CARLA CRISTINA DA SILVA

"CONSTRUCTION OF A FUNCTIONAL MAP FOR RUBBER TREE (*HEVEA BRASILIENSIS*)"

"CONSTRUÇÃO DE UM MAPA FUNCIONAL EM SERINGUEIRA (*HEVEA BRASILIENSIS*)"

CAMPINAS

2014

UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE BIOLOGIA

CARLA CRISTINA DA SILVA

"Construction of a functional map for ruber tree (*Hevea brasiliensis*)"

"Construção de um mapa funcional em seringueira (*Hevea brasiliensis*)"

Thesis presented to the Biology Institute of the University of Campinas in partial fulfillment of the requirements for the degree of doctor in GENETICS AND MOLECULAR BIOLOGY in the area PLANT BREEDING AND GENETICS.

Tese apresentada ao Instituto de Biologia da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do Título de Doutor em GENÉTICA E BIOLOGIA MOLECULAR na área de GENÉTICA VEGETAL E MELHORAMENTO.

Orientadora: Profa. Dra. Anete Pereira de Souza

ESTE EXEMPLAR CORRESPONDE À VERSÃO FINAL DE TESE DEFENDIDA PELA ALUNA CARLA CRISTINA DA SILVA, E ORIENTADA PELA PROF(A). DR(A). ANETE PEREIRA DE SOUZA.

Campinas, 2014

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Biologia
Mara Janaina de Oliveira - CRB 8/6972

Da Silva, Carla Cristina, 1978-
D26c Construção de um mapa funcional em seringueira (*Hevea brasiliensis*) / Carla Cristina da Silva. – Campinas, SP : [s.n.], 2014.

Orientador: Anete Pereira de Souza.
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Biologia.

1. Seringueira. 2. Marcadores genéticos. 3. Melhoramento genético. 4. Etiquetas de sequências expressas. I. Souza, Anete Pereira de,1962-. II. Universidade Estadual de Campinas. Instituto de Biologia. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Construction of a functional map for rubber tree (*Hevea brasiliensis*)
Palavras-chave em inglês:
Rubber plants
Genetic markers
Breeding
Expressed sequence tags
Área de concentração: Genética Vegetal e Melhoramento
Titulação: Doutora em Genética e Biologia Molecular
Banca examinadora:
Anete Pereira de Souza [Orientador]
Mário Luiz Teixeira de Morais
Maria Bernadete Lovato
Ananda Virginia de Aguiar
Flávio Maes dos Santos
Data de defesa: 18-06-2014
Programa de Pós-Graduação: Genética e Biologia Molecular

Campinas, **18** de **junho** de **2014**

BANCA EXAMINADORA

Profa. Dra. Anete Pereira De Souza (orientadora)

Assinatura

Prof. Dr. Mário Luiz Teixeira De Morais

Assinatura

Profa. Dra. Maria Bernadete Lovato

Assinatura

Dra. Ananda Virginia Aguiar

Assinatura

Prof. Dr. Flávio Antônio Maës Dos Santos

Assinatura

Prof. Dr. Michel Georges Albert Vincentz

Assinatura

Prof. Dr. Rafael Silva Oliveira

Assinatura

Dr Rodrigo Gazaffi

Assinatura

Abstract

Rubber tree (*Hevea brasiliensis*), native species of the Amazon, is world's major source of natural rubber. Rubber tree breeding programs have been fundamental for the selection of desirable traits. However, the breeding cycle is time consuming (around 30 years), which makes the development of new techniques for early evaluation a necessity. cDNA libraries and Expressed Sequence Tags (ESTs) are very important tools in molecular biology: they enable the identification of genes preferentially expressed in tissues or cellular types and are also a valuable resource of polymorphic markers, powerful instruments for genotyping and molecular mapping. The use of EST-derived markers allows the construction of functional maps, wherein expressed genes or regions near genes are positioned. This type of mapping is important for gene-trait association studies and candidate genes identification.

The present study aimed at the construction of cDNA libraries from different tissues (panel, latex and leave) and treatments (cold exposure and *Microcyclus ulei* controlled infection) of rubber tree for the development of EST sequences and gene-targeted molecular markers, to raise the saturation of a microsatellite-based integrated genetic map previously constructed in our laboratory, in which 18 quantitative trait loci (QTLs) related to growth traits were identified.

Sequencing of 10,464 clones generated 8,551 high quality ESTs that were clustered into 5,211 unigenes. Among these, 3,582 (68.7%) showed similarity to a hypothetical or expressed protein. A total of 173 EST-SSR and 43 SNP markers were developed for *H. brasiliensis*. 150 SSRs (87%) could be associated with functional genes, and 98.8% were transferred to other *Hevea* species, suggesting that the genus is a complex formed by different species. SNP markers were identified in 13 ESTs that showed

similarity to stress response, development and latex biosynthesis proteins. Six sequences were highly abundant in the cold exposure libraries and expression analyses demonstrated that five sequences were up-regulated during the exposure, with emphasis to two sequences with more than 70-fold increase in expression.

From the developed EST-SSRs, 46 were genotyped in the segregating F1 population comprised of 270 plants. These markers were added to the genetic map, which know contains a total of 330 markers. The OneMap software was used for the map construction that now has 3,068.9 cM and 22 linkage groups. Five loci were mapped into QTLs, and transcripts of three of them present similarity to proteins involved in stress response and developmental processes. These loci may be candidate genes for studies related to rubber tree growth traits.

To our knowledge, this is the first work in rubber tree that combines analyses of ESTs from different tissues and treatments, and to analyze sequences under cold stress, in several *H. brasiliensis* genotypes. The new positioned markers may help in the identification of genes of interest and QTLs for other agronomic important traits. The several gene-targeted markers developed here will be used in the mapping and positioning of possible genes in other mapping populations that are now being evaluated at Genetics and Molecular Analysis Laboratory.

Resumo

A seringueira (*Hevea brasiliensis*), espécie nativa da Amazônia, é a maior fonte de borracha natural do mundo. Programas de melhoramento genético da seringueira têm sido cruciais para a obtenção de caracteres desejáveis. Entretanto, o ciclo de melhoramento da seringueira é muito longo (cerca de 30 anos), tornando-se essencial o desenvolvimento de novas técnicas de avaliação precoce. As bibliotecas de cDNA e *Expressed Sequence Tags* (ESTs) são ferramentas muito importantes em biologia molecular: possibilitam identificar genes preferencialmente expressos em tecidos ou tipos celulares e também são valiosas fontes de marcadores polimórficos, instrumentos poderosos para genotipagem e mapeamento molecular. O uso de marcadores derivados de ESTs permite construir mapas funcionais, nos quais são posicionados genes transcritos ou regiões próximas aos genes. Este tipo de mapeamento é importante para estudos de associação gene-característica, e identificação de genes candidatos.

Este trabalho objetivou a construção de bibliotecas de cDNA de diferentes tecidos (painel, látex e folha) e tratamentos (exposição ao frio e infecção controlada por *Microcyclus ulei*) de seringueira para desenvolver sequências EST e marcadores moleculares gene-direcionados a partir destas sequências, para aumentar a saturação de um mapa integrado baseado em microssatélites, no qual identificaram 18 *quantitative trait loci* (QTLs) para características de crescimento, construído previamente em nosso laboratório.

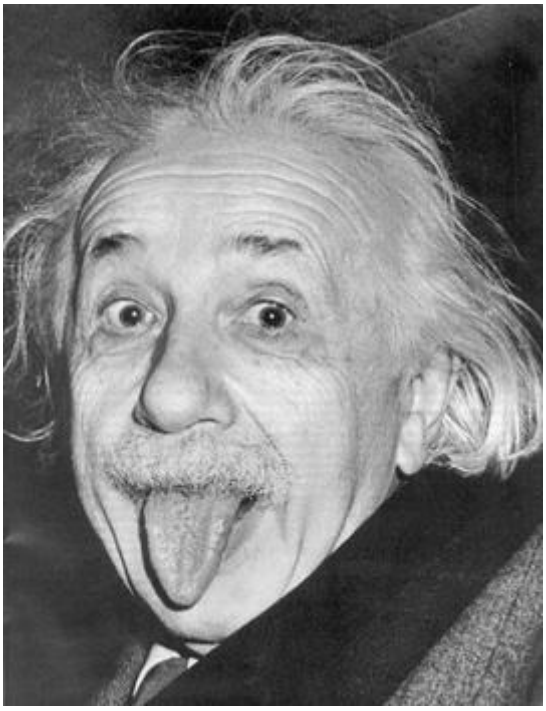
Foram sequenciados 10.464 clones, gerando 8.551 ESTs de alta qualidade que agrupadas formaram 5.211 unigenes. Destes, 3.582 (68,7%) apresentam similaridade com uma proteína hipotética ou expressa. Foram desenvolvidos 173 marcadores EST-SSR e 43 marcadores SNP para *H. brasiliensis*. 150 EST-SSRs (87%) podem estar associados a

genes funcionais, e 98,8% foram transferidos para outras espécies de *Hevea*, sugerindo que o gênero seja um complexo formado pelas diferentes espécies. Os SNPs foram identificados em 13 ESTs similares a proteínas de resposta a estresse, desenvolvimento e síntese de látex. Seis sequências foram abundantes nas bibliotecas de exposição ao frio e análises de expressão demonstraram que a expressão de cinco sequências aumentou durante o experimento, principalmente a expressão de duas sequências que foi aumentada mais de 70 vezes.

Dos EST-SSRs desenvolvidos, 46 foram genotipados na população segregante F_1 com 270 indivíduos, e estes marcadores foram adicionados ao mapa genético de seringueira, totalizando 330 marcadores. O programa OneMap foi usado para a construção do mapa que possui 3.068,9 cM de extensão e 22 grupos de ligação (LGs). Cinco locos foram mapeados em regiões QTL, e os transcritos de três são similares a proteínas de resposta a estresse e desenvolvimento. Estes locos podem ser genes candidatos para estudos relacionados a características de crescimento em seringueira.

Até o momento, este é o primeiro trabalho em seringueira que combina análises de ESTs de diferentes tecidos e tratamentos, e análises sobre a exposição a baixas temperaturas, em vários genótipos de seringueira. Os novos marcadores adicionados ao mapa poderão auxiliar na identificação de genes de interesse e de QTLs para outras características de importância agronômica. Os vários marcadores gene-direcionados desenvolvidos serão utilizados para mapeamento e posicionamento de possíveis genes em outras populações de mapeamento que estão sendo avaliadas no Laboratório de Análise Genética e Molecular.

Sumário



Estudar, e buscar pela verdade e beleza, é uma esfera de atividade na qual nos é permitido continuar crianças por toda a vida.

(Albert Einstein)

Agradecimentos

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), pelo apoio financeiro no projeto e pela bolsa concedida.

À professora Anete Pereira de Souza, por ter me aceitado como aluna, pela orientação constante durante todo o trabalho, e pela paciência sem fim que tem comigo. Obrigada pelas ideias que enriqueceram este trabalho, pelos conselhos nos momentos em que me sentia perdida, por manter um ambiente de trabalho do qual não se sente vontade de se despedir, e pelas goiabadas que alegravam as minhas tardes na hora do café (desculpe, não resisti).

Ao professor Paulo de Souza Gonçalves, por transferir a mim o máximo possível de seu imenso conhecimento e todo o seu entusiasmo pela seringueira. Acredito que, sem sua ajuda, não conseguiria desenvolver este trabalho.

Ao professor Antônio Augusto Franco Garcia, por esclarecer minhas inúmeras dúvidas estatísticas com tanta gentileza e paciência.

A Carlos R. R. Mattos (Michelin/Bahia) e Dr. Dominique Garcia (CIRAD) pelos conselhos e colaboração preciosa nos experimentos de infecção controlada.

À empresa Michelin, pela população de mapeamento.

À Embrapa pelo material genético utilizado nas análises.

Aos membros da pré-banca, pela disponibilidade e ajuda na pré-avaliação da tese.

Aos membros da banca de defesa, pela disponibilidade em contribuir para que esse trabalho fosse melhorado.

Às secretárias Sandra e Tânia pela atenção, profissionalismo e carinho em me atender em todos os momentos que precisei.

Aos colegas e amigos do Laboratório de Análise Genética e Molecular/CBMEG e Barracão, pelas inúmeras vezes que pedi ajuda e sempre se prontificaram a me ajudar, mesmo que às vezes tivessem que desenhar. Pela paciência que tiveram comigo, o carinho com que fui recebida no laboratório, por rirem das minhas piadas sem graça e por rirem do que eu falei, mesmo não sendo piada. É tão legal trabalhar com vocês que parece que eu cheguei mês passado no LAGM. Gente... cinco anos passaram depressa demais...

Aos funcionários, Carlão, Aline, Alessandra e Pathy, que com sua ajuda conseguem transformar uma hora de nosso tempo em duas!

Ao Juverlande, por aguentar pacientemente todas as vezes que fui atrás dele pedindo ajuda para resolver algum problema que, óbvio, ele resolveu em seguida. Obrigada pelos papos durante os cafezinhos, pelos “causos” que sempre alegravam o dia, e por tomar conta de nós como um pai, inclusive com merecidos puxões de orelha.

Ao Danilo, por respirar fundo toda vez que me via aparecer na porta e me responder com a maior paciência todas as dúvidas que tive ao longo da minha estadia no laboratório. Com certeza já tem um lugar guardadinho no céu.

À Taty, por conseguir entender como eu funciono logo na primeira semana que nos conhecemos e me apresentar para o então mundo novo dos marcadores moleculares e mapeamento. Saudades de dar risada com você...

À Lívia e à Camila, por não ter nem como explicar a paciência que tiveram comigo, a capacidade de me ensinar e me fazer entender como é lindo ver os alelos segregando no gel de acrilamida e mais lindo ainda ver os locos posicionados um do lado do outro em seus respectivos grupos de ligação. Obrigada por tornarem divertidas nossas viagens de coleta, suportarem elegantemente as minhas gafes, e por não me esquecerem dormindo no avião.

Aos meus amigos, em especial, à Carla, Kelly, Juliana e Cilb, minhas amigas de todas as horas há mais de 15 anos.

À minha família, pelo amor incondicional que sempre demonstraram e o apoio que sempre me deram, mesmo durante nossas constantes confusões.

À Deus, que considerou que eu merecia esta oportunidade, sem eu ter conhecimento do porquê.

À música, por existir e ter a capacidade de me acalmar e levar meus pensamentos a lugares onde as formas e as cores são mais vivas, estimulando meu raciocínio e trazendo leveza à minha alma. Sua presença constante, com certeza, me ajudou até o último minuto deste trabalho.

A imaginação frequentemente nos levará a mundos que nunca existiram, mas sem isso, não iríamos a lugar algum.

(Carl Sagan)

Dedico este trabalho à minha família, pela paciência, carinho e amor que sempre me dedicaram. Sem eles, com certeza, eu não teria aprendido que dias melhores sempre virão.

Introdução

A seringueira (*Hevea brasiliensis*), espécie nativa da Amazônia, pertence à família Euphorbiaceae e é a maior fonte de borracha natural do mundo, matéria-prima de grande importância em diversos setores industriais. Ao contrário da maioria das plantas cultivadas, é uma espécie que ainda está sendo domesticada, e sua história é tão curta que quase não mudou o seu aspecto de árvore nativa encontrada na Amazônia. Apesar de o Brasil ser o centro de origem e de diversidade da espécie, e principal produtor e exportador de borracha no final do século XIX, o Brasil é importador dessa matéria prima desde o início do século passado. Em 2012, a produção mundial de borracha natural atingiu 11.327.000 toneladas, das quais o Brasil contribuiu, apenas, com 171.500 toneladas (1,5%). Nesse ano o consumo nacional foi de 343.400 toneladas, ou seja, o Brasil importou a maior parte da borracha consumida no país.

A grande limitação para o cultivo da seringueira em sua região de origem é a ocorrência do mal-das-folhas, causada pelo fungo *Microcyclus ulei*, que é favorecida pela alta temperatura e umidade características da Amazônia. Desta maneira, o cultivo da seringueira se estendeu a áreas de diferentes características climáticas, denominadas regiões de escape, criando a necessidade de novos cultivares adaptados a essas áreas. A expansão do cultivo da seringueira no Brasil está diretamente ligada à identificação de clones adaptados às áreas de escape. Programas de melhoramento genético da seringueira têm sido fundamentais para a obtenção de caracteres desejáveis. A seringueira por ser uma cultura perene, tem um longo ciclo de melhoramento, sendo necessários cerca de 30 anos para obter-se uma variedade melhorada. Com o intuito de maximizar a seleção e reduzir esse período, mudanças que possam ser efetuadas sem comprometer o ganho genético, pode resultar em um programa mais eficiente.

As ferramentas da biologia molecular podem ser a chave para tornar mais eficiente e rápida a obtenção de novos cultivares. A geração de marcadores moleculares e mapas genéticos podem possibilitar a detecção precoce de um genótipo de interesse a partir de seleção assistida (MAS) e seleção genômica ampla (GWS), assim reduzindo o tempo e os custos. O uso de técnicas moleculares em seringueira teve um início bastante tímido, e somente nos últimos anos a utilização dessas técnicas se tornou mais abrangente.

Neste contexto, o principal objetivo deste trabalho foi o desenvolvimento de *Expressed Sequence Tags* (ESTs) de diferentes tecidos e tratamentos em seringueira e o desenvolvimento de marcadores gene-direcionados a partir destes ESTs, com a finalidade de aumentar a saturação e posicionar possíveis genes em um mapa genético molecular de seringueira construído em nosso laboratório. Para tanto, foi necessária a construção de sete bibliotecas de cDNA, o sequenciamento e análise dos *reads*, e a identificação de polimorfismos em sequências que apresentavam similaridade com proteínas.

A adição dos novos marcadores, juntamente com uma maior saturação do mapa, abre novas possibilidades para a identificação de genes envolvidos em características agronômicas de interesse, QTLs e para realizar MAS e GWS. Os marcadores desenvolvidos podem ser usados em diferentes populações de mapeamento, assim como na caracterização de germoplasma, estudos sobre variabilidade genética e estrutura populacional em *H. brasiliensis* e em outras espécies de *Hevea*. Além disso, as sequências EST obtidas podem ser usadas para auxiliar na análise de outras bibliotecas de cDNA e, por serem longas, na montagem de transcriptomas de seringueira.

Este trabalho foi desenvolvido em colaboração com o Centro de Seringueira do Instituto Agronômico de Campinas (IAC), Departamento de Genética da ESALQ/USP e a empresa Michelin do Brasil.

Capítulo I

Revisão bibliográfica

O contexto e a importância da cultura da seringueira

A seringueira (*Hevea* spp.) é uma das principais espécies florestais nativas do Brasil, é perene, pertence à família Euphorbiaceae e a classificação atual do gênero *Hevea* conduz a 11 espécies intercruzáveis (Pires *et al.* 2002; Gonçalves *et al.* 1990). As espécies conhecidas são: *H. benthamiana* Müll-Arg., *H. brasiliensis* (Willd. ex Adr. de Juss.) Muell-Arg., *H. camargoana* Pires, *H. camporum* Ducke, *H. guianensis* Aubl., *H. microphylla* Ule, *H. nitida* Mart. ex Müll-Arg., *H. paludosa* Ule, *H. pauciflora* (Spruce ex Benth.) Müll-Arg., *H. rigidifolia* (Spruce ex Benth.) Müll-Arg., e *H. spruceana* (Benth.) Müll-Arg (Gonçalves *et al.*, 1997). O rio Negro, na confluência com o rio Amazonas, é o centro primário de diversidade genética do gênero *Hevea*, sendo as áreas nas proximidades do município de Borba, no baixo rio Madeira, o centro secundário (Wycherley, 1977). A ocorrência natural das espécies de *Hevea* está circunscrita à região Amazônica, sendo encontradas no Brasil, Bolívia, Peru, Colômbia, Equador, Guianas, Suriname e Venezuela (Webster e Paardekooper, 1989; Priyadarshan e Clement-Demange, 2004) (Fig. 1).

A seringueira é uma dicotiledônea monóica, as flores são unissexuadas, pequenas, amarelas e dispostas em racimo. As folhas são longamente pecioladas e repartidas em três folíolos. O fruto é uma cápsula grande que geralmente apresenta três sementes (Paiva, 1992). A polinização da seringueira é entomófila, sendo que pequeninos insetos da família Ceratopogonidae (Heleidae) são os principais responsáveis pela polinização natural, operando em curtas distâncias para polinização cruzada (IPEF, 2007), e as sementes são dispersas pelos rios.

Estudos sobre o gênero tiveram início em 1775, com a descrição de *H. guianensis* por Aublet (Priyadarshan e Clement-Demange, 2004). Em decorrência da sobreposição geográfica das espécies em sua área natural de ocorrência, a hibridização é frequente devido à falta de

barreiras reprodutivas entre as espécies. Desta maneira, é provável que as espécies de *Hevea* ainda estejam em processo de especiação (Pires *et al.*, 2002; Gonçalves e Fontes, 2012).

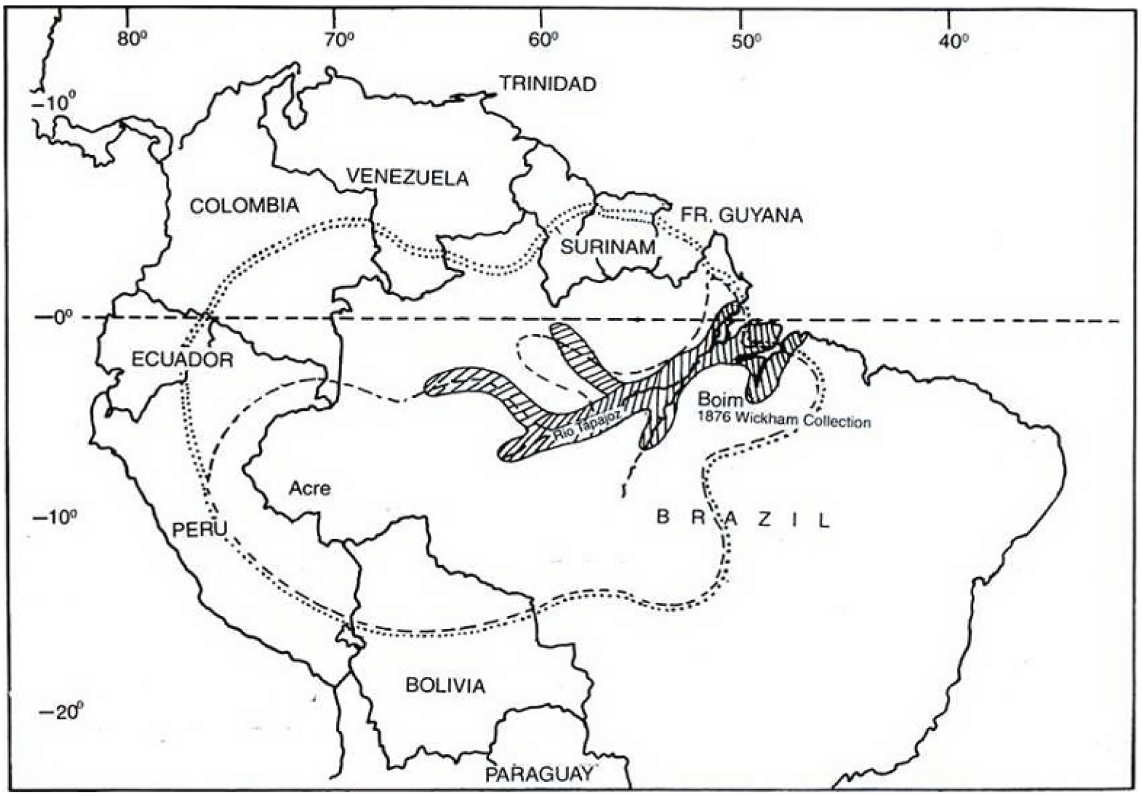


Figura 1 – Ocorrência natural do gênero *Hevea* e da espécie *H. brasiliensis* (Lam *et al.*, 2012).

····· Distribuição do gênero *Hevea*

- - - - - Distribuição de *H. brasiliensis*

Hevea é considerado um gênero diploide (2n=2x=36), todavia alguns estudos citogenéticos apontam que a seringueira seria um anfidiplóide (2n=4x=36) que se estabilizou durante o curso da evolução (Ong, 1975; Priyadarshan e Clement-Demange, 2004). Estudos de hibridização *in situ* localizaram dois locos distintos para o rDNA 18S-25S e um loco para o rDNA 5S no genoma de *Hevea*, o que indicaria a possibilidade de uma origem alotetraplóide para seringueira, com a perda de um loco rDNA 5S no processo (Leitch *et al.*, 1998). Entretanto, análises com marcadores moleculares demonstram que *Hevea* se comporta como um diploide (Lespinasse *et al.*, 2000a; Le Guen *et al.*, 2011; Souza *et al.*, 2013), e embora algumas

observações possam sugerir uma origem a partir de duas espécies ancestrais, nenhum ancestral diploide com 2n = 18 foi encontrado (Lespinasse *et al.*, 2000a).

. Dentre as espécies de seringueira se destaca a *H. brasiliensis* por ser a principal fonte de borracha natural existente. A borracha natural é matéria-prima estratégica para mais de 40 mil produtos, sendo possuidora de elasticidade, resistência ao desgaste, plasticidade, impermeabilidade para líquidos e gases e propriedades isolantes de eletricidade (Mooibroek e Cornish, 2000). De acordo com dados do International Rubber Study Group (IRSG, 2013), em 2012 a produção mundial de borracha natural atingiu 11,327 milhões de toneladas para um consumo de 11,005 milhões de toneladas. Deste montante, o Brasil contribuiu apenas com 171,5 mil toneladas (1,5%), enquanto o consumo nacional interno foi de 343,4 mil toneladas (IRSG, 2013), ou seja, o país importou a maior parte da borracha natural que consumiu, e as projeções de consumo para 2030 no Brasil ultrapassam a marca de um milhão de toneladas (Cortez, 2005).

Embora a bacia Amazônica ofereça ótimas condições climáticas para o desenvolvimento da cultura, a ocorrência de uma doença causada pelo fungo *Microcyclus ulei*, o mal das folhas também chamada de SALB (*South American leaf bligtht*), limita a produção de borracha nessa região (Pushparajah, 2001). Por consequência, a produção de borracha natural tem se expandido para áreas de escape. Entre os locais proeminentes dessa expansão encontra-se o norte da Índia, Vietnã, sul da China e o planalto sul do Brasil. Além das novas condições de desenvolvimento da cultura, essas novas áreas possuem condições de estresse como baixas temperaturas, altitude elevada, vento e outras doenças (Pushparajah, 1983; Priyadarshan e Gonçalves 2003). O melhoramento genético vem buscando clones adaptados a essas novas áreas de escape (Priyadarshan e Gonçalves 2003). Entretanto, o ciclo de melhoramento genético da seringueira, que vai do cruzamento até o final da produção leva de 20 a 30 anos para se concretizar. Além do longo ciclo de vida, da falta de sincronismo no florescimento dos genitores e o baixo sucesso na

polinização controlada em seringueira, as progênies passam por testes precoces de produção, vigor e tolerância a doenças para serem testados em Experimentos de Avaliação em Pequena Escala (EAPEs), seguidos por Experimentos de Avaliação em Grande Escala (EAGEs) (Fig. 2) (Gonçalves e Fontes, 2012). Sendo assim, o desenvolvimento de novas técnicas de avaliação precoce que possibilitem diminuir e aperfeiçoar as avaliações para seringueira é extremamente necessário. Neste ponto, as técnicas de biologia molecular podem auxiliar na obtenção de novos cultivares de maneira mais rápida e eficiente.

Bibliotecas de cDNA e *Expressed Sequence Tags*

A clonagem de cDNA é uma das mais importantes ferramentas em biologia molecular. Grande parte do nosso conhecimento sobre a natureza de transcritos e de proteínas se deve à capacidade de preparação de cópias de cDNA a partir de RNA e da criação de bibliotecas de cDNA com a clonagem dessas cópias. As bibliotecas de cDNA são derivadas, em sua maior parte, de um *pool* de RNAs mensageiros (mRNAs) originário de diferentes tecidos ou até de uma única célula. Geralmente a síntese de cDNA se inicia pelas caudas poli(A), na extremidade 3', por *primers* oligo(dT) contendo de 12 a 18 repetições de timina. Os *primers* oligo(dT) ainda podem conter sítios de restrição para endonucleases, facilitando a clonagem dos cDNAs através da digestão dos fragmentos (revisado em Harbers, 2008). Para a síntese da segunda fita de cDNA é necessária a criação de um sítio de iniciação na extremidade 5' do cDNA, tendo várias técnicas sido desenvolvidas para este fim (Kato *et al.*, 1994; Maruyama e Sugano, 1994; Vettore *et al.*, 2001). Os cDNAs dupla-fita são clonados em plasmídios (Vettore *et al.*, 2001; Tobias *et al.*, 2005) ou bacteriófagos (Lin *et al.*, 2005), e os clones são propagados usualmente em *Escherichia coli*.

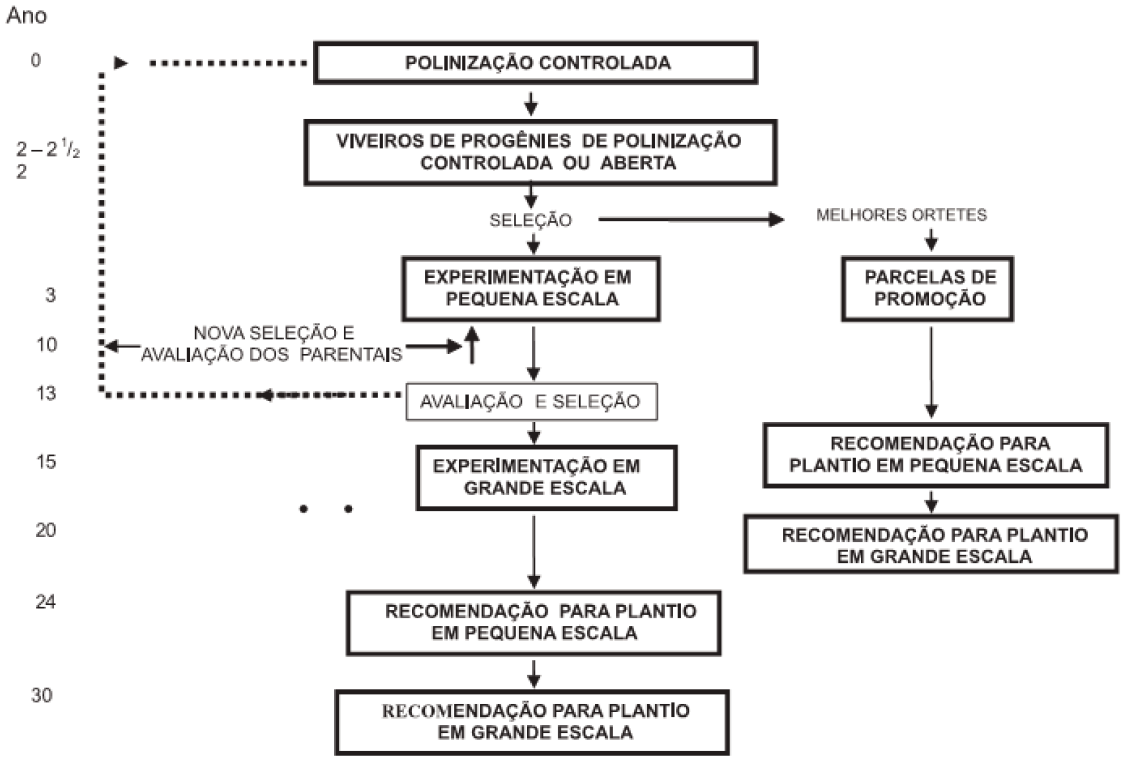


Figura 2 - Ciclo utilizado em programas de seleção e melhoramento genético da seringueira (Gonçalves e Fontes, 2012).

O sequenciamento parcial de cDNAs, através de suas extremidades 5' e 3', gera as chamadas *Expressed Sequence Tags* (ESTs) (Sterky e Lundeberg 2000). As ESTs são uma ferramenta poderosa para a identificação de genes que são preferencialmente expressos em certos tecidos ou tipos celulares (Lin *et al.*, 2005; Tobias *et al.*, 2005) e para a anotação de sequências de genomas completos (genoma expresso) em diversos organismos (revisado em Harbers, 2008). Em plantas, os projetos de sequenciamento de ESTs geraram uma enorme quantidade de sequências de DNA informativas, além de caracterizar genes por completo (Clepet *et al.*, 2011; Tang *et al.*, 2013; Lin *et al.*, 2013). A comparação das sequências de ESTs às de genes conhecidos depositados em bancos de dados de sequências de DNA, tal como o NCBI (http://ncbi.nlm.nih.gov/dbEST/), permite, via homologia, associar a cada EST uma provável função. A identificação de genes sofreu um avanço possibilitado pelo uso das ESTs como

evidências biológicas para centenas de genes preditos e como fonte para a descoberta de novos genes ou de transcritos alternativos. Apesar de a técnica de clonagem de cDNA ser usada há muitos anos, os projetos em larga escala provavelmente ainda não possuem todos os transcritos representados nas bibliotecas. A fração do genoma ativamente transcrita é muito maior do que se esperava com base nas anotações de genomas totais (Carninci, 2007; Kapranov *et al.*, 2007).

As bibliotecas de cDNA proporcionam recursos para estudos funcionais de novos transcritos, a descoberta de novos genes e para análises de transcriptoma (Harbers, 2008), que em seringueira começaram a ser realizados apenas nos últimos anos (Xia *et al.* 2011; Triwitayakorn *et al.* 2011; Li *et al.* 2012; Salgado *et al.*, 2014; Mantello *et al.*, *in press*). A produção de ESTs em larga escala cria a possibilidade de identificação de genes envolvidos em diferentes processos celulares levando a um melhor entendimento dos mecanismos moleculares atuantes. Em plantas de interesse agronômico, é possível a identificação de genes relacionados com características fisiológicas e bioquímicas economicamente importantes, podendo-se fazer uma ligação direta entre fenótipo e genótipo. Além disso, as bibliotecas de cDNA são uma valiosa fonte de marcadores moleculares polimórficos baseados em PCR, os quais são instrumentos poderosos para genotipagem e mapeamento molecular.

Marcadores moleculares desenvolvidos a partir de ESTs

As variações encontradas nas sequências de DNA são uma das mais importantes características para estudos genéticos em qualquer espécie, e os marcadores moleculares se apresentam como ferramentas convenientes para se examinar essas variações genéticas. Em plantas, os marcadores são usados na pesquisa básica e em melhoramento para caracterização de germoplasma, isolamento de genes, para acompanhar a introgressão de alelos favoráveis e proteção da variabilidade (revisado em Andersen e Lübberstedt, 2003). Vários marcadores foram

desenvolvidos em diferentes espécies de plantas [*Restriction Fragment Lenght Polymorphisms* (RFLPs), *Variable Number of Tandem Repeats* (minissatélites ou VNTR), *Amplified Fragment Length Polymorphisms* (AFLPs), Random Amplified Polymorphic DNA (RAPDs)], sendo os microssatélites ou SSRs (*Simple Sequence Repeats*) e os SNPs (*Single Nucleotide Polymorphisms*) os mais utilizados (Borém e Caixeta, 2006).

Os marcadores moleculares foram primeiramente desenvolvidos a partir de sequências genômicas anônimas (RDMs – *Random DNA Markers*), sendo fenotipicamente neutros, e têm sido usados com sucesso em vários estudos de caracterização genética (revisado em Andersen e Lübberstedt, 2003). Entretanto, a ligação genética entre um RDM específico e o loco de um alelo-alvo pode ser quebrada por recombinação, o que limita o uso destes marcadores como uma ferramenta diagnóstica (Rafalski e Tingey, 1993).

A caracterização funcional de genes, os ESTs e os projetos de sequenciamento de genomas facilitam o desenvolvimento de marcadores moleculares de regiões transcritas do genoma (Varshney *et al.*, 2005a). Os SSRs (Varshney *et al.*, 2005b) e os SNPs (Rafalski, 2002) são os marcadores mais importantes e populares que podem ser desenvolvidos a partir de ESTs. Os ESTs que deram origem aos marcadores podem ser submetidos a buscas por homologia (BLASTX) nos bancos de dados de proteínas (i.e.: SWISSPROT), podendo-se deduzir suas prováveis funções. Estes marcadores são chamados gene-direcionados (GTMs – *Gene Targeted Markers*) por se encontrarem dentro dos genes, mas não estarem relacionados com características fenotípicas. Quando estão envolvidos com a variação fenotípica de uma característica, são chamados de marcadores funcionais (FMs – *Functional Markers*) (Andersen e Lübberstedt, 2003). Se um marcador FM estiver geneticamente associado a uma característica de interesse, é provável que ele afete diretamente a característica. Além disso, os ESTs que apresentam homologia com genes candidatos podem ser utilizados de forma específica para o mapeamento

genético (Cato *et al.* 2001). Vários possíveis FMs têm sido desenvolvidos para diferentes espécies de plantas (Liu *et al.*, 2012; Xia *et al.*, 2012; Azmach *et al.*, 2013).

A utilização de marcadores moleculares derivados de ESTs permite a construção de mapas funcionais, nos quais são posicionados genes transcritos, que podem apresentar associação significativa com características agronômicas de interesse. Em longo prazo, o desenvolvimento de marcadores alelo-específicos para os genes que controlam características de valor agronômico será um avanço para o melhoramento vegetal.

1. EST-SSRs

Os microssatélites, ou SSRs, representam uma sequência de DNA composta por pequenos motivos de 1 a 6 pares de bases repetidos em *tandem*. O número de repetições do microssatélite pode variar, sendo que sua sequência adjacente costuma ser conservada entre indivíduos diferentes da mesma espécie ou de espécies próximas. Estes marcadores são altamente polimórficos e abundantes, apresentando grande reprodutibilidade, herança codominante e boa cobertura do genoma, sendo ideais para estudos de genética e melhoramento em plantas (Powell *et al.*, 1996). Os microssatélites também são usados para a integração de mapas genéticos e físicos em plantas, fornecendo meios para se associar variações fenotípicas a variações genotípicas (revisado em Varshney *et al.*, 2005b). Em função das vantagens que esse marcador apresenta, ele se tornou o marcador genético mais utilizado nos diferentes estudos genéticos para os quais são utilizados marcadores moleculares (Schlotterer, 2000).

