

CARACTERIZAÇÃO E MAPEAMENTO DE RGAS EM *Arachis*

RESUMO

O amendoim (*Arachis hypogaea*) está entre as principais oleaginosas cultivadas em todo o mundo. É bastante susceptível a diversos estresses bióticos e abióticos, para os quais muitas das espécies silvestres são resistentes. O processo de introgressão de genes úteis de espécies silvestres em amendoim pode ser acelerado pelo uso de seleção assistida por marcadores moleculares, por meio da identificação de marcadores proximamente ligados a genes de interesse como os de resistência a doenças. QTLs (*Quantitative Trait locus*) para resistência à mancha foliar causada pelo fungo *C. personatum* foram mapeados em mapa genético de referência construído para o cruzamento entre *A. duranensis* e *A. stenosperma*. Utilizou-se o RGA (*Resistance Genes Analog*) S1_A_36 que cossegrega com um destes QTLs para identificar clones positivos nas bibliotecas BAC (*Bacterial Artificial Chromosome*) de espécies silvestres do gênero *Arachis*. A subclonagem e o sequenciamento do BAC que contém este RGA possibilitaram a caracterização da sequência, identificando, com o auxílio do Programa de predição RiceGAAS, 20 ORFs distribuídas em oito *contigs*. Destas, cinco ORFs estavam relacionadas a genes de resistência pertencentes à família TIR-NBS-LRR e quatro ORFs a elementos de transposição, sendo um transposon e três retrotransposons LTR do tipo Ty1-Copia. A identificação de marcadores mais robustos e transferíveis, associados a locos que conferem resistência, possibilitará a seleção assistida nos programas de melhoramento genético, o que acelerará o processo de introgressão de genes de resistência no amendoim.

ABSTRACT

Peanut (*Arachis hypogaea*) is one of the major oilseed crops worldwide. It is very susceptible to various biotic and abiotic stresses, to which many wild species are resistant. Introgression of useful genes from wild species in peanut can be accelerated by the use of molecular marker assisted selection (MAS), through the identification of markers closely linked to genes of interest such as resistances to diseases. QTLs for resistance to leaf-spot caused by the fungus *C. personatum* were mapped in the reference genetic map from the cross between *A. duranensis* and *A. stenosperma*. The RGA S1_A_36 that co-segregated with one QTL was used to identify positive clones in BAC libraries of two wild species of *Arachis*. The subcloning and sequencing of a full BAC containing the RGA enabled the characterization of this sequence, and its annotation using the gene prediction program RiceGAAS. In total 20 ORFs were found distributed in eight *contigs*. Of these, five ORFs contained resistance genes analogs of the TIR-NBS-LRR family and four contained transposable elements, one DNA transposon and three retrotransposons of the LTR Ty1-Copia type. The identification of such markers associated with resistance loci is of great importance as it has the potential of speeding up the process of introgression of resistance genes in peanut.

1. INTRODUÇÃO

O gênero *Arachis* L., família Fabaceae, é nativo da América do Sul. O Brasil central é seu provável centro de origem, em uma região que se estende do sudoeste do estado de Mato Grosso do Sul até o sul de Goiás (Valls, 2000). Sua distribuição natural é restrita ao Brasil, Bolívia, Paraguai, Argentina e Uruguai (Valls & Simpson, 1994). Este gênero contém 80 espécies descritas e 64 destas ocorrem no Brasil, sendo 47 exclusivas da flora brasileira (Valls & Simpson, 2005). De acordo com dados citogenéticos, existem três tipos de genomas, A, B e D, que diferenciam as espécies em relação aos cromossomos. A maioria das espécies do gênero é diploide, algumas aneuploides e cinco espécies tetraploides incluindo o amendoim, a espécie cultivada, *Arachis hypogaea* (Krapovickas & Gregory, 1994; Lavia, 1998; Valls & Simpson, 2005).

