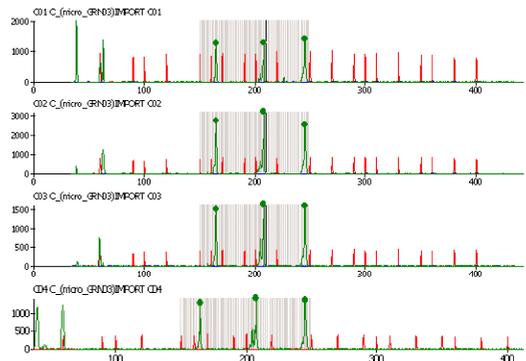
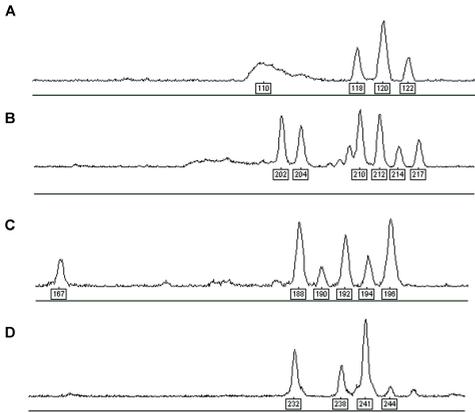


## Microssatélites e o Guaranazeiro: Diversidade Submersa em Um Mar de Alelos





ISSN 1517-2457

Dezembro, 2013

*Empresa Brasileira de Pesquisa Agropecuária  
Embrapa Amazônia Ocidental  
Ministério da Agricultura, Pecuária e Abastecimento*

# ***Boletim de Pesquisa e Desenvolvimento 18***

## **Microssatélites e o Guaranazeiro: Diversidade Submersa em Um Mar de Alelos**

*Paula Cristina da Silva Angelo  
Maria do Perpétuo Socorro Lira  
Manuella Villar Amado  
Ana Takagaki Yamaguishi  
Gilvan Ferreira da Silva  
Jorge Ivan Rebelo Porto  
Spartaco Astolfi-Filho  
Andre Luiz Atroch*

Embrapa Amazônia Ocidental  
Manaus, AM  
2013

Exemplares desta publicação podem ser adquiridos na:

**Embrapa Amazônia Ocidental**

Rodovia AM-010, Km 29, Estrada Manaus/Itacoatiara

Caixa Postal 319

Fone: (92) 3303-7800

Fax: (92) 3303-7820

<http://www.cpaa.embrapa.br>

**Comitê de Publicações da Unidade**

Presidente: *Celso Paulo de Azevedo*

Secretária: *Gleise Maria Teles de Oliveira*

Membros: *André Luiz Atroch, Edsandra Campos Chagas, Jony Koji Dairiki, José Clério Rezende Pereira, Kátia Emídio da Silva, Lucinda Carneiro Garcia, Maria Augusta Abtibol Brito, Maria Perpétua Beleza Pereira, Rogério Perin, Ronaldo Ribeiro de Moraes e Sara de Almeida Rios.*

Revisor de texto: *Maria Perpétua Beleza Pereira*

Normalização bibliográfica: *Maria Augusta Abtibol Brito de Sousa*

Diagramação: *Gleise Maria Teles de Oliveira*

Fotos da capa: *Murilo Rodrigues de Arruda*

**1ª edição**

1ª impressão (2013): 300

**Todos os direitos reservados.**

A reprodução não autorizada desta publicação, no todo ou em parte, constitui violação dos direitos autorais (Lei nº 9.610).

**CIP-Brasil. Catalogação-na-publicação.**

**Embrapa Amazônia Ocidental.**

---

Microsatélites e o guaranazeiro: diversidade submersa em um mar de alelos / Paula Cristina da Silva Angelo... [et al.]. – Manaus: Embrapa Amazônia Ocidental, 2013.

44 p. : il. color. - (Boletim de Pesquisa e Desenvolvimento / Embrapa Amazônia Ocidental, ISSN 1517- 2457; 18).

ISSN 1517-2457

1. *Paullinia cupana*. 2. Guaraná. 3. Marcadores moleculares. I. Angelo, Paula Cristina da Silva. II. Lira, Maria do Perpétuo Socorro. III. Amado, Manuella Villar. IV. Ciampi, Ana Yamaguishi. V. Silva, Gilvan Ferreira da. VI. Porto, Jorge Ivan Rebelo. VII. Astolfi-Filho, Spartaco. VIII. Atroch, André Luiz. IX. Série.

# Sumário

<b>Resumo.....</b>	<b>5</b>
<b>Abstract.....</b>	<b>7</b>
<b>Introdução.....</b>	<b>9</b>
<b>Material e Métodos.....</b>	<b>11</b>
<b>Resultados e Discussão.....</b>	<b>14</b>
<b>Conclusões.....</b>	<b>34</b>
<b>Referências.....</b>	<b>35</b>
<b>Anexo I.....</b>	<b>42</b>
<b>Anexos II.....</b>	<b>43</b>



# Microssatélites e o Guaranazeiro: Diversidade Submersa em Um Mar de Alelos

---

*Paula Cristina da Silva Angelo<sup>1</sup>  
Maria do Perpétuo Socorro Lira<sup>2</sup>  
Manuella Villar Amado<sup>3</sup>  
Ana Takagaki Yamaguishi<sup>4</sup>  
Gilvan Ferreira da Silva<sup>5</sup>  
Jorge Ivan Rebelo Porto<sup>6</sup>  
Spartaco Astolfi-Filho<sup>7</sup>  
Andre Luiz Atroch<sup>8</sup>*

## Resumo

Microssatélites podem gerar bons marcadores moleculares codominantes para análise de diversidade e determinação de taxa de autofecundação utilizando progênies de polinização aberta oriundas de genótipos selecionados. O enriquecimento de bibliotecas genômicas para microssatélites estava em andamento quando a poliploidia do guaranazeiro foi definitivamente documentada: é um aloploiploide com

---

<sup>1</sup>Bióloga, D.Sc. em Ciências Biológicas/Genética, pesquisadora da Embrapa Amazônia Ocidental, Manaus, AM, paula.angelo@cpaa.embrapa.br

<sup>2</sup>Engenheira florestal, M.Sc. em Ciências Biológicas (Botânica), assistente da Fundação Amazonas Sustentável, Manaus, AM.

<sup>3</sup>Bióloga, D.Sc. em Biotecnologia, professora do Instituto Federal de Educação, Ciência e Tecnologia do Espírito Santo, Unidade Vitória, Vitória, ES.

<sup>4</sup>Bióloga, D.Sc. em Ciências Biológicas (Genética), pesquisadora aposentada da Embrapa Recursos Genéticos e Biotecnologia, Brasília, DF.

<sup>5</sup>Biólogo, D.Sc. em Microbiologia Agrícola, pesquisador da Embrapa Amazônia Ocidental, Manaus, AM, gilvan.silva@embrapa.br

<sup>6</sup>Biólogo, D.Sc. em Biologia de Água Doce e Pesca Interior, pesquisador do Instituto Nacional de Pesquisas da Amazônia, Manaus, AM.

<sup>7</sup>Biólogo, D.Sc. em Ciências Biológicas (Biofísica), professor da Universidade Federal do Amazonas, Manaus, AM.

<sup>8</sup>Engenheiro agrônomo, D.Sc. em Genética, Conservação e Biologia Evolutiva, pesquisador da Embrapa Amazônia Ocidental, Manaus, AM, andre.atroch@embrapa.br

210 cromossomas. Os experimentos foram concluídos com sondas  $(CA)_{12}$ ,  $(CT)_{12}$  e  $(TC)_{14}$  e foi realizada busca por blocos repetitivos no banco de ESTs (*expressed sequence tags*) de frutos com sementes. Grande número de blocos repetitivos foi identificado, e isso foi considerado consequência da poliploidização recente. No entanto, blocos com oito ou mais repetições foram muito raros. A saturação das sondas por blocos curtos explica a escassez dos longos nas bibliotecas genômicas, mas não a raridade destes nas ESTs. Quinze cultivares clonais foram genotipadas em dez dos *loci* identificados, em triplicata, por causa da complexidade dos padrões de alelos observados. Pela mesma razão, nas análises subsequentes procedeu-se como naquelas para marcadores dominantes. Um dos *loci* é relacionado ao tempo para germinação e apresentou arranjo estrutural bastante peculiar. Concluiu-se que os alelos descritos neste trabalho podem distinguir cultivares de guaranazeiro. Os *fingerprintings* podem ser adicionados aos traços morfológicos e agrônômicos como características descritivas.

**Palavras-chave:** Amazônia, Sapindaceae, guaraná, poliploide, SSR.

# Microsatellites and the Guarana Plant: Diversity in a Sea of Alleles

---

## Abstract

*Microsatellites can produce good co-dominant molecular markers for diversity analysis and determination of self pollination rates on progenies originating from open pollination of selected genotypes. The enrichment of genomic libraries for microsatellites was on course when the polyploidy of the guarana plant was definitely reported: it is a polyploid with 210 chromosomes. The experiments were concluded using (CA)<sub>12</sub>, (CT)<sub>12</sub> and (TC)<sub>14</sub> probes and the ESTs (expressed sequence tags) in the data bank of seeded fruits were screened as well. A huge number of repetitive blocks was identified, and this was considered a consequence of the recent polyploidization. However, blocks with eight or more repeats were very rare. Saturation of the probes with short blocks can explain their scarcity in genomic libraries, but cannot explain why they were rare in the ESTs. Fifteen clonal cultivars were genotyped in 10 of the loci identified, with three replicates, due to the complexity of the observed allele patterns. For the same reason the subsequent analyses were conducted as proper for dominant markers. One of the assessed loci is related to the time needed for seeds germination and displayed a very peculiar structural organization. It was concluded that alleles*

*described in the present work can distinguish guarana cultivars. The fingerprintings can be added, as descriptors, to morphological and agronomic characteristics.*

*Key words: Amazon, Sapindaceae, guarana plant, polyploid, SSR.*

## Introdução

O uso de marcadores moleculares tem sido uma ferramenta importante para auxiliar os programas de melhoramento genético de plantas. Segundo Ferreira e Grattapaglia (1996), as aplicações de marcadores moleculares no melhoramento de plantas podem ser divididas em aplicações de curto, médio e longo prazo. Em curto prazo, a identificação e discriminação de genótipos; a médio e longo prazo, os marcadores permitem quantificar a variabilidade genética existente ao nível de DNA e correlacioná-la com a expressão fenotípica. Essa informação molecular integrada às metodologias de seleção e recombinação de genótipos permite obter avanços genéticos nos programas de melhoramento de forma mais rápida.

Em plantas perenes, marcadores moleculares têm sido usados para avaliar o grau de heteroziguidade em busca de vigor híbrido (GRATTAPAGLIA, 2001). Além disso, a seleção precoce em culturas perenes e a seleção indireta para caracteres de difícil avaliação são facilitadas pelo uso de marcadores moleculares, quando estão associados às características selecionadas.

Utilizando marcadores moleculares RAPD e caracteres relacionados à produção de frutos, Sousa (2003) avaliou parâmetros genéticos e a divergência genética em clones de guaranazeiro constituintes do Banco Ativo de Germoplasma da Embrapa Amazônia Ocidental e de 27 clones elite da rede estadual de avaliação e seleção de clones de guaranazeiro. No entanto, marcadores codominantes, como os microssatélites, podem ter utilização mais ampla e ser empregados mais eficientemente para as análises de fluxo gênico e identificação de taxas de autofecundação, por exemplo.

Tratava-se de encontrar *loci* de microssatélites com número útil de repetições, quando foi confirmada a poliploidia do guaranazeiro cultivado, *Paullinia cupana* (Kunth) var. *sorbilis* (Mart.) Ducke, que tem

210 cromossomas, é um alopoliploide (FREITAS et al., 2007) e está incluído no clado X (BUERKI et al., 2009)/tribo Paullinieae (HARRINGTON et al., 2005), da família Sapindaceae. O evento que deu origem à variedade *sorbilis* é muito provavelmente recente (ATROCH et al., 2009) e bastante extremo, visto que as outras plantas do mesmo gênero que foram genotipadas apresentam número básico de cromossomas igual a 12 e  $2n = 24$  (SOLIS-NEFFA et al., 2001). Esse grau de poliploidia, somado ao desconhecimento dos ancestrais, para que se possa determinar quantas vezes cada um dos dois genomas está representado nas células do guaranazeiro, dificulta a utilização dos microsatélites para análises de diversidade em populações de polinização aberta. Dificuldades semelhantes àquelas encontradas para as análises no guaranazeiro são registradas para se trabalhar com marcadores em cana-de-açúcar, espécie poliploide com cerca de 130 cromossomas nas cultivares atuais. As cultivares de cana-de-açúcar são formadas por cruzamentos entre espécies do mesmo gênero, com números de cromossomas muito diferentes, o que gera poliploides e aneuploides com padrões complexos de segregação nas meioses e contribui para o aparecimento de números de cópias (doses) diferentes de cada alelo em diferentes cultivares. Marcadores moleculares de dose única e modelos de segregação do tipo  $O1 \times O0$  (presença e ausência dos alelos dose única, sendo presença entendida como caráter dominante nos parentais de cruzamentos controlados) têm sido utilizados para mapeamento de QTLs (PASTINA et al., 2012). Esse exercício muito laborioso, com certeza, deve ser apoiado pelo conhecimento da ancestralidade e pela disponibilidade de montagens provisórias do genoma completo de *Saccharum officinalis*. De qualquer forma, os resultados de experimentos com marcadores microsatélite em poliploides complexos podem contribuir para inferir sobre a representatividade de cada um dos genótipos ancestrais e, talvez, sejam úteis para acelerar a seleção de características agrônômicas em populações oriundas de cruzamentos controlados, nos casos em que ocorrer a cosegregação.

