

SAMUEL MAZZINGHY ALVARENGA

**CARACTERIZAÇÃO DE SEQÜÊNCIAS EXPRESSAS DO
GENOMA CAFÉ POTENCIALMENTE RELACIONADAS
COM A RESISTÊNCIA A DOENÇAS**

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

**VIÇOSA
MINAS GERAIS – BRASIL
2007**

**Ficha catalográfica preparada pela Seção de Catalogação e
Classificação da Biblioteca Central da UFV**

T

A473c
2007

Alvarenga, Samuel Mazzinghy , 1981-
Caracterização de seqüências expressas do genoma café
potencialmente relacionadas com a resistência a doenças /
Samuel Mazzinghy Alvarenga. – Viçosa, MG, 2007.
xiv, 107f. : il. (algumas col.) ; 29cm.

Orientador: Ney Sussumu Sakiyama.
Dissertação (mestrado) - Universidade Federal de
Viçosa.
Inclui bibliografia.

1. Café - Genética. 2. Genética molecular. 3. Marcadores
genéticos. 4. Seqüenciamento de nucleotídeo. 5. Café -
Resistência a doenças e pragas. 6. Bioinformática.
I. Universidade Federal de Viçosa. II. Título.

CDD 22.ed. 633.732

SAMUEL MAZZINGHY ALVARENGA

**CARACTERIZAÇÃO DE SEQÜÊNCIAS EXPRESSAS DO
GENOMA CAFÉ POTENCIALMENTE RELACIONADAS
COM A RESISTÊNCIA A DOENÇAS**

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

APROVADA: 16 de julho de 2007

Pesq. Eveline Teixeira Caixeta
(Co-Orientadora)

Pesq. Eunize Maciel Zambolim
(Co-Orientadora)

Prof. Cosme Damião Cruz

Pesq. Felipe Rodrigues da
Silva

Prof. Ney Sussumu Sakiyama
(Orientador)

“Você faz suas escolhas, e as suas escolhas fazem você”.

Steve Beckman

Aos meus pais, Aurelino Ferreira Alvarenga e Maria de Jesus
Mazzinghy Alvarenga, dedico.

AGRADECIMENTOS

À Deus, pela vida em abundância, pela saúde, pela força que tem me dado e pela sua presença em TODOS os momentos.

À Universidade Federal de Viçosa e ao Programa de Pós Graduação em Genética e Melhoramento, pela oportunidade de realizar esse curso.

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

Ao Professor Ney Sakiyama, pela orientação, pelas nossas longas e ricas conversas, pelo exemplo de vida e por ter acreditado em mim.

À Doutora Eveline Caixeta, pela amizade, pelos ensinamentos, por sua paciência, e por ter confiado a mim a execução desse trabalho.

À Doutora Eunize Zambolim, pelos conselhos, pela amizade e por estar sempre pronta a ajudar.

Ao Professor Cosme, pelas críticas e sugestões no aperfeiçoamento desse trabalho.

Ao Doutor Felipe Rodrigues da Silva, com quem pude aprender muito sobre Bioinformática e Genômica, pela suas sugestões e colaboração com esse trabalho.

Aos amigos do Laboratório de Biotecnologia do Cafeeiro (BIOCAFÉ), onde tive a oportunidade de realizar esse trabalho, pela amizade, brincadeiras

e festas, pelo convívio super agradável, pelos ensinamentos e pelo apoio durante esse tempo.

Aos meus grandes amigos Paulinho Monteiro, Marlon Pereira, Eduardo Franklin, Felipe Louback e Fernanda Lelis, pela amizade incondicional, por me fazerem acreditar em mim e pelo apoio nas horas difíceis.

Aos meus amigos da Igreja Presbiteriana de Viçosa (IPV), especialmente Márcio Veríssimo, Alisson Xavier, Frederico Prado, Jeanne Scardini, Ana Carolina Camargo, Juliano Pinheiro, e Paulo Alexandre Lobato, pelo convívio, pelas boas influências, pelas orações e por me fazerem sentir especial e importante.

Ao Zilbinho e à Marô, meus discipuladores, pelas conversas, pelos conselhos, pelo carinho e pela amizade.

Aos meus amigos do curso de Genética e Melhoramento, em especial Maíra Freire, Salvador Lima, Janaína Melo e Ramon de Almeida, pelo companheirismo, pela ajuda nas disciplinas cursadas e pela troca de experiências.

A todos os meus companheiros de república, Ti-Burga, Lulu, Andrezón, Tonhão e Fabão, pela amizade e pelos momentos de descontração.

Aos meus pais, pelo apoio, pelo compromisso comigo, por nunca deixarem faltar nada, pelos conselhos, pelas orações e pela atenção.

Ao meu irmão Áquila e sua esposa Mirella, pelo carinho, preocupação e companheirismo.

A todos que, de alguma forma, contribuíram para a realização desse trabalho.

Muito obrigado !

BIOGRAFIA

Samuel Mazzinghy Alvarenga, filho de Aurelino Ferreira Alvarenga e Maria de Jesus Mazzinghy Alvarenga, nasceu em Teófilo Otoni, Minas Gerais, no dia cinco de Agosto de 1981.

Em Março de 2000 ingressou no curso de Ciências Biológicas na Universidade Federal de Viçosa, graduando-se em janeiro de 2005, como Bacharel em Ciências Biológicas.

Em Agosto de 2005 ingressou no Programa de Pós-Graduação, em nível de Mestrado, em Genética e Melhoramento da UFV, submeteu-se à defesa de dissertação em Julho de 2007.

SUMÁRIO

LISTA DE ABREVIATURAS.....	IX
RESUMO.....	XI
ABSTRACT.....	XIII
1. INTRODUÇÃO GERAL.....	1
1.1. GENÔMICA EM PLANTAS	3
1.2. BIOINFORMÁTICA	5
1.2.1. BANCOS DE DADOS.....	6
1.2.2. BUSCA DE SEQÜÊNCIAS.....	7
1.2.3. BIOINFORMÁTICA NO PROJETO BRASILEIRO DO GENOMA CAFÉ.....	8
1.3. MARCADORES MOLECULARES NO MELHORAMENTO DE PLANTAS	11
1.3.1. TIPOS DE MARCADORES MOLECULARES.....	11
1.3.2. MARCADORES MOLECULARES NO MELHORAMENTO DO CAFEIRO	13
1.4. GENES DE RESISTÊNCIA A DOENÇAS.....	15
1.4.1. REGIÃO NBS.....	19
1.4.2. REGIÃO LRR	20
2. REFERÊNCIAS BIBLIOGRÁFICAS.....	22
CAPÍTULO 1	45
IDENTIFICAÇÃO DE SEQÜÊNCIAS EXPRESSAS DO GENOMA DO CAFÉ POTENCIALMENTE ASSOCIADAS COM A RESISTÊNCIA A DOENÇAS ...	45
1. INTRODUÇÃO	46
2. MATERIAL E MÉTODOS.....	47
2.1. Base de dados	47

2.2. Estratégias de mineração	47
2.2.1. Electronic Northern	47
2.2.2. Palavras-Chave	49
2.2.3. Seqüências Publicadas	51
3. RESULTADOS E DISCUSSÃO.....	52
3.1 Electronic Northern	52
3.2. Palavras-chave	56
3.3. Seqüências publicadas	84
4. CONCLUSÃO.....	87
5. REFERÊNCIAS BIBLIOGRÁFICAS.....	88
CAPÍTULO 2	94
MARCADORES MOLECULARES DERIVADOS DE SEQÜÊNCIAS EXPRESSAS DO GENOMA DO CAFÉ POTENCIALMENTE ENVOLVIDAS NA RESISTÊNCIA À FERRUGEM	94
1. INTRODUÇÃO	95
2. MATERIAL E MÉTODOS.....	96
2.1. Material Vegetal e Extração de DNA	96
2.2. Seqüências do Banco de Dados do PBGC.....	97
2.3. Desenvolvimento de Marcadores Moleculares	97
3. RESULTADOS E DISCUSSÃO.....	98
4. CONCLUSÃO.....	104
5. REFERÊNCIAS BIBLIOGRÁFICAS.....	105

LISTA DE ABREVIATURAS

- AFLP:** *Amplified Fragment Length Polymorphism.*
- BAC:** *Bacterial Artificial Chromosome.*
- BIOCAFÉ:** Laboratório de Biotecnologia do Cafeeiro.
- BLAST:** *Basic Local Alignment Search Tool.*
- CAPES:** Coordenação de Aperfeiçoamento de Pessoal de Nível Superior.
- CBD:** *Coffee Berry Disease.*
- CBP&D-Café:** Consórcio Brasileiro de Pesquisa e Desenvolvimento do Café.
- CC:** *Coiled Coil.*
- CENARGEN:** Embrapa Recursos Genéticos e Biotecnologia.
- CIFC:** Centro de Investigação das Ferrugens do Cafeeiro.
- dbEST:** *EST database.*
- EMBRAPA:** Empresa Brasileira de Pesquisa Agropecuária.
- EPAMIG:** Empresa de Pesquisa Agropecuária de Minas Gerais.
- EST:** *Expressed Sequence Tag.*
- FAPESP:** Fundação de Amparo à Pesquisa do Estado de São Paulo.
- FISH:** *Fluorescent In Situ Hybridization.*
- GENOLYPTUS:** Rede Brasileira de pesquisa do Genoma do *Eucalyptus*.
- GISH:** *Genomic In Situ Hybridization.*
- HSP:** *Heat Shock Protein.*
- IPV:** Igreja Presbiteriana de Viçosa.
- ITS:** *Internal Transcribed Spacer.*
- LGE:** Laboratório de Genômica e Expressão.
- LRR:** *Leucine Rich Repeats.*
- LZ:** *Leucine Zipper.*

NBS: *Nucleotide Binding Site.*

NCBI: *National Center for Biotechnology Information.*

NR: *Non Redundant.*

OMIM: *Online Mendelian Inheritance in Man.*

ONSA: *Organization for Nucleotide Sequencing and Analysis.*

PBGC: *Projeto Brasileiro do Genoma Café.*

PCR: *Polymerase Chain Reaction.*

PFP: *Hypertext Preprocessor.*

PlantGDB: *Plant Genome DataBase.*

QTL: *Quantitative Trait Loci.*

RAPD: *Random Amplified Polymorphic DNA.*

RFLP: *Restriction Fragment Length Polymorphism.*

RGA: *Resistance Gene Analog.*

SCAR: *Sequence Characterized Amplified Regions.*

SNP: *Single Nucleotide Polymorphism.*

SRAP: *Sequence Related Amplified Polymorphism.*

SSAP: *Sequence Specific Amplification Polymorphism.*

SSR: *Simple Sequence Repeat.*

STS: *Sequence Tagged Sites.*

SUCEST: *Sugar Cane EST Genome Project.*

TIR: *Toll/Interleukin-1 Receptor-like.*

TMV: *Tobacco Mosaic Vírus.*

TRAP: *Target Region Amplification Polymorphism.*

UFV: *Universidade Federal de Viçosa.*

VNTR: *Variable Number of Tandem Repeats.*

RESUMO

ALVARENGA, Samuel Mazzinghy, M.Sc., Universidade Federal de Viçosa, julho de 2007. **Caracterização de seqüências expressas do genoma café potencialmente relacionadas com a resistência a doenças.** Orientador: Ney Sussumu Sakiyama. Co-orientadoras: Eveline Teixeira Caixeta e Eunize Maciel Zambolim.

Seqüências potencialmente envolvidas na resistência do cafeeiro a doenças foram identificadas, por meio de análise *in silico*, a partir das informações geradas pelo Projeto Brasileiro do Genoma Café (PBGC). Para isso foram usadas três estratégias. Inicialmente, palavras-chave correspondentes a termos relacionados aos mecanismos de resistência de plantas a patógenos foram identificadas na literatura e utilizadas como “iscas” para a mineração dos dados. Com o auxílio de ferramentas disponíveis na plataforma de bioinformática do PBGC, foram identificadas ESTs (*Expressed Sequence Tags*) relacionadas a cada uma destas palavras. Outra estratégia utilizada foi a busca por similaridades entre algumas seqüências públicas envolvidas com a resistência do cafeeiro a doenças com as seqüências do PBGC, por meio do programa BLAST. Utilizou-se, também, o *Electronic Northern*, uma ferramenta desenvolvida pelo Laboratório de Genômica e Expressão (LGE). A mineração, usando as três estratégias, identificou 14.060 seqüências do PBGC. Essas seqüências apresentaram similaridade com proteínas conhecidamente relacionadas com o processo de defesa da planta contra doenças como, por exemplo, quitinase, proteína quinase, citocromo P450, proteína de resistência a doenças, proteína relacionada com patogênese, proteínas com domínio LRR e NBS, proteínas induzidas por hipersensibilidade, entre outras. Os processos biológicos com os quais essas seqüências estão envolvidas incluíram metabolismo, transporte, regulação da transcrição, enovelamento de proteínas, biossíntese entre outros. A análise global baseada em ontologia de função molecular das seqüências obtidas mostrou que os genes estão envolvidos com metabolismo, resposta a estímulos externos, diferenciação celular, ligação a ácidos nucleicos, ligação a nucleotídeos, resposta de defesa, apoptose entre

outras. Visando verificar o envolvimento destas seqüências com a resistência do cafeeiro à ferrugem foram desenhados 40 *primers* para amplificar algumas das seqüências mineradas. Os *primers* foram desenhados com o programa computacional *Primer3* e a estabilidade desses foi verificada por meio do programa *PrimerSelect*. Diferentes concentrações dos componentes da reação de PCR foram analisadas. Utilizando as condições de reação e amplificação otimizadas, os 40 *primers* foram testados em 12 genótipos resistentes e 12 susceptíveis a *Hemileia vastatrix*, fungo causador da ferrugem. Vinte e nove destes 40 *primers* resultaram em bandas únicas e bem definidas, sendo um polimórfico. Este trabalho permitiu obter, até o momento, um marcador molecular polimórfico entre os indivíduos resistentes e susceptíveis. Esse marcador, denominado CARF 005, amplifica uma região do DNA que corresponde a uma ORF parcial de *Coffea arabica* que codifica uma proteína de resistência a doenças.

ABSTRACT

ALVARENGA, Samuel Mazzinghy, M.Sc., Universidade Federal de Viçosa, July, 2007. **Characterization of expressed sequences from coffee genome potentially related with the resistance to diseases.** Adviser: Ney Sussumu Sakiyama. Co-advisers: Eveline Teixeira Caixeta and Eunize Maciel Zambolim.

Sequences potentially involved with the coffee resistance were identified, by *in silico* analyses, using information generated by the Projeto Brasileiro do Genoma Café (PBGC). For that, three strategies were used. Initially, keywords related to the plants resistance mechanism to pathogens were searched in scientific literature and used as drivers for data mining. Using the available tools at the PBGC bioinformatics platform, ESTs (Expressed Sequence Tags) related to each one of these words were identified. The search for similarities between some published sequences and sequences from the PBGC, by using the BLAST program was another strategy. The Electronic Northern, a tool developed by the Laboratório de Genômica e Expressão (LGE), was also used. Those strategies allowed the identification of 14,060 sequences of the PBGC. These sequences were similar to proteins known to be related to de plant disease defense process, for instance chitinase, kinase protein, cytochrome P450, disease resistance protein, pathogenesis related protein, LRR and NBS proteins, hypersensibility induced protein among others. The biological processes with witch these sequences are involved included metabolism, transport, transcription regulation, protein folding, biosynthesis and others. The ontology-based global analysis for molecular function showed that the genes are involved with metabolism, external stimulus response, cellular differentiation, nucleic acid binding, nucleotide binding, defense response, apoptosis and others. Aiming to verify the involvement of these sequences with the coffee tree resistance to leaf rust, 40 primers were designed to amplify the mined sequences. The primers were synthesized using the computational program Primer3 and their stability was tested by the program PrimerSelect. Different PCR conditions were tested. Using optimized reaction and

amplification conditions, those 40 primers were tested in 12 resistant and 12 susceptible genotypes to *Hemileia vastatrix*, fungus that causes coffee leaf rust. Twenty nine of those resulted in unique and sharp bands, and only one of these was polymorphic. The 40 primers permitted to find one molecular marker polymorphic between the resistant and susceptible genotypes. This marker amplifies a region of the DNA which corresponds to a *Coffea Arabica* ORF for disease resistance protein.

1. INTRODUÇÃO GERAL

O café, produzido em mais de 60 países, é um dos mais importantes produtos de exportação mundial, movimentando US\$10-12 bilhões por ano. Em 2004, os cinco maiores países produtores responderam por 64,5% da produção mundial. O Brasil, na primeira posição, respondeu por uma participação de 31,80%, três vezes maior do que a dos concorrentes mais próximos: Vietnã (10,49%), Indonésia (9,10%) e Colômbia (8,78%) (Siqueira, 2005).

O cafeeiro pertence ao subgênero *Coffea*, família *Rubiaceae*, formado por cerca de 100 espécies conhecidas. Das espécies cultivadas, *Coffea arabica* (café arábica) e *Coffea canephora* (café robusta) são as mais importantes economicamente, sendo o primeiro responsável por 70% da produção mundial e 99% da produção da América Latina (Dolzan, 2005).

A produção do café, no entanto, é freqüentemente afetada por doenças e pragas que ocorrem durante praticamente todo o ciclo. Dentre as doenças que ocorrem no cafeeiro, a ferrugem é a mais importante, por causar grandes prejuízos para a cafeicultura. Ela ocorre em todas as regiões produtoras de café no Brasil e na América Central (Carvalho e Chalfoun, 2005). Na ausência de controle químico, a ferrugem do cafeeiro pode causar perdas em torno de 35% na produção (Zambolim et al., 2005).

O agente etiológico da ferrugem é o fungo biotrófico *Hemileia vastatrix* Berk. et Br., que sobrevive na forma de raças. No Centro de Investigação das Ferrugens do Cafeeiro (CIFC), Oeiras, Portugal, foram biologicamente caracterizadas mais de quarenta e cinco raças que atacam os cafeeiros

(Várzea e Marques, 2005), enquanto que no Brasil foram encontradas 12 (Cardoso et al., 1988).

Atualmente, a principal forma de controle da ferrugem é por meio de fungicidas. Embora potencialmente eficiente, o controle químico é um método oneroso para ser adotado pelos produtores, sem falar nos riscos aos quais eles e o ambiente estão sujeitos. Nesse contexto, o desenvolvimento de cultivares resistentes se torna o melhor método de controle, pois é econômico, eficiente e não causa danos ao meio ambiente (Zambolim et al., 2005).

Todavia, os grandes esforços despendidos para a obtenção de cultivares resistentes por meio dos programas convencionais de melhoramento têm mostrado um progresso lento (Vieira et al., 2005). Várias dificuldades podem ser observadas, como por exemplo, o ciclo longo do cafeeiro; a baixa variabilidade genética dentro das espécies de interesse comercial; o alto custo dos ensaios de campo; a necessidade de avaliar vários anos de produções consecutivas para conhecer a capacidade produtiva a longo prazo; e a falta de marcadores genéticos disponíveis. Além disso, existe ainda a possibilidade da resistência do cultivar ser suplantada pela grande variabilidade genética do patógeno (Zambolim et al., 2005) e a capacidade do fungo de aumentar o nível de agressividade (Eskes, 1989).

Desta forma, existe a necessidade do desenvolvimento e utilização de métodos alternativos para auxiliar o melhoramento do café, proporcionando maior eficiência, velocidade e direcionamento para os programas de melhoramento (Vieira et al., 2005). Nesse sentido, as pesquisas em biotecnologia estão causando grande impacto na agricultura em todo o mundo. O recente desenvolvimento da tecnologia genômica tem gerado muitas informações e criado bancos de dados de seqüências de DNA, que possibilitam a identificação dos fatores genéticos determinantes e/ou associados com características de interesse agrônomo. O seqüenciamento de cDNAs em larga escala para produzir Etiquetas de Seqüências Expressas (ESTs - *Expressed Sequence Tags*) e a comparação dessas seqüências com as disponibilizadas em banco de dados tem se tornado a opção para obtenção rápida de dados sobre a capacidade codificadora de genomas e para identificar novos genes.

1.1. GENÔMICA EM PLANTAS

A genômica em plantas sofreu uma revolução com a liberação da seqüência do genoma completo de *Arabidopsis thaliana* pela *Arabidopsis Genome Initiative* em 2000, quatro anos antes do previsto (*The Arabidopsis Genome Initiative*). O projeto de seqüenciamento do genoma da *Arabidopsis* começou em 1995-1996, durante um período transitório na análise de seqüência em genomas. O sucesso da abordagem clone-a-clone para seqüenciar um genoma complexo era evidente nessa época tendo como base o projeto de seqüenciamento do nematóide *Caenorhabditis elegans* (Martienssen e McCombie, 2001).

A conclusão do seqüenciamento do genoma completo do arroz (*Oryza sativa*) por um consórcio público foi anunciada em 2002. Este trabalho foi complementado pelos projetos de seqüenciamento realizados pela Syngenta (Goff et al., 2002), Monsanto (Barry, 2001) e o Instituto de Genômica de Pequim, que seqüenciou a subespécie *indica*. Devido a similaridades no nível genômico entre arroz e outras culturas importantes (Moore et al., 1995), a conclusão do genoma do arroz teve um impacto significativo na biotecnologia vegetal e na bioinformática.

O seqüenciamento completo de espécies com genomas grandes é demorado, oneroso e trabalhoso, apesar das técnicas modernas. O seqüenciamento parcial de clones de cDNA, por ser rápido e apresentar boa relação de custo benefício para a obtenção de dados em genomas, tem se tornado um segmento crescente nas bases de dados públicas (Wolfsberg e Landsman, 1997). Em plantas, o seqüenciamento de ESTs foi inicialmente feito para as espécies modelos *Arabidopsis* (Höfte et al., 1993) e arroz (Yamamoto e Sasaki, 1997). Posteriormente, grande quantidade de seqüências ESTs de outras espécies, incluindo pinheiro (*Pinus taeda* L., Allona et al., 1998), milho (*Zea mays* L., Gai et al., 2000), soja (*Glycine max* (L.) Merr., Shoemaker et al., 2002), trigo (*Triticum aestivum* L., Lazo et al., 2001), batata (*Solanum tuberosum*, Ronning et al., 2003), maçã (*Malus domestica*, Newcomb et al., 2006), uva (*Vitis vinifera*, Silva et al., 2005), algodão (*Gossypium hirsutum*, Udal et al., 2006), cacau (*Theobroma cacao*, Gesteira et al., 2007) e várias outras foram depositadas no dbEST, o banco de dados de seqüências ESTs.

A relação completa pode ser encontrada acessando <http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html>.

No Brasil, alguns projetos de seqüenciamento baseados na metodologia de geração de ESTs foram concluídos ou estão em andamento.

O projeto genoma da cana-de-açúcar, (SUCEST – *Sugar Cane EST Genome Project*), realizado pelo grupo brasileiro ONSA (*Organization for Nucleotide Sequencing and Analysis*) resultou em uma base de dados contendo 238.000 ESTs de 37 bibliotecas. O projeto foi voltado para o estudo dos genes relacionados com o metabolismo da sacarose, resistência a pragas e tolerância a condições desfavoráveis de clima e solo (Grivet e Arruda, 2001).

Criado em 2001, o Genolyptus (Rede Brasileira de pesquisa do Genoma do *Eucalyptus*) ainda está em andamento. O objetivo do projeto é entender a genética molecular e quantitativa de *Eucalyptus* voltada para o controle das características de importâncias econômicas silvicultural e industrial, com maior enfoque no processo de formação da madeira (Grattapaglia, 2003).

O Projeto Brasileiro do Genoma da Banana teve início em 2002, com o objetivo de desenvolver as bases do programa de genômica e biotecnologia de banana. Este projeto resultou na criação do DATAMusa, hoje o segundo maior banco de dados de genômica de banana no mundo, que é composto de informações de Genômica Estrutural (BACs), de Seqüências Expressas e de Análogos de Genes de Resistência (RGAs) (Santos et al., 2003).

A cultura cafeeira também está se beneficiando da revolução genômica. Por iniciativa do Consórcio Brasileiro de Pesquisa e Desenvolvimento do Café (CBP&D-Café), coordenado pela Embrapa Café, em colaboração com a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) e Embrapa Recursos Genéticos e Biotecnologia (Cenargen), teve início em fevereiro de 2002 o Projeto Brasileiro do Genoma Café (PBGC). O projeto foi criado para desenvolver ferramentas úteis para a descoberta de genes e análise genética funcional em café e espécies relacionadas e ajudar no avanço de conhecimento da organização e evolução do genoma do café. A base de dados das ESTs de *C. arabica*, *C. canephora* e *C. racemosa* gerada pelo projeto resultou na identificação de mais de 30 mil unigenes (seqüências que representam um único gene) diferentes. Esta informação básica fornece um recurso valioso para estudos da biologia e fisiologia do café que poderá resultar

no isolamento e a caracterização de genes agronômicos importantes para o melhoramento genético de *Coffea*. Os genes identificados poderão ser úteis para acelerar a escolha de variedades mais produtivas, tolerantes à seca e resistentes ao ataque de pragas e doenças. Também terá utilidade para obtenção de novos cultivares com qualidades superiores em aroma e sabor e com melhores características nutritivas e farmacêuticas, com vistas à maior satisfação dos consumidores e à conquista do mercado com produtos de maior qualidade e valor agregado (Vieira et al., 2006).

1.2. BIOINFORMÁTICA

Com a expansão das pesquisas em genômica, possibilitada pelo desenvolvimento de novas tecnologias de obtenção de dados em larga escala (*High Throughput Technology*), a quantidade de informação biológica vem crescendo exponencialmente. Essa grande quantidade de dados de seqüências de genomas disponível tende a aumentar (Winslow e Boguski, 2003). No mês de agosto de 2007, o seqüenciamento de mais de 2.863 projetos genomas havia sido finalizado ou estava em progresso (*Genomes Online Database* – <http://www.genomesonline.org>), e os resultados podem ser encontrados em vários bancos de dados públicos.

Esse crescimento está exigindo recursos computacionais cada vez mais sofisticados. A bioinformática é uma disciplina que tem auxiliado a suprir essas necessidades pelo desenvolvimento de novas ferramentas para o gerenciamento de grande quantidade de informações. Além do armazenamento, existe a necessidade de análise dos dados, o que torna indispensável a utilização de plataformas computacionais eficientes para a interpretação dos resultados (Prosdocimi et al., 2002; Zatz, 2002).

Desde o advento do GenBank em 1982 e do Projeto Genoma Humano em 1990, a bioinformática se tornou um sinônimo de análise e gerenciamento de dados de seqüência de DNA e proteína (Boguski, 1999). A tendência da bioinformática é continuar sendo desenvolvida com o surgimento de novos dados. O GenBank vem dobrando de tamanho a cada 18 meses e atualmente contém mais de 145 bilhões de bases (Benson et al., 2007). Portanto, a

habilidade em navegar e realizar buscas de seqüências nos bancos de dados é fundamental para os pesquisadores que trabalham na área de genômica.

1.2.1. BANCOS DE DADOS

Devido a imensa quantidade de dados de seqüências de DNA gerados por vários laboratórios no mundo, faz-se necessário organizá-los de maneira acessível, para evitar redundância na pesquisa científica e possibilitar as análises pelo maior número possível de pesquisadores (Roos, 2001). A construção de bancos de dados para armazenamento de informações de seqüências de DNA e de genomas inteiros, proteínas e suas estruturas tridimensionais, bem como vários outros produtos da era genômica, tem sido um grande desafio.

O Centro Nacional para Informação Biotecnológica dos EUA (NCBI) é considerado a central de informações genômicas. O NCBI foi criado em 1988 para desenvolver sistemas de informações para a biologia molecular (Wheeler et al., 2007). Vários outros bancos de dados similares estão distribuídos por países da Europa (Kulikova et al., 2007) e Japão (Tateno et al., 2002), que trocam dados em um intervalo de 24 horas com o NCBI (Pruitt et al., 2005; Benson et al., 2007). O GenBank é o principal banco de dados do NCBI e armazena todas seqüências de DNA (de seqüências pequenas a genomas inteiros), RNA e proteínas disponíveis publicamente (Benson et al., 2007). Além do GenBank, outros bancos do NCBI apresentam as informações organizadas de diferentes maneiras. Por exemplo, o banco de dados UniGene agrupa todas as seqüências parciais do transcriptoma de um organismo formando *clusters*, onde cada um representa a seqüência consenso de um gene (Pontius et al., 2003). Também no NCBI, o banco de dados RefSeq reúne somente as seqüências de referência, ou seja, a mais representativa de um transcrito, editada e inspecionada por um curador (Pruitt et al., 2005). O RefSeq é, freqüentemente, o melhor banco de dados para evitar a redundância natural num universo com muitas informações. Para acessar o RefSeq e outros bancos de seqüências curadas foi desenvolvida a ferramenta LocusLink no NCBI (Pruitt et al., 2000). Outros bancos são específicos de um organismo, como o

OMIM (*Online Mendelian Inheritance in Man*) que foi criado para catalogar todos os genes e alelos relacionados a doenças e outras características humanas, bem como fornecer o detalhamento técnico e bibliografia correspondente a cada característica (Hamosh et al., 2000). O PlantGDB (*Plant Genome DataBase*) é um banco de dados de seqüências de várias espécies de plantas. O objetivo do PlantGDB é estabelecer uma base para a identificação de conjuntos de genes comuns a todas as plantas ou genes específicos de determinadas espécies. Isso é feito por meio da integração de várias ferramentas de bioinformática que podem facilitar a predição de genes e a comparação entre as espécies. Estes bancos de dados, ditos secundários, são tão importantes quanto o GenBank.

Nos bancos de dados há também uma grande variedade de informações sobre estruturas moleculares, expressão gênica diferencial, diversidade genética, evolução e outras que podem ser obtidas pela bioinformática. O desenvolvimento de procedimentos pelos quais esses dados podem ser inseridos e extraídos de bancos de dados secundários é um dos grandes desafios dos pesquisadores. Apesar da existência de várias ferramentas disponíveis no NCBI e em outras instituições, ainda existem muitos campos para o desenvolvimento de procedimentos específicos (Santos e Ortega, 2003).

1.2.2. BUSCA DE SEQÜÊNCIAS

Várias ferramentas desenvolvidas pela bioinformática permitem o acesso e análise dos bancos de dados. A ferramenta mais conhecida para a comparação de seqüências de DNA com as dos bancos de dados genômicos é o BLAST ou *Basic Local Alignment Search Tool* (Altschul et al., 1990). Este programa compara uma seqüência de DNA ou de uma proteína com todas as seqüências genômicas de domínio público.

O programa BLAST não faz uma comparação da extensão total das seqüências analisadas, mas apenas identifica, no banco de dados, a presença de uma seqüência suficientemente semelhante com a pesquisada. Dessa forma, descarta rapidamente os resultados não relevantes, e estende a

vizinhança da região de similaridade detectada até um determinado valor de corte (Pertsemlidis e Fondon, 2001).

Esta busca resulta em seqüências (DNA ou proteínas) depositadas com maior similaridade. Desta forma, várias regiões de DNA podem ser anotadas por meio do BLAST. O resultado dessa busca pode ser usado para atribuir uma função a qualquer segmento de DNA que apresenta similaridade significativa com outras seqüências de DNA ou proteínas previamente depositadas no GenBank com função conhecida experimentalmente (Santos e Ortega, 2003).

1.2.3. BIOINFORMÁTICA NO PROJETO BRASILEIRO DO GENOMA CAFÉ

O Projeto Brasileiro do Genoma Café (PBGC) gerou 214.964 ESTs de *Coffea arabica*, *C. canephora* e *C. racemosa*. Estas seqüências são oriundas de clones de 37 bibliotecas de cDNA provenientes de diferentes tecidos de cafeeiros em diferentes estádios de desenvolvimento e/ou sob estresses bióticos e abióticos (Vieira et al., 2006).

As seqüências geradas foram processadas, com remoção de seqüências provenientes de rRNA e eliminação de porções contendo seqüências de vetor, adaptador, caudas de poliA e bases de baixa qualidade, e agrupadas para a formação do conjunto UniGene do Café (Sales et al., 2007).

Esses dados foram armazenados em duas plataformas de bioinformática. Uma delas foi desenvolvida pelo Laboratório de Genômica e Expressão (LGE - <http://www.lge.ibi.unicamp.br/cafe/>) e está disponível desde o segundo semestre de 2004. A outra plataforma, disponibilizada em março de 2007, foi criada pela Embrapa Recursos Genéticos e Biotecnologia (Cenargen - <https://alanine.cenargen.embrapa.br/cafEST/>).

Essas plataformas foram desenvolvidas com o objetivo de armazenar os dados gerados e dispor de medidas de segurança para garantir a integridade dos mesmos. Além disso, por meio dessas plataformas, os dados foram organizados de forma a permitir aos seus usuários a obtenção e manipulação das informações de maneira simplificada (Vieira et al., 2006).

Apesar do conjunto de dados ser o mesmo nas duas plataformas, os tratamentos empregados na sua organização não foram idênticos. Algumas

diferenças são grandes, como por exemplo, o modo de tratamento dos dados. Na plataforma do LGE as ESTs foram agrupadas separadamente por espécies, resultando em 14.886 *contigs* e 24.426 *singletons* para *C. arabica*, 2.147 *contigs* e 4.622 *singletons* para *C. canephora*, e 949 *contigs* e 3.107 *singletons* para *C. racemosa* (Vieira et al., 2006). No caso da plataforma do Cenargen, as ESTs das três espécies foram tratadas conjuntamente. Outra diferença foi a estratégia de limpeza das seqüências antes do agrupamento.

Depois de agrupadas, as seqüências consenso de cada UniGene foram comparadas com seqüências protéicas presentes no GenBank utilizando-se o software BlastX, contra NR, considerando o *e-value* 10^{-5} como limite para identidade (Vieira et al., 2006).

A interface *Web* foi desenvolvida nas linguagens PHP e *Perl* rodando sobre o Apache para permitir aos usuários o acesso aos dados de maneira simplificada e rápida. A escolha destas ferramentas se deve ao fato de serem todas de código livre, permitindo personalizações, se necessárias, e por não agregarem nenhum vínculo de licença (Sales et al., 2007).

A vantagem de ter os dados disponíveis *online* é que qualquer um, com conexão à internet, pode ter acesso. Entretanto esta vantagem é contrabalançada pelo risco de panes quando se dispõe de apenas um sistema central de armazenamento de dados. Esta é mais uma razão de existirem as duas plataformas de acesso aos dados do PBGC. Além disso, a possibilidade de gerenciar os dados por meio de dois grupos de bioinformática permite que o usuário escolha aquela com o qual ele melhor se identifica (Vieira et al., 2006).

- A plataforma do LGE dispõe das seguintes funcionalidades:
 - *Gene Projects*: é um sistema de gerenciamento e manipulação de seqüências. É possível incluir ESTs de interesse em cada projeto criado por meio de buscas por ESTs (*Reads Search*), palavras-chave (*Keyword Search*), similaridade com alguma seqüência (*BLAST Search*) ou a partir de um determinado padrão (*Pattern Search*). As seqüências que atendam aos requisitos desejados pelo usuário podem ser armazenadas em projetos. Depois de armazenadas podem ser agrupadas, e os *contigs* formados podem ser submetidos a uma nova comparação com o GenBank para se verificar uma nova anotação (ou não).

- *Electronic Northern*: Essa ferramenta é utilizada para calcular o nível de expressão de um gene em diferentes bibliotecas, e se baseia no número de ESTs de um dado gene que foi identificado em diferentes bibliotecas de cDNA.
- *Microsatellites*: Ferramenta que apresenta microssatélites nos *clusters* do genoma café. A tela mostra o *contig*, a posição do microssatélite, o número de cópias, a composição percentual de cada nucleotídeo do microssatélite e o padrão da repetição.
- SNPs: Ferramenta que apresenta SNPs (*Single Nucleotide Polymorphism* – Polimorfismo de Um Nucleotídeo) em *clusters* do genoma café. Por essa ferramenta são apresentados o *contig*, o número de ESTs que forma esse *contig*, o número de SNPs presente e a anotação do BLAST.
- *Search*: Ferramenta que permite a busca de seqüências por palavras-chave ou por BLAST.
- A plataforma do Cenargen dispõe das seguintes funcionalidades:
 - Busca: Sistemas de busca por *cluster*, *read*, palavra-chave ou biblioteca específica.
 - Teste Exato de Fisher (*Northern Blot* Eletrônico): Determina diferenças de expressão de um unigene, estatisticamente significativas, em diferentes bibliotecas.
 - BLAST Local: Ferramenta de busca de seqüências por similaridade.
 - Sistemas de Etiquetas: criação de Etiquetas para organizar as seqüências da maneira que o usuário preferir. As etiquetas têm comportamento similar à estrutura de diretórios encontrada em vários sistemas operacionais. Diferentemente do sistema de pastas ou projetos, entretanto, o usuário pode marcar uma mesma seqüência com uma ou mais etiquetas.

Sales et al. (2007) descrevem em detalhes as funcionalidades da plataforma de bioinformática do Cenargen. Além disso, essa plataforma possui um manual de ajuda que pode ser acessado pelo endereço eletrônico <https://alanine.cenargen.embrapa.br/Manual_Rapido_Genoma-Cafe.htm>.

1.3. MARCADORES MOLECULARES NO MELHORAMENTO DE PLANTAS

As análises *in silico* (em computadores) são necessárias para minerar os dados (*data mining*) gerados pelos projetos de sequenciamento. O objetivo dessas análises é de encontrar seqüências de genes relacionadas com processos biológicos de interesse. No entanto, o envolvimento do gene no processo biológico de interesse deve ser confirmado por metodologias de genômica funcional. Uma delas é a tecnologia de marcadores moleculares, que permite a detecção de variabilidade do DNA.

Um marcador molecular é definido como qualquer fenótipo molecular oriundo de um gene expresso ou de um segmento específico de DNA (Ferreira e Grattapaglia, 1998). Segundo Milach (1998), marcadores moleculares são características de DNA que diferenciam dois ou mais indivíduos e são herdados geneticamente. Os marcadores moleculares têm sido empregados no melhoramento de plantas para diversos fins, constituindo-se uma ferramenta importante, em virtude das vantagens que apresentam, como por exemplo, maior rapidez na obtenção de resultados e maior confiabilidade dos dados obtidos (Ferreira, 2003).

Alguns dos exemplos da utilização dos marcadores moleculares no melhoramento de plantas incluem: estudos de diversidade genética, caracterização de bancos de germoplasma, genealogia, construção de mapas, mapeamento comparativo, mapeamento gênico, seleção de genitores, certificação de cruzamentos, predição de fenótipos, *fingerprinting*, análises de pureza genética de sementes, melhoramento assistido, proteção varietal, mapeamento físico de genomas, integração de mapas genéticos e físicos, clonagem posicional, estudos de desequilíbrio de ligação, mapeamento de associação, filogenia, estudo de pedigree, isolamento de genes e diagnose (Borém e Caixeta, 2006).

1.3.1. TIPOS DE MARCADORES MOLECULARES

Os tipos distintos de marcadores moleculares atualmente disponíveis diferenciam-se pela tecnologia utilizada para revelar variabilidade ao nível de

DNA, e variam de acordo com a habilidade de detectar diferenças entre indivíduos, custo, facilidade de uso, consistência e repetibilidade. Os principais tipos de marcadores moleculares podem ser classificados em dois grupos, conforme a metodologia utilizada para identificá-los: hibridização ou amplificação do DNA via PCR. Entre os identificados por hibridização estão os marcadores RFLP - "*Restriction Fragment Length Polymorphism*" (Bostein et al., 1980) e minissatélites ou locos VNTR - "*Variable Number of Tandem Repeats*" (Jeffreys et al., 1985). Os revelados por amplificação incluem os marcadores do tipo RAPD - "*Random Amplified Polymorphic DNA*" (Williams et al., 1990), SCAR - "*Sequence Characterized Amplified Regions*", STS - "*Sequence Tagged Sites*" (Paran e Michelmore, 1993), Microssatélite (Litt e Luty, 1989) e AFLP - "*Amplified Fragment Length Polymorphism*" (Vos et al., 1995). Além desses, existem vários outros marcadores moleculares disponíveis atualmente.

