

5.01.99 - Agronomia
Varição interespecífica de SNP em espécies do gênero *Spondias*

Rychard R. L. Leite^{1*}, Cicero C. S. Almeida²

1. Estudante de Agronomia da UFAL

2. Prof. Dr. Engenheiro Agrônomo - Departamento de Agronomia. Arapiraca UFAL / Orientador

Resumo:

O gênero *Spondias*, pertence à família Anacardiaceae, possui espécies que apresentam grande importância econômica e social, com algumas espécies sendo utilizadas na agroindústria. No Brasil, as espécies de ocorrência são: *S. tuberosa* (umbu), *S. venulosa* (cajazinho), *S. bahiensis* (umbucajá), *S. purpurea* (siriguela), *S. dulcis* (cajarana) e *S. mombim* (cajá). Apesar da grande importância, este gênero é negligenciado quanto a estudos genéticos/genômicos, ocasionando uma lacuna de informações científicas importantes para o esclarecimento das relações filogenéticas deste grupo e manejo de recursos genéticos. Alguns estudos têm sido conduzidos com o desenvolvimento de marcadores moleculares microssatélites e relação genética entre as espécies usando bandeamento cromossômico e filogenia usando DNA barcoding. Essas espécies são negligenciadas no âmbito acadêmico, entretanto, são de elevada importância econômica e social para a região do Nordeste do Brasil, pelo fato de serem utilizadas tanto para consumo humano como para a produção de produtos típicos da região Nordeste. O presente projeto tem como objetivo analisar polimorfismo de único nucleotídeo e desenvolver marcadores moleculares SNP para espécies do gênero *Spondias*. Para isso, foram coletados indivíduos de umbu (*S. tuberosa*), umbucaja (*S. bahiensis*) e cajá (*S. mombim*) no estado de Alagoas, Brasil. A extração de DNA foi realizada em material vegetal vivo, usando o método de extração CTAB. O sequenciamento foi realizado no Laboratório Central de Tecnologias de Alto Desempenho em Ciências da Vida (LacTad) na Universidade Estadual de Campinas, Brasil. Para análise de SNP, foram montados contigs utilizando o software Ray. Os contigs foram utilizados para mapeamentos dos reads utilizando o software Bowtie e os reads mapeados foram utilizados para detectar os SNPs. Para elevar a acurácia das análises, quatro diferentes abordagens foram utilizadas: GATK (Genome Analysis Toolkit), FreeBayes, VarScan e Samtools. O gênero *Spondias* apresenta uma taxonomia não clara e muitas espécies podem ter surgido por hibridização. Análise detalhada utilizando seis contigs mostra que *S. bahiensis* possui elevado número de SNPs e que os possíveis parentais possuem poucos SNP.

Autorização legal: Não se aplica

Palavras-chave: Evolução; Melhoramento vegetal; Recursos genéticos

Apoio financeiro: CNPq

Trabalho selecionado para a JNIC pela instituição: UFAL

Introdução:

O gênero *Spondias* é o mais importante gênero da família Anacardiaceae, com diversas espécies com potencial para exploração agroindustrial. O gênero contém 18 taxa, sendo 10 de ocorrência em regiões Neotropical (Mitchel e Daly, 2015), entretanto, Machado et. al. (2015), descreve uma nova espécie (*S. bahiensis*) conhecida como o híbrido umbucajá, sendo acrescentada como uma nova espécie. Entre as espécies de ocorrência no Brasil, algumas são cultivadas, como *S. mombim* L (cajá), *S. tuberosa* Arruda Câmara (umbu), *S. purpurea* que foi introduzida da América Central e *S. bahiensis* (umbucajá). Outras espécies são encontradas no Nordeste do Brasil, cajarana (*S. cytherea* Sonner ou *S. dulcis*) e o cajazinho (*S. venulosa*) (Almeida et al, 2007, Silva et. al, 2015, Mitchel e Daly, 2015). Recentes estudos, indicam que o grupo possui uma filogenia contendo espécies muito próximas, mostrando que as espécies *S. tuberosa* e *S. venulosa* são a mais derivadas no gênero, enquanto *S. purpurea* e *S. dulcis* se situam em um clado basal (Silva et. al, 2015; Machado et. al. 2015).

