

Novos marcadores SNP desenvolvidos a partir do transcriptoma de amendoim forrageiro

Jônatas Chagas de Oliveira⁽¹⁾, Eduardo Fernandes Formighieri⁽²⁾, Ana Letycia Basso Garcia⁽³⁾, Carla Cristina da Silva⁽⁴⁾, Gabriel Rodrigues Alves Margarido⁽⁵⁾ e Tatiana de Campos⁽⁶⁾

⁽¹⁾ Técnico, Universidade Federal do Acre, Rio Branco, AC. ⁽²⁾ Pesquisador, Embrapa Agroenergia, Brasília, DF. ⁽³⁾ Estudante de doutorado, Escola Superior de Agricultura Luiz de Queiróz, Piracicaba, SP. ⁽⁴⁾ Bióloga, doutora em Genética e Biologia Molecular, profissional autônoma, Campinas, SP. ⁽⁵⁾ Professor, Escola Superior de Agricultura Luiz de Queiróz, Piracicaba, SP. ⁽⁶⁾ Pesquisadora, Embrapa Acre, Rio Branco, AC.

Resumo – O amendoim forrageiro (*Arachis pintoi*) tem-se destacado como fonte alternativa de fertilização, como cobertura verde, por aumentar a produtividade da forragem e, ao mesmo tempo, contribuir para uma pecuária mais sustentável. O uso de ferramentas genômicas no programa de melhoramento pode acelerar os ganhos genéticos e reduzir o tempo de desenvolvimento de novas cultivares. Porém, o número limitado de marcadores e a ausência de análises baseadas em polimorfismos de nucleotídeo único (SNPs) têm sido fatores limitantes à aplicação dessa abordagem no programa de melhoramento da espécie. Assim, o objetivo deste estudo foi desenvolver marcadores SNPs a partir do transcriptoma foliar de *A. pintoi*. As bibliotecas foram construídas a partir de duas cultivares: Belomonte e Amarillo MG-100. Foi identificado um total de 374.385 SNPs funcionais, dos quais 39,3% foram identificados exclusivamente com o programa GATK e 8,1% pelo programa SAMtools. Por meio da anotação funcional foi possível identificar genes envolvidos tanto na resposta a estresses bióticos e abióticos como no processo da fotossíntese. Esses são os primeiros marcadores SNPs funcionais desenvolvidos para *A. pintoi* e permitirão estudos de expressão gênica, genes candidatos para edição, construção de mapas genéticos, entre outros que fornecerão informações cruciais para programas de melhoramento de *Arachis*.

Termos para indexação: *Arachis pintoi*, RNA-Seq, locos funcionais.

New SNP markers developed from the forage peanut transcriptome

Abstract – Forage peanut (*Arachis pintoi*) has been highlighted as an alternative source of fertilization, as green cover, as it increases forage productivity and, at the same time, contributes to more sustainable livestock farming. The use of genomic tools in breeding programs can accelerate genetic gains and reduce the development time of new cultivars. However, the limited number of markers and the lack of analyzes based on single nucleotide polymorphisms (SNPs) have been limiting factors for the application of this approach in the species improvement program. Therefore, the objective of this study was to develop SNP markers from the *A. pintoi* leaf transcriptome. The libraries were constructed from two cultivars: Belomonte and Amarillo MG-100. A total of 374,385 functional SNPs were identified, of which 39.3% were identified exclusively with the GATK program and 8.1% with the SAMtools program. Through functional annotation it was possible to identify genes involved both in the response to biotic and abiotic stresses and in the photosynthesis process. These are the first functional SNP markers developed for *A. pintoi* and will allow studies of gene expression, candidate genes for editing, construction of genetic maps, among others that will provide crucial information for *Arachis* breeding programs.

Index terms: *Arachis pintoi*, RNA-Seq, functional loci.