Os marcadores microssatélites, desenvolvidos a partir de bibliotecas genômicas, podem pertencer a qualquer região do genoma, transcrita ou não, e raramente há informações disponíveis quanto às suas funções. Em contraste, os EST-SSRs geralmente possuem funções putativas e são marcadores gene-direcionados, com o potencial de se tornarem FMs, caso os polimorfismos dos

motivos repetidos afetem a função do gene no qual ele se encontra (Andersen e Lübberstedt, 2003; Varshney *et al.*, 2005b). São encontrados em alta densidade nas regiões expressas, sendo estimado em média um SSR a cada 6Kb em plantas (Varshney *et al.*, 2002). Em comparação com os SSRs genômicos, os EST-SSRs apresentam níveis de polimorfismo inferiores, devido à maior pressão seletiva contra mutações nas regiões expressas, entretanto a qualidade de definição e as taxas de transferibilidade entre diferentes espécies são maiores (Varshney *et al.* 2005b; Varshney *et al.* 2005c; Feng *et al.* 2009).

2. SNPs

Os SNPs são muito abundantes no genoma e, devido à sua frequência e distribuição, surgem como importantes marcadores genéticos para a obtenção de mapas de alta resolução (Rafalski, 2002; Andersen and Lübberstedt 2003; Gaur *et al.* 2012) e para estudos de associação baseados em genes candidatos ou possivelmente em genomas completos (Rafalski, 2002). SNPs são trocas de bases únicas que ocorrem em uma posição específica do genoma. Por definição, o alelo menos frequente tem uma abundância de 1% ou mais, caso contrário é referido como uma mutação (Brookes, 1999). Os SNPs apresentam natureza bialélica (em diploides) e podem ser encontrados tanto em regiões expressas quanto em não-expressas (Borém e Caixeta, 2006). Estão espaçados, em média, a cada 500-1000 bases em humanos (Tost e Gut, 2002). Em espécies de plantas, essa distância pode ser bem menor: em amendoeira, pode ser encontrado um SNP a cada 114pb (Wu *et al.* 2008), e em eucalipto, há um SNP a cada 192pb (Novaes *et al.* 2008). Também há as inserções/deleções (indels) de bases únicas que, apesar de não serem estritamente SNPs, elas são quase sempre incluídas nessa definição (Brookes 1999).

Em plantas, os SNPs podem ser usados como marcadores genéticos para muitas aplicações em melhoramento, estudos populacionais, *fingerprint* de germoplasma, mapeamento

genético, associações genótipo/fenótipo e para estudos de clonagem posicional. Tais marcadores têm sido fundamentais para análise de genes e para o entendimento das bases genético-molecular de importantes características vegetais. O desenvolvimento de marcadores SNPs identificados a partir de sequências de ESTs representa um valioso sistema de marcadores no mapeamento de genes candidatos e na identificação da base genética de QTLs de características de importância agronômica.

Mapeamento genético

A base para a construção de mapas genéticos é o conceito de ligação gênica, que leva à percepção de que a frequência de recombinação pode ser usada como medida de distância entre dois genes (Sturtevant, 1913; Coelho e Silva, 2002). O entendimento desta característica e a disponibilidade de marcadores moleculares, aliados a métodos estatísticos, tem possibilitado a construção de mapas genéticos para várias espécies de plantas de interesse econômico, incluindo as que apresentam longo ciclo de vida (Carneiro e Vieira, 2002). Os mapas genéticos são considerados uma das aplicações de maior impacto dos marcadores moleculares para a análise genética e no melhoramento genético vegetal. Os mapas permitem a análise de genomas, desmembrar as características complexas nos vários genes que as determinam, localizar regiões genômicas que afetam caracteres agronômicos e quantificar seus efeitos na característica de interesse. Para programas de melhoramento genético, a inclusão desses dados é de extrema importância (Cruz e Silva, 2009). Três pontos são extremamente importantes para a construção de um mapa de ligação: a população de mapeamento, a seleção de marcadores polimórficos e a análise de ligação entre os marcadores.

Para a população de mapeamento, a escolha dos genitores é o principal passo: quanto mais próximos geneticamente forem os indivíduos, mais difícil será a identificação de polimorfismos

(Paterson *et al.*, 1981). Para maximizar a probabilidade de detecção de locos polimórficos, o ideal é que os genitores não possuam pais em comum e, no caso de espécies com autofecundação, pode-se fazer uso de cruzamentos interespecíficos (Tanksley, 1993). Em mapeamento, a utilização de populações segregantes com alto desequilíbrio de ligação é muito importante, uma vez que a principal causa do desequilíbrio é a ligação física dos locos. Este desequilíbrio é bastante pronunciado em populações de cruzamentos controlados, desta maneira a detecção de ligação entre dois locos fisicamente próximos também é elevada (Tanksley, 1993; Falconer e Mackay, 1996; Lynch e Walsh, 1998).

As populações mais usadas para a construção de mapas de ligação são gerações F_2, populações de retrocruzamentos e cruzamentos entre heterozigotos e linhagens duplo-haploide (Tanksley, 1993). Entretanto, para espécies de ciclo longo, como a seringueira, a obtenção dessas linhagens é inviável. Para possibilitar o mapeamento genético nessas espécies, diferentes estratégias foram desenvolvidas, sendo o duplo pseudo *testcross* uma delas. A estratégia consiste na construção de dois mapas individuais para cada genitor com base em marcadores que segregam em um pai, mas não no outro. Esta técnica foi utilizada, por exemplo, na construção de mapas genéticos para duas espécies de eucalipto (Grattapaglia e Sederoff, 1994) e álamo (Zang *et al.*, 2004). Outra estratégia, considerada mais efetiva, faz uso de métodos de máxima verossimilhança para estimar simultaneamente a ligação e as fases de ligação (Wu *et al.*, 2002). Este método foi utilizado com sucesso no mapeamento em cana-de-açúcar (Garcia *et al.*, 2006; Oliveira *et al.*, 2007; Palhares *et al.*, 2012) e maracujá (Oliveira *et al.*, 2008).

Marcadores moleculares e mapeamento em seringueira

1. Marcadores moleculares

Marcadores do tipo minissatélites (Besse *et al.*, 1993a), RFLPs (Besse *et al.*, 1993b), mtDNA RFLPs (Luo *et al.*, 1995), RAPDs (Venkatachalam *et al.,* 2001) e AFLPs (Lespinasse *et al.,* 2000a) foram desenvolvidos e utilizados em *Hevea.* Low *et al.* (1996) descreveram pela primeira vez marcadores do tipo microssatélite em seringueira, através de pesquisas em bancos de dados de algumas sequências do gênero, e esses marcadores demonstraram o alto nível de polimorfismo em várias espécies de *Hevea*. Os trabalhos seguintes se basearam na construção de bibliotecas enriquecidas em sequências repetitivas, levando à identificação de um grande número de marcadores microssatélites (Atan *et al.*, 1996; Seguin *et al.*, 2003; Roy *et al.*, 2004; Saha *et al.*, 2005; Souza *et al.*, 2009; Le Guen *et al.*, 2010; Mantello *et al.*, 2012). Yu *et al.* (2011) desenvolveram marcadores SSR através de um biblioteca genômica construída pelo método SAM (*Selectively Amplified Microsatellite*), que gera *fingerprints* de SSR multi-loco e as bandas obtidas são sequenciadas para a detecção de regiões microssatélite.

Feng *et al.* (2009) desenvolveram os primeiros marcadores EST-SSR para seringueira a partir de bancos de dados de domínio público. Estes marcadores, além de terem amplificado em diferentes espécies de *Hevea*, também apresentaram transferibilidade para outros gêneros. Após a publicação deste trabalho, novos marcadores EST-SSR para seringueira foram desenvolvidos apenas a partir de 2012 (Li *et al.*, 2012; Cubry *et al.*, 2014), provavelmente devido ao número limitado de sequências EST disponíveis nesse período.

Pootakham *et al.* (2011) relataram o primeiro trabalho de desenvolvimento de marcadores SNP em seringueira, onde um total de 5.883 supostas posições polimórficas foram descobertas em sequências EST, e apenas dez marcadores SNP bialélicos foram validados. Depois deste

trabalho, foram necessários três anos para que novos marcadores SNP para *H. brasiliensis* fossem desenvolvidos (Salgado *et al.*, 2014; Silva *et al.*, 2014; Mantello *et al.*, *in press*).

2. Mapeamento

A construção de um mapa genético em *Hevea* requer metodologia específica devido à alta heterozigozidade da cultura e seu ciclo de vida. O cruzamento entre dois parentais heterozigotos em *Hevea* pode produzir informação sobre quatro alelos, que segregaram futuramente. O primeiro mapa genético de ligação molecular de *Hevea brasiliensis* foi formulado com a ajuda de marcadores RFLPs, AFLPs, microssatélites e isoenzimas (Lespinasse *et al.,* 2000a).

Os genitores usados foram PB 260 (PB 5/51 x PB 49) e RO 38 (F 4542 x AVROS 363), e um duplo pseudo *testcross* foi realizado, resultando na construção de mapas individuais para cada genitor. Os marcadores homólogos que segregaram em ambos os pais foram usados para construir um único mapa. O mapa sintético de F_1 possui 717 marcas distribuídas entre 18 grupos de ligação (LGs): 301 RFLPs, 388 AFLPs, 18 microssatélites e 10 enzimas. O tamanho genético dos 18 cromossomos foi razoavelmente homogêneo, sendo de 120 cM o tamanho médio por cromossomo. Foi observado o agrupamento de muitos marcadores AFLP em determinadas regiões, o que pode ser consequência de reduzida frequência de recombinação nestas áreas. Contudo, os marcadores RFLP foram bem distribuídos entre os 18 LGs. Uma vez que o clone PB 260 é suscetível e o clone RO 38 é resistente à infecção por *M. ulei*, esta população de mapeamento e os mapas construídos foram utilizados para a detecção de QTLs de resistência a SALB: oito QTLs para tolerância foram identificados (Lespinasse *et al.*, 2000b) e um grande QTL (*M13-lbn*) responsável por 36 a 89% da variância fenotípica para resistência foi localizado no LG 13 do mapa de RO 38 (Le Guen *et al.*, 2003).

Em 2011, dois novos mapas genéticos foram construídos para seringueira, e a estratégia de pseudo *testcross* foi utilizada para a construção dos mapas. Le Guen *et al.* (2011), com o intuito de mapear QTLs para resistência a *M. ulei*, desenvolveu um mapa genético para o cruzamento entre os clones PB 260 e MDF 180 (resistente ao fungo), utilizando marcadores SSR e AFLP. O mapa sintético possui 18 LGs contendo 383 marcas no total, e 2.441 cM de tamanho. Em MDF 180, um gene qualitativo (*M15md*) foi localizado no LG 15 e quatro QTLs menores foram identificados, sendo que dois deles apresentam interação epistática com o gene. O segundo mapa foi construído por Triwitayakorn *et al.* (2011) para uma população resultante do cruzamento entre os clones RRIM 600 (Tjir 1 x PB 86) e RRII 105 (Tjir 1 x Gl 1). O mapa possui 97 marcadores SSR distribuídos em 23 LGs, e tamanho total de 842,9 cM.

Em 2013, Souza *et al.*construíram um mapa genético integrado de uma população F_1 segregante resultante do cruzamento de PB 217 (PB 5/51 x PB 69) e PR 255 (Tjir 1 x PR 107), utilizando a estratégia de análise multiponto que permite a avaliação conjunta de diferentes tipos de marcadores com diferentes padrões de segregação. O mapa possui 284 locos microssatélite, distribuídos em 23 LGs com tamanho total de 2.792,8 cM e densidade de um marcador a cada 9,8 cM. O mapa foi, então, utilizado para a identificação de QTLs envolvidos com caracteres de crescimento (altura e circunferência) em dois diferentes períodos: inverno e verão. Foram identificados 18 QTLs para altura e circunferência, sendo que os QTLs detectados durante o verão eram diferentes dos QTLs detectados durante o inverno, o que sugere a presença de regulação gênica diferencial entre as estações. Este trabalho foi o primeiro a identificar QTLs em seringueira para características não relacionadas à resistência ao *M. ulei*.

Capítulo II

Objetivos

Geral:

Construção de um mapa funcional em seringueira (*H. brasiliensis*) para auxiliar na localização de genes e regiões que controlam características de interesse ao melhoramento genético da seringueira.

Específicos:

- Construção de bibliotecas de cDNA a partir de diferentes tecidos e tratamentos;
- Sequenciamento de, pelo menos, 10.000 clones para geração de um banco de sequências EST de seringueira;
- Análise e anotação das sequências EST obtidas;
- Identificação de SSRs e SNPs nas sequências expressas presentes nas bibliotecas de cDNA;
- Desenvolvimento de marcadores EST-SSRs e SNPs gene-direcionados;
- Caracterização dos marcadores SNPs e EST-SSRs através da genotipagem de clones divergentes fenotipicamente e os genitores da população de mapeamento;
- Genotipagem da população F_1 segregante;
- Construção do mapa funcional para *H. brasiliensis*.

Capítulo III

Capítulo III

"Leaf-, panel- and latex-expressed sequenced tags from the rubber tree (*Hevea brasiliensis*) under cold-stressed and suboptimal growing conditions: the development of gene-targeted functional markers for stress response."

Carla Cristina da Silva

Camila Campos Mantello

Tatiana de Campos

Lívia Moura de Souza

Paulo de Souza Gonçalves

Anete Pereira de Souza

Molecular Breeding (2014)

Mol Breeding
DOI 10.1007/s11032-014-0095-2

Leaf-, panel- and latex-expressed sequenced tags from the rubber tree (*Hevea brasiliensis*) under cold-stressed and suboptimal growing conditions: the development of gene-targeted functional markers for stress response

Carla C. Silva · Camila C. Mantello · Tatiana Campos · Livia M. Souza · Paulo S. Gonçalves · Anete P. Souza

Received: 17 December 2013 / Accepted: 17 April 2014
© The Author(s) 2014. This article is published with open access at Springerlink.com

Abstract *Hevea brasiliensis* is a native species of the Amazon Basin of South America and the primary source of natural rubber worldwide. Due to the occurrence of South American Leaf Blight disease in this area, rubber plantations have been extended to suboptimal regions. Rubber tree breeding is time-consuming and expensive, but molecular markers can serve as a tool for early evaluation, thus reducing time and costs. In this work, we constructed six different cDNA libraries with the aim of developing gene-targeted molecular markers for the rubber tree. A total of 8,263 reads were assembled, generating 5,025 unigenes that were analyzed; 912 expressed sequence tags (ESTs) represented new transcripts, and two sequences were highly up-regulated by cold stress. These unigenes were scanned for microsatellite (SSR) regions and single nucleotide polymorphisms (SNPs). In total, 169 novel EST-SSR markers were developed; 138 loci were polymorphic in the rubber tree, and 98 % presented transferability to six other *Hevea* species. Locus duplication was observed in *H. brasiliensis* and other species. Additionally, 43 SNP markers

Electronic supplementary material The online version of this article (doi:10.1007/s11032-014-0095-2) contains supplementary material, which is available to authorized users.

C. C. Silva (✉) · C. C. Mantello · L. M. Souza · A. P. Souza (✉)
Centro de Biologia Molecular e Engenharia Genética (CBMEG), Universidade Estadual de Campinas (UNICAMP), Cidade Universitária Zeferino Vaz, CP 6010, Campinas, SP CEP 13083-970, Brazil
e-mail: silvacbio@yahoo.com.br

A. P. Souza
e-mail: anete@unicamp.br

C. C. Mantello
e-mail: camila.mantello@gmail.com

L. M. Souza
e-mail: liviamoura31@gmail.com

T. Campos
Centro de Pesquisa Agroflorestal do Acre (CPAFAC), Embrapa, Rodovia BR-364, km 14, CP 321, Rio Branco, AC CEP 69900-970, Brazil
e-mail: tatiana.campos@embrapa.br

P. S. Gonçalves
Instituto Agronômico de Campinas (IAC), CP 28, Campinas, SP CEP 13012-970, Brazil
e-mail: paulog@iac.sp.gov.br

A. P. Souza
Departamento de Biologia Vegetal, Instituto de Biologia, Universidade Estadual de Campinas (UNICAMP), Cidade Universitária Zeferino Vaz, CP 6109, Campinas, SP CEP 13083-970, Brazil

Published online: 29 April 2014

in 13 sequences that showed similarity to proteins involved in stress response, latex biosynthesis and developmental processes were characterized. cDNA libraries are a rich source of SSR and SNP markers and enable the identification of new transcripts. The new markers developed here will be a valuable resource for linkage mapping, QTL identification and other studies in the rubber tree and can also be used to evaluate the genetic variability of other *Hevea* species, which are valuable assets in rubber tree breeding.

Keywords *Hevea* · Rubber tree · cDNA library · Molecular markers · Microsatellite · SNPs

Introduction

Rubber tree [*Hevea brasiliensis* (Willd. ex Adr. de Juss.) Muell-Arg.], a native species of the Amazon rainforests, is the primary commercial source of natural rubber. *H. brasiliensis* is a diploid ($2n = 36$, $n = 18$), perennial, cross-pollinated and monoecious tropical tree that belongs to the Euphorbiaceae family. The genus *Hevea* is composed of 11 inter-crossable species, of which *H. brasiliensis* is the most economically important (Pires et al. 2002; Gonçalves et al. 1990). Natural rubber is used in a large variety of products due to its flexibility, resistance, plasticity, impermeability and insulating properties (Mooibroek and Cornish 2000). In 2012, 11.327 million tons of natural rubber were produced worldwide to meet a demand of 11.005 million tons (International Rubber Study Group (IRSG) 2013); by the year 2020, economists predict that the demand for natural rubber will surpass its production by thousands of tons (Burger and Smith 1997).

The Amazon Basin presents a suitable climate for crop development, but the occurrence of South American Leaf Blight (SALB), which is caused by the fungus *Microcyclus ulei* (P. Henn) v. Arx, limits rubber production in the area (Pushparajah 2001). This problem has led to the expansion of rubber plantations to suboptimal areas, such as northeastern India, Vietnam, southern China and the southern plateau of Brazil. In addition to new conditions for crop development, these new areas of production often present stressful conditions, such as low temperatures, high altitudes, typhoons and dry periods, and all of these factors affect latex production (Pushparajah 1983; Priyadarshan and Gonçalves 2003). Rubber breeding programs aim to identify clones that are adapted to these stress conditions (Priyadarshan and Gonçalves 2003). However, rubber tree breeding is time-consuming and expensive because it can take more than 20 years to develop a new variety (Gonçalves and Fontes 2012). The generation of molecular markers can enable the early detection of the target genotype by marker-assisted selection (MAS), thus reducing the length of the breeding period and its costs.

Expressed sequence tags (ESTs) are a powerful tool for genetic studies such as gene identification, tissue expression profiling and gene mapping; ESTs are also a rich source of molecular markers (Varshney et al. 2005a; Harbers 2008). Microsatellite regions (SSRs; Varshney et al. 2005a) and single nucleotide polymorphisms (SNPs; Rafalski 2002) are the most important and most widely used markers that can be developed from ESTs. Unlike genomic microsatellites, EST-SSRs are likely to be embedded in functional gene sequences because they are gene-targeted markers that have the potential to become functional markers (Andersen and Lübberstedt 2003; Varshney et al. 2005a). Although EST-derived SSR markers are less polymorphic than genomic SSRs, they are better defined (Varshney et al. 2005a) and exhibit higher rates of transferability to related species (Varshney et al. 2005b; Feng et al. 2009). SNPs are highly abundant in the genome and can be used for the construction of high-resolution maps due to their frequency and distribution (Andersen and Lübberstedt 2003; Gaur et al. 2012). EST-derived SNP markers are also gene-targeted and can be directly involved in the expression of a desirable trait; therefore, they are a tool for marker-assisted selection (Liu et al. 2012; Xia et al. 2012) and are useful for genetic studies such as candidate gene mapping and gene-based association studies.

H. brasiliensis EST studies have been carried out over the last 10 years (Ko et al. 2003; Chow et al. 2007), but large-scale EST studies of the rubber tree have only been initiated in the last few years (Xia et al. 2011; Triwitayakorn et al. 2011; Li et al. 2012). The development of EST-derived SSR markers for the rubber tree has also only recently been reported (Feng et al. 2009; Triwitayakorn et al. 2011; Li et al. 2012), and only ten SNP markers have been reported thus far (Pootakham et al. 2011). In the present work, cDNA libraries of cold-stressed clones and different tissues

from the rubber tree were constructed for the development of EST-SSR and SNP markers.

Methods

Plant materials

Clonal graftings of *Hevea brasiliensis* clones PB 217, PR 255, GT 1 and IAN 873 were subjected to a 24-h cold treatment in a Thermo Forma Diurnal Growth Chamber (model 3740; Thermo Scientific Inc., USA) and maintained at 8 °C with a 12-h photoperiod. This treatment was performed to promote the expression of genes involved in cold response and for the development of molecular markers related to this stress condition. Clones PB 217 (high rubber yield potential and cold sensitive) and PR 255 (tolerant to injury and cold) are the parents of a mapping population (Souza et al. 2013), and clones GT 1 and IAN 873 showed cold tolerance in the field (Gonçalves PS, personal communication). The leaves were wrapped in tinfoil prior to collection to prevent transcript redundancy. The leaves were sampled at intervals of 0, 6, 10 and 24 h; immediately frozen in liquid nitrogen; and stored at −80 °C until use. For the panel and latex libraries, samples were collected from 16-year-old tree clones of PB 217, PR 255, GT 1, PB 235, RRIM 701 and IAN 873, and leaves of the same clones were collected from the rubber tree germplasm. Clones GT 1, PB 235 and RRIM 701 are the parents of two mapping populations that are being evaluated in our laboratory, and all clones used are recommended for cultivation in São Paulo State (Gonçalves PS, personal communication). This number of clones was used to increase the chance of detecting SNPs related to stress conditions for mapping in the F1 populations under evaluation. All of the samples were frozen immediately on dry ice and stored at −80 °C prior to RNA extraction.

To characterize the microsatellite markers, 18 contrasting *H. brasiliensis* genotypes were selected, including clones PB 217, PR 255, GT 1, PB 235 and RRIM 701. Furthermore, six species from the genus *Hevea* were included to assay the transferability of the SSR markers. The SNP markers were characterized using the clones listed above in addition to 18 other *H. brasiliensis* genotypes. Thus, a total of 36 clones were used to validate the polymorphic positions (Online Resource 1—Table S1). All of the *H. brasiliensis* samples were collected at the Rubber Research Center of the Agronomic Institute of Campinas (IAC), Votuporanga, São Paulo, Brazil (latitude: 20°25′22″S; longitude: 49°58′22″W), which is a suboptimal region for rubber plantations. The samples of the other species were obtained from the West Amazon Agroforestry Research Center (Embrapa Amazônia Ocidental), Amazonas, Brazil.

RNA preparation, cDNA library construction and EST generation

Total RNA was extracted from the leaves, panel and latex following the protocol described by Chang et al. (1993) and treated with RNAse-free DNAse I (Qiagen Inc., USA). Equal amounts of total RNA were pooled (up to 5 μg) according to tissue (leaf, panel and latex) and time of sampling (6-, 10- and 24-h cold treatment). The In-fusion SMARTer cDNA Library Construction kit (Clontech Laboratories Inc., USA) was used to construct the cDNA libraries according to the manufacturer's instructions. The ligation mixtures were transformed into electrocompetent *Escherichia coli* DH10B cells, and colonies were selected using LB-ampicillin plates containing IPTG (isopropylthio-β-galactoside) and X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside). The insert fragment sizes of 15 positive clones from each library were analyzed by PCR amplification using M13 primers.

Sequencing (10-μL reaction mixtures) was carried out from the 5′ end of the inserts using M13 primers and the Big Dye Terminator 3.1 Cycle Sequencing kit (Applied Biosystems Inc., USA). The sequencing reactions were analyzed in a 3500XL DNA ABI PRISM Automatic Sequencer (Applied Biosystems Inc., USA).

DNA extraction

Genomic DNA was extracted from the lyophilized leaf tissues using a modified CTAB method (Doyle and Doyle 1987). The quality and concentration of the DNA were assessed by 1 % agarose gel electrophoresis.

EST sequence processing and analysis

PHRED (Ewing and Green 1998) was used to perform vector and poly(A) removal and to trim low-quality

segments. CLC Genomics Workbench 4 (CLC bio A/S, Denmark) and ChromasPro 1.5 (Technelysium Pty Ltd., Australia) software were used to assemble the high-quality EST sequences into contigs and singletons. A similarity comparison was performed with the ESTs (contigs and singletons) using the BLAST2GO program (Conesa et al. 2005) to search the National Center for Biotechnology Information (NCBI) non-redundant (nr) protein database. Additionally, the BLAST2GO program was used with default parameters to obtain the Gene Ontology (GO; Ashburner et al. 2000) terms for the molecular function, biological process and cellular component categories and to identify the metabolic pathways using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Ogata et al. 1999). Open reading frames (ORFs) were predicted using the OrfPredictor program (Min et al. 2005). All processed EST sequences were deposited into the NCBI dbEST database under accession numbers JZ536145 to JZ544407.

Quantitative RT-PCR analysis

Expression analysis of the cold-stressed cDNA libraries most represented unigenes was performed by quantitative RT-PCR. For the analysis of individual samples, 1 μg of total RNA was used for cDNA synthesis. For the combined samples analysis, equal amounts of total RNA were pooled according to time of sampling (0-, 6-, 10- and 24-h cold treatment), up to a total of 1 μg for cDNA synthesis with a QuantiTect Reverse Transcription Kit (Qiagen Inc., USA), which includes a genomic DNA removal step. The cDNAs were diluted (1:20) in nuclease-free water, and 1 μL was used for the qPCR.

Quantitative RT-PCR was conducted in a CFX384 Real-Time PCR Detection System (Bio-Rad Laboratories Inc., USA) with Maxima SYBR Green qPCR Master Mix (Thermo Scientific Inc., USA) following the manufacturers' instructions and at a final primer concentration of 0.3 μM. The reaction conditions were as follows: 95 °C for 10 min; 40 cycles at 95 °C for 30 s, 60 °C for 30 s and 72 °C for 30 s. No template controls for any primer pair were included, and each reaction was performed in triplicate.

The evaluated sequences were similar to NAD(P)H-quinone oxidoreductase subunit H (NADH), chloroplast photosystem II 10 kDa polypeptide (PsbR), a hypothetical protein (HYPOT), ATP synthase CF0 C subunit (CF0) and indole-3-acetic acid-induced protein (ARG2-1 and ARG2-2). The glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and the eukaryotic translation initiation factor (eIF2; Li et al. 2011) genes were used as reference genes, and the 0-h cold treatment and PB 217 0-h cold treatment samples were used as the controls for gene expression normalization of the combined samples and individual sample analyses, respectively. The presence of single amplicons in the PCR products was confirmed by analyzing their melting curves at temperatures ranging from 65 to 95 °C. The baseline and Cq values were automatically determined, and expression analysis was performed using CFX Manager 2.1 software (Bio-Rad Laboratories Inc., USA). All primer sequences except eIF2 are described in Online Resource 1—Table S2.

Search for putative molecular markers

SSR mining was performed using the SciRoKo software (version 3.3; Kofler et al. 2007) with the "Perfect; MISA-mode" search mode and default settings. A sequence was defined as a microsatellite region if it exhibited the following characteristics: six repeats for dinucleotides; five repeats for trinucleotides; and four repeats for tetra-, penta- and hexanucleotides.

Contigs with a minimum of fourfold coverage were utilized for SNP mining using the CLC Genomics Workbench software (CLC bio A/S, Denmark). The minimum quality of the central base and the average quality of the surrounding bases were set at 20, and putative SNPs were annotated when the less-represented allele was present in at least two EST sequences up to a minimum frequency of 10 %. The candidate SNPs were classified according to the type of single-base substitution and visually localized into the probable exonic and untranslated (UTR) regions if the EST had a BLASTX hit.

EST-SSR marker characterization and species transferability

PRIMER3 software (Rozen and Skaletsky 2000) was used to design EST-SSR primer pairs from the flanking sequences. The target amplicon size was set to 100–300 bp. The optimal annealing temperature was set to 60 °C, and the optimal primer length was set to 20 bp. For SSR genotyping and characterization, we

used a 4300 DNA Analyzer (LI-COR Biosciences, USA), an Advance FS96 dsDNA Fragment Analyzer (Advanced Analytical Technologies Inc., USA) and 6 % denaturing polyacrylamide gels with silver staining (Creste et al. 2001).

For the analysis performed with the 4300 DNA Analyzer (LI-COR Biosciences, USA), the M13F sequence was added to the 5′ end of the forward primer of 115 primer pairs. PCR amplification was performed as described by Le Guen et al. (2011) with the following modifications: (TD1) ten amplification cycles with a 0.5 °C decrease in annealing temperature per cycle, starting at 57 °C (95 °C for 1 min, 57 °C for 30 s and 72 °C for 1 min); (TD2) ten amplification cycles with a 1 °C decrease in annealing temperature per cycle, starting at 65 °C (95 °C for 1 min, 65 °C for 30 s and 72 °C for 1 min); and (TD3) ten amplification cycles with a 1 °C decrease in annealing temperature per cycle, starting at 62 °C (95 °C for 1 min, 62 °C for 30 s and 72 °C for 1 min).

For the SSR characterization using an Advance FS96 dsDNA Fragment Analyzer (Advanced Analytical Technologies Inc., USA) and silver-stained 6 % denaturing polyacrylamide gels (Creste et al. 2001), the amplification reactions were performed as follows: denaturation at 95 °C for 3 min followed by 35 amplification cycles (95 °C for 1 min, specific annealing temperature for 45 s and 72 °C for 1 min) and a final extension step at 72 °C for 5 min. All of the amplification products were verified by 3 % agarose gel electrophoresis.

The allelic polymorphic information content (PIC) of the SSRs was calculated using the following formula:

$$\mathrm{PIC} = 1 - \sum_{i=1}^{n} p_i^2 - \sum_{i=1}^{n} \sum_{j=i+1}^{n} 2p_i^2 p_j^2$$

where n is the number of alleles of the marker among the set of genotypes used for the characterization of the SSR polymorphism and p_i and p_j are the frequencies of alleles i and j, respectively. The TFPGA program (Miller 1997) was used to calculate the expected and observed heterozygosities.

SNP marker validation and characterization

Sequences that showed similarity to known proteins were chosen for the validation of SNPs in 36 *H. brasiliensis* genotypes (Online Resource 1—Table S1). Primer pairs were designed using PRIMER3 software (Rozen and Skaletsky 2000) to validate at least one SNP present in the EST sequence. The target amplicon size depended on the position and number of putative SNP(s). The optimal annealing temperature was set to 60 °C, and the optimal primer length was set to 20 bp. PCR amplification was carried out in 20-μL reaction mixtures, each containing 25 ng of genomic DNA, 2 μM $MgCl_2$, 0.2 μM dNTPs, 1× reaction buffer, 0.2 μM of each primer and 0.5 U of Pfu DNA polymerase (recombinant; Thermo Scientific Inc., USA), which was used to reduce the rate of nucleotide incorporation errors. PCR was performed using the following thermal cycling conditions: denaturation at 95 °C for 3 min followed by 35 amplification cycles (95 °C for 30 s, specific annealing temperature for 30 s and 72 °C for 2 or 3 min) with a final extension step at 72 °C for 10 min. The PCR products were verified by 1.5 % agarose gel electrophoresis.

For sequencing, the amplicons were purified using a solution containing 20 % PEG8000 (w/v) and 2.5 M NaCl solution. The sequencing of the amplicons was carried out bidirectionally (forward and reverse) in a 10-μL reaction mixture using the Big Dye Terminator 3.1 Cycle Sequencing kit (Applied Biosystems Inc., USA). The sequencing reactions were analyzed using a 3500XL DNA ABI PRISM Automatic Sequencer (Life Technologies Corporation, USA). The chromatograms were aligned to the reference sequence using ChromasPro 1.5 software, and the SNPs were identified as overlapping nucleotide peaks. The expected and observed heterozygosities of the polymorphic positions and their PIC values were calculated using the same methods used for EST-SSR marker characterization.

Results and discussion

Library construction, characterization, sequencing and contig assembly

To develop gene-targeted molecular markers, six standard cDNA libraries were constructed from the leaves of cold-stressed and panel, latex and leaf tissues of different rubber tree clones (see "Methods"). Colony PCR revealed that the cDNA inserts ranged from 200 bp to 2.8 kb in length. A total of 10,080

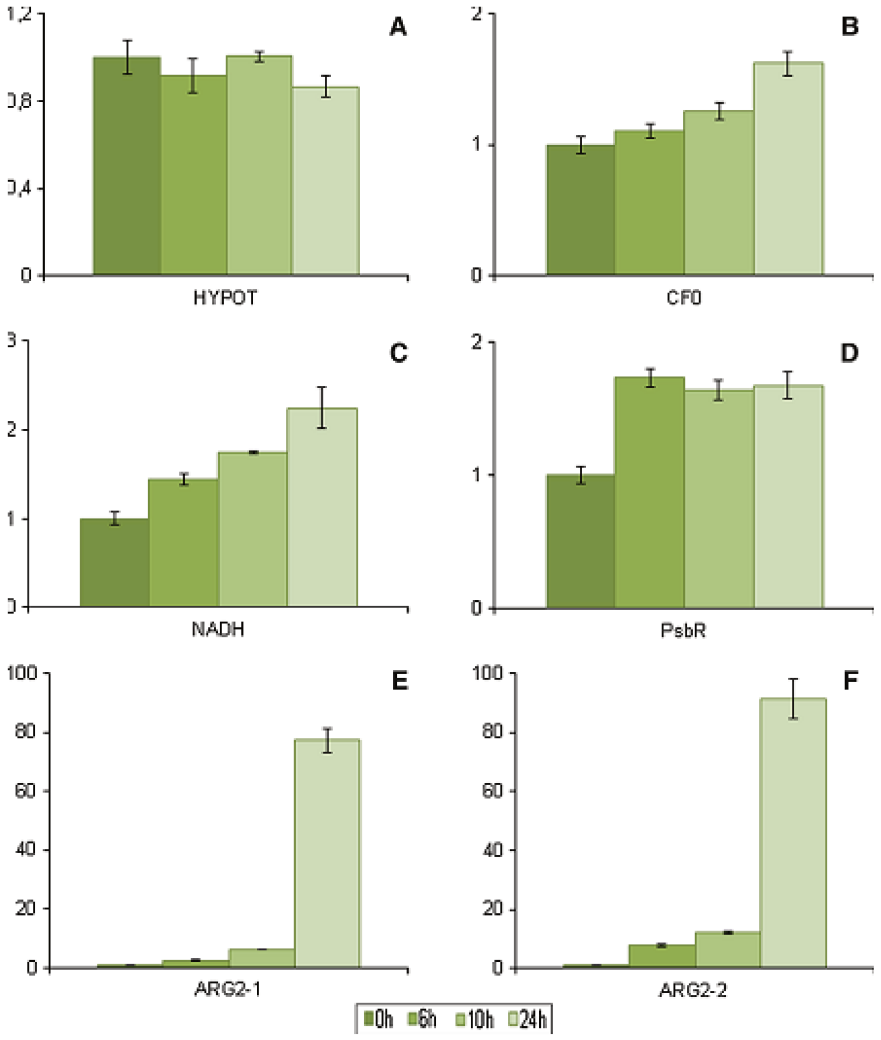


Fig. 1 Expression analysis of the most highly represented sequences in the cold-stressed leaf libraries **a** hypothetical protein (HYPOT), **b** ATP synthase CF0 C subunit (CF0), **c** NAD(P)H-quinone oxidoreductase subunit H (NADH), **d** chloroplast photosystem II 10 kDa polypeptide (PsbR), **e** and **f** indole-3-acetic acid-induced proteins (ARG2-1 and ARG2-2)

clones of the cDNA libraries, consisting of all the clones from the cold-stressed leaf libraries (cold-6 h and cold-10 h: 1,824 clones each; cold-24 h: 2,496 clones) and randomly chosen clones from the panel, latex and leaf libraries (1,536, 1,632 and 768 clones, respectively), were subjected to sequencing. After removing the vector, adaptor, low-quality and short sequences (<150 bp) as well as all possible contaminating sequences, a total of 8,263 (82 %) EST sequences were generated, with an average length of 664 bp. The percentage of redundant sequences was approximately 41 %.

CLC Genomics Workbench 4 (CLC bio A/S, Denmark) and Chromas Pro 1.5 (Technelysium Pty Ltd., Australia) were used to assemble the ESTs, generating 5,025 unigenes composed of 816 contigs and 4,209 singletons. The majority of these sequences (3,640; 72.4 %) had lengths of between 200 bp and 1 kb. The average length of the unigene sequences was 715 bp, which is longer than the *Hevea brasiliensis* ESTs from the panel (Li et al. 2012), leaf and latex (Xia et al. 2011) sequences obtained from RNA-seq experiments and longer than the ESTs obtained from *M. ulei*-infected leaves (Garcia et al. 2011). The

contigs were formed from between two (413 contigs) and 162 (one contig) reads, and the average number of reads per contig was 4.97 (Online Resource 1—Table S3). Among the 5,025 sequences, 4,991 (99.3 %) contained ORFs, and after read assembly, the sequence redundancy decreased to 1.5 %.

The most highly represented genes in the EST sequences

EST assembly was performed for each library prior to the assembly of all sequences, and the number of reads present in the generated contigs was evaluated. In the cold-6 h library, two contigs formed by 33 reads each were the most highly expressed based on the number of ESTs in a contig. Both contigs exhibited similarity to proteins related to photosynthesis (NAD(P)H-quinone oxidoreductase subunit H, *e*-value $7e^{-148}$, and chloroplast photosystem II 10 kDa polypeptide (PsbR), *e*-value $2e^{-65}$). A hypothetical protein (*e*-value $2.83e^{-14}$) was the most highly expressed sequence in the cold-10 h library (36 reads) and the second most highly expressed in the cold-24 h library (37 reads) and was highly represented in the leaf library (63 reads), which suggests that this transcript might be important for the leaf tissue, although no probable function has been described for this transcript. The second most highly represented sequence in the cold-10 h library was identified by 25 reads and was similar to ATP synthase CF0 C subunit (*e*-value $3e^{-28}$), which is also involved in photosynthesis.