Os objetivos dos experimentos aqui descritos foram identificar microssatélites em guaranazeiro, desenhar *primers* para amplificar esses microssatélites, validar e revalidar *primers* genotipando plantas de cultivares clonais recomendadas para plantio e avaliar o potencial de alguns *loci* como marcadores para distinguir cultivares e para adicionar características descritivas aos morfológicos e agronômicos que já estão vigentes.

## Material e Métodos

Foram realizados o enriquecimento de bibliotecas genômicas Sau3A I e Mse I com sondas (CA)<sub>12</sub>, (CT)<sub>12</sub>, (CA)<sub>12</sub> + (CT)<sub>12</sub> e (TC)<sub>14</sub>, seguindo o protocolo desenvolvido por Tenzer et al. (1999) e modificado por Gautschi et al. (2001) e Farias et al. (2003), e a busca por blocos de repetições no banco de ESTs (*expressed sequences tags*) de frutos e sementes de guaranazeiro (<https://helix.biomol.unb.br/GR/>), utilizando os aplicativos Staden/TROLL (MARTINS et al., 2006), em projeto coordenado pela Embrapa Amazônia Ocidental, com a colaboração da Universidade Federal do Amazonas (Ufam) e do Instituto Nacional de Pesquisas da Amazônia (Inpa). Os projetos foram financiados pela Fundação de Amparo à Pesquisa do Estado do Amazonas (Fapeam) (Auxílio nº 924/2003) e por recursos próprios da Embrapa, por meio do Programa de Melhoramento do Guaranazeiro (PA 02.10.02.006.00.04 e aqueles que o antecederam desde 2004).

Os *primers* foram desenhados utilizando o aplicativo Primer 3. Para validação dos *primers*, daqui em diante designados iniciadores, desenhados, estes foram utilizados para amplificar o DNA genômico de 15 cultivares clonais de guaranazeiro desenvolvidas pela Embrapa: 189 (BRS-CG 189), 372, 388 (cultivar Mundurucânia), 505, 608 (BRS-CG 608), 610, 611, 612, 624, 626 (cultivar Luzeia), 648 (BRS-CG 648), 861 (cultivar Cereçaporanga), 871 (cultivar Maués), 882 (BRS-CG 882), 850 (BRS-CG 850). O DNA dos clones foi extraído de folhas utilizando o kit YGP100 (BMC *Technologies*). Entre 27 pares de iniciadores, dez foram selecionados para validação.

As reações de PCR, em volume total de 13  $\mu\text{L}$ , continham 2,5 mM de mistura equimolar dos dNTPs, 2,5 mM de  $\text{MgCl}_2$ , 0,25  $\mu\text{g}/\mu\text{L}$  de BSA (*bovine serum albumine*), 150 nM de cada iniciador, 150 nM de *primer tail* marcado com fluoróforo FAM, 15 ng de DNA, 1 U de Taq DNA polimerase, tampão para Taq polimerase (Phoneutria) diluído para 1X. À extremidade 5' do iniciador F de cada par incluído na Tabela 1, foi adicionada a sequência TGTAACGACGGCCAGT, complementar à *tail* marcada com o fluoróforo. O termociclador foi programado para 3 min a 95 °C; 45 s a 57 °C; 30 s a 72 °C; 20 s a 95 °C, por 38 ciclos; síntese final por 30 minutos. Reações foram aplicadas em géis de agarose a 2,5% sem adição de brometo de etídeo, para o acompanhamento da incorporação da marcação fluorescente das *primer tails* aos produtos de amplificação, mediante exposição à luz ultravioleta. Para as genotipagens automáticas realizadas em sequenciador ABI 3100, na Embrapa Recursos Genéticos e Biotecnologia, em outubro de 2011, as reações de amplificação foram repetidas por três vezes.

Os picos/alelos foram identificados automaticamente nos cromatogramas de genotipagem, utilizando a ferramenta *label peaks* do aplicativo Genotyper (Applied Biosystems). A ferramenta *filter labels* foi programada para aplicação de dois critérios diferentes para eliminação de picos dos cromatogramas: **critério 1**) picos/alelos que diferiram dos picos próximos por 0,1 a 0,8 par de bases eliminados; e **critério 2**) picos/alelos que diferiram de picos próximos por 0,1 a 1,5 par de bases eliminados. Utilizando o critério 2, ficaram preservados apenas picos distando entre si por números de bases múltiplos de dois, porque a maioria dos *loci* acessados continha repetições de dinucleotídeos. O número de picos/alelos eliminados e preservados pela aplicação dos critérios 1 e 2 sobre os resultados obtidos na primeira réplica das genotipagens foi comparado. O número de picos/alelos eliminados pela aplicação do critério 2 foi anotado. Os dois critérios também foram avaliados qualitativamente observando-se a eliminação de artefatos e a preservação dos fragmentos utilizados como padrões de tamanho, que foram preparados e marcados com fluoróforos ROX por PCR. Os

cromatogramas referentes às cultivares clonais que foram submetidas à aplicação dos critérios 1 e 2 descritos acima foram comparados com resultados anteriores obtidos por genotipagens realizadas na Ufam (ANGELO, 2007, Anexo II). A leitura dos cromatogramas, a anotação automática do comprimento dos fragmentos/picos (em número de pares de bases) e a aplicação dos critérios 1 e 2 foram realizadas para cada uma das três réplicas. Além disso, foram preservados para análise apenas alelos presentes em pelo menos duas das réplicas.

A presença/ausência de alelos foi tabulada sob forma de dados binários. Para avaliar a capacidade de distinção dos genótipos, a contribuição de cada alelo foi obtida utilizando o aplicativo *pars*, atribuindo valor 1 a cada um dos picos/alelos e valor 0 aos demais, até que todos tivessem sido testados quanto ao poder de distinção, individualmente. A comparação entre os filogramas produzidos foi realizada utilizando o aplicativo *treedist* e o exame visual dos filogramas para cada alelo ou fragmento. O agrupamento das cultivares foi realizado pelo método do vizinho mais próximo (aplicativos *restdist* e *neighbor*), utilizando a distância de Nei e Li (1979), com site length = 20 e duas mil permutas de *bootstrap*. Fragmentos amplificados com intensidade muito baixa mas que se repetiram com as mesmas características em pelo menos dois dos cromatogramas de cada *loci* para cada cultivar foram mantidos nas análises, mas a estes foi dado um peso equivalente à metade do peso dado aos alelos com intensidade resolutive (peso 1 em 1.000 permutas e peso 0 nas demais; aplicativo *seqboot*). Esses aplicativos fazem parte do pacote PHYLIP, versão 3.6 (FELSENSTEIN, 2004).

Para identificação por homologia das sequências de ESTs que possuíam microssatélites foram utilizados os aplicativos BLAST, as sequências disponíveis no site do NCBI e do TAIR (<http://www.ncbi.nlm.nih.gov/>) e o BioEdit versão 7. O aplicativo Splice Predictor foi utilizado para estimar a existência de sítios de *splicing*, com base nas sequências de sítios canônicos em genes de *Arabidopsis thaliana* e também de sítios

cromatogramas referentes às cultivares clonais que foram submetidas à aplicação dos critérios 1 e 2 descritos acima foram comparados com resultados anteriores obtidos por genotipagens realizadas na Ufam (ANGELO, 2007, Anexo II). A leitura dos cromatogramas, a anotação automática do comprimento dos fragmentos/picos (em número de pares de bases) e a aplicação dos critérios 1 e 2 foram realizadas para cada uma das três réplicas. Além disso, foram preservados para análise apenas alelos presentes em pelo menos duas das réplicas.

A presença/ausência de alelos foi tabulada sob forma de dados binários. Para avaliar a capacidade de distinção dos genótipos, a contribuição de cada alelo foi obtida utilizando o aplicativo *pars*, atribuindo valor 1 a cada um dos picos/alelos e valor 0 aos demais, até que todos tivessem sido testados quanto ao poder de distinção, individualmente. A comparação entre os filogramas produzidos foi realizada utilizando o aplicativo *treedist* e o exame visual dos filogramas para cada alelo ou fragmento. O agrupamento das cultivares foi realizado pelo método do vizinho mais próximo (aplicativos *restdist* e *neighbor*), utilizando a distância de Nei e Li (1979), com site length = 20 e duas mil permutas de *bootstrap*. Fragmentos amplificados com intensidade muito baixa mas que se repetiram com as mesmas características em pelo menos dois dos cromatogramas de cada *loci* para cada cultivar foram mantidos nas análises, mas a estes foi dado um peso equivalente à metade do peso dado aos alelos com intensidade resolutive (peso 1 em 1.000 permutas e peso 0 nas demais; aplicativo *seqboot*). Esses aplicativos fazem parte do pacote PHYLIP, versão 3.6 (FELSENSTEIN, 2004).

Para identificação por homologia das sequências de ESTs que possuíam microssatélites foram utilizados os aplicativos BLAST, as sequências disponíveis no site do NCBI e do TAIR (<http://www.ncbi.nlm.nih.gov/>) e o BioEdit versão 7. O aplicativo Splice Predictor foi utilizado para estimar a existência de sítios de *splicing*, com base nas sequências de sítios canônicos em genes de *Arabidopsis thaliana* e também de sítios

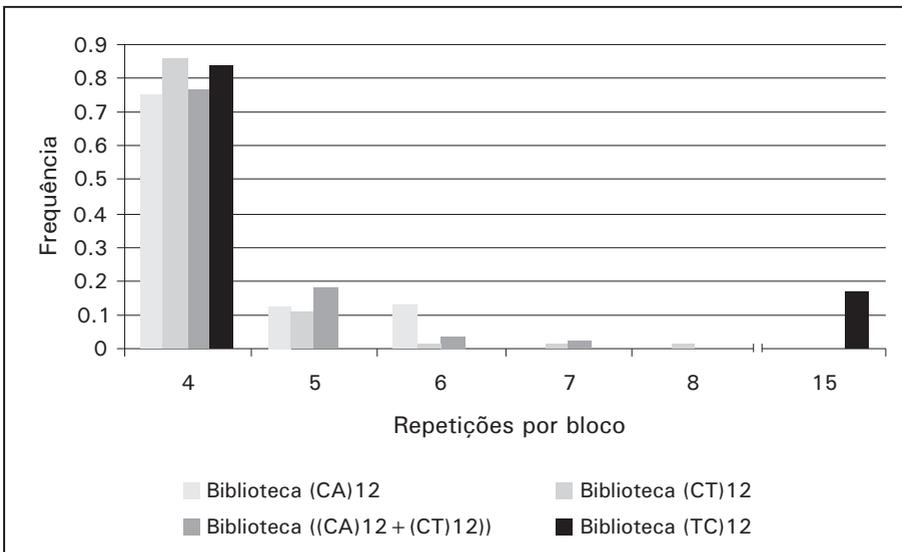
canônicos em humanos (<https://sites.google.com/a/usd.edu/bioinformatics/projects/splice predictor>). Os aplicativos Softberry (<http://linux1.softberry.com/cgi-bin/programs>) alimentados com dados de *A. thaliana* e os bancos de dados do projeto *Arabidopsis Small RNA Project* – ASRP (<http://asrp.danforthcenter.org/>) foram utilizados para a identificação de sequências regulatórias e predição de ocorrência de alvos e fontes de smRNAs (*small RNAs*) na região 3' – não traduzida do gene MFT.

## Resultados e Discussão

Das sequências com qualidade suficiente para análise (688) que resultaram do enriquecimento de bibliotecas genômicas, 28% apresentaram blocos perfeitos com pelo menos quatro repetições de dinucleotídeos e não houve diferença significativa entre bibliotecas quanto a essas características (Figura 1). Os blocos perfeitos com quatro repetições foram significativamente mais frequentes (80% aproximadamente, em média) do que blocos com cinco repetições (13%) e com seis, sete e oito repetições reunidos (menos de 10% cada).

