UNIVERSIDADE FEDERAL DO ACRE – UFAC
PROGRAMA DE PÓS-GRADUAÇÃO EM
BIODIVERSIDADE E BIOTECNOLOGIA DA AMAZONIA
LEGAL

ANÁLISE DO GENOMA FUNCIONAL DE *Arachis pintoi* E DESENVOLVIMENTO DE NOVOS MARCADORES MOLECULARES

JÔNATAS CHAGAS DE OLIVEIRA

Rio Branco – AC
2020

JÔNATAS CHAGAS DE OLIVEIRA

ANÁLISE DO GENOMA FUNCIONAL DE *Arachis pintoi* E DESENVOLVIMENTO DE NOVOS MARCADORES MOLECULARES

Tese de Doutorado apresentada ao Curso de Doutorado do Programa de Pós-graduação em Biodiversidade e Biotecnologia da Rede BIONORTE, na Universidade Federal do Acre, como requisito para a obtenção do Título de Doutor em Biodiversidade e Conservação.

Orientadora: Tatiana de Campos
Co-orientadora: Carla Cristina da Silva

Rio Branco – AC
OUTUBRO/2020

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica elaborada pela Biblioteca Central da UFAC

O482a Oliveira, Jônatas Chagas de, 1987 -
Análise do genoma funcional de *Arachis pintoi* e desenvolvimento de novos marcadores moleculares / Jônatas Chagas de Oliveira; orientador: Dra. Tatiana de Campos e Coorientadora: Dra. Carla Cristina da Silva. – 2020.
94 f.: il.; 30 cm.

Tese (Doutorado) – Universidade Federal do Acre, Programa de Pós-Graduação em Rede de Biodiversidade e Biotecnologia da Rede BIONORTE, Rio Branco, 2020.
Inclui referências bibliográficas.

1. Amendoim forrageiro. 2. RNA-Seq. 3. Transcriptoma. I. Campos, Tatiana de (orientadora). II. Silva, Carla Cristina da (Coorientadora). III. Título.

CDD: 660

Bibliotecário: Uéliton Nascimento Torres CRB-11º/1074.

JÔNATAS CHAGAS DE OLIVEIRA

ANÁLISE DO GENOMA FUNCIONAL DE *Arachis pintoi* E DESENVOLVIMENTO DE NOVOS MARCADORES MOLECULARES

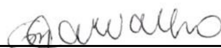
Tese de Doutorado apresentada ao Curso de Doutorado do Programa de Pós-graduação em Biodiversidade e Biotecnologia da Rede BIONORTE, na Universidade Federal do Acre, como requisito parcial para a obtenção do Título de Doutor em Biodiversidade e Conservação.

Aprovado em: 27/10/2020

Banca Examinadora

Dra. Tatiana de Campos (Orientadora)
Empresa Brasileira de Pesquisa Agropecuária – Embrapa Acre

Dra. Berenice Kussumoto de Alcântara da Silva
Universidade Federal do Acre



Dra. Clarice Maia Carvalho
Universidade Federal do Acre

Dra. Leila Priscila Peters
Universidade Federal do Acre

Dra. Elisa Ferreira Moura Cunha
Empresa Brasileira de Pesquisa Agropecuária – Embrapa Amazônia Oriental

À minha família e minha esposa Josiany, dedico.

AGRADECIMENTOS

Agradeço, primeiramente, a Deus que me presenteou com o dom da vida mais de uma vez e, de forma especial, me concedeu força, saúde, inteligência e fé, sem as quais, não conseguiria chegar a conclusão deste trabalho, além de suprir todas as minhas necessidades e abrir as portas no momento certo.

À minha família pelo amor, apoio, suporte e incentivo durante todo o período de realização desse trabalho.

À minha esposa Josiany por sua paciência, amor, motivação, e incentivo em todos os momentos da vida, especialmente de minha jornada da pós-graduação (mestrado e doutorado).

À Dra. Tatiana de Campos, minha orientadora, não só pela orientação nesse trabalho, mas pela confiança em meu trabalho desde o mestrado até a conclusão do processo de doutoramento. Posso afirmar sem dúvidas que cresci muito com todos os conhecimentos compartilhados, as correções e “broncas” recebidas e, acima de tudo, essa amizade que me inspirou e motivou durante todo o caminho.