The cold-24 h library also presented highly represented (≥20 reads) sequences similar to proteins involved in photosynthesis, but the most highly expressed sequence in this library (with 44 reads) was similar to indole-3-acetic acid-induced protein (ARG2; *e*-value $7e^{-31}$). There was also a second contig formed by seven reads that matched the same protein. This sequence was also present at a very low frequency in the cold-6 h, cold-10 h and panel libraries (two, four and two reads, respectively). This protein is associated with the stress response in plants (Yamamoto et al. 1992; Seki et al. 2001).

Sequences similar to the rubber elongation factor protein (REF; *e*-value $6e^{-91}$; 88 reads) and pro-hevein (54 reads; *e*-value $2e^{-144}$) were the most highly represented ESTs in the latex library. REF and small rubber particle protein (SRPP), which was represented by two contigs composed of 36 and 26 reads, are believed to be involved in latex biosynthesis (Dennis and Light 1989; Oh et al. 1999) and are highly expressed in latex and laticifers (Ko et al. 2003; Chow et al. 2007). Pro-hevein is believed to be involved in the defense response because it is able to bind to chitin and inhibit fungal growth (Van Parijs et al. 1991); pro-hevein is also abundant in latex and laticifers (Ko et al. 2003; Chow et al. 2007). The panel library sequences seemed to be less redundant than the sequences from the other libraries because most of the contigs of the panel library were formed by fewer than seven reads. The most represented sequences in this library presented similarity to non-specific lipid transfer protein (17 reads; *e*-value $1e^{-40}$) and metallothionein 3-like protein (15 reads; *e*-value $1e^{-24}$).

The contigs were also analyzed after the assembly of all ESTs. Because a majority of the ESTs originated from the leaf tissues, reads exhibiting similarity to chloroplast sequences, such as the proteins of photosystems I and II, were highly abundant, as expected. These reads constituted 13 of the 27 most highly expressed unigenes, considering the number of ESTs in a contig (≥20 reads). These 27 unigenes accounted for 15.1 % of the 8,263 sequences obtained. The following contigs did not show similarity to any protein in the GenBank database but were similar to other EST sequences: contig 366, which had 47 reads from the cold-stressed leaf libraries; contig 42; and contig 142. Sequences similar to REF, SRPP and pro-hevein were also highly represented, mostly due to the sequences from the latex library. All of the 27 most expressed unigenes are described in Online Resource 1—Table S4.

Expression analysis of the most highly represented genes in the cold-stressed leaf libraries

Quantitative RT-PCR analysis was performed to examine the expression of the sequences that presented the highest number of reads in each of the cold-stressed leaf libraries (6, 10 and 24 h) plus the additional sequence also similar to the ARG2 protein. The expression of the hypothetical protein did not differ among the combined samples (Fig. 1a); however, there was a 5.4-fold decrease in the expression of the clone GT 1 6 h sample (Fig. S1A). Clone GT 1 appeared to maintain a high level of expression of this sequence when compared to the other clones, and low temperature had an effect on its transcription. Because

no function has been assigned to this protein, the processes that it may be involved in remain to be investigated. All other sequences evaluated in the combined samples presented some level of up-regulation. ATP synthase CF0 C subunit is a membrane component of the chloroplast ATP synthase complex (Seelert et al. 2000). A 1.6-fold increase in its expression at 24-h cold exposure was observed in comparison with the 0 h sample (Fig. 1b). Chilling-stress impairs the function of the ATP synthase complex through the production of reactive oxygen species (ROS; Prasad et al. 1994; Buchert et al. 2012). The increase in the expression of this subunit may be due to the required replenishment of novel ATP synthase complexes in chloroplasts. Clones PB 217 and PR 255 presented an increase in the expression of this sequence from the 10 and 6 h samples, respectively. There was a 1.8-fold decrease in its expression in the GT 1 sample at 6 h. The expression level remained similar thenceforth, and clone IAN 873 did not show a significant difference among samples (Fig. S1B). Clones GT 1 and IAN 873 may exhibit a better protection of their photosynthetic apparatus against oxidative stress than clones PB 217 and PR 255. This protection may also be related to their tolerance to low temperatures.

The NAD(P)H-quinone oxidoreductase subunit H showed a gradual increase in its expression among all combined samples (1.4-fold in 6 h, 1.7-fold in 10 h and 2.2-fold in 24 h; Fig. 1c) and was also up-regulated in the individual clone samples (Fig. S1C). The NAD(P)H-quinone oxidoreductase complex seems to be important for cold-stress response. In tobacco, mutants with deleted subunits of this complex present a more severe phenotype under chilling stress than wild-type plants (Li et al. 2004; Wang et al. 2006). This expression increase might be a response to cope with the low temperature.

PsbR protein is important for the proper assembly of the oxygen-evolving complex of the photosystem II (PSII) complex (Suorsa et al. 2006) and demonstrated a 1.7-fold increase in its expression in leaf tissues after 6 h low temperature exposure, which was maintained in the other combined samples (Fig. 1d). A similar pattern was observed in the clones' individual samples, except for IAN 873, which presented a 1.6-fold increase in expression in the 6 h sample, whereas the 10 and 24 h samples showed the same expression level as the 0 h sample (Fig. S1D). *Arabidopsis thaliana PsbR* mutants present higher PSII excitation pressure than wild-type plants (Suorsa et al. 2006), which is also caused by low temperature (Huner et al. 1998). The increase in *PsbR* expression after chilling stress may be related to the photoprotection of the PSII complex.

ARG2 is a late embryogenesis abundant (LEA)-type protein, a group of hydrophilic proteins present in a wide range of plant species that are induced by water deficit caused by desiccation, cold and osmotic stress (Wang et al. 2003). Two sequences that presented similarity to the ARG2 protein were evaluated by qPCR: ARG2-1 (24 h—44 reads), for which SNP markers were developed (see below), and ARG2-2 (24 h—7 reads); both sequences were up-regulated. ARG2-1, in comparison with the 0 h combined sample, presented an increase in expression of 2.6-fold in the 6 h sample, 6.3-fold in the 10 h sample and 77-fold in the 24 h sample (Fig. 1e). An analysis of the individual samples demonstrated a large difference in the up-regulation of this sequence between the PB 217 clone 24 h sample (38-fold) and the other clones' 24 h samples (>100-fold) (Fig. S1E). The expression of ARG2-2 was up-regulated in the combined samples by 7.7-fold, 12-fold and 91-fold after 6, 10 and 24 h of cold exposure (Fig. 1f), respectively. This sequence was also less up-regulated in the PB 217 clone 24 h sample (80-fold) when compared to the other clones' 24 h samples. Clone PR 255 presented a 395-fold expression increase in ARG2-2 after 24 h of cold exposure, while GT 1 showed a 249-fold up-regulation and IAN 873 presented a 171-fold increase in ARG2-2 expression (Fig. S1F).

ARG2 transcription increased in the presence of indole-3-acetic acid (IAA; Yamamoto et al. 1992). It had been previously observed that the IAA levels increase in *A. thaliana* (Gray et al. 1998) and rice (Du et al. 2013) under temperature stress. Our results suggest that these sequences were up-regulated due to the prolonged exposure of the rubber tree clones to low temperatures. Similarly, the increase in the expression of these sequences might be due to increased levels of IAA in the clones. In addition, the higher expression of these sequences in clones PR 255, GT 1 and IAN 873 might be related to their better tolerance to low temperatures when compared to clone PB 217. These results may require further detailed analysis because only one individual of each clone was evaluated. Nevertheless, these data demonstrate that the

individuals used in our work exhibit different responses to cold stress.

Functional annotation

To identify unigenes that were likely to encode proteins with known functions, sequences were subjected to BLASTX analysis against the GenBank non-redundant protein database using BLAST2GO software (Conesa et al. 2005). A total of 3,456 (68.8 %) unigenes showed significant similarity (*e*-value $<1e^{-06}$) to at least one unknown, hypothetical or expressed protein, and 3,404 (98.5 %) unigenes had *e*-values less than $1e^{-10}$. Proteins from *Ricinus communis* accounted for the majority of BLASTX hits for these sequences (1,771), followed by proteins from *Populus trichocarpa* (709), *Vitis vinifera* (240) and *H. brasiliensis* (236) (Online Resource 1—Fig. S2). *R. communis* also belongs to the Euphorbiaceae family, and the GenBank database contains information on a large number of *R. communis* and *P. trichocarpa* proteins (68,409 and 104,560 proteins, respectively, as of November 2013); this abundance of data may explain the similarity between and number of hits for *R. communis* and *P. trichocarpa* proteins. In comparison, the GenBank database contains fewer rubber tree proteins (1,120), which may have led to the limited number of *H. brasiliensis* protein hits. The remaining 1,569 (31.2 %) sequences that did not show a significant similarity to any protein in the database and therefore could not be annotated were subjected to BLASTN analysis (*e*-value $<1e^{-06}$). Of these, 657 ESTs were similar to ESTs that are present in the GenBank database, leaving 912 (18.2 %) sequences with no hits. The sequence length may affect the annotation success of reads. Among the sequences without hits, 358 ESTs had lengths shorter than 400 bp, accounting for 41.7 % of the analyzed unigenes. In contrast, 9.6 % (196) of the sequences longer than 800 bp did not match any sequence in the GenBank database. These ESTs may be considered to be novel or specific transcripts of *H. brasiliensis*.

A functional classification of the unigenes was performed according to the GO database using the BLAST2GO program. The terms were organized into three main categories: molecular function (MF), biological process (BP) and cellular component (CC). Of the 3,456 sequences analyzed, 2,503 (72.4 %) were annotated with 8,232 terms, including 3,867 MFs, 2,743 BPs and 1,622 CCs. In the MF category, binding (1,410; 36.5 %) and catalytic activity (1,283; 33.2 %) were the largest categories, followed by structural molecule activity (189; 4.9 %; Fig. 2a). Cellular metabolic process (1,124; 41.0 %), biosynthetic process (585; 21.3 %) and gene expression (381; 13.9 %) were the most highly represented categories in the BP category (Fig. 2b). In the CC category, the most represented categories were cytoplasm (669; 41.2 %), membrane (446; 27.5 %) and the protein complex and non-membrane-bounded organelle, which shared the same number of sequences (215; 13.3 %; Fig. 2c).

In addition to the GO terms, 1,050 enzyme commission (EC) numbers were attributed to 910 unigenes. To establish associations with biological pathways, KEGG pathway analysis was used to map the 1,050 EC numbers into 116 KEGG pathways. Some sequences were mapped to more than one pathway. The majority of the mapped unigenes were related to carbohydrate metabolism pathways, and starch and sucrose metabolism (48 sequences, 22 enzymes) and glycolysis/gluconeogenesis (36 sequences, 16 enzymes) were the most highly represented in this group. The purine metabolism pathway had the largest number of unigenes (51), representing 18 enzymes. Other highly represented categories included carbon fixation in photosynthetic organisms (45 sequences, 18 enzymes), in which the majority of the mapped sequences originated from libraries of leaf tissues; nitrogen metabolism (40, 12); and oxidative phosphorylation (34, 7). Enzymes involved in secondary metabolism-related pathways were also identified; a majority of these were involved in phenylpropanoid biosynthesis (29, 4) and flavonoid biosynthesis (26, 13). Sixteen of the sequences that were mapped to the phenylpropanoid biosynthesis pathway were annotated as lactoperoxidase (EC 1.11.1.7), which is involved in lignin biosynthesis and stress response (Dixon and Paiva 1995), and the majority of the sequences mapped to the flavonoid biosynthesis pathway were identified from the panel library. The 16 most represented pathways (≥15 unigenes) are listed in Online Resource 1—Table S5. Sequences that showed similarity to genes involved in the 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway (six sequences, five enzymes; Sando et al. 2008), which is believed to be one of the metabolic pathways involved in rubber biosynthesis, were also identified.

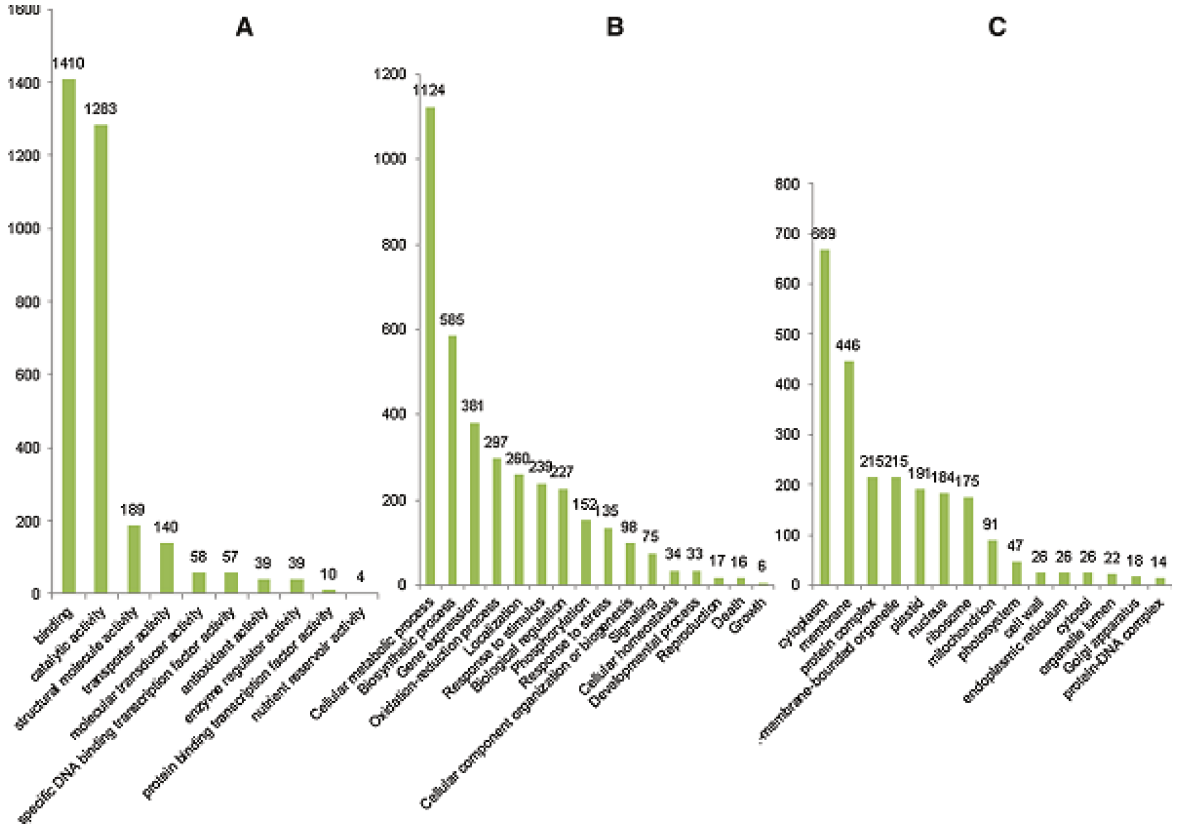


Fig. 2 Functional category distribution of the annotated unigenes **a** molecular function (MF), **b** biological process (BP) and **c** cellular component (CC)

The characterization and development of the EST-derived SSR markers

SciRoKo software (version 3.3; Kofler et al. 2007) was used in MISA mode to identify microsatellite regions in the 5,025 sequences analyzed. A total of 588 microsatellite regions were identified, and 527 unigenes (10.5 %) contained at least one of the defined motifs (di-, tri-, tetra-, penta- or hexanucleotides). Of these sequences, 58 contained more than one SSR region. Nineteen microsatellites were present as compound SSRs; however, each motif was considered as a single repeat. On average, one SSR was found per 6.1 kb in the 3,578,774 bp of EST sequences that were searched, which is higher than the frequency reported for rice (one SSR per 3.4 kb; Cardle et al. 2000) and castor bean (1.23 kb; Pranavi et al. 2011) and similar to that of *Jatropha curcas* (6.0 kb; Yadav et al. 2011) but lower than that of bread wheat (9.2 kb; Gupta et al. 2003), soybean (7.4 kb), maize (8.1 kb), tomato (11.1 kb), poplar (14.0 kb) and cotton (20.0 kb; Cardle et al. 2000). The proportion of SSR unit sizes was not evenly distributed. Among the 588 SSR regions found, dinucleotide (302, 51.4 %) motifs were the most frequent, followed by tri- (205, 34.9 %), tetra- (39, 6.6 %), penta- (27, 4.6 %) and hexanucleotide motifs (15, 2.6 %). Several previous EST studies have shown that trinucleotide repeats are the most abundant microsatellite type in the expressed sequences of many plants (Cardle et al. 2000; Gupta et al. 2003; Clepet et al. 2011), whereas other studies showed that dinucleotide motifs were more frequent (Pranavi et al. 2011; Yadav et al. 2011). In our analysis, dinucleotide repeats were the most highly represented SSR motif in the unigenes, in agreement with previous studies of rubber tree ESTs (Feng et al. 2009; Triwitayakorn et al. 2011; Li et al. 2012). Nevertheless, these differences in the distribution and

frequency of SSR regions among the different crops may be due to the SSR search criteria, the number of total ESTs and bases searched and the software tools used (Varshney et al. 2005a), making a direct comparison of the abundance and frequency of SSR motifs difficult.

The most frequent type of dinucleotide motif was AG/TC (167, 55.3 %), followed by AT/TA (107, 35.4 %) and AC/GT (27, 8.9 %). The AAG/TTC motif (82, 40.0 %) was the most common trinucleotide repeat (Online Resource 1—Table S6). Among the dinucleotide repeats, the GC/CG motif was identified only once. This repeat motif appears to be rare in most plant ESTs; GC repeats were found at a very low frequency in previous studies (Pranavi et al. 2011; Yadav et al. 2011; Clepet et al. 2011), including those involving the rubber tree (Feng et al. 2009; Triwitayakorn et al. 2011; Li et al. 2012). This low frequency of CG repeats and CCG repeats in EST sequences may be due to the methylation of CpG sequences, which may inhibit transcription (Lister et al. 2008).

SSR-containing sequences that showed similarity to proteins in the GenBank database were preferentially chosen for the development of microsatellite markers. A total of 211 primer pairs were designed based on 202 SSR-containing sequences; of these primer pairs, 18 were designed from nine ESTs bearing two different SSR regions. These sequences were subjected to a BLASTN (e-value $<1e^{-06}$) search against the GenBank database to remove possible redundancies with published SSRs. Only one sequence, bearing a dinucleotide motif, was identical to a previously published locus, and this sequence was thus removed from this study. Of the 210 non-redundant primer pairs, 83, 97, 19, 3 and 8 primer pairs were designed to amplify di-, tri-, tetra-, penta- and hexanucleotide motifs, respectively. The M13 tail was added to the 5′ end of 115 forward primers for the fluorescence analysis of these loci using a 4300 DNA Analyzer (LI-COR Biosciences, USA). The other 95 primer pairs were analyzed via silver-stained 6 % acrylamide gel electrophoresis (Creste et al. 2001) or capillary electrophoresis using an Advance FS96 dsDNA Fragment Analyzer (Advanced Analytical Technologies Inc., USA). After the primers for fluorescence analysis were tested in different touchdown programs and the primers for acrylamide and capillary electrophoresis analysis were tested at different annealing temperatures, 196 primer pairs (93.3 %) produced amplicons, as shown by 3 % agarose gel electrophoresis. Of the 196 working primer pairs, 178 amplified PCR products of the expected sizes and 18 produced larger PCR products than expected. Of these 18 products, ten were amplicons that ranged from 500 bp to 1 kb. Because the primers were designed based on expressed sequences and genomic DNA was used for amplification, the existence of these larger PCR products suggests the presence of intronic regions in the genomic sequences.

Eighteen *H. brasiliensis* genotypes (Online Resource 1—Table S1) were used to assess the polymorphism of the 186 primer pairs that produced amplicons smaller than 400 bp. Seventeen of these 186 primer pairs showed non-specific amplification and could not be evaluated; thus, 169 primer pairs were analyzed, and 137 were polymorphic among the 18 genotypes tested. Among the 161 ESTs used to design these primer pairs, 141 were annotated as known or uncharacterized proteins; thus, 147 SSR loci (87.0 %) may be associated with possible functional genes.

One of the polymorphic primer pairs (EHBp-23) amplified two distinct polymorphic regions, resulting in 138 polymorphic loci. The EST used to design this primer pair was obtained from the panel library and showed similarity at the nucleotide sequence level to the protein aquaporin, which is involved in water transport and belongs to the large major intrinsic protein (MIP) family of transmembrane channels. Several genes encoding aquaporins have been discovered in plants (Chaumont et al. 2005). Because aquaporins are a highly conserved group of proteins, the two loci amplified may represent different genes that encode aquaporins in *H. brasiliensis*.

The expected (H_e) and observed (H_o) heterozygosities and polymorphic information content (PIC) values could only be calculated for 136 of the 138 polymorphic loci because two loci contained duplicated alleles in several genotypes (see below). H_e and H_o ranged from 0.0556 to 0.89 (average 0.4648) and 0–1 (average 0.3622), respectively, and the mean number of alleles was 3.7 (2–10 alleles). Nevertheless, at several loci, an exclusive allele was present in only one or two of the genotypes tested, and in most cases, clones RRIM 809, RRIC 100, IAC 306 or RRII105 were the bearers of this allele. Expressed regions show a greater level of DNA sequence conservation

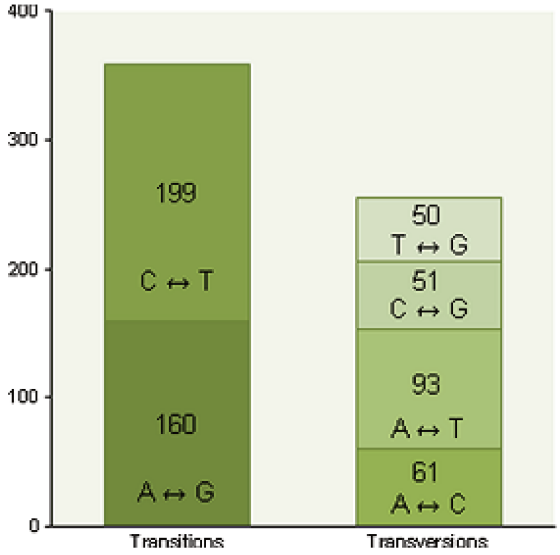


Fig. 3 Distribution of the putative SNPs identified in this study

(Varshney et al. 2005a), which explains both the lower number of alleles observed in EST-SSRs compared to genomic SSR markers (Souza et al. 2009; Mantello et al. 2012) and the presence of rare alleles. The PIC values ranged from 0.0526 to 0.8512, with an average of 0.4036, indicating that this group of EST-SSR markers presents a moderate level of informativeness. Although this group of markers presented a low to moderate level of polymorphism, these markers will be useful for genetic mapping, determining the linkage between markers and genes for important traits, QTL mapping, marker-assisted selection and functional analysis of candidate genes in the rubber tree, among other information. All 169 primer pairs are fully described in Online Resource 2.

The two loci that could not be analyzed exhibited duplicated alleles in several rubber tree accessions. Of the 18 genotypes of *H. brasiliensis* used in this study, nine showed duplicated alleles for the SSR region amplified by primer pair EHBc-103 and 15 contained duplicated alleles for the region amplified by primer pair EHBp-27, resulting in six and ten different alleles, respectively. Plants of the *Hevea* genus exhibit diploid behavior, mainly forming bivalents during meiosis (Bouharmont 1960; Majumder 1964; Ong 1975); however, cytogenetics studies have revealed two loci on different chromosomes bearing the same rDNA sequence, suggesting a possible allotetraploid origin (Leitch et al. 1998). Although the species has a diploid

Table 1 Summary statistics of the SNP analysis

	Numbers
Number of contigs used	121
Number of contigs with putative SNPs	104
Total contig length	109,512 bp
Average number of reads per contig	19.9
Number of identified putative SNPs	614
Average number of SNPs per contig	5.03
SNP frequency	1/178 bp
Localization	
5′ UTR	45
Exon	376
3′ UTR	150
No-hit	43

genome, molecular marker analyses revealed locus duplication in *H. brasiliensis* (Lespinasse et al. 2000; Mantello et al. 2012). These duplicated loci are likely due to the allotetraploid origin of the species (Lespinasse et al. 2000). The EST sequences used to design these primer pairs showed similarity to proteins that mapped to different chromosomes in *Glycine max*, a diploidized tetraploid (Shoemaker et al. 1996). Although no potential ancestor has yet been described for the rubber tree (Leitch et al. 1998), our results support the hypothesis of a polyploid origin followed by a diploidization event.

Cross-species transferability

The transferability of the 169 primer pairs to other *Hevea* species was also evaluated; 167 (98.8 %) primers successfully produced amplicons in at least one of the six species tested: 164 (97.0 %) were amplified in *H. guianensis*, 158 (93.5 %) in *H. nitida* and *H. benthamiana*, 157 (92.9 %) in *H. rigidifolia*, 156 (92.3 %) in *H. pauciflora* and 148 (87.6 %) in *H. camargoana* (Online Resource 2). In addition to this high transferability, the number of alleles per locus increased to 5.6 when compared with *H. brasiliensis* alone, revealing the presence of novel alleles. As expected, the transferability rates of the EST-SSR markers were higher than those of the genomic SSR markers (Varshney et al. 2005a, b; Feng et al. 2009; Mantello et al. 2012). The EST-SSR markers are likely related to gene units, and as such, their potential for inter-specific transferability is greater (Gupta et al.

Table 2 Validated and characterized SNP markers in the rubber tree

Name	H_e	H_o	PIC	BLASTX hit	Primer sequence (5′–3′)	Expected length (bp)	Observed length (bp)	Ta (°C)
Hb-SNP1-292	0.4917	0.5882	0.3671	Copper chaperone ($4e^{-37}$) [*Populus alba* × *Populus glandulosa*]	F: TGATTTGAAGGAGCAAAAGG R: GGCATACGACCATAAAGCAC	353	~350	60
Hb-SNP1-349	0.4860	0.5588	0.3642					
Hb-SNP1-362	0.1925	0.2121	0.1716					
Hb-SNP1-455	0.3566	0.3939	0.2896					
Hb-SNP1-459	0.4909	0.5758	0.3633					
Hb-SNP1-469	0.4909	0.5758	0.3633					
Hb-SNP2-497	0.5034	0.2581	0.3726	Membrane steroid-binding protein ($2e^{-86}$) [*Arabidopsis thaliana*]	F: ATGGACCTGGTGGACCTTAT R: CACCAAGTACATGCATCCAA	425	~430	61.6
Hb-SNP2-569	0.4779	0.3939	0.3599					
Hb-SNP2-624	0.5055	0.1111	0.3742					
Hb-SNP3-531	0.4571	0.6857	0.3491	Class IV endochitinase ($7e^{-94}$) [*Vitis vinifera*]	F: TTCTAAACGGGAAGTTGCTG R: ATTGGCGTACGTGCATTTAT	512	~600	63.2
Hb-SNP3-535	0.486	0.7941	0.3642					
Hb-SNP4-387	0.487	0.5714	0.3648	NAC domain protein ($1e^{-65}$) [*Populus trichocarpa*]	F: TTCAGTACCGAAGTTGCACA R: AACCCACCCTTAAAACTACCA	432	~430	60
Hb-SNP4-572	0.5056	0.5429	0.3742					
Hb-SNP4-590	0.5004	0.4118	0.3715					
Hb-SNP5-458	0.4539	0.25	0.3457	Lipid transfer protein precursor ($4e^{-47}$) [*Gossypium hirsutum*]	F: GCTTGAAAAGCTCTGCTGCT R: TGGGCTCTCTAACACCCATT	221	~400	63.2
Hb-SNP6-84	0.4539	0.25	0.3457	Pro-hevein ($5e^{-145}$) [*H. brasiliensis*]	F: AATTGGGAAGAAATGGGAAG R: TGGCTCAAATGCCATTATTT	804	~880	60
Hb-SNP6-452	0.4539	0.25	0.3457					
Hb-SNP6-774	0.3398	0.2813	0.2957					
Hb-SNP7-167	0.3566	0.3333	0.2896	Indole-3-acetic acid-induced protein ARG2, putative ($1e^{-31}$) [*Ricinus communis*]	F: GCTTCTTCCTTCCTTGTTCC R: TTTCATTCACAAGCTCAGCA	696	~700	60
Hb-SNP7-251	0.0294	0.0294	0.0286					
Hb-SNP7-273	0.4917	0.6471	0.3671					
Hb-SNP7-544	0.3951	0.4706	0.3135					
Hb-SNP7-556	0.3951	0.4706	0.3135					
Hb-SNP7-562	0.4087	0.4412	0.3216					
Hb-SNP7-586	0.3951	0.4706	0.3135					
Hb-SNP8-475	0.1874	0.2059	0.1676	GDP-L-galactose phosphorylase ($3e^{-131}$) [*Malpighia glabra*]	F: AAGCTCTTGGGGAAGTGAGT R: CAAGTCCTGAGCATCGTTCT	242	~250	63.2

Table 2 continued

Name	H_e	H_o	PIC	BLASTX hit	Primer sequence (5′–3′)	Expected length (bp)	Observed length (bp)	Ta (°C)
Hb-SNP9-720	0.1549	0.1667	0.1411	Rubber elongation factor ($5e^{-100}$) [*H. brasiliensis*]	F: GCATTGTTCCTCCAATTGTC R: TTGGCCATTTATTCCCATTA	308	~300	60
Hb-SNP10-149	0.3176	0.12	0.289	Major allergen Pru ar, putative ($6e^{-71}$) [*Ricinus communis*]	F: AAATTTTTGTTTAGACTCGCTCT R: AAGCCATCATGGGTGTTTTA	832	~900	57.5
Hb-SNP10-152	0.1502	0	0.1364					
Hb-SNP10-221	0.1502	0	0.1364					
Hb-SNP10-267	0.1448	0	0.1319					
Hb-SNP10-302	0.1448	0	0.1319					
Hb-SNP10-332	0.1448	0	0.1319					
Hb-SNP11-60	0.4329	0.3478	0.3338	Small rubber particle protein ($1e^{-63}$) [*H. brasiliensis*]	F: TTGGAATTTGTACAAGCGACT R: CAAACACCTTGGCAATTCTC	400	~700	63.2
Hb-SNP12-225	0.2967	0.2258	0.2493	Latex cystatin ($1e^{-53}$) [*H. brasiliensis*]	F: GAAAGTGGTGAATGCAAAGC R: AGATGTAACCATTCATAAA TATCCA	368	~1,500	60
Hb-SNP12-360	0.2544	0.0417	0.2181					
Hb-SNP12-417	0.3883	0.2727	0.3093					
Hb-SNP12-430	0.2821	0.1515	0.2392					
Hb-SNP12-439	0.3075	0.2273	0.2417					
Hb-SNP12-453	0.4188	0.2581	0.3272					
Hb-SNP13-258	0.4543	0.2647	0.3475	Rubber elongation factor protein ($3e^{-91}$) [*H. brasiliensis*]	F: CATCCATCCATCCGAATTT R: TCAAGGACGCATCTATCCA	466	~470	60
Hb-SNP13-309	0.4611	0.2121	0.351					

2003; Andersen and Lübberstedt 2003; Varshney et al. 2005a). Our results indicate that the SSR flanking regions are conserved among the *Hevea* species, consistent with other studies (Feng et al. 2009; Souza et al. 2009; Mantello et al. 2012). Additionally, the *Hevea* genus is considered to be a complex of species without genetic reproductive barriers between them. This characteristic of the genus has been used in rubber tree breeding programs for inter-specific crosses, mostly with the aim of developing clones that are tolerant or resistant to SALB (Gonçalves and Fontes 2012). These loci could be used to follow gene introgressions in the resulting hybrids and to provide markers for comparative mapping and for population structure and genetic analyses of these species.

As described previously (Souza et al. 2009; Mantello et al. 2012), the other *Hevea* species also presented duplicated alleles. The primer pair EHBc-103 amplified more than two alleles in all species, except *H. pauciflora*, and *H. rigidifolia* was the only one that did not have more than two alleles in the region amplified by EHBp-27. Taken together with previous results, our results suggest that the other *Hevea* species analyzed also have duplicated loci in their genomes.

Development and characterization of the SNP markers

For SNP discovery, CLC Genomics Workbench software (CLC Bio A/S) was used to assemble the 8,263 EST sequences obtained, generating 816 contigs. From these contigs, 121 (composed of 2,429 reads with a total length of 109,512 bp) had coverage equal to or greater than four and were therefore analyzed for SNP identification. A total of 614 putative SNPs (359 transitions and 255 transversions) were identified in 104 contigs. Transitions are the most common SNP variant in several plants (Wu et al. 2008; Novaes et al. 2008; Clepet et al. 2011; Gaur et al. 2012), including the rubber tree (Pootakham et al. 2011). The most frequent variation was C ↔ T, and the least frequent variation was G ↔ T (Fig. 3). On average, an SNP was identified every 178 bp, which is comparable to the rate observed in other plant species, such as almond (1 SNP/114 bp; Wu et al. 2008) and *Eucalyptus grandis* (1 SNP/192 bp; Novaes et al. 2008). However, the SNP frequency in *H. brasiliensis* was reported to be approximately eightfold higher (1 SNP/1.5 kb; Pootakham et al. 2011) than the frequency determined in the present work. The sequences of two clones that shared a parent were analyzed in the previous study (Pootakham et al. 2011), whereas in the current work, ESTs from six clones with different parents were used for assembly and SNP mining. Most likely, this difference in SNP frequency was due to the different numbers and genotypes of *H. brasiliensis* used. Most of the putative SNPs identified (376, 61.2 %) were located in exonic regions, whereas 45 (7.3 %) were identified in 5′ UTRs, 150 (24.4 %) were in 3′ UTRs and 43 (7.0 %) were located in "no-hit" sequences (Table 1).

Of the 104 contigs in which SNPs were identified, 16 sequences that showed similarity to proteins involved in the stress response, rubber biosynthesis and developmental processes were chosen to validate an SNP subset. Sixteen primer pairs flanking 61 putative SNPs were designed with expected amplicon sizes of approximately 200–800 bp. Thirteen primer pairs amplified products in the 36 *H. brasiliensis* genotypes used (Online Resource 1—Table S1), and the amplicons observed by 1.5 % agarose gel electrophoresis ranged from 250 bp to 1.5 kb, with five loci showing PCR products that were longer than the expected length. Sequencing and a comparison of the chromatograms to the original sequence revealed the presence of intronic regions in these loci; the smallest intron was approximately 80 bp in length, and the longest was approximately 1.1 kb. A total of 46 putative SNPs were evaluated by visual inspection of overlapping nucleotide peaks in the chromatograms, and 43 positions (91.5 %) were validated in the 36 genotypes used. The majority of these polymorphic positions (23 SNPs) were located in probable 3′ UTRs, 18 were in exonic regions (12 non-synonymous SNPs and six synonymous SNPs), and one was located in a probable 5′ UTR. One of the non-synonymous polymorphic positions (Hb-SNP2-624) was not a true SNP but, rather, an 18-bp duplication that was considered an SNP in the CLC Genomics Workbench software alignment analysis. In the predicted translation, this duplication causes a repetition of six amino acids in the protein sequence.

The 43 polymorphic positions presented H_e values ranging from 0.0294 to 0.5056, with an average of 0.3566. H_o varied between 0 and 0.7941, with an average of 0.3256. The average PIC value was 0.2807, and the PIC values ranged from 0.0286 to 0.3742

(Table 2); these markers were therefore less informative than the EST-SSR markers developed in this work. This difference between SNP and SSR markers was also reported in other species (Jones et al. 2007; Emanuelli et al. 2013). Although SNPs are the most abundant variation found in plant genomes, they are usually limited to two alleles per locus, even when considering that a SNP locus theoretically has four different alleles. This limitation can be overcome by using multiple SNP loci to construct haplotypes, which may increase the genetic diversity and information content of these markers (Rafalski 2002; Jones et al. 2007; Emanuelli et al. 2013). Nine of the 13 loci analyzed here contain two or more SNPs in their sequence, and these SNPs together can be considered a haplotype for the locus, thereby compensating for the low informativeness of a single SNP.

The SNP markers developed in this work may be powerful tools for genetic and QTL mapping because they are likely located in sequences that encode proteins related to the stress response and developmental processes in the rubber tree. Some of these SNPs might also be associated with desired traits and could therefore be used as functional markers for marker-assisted selection in *H. brasiliensis* breeding programs.

Conclusions

The use of EST sequences for the development of molecular markers enables the generation of gene-associated markers, thereby providing a means for the construction of more informative high-density genetic maps. Although cDNA libraries yield a lower number of sequences than NGS technologies, our work shows that these libraries remain a rich source of SSR and SNP markers and can reveal the existence of unknown transcripts. The EST-SSR and SNP markers developed here are a valuable resource for genetic diversity studies, linkage mapping, QTL identification, gene-based association studies, functional analysis of candidate genes and marker-assisted selection in rubber tree genetic studies and breeding programs. These markers are also a powerful tool for evaluating the genetic variability of other *Hevea* species, which are a valuable asset for the genetic improvement of cultivated *H. brasiliensis* clones.