A porcentagem média de enriquecimento (blocos complementares às sondas) das bibliotecas foi de 13,8%. No entanto, o enriquecimento com sondas repetitivas (TC)<sub>n</sub> e (AC)<sub>n</sub> é geralmente bem-sucedido para outras espécies vegetais como pequi (COLLEVATTI et al., 1999), carvalho (HODGETTS et al., 2001) e *Orphys araneola*, uma Orchidaceae (SOLIVA et al., 2000). Em bibliotecas enriquecidas de *Eucalyptus* as classes com maior frequência de blocos, que constituíram juntas aproximadamente 60%, foram aquelas com 16 a 20 e 21 a 25 repetições, e 80% dos blocos selecionados eram perfeitos (BRONDANI et al., 1998). Para as análises de diversidade em kiwi (HUANG et al., 1998), coqueiro (PERERA et al., 1999), oliveira (CARRIERO et al., 2002) e maçaranduba (AZEVEDO et al., 2004), foram utilizados microsatélites com números mínimos de 15, 8, 13 e 9 repetições de dinucleotídeos, respectivamente,

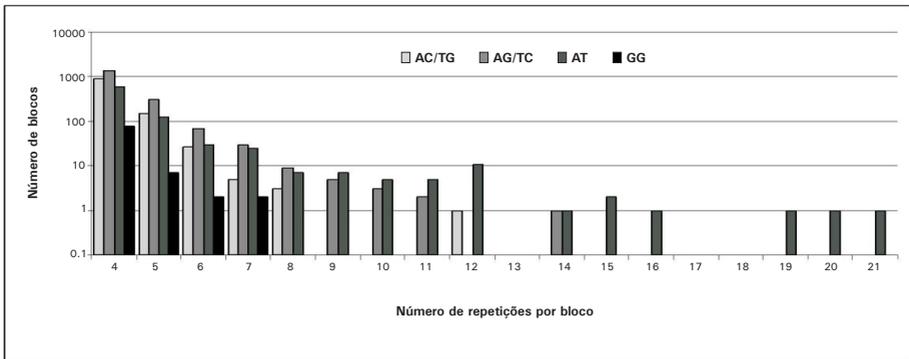
e número máximo sempre superior a 20 repetições, em arranjos perfeitos, e também arranjos compostos. Parece haver evidências de que variações em *loci* com números grandes de repetições são mais frequentes do que em *loci* com poucas repetições (FELSENSTEIN, 2004b). Portanto, apesar de terem sido construídas quatro bibliotecas genômicas enriquecidas em sequências repetitivas, não foram identificados *loci* com a estrutura ideal para avaliar a diversidade de populações de polinização aberta em número sequer aproximado ao que é relatado para outras espécies.



**Figura 1.** Frequência relativa (número de blocos em cada classe/total de fragmentos clonados em cada biblioteca) dos blocos repetitivos de dinucleotídeos por classes de número de repetições, selecionados em bibliotecas genômicas de guaranaizeiro por busca utilizando os aplicativos Staden/TROLL.

Entre ESTs de frutos com sementes de guaranaizeiro foram identificados 4.999 blocos repetitivos, sendo 3.814 blocos de dinucleotídeos (Figura 2), 1.093 blocos de trinucleotídeos e 92 blocos de tetranucleotídeos. Os dinucleotídeos merecerão mais atenção neste trabalho porque a

maioria dos *loci* utilizados para genotipagens continha blocos de repetições de dinucleotídeos. Entre os dinucleotídeos, os blocos mais frequentes foram compostos de repetições AG/TC, seguidos por repetições de AC/TG, TA e GC (Figura 2).



**Figura 2.** Número de blocos repetitivos perfeitos de dinucleotídeos, identificados no banco de ESTs de frutos e sementes de guaranazeiro por busca utilizando os aplicativos Staden/TROLL.

Frequências baixas de blocos perfeitos de dinucleotídeos podem ser relacionadas ao tamanho dos genomas. As frequências de blocos perfeitos de dinucleotídeos com pelo menos três repetições no DNA genômico (43,4 por Mpb) e em ESTs (61,0 por Mpb) de trigo, que tem genoma haploide de 5.600 Mbps, foram menores que aquelas observadas no genoma com 125 Mpbs de *Arabidopsis thaliana*: 77,7 e 137,0 por Mpbs, em DNA genômico e ESTs, respectivamente. Supondo que os microsatélites teriam surgido antes das expansões genômicas que acometeram aparentemente todas as espécies vegetais nos últimos cinco milhões de anos (portanto, começando no Período Terciário), eles seriam mais frequentes em genomas menores, com menores quantidades de *junk* DNA. É que, nos períodos evolutivos posteriores, as expansões dos genomas das plantas e o aparecimento do *junk* DNA teriam sido muito influenciados pela amplificação de retrotransposons. Esses elementos móveis são eliminados com mais eficiência de regiões

hipometiladas, não repetitivas e ricas em sequências transcritas dos genomas. Por isso, as regiões transcritas teriam sido menos afetadas pela amplificação dos retrotransposons do que as regiões repetitivas, hipermetiladas dos genomas. Se os microssatélites já existiam nas regiões transcritas antes das expansões por poliploidização que começaram no Período Terciário, explica-se a correlação positiva entre a frequência de microssatélites e a porcentagem de DNA cópia única ou com número baixo de cópias, melhor representado nas ESTs do que no genoma (MORGANTE et al., 2002). Para o guaranazeiro, com genoma estimado em 22 Gpbs (22,8 pg por núcleo diploide, segundo Freitas et al., 2007), o que foi dito acima poderia explicar, pelo menos parcialmente, a dificuldade no isolamento dos microssatélites. Mas, no entanto, para a análise do transcriptoma de frutos com sementes foram acessados 15.387 ESTs com comprimento médio de 773 bases (ANGELO et al., 2008). Então, estimando grosseiramente, foram encontrados cerca de 320 blocos com pelo menos quatro repetições de dinucleotídeos por Mpb no guaranazeiro. Isto é, 2,3 vezes os 137 blocos por Mpb registrados para *Arabidopsis*, incluídos aí blocos com três repetições, e *Arabidopsis* tem genoma cerca de 6 mil vezes menor do que o guaranazeiro e contradiz frontalmente o que seria esperado se a correlação negativa com o tamanho dos genomas fosse absoluta. No entanto, é preciso pensar que a poliploidização foi um evento muito mais recente que expansões terciárias que possam ter sofrido os genomas das espécies que deram origem ao guaranazeiro. E, se os genomas dos dois ancestrais do guaranazeiro foram amplificados em curtíssimo espaço de tempo após o advento da hibridação inicial, obviamente incluindo as regiões transcritas, então os microssatélites que naqueles genomas já existiam passaram a existir em número igualmente amplificado em uma mesma célula aloploiploide da variedade cultivada.

A mesma tendência foi observada entre ESTs. Um número muito reduzido de blocos com oito ou mais repetições de dinucleotídeos em guaranazeiro foi identificado (Figura 2). A frequência relativa de blocos

com número maior ou igual a oito foi 0,77% (66 blocos repetitivos/ 8.597 sequências examinadas) no banco de ESTs e 0,29% (2/688) nas bibliotecas genômicas, valores que não diferem estatisticamente ( $P > 0,05$ , teste Z). *Arabidopsis*, soja, trigo, milho e arroz apresentaram, todos, número médio de repetições de dinucleotídeos por bloco de cerca de nove (MORGANTE et al., 2002). Mas o simples exame da Figura 2 permite inferir que uma frequência muito mais alta de blocos com número de repetições acima de oito seria necessária para aproximar a média do guaranazeiro das nove repetições nas outras espécies. Isso reforça a afirmativa feita anteriormente de que muito poucos blocos ideais para a análise de diversidade em populações de polinização aberta foram identificados. Mas a explicação para essa característica do guaranazeiro não é fácil. Se tem alguma relação com a poliploidização, não é uma relação óbvia. A existência de muitos blocos com poucas repetições ficou evidente até mesmo pelo exame visual dos fragmentos clonados após enriquecimento (Figura 3) e pode-se supor que estes blocos curtos, que são muito mais frequentes e devem ser muitíssimo numerosos no genoma com cerca de 22 Gpbs do guaranazeiro, tenham saturado as sondas utilizadas para o enriquecimento de bibliotecas genômicas, impedindo a hibridização com blocos de número mais alto de repetições. Em um genoma tão grande, pode-se considerar que nem mesmo experimentos de reenriquecimento devem surtir efeitos muito significativos. No entanto, a influência dos métodos de enriquecimento não pode ser aventada como razão para explicar a raridade de blocos com oito e mais repetições em ESTs. Teoricamente, pode-se pensar que blocos com número muito alto de repetições podem interferir mais ubiquamente na conservação evolutiva de segmentos codificadores do genoma e por isso seriam mais raros aí. Mas a diferença com relação às espécies citadas acima é marcante e precisa ser analisada com mais profundidade.

Entre 27 pares de iniciadores desenvolvidos, 10 foram selecionados para a validação e utilização nas genotipagens (Tabela 1).

com número maior ou igual a oito foi 0,77% (66 blocos repetitivos/ 8.597 sequências examinadas) no banco de ESTs e 0,29% (2/688) nas bibliotecas genômicas, valores que não diferem estatisticamente ( $P > 0,05$ , teste Z). *Arabidopsis*, soja, trigo, milho e arroz apresentaram, todos, número médio de repetições de dinucleotídeos por bloco de cerca de nove (MORGANTE et al., 2002). Mas o simples exame da Figura 2 permite inferir que uma frequência muito mais alta de blocos com número de repetições acima de oito seria necessária para aproximar a média do guaranazeiro das nove repetições nas outras espécies. Isso reforça a afirmativa feita anteriormente de que muito poucos blocos ideais para a análise de diversidade em populações de polinização aberta foram identificados. Mas a explicação para essa característica do guaranazeiro não é fácil. Se tem alguma relação com a poliploidização, não é uma relação óbvia. A existência de muitos blocos com poucas repetições ficou evidente até mesmo pelo exame visual dos fragmentos clonados após enriquecimento (Figura 3) e pode-se supor que estes blocos curtos, que são muito mais frequentes e devem ser muitíssimo numerosos no genoma com cerca de 22 Gpbs do guaranazeiro, tenham saturado as sondas utilizadas para o enriquecimento de bibliotecas genômicas, impedindo a hibridização com blocos de número mais alto de repetições. Em um genoma tão grande, pode-se considerar que nem mesmo experimentos de reenriquecimento devem surtir efeitos muito significativos. No entanto, a influência dos métodos de enriquecimento não pode ser aventada como razão para explicar a raridade de blocos com oito e mais repetições em ESTs. Teoricamente, pode-se pensar que blocos com número muito alto de repetições podem interferir mais ubiquamente na conservação evolutiva de segmentos codificadores do genoma e por isso seriam mais raros aí. Mas a diferença com relação às espécies citadas acima é marcante e precisa ser analisada com mais profundidade.

Entre 27 pares de iniciadores desenvolvidos, 10 foram selecionados para a validação e utilização nas genotipagens (Tabela 1).

```

TGCCAGTGAA TCTCCTGTAG TGGAAAGAGAC TCCCTAGGGC TGTCATATTG CTATCGGTGA
GCATTGTTAC TATTAGTGTG TCCCGACTTA TCGTTCGGAA TAGGGCGGTA ATCTCACTGT
GGTCCATTTG CCTTTGT-CC CGGGAAGGTT TTTATAAAAC GCAACGTCGT TAGACCAAGA
GTTAGATAAG AATAAGAata tata tGAAAG CGCTTGATTA TACAGCAAGA AAGATAATTA
TTATTTAATA CCGTGATTTA CTTCGCGTAG TTGAGGGAGG GTGAGAAGAG ATACAAGAGG
GGGAGGGTCA GGTGTGTGCT AGGTGGTGCG GGGAGGGgag agagagaAAG TGTGTATGAT
GGAGTGTGGT GGTGAGTGT CGTATCCGTG TAGGGATGCC TATATGGTGG GGTGGTATGG
GAGGAGGTGT GGTGAAGTAG CTGATGTGAA GGAGGATAGG TGGGAAGGG GAGAAGTGAA
GGTGGTGAG GAGAAAAAAA ATATATAATA GTTAAatata tatTAATAAA AATAACAAAA
AATCATACAG AGATCACAAA AGAGCTCAAC AAACATAAT AGTCAACTTA TTACTCTCTT
CAATTATATA TTTAACACAA ATACATTAAT AACTATATAT TTTAATAACA TGTCTTGTAT
AAAAAAACA TACAGACAAA ATCATACACC TATTAAATCA TATAACACAA TAAAAATAAA
    
```

**Figura 3.** Sequência de fragmento selecionado por enriquecimento de bibliotecas genômicas de guaranazeiro com sondas repetitivas CA.