Introdução

O amendoim forrageiro (*Arachis pinto* Krapov. & W.C. Greg.) tem-se destacado como cobertura verde consorciada com culturas comerciais, pois contribui para a manutenção da umidade do solo e ciclagem de nutrientes. Além disso, seu uso em pastagens consorciadas com gramíneas tem auxiliado no aumento expressivo do ganho de peso diário em bovinos, crescimento da produtividade e redução no tempo de recria a pasto de 20 para 13 meses pós-desmama (Sales et al., 2020).

Atualmente, o mercado tem apresentado uma demanda pelo desenvolvimento de novas cultivares. Diante disso, as ferramentas genômicas são apontadas como uma alternativa, pois podem contribuir na redução do tempo necessário ao desenvolvimento de uma cultivar. Dentre os requisitos necessários à aplicação da genômica está a disponibilidade de um grande número de marcadores moleculares, os quais são obtidos a partir da análise do genoma da espécie de interesse. Para espécies que não possuem genoma disponível, o sequenciamento de RNA (RNA-Seq) é apontado como uma alternativa, pois reduz o custo e tempo de desenvolvimento de marcadores moleculares, tais como os SNPs (polimorfismos de nucleotídeo único), além de permitir a coleta de informações que contribuem para a associação do fenótipo com eventos de *splicing* alternativos ou alterações na expressão gênica (Wit et al., 2015). Estudos de associação entre características de interesse e marcadores SNPs têm sido aplicados com sucesso em diversas culturas, demonstrando o grande potencial dessas abordagens para o melhoramento.

Os SNPs são ferramentas muito úteis, especialmente no melhoramento genético, pois têm natureza codominante, são altamente polimórficos e bialélicos, possuem baixa taxa de mutação, estão distribuídos por todo o genoma, além da facilidade de genotipagem, que oferece um bom desempenho considerando a consistência, acurácia e maior facilidade no uso dos dados pelo melhorista (Rafalski, 2002).

Os estudos com RNA-Seq no gênero *Arachis* estão concentrados, principalmente, no amendoim cultivado (*Arachis hypogaea* L.) e nas espécies mais próximas filogeneticamente (*Arachis ipaënsis* Krapov. & W.C. Greg.; *Arachis duranensis* Krapov. & W.C. Greg.), permitindo a identificação de milhares de marcadores SNPs, que têm sido aplicados no melhoramento do amendoim cultivado (Pandey et al., 2017). Já em *A. pinto* foram obtidas pequenas sequências do genoma, das quais foram desenvolvidos

25 marcadores microssatélites. Esses marcadores foram utilizados em estudos de caracterização do germoplasma de amendoim forrageiro, determinação da taxa de cruzamento e identificação de híbridos (Azêvedo et al., 2016; Campos et al., 2016; Oliveira et al., 2019). Porém, ainda não há marcadores SNPs desenvolvidos para a espécie. Assim, o objetivo deste estudo foi identificar novos marcadores SNPs a partir do transcriptoma foliar de *A. pinto*.

Material e métodos

As análises foram realizadas com as sequências (*reads*) e o transcriptoma foliar de *A. pinto* desenvolvidos por Oliveira (2020), o qual utilizou duas cultivares: Amarillo MG100 e Belomonte. O alinhador STAR (versão STAR-2.7.3a) foi usado para alinhar as *reads* ao transcriptoma, a fim de obter as transcrições classificadas (arquivos BAM) para cada cultivar (Dobin et al., 2013). Depois disso, a descoberta de variantes SNPs foi realizada utilizando as configurações padrão no pipeline Genome Analysis Tool Kit (GATK, versão 4.1.4.0) (Poplin et al., 2018) e no pipeline SAMtools (Li, 2011). Para chamada de SNPs, o transcriptoma foi usado como referência. O critério para identificação dos SNPs foi: se alguma cultivar tivesse um alelo diferente da referência, a posição era chamada de provável SNP. Os supostos SNPs identificados pelos programas GATK e SAMtools foram filtrados para QD < 2 e QUAL < 20, respectivamente. Para determinar quais SNPs foram encontrados por cada programa e por ambos, foi aplicado o comando “BCFtools isec” no programa BCFtools (Li, 2011), e esses resultados foram utilizados para construção de um diagrama de Venn usando o pacote *venndiagram* R (Chen; Boutros, 2011). Foi realizada a anotação funcional dos transcritos que deram origem aos SNPs utilizando a base não redundante (nr) do programa Blast2GO, aplicando a opção BlastX e as configurações padrão.