Entre as espécies do gênero, o umbuzeiro (*S. tuberosa*), é uma planta nativa do Nordeste do Brasil, seus frutos são utilizados como alimento para o homem, animais domésticos e silvestres no semiárido brasileiro. Sua importância é elevada pelo fato de disponibilizar frutos na época falta de alimento, além disso, segundo Cavalcanti et al. (2000) existem cerca de 48 produtos derivados do umbu, sendo os frutos o mais importante produto, podendo ser consumido *in natura*, na forma de suco, geleia, doce, sorvete e na tradicional umbuzada (polpa do fruto cozido com açúcar e leite). Os frutos são ricos em ácido ascórbico, potássio, cálcio, magnésio e fibra e possuem valor na medicinal popular no controle do colesterol, diabetes e congestão (Lins Neto et al. 2010). *Spondias tuberosa* é conhecida por diversos nomes no Brasil, como imbu e umbu, ambu, embu, imbu, ombu. Essa espécie ocorre do Norte de Minas gerais e se estende por todo o Nordeste do Brasil, com exceção do Estado do Maranhão e estudos de filogeografia indicam que a espécie possui grupos populacionais com uma distribuição na região central da Caatinga e que pequenas populações foram formadas

em regiões de instabilidades no bioma (dados em publicação). O umbucajá tem sido encontrado em ambiente modificado pelo homem, geralmente perto de residências, não sendo observadas populações naturais. Essa espécie foi inicialmente descrita como híbrido entre *S. tuberosa* e *S. mombin*, entretanto, análises citogenéticas (Almeida et al., 2007) e moleculares (Silva et al., 2015; Machado et al., 2015) refutam a hipótese híbrida e sugerem ser uma nova espécie. Machado et al. (2015) descreveram o umbucajá como uma nova espécie e denominada de *S. bahiensis*.

Metodologia:

Material vegetal, isolamento de DNA e sequenciamento de DNA

Foram coletados indivíduos de umbu (*S. tuberosa*), umbucaja (*S. bahiensis*) e cajá (*S. mombin*) no Estado de Alagoas, Brasil. A identificação de cada amostra foi validada usando DNA barcoding para o grupo, usando as regiões plastidiais *rbcL*, *trnH-psbA* e *matK*, seguindo as instruções do BOLD (Barcoding Database) e conforme descrito em Silva et al. (2015). A extração de DNA foi realizada em material vegetal vivo, usando aproximadamente 2 cm² de folhas, usando o método de extração CTAB (SAGHAIMAROOF et al., 1984).

A análise do DNA extraído foi realizada por meio de visualização em gel 1% e quantificação em espectrofotômetro. As amostras de DNA foram fragmentadas para obter fragmentos de aproximadamente 200 500 pares de base (pb) e construída as bibliotecas para sequenciamento. Os fragmentos foram ligados com adaptadores usando Illumina Nextera kit e sequenciados *single reads* com tamanho de 100nt usando a plataforma Illumina HiSeq2500. O sequenciamento foi realizado no Laboratório Central de Tecnologias de Alto Desempenho em Ciências da Vida (LacTad) na Universidade Estadual de Campinas, Brasil.

Montagem de novo e seleção de contigs com genes

Para análise de SNP foram montados *contigs* utilizando o softwar Ray (Boisvert et al., 2012), seguindo os parâmetros: tamanho mínimo dos contigs de 100 nt e cobertura mínima de 8x. Os resultados revelaram aproximadamente 180 mil contigs para as espécies, entretanto, esses contigs incluem tanto sequências codificadoras, como não codificadoras (DNA repetitivo). Nesse contexto foi utilizado o softwar augustus para predição de genes e identificação dos contigs que correspondem a genes. Para *S. tuberosa* foram detectados 41.647 mil genes correspondendo a gene, enquanto em *S. bahiensis* foram observados mais de 32.290 mil genes. A partir disso procedeu-se a análise de SNP.