**Tabela 1.** Iniciadores utilizados para a genotipagem de cultivares clonais de guaranazeiro recomendadas para plantio no Amazonas, no ano de 2011.

Primer	Repetições	Sequências dos primers (5´-3´)	Origem
GRN01	(CA) <sub>6</sub> .(CA) <sub>3</sub>	AGAACTGGTCCAACCGTCTC e CGTGAAAGGTCTGAGTGAAGC	B GEN
GRN04	CA(CCA) <sub>4</sub> CA(CCA)	CATCCATTGTCACCTCTTGG e TGGCATGAGACAATTTGTGG	B GEN
GRN05	(TA) <sub>5</sub> .(AT) <sub>5</sub>	GATCAGGTGCCACTCCAAGT e TTGTGTGAGTCCTCAGCCTCT	B GEN
GRN07	(TA) <sub>16</sub>	GGTCTTTTTCAGCGCAGTTG e GGATAGAGCACCGAGAGAC	BC ESTS
GRN09	(AT) <sub>8</sub> C(TA) <sub>12</sub>	CTGAATGCTGTCCAAGCA e GCAGCCTTCCCATTTTACC	BC ESTS
GRN12	(TG) <sub>5</sub> (TA) <sub>7</sub>	TGTCAAACCCCTTTATGTTT e CAATGGTGCCAGTAAATACA	BC ESTS
GRN13	(CT) <sub>5</sub> (TA) <sub>7</sub>	TGAAATATATGTGGCATGA e TGCATAAAATCACCAATCA	BC ESTS
GRN14	(TC) <sub>11</sub>	ATGATCATCAGCATCATGG e TATCCAACCTCAATTCCCAGA	BC ESTS
GRN17	(CT) <sub>9</sub>	TGCGACTGTGAGTGAGCTA e TTGAGATGAACTCAGCCAAC	BC ESTS
GRN20	(CGC) <sub>5</sub> .(CGC) <sub>3</sub>	TGGGAGGAGAGAACCGTACA e GCCTTCCCTTCTATCAAGC	BC ESTS

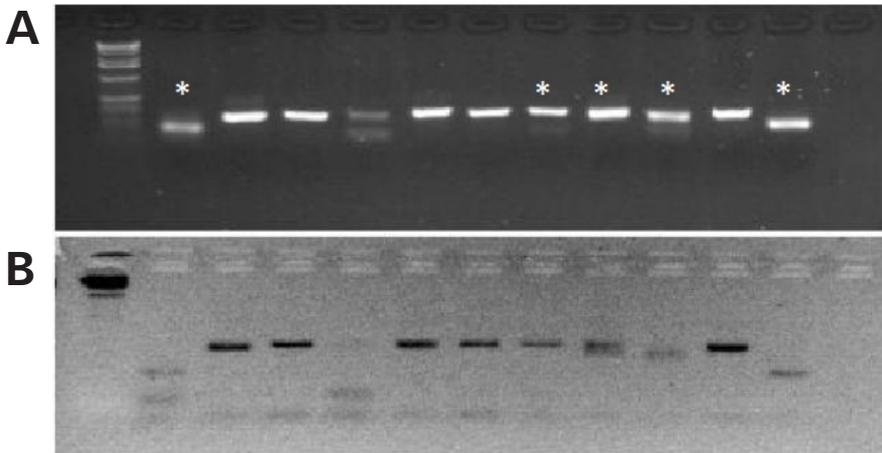
OBS: B GEN = enriquecimento de biblioteca genômica e BC ESTs = busca em banco de ESTs.

A utilização de *primer tail* diminui os gastos com relação à marcação de um dos iniciadores em cada par, mas provoca o aumento do número de variáveis que interferem na obtenção de reações com a qualidade necessária para uma boa genotipagem. Houve situações em que a amplificação dos fragmentos ocorreu com a intensidade esperada (Figura 4A), mas a incorporação da *tail* marcada com o fluoróforo não seguiu o mesmo padrão (Figura 4B, especialmente as reações marcadas

com asteriscos). Nestes casos, em que a incorporação da fluorescência falha ao serem utilizadas as condições nas quais os fragmentos contendo blocos repetitivos são amplificados, é necessário realizar mais testes de ajuste. Buscar combinações ideais de temperaturas de anelamento e concentrações de iniciadores e dNTPs (não fluorescentes) com as *tails* portando os fluoróforos que são adicionadas nas mesmas reações, de maneira a favorecer estequiométrica e bioquimicamente a marcação fluorescente dos fragmentos amplificados, pelo menos nos últimos ciclos de amplificação, demanda tempo e reagentes. Nos experimentos com iniciadores para microssatélites de guaranazeiro, alguns *loci* foram eliminados das análises porque as condições ideais para que ocorressem consecutivamente a amplificação e marcação dos fragmentos portando microssatélites não foram encontradas entre aquelas que foram testadas. A incorporação da fluorescência aos fragmentos amplificados é importante para preservar a reprodutibilidade das genotipagens. Cromatogramas reprodutíveis podem fornecer informações que vão além do comprimento e número dos alelos, como área sob os picos e altura. Conclui-se, então, que quando houver um número limitado de pares de iniciadores selecionados para a realização de experimentos de genotipagem, como ocorreu com o guaranazeiro, que seja mais vantajoso marcar um iniciador de cada par com fluoróforos e eliminar os ajustes necessários para a incorporação das *tails*.

A aplicação do critério 2 levou à eliminação de picos candidatos a alelos em sete dos dez *loci*, porque estavam distanciados dos vizinhos por menos de duas bases. O *locus* mais afetado pela diferença entre os critérios 1 e 2 foi o GRN14. Nos outros nove *loci*, foi eliminado das análises apenas um ou nenhum dos picos que estavam dentro da faixa de comprimento esperada, na área da concentração de picos em cada *loci*, e que haviam sido identificados automaticamente nos cromatogramas (Tabela 2). As diferenças não foram muito drásticas, e então, sem que fossem realizadas análises adicionais, considerou-se mais útil utilizar os resultados produzidos pela aplicação do critério 1, porque a ocorrência de variantes alélicas por diferença de apenas uma

base em *loci* portando blocos de dinucleotídeos não é inédita, tendo sido registrada em kiwi (HUANG et al., 1998) e oliveira (RALLO et al., 2000). Tratando-se do guaranazeiro, que é um aloploiploide, até mesmo diferenças resultantes de pequenos deslocamentos do sítio de anelamento dos iniciadores em cromossomas homoeólogos são teoricamente admissíveis e podem produzir alelos. Variações no número de repetições incluídas nos blocos repetitivos, mas também variações (inserções/deleções) nos flancos dos blocos repetitivos foram relatadas recentemente em *loci* transferidos de *Brassica rapa* para outras espécies do mesmo gênero (RAMCHIARY et al., 2011). Em *Lychia*, os *loci* em que um padrão de ocorrência de alelos diferente do diploide foi observado – LMY1, 2 e 10 – foram eliminados da análise de diversidade (VIRUEL; HORMAZA, 2004). Esse tipo de procedimento, caso fosse aplicado, teria inviabilizado o trabalho aqui apresentado.



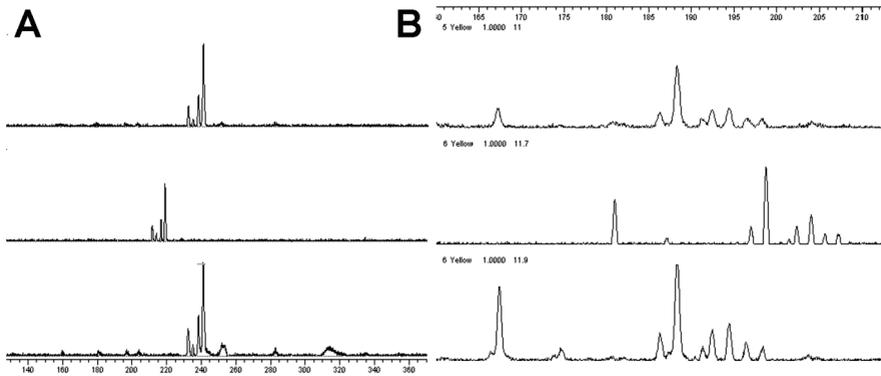
**Figura 4.** Marcador 1 kb *ladder* (Invitrogen) e fragmentos de DNA de cultivar de guaranazeiro amplificados por PCR. A ordem de aplicação das reações nos dois géis foi a mesma e as reações são idênticas. A – gel de agarose a 2,5% corado com brometo de etídio (0,5 µg/mL). B – gel de agarose sem adição de brometo de etídio para avaliar a incorporação de *primer tail* marcada com o fluoróforo. Os asteriscos indicam as reações em que a incorporação da fluorescência foi menos eficiente.

**Tabela 2.** Número de picos observados dentro da faixa de tamanho esperada para os fragmentos amplificados com iniciadores para microssatélites que foram preservados ou eliminados pela aplicação de critérios referentes à distância entre picos vizinhos, nos cromatogramas resultantes da genotipagem automática de 15 cultivares clonais de guaranazeiro.

<i>Locus</i>	Alelos		Diferença entre os critérios
	Critério 1	Critério 2	
GRN01	8	8	0
GRN04	7	6	1
GRN05	4	3	1
GRN07	6	6	0
GRN09	5	5	0
GRN12	2	1	1
GRN13	5	4	1
GRN14	23	17	6
GRN17	10	9	1
GRN20	9	8	1
<b>Total</b>	<b>79</b>	<b>67</b>	<b>12</b>
<b>Média</b>	<b>7,9</b>	<b>6,7</b>	<b>1,2</b>

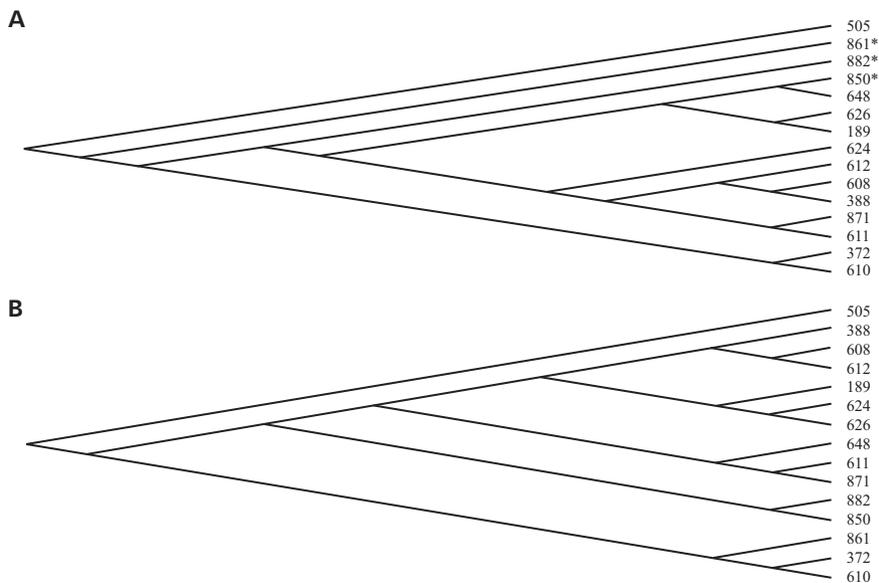
O próximo passo para a seleção de alelos úteis para a distinção ou discriminação das cultivares clonais de guaranazeiro foi avaliar sua reprodutibilidade. As genotipagens foram realizadas com réplicas porque a maioria dos *loci* de microssatélites apresentou padrão de alelos muito complexo. Pretendeu-se com as réplicas eliminar, pelo menos um pouco, a interferência dos artefatos gerados durante as reações de amplificação ou durante a eletroforese dos fragmentos (picos/alelos) amplificados. Contando todos os *loci*, 72% dos cromatogramas repetiram-se em duas dentre as três genotipagens de cada cultivar, o que significa dizer que nessa percentagem dos cromatogramas foram observados os mesmos picos por *loci* e cultivar, em pelo menos duas ocasiões (Figura 5). O

registro de dados foi feito por cultivar e os picos que não contribuíram para escore de reprodutibilidade de 2/3 foram registrados como dados faltantes. Os piores *loci* nesse sentido foram o GRN05 e o GRN12, com muitos picos observados em apenas uma das três réplicas e eliminados das análises para nove cultivares.



**Figura 5.** Cromatogramas representativos de genotipagens de cultivares clonais de guaranazeiro. A – três réplicas da genotipagem da cultivar 388 com os iniciadores GRN20. B – três réplicas da genotipagem da cultivar 612 com os iniciadores GRN17.