Acknowledgments The authors thank the Brazilian Agricultural Research Corporation (EMBRAPA—Amazônia Ocidental) and the Agronomic Institute for donating the analyzed *Hevea* germplasm. This work was supported by grants from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, proc. 2007/50562-4; 2012/50491-8) and scholarships to CCS (2009/52975-0), CCM (2008/55974-1, 2011/50188-0) and LMS (2012/05473-1). PSG and APS are recipients of a research fellowship from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

Conflict of interest None.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

Andersen JR, Lübberstedt T (2003) Functional markers in plants. Trends Plant Sci 8(11):554–560

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene ontology: tool for the unification of biology. The gene ontology consortium. Nat Genet 25(1):25–29

Bouharmont J (1960) Recherches taxonomiques et caryologiques chez quelques espèces du genre Hevea. In: INEAC (ed) Série Scientifique, 85th edn, pp 64

Buchert F, Schober Y, Rompp A, Richter ML, Forreiter C (2012) Reactive oxygen species affect ATP hydrolysis by targeting a highly conserved amino acid cluster in the thylakoid ATP synthase γ subunit. Bioch Bioph Acta 1817:2038–2048

Burger K, Smith HP (1997) The natural rubber market. Review, analysis, politics and outlook. Woodhead Publishing Ltd., Cambridge

Cardle L, Ramsay L, Milbourne D, Macaulay M, Marshall D, Waugh R (2000) Computational and experimental characterization of physically clustered simple sequence repeats in plants. Genetics 156(2):847–854

Chang S, Puryear J, Cairney J (1993) A single and efficient method for isolating RNA from pine trees. Plant Mol Biol 11:113–116

Chaumont F, Moshelion M, Daniels MJ (2005) Regulation of plant aquaporin activity. Biol Cell 97(10):749–764

Chow KS, Wan KL, Isa MNM, Bahari A, Tan SH, Harikrishna K, Yeang HY (2007) Insights into rubber biosynthesis from transcriptome analysis of *Hevea brasiliensis* latex. J Exp Bot 58(10):2429–2440

Clepet C, Joobeur T, Zheng Y, Jublot D, Huang M, Truniger V, Boualem A, Hernandez-Gonzalez ME, Dolcet-Sanjuan R, Portnoy V, Mascarell-Creus A, Caño-Delgado AI, Katzir N, Bendahmane A, Giovannoni JJ, Aranda MA, Garcia-Mas J, Fei Z (2011) Analysis of expressed sequence tags

generated from full-length enriched cDNA libraries of melon. BMC Genom 12:252. doi:10.1186/1471-2164-12-252

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21(18):3674–3676. http://www.blast2go.com

Creste S, Tulmann A, Figueira A (2001) Detection of single sequence repeat polymorphism in denaturating polyacrylamide sequencing gels by silver staining. Plant Mol Biol Rep 19:299–306

Dennis MS, Light DR (1989) Rubber elongation factor from *Hevea brasiliensis*: identification, characterization, and role in rubber biosynthesis. J Biol Chem 264(31):18608–18617

Dixon RA, Paiva NL (1995) Stress-induced phenylpropanoid metabolism. Plant Cell 7(7):1085–1097

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:11–15

Du H, Liu H, Xiong L (2013) Endogenous auxin and jasmonic acid levels are differentially modulated by abiotic stresses in rice. Front Plant Sci 4:397. doi:10.3389/fpls.2013.00397

Emanuelli F, Lorenzi S, Grzeskowiak L, Catalano V, Stefanini M, Troggio M, Myles S, Martinez-Zapater JM, Zyprian E, Moreira FM, Grando MS (2013) Genetic diversity and population structure assessed by SSR and SNP markers in a large germplasm collection of grape. BMC Plant Biol 13:39. doi:10.1186/1471-2229-13-39

Ewing B, Green P (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res 8(3):186–194

Feng SP, Li WG, Huang HS, Wang JY, Wu YT (2009) Development, characterization and cross-species/genera transferability of EST-SSR markers for rubber tree (*Hevea brasiliensis*). Mol Breed 23(1):85–97

Garcia D, Carels N, Koop DM, Araujo de Sousa L, Junior SA, Pujade-Renaud V, Mattos CRR, Cascardo JCM (2011) EST profiling of resistant and susceptible Hevea infected by Microcyclus ulei. Physiol Mol Plant Pathol 76(2):126–137

Gaur R, Azam S, Jeena G, Khan AW, Choudhary S, Jain M, Yadav G, Tyagi AK, Chattopadhyay D, Bhatia S (2012) High-throughput SNP discovery and genotyping for constructing a saturated linkage map of chickpea (*Cicer arietinum* L.). DNA Res 19:357–373

Gonçalves PS, Fontes JRA (2012) Domestication and breeding of rubber tree. In: Borem A, Lopes MTG, Clement CR, Noda H (eds) Domestication and breeding: Amazonian species. UFV, Viçosa, pp 393–420

Gonçalves P, Cardoso M, Ortolani AA (1990) Origem, variabilidade e domesticação da *Hevea*; uma revisão. Pesq Agrop Bras 25(2):135–156

Gray WM, Ostin A, Sandberg G, Romano CP, Estelle M (1998) High temperature promotes auxin-mediated hypocotyl elongation in Arabidopsis. PNAS 95:7197–7202

Gupta PK, Rustgi S, Sharma S, Singh R, Kumar N, Balyan HS (2003) Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. Mol Genet Genomics 270(4):315–323

Harbers M (2008) The current status of cDNA cloning. Genomics 91:232–242

Huner NPA, Oquist G, Sarhan F (1998) Energy balance and acclimation to light and cold. Trends Plant Sci 3(6):224–230

International Rubber Study Group (IRSG) (2013) Statistical summary of world rubber situation. Quarterly statistics. http://www.rubberstudy.com/documents/WebSiteData_2.0.pdf

Jones ES, Sullivan H, Bhattramakki D, Smith JS (2007) A comparison of simple sequence repeat and single nucleotide polymorphism marker technologies for the genotypic analysis of maize (*Zea mays* L.). Theor Appl Genet 115(3):361–371

Ko JH, Chow KS, Han KH (2003) Transcriptome analysis reveals novel features of the molecular events occurring in the laticifers of *Hevea brasiliensis* (para rubber tree). Plant Mol Biol 53(4):479–492

Kofler R, Schlötterer C, Lelley T (2007) SciRoKo: a new tool for whole genome microsatellite search and investigation. Bioinformatics 23(13):1683–1685. http://www.kofler.or.at/bioinformatics/SciRoKo/index.html

Le Guen V, Gay C, Xiong TC, Souza LM, Rodier-Goud M, Seguin M (2011) Development and characterization of 296 new polymorphic microsatellite markers for rubber tree (*Hevea brasiliensis*). Plant Breed 130:294–296

Leitch AR, Lim KY, Leitch IJ, O'Neill M, Chye M, Low F (1998) Molecular cytogenetic studies in rubber, *Hevea brasiliensis* Muell. Arg. (Euphorbiaceae). Genome 41:464–467

Lespinasse D, Rodier-Godier M, Grivet L, Leconte A, Legnate H, Seguin M (2000) A saturated genetic linkage map of rubber tree (*Hevea* spp.) based on RFLP, AFLP, microsatellite, and isozyme markers. Theor Appl Genet 100:127–138

Li XG, Duan W, Meng QW, Zou Q, Zhao SJ (2004) The function of chloroplastic NAD(P)H dehydrogenase in tobacco during chilling stress under low irradiance. Plant Cell Physiol 45(1):103–108

Li H, Qin Y, Xiao X, Tang C (2011) Screening of valid reference genes for real-time RT-PCR data normalization in *Hevea brasiliensis* and expression validation of a sucrose transporter gene HbSUT3. Plant Sci 181:132–139

Li D, Deng Z, Qin B, Liu X, Men Z (2012) De novo assembly and characterization of bark transcriptome using Illumina sequencing and development of EST-SSR markers in rubber tree (*Hevea brasiliensis* Muell. Arg.). BMC Genom 13:192. doi:10.1186/1471-2164-13-192

Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell 133(3):523–536

Liu Y, He Z, Appels R, Xia X (2012) Functional markers in wheat: current status and future prospects. Theor Appl Genet 125:1–10

Majumder SK (1964) Chromosome studies of some species of *Hevea*. J Rubber Res Inst Malays 18:269–273

Mantello CC, Suzuki FI, Souza LM, Gonçalves PS, Souza AP (2012) Microsatellite marker development for the rubber tree (*Hevea brasiliensis*): characterization and cross-amplification in wild *Hevea* species. BMC Res Notes 5:329. doi:10.1186/1756-0500-5-329

Miller MP (1997) Tools for population genetic analysis (TFPGA) 1.3: a windows program for the analysis of allozyme and molecular population genetic data. Computer Software distributed by the author

Min XJ, Butler G, Storms R, Tsang A (2005) OrfPredictor: predicting protein-coding regions in EST-derived sequences. Nucleic Acids Res 33(Suppl 2):W677–W680. doi:10.1093/nar/gki394

Mooibroek H, Cornish K (2000) Alternative sources of natural rubber. Appl Microbiol Biotechnol 53:335–365

Novaes E, Drost DR, Farmerie WG, Pappas GJ Jr, Grattapaglia D, Sederoff RR, Kirst M (2008) High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome. BMC Genom 9:312. doi:10.1186/1471-2164-9-312

Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M (1999) KEGG: Kyoto Encyclopedia of genes and genomes. Nucleic Acids Res 27(1):29–34

Oh SK, Kang H, Shin DH, Yang J, Chow KS, Yeang HY, Wagner B, Breiteneder H, Han KH (1999) Isolation, characterization, and functional analysis of a novel cDNA clone encoding a small rubber particle protein from *Hevea brasiliensis*. J Biol Chem 274(24):17132–17138

Ong SN (1975) Chromosome morphology at pachytene stage in *Hevea brasiliensis*: a preliminary report. In: Sripathi B (ed) Proceedings of the international rubber conference. Rubber Research Institute of Malaysia, Kuala Lumpur, pp 3–12

Pires JM, Secco RS, Gomes JI (2002) Taxonomia e fitogeografia das seringueiras *Hevea* spp. Embrapa amazônia oriental, Belém

Pootakham W, Chanprasert J, Jomchai N, Sangsrakru D, Yoocha T, Therawattanasuk K, Tangphatsornruang S (2011) Single nucleotide polymorphism marker development in the rubber tree, Hevea brasiliensis (Euphorbiaceae). Am J Bot: e337–e338. doi:10.3732/ajb.1100228

Pranavi B, Sitaram G, Yamini KN, Dinesh Kumar V (2011) Development of EST-SSR markers in castor bean (*Ricinus communis* L.) and their utilization for genetic purity testing of hybrids. Genome 54(8):684–691

Prasad TK, Anderson MD, Martin BA, Stewart CR (1994) Evidence for chilling-induced oxidative stress in maize seedlings and a regulatory role for hydrogen peroxide. Plant Cell 6:65–74

Priyadarshan PM, Gonçalves PS (2003) *Hevea* gene pool for breeding. Genet Resour Crop Evol 50:101–114

Pushparajah E (1983) Problems and potentials for establishing *Hevea* under difficult environmental conditions. Plant 59:242–251

Pushparajah E (2001) Natural rubber. In: Last FT (ed) Tree crop ecosystems. Elsevier Science, Amsterdam, pp 379–407. [Goodall DW (Series Editor): Ecosystems of the World Series, Vol 19]

Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol 5:94–100

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S, Totowa NJ (eds) Bioinformatics methods and protocols: methods in molecular biology. Humana Press, New York, pp 365–386

Sando T, Takeno S, Watanabe N, Okumoto H, Kuzuyama T, Yamashita A, Hattori M, Ogasawara N, Fukusaki E, Kobayashi A (2008) Cloning and characterization of the 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway genes of a natural-rubber producing plant, *Hevea brasiliensis*. Biosci Biotechnol Biochem 72(11):2903–2917

Seelert H, Poetsch A, Dencher NA, Engel A, Stahlberg H, Muller DJ (2000) Proton-powered turbine of a plant motor. Nature 405:418–419

Seki M, Narusaka M, Abe H, Kasuga M, Yamaguchi-Shinozaki K, Carninci P, Hayashizaki Y, Shinozaki K (2001) Monitoring the expression pattern of 1300 Arabidopsis genes under drought and cold stresses by using a full-length cDNA microarray. Plant Cell 13:61–72

Shoemaker RC, Polzin K, Labate J, Specht J, Brummer EC, Olson T, Young N, Concibido V, Wilcox J, Tamulonis JP, Kochert G, Boerma HR (1996) Genome duplication in soybean (Glycine subgenus soja). Genetics 144(1):329–338

Souza LM, Mantello CC, Santos MO, Gonçalves PO, Souza AP (2009) Microsatellites from rubber tree (*Hevea brasiliensis*) for genetic diversity analysis and cross amplification in six *Hevea* wild species. Cons Genet Res 1:75–79

Souza LM, Gazaffi R, Mantello CC, Silva CC, Garcia D, Le Guen V, Cardoso SEA, Garcia AAF, Souza AP (2013) QTL mapping of growth-related traits in a full-sib family of rubber tree (Hevea brasiliensis) evaluated in a sub-tropical climate. PLoS One 8(4):e61238. doi:10.1371/journal.pone.0061238

Suorsa M, Sirpio S, Allahverdiyeva Y, Paakkarinen V, Mamedov F, Styring S, Aro EM (2006) PsbR, a missing link in the assembly of the oxygen-evolving complex of plant photosystem II. J Biol Chem 281:145–150

Triwitayakorn K, Chatkulkawin P, Kanjanawattanawong S, Sraphet S, Yoocha T, Sangsrakru D, Chanprasert J, Ngamphiw C, Jomchai N, Therawattanasuk K, Tangphatsornruang S (2011) Transcriptome sequencing of *Hevea brasiliensis* for development of microsatellite markers and construction of a genetic linkage map. DNA Res 18(6):471–482

Van Parijs J, Broekaert WF, Goldstein IJ, Peumans J (1991) Hevein: an antifungal protein from rubber-tree (*Hevea brasiliensis*) latex. Planta 183:258–264

Varshney RK, Graner A, Sorrells ME (2005a) Genic microsatellite markers in plants: features and applications. Trends Biotechnol 23(1):48–55

Varshney RK, Sigmund R, Börner A, Korzun V, Stein N, Sorrells ME, Langridge P, Graner A (2005b) Interspecific transferability and comparative mapping of barley EST-SSR markers in wheat, rye and rice. Plant Sci 168(1):195–202

Wang W, Vinocur B, Altman A (2003) Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. Planta 218:1–14

Wang P, Duan W, Takabayashi A, Endo T, Shikanai T, Ye JY, Mi H (2006) Chloroplastic NAD(P)H dehydrogenase in tobacco leaves functions in alleviation of oxidative damage caused by temperature stress. Plant Physiol 141:465–474

Wu SB, Wirthensohn MG, Hunt P, Gibson JP, Sedgley M (2008) High resolution melting analysis of almond SNPs derived from ESTs. Theor Appl Genet 118(1):1–14

Xia Z, Xu H, Zhai J, Li D, Luo H, He C, Huang X (2011) RNA-Seq analysis and de novo transcriptome assembly of *Hevea brasiliensis*. Plant Mol Biol 77:299–308

Xia Y, Ning Z, Bai G, Li R, Yan G, Siddique KHM, Baum M, Guo P (2012) Allelic variations of a light harvesting chlorophyll A/B-binding protein gene (Lhcb1) associated with agronomic traits in barley. PLoS ONE 7(5):e37573. doi:10.1371/journal.pone.0037573

Yadav HK, Ranjan A, Asif MH, Mantri S, Sawant SV, Tuli R (2011) EST-derived SSR markers in *Jatropha curcas* L.: development, characterization, polymorphism, and transferability across the species/genera. Tree Genet Genomes 7:207–219

Yamamoto KT, Mori H, Imaseki H (1992) Novel mRNA sequences induced by indole-3-acetic acid in sections of elongating hypocotyls of mung bean (*Vigna radiata*). Plant Cell Physiol 33:13–20

Online files

Online Resource 1

Table S1 – *Hevea brasiliensis* genotypes used for the characterizations of EST-SSR and SNP markers

Genotypes[1]/markers SNPs	Parental	Origin
RRIM 600	Tjir 1 x PB 86	Malaysia
RRIM 606	Tjir 1 x PB 49	Malaysia
RRIM 729	RRIM 623 (PB 49 x Pil B 84) x Fx 25 (F 351 x AVROS 49)	Malaysia
RRIM 805	self pollination of RRIM 628 [Tjir 1 x RRIM 527 (Pil B 50 x Pil B 84)]	Malaysia
RRIM 915	RRIM 605 (Tjir 1 x PB 49) x PB 5/51 (PB 56 x PB 24)	Malaysia
RRIM 937	PB 5/51 (PB 56 x PB 24) x RRIM 703 [RRIM 600(Tjir 1 x PB 86) x RRIM 500 (Pil B 84 x Pil A 44)]	Malaysia
PB 260	PB 5/51 (PB 56 x PB 24) x PB 49	Brazil
PB 346	PB 235 [(PB 5/51(PB 56 x PB 24) x PB S/78] x PB 218	Brazil
Fx 4098	PB 86 x FB 110	Brazil
RO 38	Primary clone	Brazil
CMB 104	IRCA 109 [PB 5/51 (PB 56 x PB 24) x RRIM 600((Tjir 1 x PB 86)] x PFB-5	Brazil
CMB 114	IRCA 109 (PB 5/51 (PB 56 x PB 24) x RRIM 600((Tjir 1 x PB 86)] x PFB-5	Brazil
IAC 307	AVROS 1328 (AVROS 214 x AVROS 317) x PR 107	Brazil
IAC 309	RRIM 626 (Tjir 1 x RRIM 600) x Fx 25 (F 351 x AVROS 49)	Brazil
IAC 500	RRIM 600 (Tjir 1 x PB 86)	Brazil
IRCA 230	GT 1 x PB 5/51 (PB 56 x PB 5/78)	Ivory Coast
IRCA 707	PB 235 [PB 5/51(PB 56 x PB 24) x PB S/78)] x MDF 38	Ivory Coast
IRCA 1159	PB 260 [(PB 5/51 (PB 56 x PB 24) x PB 49] x RO 38	Ivory Coast
EST-SSRs and SNPs		
GT 1	Primary clone	Indonesia
PB 235	PB 5/51 (PB 56 x PB 24) x PB S/78	Brazil
PB 217	PB 5/51 (PB 56 x PB 24) (PB 56 x PB 24) x PB 69	Brazil
PB 233	PB S/15 x PB 5/63 (PB 56 x PB 24)	Brazil
PB 311	RRIM 600 (Tjir 1 x PB 86) x PB 235 [PB 5/51 (PB 56 x PB 24) x PB S/78]	Brazil
PC 140	PB 5/51 (PB 56 x PB 24) x RRIM 703 [RRIM 600 (Tjir 1 x PB 24) x RRIM 500 (Pil B 84 x Pil A 44)	Malaysia
PR 255	Tjir 1 x PR 107	Indonesia
RRIM 701	44/553 x RRIM 501(Pil A 44 x Lun N)	Malaysia
RRIM 809	RRIM 600 (Tjir 1 x PB 86) x RRIM 623 (PB 49 x Pil B 84)	Malaysia
RRIM 728	GT 1 x RRIM 623((PB 49 x Pil B 84)	Malaysia

RRIM 913	PB 5/51 (PB 56 x PB 24) x RRIM 623 (PB 49 x Pil B 84)	Malaysia
RRII 105	Tjir 1 x Gl1	India
RRIC 100	RRIC 52 x PB 86	Sri Lanka
IAC 306	AVROS 49 x RRIM 509 (Pil A 44 x Lun N)	Brazil
IAC 313	RRIM 626 [Tjir 1 x RRIM 600(Tjir 1 x PB 86)] x Fx 25 (F 351 x AVROS 49)	Brazil
IAC 318	RRIM 600 (Tjir 1 x PB 86) x Fx 3899 (F 4542 x AVROS 363)	Brazil
IRCA 27	F 4542 x AVROS 363	Ivory Coast
IRCA 209	GT 1 x RRIM 605 (Tjir 1 x PB 49)	Ivory Coast
EST-SSRs		
Hevea guianensis	Wild species	Brazil
Hevea. rigidifolia	Wild species	Brazil
Hevea nítida	Wild species	Brazil
Hevea pauciflora	Wild species	Brazil
Hevea benthamiana	Wild species	Brazil
Hevea camargoana	Wild species	Brazil

(1)IAC – Instituto Agronômico de Campinas; IAN: Instituto Agronômico do Norte; F: Ford (clone primário); Fx: Ford cross; AVROS: Algemene Veriniging Rubberplanters Oostkust Sumatra; Tjir: Tjirandji; RO: Rondônia; RRIM: Rubber Research Institute of Malaysia; Pil: Pilmor; Lun: Lunderston; PB: Prang Besar; GT: Godang Tapen; PR: Proesfstation voor Rubber; RRII: Rubber Research Institute of India; RRIC: Rubber Research Institute of Ceylon; FB: Ford Belém; F: Ford; Gl: Glenshield; IRCA: Institut des Recherches sur e Caoutchouc; PC: Promotion Clone; CMB - Cirad Michelin Brasil; PFB: Pé Franco de Belterra.

Table S2 – Quantitative RT-PCR primer sequences.

Primer	Primer Sequence (5' – 3')	Amplicon Length (bp)
CF0	F - ATTGATTTCTGCCGCTTCC	121
	R - GGGTTGTCTTGCGATACCTTC	
HYPOT	F - TGGAAGGCTAGGGGTTATAGTC	101
	R - TTCCATAAAGGAGCCGAATG	
PsbR	F - ATATGGAGCCAACGTGGATG	109
	R - CACTGCCCATATTGCCAAAC	
NADH	F - GTTCTTTACTTGGGCGGTTG	100
	R - CAATTGTTGTTCCAAAGACTCC	
ARG2-1	F – AACTGCGGAGGAGAAGATTG	93
	R – TCCTCCGCTAAATTCTCTGG	
ARG2-2	F - TGGCTCGCTCTTTCTCAAAC	112
	R - AGGCTGGACACAACTCCTTG	
GAPDH	F - AGTGCACCGATGTTTCTTCC	143
	R - CTTGCTGGCTACCAAATGAG	

CF0 - ATP synthase CF0 C subunit, HYPOT - hypothetical protein, PsbR - chloroplast photosystem II 10 kDa polypeptide, NADH - NAD(P)H-quinone oxidoreductase subunit H, ARG2-1 e 2 - indole-3-acetic acid-induced proteins.

Table S3 – Summary statistics of the EST sequences generated

	Numbers
Total number of clones sequenced	10,080
Number of high-quality EST sequences	8,263 (82%)
Redundancy of ESTs	41%*
Contigs	816
Number of ESTs in contigs (range)	2 – 162

Total number of ESTs in contigs	4,054 (49.1%)
Average number of ESTs per contig	4.97
Average contig length (bp)	826
Singletons	4,209 (50.9%)
Singleton average length (bp)	696
Total unigene sequences	5,025
Redundancy of unigenes	1.5%*

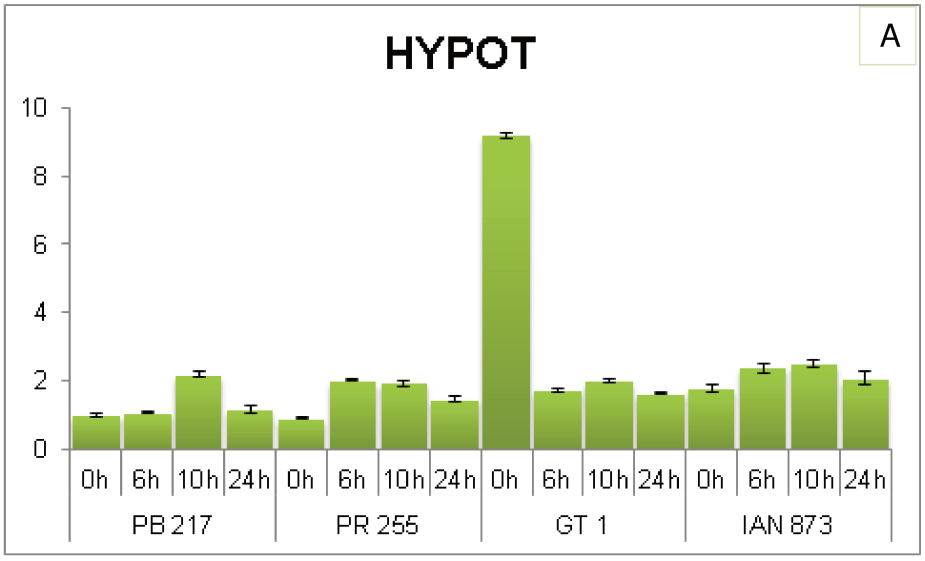
* "No-hit" sequences were excluded from the analysis.

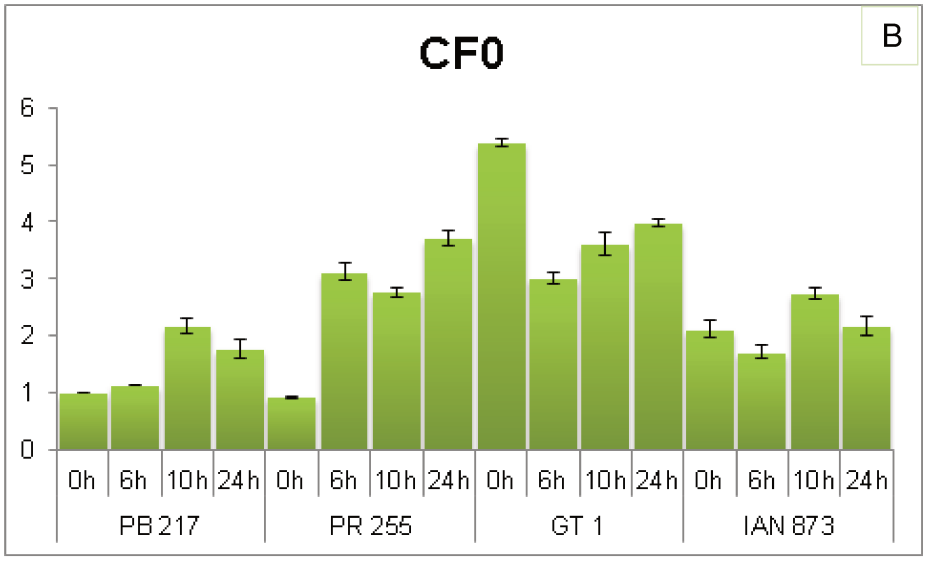
Table S4 – The most highly represented sequences in the cDNA libraries.

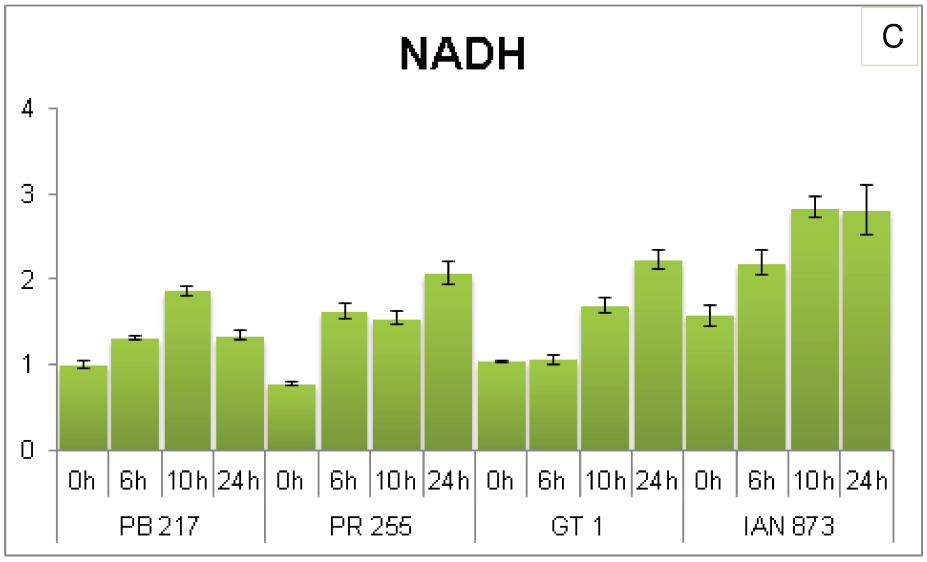
Contig	N° of ESTs	Length (bp)	Sequence description (BlastX)	E-value	% identity
98	162	803	hypothetical protein M569_00588, partial [*Genlisea aurea*]	$2e^{-14}$	82
130	88	913	Rubber elongation factor protein [*H. brasiliensis*] (sp|P15252.2|REF_HEVBR)	$6e^{-91}$	100
66	78	2364	Photosystem I assembly protein Ycf4 (*H. brasiliensis*); Photosystem I subunit VIII (*H. brasiliensis*)	$1e^{-93}$; $3e^{-13}$	99; 100
127	74	665	Chloroplast photosystem II 10 KDa polypeptide (*Jatropha curcas*)	$2e^{-65}$	82
101	58	883	Ribulose-1,5-bisphosphate carboxylase small subunit (*H. brasiliensis*)	$4e^{-122}$	99
52	54	1040	Pro-hevein [*H. brasiliensis*] (P02877.2|HEVE_HEVBR)	$2e^{-144}$	99
348	50	2646	ATP Synthase CF0 A subunit [*H. brasiliensis*] (ADO33548.1)	$6e^{-127}$	99
			ATP synthase CF0 C subunit [Nicotiana tabacum] (NP_054483.1)	$4e^{-28}$	100

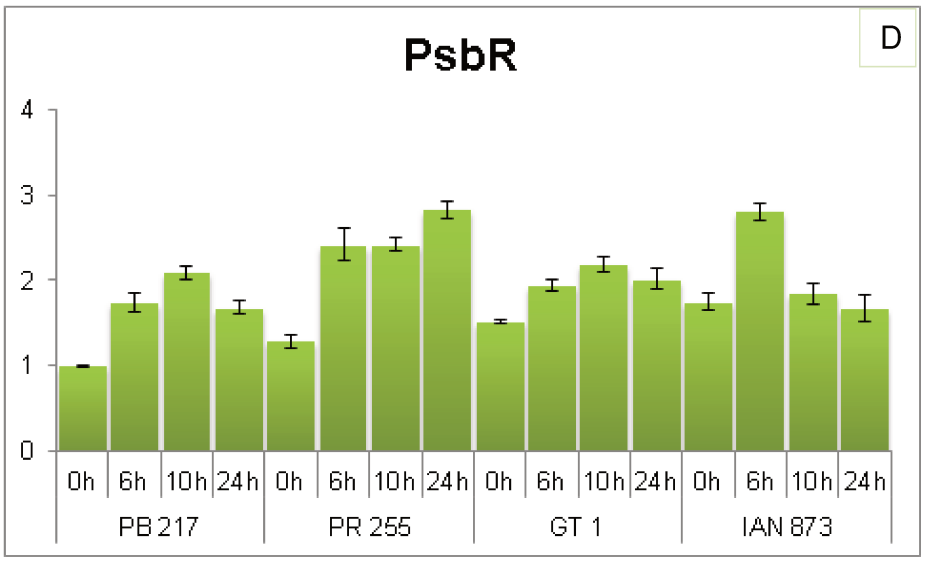
81	47	1094	Indole-3-acetic acid induced protein ARG2, putative [*Ricinus communis*] (EEF39286.1)	$7e^{-31}$	70
364	47	2344	Photosystem II protein K (*Camellia sinensis*); Photosystem II protein I (*Magnolia kwangsiensis*)	$3e^{-20}$; $3e^{-18}$	91; 88
366	47	1194	*Manihot esculenta* EST (DB938126.1)	0	92 (BlastN)
152	46	2585	NAD(P)H-quinone oxidoreductase subunit H [*Medicago truncatula*] (AES86356.1)	$7e^{-148}$	90
246	44	2724	ATP-dependent Clp protease proteolytic subunit [*H. brasiliensis*] (ADO33581.1)	$3e^{-57}$	97
126	39	655	Metallothionein [*Hevea brasiliensis*] (ACT68013.1)	$9e^{-27}$	98
284	37	1190	ATP Synthase CF0 C subunit [*Nicotiana tabacum*] (NP_054483.1)	$3e^{-28}$	100
60	36	1028	Small rubber particle protein [*H. brasiliensis*] (sp|O82803.1|SRPP_HEVBR)	$2e^{-138}$	100
491	35	606	Probable non-specific lipid-transfer protein AKCS9-like [*Glycine Max*] (XP_003555069.1)	$6e^{-25}$	61
177	34	1953	Ribossomal protein L16 [*Ximenia americana*] (ADD30329.1)	$9e^{-38}$	67
144	34	1904	NADH dehydrogenase subunit I [*Manihot esculenta*] (YP_001718490.1);	$3e^{-111}$	99
			NAD(P)H-quinone oxidoreductase subunit [*Medicago truncatula*] (XP_003621687.1)	$4e^{-57}$	62
185	33	2126	Ribossomal protein L33 (*H. brasiliensis*); Photosystem I subunit IX (*Populus trichocarpa*)	$2e^{-35}$; $1e^{-14}$	100; 100
			Ribosomal protein S18 [*Ricinus communis*] (AEJ82577.1)	$3e^{-39}$	97
288	31	435	metallothionein 3-like protein [*Hevea brasiliensis*] (ADR30789.1)	$1e^{-24}$	91
71	30	850	REF-like stress related protein 2 [*H. brasiliensis*]	$1e^{-61}$	99

			(gb|AAP46160.1)		
42	26	496	*H. brasiliensis* EST (JG294355.1)	0	99 (BlastN)
89	25	386	*Hevea brasiliensis* latex abundant protein 1 (LAP1) mRNA, complete cds (gb|AY221986.1)	0	100 (BlastN)
20	24	1211	NADH dehydrogenase subunit 4 (*H. brasiliensis*)	0	99
27	24	690	Hypothetical protein (*Medicago truncatula*)	$1e^{-16}$	91
493	23	476	Conserved hypothetical protein [*Ricinus communis*] (ref|XP_002513861.1)	$8e^{-09}$	59
142	20	598	*H. brasiliensis* EST (gb|JG005167.1)	0	99 (BlastN)




B
CF0
6
5
4
3
2
1
0
0h 6h 10h 24h 0h 6h 10h 24h 0h 6h 10h 24h 0h 6h 10h 24h
PB 217
PR 255
GT 1
IAN 873


C
NADH
4
3
2
1
0
0h 6h 10h 24h 0h 6h 10h 24h 0h 6h 10h 24h 0h 6h 10h 24h
PB 217
PR 255
GT 1
IAN 873


D
PsbR
4
3
2
1
0
0h 6h 10h 24h 0h 6h 10h 24h 0h 6h 10h 24h 0h 6h 10h 24h
PB 217
PR 255
GT 1
IAN 873

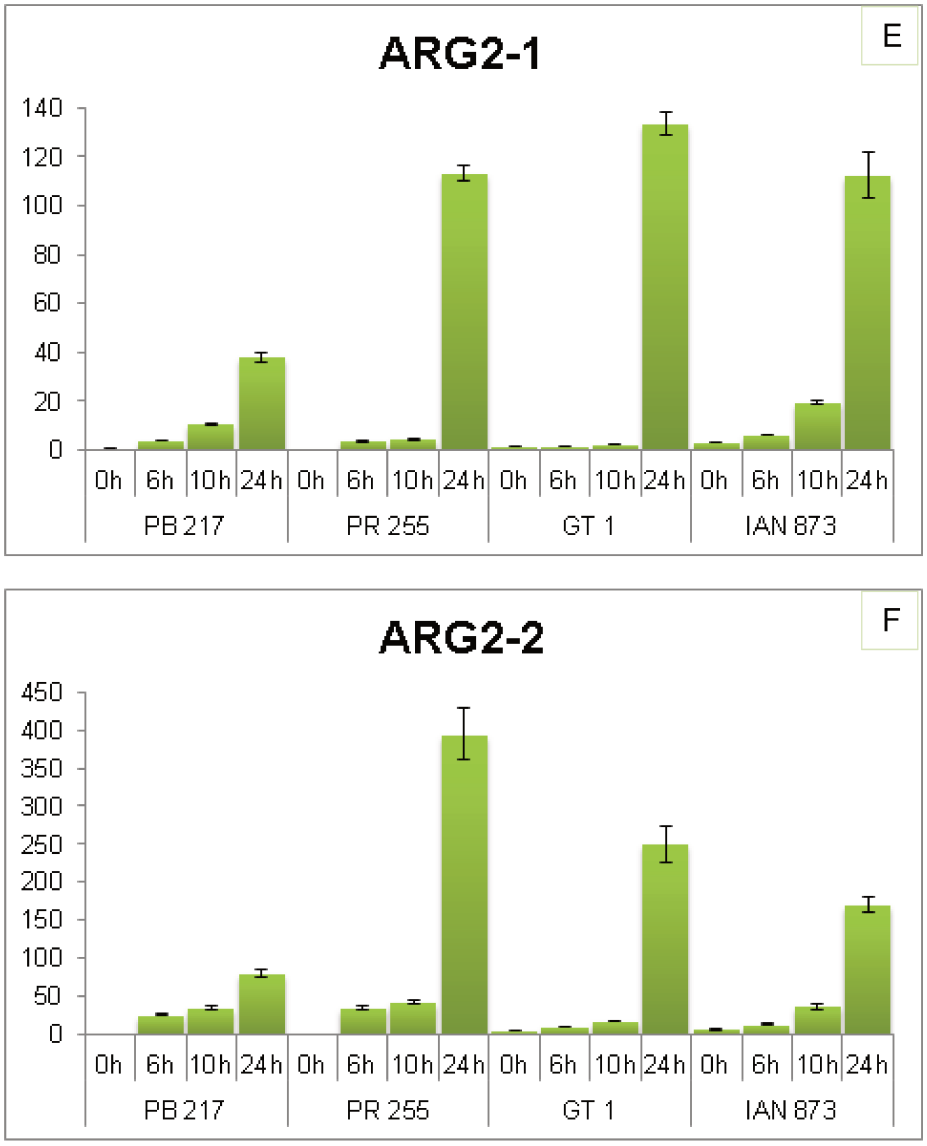


Figure S1 – Expression analysis of the most highly represented sequences in the cold-stressed leaves libraries in the clones PB 217, PR 255, GT 1 and IAN 873. A – hypothetical protein (HYPOT), B - ATP synthase CF0 C subunit (CF0), C - NAD(P)H-quinone oxidoreductase subunit H (NADH), D - chloroplast photosystem II 10 kDa polypeptide (PsbR), E and F - indole-3-acetic acid-induced proteins (ARG2-1 and ARG2-2).