À Dra. Carla Cristina da Silva, cuja coorientação foi essencial tanto na construção como realização desse trabalho.

À Universidade Estadual de Campinas, especialmente na pessoa da Dra. Anete Pereira de Souza, por permitir a realização de etapas fundamentais desse trabalho em seus laboratórios além de contribuir na discussão e construção dos artigos que resultaram de nossos experimentos.

À Escola Superior de Agricultura Luiz de Queiróz, na pessoa do Dr. Gabriel Rodrigues Alves Margarido e seus alunos, Ana Letycia Basso e Victor Hugo de Mello Pessoa, do Laboratório de Bioinformática Aplicado à Bioenergia, pela excelente recepção, cuidado e o aprendizado extremamente importante às análises e elaboração desse trabalho. As experiências vividas com vocês certamente contribuíram para tornar-me não somente um profissional melhor, mas também uma pessoa melhor.

À Embrapa Acre por permitir e contribuir com os meios necessários para a realização dos experimentos.

Aos professores do PPG-Bionorte pela dedicação em nos ensinar e pelos conhecimentos essenciais para minha capacitação profissional.

Aos colegas do curso de doutorado, em especial a Marilene, Diego e Ana Paula pela amizade, companheirismo e ajuda mútua construída nesse processo tão árduo, os quais quero levar por toda a minha vida.

Aos amigos do LabMol da Embrapa: Clemeson, Susana Melo, Hellen Azêvedo, André Lucas, Nathalia, Caroline, Fabiana, Polinar e Emilly. A caminhada foi mais fácil com vocês ao meu lado contribuindo não somente na execução dos projetos, mas compartilhando momentos felizes e tristes, estando à disposição sempre que precisei.

Ao corpo técnico da Embrapa por auxiliar nos experimentos, análises e construção dessa tese, especialmente o Dr. Eduardo Fernandes Formighieri, Luciélio Manoel da Silva e Renata Beltrão Teixeira Yomura.

À todos aqueles que não mencionei o nome, mas que participaram desse processo de forma direta ou indireta: Muito obrigado!

“Melhor é o fim das coisas do que o seu princípio”. Eclesiastes 7:8.

RESUMO

O amendoim forrageiro (*Arachis pintoi* Krapov. e W.C. Greg.) tem apresentado resultados importantes na pecuária por seu alto valor nutritivo, o que tem contribuído no aumento da produção de leite, ganho de peso no gado de corte e redução no tempo de abate em até nove meses. Além disso, seu uso como cobertura verde em consórcios com culturas comerciais tem contribuído na redução da erosão, manutenção da umidade do solo, ciclagem de nutrientes, fixação biológica de nitrogênio e baixas emissões de óxido nitroso. O uso de sequenciamento nova geração (NGS) pode auxiliar no conhecimento da biologia da espécie e desenvolvimento de novas cultivares. O objetivo do presente trabalho foi analisar o genoma funcional de *Arachis pintoi* por meio de dados de RNA-Seq e desenvolver novos marcadores moleculares *single nucleotide polymorphisms* (SNP) e microssatélites (SSR). Foram realizadas modificações no protocolo de cloreto de lítio para isolamento de RNA total que ajudaram a eliminar problemas associados com a presença de altas concentrações de metabólitos secundários em folhas de amendoim forrageiro e sem sinais de degradação. O RNA de duas cultivares (Belomonte e Amarillo MG-100) foi extraído e utilizado para montagem das bibliotecas, as quais foram sequenciadas e tiveram seu genoma funcional montado com o programa Trinity e metodologia *de novo*. O sequenciamento gerou aproximadamente 223 milhões de *pair-end reads* de alta qualidade. Um total de 98.432 transcritos com comprimento médio de 1.102,2 bp foram obtidos. O valor de N50 e a taxa de alinhamento completo foram 1.924 bp e 95%, respectivamente. A função molecular (36,0%) e o processo biológico (35,8%) apresentaram a maioria dos termos da Ontologia Genética atribuídos, enquanto em componente celular (28,2%) foi atribuído o menor número. Uma busca por SNPs funcionais e marcadores SSR identificou 374.385 e 4.461, respectivamente. Cento e oitenta e seis marcadores SSR foram selecionados para validação, dos quais 63 (33,87%) foram polimórficos com média de 7,37 alelos por loco. Os marcadores apresentaram elevados valores médios para conteúdo de informação polimórfica (PIC = 0,70) e poder discriminatório (D = 0,80). O marcador Ap(CT)75 foi capaz de realizar o *fingerpriting* das cultivares Belomonte, Amarillo MG-100, Alqueire 1, e BRS Mandobi, identificando perfis únicos. Trinta e três SSRs foram avaliados quanto à transferibilidade para amendoim e outras seis espécies silvestres do gênero *Arachis*, o que resultou em amplificação cruzada variável (63,64 a 100%). O loco Ap(CT)68 permitiu a certificação entre 229 cruzamentos. O polimorfismo dos alelos foi detectado em gel de agarose, permitindo a redução de custos e tempo necessário para o processo de identificação do híbrido. O conjunto dos resultados apresentados neste trabalho permitirão a aplicação de técnicas

