

Repetições de Dinucleotídeos em ESTs de Frutos e Sementes de Guaraná

*Natalia Dayane Moura Carvalho
Susan Karoline Barbosa Soares
Enedina Nogueira Assunção
Paula Cristina da Silva Angelo
Spartaco Sastolfi Filho
André Luiz Atroch*

Resumo

Loci microssatélite tem sido buscado em bibliotecas genômicas de guaranazeiro. Esta busca, no entanto, ainda não produziu resultados satisfatórios. O objetivo deste trabalho foi avaliar as repetições de dinucleotídeos em 5.969 unidades transcricionais (ESTs "singlets" and "contigs") de frutos e sementes de guaranazeiro mantidos no banco de dados da REALGENE e comparar com aqueles encontrados em bibliotecas genômicas enriquecidas para repetições AG/TC AC/TG. As seqüências (AC/TG)₈₀, (AG/TC)₈₀, (AT)₈₀ e (CG)₈₀ foram usadas como sondas para o BioEdit "local Blast". Aquelas que continham blocos de repetições perfeitos foram selecionadas. As freqüências encontradas para os diferentes tipos de repetições foram submetidas a ANOVA. O número de repetições - 7 a 21 - por bloco foi analisado pelo ². Essas análises revelaram que 1,91% das unidades transcricionais apresentavam blocos de repetições de dinucleotídeos. O tipo mais freqüente ($P < 0,001$) de repetição foi TA (0,97%), seguido por AG/TC (0,77%). Seqüências contendo blocos de 7 repetições (0,92%) foram as mais freqüentes. Não houve discrepância entre os resultados das buscas por microssatélites compostos por blocos de dinucleotídeos em bibliotecas genômicas e no banco de ESTs.

Termos para indexação: diversidade genética, marcadores moleculares, *Paullinia cupana*.

Dinucleotide repeats in ESTs from guarana seeded fruits

Abstract

Microsatellite *loci* have been searched in guarana plant genomic libraries. This search, nevertheless, has not yet produced satisfactory results. The aim of this work was to evaluate dinucleotide repeats in 5,969 guarana fruits and seeds transcriptional units (ESTs singlets and contigs) recorded in the REALGENE databank and to compare to the genomic survey. The sequences (AC/TG)₈₀, (AG/TC)₈₀, (AT)₈₀ and (CG)₈₀ were used like probes as queries to run BioEdit local Blast. Those containing perfect repeat blocks were selected. Recorded frequencies for different repeats were analyzed by ANOVA. The number of repeats - 7 to 21- per block was analyzed by the χ^2 test. These analyses revealed that 1.91% transcriptional units presented dinucleotide repeat blocks. The most frequent ($P < 0.001$) repeat was TA (0.97%) followed by AG/TC (0.77%). Sequences containing 7 repeats blocks (0.92%) were the most frequent. Results from the genomic libraries and the EST databank were similar.

Index terms: genetic diversity, molecular markers, *Paullinia cupana*.

Introdução

O guaranazeiro (*P. cupana* var. *sorbilis*) é uma Sapindaceae nativa da Floresta Amazônica, comercialmente cultivada somente no Brasil. Da produção anual de cerca de 4.000 toneladas 70% é consumido pelas indústrias de refrigerantes. O restante é vendido principalmente como pó (das sementes) para laboratórios, farmácias e outras lojas de produtos alternativos.

A Embrapa Amazônia Ocidental é a única instituição de pesquisa que tem mantido continuamente um Programa de Melhoramento e um banco de germoplasma da espécie desde a década de 70. Como parte deste Programa de Melhoramento do Guaranazeiro foram desenvolvidos os clones BRS e uma tecnologia para clonagem através de estacas. O banco de germoplasma contém 246 acessos e a diversidade genética entre 100 deles foi analisada utilizando marcadores RAPD. Marcadores microssatélite deverão contribuir no aprofundamento deste conhecimento, e, também, na avaliação de taxas de autofecundação.