O amendoim está entre as principais oleaginosas cultivadas no mundo e as maiores causas de perda na produção são a seca e fungos causadores de manchas foliares como a cercosporiose (Godoy et al., 1994). Embora estejam disponíveis acessos de *A. hypogaea* com diferentes níveis de resistência, o melhoramento das cultivares é dificultado devido à baixa variabilidade genética em relação a características agrônomicas de interesse. Por outro lado, espécies silvestres do gênero *Arachis* possuem resistência a diferentes fatores bióticos e abióticos, apresentando uma maior variabilidade genética (Hilu & Stalker, 1995; Moretzsohn et al., 2004; Milla et al., 2005).

Por constituírem importante fonte de genes úteis e apresentarem alto polimorfismo, espécies silvestres vêm sendo introduzidas em programas de melhoramento. O processo de introgressão de genes pode ser acelerado pelo uso da seleção de plantas assistida por marcadores moleculares e para tal, é necessário identificar marcadores proximamente ligados a genes de interesse como resistência a doenças (RGA). Muitos destes genes são caracterizados pela presença do domínio NBS-LRR que, em plantas, tem sua função associada à resistência a doenças.

Primers degenerados para região NBS foram produzidos para diferentes espécies de *Arachis* visando à identificação de sequências do tipo RGA (Bertioli *et al.*, 2003). Alguns desses RGAs foram posicionados em um mapa genético construído para uma população de genoma AA (*A. duranensis* x *A. stenosperma*) (Moretzsohn *et al.*, 2005). QTLs para resistência à mancha foliar causada pelo fungo *Cercosporidium personatum* foram posicionados no mesmo mapa, sendo que o RGA S1_A_36 apresentou cossegregação com este QTL, no Grupo de Ligação 4. Utilizou-se esta sequência como sonda para identificar clones positivos nas bibliotecas BAC construídas para as prováveis espécies silvestres parentais do amendoim, *A. duranensis* (genoma AA) e *A. ipaënsis* (genoma BB) (Guimarães *et al.*, 2008). Foi encontrado um clone positivo na biblioteca BAC da espécie silvestre *A. duranensis* que, em virtude de sua localização na biblioteca, foi denominado AD25F09 (AD=*Arachis duranensis*, placa 25, posição F09). A subclonagem e o sequenciamento do clone BAC que contém este RGA possibilitarão a caracterização da sequência, além da identificação de microssatélites na região designada, provendo marcadores mais próximos ao QTL já identificado. Esses marcadores podem ser utilizados no processo de seleção assistida (SAM) visando o melhoramento do amendoim, permitindo também que vários genes sejam agrupados em um mesmo genótipo (piramidação dos genes de resistência).

O sequenciamento aleatório aliado a uma estratégia de clonagem randômica (*shotgun*) é interessante e possui um relevante poder informativo. Nesta estratégia, como demonstrada na Figura 1, todo o DNA genômico, ou alguns trechos deste, como BACs, são sonificados, nebulizados ou cortados com enzimas de restrição, e os fragmentos resultantes são subclonados e sequenciados. As sequências de nucleotídeos dos fragmentos são então montadas e analisadas, com o intuito de descrever a estrutura e composição nucleotídica. Dentre as informações provenientes do sequenciamento aleatório podem-se citar a distribuição de microssatélites, elementos transponíveis e retrotransponíveis, genes, pseudogenes, regiões promotoras, conteúdo nucleotídico, tRNAs, e outras.

