UTILIZAÇÃO DO POLIMORFISMO DE SEQUÊNCIAS DE DNA DE UMA COLEÇÃO DE SORGO *(SORGHUM BICOLOR* L. MOENCH) PARA ESTUDOS DE FILOGEOGRAFIA E ESTUDOS DE ASSOCIAÇÃO.

Lúcio Flávio Alencar Figueiredo
Departamento de Botânica, 70910-900, Universidade de Brasília, Brasil
Autor para correspondência: lucioalencar@unb.br

Bassirou Sine
Centre d'Etude Régional pour l'Amélioration de l'Adaptation à la Sécheresse (Ceraas), Thiès, Senegal

Caroline Calatayud
Cirad, UMR DAP, TA A96/3, Avenue Agropolis

Jacques Chantereau
Cirad, UPR AIVA, 34398 Montpellier, France

Christian Mestres
Cirad, UPR QualiSud, 34398 Montpellier, France

Geneviève Fliedel
Cirad, UPR QualiSud, 34398 Montpellier, France

Xavier Perrier
Cirad, UMR DAP, TA A96/3, Avenue Agropolis

Claire Billot
Cirad, UMR DAP, TA A96/3, Avenue Agropolis

Jean-François Rami
Cirad, UMR DAP, TA A96/3, Avenue Agropolis

Jean-Christophe Glaszmann
Cirad, UMR DAP, TA A96/3, Avenue Agropolis

Monique Deu
Cirad, UMR DAP, TA A96/3, Avenue Agropolis

Brigitte Courtois
Cirad, UMR DAP, TA A96/3, Avenue Agropolis

RESUMO - Sequências do DNA de uma região em comum de múltiplos indivíduos de uma população podem revelar polimorfismos que permitem o estudo da evolução do genoma e suas histórias de seleções, migrações e recombinações. Adicionando-se a esses polimorfismos da população informações quanto à origem geográfica, viabiliza-se um estudo filogeográfico que permite a elaboração de cenários de evolução após a domesticação; bem como quais eventos (mutações ou recombinações) mais influenciaram a diversidade genética (DG) desta população, para aquela região do DNA em estudo. Por outro lado, a associação desses polimorfismos das sequências com as características de interesse estabelece um estudo de associação. Neste contexto, este trabalho visou mostrar a aplicação do polimorfismo de sequências de DNA em um estudo de filogeografia e de associação. Para ambos os estudos foram utilizados dezenove segmentos de seis genes candidatos (*Sh2*, *Bt2*, *SssI*, *Ae1*, *Wx* e *O2*) em uma coleção de sorgo (*Sorghum bicolor* L. Moench), composta por 210 acessos. Os genes *Sh2*, *Bt2*, *SssI*, *Ae1* e *Wx* estão envolvidos na síntese do amido e o gene *O2* na síntese de proteínas de reserva. A DG foi menor nos genes *Sh2*, *Bt2* e *SssI*; e maior para *Ae1*, *Wx* e *O2*. Os testes estatísticos revelaram um elevado número de associações falso-positivas em função da estruturação da coleção, porém foram mantidas várias associações corroboradas pela co-localização com QTLs de sorgo e de milho.

Palavras-chave: genes candidatos, SNPs, NIR, estrutura da população e haplótipo.

INTRODUÇÃO

O polimorfismo de sequências de DNA, ou polimorfismo molecular, é bem mais informativo que outros tipos de polimorfismos (isoenzimas, microssatélites, morfológicos). Ele se apresenta nas mutações na forma de substituições, inserções-deleções e elementos transponíveis (Veuille, 2002). O uso de sequências que pertencem a um determinado gene é uma boa estratégia para os estudos de filogeografia, como também para os estudos de associação (EA), devido às informações existentes nesses genes, como: número de cópias, localização no genoma, similaridades com sequências depositadas, como e quanto ele pode afetar ou controlar uma determinada característica (Ferreira & Grattapaglia, 2009).

No presente trabalho, um dos critérios utilizados na definição dos seis genes candidatos (*Shrunken2* - *Sh2*, *Brittle2* - *Bt2*, *Soluble starch synthase1* - *SssI*, *Amylose extender1* - *Ae1*, *Waxy* - *Wx* e *Opaque2* - *O2*) para o estudo filogeográfico e para o EA foi a participação destes genes em uma rota metabólica

conhecida, no caso, a síntese de amido. Um segundo critério foi a co-localização desses genes candidatos com locos de caracteres quantitativos (QTLs) de qualidade do grão em populações de sorgo (Rami *et al.* 1998). A população utilizada foi uma coleção de sorgo cultivado composta por 210 acessos.