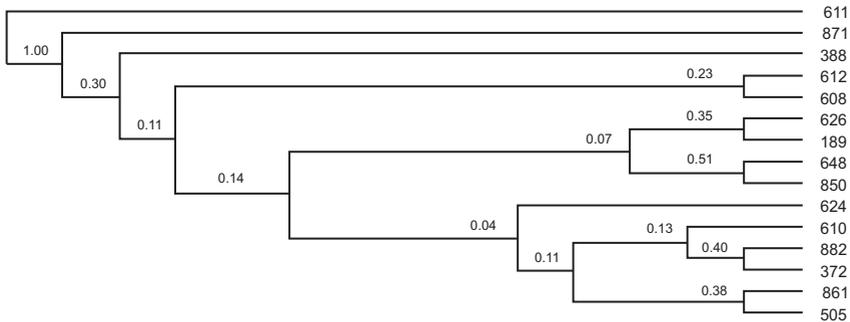
Alguns picos (63 picos marcados por “?” no Anexo I) satisfizeram o escore de reprodutibilidade, mas apresentaram outras características, como intensidade muito baixa ou áreas com formas incomuns, consideradas peculiares o suficiente para que se necessite de mais réplicas de verificação. No entanto, por causa do seu valor resolutivo, os fragmentos/alelos onde isso foi verificado foram mantidos nas análises de distinção entre cultivares, com peso menor do que os alelos com intensidade alta e reprodutíveis, simultaneamente. Esses picos foram úteis, principalmente, para distinguir as cultivares 850, 861 e 882, mas contribuíram para reduzir o número de cultivares agrupadas em outros clados (Figura 6) e devem ser observados nos experimentos futuros, para serem ou não eliminados definitivamente.



**Figura 6.** Filogramas de agrupamento de cultivares clonais de guaranazeiro realizados pelo método do vizinho mais próximo, segundo a presença **(A)** e ausência **(B)** de alguns picos de cromatogramas que precisam ser reavaliados nos próximos experimentos. Os agrupamentos foram realizados após 1.000 rearranjos de *bootstrap*.

Por fim, ficaram preservados para as análises 59 alelos. O tamanho e a frequência de cada alelo estão registrados no quadro do anexo I. Dois alelos exclusivos foram encontrados, L14215 e L20230, para as cultivares 626 e 608, respectivamente (anexo I), mas não se pode informar, por enquanto, quantas cópias de cada um ocorreu em cada uma das cultivares. Não foram observadas, entre as 15 avaliadas, cultivares completamente idênticas, ou seja, nenhuma cultivar apresentou perfil de genotipagem idêntico a qualquer outra (Figura 7). Os valores de *bootstrap* registrados para cada partição do agrupamento foram relacionados com a capacidade de distinção entre cultivares agrupadas em um mesmo clado. Quanto maior o índice de *bootstrap*, um

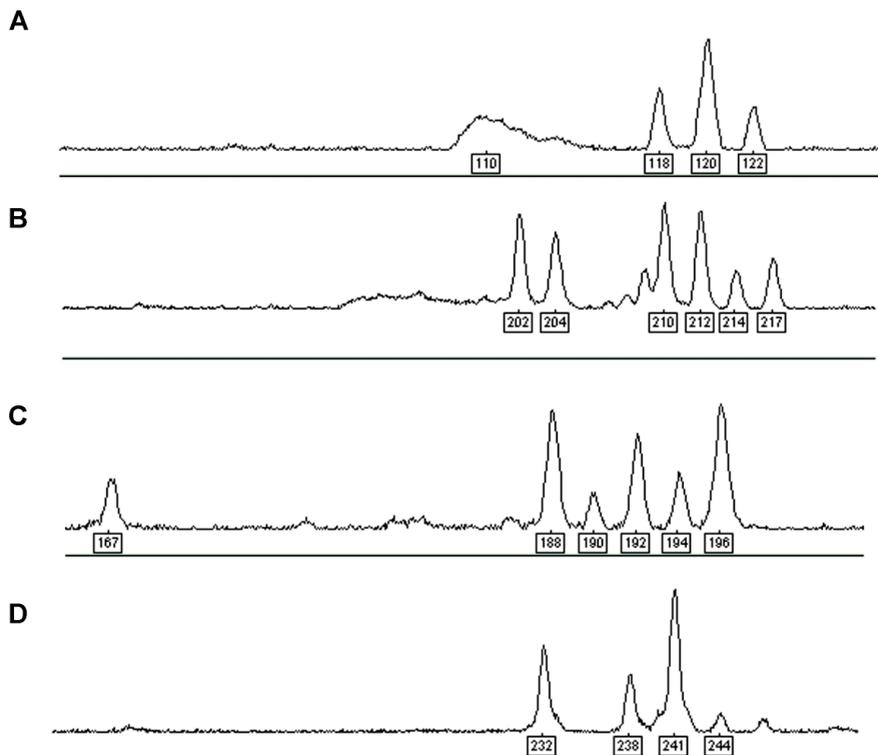
maior número de picos/alelos identificados neste trabalho será necessário para distinguir ou identificar membros de um mesmo clado. Então, para distinguir a cultivar 850 da 648 pode ser necessário amplificar todos os 10 *loci* de microssatélites e observar a presença e ausência de um grande número de picos/alelos. Por outro lado, será menos difícil distinguir a 626 da 189, e muito fácil distinguir os pares 648 + 850 e 626 + 189. Os picos/alelos que mais contribuíram para distinguir cultivares foram L12259, L12265, L14208, L17190, L17198, L20235 e L20238.



**Figura 7.** Agrupamento de cultivares clonais de guaranazeiro realizado pelo método do vizinho mais próximo, segundo a presença e ausência de 59 alelos de microssatélites presentes em 10 *loci*. Os valores próximos às partições do agrupamento representam porcentagens de *bootstrap* em 2.000 rearranjos. As cultivares clonais são identificadas por seus códigos de acesso no Banco de Germoplasma do Guaranazeiro.

Por se tratar de um poliploide, observou-se em alguns *loci* de microssatélites até mais de cinco picos/alelos por planta (Figura 8). Para alguns *loci*, o número máximo de alelos observados por planta foi dois ou quatro, que são múltiplos de dois (Figuras 5A e 8A); outros apresentaram número máximo igual a três ou seis alelos (GRN03 e Figura 8B), que são múltiplos de três. O *locus* GRN03 (vide anexo II deste trabalho) é multialélico e monomórfico, e os mesmos três alelos ocorreram para todos os indivíduos genotipados. Em um terceiro padrão

de organização, o número máximo de alelos por indivíduo é uma soma de dois e três (Figura 8C). Há indícios, no cariógrama, de que um grupo de cromossomos que contribuíram para a variedade *sorbilis*, o poliploide cultivado atualmente, estaria multiplicado quatro vezes e o outro, provavelmente, seis vezes. Então, talvez, padrões com número máximo de alelos múltiplo de dois sejam gerados por amplificação de regiões presentes nos cromossomas de um dos grupos e padrões em que o número máximo de alelos seja múltiplo de três comportem alelos presentes no outro grupo, enquanto os *loci* que são composições desses dois padrões talvez comportem combinações de alelos diferentes presentes em nos dois grupos. Mas essas possibilidades precisam ser muito melhor investigadas e avaliadas, considerando-se outra característica muito interessante dos cromatogramas que é a área abaixo dos picos ou a altura relativa dos picos que representam os alelos. Para tal, será necessário replicar novamente as genotipagens utilizando iniciadores marcados individualmente com fluoróforos, de maneira a obter picos com intensidade mais alta nos cromatogramas e valores de desvio estatisticamente confiáveis de áreas e/ou alturas relativas. Trabalhar com citogenética de hibridização *in situ* para verificar a possibilidade de que picos mais altos representem alelos que ocorrem em número maior de vezes no genoma do guaranazeiro também traria contribuição importantíssima. Diferenças entre a altura dos picos em cromatogramas de cultivares diferentes estão exemplificadas pela genotipagem da 388 e da 612 com iniciadores GRN20 (Figuras 5A e 8D). A descoberta das espécies que são as ancestrais da variedade *sorbilis*, provavelmente diploides, poderia fornecer o suporte para todas essas análises. Um trabalho bastante completo foi realizado utilizando cruzamentos controlados entre espécies diferentes da seção *Caninae* DC. de *Rosa* L., em que se demonstrou a contribuição de cada cópia de um dado alelo, presente em doses únicas nos parentais, para a altura dos picos nos cromatogramas dos híbridos interespecíficos, onde o alelo entrou em homozigose e ganhou aproximadamente o dobro da altura (NYBOM et al., 2004).



**Figura 8.** Cromatogramas correspondentes a genotipagens automáticas de cultivares clonais de guaranazeiro. Cada cromatograma representa o genótipo de um único indivíduo com apenas um par de iniciadores, sendo: **A** – GRN01; **B** – GRN14, **C** – GRN17; e **D** – GRN20.

Dos dez *loci* amplificados para as genotipagens mais recentes, sete estão em ESTs. Os ortólogos das ESTs de guaranazeiro, identificados por homologia estrutural com sequências depositadas no GenBank, estão listados na Tabela 3.

**Tabela 3.** Identidade das ESTs de guaranazeiro que portam blocos de microsatélites, definida por homologia estrutural com sequências depositadas no GenBank.

Locus	Sequências	Ortólogo				E <sup>VALUE</sup>
		Gene	Espécie	Acesso		
GRN07	Contig 0660	<i>Mother of flowering locus T</i>	<i>Fragaria vesca</i>	XP4299541.1	2 x e <sup>-89</sup>	
GRN09	Contig 2123	<i>Endo-xyloglucan transferase</i>	<i>Gossypium hirsutum</i>	BAA21107.1	1 x e <sup>-136</sup>	
GRN12	AO-19s-C12	<i>Cyclase</i> (catabolismo de triptofano)	<i>Arabidopsis thaliana</i>	EFH43390.1	2 x e <sup>-39</sup>	
GRN13	AO-21s-F07	<i>Gibberelin oxidase</i>	<i>Pisum sativum</i>	ABI64150.1	4 x e <sup>-42</sup>	
GRN14	AO-14s-C06	<i>Unnamed</i> (TCP transcription factor)	<i>Vitis vinifera</i>	CAO48409.1	8 x e <sup>-26</sup>	
GRN17	IA-175g-H08	<i>Anfipatic channel</i> (membrana)	<i>Jatropha curcas</i>	ADU56185.1	6 x e <sup>-24</sup>	
GRN20	Contig 2098	<i>Zinc finger protein C2H2</i>	<i>Ricinus communis</i>	EEF34900.1	9 x e <sup>-49</sup>	

Há registro de alterações na regulação e na estrutura de sequências expressas que são causadas por variações no número de repetições incluídas em blocos de microsatélites. Nove microsatélites identificados em cevada são ligados com *loci* relacionados ao tempo necessário para o florescimento, sendo que os padrões de segregação de dois deles combinados explicaram 15% da variação para essa característica, quando avaliada sob dias curtos. Um dos dois *loci*, o HVM40, é composto de (GA)<sub>6</sub>(GT)<sub>4</sub>(GA)<sub>7</sub> e o outro, o HVM03, é composto por variações no número de repetições de AT (LIU et al., 1996; IVANDIC et al., 2002). Outras combinações entre *loci* explicaram mais de 30% da variabilidade quanto ao tempo necessário para o florescimento (LIU et al., 1996).

Em guaranazeiro, o *locus* GRN07 foi identificado em uma EST homóloga ao gene *MOTHER OF FLOWERING LOCUS T AND TFL1 (MFT)* de *Fragaria vesca* (Tabela 3). O homólogo desse gene em *Arabidopsis* codifica para um fator de transcrição que pode ser expresso em sementes e interferir com o tempo necessário para a germinação, interagindo com o ácido abscísico e com a giberelina, que são hormônios vegetais (XI et al., 2010).

A longa região 3'-UTR (*untranslated region*, ou região não traduzida) do MFT tem várias características que lhe garantem papel regulatório pós-transcricional (Figura 9). Além da presença do elemento repetitivo (TA)<sub>16</sub>, é alvo de *small RNAs* (*smRNAs*) preditos e homólogos a siRNAs identificados em *A. thaliana*, e possui 27 tetranucleotídeos identificados como PREs (*putative recognition elements*), ou seja, elementos importantes para o reconhecimento e interação entre proteínas e RNAs que compõem complexos de regulação pós-transcricional, como microRNAs, por exemplo. Estas três características foram definidas pela análise de regiões não traduzidas em seis famílias de dicotiledôneas, e as sequências consenso identificadas estão supostamente preservadas há 70 milhões de anos, apesar de não codificarem produtos proteicos. Além disso, essa classe de regulação da expressão gênica é frequentemente associada com ORFs principais que codificam fatores de transcrição (VAUGHN et al., 2012), neste caso o *MFT*. Variações no bloco de repetições TA propriamente dito, e especialmente na sua extremidade 3' (Figura 9), poderiam influenciar a maturação do transcrito, porque sequências consenso para os sinais distais de poliadenilação variam em torno do tema UUUGUA e sinais proximais assemelham-se, bastante livremente, a AAUGAA. A eficiência da poliadenilação, o comprimento e a idade das caudas poli(A) são diretamente relacionadas com a estabilidade e funcionalidade dos transcritos (GRIERSON; COVEY, 1988; ROTHNIE, 1996).