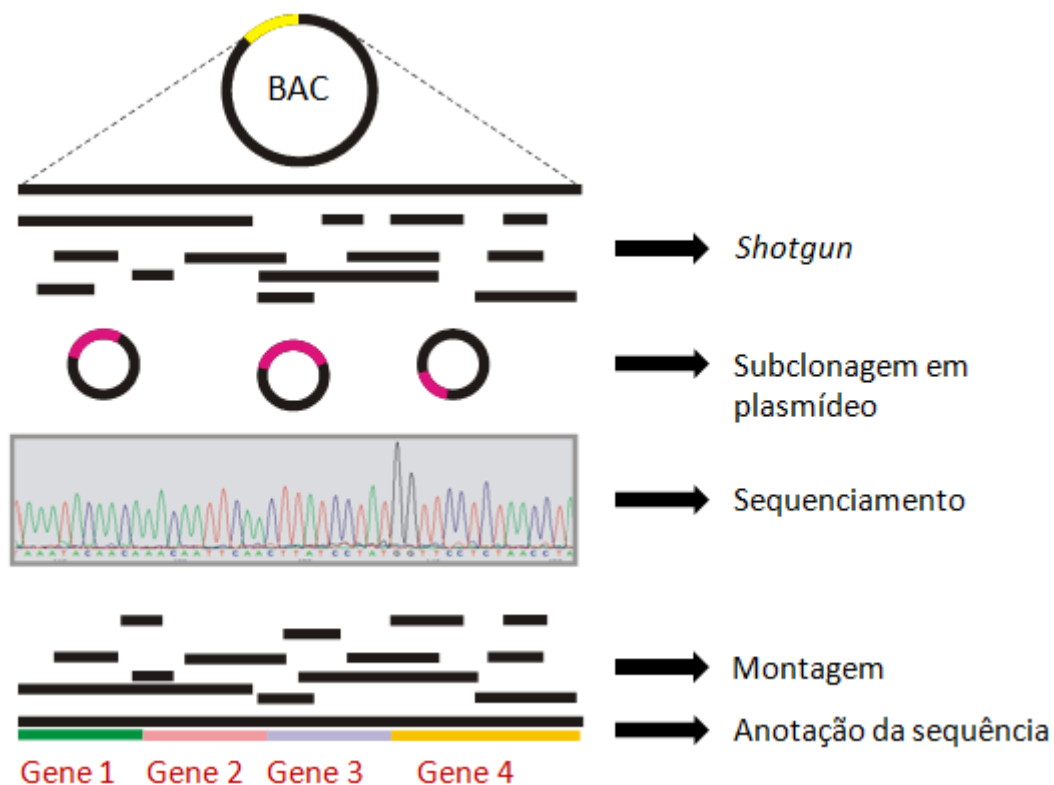


Figura 1: Método de subclonagem de BAC por meio da técnica de shotgun.

2. MATERIAL E MÉTODOS

Os experimentos foram conduzidos no Laboratório de Biotecnologia (Pós-graduação da Universidade Católica de Brasília) e na Embrapa Recursos Genéticos e Biotecnologia. As plantas utilizadas neste trabalho foram fornecidas pelo Banco Ativo de Germoplasma (BAG) de espécies de *Arachis* situado na Embrapa Recursos Genéticos e Biotecnologia, Brasília, DF. Os clones BAC utilizados foram selecionados a partir de biblioteca BAC construídas no CIRAD (Montpellier – França) (Guimarães *et al.*, 2008).

2.1 Extração do DNA genômico de *Arachis* e população de mapeamento

O protocolo baseado em CTAB publicado por Ferreira & Grattapaglia, (1998) foi utilizado neste trabalho. Para os experimentos de mapeamento realizados, utilizaram-se o DNA genômico dos parentais, F₁ e F₂ (93 indivíduos) da população de

mapeamento AA (Moretzsohn *et al.*, 2005); os parentais da população de mapeamento BB (Moretzsohn *et al.*, 2009) e quatro indivíduos tetraploides, sendo dois deles sintéticos.

2.2 Subclonagem do BAC AD25F09

Para realizar a subclonagem foram utilizadas duas estratégias distintas para a fragmentação do BAC AD25F09. A primeira consistiu na utilização de três conjuntos de enzimas de restrição para uma geração de diferentes perfis, o que ajudaria na sobreposição das sequências. Realizou-se a segunda estratégia de fragmentação por meio do processo de nebulização.