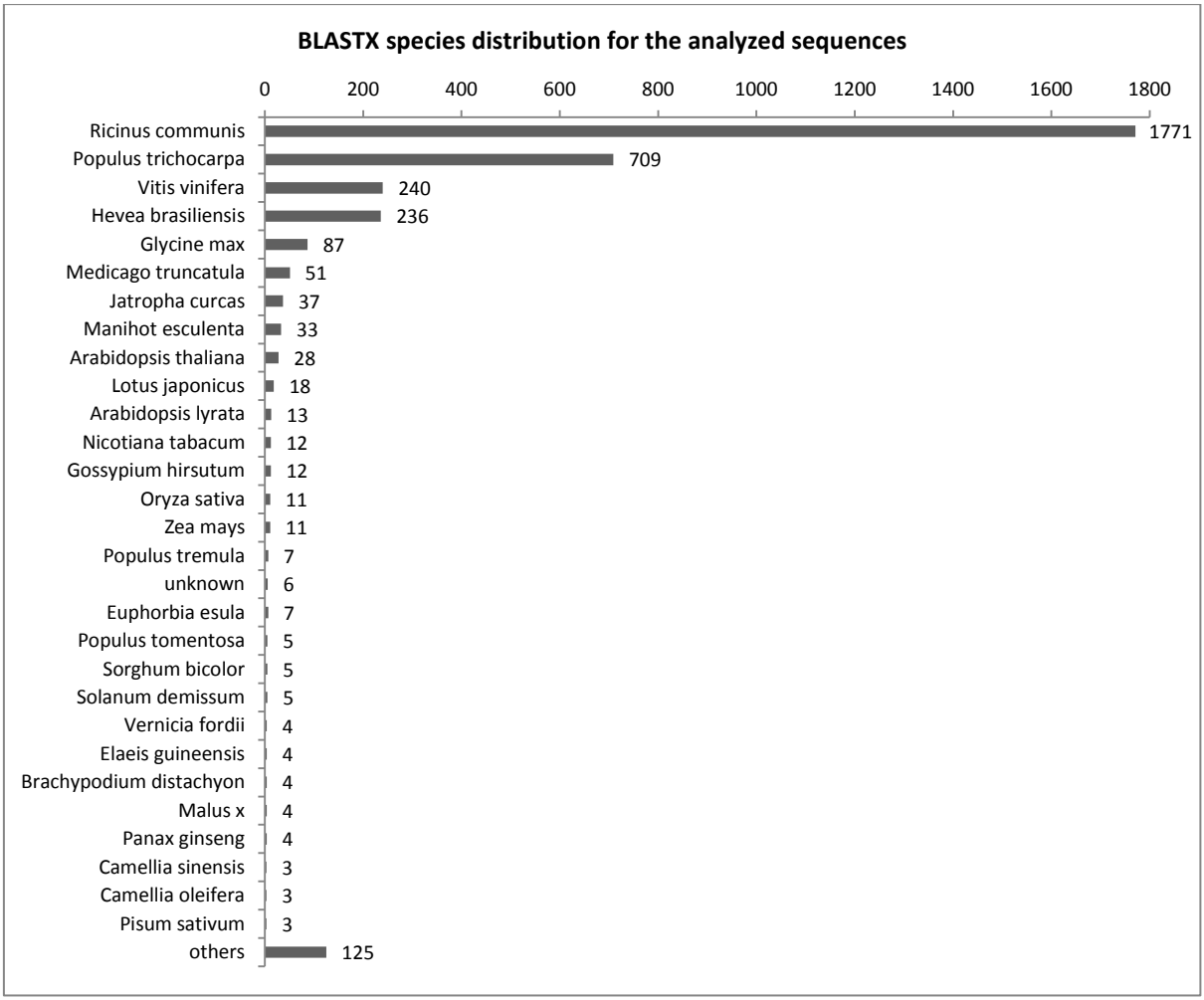


Figure S2 – BLASTX species distribution for the analyzed sequences

Table S5 – The most abundant KEGG pathways represented by the annotated unigenes

	Pathways	N° of sequences	N° of enzymes
1	Carbohydrate metabolism	158	81
2	Purine metabolism	51	18
3	Carbon fixation in photosynthetic organisms	45	18
4	Nitrogen metabolism	40	12
5	Oxidative phosphorylation	34	7
6	Phenylpropanoid biosynthesis	29	4
7	Pyrimidine metabolism	27	11
8	Phenylalanine metabolism	27	10
9	Flavonoid biosynthesis	26	13

10	Cysteine and methionine metabolism	22	10
11	Glycerolipid metabolism	20	12
12	Glycrophospholipid metabolism	19	12
13	Arginine and proline metabolism	18	10
14	Glutathione metabolism	15	7
15	Glycine, serine and threonine metabolism	15	8
16	Thiamine metabolism	15	5

Table S6 – The frequency of the identified SSR motifs.

Motif	N° of SSRs	Total (%)
AG/CT	167	28.4
AT/TA	107	18.2
AAG/CTT	82	13.9
AAT/ATT	50	8.5
AC/TG	27	4.6
ATC/GAT	24	4.1
AAAT/ATTT	23	3.9
AGC/GCT	18	3.1
AGG/CCT	11	1.9
AAAAT/ATTTT	11	1.9
ACC/GGT	8	1.4
AAAAG/CTTTT	7	1.2
AAAG/CTTT	5	0.9
AATT/TTAA	5	0.9
AAC/TTG	5	0.9
AAAC/GTTT	4	0.7

AAAAAT/ATTTTT	4	0.7
CCG/CGG	3	0.5
AACTC/GAGTT	3	0.5
AAATT/AATTT	3	0.5
ACG/CGT	2	0.3
ACT/AGT	2	0.3
AATCGG/CCGATT	2	0.3
CG/GC	1	0.2
AACC/GGTT	1	0.2
ATAG/CTAT	1	0.2
ATCTC/GAGAT	1	0.2
AAAAC/GTTTT	1	0.2
AAAGG/CCTTT	1	0.2
AATCAC/GTGATT	1	0.2
AGAGGC/GCCTCT	1	0.2
AAAAAG/CTTTTT	1	0.2
AAAAGG/CCTTTT	1	0.2
AATGGC/GCCATT	1	0.2
ACCTCC/GGAGGT	1	0.2
AATTCC/GGAATT	1	0.2
AGCTCC/GGAGCT	1	0.2
AATTAT/ATAATT	1	0.2

Online Resource 2

Primer	Motif	Primer sequences (5' - 3')	Ta (°C)	Expected length (bp)	Observed length (range, bp)	N° of alleles	He	Ho	PIC	Cross-species amplification	Analysis method	BLASTX hit
EHBp-1	$(CA)_7$	F - GCATCCTCAATCTCTCAAAG R - AGTGTAGCAGCTGTAGCAAGT	60	210	208-215	4	0.6578	0.4	0.607	all	Acrilamide	no hit
EHBp-2	$(TAAT)_4$	F - CACGACGTTGTAAAACGACCATAAAAATGAATCACACTCACG R - ACCACCATCCTCCACAACTC	TD1	209	209-220	3	0.3711	0.4667	0.3227	all	Fluorescence	Cucumber peeling cupredoxin
EHBp-3	$(CT)_6$	F - CACGACGTTGTAAAACGACCTTGTTTATTTGCCGCTGCT R - TGTTCTGCCAACCCACTGTA	TD1	242	250-255	2	0.4911	0.3333	0.3705	H. gui	Fluorescence	S-adenosylmethionine synthase 2
EHBp-4	$(GAT)_7$	F - GGTTATGGGTGTTTTGCAAGT R - CCAACTACACCCACCAACAA	63	236	236-239	2	0.51	0.5	0.3737	all	Acrilamide	no hit
EHBp-5	$(CT)_8$	F - GGGTGTAGATGAGGACAGCAA R - GCAACCGTGCAAGGATTTAT	63	225	219-231	3	0.63	0.67	0.541	H. gui, H. nit, H. rig, H. pau, H. bent	Acrilamide	no hit
EHBp-6	$(AG)_{14}$	F - GATACCCACCCCAAGAACAA R - TGCCAAAGCTGATGGTTTTA	63	225	209-229	7	0.83	0.76	0.7852	all	Acrilamide	hypothetical protein
EHBp-7	$(TTC)_5$	F - CAGGTCCAGATCTGCTGACA R - CAAAAAGGATCGGAATTGGA	63	228	230-233	2	0.06	0.06	0.0526	all	Acrilamide	no hit
EHBp-8	$(GAA)_7$	F - CCACCAGGTGCAACTTGATA R - CCATTGCTTCAGACATAGCA	63	248	245-257	4	0.55	0.67	0.4438	all	Acrilamide	predicted protein (LIGHT SENSITIVE HYPOCOTYLS 3)

EHBp-9	$(AAAG)_6$	F - GTCTCAAGCTCAATAGGAGAAT R - GCGCAAATCATGAAAGAAG	63	170	170-176	3	0.38	0.22	0.3402	all	Acrilamide	rapid alkalinization factor 1
EHBp-10	$(AATT)_5$	F - GCATCAAGTGCCACGAGTT R - CAAAAGGCCAAGGCACTAAT	63	231	231-239	3	0.16	0.06	0.1494	H. gui, H. rig, H. nit, H. bent, H. cam	Acrilamide	no hit
EHBp-11	$(TAAAAA)_4$	F - TGATGCAACGTTAACACACG R - TGGAAGATAAAGGGCAAATCA	63	238	238	1	0	0	0	H. gui, H. rig, H. nit, H. bent	Acrilamide	conserved hypothetical protein
EHBp-12	$(CAC)_6$	F - TCACTGCCTTCCAATTCTCC R - CATCCTTGCCCTTTTTCAAG	62	206	200-209	2	0.36	0.22	0.2859	all	Acrilamide	probable plastid-lipid-associated protein 4, chloroplastic
EHBp-13	$(TTTA)_4$	F - CGGAACTACTTCCAAAACTTGA R - GAATTCCCAATAGCGTTTCG	63	240	240-244	2	0.29	0.22	0.2392	all	Acrilamide	Lipoxygenase 1
EHBp-14	$(ACT)_5$	F - CCTTCCATCTCCAGCATAACA R - CTTCGATTCCCCGAGATACA	63	247	247-253	3	0.64	0	0.5443	all	Acrilamide	no hit
EHBp-15	$(AG)_{13}$	F - TGCAAGTGAGAAAGCAAAGAA R - GAAACGGCGAAATATTGCAT	57	201	199-213	5	0.7	0.89	0.6352	all	Acrilamide	acyl-carrier-protein
EHBp-16a	$(TC)_{12}$	F - GCGAAAATTGGTTTTTGGTT R - TCAAGCTACACACAGAGAGAAA	55	204	206-228	4	0.41	0.33	0.3538	H. gui, H. nit, H. pau, H. bent, H. cam	Acrilamide	no hit
EHBp-16b	$(TAT)_5$	F - AAAAAGGAAGCTTTTGCTTAGT R - CAAGGAAACAGTTCATGCAAA	63	215	215	1	0	0	0	all	Acrilamide	
EHBp-17	$(TCT)_5$	F - CCTCCTCAAAACCCCTCAA R - ATGGAAAGGGAGCGAAAGAT	63	202	204-216	4	0.68	0.76	0.5972	all	Acrilamide	60S ribosomal protein L34

EHBp-18	$(TC)_{13}$	F - GGACCAAACGCTCACTGTCT R - GGGAGAAAATGCAGATCCAA	63	212	204-224	5	0.73	1	0.66	all	Acrilamide	vesicle-associated membrane protein
EHBp-19	$(ATA)_6$	F - TGTGCAACACATACAAGCTGA R - TTGGACATTTGGCGTTGTTA	63	179	179	1	0	0	0	all	Acrilamide	no hit
EHBp-20	$(GCT)_5$	F - CACGACGTTGTAAAACGACGGACAGAGCAAAGTGGGATG R - TTTTGAACACAAGGAACCTACAG	TD3	169	300-303	2	0.0556	0.0556	0.0526	all	Fluorescence	acyl-CoA-binding protein
EHBp-21	$(TC)_7$	F - CACGACGTTGTAAAACGACTTCATCTTCTCAGGCCAAAA R - ACTGAAACCCACCTCACCAA	TD2	207	207-223	6	0.7077	0.2353	0.6419	all	Fluorescence	Jasmonate-zim-domain protein 3, putative isoform 1
EHBp-22	$(TAT)_6$	F - CACGACGTTGTAAAACGACGCAATTAAGGATGGCAAGGA R - CAATTGAAATCAAATGAACCA	TD2	171	171-174	2	0.4365	0.3889	0.3343	H. gui, H. nit, H. pau, H. bent, H. cam	Fluorescence	pathogenesis-related protein STH-2
EHBp-23	$(TTTG)_4$	F - CACGACGTTGTAAAACGACTTTCATTGGAGCACTTGCAG R - AAAGAGTCATCATCATAATCCTCTCA	TD2	267	263-267	3	0.5952	0.4444	0.5056	all	Fluorescence	aquaporin
					229-239	4	0.6175	0.6111	0.5576			
EHBp-24	$(AAG)_5$	F - CACGACGTTGTAAAACGACATTCCTGCAAGAGGCTGAGA R- TCCAGGTTTTTCTTCGTGG	TD2	210	207-210	2	0.0556	0.0556	0.0526	all	Fluorescence	AUX/IAA transcriptional regulator family protein isoform 1
EHBp-25	$(CAG)_5$	F - CACGACGTTGTAAAACGACTCCAAGTTCCATTCCCAAAC R - TAACGGTGTCAGCCAATGAA	TD2	253	248-263	3	0.5952	0.0556	0.496	all	Fluorescence	WRKY transcription factor
EHBp-26	$(ATT)_5$	F - CACGACGTTGTAAAACGACCCCTCCTATGTTTGCAGAAG R - TCGGAACAGATATATTCTCACAATTC	TD2	257	257	1	0	0	0	all	Fluorescence	eukaryotic translation initiation factor 2
EHBp-27	$(AAG)_7$	F - TCTTCAAAGCTTCAACAGCAAC R - CTTCGCTCCAAAAACTCCAT	63	249	240 - 286	10	n/a	n/a	n/a	all	Acrilamide	probable beta-D-xylosidase 5-like

EHBla-1	$(ATT)_5$	F - CTTTCAGCCAAGTTGCATCA R - ATCTCGTCGGCATCAACTTC	63	172	172-175	2	0.11	0	0.0994	all	Acrilamide	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 1
EHBla-2	$(AAT)_{17}$	F - CGACTGCTGCTGTGACAAAT R - CAGGAACACAACGTCTTCTTTT	63	212	182-218	8	0.73	0.88	0.6734	H. gui, H. nit, H. pau, H. bent	Acrilamide	conserved hypothetical protein
EHBla-3	$(GAT)_5$	F - TGGTTAATCCCATGATGCAA R - CCTGAGCCATATGGATGGAC	60	220	220	1	0	0	0	all	Acrilamide	putative clathrin assembly protein At5g35200 isoform 1
EHBla-4	$(GA)_7$	F - CGGCAAATCATTTTGGTCTT R - AACTGGTGGAGGATTTGCAC	63	245	386-392	2	0.51	0	0.3719	all	Capillary Electrophoresis	ENTH/VHS/GAT family protein isoform 1
EHBla-5a	$(CT)_8$	F - CACGACGTTGTAAAACGACGCAAAGCCTTAAACGCAAAG R - CCTTCTCTCTTCTTCTGGGCTA	TD2	123	123-135	3	0.1603	0.0556	0.1494	all	Fluorescence	NC domain-containing protein-related
EHBla-5b	$(GAA)_{12}$	F - CACGACGTTGTAAAACGACGCCCAGAAGAAGAGAGAAGGA R - CCCTGGCTTCAACTCATCTC	TD2	127	115-136	6	0.7094	0.8235	0.636	all	Fluorescence	
EHBla-6	$(AAG)_{10}$	F - CACGACGTTGTAAAACGACTTGATGGCCAATTATCAGCA R - CTTTCTTCGTCGTTGCCTTC	TD2	223	214-238	6	0.7016	0.7778	0.6283	H. gui, H. rig, H. nit, H. cam	Fluorescence	conserved hypothetical protein
EHBla-7	$(TC)_{12}$	F - CACGACGTTGTAAAACGACGCAAAGACAAATTAACaATCAACG R - AAAGGAAAATCGCTGGATCA	TD2	166	154-166	4	0.7206	0.8889	0.6457	all	Fluorescence	membrane-anchored ubiquitin-fold protein 3
EHBla-8	$(AT)_8$	F - CACGACGTTGTAAAACGACGGATTGGGGGAGAGAAAACT R - CAAAGCCTTCATCTTCCGACT	TD2	177	174-180	2	0.4365	0.1667	0.3343	H. gui, H. rig, H. nit, H. pau, H. bent	Fluorescence	conserved hypothetical protein
EHBla-9	$(AG)_{10}C(GA)_{10}$	F - CACGACGTTGTAAAACGACTGGGGCATGAAAAAGCTATC R - AAGGTATTGCGTGCCATTTC	TD2	151	131-157	5	0.6349	0.6667	0.5572	all	Fluorescence	conserved hypothetical protein
EHBla-10	$(CTT)_6$	F - CACGACGTTGTAAAACGACGCCATGGCAGCAACATACTA	TD2	196	190-217	4	0.554	0.2778	0.4438	all	Fluorescence	2-methyl-6-geranylgeranylbenzoquin

		R - TGACCACACAGGTGGAAACT										one methyltranferase
EHBla-11	$(AT)_8$	F - CACGACGTTGTAAAACGACGGACAGATCTGGAGGATTCG R - AACCAAGTTTTAAAGCAATTG	TD2	213	217-229	5	0.4952	0.3333	0.446	H. gui, H. rig, H. nit, H. pau	Fluorescence	CSL zinc finger domain-containing protein
EHBla-12	$(AATT)_4$	F - CACGACGTTGTAAAACGACTGTGACAAATCTTTTCAAATTTCACT R - TCTTGTCGGAAACTGCCTTT	TD2	245	325	1	0	0	0	H. gui, H. rig, H. nit, H. pau, H. bent	Fluorescence	EG45-like domain containing protein-like
EHBla-13	$(AG)_7$	F - CACGACGTTGTAAAACGACAGCAAAAAGAATTTTCAGTCTCT R - TTCTTTGACGCTGGACCTCT	TD2	204	202-204	2	0.5127	0.3889	0.3742	all	Fluorescence	VAMP-like protein YKT61-like
EHBla-14	$(AG)_{12}$	F - CACGACGTTGTAAAACGACCAACTCACCAGTTGCTGTCG R - TGAATCAAAACCAGGAGGTG	TD2	167	157-167	3	0.2079	0.2222	0.19	H. gui, H. nit, H. pau, H. bent, H. cam	Fluorescence	balbiani ring 1-related family protein
EHBla-15	$(GA)_{13}$	F - CACGACGTTGTAAAACGACGCTGCTTGAACACGAAACTG R - ACAGCGGATTCTCTTGGAAA	TD3	227	227-239	4	0.5444	0.6667	0.456	all	Fluorescence	conserved hypothetical protein
EHBla-16	$(AGC)_6$	F - CACGACGTTGTAAAACGACCAGGGGCAGGAAATCATCTA R - CTGAAATGCGCCTTCTCTTC	TD2	282	273-282	3	0.6365	0.3889	0.5462	all	Fluorescence	ATP binding protein, putative
EHBla-17	$(TC)_8$	F - CACGACGTTGTAAAACGACGCCGCTACAAGCAATAGAGC R - CATGATTGAGCGTTGGATTG	TD2	157	157-163	3	0.5095	0.3333	0.3972	all	Fluorescence	Signal recognition particle 14 kDa protein
EHBla-18	$(GT)_7$	F - CACGACGTTGTAAAACGACCCGTCAACAGCTTCTTCACA R - AACCCATTTCCACCTTTATTTT	TD3	223	223-227	3	0.2656	0.1765	0.2364	all	Fluorescence	B12D protein
EHBla-19	$(TA)_6$	F - CACGACGTTGTAAAACGACAATGAAGCACGGGGAAGTTA R - ATGTTGCCAGGGACAATACA	TD2	179	179-181	2	0.0556	0.0556	0.0526	all	Fluorescence	mannan endo-1,4-beta-mannosidase 7
EHBla-20	$(GCTGGA)_4$	F - CACGACGTTGTAAAACGACTGCTTGTCTTGATGCCTTTG R - GGAAGAGCCACCAAAACTC	TD2	151	139-157	4	0.6254	0.5	0.5411	all	Fluorescence	Werner-syndrome like protein

EHBc-1a	$(CT)_6$	F - CACGACGTTGTAAAACGACGAAGGGACCTGTGGGAAAAC R - GCAAGAGCAAGGATCGCTAA	TD1	245	239-245	2	0.5	1	0.375	all	Fluorescence	chlorophyll a-b binding protein CP26, chloroplastic-like
EHBc-1b	$(AATT)_4$	F - CACGACGTTGTAAAACGACTCACCTATCCATCTCCTTTCTCA R - CCACTAAACAAAATCCCCTAACC	TD1	206	206	1	0	0	0	all	Fluorescence	
EHBc-2	$(GAAA)_4$	F - CACGACGTTGTAAAACGACCACAAACAAAAGAATAACCAATCTG R - TTATTGGCTCGAGGCTCTTG	TD1	264	264	1	0	0	0	all	Fluorescence	bifunctional nuclease in basal defense response 1
EHBc-3	$(AT)_6$	F - CACGACGTTGTAAAACGACTTGCAAGCAGAATGATGGAG R - TGCAATATAAGCCATACAACATCC	TD1	174	174	1	0	0	0	H. bent	Fluorescence	NDH dependent flow 6 isoform 1
EHBc-4	$(GAT)_6$	F - GCAACTCAAAGAAGGGTCCA R - TGAGGAGCCTTCTGCCTTTA	63	269	269-275	3	0.45	0	0.3957	all	Acrilamide	predicted protein
EHBc-5	$(TCA)_8$	F - CCGGAGACTTCTTTGCTGAC R - CGAGGAAGAAAGGCAATGTC	63	246	146-246	6	0.71	0.72	0.6466	all	Acrilamide	eukaryotic translation initiation factor
EHBc-6	$(TTC)_{10}$	F - GCTGGGTCGATGAGATCTGT R - CACCAAATAATCACCGTCCA	62	169	169-193	5	0.57	0.33	0.5015	all	Acrilamide	no hit
EHBc-7	$(TTC)_6$	F - GAGCGTGATGAGATGGATCA R - GGGAGGGGATGCATAGAT	60	168	168	1	0	0	0	all	Acrilamide	ubiquitin-like modifier-activating enzyme 5-like
EHBc-8	$(TCT)_5$	F - GAGAGGGTGCTTCTTTGTGC R - CCAAACCCAAAAACCAGAAA	60	187	187-208	5	0.69	0.61	0.6099	all	Acrilamide	no hit
EHBc-9	$(GA)_{10}$	F - TGCCGGACACACTTGTAGTT R - GGGAATGACGTGGACTTGTT	63	221	227-233	3	0.16	0.17	0.1495	all	Acrilamide	conserved hypothetical protein

EHBc-10	$(CTT)_5$	F - AGGGGGACACAAAAACCAG R - TGGACCGCTATGAGAAACAA	63	220	205-226	3	0.51	0.44	0.3972	all	Acrilamide	no hit
EHBc-11	$(TTC)_6$	F - GGAGGTCGTCCATAAGGTAGC R - CCTTCCTCTTCTGCCTTGAA	63	201	201-228	5	0.6	0.5	0.5296	all	Acrilamide	CTP synthase 1a
EHBc-12	$(TTG)_6$	F - TGGATGAGTGCCTTTACTCG R - GACGTACCAACGTGGGATAAA	63	158	161-170	4	0.54	0.56	0.4561	all	Acrilamide	no hit
EHBc-13	$(AT)_8$	F - TGCATGAGTAGGGAATGGTG R - ATTCCGTCAACCAATGATCC	57	240	238-242	3	0.41	0.06	0.3633	all	Acrilamide	no hit
EHBc-14	$(AT)_{12}$	F - AGAGAGGTGGTGAAGCTGGA R - AAACAAGTGCCGGAATTCAT	63	182	174-184	4	0.36	0.24	0.3165	all	Acrilamide	CBL-interacting protein kinase 16
EHBc-15	$(CCTCTG)_4$	F - CACGACGTTGTAAAACGACCTGCTCACACAAGTGGCA R - TGGGGCTTGCGTACTTTTAC	TD2	270	264-276	3	0.4902	0.1765	0.3872	all	Fluorescence	negative cofactor 2 transcriptional co-repressor, putative
EHBc-16	$(AAGAA)_5$	F - CACGACGTTGTAAAACGACTGGACTGGGCCTTAATGTTT R - GGTCCAATGACCCAGAGAAA	TD2	218	218-228	2	0.1079	0.1111	0.0994	all	Fluorescence	F-box protein PP2-A12
EHBc-17	$(TCC)_7$	F - CACGACGTTGTAAAACGACCCCGGCGAATCTATGTCT R - ATAGCATGGCCACAAGGAAG	TD3	271	271	1	0	0	0	all	Fluorescence	WD repeat domain phosphoinositide-interacting protein
EHBc-18	$(AGA)_5$	F - CACGACGTTGTAAAACGACTAAGCTCCGCTTGCAGTACA R - AAGAAGGTGGGTTTTCCTCTG	TD2	162	162	1	0	0	0	all	Fluorescence	isomerase peptidyl-prolyl cis-trans isomerase
EHBc-19	$(GAA)_7$	F - CACGACGTTGTAAAACGACCAAAAATGGTGGAAGGGAGA R - CAAGTTTGCCATTAGGTATTAAA	TD2	140	137-140	2	0.1079	0	0.0994	all	Fluorescence	photosystem II stability/assembly factor HCF136, chloroplastic-like
EHBc-20	$(CT)_{17}$	F - CACGACGTTGTAAAACGACTCCCTTTCTCAAGGTGCTTC	TD2	230	220-238	5	0.7937	0.6111	0.7327	all	Fluorescence	histidine-containing phosphotransfer protein 1

		R - CCCATGTTCTCTGCATCTGA										isoform 1
EHBc-21	$(CT)_9$	F - CACGACGTTGTAAAACGACTTCATCTTCCTCCTCTCTCCTG R - GCTGAGAAGCTCTCCATCGT	TD2	223	219-227	5	0.4286	0.3333	0.393	all	Fluorescence	xyloglucan endotransglucosylase/hydrolase protein 9
EHBc-22	$(TAT)_5$	F - CACGACGTTGTAAAACGACAAACCACATTCCGACTCCTG R - GGATGCTTACTGACGCACAA	TD2	268	244-268	3	0.2656	0.2941	0.2364	H. gui, H. rig, H. nit, H. pau	Fluorescence	RING-H2 finger protein ATL48-like
EHBc-23	$(CCAATT)_5$	F - CACGACGTTGTAAAACGACCCGTTTCTCTCCGTCTTTGT R - TTGGGTAATAGTCCGCTTGG	TD2	250	254-266	3	0.4381	0.4444	0.3709	all	Fluorescence	conserved hypothetical protein
EHBc-24	$(GAG)_7$	F - CACGACGTTGTAAAACGACGAAATCGCCGGTAAGTTTGA R - ATCCATTGGGATTCAAGCAA	TD2	200	194-200	3	0.1603	0.1667	0.1494	all	Fluorescence	conserved hypothetical protein
EHBc-25	$(GAT)_5$	F - CACGACGTTGTAAAACGACACTCCTCCAGCTGCTGATTC R - CGTTCTTCTGCAGCCTTCTT	TD2	125	380	1	0	0	0	all	Fluorescence	Elongation factor 1-delta 2
EHBc-26	$(TTTTAT)_4$	F - CACGACGTTGTAAAACGACTCCCCAGGCTCAAATGATTA R - AACTCTCGTTTTCGCACTCAA	TD2	236	236	1	0	0	0	all	Fluorescence	ABC transporter C family member 2 isoform 2
EHBc-27	$(AG)_{19}$	F - CACGACGTTGTAAAACGACTGCCACTTCCAAGAGCTACC R - ACACCCATTTTCCCTTTCCT	TD2	180	158-192	7	0.7698	0.7222	0.7173	all	Fluorescence	senescence-associated family protein
EHBc-28	$(TC)_6$	F - CACGACGTTGTAAAACGACGAGTCTTTTTGATTGTTTGTCTGC R - GGCAATTGCATTGATTCTGA	TD1	212	330	1	0	0	0	all	Fluorescence	GroES-like family protein
EHBc-29	$(AAAT)_4$	F - CACGACGTTGTAAAACGACGATCAAAGGAAGGTGGCAAT R - AAGCCCAAGTCCCAAACATA	TD1	196	160; 196	2	0.42	0.3333	0.3318	H. gui, H. rig, H. nit, H. pau, H. bent	Fluorescence	no hit
EHBc-30	$(GGGT)_4$	F - CACGACGTTGTAAAACGACAAAATCTCCTGCGATGCAAT R - TGCTCTGAACCCACCTTACC	TD1	250	250	1	0	0	0	all	Fluorescence	conserved hypothetical protein

EHBc-31	$(TTC)_5$	F - CACGACGTTGTAAAACGACCCTCAGAGGAGGCAATTCAC R - CCCATCGATTTAGCCTCTCA	TD1	246	246-260	3	0.1267	0.1333	0.1228	all	Fluorescence	patellin-6
EHBc-32	$(TAT)_6$	F - TTGGCTACCTACCCAGATGC R - ATGTTCCTTGTGCTCCCAAC	63	225	225-258	7	0.76	0.61	0.6986	all	Acrilamide	conserved hypothetical protein
EHBc-33	$(TTA)_9$	F - TCCTGCAACAGAACAACACC R - ATGAAAACAATGCACCCACT	63	248	236-248	3	0.56	0.5	0.4813	all	Acrilamide	no hit
EHBc-34	$(CT)_{10}$	F- ATTCTGGTGGAAATCGAACG R - AAGGCGAGCAAGAAAACTGT	63	236	234-270	6	0.66	0.61	0.5758	all	Acrilamide	metallothionein
EHBc-35	$(CT)_6$ - $(AG)_{12}$	F- GTGAGAAGTGGGATAACG R - TCGTACAAAATCTTCTCGTG	57	190	190-196	2	0.16	0.17	0.1411	all	Acrilamide	R3H domain-containing protein
EHBc-36	$(GAA)_6$	F - GCCCAAAGAGGAAAATGAGA R - TATGCAACCAATGGGCTTTT	63	198	198-207	2	0.11	0.11	0.0995	all	Acrilamide	no hit
EHBc-37	$(AGA)_9$	F - CAGAAGGGGATTTTGATTGG R - CCGTGGAAGAAAAGAACGAG	63	250	250-259	3	0.43	0.33	0.3491	all	Acrilamide	conserved hypothetical protein
EHBc-38a	$(ATAA)_4$-$(CT)_6$	F - GCGAAAATTGGTTTTTGGTT R - TCAAGCTACACACAGAGAGAAA	58	191	191-199	5	0.76	0.11	0.69	all	Capillary Electrophoresis	no hit
EHBc-38b	$(GTTTT)_4$	F - AAAAAGGAAGCTTTTGCTTAGT R - CAAGGAAACAGTTCATGCAAA	58	169	169-174	2	0.06	0.06	0.0526	all	Capillary Electrophoresis	
EHBc-39	$(AT)_7$	F - GCATTAAGATAGGTGCGATTAC R - AAGTGAAATTGCCGTGATTG	60	214	212-235	10	0.89	0.94	0.8512	all	Capillary Electrophoresis	photosystem II core complex proteins psbY, chloroplast precursor
EHBc-40	$(GGC)_5$	F - CACCTTGCGGAATTTAGGAT	62	195	186-192	3	0.64	0.06	0.5455	H. gui, H. rig, H. pau	Capillary Electrophoresi	serine/threonine-protein kinase HT1-like

		R - CTGCTGCTGCTCATTTTCTG									s	
EHBc-41	$(AT)_7$	F - GCCAAAGCCAAATATTAACCA R - TGATCCGGATGGTAATGGAT	60	274	274-286	8	0.82	0.67	0.7631	all	Capillary Electrophoresis	Lactoylglutathione lyase
EHBc-42	$(ACAA)_4$	F - ATTTGGGCAAAGTGATGAAA R - TGATCCAACTGGAGCAACTATG	63	193	189-201	4	0.62	0	0.5393	all	Capillary Electrophoresis	lycopene epsilon cyclase, chloroplastic
EHBc-43	$(GAA)_5$	F - ACCTTCAAGCGTCGAACAAT R - AAGGATTTCCATCGGACAAA	63	222	367-394	7	0.51	0.44	0.4797	H. gui, H. rig, H. pau, H. bent, H. cam	Capillary Electrophoresis	conserved hypothetical protein
EHBc-44	$(AATA)_4$	F - CACGACGTTGTAAAACGACCTGCATTTCCTTGGGAGAA R - GAGCACGACTAATAACAAATCGAA	TD2	193	190-193	2	0.0556	0.0556	0.0526	all	Fluorescence	60S ribosomal protein L34, putative
EHBc-45	$(ATT)_{11}$	F - CACGACGTTGTAAAACGACATGCCACCACCCTTCAAGTA R - TGAGGGGACAAAAATTCCAA	TD3	289	280-310	8	0.7841	0.7778	0.7292	all	Fluorescence	predicted protein
EHBc-46	$(TG)_9$	F - CACGACGTTGTAAAACGACTGCTAATGATGCTGGTGGAC R - CATCATGCGTGCCATAAAAT	TD2	225	211-229	5	0.5921	0.5556	0.5031	H. gui, H. rig, H. nit, H. pau, H. bent	Fluorescence	guanosine-3',5'-bis(diphosphate) 3'-pyrophosphohydrolase, putative
EHBc-47	$(TGC)_5$	F - CACGACGTTGTAAAACGACATTTTGTGGGATGCTGTGC R - GCCTCAGGCCTAAAACTCTACA	TD2	190	190	1	0	0	0	all	Fluorescence	conserved hypothetical protein
EHBc-48	$(CT)_{14}$	F - CACGACGTTGTAAAACGACTCACCGAAGTTCATCGACAG R - TTCCATTGATTGAGCATCCA	TD3	272	266-280	5	0.7576	0.5294	0.6889	none	Fluorescence	Tetraspanin family protein
EHBc-49	$(CAC)_5$	F - CACGACGTTGTAAAACGACTGGATTGATGCTAGCTGTGG R - GGCAACATATCCCCATCATT	TD3	196	187-196	2	0.5	0.5	0.368	all	Fluorescence	transcription activator GLK1-like
EHBc-50	$(TCT)_{11}$	F - CACGACGTTGTAAAACGACCAATAATGGATGCAATCAGC R - TTTGTTGCTGCACTCAGGAC	TD2	245	230-245	6	0.8079	0.8333	0.7507	all	Fluorescence	rhodanese-like domain-containing family protein

EHBc-51	$(TA)_8$	F - CACGACGTTGTAAAACGACTGCAGACCTCATTTCCTTTG R - ACAACAAATGCAAAACAAGC	TD2	215	210-245	5	0.369	0.2941	0.3439	all	Fluorescence	conserved hypothetical protein
EHBc-52	$(TTTC)_5$	F - TTTTGTTCTTTCCGTTTCTGG R - CAAGCGATGCAGGAGAAAA	56	151	155	1	0	0	0	all	Acrilamide	RNA-binding protein
EHBc-53	$(CTT)_6$	F - CTTTGTCGGATCCGCATTAC R - CAACGTCAAAATCACCATGC	62	184	177-180	2	0.4365	0.1667	0.3343	all	Acrilamide	LMBR1 domain-containing protein 2 homolog A-like
EHBc-54	$(TATT)_4$	F - CACGACGTTGTAAAACGACACCGTTGTGCTTGTGGATA R - TTGCTGCTGGATTTTCTTCA	TD2	210	210	1	0	0	0	all	Fluorescence	Rhomboid-related intramembrane serine protease family protein, putative isoform 1
EHBc-55	$(GAT)_6$	F - CACGACGTTGTAAAACGACGAAGAAACGGCCAAGAAGAA R - CATCTCTTGAAAATGCCGAAA	TD2	181	157-181	3	0.1603	0.166	0.1494	all	Fluorescence	conserved hypothetical protein
EHBc-56	$(GAA)_5$	F - CACGACGTTGTAAAACGACGGGTAGCCTTTGCCTAAACC R - GGTGGTGATCCCTCTTCTT	TD2	173	173	1	0	0	0	all	Fluorescence	conserved hypothetical protein
EHBc-57	$(AG)_{16}$	F - CACGACGTTGTAAAACGACACCTTCCCGGCTTTAATTG R - GAGAGGGAGATGATGGGTGA	TD2	284	276-294	6	0.6532	0.6875	0.5661	all	Fluorescence	conserved hypothetical protein
EHBc-58	$(TAT)_6$	F - CACGACGTTGTAAAACGACACAACCCTGTGGCTGTGACT R - CATGTGCCCTTCTTGTGAGA	TD2	186	186-189	2	0.0556	0.0556	0.0526	all	Fluorescence	homeobox-leucine zipper protein HAT14-like
EHBc-59	$(TC)_{12}$	F - CACGACGTTGTAAAACGACCTCGCGCTCTCTGAATCTTT R - GGCCTTATTGAAACCCAACA	TD2	216	216-226	4	0.2587	0.0556	0.2412	all	Fluorescence	nudix hydrolase 25-like
EHBc-60	$(AG)_{13}$	F - CACGACGTTGTAAAACGACGAGAGGTAGATAGGCGGGAAA R - GGGTGATCCCAGAAACAGAA	TD1	267	267-280	4	0.5933	0.4	0.5362	none	Fluorescence	conserved hypothetical protein

EHBc-61	$(TTC)_5$	F - CACGACGTTGTAAAACGACGGAGTTGAATCGGGTTTCTG R - GGAAGCATGGGATTGAAAAA	TD1	231	231	1	0	0	0	all	Fluorescence	Lorelei-like-gpi-anchored protein 1
EHBc-62	$(CT)_{10}$	F - CACGACGTTGTAAAACGACTTCTTTGGTGGAATAGTTGGTTT R - TGCAAGAGGGAGGAAGACTC	TD1	213	213-215	2	0.4978	0.5333	0.3739	all	Fluorescence	pectin acetylesterase
EHBc-63	$(GA)_6$	F - CACGACGTTGTAAAACGACCTCAGTGGAGAGCGAAAATTG R - GCAATATCACATCCGTCATCC	TD1	253	253	1	0	0	0	all	Fluorescence	PPPDE peptidase domain-containing protein
EHBc-64	$(GAA)_6$	F - TCTCCACTGCGGAGTCTTTT R - ACCCCTAGGACCTGGTCAGT	63	264	264-270	3	0.66	0.22	0.5675	H. gui, H. rig, H. pau, H. bent, H. cam	Acrilamide	GATA domain class transcription factor
EHBc-65	$(AC)_7(AT)_6$	F - CCCCCTTCAGCCTTCAATA R - TCCCCACTTCTCAAACAAGC	60	241	241-249	3	0.53	0.33	0.4264	H. gui, H. rig, H. nit	Acrilamide	cysteine-type peptidase, putative
EHBc-66	$(ACA)_5G(CAG)_5$	F - ACTTCTTGACCCACCCTCCT R - GGCATCCTTATCCTCCTTCC	63	250	250-259	3	0.63	0.89	0.5267	H. gui, H. nit, H. pau, H. bent, H. cam	Capillary Electrophoresis	TCP domain class transcription factor
EHBc-67	$(CGATTC)_7$	F - TCGGACGGATGAAAAAGTTC R - CAGGGTCATTTCCGTCATTT	63	241	217-241	4	0.61	0.67	0.5122	all	Acrilamide	BTB and TAZ domain protein 2 isoform 2
EHBc-68	$(ATT)_8$	F - GTTGCGTTTTCTCGCAATTT R - AGCCAAACAAGCACAGGAAT	55	234	234-237	2	0.49	0	0.3624	H. gui, H. bent, H. cam	Acrilamide	no hit
EHBc-69	$(AT)_7(AG)_{11}G(GA)_8$	F - AAAAAGGGAAAAATTTAGCAA R - TCGCTTCAGTTCATCTGGTG	55	177	177-201	6	0.76	0.5	0.7011	all	Acrilamide	Hydroxyproline-rich glycoprotein family protein
EHBc-70	$(CTT)_{10}$	F - GGGGACGAGAAGCAAACATA R - CACCCTTTTGAAAGCGAAAT	60	162	150-171	4	0.63	0.61	0.5517	all	Acrilamide	no hit
EHBc-71	$(ATG)_6$	F - CCACTGGCTCCAATCAAAAT	63	264	332-359	4	0.53	0	0.4846	all	Capillary Electrphosresi	exocyst complex component SEC3A