Com o avanço da genômica, alguns tipos de marcadores moleculares têm sido beneficiados, pois o seqüenciamento do DNA tem facilitado o desenvolvimento e a identificação dos mesmos. Dentre estes marcadores existem os denominados marcadores *Expressed Sequence Tag-Polymerase Chain Reaction* (EST-PCR). Alguns tipos de marcadores EST-PCR incluem: SRAP (*Sequence Related Amplified Polymorphism*) (Li e Quiros, 2001), SSAP (*Sequence Specific Amplification Polymorphism*) (Nagy e Lelley, 2003) e TRAP (*Target Region Amplification Polymorphism*) (Hu e Vick, 2003). Os marcadores SRAP usam pares de *primers* com núcleos ricos em AT ou GC para amplificar fragmentos intragênicos na detecção de polimorfismos. Os marcadores SSAP exploram o polimorfismo de regiões LTR (*Long Terminal Repeat*) de retrotransposons ao longo do genoma. A técnica TRAP revela fragmentos polimórficos em torno de seqüências alvo de genes candidatos putativos.

Existem várias vantagens em realizar estudos genéticos com marcadores EST-PCR ou outros baseados em cDNA em relação aos que utilizam *primers* aleatórios. Eles marcam genes expressos e, portanto são particularmente úteis para mapeamento de QTLs. Se um marcador EST-PCR é encontrado ligado a um QTL é possível que o gene, a partir do qual o marcador EST-PCR foi derivado, controle a característica em questão. Pelo fato de serem originados de regiões codificadoras, que são mais conservadas entre

populações e espécies do que as não codificadoras, os marcadores EST-PCR são úteis para o estudo de mapeamento comparativo. Além disso, os EST-PCR podem ser herdados codominantemente, o que permite a identificação de dois alelos diferentes em locos heterozigotos de organismos diplóides (Rowland et al., 2003). Os EST-PCR foram desenvolvidos para várias espécies vegetais, incluindo pinheiro - *Pinus taeda* L. - (Temesgen et al., 2001), píceia - *Picea mariana* - (Perry e Bousquet, 1998) e *Picea abies* (Schubert et al., 2001), criptoméria - *Cryptomeria japonica* - (Tsumura et al., 1997), mirtilo - *Vaccinium sp.* - (Rowland et al., 2003), girassol - *Helianthus annuus* - (Pashley et al., 2006), amendoim - *Arachis hypogaea* L. - (Luo et al., 2005), cana de açúcar - *Saccharum sp.* - (Alwala et al., 2006), feijão - *Phaseolus vulgaris* L. - (Miklas et al., 2006) e muitas outras.

Outro tipo de marcador molecular beneficiado com o grande número de projetos de seqüenciamento em larga escala são os SNPs (*Single Nucleotide Polymorphism*). Esses marcadores se baseiam na variação de um único nucleotídeo em regiões codificadoras ou não do genoma. Os SNPs são as variações mais abundantes no genoma. Eles podem contribuir diretamente para um fenótipo ou podem estar associados a ele como resultado de um desequilíbrio de ligação (Nasu et al., 2002).

1.3.2. MARCADORES MOLECULARES NO MELHORAMENTO DO CAFEIEIRO

Como o café é uma cultura perene, de ciclo longo, e em que o melhoramento é efetuado ao longo de várias gerações, o desenvolvimento de variedades melhoradas é bastante demorado. Desta forma, torna-se importante a incorporação de técnicas avançadas de biotecnologia aos métodos tradicionais de melhoramento (Fontes, 2001). A utilização de marcadores moleculares pode auxiliar o melhoramento dessa espécie de forma bem ampla. As avaliações com esse tipo de tecnologia podem, por exemplo, ser realizadas em estádios iniciais de desenvolvimento da planta, com a obtenção de resultados similares aos que seriam alcançados na avaliação de plantas adultas (Oliveira et al., 2003).

Relações taxonômicas dentro do gênero *Coffea* foram estudadas usando marcadores específicos para DNA mitocondrial e do cloroplasto (Berthou et al. 1983; Orozco-Castillo et al. 1996). Estudos de filogenia molecular das espécies de *Coffea* foram realizados com marcadores para detectar variação na seqüência da região ITS2 (*Internal Transcribed Spacer*) e do DNA do cloroplasto (Lashermes et al., 1996; Cros et al., 1998).

Para verificar a constituição do genoma e a origem de *C. arabica*, estudos utilizando marcadores RFLP, FISH (*Fluorescent In Situ Hybridization*) e GISH (*Genomic In Situ Hybridization*) demonstraram que esta espécie é um alotetraplóide (Raina et al., 1998; Lashermes et al., 1999), resultante da hibridação natural de *C. eugenioides* com *C. canephora* (Lashermes et al., 1999).

Estudos com marcadores RAPD detectaram diversidade genética entre acessos de *C. arabica*, cultivares etíopes e acessos derivados de Típica e Bourbon (Orozco-Castillo et al., 1994; Lashermes et al., 1996; Anthony et al., 2001). Marcadores RAPD foram utilizados para caracterização e estimação de distâncias genéticas entre acessos do banco de germoplasma do Programa de Melhoramento da UFV/EPAMIG (Cabral et al., 2002; Alvarenga et al., 2005) e do IAPAR (Sera et al., 2003).

Steiger et al. (2002), usando marcadores AFLP, não conseguiram distinguir cultivares de *C. arabica*. Entretanto, a avaliação da diversidade genética entre esses cultivares forneceu informações necessárias para estimar o potencial dos marcadores para o melhoramento do cafeeiro. Também utilizando marcadores AFLP, Lashermes et al. (2000) distinguiram um grupo formado por genótipos derivados de Híbrido de Timor de acessos de *C. arabica*. Neste trabalho foi observado que a diversidade genética observada nos genótipos derivados de Híbrido de Timor era aproximadamente o dobro do que em *C. arabica*.

Maluf et al. (2005) identificaram polimorfismos entre cultivares de *Coffea* utilizando marcadores RAPD, SSR e AFLP. A variabilidade genética detectada por esses marcadores foi muito semelhante, ainda que reduzida. Neste estudo, os marcadores do tipo RAPD e SSR foram mais eficientes em análises de parentesco e o agrupamento das linhagens correspondeu à sua origem

genealógica. No entanto, nenhum dos métodos testados permitiu a identificação individual de linhagens.

Alguns marcadores moleculares ligados a genes controlando características qualitativas foram identificados em café, como auto-incompatibilidade em *C. canephora* (Lashermes et al., 1996), resistência ao CBD – *Coffee Berry Disease* - (Agwanda et al., 1997), resistência à ferrugem (Tedesco et al., 2000; Moreno et al., 2000; Prakash et al., 2004), porte de planta (Ruas et al., 2000) e resistência a *M. exigua* (Noir et al., 2003). Os marcadores ligados ao gene de resistência a *M. exigua* foram utilizados, posteriormente, para avaliar a introgressão do gene em linhagens possivelmente resistentes a esse patógeno (Diniz et al., 2005).

Em estudo realizado para identificar marcadores do tipo microssatélite (SSR) associados à resistência ao bicho-mineiro, 32 seqüências expressas em folhas de cafeeiro e relacionadas a mecanismos de defesa foram encontradas. Embora nenhuma marca molecular tenha sido identificada, o estudo exploratório pode ser importante para o entendimento de eventos de natureza genética envolvidos em cruzamentos interespecíficos, especialmente em espécies poliplóides como o cafeeiro (Pinto et al., 2007).

1.4. GENES DE RESISTÊNCIA A DOENÇAS

Dentre as várias aplicações dos marcadores moleculares estão os estudos dos genes de resistência de plantas a doenças.

Devido à inexistência de um sistema circulatório e de anticorpos, as plantas precisam ter, em cada uma das suas células, um mecanismo de defesa (Hammond-Kosack e Jones, 1997). As plantas normalmente são resistentes à maioria dos patógenos. Portanto, a habilidade do fitopatógeno em causar doença é geralmente a exceção, não a regra. Isto porque as plantas possuem uma habilidade natural de reconhecer potenciais patógenos invasores e preparar uma defesa bem sucedida. De maneira oposta, patógenos bem sucedidos causam doença porque são capazes de evitar o reconhecimento ou conter os mecanismos de defesa da planta ou as duas coisas (Staskawicz, 2001). Em trabalho clássico, Flor, em 1942, propôs a existência do sistema de

reconhecimento gene-a-gene, com interação específica: uma planta com o gene dominante de resistência *R* reconhece um patógeno com o gene dominante de avirulência *Avr* correspondente. Assim, a presença do gene *Avr* torna o patógeno não-virulento se a planta tiver o gene *R* apropriado. Se o gene não existe na planta e no agente patogênico, não há reconhecimento nem resistência e a doença se instala. O gene *R* codifica o receptor, que por sua vez reconhece a molécula elicitora gerada direta ou indiretamente pela ação do gene *Avr*, ativando os mecanismos de defesa (Jia et al., 2000; Deslandes et al., 2003; Belkhadir et al., 2004).

Em muitos casos, um único gene *R* pode conferir resistência para uma ou mais raças de um determinado patógeno quando transferido para uma planta suscetível de mesma espécie. Por esta razão, os genes *R* têm sido explorados em programas de melhoramento de plantas (McDowell e Woffenden, 2003).

A identificação e isolamento de genes *R* em plantas têm aumentado muito nos últimos anos (Richter e Ronald, 2000). O seqüenciamento genômico de várias espécies de plantas aliado a busca de mutantes em larga escala tem permitido aumentar, sobremaneira, a quantidade de informações relacionadas ao isolamento e identificação das funções dos produtos gênicos, regulação das vias de biossíntese e de transdução de sinal (Staskawicz et al., 1995). Os genes *R* codificam pelo menos cinco tipos diferentes de proteínas (proteínas *R*) (Dangl e Jones, 2001), que podem ser classificadas de acordo com as suas características estruturais.

O primeiro gene *R* clonado foi o *Pto* do tomate, que codifica uma proteína serina/treonina quinase e confere resistência às raças de *Pseudomonas syringae* que carregam o gene de avirulência *AvrPto* (Martin et al., 1993). Este gene pertence à primeira classe das proteínas *R*.

Na classe 2 estão genes como o *Cf9* do tomate que confere resistência a *Cladosporium fulvum* e codifica proteínas transmembranas, cuja região N-terminal extracelular é composta de repetições ricas de leucinas (LRR – *Leucine Rich Repeats*) (Jones et al., 1994). No entanto essas proteínas não possuem uma região intracelular significativa, que poderia constituir um componente sinalizador que ativaria o mecanismo de defesa (Ellis e Jones, 1998).

A classe 3, representada pelo gene *Xa21* (resistência a *Xanthomonas oryzae*) do arroz, inclui proteínas com um formato clássico de receptor-quinase, uma região LRR extracelular, uma região transmembrana e um domínio serina/treonina quinase citosólico (Song et al., 1995; Ellis e Jones, 1998).

Entretanto, a maioria dos genes de resistência codifica proteínas com um domínio N-terminal variável, seguido por um sítio de ligação de nucleotídeo (NBS – *Nucleotide Binding Site*) tripartida e uma região LRR (Baker et al., 1997; Ellis et al., 2000; Dangl e Jones, 2001). Baseando-se no seu domínio N-terminal, essas proteínas podem ser subdivididas em dois grupos. O grupo TIR-NBS-LRR (classe 4) exibe uma seqüência similar à da proteína Toll da *Drosófila* e ao do receptor interleucina-1 de mamíferos. A subclasse TIR-NBS-LRR inclui o gene *N* do tabaco (que confere resistência ao *Tobacco Mosaic Virus* – TMV), os genes *L6* e *M* do linho (resistência à ferrugem), e os genes *RPP5*, *RPP1* (resistência ao fungo *Peronospora parasitica*) e *RPS4* (resistência a *Pseudomonas syringae*) de *Arabidopsis* (Ellis e Jones, 1998; Gassman et al., 1999).

O outro grupo (classe 5) é o de proteínas não TIR-NBS-LRR, que não possuem a seqüência TIR (Meyers et al., 1999; Pan et al., 2000; Cannon et al., 2002; Richly et al., 2002) e/ou contêm um domínio zíper de leucina (LZ – *Leucine Zipper*) ou *coiled coil* (CC) potencial nas suas regiões N-terminais (Pan et al., 2000). CCs são estruturas formadas por duas a cinco hélices, que possuem um empacotamento distinto de cadeias laterais de aminoácidos na interface hélice-hélice (Lupas, 1996). A estrutura CC exibe uma organização de repetições de sete resíduos de aminoácidos com as cadeias laterais hidrofóbicas de dois deles formando uma interface para interações entre *coils* (Young, 2000).

A classe CC-NBS-LRR inclui os genes *RPM1* e *RPS2* e a classe LZ-NBS-LRR inclui os genes *RPS5* de *Arabidopsis* e o gene *Prf* do tomate, que conferem resistência a *P. syringae*, o gene *RPP8*, também de *Arabidopsis* (resistência a *Peronospora parasitica*), *Mi* do tomate (resistência ao nematóide *Meloidogyne incognita*) e o *Rx1* da batata (resistência ao *Potato Virus X* – PVX) (Ellis e Jones, 1998; Bendahmane et al., 1999). Os genes *I2* do tomate, *Rp1-D* do milho, *RGC2* da alface e o *Bs2* da pimenta não contêm a região TIR nem a LZ (Tai et al., 1999). Além das regiões mencionadas acima, os produtos de

alguns genes R possuem ainda um domínio conservado de função desconhecida chamado GLPLAL entre as regiões NBS e LRR (Dinesh-Kumar et al., 2000). A figura 1 resume as 5 classes acima descritas.

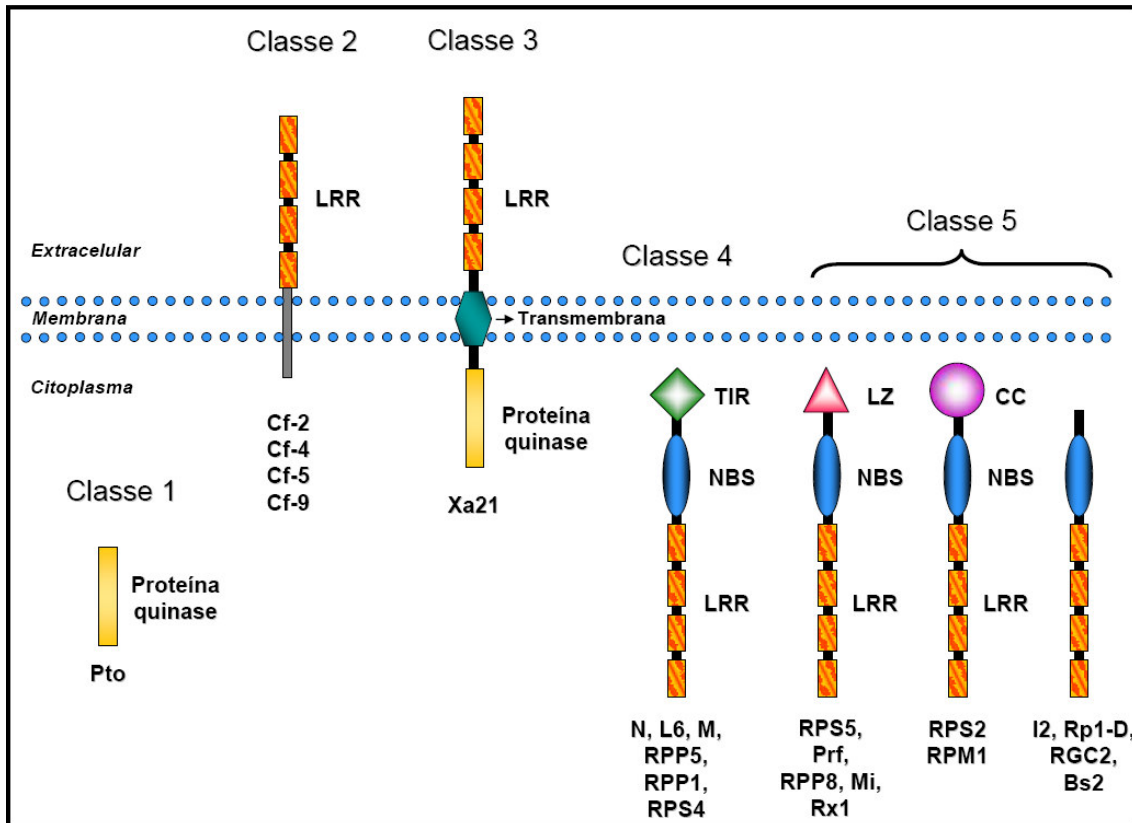


Figura 1: Representação esquemática das 5 classes de proteínas de resistência. Adaptado de Lehman, 2002; Hammond-Kosack e Parker, 2003; Michelmore e Meyers, 1998; Toyoda et al., 2002; McDowell e Woffenden, 2003 e Belkhadir et al., 2004.

Para o cafeeiro, já foram encontradas ESTs com similaridade com proteínas R da classe NBS-LRR, sugerindo que várias classes dos genes R estão presentes no genoma desta espécie. Utilizando vários *primers* degenerados para dois motivos conservados entre as regiões NBS dos genes R de diferentes espécies, Noir et al. (2001) amplificaram e clonaram seqüências de nove motivos distintos de proteínas do tipo NBS em café. Neste mesmo trabalho, foram identificados 18 RGAs (*Resistance Gene Analog*) em um único acesso de cafeeiro, sugerindo a presença de grande número de genes R no genoma do café.

Em 2004, Fernandez et al. identificaram duas ESTs de café com alta similaridade às proteínas DND1 e NDR1, que codificam componentes de sinalização de resistência em *Arabidopsis thaliana*. A proteína DND1 (*defence*,

no death – defesa, não morte) é um canal iônico que está envolvido na via de sinalização da reação de hipersensibilidade. A proteína NDR1 (*non race-specific disease resistance* – resistência a doenças, raça não específica) é um componente chave na via de sinalização de muitas proteínas da classe CC-NBS- LRR (Silva et al., 2006).

Guzzo (2004) isolou genes de acessos de cafeeiros apresentando funções relacionadas à defesa de plantas contra fitopatógenos. Estes genes estão envolvidos em diferentes processos como: resposta de hipersensibilidade, morte celular programada, síntese de proteínas antimicrobianas e degradação controlada de proteínas, dentre outros.

Lin et al (2005) encontraram genes de cafeeiro com funções putativas relacionadas com resistência a doenças, tais como proteínas da classe TIR-NBS-LRR, e com patogênese, entre outras.

1.4.1. REGIÃO NBS

O domínio NBS é encontrado em uma grande variedade de proteínas com função de ligação de ATP ou GTP em diversos organismos (Traut, 1994 e Saraste et al., 1990). Além disso, o domínio NBS compartilha uma homologia com as regiões NBS dos produtos dos genes *CED4* de *Caenorhabditis elegans* e dos genes *Apaf-1*, *FLASH*, *CARD4* e *Nod1*, que estão envolvidos na imunidade natural e na apoptose em animais (Saraste et al., 1990; Li et al., 1997; Van Der Biezen e Jones, 1998; Aravind et al., 1999). Essa homologia sugere um mecanismo conservado de morte celular programada em plantas e animais, inclusive em mamíferos. Acredita-se que a região N-terminal, juntamente com o domínio NBS, participa na ativação dos componentes da transdução de sinais levando a respostas de resistência patógeno-específicas (Aarts et al., 1998; Feys e Parker, 2000; Van Der Biezen et al., 2000). Os domínios NBS são encontrados também em muitas famílias de proteínas, incluindo as do grupo RAS, ATPases, fatores de alongação e proteínas G (Saraste et al., 1990). Essas proteínas servem como chaves moleculares e são fundamentais para muitos eventos celulares eucarióticos fundamentais como crescimento celular, diferenciação, organização do citoesqueleto, transporte de

vesículas, apoptose e defesa (Van Der Biezen e Jones, 1998, Aravind et al., 1999; Bourne et al., 1991).

O motivo *P-loop* [GXXXXGK(T/S)] dessas proteínas está envolvido em interações com fosfatos e íons Mg^{2+} (Saraste et al., 1990). O domínio quinase 2, que supostamente funciona numa reação de fosfotransferase, contém quatro aminoácidos hidrofóbicos consecutivos seguidos de um aspartato, que coordena os íons metálicos divalentes Mg^{2+} . O domínio quinase 3a está envolvido na ligação de purina ou ribose e contém uma tirosina ou uma arginina. Além desses domínios discutidos acima, os domínios NKXD, DXXG e (C/S)AX são altamente conservados na superfamília GTPase (Bourne et al., 1991). A ausência desses domínios nas proteínas R e suas homologias com *CED4/Apaf-1* sugerem que as proteínas R podem se ligar ao ATP e/ou agir como ATPases.

1.4.2. REGIÃO LRR

Este domínio consiste de nove a 40 unidades de repetições imperfeitas, cada uma delas com aproximadamente 25 aminoácidos. Na região central de cada repetição está uma estrutura fita β /volta β , que é hipervariável e tem a seqüência consenso XX(L)X(L)XXXX, onde **L** corresponde a leucinas (ou outros aminoácidos alifáticos) conservadas e **X** denota os aminoácidos hipervariáveis flanqueadores (Parniske et al., 1997; McDoewell et al., 1998). Nesta estrutura, diferentes repetições supostamente se ajustam e formam uma folha β paralela exposta ao solvente (Kobe e Deisenhofer, 1995). Como é exposta ao solvente, a superfície hipervariável poderia facilitar a interação da proteína R com o seu fator Avr (elicitor) relacionado e poderia fornecer diferentes especificidades de reconhecimento para fatores Avr alterados (Parniske et al., 1997). De fato, em vários sistemas planta-patógeno, a variação da seqüência na região LRR, particularmente na estrutura fita β /volta β , tem mostrado ser responsável por diferentes especificidades de reconhecimento ou de resistência (Parniske et al., 1997; Thomas et al., 1997; Botella et al., 1998; Wang et al., 1998; Ellis et al., 1999). No entanto, outras regiões da proteína R,

como a TIR da proteína do linho (Ellis et al., 1999), podem também contribuir para a especificidade do reconhecimento.

As proteínas de genes R que possuem a região LRR podem ser subdividas em duas grandes classes: aquelas que contêm uma região N-terminal LRR extracelular e uma âncora de membrana, e aquelas que são citoplasmáticas. Exemplos da primeira classe incluem o gene *Cf-2*, que confere resistência a *Cladosporium fulvum* em tomate (Dixon et al., 1996), *Xa21*, que confere resistência a *Xanthomonas oryzae* em arroz (Song et al., 1995), e *Hs1^{pro-1}*, um gene que confere resistência ao nematóide *Heterodera schachtii* (Cai et al., 1997). As proteínas citoplasmáticas incluem aquelas codificadas pelo gene *N* do tabaco, que confere resistência ao *Tobacco Mosaic Virus*, os genes *RPS2* e *RPM1* de arábida, que conferem resistência a raças específicas de *Pseudomonas spp*, os genes *L6* e *M* do linho, que conferem resistência à ferrugem, causada pelo fungo *Melampsora lini*, e o gene *RPP5* de arábida, que confere resistência ao fungo *Peronospora parasitica* (Bent et al., 1994; Mindrinos et al., 1994; Grant et al., 1995; Ellis et al., 1997; Parker et al., 1997).

2. REFERÊNCIAS BIBLIOGRÁFICAS

- Aarts, N., Metz, M., Holub, E., Staskawicz, B. J., Daniels, M. J., Parker, J. E. Different requirements for EDS1 and NDR1 by disease resistance genes define at least two R gene-mediated signaling pathways in Arabidopsis. Proceedings of the National Academy of Sciences of the United States of America, v.95, n.17, p.10306-10311. 1998.
- Aggarwal, R. K., Hendre, P. S., Varshney, R. K., Bhat, P. R., Krishnakumar, V., Singh, L. Identification, characterization and utilization of EST-derived genic microsatellite markers for genome analyses of coffee and related species. Theoretical and Applied Genetics, v.114, p.359-372. 2007.
- Agwanda, C. O., Lashermes, P., Trouslot, P., Combes, M. C., Charrier, A. Identification of RAPD markers for resistance to coffee berry disease, *Colletotrichum kahawae*, in arabica coffee. Euphytica, v.97, p.241-248. 1997.
- Allona, I., Quinn, M., Shoop, E., Swope, K., Cyr, S. S., Carlis, J., Riedl, J., Retzel, E., Campbell, M. M., Sederoff, R., Whetten, R. W. Analysis of xylem formation in pine by cDNA sequencing. Proceedings of the National Academy of Sciences of the United States of America, v.95, p.9693-9698. 1998.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D. J. Basic local alignment search tool. Journal of Molecular Biology, v.215, n.3, p.403-410. 1990.
- Alvarenga, S. M., Caixeta, E. T., Oliveira, A. C. B., Rufino, R. J. N., Brito, G. G., Pereira, A. A., Zambolim, L., Sakiyama, N. S. Análise da diversidade genética de cafeeiros do banco de germoplasma da UFV/EPAMIG utilizando marcadores RAPD. IV Simpósio de Pesquisa dos Cafés do Brasil. Londrina, PR, 2005.
- Alwala, S., Suman, A., Arro, J. A., Veremis, J. C., Kimbeng, C. A. Target region

- amplification polymorphism (TRAP) for assessing genetic diversity in sugarcane germplasm collections. Crop Science, v.46, p.448-455. 2006.
- Alwala, S., Kimbeng, C. A., Gravois, K. A., Bischoff, K. P. TRAP, a new tool for sugarcane breeding: comparisson with AFLP and coefficient of parentage. Journal American Society Sugar Cane Technologists, v.26, p.62-86. 2006.
- Anthony, F., Bertrand, B., Quiros, O., Wilches, A., Lashermes, P., Berthaud, J., Charrier, A. Genetic diversity of wild coffee (*Coffea arabica* L.) using molecular markers Euphytica, v.18, n.1, p.53-65. 2001.
- Aravind, L. D., V. M, Koonin, E. V. The domains of death: evolution of the apoptosis machinery. Trends in Biochemical Sciences, v.24, n.2, p.47-53. 1999.
- Baker, B., Zambryski, P., Staskawicz, B., Dinesh-Kumar, S. P. Signaling in plant-microbe interactions. Science, v.276, p.726-733. 1997.
- Barry, G. F. The use of the Monsanto draft rice genome sequence in research. Plant Physiology, v.125, p.1164-1165. 2001.
- Belaj, A., Trujillo, I., Rosa, R., Rallo, L., Giménez, M. J. Polymorphism and discrimination capacity of randomly amplified polymorphic markers in an olive germplasm bank. Journal of the American Society for Horticultural Science, v.126, n.1, p.64-71. 2001.
- Belkhadir, Y., Subramaniam, R., Dangl, J. Plant disease resistance protein signaling: NBS-LRR proteins and their partners. Current Opinion in Plant Biology, v.7, p.391-399. 2004.
- Bendahmane, A., Kanyuka, K., Baulcombe, D. C. The Rx gene from potato controls separate virus resistance and cell death responses Plant Cell, n.11, p.781-791. 1999.
- Bennetzen, J. Opening the door to comparative plant biology. Science, v.296, p.60-63. 2002.
- Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., Wheeler, D. L. GenBank. Nucleic Acids Research, v.35, p.21-25. 2007.
- Bent, A. F., Kunkel, B. N., Dahlbeck, D., Brown, K. L., Schmidt, R., Giraudat, J., Leung, J., Staskawicz, B. J. RPS2 of *Arabidopsis thaliana*: a leucine-rich repeat class of plant disease resistance genes. Science, v.265, p.1856-1860. 1994.
- Berthaud, J., Charrier, A. Genetic resources of Coffea. In: R. J. Clarke, Macrae, R (Ed.). Coffee. London: Elsevier Applied Science, v.4, 1988. Genetic resources of Coffea, p.1-42
- Berthou, F., Trouslot, P., Hamon, S., Vedel, F., Quetier, F. Electrophoretic

- analysis of the biochemical polymorphism of coffee trees - enzymatic variation in 18 wild populations - variation of the mitochondrial DNA in the species *Coffea canephora*, *Coffea eugenioides* and *Coffea arabica*. IX Colloque Scientifique International sur le Café. Londres: ASIC (Paris), 1980. 16 p.
- Berthou, P. Y., Correia, F., Prates, S., Taugourdeau, J. Essay de synthèse du cretace de l'Algarve. Biostratigrafie, paleogeografie et sedimentation argileuse. Bulletin d'Information des Geologie du Bassin de Paris, v.20, n.2, p.3-24. 1983.
- Bevan, M., Mayer, K., White, O., Eisen, J. A., Preuss, D., Bureau, T., Salzberg, S. L., Mewes, H. Sequence and analysis of the Arabidopsis genome. Current Opinion in Plant Biology, v.4, p.105-110. 2001.
- Bevan, M., Walsh, S. The Arabidopsis genome: A foundation for plant research. Genome Research, v.15, p.1632-1642. 2005.
- Bhat, P. R., Krishnakumar, V., Hendre, P. S., Rajendrakumar, P., Varshney, R. K., Aggarwal, R. K. Identification and characterization of expressed sequence tags-derived simple sequence repeats markers from robusta coffee variety 'CxR' (an interspecific hybrid of *Coffea canephora* × *Coffea congensis*). Molecular Ecology Notes, v.5, n.1, p.80-83. 2005.
- Binneck, E. As ômicas: Integrando a bioinformação. Biotecnologia, Ciência & Desenvolvimento, v.32, p.28-37. 2004.
- Boguski, M. S. Biosequence exegesis. Science, v.286, p.453-455. 1999.
- Borém, A., Giúdice, M., Sdiyama, T. Melhoramento Genômico. Viçosa, MG. 2003. 224 p.
- Borém, A. Aplicação dos Marcadores Moleculares no Melhoramento de Plantas. In: A. Borém, Caixeta, E. T. (Ed.). Marcadores Moleculares. Viçosa, 2006. Aplicação dos Marcadores Moleculares no Melhoramento de Plantas, p.79-84
- Borém, A., Caixeta, E. T. Marcadores Moleculares. Viçosa, MG. 2006. 374 p.
- Botella, M. A., Parker, J. E., Frost, L. N., Bittner-Eddy, P. D., Beynon, J. L., Daniels, M. J., Holub, E. B., Jones J. D. G. Three genes of the arabidopsis RPP1 complex resistance locus recognize distinct *Peronospora parasitica* avirulence determinants Plant Cell, v.10, p.1847-1860. 1998.
- Botstein, D., White, R.L., Skolnick, M., Davis, R.W. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. American Journal of Human Genetics, v.32, n.3, p.314-331. 1980.
- Bourne, H. R., Sanders, D. A., McCormick, F. The GTPase superfamily: conserved structure and molecular mechanism. Nature, v.349, n.6305,

p.117-127. 1991.

Brent, M. R. Genome annotation past, present, and future: How to define an ORF at each locus. Genome Research, v.15, p.1777-1786. 2005.

Cabral, T. A. T. Padrões moleculares, diversidade genética e mapa parcial de ligação do cafeeiro. Genética e Melhoramento, Universidade Federal de Viçosa, Viçosa, MG, 2001. 112 p.

Cabral, T. A. T., Sakiyama, N. S., Zambolim, L., Pereira, A. A., Barros, E. G., Sakiyama, C. C. H. Reproducibility of the RAPD marker and its efficiency in coffee tree genotype grouping analysis. Crop Breeding and Applied Biotechnology, v.2, n.1, p.121-129. 2002.

Cai, D., Kleine, M., Kifle, S., Harloff, H., Sandal, N. N., Marcker, K. A., Klein-Lankhorst, R. M., Salentijn, E. M. J., Lange, W., Stiekema, W. J., Wyss, U., Grundler, F. M. W., Jung, C. Positional cloning of a gene for nematode resistance in sugar beet. . Science, v.275, p.832-834. 1997.

Cannon, S. B., Zhu, H., Baumgarten, A. M., Spangler, R., May, G., Cook, D. R., Young, N. D. Diversity, distribution, and ancient taxonomic relationships within the TIR and non-TIR NBS-LRR resistance gene subfamilies. Journal of Molecular Evolution, v.54, n.4, p.548-562. 2002.

Cardoso, R. M. L., Zambolim, L., Chaves, G. M. Ocorrência no Brasil da raça XVI de *Hemileia vastatrix* Berk. et Br. coletada do germoplasma de *Coffea arabica* L. no Estado de Minas Gerais. Fitopatologia Brasileira, v.12, p.343-346. 1988.

Carvalho, V. L., Chalfoun, S. M. A ferrugem do cafeeiro. <http://www.coffeebreak.com.br/ocafezal.asp?SE=8&ID=158>. 2005.

Cros J, C., M. C., Trouslot, P., Anthony, F., Hamon, S., Charrier, A., Lashermes, P. Phylogenetic analysis of chloroplast DNA variation in *Coffea* L. Molecular Phylogenetics and Evolution, v.9, n.1, p.109-117. 1998.

Dangl, J. L., Jones, J. D. Plant pathogens and integrated defense responses to infection. Nature, v.411, p.826-833. 2001.

Deslandes, L. O., J.; Peeters, N.; Feng, D. X.; Khounlotham, M.; Boucher, C.; Somssich, I.; Genin, S., Marco, Y. Physical interaction between RRS1-R, a protein conferring resistance to bacterial wilt, and PopP2, a type III effector targeted to the plant nucleus. Proceedings of the National Academy of Sciences of the United States of America, v.13, p.8024-8029. 2003.

Dinesh-Kumar, S. P., Tham, W, Baker, B. J. Structure-function analysis of the tobacco mosaic virus resistance gene N. Proceedings of the National Academy of Sciences of the United States of America, v.97, n.26, p.14789-14794. 2000.

- Diniz, L. E. C. Mapeamento físico da região cromossômica próxima ao gene de resistência a *Meloidogyne exigua* (Mex-1) em *Coffea arabica* L. Tese. Genética e Melhoramento, Universidade Federal de Viçosa, Viçosa, 2004. 115 p.
- Diniz, L. E. C., Sakiyama, N. S., Lashermes, P., Caixeta, E. T., Oliveira, A. C. B., Zambolim, E. M., Loureiro, M. E., Pereira, A. A., Zambolim, L. Analysis of AFLP markers associated to the Mex-1 resistance locus in Icatu progenies. Crop Breeding and Applied Biotechnology, v.5, p.387-393. 2005.
- Dita, M. A., Rispaill, N., Prats, E., Rubiales, D., Singh, K. B. Biotechnology approaches to overcome biotic and abiotic stress constraints in legumes. Euphytica, v.147, p.1-24. 2006.
- Dixon, M. S., Jones, D. A., Keddie, J. S., Thomas, C. M., Harrison, K., Jones, J. D. The tomato Cf-2 disease resistance locus comprises two functional genes encoding leucine-rich repeat proteins. Cell, v.84, n.3, Feb p.451-459. 1996.
- Dolzan, M. 14 de abril é o dia internacional do café. <http://www.unificado.com.br/calendario/04/cafe.htm>. 2005.
- Durick, K., Mendlein, J., Xanthopoulos, K. G. Hunting with traps: genome-wide strategies for gene discovery and functional analysis. Genome Research, v.9, p.1019-1025. 1999.
- Dvorak, J., Akhunov, E. D., Akhunov, A. R., Deal, K. R., Luo, M. Molecular characterization of a diagnostic DNA marker for domesticated tetraploid wheat provides evidence for gene flow from wild tetraploid wheat to hexaploid wheat. Molecular and Molecular Evolution, v.23, p.1386-1396. 2006.
- Ecker, J. R., Cook, D. Genome studies and molecular genetics Unwrapping new layers of complexity in plant genomes. Current Opinion in Plant Biology, v.7, p.99-101. 2004.
- Edwards, D., Batley, J. Plant bioinformatics: from genome to phenome. TRENDS in Biotechnology, v.22, n.5, p.232-237. 2004.
- Ellis, J., Lawrence, G., Ayliffe, M., Anderson, P., Collins, N., Finnegan, J., Frost, D., Luck, J., Pryor, T. Advances in the molecular genetic analysis of the flax-flax rust interaction. Annual Review of Phytopathology, v.35, p.271-91. 1997.
- Ellis, J., Jones, D. Structure and function of proteins controlling strain-specific pathogen resistance in plants. Current Opinion in Plant Biology, v.1, p.288-293. 1998.
- Ellis, J., Dodds, P., Pryor, T. Structure, function and evolution of plant disease resistance genes. Current Opinion in Plant Biology, v.3, p.278-284. 2000.
- Ellis, J. G., Lawrence, G. J., Luck, J. E., Dodds, P. N. Identification of regions in

- alleles of the flax rust resistance gene L that determine differences in gene-for-gene specificity Plant Cell, v.11, n.3, p.495-506. 1999.
- Esques, A. B. Resistance. In: A. C. Kushallapa, Esques, A. B. (Ed.). Coffee Rust: epidemiology, resistance and management. Boca Raton, Florida: CRC Press, 1989. Resistance, p.171-291
- Fazuoli, L. C., Oliveira, A. C. B., Toma-Braghini, M., Silvarolla, M. B. Identification and use of sources of durable resistance to coffee leaf rust at IAC. In: L. Zambolim, Zambolim, E. M., Várzea, V. M. P. (Ed.). Durable resistance to coffee leaf rust. Viçosa UFV, DFP 2005. Identification and use of sources of durable resistance to coffee leaf rust at IAC, p.137-186
- Fernandez, D., Santos, P., Agostini, C., Bon, M., Petitot, A., Silva, M. C., Guerra-Guimarães, L., Ribeiro, A., Argout, X., Nicole, M. Coffee (*Coffea arabica* L.) genes early expressed during infection by the rust fungus (*Hemileia vastatrix*). Molecular Plant Pathology, v.5, n.6, p.527-536. 2004.
- Ferreira, M. A. J. F. Utilização das técnicas de marcadores moleculares na genética de populações, na genética quantitativa e no melhoramento de plantas. Boa Vista: Embrapa Roraima, v.1. 2003. 63 p. (Embrapa Roraima. Documentos, 1).
- Ferreira, M. E., Grattapaglia, D. Introdução ao uso de marcadores moleculares em análise genética. Brasília: EMBRAPA-CENARGEN. 1998. 220 p.
- Ferriol, M., Picó, B., Córdova, P. F., Nuez, F. Molecular diversity of a germplasm collection of squash (*Cucurbita moschata*) determined by SRAP and AFLP markers. Crop Science, v.44, p.653-664. 2004.
- Feys, B. J., Parker, J. E. Interplay of signaling pathways in plant disease resistance. Trends in Genetics, v.16, n.10, p.449-455. 2000.
- Flavell, R. B. Plant biotechnology Moral dilemmas Current Opinion in Plant Biology, v.3, p.143-146. 2000.
- Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J., Dougherty, B. A., Merrick, J. M., Mckenney, K., Sutton, G., Fitzhugh, W., Fields, C., Gocayne, J. D., Scott, J., Shirley, R., Liu, L., Glodek, A., Kelley, J. M., Weidman, J. F., Phillips, C. A., Spriggs, T., Hedblom, E., Cotton, M. D., Utterback, T. R., Hanna, M. C., Nguyen, D. T., Saudek, D. M., Brandon, R. C., Fine, L. D., Fritchman, J. L., Fuhrmann, J. L., Geoghagen, N. S. M., Gnehm, C. L., Mcdonald, L. A., Small, K. V., Fraser, C. M., Smith, H. O., Venter, J. C. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd.(first genome of free-living organism ever completely sequenced)(includes fold-out genome map)(Cover Story). Science, v.269, n.5223, p.496-513. 1995.
- Flor, H. H. Inheritance of pathogenicity in *Melampsora lini*. Phytopathology, v.32, p.653-669. 1942.