2.2.1 Estratégia de subclonagem 1 – Fragmentação por Enzimas de Restrição

Em experimentos distintos, o BAC (aproximadamente 10 µg) foi digerido com 2 U de enzima de restrição (*HindIII*, *BamHI* e *Sau3AI* (Invitrogen) de acordo com o protocolo do fabricante e em seguida precipitado. Aproximadamente 20 µg do vetor pBlueScript SK- foi linearizado utilizando-se 1 U de enzima de restrição (*HindIII* ou *BamHI*) (Invitrogen) e desfosforilado e em seguida submetido a uma eletroforese em gel de agarose 1% corado com brometo de etídeo. A banda linearizada foi retirada do gel e purificada com o auxílio do Kit *GENE CLEAN II* (Q-BIO gene). Para reação de ligação do inserto ao vetor, foram utilizados 6 µL do BAC (aproximadamente 60 ng/ µL); 0,5 µL de plasmídeo (aproximadamente 50 ng/µL); 2 µL de 5X Rapid Ligation Buffer-T4 DNA ligase (Invitrogen); 1 U de T4 ligase (Invitrogen) e 0,5 µL de água estéril para 10 µL de volume final. A reação foi incubada durante 24 horas, a 16° C.

2.2.2 Estratégia de Subclonagem 2 – Fragmentação por nebulização

O método de fragmentação mecânica utilizado (nebulização) difere do enzimático, principalmente por gerar fragmentos de DNA cujas extremidades precisam ser reparadas por meio da utilização das enzimas *T4 polimerase* e

fragmento *Klenow*. Após a nebulização de aproximadamente 20 µg do BAC e o tratamento dos fragmentos resultantes, estes foram ligados ao vetor aberto pCR 4Blunt-TOPO, por meio da técnica proposta pelo *TOPO[®] Shotgun Subcloning Kit*, que utiliza propriedades da topoisomerase (Shuman, 1994).

2.3 Extração de subclones derivados do BAC AD25F09 e sequenciamento

As reações de transformação foram realizadas utilizando-se a cepa XL1-Blue da bactéria *Escherichia coli*, da mesma forma para ambos os experimentos. A extração de DNA plasmidial (método de lise alcalina) foi feita individualmente (Ahn *et al.*, 2000) e em placas de 96 poços (Sambrook & Russel, 2001). As reações de sequenciamento foram realizadas utilizando o Kit BigDye Terminator cycle Sequencing (Applied Biosystems, CA, USA) em um sequenciador automático ABI377 (Applied Biosystems) e utilizando os *primers* universais T7/T3 e M13F/M13R.

2.4 Análise dos dados de sequenciamento

As sequências foram processadas utilizando-se o programa Staden Package (Staden *et al.*, 1996) e a ferramenta *Cap3* (Huang & Madan, 1999) dentro da interface do Staden, para realizar a compilação, retificação, leitura e adição de sequências.

Para a identificação das ORFs, utilizou-se o sistema de anotação automática do arroz, o RiceGAAS (*Rice Genome Automated Annotation System* - <http://ricegaas.dna.affrc.go.jp/>), (Sakata *et al.*, 2002), o qual é composto por programas de predição e análises de códigos de proteínas e estruturas de genes. Com o intuito de gerar informações mais seguras acerca da sequência do BAC, realizou-se uma compilação de análises, incluindo comparações com sequências GSS (*Genome Survey Sequence*) do gênero *Arachis* depositadas no Banco de Dados do NCBI (*National Center for Biotechnology Information* - (<http://www.ncbi.nlm.nih.gov/>), além de sequências ESTs (*Expressed Sequence Tags*) de *Arachis*, podendo indicar genes putativos nas sequências anônimas.

3. RESULTADOS E DISCUSSÃO

3.1 Subclonagem e caracterização do BAC AD25F09

O método de subclonagem que utilizou a fragmentação do BAC por nebulização, apesar de mais difícil quando comparado com o enzimático, foi mais eficiente, otimizando tempo e produzindo cinco vezes mais colônias do que o método que utilizou a restrição enzimática, gerando, dessa forma, 2.592 sequências, como observado na Tabela 1.

Tabela 1: Número de colônias recombinantes e sequências obtidas por meio das diferentes estratégias de subclonagem.