O sorgo pertence à família Poaceae e tem origem no nordeste da África. Sua domesticação ocorreu em três fases. A primeira fase ocorreu há 5000 anos no sudeste do Saara com espécies selvagens. Este processo formou um sorgo bicolor primitivo que migrou para diferentes regiões da África, diferenciando-se em diversas raças por efeito de fundação, seleção e hibridação com sorgos selvagens locais. A segunda fase ocorreu há cerca de 3000 anos, quando o sorgo africano migrou em direção à Ásia. A terceira fase da domesticação foi quando o sorgo chegou ao continente europeu, no século XVI (Doggett, 1988). A subespécie cultivada possui cinco raças: Bicolor, Caudatum, Durra, Guinea e Kafir; e suas combinações. Todas essas informações são importantes para o estudo filogeográfico que analisa a história das relações evolutivas de um grupo e sua distribuição geográfica, e posteriormente para os EA.

O sorgo é o quinto cereal mais produzido no mundo. Dessa produção, cerca de 40% é destinada à alimentação humana na forma de diferentes preparações na África e na Ásia. Características químicas (conteúdo de proteínas - PR, conteúdo de lipídeos - LI, conteúdo de amilose - AM) e físicas (dureza do grão - HD, textura do endosperma - ET) estão envolvidas com a qualidade do grão e consequentemente no produto final na forma de diversas preparações que estão ligadas aos aspectos culturais, religiosos, éticos e econômicos. As características físicas como HD e ET são as mais importantes para o consumidor.

O EA em plantas correlaciona o polimorfismo genético com características fenotípicas que estejam relacionadas com estes polimorfismos. A vantagem do EA em relação à detecção por QTLs é que ela não requer uma população de mapeamento. Assim, o EA investiga o desequilíbrio de ligação (DL) genético, ou seja, as ligações estatísticas entre locos. Esta estratégia do EA explora a diversidade alélica, tendo uma melhor resolução quando o DL não é muito forte. Os componentes essenciais no EA são: i) uma população grande e não estruturada, visando evitar falsas associações de um marcador com determinado fenótipo; ii) a caracterização do fenótipo de interesse da população, que apresente variabilidade; iii) o polimorfismo de uma determinada região em comum do DNA da população que esteja envolvida com o fenótipo em estudo; iv) método estatístico para analisar as associações, onde o DL do genoma, estrutura e tamanho da população são de grande importância (Ferreira & Grattapaglia, 2009).

MATERIAL E MÉTODOS

A composição da coleção de sorgo com 210 acessos foi constituída com base em marcadores moleculares (76 sondas RFLP), classes de fotoperíodo e origem geográfica (Deu *et al.* 2006). Ela foi estabelecida pelo Cirad e pelo International Crops Research Institute for the Semi-Arid Tropics (Icrisat). Os grãos analisados foram cultivados no Ceraas em delineamento descrito por Sine (2003). A fenotipagem dos grãos está descrita em de Alencar Figueiredo *et al.* (2006). Visando minimizar custo e tempo na fenotipagem, foi utilizada a tecnologia da espectroscopia no infravermelho próximo (NIR) para a análise das características físicas (HD, ET) e químicas (PR, AM, LI).

O marcado molecular utilizado foi o polimorfismo nucleotídico único (SNP). A definição dos iniciadores (*primers*), amplificação, sequenciamento, anotação e depósito das sequências dos dezenove segmentos dos seis genes candidatos estão descritos em de Alencar Figueiredo *et al.* (2008). A extensão desses genes varia de 4.3 (*O2*) a 23.5 Kpb (*Ae1*) e o número de exons de seis (*O2*) a 22 (*Ae1*). A análise da diversidade nucleotídica foi realizada com 53 acessos com dados completos para todos os segmentos, utilizando os programas Tassel (Bradbury *et al.* 2007) e dnaSP (Rozas *et al.* 2003). A análise dos haplótipos foi conduzida com o programa Network (http://www.fluxus-technology.com/sharenet.htm).