		R - CAGCCTTGCTGTCATTCTCA									s	isoform 1
EHBc-72	$(AACC)_4$	F - CACGACGTTGTAAAACGACTTGCTCTCCCTCTACACCTACA R - AAGCGTTTGTCGTCTCCAAG	TD2	266	266	1	0	0	0	all	Fluorescence	conserved hypothetical protein
EHBc-73	$(AG)_9$	F - CACGACGTTGTAAAACGACGCATTTCCCTCAGAAAAAGTAGA R - GCATGAGGACTAGGGCTGAG	TD2	228	224-238	3	0.5524	0.4444	0.4408	all	Fluorescence	CASP-like protein
EHBc-74a	$(TCT)_6$	F - CACGACGTTGTAAAACGACCACAGCTCTTGTATCATCACAGC R - CCTCTGGCCTCTTTCTTCCT	TD2	170	167-170	2	0.3222	0.2778	0.2642	all	Fluorescence	conserved hypothetical protein
EHBc-74b	$(AAG)_6$	F - CACGACGTTGTAAAACGACTCCTCTTCCCTCCTTTCCTC R - GGCGACAAGGAGAAACTCTG	TD2	246	243-247	4	0.4905	0.5	0.4367	all	Fluorescence	
EHBc-75	$(TCT)_7$	F - CACGACGTTGTAAAACGACGGGGTGGCAGTTAATTCTGA R - CCCACTTGAGACTCCACAAA	TD2	226	223-229	3	0.1603	0.1667	0.1494	all	Fluorescence	conserved hypothetical protein
EHBc-76	$(TTTG)_4$	F - CGATCCTTGTCGAAGCAAA R - ATCCATCGTCATCGTCATCA	62	208	208	1	0	0	0	all	Acrilamide	double-stranded RNA binding protein, putative
EHBc-77	$(TA)_{10}$	F - AAGCTTGCCGTTGAAAAGAA R - GATCATTACATTGCCCATCAA	62	256	246-252	3	0.2556	0.1667	0.2335	all	Acrilamide	ATP synthase subunit b', chloroplastic-like
EHBc-78	$(GAA)_7$	F - TCTTCACCTCGCCTTCTCTG R - CCATTAAACCTGCATCAGCA	62	227	221-233	3	0.1603	0.0556	0.1494	H. gui, H. rig, H. pau, H. bent	Acrilamide	chaperone protein dnaJ 20, chloroplastic
EHBc-79	$(GCT)_6$	F - CACGACGTTGTAAAACGACCAAGGGGCACAGATTCACTT R - TATCCGAGGGCTACAAGCAC	TD2	136	136-145	4	0.4905	0.2778	0.4067	all	Fluorescence	DNA double-strand break repair rad50 ATPase, putative
EHBc-80	$(ATTTT)_4$	F - CACGACGTTGTAAAACGACGCTCATCCCTGGAGACTCTG R - CCCCAAGAATTCAAAGTCCA	TD2	231	231	1	0	0	0	H. pau	Fluorescence	xyloglucan endotransglycosylase/hydrolase

EHBc-81	$(TGA)_5$	F - TGTTCTGATCGATTCCTCCA R - CACGCCCATTTCTGACTACA	62	178	178-181	2	0.1571	0.0556	0.1411	all	Acrilamide	r3h domain containing protein, putative
EHBc-82	$(ATG)_6$	F - CACGACGTTGTAAAACGACAGGGAGTGATGTTGCTCACC R - TGAAGGAATCATGGGGAAGT	TD2	142	142	1	0	0	0	all	Fluorescence	leucine-rich repeat containing protein, putative
EHBc-83a	$(CTT)_5$	F - CACGACGTTGTAAAACGACCCCAAGAAAACAATCCAACC R - GCTAAAGCTTGATGTGGAAGATG	TD2	184	184	1	0	0	0	all	Fluorescence	predicted protein
EHBc-83b	$(TTC)_5$	F - CACGACGTTGTAAAACGACACAAATTTGCACCCAGGTT R - GTGGAAAAAGGGCAAAACAA	TD2	232	232	1	0	0	0	all	Fluorescence	
EHBc-84a	$(TCT)_6$	F - CACGACGTTGTAAAACGACGAAGTCTTCAGCAACATCATCTTC R - CAAGCACCAAAACCACCTTT	TD2	217	217-223	2	0.0667	0.0667	0.0624	H. gui, H. rig, H. nit, H. pau	Fluorescence	conserved hypothetical protein
EHBc-84b	$(CTT)_5$-$(CAT)_7$	F - AAAGGTGGTTTTGGTGCTTG R - GTCCATTCGCATGTGATGTC	62	154	151-160	3	0.3794	0.4444	0.3368	all	Acrilamide	
EHBc-85	$(CT)_7$	F - CACGACGTTGTAAAACGACAGTATCTCCCTCGCTCTG R - GCGATGCCCTTCAACATTAT	TD2	157	400	1	0	0	0	H. gui, H. rig, H. nit, H. bent, H. cam	Fluorescence	Monothiol glutaredoxin-4
EHBc-86	$(TTA)_6$	F - CACGACGTTGTAAAACGACAAACAGCCTCTCTGCATGGT R - AATTGCATTTGCCGTAAACC	TD2	259	253-259	2	0.1079	0	0.0994	all	Fluorescence	ATP binding protein, putative
EHBc-87	$(AGC)_6$	F - CACGACGTTGTAAAACGACCCCTCTACATCTGGCTCTGC R - TGTTGGGAAGTGACATGGAA	TD2	281	281-335	2	0.0556	0.0556	0.0526	all	Fluorescence	protein TIME FOR COFFEE-like
EHBc-88	$(TTATAA)_4$	F - CACGACGTTGTAAAACGACTGGACCGCCTGTATAACTCC R - GACGCTAATTTACACCGACCA	TD2	273	267-282	3	0.1095	0.1111	0.1037	all	Fluorescence	zinc finger protein, putative
EHBc-89	$(TC)_9$	F CACGACGTTGTAAAACGACAGGCAACAAATCACCGAATC	TD2	228	228-246	5	0.4698	0.5	0.434	all	Fluorescence	probable ADP-ribosylation factor

		R - TATTGGCTCCAGCACCTCTT										GTPase-activating protein AGD8-like
EHBc-90	$(TTC)_7$	F - CACGACGTTGTAAAACGACCAAGCCGATCAGATCCAAAT R - TTCTGAAGCTCTTCGAGATCC	TD2	150	150-156	3	0.6578	0.7059	0.5629	all	Fluorescence	conserved hypothetical protein
EHBc-91	$(TTG)_5$	F - CACGACGTTGTAAAACGACGATGTGGTTTGCCTCCTTGT R - CACAAGTTTCGTCTGCCAGT	TD2	283	283	1	0	0	0	all	Fluorescence	ABC transporter I family member 17
EHBc-92	$(CT)_9$	F - CACGACGTTGTAAAACGACTCGTTCATCCACCTCATCAA R - GGAAGTCAGCTTACCGACCA	TD3	278	276-294	5	0.4889	0.5	0.434	all	Fluorescence	probable receptor-like protein kinase At5g15080-like
EHBc-93	$(GAAAAG)_4$	F - CACGACGTTGTAAAACGACTTTCCTTCCCATCTCTTTCG R - TTGATTTCCCAACAGCTGAA	TD2	192	186-192	2	0.5143	0.1111	0.375	all	Fluorescence	ras-related protein RABA6b
EHBc-94	$(ATG)_6$	F - CACGACGTTGTAAAACGACGGTGGCGGCTAATATGGTTA R - CCCAGTCCAAAGGAACATTG	TD2	262	256-262	2	0.2857	0.2222	0.2392	all	Fluorescence	adenylate translocator
EHBc-95	$(AT)_{11}$	F - CACGACGTTGTAAAACGACGGCATTTCCCTCTCCTAAG R - CGATACAAAGTCCCAATGGA	TD2	202	200-218	4	0.527	0.4444	0.473	H. gui, H. nit, H. pau	Fluorescence	sigma factor sigB regulation protein rsbQ isoform
EHBc-96	$(TC)_{14}$	F - CACGACGTTGTAAAACGACCAAAGCAAGAGCAAGCCAGT R - GGCCTCAAAATCTCAGTACCC	TD2	183	177-193	7	0.7905	0.8333	0.7364	all	Fluorescence	conserved hypothetical protein
EHBc-97	$(CT)_{17}$	F - CACGACGTTGTAAAACGACACACAAGTGCCATCCTCTCC R - CTTGGGAATCAAACGATGCT	TD2	188	171-195	7	0.7556	0.3333	0.6902	H. gui, H. rig, H. nit, H. pau, H. bent	Fluorescence	conserved hypothetical protein
EHBc-98	$(TC)_{19}(TA)_9$	F - CACGACGTTGTAAAACGACCATTATCAAATCGCCGATCA R - TCATTGCAAATCGAAAAGGA	TD3	278	274-298	5	0.7487	0.4706	0.6773	all	Fluorescence	BTB/POZ and TAZ domain-containing protein 1-like
EHBc-99	$(TTC)_5$	F - CACGACGTTGTAAAACGACGAATCTAATATCCTACATTCTCATTCC R - CGTTTATGTGCTTCCCTGGT	TD2	251	247-268	5	0.681	0.7222	0.6	all	Fluorescence	microtubule-associated protein RP/EB family member 1-like isoform 2

EHBc-100	$(ATT)_5$	F - CACGACGTTGTAAAACGACCGTTGGTGTCAATGTGAAGC R - TTCAATGAGTGATCTACATGCAA	TD2	274	274-277	2	0.4278	0	0.329	all	Fluorescence	conserved hypothetical protein
EHBc-101	$(TA)_9$	F - CACGACGTTGTAAAACGACGGCACTTTCACTGATGCAAA R - TTTATTCACCCTTCAGCAGAT	TD2	175	169-175	3	0.5187	0.1765	0.4372	all	Fluorescence	predicted protein
EHBc-102	$(TCG)_5$	F - GAGTCGGTGTCGGAGTCATT R - GCCAAACATGCGATAAATCTT	62	160	160	1	0	0	0	all	Acrilamide	reticulon-like protein B2-like isoform 1
EHBc-103	$(TC)_8TA(TC)_8$	F - CACGACGTTGTAAAACGACTTCTGCTTCAAAACCAAACAA R - GAGAGAGGTGGCAAAGAGGA	TD1	212	180-210	6	n/a	n/a	n/a	all	Fluorescence	F-box protein PP2-A13-like
EHBle-1	$(AAAT)_4$	F - CATATGCTCCCGGATAAAGG R - CCTGCCCCTTTCCATTCTAT	63	194	194	1	0	0	0	all	Acrilamide	B3 domain-containing protein
EHBle-2a	$(TTC)_6$	F - GGTGTTGAGGAATGCTGTTG R - TGTGTTTTGTTGGCCTTTTG	63	262	262-274	2	0.06	0.06	0.0526	all	Acrilamide	no hit
EHBle-2b	$(AAG)_6$	F - CACACAGCCCTACATGCAAT R - GGGGTTTGTTGTGCTGTTCT	63	223	223-235	3	0.59	0.56	0.5095	all	Acrilamide	
EHBle-3	$(CT)_{11}$	F - CACGACGTTGTAAAACGACTCTATCTAGATCATCCACACAACC R - TGGACGAGAACGAGAGAGAA	TD2	103	103-113	4	0.554	0.6111	0.4438	H. gui, H. rig, H. nit, H. bent, H. cam	Fluorescence	actin depolymerizing factor 4
EHBle-4	$(AT)_7$	F - GTGAGGTACGGCATGTAG R - TGCAGATTAATTTTCCCATTGA	62	251	198-252	4	0.3209	0	0.2973	H. rig, H. bent	Acrilamide	probable pectinesterase/pectinesterase inhibitor
EHBle-5	$(GT)_{11}$	F - CACGACGTTGTAAAACGACTGGCCTTGTCCATTGAACT R - GAACTTTGCAGCGGATTGTT	TD3	219	215-223	5	0.6881	0.7059	0.6228	all	Fluorescence	Sulfiredoxin-1, putative
EHBa-1	$(TGC)_7$	F - AGGCCTGTATCAATGGCAAC	62	257	338-356	5	0.7	0.18	0.6519	all	Capillary electrophoresi	Jasmonate-zim-domain protein 12, putative

		R - CTTGTACTGCTTTGCGGTGA									s	isoform 1
EHBa-2	$(AG)_7$	F - GCGACCTGTTCTCTCTCACC R - TGGAAGCTAAGGATCCAAATG	60	261	261-265	3	0.11	0.06	0.1037	H. gui, H. rig, H. nit, H. bent, H. cam	Acrilamide	conserved hypothetical protein
EHBa-3	$(TAA)_6$	F - CACGACGTTGTAAAACGACCGCTTGCTTAGCAGATGATG R - TGGAGCCAATGAATAAATTTC	63	206	221-279	5	0.2529	0.2	0.2378	all	Acrilamide	predicted protein
EHBa-4	$(ATCT)_5$	F - CACGACGTTGTAAAACGACGCACGGTAAGGCTAGGGTTT R - TGAGGAAGGAACGGAAGAGA	TD2	104	111-119	3	0.5317	0.4444	0.4091	all	Fluorescence	wound-induced basic protein
EHBa-5	$(GCT)_5$	F - CACGACGTTGTAAAACGACGGCAAATCAAAGGAACAAGC R - GAAACAACATTACACCACTCCAGA	TD2	205	205-208	2	0.0556	0.0556	0.0526	all	Fluorescence	acyl-CoA-binding protein
EHBa-6	$(TAT)_5$	F - CACGACGTTGTAAAACGACTTTGTTGAAGACGTCTGTTTTGA R - AAAGAAATACCACCAAGATGACA	TD2	267	267-288	2	0.5143	0.5556	0.375	all	Fluorescence	stress-induced hydrophobic peptide

Primer pairs number followed by letters a and b - belong to the same EST; H. gui - *Hevea guianensis*, H. rig - *Hevea rigidifolia*, H. nit - *Hevea nitida*, H. pau - *Hevea pauciflora*, H. bent – *Hevea benthamiana*, H. cam - *Hevea camargoana*; n/a - not applicable

p - panel library EST, la - latex library EST, c - cold-stressed library EST, le - leaf library EST, a - contig

Capítulo IV

Capítulo IV

"Saturation of a *Hevea brasiliensis* microsatellite-based linkage map and identification of candidate genes for growth-related traits."

Saturation of a *Hevea brasiliensis* microsatellite-based linkage map and identification of candidate genes for growth-related traits

Hevea brasiliensis is a native species of the Amazon basin and world's most important source of natural rubber. Although the Amazon region provides optimal conditions for rubber plantation, the occurrence of South American leaf blight (SALB) disease limits the rubber production in this region. Therefore, rubber plantation has been extended to suboptimal areas. Breeding programs have been searching for clones adapted to these suboptimal areas, but rubber tree breeding is time-consuming and expensive. The use of molecular markers and genetic maps in breeding programs can accelerate this process. In this work, novel microsatellite (EST-SSR) markers were added to a *H. brasiliensis* integrated genetic map. This new map consists of 330 markers distributed in 22 linkage groups (LG), with a total length of 3068.9cM and an average genetic distance of 9.3cM between adjacent markers. Groups LG11a and LG11b, in the previous map, were united into one group (LG11) with the addition of the new markers. The largest group covers a distance of 306.4cM (LG14) and the smallest covers 8.8cM (LG2b). Five loci were mapped into growth-related QTL regions and three of them had their transcripts annotated as proteins that are involved in stress responses. These loci could be candidate genes for the variation observed in the girth and height traits. The addition of new markers provided a denser rubber tree genetic map and the identification of loci that could be involved in growth traits. This new map may help the positioning of genes of interest; identification of QTLs related to other important agronomic characteristics and marker-assisted selection in the rubber tree program.

Keywords: rubber tree, linkage map, EST-SSRs, *Hevea brasiliensis*.

Introduction

Hevea brasiliensis [(Wild. ex Adr. de Juss.) Muell-Arg.] is world's most important rubber producing species and is a diploid (2n = 36) perennial tree. The *Hevea* genus comprises eleven inter-crossable species that are native to the Amazon basin and belongs to the Euphorbiaceae family (Pires *et al.*, 2002; Gonçalves *et al.*, 1990). Even though the Amazon provides optimal conditions for the rubber tree development, the presence of the fungus *Microcyclus ulei* (P. Henn) v. Arx, which causes the South American leaf bligtht (SALB) disease, restrains the rubber plantation in the region (Pushparajah, 2001). As a consequence, rubber production was extended to several suboptimal environments worldwide, in an attempt to evade the disease. Nevertheless these new areas present various climatic constraints, such as low temperatures and dry periods (Priyadarshan and Goncalves, 2003). Breeding programs have been searching for clones adapted to these suboptimal areas, but rubber tree breeding is time-consuming and expensive (Gonçalves and Fontes, 2012).

The development of molecular markers and molecular markers-based maps provided new tools for applied genetics and breeding programs. They allow the analysis of genetic diversity, characterization of genotypes, identification of quantitative trait loci (QTLs) and genes of interest (Besse *et al.*, 1994; Lespinasse *et al.*, 2000; Le Guen *et al.*, 2003; Saha *et al.*, 2005; Souza *et al.*, 2009; Feng *et al.*, 2009; Le Guen *et al.*, 2011; Mantello *et al.*, 2012; Souza *et al.*, 2013). The association of markers with characteristics of interest enables the application of marker-assisted selection (MAS); and high-density

genetic maps combined with phenotypic data allows the application of genome wide selection (GWS). Both thechniques are of great importance for perennial trees with a long life cycle, such as the rubber tree, since they could reduce breeding programs time and costs (Neale and Kramer, 2011; Heffner *et al.*, 2009).

Due to *H. brasiliensis* extended juvenile period and inbreeding depression, the development of homozygous inbred lines is fairly difficult. As such, traditional methods for constructing linkage maps (i.e.: F_2 population) are not suitable for rubber tree (Grattapaglia and Sederoff, 1994). The first rubber tree genetic map was constructed using a double "pseudo-testcross" strategy, using a full-sib cross (or F_1 population). The map contains 717 markers, mostly RFLPs and AFLPs, distributed in 18 linkage groups (LGs) (Lespinasse *et al.*, 2000). In 2011, another two genetic maps were constructed, both using the "pseudo-testcross" strategy, with only SSR markers (Triwitayakorn *et al.*, 2011) and SSR and AFLP markers (Le Guen *et al.*, 2011). Souza *et al.* (2013) developed an integrated rubber tree map comprised of 284 SSR markers (226 genomic SSRs and 58 EST-SSRs), distributed in 23 LGs, using OneMap software (Margarido et al., 2007) and multipoint analysis based on the Hidden Markov Method (Mollinari et al., 2009). The map has a total length of 2792.8cM with a density of one marker per 9.83cM. In addition, 18 QTLs for growth-related traits during summer and winter periods were detected in 11 LGs.

In this work, novel EST-SSR markers developed by Silva *et al.* (2014; unpublished data) were added to the integrated map of Souza *et al.* (2013) using the same methodology for the construction of the genetic map. Loci that were mapped into QTL regions could be candidate genes for the variation observed in the growth-related traits.

Methods

Mapping population

The mapping population was the same used by Souza *et al.* (2013): 270 individuals derived from a cross among clones PB 217 (high rubber yield potential) and PR 255 (tolerant to injury and low temperatures). The progeny was obtained by controlled pollination and multiple replicates were generated by bud graffiting onto rootstocks (GT 1 seedlings) for the field experiment. The mapping population was planted in the Edouard Michelin Plantation (Mato Grosso state, Brazil, 17° 23' 59.60'' S and 54° 44' 53.93'' W, altitude 519 m) in 2006.

DNA extraction and Molecular marker analysis

The mapping population DNA samples were the same used in the work by Souza *et al.* (2013). The polymorphisms from a total of 173 EST-SSR markers (Silva *et al.*, 2014; unpublished data) were screened for both the PB217 and PR255 parents. Alleles were visualized by three alternative techniques: a 4300 DNA Analyzer (LI-COR Biosciences, USA), an Advance FS96 dsDNA Fragment Analyzer (Advanced Analytical Technologies Inc., USA) and 6% denaturing polyacrylamide gels with silver staining (Creste *et al.* 2001). All PCR reactions were performed according to Silva *et al.* (2014).

Linkage Map

A chi-square test was performed to verify the segregation of the markers. Linkage analysis was performed using the software OneMap (Margarido *et al.*, 2007) and Hidden Markov Models-based multipoint technology (Wu *et al.*, 2002). Cosegregation groups were determined with a 4.5 LOD score and a 0.4 recombination fraction. The markers order was established with algorithms compare (for groups with up to six markers) and order (for

groups with more than six markers) (Lander *et al.*, 1987; Mollinari *et al.*, 2009). The Kosambi map function was used to convert the recombination fractions between the markers to centiMorgan units (Kosambi, 1944).

Results and Discussion

Marker polymorphism and segregation analysis

From the 173 EST-SSR markers developed by Silva *et al.* (2014; unpublished data), 54 (31.4%) exhibited polymorphism between the parents of the mapping population. This percentage of EST-SSRs polymorphism among the parents was smaller than the observed for genomic microsatellite markers (54.5%) and a little smaller than EST-SSR markers (39%) (Souza *et al.*, 2013). Although EST-SSR markers present a lower polymorphism level than genomic SSR makers, they have a higher probability of being inserted in functional gene sequences, creating the possibility of mapping and identifying candidate genes (Varshney *et al.*, 2005). From the 54 polymorphic markers identified, a total of 46 were selected to genotype the 270 individuals of the mapping population.

A chi-squared test was performed on the genotyped loci and 28 (60.9%) presented a 1:1 segregation ratio, six (13%) showed a 1:2:1 segregation ratio and 12 (26.1%) exhibited a segregation ratio of 1:1:1:1. Four markers (EHBp23b, EHBp17, EHBp15 and EHBc90; 8.7%) showed distorted segregation ($P \leq 0.005$, chi-squared test); however these markers did not caused distortions in the LGs. These four loci transcripts presented similarity to proteins aquaporin (EHBp23b), 60S ribosomal protein L34 (EHBp17), a nuclear-encoded mitochondrial acyl carrier protein (EHBp15) and a conserved hypothetical protein (Silva *et al.*, 2014) bearing a mitochondrial ribosome subunit S26 domain (MRP-S26) (EHBc90). Three of them are reported to be related to important processes in plants.

Aquaporins are transmembrane channels involved in water transport (Chaumont *et al.*, 2005), shown to be essential in leaf and petal movements (Azad *et al.*, 2004; Siefritz at al., 2004), reproduction (Bots *et al.*, 2005), seed germination (Vander Willigen *et al.*, 2006), and salt and drought stresses tolerance (Xu *et al.*, 2014). The ribosomal protein L34 presents a higher expression in stems and roots than in leaves and flowers (Gao et al, 1994), its promoter activity is higher in actively growing tissues and is enhanced in response to wounding and plant growth regulators (Dai *et al.*, 1996). The mitochondrial acyl carrier protein is involved in the fatty acid biosynthetic process (Yasuno *et al.*, 2004), which is an essential component of plant cellular membranes, cutin and suberin (Beisson *et al.*, 2007). These skewed segregations detected may be the result of several biological factors, such as chromosome loss, locus duplication and gamete selection (Liebhard *et al.*, 2003; Moyle and Graham, 2006). In the case of gamete selection, these loci, or possibly other loci linked to them, may be influencing pollen and zygotes viability.

Linkage map

Forty-six EST-SSR markers were used to enhance the coverage of the microsatellite-based rubber tree integrated genetic map constructed by Souza *et al.* (2013). The new map has 330 markers assigned to 22 LGs, spanning a total of 3068.9cM, which is 276.1cM larger than the original map. Loci that presented similarity to proteins involved in response to stress (i.e.: EHBp13) and developmental processes (i.e.: EHBc20) (Silva *et al.*, 2014) are among the new mapped loci. The larger LG has 306.4cM in length (LG14) and the smallest presents a length of 8.8cM (LG2b). Groups LG11a and LG11b, in the previous map, were united into one group (LG11) with the mapping of marker EHBle3, that presents similarity to actin depolymerizing factor 4 protein, between markers A2535 and HBE30.

A primer pair that presented a polyploid pattern of segregation (EHBp27) (Silva *et al.*, 2014) was most probably amplifying three different genomic regions (Fig.1). Thereby two of these loci (EHBp27a and c) were genotyped in the mapping population and the linkage analysis mapped both into LG11 with a distance of only 0.8cM. The *Ricinus communis* proteins with which these loci presented similarity (RCOM_0175580 and RCOM_0175590) are also close together: the distance between them is approximately 1,840 bp (Chan *et al.*, 2010). The EST used to design this primer pair bears a Rho (Ras-like) small GTPase putative conserved domain, and in plants, Rho-related GTPases are involved in the regulation of hormone functions and developmental processes and form a family with several members (Yang, 2002). The two distinct genomic regions amplified by primer pair EHBp23 (Silva *et al.*, 2014) were mapped as well: LG3 bears the locus EHBp23a and locus EHBp23b was placed in LG13. The genetic map is presented in Fig. 2 and statistics for each LG are presented in Table 1.

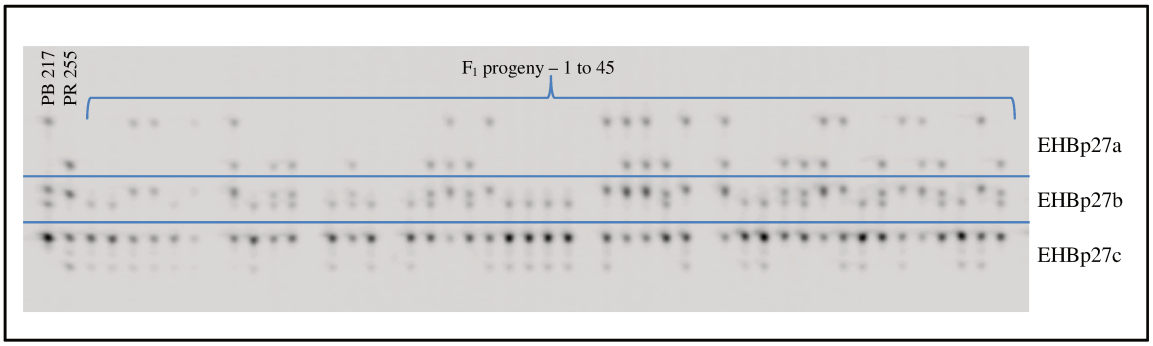


Figure 1 – Genotyping of the F_1 population (plants 1 to 45) with primer pair EHBp27. The lines separate the probable different genomic regions named EHBp27a, b and c. PB 217 and PR 255 – F_1 parents.

TABLE 1 – The distribution of the markers in the linkage groups.

Linkage Group	N° of markers	Length	Marker density
LG1	25	166.4	6.66
LG2a	14	98.6	7
LG2b	3	8.8	2.9
LG3	19	177.3	9.3

LG4	11	115.1	10.46
LG5	22	193.7	8.8
LG6a	14	82.1	5.9
LG6b	3	38.9	13
LG7	19	197.4	10.4
LG8	25	194	7.8
LG9	14	117	8.4
LG10	33	252.8	7.7
LG11	8	160.3	20
LG12	12	201.8	16.8
LG13	15	167.4	11.2
LG14	25	306.4	12.3
LG15	16	155.1	9.7
LG16	18	209.3	11.6
LG17a	7	54.5	7.8
LG17b	5	23.4	4.68
LG18	16	138.3	8.6
LG19	4	10.3	2.56

The LGs were classified from one to 19 and groups 2, 6 and 17 were still divided into subgroups “a” and “b”, although new markers were added to these groups. Markers density was slightly higher (1/9.3cM) than the original map (1/9.8cM); however it was lower than Lespinasse *et al.* (2000) map (1/3cM), which was saturated with several RFLP and AFLP markers. Nevertheless, although the LGs marker distribution was not uniform, the new markers were able to resolve three of the five gaps present in the Souza *et al.* (2013) map (LGs 11, 13 and 15). The two gaps that remained are in LG12 (66.3cM) and LG 16 (43.3cM) (Fig. 2). The gaps observed may be the result of low level of microsatellite polymorphisms in these regions, and could be associated with heterochromatic regions near the centromeres. In order to confirm this hypothesis in rubber tree, cytogenetics analyses need to be carried out to demonstrate this association. Another reason for the gaps is that the recombination events or mapped loci are not equally distributed through the genome.

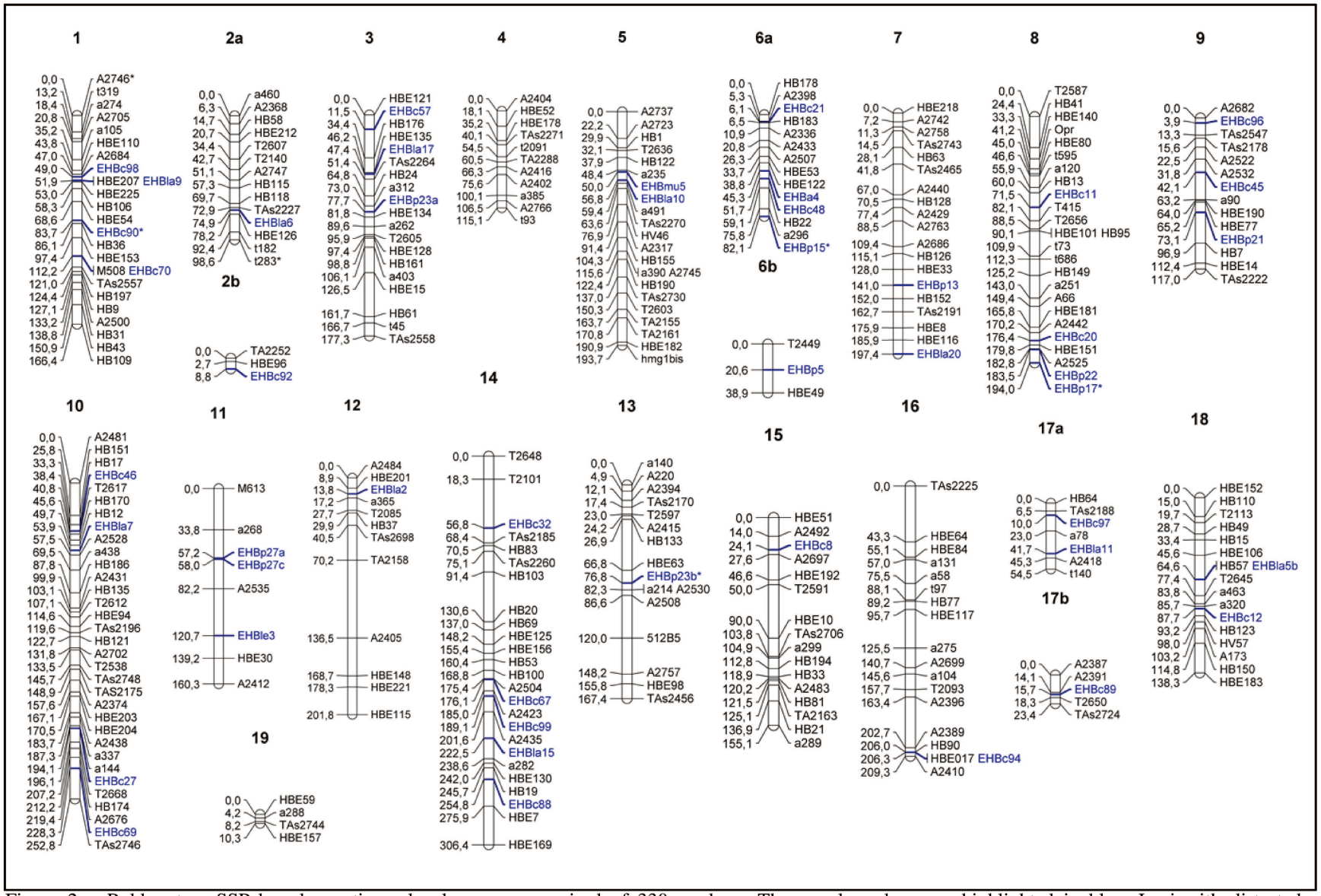


Figure 2 – Rubber tree SSR-based genetic molecular map comprised of 330 markers. The novel markers are highlighted in blue. Loci with distorted segregation are marked with an asterisk.

The low marker density in some LGs may as well be due to highly homozygous regions subjected to lower recombination rates (Castiglioni et al., 1999).

Eighteen linkage groups were expected for *H. brasiliensis* (2n = 36) linkage map, however this number was exceeded by four LGs. Since some LGs are small, this exceeding number can be due to incomplete coverage of the genome by the microsatellite loci. A similar result was obtained for a rubber tree linkage map by Triwitayakorn *et al.* (2011) that used only SSR and EST-SSR markers. Nevertheless Lespinasse *et al.* (2000) and Le Guen *et al.* (2011) showed that using several types of markers it is possible to obtain 18 linkage groups, indicating that different markers may be necessary to fill in the gaps between the groups.

Although the new map most likely presents an incomplete coverage, it may help the positioning of genes of interest, the identification of QTLs related to other important agronomic characteristics and MAS. The addition of new markers to the genetic map and the phenotyping of the mapping population may possibilitate the use of GWS in the rubber tree program as well.

Candidate genes for growth-related traits

Souza *et al.* (2013) identified 18 QTLs in 11 linkage groups for height and girth growth during summer and winter periods. Among the 46 new loci mapped in this work, five were positioned into six QTLs regions: summer girth (SG) 1, total height (TH) 1, SG.2, total girth (TG) 1, TH.2 and summer height (SH) 6 (Fig. 3). SG.1 and TH.1 were identified in LG5 between markers a235 and a491 (Souza *et al.*, 2013) and markers EHBmu5 and EHBla10 were mapped between those markers.

The transcript from which EHBmu5 primer pair was designed showed similarity to heat shock factor-binding protein (HSBP) 1-like (Silva CC, unpublished results). Heat shock proteins (HSP) are known to be an early adaptive response to several types of stresses (Wang *et al.*, 2004) and *hsps* transcription is controlled by heat stress transcription factors (HSF) (Baniwal *et al.*, 2004). HSBP1 acts as a negative regulator of the heat shock response by binding to HSF1 in humans (Satyal *et al.*, 1998) and proteins similar to HSBP1 were discovered in plants. In maize, a mutated *empty pericarp 2* (*emp2*) gene, which encodes a HSBP-like protein, is embryo lethal and its lethality was attributed to the unattenuation of the heat shock response with up-regulation of *hsp* genes in the mutant (Fu *et al.*, 2002) and the interaction of EMP2 protein with some HSFs was reported (Fu *et al.* 2006). Similar results were reported for *Arabidopsis thaliana* AtHSBP protein: *AtHSBP*-knockout lines presented seed abortion and AtHSBP negatively regulates the heat stress response (Hsu *et al.*, 2010). The EHBmu5 locus product may be involved in the heat stress response in rubber tree, which is seemingly important during plant development. The positioning of this marker into a growth-related QTL region suggests that this locus could be considered a candidate gene for growth-related traits.

The EHBla10 locus transcript showed similarity to 2-methyl-6-geranylgeranyl benzoquinone methyltranferase (MGGBQ-MT) (Silva *et al.*, 2014), which is a protein from the vitamin E biosynthesis pathway (Farré *et al.*, 2012). Vitamin E is a group of lipophilic antioxidants that are essential for seed longevity, by reducing the accumulation of lipid oxidation products, and are also involved in the protection of chloroplasts against photooxidative stress in *A. thaliana* (Sattler *et al.*, 2004; Havaux *et al.*, 2005). The EHBla10 locus was mapped close to the EHBmu5 locus, and its product could also be

related to stress responses and developmental processes in rubber tree, therefore this locus could be a good candidate for growth-related traits further analyses.

SH.6 was detected in LG9 between markers HBE77 and HB7 (Souza *et al.*, 2013). Marker EHBp21 was mapped inside this QTL region and its transcript presented similarity to jasmonate-ZIM-domain (JAZ) protein 3 (Silva *et al.*, 2014). Jasmonic acid (JA) is involved in regulation of plant secondary metabolites biosynthesis, response to wounding (i.e.: herbivory) and necrotrophic pathogens, reproductive organs development, leaf senescence and growth inhibition (Wasternack and House, 2013). JAZ proteins are negative regulators of the JA signaling pathway by interacting with transcriptional activators of JA-responsive genes, which includes the *JAZ* genes (Chini *et al.*, 2007; Yan *et al.*, 2007; Chung *et al.*, 2008). The feedback loop created by JA induction of *JAZ* genes transcription must serve to attenuate the transcriptional response soon after it is initiated, providing a mechanism for the adjustment of the expression of JA-responsive genes (Chini *et al.*, 2007; Chung *et al.*, 2008). JAZ proteins are redundant in their negative regulation functions, however mutations that prevent JAZ protein degradation and the overexpression of JAZ proteins lead to a jasmonate-insensitive phenotype (i.e.: repression of JA-mediated growth inhibition) (Chini *et al.*, 2007; Yan *et al.*, 2007). EHBp21 locus product may be involved in JA-mediated defense and developmental processes, therefore this locus may be related to growth-related traits in rubber tree as well.