Subclonagem do BAC AD25F09		
Estratégia 1: Enzimas de restrição	Nº de colônias	Nº de sequências
BAC – <i>Hind</i> III / Vetor – <i>Hind</i> III	305	352
BAC – <i>Sau</i> 3AI / Vetor – <i>Bam</i> HI	60	100
BAC – <i>Bam</i> HI / Vetor – <i>Bam</i> HI	66	66
Estratégia 2: Nebulização	1.536	2.592
TOTAL	1.967	3.110

Ao reunir as duas metodologias de fragmentação do BAC, obtiveram-se 3110 sequências que foram agrupadas em 505 contigs (203.694 pb), na tentativa de remontar o BAC inteiro. O grande número de sequências obtidas por meio da subclonagem foi bastante útil para a análise, pois uma grande parte do BAC pôde ser remontada utilizando-se 102.256 pb (referentes aos oito maiores contigs), ou seja, metade do número de pares de base obtida.

Os oito maiores *contigs* com tamanhos variando entre 3 e 28 Kb foram utilizados neste estudo para a predição de possíveis genes encontrados no BAC. Com o auxílio do Programa de predição RiceGAAS, foram identificadas 20 ORFs distribuídas na sequência de oito *contigs* referentes à subclonagem e ao sequenciamento do BAC AD25F09. Exatamente 50% dessas regiões foram identificadas como proteínas hipotéticas, 25% com proteínas relacionadas à resistência, 20% com proteínas relacionadas a elementos transponíveis e 5% relacionadas a outros tipos de proteínas (Figura 2).

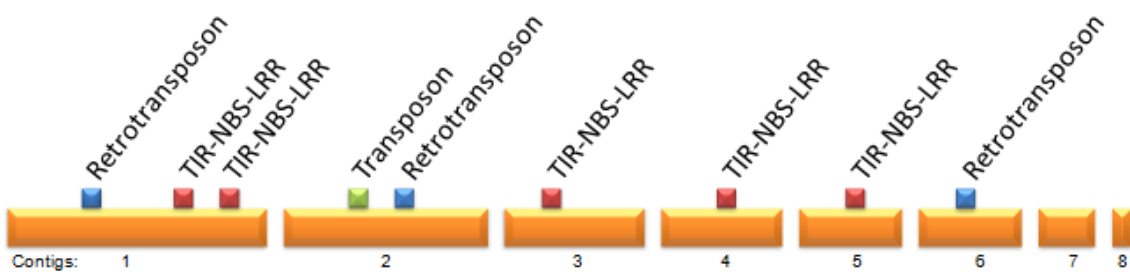


Figura 2: Análise da predição do conteúdo gênico do BAC AD25F09.

Foram identificadas neste trabalho, quatro ORFs relacionadas a elementos de transposição, sendo uma correspondente a um transposon e três a retrotransposons. A partir de comparações com o banco de dados GenBank e predição de função por meio do programa AutoPredgeneset, os retrotransposons identificados pertenciam ao grupo LTR do tipo Ty1-Copia, distinto de outro grupo principalmente em virtude da ordem dos genes. A identificação de sequências repetitivas flanqueando a ORF (representadas por setas flanqueando a ORF 2 situada no *Contig* 1 – Gráfico 2) revela o fato destas possivelmente serem as porções LTR do elemento, portanto um elemento inteiro pode ter sido caracterizado. Comparações com ESTs revelam que estes podem estar transcricionalmente pouco ativos neste BAC. A existência de elementos transponíveis dentro de genes pode acarretar na perda da função gênica. O BAC subclonado possui vários genes de resistência, porém, intercalados por elementos transponíveis. A co-localização do RGA contido no BAC com um QTL revela o fato de que esses genes possivelmente podem estar ativos.