- Fontes, J. R. M. Heterose, capacidade combinatória e divergência genética estimada por análise de marcadores RAPD em cruzamentos entre cafeeiros catuaí (*Coffea arabica* L.) e híbridos de timor. Fitotecnia, Universidade Federal de Viçosa, Viçosa, 2001. 140 p.
- Fridman, E., Pichersky, E. Metabolomics, genomics, proteomics, and the identification of enzymes and their substrates and products. Current Opinion in Plant Biology, v.8, p.242-248. 2005.
- Gai, X. W., Lal, S., Xing, L.Q., Brendel, V., and Walbot, V. Gene discovery using the maize genome database ZmDB. Nucleic Acids Research, v.28, p.94-96. 2000.
- Gassmann, W., Hinsch, M. E., Staskawicz, B. J. The Arabidopsis RPS4 bacterial-resistance gene is a member of the TIR-NBS-LRR family of disease-resistance genes. Plant Journal, v.20, n.3, Nov p.265-277. 1999.
- Gelbart, W. M. Databases in genomic research. Science, v.282, p.659-661. 1998.
- Geromel, C., Ferreira, L. P., Carmelo Guerreiro, S. M. C., Cavalari, A. A., Pot, D., Pereira, L. P. P., Leroy, T., Vieira, L. G. E., Mazzafera, P., Marraccini, P. Biochemical and genomic analysis of sucrose metabolism during coffee (*Coffea arabica*) fruit development. Journal of Experimental Botany, v.57, n.12, p.3243-3258. 2006.
- Gesteira, A. S., Micheli, F., Carels, N., Silva, A. C., Gramacho, K. P., Schuster, I., Macedo, J. N., Pereira, G. A. G., Cascardo, J. C. M. Comparative Analysis of Expressed Genes from Cacao Meristems Infected by *Moniliophthora perniciosa*. Annals of Botany, v.100, p.129-140. 2007.
- Goff, S. A., Ricke, D., Lan, T., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., Halley, D., Hutchison, D., Martin, C., Katagiri, F., Lange, B. M., Moughamer, T., Xia, Y., Budworth, P., Zhong, J., Miguel, T., Paszkowski, U., Zhang, S., Colbert, M., Sun, W., Chen, L., Cooper, B., Park, S., Wood, T. C., Mao, L., Quail, P., Wing, R., Dean, R., Yu, Y., Zharkikh, A., Shen, R., Sahasrabudhe, S., Thomas, A., Cannings, R., Gutin, A., Pruss, D., Reid, J., Tavtigian, S., Mitchell, J., Eldredge, G., Scholl, T., Miller, R. M., Bhatnagar, S., Adey, N., Rubano, T., Tusneem, N., Robinson, R., Feldhaus, J., Macalma, T., Oliphant, A., Briggs, S. A draft sequence of the rice genome genome (*Oryza sativa* L. ssp japonica). Science, v.269, n.5565, p.92-100. 2002.
- Gostimsky, S. A., Kokaeva, Z. G., Konovalov, F. A. Studying plant genome variation using molecular markers. Russian Journal of Genetics, v.41, n.4, p.378-388. 2005.
- Gouveia, M. M. C., Ribeiro, A., Várzea, V. M. P., Rodrigues-Jr, C. J. Genetic diversity in *Hemileia vastatrix* based on RAPD markers. Mycologia, v.97,

- n.2, p.396-404. 2005.
- Grant, M. R., Godiard, L., Straube, E., Ashfield, T., Lewald, J., Sattler, A., Innes, R. W., Dangl, J. L. Structure of the Arabidopsis RPM1 gene enabling dual specificity disease resistance. Science, v.269, p.843-846. 1995.
- Grattapaglia, D. Genolyptus. In: A. Borém, Giúdice, M., Sediyaama, T. (Ed.). Melhoramento Genômico. Viçosa, 2003. Genolyptus, p.51-72
- Grivet, L., Arruda, P. Sugarcane genomics: depicting the complex genome of an important tropical crop. Current Opinion in Plant Biology, v.5, p.122-127. 2001.
- Guo, W., Cai, C., Wang, C., Han, Z., Song, W., Wang, K., Niu, X., Wang, C., Lu, K., Shi, B., Zhang, T. A microsatellite-based, gene-rich linkage map reveals genome structure, function, and evolution in *Gossypium*. Genetics: Published Articles Ahead of Print, published on April 3, 2007 as [10.1534/genetics.107.070375](https://doi.org/10.1534/genetics.107.070375). 2007.
- Guzzo, S. D. Aspectos bioquímicos e moleculares da resistência sistêmica adquirida em cafeeiro contra *Hemileia vastatrix*. Centro de Energia Nuclear na Agricultura, Universidade de São Paulo, São Paulo, 2004. 236 p.
- Hammond-Kosack, K. E., Jones, J. D. G. Resistance gene-dependent plant defense responses. The Plant Cell, v.8, p.1773-1791. 1996.
- Hammond-Kosack, K. E., Jones, D. G. Plant disease resistance genes. Annual Review of Plant Physiology and Plant Molecular Biology v.48, p.575-607. 1997.
- Hammond-Kosack, K. E., Parker, J. E. Deciphering plant-pathogen communication: fresh perspectives for molecular resistance breeding. Current Opinion in Biotechnology, v.14, p.177-193. 2003.
- Hamosh, A., Scott, A. F., Amberger, J., Valle, D., McKusick, V. A. Online Mendelian Inheritance in Man (OMIM). Human Mutation, v.15, p.57-61. 2000.
- Höfte, H., Desprez, T., Amselem, J., Chiapello, H., Caboche, M., Moisan, A., Jourjon, M. F., Charpentreau, J. L., Berthomieu, P., Guerrier, D., Giraudat, J., Quigley, F., Thomas, F., Yu, D. Y., Mache, R., Raynal, M., Cooke, R., Grellet, F., Delseny, M., Parmentier, Y., Marcillac, G., Gigot, C., Fleck, J., Philipps, G., Axelos, M., Bardet, C., Tremousaygue, D., Lescure, B. An inventory of 1152 expressed sequence tags obtained by partial sequencing of cDNAs from *Arabidopsis thaliana*. Plant Journal, v.4, p.1051-1061. 1993.
- Hu, J., Vick, B. A. Target Region Amplification Polymorphism: A Novel Marker Technique for Plant Genotyping. Plant Molecular Biology Reporter, v.21, p.289-294. 2003.

- Hulbert, S. H., Webb, C. A., Smith, S. M., Sun, Q. Resistance gene complexes: evolution and utilization. Annual Review of Phytopathology, v.39, p.285-312. 2001.
- Jansson, S., Buckler, E. S. Genome studies and molecular genetics: Genomics — deeper and wider in order to understanding plant diversity. Current Opinion in Plant Biology, v.10, p.107-108. 2007.
- Jeffreys, A. J. W., V.; Thein, S.L. Hypervariable minisatellite regions in human DNA. Nature, v.314, p.67-73. 1985.
- Jia, Y., Mcadams, S. A., Bryan, G. T., Hershey, H. P., Valent, B. Direct interaction of resistance gene and avirulence gene products confers rice blast resistance. European Molecular Biology Organization Journal, v.19, p.4004-4014. 2000.
- Jones, D. A., Thomas, C. M., Hammond-Kosack, K. E., Balint-Kurti, P. J., Jones, J. D. G. Isolation of the tomato Cf-9 gene for resistance to *Cladosporium fulvum* by transposon tagging. Science, v.266, p.789-793. 1994.
- Jones, J., D. G. Putting knowledge of plant disease resistance genes to work. Current Opinion in Plant Biology, v.4, p.281-287. 2001.
- Kaiser, J. From genome to functional genomics. Science, v.288, n.5472, p.1715. 2005.
- Kobe, B., Deisenhofer, J. Proteins with leucine-rich repeats. Current Opinion in Structure Biology, v.8, n.5, p.409-416. 1995.
- Kulikova, T., Aldebert, P., Althorpe, N., Baker, W., Baldwin, A., Bates, K., Browne, P., Van Den Broek, A., Castro, M., Cochrane, G., Duggan, K., Eberhardt, R., Faruque, N., Garcia-Pastor, M., Harte, N., Kanz, C., Leinonen, R., Lin, Q., Lombard, V., Lopez, R., Mancuso, R., Mchale, M., Nardone, F., Silventoinen, V., Stoehr, P., Stoesser, G., Tuli, M. A., Tzouvara, K., Vaughan, R., Wu, D., Zhu, W., Apweiler, R. The EMBL nucleotide sequence database. Nucleic Acids Research, v.32, p.D27-D30. 2004.
- Kulikova, T., Akhtar, R., Aldebert, P., Althorpe, N., Andersson, M., Baldwin, A., Bates, K., Bhattacharyya, S., Bower, L., Browne, P., Castro, M., Cochrane, G., Duggan, K., Eberhardt, R., Faruque, N., Hoad, G., Kanz, C., Lee, C., Leinonen, R., Lin, Q., Lombard, V., Lopez, R., Lorenc, D., McWilliam, H., Mukherjee, G., Nardone, F., Pastor, M. P. G., Plaister, S., Sobhany, S., Stoehr, P., Vaughan, R., Wu, D., Zhu W., Apweiler, R. EMBL nucleotide sequence database in 2006. Nucleic Acids Research, v.35, p.16-20. 2007.
- Lashermes, P., Couturon, E., Moreau, N., Paillard, M., Louarn, J. Inheritance and genetic mapping of self-incompatibility in *Coffea canephora* Pierre. Theoretical and Applied Genetics, v.93, p.458-462. 1996.

- Lashermes, P., Andrzejewski, S., Bertrand, B., Combes, M. C., Dussert, S., Graziosi, G., Trouslot, P., Anthony, F. Molecular analysis of introgressive breeding in coffee (*Coffea arabica* L.). Theoretical and Applied Genetics, v.100, p.139-146. 2000.
- Lashermes, P., Paczek, V., Trouslot, P., Combes, M. C., Couturon, E., Charrier, A. Single-locus inheritance in the allotetraploid *Coffea arabica* L. and interspecific hybrid *C. arabica* X *C. canephora*. The Journal of Heredity, v.91, n.1, p.81-85. 2000.
- Lashermes, P. C., M-C., Robert, J., Trouslot, P., D'hont, A., Anthony, F., Charrier, A. Molecular characterisation and origin of the *Coffea arabica* L. genome. Molecular and General Genetics v.261, p.259-266. 1999.
- Lazo, G. R., Chao, S., Hummel, D. D., Edwards, H., Crossman, C. C., Lui, N., Matthews, D. E., Carollo, V. L., Hane, D. L., You, F. M., Butler, G. E., Miller, R. E., Close, T. J., Peng, J. H., Lapitan, N. L. V., Gustafson, J. P., Qi, L. L., Echalié, B., Gill, B. S., Dilbirligi, M., Randhawa, H. S., Gill, K. S., Greene, R. A., Sorrells, M. E., Akhunov, E. D., Dvorak, J., Linkiewicz, A. M., Dubcovsky, J., Hossain, K. G., Kalavacharla, V., Kianian, S. F., Mahmoud, A. A., Miftahudin, Ma, X. F., Conley, E. J., Anderson, J. A., Pathan, M. S., Nguyen, H. T., McGuire, P. E., Qualset, C. O., Anderson, O. D. Development of an Expressed Sequence Tag (EST) Resource for Wheat (*Triticum aestivum* L.): EST Generation, Unigene Analysis, Probe Selection and Bioinformatics for a 16,000-Locus Bin-Delineated Map. Genetics, v.168, p.585-593. 2004.
- Lehman, P. Structure an evolution of plant disease resistance genes Journal Of Applied Genetics, v.43, n.4, p.403-414. 2002.
- Leister, R. T., Ausubel, F. M., Katagiri, F. Molecular recognition of pathogen attack occurs inside of plant cells in plant disease resistance specified by the Arabidopsis genes RPS2 and RPM1. Proceedings of the National Academy of Sciences of the United States of America, v.93, p.15497-15502. 1996.
- Li, G., Quiros, C. F. Sequence-related amplified polymorphism (SRAP), a new marker system based on a simple PCR reaction: its application to mapping and gene tagging in Brassica. Theoretical and Applied Genetics, v.103, p.455-461. 2001.
- Li, L., Hong, R., Hastings, W. Three functional luciferase domains in a single polypeptide chain. Proceedings of the National Academy of Sciences of the United States of America, v.94, n.17, p.8954-8958. 1997.
- Liang, F., Holt, I., Perteá, G., Karamycheva, S., Salzberg, S. L., Quackenbush, J. An optimized protocol for analysis of EST sequences. Nucleic Acids Research, v.28, n.18, p.3657-3665. 2000.
- Lin, C., Mueller, L. A., Carthy, J. M., Cruzillat, D., Pétiard, V., Tanksley, S. D. Coffee and tomato share common gene repertoires as revealed by deep

- sequencing of seed and cherry transcripts. Theoretical and Applied Genetics, v.112, n.1, p.114-130. 2005.
- Litt, M., Luty, J.A 44:397-401. An hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. American Journal of Human Genetics, v.44, p.397-401. 1989.
- Luo, M., Dang, P., Guo, B. Z., He, G., Holbrook, C. C., Bausher, M. G., Lee, R. D. Generation of expressed sequence tags (ESTs) for gene discovery and marker development in cultivated peanut. Crop Science, v.45, p.346-353. 2005.
- Lupas, A. Coiled coils: new structures and new functions. Trends in Biochemical Science, v.21, n.10, p.375-82. 1996.
- Madlung, A., Comai, L. The effect of stress on genome regulation and structure. Annals of Botany, v.94, p.481-495. 2004.
- Maglott, D., Fatz, K. S., Sicotte, H., Pruitt, K. D. NCBI's LocusLink and RefSeq. Nucleic Acids Research, v.28, n.1, p.126-128. 2000.
- Maglott, D., Ostell, J., Pruitt, K. D., Tatusova, T. Entrez Gene: gene-centered information at NCBI. Nucleic Acids Research, v.35, p.D26–D31. 2007.
- Maluf, M. P., Silvestrini, M., Ruggiero, L. M. C., Filho, O. G., Colombo, C. A. Genetic diversity of cultivated *Coffea arabica* inbred lines assessed by RAPD, AFLP and SSR marker systems. Scientia Agricola (Piracicaba, Braz.), v.62, n.4, p.366-373. 2005.
- Manifesto, M. M., Schlatter, A. R., Hopp, H. E., Suárez, E. Y., Dubcovsky, J. Quantitative evaluation of genetic diversity in wheat germplasm using molecular markers. Crop Science, v.41, p.682-690. 2001.
- Margis-Pinheiro, M., Sandroni, M., Lummerzheim, M. Oliveira, D. E. A defesa das plantas contra as doenças. Ciência Hoje. 147 1999.
- Martienssen, R., Mccombie, W. R. The First Plant Genome. Cell, v.105, p.571-574. 2001.
- Martin, G., Brommonschenkel, S., Chunwongse, J., Frary, A., Ganai, M., Spivey, R., Wu, T., Earle, E., Tanksley, S. Map-based cloning of a protein kinase gene conferring disease resistance in tomato. Science, v.262, p.1432-1436. 1993.
- Martin, G. Functional analysis of plant disease resistance genes and their downstream effectors. Current Opinion in Plant Biology, v.2, p.273-279. 1999.
- Martin, G., Bogdanove, A., Sessa, G. Understanding the functions of plant disease resistance proteins. Annual Review of Plant Biology, v.54, p.23-61.

2003.

- Matthews, B. F., Devine, T. E., Weisemann, J. M., Beard, H. S., Lewers, K. S., Macdonald, M. H., Park, Y., Maiti, R., Lin, J., Kuo, J., Pedroni, M. J., Cregan, P. B., Saunders, J. A. Incorporation of sequenced cDNA and genomic markers into the soybean genetic map. Crop Science, v.41, p.516-521. 2001.
- Mcdowell, J. M., Dhandaydham, M., Long, T. A., Aarts, M. G. M., Goff, S., Holub, E. B, Dangl, J. L. Intragenic recombination and diversifying selection contribute to the evolution of downy mildew resistance at the RPP8 locus of Arabidopsis. Plant Cell, v.10, n.11, p.1861-1874. 1998.
- Mcdowell, J. M., Woffenden, B. J. Plant disease resistance genes: recent insights and potential applications. Trends in Biotechnology, v.21, n.4, p.178-83. 2003.
- McGinnis, S., Madden, T. L. BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Research, v.32, p.w20-w25. 2004.
- Meinke, D., Tanksley, S. Genome studies and molecular genetics: The maturation and specialization of plant genomics. Current Opinion in Plant Biology, v.3, p.95-96. 2000.
- Meinke, D. W., Cherry, J. M., Dean, C., Rounsley, S. D., Koornneef, M. *Arabidopsis thaliana*: A model plant for genome analysis. Science, v.282, p.662-682. 1998.
- Melchers, L. S., Stuiver, M. H. Novel genes for disease-resistance breeding. Current Opinion in Plant Biology v.3, p.147-152. 2000.
- Meyers, B. C., Dickerman, A. W., Michelmore, R. W., Sivaramakrishnan, S., Sobral, B. W., Young, N. D. Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. Plant Journal, v.20, n.3, p.317-332. 1999.
- Meyers, B. C., Kozik, A., Griego, A., Kuang, H., Michelmore, R. W. Genome-wide analysis of NBS-LRR-encoding genes in arabidopsis. The Plant Cell, v.15, p.809-834. 2003.
- Meyers, B. C., Kaushik, S., Nandety, R. S. Evolving disease resistance genes. Current Opinion in Plant Biology, v.8, p.129-134. 2005.
- Michelmore, R. Genomic approaches to plant disease resistance. Current Opinion in Plant Biology, v.3, p.125-131. 2000.
- Michelmore, R. W., Paran, I., Kesseli, R. V. Identification of markers linked to disease-resistance genes by bulked segregant analysis: A rapid method to detect markers in specific genomic regions by using segregating populations. Proceedings of the National Academy of Sciences of the United

- States of America, v.88, p.9828-9832. 1991.
- Michelmore, R. W., Meyers, B. C. Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. Genome Research, v.8, p.1113-1130. 1998.
- Miklas, P. N., Hub, J., Grünwaldc, N. J., Larsena, K. M. Potential application of TRAP (Targeted Region Amplified Polymorphism) markers for mapping and tagging disease resistance traits in common bean Crop Science, v.46, p.910-916. 2006.
- Milach, S. C. K. Marcadores moleculares em plantas. Porto Alegre. 1998. 141 p. (UFRGS)
- Mindrinos, M., Katagiri, F., Yu, G. L., Ausubel, F. M. The *A. thaliana* disease resistance gene RPS2 encodes a protein containing a nucleotide-binding site and leucine-rich repeats. Cell, v.78, n.6, p.1089-1099. 1994.
- Mitchell-Olds, Y., Clauss, M. J. Plant evolutionary genomics. Current Opinion in Plant Biology, v.5, p.74-79. 2002.
- Mlot, C. Plant biology in the genome era. Science, v.281, n.5375, p.331-332. 1998.
- Moffett, P., Farnham, G., Peart, J., Baulcombe, D. C. Interaction between domains of a plant NBS-LRR protein in disease resistance-related cell death. The EMBO Journal, v.21, n.17, p.4511-4519. 2002.
- Moloney, M., Peacock, J. Plant biotechnology. Current Opinion in Plant Biology, v.8, p.163-164. 2005.
- Moore, G., Devos, K. M., Wang, Z., Gale, M. D. Grasses, line up and form a circle. Current Biology, v.5, n.7, p.737-739. 1995.
- Moreno, G., Peña, M., Acuña, R., Aponte, M. E., Herrera, J. C., Molina, D., Gaitán, A., Cristancho, M., Moncada, P., Góngora, C., Cadena, G. Aplicaciones de la biotecnología al mejoramiento genético del café, en Colombia. In: I. Iapar (Ed.). International Seminar on Biotechnology in the Coffee Agroindustry Londrina Curitiba : UFPR, 2000. Aplicaciones de la biotecnología al mejoramiento genético del café, en Colombia, p.47-53
- Muñoz, L. C., Blair, M. W., Duque, M. C., Tohme, J., Roca, W. Introgression in common bean X tepary bean interspecific congruity-backcross lines as measured by AFLP markers. Crop Science, v.44, p.637-645. 2004.
- Nagy, E. D., Lelley, T. Genetic and physical mapping of sequence-specific amplified polymorphic (SSAP) markers on the 1RS chromosome arm of rye in a wheat background. Theoretical and Applied Genetics, v.107, p.1271-1277. 2003.

- Nasu, S., Suzuki, J., Ohta, R., Hasegawa, K., Yui, R., Kitazawa, N., Monna, L., Minobe, Y. Search for and Analysis of Single Nucleotide Polymorphisms (SNPs) in Rice (*Oryza sativa*, *Oryza rufipogon*) and Establishment of SNP Markers. DNA Research, v.9, p.163-171. 2002.
- Newcomb, R. D., Crowhurst, R. N., Gleave, A. P., Rikkerink, E. H. A., Allan, A. C., Beuning, L. L., Bowen, J. H., Gera, E., Jamieson, K. R., Janssen, B. J., Laing, W. A., McCartney, S., Nain, B., Ross, G. S., Snowden, K. C., Souleyre, E. J. F., Walton, E. F., Yauk, Y. K. Analyses of Expressed Sequence Tags from Apple. Plant Physiology, v.144, p.147-166. 2006.
- Nogueira, F. T. S., Rosa-Jr, V. E., Menossi, M., Ulian, E. C., Arruda, P. RNA expression profiles and data mining of sugarcane response to low temperature. Plant Physiology, v.132, p.1811-1824. 2003.
- Noir, S., Combes, M. C., Anthony, F., Lashermes, P. Origin, diversity and evolution of NBS-type disease-resistance gene homologues in coffee trees (*Coffea* L.). Molecular Genetics Genomics, v.265, p.654-662. 2001.
- Noir, S., Anthony, F., Bertrand, B., Combes, M. C., Lashermes, P. Identification of a major gene (Mex-1) from *Coffea canephora* conferring resistance to *Meloidogyne exigua* in *Coffea arabica*. Plant Pathology, v.52, p.97-103. 2003.
- Noir, S., Patheyron, S., Combes, M-C., Lashermes, P., Chalhou, B. Construction and characterisation of a BAC library for genome analysis of the allotetraploid coffee species (*Coffea arabica* L.). Theoretical and Applied Genetics, v.109, p.225-230. 2004.
- Oliveira, A. C. B., Caixeta, E. T., Zambolim, E. M., Diniz, L. E. C. Avanços Tecnológicos em Biologia Molecular: Marcadores de DNA no Melhoramento do Cafeeiro. In: L. Zambolim (Ed.). Produção integrada de café. Viçosa: UFV/DFP, 2003. Avanços Tecnológicos em Biologia Molecular: Marcadores de DNA no Melhoramento do Cafeeiro, p.247-278
- Ori, N., Eshed, Y., Paran, I., Presting, G., Aviv, D., Tanksley, S., Zamir, D., Fluhra, R. The 12C family from the wilt disease resistance locus 12 belongs to the nucleotide binding, leucine-rich repeat superfamily of plant resistance genes. The Plant Cell, v.9, p.521-532. 1997.
- Orozco-Castillo, C., Chalmers, K. J., Waugh, R., Powell, W. Detection of genetic diversity and selective gene introgression in coffee using RAPD markers. Theoretical and Applied Genetics, v.87, p.934-940. 1994.
- Orozco-Castillo, C., Chalmers, K. J., Powell, W., Waugh, R. RAPD and organelle specific PCR re-affirms taxonomic relationships within the genus *Coffea*. Plant Cell Reports, v.15, n.5, p.337-341. 1996.
- Paillard, M. L., P., Pétiard, V. Construction of a molecular linkage map in coffee. Theoretical and Applied Genetics, v.93, p.41-47. 1996.

- Pan, Q., Wendel, J., Fluhr, R. Divergent evolution of plant NBS-LRR resistance gene homologues in dicot and cereal genomes. Journal of Molecular Evolution, v.50, n.3, p.203-213. 2000.
- Paran, I. M., R.W. Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. Theoretical and Applied Genetics, v.85, p.985-993. 1993.
- Parker, J. E., Coleman, M. J., Szabo, V., Frost, L. N., Schmidt, R., Vanderbiezen, E. A., Moores, T., Dean, C., Daniels, M. J., Jones, J. D. G. The Arabidopsis downy mildew resistance gene RPP5 shares similarity to the toll and interleukin-1 receptors with N and L6. Plant Cell, v.9, n.6, p.879-894. 1997.
- Parniske, M., Hammondkosack, K. E., Golstein, C., Thomas, C. M., Jones, D. A., Harrison, K., Wulff, B. B. H., Jones, J. D. G. Novel disease resistance specificities result from sequence exchange between tandemly repeated genes at the Cf-4/9 locus of tomato. Cell, v.91, n.6, p.821-832. 1997.
- Pashley, C. H., Ellis, J. R., Mccauley, D. E., Burke, J. M. EST databases as a source for molecular markers: Lessons from *Helianthus*. The Journal of Heredity, v.97, n.4, p.381-388. 2006.
- Pei, X., Li, S., Jing, Y., Zhang, Y., Wang, Z., Jia, S. Isolation, characterization and phylogenetic analysis of the resistance gene analogues (RGAs) in banana (*Musa* spp.). Plant Science. 2007.
- Perrin, R. M., Wigge, P. A. Genome studies and molecular genetics: Plant biotechnology - Web alert. Current Opinion in Plan Biology, v.5, p.89-90. 2002.
- Perry, D. J., Bousquet, J. Genetic diversity and mating system of post-fire and porst-harvest black spruce: an investigation using codominant sequence-tagged-site (STS) markers. Canadian Journal of Forest Resources, v.31, p.32-40. 2001.
- Pertsemlidis, A., Fondon, J. Having a BLAST with bioinformatics (and avoiding BLASTphemy). Genome Biology, v.2, n.10, p.1-10. 2001.
- Pickeral, O., Boguski, M. S. The bioinformatics bookshelf: teach yourself computational biology. Cell, v.96, n.4, p.451-455. 1999.
- Picoult-Newberg, L., Ideker, T. E., Pohl, M. G., Taylor, S. L., Donaldson, M. A., Nickerson, D. A., Boyce-Jacino, M. Mining SNPs from EST databases. Genome Research, v.9, p.167-174. 1999.
- Pinto, F. O., Maluf, M. P., Guerreiro-Filho, O. Study of simple sequence repeat markers from coffee expressed sequences associated to leaf miner resistance. Pesquisa Agropecuária Brasileira, v.42, n.3, p.377-384. 2007.

- Poncet, V., Hamon, P., Sauvage De Saint Marc, M. B., Bernard, T., Hamon, S., Noirot, M. Base composition of *Coffea* AFLP sequences and their conservation within the genus. The Journal of Heredity, v.96, n.1, p.59-65. 2005.
- Pontius, J. U., Wagner, L., Schuler, G. D. UniGene: a unified view of the transcriptome. In: J. Mcentyre, Ostell, J. (Ed.). The NCBI Handbook. National Center for Biotechnology Information, Bethesda, MD. , 2003. UniGene: a unified view of the transcriptome.
- Prakash, N. S., Marques, D. V., Varzea, V. M. P., Silva, M. C., Combes, M. C. Lashermes, P. Introgression molecular analysis of a leaf rust resistance gene from *Coffea liberica* into *C. arabica* L. Theoretical and Applied Genetics, v.109, p.1311-1317. 2004.
- Prosdoci, P., Cerqueira, G. C., Binneck, E., Silva, A. F., Reis, A. N., Junqueira, A. C. M., Santos, A. C. F., Júnior, A. N., Wust, C. I., Filho, F. C., Kessedjian, J. L., Petretski, J. H., Camargo, L. P., Ferreira, R. G. M., Lima, R. P., Pereira, R. M., Jardim, S., Sampaio, V. S., Folgueras-Flatschart, A. V. Bioinformática: Manual do usuário. Biociência, Ciência & Desenvolvimento, v.29, p.12-25. 2002.
- Pruitt, K., Maglott, D. R. RefSeq and LocusLink: NCBI gene-centered resources. Nucleic Acids Research, v.29, n.1, p.137-140. 2001.
- Pruitt, K. D., Katz, K. S., Sicotte, H., Maglott, D. R. Introducing RefSeq and LocusLink: curated human genome resources at the NCBI. Trends in Genetics, v.16, p.44-47. 2000.
- Pruitt, K. D., Tatusova, T., Maglott, D. R. NCBI reference sequence project: update and current status. Nucleic Acids Research, v.31, n.1, p.34-37. 2003.
- Pruitt, K. D., Tatusova, T., Maglott, D. R. NCBI reference sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Research, v.33, p.501-504. 2005.
- Rabinowicz, P. D., Bennetzen, J. L. The maize genome as a model for efficient sequence analysis of large plant genomes. Current Opinion in Plant Biology, v.9, p.149-156. 2006.
- Raina, S. N., Mukai, Y., Yamamoto, M. In situ hybridization identifies the diploid species of *Coffea arabica* (Rubiaceae). Theoretical and Applied Genetics, v.97, p.1204-1209. 1998.
- Richly, E., Kurth, J., Leister, D. Mode of amplification and reorganization of resistance genes during recent *Arabidopsis thaliana* evolution. Molecular Biology Evolution, v.19, p.76-84. 2002.
- Richmond, T., Somerville, S. Chasing the dream: plant EST microarrays.

- Current Opinion in Plant Biology, v.3, p.108-116. 2000.
- Richter, T. E., Ronald, P. C. The evolution of disease resistance genes. Plant Molecular Biology, v.42, n.1, p.195-204. 2000.
- Ronald, P. C. Resistance gene evolution. Current Opinion in Plant Biology, v.1, p.294-298. 1998.
- Roos, D. S. Bioinformatics – trying to swim in a sea of data. Science, v.291, p.1260-1261. 2001.
- Rose, E. A. Applications of the polymerase chain reaction to genome analysis. The FASEB Journal, v.5, p.46-54. 1991.
- Rossi, M., Araujo, P. G., Paulet, F., Garsemeur, O., Dias, V. M., Chen, H., Sluys, V., D'hont, A. Genomic distribution and characterization of EST-derived resistance genes analogs (RGAs) in sugarcane. Molecular Genetics and Genomics, v.269, p.406-419. 2003.
- Rouz, P., Pavyt, N., Rombauts, S. Genome annotation: which tools do we have for it? Current Opinion in Plant Biology, v.2, p.90-95. 1999.
- Rowland, L. J., Mehra, S., Dhanaraj, A. L., Ogden, O. L., Slovin, J. P. Development of EST-PCR markers for DNA fingerprinting and genetic relationship studies in blueberry (*Vaccinium*, section *Cyanococcus*). Journal of the American Society for Horticultural Science, v.128, n.5, p.682-690. 2003.
- Rowland, L. J., Dhanaraj, A. L., Polashock, J. J., Arora, R. Utility of Blueberry-derived EST-PCR. HortScience, v.38, n.7, p.1428-1432. 2003.
- Ruas, P. M., Ruas, C. F., Azevedo, L. B., Carvalho, V. P., Ruas, E. A., Sera, T. Identificação de marcador molecular ligado ao porte de planta em *C. Arabica*. I Simpósio de Pesquisa de Cafés do Brasil. Poços de Caldas/MG, 2000. 153-155 p.
- Rudd, S., Schoof, H., Mayer, K. PlantMarkers—a database of predicted molecular markers from plants. Nucleic Acids Research, v.33, p.d628-d632. 2005.
- Sales, R. M. O. B., Barbosa, A. V., Silva, F. R. O banco de dados do “Projeto Genoma Café” na EMBRAPA Recursos Genéticos e Biotecnologia (CENARGEN). V Simpósio de Pesquisa dos Cafés do Brasil. Águas de Lindóia, SP, 2007.
- Salmeron, J., Herrera-Estrella, L. R. Plant biotechnology Fast-forward genomics for improved crop production. Current Opinion in Plant Biology, v.9, p.177-179. 2006.
- Santos, C. M. R., Martins, N. F., Silva, F. R., Costa, M. M., Souza-Jr., M. T. O

- Projeto Genoma da Bananeira – uma iniciativa mundial para o benefício dos pequenos produtores. Revista Brazila Esperantisto, v.325. 2003.
- Santos, F. R., Ortega, J. M. Bioinformática. In: A. Borém, Del Giúdice, M., Sedyjama, T. (Ed.). Melhoramento Genômico. Viçosa, MG: UFV, v.1, 2003. Bioinformática, p.165-185
- Saraste, M., Sibbald, P. R., Wittinghofer, A. The P-loop--a common motif in ATP- and GTP-binding proteins. Trends in Biochemical Sciences, v.15, n.11, p.430-434. 1990.
- Sasaki, T., Burr. B. International Rice Genome Sequencing Project: the effort to completely sequence the rice genome. Current Opinion in Plan Biology, v.3, p.138-141. 2000.
- Schubert, R., Mueller-Starck, G., Riegel, R. Development of EST-PCR markers and monitoring their intrapopulational genetic variation in *Picea abies* (L.) Karst. Theoretical and Applied Genetics, v.103, p.1223-1231. 2001.
- Sera, T., Ruas, P. M., Ruas, C. F., Diniz, L. E. C., Carvalho, V. P., Rampim, L., Ruas, E. A., Silveira, S. R. Genetic polymorphism among 14 elite *Coffea arabica* L. cultivars using RAPD markers associated with restriction digestion. Genetics and Molecular Biology, v.26, n.1, p.59-64. 2003.
- Sera, T., Sera, G. H., Ito, D. S., Doi, D. S. Coffee breeding for durable resistance to leaf rust disease at Instituto Agronômico do Paraná. In: L. Zambolim, Zambolim, E. M., Várzea, V. M. P (Ed.). Durable resistance to coffee leaf rust. Viçosa : UFV, DFP, 2005. Coffee breeding for durable resistance to leaf rust disease at Instituto Agronômico do Paraná, p.187-214
- Shanmugam, V. Role of extracytoplasmic leucine rich repeat proteins in plant defence mechanisms. Microbiological Research, v.160, n.1, p.83-94. 2005.
- Shoemaker, R., Keim, P., Vodkin, L., Retzel, E., Clifton, S. W., Waterston, R., Smoller, D., Coryell, V., Khanna, A., Erpelding, J., Gai, X., Brendel, V., Raph-Schmidt, C., Shoop., Vielweber, C. J., Schmatz, M., Pape. D., Theising, B., Martin, J., Dante, M., Wylie, T., Granger, C. A Compilation of Soybean ESTs: Generation and Analysis. Genome, v.45, p.329-338. 2002.
- Silva, A. B., Wanderley-Nogueira, A. C., Silva, R. R. M., Berlarmino, L. C., Soares-Cavalcanti, N. M., Benko-Iseppon, A. M. In silico survey of resistance (R) genes in *Eucalyptus* transcriptome. Genetics and Molecular Biology, v.28, n.3, p.562-574. 2005.
- Silva, F. G., Iandolino, A., Al-Kayal, F., Bohlmann, M. C., Cushman, M. A., Lim, H., Ergul, A., Figueroa, R., Kabuloglu, E. K., Osborne, C., Rowe, J., Tattersall, E., Leslie, A., Xu, J., Baek, J., Cramer, G. R., Cushman, J. C., Cook, D. R. Characterizing the grape transcriptome. Analysis of expressed sequence tags from multiple *Vitis* species and development of a compendium of gene expression during berry development. Plant

- Physiology, v.139, p.574-597. 2005.
- Silva, M. C., Várzea, V., Guerra-Guimarães, L., Azinheira, H. G., Fernandez, D., Petitot, A., Bertrand, B., Lashermes, P., Nicole, M. Coffee resistance to the main diseases: leaf rust and coffee berry disease. Brazilian Journal of Plant Physiology, v.18, n.1, p.119-147. 2006.
- Siqueira, T. V. A cultura do Café: 1961-2005. BNDES Setorial. Rio de Janeiro, p.205-270. 2005
- Snowdon, R. J., Friedt, W. Molecular markers in *Brassica* oilseed breeding: current status and future possibilities. Plant Breeding, v.123, p.1-8. 2004.
- Somerville, C., Somerville, S. Plant functional genomics. Science, v.285, p.380-383. 1999.
- Song, W. Y., Wang, G. L., Chen, L. L., Kim, H. S., Pi, L. Y., Holsten, T., Gardner, J., Wang, B., Zhai, W., Zhu, L., Fauquet, C., Ronald, P. A receptor kinase-like protein encoded by the rice disease resistance gene, Xa21. Science, v.270, p.1804-1806. 1995.
- Staskawicz, B. J., Ausubel, F. M., Maker, B. J., Ellis, J. G., Jones, J. D. G. Molecular genetics of plant disease resistance. Science, v.268, p.661-667. 1995.
- Staskawicz, B. J. Genetics of plant-pathogen interactions specifying plant disease resistance. Plant Physiology, v.125, p.73-76. 2001.
- Steiger, D. L., Nagai, C. Moore, P. H., Morden, C. W., Osgood, R. V., Ming, R. AFLP analysis of genetic diversity within and among *Coffea arabica* cultivars. Theoretical and Applied Genetics, v.105, p.209-215. 2002.
- Stein, L. Genome annotation: from sequence to biology. Nature Reviews Genetics, v.2, p.493-503. 2001.
- Sterck, L., Rombauts, S., Vandepoele, K., Rouzé, P., Peer, Y. V. How many genes are there in plants (. . . and why are they there)? Current Opinion in Plant Biology, v.10, p.199-203. 2007.
- Stokstad, E. Genomes highlight plant pathogens' powerful arsenal. Science, v.313, p.1217. 2006.
- Tai, T. H., Dahlbeck, D., Clark, E. T., Gajiwala, P., Pasion, R., Whalen, M. C., Stall, R. E., Staskawicz, B. J. Expression of the Bs2 pepper gene confers resistance to bacterial spot disease in tomato. Proceedings of the National Academy of Sciences of the United States of America, v.96, n.4, p.14153-14158. 1999.
- Tateno, Y., Imanishi, T., Miyazaki, S., Fukami-Kobayashi, K., Saitou, N., Sugawara, H., Gojobori, T. DNA Data Bank of Japan (DDJB) for genome-

- scale research in life science. Nucleic Acids Research, v.30, n.1, p.27-30. 2002.
- Tedesco, N. S., Sakiyama, N. S., Zambolim, L., Teixeira, T. A., Pereira, A. A. Marcadores OPF-151280 e OPC-091120 ligados ao gene de resistência à ferrugem do cafeeiro (Hemileia vastatrix Berk. et Br.). I Simpósio de Pesquisa dos Cafés do Brasil. Poços de Caldas, 2000. 169-171 p.
- Temesgem, B., Brown, G. R., Harry, D. E., Kinlaw, C. S., Sewell, M. M., Neal, D.B. Genetic mapping of expressed sequence tag polymorphism (ESTP) markers in loblolly pine (*Pinus taeda* L.). Theoretical and Applied Genetics, v.102, p.664-675. 2001.
- The-Arabidopsis-Genome-Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature, v.408, p.796-815. 2000.
- Thomas, C. M., Jones, D. A., Parniske, M., Harrison, K., Balint-Kurti, P. J., Hatzixanthis, K., Jones, J. D. G. Characterization of the tomato Cf-4 gene for resistance to *Cladosporium fulvum* identifies sequences that determine recognition specificity in Cf-4 and Cf-9. Plant Cell, v.9, n.12, p.2209-2224. 1997.
- Torto, T. A., Li, S., Styer, A., Huitema, E., Testa, A., Gow, N. A. R., West, P. V., Kamoun, S. EST mining and functional expression assays identify extracellular effector proteins from the plant pathogen *Phytophthora*. Genome Research, v.13, p.1675-1685. 2003.
- Toyoda, K., Collins, N. C., Takahashi, A., Shirasu, K. Resistance and susceptibility of plants to fungal pathogens. Transgenic Research, v.11, p.567-582. 2002.
- Traut, T. W. The functions and consensus motifs of nine types of peptide segments that form different types of nucleotide-binding sites. European Journal of Biochemistry, v.229, p.9-19. 1994.
- Tsumura, Y., Suyama, Y., Yoshimura, K., Shirato, N., Mukai, Y. Sequence-tagged-sites (STSs) of cDNA clones in *Cryptomeria japonica* and their evaluation as molecular markers in conifers. Theoretical and Applied Genetics, v.94, p.764-772. 1997.
- Udall, J. A., Swanson, J. M., Haller, K., Rapp, R. A., Sparks, M. E., Hatfield, J., Yu, Y., Wu, Y., Dowd, C., Arpat, A. B., Sickler, B. A., Wilkins, T. A., Guo, J. Y., Chen, X. Y., Scheffler, J., Taliencio, E., Turley, R., Mcfadden, H., Payton, P., Klueva, N., Allen, R., Zhang, D., Haigler, C., Wilkerson, C., Suo, J., Schulze, S. R., Pierce, M. L., Essenberg, M., Kim, H., Llewellyn, D. J., Dennis, E. S., Kudrna, D., Wing, R., Paterson, A. H., Soderlund, C., Wendel, J. F. A global assembly of cotton ESTs. Genome Research, v.16, p.441-450. 2006.
- Udall, J. A., Swanson, J. M., Haller, K., Rapp, R. A., Sparks, M. E., Hatfield, J.,