moleculares modernas como associação genômica, expressão diferencial e seleção genômica, os quais contribuirão significativamente para o conhecimento da biologia de *A. pintoi* e avanço do programa de melhoramento de *Arachis*.

Palavras-chave: Amendoim forrageiro, RNA-Seq, transcriptoma, SNP, SSR.

ABSTRACT

Forage peanuts (*Arachis pintoi* Krapov. and W.C. Greg.) has shown important results in livestock due to the high nutritional value, which has contributed to the increase in milk production, weight gain in beef cattle, and reduction in slaughter time by up to nine months. In addition, its use as green cover in associations with commercial crops has contributing to the erosion reduction, maintenance of soil moisture, nutrient cycling, biological nitrogen fixation and low nitrous oxide emissions. The use of next generation sequencing (NGS) can assist in understanding the biology of the species and the development of new cultivars. The objective of the present work was to analyze the functional genome of *Arachis pintoi* using RNA-Seq data and to develop new single nucleotide polymorphisms (SNP) and microsatellite (SSR) molecular markers. Modifications in lithium chloride protocol were performed for total RNA isolation that helped eliminate the problems associated with the presence of high concentrations of secondary metabolites in forage peanut leaves, and lacked signs of degradation. The RNA of two cultivars (Belomonte and Amarillo MG-100) was extracted and used to assemble the libraries, which were sequenced and had their transcriptome assembled with Trinity program and *de novo* methodology. The sequencing output generated approximately 223 million high-quality pair-end reads. A total of 98,432 transcripts with an average length of 1,102.2 bp were obtained. The N50 value and complete alignment rate were 1,924 bp and 95 %, respectively. The molecular function (36.0%) and biological process (35.8%) presented the majority of the Gene Ontology terms assigned, whereas in the cellular component (28.2%) were assigned the smaller number. A search for functional SNPs and SSRs markers identified 374,385 and 4,461, respectively. One hundred eighty-six SSR markers were selected for validation, of which 63 (33.87%) were polymorphic with average of 7.37 alleles per locus. The markers presented high average for polymorphic information content (PIC = 0.70) and discriminatory power (D = 0.80). The Ap(CT)75 marker was able to identify unique profiles among Belomonte, Amarillo MG-100, Alqueire 1, and BRS Mandobi. Thirty-three SSRs were evaluated for the transferability to peanut and other six wild species from *Arachis* genus, which resulted in variable cross-species amplification (63.64 to 100%). The Ap(CT)68 locus allowed the certification among 229 crosses. The allele polymorphism was detected on agarose gel, allowing the reduction of costs and time required for the hybrid identification process. The set of results presented in this work will allow the application of modern molecular techniques such as genomic association, differential expression and genomic selection, which

will significantly contribute to the knowledge of the biology of *A*. *pintoi* and the advancement of the peanut breeding program.

Keywords: Forage peanut, RNA-Seq, transcriptome, SNP, SSR.