A estrutura da população foi analisada pelo programa *Structure* (Pritchard *et al.* 2000a,b), que sugere um número de sub-populações utilizando marcadores neutros, neste caso as 76 sondas de RFLP utilizadas por Deu *et al.* (2006). O percentual de mistura de cada indivíduo foi também calculado pelo programa *Structure* e foi utilizado como co-variável nos testes de associação. O número de classes baseou-se nos critérios de Evano *et al.* (2005). SNPs com frequência menor que 1% e haplótipos com menos de cinco indivíduos foram eliminados. Os testes de associação foram realizados no programa *Tassel* (Bradbury *et al.* 2007), no qual três modelos de análise de variância foram testados: i) somente o modelo linear geral (GLM), ii) o GLM utilizando o percentual de mistura do genoma de cada acesso (Matriz Q) e iii) o modelo linear misto (MLM) usando a matriz Q e o coeficiente de Kinship (matriz K).

RESULTADOS E DISCUSSÃO

Foram depositadas 3562 sequências (EU387138–EU390699), referentes aos analisados, perfazendo 1.7 Mpb. Para os genes *Sh2*, *Bt2*, *SssI* e *Ae1* foram analisados dois segmentos por gene. Para o gene *Wx* foram analisados quatro segmentos e para o gene *O2*, seis segmentos. As regiões sequenciadas concatenadas dos respectivos genes correspondem a uma extensão de 1331, 1008, 1525, 1377, 2267 e 3771 pb. Cerca de 90% do gene *O2* foi sequenciado (de Alencar Figueiredo *et al.* 2008).

Nos dezenove segmentos dos seis genes foi identificado um total de 141 polimorfismos e 29 inserções e deleções. A diversidade genética (DG) de cada gene foi analisada por meio de haplótipos, aos quais foram adicionadas a origem geográfica, em função das regiões primárias e secundárias da domesticação, e a DG baseada em marcadores RFLPs (Deu *et al.* 2006; de Alencar Figueiredo *et al.* 2008). As mutações afetaram a diversidade dos haplótipos mais do que as recombinações. Os genes *Sh2*, *Bt2* e *SssI* apresentaram uma menor diversidade do que os outros (*Ae1*, *Wx* e *O2*). A menor diversidade desses genes pode ser devida a uma grande conservação e indícios de fixação

em processo de seleção anterior ao início da agricultura ou no momento da separação das subespécies selvagens e cultivadas (Myers *et al.* 2000; Manicacci *et al.* 2007). Uma nova DG foi identificada em alguns haplótipos ausentes em regiões primárias de origem geográfica e presentes em regiões secundárias.

A nova estrutura da população (Pritchard *et al.* 2000ab; Evanno *et al.* 2005) teve uma sobreposição muito boa com aquela definida por Deu *et al.* (2006) com um reagrupamento de grupos (clusters) próximos. A nova organização dos acessos foi em seis grupos (clusters) no lugar de dez. Considerando a matriz Q (percentual de mistura da população) cerca de 70% das associações foi eliminada em relação ao modelo 1 (GLM). Considerando as matrizes Q+K cerca de 40% das associações foram eliminadas em relação ao modelo GLM + matriz Q. Esta grande quantidade de associações falso-positivas está de acordo com os resultados com milho (Thornsberry *et al.* 2001) e sorgo (Brown *et al.* 2008). Várias associações foram corroboradas pela co-localização com QTLs de sorgo e de milho.

Características físicas apresentaram maior número de associações que as químicas. A ausência de associações do gene *Wx* com o AM se deve à falta de mutantes em amilose na população utilizada. A ausência de associações do gene *O2* com o PR se deve à co-localização do gene *O2* com o conteúdo de kafirinas, e não ao de PR, assim seria melhor conduzir o EA com esta fração proteíca. Associações fortes identificadas somente com os segmentos, e não com eles conectados, sugerem a necessidade de sequenciamento para todo o gene candidato, mesmo para populações estruturadas e com um forte DL.

REFERÊNCIAS BIBLIOGRÁFICAS

BRADBURY, P. J.; ZHANG, Z.; KROON, D. E.; CASSTEVENS, T. M.; RAMDOSS, Y.; BUCKLER, E. S. TASSEL: software for association mapping of complex traits in diverse samples. **Bioinformatics** 23(19): 2633-2635, 2007.