- Yu, Y., Wu, Y., Dowd, C., Arpat, A. B., Sickler, B. A., Wilkins, T. A., Guo, J. Y., Chen, X. Y., Scheffler, J., Taliercio, E., Turley, R., Mcfadden, H., Payton, P., Klueva, N., Allen, R., Zhang, D., Haigler, C., Wilkerson, C., Suo, J., Schulze, S. R., Pierce, M. L., Essenberg, M., Kim, H., Llewellyn, D. J., Dennis, E. S., Kudrna, D., Wing, R., Paterson, A. H., Soderlund, C., Wendel, J. F. A global assembly of cotton ESTs. Genome Research, v.16, p.441-450. 20069.
- Van Der Biezen, E. A., Jones, J. D. The NB-ARC domain: a novel signaling motif shared by plant resistance gene products and regulators of cell death in animals. Current Biology, v.8, n.7, Mar, p.R226-227. 1998.
- Van Der Biezen, E. A., Sun, J., Coleman, M. J., Bibb, M. J., Jones, J. D. G. Arabidopsis RelA/SpoT homologs implicate (p)ppGpp in plant signaling Proceedings of the National Academy of Sciences of the United States of America, v.97, n.7, p.3747-3752. 2000.
- Van Der Linden, C. G., Wouters, D. C. A. E., Mihalka, V., Kochieva, E. Z., Smulders, M. J. M., Vosman, B. Efficient targeting of plant disease resistance loci using NBS profiling. Theoretical and Applied Genetics, v.109, p.384-393. 2004.
- Várzea, V. M. P., Marques, D. V. Population variability of *Hemileia vastatrix* vs. coffee durable resistance. In: L. Zambolim, Zambolim, E. M., Várzea, V. M. P. (Ed.). Durable resistance to coffee leaf rust. Viçosa : UFV, DFP, 2005. Population variability of *Hemileia vastatrix* vs. coffee durable resistance, p.53-74
- Vettore, A. L., Silva, F. R., Kemper, E. L., Souza, G. M., Silva, A. M., Ferro, M. I. T., Henrique-Silva, F., Giglioti, E. A., Lemos, M. V. F., Coutinho, L. L., Nobrega, M. P., Carrer, H., França, S. C., Bacci-Jr, M., Goldman, M. H. S., Gomes, S. L., Nunes, L. R., Camargo, L. E. A., Siqueira, W. J., Sluys, M. V., Thiemann, O. H., Kuramae, E. E., Santelli, R. V., Marino, C. L., Targon, M. L. P. N., Ferro, J. A., Silveira, H. C. S., Marini, D. C., Lemos, E. G. M., Monteiro-Vitorello, C. B., Tambor, J. H. M., Carraro, D. M., Roberto, P. G., Martins, V. G., Goldman, G. H., Oliveira, R. C., Truffi, D., Colombo, C. A., Rossi, M., Araujo, P. G., Sculaccio, S. A., Angella, A., Lima, M. M. A., Rosa-Jr, V. E., Siviero, F., Coscrato, V. E., Machado, M. A., Grivet, L., Mauro, S. M. Z., Nobrega, F. G., Menck, C. F. M., Braga, M. D. V., Telles, G. P., Cara, F. A. A., Pedrosa, G., Meidanis, J., Arruda, P. Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. Genome Research, v.13, p.2725-2735. 2003.
- Vieira, L. G., Andrade, A. C., Colombo, C.A., Pereira, G. A. Coffee genome project: a resource for functional genomics. In: L. Zambolim, Zambolim, E. M., Várzea, V. M. P. (Ed.). Durable resistance to coffee leaf rust. Viçosa : UFV, DFP, 2005. Coffee genome project: a resource for functional genomics, p.363-396
- Vieira, L. G. E., Andrade, A. C., Colombo, C. A., Moraes, A. H. A., Metha, A.,

- Oliveira, A. C., Labate, C. A., Marino, C. L. e C. B. Monteiro-Vitorello, Monte, D. C., Giglioti, E., Kimura, E. T., Romano, E., Kuramae, E. E., Lemos, E. G. M., Almeida, E. R. P., Jorge, E. C., Albuquerque, E. V. S., Silva, F. R., Vinecky, F., Sawazaki, H. E., Dorry, H. F. A., Carrer, H., Abreu, I. N., Batista, J. A. N., Teixeira, J. B., Kitajima, J. P., Xavier, K. G., Lima, L. M., Camargo, L. E. A., Pereira, L. F. P., Coutinho, L. H., Lemos, M. V. F., Romano, M. R., Machado, M. A., Costa, M. M. C., Sá, M. F. G., Goldman, M. H. S., Ferro, M. I. T., Tinoco, M. L. P., Oliveira, M. V., Sluys, M. V., Shimizu, M. M., Maluf, M. P., Eira, M. T. S., Filho, O. G., Arruda, P., Mazzafera, P., Mariani, P. D. S. C., Oliveira, R. L. B. C., Harakava, R., Balbao, S. F., Tsai, S. M., Mauro, S. M. Z., Santos, S. N., Siqueira, W. J., Costa, G. G. L., Formighieri, E. F., Carazzolle, M. F., Pereira, G. A. G. Brazilian coffee genome project: an EST-based genomic resource. Brazilian Journal of Plant Physiology, v.18, n.1, p.95-108. 2006.
- Vos, P., Hogers, R., Bleeker, M., Reijans, M., Van-De-Lee, T., Hornes, M., Friters, A., Pot, J., Paleman, J., Kuiper, M., Zabeau, M AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research, v.23, n.21, p.4407-4414. 1995.
- Wang, G., Ruan, D., Song, W., Sideris, S., Chen, L., Pi, L., Zhang, S., Zhang, Z., Fauquet, C., Gaut, B. S., Whalen, M. C., Ronald, P. C. Xa21D encodes a receptor-like molecule with a leucine-rich repeat domain that determines race-specific recognition and is subject to adaptive evolution Plant Cell, v.10, n.5, p.765-779. 1998.
- Wheeler, D. L., Smith-White, B., Chetvernin, V., Resenchuk, S., Dombrowski, S. M., Pechous, S. W., Tatusova, T., Ostell, J. Plant genome resources at the National Center for Biotechnology Information. Plant Physiology, v.138, p.1280-1288. 2005.
- Wheeler, D. L., Barrett, T., Benson, D. A., Bryant, S. H., Canese, K., Chetvernin, V., Church, D. M., Dicuccio, M., Edgar, R., Federhen, S., Geer, L. Y., Kapustin, Y., Khovayko, O., Landsman, D., Lipman, D. J., Madden, T. L., Maglott, D. R., Ostell, J., Miller, V., Pruitt, K. D., Schuler, G. D., Sequeira, E., Sherry, S. T., Sirotkin, K., Souvorov, A., Starchenko, G., Tatusov, R. L., Tatusova, T. A., Wagner, L., Yaschenko, E. . Database resources of the National Center for Biotechnology Information. Nucleic Acids Research, v.35, p.5-12. 2007.
- Williams, J. C. G. K., A. R.; Livak, K. J.; Rafalski, J. A., Tingey, S. V. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Research, v.18, p.6531-6535. 1990.
- Winslow, R. L., Boguski, M. S. Genome informatics: current status and future prospects. Circulation Research – Journal of The American Heart Association, v.92, p.953-961. 2003.
- Wolfsberg, T. G., Landsman, D. A comparison of expressed sequence tags (ESTs) to human genomic sequences. Nucleic Acids Research, v.25,

- p.1626-1632. 1997.
- Wolfsberg, T. G., Wetterstrand, K. A., Guyer, M. S., Collins, F. S., Baxeavanis, A. D. A user's guide to the human genome. Nature Genetics v.32, p.1-79. 2002.
- Yamamoto, K., Sasaki, T. L. Large-scale EST sequencing in rice. Plant Molecular Biology, v.35, p.135-144. 1997.
- Yang, W., Scheffler, B. E., Weston, L. A. Recent developments in primer design for DNA markers in higher plants. HortScience, v.41, n.4, p.1006. 2006.
- Young, N. D. The genetic architecture of resistance. Current Opinion in Plant Biology, v.3, n.4, p.285-290. 2000.
- Zambolim, L. Produção Integrada de Café. Viçosa: DFP/UFV. 2003. 4170 p.
- Zambolim, L., Zambolim, E. M., Várzea, V. M. P. Durable Resistance to Coffee Leaf Rust. Viçosa. 2005. 540 p.
- Zambolim, L., Zambolim, E. M., Do Vale, F. X. R., Pereira, A. A., Sakiyama, N. S., Caixeta, E. T. Physiological races of *Hemileia vastatrix* Berk. Et Br. In Brazil – Physiological variability, current situation and future prospects. In: L. Zambolim, Zambolim, E. M., Várzea, V. M. P. (Ed.). Durable resistance to coffee leaf rust. Viçosa : UFV, DFP, 2005. Physiological races of *Hemileia vastatrix* Berk. Et Br. In Brazil – Physiological variability, current situation and future prospects, p.75-98
- Zatz, M. M. Bioinformatics training in the USA. Briefings in Bioinformatics, v.3, n.4, p.353-360. 2002.
- Zhang, J., Zhang, L., Coombes, K. R. Gene sequence signatures revealed by mining the UniGene affiliation network. Bioinformatics, v.22, n.4, p.385-391. 2006.

CAPÍTULO 1

IDENTIFICAÇÃO DE SEQÜÊNCIAS EXPRESSAS DO GENOMA DO CAFÉ POTENCIALMENTE ASSOCIADAS COM A RESISTÊNCIA A DOENÇAS

1. INTRODUÇÃO

O seqüenciamento de ESTs (*Expressed Sequence Tags*) é um método rápido e apresenta boa relação de custo benefício para a obtenção de dados de interesse em genomas. Por esta razão, esse método tem se tornado um segmento crescente nas bases de dados públicas (Wolfsberg e Landsman, 1997). Em plantas, o seqüenciamento de ESTs foi inicialmente feito para as espécies modelo *arabidopsis* (Höfte et al., 1993) e arroz (Yamamoto e Sasaki, 1997). Posteriormente, grande quantidade de seqüência ESTs de outras espécies, incluindo milho (*Zea mays* L., Gai et al., 2000), soja [*Glycine max* (L.) Merr., Shoemaker et al., 2002], trigo (*Triticum aestivum* L., Lazo et al., 2001), batata (*Solanum tuberosum*, Ronning et al., 2003) e algodão (*Gossypium hirsutum*, Udal et al., 2006), além de várias outras foram depositadas no dbEST.

As informações geradas pelo seqüenciamento desses genomas têm facilitado e acelerado a identificação de genes responsáveis por características agrônômicas desejáveis, possibilitando a manipulação subsequente de genes de interesse por meio de técnicas de genética molecular.

A cultura do café também está sendo beneficiada pela revolução genômica. No Projeto Brasileiro do Genoma Café (PBGC) foram identificados cerca de 33.000 unigenes a partir de 214.964 ESTs de 37 bibliotecas de tecidos em diferentes estados fisiológicos de *Coffea arabica*, *C. canephora* e *C. racemosa*. Estas ESTs foram agrupadas, resultando em 17.982 *contigs* e 32.155 *singlets* (Vieira et al., 2006). Todos estes dados, além de várias ferramentas genéticas e genômicas que auxiliam os estudos e pesquisas, foram disponibilizados pelo Laboratório de Genômica e Expressão (LGE) e pela Embrapa Recursos Genéticos e Biotecnologia (Cenargen).

O objetivo principal de um trabalho de genômica é identificar genes responsáveis por características importantes. Os genes de resistência a doenças estão entre as classes de genes mais importantes para o melhoramento, sendo responsáveis por respostas imunes específicas incluindo o reconhecimento de patógenos e ativação de mecanismos de defesa.

Desta forma, este trabalho foi realizado com o objetivo de identificar, por meio da análise *in silico*, seqüências do banco de dados do PBGC

potencialmente envolvidas com a resistência do cafeeiro a doenças. Estas ESTs serão úteis para desenvolver novos e eficientes marcadores moleculares para assistirem os programas de melhoramento do cafeeiro.

2. MATERIAL E MÉTODOS

2.1. Base de dados

Para a mineração das seqüências potencialmente envolvidas com a resistência do cafeeiro a patógenos, analisou-se as ESTs do PBGC por meio da plataforma de bioinformática do LGE (<http://www.lge.ibi.unicamp.br/cafe/>).

2.2. Estratégias de mineração

Para o trabalho de mineração foram utilizadas três diferentes estratégias, aqui denominadas de *Electronic Northern*, Palavras-Chave e Seqüências Publicadas.

2.2.1. *Electronic Northern*

Esta estratégia foi baseada nos dados fornecidos pela ferramenta *Electronic Northern*. Todas as 214.964 ESTs provenientes das 37 bibliotecas de cDNA (Tabela 1.1) foram agrupadas e os *clusters* (*contigs* e *singlets*) formados. A porcentagem normalizada de ESTs das bibliotecas que constituem cada um dos *contigs* foi descrita. Nesta estratégia, foram selecionados para análise apenas os *clusters* constituídos por pelo menos uma EST da biblioteca RM1 (biblioteca de folhas de *Coffea arabica* infectadas com bicho mineiro e ferrugem), pois esta biblioteca é uma fonte potencial de seqüências envolvidas no processo de resistência da planta contra doenças.

Tabela 1.1: Bibliotecas do Projeto Brasileiro do Genoma Café com os respectivos números de *reads* aceitos.

<i>Coffea arabica</i>		
Biblioteca	Descrição	Reads aceitos
AR1	Folhas com tratamento de ácido araquidônico	2.364
BP1	Bion positive - plântulas inteiras + células	6.218
CA1	Calo	4.278
CB1	Células em suspensão com bion e brassinosteróides	8.237
CL2	Calo com e sem bion - curta	9.294
CS1	Células em suspensão com sais	7.804
EA1	Calos embriogênicos	4.847
EB1	Embrião de frutos	138
EM1	Embrião de sementes germinando	139
FB1	Botão floral estágio 1 e botão floral estágio 2 - longa	7.537
FB2	Botão floral estágio 1 e botão floral estágio 2 - curta	5.429
FB4	Botão floral estágio 3 e botão floral estágio 4 - curta	2.705
FR1	Botão floral nº 6, chumbinho nº 1, frutos estádios 1 e 2 - longa	3.369
FR2	Botão floral nº 6, chumbinho nº 1, frutos estádios 1 e 2 - curta	5.474
IA1	Calos embriogênicos	
IA2	Linhagem embriogênica (folha) com indução de 2,4D	2.052
IC1	Linhagem não embriogênica (folha) sem indução de 2,4D	2.227
LP1	Plântulas com tratamento de ácido araquidônico	2.301
LV4	Folhas de ponteiro ortotrópicas sem bion - longa	4.742
LV5	Folhas de ponteiro ortotrópicas sem bion - curta	6.724
LV8	Folhas plagiotrópicas de plantas adultas sem bion - longa	7.320
LV9	Folhas plagiotrópicas de plantas adultas sem bion - curta	3.017
NS1	Raízes com nematóide	321
PA1	Linhagem embriogênica (calos primários)	1.761
PC1	Linhagem não embriogênica (folhas) com indução de 2,4D	2.204
RM1	Folhas com ferrugem e bicho mineiro	3.512
RT3	Raiz sem bion - longa	356
RT5	Raiz com bion - longa	1.507
RT8	Raiz e células em suspensão na presença de alumínio - curta	5.074
RX1	Ramos infectados com <i>Xylella</i>	7.117
SH2	Estresse hídrico no campo (pool de tecidos)	5.185
SI3	Sementes germinando (inteira)	7.493
SS1	Condições normais - irrigado (pool de tecidos)	702
<i>Coffea canephora</i>		
Biblioteca	Descrição	Reads aceitos
EC1	Calos embriogênicos	6.632
SH1	Folhas - estresse hídrico - curta	89
<i>Coffea racemosa</i>		
Biblioteca	Descrição	Reads aceitos
FR4	Fruto	5.436
FV2	Fruto verde - estágio 1, 2 e 3 - curta	5.080

A ferramenta *Electronic Northern* está disponível na plataforma de bioinformática do LGE. No lado esquerdo da página da plataforma estão listadas as ferramentas e *links* de serviços. Para acessar a ferramenta *Electronic Northern*, basta clicar sobre o link com esse nome. Uma nova página

se abre, onde o usuário deve escolher as espécies com as quais deseja trabalhar. As opções são: CA - *Coffee Arabica*, CC – *Coffee Canephora*, CR – *Coffee Racemosa* e *All Species*.

Assim que é feita a escolha, é necessário fornecer o *login* e a senha do pesquisador responsável. Na página da ferramenta é possível organizar as seqüências (*Singlets* ou *Contigs*) pelo nome, por meio da ordem alfabética; por biblioteca, por meio da percentagem normalizada decrescente de ESTs da biblioteca escolhida, presente em cada *contig*; por gene, por meio da ordem crescente do número de identificação do gene (gi) ao qual o *contig* está relacionado; e pelo componente celular, processo biológico ou função molecular, por meio da ordem decrescente da numeração do GO.

2.2.2. Palavras-Chave

Realizou-se uma busca na literatura científica por termos conhecidamente relacionadas com o mecanismo de resistência a doenças. Estas palavras-chave foram utilizadas como “iscas” para mineração dos dados, e as seqüências obtidas foram depositadas no sistema de gerenciamento e manipulações de seqüências, o *Gene Projects*. Cada projeto recebe o nome de uma palavra-chave e armazena as ESTs a ela relacionadas.

Os projetos criados foram: LRR, NBS, *Resistance*, *Chitinase*, *Cytochrome P450*, *Glucanase*, HSP (*Heat Shock Protein*), *Thaumatococcus*, *Chalconesynthase*, *Hypersensitive*, *Pathogenesis*, *Polyphenoloxidase*, *Importin*, *Glucosyltransferase* e *Phytoalexin*. As ESTs de cada projeto foram agrupadas formando os *contigs*, que foram comparados com seqüências do GenBank pelo programa BLAST.

O sistema *Gene Projects* pode ser acessado clicando sobre o *link* presente no lado esquerdo da plataforma de bioinformática do LGE. Uma nova página se abre, onde o usuário deve escolher as espécies com as quais deseja trabalhar. As opções são: CA - *Coffee Arabica*, CC – *Coffee Canephora* e CR – *Coffee Racemosa*. Assim que é feita a escolha, é necessário fornecer o *login* e a senha do pesquisador responsável, e posteriormente, o *login* e a senha do usuário.

Na página de Inclusão e Manutenção de Projetos é possível criar e editar projetos. Para criar um projeto basta clicar em “criar novo projeto”. Deve-se dar um título e uma breve descrição para o projeto, além de fornecer o e-mail do usuário.

Para acrescentar seqüências ao projeto, deve-se selecionar o projeto com o qual o usuário deseja trabalhar, na página de Inclusão e Manutenção de Projetos. Na página do projeto em questão, clicar em “acrescentar novos reads”. É possível buscar as ESTs desejadas por meio de cinco estratégias: *Reads Search*, *Keyword Search*, *BLAST Search*, *All Reads* e *Pattern Search*. Para a estratégia de Palavras-Chave foi utilizada a busca por palavras-chave (*Keyword Search*). Após selecionar essa estratégia, basta digitar o termo de busca e clicar em “*find now*”. É possível também fazer uma pesquisa booleana, utilizando os operadores AND, OR e NOT, em caso de busca por mais de um termo simultaneamente.

A página de respostas apresentará todas as seqüências que contenham o termo de busca na anotação do BLAST. Nessa página pode-se selecionar as seqüências uma a uma, clicando sobre os *boxes* laterais, ou pode-se selecionar todas de uma só vez, clicando na opção “selecionar todas” na parte inferior da página. Após a seleção deve-se enviar as ESTs selecionadas para a página de busca, clicando em “enviar reads”.

Depois de enviar as ESTs selecionadas para a página de busca, é necessário exportá-las para o projeto em questão. A seleção das ESTs pode ser feita uma a uma, clicando sobre os *boxes* laterais, ou clicando na opção “select all”. Posteriormente, deve-se clicar em “exportar reads selecionados”. Após a exportação das ESTs para o projeto, basta clicar em “voltar” para acessar a página do projeto onde estarão armazenadas as ESTs que foram exportadas.

Na página do projeto é possível agrupar as ESTs, afim de obter seqüências consenso, maiores e mais representativas. Para isso, deve-se selecionar as seqüências a serem agrupadas, uma a uma ou todas, clicando em “selecionar todos” e depois clicar em “clusterizar”. Para verificar o resultado do agrupamento, basta clicar em “visualizar clusterização”.

Na página dos *clusters* pode-se visualizar os *contigs* e os *singlets* formados pelo agrupamento. Clicando em “view” o usuário pode visualizar a

montagem do *contig* formado. O *hyperlink* da coluna “reads” permite a verificação das ESTs que constituem cada *contig*. É possível organizar as seqüências da página de *clusters* pelo nome, *score*, *e-value*, percentagem de erro ou comprimento total da seqüência. É possível ainda, fazer uma busca de similaridade para os *contigs* formados, clicando em “iniciar blastagem”.

2.2.3. Seqüências Publicadas

Nesta estratégia foi realizada uma busca por similaridade entre seqüências publicadas que codificam genes de cafeeiros relacionados à resistência a patógenos (Chen et al., 2003; Fernandez et al., 2004; Lin et al., 2005; Ganesh et al., 2006) com as do PBGC. As ESTs encontradas no PBGC foram, posteriormente, submetidas a uma comparação com o banco de dados NR (*non redundant*) do NCBI, por meio do programa BLAST, para verificar a sua anotação.

A estratégia Seqüências Publicadas foi realizada utilizando a ferramenta “Search”, disponível na plataforma de bioinformática do LGE. O *link* para a ferramenta está presente no lado esquerdo da plataforma. As opções dessa ferramenta são: BLAST Local, *Keyword Search in BLAST Results* e *All BLAST Results*. Para a estratégia Seqüências Publicadas foi utilizada a opção BLAST Local. Depois de escolher a opção é necessário fornecer o *login* e a senha do pesquisador responsável para acessar a ferramenta.

Na página do BLAST é possível escolher o tipo de programa a ser utilizado: *blastn*, *blastp*, *blastx*, *tblastn* e *tblastx*. É possível também escolher a base de dados do PBGC com a qual a seqüência publicada será comparada.

Para fazer a busca, basta colar a seqüência de interesse no campo de busca e, após escolher o programa e a base de dados, clicar em “search”. A página de resposta é bem similar à do BLAST do NCBI, apresentando um resultado gráfico dos alinhamentos e mostrando as seqüências da base de dados do PBGC que apresentam similaridade com a seqüência de interesse, com seus respectivos valores de *score* e *e-value*.

3. RESULTADOS E DISCUSSÃO

3.1 *Electronic Northern*

Após o agrupamento de todas as seqüências do genoma, foram formados 2.206 *clusters* que continham pelo menos uma ESTs da biblioteca RM1. Deste total, 602 foram formados apenas por ESTs da biblioteca RM1(100% RM1), 21 por seqüências com 99% a 75% e 213 de 74% a 50% das ESTs provenientes da RM1 (Tabela 1.2).

Tabela 1.2: Número de *clusters* formados por ESTs da Biblioteca RM1

Estratégia <i>Electronic Northern</i>	Número de seqüências	% em relação a RM1
Seqüências com ESTs da RM1	2.206	100
Seqüências 50% a 74% de ESTs da RM1	213	9,65
Seqüências 75% a 99% de ESTs da RM1	21	0,95
Seqüências 100% dos ESTs da RM1	602	27,29

Como essas seqüências são de genes expressos no momento de infecção da planta por bicho mineiro e ferrugem, existe a possibilidade de que elas tenham uma função no processo de infecção e ao mecanismo de defesa da planta contra esses estresses bióticos. De fato, vários *clusters* formados apenas por ESTs da biblioteca RM1 apresentaram similaridade com proteínas reconhecidamente relacionadas com esse processo como, por exemplo, LRR, proteína quinase, citocromo P450, proteína de susceptibilidade e apoptose celular (importina - alfa), proteína putativa de ligação ao DNA, entre outras (Tabela 1.3).

Tabela 1.3: Anotação dos *clusters* formados apenas por ESTs da biblioteca RM1

Anotação	Nº <i>Clusters</i>	
<i>No hits found</i>	135	
Proteínas não-conhecidas	{ Proteína Expressa	48
	{ Proteína Hipotética	113
	{ Proteína Desconhecida	21
Proteína Quinase	8	
Transportador ABC (Cassete de ligação a ATP)	6	
Proteína LRR	4	
Citocromo P450	3	
Outros	264	
Total	602	

Das 602 seqüências formadas exclusivamente por ESTs da biblioteca RM1, 317 (52,65%) não têm similaridade com nenhum gene presente no banco de dados (Figura 1.1). Essas seqüências podem ser únicas de cafeeiro, ou de algum gene que ainda não foi descrito. São, portanto dados novos e devem ser analisados minuciosamente.

Os *clusters* formados por 50 a 99% de ESTs da biblioteca RM1 também exibiram anotação de proteínas envolvidas no processo de defesa da planta contra doenças. Alguns exemplos foram proteína dedo de zinco, proteína de resistência a doença, proteína quinase putativa, proteína de membrana, chalcone redutase, entre outras.

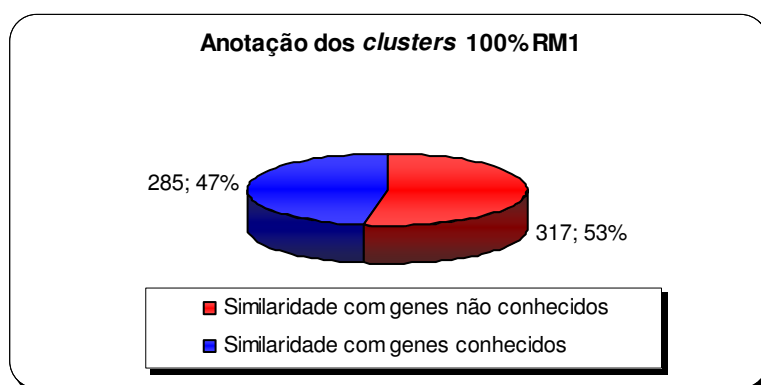


Figura 1.1. Anotação dos *clusters* formados exclusivamente por ESTs da biblioteca RM1.

Os processos biológicos com os quais os *clusters* 100% RM1 estão envolvidos incluem metabolismo, transporte, regulação da transcrição, enovelamento de proteínas, biossíntese, entre outros (Tabela 1.4).

Tabela 1.4: Processos biológicos dos *clusters* formados exclusivamente por ESTs da biblioteca RM1

Processo Biológico	Nº Clusters
Não definido	281
Desconhecido	51
Transporte	105
Metabolismo	44
Biossíntese	32
Fosforilação de proteínas e aminácidos	20
Regulação da transcrição, DNA-dependente	17
Proteólise e Peptidólise	14
Enovelamento de Proteína	7
Outros	31
Total	602

Mais da metade dos *clusters* 100% RM1 (332; 55,15%) possuem processo biológico desconhecido (Figura 1.2).

As funções moleculares apresentadas pelos *clusters* 100% RM1 incluíram atividade catalítica, ligação a ácidos nucleicos, ligação a nucleotídeos entre outros (Tabela 1.5). A função de ligação a nucleotídeos pode estar relacionada com o mecanismo de resistência da planta, uma vez que o domínio NBS de proteínas de resistência a doenças é encontrado em uma grande variedade de proteínas com a função de ligação a ATP ou GTP em diversos organismos (Traut, 1994 e Saraste et al., 1990).

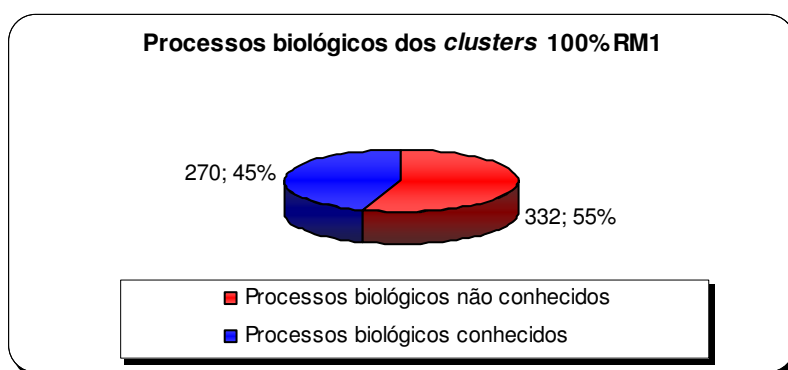


Figura 1.2. Processos biológicos dos *clusters* formados exclusivamente por ESTs da biblioteca RM1.

Tabela 1.5: Funções moleculares dos *clusters* formados exclusivamente por ESTs da biblioteca RM1

Função Molecular	Nº Clusters
Não definida	259
Desconhecida	23
Atividade Catalítica	29
Ligação a DNA	20
Ligação a Ácidos Nucleicos	13
Ligação a Nucleotídeos	12
Ligação a RNA	4
Outras	242
Total	602

De modo semelhante à anotação e ao processo biológico, aproximadamente metade (282; 46,84%) dos *clusters* 100% RM1 apresentaram função molecular desconhecida (Figura 1.3).

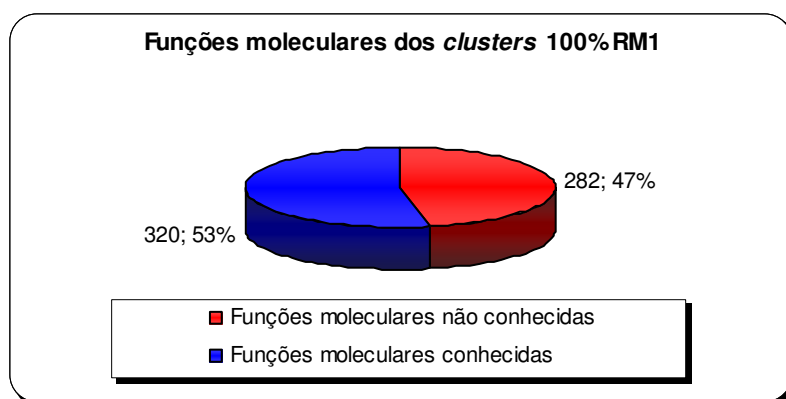


Figura 1.3. Funções moleculares dos *clusters* formados exclusivamente por ESTs da biblioteca RM1.

A maior parte dos *clusters* 100% RM1 que possuem componente celular definido está relacionada com a membrana (66 *clusters*). Foram encontrados também *clusters* relacionados com o cloroplasto, núcleo e citoplasma, entre outros (Tabela 1.6).

Tabela 1.6: Componentes celulares dos *clusters* formados exclusivamente por ESTs da biblioteca RM1

Componente celular	Nº <i>Clusters</i>
Não definido	393
Desconhecido	30
Membrana	66
Cloroplasto	36
Citoplasma	10
Núcleo	10
Mitocôndria	8
Outros	49
Total	602

Com relação ao organismo homólogo, foram encontrados *clusters* 100% RM1 similares a seqüências de *Arabidopsis thaliana*, *Oryza sativa*, *Nicotiana* sp, *Solanum tuberosum*, *Lycopersicon esculentum*, entre outras espécies (Tabela 1.7). Esta análise mostrou ainda que existem *clusters* similares a seqüências de bactérias (*Pseudomonas* sp. e *Bordetella* sp.) e até mesmo à mosca das frutas (*Drosophila melanogaster*). Esses resultados indicam que pode ter ocorrido contaminação. Uma outra hipótese é a existência de genes conservados entre essas espécies.

Tabela 1.7: Espécies que apresentaram seqüências similares aos *clusters* formados exclusivamente por ESTs da biblioteca RM1

Organismo homólogo	Nº Clusters
<i>Arabidopsis thaliana</i>	177
<i>Oryza sativa</i>	39
<i>Pseudomonas</i> sp.	43
<i>Nicotiana</i> sp.	9
<i>Bordetella</i> sp.	6
<i>Lycopersicon esculentum</i>	3
<i>Solanum tuberosum</i>	3
<i>Drosophila melanogaster</i>	2
Outros	320
Total	602

3.2. Palavras-chave

Foram mineradas um total de 11.793 ESTs nos 15 projetos criados (Figura 1.4). Os projetos *CytochromeP450*, *Resistance* e *Chitinase* foram os que tiveram o maior número de ESTs, com 2.441 (20,70%), 1.864 (15,81%) e 1.855 (15,73%) respectivamente (Tabela 1.8). Os projetos menos abundantes em ESTs foram *Phytoalexin*, *Polyphenoloxidase* e *Hypersensitive*, com 12 (0,10%), 67 (0,57%) e 86 (0,73%) respectivamente.

Após o agrupamento das ESTs presentes em cada um destes projetos (Tabela 1.8), os *contigs* formados foram submetidos a uma análise comparativa, por meio do BLAST. A anotação de cada *contig* foi analisada, visando a obtenção do maior número de informações relevantes sobre os prováveis genes de resistência a doenças.

Com a palavra-chave LRR (*Leucine Rich Repeat* – Repetições Ricas em Leucina) foi possível minerar 825 ESTs.

O domínio LRR possui, entre outras funções, a de facilitar a interação da proteína R com o seu fator Avr (elicitador) relacionado podendo fornecer diferentes especificidades de reconhecimento para fatores Avr alterados (Parniske et al., 1997). De fato, em vários sistemas planta-patógeno, a variação da seqüência na região LRR, mostrou ser responsável por diferentes especificidades de reconhecimento ou de resistência (Parniske et al., 1997; Thomas et al., 1997; Botella et al., 1998; Wang et al., 1998; Ellis et al., 1999).

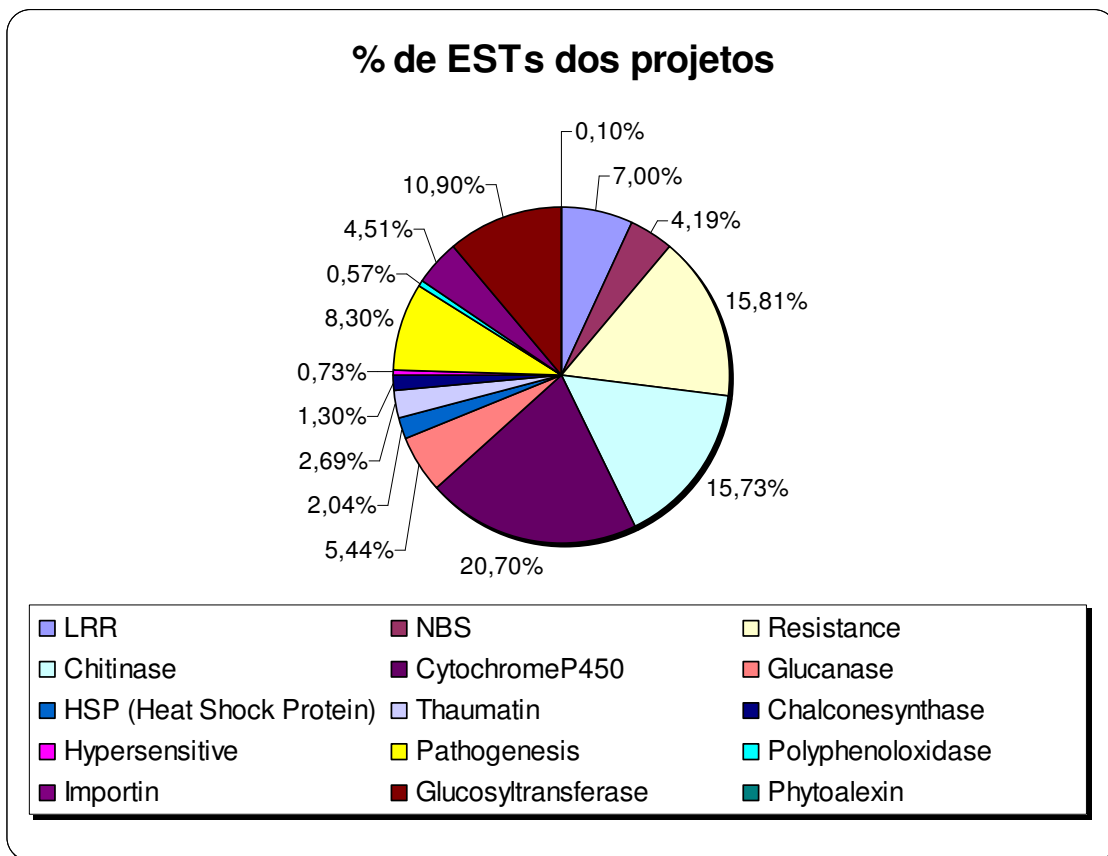


Figura 1.4: Porcentagem relativa do número de ESTs dos 15 projetos criados pela estratégia Palavras-Chave.

Tabela 1.8: Número de ESTs de cada um dos 15 projetos criados a partir de palavras-chave e número de *contigs* e *singlets* formados como resultado do agrupamento.

Projeto	ESTs	Contigs	Singlets
Chalconesynthase	153	5	8
Chitinase	1.855	47	48
CytochromeP450	2.441	235	202
Glucanase	642	92	68
Glucosyltransferase	1.286	130	160
HSP (Heat Shock Protein)	240	31	27
Hypersensitive	86	8	4
Importin	532	11	14
LRR	825	160	243
NBS	494	105	142
Pathogenesis	979	63	37
Polyphenoloxidase	67	4	3
Phytoalexin	12	3	2
Resistance	1864	347	416
Thaumatococcus	317	16	7
TOTAL	11.793	1.257	1.381

Foram encontradas ESTs com similaridade a proteínas de resistência a doenças, proteínas NBS-LRR, proteínas quinases, dentre outras (Tabela 1.9).

Tabela 1.9: ESTs encontradas no projeto LRR

Anotação	Nº ESTs
Proteína de Resistência a Doenças	249
Proteína NBS-LRR	120
Proteína LRR	116
Proteína de Resistência	62
Receptor LRR de Proteína Quinase	62
Rad-51 (Fifty one) like, Short RFS-1, f-box protein Fbl2 (51.6 kD) (rfs-1Co)	20
Coronatina-insensível 1 (COI1), AtFBL2	19
Proteína Hipotética F6H11.60	12
Proteína EILP receptora de LRR Elicitora-indutível	11
Proteína F-box e LRR	10
Outras	144
Total	825

As ESTs encontradas nesse projeto tiveram similaridade com genes de espécies vegetais como *Arabidopsis thaliana*, *Oryza sativa*, *Nicotiana tabacum*, *Glycine max* e *Phaseolus vulgaris*, e também animais como *Rattus norvegicus* e *Homo sapiens* (Tabela 1.10).