Utilizando a técnica de sequenciamento por meio da fragmentação randômica O sequenciamento do BAC AD25F09 que contém um RGA que cossegrega com um QTL relacionado à mancha foliar, mostrou-se bastante eficiente, pois possibilitou acessar e descrever aspectos importantes, além de ter mostrado ser uma estratégia alternativa para se obter um panorama geral de parte do genoma de *Arachis*. Este trabalho pode, ainda, tornar-se importante para a ancoragem do mapa genético com um futuro mapa físico a ser desenvolvido para *Arachis*. O mapeamento físico de *Arachis* possibilitará a localização cromossômica de genes de interesse, auxiliando no processo de melhoramento assistido, além de

possibilitar a análise genômica comparativa entre *Arachis* e outras espécies, principalmente às leguminosas.

Os programas de melhoramento de *A. hypogaea* têm conduzido a avanços significativos, entretanto, a inclusão de genes oriundos de espécies selvagens relacionadas possibilitaria o aumento de variabilidade genética para várias características agrônômicas de interesse, principalmente a resistência a doenças que afetam o amendoim. A futura identificação de marcadores associados a locos que conferem resistência permitirão a seleção assistida por marcadores nos programas de melhoramento genético do amendoim cultivado, o que acelerará o processo de introgressão dos genes de resistência pelo método de retrocruzamento. Uma vez que serão mapeados genes de resistência tanto em espécies possuidoras de genoma A como em espécies de genoma B, plantas AA com genes de resistência poderão ser cruzadas com plantas BB também resistentes, obtendo-se plantas AB. Estas, após poliploidização, poderão ser cruzadas com *A. hypogaea*, para introgressão desses genes.

4. CONCLUSÕES

- As metodologias de subclonagem do BAC AD25F09 utilizadas neste trabalho foram satisfatórias e garantiram uma grande porcentagem do BAC sequenciada.
- A análise do conteúdo gênico da sequência do BAC contendo o RGA identificou cinco regiões relacionadas aos genes de resistência pertencentes à família TIR-NBS-LRR, além de um transposon e três retrotransposons LTR do tipo Ty1-copia.
- Seria importante a comparação do mapa genético com um mapa físico.
- Para S.A.M. será necessário **mapear** marcadores proximamente ligados aos RGAs.
- A identificação de genes relacionados com a resistência ao fungo foliar *C. personatum*, além de outras doenças, é fundamental para auxiliar os programas de melhoramento. Assim, espera-se no futuro que a seleção de genótipos resistentes seja auxiliada pelas informações moleculares, permitindo também a piramidação destes genes e, como conseqüência, a diminuição dos custos para o agricultor com

inseticidas e os danos causados ao meio ambiente pela aplicação destes defensivos agrícolas.