BROWN, P.J.; ROONEY, W.L.; FRANKS, C.; KRESOVICH, S. Efficient mapping of plant height quantitative trait loci in a sorghum association population with introgressed dwarfism genes. **Genetics** 180(1): 629-637, 2008.

DE ALENCAR FIGUEIREDO, L. F.; CALATAYUD, C.; DUPUIS, C.; BILLOT, C.; RAMI J-F.; BRUNEL, D.; PERRIER, X.; COURTOIS, B.; DEU, M.; GLASZMANN, J-C. Phylogeographic evidence of crop neodiversity in sorghum. **Genetics** 179(2): 997-1008, 2008.

DE ALENCAR FIGUEIREDO, L. F.; DAVRIEUX, F.; FLIEDEL, G.; RAMI, J-F.; CHANTEREAU J.; DEU, M.; COURTOIS, B.; MESTRES, C. Development of NIRS equations for food grain quality traits through exploitation of a core collection of cultivated sorghum. **Journal of Agricultural and Food Chemistry** 54(22): 8501-8509, 2006.

DE ALENCAR FIGUEIREDO, L. F.; SINE, B.; CHANTEREAU, J.; MESTRES, C.; FLIEDEL, G.; RAMI J-F.; GLASZMANN, J-C.; DEU, M; COURTOIS, B. Variability of grain quality in sorghum: association with polymorphism in *Sh2*, *Bt2*, *SssI*, *Ae1*, *Wx* and *O2*. **Theoretical and Applied Genetics** 121(6):1171-1185, 2010.

DEU, M.; RATTUNDE, F.; CHANTEREAU, J. A global view of genetic diversity in cultivated sorghums using a core collection. **Genome** 49(2): 168-180, 2006.

DOGGET, H. *Sorghum*. 2 ed. Longman scientific; technical: London, 1988; p. 1-512.

EVANNO, G.; REGNAUT, S.; GOUDET, J. Detecting the number of clusters of individuals using the software Structure: a simulation study. **Molecular Ecology** 14(8): 2611-2620, 2005.

FERREIRA, M. E.; GRATTAPAGLIA, D. Genética de associação em plantas. In: **Genética de Associação em Plantas**, 2ª ed. UFV: Viçosa, 327-370, 2009.

MANICACCI, D.; FALQUE, M.; LE GUILLOU, S.; PIEGU, B.; HENRY, A-M; LE GUILLOUX, M.; DAMERVAL, C.; DE VIENNE, D. Maize *Sh2* gene is constrained by natural selection but escaped domestication. **Journal Evolutionary Biology** 20(2): 503-516, 2007.

MYERS, A. M.; MORELL, M. K.; JAMES, M. G.; BALL, S. G. Recent progress toward understanding biosynthesis of the amylopectin crystal. **Plant Physiology** 122(4): 989-998, 2000.

PRITCHARD, J. K.; STEPHENS, M.; DONNELLY, P. Inference of population structure using multilocus genotype data. **Genetics** 155(2): 945-959, 2000a.

PRITCHARD, J, K.; STEPHENS, M.; ROSENBERG, N. A.; DONNELLY, P. Association mapping in structured populations. **American Journal of Human Genetics** 67(1): 170-181, 2000b.

RAMI, J.F.; DUFOUR, P.; TROUCHE, G.; FLIEDEL, G.; MESTRES, C.; DAVRIEUX, F.; BLANCHARD, P.; HAMON, P. Quantitative trait loci for grain quality, productivity, morphological and agronomical traits in sorghum (*Sorghum bicolor* L. Moench). **Theoretical and Applied Genetics** 97(4): 605-616, 1998.

SINE, B. **Evaluation d'une core collection de sorgho en conditions de stress hydrique pré-floral**. Dakar: Master University. Cheikh Anta Diop, 2003.

THORNSBERRY, J. M.; GOODMAN, M. J.; DOEBLEY, J.; KRESOVICH, S.; NIELSEN, D.; BUCKLER, E. D. Dwarf8 polymorphisms associate with variation in flowering time. **Nature Genetics** 28(3): 286-289, 2001.

VEUILLE, M. Génétique des populations. Polymorphisme moléculaire et théorie de la coalescence. http:/www.snv.jussieu.fr/veuille/cwv02b.pdf., 2002-2003