LG9 also bears QTL TH.2, present between markers A2532 and a90 (Souza *et al.*, 2013), and marker EHBc45 was mapped into that region. The EHBc45 locus product showed similarity to a predicted protein detected in several plants (Silva *et al.*, 2014). Although this protein seems to be conserved among different plant species, there is no biological function attributed to it. A similar situation was presented by EHBp5 locus,

which was mapped in QTLs SG.2 and TG.1 present in LG6b among markers T2449 and HBE49. The EST from which EHBp5 primer pair was designed showed no similarity to any protein or EST sequence in the GenBank database, however this EST length is only 217 bp. The target sequence length may affect the identification of similarity to known proteins or transcripts (Silva *et al.*, 2014), as such it is possible that this short EST may belong to an untranslated region (UTR) of a known sequence. Even though both loci cannot be associated to known biological processes, they may be linked to genes that are involved in developmental processes in rubber tree.

Quantitative traits, such as height and girth, are dependent on a combination of individual effects of several genes and are influenced by environment fluctuations as well (Griffing and Scholl, 1991; Neale and Kramer, 2011). The analysis of candidate genes has been a powerful approach to dissect the genetics of these complex traits (Neale and Kramer, 2011; Fabrini *et al.*, 2012; Prunier *et al.*, 2013; Celton *et al.*, 2014). The loci identified here open new possibilities for further molecular investigations. These loci may be associated to height and girth growth in rubber tree by being physically close to the functional variant or even be the functional variant themselves. An analysis of these loci effects based on the population phenotypic data may help to implement the GWS approach into rubber tree breeding programs.

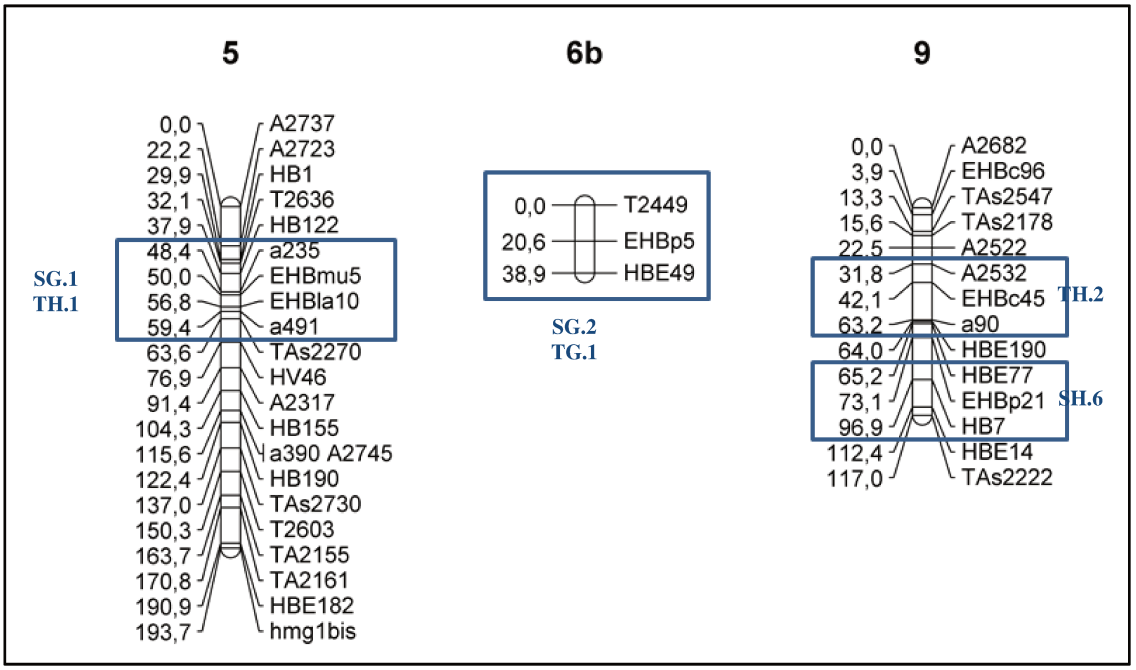


Figure 3 – LGs bearing QTL regions into which loci analyzed in this work were mapped.

Acknowledgments

This work was supported by grants from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, proc. 2007/50562-4; 2012/50491-8) and scholarships to CCS (2009/52975-0), CCM (2008/55974-1, 2011/50188-0) and LMS (2012/05473-1). PSG and APS are recipients of a research fellowship from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

References

Azad AK, Sawa Y, Ishikawa T, Shibata H (2004). Phosphorylation of plasma membrane aquaporin regulates temperature-dependent opening of tulip petals. Plant Cell Physiol. 45:608–617.

Baniwal SK, Bharti K, Chan KY, Fauth M, Ganguli A, Kotak S, Mishra SK, Nover L, Port M, Scharf KD, Tripp J, Weber C, Zielinski D, von Koskull-Döring P (2004). Heat stress response in plants: a complex game with chaperones and more than twenty heat stress transcription factors. J Biosci 29(4):471-87.

Beisson F, Li Y, Bonaventure G, Pollard M, Ohlrogge JB (2007). The acyltransferase GPAT5 is required for the synthesis of suberin in seed coat and root of Arabidopsis. Plant Cell 19(1):351-68.

Besse P, Seguin M, Lebrun P, Chevallier MH, Nicolas D, Lanaud C (1994). Genetic diversity among wild and cultivated populations of Hevea brasiliensis assessed by nuclear RFLP analysis. Theor Appl Genet 88:199-207.

Bots M, Vergeldt F, Wolters-Arts M, Weterings K, van As H, Mariani C (2005). Aquaporins of the PIP2 class are required for efficient anther dehiscence in tobacco. Plant Physiol. 137:1049–1056.

Castiglioni P, Ajmone-Marsan P, Van Wijk R, Motto M (1999). AFLP markers in a molecular linkage map of maize: codominant scoring and linkage group distribution. Theor Appl Genet 99:425–431.

Celton JM, Kelner JJ, Martinez S, Bechti A, Khelifi Touhami A, James MJ, Durel CE, Laurens F, Costes E (2014). Fruit self-thinning: a trait to consider for genetic improvement of apple tree. PLoS One 9(3):e91016. doi: 10.1371/journal.pone.0091016

Chan AP, Crabtree J, Zhao Q, Lorenzi H, Orvis J, Puiu D, Melake-Berhan A, Jones KM, Redman J, Chen G, Cahoon EB, Gedil M, Stanke M, Haas BJ, Wortman JR, Fraser-Liggett CM, Ravel J, Rabinowicz PD (2010). Draft genome sequence of the oilseed species Ricinus communis. Nat Biotechnol 28(9):951-6.

Chaumont F, Moshelion M, Daniels MJ (2005). Regulation of plant aquaporin activity. Biol Cell 97(10):749-764.

Chini A, Fonseca S, Fernández G, Adie B, Chico JM, Lorenzo O, García-Casado G, López-Vidriero I, Lozano FM, Ponce MR, Micol JL, Solano R (2007). The JAZ family of repressors is the missing link in jasmonate signalling. Nature 448:666-71.

Chung HS, Koo AJ, Gao X, Jayanty S, Thines B, Jones AD, Howe GA (2008). Regulation and function of Arabidopsis JASMONATE ZIM-domain genes in response to wounding and herbivory. Plant Physiol 146(3):952-64.

Creste S, Tulmann A, Figueira A (2001). Detection of single sequence repeat polymorphism in denaturating polyacrylamide sequencing gels by silver staining. Plant Mol Biol Rep 19:299-306.

Dai Z, Gao J, An K, Lee JM, Edwards GE, An G (1996). Promoter elements controlling developmental and environmental regulation of a tobacco ribosomal protein gene L34. Plant Mol Biol. 32(6):1055-65.

Fabbrini F, Gaudet M, Bastien C, Zaina G, Harfouche A, Beritognolo I, Marron N, Morgante M, Scarascia-Mugnozza G, Sabatti M (2012). Phenotypic plasticity, QTL mapping and genomic characterization of bud set in black poplar. BMC Plant Biol 12:47. doi: 10.1186/1471-2229-12-47.

Farré G, Zorrilla U, Berman J, Zhu C, Christou P, Capell T (2012). Increasing the vitamin E content of food by in-plant production. In: Hemming D (ed). CAB Reviews 7, n° 30. CAB International, Oxfordshire, UK, pp. 137-146.

Feng SP, Li WG, Huang HS, Wang JY, Wu YT (2009). Development, characterization and cross-species/genera transferability of EST-SSR markers for rubber tree (*Hevea brasiliensis*). Mol Breeding 23(1):85-97.

Fu S, Meeley R, Scanlon MJ (2002). *empty pericarp2* encodes a negative regulator of the heat shock response and is required for maize embryogenesis. Plant Cell 14(12):3119-32.

Fu S, Rogowsky P, Nover L, Scanlon MJ (2006). The maize heat shock factor-binding protein paralogs EMP2 and HSBP2 interact non-redundantly with specific heat shock factors. Planta 224(1):42-52.

Gao J, Kim SR, Chung YY, Lee JM, An G (1994). Developmental and environmental regulation of two ribosomal protein genes in tobacco. Plant Mol Biol. 25(5):761-70.

Gonçalves P, Cardoso M, Ortolani AA (1990). Origem, variabilidade e domesticação da *Hevea*; uma revisão. Pesq Agrop Bras 25(2):135-156.

Gonçalves PS, Fontes JRA (2012). Domestication and breeding of rubber tree. In: Borem A, Lopes MTG, Clement CR, Noda H (ed) Domestication and Breeding: Amazonian Species. UFV, Viçosa, Brasil, pp 393-420.

Grattapaglia D, Sederoff R (1994). Genetic linkage maps of Eucalyptus grandis and E. urophylla using a pseudo-testcross mapping strategy and RAPD markers. Genetics 137: 1121-1137.

Griffing B, Scholl RL (1991). Qualitative and Quantitative Genetic Studies of *Arabidopsis thaliana*. Genetics 129:605-609.

Havaux M, Eymery F, Porfirova S, Rey P, Dörmann P (2005). Vitamin E protects against photoinhibition and photooxidative stress in *Arabidopsis thaliana*. Plant Cell 17(12):3451-69.

Heffner EL, Sorrels ME, Jannink J-L (2009). Genomic Selection for Crop Improvement. Crop Sci 49(1):1-12.

Hsu SF, Lai HC, Jinn TL (2010). Cytosol-localized heat shock factor-binding protein, AtHSBP, functions as a negative regulator of heat shock response by translocation to the nucleus and is required for seed development in Arabidopsis. Plant Physiol 153(2):773-84.

Kosambi DD (1944). The estimation of map distances from recombinant values. Annals of Eugenics 12:172-175.

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln Newburg EE (1987). MapMaker: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181.

Le Guen M, Lespinasse D, Oliver G, Rodier-Goud M, Pinard F, Seguin M (2003). Molecular mapping of genes conferring field resistance to South American Leaf Blight (*Microcyclus ulei*) in rubber tree. Theor Appl Genet 108:160-167.

Le Guen, V., Garcia, D., Doare, F., Mattos CR, Condina V, Couturier C, Chambon A, Weber C, Espéout S, Seguin M (2011). A rubber tree's durable resistance to *Microcyclus ulei* is conferred by a qualitative gene and a major quantitative resistance factor. Tree Genet Genomes doi:10.1007/s11295-11011-11038-11297.

Lespinasse D, Rodier-Godier M, Grivet L, Leconte A, Legnate H, Seguin M (2000). A saturated genetic linkage map of rubber tree (*Hevea* spp.) based on RFLP, AFLP, microsatellite, and isozyme markers. Theor Appl Genet. 100:127-138.

Li H, Qin Y, Xiao X, Tang C (2011). Screening of valid reference genes for real-time RT-PCR data normalization in Hevea brasiliensis and expression validation of a sucrose transporter gene HbSUT3. Plant Sci 181:132-139.

Liebhard R, Koller B, Gianfranceschi L, Gessler C (2003). Creating a saturated reference map for the apple (Malus x domestica Borkh.) genome. Theor Appl Genet 106(8):1497-508.

Mantello CC, Suzuki FI, Souza LM, Gonçalves PS, Souza AP (2012). Microsatellite marker development for the rubber tree (*Hevea brasiliensis*): characterization and cross-amplification in wild *Hevea* species. BMC Res Notes. 5:329. doi: 10.1186/1756-0500-5-329.

Margarido GRA, Souza AP, Garcia AAF (2007). OneMap: software for genetic mapping in outcrossing species. Hereditas 144:78-79.

Mollinari M, Margarido GRA, Vencovsky R, Garcia AAF (2009). Evaluation of algorithms used to order markers on genetic maps. Heredity 103:494–502.

Moyle LC, Graham EB (2006). Genome-wide associations between hybrid sterility QTL and marker transmission ratio distortion. Mol. Biol. Evol. 23(5):973–980.

Neale DB, Kremer A (2011). Forest tree genomics: growing resources and applications. Nature Reviews 12:111-122.

Pires JM, Secco RS, Gomes JI (2002). Taxonomia e fitogeografia das seringueiras *Hevea* spp. Embrapa amazômia oriental, Belém, Brazil.

Priyadarshan PM, Gonçalves PS (2003). *Hevea* gene pool for breeding. Genet Resour Crop Evol 50:101-114.

Prunier J, Pelgas B, Gagnon F, Desponts M, Isabel N, Beaulieu J, Bousquet J (2013). The genomic architecture and association genetics of adaptive characters using a candidate SNP approach in boreal black spruce. BMC Genomics 14:368. doi: 10.1186/1471-2164-14-368.

Pushparajah E (2001). Natural rubber. In: Last FT (ed) Tree Crop Ecosystems. Elsevier Science, Amsterdam, pp 379-407. [Goodall DW (Series Editor): Ecosystems of the World Series, Vol 19].

Saha T, Bindu Roy C, Nazeer MA (2005). Microsatellite variability and its use in the characterization of cultivated clones of *Hevea brasiliensis*. Plant Breed 124:86-92.

Sattler SE, Gilliland LU, Magallanes-Lundback M, Pollard M, DellaPenna D (2004). Vitamin E is essential for seed longevity and for preventing lipid peroxidation during germination. Plant Cell 16(6):1419-32.

Satyal SH, Chen D, Fox SG, Kramer JM, Morimoto RI (1998). Negative regulation of the heat shock transcriptional response by HSBP1. Genes Dev 12(13):1962-74.

Siefritz F, Otto B, Bienert GP, Van Der Krol A, Kaldenhoff R (2004). The plasma membrane aquaporin NtAQP1 is a key component of the leaf unfolding mechanism in tobacco. Plant J. 37, 147–155.

Silva CC, Mantello CC, Campos T, Souza LM, Gonçalves PS, Souza AP (2014). Leaf-, panel- and latex-expressed sequenced tags from the rubber tree (*Hevea brasiliensis*) under cold-stressed and suboptimal growing conditions: the development of gene-targeted functional markers for stress response. Mol Breed DOI 10.1007/s11032-014-0095-2.

Souza LM, Mantello CC, Santos MO, Gonçalves PO, Souza AP (2009). Microsatellites from rubber tree (*Hevea brasiliensis*) for genetic diversity analysis and cross amplification in six *Hevea* wild species. Cons Genet Res 1:75–79.

Souza LM, Gazaffi R, Mantello CC, Silva CC, Garcia D, Le Guen V, Cardoso SEA, Garcia AAF, Souza AP (2013). QTL mapping of growth-related traits in a full-sib family of rubber tree (*Hevea brasiliensis*) evaluated in a sub-tropical climate. PLoS One 8(4):e61238. doi/10.1371/journal.pone.0061238.

Triwitayakorn K, Chatkulkawin P, Kanjanawattanawong S, Sraphet S, Yoocha T, Sangsrakru D, Chanprasert J, Ngamphiw C, Jomchai N, Therawattanasuk K, Tangphatsornruang S (2011). Transcriptome sequencing of *Hevea brasiliensis* for development of microsatellite markers and construction of a genetic linkage map. DNA Res 18(6):471–482.

Vander Willigen C, Postaire O, Tournaire-Roux C, Boursiac Y, Maurel C (2006). Expression and inhibition of aquaporins in germinating Arabidopsis seeds. Plant Cell Physiol. 47:1241–1250.

Varshney RK, Graner A, Sorrells ME (2005). Genic microsatellite markers in plants: features and applications. Trends Biotechnol 23(1):48-55.

Wang W, Vinocur B, Shoseyov O, Altman A (2004). Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. Trends Plant Sci 9(5):244-52.

Wasternack C, Hause B (2013). Jasmonates: biosynthesis, perception, signal transduction and action in plant stress response, growth and development. An update to the 2007 review in Annals of Botany. Ann Bot. 111(6):1021-58.

Wu R, Ma CX, Painter I, Zeng ZB (2002). Simultaneous maximum likelihood estimation of linkage and linkage phases in outcrossing species. Theor Pop Biol 61:349-363.

Xu Y, Hu W, Liu J, Zhang J, Jia C, Miao H, Xu B, Jin Z (2014). A banana aquaporin gene, MaPIP1;1, is involved in tolerance to drought and salt stresses. BMC Plant Biol. 14:59. doi: 10.1186/1471-2229-14-59.

Yang Z (2002). Small GTPases: versatile signaling switches in plants. Plant Cell 14 Suppl:S375-88.

Yan Y, Stolz S, Chételat A, Reymond P, Pagni M, Dubugnon L, Farmer EE (2007). A downstream mediator in the growth repression limb of the jasmonate pathway. Plant Cell 19(8):2470-83.

Yasuno R, von Wettstein-Knowles P, Wada H (2004). Identification and molecular characterization of the beta-ketoacyl-[acyl carrier protein] synthase component of the Arabidopsis mitochondrial fatty acid synthase. J Biol Chem. 279(9):8242-51.

Capítulo V

Resultados complementares

A necessidade de novos cultivares de seringueira (*Hevea brasiliensis*) adaptáveis a diferentes regiões ecológicas, denominadas de regiões de escape para o cultivo da *H. brasilienses*, constitui um ponto importantíssimo para o sucesso da heveicultura no Brasil e no mundo. O ciclo de melhoramento genético da seringueira, que vai do cruzamento para obtenção das progênies recombinantes até o final da produção, pode demorar aproximadamente 30 anos para se concretizar. Desta forma, torna-se fundamental o desenvolvimento de novas técnicas de avaliação precoce, que possibilitem diminuir e otimizar as avaliações para essa cultura. Neste contexto, o projeto tinha como objetivo a construção de bibliotecas de cDNA, a geração de sequências EST (*Expressed Sequence Tag*) e o desenvolvimento de marcadores microssatélites gene-direcionados, para saturar e posicionar possíveis genes em um mapa genético-molecular de uma população F_1 de *Hevea brasiliensis*.

No período em que o projeto foi elaborado (agosto/2009), havia poucas sequências EST de seringueira depositadas em bancos públicos de sequências. Desta maneira, a construção de bibliotecas de cDNA de diferentes tecidos e tratamentos foi a estratégia escolhida para a geração de ESTs. Ao todo foram construídas sete bibliotecas de acordo com o tecido ou tratamento utilizado: “painel”, “látex”, “folha”, “frio-6h”, “frio-10h”, “frio-24h” e “M.ulei”. Os resultados obtidos para as bibliotecas, bem como os marcadores moleculares desenvolvidos a partir de suas sequências, com exceção de “M.ulei”, estão descritos no artigo presente no capítulo III. Apresentamos a seguir os resultados complementares deste trabalho sobre a biblioteca de cDNA “M.ulei”.

Microcyclus ulei é um fungo ascomiceto que ataca os folíolos jovens de seringueira, causando perda das folhas e podendo acarretar em morte da planta. A disseminação dos conídios se dá através dos ventos e da chuva e as folhas infectadas ficam com aspecto de lixa (Gasparotto *et al.*, 1990). Há cinco níveis principais na interação hospedeiro-patógeno: 1) germinação e penetração, que acontece nas primeiras 24h após o contato dos esporos com os tecidos foliares jovens; 2) infecção inicial, que consiste na colonização dos tecidos adjacentes aos feixes vasculares pela hifa através de seu crescimento intercelular (24h-3 dias); 3) colonização da folha, crescimento intercelular e ramificação da hifa acontecem nos próximos 3-5 dias pelo crescimento de hifas laterais perpendiculares à hifa principal; 4) as hifas laterais direcionam seu crescimento à superfície dos folíolos e, aproximadamente, 24h após sua formação, estas hifas saem pela superfície abaxial, formando conidióforos; e, em casos onde houve baixa incidência de infecção, 5) há a formação de estruturas estromáticas, como pseudotécios e ascósporos (Lieberei, 2007).

Um Ensaio de infecção controlada por *Microcyclus ulei* foi realizado, nas Plantações Michelin Bahia (Igrapiúna/BA), em plantas enxertadas dos clones Fx 2784, PFB 5, FDR 5788 (resistência/tolerância ao *M. ulei*), RRIM 600 (suscetível ao *M. ulei*), F 4542 (*H. benthamiana*) e PA 31 (*H. pauciflora*) (utilizadas no melhoramento da seringueira com intuito de introdução de genes de resistência em *H. brasiliensis* através de cruzamentos). Conídios do isolado PMB136 de *M. ulei* foram coletados de uma lesão, com a utilização de um pincel estéril, de uma folha previamente infectada de RRIM 600 aproximadamente 30 minutos antes da infecção (Fig. 1A), e colocados em solução de Tween80 0,05%. Foi feita a contagem do número de esporos em câmara de Neubauer, sendo a concentração final 230.000 conídios/mL. Uma alíquota da suspensão de conídios (50µL) foi plaqueada em

meio ágar-ágar para teste de germinação e aproximadamente 98% dos conídios germinaram.

As plantas enxertadas foram transferidas da casa telada para a câmara úmida (25°C, 99,9% de umidade) e um compressor de ar (FANEM) foi utilizado para fazer a inoculação dos esporos na face abaxial das folhas jovens (estágio B2 de crescimento) (Fig. 1B e C, respectivamente). Foram inoculadas duas folhas por clone. As plantas foram mantidas no escuro nas primeiras 24h (para manter a umidade), passando em seguida para um regime de 12h de luz por dia. Os folíolos infectados foram coletados com 48h e cinco dias após a infecção. Folíolos sadios com cinco dias de crescimento também foram amostrados. Folíolos com nove dias de infecção dos clones RRIM 600, PA 31 e F 4542, que já estavam presentes na câmara úmida quando do momento do ensaio, também foram coletados, sendo que foram selecionados apenas, com a utilização de um bisturi, os tecidos das regiões que apresentavam lesões causadas pelo fungo e tecidos adjacentes às lesões.

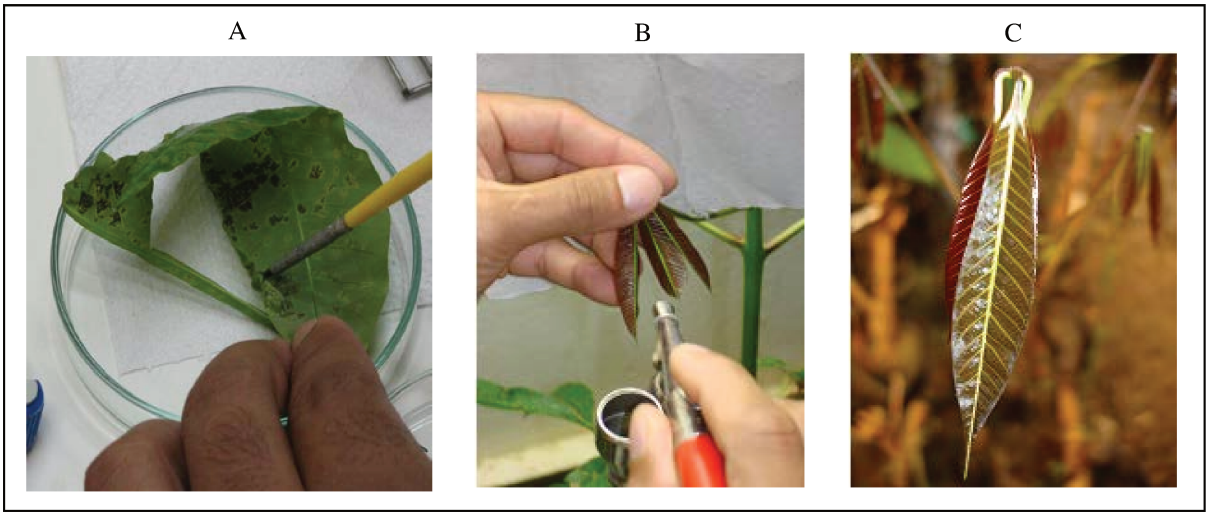


Figura 1 – Coleta dos conídios de *M. ulei* (A) para inoculação na face abaxial dos folíolos (B) em estágio B2 (C).

Nenhum dos clones, com exceção de RRIM 600 (Fig. 2A), apresentou sintomas visíveis da infecção nos primeiros dois dias. O aparecimento de lesões nos folíolos de RRIM 600 era esperado neste período, pois o clone é altamente suscetível ao fungo. No

quinto dia pós-infecção, todos os clones apresentaram algum tipo de lesão (Fig. 2B), sendo os folíolos de RRIM 600 os mais afetados, havendo inclusive deformação dos folíolos (Fig. 2C e D).

A extração de RNA total das amostras foi realizada, entretanto, apesar dos folíolos terem sido transportados em RNA Later (Qiagen), os RNAs totais das seguintes amostras estavam degradados: F 4542-2°dia, Fx 2784-sadia, PA 31-sadia, F 4542-sadia, e todas as amostras do 5° dia, com exceção do clone FDR 5788 (Fig. 3). Para a construção da biblioteca de cDNA foram utilizados os RNAs totais das amostras coletadas no 2° dia de infecção juntamente com o RNA total da amostra FDR 5788-5°dia.

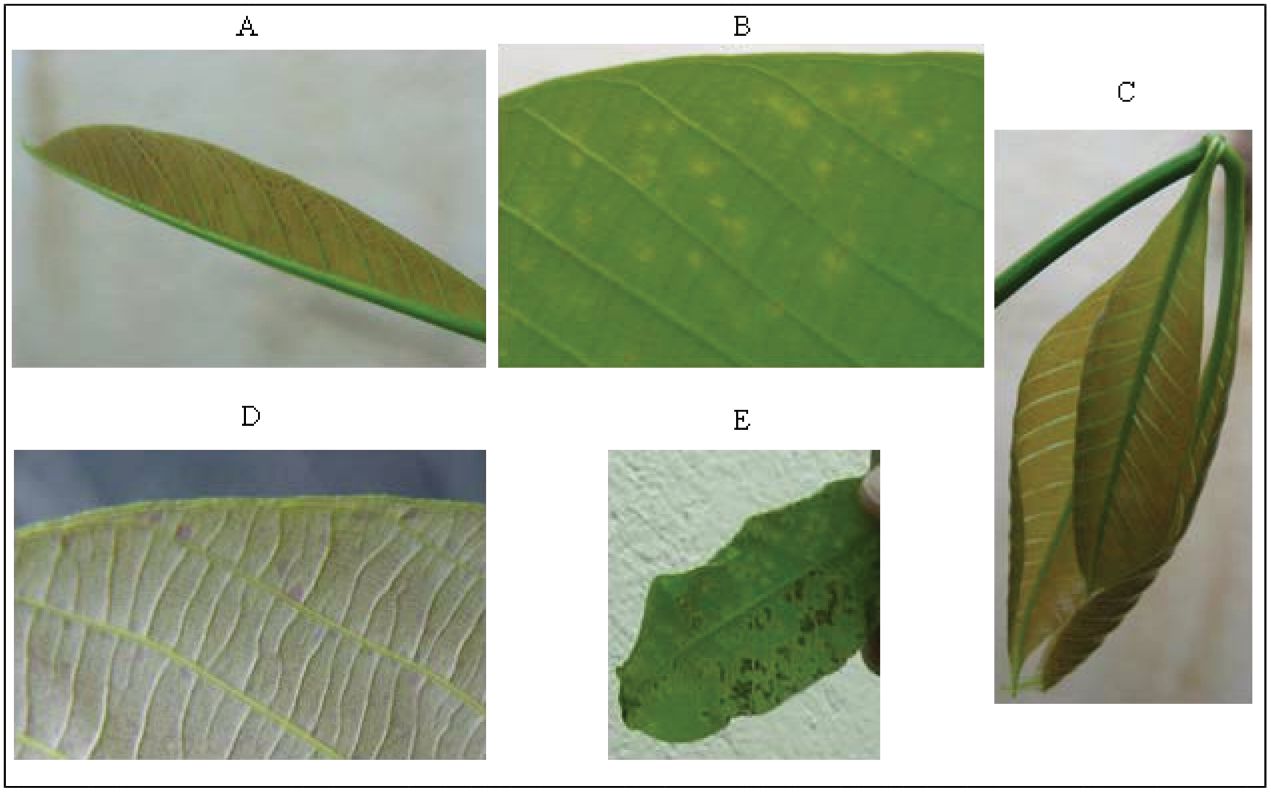


Figura 2 – Fases amostradas do ensaio de infecção controlada por *M. ulei*. (A) Folíolo de RRIM 600, 2 dias após a infecção; (B) Parte de um folíolo de F4542 e (D) RRIM 600, 5 dias após a infecção; (C) Folíolos deformados de RRIM 600, 5 dias após a infecção; (E) Folíolo de RRIM 600, 9 dias após a infecção.

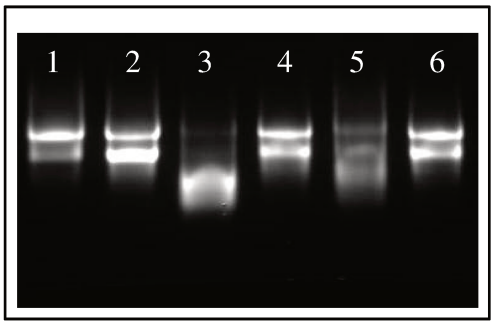


Figura 3 – RNA total extraído dos folíolos jovens do ensaio de infecção controlada. Os RNAs totais das amostras PA 31-5º dia (3) e F4542-2º dia (5) se mostraram degradados. 1 - PFB5 sadia; 2 - PFB5 2ºdia; 4 - F 4542 9ºdia; 6 - FDR5788 5ºdia.

A biblioteca de transcritos de folhas jovens infectadas por *M. ulei* possui aproximadamente 2.500 clones. Foi feito PCR de 15 colônias aleatórias para a verificação do tamanho dos insertos, e estes variaram entre 250 pb e 1.200 pb, com tamanho médio de 584 pb. Foram sequenciados 384 clones desta biblioteca, que geraram 288 *reads* de alta qualidade (75,0%), totalizando 8.551 ESTs. O alinhamento destes *reads* com as sequências das outras bibliotecas gerou 5.211 sequências EST, consistindo em 828 *contigs* e 4.383 *singletons*. Com a adição dos *reads* desta biblioteca, o número de sequências anotadas aumentou para 2.595 e 40 novas sequências não apresentaram similaridade com qualquer proteína ou EST no banco de dados GenBank.

Não houve mudanças com relação às categorias do banco de dados Gene Ontology (GO) mais representadas (função molecular: *binding* - 1.455; 36,6%, *catalytic activity* - 1313; 33,0%, *structural molecule activity* - 202; 5,1%; processo biológico: *cellular metabolic process* - 1163; 41,2%, *biosynthetic process* - 608; 21,6%, *gene expression* - 398; 14,1%; componente celular: *cytoplasm* - 700; 41,3%, *membrane* - 468; 27,6%, *non-membrane-bounded organelle* - 234; 13,8%), o mesmo ocorrendo com as espécies com as quais os ESTs apresentaram maior número de *hits* no BLASTX (*Ricinus communis* - 1.840, *Populus trichocarpa* - 729, *Vitis vinifera* - 251, *H. brasiliensis* - 247). Entre os *singletons*, 16 sequências apresentaram regiões SSR: sete di-, cinco tri-, um tetra- e três pentanucleotídeos.

Cinco ESTs contendo motivos SSR foram selecionados para o desenho de *primers*: três apresentaram similaridade com proteínas e dois apresentaram similaridade com sequências EST do banco de dados GenBank (Tabela 1). A definição da temperatura de *annealing* ótima para cada par foi feita através de PCR gradiente. O polimorfismo dos microssatélites desenvolvidos foi avaliado em 18 genótipos de *H. brasiliensis*, coletados em janeiro de 2008 e abril de 2009 no Centro Experimental de Votuporanga/SP (APTA/IAC), e a transferibilidade foi avaliada em seis espécies do gênero *Hevea* que foram cedidas pela Embrapa Ocidente situada em Manaus/AM (Tabela 2).

TABELA 1 – Motivo das regiões SSR encontradas, tamanho esperado do produto de amplificação e anotação das sequências utilizadas para o desenho dos *primers*.

Primer	**Motivo**	**Tamanho Esperado**	**Anotação**
EHBmu1	$(AAG)_5$	203	*H. brasiliensis* EST (JG009018.1)
EHBmu2	$(TCT)_5$	259	alcohol dehydrogenase, putative [*Ricinus communis*] (8258312 RCOM_1497750)
EHBmu3	$(AAG)_7$	110	*Manihot esculenta* EST (FG805548.1)
EHBmu4	$(AT)_8$	188	ubiquitin-conjugating enzyme [*Hevea brasiliensis*] (AEH05974.1)
EHBmu5	$(CT)_{21}$	267	heat shock factor-binding protein 1-like [*Glycine max*] (XP_003520737.1)

TABELA 2: Relação de genótipos coletados em Votuporanga-SP e Manaus-AM.

Genótipos APTA/IAC		Espécies / Manaus (AM)
GT 1	PB 233	*H. guianensis*
RRIM 701	PB 311	*H. rigifolia*
PB 235	PC 140	*H. nitida*
PB 217	RRIC 100	*H. pauciflora*
PR255	IAC 313	*H. benthamiana*
RRIM 809	IAC 318	*H. camargoana*
RRIM 728	IRCA 27	
RRIM 913	IRCA 209	
RRII 105	IAC 306	

Os cinco pares de *primers* amplificaram bandas definidas com as T_{as} específicas utilizadas, mas o par EHBmu2 possuía *amplicon* de tamanho superior a 400pb, sendo por isso não caracterizado. O par EHBmu1 possui fragmento maior que o esperado (Tabela 3) e foi caracterizado no Advance FS96 dsDNA Fragment Analyzer (Advanced Analytical). Os pares EHBmu3, 4 e 5 apresentaram fragmentos nos tamanhos esperados (Tabela 3) e foram caracterizados em géis de acrilamida 6% corados com prata (Creste *et al.*, 2001). O loco EHBmu3 foi monomórfico entre os genótipos avaliados. Os demais locos apresentaram heterozigosidades esperada (H_e) e observada (H_o) que variaram de 0,06 a 0,76 e 0 a 0,5, respectivamente, e os valores do conteúdo de informação polimórfica (PIC) variaram entre 0,0526 a 0,6912 (Tabela 3).

TABELA 3 – Caracterização dos locos SSR em genótipos de *H. brasiliensis.*

Primer	**Tamanho Esperado**	**Tamanho do Alelo**	**Nº de Alelos**	**H_e**	**H_o**	**PIC**
EHBmu1	203	316-343	5	0,74	0	0,6667
EHBmu3	110	114	1	0	0	0
EHBmu4	188	173-175	2	0,06	0,06	0,0526
EHBmu5	267	255-279	6	0,76	0,50	0,6912

A transferibilidade destes pares de *primers* para outras espécies de *Hevea* também foi avaliada. Todos os quatro marcadores EST-SSR apresentaram amplificação nas seis espécies de *Hevea* utilizadas. Novamente a alta transferibilidade indica que as regiões flanqueadoras aos microssatélites são conservadas entre as espécies de *Hevea* (Saha *et al.*, 2005; Feng *et al.*, 2009; Mantello *et al.*, 2012). Além disso, o gênero *Hevea* é considerado um complexo de espécies, não existindo uma barreira reprodutiva entre elas (Gonçalves e Fontes, 2012), o que explica a alta transferibilidade dos marcadores. Devido à resistência

que estas espécies apresentam ao *M. ulei*, elas são utilizadas em programas de melhoramento que buscam clones de *H. brasiliensis* resistentes ao fungo (Gonçalves e Fontes, 2012). Estes locos, juntamente com os marcadores desenvolvidos apresentados no capítulo III, podem ser usados para acompanhar a introgressão de genes, mapeamento comparativo e para estudos de estrutura populacional e análises genéticas destas espécies.

Para todos os genótipos avaliados, o par EHBmu1 apresentou *amplicon* maior que o esperado. Uma vez que o desenho dos *primers* foi feito usando um transcrito como molde e o *template* utilizado para as amplificações foi o DNA genômico dos acessos de *Hevea*, é possível que a região amplificada a partir do DNA genômico dos genótipos de *Hevea* utilizados possua um íntron, o que explicaria a diferença entre os tamanhos esperado e observado do *amplicon*.

A ocorrência de mal-das-folhas na região amazônica, e que tem se expandido para outras regiões como a Bahia (Gasparotto *et al.*, 1990; Mattos *et al.*, 2003), é a principal razão da necessidade de áreas de escape para o plantio de seringueira. Um melhor entendimento da doença, e de como selecionar clones mais tolerantes a ela, pode ser conseguido com análises das respostas moleculares da interação entre planta e patógeno. Uma análise completa da biblioteca de cDNA de tecidos infectados construída neste trabalho pode ajudar na obtenção de respostas sobre os fatores que garantem maior tolerância ao fungo, como a identificação de sequências que podem estar relacionadas com a tolerância ou a susceptibilidade da seringueira ao *M. ulei*.

No capítulo III também foram apresentados os resultados das análises de expressão por PCR em tempo real das sequências mais abundantes nas bibliotecas de cDNA dos experimentos de exposição à baixas temperaturas. A seguir apresentamos resultados complementares a esta parte do trabalho.