A análise dos *contigs* revelou a presença de genes de resistência a doenças com domínios conservados pfam00931NB-ARC (*Nucleotide Binding in APAF-1, R gene products, e CED-4*) e COG4886 LRR. Dos 160 *contigs* formados, 58 apresentaram *e-value* menor que 1^{-10} . As informações destes *contigs* estão apresentadas na Tabela 1.11.

O *contig* 158, anotado como *resistance protein* KR4, *Glycine max*, foi o de menor tamanho, com 476 pb. O *contig* 49, anotado como *putative NBS-LRR type disease resistance protein*, de *Prunus persica*, foi o de tamanho maior, com 3.281 pb.

Quatro *contigs* foram formados por pelo menos uma EST da biblioteca RM1 (48, 95, 117 e 143). O *contig* 48 foi anotado como *Disease resistance protein (CC-NBS-LRR class), putative*, de *Arabidopsis thaliana*, apresentou 2.255 pb e foi formado por 13 ESTs, sendo que apenas uma delas é da biblioteca RM1. O *contig* 95, anotado como *Disease resistance-like protein*, de *Coffea arabica*, com 778 pb, foi formado por duas ESTs da biblioteca RM1.

Tabela 1.10: Espécies que apresentaram seqüências similares às ESTs encontradas no projeto LRR

Organismo	Nº ESTs
<i>Arabidopsis thaliana</i>	287
<i>Solanum</i> sp.	120
<i>Oryza sativa</i>	102
<i>Hordeum vulgare</i>	37
<i>Nicotiana tabacum</i>	29
<i>Rattus norvegicus</i>	19
<i>Phaseolus vulgaris</i>	18
<i>Glycine max</i>	17
<i>Manihot esculenta</i>	15
<i>Homo sapiens</i>	10
Outros	171
Total	825

O *contig* 117 foi anotado como *putative disease related protein 2*, de *Oryza sativa*, com 1.886 pb e formado por quatro ESTs, sendo uma da biblioteca RM1.

O *contig* 143, anotado como *NBS-LRR disease resistance protein homolog*, de *Hordeum vulgare*, com 841 pb, foi formado por três ESTs, sendo uma da biblioteca RM1.

Usando a palavra-chave NBS (*Nucleotide Binding Site* – Sítio de Ligação a Nucleotídeo) foram encontradas 494 ESTs do PBGC. Dentre elas, destacam-se as anotações: proteínas de resistência, proteínas NBS-LRR e proteínas de resistência a doenças (Tabela 1.12), que corresponderam a 84,81% do número total de ESTs presentes nesse projeto.

Acredita-se que o domínio NBS participa da ativação dos componentes da transdução de sinais, levando a respostas de resistência patógeno-específicas (Aarts et al., 1998; Feys e Parker, 2000; Van Der Biezen et al., 2000).

As ESTs encontradas nesse projeto apresentaram similaridade com genes de *Arabidopsis thaliana*, *Oryza sativa*, *Glycine max*, *Solanum* sp. e *Phaseolus vulgaris*, entre outras (Tabela 1.13).

Tabela 1.11: Relação de *Contigs* com *e-value* menor que 1^{-10} formados pelo agrupamento do projeto LRR e o respectivo número de ESTs de cada biblioteca.

Contig	Definição	Organismo	Domínio 1	Domínio 2	# pb	# ESTs	C	L	L	L	L	C	F	F	F	F	F	R	B	R	C	C	F	E	E	I	P	I	R	P	S	S	S		
							A	V	V	V	V	L	R	R	B	B	B	B	T	P	M	B	S	R	A	C	C	C	A	X	A	I	H	S	
4	very similar to disease resistance proteins	<i>Arabidopsis thaliana</i>	pfam00931NB-ARC domain		985	2				1														1											
6	NBS/LRR disease resistance protein RPM1	<i>Arabidopsis lyrata</i>	pfam00931NB-ARC domain	COG4886:Lrr protein	1121	2																													
8	similar to NBS-LRR type resistance gene	<i>Oryza sativa</i>	pfam00931NB-ARC domain	COG4886:Lrr protein	946	2										1																			
9	disease resistance protein (CC-NBS-LRR class), putative	<i>Arabidopsis thaliana</i>	pfam00931NB-ARC domain	pfam01576:Myosin tail	953	2	1					1																							
11	putative resistance complex protein I2C-2	<i>Solanum demissum</i>	pfam00931NB-ARC domain		762	2																													
12	putative late blight resistance protein	<i>Solanum demissum</i>	pfam00931NB-ARC domain		858	2							2																						
13	putative RGH1A	<i>Oryza sativa</i>	pfam00931NB-ARC domain		782	2										2																			
21	fom-2 protein	<i>Cucumis melo</i>			1579	9				4									1															4	
23	potato resistance-like protein I2GA-SH23-1	<i>Solanum tuberosum</i>	pfam00931NB-ARC domain	COG4886:Lrr protein	996	2	1																												
24	fom-2 protein	<i>Cucumis melo</i>	pfam00931NB-ARC domain		1852	3							1			1																		1	
25	disease resistance protein	<i>Arabidopsis thaliana</i>	pfam00931NB-ARC domain		867	4							4																						
27	disease resistance protein	<i>Arabidopsis thaliana</i>	pfam00931NB-ARC domain	COG4886:Lrr protein	912	4				2					2																				
30	putative leucine-rich repeat disease resistance protein	<i>Arabidopsis thaliana</i>	pfam00931NB-ARC domain		1335	7				4			1	1						1															
33	Disease resistance protein BS2	<i>Capsicum chacoense</i>	pfam00931NB-ARC domain		2540	15	1		4										2														1		
35	putative disease resistance protein RPH8A	<i>Oryza sativa</i>	pfam00931NB-ARC domain	COG4886:Lrr protein	1857	7		1	1						1	4																			
37	disease resistance protein RPH8A (RPP8 homolog A)	<i>Arabidopsis thaliana</i>	pfam00931NB-ARC domain		2164	5					1		1							1							1						1		
40	blight resistance protein RGA1	<i>Solanum bulbocastanum</i>	pfam00931NB-ARC domain	COG4886:Lrr protein	1540	2	1																											1	
43	disease resistance protein, putative	<i>Arabidopsis thaliana</i>	pfam00931NB-ARC domain		561	2	1																											1	
44	putative disease resistance protein RPH8A	<i>Oryza sativa</i>	pfam00931NB-ARC domain	COG4886:Lrr protein	1981	6		2		3							1																		
46	putative disease resistance protein	<i>Arabidopsis thaliana</i>	pfam00931NB-ARC domain		1500	2				1	1																								
47	disease resistance-like protein	<i>Coffea canephora</i>	pfam00931NB-ARC domain		1025	3							2																				1		
48	disease resistance protein (CC-NBS-LRR class), putative	<i>Arabidopsis thaliana</i>	pfam00931NB-ARC domain		2255	13	1						1	4		1																		2	
49	putative NBS-LRR type disease resistance protein	<i>Prunus persica</i>	pfam00931NB-ARC domain		3281	15					4		1		1																			2	

*Bibliotecas: AR1 - Folhas com tratamento de ácido araquidônico. BP1 - Bion positive - plântulas inteiras + células. CA1 – Calo. CB1 - Células em suspensão com bion e brassinosteróides. CL2 - Calo com e sem bion – curta. CS1 - Células em suspensão com sais. EA1 - Calos embriogênicos. EB1 - Embrião de frutos. EM1 - Embrião de sementes germinando. FB1 - Botão floral estágio 1 e botão floral estágio 2 – longa. FB2 - Botão floral estágio 1 e botão floral estágio 2 – curta. FB4 - Botão floral estágio 3 e botão floral estágio 4 – curta. FR1 - Botão floral nº 6, chumbinho nº 1, frutos estádios 1 e 2 – longa. FR2 - Botão floral nº 6, chumbinho nº 1, frutos estádios 1 e 2 – curta. IA1 - Calos embriogênicos. IA2 - Linhagem embriogênica (folha) com indução de 2,4D. IC1 - Linhagem não embriogênica (folha) sem indução de 2,4D. LP1 - Plântulas com tratamento de ácido araquidônico. LV4 - Folhas de ponteiro ortotrópicas sem bion – longa. LV5 - Folhas de ponteiro ortotrópicas sem bion – curta. LV8 - Folhas plagiotrópicas de plantas adultas sem bion – longa. LV9 - Folhas plagiotrópicas de plantas adultas sem bion – curta. NS1 - Raízes com nematóide. PA1 - Linhagem embriogênica (calos primários). PC1 - Linhagem não embriogênica (folhas) com indução de 2,4D. RM1 - Folhas com ferrugem e bicho mineiro. RT3 - Raiz sem bion – longa. RT5 - Raiz com bion – longa. RT8 - Raiz e células em suspensão na presença de alumínio – curta. RX1 - Ramos infectados com *Xylella*. SH2 - Estresse hídrico no campo (pool de tecidos). SI3 - Sementes germinando (inteira). SS1 - Condições normais - irrigado (pool de tecidos). EC1 - Calos embriogênicos (*C. canephora*). SH1 - Folhas - estresse hídrico – curta (*C. canephora*). FR4 – Fruto (*C. racemosa*). FV2 - Fruto verde - estágio 1, 2 e 3 – curta (*C. racemosa*).

Tabela 1.11 continuação

Contig	Definição	Organismo	Domínio 1	Domínio 2	# pb	# ESTs	C A	L V	L V	L V	L V	C L	F R	F R	F B	F B	F B	F B	R T	B P	R M	C B	C S	F R	E A	E C	I C	P C	I A	R X	P A	S I	S H	S S					
53	blight resistance protein T118	<i>Solanum tarijense</i>	pfam00931NB-ARC domain	COG4886:Lrr protein	768	4																																	
54	putative disease resistance gene homolog	<i>Oryza sativa</i>	pfam00931NB-ARC domain	COG4886:Lrr protein	1619	5											1					2																	
60	putative NBS-LRR type resistance protein	<i>Oryza sativa</i>	pfam00931NB-ARC domain	COG4886:Lrr protein	900	2																		1															
62	disease resistance gene homolog 9N	<i>Brassica napus</i>	pfam00931NB-ARC domain	COG4886:Lrr protein	917	2																						1											
64	putative late blight resistance protein	<i>Solanum demissum</i>			693	2																																	
65	resistance protein KR4	<i>Glycine max</i>	pfam00931NB-ARC domain	COG4886:Lrr protein	759	2																																	
69	fom-2	<i>Cucumis melo</i>	pfam00931NB-ARC domain		1022	3	2						1																										
73	RGC1	<i>Solanum tuberosum</i>	pfam00931NB-ARC domain		1239	5	1							3										1															
76	prf	<i>Lycopersicon pimpinellifolium</i>	pfam00931NB-ARC domain		744	2																	2																
77	fom-2 protein	<i>Cucumis melo</i>			1384	2				1													1																
83	putative disease resistance protein RPH8A	<i>Oryza sativa</i>	pfam00931NB-ARC domain	COG4886:Lrr protein	2376	4	3						1																										
86	putative disease resistance protein Prf	<i>Solanum demissum</i>	pfam00931NB-ARC domain		1571	7								1	4		2																						
87	NBS/LRR disease resistance protein RPM1	<i>Arabidopsis lyrata</i>	pfam00931NB-ARC domain	COG4886:Lrr protein	694	2			2																														
95	disease resistance-like protein	<i>Coffea arabica</i>	pfam00931NB-ARC domain		778	2																2																	
96	resistance protein Tsu5	<i>Arabidopsis thaliana</i>	pfam00931NB-ARC domain		931	2	1																																
97	disease resistance-like protein	<i>Coffea arabica</i>	pfam00931NB-ARC domain		1047	2	2																																
98	NBS-LRR protein	<i>Solanum tuberosum</i>	pfam00931NB-ARC domain		593	2						2																											
101	disease resistance protein I2	<i>Lycopersicon esculentum</i>	pfam00931NB-ARC domain		715	2																	2																
106	putative disease resistance protein	<i>Solanum demissum</i>	pfam00931NB-ARC domain		729	2					2																												
109	disease resistance-like protein	<i>Coffea arabica</i>	pfam00931NB-ARC domain		1131	5	3			1															1														
112	putative NBS-LRR resistance protein	<i>Oryza sativa</i>	pfam00931NB-ARC domain		1256	3							1																										
117	putative disease related protein 2	<i>Oryza sativa</i>	pfam00931NB-ARC domain	COG4886:Lrr protein	1886	4										1						1	1																
123	putative disease resistance gene homolog	<i>Oryza sativa</i>	pfam00931NB-ARC domain	COG4886:Lrr protein	779	4			1							2									1														
124	disease resistance protein-like	<i>Arabidopsis thaliana</i>	pfam00931NB-ARC domain	COG4886:Lrr protein	936	2			1						1																								
126	putative late blight resistance protein	<i>Solanum demissum</i>	pfam00931NB-ARC domain		786	2				2																													

*Bibliotecas: AR1 - Folhas com tratamento de ácido araquidônico. BP1 - Bion positive - plântulas inteiras + células. CA1 – Calo. CB1 - Células em suspensão com bion e brassinosteróides. CL2 - Calo com e sem bion – curta. CS1 - Células em suspensão com sais. EA1 - Calos embriogênicos. EB1 - Embrião de frutos. EM1 - Embrião de sementes germinando. FB1 - Botão floral estádio 1 e botão floral estádio 2 – longa. FB2 - Botão floral estádio 1 e botão floral estádio 2 – curta. FB4 - Botão floral estádio 3 e botão floral estádio 4 – curta. FR1 - Botão floral nº 6, chumbinho nº 1, frutos estádios 1 e 2 – longa. FR2 - Botão floral nº 6, chumbinho nº 1, frutos estádios 1 e 2 – curta. IA1 - Calos embriogênicos. IA2 - Linhagem embriogênica (folha) com indução de 2,4D. IC1 - Linhagem não embriogênica (folha) sem indução de 2,4D. LP1 - Plântulas com tratamento de ácido araquidônico. LV4 - Folhas de ponteiro ortotrópicas sem bion – longa. LV5 - Folhas de ponteiro ortotrópicas sem bion – curta. LV8 - Folhas plagiotrópicas de plantas adultas sem bion – longa. LV9 - Folhas plagiotrópicas de plantas adultas sem bion – curta. NS1 - Raízes com nematóide. PA1 - Linhagem embriogênica (calos primários). PC1 - Linhagem não embriogênica (folhas) com indução de 2,4D. RM1 - Folhas com ferrugem e bicho mineiro. RT3 - Raiz sem bion – longa. RT5 - Raiz com bion – longa. RT8 - Raiz e células em suspensão na presença de alumínio – curta. RX1 - Ramos infectados com *Xylella*. SH2 - Estresse hídrico no campo (pool de tecidos). SI3 - Sementes germinando (inteira). SS1 - Condições normais - irrigado (pool de tecidos). EC1 - Calos embriogênicos (*C. canephora*). SH1 - Folhas - estresse hídrico – curta (*C. canephora*). FR4 – Fruto (*C. racemosa*). FV2 - Fruto verde - estádio 1, 2 e 3 – curta (*C. racemosa*).

Tabela 1.11 continuação

Contig	Definição	Organismo	Domínio 1	Domínio 2	# pb	# ESTs	C A 1	L V 4	L V 5	L V 8	L V 9	C L 2	F R 1	F R 2	F B 1	F B 2	F B 4	R T 8	B P 1	R M 1	C B 1	C S 1	F R 4	E A 1	E C 1	I C 1	P C 1	I A 2	R X 1	P A 1	S I 3	S H 2	S S 1	
135	putative NBS-LRR disease resistance protein	<i>Oryza sativa</i>	pfam00931NB-ARC domain		814	3												3																
136	late blight resistance protein-like	<i>Solanum tuberosum</i>	pfam00931NB-ARC domain		803	2												2																
137	disease resistance-like protein	<i>Coffea arabica</i>	pfam00931NB-ARC domain		1237	3	1												2															
138	blight resistance protein T118	<i>Solanum tarijense</i>	pfam00931NB-ARC domain	COG4886:Lrr protein	1124	2	2																											
139	Prf	<i>Lycopersicon esculentum</i>	pfam00931NB-ARC domain		1366	2									1														1					
143	NBS-LRR disease resistance protein homologue	<i>Hordeum vulgare</i>	pfam00931NB-ARC domain	COG4886:Lrr protein	841	3	1													1												1		
145	disease resistance protein (CC-NBS-LRR class), putative	<i>Arabidopsis thaliana</i>	pfam00931NB-ARC domain		1906	5					1		2										1									1		
147	Disease resistance protein BS2	<i>Capsicum chacoense</i>	pfam00931NB-ARC domain		2019	6													2				1									1		2
152	putative disease resistance protein	<i>Arabidopsis thaliana</i>	pfam00931NB-ARC domain		1580	3													1													1	1	
157	putative disease resistance protein Prf	<i>Solanum demissum</i>	pfam00931NB-ARC domain		1745	6												1			1	2					1						1	
158	resistance protein KR4	<i>Glycine max</i>	pfam00931NB-ARC domain	COG4886:Lrr protein	476	3						1																						2

*Bibliotecas: AR1 - Folhas com tratamento de ácido araquidônico. BP1 - Bion positive - plântulas inteiras + células. CA1 – Calo. CB1 - Células em suspensão com bion e brassinosteróides. CL2 - Calo com e sem bion – curta. CS1 - Células em suspensão com sais. EA1 - Calos embriogênicos. EB1 - Embrião de frutos. EM1 - Embrião de sementes germinando. FB1 - Botão floral estágio 1 e botão floral estágio 2 – longa. FB2 - Botão floral estágio 1 e botão floral estágio 2 – curta. FB4 - Botão floral estágio 3 e botão floral estágio 4 – curta. FR1 - Botão floral nº 6, chumbinho nº 1, frutos estádios 1 e 2 – longa. FR2 - Botão floral nº 6, chumbinho nº 1, frutos estádios 1 e 2 – curta. IA1 - Calos embriogênicos. IA2 - Linhagem embriogênica (folha) com indução de 2,4D. IC1 - Linhagem não embriogênica (folha) sem indução de 2,4D. LP1 - Plântulas com tratamento de ácido araquidônico. LV4 - Folhas de ponteiro ortotrópicas sem bion – longa. LV5 - Folhas de ponteiro ortotrópicas sem bion – curta. LV8 - Folhas plagiotrópicas de plantas adultas sem bion – longa. LV9 - Folhas plagiotrópicas de plantas adultas sem bion – curta. NS1 - Raízes com nematóide. PA1 - Linhagem embriogênica (calos primários). PC1 - Linhagem não embriogênica (folhas) com indução de 2,4D. RM1 - Folhas com ferrugem e bicho mineiro. RT3 - Raiz sem bion – longa. RT5 - Raiz com bion – longa. RT8 - Raiz e células em suspensão na presença de alumínio – curta. RX1 - Ramos infectados com *Xylella*. SH2 - Estresse hídrico no campo (pool de tecidos). SI3 - Sementes germinando (inteira). SS1 - Condições normais - irrigado (pool de tecidos). EC1 - Calos embriogênicos (*C. canephora*). SH1 - Folhas - estresse hídrico – curta (*C. canephora*). FR4 – Fruto (*C. racemosa*). FV2 - Fruto verde - estágio 1, 2 e 3 – curta (*C. racemosa*).

Tabela 1.12: ESTs encontradas no projeto NBS

Anotação	Nº ESTs
Proteína de Resistência a Doenças	239
Proteína NBS-LRR	119
Proteína de Resistência	61
Outros	75
Total	494

Tabela 1.13: Espécies que apresentaram seqüências similares às ESTs encontradas no projeto NBS

Organismo	Nº ESTs
<i>Arabidopsis thaliana</i>	171
<i>Solanum</i> sp.	119
<i>Oryza sativa</i>	86
<i>Hordeum vulgare</i>	36
<i>Phaseolus vulgaris</i>	18
<i>Manihot esculenta</i>	15
<i>Glycine max</i>	4
Outros	45
Total	494

Como no projeto LRR, a análise dos *contigs* do projeto NBS também revelou a presença de genes potencialmente envolvidos na resistência a doenças com os domínios conservados pfam00931NB-ARC e COG4886 LRR. Dos 105 *contigs* formados, 46 apresentaram *e-value* menor que 1^{-10} (Tabela 1.14).

O *contig* 32, anotado como *disease resistance protein, putative*, de *Arabidopsis thaliana*, foi o de menor tamanho, com 561 pb. O *contig* 54, anotado como *putative disease resistance protein RPH8A*, de *Oryza sativa*, foi o de maior tamanho, com 2.376 pb.

Apenas os *contigs* 37, 62 e 78 foram formados por pelo menos uma EST da biblioteca RM1. O *contig* 37 foi anotado como *Disease resistance protein (CC-NBS-LRR class), putative*, de *Arabidopsis thaliana*, com 2.255 pb e formado por 13 ESTs, sendo que apenas uma delas pertenceu à biblioteca RM1. O *contig* 62 foi anotado como *Disease resistance-like protein*, de *Coffea arabica*, com 778 pb de tamanho e formado por duas ESTs da biblioteca RM1. O *contig* 78 foi anotado como *putative disease related protein 2*, de *Oryza sativa*, apresentou 1.886 pb e foi formado por quatro ESTs, sendo uma da biblioteca RM1.

Tabela 1.14: Relação de *Contigs* com *e-value* menor que 1^{-10} formados pelo agrupamento do projeto NBS e o respectivo número de ESTs de cada biblioteca.

Contig	Definição	Organismo	Domínio 1	Domínio 2	# pb	# ESTs	A	L	L	L	C	F	F	F	F	R	B	R	B	C	C	F	E	I	I	R	P	S	S
							1	4	5	8	9	2	1	2	2	1	2	2	4	8	1	1	1	4	1	4	1	1	2
4	very similar to disease resistance proteins	<i>Arabidopsis thaliana</i>	pfam00931NB-ARC domain		985	2				1																			
6	NBS/LRR disease resistance protein RPM1	<i>Arabidopsis lyrata</i>	pfam00931NB-ARC domain	COG4886 LRR protein	1121	2														2									
8	similar to NBS-LRR type resistance gene	<i>Oryza sativa</i>	pfam00931NB-ARC domain	COG4886 LRR protein	946	2								1							1								
9	disease resistance protein (CC-NBS-LRR class), putative	<i>Arabidopsis thaliana</i>	pfam00931NB-ARC domain	pfam01576 Myosin tail	953	2	1				1																		
11	putative resistance complex protein I2C-2	<i>Solanum demissum</i>	pfam00931NB-ARC domain		762	2														1	1								
12	putative late blight resistance protein	<i>Solanum demissum</i>	pfam00931NB-ARC domain		858	2					2																		
13	putative RGH1A	<i>Oryza sativa</i>	pfam00931NB-ARC domain		782	2									2														
17	Fom-2 protein	<i>Cucumis melo</i>			1579	9			4							1												4	
19	Fom-2 protein	<i>Cucumis melo</i>			1852	3					1	1																	1
20	disease resistance protein	<i>Arabidopsis thaliana</i>	pfam00931NB-ARC domain		867	4					4																		
21	disease resistance protein	<i>Arabidopsis thaliana</i>	pfam00931NB-ARC domain	COG4886 LRR protein	912	4			2			2																	
22	Disease resistance protein RPH8A (RPP8 homolog A)	<i>Arabidopsis thaliana</i>	pfam00931NB-ARC domain		1443	3	1	2																					
23	putative late blight resistance protein	<i>Solanum demissum</i>	pfam00931NB-ARC domain		1266	4	2		1							1													
30	blight resistance protein RGA1	<i>Solanum bulbocastanum</i>	pfam00931NB-ARC domain	COG4886 LRR protein	1540	2	1													1									
32	disease resistance protein, putative	<i>Arabidopsis thaliana</i>	pfam00931NB-ARC domain		561	2	1														1								
35	putative disease resistance protein	<i>Arabidopsis thaliana</i>	pfam00931NB-ARC domain		1500	2			1	1																			
36	disease resistance-like protein	<i>Coffea canephora</i>	pfam00931NB-ARC domain		1025	3				2																		1	
37	disease resistance protein (CC-NBS-LRR class), putative	<i>Arabidopsis thaliana</i>	pfam00931NB-ARC domain		2255	13	1				1	4	1						1	1	1	1	1					2	
38	putative NBS-LRR type disease resistance protein	<i>Prunus pérsica</i>	pfam00931NB-ARC domain	COG4886 LRR protein	3281	15			4	1	1		1	2	3				1								2		
40	putative disease resistance gene homolog	<i>Oryza sativa</i>	pfam00931NB-ARC domain	COG4886 LRR protein	1619	5								1					2							1	1		
43	putative NBS-LRR type resistance protein	<i>Oryza sativa</i>	pfam00931NB-ARC domain	COG4886 LRR protein	900	2																1						1	
44	disease resistance gene homolog 9N	<i>Brassica napus</i>	pfam00931NB-ARC domain	COG4886 LRR protein	917	2																		1		1			
45	putative late blight resistance protein	<i>Solanum demissum</i>			693	2																					2		
48	FOM-2	<i>Cucumis melo</i>	pfam00931NB-ARC domain		1022	3	2				1																		
49	RGC1	<i>Solanum tuberosum</i>	pfam00931NB-ARC domain		1239	5	1					3										1							
51	Prf	<i>Lycopersicon pimpinellifolium</i>	pfam00931NB-ARC domain		744	2															2								
54	putative disease resistance protein RPH8A	<i>Oryza sativa</i>	pfam00931NB-ARC domain	COG4886 LRR protein	2376	4	3					1																	
56	putative disease resistance protein Prf	<i>Solanum demissum</i>	pfam00931NB-ARC domain		1961	8						1	4	2															1
62	disease resistance-like protein	<i>Coffea arábica</i>	pfam00931NB-ARC domain		778	2													2										
63	resistance protein Tsu5	<i>Arabidopsis thaliana</i>	pfam00931NB-ARC domain		931	2	1																					1	
64	disease resistance-like protein	<i>Coffea arábica</i>	pfam00931NB-ARC domain		1047	2	2																						
65	NBS-LRR protein	<i>Solanum tuberosum</i>	pfam00931NB-ARC domain		593	2					2																		
67	disease resistance protein I2	<i>Lycopersicon esculentum</i>	pfam00931NB-ARC domain		715	2															2								
72	disease resistance-like protein	<i>Coffea arábica</i>	pfam00931NB-ARC domain		1131	5	3			1													1						
75	putative NBS-LRR type resistance protein	<i>Oryza sativa</i>	pfam00931NB-ARC domain		1256	3						1														2			
78	putative disease related protein 2	<i>Oryza sativa</i>	pfam00931NB-ARC domain	COG4886 LRR protein	1886	4									1				1	1									
81	putative disease resistance gene homolog	<i>Oryza sativa</i>	pfam00931NB-ARC domain	COG4886 LRR protein	779	4			1					2									1						
82	disease resistance protein-like	<i>Arabidopsis thaliana</i>	pfam00931NB-ARC domain	COG4886 LRR protein	936	2			1				1																
83	UVB-resistance protein UVR8	<i>Arabidopsis thaliana</i>	COG5184 ATS1		802	4	1	1				1																	

*Bibliotecas: AR1 - Folhas com tratamento de ácido araquidônico. BP1 - Bion positive - plântulas inteiras + células. CA1 - Calo. CB1 - Células em suspensão com bion e brassinosteróides. CL2 - Calo com e sem bion - curta. CS1 - Células em suspensão com sais. EA1 - Calos embriogênicos. EB1 - Embrião de frutos. EM1 - Embrião de sementes germinando. FB1 - Botão floral estágio 1 e botão floral estágio 2 - longa. FB2 - Botão floral estágio 1 e botão floral estágio 2 - curta. FB4 - Botão floral estágio 3 e botão floral estágio 4 - curta. FR1 - Botão floral nº 6, chumbinho nº 1, frutos estádios 1 e 2 - longa. FR2 - Botão floral nº 6, chumbinho nº 1, frutos estádios 1 e 2 - curta. IA1 - Calos embriogênicos. IA2 - Linhagem embriogênica (folha) com indução de 2,4D. IC1 - Linhagem não embriogênica (folha) sem indução de 2,4D. LP1 - Plântulas com tratamento de ácido araquidônico. LV4 - Folhas de ponteiro ortotrópicas sem bion - longa. LV5 - Folhas de ponteiro ortotrópicas sem bion - curta. LV8 - Folhas plagiótropicas de plantas adultas sem bion - longa. LV9 - Folhas plagiótropicas de plantas adultas sem bion - curta. NS1 - Raízes com nematóide. PA1 - Linhagem embriogênica (calos primários). PC1 - Linhagem não embriogênica (folhas) com indução de 2,4D. RM1 - Folhas com ferrugem e bicho mineiro. RT3 - Raiz sem bion - longa. RT5 - Raiz com bion - longa. RT8 - Raiz e células em suspensão na presença de alumínio - curta. RX1 - Ramos infectados com *Xylella*. SH2 - Estresse hídrico no campo (pool de tecidos). SI3 - Sementes germinando (inteira). SS1 - Condições normais - irrigado (pool de tecidos). EC1 - Calos embriogênicos (*C. canephora*). SH1 - Folhas - estresse hídrico - curta (*C. canephora*). FR4 - Fruto (*C. racemosa*). FV2 - Fruto verde - estágio 1, 2 e 3 - curta (*C. racemosa*).

Tabela 1.14: continuação

Contig	Definição	Organismo	Domínio 1	Domínio 2	# pb	# ESTs	C	L	L	L	L	C	F	F	F	F	F	R	B	R	C	C	F	E	I	I	R	P	S	S	S					
							A	V	V	V	V	L	R	R	B	B	T	P	M	B	S	R	A	C	A	C	A	X	A	I	A	I	H	S		
88	putative NBS-LRR disease resistance protein	<i>Oryza sativa</i>	pfam00931NB-ARC domain		814	3												3																		
89	late blight resistance protein-like	<i>Solanum tuberosum</i>	pfam00931NB-ARC domain		803	2												2																		
90	disease resistance-like protein	<i>Coffea arabica</i>	pfam00931NB-ARC domain		1237	3	1													2																
91	blight resistance protein T118	<i>Solanum torajense</i>	pfam00931NB-ARC domain	COG4886 LRR protein	1124	2	2																													
92	Prf	<i>Lycopersicon esculentum</i>	pfam00931NB-ARC domain		1366	2									1												1									
96	disease resistance protein (CC-NBS-LRR class), putative	<i>Arabidopsis thaliana</i>	pfam00931NB-ARC domain		1906	5				1	2												1										1			
97	disease resistance protein BS2	<i>Capsicum chacoense</i>	pfam00931NB-ARC domain		2019	6													2				1									1			2	
99	putative disease resistance protein	<i>Arabidopsis thaliana</i>	pfam00931NB-ARC domain		1580	3												1															1	1		

*Bibliotecas: AR1 - Folhas com tratamento de ácido araquidônico. BP1 - Bion positive - plântulas inteiras + células. CA1 – Calo. CB1 - Células em suspensão com bion e brassinosteróides. CL2 - Calo com e sem bion – curta. CS1 - Células em suspensão com sais. EA1 - Calos embriogênicos. EB1 - Embrião de frutos. EM1 - Embrião de sementes germinando. FB1 - Botão floral estágio 1 e botão floral estágio 2 – longa. FB2 - Botão floral estágio 1 e botão floral estágio 2 – curta. FB4 - Botão floral estágio 3 e botão floral estágio 4 – curta. FR1 - Botão floral nº 6, chumbinho nº 1, frutos estádios 1 e 2 – longa. FR2 - Botão floral nº 6, chumbinho nº 1, frutos estádios 1 e 2 – curta. IA1 - Calos embriogênicos. IA2 - Linhagem embriogênica (folha) com indução de 2,4D. IC1 - Linhagem não embriogênica (folha) sem indução de 2,4D. LP1 - Plântulas com tratamento de ácido araquidônico. LV4 - Folhas de ponteiro ortotrópicas sem bion – longa. LV5 - Folhas de ponteiro ortotrópicas sem bion – curta. LV8 - Folhas plagiotrópicas de plantas adultas sem bion – longa. LV9 - Folhas plagiotrópicas de plantas adultas sem bion – curta. NS1 - Raízes com nematóide. PA1 - Linhagem embriogênica (calos primários). PC1 - Linhagem não embriogênica (folhas) com indução de 2,4D. RM1 - Folhas com ferrugem e bicho mineiro. RT3 - Raiz sem bion – longa. RT5 - Raiz com bion – longa. RT8 - Raiz e células em suspensão na presença de alumínio – curta. RX1 - Ramos infectados com *Xylella*. SH2 - Estresse hídrico no campo (pool de tecidos). S13 - Sementes germinando (inteira). SS1 - Condições normais - irrigado (pool de tecidos). EC1 - Calos embriogênicos (*C. canephora*). SH1 - Folhas - estresse hídrico – curta (*C. canephora*). FR4 – Fruto (*C. racemosa*). FV2 - Fruto verde - estágio 1, 2 e 3 – curta (*C. racemosa*).

Utilizando a palavra *Resistance* como “isca” foi possível identificar 1.864 ESTs. Essa palavra representou um termo geral para a obtenção de ESTs relacionadas com o mecanismo de defesa da planta. Nesse projeto foram encontradas ESTs com similaridade a várias classes de proteínas de resistência a doenças, como CC-NBS-LRR, TIR-NBS-LRR e a genes de resistência a nematóides e vírus (Tabela 1.15).

As ESTs encontradas nesse projeto possuíam similaridade com genes de espécies vegetais como *Arabidopsis thaliana*, *Oryza sativa*, *Glycine max*, *Triticum aestivum* e *Hordeum vulgare*. Além de espécies vegetais, foram encontradas ESTs similares a genes de bactérias como *Escherichia coli* e *Neisseria meningitidis*, e também de humanos (Tabela 1.16).

Outros dois termos gerais utilizados para a obtenção de ESTs foram *Hypersensitive Reaction* (HR – Reação de Hipersensibilidade) e *Pathogenesis* (Patogênese). A HR consiste em uma resposta celular extrema da planta, podendo levar a um alto grau de resistência a doença. A HR resulta na morte repentina de um número limitado de células do hospedeiro que circunda os sítios de infecção do patógeno. A reação é considerada como uma resposta de defesa induzida, culminando com a interrupção do crescimento e multiplicação do patógeno nos tecidos da planta. A resposta ocorre em função do reconhecimento da infecção, por parte do hospedeiro, como consequência da incompatibilidade entre a planta e o patógeno (Pascholati e Leite, 1994).

No projeto *Hypersensitive* foram encontradas várias ESTs de proteínas relacionadas com a reação de hipersensibilidade (Tabela 1.17). As ESTs do desse projeto apresentaram similaridade com genes de *Arabidopsis thaliana*, *Oryza sativa*, *Phaseolus vulgaris*, *Hordeum vulgare*, entre outros (Tabela 1.18).

No projeto *Pathogenesis* foram encontradas ESTs de proteínas relacionadas com a patogênese, quitinases, proteínas da família *thaumatin*, entre outras (Tabela 1.19). As ESTs desse projeto foram similares a genes de *Arabidopsis thaliana*, *Oryza sativa*, *Nicotiana tabacum*, *Solanum tuberosum*, *Lycopersicon esculentum*, *Vitis vinifera*, entre outras (Tabela 1.20).

Tabela 1.15: ESTs encontradas no projeto *Resistance*

Anotação	Nº ESTs
Proteína Prf de Resistência a Doenças	107
Proteína de Resistência a Drogas Pleiotrópicas	73
Proteína Associada à Senescencia	16
Proteína BS2 de Resistência a Doenças	26
Proteína Hipotética	103
Proteína de Resistência a Requeima	16
Proteína de Resistência Multidrogas	214
Proteína de Resistência a Bacteriocina, Putativa	18
Proteína I2 de Resistência a Doenças	20
Proteína de Resistência a Nematóides	15
Proteína Putativa Mlo (Resistência a Patógenos)	20
Proteína Regulatória de Resistência – Putativa	8
Proteína Cf2.1 de Resistência a Doenças	18
Proteína de Resistência a Doenças (classe CC-NBS-LRR), Putativa	59
Proteína de Resistência a Doenças (classe TIR-NBS-LRR), Putativa	18
Proteína de Resistência a Doenças (classe TIR), Putativa	16
Proteína de Resistência a Doenças (classe NBS-LRR), Putativa	30
Proteína de Resistência a Doenças	8
Proteína de Resistência a Doenças LRR	18
Família de Proteína de Resistência a Doenças	23
Proteína de Resistência a Doenças, Putativa	17
Proteína RPP8 de Resistência a Doenças	7
Proteína RPM1 de Resistência a Doenças	14
Proteína Ribossomal S14; 40S; Resistência a Emetina	25
dTDP-glicose 4-6-dehidratase Homologo D18	50
Contém ESTs AU101213(E2609), C72987(E2609)~Proteína de Resistência a Doenças	27
Proteína de Resistência Natural Associada a Macrófago	12
Proteína UVR8 de Resistência-UVB Putativa	27
Transportador ATPase de Glutathione S-conjugado	7
Proteína de Resistência a Tospovirus	14
Proteína Relacionada com Resistência a Biostress	24
Proteína de Resistência a Bacteriocina, Putativa	18
Proteína Relacionada com Resistência ao TMV	10
Proteína de Resistência a Acriflavina	24
Proteína de Resistência a Tetraciclina	17
[Segmento 1 de 2] Redutase Mercúrica (Hg(II) Redutase)	21
Proteína de Resistência a Doenças, Similar	36
Outros	688
Total	1.864

Tabela 1.16: Espécies que apresentaram seqüências similares às ESTs encontradas no projeto *Resistance*

Organismo	Nº ESTs
<i>Arabidopsis thaliana</i>	539
<i>Oryza sativa</i>	245
<i>Lycopersicon esculentum</i>	184
<i>Nicotiana tabacum</i>	63
<i>Homo sapiens</i>	46
<i>Triticum aestivum</i>	44
<i>Escherichia coli</i>	37
<i>Coffea</i> sp.	31
<i>Solanum</i> sp.	28
<i>Hordeum vulgare</i>	26
<i>Neisseria meningitidis</i>	19
<i>Glycine max</i>	16
Outros	586
Total	4.815

Tabela 1.17: ESTs encontradas no projeto *Hypersensitive*

Anotação	Nº ESTs
Proteína de Resposta Induzida por Hipersensibilidade	45
Proteína de Reação Induzida por Hipersensibilidade 4	26
Proteína de ligação a Ca ⁺² Associada à Reação de Hipersensibilidade	6
Proteína UV de Hipersensibilidade (UVH3)	5
Outros	4
Total	86

Tabela 1.18: Espécies que apresentaram seqüências similares às ESTs encontradas no projeto *Hypersensitive*

Organismo homólogo	Nº ESTs
<i>Arabidopsis thaliana</i>	32
<i>Hordeum vulgare</i>	29
<i>Cucumis sativus</i>	7
<i>Oryza sativa</i>	7
<i>Phaseolus vulgaris</i>	6
Outros	5
Total	86

Tabela 1.19: ESTs encontradas no projeto *Pathogenesis*

Anotação	Nº ESTs
Proteína Relacionada com Patogênese 10	295
Precursor 1b de Proteína Relacionada com Patogênese	14
Proteína Relacionada com Patogênese e Stress	79
Proteína Relacionada com Patogênese, Similar	22
Provável Glutathione S-transferase (Proteína Relacionada com Patogênese 1)	7
Família de Taumatina	8
Proteína Relacionada com Patogênese F20M13.220	15
Proteína Relacionada com Patogênese PR10A	121
Ativador Transcricional PTI5 de Genes Relacionados com Patogênese	7

Tabela 1.19: Continuação

Anotação	Nº ESTs
Quitinase (EC 3.2.1.14) / Lisozima (EC 3.2.1.17) Precursor PZ, Relacionado com Patogênese	24
Proteína Similar à Taumatina (Proteína Relacionada com Patogênese), Putativa	28
Proteína Relacionada com Patogênese 5-1	168
Proteína Relacionada com Patogênese 3	10
Proteinase Similar à Subtilisina (EC 3.4.21.-) Precursor P69B, Relacionado com Patogênese	15
Precursor Principal de Proteína Relacionada com Patogênese R (Proteína Similar a Taumatina E22)	72
Outros	94
Total	979

Tabela 1.20: Espécies que apresentaram seqüências similares às ESTs encontradas no projeto *Pathogenesis*

Organismo	Nº ESTs
<i>Vitis vinifera</i>	306
<i>Helianthus annuus</i>	168
<i>Datisca glomerata</i>	109
<i>Arabidopsis thaliana</i>	84
<i>Fagus sylvatica</i>	79
<i>Oryza sativa</i>	40
<i>Nicotiana tabacum</i>	40
<i>Lycopersicon esculentum</i>	17
<i>Solanum tuberosum</i>	14
Outros	122
Total	979

O projeto contendo a palavra *chitinase* foi um dos maiores, sendo formado por 1.855 ESTs (aproximadamente 17% do total). As quitinases são enzimas de um grande grupo, com estruturas e funções bastante diversas, algumas das quais relacionadas com a resistência de várias espécies de plantas a patógenos (Sahai e Manocha, 1993; Jackson e Taylor, 1996). Estas enzimas estão entre os genes mais expressos em café. Dentro do mecanismo de defesa a patógenos, o modo de ação das quitinases é bem semelhante ao das glucanases, ou seja, agem na parede celular do fungo, enfraquecendo-a por meio da hidrólise da quitina (um polímero de N-acetilglicosamina). Esse processo resulta na lise celular e conseqüentemente em morte (Lin et al., 2005). As ESTs encontradas nesse projeto e as espécies homólogas estão relacionadas nas tabelas **1.21** e **1.22**, respectivamente.