Microssatélites são blocos compostos por repetições de 1 a 6 nucleotídeos em "tandem", ou seja, arrançados consecutivamente, encontrados no genoma de organismos eucariontes e procariontes. Em guaranazeiro, a busca por estes blocos em bibliotecas genômicas não tem produzido os resultados esperados. Bibliotecas genômicas enriquecidas para repetições AG/TC - AC/TG produziram 6,7% de fragmentos Sau3A1 contendo principalmente blocos com menos de 11 repetições e imperfeitos; destes, 75% eram complementares às sondas utilizadas e 25% eram repetições de TA. "Primers" específicos foram sintetizados para definir e testar cinco desses *loci*, mas geraram padrões complexos de polimorfismo, considerados associados às imperfeições nos blocos.

O objetivo deste trabalho foi avaliar os blocos de dinucleotídeos repetidos, presentes em 5.969 unidades transcricionais (ESTs "singlets" and "contigs") de frutos e sementes de guaraná, mantidos no banco de dados da Realgene (www.biomol.unb.br/GR/) e comparar com aqueles encontrados pelo "screening" de bibliotecas genômicas enriquecidas.

Material e Métodos

A busca por blocos de dinucleotídeos foi realizada utilizando seqüências (AC/TG)₈₀; (AG/TC)₈₀; (AT)₈₀ e (CG)₈₀ como sondas para a ferramenta "local Blast" do aplicativo BioEdit. Os blocos perfeitos de repetições (sem interrupções) foram selecionados e a significância estatística das diferenças na frequência dos quatro tipos de repetições foi analisada (Kruskal-Wallis ANOVA - Sigma Stat v.02). Os blocos foram organizados em classes de acordo com o número de repetições - 7 a 21 - e a significância estatística da distribuição foi testada pelo chi-quadrado (aplicativo Genes - Universidade Federal de Viçosa).

A similaridade entre os acessos foi calculada usando o Coeficiente de Similaridade Geral de Gower. Os acessos foram agrupados pelo método hierárquico das médias das distâncias (UPGMA - Unweighted Pair-Group Method Using an Arithmetic Average) e a análise da dispersão gráfica da similaridade entre eles realizada pelo método da Análise de Coordenadas Principais (PCA). As análises foram realizadas utilizando o programa computacional Multi-Variate Statistical Package (MVSP v.3.13).

Pares de "primers" flanqueadores de quatro dos microsatélites identificados em ESTs de frutos/sementes foram sintetizados e utilizados para amplificar o DNA de plantas de guaranazeiro. As reações de PCR ("polymerase chain reaction") foram realizadas conforme o esquema abaixo, em volume final de 25 μ l: 20 ng por reação de DNA; 4 pg de cada "primer"; 1 U de Taq polimerase; 2 mM de cloreto de magnésio.

O termociclador foi programado para realizar os seguintes ciclos: 95 °C, por 5 minutos; 35 x (95 °C, por 30 segundos; 55 °C, por 30 segundos; 72 °C, por 45 segundos); 72 °C, por 7 minutos.

Resultados e Discussão

A busca por repetições de dinucleotídeos *in tandem* no banco de ESTs (unidades transcricionais) de frutos/sementes de guaranazeiro revelou que 1,91% das seqüências apresentava microsatélites.

O tipo mais freqüente de repetição ($P < 0,001$) foi o TA (0,97%), seguido do AG/TC (0,77%); AC/TG (0,15%) e GC (0,02%) - Figura 1.

Blocos contendo 7 repetições (0,92%) foram mais freqüentes, seguidos por blocos com 8 (0,28%); 9 (0,18%); 12 (0,17%); 10 e 11 (0,12%) repetições. A freqüência de blocos em cada classe foi significativamente diferente do esperado ($P = 0,05$), com exceção para os blocos com 8 repetições. Um bloco (TA)₄₀ foi encontrado. Blocos com menos de 7 repetições não foram analisados.

A análise da seqüência de 960 clones produzidos pelo enriquecimento de bibliotecas genômicas com sondas (AG)₁₂ e (AC)₁₂ + (AG)₁₂ (AMADO et al., 2005; ANGELO et al., 2005) indicou que entre as 174 que apresentaram pelo menos 100 bases com qualidade (PHRED) acima de 20 não havia microsatélites com mais de 11 repetições. Das seqüências selecionadas da biblioteca enriquecida com (AG)₁₂ 6,1% continham microsatélites, sendo 75% deles do tipo esperado, ou seja, complementares a TC. Da biblioteca enriquecida com TC + AC 7,3% das seqüências continham microsatélites, sendo 75% destes do tipo esperado. Nos dois casos 25% dos arranjos eram de repetições TA, não esperadas.