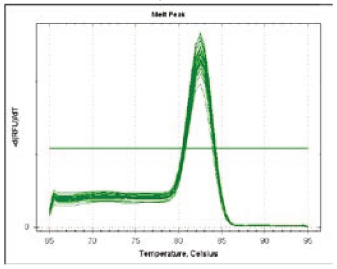
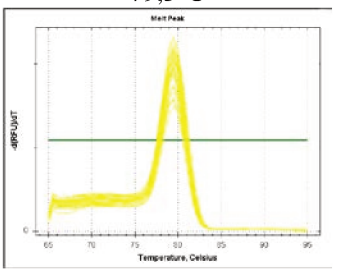
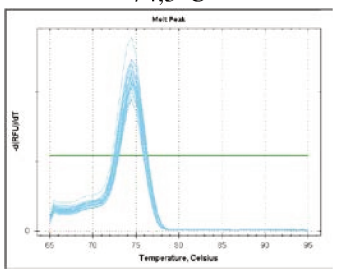
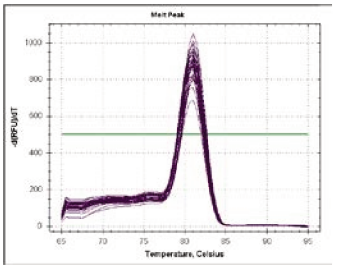
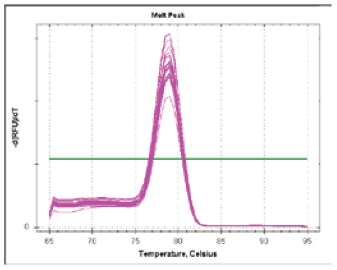
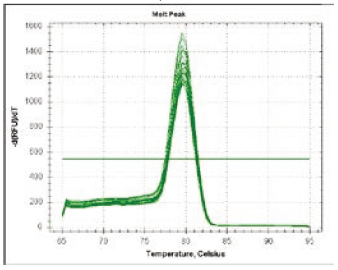
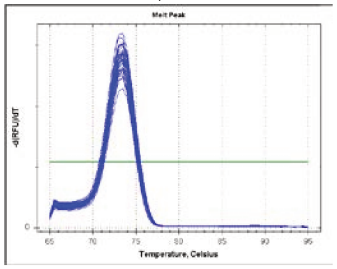
A expressão das sequências mais abundantes, que apresentaram similaridade com as proteínas NAD(P)H-quinone oxidoreductase subunit H (NADH), chloroplast photosystem II 10 kDa polypeptide (PsbR), uma proteína hipotética (HYPOT), ATP synthase CF0 C subunit (CF0) e indole-3-acetic acid-induced protein (ARG2-1 e ARG2-2), foi avaliada por quantificação relativa com o uso dos genes *glyceraldehyde-3-phosphate dehydrogenase* (*GAPDH*) e *eukaryotic translation initiation factor* (*eIF2*) como genes referência para a normalização das análises. A escolha pela combinação destes genes como referência foi feita com base em resultados obtidos através da análise de oito genes candidatos: *eIF2*, *YSL8*, *UBC2a*, *UBC4*, *18SrRNA*, *RH2b*, *RH8* (Li *et al.*, 2011) e *GAPDH*.

A expressão dos genes candidatos foi avaliada nas amostras de RNA total combinado e nas amostras de RNA total individuais. A estabilidade da expressão de cada gene foi analisada em duplas pelo software CFX Manager 2.1 (Bio-Rad Laboratories Inc., USA), e os genes *GAPDH* e *eIF2* apresentaram os menores valores médios de coeficiente de variância (CV) e valor M (*gene stability value*), tanto para as amostras combinadas (CV = 0,0477; M = 0,1376) quanto para as amostras individuais (CV = 0,1515; M = 0,4394) (Hellemans *et al.*, 2007). A estabilidade dos genes de referência depende das condições experimentais em que estes estão sendo utilizados (i.e.: genótipos diferentes) (Hellemans *et al.*, 2007; Li *et al.*, 2011); desta maneira, a diferença observada entre os valores de CV e M se deve às diferenças entre as amostras: as amostras de RNA total combinado dos genótipos resultou em uma condição mais homogênea que a das amostras individuais.

Para avaliar a eficiência e a especificidade dos pares de *primers* desenhados para as sequências, análises de curva padrão e de curva de *melting* foram realizadas. Todos os pares apresentaram eficiência de amplificação entre 90 e 110%, a grande maioria apresentou coeficiente de determinação (R^2) igual ou superior a 0,99 e todos apresentaram apenas um

pico nas análises de *melting* (http://www.bio-rad.com/webroot/web/pdf/lsr/literature/Bulletin_5859.pdf) (Tabela 5).

TABELA 5 – Eficiência e especificidade dos pares de *primers* desenhados para as análises de expressão.

Primer	R^2/ Efic.	Temperatura de *Melting*	*Primer*	R^2/ Efic.	Temperatura de *Melting*
CF0	0,99/ 102,6%	82,5°C 	ARG2-1	0,997/ 95,5%	79,5°C 
HYPOT	0,99/ 103,5%	74,5°C 	ARG2-2	0,992/ 102,4%	81°C 
PsbR	0,992/ 95,4%	79°C 	GAPDH	0,983/ 106,5%	79,5°C 
NADH	0,99/ 108,6%	73,5°C 			

CF0 - ATP synthase CF0 C subunit, HYPOT – proteína hipotética, PsbR - chloroplast photosystem II 10 kDa polypeptide, NADH - NAD(P)H-quinone oxidoreductase subunit H, ARG2-1 e 2 - indole-3-acetic acid-induced proteins. R^2 a 60°C. Efic. = Eficiência.

Os resultados obtidos e descritos nos capítulo III sugerem que os clones de seringueira apresentam respostas diferentes quando submetidos ao mesmo tipo de estresse. Essas sequências podem agora ser também avaliadas em outros clones, e abrem a possiblidade para a avaliação de outros genes que estejam envolvidos nos mesmos processos biológicos.

No capítulo III também foram descritos dois pares de primer EST-SSR (EHBc103 e EHBp27) que amplificaram alelos duplicados em vários genótipos utilizados na caracterização dos marcadores. Mantello *et al.* (2012) também relataram a presença de alelos duplicados em seringueira. O par de *primer* EHBp27, em especial, apresentou um padrão de alelos similar ao de plantas poliploides, com genótipos contendo até seis marcas (Fig. 4). Outro par de *primer* que foi desenhado a partir de um EST que não se alinhou com a sequência molde de EHBp27 (identidade ≥ 80%), e no qual foi inserido a cauda M13, apresentou o mesmo tipo de amplificação e os mesmos alelos. Uma vez que *H. brasiliensis* é uma planta diploide e se comporta como um diploide (Bouharmont 1960; Majumder 1964; Ong 1975), provavelmente três diferentes locos foram amplificados por EHBp27. Foi possível o mapeamento de dois locos, que foram posicionados no mesmo grupo de ligação e com apenas 0,8cM de distância entre eles (Cap. IV). As outras espécies de *Hevea* também apresentaram padrão poliploide para este *primer*. O par de *primer* EHBp23 também amplificou mais de um loco, sendo que, neste caso, eles eram bastante distintos (Fig. 5). Ambos os locos foram analisados na população de mapeamento (Fig. 6) e foram posicionados em grupos de ligação diferentes.

No capítulo IV, foi apresentado o novo mapa genético-molecular para seringueira com 330 locos mapeados. A população de mapeamento foi analisada com 54 novos marcadores microssatélites, entretanto oito locos foram descartados sendo mapeados 46

novos locos. Entre os locos retirados do mapeamento, cinco causaram deformações nos grupos de ligação em que foram posicionados. Os três locos restantes não puderam ser avaliados devido às limitações do equipamento Advance FS96 dsDNA Fragment Analyzer (Advanced Analytical Technologies Inc., USA) que em vários indivíduos apresentava três picos de leitura.

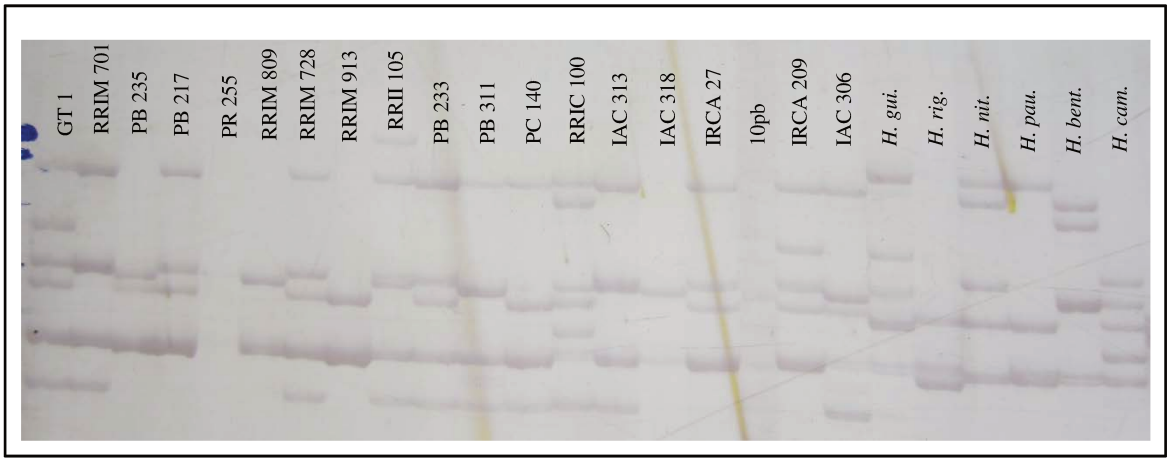


Figura 4 – Caracterização do EST-SSR EHBp27 em 18 genótipos de *H. brasiliensis* e avaliação da transferibilidade deste marcador para outras espécies de *Hevea*.

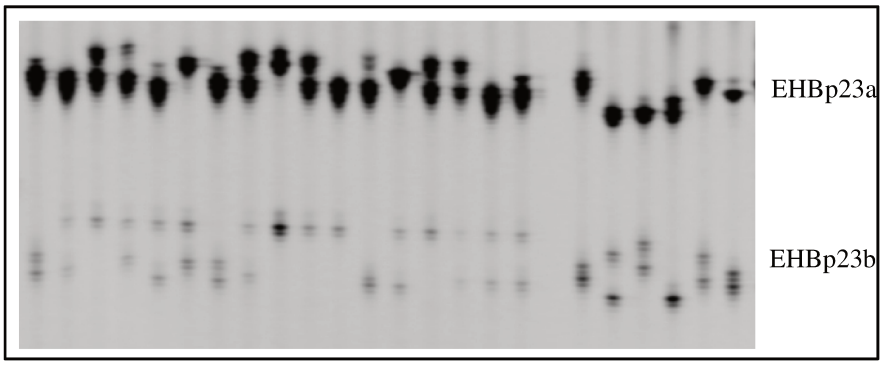


Figura 5 – Caracterização do EST-SSR EHBp23 em 18 genótipos de *H. brasiliensis* e avaliação da transferibilidade deste marcador para outras espécies de *Hevea*. Obs: mesma ordem da fig. 4.

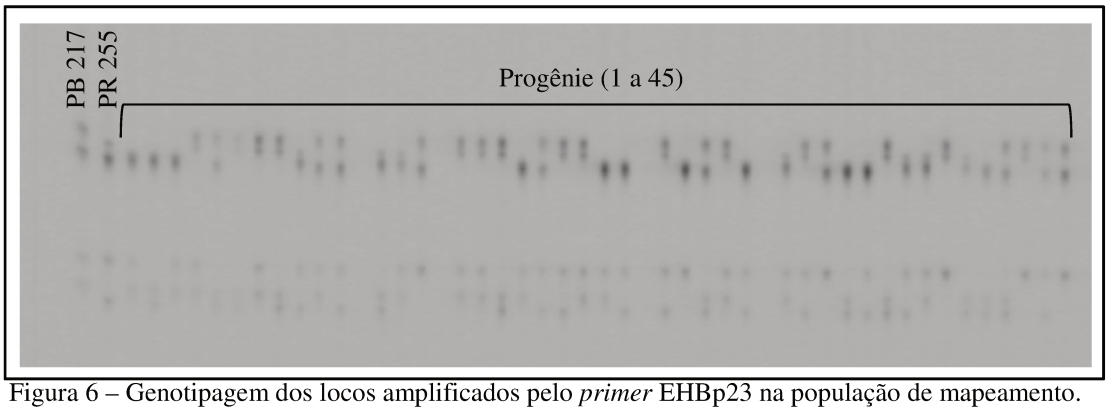


Figura 6 – Genotipagem dos locos amplificados pelo *primer* EHBp23 na população de mapeamento.

Capítulo VI

Conclusões

- Apesar das bibliotecas de cDNA gerarem um número menor de sequências, quando comparadas a tecnologias *Next Generation Sequencing* (NGS), elas ainda são uma fonte rica em marcadores moleculares e podem revelar a existência de transcritos ainda não conhecidos.
- A transferibilidade dos marcadores EST-SSR para outras espécies de *Hevea* atingiu quase 100%. Estes marcadores podem ser utilizados para avaliar a variabilidade genética e investigar a relação de parentesco entre as diferentes espécies. A alta transferibilidade sugere que o genoma das diferentes espécies ainda é extremamente parecido, o que explicaria o intercruzamento e a alta frequência de híbridos naturais. Estas espécies são importantes para o melhoramento genético de *H. brasiliensis*, e a introgressão de genes pode ser acompanhada através destes marcadores.
- Utilizando apenas marcadores microssatélites, foi possível resolver um dos grupos que estava dividido (LG11) e três regiões com distâncias maiores que 40cM. Não descartando a necessidade de outros tipos de marcadores, é possível que a inclusão de novos marcadores SSR consiga resolver os *gaps* restantes e unir os grupos que ainda estão separados.
- Os clones utilizados possuem diferentes repostas ao estresse por baixas temperaturas, com uma maior diferença para os genes ARG2, o que pode estar relacionado com a melhor tolerância às baixas temperaturas dos clones PR 255, GT 1 e IAN 873 quando comparados com o clone PB 217.

- Os locos mapeados dentro de regiões QTL podem estar envolvidos com as características de crescimento analisadas, ou então ligados a genes que efetivamente influenciam o fenótipo.

Perspectivas

Nos últimos anos, a pesquisa em seringueira avançou consideravelmente, principalmente na área da biologia molecular. Entretanto, há muito ainda para ser feito com relação ao mapeamento de características agronomicamente importantes e para o entendimento das diferentes interações genéticas e moleculares da seringueira em resposta a diferentes situações ambientais e em relação ao seu próprio genótipo. Em vista disto, duas populações F_1 de *H. brasiliensis* foram desenvolvidas em colaboração com o Programa Seringueira do IAC, em que o genitor materno (GT 1) é resistente ao frio e os dois genitores paternos (PB 235 e RRIM 701) apresentam alto potencial produtivo. As populações já estão sendo avaliadas quanto a diferentes características fenotípicas e também já estão sendo genotipadas com os marcadores desenvolvidos neste trabalho e outros marcadores desenvolvidos em nosso laboratório. O mapeamento e a fenotipagem levarão à identificação de novos QTLs envolvidos em importantes caracteres agronômicos em seringueira.

A população de mapeamento utilizada neste trabalho se encontra na cidade Itiquira, na antiga Plantações Edouard Michelin de Mato Grosso. No momento, análises fenotípicas com relação à produção estão sendo feitas na população, como também a genotipagem de SNPs por espectrometria de massa (Sequenom). Entretanto, em 2009, 80% da área explorada para plantação e pesquisa da seringueira foi vendida ao Grupo Maggi. Até o momento esta população é mantida através de um acordo entre o Grupo Maggi e a Michelin, por um período de alguns anos. Desta maneira, apesar da população ainda oferecer campo para maiores pesquisas e apresentar variabilidade para entrar em processo de seleção, ela ainda está ameaçada.

Outras linhas de pesquisa em seringueira também estão sendo desenvolvidas em nosso laboratório. Trabalhos de sequenciamento e análise do trancriptoma de seringueira, tanto de *H. brasiliensis* quanto de outras espécies de *Hevea*; desenvolvimento de novos marcadores SNP, tanto genômicos quanto expressos; genotipagem, análise de diversidade genética e caracterização do desequilíbrio de ligação de germoplasmas *ex situ* presentes em diferentes centros de pesquisa; análises de expressão diferencial, por RNA-seq, de diferentes clones (mudas) de *H. brasiliensis* submetidos a baixas temperaturas.

Ao contrário da grande maioria das culturas, a seringueira é uma planta que ainda está sendo domesticada e, talvez por isto, trabalhos mais complexos sobre a estrutura do seu genoma e como este é expresso começaram a ser feitos relativamente há pouco tempo. Este cenário demonstra que muito ainda precisa ser feito para entender a seringueira, e que é possível que o trabalho nunca se esgote.

Referências

Referências

Andersen JR, Lübberstedt T (2003). Functional markers in plants. Trends Plant Sci. 8(11):554-560.

Atan S, Low FC, Saleh NM (1996). Construction of a microsatellite enriched library from *Hevea brasiliensis*. J Nat Rubb Res 11: 247-255.

Azmach G, Gedil M, Menkir A, Spillane C (2013). Marker-trait association analysis of functional gene markers for provitamin A levels across diverse tropical yellow maize inbred lines. BMC Plant Biol 13:227. doi: 10.1186/1471-2229-13-227.

Besse P, Seguin M, Lebrun P, Lanaud C (1993a). Ribosomal DNA variations in wild and cultivated rubber tree (*Hevea brasiliensis*). Genome 36(6):1049-57.

Besse P, Lebrun P, Seguin M, Lanaud C (1993b). DNA fingerprints in *Hevea brasiliensis* (rubber tree) using human minissatelite probes. Heredity 70:237-244.

Borém A, Caixeta ET (2006). Marcadores moleculares. Viçosa, MG, 374p.

Bouharmont J (1960). Recherches taxonomiques et caryologiques chez quelques espèces du genre *Hevea*. In: INEAC (ed) Série Scientifique, 85th edition, Pp 64.

Brookes AJ (1999). The essence of SNPs. Gene 234:177-186.

Carneiro MS, Vieira MLC (2002). Mapas genéticos em plantas. Bragantia 61:89-100.

Carninci P (2007). Constructing the landscape of the mammalian transcriptome. J Exp Biol 210:1497-1506.

Cato SA, Gardner RC, Kent J, Richardson TE (2001). A rapid PCR-based method for genetically mapping ESTs. Theor Appl Genet 102:296-306.

Clepet C, Joobeur T, Zheng Y, Jublot D, Huang M, Truniger V, Boualem A, Hernandez-Gonzalez ME, Dolcet-Sanjuan R, Portnoy V, Mascarell-Creus A, Caño-Delgado AI, Katzir N, Bendahmane A, Giovannoni JJ, Aranda MA, Garcia-Mas J, Fei Z (2011) Analysis of expressed sequence tags generated from full-length enriched cDNA libraries of melon. BMC Genomics 12:252. doi:10.1186/1471-2164-12-252

Coelho ASG, Silva HD (2002). Construção de mapas genéticos e mapeamento de QTLs. ESALQ, Piracicaba, 77p. (apostila).

Cortez JV (2005). Reflexões sobre as projeções de consumo de borracha natural no Brasil até 2030. Disponível em http://www.apabor.org.br/sitio/artigos/html/20050824-1.htm

Creste S, Tulmann A, Figueira A (2001). Detection of single sequence repeat polymorphism in denaturating polyacrylamide sequencing gels by silver staining. Plant Mol Biol Rep 19:299-306.

Cruz CD, Silva LC (2009). Análise de marcadores moleculares. In: Bórem A, Caixeta ET (Eds). Marcadores moleculares.Viçosa, Brasil, 2:371-442.

Cubry P, Pujade-Renaud V, Garcia D, Espeout S, Le Guen V, Granet F, Seguin M (2014). Development and characterization of a new set of 164 polymorphic EST-SSR markers for diversity and breeding studies in rubber tree (*Hevea brasiliensis* Müll. Arg.). Plant Breed doi: 10.1111/pbr.12158.

Falconer DS, Mackay TFC (1996). Introduction to Quantitative Genetics. 4.ed. London: Longman Scientific Technical, 464p.

Feng SP, Li WG, Huang HS, Wang JY, Wu YT (2009). Development, characterization and cross-species/genera transferability of EST-SSR markers for rubber tree (*Hevea brasiliensis*). Mol Breeding 23(1):85-97.

Garcia AAF, Kido EA, Meza AN, Souza HMB, Pinto LR, Pastina MM, Leite CS, Silva JAG Da, Ulian EC, Figueira AVO, Souza AP (2006). Development of an integrated genetic map of a sugarcane (*Saccharum* spp.) commercial cross, based on a maximum-likelihood approach for estimation of linkage and linkage phases. Theor Appl Genet 112:298-314.

Gasparotto L, Ferreira FA, Lima MIPM, Pereira JCR, Santos AF (1990). Enfermidades da Seringueira no Brasil. In: Série Embrapa Circular Técnica 3. Embrapa-CPAA, Manaus, Brasil, pp 12-44.

Gaur R, Azam S, Jeena G, Khan AW, Choudhary S, Jain M, Yadav G, Tyagi AK, Chattopadhyay D, Bhatia S (2012) High-throughput SNP discovery and genotyping for constructing a saturated linkage map of chickpea (*Cicer arietinum* L.). DNA Res 19:357–373.

Gonçalves P, Cardoso M, Ortolani AA (1990) Origem, variabilidade e domesticação da *Hevea*; uma revisão. Pesq Agrop Bras 25(2):135-156.

Gonçalves PS, Ortolani AA, Cardoso M (1997). Melhoramento genético da seringueira: uma revisão, 55p. (Documentos IAC 54).

Gonçalves PS, Fontes JRA (2012). Domestication and breeding of rubber tree. In: Borem A, Lopes MTG, Clement CR, Noda H (ed) Domestication and Breeding: Amazonian Species. UFV, Viçosa, Brasil, pp 393-420.

Grattapaglia D, Sederoff R (1994). Genetic linkage maps of Eucalyptus grandis and E. urophylla using a pseudo-testcross mapping strategy and RAPD markers. Genetics 137: 1121-1137.

Harbers M (2008). The current status of cDNA cloning. Genomics 91:232-242.

Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J (2007). qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. Genome Biol 8(2):R19. doi:10.1186/gb-2007-8-2-r19.

Instituto de Pesquisa e Estudos Florestais (IPEF) (2007). *Hevea brasiliensis* (seringueira). Disponível em http://www.ipef.br/identificacao/hevea.brasiliensis.asp.

International Rubber Study Group (IRSG) (2013). Rubber statistical bulletin, v.67 (10/12).

Kapranov P, Willingham AT, Gingeras TR (2007). Genome-wide transcription and the implications for genome organization. Nat Rev Genet 8:413-423.

Kato S, Sekine S, Oh SW, Kim NS, Umezawa Y, Abe N, Yokoyama-Kobayashi M, Aoki T (1994). Construction of a human full-length cDNA bank. Gene 150:243-250.

Lam LV, Thanh T, Trang LTT, Truong VV, Lam HB, Tuy LM (2012). *Hevea* Germplasm in Vietnam: Conservation, Characterization, Evaluation and Utilization. In: Caliskan M (ed) Genetic Diversity in Plants. ISBN: 978-953-51-0185-7, InTech, disponível em: http://www.intechopen.com/books/genetic-diversity-in-plants/hevea-germplasm-in-vietnam-conservation-characterization-evaluation-and-utilization

Le Guen M, Lespinasse D, Oliver G, Rodier-Goud M, Pinard F, Seguin M (2003). Molecular maping of genes confering field resistence to South American Leaf Blight (*Microcyclus ulei*) in rubber tree. Theor Appl genet 108:160-167.

Le Guen V, Gay C, Xiong TC, Souza LM, Rodier-Goud M, Seguin M (2010). Development and characterization of 296 new polymorphic microsatellite markers for rubber tree (*Hevea brasiliensis*). Plant Breed 130:294–296.

Le Guen, V., Garcia, D., Doare, F., Mattos CR, Condina V, Couturier C, Chambon A, Weber C, Espéout S, Seguin M (2011). A rubber tree's durable resistance to *Microcyclus ulei* is conferred by a qualitative gene and a major quantitative resistance factor. Tree Genet Genomes doi:10.1007/s11295-11011-11038-11297.

Leitch AR, Lim KY, Leitch IJ, O'Neill M, Chye M, Low F (1998). Molecular cytogenetic studies in rubber, *Hevea brasiliensis* Muell. Arg. (Euphorbiaceae). Genome 41:464–467.

Lespinasse D, Rodier-Godier M, Grivet L, Leconte A, Legnate H, Seguin M (2000a). A saturated genetic linkage map of rubber tree (*Hevea* spp.) based on RFLP, AFLP, microsatellite, and isozyme markers. Theor Appl Genet 100:127-138.

Lespinasse D, Grivet L, Troispoux V, Rodier-Gould M, Pinard F, Segun M (2000b). Identification of QTLs involved in the resistence to South American Leaf Blight (*Microcyclos ulei*) in the rubber tree. Theor Appl Genet 100:975-984.

Li D, Deng Z, Qin B, Liu X, Men Z (2012). De novo assembly and characterization of bark transcriptome using Illumina sequencing and development of EST-SSR markers in rubber tree (*Hevea brasiliensis* Muell. Arg.). BMC Genomics 13:192. doi: 10.1186/1471-2164-13-192.

Li H, Qin Y, Xiao X, Tang C (2011). Screening of valid reference genes for real-time RT-PCR data normalization in Hevea brasiliensis and expression validation of a sucrose transporter gene HbSUT3. Plant Sci 181:132-139.

Lieberei R (2007). South American Leaf Blight of the rubber tree (*Hevea* spp.): new steps in plant domestication using physiological features and molecular markers. Ann Bot 100:1125–1142.

Lin C, Mueller LA, McCarthy J, Crouzillat D, Pétiard V, Tanskley SD (2005). Coffee and tomato share common gene repertoires as revealed by deep sequencing of seed and cherry transcripts. Theor Appl Genet 112:114-130.

Lin M, Lai D, Pang C, Fan S, Song M, Yu S (2013). Generation and analysis of a large-scale expressed sequence Tag database from a full-length enriched cDNA library of developing leaves of *Gossypium hirsutum* L. PLoS One 8(10):e76443. doi: 10.1371/journal.pone.0076443

Liu Y, He Z, Appels R, Xia X (2012). Functional markers in wheat: current status and future prospects. Theor Appl Genet 125:1–10.

Low FC, Atan S, Jaafar H, Tan H (1996). Recent advances the development of molecular markers for *Hevea* studies. J Nat Rubber Res 11:32-44.

Luo H, Boutry M (1995). Phylogenetic relationships within *Hevea brasiliensis* as deduced from a polymorphic mitocondrial DNA region. Theor Appl Genet 91:876-884.

Lynch M, Walsh B (1998). Genetics and analysis of quantitative traits. Sinauer Associates, Sunderland, MA. 980p.

Majumder SK (1964). Chromosome studies of some species of *Hevea*. J. Rubber Res. Inst. Malays 18:269–273.

Mantello CC, Suzuki FI, Souza LM, Gonçalves PS, Souza AP (2012). Microsatellite marker development for the rubber tree (*Hevea brasiliensis*): characterization and cross-amplification in wild *Hevea* species. BMC Res Notes. 5:329. doi: 10.1186/1756-0500-5-329.

Mantello CC, Cardoso-Silva CB, Da Silva CC, Souza LM, Scallopi Jr EJ, Gonçalves PS, Vicentini R, Souza AP (2014). De novo assembly and transcriptome analysis of the rubber tree (*Hevea brasiliensis*) and SNP markers development for rubber biosynthesis pathways. PLoS One *in press*.

Maruyama K, Sugano S (1994). Oligo-capping: a simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides. Gene 138:171-174.

Mattos CRR, Garcia D, Pinard F, Le Guen V (2003). Variabilidade de isolados de *Microcyclus ulei* no sudeste da Bahia. Fitopatologia Brasileira 28:502-507.

Mooibroek H, Cornish K (2000). Alternative sources of natural rubber. Appl Microbiol Biotechnol 53:335-365.

Novaes E, Drost DR, Farmerie WG, Pappas GJ Jr, Grattapaglia D, Sederoff RR, Kirst M (2008). High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome. BMC Genomics 9:312. doi: 10.1186/1471-2164-9-312.

Oliveira KM, Pinto LR, Marconi TG, Margarido GRA, Pastina MM, Teixeira LHM, Figueira AV, Ulian EC, Garcia AAF, Souza AP (2007). Functional integrated genetic linkage map based on EST-markers for a sugarcane (*Saccharum* spp.) commercial cross. Mol Breed 21:1-20.

Oliveira EJ, Vieira MLC, Garcia AAF, Munhoz CF, Margarido GRA, Consoli L, Matta FP, Moraes MC, Zucchi MI, Fungaro MHP (2008). An integrated molecular map of yellow passion fruit based on simultaneous maximum-likehood estimation of linkage and linkage phases. J Am Soc Hortic Sci 133:35-44.

Paiva JR (1992). Variabilidade enzimática em populações naturais de seringueira. 145p. Tese (Ph.D) - ESALQ-USP, Piracicaba.

Palhares AC, Rodrigues-Morais TB, Van Sluys MA, Domingues DS, Maccheroni W Jr, Jordão H Jr, Souza AP, Marconi TG, Mollinari M, Gazaffi R, Garcia AA, Vieira ML (2012). A novel linkage map of sugarcane with evidence for clustering of retrotransposon-based markers. BMC Genet 13:51. doi: 10.1186/1471-2156-13-51.

Paterson AH, Lander ES, Hewitt JD, Peterson S, Lincoln SE, Tanksley SD (1981). Resolution of quantitative traits into mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 335:721-726.

Pires JM, Secco RS, Gomes JI (2002). Taxonomia e fitogeografia das seringueiras *Hevea* spp. Embrapa amazômia oriental, Belém, Brazil

Pootakham W, Chanprasert J, Jomchai N, Sangsrakru D, Yoocha T, Therawattanasuk K, Tangphatsornruang S (2011). Single nucleotide polymorphism marker development in the rubber tree, *Hevea brasiliensis* (Euphorbiaceae). Am J Botany e337-e338. doi: 10.3732/ajb.1100228

Powell W, Machray GC, Provan J (1996). Polymorphism revealed by simple sequence repeats. Trends Plant Sci 1:215-222.

Priyadarshan PM, Gonçalves PS (2003). *Hevea* gene pool for breeding. Genet Resour Crop Evol 50:101-114.

Priyadarshan PM, Clement-Demange A (2004). Breeding *Hevea* Rubber: formal and molecular genetics. Adv Genet 52:51-105.

Pushparajah E (1983). Problems and potentials for establishing *Hevea* under difficult environmental conditions. The Planter 59:242-251.

Pushparajah E (2001). Natural rubber, p. 379-407. In: Last, F.T. (Ed.) Tree Crop Ecosystems (Vol 19: Ecosystems of the World Series). Elsevier Science, Amsterdam, The Netherlands.

Rafalski JA, Tingey SV (1993). Genetic diagnostics in plant breeding: RAPDs, microsatellites and machines. Trends Genet 9:275-280.

Rafalski A (2002). Applications of single nucleotide polymorphisms in crop genetics. Curr Opinion Plant Biol 5:94-100.

Ritter E, Gebhardt C, Salamini F (1990). Estimation of recombination frequencies and construction of RFLP linkage maps in plants. Genetics 125:645-654.

Roy CB, Nazeer MA, Saha T (2004). Identification of simple sequence repeats in rubber (*Hevea* brasiliensis). Curr Sci (Bangalore) 87:807-11.

Saha T, Bindu Roy C, Nazeer MA (2005). Microsatellite variability and its use in the characterization of cultivated clones of *Hevea brasiliensis*. Plant Breed 124:86-92.

Salgado LR, Koop DM, Pinheiro DG, Rivallan R, Le Guen V, Nicolás MF, de Almeida LG, Rocha VR, Magalhães M, Gerber AL *et al.* (2014). De novo transcriptome analysis of Hevea brasiliensis tissues by RNA-seq and screening for molecular markers. BMC Genomics 15(1):236. doi: 10.1186/1471-2164-15-236.

Schlötterer C (2000). Evolutionary dynamics of microsatellite DNA. Cromosoma 109:365-371.

Silva CC, Mantello CC, Campos T, Souza LM, Gonçalves PS, Souza AP (2014). Leaf-, panel- and latex-expressed sequenced tags from the rubber tree (*Hevea brasiliensis*) under cold-stressed and suboptimal growing conditions: the development of gene-targeted functional markers for stress response. Mol Breed DOI 10.1007/s11032-014-0095-2.

Souza LM, Mantello CC, Santos MO, Gonçalves PO, Souza AP (2009) Microsatellites from rubber tree (*Hevea brasiliensis*) for genetic diversity analysis and cross amplification in six *Hevea* wild species. Cons Genet Res 1:75–79.

Souza LM, Gazaffi R, Mantello CC, Silva CC, Garcia D, Le Guen V, Cardoso SEA, Garcia AAF, Souza AP (2013) QTL mapping of growth-related traits in a full-sib family of rubber tree (*Hevea brasiliensis*) evaluated in a sub-tropical climate. PLoS One 8(4):e61238. doi/10.1371/journal.pone.0061238.

Sterky F, Lundeberg J (2000). Sequence analysis of genes and genomes. J Biotechnol 76:1-31.

Sturtevant AH (1913). The linear arrangement of six sex-linked factors in Drosophila, as shown by their mode of association. J Exper Zool 14:43-59.

Ong SN (1975). Chromosome morphology at pachytene stage in *Hevea brasiliensis*: a preliminary report. In: Sripathi B (ed) Proceedings of the International Rubber Conference. Rubber Research Institute of Malaysia, Kuala Lumpur, pp 3–12.

Tang J, Wang F, Wang Z, Huang Z, Xiong A, Hou X (2013). Characterization and co-expression analysis of WRKY orthologs involved in responses to multiple abiotic stresses in Pak-choi (*Brassica campestris* ssp. chinensis). BMC Plant Biol 13:188. doi: 10.1186/1471-2229-13-188.

Tanksley SD (1993). Mapping polygenes. Ann Rev Genet 27:205-233.

Tobias CM, Twigg P, Hayden DM, Vogel KP, Mitchell RM, Lazo GR, Chow EK, Sarath G (2005). Analysis of expressed sequence tags and the identification of associated short tandem repeats in switchgrass Theor Appl Genet 111:956-964.

Tost J, Gut IG (2002). Genotyping single nucleotide polymorphisms by mass spectrometry. Mass Spectrometry Rev 21:388-418.

Triwitayakorn K, Chatkulkawin P, Kanjanawattanawong S, Sraphet S, Yoocha T, Sangsrakru D, Chanprasert J, Ngamphiw C, Jomchai N, Therawattanasuk K, Tangphatsornruang S (2011). Transcriptome sequencing of *Hevea brasiliensis* for development of microsatellite markers and construction of a genetic linkage map. DNA Res 18(6):471–482.

Varshney RK, Thiel T, Stein N, Langridge P, Graner A (2002). In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species. Cell Mol Biol Lett 7(2A):537-546.

Varshney RK, Graner A, Sorrells ME (2005a). Genomics-assisted breeding for crop improvement. Trends Plant Sci 10(12):621-630.

Varshney RK, Graner A, Sorrells ME (2005b). Genic microsatellite markers in plants: features and applications. Trends Biotechnol 23(1):48-55.

Varshney RK, Sigmund R, Börner A, Korzun V, Stein N, Sorrells ME, Langridge P, Graner A (2005c). Interspecific transferability and comparative mapping of barley EST-SSR markers in wheat, rye and rice. Plant Sci 168(1):195-202.

Venkatachalam P, Sailasree R, Priya P, Saraswathyamma CK, Thulaseedharan A (2001). Identification of a DNA marker associated with dwarf trait in *Hevea brasiliensis* (Muell) Arg. through randon amplified polymorphic DNA analysis. In: Sainte-Beuve J (ed) Proc Annu IRRDB Mtg. CIRAD, Montpellier, france, pp 73-81.

Vettore AL, da Silva FR, Kemper EL, Arruda P (2001). The libraries that made SUCEST. Genet Mol Biol 24(1-4):1-7.

Webster CC, Paardekooper EC (1989). Botany of the rubber tree, p. 57-84. In: Webster, CC; Baulkwill WJ (Eds). Rubber. Longman Scientific and Technical, Essex, England, UK.

Wu R, Ma CX, Painter I, Zeng ZB (2002). Simultaneous maximum likelihood estimation of linkage and linkage phases in outcrossing species. Theor Pop Biol 61:349-363.

Wu SB, Wirthensohn MG, Hunt P, Gibson JP, Sedgley M (2008). High resolution melting analysis of almond SNPs derived from ESTs. Theor Appl Genet 118(1):1-14.

Xia Z, Xu H, Zhai J, Li D, Luo H, He C, Huang X (2011). RNA-Seq analysis and de novo transcriptome assembly of *Hevea brasiliensis*. Plant Mol Biol 77:299-308.

Xia Y, Ning Z, Bai G, Li R, Yan G, Siddique KHM, Baum M, Guo P (2012). Allelic variations of a light harvesting chlorophyll A/B-binding protein gene (Lhcb1) associated with agronomic traits in barley. PLos One 7(5):e37573. doi: 10.1371/journal.pone.0037573.

Wycherley PR (1977). The genus *Hevea*. In: Workshop on international collaboration *Hevea* breeding and the collection and establishment of materials from neotropic, Kuala Lumpur, 12p.

Yu F, Wang BH, Feng SP, Wang JY, Li WG, Wu YT (2011). Development, characterization, and cross-species/genera transferability of SSR markers for rubber tree (*Hevea brasiliensis*). Plant Cell Rep 30:335-344.

Zhang D, Zhang Z, Yang K, Li B (2004). Genetic mapping in (*Populus tomentosa* x *Populus bolleana*) and *P. tomentosa* Carr. using AFLP markers. Theor Appl Genet 108(4):657-62.

Anexos

Anexo A – Documento de aprovação do CIBio

Cidade Universitária "Zeferino Vaz",
5 de maio de 2014.

CIBio: 03/2014

Identificação:
Doutorado: Carla Cristina da Silva, CPG-GBM UNICAMP
Projeto: Construção de um mapa funcional em seringueira (*Hevea brasiliensis*)

Parecer:
Projeto aprovado pela CIBio/CBMEG em 06/03/2009 sob número 01/2009
(em andamento)
Coordenador: Profa. Dra. Anete Pereira de Souza

Profa. Dra. Edi Lúcia Sartorato
Presidente da CIBio/CBMEG - UNICAMP