Como alvo de smRNA, a expressão do *MFT* poderia ser regulada no núcleo por interferência com o processo de poliadenilação. A regulação também poderia se dar no citoplasma, por inibição da tradução, à semelhança do que foi registrado para o gene *SPL3* de *A. thaliana* (GANDIKOTA et al., 2007). A repressão coordenada da expressão do *MFT* e da malato desidrogenase citosólica por um mesmo smRNA (Figura 9) poderia assegurar a síntese e o acúmulo de sacarose em detrimento da glicólise, enquanto o embrião está dormente e o fruto é dreno de fotossintatos. A expressão da malato desidrogenase deve ser inibida na transcrição, talvez por metilação e heterocromatinização, porque nesse gene o sítio de acoplamento do smRNA fica na região codificadora da enzima.

```

AAA TAT ATC TTC TCT CTT GTG TTT TAG CCT CTT TGC TCT CTG TTT GTG TTT GTA ATG 57
GGA TCA ATG GGA ACA CCG GTG GAT CCA TTA GTG GTG GGA AGA ATA ATC GAC GAA GTA 114
      M G T P V D P L V V G R I I D E V 17
GTC GAT TTG TTC ATG CCT AGT GTG AAC ATG AGT GCC TAC TTT AAG AAC AAA CAC GTC 171
V D L F M P S V N M S A Y F K N K H V 36
ACG AAC GGA TGC GAC ATA AAG CCG TCC ATG GCT TCT GAG CCT CCC CGG ATC ACT TTT 228
T N G C D I K P S M A S E P P R I T F 55
TCT GGG AAA CCT ATG GAT CTC TAC ACA CTG ATC ATG ACC GAC CCT GAT GCG CCC AGT 285
S G K P D E L Y T L I M T D P D A P S 74
CCT AGC GAG CCC AGC ATG AGA GAA TGG ATC CAT TGG ATC GTG GTT GAC ATA CCA GGG 342
P S E P S M R E W I H W I V V D I P G 93
GGA GCT AGT CCT TCT CAA GGT ATT GAA ATT CTG CCG TAC ATG GCT CCC AAA CCC CCA 399
G A S P S Q G I E I L P Y M A P K P P 112
ATT GGC ATC CAC CGG TTT ATT CTG ATA CTG TTC AAA CAG AAG TCG AGG CTG GGG ATG 456
I G I H R F I L I L F K Q K S R L G G M 131
ATG GAA CCC CCG ATG AAC CGC GCG AAT TTC AAC ACA CGG TTC TTT TCA GCG CAG TTG 513
M E P P M N R A N F N T R F F S A Q L 150
GAC CTG GGC CTE CCA GCG GCA TTA GTG TAT TTC AAC GCA CAT AAA GAG CCA GCA AAC 570
D L G L P A A L V Y F N A H K E P A N 169
AGG CAG TCA CAG GCT AGG AGG ACA TAG TCC CAA AAA TTG AAA GCA CCT GAT CAG CTT 627
R Q S Q A R R T * 177
CTC TGC TAC TAT CAC TAA TGA TGC CAG GCG TAT ATA TAT ATA TAT ATA TAT ATA TAT 684
ATA TAT TG ATA ATG TAT ACT CAT GGA GAG GGG CGG TAT TCT CTC GGT GCT CTA TAC 741
CAT TGA GTT CTC CTG ATC GGG TCG GGT TAT TTG TCC TTA TTT GTG TTT TGG TCT TGT 798
GTT GTC TGT TCG GAT CGT GTT GTC TCT TCG GGT CGA GTC GTT TCA Gac CAA Gat GtA 855
ATC TTT TCT TTG AGT ATG GTG TGT CCT TTG TAA TGA ACT TCT AGC TTA TTA AGT TGT 912
ATG TAG GTG AAA CTA GTT TGG TTT CTT TTA TGT ATC TGG TGG TTT CCT CTA aaa aa 968

```

**Figura 9.** Contig 660 do banco de dados de ESTs de frutos com sementes do guaranazeiro, onde está localizado o bloco de repetições de dinucleotídeos TA (BOX) do *locus* GRN07. Os aminoácidos, numerados em negrito e itálico, na proteína homóloga a “*mother of flowering T*” estão indicados sob a sequência de nucleotídeos no cDNA. Três sítios de *splicing* hipotéticos estão representados por ▼ e foram posicionados por comparação com os homólogos em *A. thaliana*. Nucleotídeos da calda poli(A) estão representados por letras minúsculas, 365 posições depois do sítio de parada de tradução. Os elementos regulatórios na longa região 3'-UTR do transcrito estão indicados por cores. Em azul: consenso para interação com fator de transcrição EIN3 ativado por etileno (SOFTBERRY). Em vermelho: alvo para smRNA homólogo em *A. thaliana* (LEE et al., 2012) e também encontrado na sequência codificadora da malato desidrogenase citosólica, o que é indício de regulação coordenada de genes envolvidos em processos complementares. Em verde: tetranucleotídeos cuja presença é altamente enriquecida em regiões 3'-UTR relacionadas com regulação via smRNAs (VAUGHN et al., 2012). Sinais de poliadenilação (ROTHNIE, 1996) hipotéticos estão sublinhados, em negrito e itálico e muitas vezes sobrepostos a outros elementos regulatórios.

Por outro lado, o acoplamento de smRNAs ou outras proteínas poderia, por exemplo, garantir que a 3'-UTR estivesse acessível à RNA polimerase II para funcionar também como origem de microRNAs, mesmo quando a região codificadora principal (*MFT*) estivesse metilada e em estado heterocromático, talvez em outros órgãos da planta. Há pelo menos um pré-miRNA predito para essa região, com início no nucleotídeo 614 e término na posição 733, que poderia gerar o miRNA maduro 5'-UGCCAGGCGUAAUAUAUAU-3'. Além disso, o bloco de repetições tem potencialidade prevista para funcionar como um TATA box no contexto do DNA e ativar a transcrição de smRNAs a 3'.

No contig 660, além do (AT)<sub>n</sub> incluído no *locus* GRN07, há vários blocos com duas ou três repetições apenas, como (CT)<sub>2</sub> e (AC)<sub>3</sub>, e outros que são monoblocos de As e Ts nas regiões codificadoras do *MFT*. Regiões repetitivas evoluem em taxa mais acelerada do que o restante do genoma (FELSENSTEIN, 2004b), e alguns tipos de variações em blocos de dinucleotídeos produzem perdas de aminoácidos, mas não alteram a *frame* de leitura dos códons subsequentes (Fig. 10).

AAA	CCT	GAT	GAG	CTC	TAC	ACA	CTG	ATC	ATG	ACC	GAC	CCT	GAT		
K	P	D	E	L	<u>Y</u>	T	L	I	M	T	D	P	D		
AAA	CCT	GAT	GAG	CTC		ACA	CTG	ATC	ATG	ACC	GAC	CCT	GAT	GCG	
K	P	D	E	L		<u>T</u>	L	I	M	T	D	P	D	A	
AAA	CCT	GAT	GAG	CTC			CTG	ATC	ATG	ACC	GAC	CCT	GAT	GCG	CCC
K	P	D	E	L			L	I	M	T	D	P	D	A	P

**Figura 10.** Modelos de variantes passíveis de ocorrência na região codificadora do homólogo de *flowering locus T* no guaranazeiro. Na 3' – UTR desse gene está alocado o *locus* GRN07. O bloco de dinucleotídeos composto que está destacado em tons de cinza potencializa a deleção de grupos de três nucleotídeos, em frequência mais alta do que a esperada para regiões não repetitivas, sem que os códons subsequentes fiquem alterados. Pequenas variações como essas podem passar por artefatos de genotipagem quando há a expectativa de observar apenas variantes múltiplas de dois, mas também podem tornar-se responsáveis por pequenas alterações na função dos fatores de transcrição que são codificados por um gene e conseqüentemente pela geração de padrões de florescimento levemente diferentes.

Por outro lado, o acoplamento de smRNAs ou outras proteínas poderia, por exemplo, garantir que a 3´-UTR estivesse acessível à RNA polimerase II para funcionar também como origem de microRNAs, mesmo quando a região codificadora principal (*MFT*) estivesse metilada e em estado heterocromático, talvez em outros órgãos da planta. Há pelo menos um pré-miRNA predito para essa região, com início no nucleotídeo 614 e término na posição 733, que poderia gerar o miRNA maduro 5´-UGCCAGGCGUAAUAUAUAU-3´. Além disso, o bloco de repetições tem potencialidade prevista para funcionar como um TATA box no contexto do DNA e ativar a transcrição de smRNAs a 3´.

No contig 660, além do (AT)<sub>n</sub> incluído no *locus* GRN07, há vários blocos com duas ou três repetições apenas, como (CT)<sub>2</sub> e (AC)<sub>3</sub>, e outros que são monoblocos de As e Ts nas regiões codificadoras do *MFT*. Regiões repetitivas evoluem em taxa mais acelerada do que o restante do genoma (FELSENSTEIN, 2004b), e alguns tipos de variações em blocos de dinucleotídeos produzem perdas de aminoácidos, mas não alteram a *frame* de leitura dos códons subsequentes (Fig. 10).

AAA	CCT	GAT	GAG	CTC	TAC	ACA	CTG	ATC	ATG	ACC	GAC	CCT	GAT		
K	P	D	E	L	<u>Y</u>	T	L	I	M	T	D	P	D		
AAA	CCT	GAT	GAG	CTC		ACA	CTG	ATC	ATG	ACC	GAC	CCT	GAT	GCG	
K	P	D	E	L		<u>T</u>	L	I	M	T	D	P	D	A	
AAA	CCT	GAT	GAG	CTC			CTG	ATC	ATG	ACC	GAC	CCT	GAT	GCG	CCC
K	P	D	E	L			L	I	M	T	D	P	D	A	P

**Figura 10.** Modelos de variantes passíveis de ocorrência na região codificadora do homólogo de *flowering locus T* no guaranazeiro. Na 3´ – UTR desse gene está alocado o *locus* GRN07. O bloco de dinucleotídeos composto que está destacado em tons de cinza potencializa a deleção de grupos de três nucleotídeos, em frequência mais alta do que a esperada para regiões não repetitivas, sem que os códons subsequentes fiquem alterados. Pequenas variações como essas podem passar por artefatos de genotipagem quando há a expectativa de observar apenas variantes múltiplas de dois, mas também podem tornar-se responsáveis por pequenas alterações na função dos fatores de transcrição que são codificados por um gene e conseqüentemente pela geração de padrões de florescimento levemente diferentes.

Esses blocos com poucas repetições podem funcionar como *hot spots* de variabilidade. Interessante notar que as variantes em blocos de dinucleotídeos que produzem alterações desse tipo podem ser interpretadas como artefatos quando se esperam diferenças entre alelos limitadas a adições/subtrações de duas bases. Por outro lado, invariavelmente nos modelos em que foram testadas, a eliminação ou adição de dinucleotídeos ou mononucleotídeos provocaram a formação de codons de parada e destruição das ORFs. Plantas com genes silenciados não seriam obrigatoriamente inviáveis ou fenotipicamente muito distintas, pois uma das vantagens de manter um genótipo com grau de poliploidia tão alto pode estar no tamponamento da interferência desse tipo de variação sobre o fenótipo das plantas. Mas os mecanismos que induzem alterações de duas e de três bases como *slippage* ou *crossing over* desigual entre cromossomas homólogos ou homeólogos são teoricamente os mesmos. Mais uma razão para avaliar com cuidado quais picos devem ser eliminados dos cromatogramas de genotipagem. Manter picos que diferem entre si por uma base, ainda que os *locus* avaliados contenham blocos de dinucleotídeos, como foi feito no presente trabalho, especialmente quando a espécie em foco for poliploide, pode ter sido a providência mais correta para registrar toda a variabilidade presente no genoma do guaranazeiro.

Blocos com poucas repetições podem ser utilizados no desenvolvimento de *loci* candidatos a marcadores para o padrão de germinação de sementes de cultivares clonais. No entanto, cada um deles isolado dificilmente conterá variantes (alelos) em número suficiente para análise de diversidade. Além disso, em um genoma tão replicado como o do guaranazeiro pode ser difícil avaliar o valor fenotípico de variantes alélicas raras, que podem estar presentes em apenas uma das cópias do gene em um dos cromossomas de cada conjunto de homólogos e homeólogos.

Variações nos *loci* GRN07, GRN13 e 14, que estão na 3´-UTR, podem estar ligadas com fenótipos diferentes quanto à arquitetura das plantas ou mesmo a diferenças em padrões de germinação de sementes e de

resistência a doenças (CUBAS et al., 1999; TAIZ; ZEIGER, 2010), se, por exemplo, padrões de alelos nesses *loci* tiverem sido fixados nas mesmas combinações de cromossomas homólogos e homeólogos que sejam características de cultivares com diferentes arquiteturas ou resistência a doenças.