SUMÁRIO

1. INTRODUÇÃO.. **11**

2. OBJETIVOS... **13**

2.1 Objetivo geral.. **13**

2.2 Objetivos específicos... **13**

3. REVISÃO BIBLIOGRÁFICA.. **14**

3.1 Uso de leguminosas forrageiras em pastagens consorciadas na Amazônia.. **14**

3.2 Biologia do amendoim forrageiro... **17**

3.3 Estudos genéticos com amendoim forrageiro... **19**

3.3.1 Marcadores Moleculares.. **19**

3.3.2 Estudos moleculares associados ao programa de melhoramento.................. **20**

3.3.2.1 Caracterização da diversidade genética molecular de amendoim forrageiro.. **20**

3.3.2.2 Estimativa da reprodução cruzada em amendoim forrageiro.................... **21**

4. REFERÊNCIAS.. **23**

5. CAPÍTULO I... **34**

A MODIFIED PROTOCOL FOR TOTAL RNA ISOLATION FROM FORAGE PEANUTS…..……………….........…………………………….... **34**

6. CAPÍTULO II.. **40**

***De novo* TRANSCRIPTOME ASSEMBLY AND DEVELOPMENT OF SNP MARKERS FROM *Arachis pintoi* LEAVES**……………………..……………………………………….................... **40**

7. CAPÍTULO III... **56**

DEVELOPMENT OF MICROSATELLITE MARKERS FROM AN *Arachis pintoi* TRANSCRIPTOME………………………………………………… **56**

8. CAPÍTULO IV………………………………………………………………… **87**

MOLECULAR IDENTIFICATION OF INTERSPECIFIC HYBRIDS IN FORAGE PEANUTS……………………………………………………….. **87**

9. CONCLUSÕES.. **94**

1. INTRODUÇÃO

A espécie *Arachis pintoi* Krapov. e W.C. Greg. é uma espécie diplóide (2n = 2x = 20) pertencente à família Fabaceae, secção *Caulorhizae* (LEWIS et al., 2005; JUDD et al., 2009). É originária da flora brasileira, ocorrendo desde o Planalto Central, no estado de Goiás, até o litoral da Bahia (KRAPOVICKAS; GREGORY, 2007). Tem recebido destaque por seu uso em pastagens consorciadas com gramíneas, o que contribui no expressivo aumento do ganho de peso em gado de corte, reduzindo o tempo de abate destes animais (MAIA, 2018). Em 2019, pastagens consorciadas com amendoim forrageiro geraram um impacto positivo de aproximadamente R$ 91 milhões no Estado do Acre (EMBRAPA, 2020).

Além disso, como cobertura verde em consórcio com culturas comerciais como *Cucumis sativus* (SILVA et al., 2012), *Coffea arabica* (SANTOS et al., 2014; ROSE et al., 2019), *Prunus persica* (WANG et al., 2015) e *Solanum lycopersicum* (RESENDE et al., 2020), auxilia na ciclagem de nutrientes, redução da erosão e manutenção da umidade do solo, controle de ervas daninhas, fixação biológica de nitrogênio e baixos índices de emissão de óxido nitroso, bem como favorecendo o sequestro de carbono. Apesar desses aspectos positivos, a pequena quantidade de cultivares disponíveis ao produtor e o preço elevado das sementes têm dificultado a sua aplicação em larga escala.

Os marcadores moleculares têm-se revelado uma ferramenta eficiente na caracterização molecular de espécies vegetais, fornecendo informações cruciais sobre a biologia das espécies. Dentre os marcadores moleculares, os microssatélites ou SSR (*Simple Sequence Repeats*) e SNP (*Single Nucleotide Polymorphism*) destacam-se por possuir natureza co-dominante, elevado polimorfismo, e grande distribuição ao longo do genoma (RAFALSKI, 2002; HODEL et al., 2016). No entanto, os microssatélites são multialélicos enquanto os SNPs são bialélicos.

Apesar das grandes vantagens do uso desses marcadores, o custo para o desenvolvimento deles é elevado e o processo laborioso. Nesse sentido, o uso de RNA-Seq (*RNA sequencing*) é uma alternativa que permite a identificação de marcadores funcionais com menor custo e trabalho do que as técnicas tradicionais (WIT et al., 2015; TAHERI et al., 2018). A grande vantagem dos marcadores baseados em RNA é que eles são derivados de regiões expressas do genoma, e podem ser associados a características fenotípicas de interesse agronômico (POCZAI et al., 2013).

Estudos de associação genômica ampla (GWAS) baseados em SNPs e SSRs têm sido realizados com sucesso em *Sorghum bicolor* (MORRIS et al., 2013), *Gossypium* spp. (NIE

et al., 2016), *Zea mays* (ZHANG et al., 2016), *Glycine max* (CONTRERAS-SOTO et al., 2017), e *Saccharum* spp. (BARRETO et al., 2019), demonstrando o grande potencial dessa abordagem para descobrir associações entre marcador-característica (MTAs).