Resultados e discussão

Neste estudo, foi realizada a busca dos primeiros marcadores SNPs funcionais para *A. pinto*. Um total de 358.519 e 303.996 SNPs foi identificado pelos programas GATK e SAMtools, respectivamente. Após a filtragem, restaram 344.173 (96,0%) SNPs encontrados por GATK e 227.160 (74,7%) por SAMtools. Considerando os SNPs encontrados por ambos os programas, foi identificado um total de 196.948 (48,3%). O programa GATK identificou

a maior quantidade de SNPs exclusivos (39,3%) e o SAMtools a menor quantidade (8,1%) (Figura 1).

O número de SNPs de transição (Ts) foi maior que o número de SNPs de transversão (Tv), produzindo uma relação Ts/Tv de 1,75 para ambos os programas de chamada de SNP (Tabela 1). A variação de transição mais comum foi C ↔ T, com 114.622 e 74.149 SNPs para GATK e SAMtools, respectivamente. A variação de transversão mais comum foi A ↔ T, com 36.277 SNPs (GATK) e 23.727 SNPs (SAMtools).

As variações no número de SNPs identificados em cada um dos programas podem ter ocorrido devido a diferentes abordagens utilizadas em cada pipeline (Cornish; Guda, 2015; Hwang et al., 2015). A filtragem é uma etapa necessária para evitar distorções e falsos positivos na chamada de SNP. Quando Hwang et al. (2015) filtraram variantes abaixo do índice de qualidade (QUAL < 20), encontraram concordância de, aproximadamente, 92,0% entre os programas de chamada de variantes. Considerando

que foi realizada aplicação de padrão de filtragem similar no presente estudo, espera-se que a acurácia tenha sido elevada consideravelmente, reduzindo a presença de falsos positivos nas variantes identificadas. No entanto, o principal fator limitante na acurácia da busca por SNP é a ausência de um genoma de referência para o amendoim forrageiro. Diante disso, as variantes SNPs identificadas por ambos os programas podem ser mais precisas, com menor probabilidade de falsos positivos.

A anotação funcional dos transcritos que deram origem aos SNPs foi realizada por meio da comparação com o banco de dados nr e resultou em 37.219 transcritos mapeados em 146.844 termos de ontologia genética. Dentre os termos anotados, encontram-se RAP2.4, RD21A, NDK1 e ERD15, os quais estão relacionados com mecanismos de tolerância e resposta a estresses bióticos e abióticos. Também foram identificados a plastocianina, CLPB3, tioredoxina tipo F e CAB7, que estão envolvidos nos processos da fotossíntese.

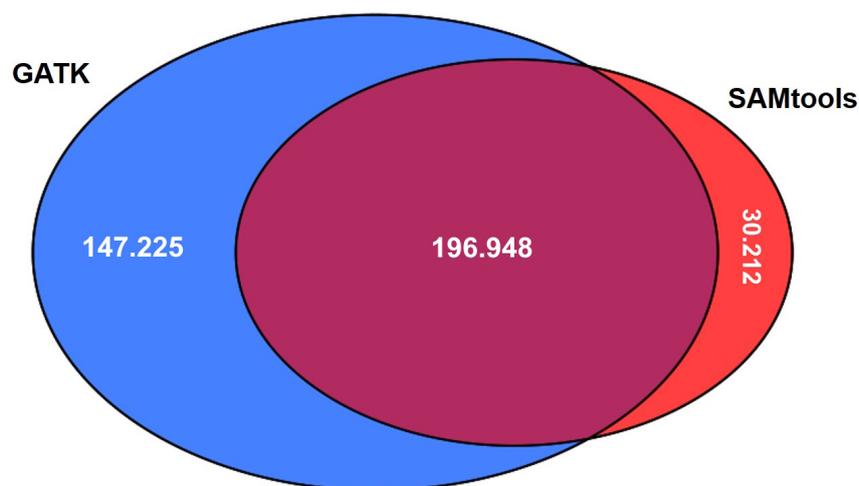


Figura 1. Diagrama de Venn mostrando o número de SNPs encontrados pelos programas GATK e SAMtools, após a filtragem.