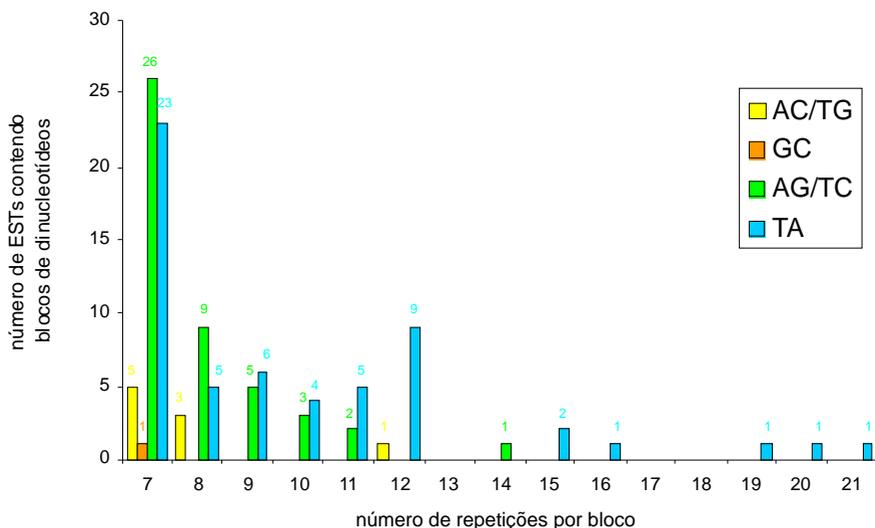


Fig. 1. Número de repetições de dinucleotídeos encontradas em ESTs de frutos e sementes de guaraná (Embrapa Amazônia Ocidental, 2006).

Quatro dessas seqüências, ainda que apresentassem como “núcleos” blocos com pequeno número de repetições (interrompidas ou não), foram utilizadas como molde para a síntese de “primers” específicos. A genotipagem de plantas de seis clones de guaranzeiro, incluindo clones selecionados de progênes diferentes, revelou no máximo três alelos por *locus*, pouca diversidade entre indivíduos e padrões pouco comuns como a existência de pelo menos três alelos em todos os indivíduos genotipados ou padrões muito complexos, com um mesmo indivíduo apresentando até cinco possíveis alelos por *locus* (AMADO et al., 2005; ANGELO et al., 2005), ou seja, pouco úteis para análise de diversidade e de taxas de autofecundação entre clones.

No entanto, o enriquecimento com TC e AC foi bem sucedido, por exemplo, para eucalipto (BRONDANI et al., 1998), “kiwi” (HUANG et al., 1998), coqueiro (PERERA et al., 1999); oliveira (CARRIERO et al., 2002), pequi (COLLEVATTI et al., 1999), carvalho (HODGETTS et al., 2001) e para *Ophrys araneola*, uma Orchidaceae (SOLIVA et al., 2000). Em eucalipto, “kiwi”, coqueiro e oliveira foram encontrados microssatélites com números mínimos de 15, 8, 13 e 9 repetições do dinucleotídeo componente do “núcleo”, respectivamente, e número máximo sempre superior a 20 repetições, em arranjos perfeitos, e,

também, arranjos compostos. Para guaranazeiro não houve diferença significativa entre o número de clones positivos selecionados pelas duas estratégias de enriquecimento (utilização de sonda AG ou de sondas AG + AC, em conjunto); e uma frequência muito baixa de microssatélites com mais de sete repetições (tanto para arranjos perfeitos quanto para arranjos compostos) foi encontrada.

Dificuldade semelhante parece ter ocorrido ao longo do processo de desenvolvimento de microssatélites para a cana-de-açúcar, espécie poliplóide para a qual a busca por microssatélites em ESTs foi relativamente mais profícua do que a análise de bibliotecas enriquecidas (SILVA, 2001). Há indícios de que o guaranazeiro também é planta poliplóide e a REALGENE dispõe de um banco de ESTs que, embora pequeno, foi analisado em busca de blocos de repetições perfeitos.