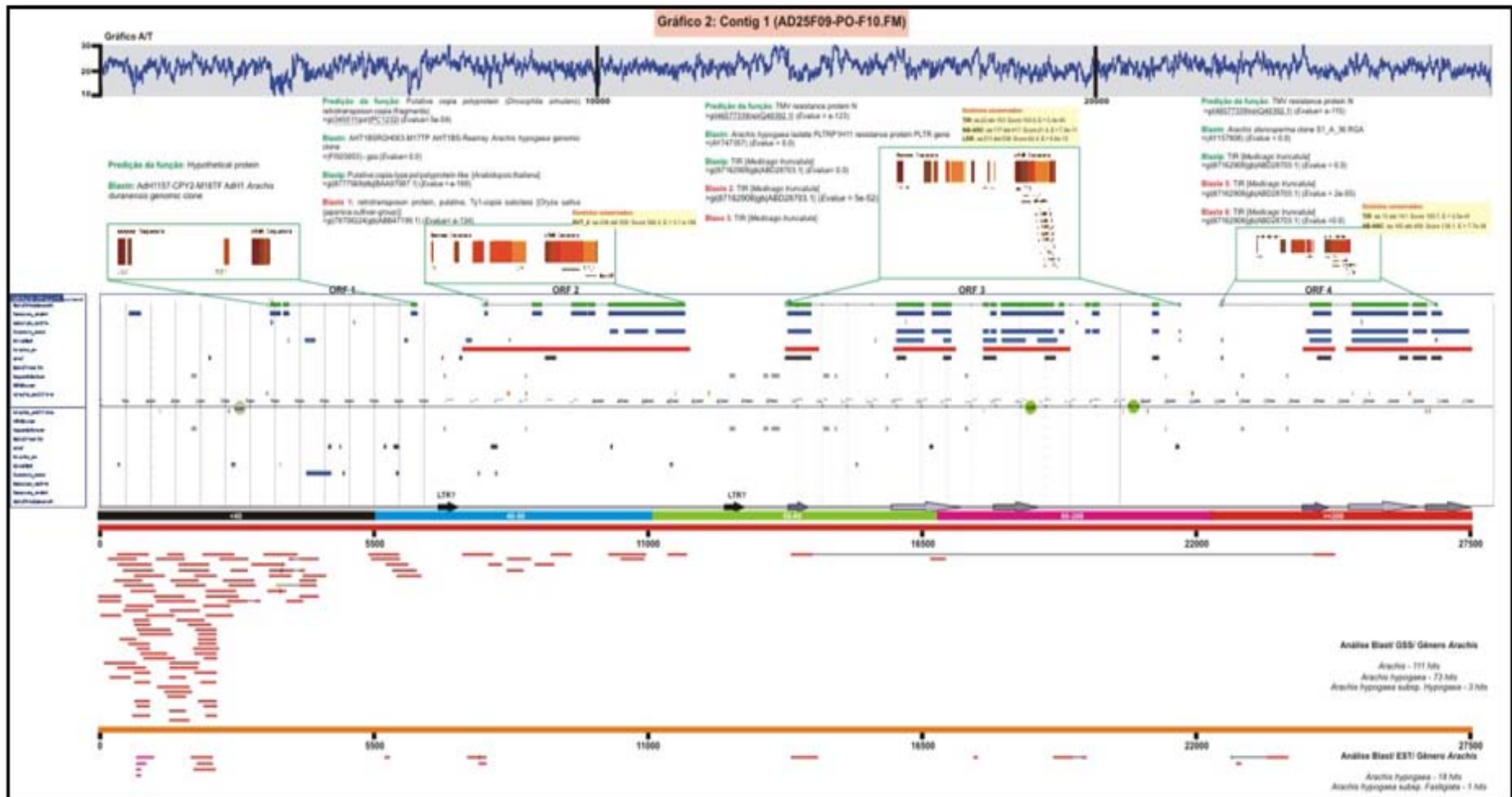


Gráfico 1: Compilação de informações acerca do sequenciamento e da anotação do *Contig 1* do BAC AD25F09. O gráfico contém comparações da sequência com outras sequências depositadas em bancos de dados, sequências GSS e ESTs, gráficos A/T e localização de microssatélites (círculos em verde).