Embora o número de marcadores SSR desenvolvidos para *A. pintoi* ainda seja pequeno e não existam SNPs descritos até o momento, estudos utilizando os microssatélites disponíveis juntamente com alguns locos heterólogos têm fornecido informações importantes sobre caracterização do banco ativo de germoplasma de amendoim forrageiro, identificação de híbridos e caracterização da taxa de cruzamento (AZÊVEDO et al., 2016; CAMPOS et al., 2016; OLIVEIRA et al., 2019).

Desta forma, o desenvolvimento de SNPs e novos SSR, especialmente derivados de genes expressos, é crucial, pois permitirá a realização de estudos de associação e seleção genômica e expressão diferencial, os quais poderão direcionar novas estratégias utilizadas no programa de melhoramento de *A. pintoi*.

2. OBJETIVOS

2.1 Objetivo Geral

O objetivo do presente trabalho foi analisar o genoma funcional de *Arachis pintoi* por meio de dados de RNA-Seq e desenvolver novos marcadores moleculares *single nucleotide polymorphisms* (SNP) e microssatélites (SSR).

2.2 Objetivos Específicos

- Identificar marcadores SNPs e SSRs a partir do transcriptoma de dois genótipos contrastantes de *A. pintoi*, utilizando ferramentas de bioinformática;
- Caracterizar 186 microssatélites utilizando 19 genótipos de *A. pintoi*, para determinação do número de alelos, tamanho de amplificação, temperatura de anelamento, heterozigosidade esperada e observada e número do PIC (conteúdo informativo de polimorfismo);
- Selecionar um conjunto de marcadores para discriminação genética de cultivares de *A. pintoi* disponíveis no mercado;
- Montar um painel de marcadores moleculares que possibilitará a identificação precoce e segura de híbridos de *A. pintoi*.

3. REVISÃO BIBLIOGRÁFICA

3.1 Uso de leguminosas forrageiras em pastagens consorciadas na Amazônia

Dentre as alternativas para uma pecuária sustentável, a intensificação do sistema produtivo na agropecuária tem-se destacado. A intensificação parte do princípio de que uma maior quantidade de animais pode ser criada em uma mesma área (SWAIN et al., 2018), o que permite o aumento da produtividade sem a necessidade da expansão das pastagens através da derrubada de novas áreas. Entretanto, para que isso seja possível é necessário o investimento em tecnologias que irão permitir o aumento da produtividade, tais como: ambiente de produção altamente controlado, uso de alimentação comercial com otimização nutricional e aplicação de técnicas avançadas de melhoramento animal, o que produzirá animais com maior peso e que estarão prontos para o abate mais rápido do que no sistema extensivo tradicional (SWAIN et al., 2018).

Algumas medidas para a implementação da intensificação da produção pecuária na Amazônia têm sido avaliadas, dentre elas destacam-se os sistemas de pastejo rotacionado, os sistemas silvipastoris com pastejo rotacionado e o uso de pastagens consorciadas com leguminosas forrageiras.

O uso de pastagens consorciadas de gramíneas com leguminosas forrageiras é apontado como uma das possibilidades até para os pequenos produtores na região Amazônica. Essa prática permite a fixação biológica do nitrogênio ao solo, evitando a degradação das pastagens e melhorando a qualidade produtiva do solo, além de agregar maior valor nutricional ao pasto (SÁ et al., 2008).

A implementação desse sistema no Estado do Acre ocorreu em 1976, pela Empresa Brasileira de Pesquisa Agropecuária (Embrapa). A puerária (*Pueraria phaseoloides*) foi a primeira leguminosa a ser adotada (VALENTIM; ANDRADE, 2005). A produtividade a partir desse consórcio foi de 4,9 a 12,5@/ha/ano. No entanto, houve baixa compatibilidade com o capim Estrela Africano sob pastejo intensivo, pois a puerária possui hábito de crescimento volúvel, competindo com a gramínea.