Análise de SNP

Os contigs foram utilizados para mapeamentos dos *reads* utilizando o software Bowtie e os reads mapeados foram utilizados para detectar os SNPs. Para elevar a acurácia das análises, quatro diferentes abordagens foram utilizadas: GATK (Genome Analysis Toolkit), FreeBayes, VarScan e Samtools. Os resultados das análises no formato VCF foram utilizados no software Geneious para filtragem e o processamento das imagens foi conduzido no software R packager. Para analisar a hipótese híbrida para a espécie *S. bahiensis*, seis contigs de novo de *S. tuberosa* foram utilizados como referência para mapear reads das espécies *S. tuberosa*, *S. bahiensis* e *S. mombin* usando software GATK (Van der Auwera et al., 2013)

Resultados e Discussão:

Os resultados mostraram de 90 mil a 113 mil SNPs nos genes de *S. tuberosa*, enquanto em *S. bahiensis* o número de SNPs foi superior, com valores variando de 122 mil a 205 mil, conforme o método utilizado. Usando o método GATK foi observado que *S. tuberosa* apresentou SNPs em apenas 51% dos genes, enquanto *S. bahiensis* foram detectados SNPs em 93% dos genes. Possivelmente, esse elevado número de genes com SNP em *S. bahiensis* é devido à origem híbrida discutida em diversos trabalhos de filogenia no gênero (ALMEIDA et al., 2007; SILVA et al., 2015; MACHADO et al., 2015).

Entre os quatro métodos de análise de SNP, podemos observar a variação do polimorfismo entre os métodos. Observando-se elevada quantidade de SNPs compartilhados em todos os métodos, indicando consistências nos resultados e confirmam a elevada quantidade de SNPs dentro do gênero.

O gênero *Spondias* apresenta uma taxonomia não clara e muitas espécies podem ter surgido por hibridização. Entre as espécies, *S. bahiensis* tem sido discutida como sendo um híbrido entre *S. tuberosa* e *S. mombin*, e recentemente tem sido descrita como uma nova espécie. As evidências no presente trabalho mostram que *S. bahiensis* apresenta uma elevada quantidade de SNPs em relação a *S. tuberosa*, sugerindo que esse elevado número de SNP pode ser devido a hibridação entre duas espécies. Análise detalhada utilizando seis *contigs* mostra que *S. bahiensis* possui elevado número de SNPs e que os possíveis parentais possuem poucos SNP.

Conclusões:

As espécies do gênero *Spondias* possui elevada quantidade de SNP em genes e que a espécie *S. bahiensis* possui origem híbrida.

Referências bibliográficas

- ALMEIDA, C. C. S.; CARVALHO, P. C. L.; GUERRA, M. Karyotype differentiation among *Spondias* species and the putative hybrid Umbu-cajá (Anacardiaceae). **Botanical Journal of the Linnean Society**. v. 155, p. 541-547, 2007.
- BOISVERT S, RAYMOND F, GODZARIDIS E, LAVIOLETTE F, CORBEIL J Ray Meta: scalable de novo metagenome assembly and profiling. **Genome Biol** 13:R122. doi: 10.1186/gb-2012-13-12-r122. 2012
- LINS NETO et al. Analysis of umbu (*Spondias tuberosa* Arruda (Anacardiaceae)) in different landscape management regimes: A process of incipient domestication?. **Environmental Monitoring and Assessment**. v. 184, p. 4489–4499, 2010
- MACHADO, M. C.; CARVALHO, P. C. L.; VAN DEN BERG, C. Domestication, hybridization, speciation, and the origins of an economically important tree crop of *Spondias* (Anacardiaceae) from the Brazilian Caatinga dry forest. **Neodiversity**. v. 8, p. 8-49, 2015.
- MITCHELL, J. D.; DALY, D. C. A revision of *Spondias* L. (Anacardiaceae) in the Neotropics. **PhytoKeys**. v. 55, p. 1-92, 2015.
- SAGHAI-MAROOF, M. A. et al. Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. **Proceedings of the National Academy of Sciences**. v. 81, p. 8014-8018, 1984.
- SILVA, J. N. et al. DNA barcoding and phylogeny in neotropical species of the genus *Spondias*. **Biochemical Systematics and Ecology**. v. 61, p. 240-243, 2015.
- VAN DER AUWERA GA, CARNEIRO M, HARTL C, POPLIN R, DEL ANGEL G, LEVY-MOONSHINE A, JORDAN T, SHAKIR K, ROAZEN D, THIBAUT J, BANKS E, GARIMELLA K, ALTSHULER D, GABRIEL S, DEPRISTO M. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. **Curr Prot bioinformatics** 43:11.10.1-11.10.33. doi: 10.1002/0471250953.bi1110s43. 2013