Tabela 1.21: ESTs encontradas no projeto *Chitinase*

Anotação	Nº ESTs
Precursor da Proteína Quitinase 3, Similar	561
Precursor da Quitinase Ácida	443
Quitinase 1	96
Provável Precursor de Quitinase (EC 3.2.1.14)	85
Quitinase (EC 3.2.1.14) Classe III, Ácida	68
Família Glicosil Hidrolase	45
Quitinase 134	40
Quitinase Homóloga	32
Precursor de Quitinase Classe VII	25
Quitinase Elicitora Indutível Putativa	11
Quitinase (EC 3.2.1.14) / Lisozima (EC 3.2.1.17) Precursor PZ, Relacionado com Patogênese	8
Outros	441
Total	1.855

Tabela 1.22: Espécies que apresentaram seqüências similares às ESTs encontradas no projeto *Chitinase*

Organismo	Nº ESTs
<i>Trichosanthes kirilowii</i>	561
<i>Glycine max</i>	127
<i>Psophocarpus tetragonolobus</i>	94
<i>Cucumis sativus</i>	85
<i>Hevea brasiliensis</i>	77
<i>Nicotiana tabacum</i>	46
<i>Arabidopsis thaliana</i>	45
<i>Coffea</i> sp.	30
<i>Gossypium hirsutum</i>	25
<i>Oryza sativa</i>	16
<i>Solanum</i> sp.	11
Outros	738
Total	1855

A análise dos *contigs* do projeto *Chitinase* revelou a presença de quitinases com os domínios conservados pfam00704 Glyco hydro 18 e pfam00187 Chitin bind 1. Todos os 47 *contigs* formados apresentaram *e-value* menor que 1^{-10} (Tabela 1.23). O *contig* 39, anotado como ORF3, de *Cucumis sativus*, foi o de menor tamanho, com 542 pb. O *contig* 47, anotado como *acidic chitinase*, de *Glycine max*, foi o de maior tamanho, com 2.662 pb. Cinco *contigs* foram formados por pelo menos uma EST da biblioteca RM1 (5, 10, 24, 31 e 47). O *contig* 5 foi anotado como *chitinase 3-like protein precursor*, de *Trichosanthes kirilowii*, com 1.335 pb e formado por 288 ESTs, sendo que apenas duas delas eram da biblioteca RM1. O *contig* 10 foi anotado como *class1 chitinase*, de *Pisum sativum*, 1.377 pb de tamanho e formado por 48 ESTs, sendo uma da biblioteca RM1. O *contig* 24, anotado como *acidic*

chitinase, de *Glycine max*, tinha 1072 pb e foi formado por oito ESTs, sendo uma da biblioteca RM1. O *contig* 31, anotado como *class1 chitinase*, de *Pisum sativum*, tinha 1.130 pb e apresentou 12 ESTs, sendo uma da biblioteca RM1. O *contig* 47 foi anotado como *acidic chitinase*, de *Glycine max*, com 2.662 pb e era formado por 421 ESTs, sendo três da biblioteca RM1.

O projeto *Cytochrome P450* foi o que apresentou maior número de ESTs, 2.120. O citocromo P450 está envolvido nas vias de oxidação celular. Os genes que codificam essas proteínas podem estar envolvidos na biossíntese de compostos relacionados com a defesa, como o gene PAD3 de *Arabidopsis*, que é requerido para a síntese de camalexina no processo de resistência a *Alternaria brassicicola* (Zhou et al., 1999). Takemoto et al., (1999) mostraram que o produto do gene *CYP82E1* (Citocromo P450) pode estar envolvido com a resistência do tabaco a *Pseudomonas syringae*. Em 2006, Qi e colaboradores mostraram que a enzima AsCYP51H10 (membro da família participou da produção de compostos antimicrobianos (avenacinas) que conferem resistência a doenças em aveia.

As ESTs obtidas nesse projeto e as espécies homólogas estão relacionadas nas tabelas 1.24 e 1.25, respectivamente.

O projeto com a palavra *Glucanase* foi formado por 642 ESTs. Essa palavra é derivada da proteína β 1,3-glucanase. Acredita-se que a atividade dessa proteína na planta ocorra na hidrólise do β 1,3-glucan presente na parede celular do fungo, resultando numa parede celular fraca. Esse enfraquecimento da parede celular do fungo resulta na lise celular e conseqüentemente, na sua morte (Selitrennikoff, 2001). Em vários trabalhos sugeriu-se que as formas extracelulares dessa proteína possuem uma função imediata na defesa das plantas, pela ação direta nas hifas invasoras dos fungos (ação fungicida), com a conseqüente liberação dos elicitores oligossacarídeos das paredes fúngicas, os quais podem levar à ativação de outros mecanismos locais ou sistêmicos de resistência nas plantas. Por sua vez, as formas intracelulares da β 1,3-glucanase parecem atuar tardiamente nas reações de defesa das plantas (Boller, 1988). A atividade antifúngica dessa proteína já foi demonstrada por vários ensaios enzimáticos celulares (Stintzi et al., 1993) e em plantas transgênicas que as expressam (Jach et al., 1995).

Tabela 1.24: ESTs encontradas no projeto *Cytochrome P450*.

Projeto <i>Cytochrome P450</i>	
Anotação	Nº ESTs
<i>No hits found</i>	69
Proteína Hipotética	51
Proteína Expressa	45
Desconhecido	11
Citocromo p450 Putativo	273
Citocromo P450 Monooxygenase	104
Família do Citocromo p450	85
Citocromo P450 ent-kaurene oxidase (GA3)	75
Proteína Ribossomal 60S L10-3 (QM/R22)	66
Trans-cinnamato 4-monooxygenase (Ácido Cinâmico 4-hidroxilase) (CA4H) (C4H) (P450C4H)	66
Aleno Óxido Sintase / Citocromo P450 74A	52
Proteína de Biogênese do Citocromo Tipo C	44
NADH-Citocromo b5 Redutase	29
Elicitor-Indutível do Citocromo P450	27
Peptidase Processadora Mitocondrial, Subunidade Alfa, Precursor Mitocondrial (Alfa-MPP) (Ubiquinol-Citocromo C Redutase Subunidade II)	24
NADPH-Citocromo P-450 Redutase	22
Citocromo P450 71D	14
Citocromo P450 98A	19
Citocromo P450 82C1	17
Proteína de 14 kDa do Complexo Ubiquinol-Citocromo C Redutase (CR14)	19
Flavonóide 3'-monooxygenase (Flavonóide 3'-hidroxilase) (Citocromo P450 75B2)	13
L-lactato Dehidrogenase	13
Proteína Citocromo P450, Similar	10
Citocromo P450 82A3	9
NADPH-citocromo P450 oxidoreductase	7
Ubiquinol-citocromo C Redutase ferro-enxofre subunidade 2, Precursor Mitocondrial (Proteína Rieske ferro-enxofre 2) (RISP2)	6
Outros	1.271
Total	2.441

Tabela 1.25: Espécies que apresentaram seqüências similares às ESTs encontradas no projeto *Chitinase*.

Organismo	Nº ESTs
<i>Arabidopsis thaliana</i>	844
<i>Oryza sativa</i>	108
<i>Zea mays</i>	108
<i>Solanum tuberosum</i>	99
<i>Nicotiana tabacum</i>	49
<i>Lycopersicon esculentum</i>	23
Outros	1.210
Total	6.882

No projeto *Glucanase* foram encontradas ESTs com similaridade a proteínas xiloglucan endotransglicosilase, celulase, beta-1,3-glucanase,

proteínas da família glicosil hidrolase, proteínas reguladoras de brassinosteróides, entre outras (Tabela 1.26). As ESTs desse projeto foram similares a genes de *Arabidopsis thaliana*, *Oryza sativa*, *Nicotiana tabacum*, *Solanum tuberosum*, *Lycopersicon esculentum*, *Zea mays*, entre outras (Tabela 1.27).

Tabela 1.26: ESTs encontradas no projeto *Glucanase*

Anotação	Nº ESTs
Xiloglucan endo-1, 4-beta-D-glucanase (EC 3.2.1.-)	78
Família Glicosil Hidrolase	55
Xiloglucan Endotransglicosilase, Putativa	51
Glucanase – relacionado	47
Precursor de Exoglucanase, Putativo	47
Proteína Expressa	42
Inibidor de Xiloglucanase, Putativo	36
Precursor de Proteína Regulada por Brassinosteróide BRU1	29
Beta-1, 3-glucanase Putativa	24
Elicitor Indutível Beta-1, 3-glucanase	18
Celulase (EC 3.2.1.4)	16
Proteína Endo-1, 3-beta-glucanase-like, Similar	11
Outros	188
Total	642

Tabela 1.27: Espécies que apresentaram seqüências similares às ESTs encontradas no projeto *Glucanase*

Organismo	Nº ESTs
<i>Arabidopsis thaliana</i>	258
<i>Oryza sativa</i>	72
<i>Lycopersicon esculentum</i>	68
<i>Solanum tuberosum</i>	38
<i>Nicotiana tabacum</i>	24
<i>Pyrus</i> sp.	21
<i>Zea mays</i>	14
Outros	147
Total	642

O projeto HSP foi constituído de 240 ESTs. As proteínas HSP (HSP – *Heat Shock Protein*) são induzidas ou têm a sua expressão aumentada para proteger animais e plantas contra estresses ambientais. A *heat shock protein* 90 (HSP90), uma chaperonina requerida para várias vias metabólicas de defesa, contribui para o acúmulo de proteínas de resistência na célula (Dreher e Callis, 2007).

As ESTs que foram obtidas nesse projeto estão relacionadas na tabela 1.28. Foram encontradas nesse projeto, ESTs similares a genes de espécies vegetais e também em humanos (Tabela 1.29). A análise dos *contigs* formados revelou a presença de várias classes de proteínas HSP, com os domínios conservados pfam00011 HSP20 e cd00298 alpha-crystallin-Hsps (Tabela 1.30).

Tabela 1.28: ESTs encontradas no projeto HSP

Anotação	Nº ESTs
Proteína Heat Shock hsp70	45
Proteína Contendo Repetições de Tetratricopeptideo (TRP)	41
Classe I de Proteína Heat Shock 18.5 kD (HSP 18.5)	34
Proteína de ligação FK506 4; 59kD; Proteína p59; Imunofilina de ligação a HSP; Peptidilprolil Cis-trans Isomerase; Rotamase	29
Proteína Heat Shock Citosólica Classe I HSP17.5	12
Classe I de Proteína Heat Shock (HSP 17.4)	10
Outros	69
Total	240

Tabela 1.29: Espécies que apresentaram seqüências similares às ESTs encontradas no projeto HSP

Organismo	Nº ESTs
<i>Arabidopsis thaliana</i>	77
<i>Homo sapiens</i>	29
<i>Castanea sativa</i>	12
Outros	122
Total	240

Todos os 31 *contigs* formados apresentaram *e-value* menor que 1^{-10} . O *contig* 3, anotado como *cytosolic class 1 small heat shock protein 3B*, de *Nicotiana tabacum*, foi o de menor tamanho, com 416 pb e o *contig* 9, anotado como *heat shock protein 70*, de *Cucumis sativus*, foi o de maior tamanho, com 2.608 pb. Apenas dois *contigs* foram formados por pelo menos uma EST da biblioteca RM1. O *contig* 8 foi anotado como *unknown protein*, de *Arabidopsis thaliana*, com 1.627 pb e formado por 40 ESTs, sendo que apenas duas delas foram da biblioteca RM1. O *contig* 12 foi anotado como *heat shock protein 70*, de *Cucumis sativus*, com 1.106 pb de tamanho e formado por 7 ESTs, sendo uma da biblioteca RM1.

Tabela 1.30: Relação de *Contigs* com *e-value* menor que 1^{-10} formados pelo agrupamento do projeto HSP e o respectivo número de ESTs de cada biblioteca.

<i>Contig</i>	Definição	Organismo	Domínio 1	Domínio 2	# pb	# ESTs	C	L	L	L	L	R	C	F	F	F	F	F	R	B	R	C	C	F	E	I	L	R	P	S	S	S
							A	V	V	V	V	T	L	R	R	B	B	B	T	P	M	B	S	R	A	A	P	X	A	A	I	X
1	heat shock protein	<i>Medicago sativa</i>	pfam00011, HSP20		721	3					3																					
2	Nthsp18p	<i>Nicotiana tabacum</i>	pfam00011, HSP20	cd00298, alpha-crystallin-Hsps	781	4																	4									
3	cytosolic class I small heat shock protein 3B	<i>Nicotiana tabacum</i>	pfam00011, HSP20		416	3				3																						
4	HSP70T-2; ATP binding	<i>Arabidopsis thaliana</i>	pfam00012, HSP70		749	2							2																			
5	heat shock protein 17.4	<i>Quercus súber</i>	pfam00011, HSP20	cd00298, alpha-crystallin-Hsps	782	4					1			2			1															
6	ROF1 (ROTAMASE FKBP 1)	<i>Arabidopsis thaliana</i>	pfam00254, FKBP_C	cd00189, TPR	1995	11	1			4	1				1	1						1			1						1	
7	small heat stress protein class CIII	<i>Lycopersicon peruvianum</i>	pfam00011, HSP20		891	5										2				2						1						
8	unknown protein	<i>Arabidopsis thaliana</i>	cd00189, TPR	smart00727, ST11	1627	40	1	2	3	4	2			1	4	10	2		1		2	1		1	1	1			1	2		1
9	heat shock protein 70	<i>Cucumis sativus</i>	pfam00012, HSP70		2608	20		2	4	4					1							2			1					1	5	
10	heat shock protein 70	<i>Cucumis sativus</i>	pfam00012, HSP70		1928	21		3		2	5			1		2				2		2	2								1	
11	unknown protein	<i>Arabidopsis thaliana</i>	pfam00011, HSP20		611	2																	1			1						
12	heat shock protein 70	<i>Cucumis sativus</i>	pfam00012, HSP70		1106	7			1							1				2	1				2							
13	Hsp22.3	<i>Glycine max</i>	pfam00011, HSP20		712	4					1				2									1								
14	peptidyl prolyl cis-trans isomerase	<i>Oryza sativa</i>	pfam00254, FKBP_C	COG0545, FkpA	806	2																	1						1			
15	high molecular weight heat shock protein	<i>Malus x domestica</i>	pfam00012, HSP70		640	2					2																					
16	Hsp70 interacting protein/thioredoxin chimera	<i>Vitis labrusca</i>	pfam00085, Thioredoxin		784	2											2															
17	cytosolic class II low molecular weight heat shock protein	<i>Prunus dulcis</i>	pfam00011, HSP20	cd00298, alpha-crystallin-Hsps	776	5				1					2									2								
18	HSP like protein	<i>Arabidopsis thaliana</i>			832	3																	1	2								
19	Nthsp18p	<i>Nicotiana tabacum</i>	pfam00011, HSP20	cd00298, alpha-crystallin-Hsps	817	18		2	1	1	1			5		1															6	1
20	Endoplasmic homolog precursor	<i>Catharanthus roseus</i>	pfam00183, HSP90		1244	5				2				1		2																
21	OSJNBa0091C07.4	<i>Oryza sativa</i>	COG0545, FkpA	cd00189, TPR	1484	8			1	1				2	3																	1
22	AC009176 putative heat-shock protein	<i>Oryza sativa</i>	COG0326, HtpG, Molecular chaperone, HSP90 family		832	2								1		1																
23	Nthsp18p	<i>Nicotiana tabacum</i>	pfam00011, HSP20		683	10					2	1			5																	2
24	17.5 kDa class I heat shock protein	<i>Glycine max</i>	pfam00011, HSP20	cd00298, alpha-crystallin-Hsps	757	3		1																	1							1
25	peptidylprolyl cis-trans isomerase	<i>Oryza sativa</i>	pfam00254, FKBP_C	COG0545, FkpA	883	4									1		2									1						
26	Nthsp18p	<i>Nicotiana tabacum</i>	pfam00011, HSP20		283	2			2																							
27	Nthsp18p	<i>Nicotiana tabacum</i>	pfam00011, HSP20	cd00298, alpha-crystallin-Hsps	846	11			3						1																	7
28	heat shock protein YegD	<i>Pseudomonas fluorescens PFO-1</i>	pfam00012, HSP70	COG0443, DnaK	749	3																	2									
29	Hsp22.3	<i>Glycine max</i>	pfam00011, HSP20		670	2																										2
30	heat shock protein	<i>Medicago sativa</i>	pfam00011, HSP20		796	2																										2
31	putative cyclophilin-40	<i>Oryza sativa</i>	cd01926, cyclophilin_ABH_like	pfam00160, Pro_isomerase	729	3							2																			1

*Bibliotecas: AR1 - Folhas com tratamento de ácido araquidônico. BP1 - Bion positive - plântulas inteiras + células. CA1 - Calo. CB1 - Células em suspensão com bion e brassinosteróides. CL2 - Calo com e sem bion – curta. CS1 - Células em suspensão com sais. EA1 - Calos embriogênicos. EB1 - Embrião de frutos. EM1 - Embrião de sementes germinando. FB1 - Botão floral estágio 1 e botão floral estágio 2 – longa. FB2 - Botão floral estágio 1 e botão floral estágio 2 – curta. FB4 - Botão floral estágio 3 e botão floral estágio 4 – curta. FR1 - Botão floral nº 6, chumbinho nº 1, frutos estádios 1 e 2 – longa. FR2 - Botão floral nº 6, chumbinho nº 1, frutos estádios 1 e 2 – curta. IA1 - Calos embriogênicos. IA2 - Linhagem embriogênica (folha) com indução de 2,4D. IC1 - Linhagem não embriogênica (folha) sem indução de 2,4D. LP1 - Plântulas com tratamento de ácido araquidônico. LV4 - Folhas de ponteiro ortotrópicas sem bion – longa. LV5 - Folhas de ponteiro ortotrópicas sem bion – curta. LV8 - Folhas plagiotrópicas de plantas adultas sem bion – longa. LV9 - Folhas plagiotrópicas de plantas adultas sem bion – curta. NS1 - Raízes com nematóide. PA1 - Linhagem embriogênica (calos primários). PC1 - Linhagem não embriogênica (folhas) com indução de 2,4D. RM1 - Folhas com ferrugem e bicho mineiro. RT3 - Raiz sem bion – longa. RT5 - Raiz com bion – longa. RT8 - Raiz e células em suspensão na presença de alumínio – curta. RX1 - Ramos infectados com *Xylella*. SH2 - Estresse hídrico no campo (pool de tecidos). SI3 - Sementes germinando (inteira). SS1 - Condições normais - irrigado (pool de tecidos). EC1 - Calos embriogênicos (*C. canephora*). SH1 - Folhas - estresse hídrico – curta (*C. canephora*). FR4 – Fruto (*C. racemosa*). FV2 - Fruto verde - estágio 1, 2 e 3 – curta (*C. racemosa*).

O projeto *thaumatin* ficou constituído de 317 ESTs. Thaumatin é uma proteína relacionada com patogênese (PR – *Pathogenesis Related*). Ela possui efeito fungicida contra um grande número de patógenos de plantas e humanos (Selitrennikoff, 2001). Embora o mecanismo preciso de ação dessa proteína ainda não seja completamente conhecido, existem várias observações que podem levar a hipótese de sua atuação na morte do fungo (Selitrennikoff, 2001). Essa proteína causa alterações na permeabilidade da célula em fungos que possuem parede celular, mas nenhum ou pouco efeito no protoplasto (Roberts e Selitrennikoff, 1990). Ela se liga ao 1,3 β -glucan e possui atividade 1,3 β -glucanase *in vitro* (Grenier et al., 1993; Trudel et al., 1998). Em tabaco causa perturbações na regulação da montagem da parede celular do fungo *Saccharomyces cerevisiae* (Yun et al., 1997; Yun et al., 1998). Essas observações dificultam a definição do mecanismo de ação dessa proteína.

Por meio desse projeto foram obtidas ESTs de proteínas relacionadas com a patogênese e thaumatinas, além de outras (Tabela 1.31). Foram obtidas ESTs com similaridade a genes de *Arabidopsis thaliana*, *Vitis vinifera* e *Nicotiana tabacum*, entre outras (Tabela 1.32).

A análise dos *contigs* formados revelou a presença de várias classes de proteínas thaumatinas, com o domínio conservado Smart 00205 THN (Tabela 1.33). Todos os 16 *contigs* formados apresentaram *e-value* menor que 1^{-10} . O *contig* 16, anotado como *thumain-like protein*, de *Arabidopsis thaliana*, foi o de menor tamanho, com 765 pb e o *contig* 1, anotado como *putative thaumatin-like protein*, de *Arabidopsis thaliana*, o de maior tamanho, com 1.952 pb. Nenhum do *contigs* eram formados por ESTs da biblioteca RM1.

O projeto *Phytoalexin* foi constituído de 12 ESTs. Todas elas apresentaram a anotação *phytoalexin*, de *Arabidopsis thaliana*. As fitoalexinas são compostos que são sintetizados por células vegetais e nelas se acumulam após a infecção microbiana (Stoessl, 1986). Entre os efeitos sobre o crescimento de fungos está a inibição da elongação do tubo germinativo e do crescimento ra-

Tabela 1.31: ESTs encontradas no projeto *Thaumatin*

Anotação	Nº ESTs
Proteína Taumatina, Similar	284
Precursor de Proteína R Relacionado com Patogênese	10
Família Taumatina	9
Outros	14
Total	317

Tabela 1.32: Espécies que apresentaram seqüências similares às ESTs encontradas no projeto HSP

Organismo	Nº ESTs
<i>Vitis vinifera</i>	121
<i>Sambucus nigra</i>	113
<i>Arabidopsis thaliana</i>	55
<i>Nicotiana tabacum</i>	9
Outros	19
Total	317

Tabela 1.33: Relação de *Contigs* com *e-value* menor que 1^{-10} formados pelo agrupamento do projeto *Thaumatococcus* e o respectivo número de ESTs de cada biblioteca.

Contig	Definição	Organismo	Domínio 1	# pb	# ESTs	C A 1	L V 5	L V 8	L V 9	F R 1	F R 2	F B 1	F B 2	F B 4	R T 8	B P 1	C B 1	C S 1	E A 1	I C 1	P C 1	I A 2	L P 1	R X 1	P A 1	S I 3	S H 2	S S 1
1	putative thaumatin-like protein	<i>Arabidopsis thaliana</i>	Smart 00205, THN	1952	4			2		1													1					
2	putative thaumatin-like protein	<i>Arabidopsis thaliana</i>	Smart 00205, THN	911	7			1	2						4													
3	thaumatin-like protein SE39b	<i>Nicotiana tabacum</i>	Smart 00205, THN	1011	4								3	1														
4	thaumatin-like protein	<i>Actinidia deliciosa</i>	Smart 00205, THN	1017	41					1									13		10	1	1		3	12		
5	putative thaumatin-like protein	<i>Solanum tuberosum</i>	Smart 00205, THN	975	56					1	14										28	12			1			
6	thaumatin-like protein	<i>Actinidia deliciosa</i>	Smart 00205, THN	892	15														1		7				2	5		
7	thaumatin-like protein	<i>Vitis vinifera</i>	Smart 00205, THN	1125	117										1						100	11		2	1	2		
8	thaumatin-like protein	<i>Sambucus nigra</i>	Smart 00205, THN	880	21																19	1			1			
9	SCUTL2	<i>Vitis vinifera</i>	Smart 00205, THN	925	10							1	2				1	6										
10	thaumatin-like protein SE39b	<i>Nicotiana tabacum</i>	Smart 00205, THN	1037	3																							
11	thaumatin, putative	<i>Arabidopsis thaliana</i>	Smart 00205, THN	1147	6		2			3		1																
12	thaumatin-like protein; 9376-10898	<i>Arabidopsis thaliana</i>	Smart 00205, THN	858	7						2		2						1	2								
13	Osmotin-like protein precursor	<i>Lycopersicon esculentum</i>	Smart 00205, THN	1135	10	1		2			2		1								1					1	1	1
14	At5g40020	<i>Arabidopsis thaliana</i>	Smart 00205, THN	890	2	1																				1		
15	thaumatin-like protein	<i>Arabidopsis thaliana</i>	Smart 00205, THN	870	4											2					1					1		
16	thaumatin-like protein; 12104-13574	<i>Arabidopsis thaliana</i>	Smart 00205, THN	765	2							3																2

*Bibliotecas: AR1 - Folhas com tratamento de ácido araquidônico. BP1 - Bion positive - plântulas inteiras + células. CA1 - Calo. CB1 - Células em suspensão com bion e brassinosteróides. CL2 - Calo com e sem bion - curta. CS1 - Células em suspensão com sais. EA1 - Calos embriogênicos. EB1 - Embrião de frutos. EM1 - Embrião de sementes germinando. FB1 - Botão floral estágio 1 e botão floral estágio 2 - longa. FB2 - Botão floral estágio 1 e botão floral estágio 2 - curta. FB4 - Botão floral estágio 3 e botão floral estágio 4 - curta. FR1 - Botão floral nº 6, chumbinho nº 1, frutos estádios 1 e 2 - longa. FR2 - Botão floral nº 6, chumbinho nº 1, frutos estádios 1 e 2 - curta. IA1 - Calos embriogênicos. IA2 - Linhagem embriogênica (folha) com indução de 2,4D. IC1 - Linhagem não embriogênica (folha) sem indução de 2,4D. LP1 - Plântulas com tratamento de ácido araquidônico. LV4 - Folhas de ponteiro ortotrópicas sem bion - longa. LV5 - Folhas de ponteiro ortotrópicas sem bion - curta. LV8 - Folhas plagiótropicas de plantas adultas sem bion - longa. LV9 - Folhas plagiótropicas de plantas adultas sem bion - curta. NS1 - Raízes com nematóide. PA1 - Linhagem embriogênica (calos primários). PC1 - Linhagem não embriogênica (folhas) com indução de 2,4D. RM1 - Folhas com ferrugem e bicho mineiro. RT3 - Raiz sem bion - longa. RT5 - Raiz com bion - longa. RT8 - Raiz e células em suspensão na presença de alumínio - curta. RX1 - Ramos infectados com *Xylella*. SH2 - Estresse hídrico no campo (pool de tecidos). SI3 - Sementes germinando (inteira). SS1 - Condições normais - irrigado (pool de tecidos). EC1 - Calos embriogênicos (*C. canephora*). SH1 - Folhas - estresse hídrico - curta (*C. canephora*). FR4 - Fruto (*C. racemosa*). FV2 - Fruto verde - estágio 1, 2 e 3 - curta (*C. racemosa*).

dial das colônias, o colapso do sistema de membranas, a perda de eletrólitos, dentre outros. As plantas resistentes produzem altos níveis de fitoalexinas, quando comparadas às susceptíveis. O aumento na concentração dessas enzimas vem acompanhado de aumentos na concentração de enzimas-chaves das vias biossintéticas envolvidas na sua produção, como exemplo a chalcone sintase (Pascholati e Leite, 1994).

A chalcone sintase é uma enzima regulatória da biossíntese de fenilpropanóides que catalizam a primeira reação da via metabólica específica da biossíntese de flavonóides e isoflavonóides. Elicitores causam um estímulo transiente rápido da transcrição de genes de chalcone sintase como um evento basal na expressão da resposta às fitoalexinas (Ryder et al., 1984). Todas as 153 ESTs obtidas no projeto *Chalconesynthase* tiveram anotação *Chalcone Synthase*.

O projeto criado utilizando a palavra *Polyphenoloxidase* foi formado por 67 ESTs. As anotações mais relevantes foram *catechol oxidase* e *polyphenol oxidase* (Figura 1.5).

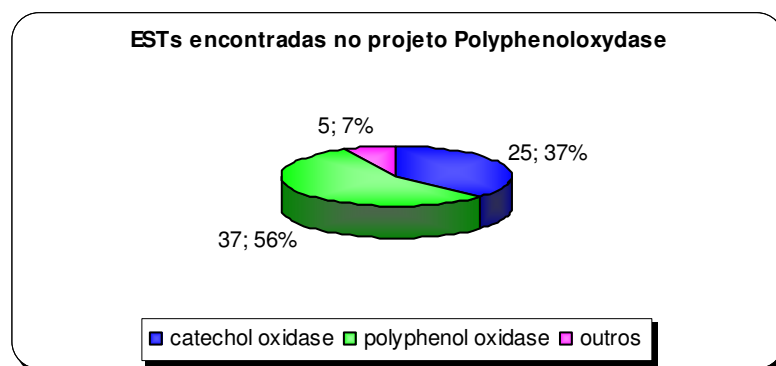


Figura 1.5: ESTs do projeto Polyphenoloxidase

Embora a função fisiológica da polifenoloxidase em células vegetais ainda não seja totalmente conhecida, a maioria dos relatos sobre essa enzima indica a função de defesa de plantas contra patógenos e pragas. O modo de ação proposto é baseado na sua capacidade de oxidar compostos fenólicos quando o tecido é danificado. Nessa situação, a ruptura dos plastídios, o compartimento celular onde a polifenoloxidase está localizada, leva a enzima a entrar em contato com compostos fenólicos liberados pela ruptura do vacúolo (Melo et al., 2006).

O projeto *Importin* foi constituído de 532 ESTs. As ESTs encontradas nesse projetos estão relacionadas na tabela 1.34 e os organismos homólogos, na tabela 1.35.

Tabela 1.34: ESTs encontradas no projeto *Importin*

Anotação	Nº ESTs
<i>No Hits Found</i>	23
Subunidade Pequena da Rubisco	440
Subunidade da Importina Alfa	17
Importina Alfa 2	13
Provável Fator de Transporte Nuclear de Importina Alfa [importado]	10
Importina Beta, Putativa	8
Proteína de Susceptibilidade a Apoptose Celualr (Re-exportador de Importina-alfa), Putativo	3
Outros	18
Total	532

Tabela 1.35: Espécies que apresentaram seqüências similares às ESTs encontradas no projeto *Importin*

Organismo	Nº ESTs
<i>Coffea arabica</i>	440
<i>Arabidopsis thaliana</i>	42
<i>Capsicum annuum</i>	14
<i>Oryza sativa</i>	7
Outros	29
Total	532

A Importina α media a importação de proteínas citosólicas para dentro do núcleo. As importinas α de plantas se ligam a fatores de virulência de *Agrobacterium tumefaciens* e vários vírus fitopatogênicos (Hermsmeier et al., 2001).

O projeto *Glucosyltransferase* gerou 1.286 ESTs. As glicosiltransferases catalisam a transferência de resíduos de glicose para vários substratos e regulam a atividade de compostos que têm funções importantes na defesa da planta contra patógenos, tal como o ácido salicílico (Chong et al., 2002).

As ESTs obtidas no projeto *Glucosyltransferase* incluíram proteínas da família das glicosiltransferases, sacarose sintase, glicosiltransferases induzidas pelo frio, entre outras (Tabela 1.36). Neste projeto foram obtidas ESTs similares a genes de *Arabidopsis thaliana*, *Oryza sativa*, *Lycopersicon esculentum*, *Nicotiana tabacum*, entre outras (Tabela 1.37).

Tabela 1.36: ESTs encontradas no projeto *Glucosyltransferase*

Anotação	Nº ESTs
<i>No Hits Found</i>	76
Proteína Expressa	22
Desconhecido	18
Sacarose Sintase (Sacarose-UDP Glucosiltransferase)	198
Família Glicosiltransferase	112
Glucosiltransferase Putativa	74
Proteína Hipotética	56
Glucosiltransferase NTGT2	32
Hidroquinona Glucosiltransferase (Arbutina Sintase)	24
Indol-3-acetate beta-glucosiltransferase - Relacionado	23
Glucosil Transferase Induzida pelo Frio	18
Indol-3-acetate beta-glucosiltransferase Putativa	18
Glucuronosil Transferase Homóloga, Relacionado com o Amadurecimento	13
Outros	602
Total	1286

Tabela 1.37: Espécies que apresentaram seqüências similares às ESTs encontradas no projeto *Glucosyltransferase*

Organismo	Nº ESTs
<i>Arabidopsis thaliana</i>	365
<i>Nicotiana tabacum</i>	108
<i>Oryza sativa</i>	88
<i>Lycopersicon esculentum</i>	53
<i>Pseudomonas</i> sp	47
<i>Solanum</i> sp	25
<i>Citrus</i> sp	17
Outros	583
Total	1286

A análise gráfica mostrou o número de ESTs de bibliotecas de interesse identificadas em cada estratégia de mineração (Figura 1.6). As bibliotecas consideradas para a análise foram a NS1 (raízes com nematóide), RM1 (folhas com bicho mineiro e ferrugem), RX1 (ramos com *Xylella*) e SS1 (tecidos em condições normais). As bibliotecas NS1, RM1 e RX1 foram escolhidas para essa análise pelo fato de terem sido submetidas a estresses bióticos e, portanto, representam uma fonte potencial de ESTs que podem estar envolvidas com o processo de defesa da planta contra esses estresses. A biblioteca SS1 serviu como controle, para identificar os genes que estão em funcionamento quando a planta está em condições normais, e quais desses genes também se expressam nas condições de estresse citadas acima.

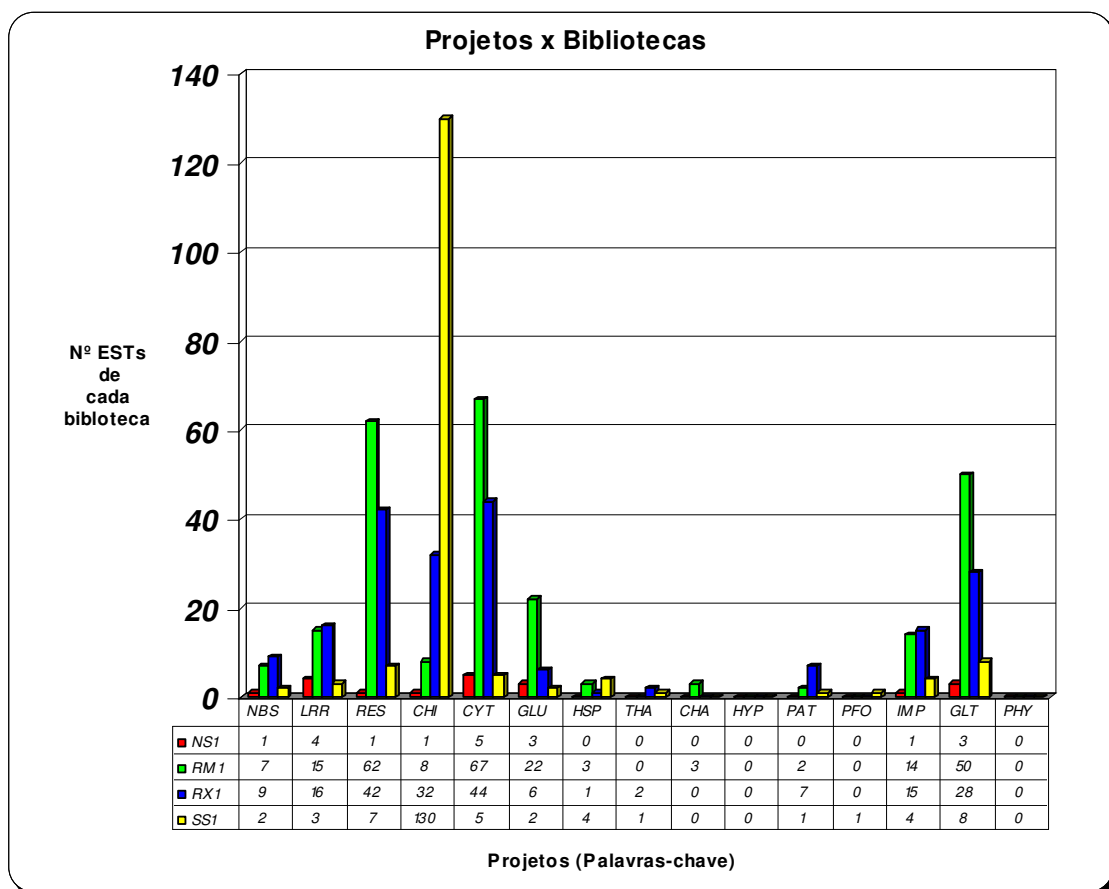


Figura 1.6: Número de ESTs das Bibliotecas NS1 (Raízes com nematóide), RM1 (Folhas com bicho mineiro e ferrugem), RX1 (Ramos com *Xylella*) e SS1 (Tecidos em condições normais) presentes nos projetos criados. RES = Resistance; CHI = Chitinase; CYT = Cytochrome P450; GLU = Glucanase; THA = Thaumatin; CHA = Chalconesynthase; HYP = Hypersensitive; PAT = Pathogenesis; PFO = Polyphenoloxidase; IMP = Importin; GLT = Glucosyltransferase e PHY = Phitoalexin.

A análise mostrou que a biblioteca RM1 foi a mais bem representada nos projetos *Resistance*, *Cytochrome P450*, *Glucanase* e *Glucosyltransferase*. Essa biblioteca também foi bem representada nos projetos NBS, LRR e *Importin*. Esse resultado confirma a importância potencial das ESTs provenientes da biblioteca RM1 no processo de defesa do cafeeiro contra bicho mineiro e ferrugem.

Semelhantemente à biblioteca RM1, a RX1 também foi bem representada nos projetos *Resistance*, *Cytochrome P450* e *Glucosyltransferase*. RX1 foi a mais bem representada nos projetos NBS, LRR, *Pathogenesis* e *Importin*. Esse resultado indica que as ESTs provenientes dessa biblioteca são importantes para o estudo da defesa do cafeeiro contra

Xylella, pois elas estão potencialmente relacionadas com o processo de resistência contra essa bactéria.

Os projetos *Resistance*, *Chitinase*, *Cytochrome P450* e *Glucosyltransferase* foram os que obtiveram o maior número de ESTs das bibliotecas de interesse.

A palavra “*resistance*” representa um termo geral no mecanismo de defesa da planta, e por essa razão, era esperado que as ESTs das três bibliotecas submetidas a estress fossem bem representadas. Entretanto, a biblioteca NS1 apresentou apenas uma EST nesse projeto, enquanto que a SS1 apresentou sete. Isso mostra que o nível de expressão de genes relacionados com esse termo pode estar bem reduzido, ou que os genes simplesmente não foram amostrados. Estudos posteriores poderão elucidar esses resultados.