Os blocos de TA e de AG/TC foram prevalentes nos dois tipos de experimentos de "screening" do genoma do guaranazeiro, como o foram também no genoma de *Arabidopsis thaliana*. No genoma do arroz foram encontrados mais blocos de GA seguidos por blocos de TA. A busca por microssatélites para essas duas últimas espécies foi feita tanto nos exons como em introns e regiões 3' e 5', não traduzidas. Repetições de dinucleotídeos foram encontradas principalmente nas regiões 5' não traduzidas das duas espécies, em frequência muito baixa nos exons e inferior à expectativa no exons de *A thaliana*, e não há ainda resultado publicado referente à variabilidade destes blocos de repetições identificados (LAWSON & ZHANG, 2006). Em humanos, o "screening" de 13.783 seqüências codificadoras de proteínas revelou em torno de 6,49 blocos de repetições por seqüência, sendo, destes, 1,58% considerados polimórficos com relação ao número de repetições.

Os pares de "primers" para repetições identificados em quatro ESTs do guaranazeiro foram sintetizados para amplificar blocos localizados nas regiões 3' não traduzidas, mais frequentes nas seqüências armazenadas no banco de dados, em função de ter sido a biblioteca construída por transcrição reversa utilizando "primers" poli(T). Os produtos de amplificação gerados por estes primers em reações de PCR contendo DNA de seis clones de guaranazeiro foram submetidos à análise inicial em géis de agarose a 3% (Fig. 2).

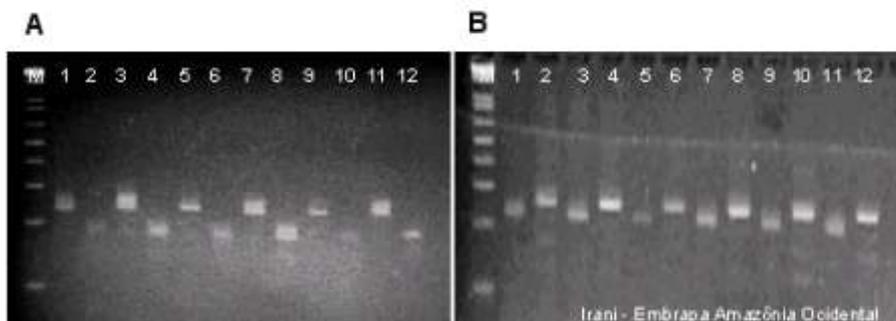


Fig. 2. Resultado da amplificação (PCR) do DNA genômico de clones de guaranazeiro com "primers" desenvolvidos a partir de ESTs de frutos/sementes. A e B - 1 e 2 = clone CMU 300; 3 e 4 = clone CMA 372; 5 e 6 = clone CMU 648; 7 e 8 = clone CMA189; 9 e 10 = clone CMU 505; 11 e 12 = clone CMU 610. A: *locus* GRN07 nas canaletas ímpares e GRN08 nas canaletas pares. B: *locus* GRN09 nas canaletas ímpares e GRN10 nas canaletas pares. M = 1 kb plus (Embrapa Amazônia Ocidental, 2006).

Estes resultados preliminares parecem, quando menos, semelhantes àqueles obtidos por genotipagem automática utilizando dois dos pares de "primers" (GRN01 e GRN04) sintetizados, tomando como modelo a seqüência de clones das bibliotecas genômicas enriquecidas: há, aparentemente, um pequeno número de alelos por *locus*. Análises mais finas em géis de poliacrilamida devem indicar a pertinência de realizar a marcação por fluorescência e proceder a genotipagem automática.

Conclusão

TA e AG/TC foram os dinucleotídeos mais frequentemente encontrados constituindo blocos de repetições *in tandem* no guaranazeiro;

A freqüência de blocos com mais de 7 repetições é baixa.

Agradecimentos

À Fapeam, pelo financiamento do projeto e pela Bolsa para Susan Karoline Barbosa Soares. Ao CNPq, pela Bolsa para Natalia Dayane Moura Carvalho. À UFAM, que foi parceira no cumprimento das metas.