Além dos iniciadores desenhados e sintetizados especificamente para o guaranazeiro, foram testados 12 pares de iniciadores desenvolvidos para *Litchi chinensis*, que também é planta da família Sapindaceae, clado Nephelieae (HARRINGTON et al., 2005). Os iniciadores foram desenvolvidos por Viruel e Hormaza (2004). Segundo os autores, quatro deles apresentaram transferibilidade para *longan* (*Euphoria longan*), que pertencem ao mesmo clado da lychia. Os experimentos foram realizados com a inclusão de indivíduos de *rambutan*, outra Nephelieae, o que contribuiu para avaliar os resultados. Apenas três pares entre os doze – LMLY5, LMLY7 e LMLY8 – geraram produtos de amplificação visualizados em agarose (ANGELO, 2007), mas os padrões de segregação dos fragmentos gerados com o DNA de sete plantas de guaranazeiro, quando observados em géis de poliacrilamida desnaturante a 6%, mostraram-se pouco reprodutíveis e muito complexos. Os resultados para o rambutan foram reprodutíveis, mas apenas um alelo foi observado em cada um dos três *loci*. Esses resultados podem ser consequência da distância evolutiva entre Nephelieae e Paullinieae.

Além dos *loci* que foram utilizados para o trabalho de genotipagem aqui descrito, foram desenvolvidos iniciadores para outros 17. O GRN02, o GRN10 e o GRN16 e alguns outros geraram apenas um alelo para todas as plantas genotipadas, como o GRN13. O GRN04 mostrou-se monoalélico e monomórfico nas genotipagens aqui analisadas, mas em genotipagens realizadas anteriormente com iniciadores marcados individualmente, quando a intensidade da fluorescência nos cromatogramas foi mais que dez vezes maior, foi registrado como multialélico e polimórfico (ANGELO, 2007; Anexo II). Como registrado

acima, o GNR03 foi multialélico e monomórfico, tendo todos os indivíduos genotipados em experimentos anteriores apresentado os mesmos três alelos (vide Anexo II). Resultados como este podem ser úteis em trabalhos futuros com cruzamentos controlados ou citogenética com marcação *in situ*. Com relação a genotipagens anteriores (ANGELO, 2007), foi registrada uma diferença de cerca de seis pares de bases para todos os alelos nos *loci* GRN01 e GRN05, mas o número e padrão de intensidade dos picos preservados para as análises nas mesmas cultivares, segundo sua ordenação por tamanho nos cromatogramas, foram os mesmos. Isso foi interpretado como consequência da utilização de *primer tail* em uma ocasião e de marcação direta de um dos iniciadores de cada par, na outra. Três pares de iniciadores (GRN06, GRN11 e GRN21) não geraram fragmentos de amplificação. Os outros 12 pares (anexo II) podem ser testados com pelo menos um iniciador de cada par marcado individualmente, porque produziram fragmentos de amplificação, mas para a maior parte deles não foi possível encontrar temperaturas adequadas para a incorporação da fluorescência e detecção pelo sequenciador.

## Conclusões

- No guaranazeiro, o número de *locus* contendo blocos de dinucleotídeos repetidos oito ou mais vezes é menor do que na maioria das espécies vegetais;
- A utilização de *primer tail* não foi ideal para alcançar os resultados esperados com o conjunto de *loci* descritos neste trabalho;
- A utilização do critério de reprodutibilidade contribuiu para a valorização dos alelos com poder resolutivo; e
- Os iniciadores (*primers*) descritos neste trabalho são úteis para distinguir cultivares clonais de guaranazeiro e os *fingerprintings* gerados podem ser utilizados como características descritivas adicionais.

## Referências

ANGELO, P. C. S. **Bioteecnologias para o guaranazeiro (*Paullinia cupana* var. *sorbilis*): marcadores microssatélite e propagação in vitro**. Relatório final de projeto FAPEAM 924/2003. Manaus, 2007.

ANGELO, P.; NUNES-SILVA, C.; BRÍGIDO, M.; AZEVEDO, J.; ASSUNÇÃO, E.; SOUSA, A.; PATRÍCIO, F.; REGO, M.; PEIXOTO, J.; OLIVEIRA-JR, W.; FREITAS, D.; ALMEIDA, E.; VIANA, A.; SOUZA, A.; ANDRADE, E.; ACOSTA, P.; BATISTA, J.; WALTER, M.; LEOMIL, L.; ANJOS, D.; COIMBRA, R.; BARBOSA, M.; HONDA, E.; PEREIRA, S.; SILVA, A.; PEREIRA, J.; SILVA, M.; MARINS, M.; HOLANDA, F.; ABREU, R.; PANDO, S.; GONÇALVES, J.; CARVALHO, M.; LEAL-MESQUITA, E.; SILVEIRA, M.; BATISTA, W.; ATROCH, A.; FRANÇA, S.; PORTO, J.; SCHNEIDER, M.; ASTOLFI-FILHO, S. Guarana (*Paullinia cupana* var. *sorbilis*), an anciently consumed stimulant from the Amazon rain forest: the seeded-fruit transcriptome. **Plant Cell Reports**, v. 27, p. 117-124, 2008.

ATROCH, A. L.; NASCIMENTO-FILHO, F. J.; ANGELO, P. C. S.; FREITAS, D.V.; SOUSA, N. R.; RESENDE, M. D. V.; CLEMENT, C. R. Domesticação e melhoramento de guaranazeiro. In: BORÉM, A.; LOPES, M. T. G.; CLEMENT, C. R. (Ed.). **Domesticação e melhoramento: espécies amazônicas**. Viçosa, MG: Universidade Federal de Viçosa, 2009. p. 333–361.

AZEVEDO, V. C. R.; VINSON, C. C.; SILVA, V. P. da; CIAMPI, A. Y. **Desenvolvimento e aplicação de marcadores microssatélites em Maçaranduba (*Manilkara huberi*)**. Brasília, DF: Embrapa Recursos Genéticos e Biotecnologia, 2003. 3 p. (Embrapa Recursos Genéticos e Biotecnologia. Circular técnica, 25).

BUERKI, S.; FOREST, F.; ACEVEDO-RODRÍGUEZ, P.; CALLMANDER, M. W.; NYLANDER, J. A.; HARRINGTON, M.; SANMARTÍN, I.; KÜPFER, P.; ALVAREZ, N. Plastid and nuclear DNA markers reveal intricate relationships at subfamilial and tribal levels in the soapberry family (Sapindaceae). **Molecular Phylogenetics and Evolution**, v. 51, p. 238–258, 2009.

BRONDANI, R. P. V.; BRONDANI, C.; TARCHINI, R.; GRATTAPAGLIA, D. Development, characterization and mapping of microsatellite markers in *Eucalyptus grandis* and *E. urophylla*. **Theoretical and Applied Genetics**, v. 97, p. 816-27, 1998.

CARRIERO, F.; FONTANAZZA, G.; CELLINI, F.; GIORIO, G. Identification of simple sequence repeats (SSR) in olive (*Olea europaea* L.). **Theoretical and Applied Genetics**, v. 104: 301-307, 2002.

COLLEVATTI, R. G.; BRONDANI, R. V.; GRATTAPAGLIA, D. Development and characterization of microsatellite markers for genetic analysis of a Brazilian endangered tree species *Caryocar brasiliense*. **Heredity**, v. 83, p. 748-56, 1999.

CUBAS, P.; LAUTER, N.; DOEBLEY, J.; COEN, E. The TCP domain: a motif found in proteins regulating plant growth and development. **The Plant Journal**, v. 18, p. 215-222, 1999.

FARIAS, I. P.; HRBEK, T.; BRINKMANN, H.; SAMPAIO, I.; MAYER, A. Characterization and isolation of DNA microsatellite *primers* for *Arapaima gigas*, an economically important but severely over-exploited fish species of the Amazon basin. **Molecular Ecology Notes**, v. 3, p. 128-130, 2003.

FELSENSTEIN, J. **PHYLIP (Phylogeny Inference Package) version 3.6**. Seattle: Department of Genome Sciences, University of Washington, 2004.

FELSENSTEIN, J. **Inferring phylogenies**. s.l.: Sinauer Associates, 2004b.

FERREIRA, M.E.; GRATTAPAGLIA, D. **Introdução ao uso de marcadores moleculares em análise genética**. EMBRAPA-CENARGEN. 2.ed. 1996. Brasília: EMBRAPA-CENARGEN, 1996. 220p.

FREITAS, D. V.; CARVALHO, C. R.; NASCIMENTO-FILHO, F. J.; ASTOLFI-FILHO, S. Karyotype with 210 chromosomes in guarana (*Paullinia cupana 'sorbilis'*). **Journal of Plant Research**, v. 120, p. 399-404, 2007.

GANDIKOTA, M.; BIRKENBIHL, R. P.; HÖHMANN, S.; CARDON, G. H.; SAEDLER, H.; HUIJSER, P. The miRNA156/157 recognition element in the 3'-UTR of the Arabidopsis SBP box gene SPL3 prevents early flowering by translational inhibition in seedlings. **The Plant Journal**, v. 49, p. 683-693, 2007.

GAUTSCHI, O.; TSCHOPP, S.; OLIE, R. A.; LEECH, S. H.; SIMOES-WUST, A. P.; ZIEGLER, A. Activity of a novel bcl-2/bcl-xL-bispecific antisense oligonucleotide against tumors of diverse histologic origins. **Journal of The National Cancer Institute**, v. 93, p. 463-71, 2001.

GRATTAPAGLIA, D. Marcadores moleculares em espécies florestais: *Eucalyptus* como modelo. In: NASS, L. L.; VALOIS, A. C. C.; MELO, I. T.; VALADARES-INGLIS, M. C. (Ed.) **Recursos genéticos e melhoramento**: plantas. Rondonópolis: Fundação MT, 2001. p. 967-993.

GRIERSON D, COVEY SN. **Plant Molecular Biology**. Blackie & Son - Chapman & Hall, USA. 1988. pp. 22-46.

HARRINGTON, M. G.; EDWARDS, K. J.; JOHNSON, S. A.; CHASE, M. W.; GADEK, P. A. Phylogenetic inference in Sapindaceae sensu lato using plastid mat K and rbcL DNA sequences. **Systematic Botany**, v. 30, p. 366-382, 2005.

HODGETTS, R. B.; ALEKSIUK, M. A.; BROWN, A.; CLARKE, C.; MacDONALD, E.; NADEEN, S.; KHASA, D. Development of microsatellite markers for white spruce (*Picea glauca*) and related species. **Theoretical and Applied Genetics**, v. 102, p. 1252-1258, 2001.

HUANG, W. G.; CIPRIANI, G.; MORGANTE, M.; TESTOLIN, R. Microsatellite DNA in *Actinidia chinensis*: isolation, characterization and homology in related species. **Theoretical and Applied Genetics**, v. 97, p. 1269-78, 1998.

IVANDIC, V.; HACKETT, C. A.; NEVO, E.; KEITH, R.; THOMAS, W. T. B.; FORSTER, B. P. Analysis of simple sequence repeats (SSRs) in wild barley from the Fertile Crescent: associations with ecology, geography and flowering time. **Plant Molecular Biology**, v. 48, p. 511-527, 2002.

LEE, T.-F.; GURAZADA, S. G. R.; ZHAI, J.; LI, S.; SIMON, S. A.; MATZKE, M. A.; CHEN, X.; MEYERS, B. C. RNA polymerase V-dependent small RNAs in *Arabidopsis* originate from small, intergenic loci including most SINE repeats. **Epigenetics**, v. 7, n. 7, p. 781-795, Jul. 2012.

LIU, Z.-W.; BIYASHEV, R.; SAGHAI MAROOF, M. Development of simple sequence repeat DNA markers and their integration into a barley linkage map. **Theoretical and Applied Genetics**, v. 93, p. 869-876, 1996.

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

MORGANTE, M.; HANAFEY, M.; POWELL, W. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. **Nature Genetics**, v. 30, p. 194-200, 2002.

NEI, M.; LI, W.-H. Mathematical model for studying genetic variation in terms of restriction endonucleases. **Proceedings of the National Academy of Sciences**, v. 76, p. 5269-5273, 1979.

NYBOM, H.; ESSELINK, G.; WERLEMARK, G.; VOSMAN, B. Microsatellite DNA marker inheritance indicates preferential pairing between two highly homologous genomes in polyploid and hemisexual dog-roses, *Rosa L. Sect. Caninae DC.* **Heredity**, v. 92, p. 139–150, 2004.