As quitinases, relacionadas com a resistência de várias espécies de plantas a patógenos (Sahai e Manocha, 1993; Jackson e Taylor, 1996), foram bastante expressas em condições normais. Esse resultado está de acordo com o trabalho de Lin et al. (2005), o qual mostrou que as quitinases estão entre os genes mais expressos em café. Segundo esses pesquisadores, o fato das quitinases serem muito expressas e bem representadas por uma grande família em café pode ser reflexo de uma grande necessidade de resistência a fungos. Essa necessidade seria devido à natureza perene do cafeeiro e ao fato dessa planta ser uma espécie tropical, para a qual é comum que haja uma grande quantidade de patógenos.

Em plantas os citocromo P450 está envolvido numa enorme quantidade de reações biossintéticas, levando a formação de conjugados de ácidos graxos, hormônios, compostos de defesa, entre outros. Até o mês de abril de 2007 haviam sido identificadas 6.766 seqüências de Citocromo P450 em 708 famílias, sendo 2.417 seqüências de espécies vegetais (*Cytochrome P450 Homepage* - <http://drnelson.utmem.edu/CytochromeP450.html>). Dessa forma, a grande quantidade de ESTs das bibliotecas RM1 e RX1 obtidas nesse projeto devem estar preferencialmente envolvidas em outros processos biológicos, em detrimento do processo de defesa da planta contra doenças.

Nos projetos *Hypersensitive* e *Phytoalexin* não foram identificadas nenhuma EST das bibliotecas analisadas. Os genes relacionados com esses

termos não são expressos ou podem apresentar uma expressão bem reduzida, e por isso não foram amostrados. Estudos detalhados poderão explicar esses resultados.

3.3. Seqüências publicadas

Foram identificadas 66 seqüências relacionadas com o a resistência do cafeeiro a doenças nos artigos científicos consultados (Tabela 1.38). Essas seqüências foram submetidas a uma comparação com o banco de dados do PBGC, por meio do BLAST. As seqüências encontradas com *e-value* menor que 1^{-10} e *score* maior que 100 foram selecionadas para análise.

As seqüências obtidas pelo trabalho de Fernandez et al. (2004) são de genes isolados que podem estar associados com a expressão de mecanismos basais de resistência de cafeeiros a *Hemileia vastatrix*. As proteínas preditas, codificadas por essas ESTs, mostraram similaridade com proteínas de resistência a doenças, proteínas de tolerância a estresses e componentes de vias de sinalização celular.

Usando RT-PCR Ganesh et al. (2006) verificaram que o gene *CaNDR1* foi expresso em amostras tratadas com *H. vastatrix* e que o gene *CaWRKY1* foi ativado logo após a inoculação com esse patógeno. Eles verificaram ainda a ativação do gene *CaR111* em resposta a raças avirulentas de *H. vastatrix*, sugerindo que esse gene pode ser um fator interessante na via de sinalização da resistência de *C. arabica* à ferrugem.

As seqüências do trabalho de Lin et al. (2005) são provenientes de bibliotecas de folhas, sementes e frutos de *C. canephora*. As ESTs selecionadas possuem as anotações de proteínas de defesa contra doenças, como *chitinase*, *thaumatin*, *hypersensitive-induced protein*, entre outras.

Os genes *cahsp70*, *cachi 3-1*, *cachi 3-2*, *cachi 3-3*, *cachi 4-1* e *capox-1* do trabalho de Chen et al. (2003) são de uma biblioteca de cDNA construída a partir de folhas de *C. arabica* var. Dilla e Alghe (CIFC 110/5) inoculadas com *H. vastatrix*. Esses genes codificam proteínas como quitinases, HSP70, e peroxidases, todas relacionadas com a defesa da planta a patógenos.

Tabela 1.38: Seqüências publicadas utilizadas na mineração de ESTs do PBGC

Fonte	Seqüência	Score	e-value	Blast EST café - NR	score	e-value
Lin et al., 2005	SGN120121 Thaumatin, pathogenesis related SGN120244 Disease resistance protein (TIR-NBS-LRR class)	656	0.0	>gi 128378 sp P10973 NLTA_RICCO NONSPECIFIC LIPID-TRANSFER PROTEIN A (NS-LTP A) (PHOSPHOLIPID TRANSFER PROTEIN) (PLTP)	90,5	3,00E-17
Lin et al., 2005	SGN121998 Disease resistance protein	888	0.0	>gi 34907232 ref NP_914963.1 putative disease resistance protein I2 [Oryza sativa (japonica cultivar-group)]	108	1,00E-16
Lin et al., 2005	SGN124574 Leucine-rich repeat, disease resistance protein	927	0.0	>gi 7488989 pir T07589 disease resistance protein Prf - tomato gi 1513144 gb AAC49408.1 PRF	163	4,00E-39
Lin et al., 2005	SGN124466 Leucin-rich repeat, plant specific, receptor-related protein kinase	1012	0.0	>gi 34904990 ref NP_913842.1 P0456B03.2 [Oryza sativa (japonica cultivar-group)]	142	7,00E-33
Lin et al., 2005	SGN119817 Chitinase	783	0.0	>gi 15225286 ref NP_180201.1 receptor-related protein kinase, ERECTA [Arabidopsis thaliana]	180	2,00E-60
Lin et al., 2005	SGN123451 Acidic endochitinase	1510	0.0	>gi 5880843 gb AAD54934.1 AF141372_1 chitinase precursor [Petroselinum crispum]	243	1,00E-111
Lin et al., 2005	SGN119672 Hypersensitive-induced protein, band 7 protein	1279	0.0	>gi 3451147 emb CAA09110.1 chitinase [Hevea brasiliensis]	266	4,00E-70
Lin et al., 2005	gi 4835584 DSS14 acidic chitinase	1584	0.0	>gi 34484310 gb AAQ72788.1 hypersensitive-induced response protein [Cucumis sativus]	423	1,00E-120
Fernandez et al., 2004	gi 9255920 DSS13 cyclic nucleotide-gated cation channel DND1	767	0.0	>gi 116328 sp P17541 CHIA_CUCSA ACIDIC ENDOCHITINASE PRECURSOR	162	7,00E-39
Fernandez et al., 2004	gb AAB97316.1 DSS11 heat shock 70 protein	181	2E-44	>gi 15242291 ref NP_197045.1 cyclic nucleotide-regulated ion channel (CNGC2) [Arabidopsis thaliana]	142	7,00E-63
Fernandez et al., 2004	gi 15222057 DSS9 cytochrome P450, putative	815	0	>gi 15230534 ref NP_187864.1 heat shock protein hsp70 [Arabidopsis thaliana]	425	1,00E-118
Fernandez et al., 2004	gi 50980143 CA-H11_01_B05 Coffee SSH library 1	835	0	>gi 9502380 gb AAF88087.1 T12C24.27 [Arabidopsis thaliana]	253	4,00E-66
Fernandez et al., 2004	[gi:50980111]CA-H11_02_B06 Coffee SSH library 1	196	2E-49	>gi 13123983 sp Q9M367 BCN1_ARATH Beclin 1-like protein	258	1,00E-67
Fernandez et al., 2004	[gi:50979912]CA-H12_04_A09 Coffee SSH library 1	681	0	>gi 7446460 pir T14337 RAD23 protein, isoform II - carrot	168	1,00E-40
Fernandez et al., 2004	[gi:50979976]CA-H11_01_B01 Coffee SSH library 1	554	1E-157	>gi 15241023 ref NP_195785.1 Macrophage migration inhibitory factor (MIF) family [Arabidopsis thaliana]	197	3,00E-49
Fernandez et al., 2004	[gi:50979965]CA-H11_04_E06 Coffee SSH library 1	629	1E-179	>gi 11066033 gb AAG28426.1 cytosolic aconitase [Nicotiana tabacum]	409	1,00E-120
Fernandez et al., 2004	[gi:50979768]CA-H12_03_C03 Coffee SSH library 1	550	0.0	>gi 17105223 gb AAL35606.1 AF442962_1 peroxisomal multifunctional protein [Oryza sativa]	259	4,00E-90
Fernandez et al., 2004	[gi:50979797]CA-H12_04_G04 Coffee SSH library 1	248	4E-65	>gi 29367118 gb AAO72306.1 NdhK [Hordeum vulgare]	283	3,00E-79
Fernandez et al., 2004	[gi:50979698]CA-H12_03_C11 Coffee SSH library 1	975	0.0	>gi 10862871 emb CAC13956.1 glutathione reductase [Mesembryanthemum crystallinum]	418	1,00E-119
Fernandez et al., 2004	[gi:50980083]CA-H11_02_B09 Coffee SSH library 1	1033	0.0	>gi 30013657 gb AAP03871.1 oxygen evolving complex 33 kDa photosystem II protein [Nicotiana tabacum]	373	1,00E-113
Fernandez et al., 2004	[gi:50980022]CA-H11_02_E04 Coffee SSH library 1	629	1E-179	>gi 12331298 emb CAC24711.1 cytochrome P450 [Solanum tuberosum]	209	6,00E-53
Fernandez et al., 2004	[gi:50980069]CA-H11_04_H05 Coffee SSH library 1	604	1E-172	>gi 584861 sp P37118 C712_SOLME Cytochrome P450 71A2 (CYPLXXIA2) (P-450EG4)	269	2,00E-79
Fernandez et al., 2004	[gi:50980191]CA-H11_02_B08 Coffee SSH library 1	291	6E-78	>gi 38567916 emb CAE04091.3 OSJNBA008812.12 [Oryza sativa (japonica cultivar-group)]	95,9	9,00E-19
Fernandez et al., 2004	[gi:50979849]CA-H12_04_E07 Coffee SSH library 1	677	0.0	>gi 12331298 emb CAC24711.1 cytochrome P450 [Solanum tuberosum]	208	1,00E-52
Fernandez et al., 2004	[gi:50980191]CA-H11_02_B08 Coffee SSH library 1	291	1E-77	>gi 15021761 gb AAK77908.1 AF397903_1 AAA-metalloprotease FtsH [Pisum sativum]	298	6,00E-80
Fernandez et al., 2004	[gi:50979670]CA-H12_02_E08 Coffee SSH library 1	508	1E-143	>gi 320605 pir A34131 metallothionein I homolog - spotted monkey flower	87,8	2,00E-16
Fernandez et al., 2004	[gi:50980032]CA-H11_01_B06 Coffee SSH library 1	767	0.0	>gi 26985221 gb AAN86275.1 non-cell-autonomous heat shock cognate protein 70 [Cucurbita maxima]	410	1,00E-113
Fernandez et al., 2004	[gi:50979950]CA-H11_02_F10 Coffee SSH library 1	565	1E-160	>gi 462013 sp P35016 ENPL_CATRO Endoplasmic homolog precursor (GRP94 homolog)	328	8,00E-89
Fernandez et al., 2004	[gi:50980104]CA-H11_04_A11 Coffee SSH library 1	860	0.0	>gi 6934298 gb AAF31705.1 AF221856_1 heat-shock protein 80 [Euphorbia esula]	236	3,00E-61
Fernandez et al., 2004	[gi:50979967]CA-H11_01_C02 Coffee SSH library 1	504	1E-142	>gi 25285681 pir G86305 SRG1 homolog [imported] - Arabidopsis thaliana	221	1,00E-56
Fernandez et al., 2004	[gi:50980012]CA-H11_03_G11 Coffee SSH library 1	592	1E-168	>gi 7488910 pir T14329 dermal glycoprotein precursor, extracellular - carrot (fragment)	171	2,00E-41
Fernandez et al., 2004	[gi:50979858]CA-H12_03_D08 Coffee SSH library 1	425	1E-118	>gi 13992713 gb AAF01465.2 bdn1 [Boea crassifolia]	98,6	2,00E-19
Fernandez et al., 2004	[gi:50980251]CA-H11_02_B04 Coffee SSH library 1	635	0.0	>gi 16974114 emb CAC95155.1 putative resistance protein [Lycopersicon esculentum]	253	2,00E-66
Fernandez et al., 2004	[gi:50980003]CA-H11_02_D06 Coffee SSH library 1	365	1E-100	>gi 24459849 emb CAC82600.1 disease resistance-like protein [Coffea canephora]	164	2,00E-39
Fernandez et al., 2004	[gi:50979978]CA-H11_04_A05 Coffee SSH library 1	725	0	>gi 13752562 gb AAK38727.1 AF369707_1 importin alpha 2 [Capsicum annuum]	479	1,00E-134
Fernandez et al., 2004	[gi:50980127]CA-H11_03_H01 Coffee SSH library 1	938	0	>gi 15232308 ref NP_188696.1 non-race specific disease resistance protein (NDR1) [Arabidopsis thaliana]	135	5,00E-31
Fernandez et al., 2004	[gi:50980232]CA-H11_03_A12 Coffee SSH library 1	862	0	>gi 15553476 gb AAL01886.1 AF404590_1 chitinase 3-like protein precursor [Trichosanthes kirilowii]	164	2,00E-39
Fernandez et al., 2004	[gi:50980231]CA-H11_04_E02 Coffee SSH library 1	944	0	>gi 15553476 gb AAL01886.1 chitinase 3-like protein precursor [Trichosanthes kirilowii]	168	1,00E-40
Fernandez et al., 2004	[gi:50979888]CA-H12_02_G12 Coffee SSH library 1	456	1E-127	>gi 15553476 gb AAL01886.1 chitinase 3-like protein precursor [Trichosanthes kirilowii]	181	2,00E-44
Fernandez et al., 2004	[gi:50979788]CA-H12_02_B01 Coffee SSH library 1	665	0	>gi 1076693 pir S51678 chitinase (EC 3.2.1.14) class I - European elder (fragment)	379	1,00E-104
Fernandez et al., 2004	[gi:50979684]CA-H12_03_A05 Coffee SSH library 1	569	1E-161	>gi 25316444 pir E84471 probable beta-1,3-glucanase [imported] - Arabidopsis thaliana	361	8,00E-99
Fernandez et al., 2004	[gi:50980140]CA-H11_01_G04 Coffee SSH library 1	191	1E-47	>gi 25404490 pir G96670 hypothetical protein F13O11.7 [imported] - Arabidopsis thaliana	120	4,00E-26
Fernandez et al., 2004	[gi:50980185]CA-H11_04_F12 Coffee SSH library 1	635	0	>gi 14423842 sp O50001 PRU1_PRUAR Major allergen Pru ar 1	164	2,00E-39
Fernandez et al., 2004	[gi:50980055]CA-H11_04_B08 Coffee SSH library 1	475	1E-133	>gi 14423842 sp O50001 PRU1_PRUAR Major allergen Pru ar 1	164	2,00E-39
Fernandez et al., 2004	[gi:50979882]CA-H12_03_B12 Coffee SSH library 1	1073	0	>gi 1345787 sp P48387 CHS2_CAMSI CHALCONE SYNTHASE 2 (NARINGENIN-CHALCONE SYNTHASE 2)	470	1,00E-131
Fernandez et al., 2004	[gi:50979881]CA-H12_04_G05 Coffee SSH library 1	898	0	>gi 12229619 sp Q9ZRS4 CHSY_CATRO Chalcone synthase (Naringenin-chalcone synthase)	187	2,00E-46
Fernandez et al., 2004	[gi:50979757]CA-H12_03_F03 Coffee SSH library 1	733	0	>gi 7433144 pir T07101 lipoxygenase (EC 1.13.11.12) - potato	293	5,00E-59
Fernandez et al., 2004	[gi:50980154]CA-H11_01_G11 Coffee SSH library 1	725	0	>gi 25044823 gb AAM28275.1 germin-like protein [Ananas comosus]	107	2,00E-28

Tabela 1.38: continuação

Fonte	Sequência	Score	e-value	Blast EST café - NR	score	e-value
Fernandez et al., 2004	[gi:50980031]CA-H11_02_H11 Coffee SSH library 1	483	1E-135	>gi 15221473 ref NP_174357.1 FAD-linked oxidoreductase family [Arabidopsis thaliana]	231	1,00E-59
Fernandez et al., 2004	[gi:50980113] CA-H11_01_B08 Coffee SSH library 1	831	0	>gi 16518974 gb AAL25088.1 Tobacco mosaic virus helicase domain-binding protein [Nicotiana tabacum]	301	1,00E-80
Fernandez et al., 2004	[gi:50979987]CA-H11_02_E07 Coffee SSH library 1	475	1E-133	>gi 15241939 ref NP_201080.1 hypersensitive-induced response protein [Arabidopsis thaliana]	186	3,00E-46
Fernandez et al., 2004	[gi:50979913]CA-H12_02_A12 Coffee SSH library 1 gi 50980275 gb CF589193.1 CF589193 DSS23 Coffee SSH library 2	890	0	>gi 7438249 pir T03803 tumor-related protein, clone NF34 - common tobacco	146	5,00E-34
Ganesh et al., 2006	[gi:50980283]gb CO773976.1 CO773976 DSS12 Coffee SSH library 2	1027	0	ref NP_196762.1 unknown protein [Arabidopsis thaliana]	378	2,00E-103
Ganesh et al., 2006	[gi:50980281]gb CO773974.1 CO773974 DSS16 Coffee SSH library 2	938	0	ref NP_188696.1 NDR1 (NON RACE-SPECIFIC DISEASE RESISTANCE 1); signal transducer [Arabidopsis thaliana]	135	2,00E-30
Ganesh et al., 2006	[gi:50980282]gb CO773974.1 CO773974 DSS16 Coffee SSH library 2	642	0	emb CAN70362.1 hypothetical protein [Vitis vinifera]	197	3,00E-49
Chen et al., 2003	[gi:50979697]gb CF588617.1 CF588617 CA-H12_01_B07 Coffee SSH library 1	494	1E-139	gb AAF70460.1 AF240445_1 polyubiquitin [Populus tremula x Populus tremuloides]	130	2,00E-29
Chen et al., 2003	[gi:50980232]gb CF589150.1 CF589150 CA-H11_03_A12 Coffee SSH library 1	564	1E-159	emb CAJ43737.1 class III chitinase [Coffea arabica]	166	8,00E-40
Chen et al., 2003	[gi:50979788]gb CF588708.1 CF588708 CA-H12_02_B01 Coffee SSH library 1	862	0	emb CAJ43737.1 class III chitinase [Coffea arabica]	201	4,00E-50
Chen et al., 2003	[gi:40786374]dbj AB158557.1 Coffea arabica ca-pox1 mRNA for peroxidase	665	0	dbj BAF44533.1 class IV chitinase [Nicotiana tabacum]	409	1,00E-112
Chen et al., 2003	[gi:50980267]gb CF589185.1 CF589185 DSS11 Coffee SSH library 2	808	0	gb AAK52084.1 peroxidase [Nicotiana tabacum]	409	2,00E-112
Chen et al., 2003	[gi:50980267]gb CF589185.1 CF589185 DSS11 Coffee SSH library 2	815	0	ref NP_187864.1 HSP70 (heat shock protein 70); ATP binding [Arabidopsis thaliana]	418	2,00E-115

Com as três estratégias de mineração foi possível identificar 14.060 ESTs potencialmente associadas à resistência do cafeeiro a doenças (Tabela 1.39).

Tabela 1.39: Número de ESTs mineradas por cada estratégia.

Estratégias	Nº ESTs
<i>Electronic Northern</i>	2.206
Palavras-Chave	11.793
Sequências Publicadas	61
Total	14.060

4. CONCLUSÃO

Neste trabalho foram mineradas um total de 14.060 ESTs potencialmente associadas com a resistência do cafeeiro a doenças.

Essas ESTs são fonte de informações úteis para a ampliação do conhecimento sobre a resistência do cafeeiro a patógenos e pragas. Recomenda-se utilizar essas ESTs como genes candidatos para o desenvolvimento de marcadores moleculares para assistirem programas de melhoramento.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- Aarts, N., Metz, M., Holub, E., Staskawicz, B. J., Daniels, M. J., Parker, J. E. Different requirements for EDS1 and NDR1 by disease resistance genes define at least two R gene-mediated signaling pathways in Arabidopsis Proceedings of the National Academy of Sciences of the United States of America, v.95, n.17, p.10306-10311. 1998.
- Aguiar, A. T. E. Atributos químicos de espécies de café. Fitotecnia, ESALQ-USP, Piracicaba, 2005. 88 p.
- Altschul, S. F., Boguski, M. S., Gish, W., Wootton, J. C. Issues in searching molecular sequence databases. Nature Genetics, v.6, p.119-129. 1994.
- Anderson, I., Brass, B. Searching DNA databases for similarities to DNA sequences: when is a match significant? Bioinformatics, v.14, n.4, p.349-356. 1998.
- Aravind, L. D., V. M, Koonin, E. V. The domains of death: evolution of the apoptosis machinery. Trends in Biochemical Sciences, v.24, n.2, p.47-53. 1999.
- Boller, T. Extracellular location of chitinases in cucumber. Physiological and Molecular Plant Pathology, v.33, p.11-16. 1988.
- Botella, M. A., Parker, J. E., Frost, L. N., Bittner-Eddy, P. D., Beynon, J. L., Daniels, M. J., Holub, E. B., Jones J. D. G. Three genes of the arabidopsis RPP1 complex resistance locus recognize distinct *Peronospora parasitica* avirulence determinants Plant Cell, v.10, p.1847-1860. 1998.
- Chen, Z. J., Ribeiro, A., Silva, M. C., Santos, P., Guerra-Guimarães, L., Gouveia, M., Fernandez, D., Rodrigues Júnior, C. J. Heat shock-induced susceptibility of green coffee leaves and berries to *Colletotrichum gloeosporioides* and its association to PR and hsp70 gene expression.

- Physiological and Molecular Plant Pathology, v.63, n.4, p.181-190. 2003.
- Chong, J., Balyz, R., Schitt, C., Beffa, R., Fritig, B., Saindrenan, P. Downregulation of a pathogen-responsive tobacco UDP-Glc: phenylpropanoid glucosyltransferase reduces scopoletin glucoside accumulation, enhances oxidative stress, and weakens virus resistance. Plant Cell, v.14, p.1093-1107. 2002.
- Dong, Q., Schlueter, S. D., Brendel, V. PlantGDB, plant genome database and analysis tools. Nucleic Acids Research, v.32, p.D354-D359. 2004.
- Dreher, K., Callis, J. Ubiquitin, Hormones and Biotic Stress in Plants. Annals of Botany, v.99, p.787-822. 2007.
- Ellis, J. G., Lawrence, G. J., Luck, J. E., Dodds, P. N. Identification of regions in alleles of the flax rust resistance gene L that determine differences in gene-for-gene specificity Plant Cell, v.11, n.3, p.495-506. 1999.
- Fernandez, D., Santos, P., Agostini, C., Bon, M., Petitot, A., Silva, M. C., Guerra-Guimarães, L., Ribeiro, A., Argout, X., Nicole, M. Coffee (*Coffea arabica* L.) genes early expressed during infection by the rust fungus (*Hemileia vastatrix*). Molecular Plant Pathology, v.5, n.6, p.527-536. 2004.
- Feys, B. J., Parker, J. E. Interplay of signaling pathways in plant disease resistance. Trends in Genetics, v.16, n.10, p.449-455. 2000.
- Gai, X. W., Lal, S., Xing, L.Q., Brendel, V., and Walbot, V. Gene discovery using the maize genome database ZmDB. Nucleic Acids Research, v.28, p.94-96. 2000.
- Ganesh, D., Petitot, a-S., Silva, M. C., Alary, R., Lecouls, a-C., Fernandez, D. Monitoring of the early molecular resistance responses of coffee (*Coffea arabica* L.) to the rust fungus (*Hemileia vastatrix*) using real-time quantitative RT-PCR. Plant Science, v.170, p.1045-1051. 2006.
- Grenier, J., Potvin, C., Asselin, A. Barley pathogenesis-related proteins with fungal cell wall lytic activity inhibit the growth of yeasts. Plant Physiology, v.103, p.1277-1283. 1993.
- Hermsmeier, D., Schittko, U., Baldwin, I. T. Molecular interactions between the specialist herbivore *Manduca sexta* (Lepidoptera, Sphingidae) and its natural host *Nicotiana attenuata*. I. Large-scale changes in the accumulation of growth- and defense-related plant mRNAs. Plant Physiology, v.125, p.683-700. 2001.
- Höfte, H., Desprez, T., Amselem, J., Chiapello, H., Caboche, M., Moisan, A., Jourjon, M. F., Charpentreau, J. L., Berthomieu, P., Guerrier, D., Giraudat, J., Quigley, F., Thomas, F., Yu, D. Y., Mache, R., Raynal, M., Cooke, R., Grellet, F., Delseny, M., Parmentier, Y., Marcillac, G., Gigot, C., Fleck, J., Philipps, G., Axelos, M., Bardet, C., Tremousaygue, D., Lescure, B. An

- inventory of 1152 expressed sequence tags obtained by partial sequencing of cDNAs from *Arabidopsis thaliana*. Plant Journal, v.4, p.1051-1061. 1993.
- Jach, G., Gornhardt, B., Mundy, J., Logemann, J., Pinsdorf, E., Leah, R., Schell, J., Maas, C. Enhanced quantitative resistance against fungal disease by combinatorial expression of different barley antifungal proteins in transgenic tobacco. Plant Journal, v.8, n.1, p.97-109. 1995.
- Jackson, A. O., Taylor, C. B. Plant-Microbe Interactions: Life and Death at the Interface. The Plant Cell, v.8, p.1651-1668. 1996.
- Kobe, B., Deisenhofer, J. Proteins with leucine-rich repeats. Current Opinion in Structure Biology, v.8, n.5, p.409-416. 1995.
- Lazo, G. R., Chao, S., Hummel, D. D., Edwards, H., Crossman, C. C., Lui, N., Matthews, D. E., Carollo, V. L., Hane, D. L., You, F. M., Butler, G. E., Miller, R. E., Close, T. J., Peng, J. H., Lapitan, N. L. V., Gustafson, J. P., Qi, L. L., Echalié, B., Gill, B. S., Dilbirli, M., Randhawa, H. S., Gill, K. S., Greene, R. A., Sorrells, M. E., Akhunov, E. D., Dvorak, J., Linkiewicz, A. M., Dubcovsky, J., Hossain, K. G., Kalavacharla, V., Kianian, S. F., Mahmoud, A. A., Miftahudin, Ma, X. F., Conley, E. J., Anderson, J. A., Pathan, M. S., Nguyen, H. T., McGuire, P. E., Qualset, C. O., Anderson, O. D. Development of an Expressed Sequence Tag (EST) Resource for Wheat (*Triticum aestivum* L.): EST Generation, Unigene Analysis, Probe Selection and Bioinformatics for a 16,000-Locus Bin-Delineated Map. Genetics, v.168, p.585-593. 2004.
- Li, L., Hong, R., Hastings, W. Three functional luciferase domains in a single polypeptide chain. Proceedings of the National Academy of Sciences of the United States of America, v.94, n.17, p.8954-8958. 1997.
- Lin, C., Mueller, L. A., Carthy, J. M., Crouzillat, D., Pétiard, V., Tanksley, S. D. Coffee and tomato share common gene repertoires as revealed by deep sequencing of seed and cherry transcripts. Theoretical and Applied Genetics, v.112, n.1, p.114-130. 2005.
- Mcdowell, J. M., Dhandaydham, M., Long, T. A., Aarts, M. G. M., Goff, S., Holub, E. B., Dangl, J. L. Intragenic recombination and diversifying selection contribute to the evolution of downy mildew resistance at the RPP8 locus of *Arabidopsis*. Plant Cell, v.10, n.11, p.1861-1874. 1998.
- Melo, G. A., Shimizu, M. M., Mazzafera, P. Polyphenoloxidase activity in coffee leaves and its role in resistance against the coffee leaf miner and coffee leaf rust. Phytochemistry, v.67, n.3, p.277-285. 2006.
- Parniske, M., Hammond-Kosack, K. E., Golstein, C., Thomas, C. M., Jones, D. A., Harrison, K., Wulff, B. B. H., Jones, J. D. G. Novel disease resistance specificities result from sequence exchange between tandemly repeated genes at the Cf-4/9 locus of tomato. Cell, v.91, n.6, p.821-832. 1997.

- Pascholati, S. F., Leite, B. Mecanismos bioquímicos de resistência às doenças. Revisão Anual de Patologia de Plantas, v.2, p.1-51. 1994.
- Qi, X., Bakht, S., Qin, B., Leggett, M., Hemmings, A., Mellon, F., Eagles, J., Werck-Reichhart, D., Schaller, H., Lesot, A., Melton, R., Osbourn, A. A different function for a member of an ancient and highly conserved cytochrome P450 family: From essential sterols to plant defense. Proceedings of the National Academy of Sciences of the United States of America, v.103, n.49, p.18848-18853. 2006.
- Roberts, W., Selitrennikoff, C. P. Zeamatin, an antifungal protein from maize with membrane-permeabilizing activity. Journal of Genetical Microbiology, v.136, p.1771-1778. 1990.
- Ronning, C. M., Stegalkina, S. S., Ascenzi, R. A., Bougri, O., Hart, A. L., Utterbach, T. R., Vanaken, S. E., Riedmuller, S. B., White, J. A., Cho, J., Perte, G. M., Lee, Y., Karamycheva, S., Sultana, R., Tsai, J., Quackenbush, J., Griffiths, H. M., Restrepo, S., Smart, C. D., Fry, W. E., Van-Der-Hoeven, R., Tanksley, S., Zhang, P., Jin, H., Yamamoto, M. L., Baker, B. J., Buell, C. R. Comparative Analyses of Potato Expressed Sequence Tag Libraries. Plant Physiology, v.131, p.419-429. 2003.
- Ryder, T. B., Cramer, C. L., Bell, J. N., Robbins, M. P., Dixon, R. A., Lamb, C. J. Elicitor rapidly induces chalcone synthase mRNA in *Phaseolus vulgaris* cells at the onset of the phytoalexin defense response. Proceedings of the National Academy of Sciences of the United States of America, v.81, p.5724-5728. 1984.
- Sahai, A. S., Manocha, M. S. Chitinases of plants and fungi: their involvement in morphogenesis and host-parasite interaction. FEMS Microbiology Reviews v.11, p.317-338. 1993.
- Saraste, M., Sibbald, P. R., Wittinghofer, A. The P-loop--a common motif in ATP- and GTP-binding proteins. Trends in Biochemical Sciences, v.15, n.11, p.430-434. 1990.
- Selitrennikoff, C. P. Antifungal proteins. Applied and Environmental Microbiology, v.67, n.7, p.2883-2894. 2001.
- Shoemaker, R., Keim, P., Vodkin, L., Retzel, E., Clifton, S. W., Waterston, R., Smoller, D., Coryell, V., Khanna, A., Erpelding, J., Gai, X., Brendel, V., Raph-Schmidt, C., Shoop, Vielweber, C. J., Schmatz, M., Pape, D., Theising, B., Martin, J., Dante, M., Wylie, T., Granger, C. A Compilation of Soybean ESTs: Generation and Analysis. Genome, v.45, p.329-338. 2002.
- Stintzi, A., Heitz, T., Prasad, V., Wiedemann-Merdinoglu, S., Kauffmann, S., Geoffroy, P., Legrand, M., Fritig, B. Plant 'pathogenesis-related' proteins and their role in defense against pathogens. Biochimie, v.75, n.8, p.687-706. 1993.

- Stoessl, A. Secondary plant metabolites in plant disease resistance. Part II: Phytoalexins. Fitopatologia Brasileira, v.11, p.25-53. 1986.
- Takemoto, D., Hayashi, M., Doke, N., Nishimura, M., Kawakita, K. Molecular cloning of a defense-response-related cytochrome p450 gene from tobacco. Plant and Cell Physiology, v.40, n.12, p.1232-1242. 1999.
- Thomas, C. M., Jones, D. A., Parniske, M., Harrison, K., Balint-Kurti, P. J., Hatzixanthis, K., Jones, J. D. G. Characterization of the tomato Cf-4 gene for resistance to *Cladosporium fulvum* identifies sequences that determine recognitional specificity in Cf-4 and Cf-9. Plant Cell, v.9, n.12, p.2209-2224. 1997.
- Traut, T. W. The functions and consensus motifs of nine types of peptide segments that form different types of nucleotide-binding sites. European Journal of Biochemistry, v.229, p.9-19. 1994.
- Trudel, J., Grenier, J., Potvin, C., Asselin, A. Several thaumatin-like proteins bind to β -1,3-glucans Plant Physiology, v.118, p.1431-1438. 1998.
- Udall, J. A., Swanson, J. M., Haller, K., Rapp, R. A., Sparks, M. E., Hatfield, J., Yu, Y., Wu, Y., Dowd, C., Arpat, A. B., Sickler, B. A., Wilkins, T. A., Guo, J. Y., Chen, X. Y., Scheffler, J., Taliercio, E., Turley, R., Mcfadden, H., Payton, P., Klueva, N., Allen, R., Zhang, D., Haigler, C., Wilkerson, C., Suo, J., Schulze, S. R., Pierce, M. L., Essenberg, M., Kim, H., Llewellyn, D. J., Dennis, E. S., Kudrna, D., Wing, R., Paterson, A. H., Soderlund, C., Wendel, J. F. A global assembly of cotton ESTs. Genome Research, v.16, p.441-450. 2006.
- Van Der Biezen, E. A., Jones, J. D. The NB-ARC domain: a novel signaling motif shared by plant resistance gene products and regulators of cell death in animals. Current Biology, v.8, n.7, Mar, p.R226-227. 1998.
- Van Der Biezen, E. A., Sun, J., Coleman, M. J., Bibb, M. J., Jones, J. D. G. Arabidopsis RelA/SpoT homologs implicate (p)ppGpp in plant signaling Proceedings of the National Academy of Sciences of the United States of America, v.97, n.7, p.3747-3752. 2000.
- Vieira, L. G. E., Andrade, A. C., Colombo, C. A., Moraes, A. H. A., Metha, A., Oliveira, A. C., Labate, C. A., Marino, C. L. e C. B. Monteiro-Vitorello, Monte, D. C., Giglioti, E., Kimura, E. T., Romano, E., Kuramae, E. E., Lemos, E. G. M., Almeida, E. R. P., Jorge, E. C., Albuquerque, E. V. S., Silva, F. R., Vinecky, F., Sawazaki, H. E., Dorry, H. F. A., Carrer, H., Abreu, I. N., Batista, J. A. N., Teixeira, J. B., Kitajima, J. P., Xavier, K. G., Lima, L. M., Camargo, L. E. A., Pereira, L. F. P., Coutinho, L. H., Lemos, M. V. F., Romano, M. R., Machado, M. A., Costa, M. M. C., Sá, M. F. G., Goldman, M. H. S., Ferro, M. I. T., Tinoco, M. L. P., Oliveira, M. V., Sluys, M. V., Shimizu, M. M., Maluf, M. P., Eira, M. T. S., Filho, O. G., Arruda, P., Mazzafera, P., Mariani, P. D. S. C., Oliveira, R. L. B. C., Harakava, R., Balbao, S. F., Tsai, S. M., Mauro, S. M. Z., Santos, S. N., Siqueira, W. J.,

- Costa, G. G. L., Formighieri, E. F., Carazzolle, M. F., Pereira, G. A. G. Brazilian coffee genome project: an EST-based genomic resource. Brazilian Journal of Plant Physiology, v.18, n.1, p.95-108. 2006.
- Wang, G., Ruan, D., Song, W., Sideris, S., Chen, L., Pi, L., Zhang, S., Zhang, Z., Fauquet, C., Gaut, B. S., Whalen, M. C., Ronald, P. C. Xa21D encodes a receptor-like molecule with a leucine-rich repeat domain that determines race-specific recognition and is subject to adaptive evolution Plant Cell, v.10, n.5, p.765-779. 1998.
- Wolfsberg, T. G., Landsman, D. A comparison of expressed sequence tags (ESTs) to human genomic sequences. Nucleic Acids Research, v.25, p.1626-1632. 1997.
- Yamamoto, K., Sasaki, T. L. Large-scale EST sequencing in rice. Plant Molecular Biology, v.35, p.135-144. 1997.
- Yun, D., Zhao, Y., Pardo, J. M., Narasimhan, M. L., Damsz, B., Lee, H., Abad, L. R., D'urzo, M. P., Hasegawa, P. M., Bressan, R. A. Stress proteins on the yeast cell surface determine resistance to osmotin, a plant antifungal protein. Proceedings of the National Academy of Sciences of the United States of America, v.94, p.7082. 1997.
- Yun, D. J., Ibeas, J. I., Lee, H., Coca, M. A., Narasimhan, M. L., Uesono, Y., Hasegawa, P. M., Pardo, J. M., Bressan, R. A. Osmotin, a plant antifungal protein, subverts signal transduction to enhance fungal cell susceptibility. Molecular Cell, v.1, n.6, p.807-817. 1998.
- Zhou, N., Tootle, T. L., Glazebrook, J. Arabidopsis PAD3, a Gene Required for Camalexin Biosynthesis, Encodes a Putative Cytochrome P450 Monooxygenase. The Plant Cell, v.11, p.2419-2428. 1999.

CAPÍTULO 2

MARCADORES MOLECULARES DERIVADOS DE SEQÜÊNCIAS EXPRESSAS DO GENOMA DO CAFÉ POTENCIALMENTE ENVOLVIDAS NA RESISTÊNCIA À FERRUGEM

1. INTRODUÇÃO

A principal doença do cafeeiro é a ferrugem alaranjada, causada pelo fungo *Hemileia vastatrix* Berk & Br. Esta doença ocorre em todas as regiões produtoras de café do mundo. Assim, o estudo e a caracterização de fatores genéticos de resistência a este fungo são importantes para ampliar os conhecimentos da interação planta-patógeno. Esse estudo poderá ser útil para o desenvolvimento de estratégias de controle para a ferrugem, incluindo o melhoramento genético visando a obtenção de variedades resistentes.

O seqüenciamento do genoma de plantas tem facilitado e acelerado a identificação de genes desejáveis, possibilitando a sua manipulação subsequente por meio de técnicas de genética molecular. A biotecnologia do cafeeiro tem sido potencializada com o Projeto Brasileiro do Genoma Café (PBGC). Neste projeto, realizou-se o seqüenciamento do genoma funcional de *Coffea arabica*, *C. canephora* e *C. racemosa*, resultando na disponibilização de um banco de dados de mais de 200.000 ESTs (*Expressed Sequence Tags*) (Vieira et al., 2006). Os clones seqüenciados foram provenientes de 37 bibliotecas de cDNA de café. Essas ESTs, portanto, representam milhares de genes expressos em diferentes órgãos do cafeeiro, obtidos em vários estágios de desenvolvimento da planta e submetidos a condições de estresse biótico e abiótico.

Com o intuito de identificar genes relacionados com o mecanismo de defesa do cafeeiro a doenças, o banco de dados gerado pelo PBGC foi minerado por meio de análise *in silico*, utilizando-se ferramentas de bioinformática. A mineração resultou na identificação de 14.060 seqüências potencialmente envolvidas com a resistência do cafeeiro a doenças.

Para verificar se os genes ou ESTs identificados estão envolvidos ou ligados à resistência do cafeeiro à ferrugem, *primers* que os amplificam foram desenhados e testados em diferentes genótipos de café.

2. MATERIAL E MÉTODOS

2.1. Material Vegetal e Extração de DNA

Para a avaliação dos marcadores moleculares, foram utilizados 24 genótipos de cafeeiro do Banco de Germoplasma da UFV/EPAMIG. Desses, 12 foram de cafeeiros resistentes a *H. vastatrix* e 12 de cafeeiros susceptíveis a esse fungo. Os genótipos susceptíveis foram: UFV-570 (“Bourbon Amarelo”), UFV 2945 (‘Típica’), UFV 557-06 (*C. racemosa* x *C. arabica*), UFV 2145-79 (“Catuaí Vermelho”), UFV 2154-345 (“Catuaí Amarelo”), UFV 2143-193 (“Catuaí Amarelo”), UFV 2143-235 (“Catuaí Amarelo”), UFV 2148-57 (“Catuaí Amarelo”), UFV 2164-193 (“Mundo Novo”), UFV 2190-100 (“Mundo Novo”), Caturra 812 (“Caturra Amarelo”) e *C. excelsa*. Os genótipos resistentes foram os acessos de Híbrido de Timor: UFV 440-22, UFV 443-3, UFV 445-46, CIFC 832-1, CIFC 832-2, UFV 438-52, UFV 446-08, CIFC 33-1, CIFC 1343-269, CIFC 4106, CIFC 420-10 e CIFC 147-1 (Tabela 2.1).

