKFLGFMLTQRGIEANSDKCRAILNMKSPTCVKEVKQLNGRL 42
++ M LNPAK TF V + +FLG+++T+RGIEAN ++ A LNM SP KEV++L GR+
Sbjct: 643 NQYQMKLNPAKCTFGVPSGEFLGYIVTKRGIEANPNQINAFLNMPSPKNFKEVQRLTGRI 702

Query: 41 AALSRFLA 18
AAL+RF++
Sbjct: 703 AALNRFIS 710

Ah1TC4SG07

Sequences producing significant alignments:
At1g10900 phosphatidylinositol-4-phosphate 5-kinase isolog
At1g60890 putative phosphatidylinositol-4-phosphate 5-kinase

Score E
(bits) Value
69 3e-13
61 9e-11

>At1g10900 phosphatidylinositol-4-phosphate 5-kinase isolog
Length = 859

Score = 69.3 bits (168), Expect = 3e-13
Identities = 28/39 (71%), Positives = 32/39 (81%)
Frame = -1

Query: 274 YIGTWSKGLKDGKGFYPTGSKQPSLKKWCSSLNSENNG 158
Y GTWS+GLKDGKG FYP G+KQPSLKKWC SL ++ G
Sbjct: 282 YYGTWSRGLKDGKGVFYPAGTKQPSLKKWCRSLEYDDTG 320

Score = 26.9 bits (58), Expect = 1.9
Identities = 8/18 (44%), Positives = 14/18 (77%)
Frame = -1

Query: 280 EFYIGTWSKGLKDGKGF 227
+ Y G W +GL+DG+G++
Sbjct: 211 DLYDGLWKEGLQDGRGSY 228

Anexo 4

Ah1TC4SG11

Sequences producing significant alignments:

At2g12210 putative TNP2-like transposon protein

Score	E
(bits)	Value
83	2e-17

>At2g12210 putative TNP2-like transposon protein

Length = 889

Score = 83.2 bits (204), Expect = 2e-17
 Identities = 45/105 (42%), Positives = 63/105 (59%)
 Frame = -1

Query: 318 MTAAEKSMFLGVLTTKLPDGSASNISRCVHLAERKMGYKTHDAHLLHYLLPIPIKSI 139
 ++ +EK L K PDG SNISR V L + K+ G K+HD H L+ LLP+ + +
 Sbjct: 624 LSKSEKKTFCRRLFEEFKGPDGYCSNISRGVSLDCKIMGLKSHDYHVLMMQQLLPVALMGL 683
 Query: 138 LPDHVAIPLIRLSSFFRRICEKLITLEEIDRLESEIVETL*HLER 4
 LP +IRL SFF +C+++I +E I +E+EIVETL ER
 Sbjct: 684 LPKGPRTAIRLCSFFNHLQCRVIDIEVISVMEAEIVETLCMFER 728

Ah1TC5SG01

Sequences producing significant alignments:

AT4g03650 putative reverse transcriptase

Score	E
(bits)	Value
120	2e-28

>AT4g03650 putative reverse transcriptase

Length = 839

Score = 120 bits (300), Expect = 2e-28
 Identities = 54/91 (59%), Positives = 67/91 (73%)
 Frame = +1

Query: 1 DPTKVEAVMDWKQPTTITEIRSFLGLAGYRRFIKGFSQLALPMTKLTRKDTPFVWVTEPEC 180
 DP K+EA+ DW +PT TEIRSFLGLAGYRRFIKGF+ +A PMTKLT KD PFVW+PEC
 Sbjct: 627 DPEKIEAIRDWRPTNATEIRSFLGLAGYRRFIKGFASMAQPMTKLTGKDVPFVWSPEC 686

Query: 181 DDSFHTLNTQLTTATVFLVLPDPHPSEPHCD 273
 ++ F +L LT+ V LP+ +P + D
 Sbjct: 687 EEGFVSLKEMLTSTPVLALPEHGEPYMVYTD 717

Ah1TC5SG10

Sequences producing significant alignments:

At1g20390 hypothetical protein

Score	E
(bits)	Value
56	4e-09

>At1g20390 hypothetical protein

Length = 1791

Score = 55.8 bits (133), Expect = 4e-09
 Identities = 31/83 (37%), Positives = 44/83 (52%)
 Frame = +2

Query: 2 INPLKCAFGMSVGNFLGFVVHXXXXXXXXXXXXXXXXLALPAPKSKKAVQSFLGKVNYLRRF 181
 +NP KC FG++ G FLG+VV L LP+P++ + VQ G++ L RF
 Sbjct: 962 LNPTKCTFGVTSGEFLGYVVTKRGIEANPKQIRAILLELPSPRNAREVQRLTGRIAAALNRF 1021

Query: 182 ISNLSDRTRVFAPLVTLKSDSQF 250
 IS +D+ F L LK +QF
 Sbjct: 1022 ISRSTDKCLPFYNL--LKRRAQF 1042

Anexo 4

Ah1TC6SE09

Sequences producing significant alignments:
AT4g07660 putative athila transposon protein

Score E
(bits) Value
66 4e-12

>**AT4g07660 putative athila transposon protein**
Length = 724

Score = 65.9 bits (159), Expect = 4e-12
Identities = 29/44 (65%), Positives = 33/44 (74%)
Frame = +1

Query: 1 DQEKTTFTCPGVSAYRTMPFALCSALATFHRCMLSISSDMVPK 132
DQEKTTFTCP G AY+ MPF LC+A TF RCM SI SD++ K
Sbjct: 472 DQEKTTFTCPYGTFAFKRMPFGLCNAPTTFQRCMTSIFSDLIEK 515

Ah1TC6SF11

Sequences producing significant alignments:
At1g60750

Score E
(bits) Value
87 2e-18

>**At1g60750**
Length = 330

Score = 86.7 bits (213), Expect = 2e-18
Identities = 51/114 (44%), Positives = 59/114 (51%)
Frame = +1

Query: 1 DLYYQHIDTKVPIEVTV*FFSLIWXXXXGY*NNENRDRL*TSFSL**KVKMYIL**L 180
DLYYQH+IDT +PIE+T+
Sbjct: 127 DLYYQHRIDTTLPIEITI----- 144
Query: 181 DSFSFVIFINIGDLKKLVEEGKVYIGLSDASASKIRKAHAVHPITAVQIEWSL 342
G+LKKLVEEGK+KYIGLS+ASAS IR+AHAVHPITAVQIEWSL
Sbjct: 145 -----GELKKLVEEGKIKYIGLSEASASTIRRAHAVHPITAVQIEWSL 187

Ah1TC6SH12

Sequences producing significant alignments:
At2g07660 putative retroelement pol polyprotein

Score E
(bits) Value
168 4e-43

>**At2g07660 putative retroelement pol polyprotein**
Length = 949

Score = 168 bits (426), Expect = 4e-43
Identities = 76/108 (70%), Positives = 91/108 (83%)
Frame = +1

Query: 1 DPRFTSRFWGAFQKAFGTRLSLSTAYHPQTDGQSERTIQTLEDMLRACVLDQPASWDRYM 180
D RFTS+FW AFQKA GTR++LSTAYHPQTDGQSERTIQTLEDMLRACVLD +W++Y+
Sbjct: 694 DTRFTSKFWNAFQKALGTRVNLSTAYHPQTDGQSERTIQTLEDMLRACVLDWGGNWEKYL 753
Query: 181 PLVEFAYNNS*HASIGMAPYEALYGKCCQSPLCWYEPGEKGLLRPEMI 324
L+EFAYNNS ASIGM+PYEALYG+ C++PLCW GE+ L P ++
Sbjct: 754 RLIEFAYNNSFQASIGMSPYEALYGRACRTPLCWTPVGERRLFGPTIV 801