Os números nos círculos indicam o número de SNPs encontrados por um ou ambos os pipelines.

Tabela 1. Resumo dos dados dos SNPs identificados pelos programas GATK e SAMtools, após a filtragem.

Programa	Número de SNP	Transição				Transversão		
		A ↔ G	C ↔ T	A ↔ C	A ↔ T	C ↔ G	G ↔ T	
GATK	344.173	105.733	114.622	29.707	36.277	26.884	33.161	
SAMtools	227.160	71.309	74.149	20.479	23.727	18.198	20.524	

Mais pesquisas são necessárias para validar e caracterizar os SNPs descobertos, bem como ligá-los a características fenotípicas de interesse agrônomo, especialmente os genes identificados na anotação funcional. Análises de expressão diferencial podem ser importantes na seleção de SNPs para validação. Além disso, esse conjunto de dados proporcionará novas possibilidades para seleção genômica, mapeamento de ligação e GWAS no programa de melhoramento de *Arachis pintoi*.

Conclusões

- 1) As análises do transcriptoma foliar de *Arachis pintoi* permitiram a identificação dos primeiros marcadores SNPs para o amendoim forrageiro.
- 2) Os primeiros genes candidatos foram identificados, permitindo que estudos de expressão possam ser iniciados para compreender cadeias metabólicas relacionadas à fotossíntese, estresses bióticos e abióticos.

Agradecimentos

Ao governo federal e governo do estado do Acre (Fapac TO: 024/2018) pelo apoio financeiro, ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela concessão de bolsa de produtividade e à Embrapa Acre pelo financiamento e infraestrutura para condução dos experimentos.