Referências

AMADO, M. V.; ANGELO, P. C. S.; FARIAS, I. P.; ASSUNÇÃO, E. N.; LIRA, M. P. S.; ATROCH, A. L.; PORTO, J. R.; ASTOLFI-FILHO, S. Isolamento e caracterização de marcadores microssatélites em guaranazeiro (*Paullinia cupana* var. *sorbilis*). In: CONGRESSO BRASILEIRO DE GENÉTICA, 51., 2005, Águas de Lindóia. Resumos... Águas de Lindóia: SBG, 2005. 1 CD-ROM. p. 498.

ANGELO, P. C. da S.; AMADO, M. V.; LIRA, M. do. P. S.; ATROCH, A. L.; FARIAS, I. P.; ASTOLFI FILHO, S. Análise preliminar de marcadores microssatélite para guaranazeiro. In: SEMINÁRIO SOBRE PESQUISAS COM O GUARANAZEIRO NA AMAZÔNIA, 1., 2005, Manaus. Anais... Manaus: Embrapa Amazônia Ocidental, 2005. p. 159-167.

BRONDANI, R.P.V. et al. Development, characterization and mapping of microsatellite markers in *Eucalyptus grandis* and *E. urophylla*. *Theoretical and Applied Genetics*, v. 97, p. 816-827, 1998.

CARRIERO, F. et al. Identification of simple sequence repeats (SSR) in olive (*Olea europaea* L.). *Theoretical and Applied Genetics*, v. 104, p. 301-307, 2002.

CHO, Y.G. et al. Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). *Theoretical and Applied Genetics*, v. 100, p. 713-722, 2004.

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. 748-756, 1999.

FARIAS, I. P. et al. 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.

FERREIRA, M. E.; GRATTAPAGLIA, D. Introdução ao uso de marcadores moleculares em análise genética. 2 ed. Brasília, DF: EMBRAPA-CENARGEN, 1996. 220 p. (EMBRAPA-CENARGEN. Documentos, 20).

GAUTSCHI, O. et al. Activity of a novel bcl-2/bcl-xL-bispecific antisense oligonucleotide against tumors of diverse histologic origins. *J Natl Cancer Inst*, v. 93, p. 463-71, 2001.

GRATTAPAGLIA, D. Marcadores moleculares em espécies florestais: *Eucalyptus* como modelo. In: NASS, L. L. et al. Recursos genéticos e melhoramento - plantas. Rondonópolis/MT: Fundação MT, 2001.. p. 967-1010, 2001.

HODGETTS, E. B. et al. Development of microsatellite markers for white spruce (*Picea glauca*) and related species. *Theoretical and Applied Genetics*, v. 102, p. 1252-1258, 2001.

HE, G. et al. Microsatellites as DNA markers in cultivated peanut (*Arachis hypogaea* L.). *BMC Plant Biology*, 3., 2003. Disponível em: < <http://www.biomedcentral.com/1471-2229/3/3> > .

UANG, W.G. et al. Microsatellite DNA in *Actinidia chinensis*: isolation, characterization, and homology in related species. *Theoretical and Applied Genetics*, v. 97, p. 1269-1278, 1998.

MOREIRA, J. de A. N.; SANTOS, J. W. dos; OLIVEIRA, S. R. de M. Abordagens e metodologias para avaliação de germoplasma. Campina Grande: EMBRAPA-CNPA; Brasília, DF: EMBRAPA-SPI, 1994. 115 p.

PERERA, L. et al. Identification and characterization of microsatellite *loci* in coconut (*Cocos nucifera* L.) and the analysis of coconut populations in Sri Lanka. *Mol. Ecol.*, v. 8, p. 335-346, 1999.

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.

SOLIVA, M. et al. Isolation and characterization of microsatellite *loci* in the orchid *Ophrys araneola* (Orchidaceae) and a test of cross-species amplification. *Mol. Ecol.*, v. 9, p. 2178-2179, 2000.

SOUZA, A. P. Biologia molecular aplicada ao melhoramento. In: NASS, L. L. et al. Recursos genéticos e melhoramento - plantas. Rondonópolis/MT: Fundação MT, 2001. p. 939-965.

TENZER, I. et al. Identification of microsatellite markers and their application to population genetics of *Venturia inaequalis*. *Phytopathology*, v. 89, p. 748-753, 1999.