Tabela 2.1: Genótipos utilizados para o teste dos *primers* obtidos das ESTs potencialmente envolvidas com a resistência do cafeeiro a ferrugem.

Indivíduos susceptíveis a ferrugem			Indivíduos resistentes a ferrugem		
Código	Genótipo	Descrição	Código	Genótipo	Descrição
1	UFV 570	Bourbon Amarelo da China	13	UFV HT 440-22	Híbrido de Timor UFV
2	UFV 2945	Típica C-117	14	UFV HT 443-3	Híbrido de Timor UFV
3	UFV 557-06	<i>Coffea racemosa</i> (triplóide)	15	UFV HT 445-46	Híbrido de Timor UFV
4	UFV 2145-79	Catuaí Vermelho	16	UFV HT 832-1	Híbrido de Timor UFV
5	UFV 2154-345	Catuaí Amarelo	17	UFV HT 832-2	Híbrido de Timor UFV
6	Catuaí 2143-193	Catuaí Amarelo	18	UFV HT 438-52	Híbrido de Timor UFV
7	Catuaí 2143-235	Catuaí Amarelo	19	UFV HT 446-08	Híbrido de Timor UFV
8	Catuaí 2148-57	Catuaí Amarelo	20	CIFC 33-1	Híbrido de Timor CIFC
9	Mundo Novo 2164-193	Mundo Novo UFV	21	CIFC 1343-269	Híbrido de Timor CIFC
10	Mundo Novo 2190-100	Mundo Novo UFV	22	CIFC 4106	Híbrido de Timor CIFC
11	Caturra 812	Caturra Amarelo	23	CIFC 420-10	Híbrido de Timor CIFC
12	<i>Coffea excelsa</i>	<i>Coffea excelsa</i>	24	CIFC 147-1	Híbrido de Timor CIFC

Folhas jovens de cada um desses cafeeiros foram coletadas, armazenadas em ultrafreezer -80°C e liofilizadas. O DNA foi extraído seguindo-se o protocolo de Diniz et al. (2005). Após a extração, o DNA foi quantificado em espectrofotômetro e armazenado a 4°C. Para a amplificação, foi diluído em TE (Tris HCl 10 mM, EDTA 1 mM, pH 8,0) para uma concentração final de 10 ng/μl.

2.2. Seqüências do Banco de Dados do PBGC

No processo de mineração do banco de dados do PBGC, foram identificadas 14.060 seqüências potencialmente relacionadas com a resistência do cafeeiro a doenças. Estas seqüências foram classificadas em relação a maior ou menor possibilidade de estarem envolvidas com o processo de resistência do cafeeiro a patógenos. A classificação foi determinada a partir da análise da anotação automática de cada seqüência. Os fatores considerados foram: organismo, função potencial da seqüência, *e-value* menor ou igual a 1×10^{-10} , *score* igual ou maior que 100 e presença de ESTs provenientes da biblioteca RM1 nos *contigs*. A partir dessa classificação, 107 seqüências foram selecionadas e *primers* que as amplificam foram desenhados.

2.3. Desenvolvimento de Marcadores Moleculares

Para desenho dos *primers*, foi utilizado o programa computacional *Primer3* (<http://frodo.wi.mit.edu/>). As condições para o desenho foram: (1) produto de PCR com tamanho de 200 a 600pb; (2) tamanho do *primer* de 18 a 24pb; (3) *tm* de 55 a 60°C; e, (4) percentagem de CG de 45 a 55. Para os outros parâmetros, foram utilizados os valores pré-definidos pelo programa. A estabilidade destes *primers* foi verificada utilizando o programa *PrimerSelect* 6.1 (DNASTAR, Madison, WI, EUA).

A reação de PCR foi realizada em 20,0µL de solução contendo 30ng de DNA, 0,8 unidades de *Taq* DNA polimerase, tampão 1x, 1mM de $MgCl_2$, 60 µM de cada dNTP e 0,7 µM de cada *primer*. Para a amplificação, utilizou-se o procedimento *touchdown* PCR. Este procedimento consistiu em uma etapa de desnaturação inicial de 94°C, por três minutos, seguido de cinco ciclos com uma etapa de desnaturação a 94°C por 30 segundos, uma etapa de anelamento por 20 segundos e extensão a 72°C por 40 segundos. A temperatura de anelamento foi de 65°C a 60°C, reduzindo 1°C a cada ciclo. Após os cinco ciclos, procederam-se mais 30 ciclos com desnaturação a 94°C por 60 segundos, anelamento de 60°C por 20 segundos e extensão de 72°C por 40 segundos.

Os produtos das reações de amplificação foram submetidos a eletroforese em géis de agarose 1,2%, corados com brometo de etídio, visualizados em UV e foto documentados em transiluminador.

3. RESULTADOS E DISCUSSÃO

Das 14.060 seqüências potencialmente envolvidas com a resistência do cafeeiro a doenças, 107 foram selecionadas para o desenho de 40 *primers* (Tabela 2.2). Os *primers* denominados de CARF 001 a CARF 040 (Tabela 2.2) foram estáveis dentro das condições estabelecidas.

Tabela 2.2: *Primers* obtidos a partir das seqüências mineradas.

Primer	Mineração	BLAST	Score	e-value
CARF 001	Northern eletrônico RM1 77,69 %	gi 24459841 emb CAC82597.1 disease resistance-like protein [Coffea arabica]	259	2,00E-67
CARF 002	Northern eletrônico RM1 100%	gi 15225780 ref NP_180241.1 leucine-rich repeat transmembrane protein kinase, putative [Arabidopsis thaliana]	296	1,00E-78
CARF 003	Northern eletrônico RM1 100%	gi 42602161 gb AAS21681.1 receptor-like kinase [Arabidopsis thaliana]	102	1,00E-20
CARF 004	Northern eletrônico RM1 100%	gi 15226721 ref NP_179220.1 leucine-rich repeat transmembrane protein kinase, putative	131	4,00E-13
CARF 005	Projeto Palavra Chave LRR	gi 24459841 emb CAC82597.1 disease resistance-like protein [Coffea arabica]	260	2,00E-68
CARF 006	Projeto Palavra Chave NBS	gi 30689664 ref NP_195056.2 disease resistance protein (CC-NBS-LRR class), putative [Arabidopsis thaliana]	652	0,0
CARF 007	Projeto Palavra Chave CHITINASE	gi 4835586 dbj BAA77677.1 acidic chitinase [Glycine max]	268	2,00E-70
CARF 008	Lin et al., 2005 SGN120244	gi 34907232 ref NP_914963.1 putative disease resistance protein l2 [Oryza sativa (japonica cultivar-group)]	108	1,00E-16
CARF 009	Lin et al., 2005 SGN124466	gi 15225286 ref NP_180201.1 receptor-related protein kinase, ERECTA [Arabidopsis thaliana]	180	2,00E-60
CARF 010	Fernandez et al., 2004 CA-H12_04_E07	gi 12331298 emb CAC24711.1 cytochrome P450 [Solanum tuberosum]	208	1,00E-52
CARF 011	Fernandez et al., 2004 CA-H11_02_D06	gi 24459849 emb CAC82600.1 disease resistance-like protein [Coffea canephora]	164	2,00E-39
CARF 012	Projeto Palavra Chave LRR	gi 30689664 ref NP_195056.2 disease resistance protein (CC-NBS-LRR class), putative (Arabidopsis thaliana)	512	2,00E-115
CARF 013	Projeto Palavra Chave LRR	gi 50899184 ref XP_450380.1 putative disease related protein 2 (Oryza sativa)	138	1,00E-65
CARF 014	Projeto Palavra Chave LRR	gi 28555894 emb CAD45029.1 NBS-LRR disease resistance protein homologue (Hordeum vulgare)	109	3,00E-11
CARF 015	Lin et al., 2005 119817	gi 5880843 gb AAD54934.1 chitinase precursor [Petroselinum crispum]	243	1,00E-111
CARF 016	Lin et al., 2005 119672	gi 34484310 gb AAQ72788.1 hypersensitive-induced response protein [Cucumis sativus]	266	4,00E-70
CARF 017	Fernandez et al., 2004 DSS14	gi 116328 sp P17541 CHIA_CUCSA ACIDIC ENDOCHITINASE PRECURSOR	162	7,00E-39
CARF 018	Fernandez et al., 2004 CA-H11_02_B09	gi 12331298 emb CAC24711.1 cytochrome P450 [Solanum tuberosum]	209	6,00E-53
CARF 019	Fernandez et al., 2004 CA-H11_02_B04	gi 16974114 emb CAC95155.1 putative resistance protein [Lycopersicon esculentum]	253	2,00E-66
CARF 020	Northern eletrônico RM1 100%	gi 21617980 gb AAM67030.1 putative RING zinc finger protein [Arabidopsis thaliana]	106	6,00E-15
CARF 021	Northern eletrônico RM1 100%	gi 30687274 ref NP_850285.1 protein kinase family [Arabidopsis thaliana]	128	1,00E-18
CARF 022	Northern eletrônico RM1 100%	gi 18855060 gb AAL79752.1 AC096687_16 putative protein kinase [Oryza sativa]	126	2,00E-29
CARF 023	Northern eletrônico RM1 100%	gi 4240031 dbj BAA74802.1 DNA binding zinc finger protein (Pspzf) [Pisum sativum]	112	3,00E-15
CARF 024	Northern eletrônico RM1 100%	gi 31324528 gb AAP47579.1 receptor kinase Lecrk [Gossypium hirsutum]	257	2,00E-67
CARF 025	Northern eletrônico RM1 100%	gi 11358841 pir serine/threonine-protein kinase-like protein - Arabidopsis thaliana	115	1,00E-11
CARF 026	Northern eletrônico RM1 100%	gb ABO78298.1 Glycosyl transferase, family 48 [Medicago truncatula]	380	4,00E-104
CARF 027	Northern eletrônico RM1 100%	gi 15221833 ref NP_173302.1 protein kinase -related [Arabidopsis thaliana]	269	3,00E-71

Tabela 2.2: continuação

Primer	Mineração	BLAST	Score	e-value
CARF 028	Northern eletrônico RM1 100%	gi 15231381 ref NP_187364.1 membrane protein [Arabidopsis thaliana]	141	2,00E-32
CARF 029	Northern eletrônico RM1 100%	gi 29124139 gb AAO65880.1 putative DNA-binding protein [Oryza sativa (japonica cultivar-group)]	141	5,00E-32
CARF 030	Northern eletrônico RM1 100%	gi 2792155 emb CAA11226.1 chalcone reductase [Sesbania rostrata]	105	7,00E-52
CARF 031	Northern eletrônico RM1 100%	gi 30683822 ref NP_850115.1 protein kinase family [Arabidopsis thaliana]	179	2,00E-49
CARF 032	Northern eletrônico RM1 100%	gi 30683822 ref NP_850115.1 protein kinase family [Arabidopsis thaliana]	241	3,00E-62
CARF 033	Northern eletrônico RM1 100%	ref NP_189017.1 leucine-rich repeat family protein / protein kinase family protein [Arabidopsis thaliana]	363	9,00E-99
CARF 034	Northern eletrônico RM1 100%	gi 92874399 gb ABE82675.1 Tyrosine protein kinase, active site [Medicago truncatula]	100	2,00E-19
CARF 035	Northern eletrônico RM1 100%	gi 30686115 ref NP_850618.1 GATA zinc finger protein [Arabidopsis thaliana]	105	3,00E-11
CARF 036	Projeto Palavra Chave LRR	gi 24459853 emb CAC82602.1 disease resistance-like protein (Coffea arabica)	142	5,00E-33
CARF 037	Projeto Palavra Chave LRR	gi 24459845 emb CAC82610.1 disease resistance-like protein (Coffea arabica)	161	2,00E-58
CARF 038	Lin et al., 2005 SGN121998	gi 7488989 pir disease resistance protein Prf - tomato	163	4,00E-39
CARF 039	Lin et al., 2005 SGN124574	gi 27817976 dbj BAC55740.1 putative disease resistance gene homolog [Oryza sativa]	142	7,00E-33
CARF 040	Lin et al., 2005 SGN123451	gi 3451147 emb CAA09110.1 chitinase [Hevea brasiliensis]	266	4,00E-70

Dos 40 pares de *primers* obtidos, 12 (30%) foram desenhados a partir de seqüências mineradas pela estratégia “Seqüências Publicadas”, oito (20%) pela estratégia “Palavras-Chave” e 20 (50%) pela estratégia “*Electronic Northern*” (Figura 2.1).

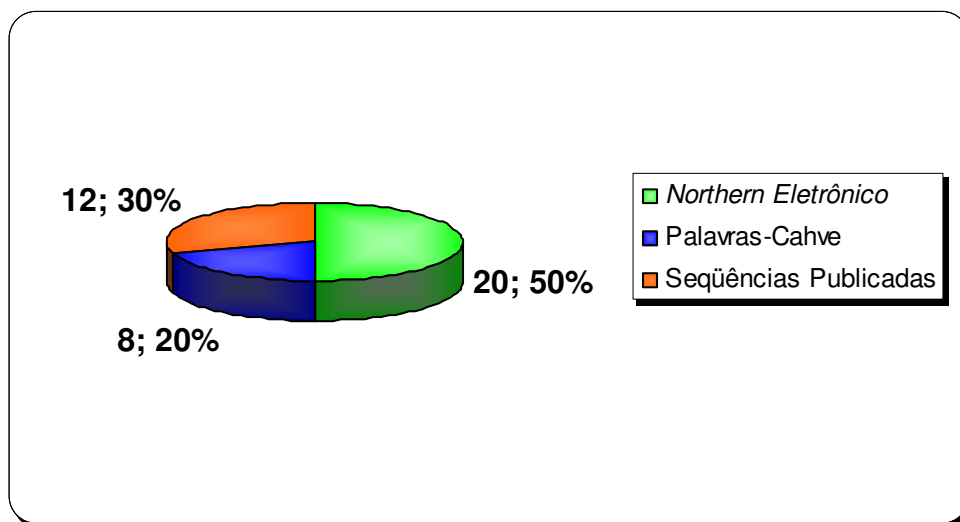


Figura 2.1: Número de pares de *primers* obtidos a partir das estratégias de mineração.

A estratégia “Seqüências Publicadas” foi a que, proporcionalmente, teve o melhor aproveitamento em relação ao número de seqüências selecionadas para desenho de *primers* (Figura 2.2). Das 61 seqüências obtidas por meio desta estratégia, 12 (19,67%) foram utilizadas para o desenho de *primers*.

Com a estratégia “Palavras-Chave” foram obtidas 11.793 ESTs, e somente 8 (0,06%) foram aproveitadas; Das 2.206 seqüências obtidas por meio da estratégia “*Electronic Northern*”, apenas 28 (0,90%) foram aproveitadas.

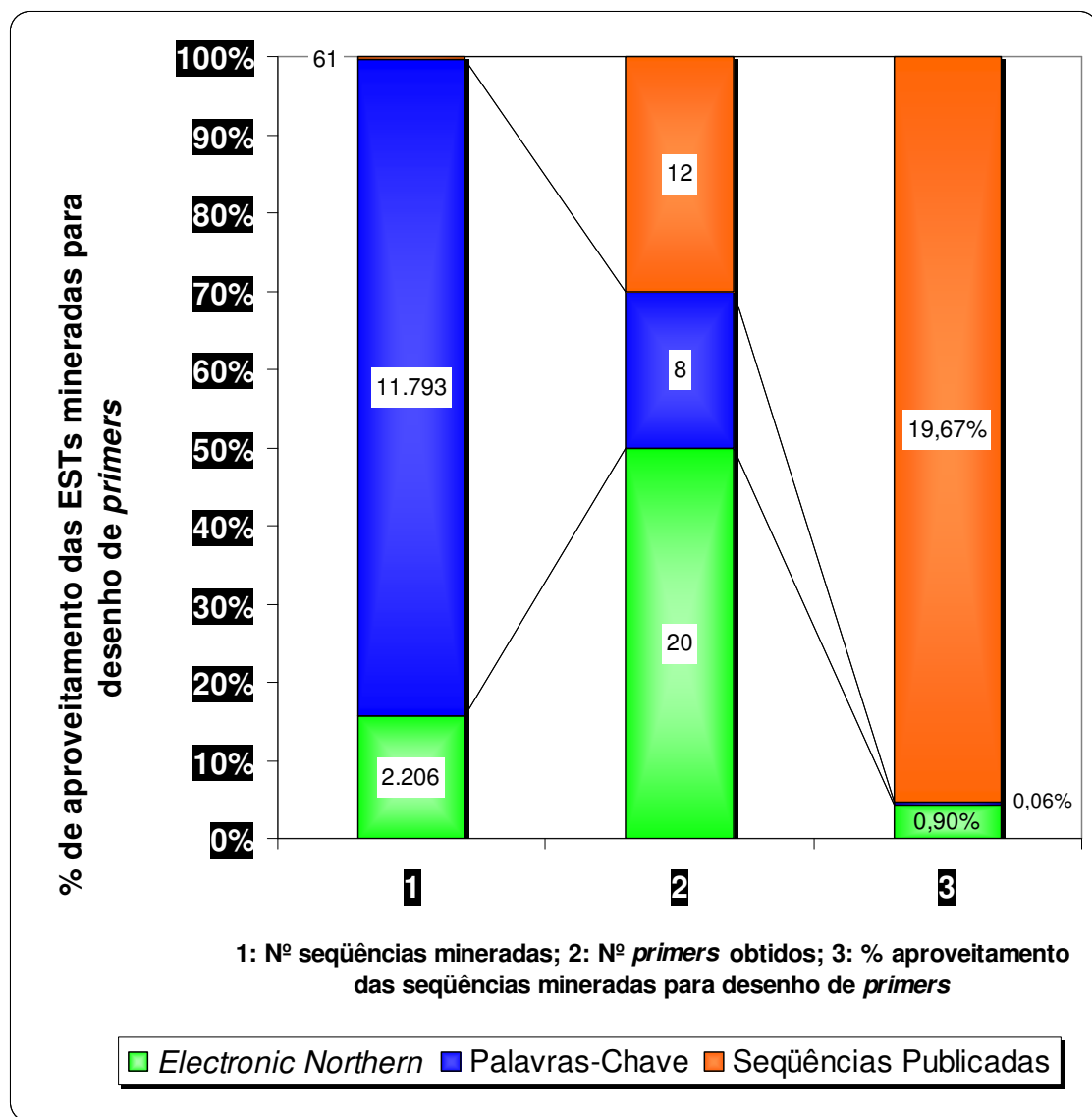


Figura 2.2: Número de seqüências de cada estratégia de mineração selecionadas para o desenho de *primers* e o aproveitamento proporcional de cada estratégia.

Após o teste de diferentes concentrações dos componentes da reação de PCR e condições de amplificação, os 40 pares de *primers* foram analisados em cafeeiros resistentes e susceptíveis à ferrugem. Desses, 11 não amplificaram nenhuma banda. Os 29 restantes geraram marcadores moleculares, visualizadas como bandas únicas e bem definidas (Figura 2.3).

Apenas um marcador, o CARF 005 resultou em fragmentos polimórficos detectáveis diretamente em gel de agarose, sendo capaz de diferenciar indivíduos resistentes e susceptíveis à ferrugem do cafeeiro (Figura 2.4). Todos os cafeeiros resistentes apresentaram o fragmento de 400 pb, enquanto nenhum susceptível apresentou a banda, com a exceção de *C. excelsa*, que apesar de ser susceptível, apresentou a banda.

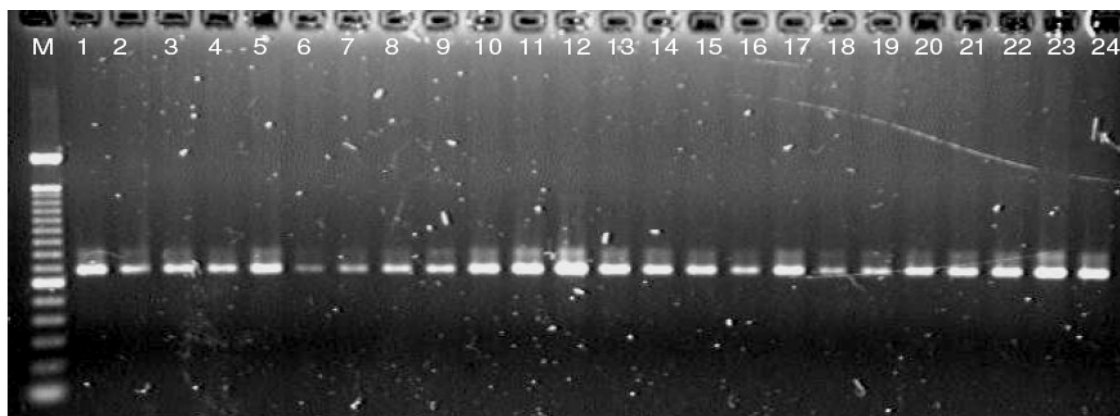


Figura 2.3: *Primer* CARF 028. Padrão de amplificação com bandas únicas e bem definidas. M= Marcador de peso molecular de 100 pb. 1-12 = Indivíduos susceptíveis. 13-24 = Indivíduos resistentes (ver Tabela 2.1).

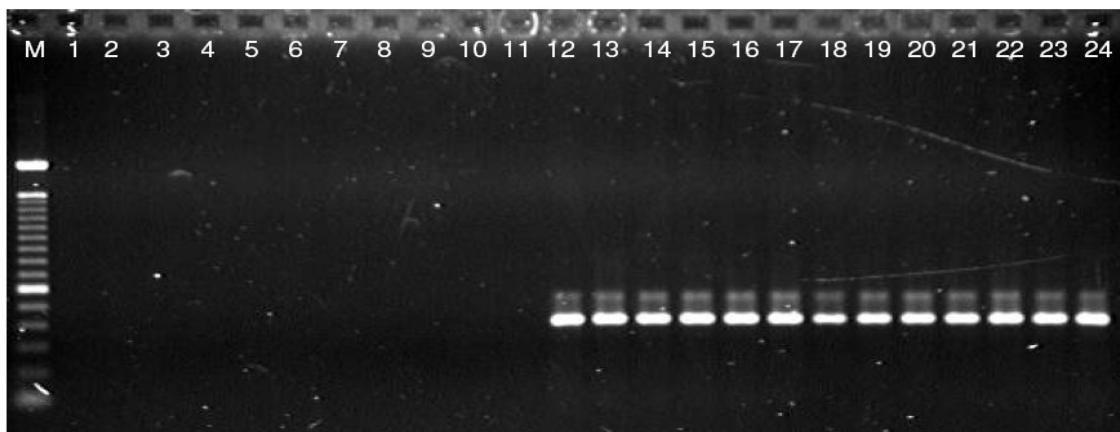


Figura 2.4: *Primer* CARF 005 – *Primer* polimórfico entre indivíduos susceptíveis e resistentes à Ferrugem. M= Marcador de peso molecular de 100 pb. 1-12 = Indivíduos susceptíveis. 13-24 = Indivíduos resistentes (ver Tabela 2.1).

O fragmento amplificado pelo *primer* CARF 005 foi seqüenciado para confirmar a sua anotação (gi|24459841|emb|CAC82597.1| disease resistance-like protein [*Coffea arabica*]). A seqüência amplificada pelo primer CARF 005 foi analisada por meio do BLASTN, otimizado para seqüências altamente similares (MEGABLAST), contra o banco de dados NR. O resultado do alinhamento mostrou que o fragmento corresponde a uma ORF parcial de

Coffea arabica (código de acesso AJ298884.1) para proteína de resistência a doenças (*Disease resistance-like protein*), com 97% de identidade, score igual a 287 e *e-value* igual a $2.4e^{-74}$ (Figura 2.5).

A seqüência nucleotídica foi convertida em seqüência de aminoácidos para uma análise em bancos de dados de proteínas. A pesquisa nos bancos InterPro, UniProtKB/TrEMBL, Pfam e PRINTS revelou similaridade da seqüência com proteína de resistência a doenças (códigos de acesso IPR000767, Q8GSU7, PF00931 e PR00364, respectivamente), com domínios conservados NB-ARC (IPR002182) e LRR (IPR011713).

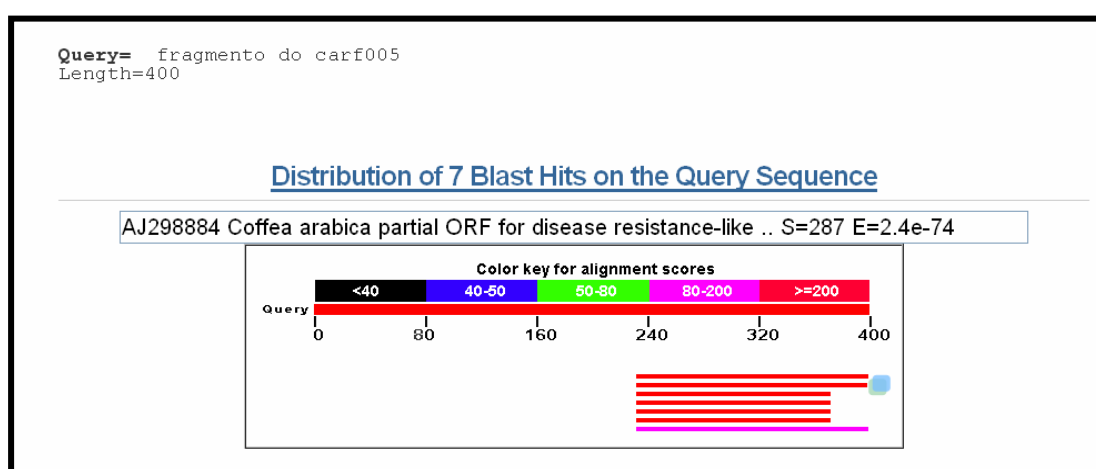


Figura 2.5: Alinhamento do fragmento amplificado pelo *primer* CARF 005 com o banco de dados NR.

Vários tipos de proteínas contendo o domínio NB-ARC estão envolvidas no processo de apoptose. O domínio NB-ARC é formado por três subdomínios: um domínio de ligação a nucleotídeo (NB – *Nucleotide Binding*), uma extensão C-terminal que forma uma estrutura de quatro hélices (ARC1) e uma hélice enovelada (ARC2) (Tameling et al., 2006). Estudos mostram que mutações no domínio NB-ARC eliminam a função de resistência das proteínas, demonstrando a relevância desse domínio (Dinesh-Kumar et al., 2000; Tao et al., 2000; Tornero et al., 2002).

De acordo com Parniske et al. (1997), o domínio LRR pode facilitar a interação da proteína R com o seu fator Avr (elicitor) relacionado e pode fornecer diferentes especificidades de reconhecimento para fatores Avr alterados. De fato, em vários sistemas planta-patógeno, a variação da seqüência na região LRR tem mostrado ser responsável por diferentes

especificidades de reconhecimento ou de resistência (Parniske et al., 1997; Thomas et al., 1997; Botella et al., 1998; Wang et al., 1998; Ellis et al., 1999).

Segundo os termos do Gene Ontology Consortium (GO – Ontologia Gênica), as seqüências participam de processos de defesa (defence response, GO:0006952), apoptose (apoptosis, GO:0006915) e tem função de ligação a ATP (ATP binding, GO:0005524).

O marcador CARF 005 poderá ser testado em populações segregantes contendo diferentes genes de resistência. Esta análise permitirá verificar a ligação desse marcador, obtido a partir das seqüências mineradas, com os diferentes genes relacionados com a resistência do cafeeiro à ferrugem.

Os marcadores baseados em cDNA, como os EST-PCR, apresentam várias vantagens. Eles marcam genes expressos, e por isso são particularmente úteis para mapeamento de QTLs. Quando o marcador EST-PCR é encontrado ligado a um QTL é possível que o gene, a partir do qual ele foi derivado, participe do controle da característica em questão. Por serem derivados de regiões codificadoras, que são mais conservadas entre populações e espécies do que as não codificadoras, os marcadores EST-PCR se tornam úteis para o estudo de mapeamento comparativo. Além disso, os marcadores EST-PCR podem ser herdados codominantemente, o que permite a identificação de dois alelos diferentes em locos heterozigotos de organismos diplóides (Rowland et al., 2003). Os marcadores EST-PCR já foram desenvolvidos para várias espécies vegetais, incluindo pinheiro [*Pinus taeda* L. (Temesgen et al., 2001)], píceas [*Picea mariana* (Perry e Bousquet, 1998) e *Picea abies* (Schubert et al., 2001)], criptoméria [*Cryptomeria japonica* (Tsumura et al., 1997)], mirtilo [*Vaccinium sp.* (Rowland et al., 2003)], girassol [*Helianthus annuus* (Pashley et al., 2006)], amendoim [*Arachis hypogaea* L. (Luo et al., 2005)], cana-de-açúcar [*Saccharum sp.* (Alwala et al., 2006)], feijão [*Phaseolus vulgaris* L. (Miklas et al., 2006)] e muitas outras.

Portanto, os marcadores moleculares obtidos a partir de ESTs são bastante informativos e permitem reconhecer a associação da marca com a sua potencial função, o que os tornam muito valiosos na análise de mapa genético. As possibilidades de sucesso da piramidação de genes de resistência à ferrugem dependem do número de genes envolvidos na resistência e a sua distribuição no genoma (ligados ou independentes).

Portanto, é importante estudar a distribuição dos possíveis genes de resistência, por exemplo, por meio da localização dos marcadores derivados de ESTs ligados a estes genes no mapa genético.

Em relação aos outros 28 pares de primers que não apresentaram polimorfismo, como CARF 028 (Figura 2.3), é possível que os fragmentos possuam pequenas diferenças dentro da seqüência e, por isso, não puderam ser observadas por eletroforese em gel de agarose. A existência desta diferença poderá ser verificada com análises de enzima de restrição e seqüenciamento para detecção de SNPs (*Single Nucleotide Polymorphisms*). Existe também a possibilidade da diferença estar no padrão de expressão dos genes de indivíduos resistentes e susceptíveis. Para a verificação da expressão diferencial, análises de RT-PCR e ensaios de macroarranjo poderão ser efetuados.

4. CONCLUSÃO

Este trabalho permitiu obter 40 *primers* para amplificação de seqüências potencialmente envolvidas com a resistência do cafeeiro a doenças.

Foi encontrado um marcador capaz de diferenciar cafeeiros resistentes dos susceptíveis à *H. vastatrix*. O *primer* CARF 005 amplifica uma região do DNA que corresponde a uma ORF parcial de *Coffea arabica* que codifica uma proteína de resistência a doenças.

Os marcadores moleculares obtidos a partir de ESTs são bastante informativos e permitem reconhecer a associação do marcador com a sua potencial função. Pretende-se, no futuro, incorporar esses marcadores em mapas genéticos, tornando a análise dos mesmos mais eficiente.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- Alwala, S., Kimbeng, C. A., Gravois, K. A., Bischoff, K. P. TRAP, a new tool for sugarcane breeding: comparisson with AFLP and coefficient of parentage. Journal American Society Sugar Cane Technologists, v.26, p.62-86. 2006.
- Alzate-Marin, A. L., Cervigni, G. D. L., Moreira, M. A., Barros, E. G. Seleção assistida por marcadores moleculares visando ao desenvolvimento de plantas resistentes a doenças, com ênfase em feijoeiro e soja. Fitopatologia Brasileira, v.30, n.4, p.333-342. 2005.
- Bent, A. F. Plant disease resistance genes: function meets structure. The Plant Cell, v.8, p.1757-1771. 1996.
- Boersma, J. G., Buirchell, B. J., Sivasithamparam, K., Yang, H. Development of a sequence-specific PCR marker linked to the Ku gene which removes the vernalization requirement in narrow-leafed lupin. Plant Breeding, v.OnlineEarly Articles. 2007.
- Bonas, U., Lahaye, T. Plant disease resistance triggered by pathogen-derived molecules: refined models of specific recognition. Current Opinion in Microbiology, v.5, n.1, p.44-50. 2002.
- Borém, A., Caixeta, E. T. Marcadores Moleculares. Viçosa, MG. 2006. 374 p.
- Botella, M. A., Parker, J. E., Frost, L. N., Bittner-Eddy, P. D., Beynon, J. L., Daniels, M. J., Holub, E. B., Jones J. D. G. Three genes of the arabidopsis RPP1 complex resistance locus recognize distinct *Peronospora parasitica* avirulence determinants Plant Cell, v.10, p.1847-1860. 1998.
- Buckler, E. S., Thornsberry, J. M. Plant molecular diversity and applications to genomics. Current Opinion in Plant Biology, v.5, p.107-111. 2002.
- Cushman, J. C., Bohnert, H. J. Genomic approaches to plant stress tolerance. Current Opinion in Plan Biology, v.3, p.117-124. 2000.
- Dinesh-Kumar, S. P., Tham, W, Baker, B. J. Structure-function analysis of the

- tobacco mosaic virus resistance gene N. Proceedings of the National Academy of Sciences of the United States of America, v.97, n.26, p.14789-14794. 2000.
- Diniz, L. E. C., Sakiyama, N. S., Lashermes, P., Caixeta, E. T., Oliveira, A. C. B., Zambolim, E. M., Loureiro, M. E., Pereira, A. A., Zambolim, L. Analysis of AFLP markers associated to the Mex-1 resistance locus in Icatu progenies. Crop Breeding and Applied Biotechnology, v.5, p.387-393. 2005.
- Doyle, J. J., Doyle, J. L. A rapid total DNA preparation procedure for fresh plant tissue. Focus, v.12, p.13-15. 1990.
- Ellis, J. G., Lawrence, G. J., Luck, J. E., Dodds, P. N. Identification of regions in alleles of the flax rust resistance gene L that determine differences in gene-for-gene specificity Plant Cell, v.11, n.3, p.495-506. 1999.
- Kobe, B., Deisenhofer, J. Proteins with leucine-rich repeats. Current Opinion in Structure Biology, v.8, n.5, p.409-416. 1995.
- Luo, M., Dang, P., Guo, B. Z., He, G., Holbrook, C. C., Bausher, M. G., Lee, R. D. Generation of expressed sequence tags (ESTs) for gene discovery and marker development in cultivated peanut. Crop Science, v.45, p.346-353. 2005.
- Mcdowell, J. M., Dhandaydham, M., Long, T. A., Aarts, M. G. M., Goff, S., Holub, E. B, Dangl, J. L. Intragenic recombination and diversifying selection contribute to the evolution of downy mildew resistance at the RPP8 locus of Arabidopsis. Plant Cell, v.10, n.11, p.1861-1874. 1998.
- Miklas, P. N., Hub, J., Grünwaldc, N. J., Larsena, K. M. Potential application of TRAP (Targeted Region Amplified Polymorphism) markers for mapping and tagging disease resistance traits in common bean Crop Science, v.46, p.910-916. 2006.
- Parniske, M., Hammondkosack, K. E., Golstein, C., Thomas, C. M., Jones, D. A., Harrison, K., Wulff, B. B. H., Jones, J. D. G. Novel disease resistance specificities result from sequence exchange between tandemly repeated genes at the Cf-4/9 locus of tomato. Cell, v.91, n.6, p.821-832. 1997.
- Pashley, C. H., Ellis, J. R., Mccauley, D. E., Burke, J. M. EST databases as a source for molecular markers: Lessons from *Helianthus*. The Journal of Heredity, v.97, n.4, p.381-388. 2006.
- Perry, D. J., Bousquet, J. Genetic diversity and mating system of post-fire and porst-harvest black spruce: an investigation using codominant sequence-tagged-site (STS) markers. Canadian Journal of Forest Resources, v.31, p.32-40. 2001.
- Rowland, L. J., Dhanaraj, A. L., Polashock, J. J., Arora, R. Utility of Blueberry-derived EST-PCR. HortScience, v.38, n.7, p.1428-1432. 2003.

- Schubert, R., Mueller-Starck, G., Riegel, R. Development of EST-PCR markers and monitoring their intrapopulational genetic variation in *Picea abies* (L.) Karst. Theoretical and Applied Genetics, v.103, p.1223-1231. 2001.
- Tameling, W. I. L., Vossen, J. H., Albrecht, M., Lengauer, T., Berden, J. A., Haring, M. A., Cornelissen, B. J. C., Takken, F. L. W. Mutations in the NB-ARC Domain of I-2 That Impair ATP Hydrolysis Cause Autoactivation. Plant Physiology, v.140, p.1233-1245. 2006.
- Tao, Y., Yuan, F., Leister, R.T., Ausubel, F.M., Katagiri, F. Mutational analysis of the Arabidopsis nucleotide binding site-leucine-rich repeat resistance gene RPS2. Plant Cell, v.12, p.2541-2554. 2000.
- Temesgem, B., Brown, G. R., Harry, D. E., Kinlaw, C. S., Sewell, M. M., Neal, D.B. Genetic mapping of expressed sequence tag polymorphism (ESTP) markers in loblolly pine (*Pinus taeda* L.). Theoretical and Applied Genetics, v.102, p.664-675. 2001.
- Thomas, C. M., Jones, D. A., Parniske, M., Harrison, K., Balint-Kurti, P. J., Hatzixanthis, K., Jones, J. D. G. Characterization of the tomato Cf-4 gene for resistance to *Cladosporium fulvum* identifies sequences that determine recognitional specificity in Cf-4 and Cf-9. Plant Cell, v.9, n.12, p.2209-2224. 1997.
- Tornero, P., Chao, R.A., Luthin, W.N., Goff, S.A., Dangl, J.L. Large-scale structure-function analysis of the Arabidopsis RPM1 disease resistance protein. Plant Cell, v.14, p.435-450. 2002.
- Vieira, L. G. E., Andrade, A. C., Colombo, C. A., Moraes, A. H. A., Metha, A., Oliveira, A. C., Labate, C. A., Marino, C. L. e C. B. Monteiro-Vitorello, Monte, D. C., Giglioti, E., Kimura, E. T., Romano, E., Kuramae, E. E., Lemos, E. G. M., Almeida, E. R. P., Jorge, E. C., Albuquerque, E. V. S., Silva, F. R., Vinecky, F., Sawazaki, H. E., Dorry, H. F. A., Carrer, H., Abreu, I. N., Batista, J. A. N., Teixeira, J. B., Kitajima, J. P., Xavier, K. G., Lima, L. M., Camargo, L. E. A., Pereira, L. F. P., Coutinho, L. H., Lemos, M. V. F., Romano, M. R., Machado, M. A., Costa, M. M. C., Sá, M. F. G., Goldman, M. H. S., Ferro, M. I. T., Tinoco, M. L. P., Oliveira, M. V., Sluys, M. V., Shimizu, M. M., Maluf, M. P., Eira, M. T. S., Filho, O. G., Arruda, P., Mazzafera, P., Mariani, P. D. S. C., Oliveira, R. L. B. C., Harakava, R., Balbao, S. F., Tsai, S. M., Mauro, S. M. Z., Santos, S. N., Siqueira, W. J., Costa, G. G. L., Formighieri, E. F., Carazzolle, M. F., Pereira, G. A. G. Brazilian coffee genome project: an EST-based genomic resource. Brazilian Journal of Plant Physiology, v.18, n.1, p.95-108. 2006.
- Wang, G., Ruan, D., Song, W., Sideris, S., Chen, L., Pi, L., Zhang, S., Zhang, Z., Fauquet, C., Gaut, B. S., Whalen, M. C., Ronald, P. C. Xa21D encodes a receptor-like molecule with a leucine-rich repeat domain that determines race-specific recognition and is subject to adaptive evolution Plant Cell, v.10, n.5, p.765-779. 1998.