5. BIBLIOGRAFIA

- Ahn SC, Baek BS, Oh T, Song CS, Chatterjee B. Rapid mini-scale plasmid isolation for DNA sequencing and restriction mapping. *BioTechniques*. 2000 Set; 29(3):466-468.
- Castelo AT, Martins W et al. TROLL-tanden repeat occurrence locator. *Bioinformatics*. 2002;(18):634-636.
- Ferreira ME, Grattapaglia D. Introdução ao uso de marcadores RAPD e RFLP em análise genética. Brasília: Embrapa-Cenargen. 1998 ;(3)220.
- Godoy IJ, Moraes SA, Pereira JCVNA, Martins ALM, Paulo EM. Cultivares de amendoim: potencial reprodutivo, resistência e resposta ao controle de manchas foliares. *Fitopatologia Brasileira*. 1994:19:288.
- Guimarães PM, Garsmeur O, Proite K, Leal-Bertioli SCM, Seijo G, Chaine C, et al. BAC libraries construction from the ancestral diploid genomes of the allotetraploid cultivated peanut. *BMC Plant Biol*. 2008;8:14.
- Hilu KW, Stalker HT. Genetic relationships between peanut and wild species of *Arachis* sect *Arachis* (Fabaceae): Evidence from RAPDs. *Plant Systematics and Evolution*. 1995;198(3-4):167-178.
- Huang X, Madan A. CAP3: A DNA sequence assembly program. *Genome Res*. 1999 Set;9(9):868-877.
- Kosambi D. The estimation of map distance from recombination values. *Annuaire of Eugenetics*. 1944;(12):172-175.
- Krapovickas A, Gregory WC. Taxonomia del genero *Arachis* (Leguminosae). *Bonplandia* (Argentina). 1994;8(1-4):1-186.
- Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L. MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics*. 1987;1(2):174-181.
- Lavia GI. Karyotypes of *Arachis palustris* and *A. praecox* (Section *Arachis*), two species with basic chromosome number $x = 9$. *Cytologia*. 1998;63(2):177-181.
- Milla SR, Isleib TG, Stalker HT. Taxonomic relationships among *Arachis* sect. *Arachis* species as revealed by AFLP markers. *Genome*. 2005 Fev;48(1):1-11.
- Moretzsohn MDC, Hopkins MS, Mitchell SE, Kresovich S, Valls JFM, Ferreira ME. Genetic diversity of peanut (*Arachis hypogaea* L.) and its wild relatives based on the analysis of hypervariable regions of the genome. *BMC Plant Biol*. 2004 Jul 14;4:11.
- Moretzsohn MC, Leoi L, Proite K, Guimarães PM, Leal-Bertioli SCM, Gimenes MA, et al. A microsatellite-based, gene-rich linkage map for the AA genome of *Arachis* (Fabaceae). *Theor Appl Genet*. 2005 Out;111(6):1060-71.
- Moretzsohn MC. Desenvolvimento e mapeamento de marcadores microssatélites e identificação de QTLs ligados à produtividade e à resistência à mancha preta em *Arachis* spp. Tese de doutorado, Universidade de Brasília, Brasília, 2006;134 p.

- Moretzsohn M, Barbosa A, Alves-Freitas D, Teixeira C, Leal-Bertioli S, Guimarães P, et al. A linkage map for the B-genome of *Arachis* (Fabaceae) and its synteny to the A-genome. *BMC Plant Biol.* 2009 Abr 7;9(1):40.
- Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* 2000;132365-386.
- Sakata K, Nagamura Y, Numa H, Antonio BA, Nagasaki H, Itonuma A, et al. RiceGAAS: an automated annotation system and database for rice genome sequence. *Nucleic Acids Res.* 2002 Jan 1;30(1):98-102.
- Shuman, S. (1994). Novel Approach to Molecular Cloning and Polynucleotide Synthesis Using Vaccinia DNA Topoisomerase. *J. Biol. Chem.* 269, 32678-32684.
- Sambrook J, Russel DW. *Molecular cloning: a laboratory manual.* 3.ed. New York: Cold Spring Harbor Laboratory. 2001;(3)
- Staden R. The Staden sequence analysis package. *Molecular Biotechnology.* 1996;5(3):233-241.
- Valls JFM, Simpson CE. Taxonomy, natural distribution and attributes of *Arachis*. In: Kerridge, PC, Hardy B. (Ed.). *Biology and agronomy of forage Arachis.* Centro Internacional de Agricultura Tropical. 1994:1-18.
- Valls JFM. Diversidade genética no gênero *Arachis* e a origem do amendoim. In: Bandel G, Aguiar-Perecin MLR, Oliveira, G.C.X. (Ed.). *Encontro sobre temas de genética, melhoramento E genômica: uma abordagem em plantas*". 2000;(17). *Anais. Piracicaba: ESALQ.* 2000:51.
- Valls JFM, Simpson CE: New species of *Arachis* L. (Leguminosae) from Brazil, Paraguay and Bolivia. *Bonplandia.* 2005;(14):35-64.
- Valls, JFM. Recursos Genéticos do Gênero *Arachis*. In: Santos, R. C. (Ed). *O agronegócio do amendoim no Brasil.* Campina Grande: Embrapa Algodão. 2005:45-70.