Referências

- AZÊVEDO, H. S. F. S.; SOUSA, A. C. B.; MARTINS, K.; OLIVEIRA, J. C.; YOMURA, R. B. T.; SILVA, L. M.; VALLS, J. F. M.; ASSIS, G. M. L.; CAMPOS, T. Genetic diversity of the forage peanut in the Jequitinhonha, São Francisco, and Paraná River valleys of Brazil. **Genetics and Molecular Research**, v. 15, n. 3, p. 1-11, Sept. 2016. DOI: <http://dx.doi.org/10.4238/gmr.15038601>.
- CAMPOS, T.; AZÊVEDO, H. S. F. S.; OLIVEIRA, J. C.; FERREIRA FILHO, J. A.; YOMURA, R. B. T.; SILVA, L. M. **Protocolo para identificação de híbridos de amendoim forrageiro utilizando marcador molecular microssatélite**. Rio Branco, AC: Embrapa Acre, 2016. 29 p. (Embrapa Acre. Documentos, 146). Disponível em: <http://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/1064915>. Acesso em: 15 maio 2023.
- CHEN, H.; BOUTROS, P. C. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. **BMC Bioinformatics**, v. 12, article number 35, Jan. 2011. DOI: <https://doi.org/10.1186/1471-2105-12-35>.
- CORNISH, A.; GUDA, C. A Comparison of variant calling pipelines using genome in a bottle as a reference. **Bio-Med Research International**, v. 2015, article ID 456479, Oct. 2015. DOI: <https://doi.org/10.1155/2015/456479>.
- DOBIN, A.; DAVIS, C. A.; SCHLESINGER, F.; DRENKOW, J.; ZALESKI, C.; JHA, S.; BATUT, P.; CHAISSON, M.; GINGERAS, T. R. STAR: ultrafast universal RNA-seq aligner. **Bioinformatics**, v. 29, n. 1, p. 15-21, Jan. 2013. DOI: <https://doi.org/10.1093/bioinformatics/bts635>.
- HWANG, S.; KIM, E.; LEE, I.; MARCOTTE, E. M. Systematic comparison of variant calling pipelines using gold standard personal exome variants. **Scientific Reports**, v. 5, article number 17875, Dec. 2015. DOI: <https://doi.org/10.1038/srep17875>.
- LI, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. **Bioinformatics**, v. 27, n. 21, p. 2987-2993, Nov. 2011. DOI: <https://doi.org/10.1093/bioinformatics/btr509>.
- OLIVEIRA, J. C. **Análise do genoma funcional de *Arachis pintoi* e desenvolvimento de novos marcadores moleculares**. 2020. 94 f. Tese (Doutorado em Biodiversidade e Conservação) – Programa de Pós-graduação em Biodiversidade e Biotecnologia da Amazônia Legal, Universidade Federal do Acre, Rio Branco, AC.
- OLIVEIRA, J. C.; RUFINO, P. B.; AZÊVEDO, H. S. F. S.; SOUSA, A. C. B.; ASSIS, G. M. L.; SILVA, L. M.; SEBBENN, A. M.; CAMPOS, T. Inferring mating system parameters in forage peanut, *Arachis pintoi*, for Brazilian Amazon conditions. **Acta Amazonica**, v. 49, n. 4, p. 277-282, out./dez. 2019. DOI: <http://dx.doi.org/10.1590/1809-4392201900200>.
- PANDEY, M. K.; AGARWAL, G.; KALE, S. M.; CLEVINGER, J.; NAYAK, S. N.; SRISWATHI, M.; CHITIKINENI, A.; CHAVARRO, C.; CHEN, X.; UPADHYAYA, H. D.; VISHWAKARMA, M. K.; LEAL-BERTIOLI, S.; LIANG, X.; BERTIOLI, D. J.; GUO, B.; JACKSON, S. A.; OZIAS-AKINS, P.; VARSHNEY, R. K. Development and evaluation of a high density genotyping 'Axiom_Arachis' array with 58K SNPs for accelerating genetics and breeding in groundnut. **Scientific Reports**, v. 7, article number 40577, Jan. 2017. DOI: <https://doi.org/10.1038/srep40577>.

POPLIN, R.; RUANO-RUBIO, V.; DEPRISTO, M. A.; FENNELL, T. J.; CARNEIRO, M. O.; VAN DER AUWERA, G. A.; KLING, D. E.; GAUTHIER, L. D.; LEVY-MOONSHINE, A.; ROAZEN, D.; SHAKIR, K.; THIBAUT, J.; CHANDRAN, S.; WHELAN, C.; LEK, M.; GABRIEL, S.; DALY, M. J.; NEALE, B.; MACARTHUR, D. G.; BANKS, E. Scaling accurate genetic variant discovery to tens of thousands of samples. **bioRxiv**, July 2018. DOI: <https://doi.org/10.1101/201178>.

RAFALSKI, A. Applications of single nucleotide polymorphisms in crop genetics. **Current Opinion in Plant Biology**, v. 5, n. 2, p. 94-100, Apr. 2002. DOI: [https://doi.org/10.1016/S1369-5266\(02\)00240-6](https://doi.org/10.1016/S1369-5266(02)00240-6).

SALES, M. F. L.; ASSIS, G. M. L. de; ANDRADE, C. M. S. de; SÁ, C. P. de; MESQUITA, A. Q. de; VALENTIM, J. F. **Recria de bovinos de corte em pastos de capim-humidícola consorciados com amendoim forrageiro no Estado do Acre**. Rio Branco, AC: Embrapa Acre, 2020. 28 p. (Embrapa Acre. Circular técnica, 79). Disponível em: <http://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/1127108>. Acesso em: 15 maio 2023.

WIT, P.; PESPENI, M. H.; PALUMBI, S. R. SNP genotyping and population genomics from expressed sequences – current advances and future possibilities. **Molecular Ecology**, v. 24, n. 10, p. 2310-2323, May 2015. DOI: <https://doi.org/10.1111/mec.13165>.