PASTINA, M. M.; MALOSETTI, M.; GAZAFFI, R.; MOLLINARI, M.; MARGARIDO, G. R. A.; OLIVEIRA, K. M.; PINTO, L. R.; SOUZA, A. P.; EEUWIJK, F. A. V.; GARCIA, A. A. F. A mixed model QTL analysis for sugarcane multiple-harvest-location trial data. **Theoretical And Applied Genetics**, v. 124, 835-849, 2012.

PERERA, L.; RUSSEL, J. R.; PROVAN, J.; POWELL, W. Identification and characterization of microsatellite *loci* in coconut (*Cocos nucifera* L.) and the analysis of coconut population in Sri Lanka. **Molecular Ecology**, v. 8, p. 335-346, 1999.

PERTEA, M.; MOUNT, S. M.; SALZBERG, S. L. A computational survey of candidate exonic splicing enhancer motifs in the model plant *Arabidopsis thaliana*. **BMC Bioinformatics**, v. 8, p. 159, 2007.

RALLO, P.; DORADO, G.; MARTÍN, A. Development of simple sequence repeats (SSRs) in olive tree (*Olea europaea* L.). **Theoretical and Applied Genetics**, v. 101, p. 984-989, 2000.

RAMCHIARY, N.; NGUYEN, V. D.; LI, X.; HONG, C. P.; DHANDAPANI, V.; CHOI, S. R.; YU, G.; PIAO, Z. Y.; LIM, Y. P. Genic microsatellite markers in *Brassica rapa*: development, characterization, mapping, and their utility in other cultivated and wild Brassica relatives. **DNA Research**, v. 18, p. 305-320, 2011.

ROTHNIE, H. M. Plant mRNA 3'-end formation. **Plant Molecular Biology**, v. 32, p. 43-61, 1996.

SILVA, J. A. G. Preliminary analysis of microsatellite markers derived from sugarcane expressed sequence tags (ESTs). **Genetics and Molecular Biology**, v. 24, p. 155-159, 2001.

SOLIS-NEFFA, V. G.; FERRUCCI, M. S. Karyotype analysis of some *Paullineae* species (Sapindaceae). **Caryologia**, v. 54, p. 371-376, 2001.

SOLIVA, M.; GAUTSCHI, B.; SALZMANN, C.; TENZER, I.; WIDMER, A. Isolation and characterization of microsatellite *loci* in the orchid *Ophrys araneola* (Orchidaceae) and a test of cross-species amplification. **Molecular Ecology**, v. 9, p. 2178-2179, 2000.

SOUSA, N. R. Variabilidade genética e estimativa de parâmetros genéticos em germoplasma de guaranazeiro. 2003. 99 f. Tese (Doutorado)- Universidade Federal de Lavras, Lavras. 2003.

TAIZ, L.; ZEIGER, E. **Plant physiology**. USA: Sinauer Associates, 2010. p. 583-619.

VIRUEL, M. A.; HORMAZA, J. L. Development, characterization and variability analysis of microsatellites in lychee (*Litchi chinensis* Sonn., Sapindaceae). **Theoretical and Applied Genetics**, v. 108, p. 896-902, 2004.

VAUGHN, J. N.; ELLINGSON, S. R.; MIGNONE, F.; ARNIM, A. V. Known and novel post-transcriptional regulatory sequences are conserved across plant families. **RNA**, v. 18, 368-384, 2012.

XI, W.; LIU, C.; HOU, X.; YU, H. MOTHER OF FT AND TFL1 Regulates Seed Germination through a Negative Feedback Loop Modulating ABA Signaling in Arabidopsis. **The Plant Cell**, v. 22, p. 1733-1748, 2010.

## Anexo I

INDIVÍDUO	861	648	850	388	626	189	608	624	505	612	871	882	610	611	372
L1116	0	0	0	0	0	0	0	0	0	0	0	0	0	0	?
L1118	1	0	0	0	0	1	0	1	1	0	0	1	1	1	1
L1120	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
L1122-123	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
L4236	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
L5252	1	1	1	1	1	1	0	0	0	0	1	1	1	1	0
L253	0	?	?	0	?	?	0	0	0	0	0	?	?	?	?
L5255	1	1	1	1	1	1	0	1	1	?	1	1	1	1	1
L5256	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
L5257	1	1	0	0	1	0	0	1	1	?	0	0	1	1	1
L7191-204	?	0	0	0	0	0	0	0	0	0	0	-	0	0	-
L7244	1	0	0	0	0	0	0	1	0	0	0	-	1	1	-
L7246	0	1	0	1	1	1	0	1	0	1	1	-	0	1	-
L7248	1	1	1	1	1	1	1	1	1	1	1	-	1	1	-
L7250	1	0	0	1	1	1	1	0	1	0	1	-	1	1	-
L7262	0	?	?	0	0	0	?	0	0	0	0	-	0	0	-
L7264	0	?	?	0	0	0	?	0	0	0	0	-	0	0	-
L7266	0	?	?	0	0	0	1	0	0	0	0	-	0	0	-
L9225	0	?	0	0	0	0	?	?	?	?	?	?	?	?	-
L9244	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
L9246	0	0	1	1	1	1	1	1	1	1	?	1	1	1	-
L9247-248	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
L9252	0	0	0	?	0	0	?	?	0	?	0	0	0	?	-
L12259	1	?	1	?	?	0	0	0	1	1	0	0	0	0	0
L12261	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
L12263	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
L12265	1	1	1	?	1	1	0	1	1	0	?	0	1	1	1
L12267	0	?	0	0	1	0	0	1	0	0	0	0	0	0	0
L13263	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
L14202	0	0	0	1	0	0	1	1	0	0	0	0	0	0	?
L14204	0	0	0	0	1	0	1	1	1	0	0	0	?	0	1
L14205	0	0	0	0	0	0	0	0	?	0	0	0	?	0	0
L14206	1	0	1	1	0	0	1	0	1	1	1	1	0	0	1
L14207	0	?	0	1	0	1	0	0	0	0	0	0	0	0	0
L14208	1	1	1	0	0	0	0	?	1	?	1	1	1	1	1
L14209	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0
L14210	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
L14212	1	1	1	0	1	1	0	1	1	0	0	1	1	0	1
L14213	0	0	0	1	0	0	?	0	0	0	0	1	0	0	0
L14214	0	0	0	0	0	0	1	1	0	1	1	1	1	0	1
L14215	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
L14217	0	0	0	0	1	1	0	1	0	0	0	0	1	0	1
L17182	1	0	0	0	0	0	0	0	0	0	1	0	-	1	?
L17186	0	0	0	0	1	0	0	0	0	1	0	0	-	0	0
L17188	1	1	1	1	1	1	1	1	1	1	1	1	-	1	1
L17190	0	0	0	0	?	0	0	1	0	0	1	?	-	0	1
L17191	0	0	0	0	0	?	0	0	0	0	0	0	-	0	0
L17192	-	1	1	1	1	1	1	1	1	1	1	1	-	1	1
L17194	-	1	1	1	1	1	1	1	1	1	?	1	-	1	0
L17196	-	1	1	1	1	1	1	1	1	1	1	?	-	1	0
L17198	1	?	-	1	1	1	0	0	1	1	?	-	-	-	0
L17204	1	0	-	0	1	-	1	1	-	0	-	-	-	1	0
L20230	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
L20232	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
L20235	1	?	1	1	1	?	1	0	0	1	1	1	1	0	1
L20238	1	?	1	1	0	1	0	1	1	?	1	1	1	1	1
L20241	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
L20244	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0
L20247	1	0	0	0	0	0	0	0	?	0	0	?	0	0	1

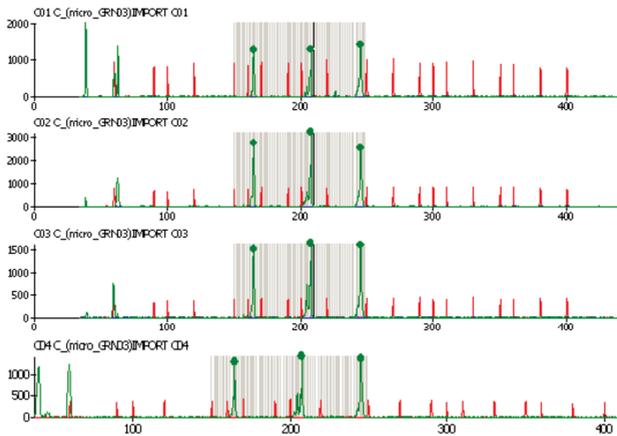
**OBS.:** 1 = presença; 0 = ausência; ? picos que necessitam réplica de verificação; - dado faltante. A primeira linha contém a designação das cultivares genotipadas e a primeira coluna os códigos dos alelos, sendo que os primeiros algarismos representam o *locus* e os três últimos o tamanho dos alelos, em pares de bases.

## Anexos II

Características de *iniciadores* ainda não utilizados para *loci* de microssatélites de guaranazeiro sequências dos iniciadores R e F, no sentido 5´-3´.

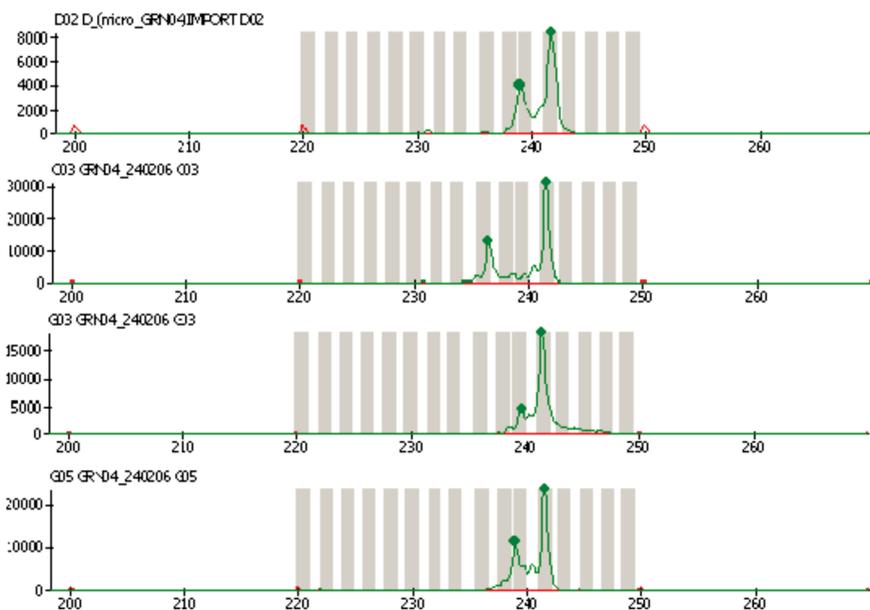
PAR	REPETIÇÕES	SEQUÊNCIAS DOS PRIMERS*	ANEL**
08	(TA) 19	GCCGAGTTTACCACGTCATT e GCCTAAAACATAACCTCAAAA	62
15	(AG) 7 (GC) 5	GCATTTTCATCTTTGTGGTTT e CAGTTCTAGGATTGTACGGC	55
18	(CTT) 8CC (TCT) 2	CCCTCCGGCGACTCCCCTTG e AAAATCATCAATCAGTGTGGTAGT	60
19	(AG) 9 (CTT) 4	CGCTGCCATGAAAACGAATAG e TGGGAGGAGAGAACCCTACA	59
22	(AT) 9	TTCAGTGGGTTTGAACCAT e GCCCACAACCTTCACAAAACC	64 OU 66
23	(AAG) 9	TAAATCCGACATGAGCGAGAC e TCATACTCTCCCTCGCCATT	57
24	(TC) 9	TAGGGTTTCTCTCGGTTTTTG e ATAAAACGTCGGCACTGGAG	62
25	(AT) 15	TGCTTCTCTGGTTGGGAGAG e TCACTACTTAGACCACCAAGC	62
26	(AT) 12	TTTGGCTTCACTCAACGAAAA e cattcataaaatgccagaaca	57
27	(AT) 12	TTGCTGGATCGGGTTAAAATC e CATTTAGAACAAACAGGACGA	57 OU 62
28	(TA) 3C (AT) 9	TCGCTCTCTGACATCCATCAA e tcaaacgcatactgctttcg	57

\*\* Temperatura de anelamento dos iniciadores.



GRN 03	(AC) 4T (TA) 3 (TG) 2	CGGGATTCTACAGACCGAAG e TCGCTTAGACCATGGAGAC	59
--------	-----------------------	--	----

Sequência dos iniciadores e genótipos de guaranazeiro para o *locus* GRN03 (ANGELO et al., 2007)



Genótipos de guaranazeiro para o *locus* GRN04 em genotipagens anteriores (ANGELO, 2007), com *primers* marcados individualmente.





**Embrapa**

---

*Amazônia Ocidental*

Ministério da  
**Agricultura, Pecuária  
e Abastecimento**

G O V E R N O F E D E R A L  
**BRASIL**  
PAÍS RICO É PAÍS SEM POBREZA