Diante desse problema, a Embrapa divulgou o uso do amendoim forrageiro (*Arachis pintoi* Krapov. e W.C. Greg.) (Figura 1), através da cultivar Belomonte. O efeito da introdução de amendoim forrageiro em pastos de braquiária (*Brachiaria humidicola* cv. Comum) representaram um aumento de 42% na produtividade sobre o ganho de peso de novilhos Nelore (URBANSKI, 2016). A média de ganho de peso diário foi de

0,263kg/animal/dia na pastagem consorciada, enquanto na pastagem exclusiva de gramínea foi de 0,186kg/animal/dia. Tais ganhos representaram a redução no tempo de abate de 37 para 29 meses.

Foto: Jônatas C. de Oliveira

Figura 1. Banco Ativo de Germoplasma de amendoim forrageiro localizado na Embrapa Acre.

O ganho em animais castrados também foi expressivo. Animais castrados do pasto consorciado apresentaram produtividade 37% superior aos castrados do pasto puro (MACHADO, 2017). E ainda, os animais não castrados do pasto consorciado apresentaram desempenho 46,8% superior aos castrados de pasto puro.

Em outro estudo realizado por Maia (2018), os animais Nelore do pasto consorciado apresentaram acréscimo de 29,25% no ganho de peso total em relação aos de pasto puro. Animais cruzados com Aberdeen Angus tiveram ganho de peso total 65,46% maior do que os animais cruzados de pasto puro. Esses ensaios demonstraram a significativa rentabilidade do uso do amendoim forrageiro em pastagens consorciadas no processo de engorda do rebanho.

Na pecuária de leite os resultados também foram positivos. Lascano (1994) relatou um aumento de 17% a 20% na produção de leite. Em consórcio entre amendoim forrageiro e capim Estrela foram observados aumentos de 1,4 kg de leite/vaca (GONZALES et al., 1996). Em uma pequena propriedade familiar no Estado do Acre, o amendoim forrageiro foi utilizado como banco de proteína e resultou no aumento na produção de leite de 3,6 para 5,2 L/vaca/dia (VALENTIM et al., 2001).

Existem apenas seis cultivares catalogadas no Registro Nacional de Cultivares do Ministério da Agricultura, Pecuária e Abastecimento (MAPA, 2020). Dentre essas cultivares, Amarillo e Belomonte são as mais utilizadas no mercado.

A cultivar Amarillo foi a primeira a ser lançada. Foi obtida a partir do primeiro acesso de *A. pintoi* coletado em 1954, por Geraldo Pinto no vale do rio Jequitinhonha, o qual chegou até o Centro Internacional de Agricultura Tropical – CIAT, onde recebeu a identificação CIAT 17434. Foi levada à Austrália, onde foi lançada com o nome de Amarillo. Esse genótipo foi liberado comercialmente na Colômbia em 1992 (Maní Forrajero Perene), no México e Honduras em 1993 (Pico Bonito), na Costa Rica em 1994 (Maní Mejorador), e no Brasil em 1995 (Matsuda Genética 100 ou MG-100), onde recebeu o BRA 013251 (VALLS, 1992; BARCELLOS et al., 2000).

A cultivar Belomonte foi lançada no Brasil em 1999, tendo sua propagação por estolões, uma vez que possui baixa produção de sementes (PEREIRA et al., 1999; PAGANELLA; VALLS, 2002). Em estudos realizados no estado do Acre, a cultivar Belomonte apresentou uma produção de matéria seca em torno de 15 a 21 $t.ha^{-1}.ano^{-1}$, cerca de 19% de proteína bruta e 60 a 70% de digestibilidade, além possuir uma boa compatibilidade em consorciação com gramíneas, apresentando ganho de peso $vivo.dia^{-1}$ em bovinos maior do que o monocultivo de gramínea adubada com nitrogênio (VALENTIM et al., 2000; 2001).

A cultivar BRS Mandobi foi obtida por meio de seleção massal, através de uma rede de avaliação de acessos instalada em 1999 (ASSIS; VALENTIM, 2009). Foi registrada em 2008 no Registro Nacional de Cultivares, e foi protegida segundo as normas do Ministério da Agricultura, Pecuária e Abastecimento em 2011. É uma cultivar bem adaptada à região tropical e equatorial, com boa produtividade de biomassa, bom estabelecimento, boa tolerância a solos bem drenados ou com baixa permeabilidade (ASSIS, 2011). A principal característica dessa cultivar é a elevada produção de sementes, aproximadamente 3 $t.ha^{-1}.ano^{-1}$ de sementes puras, após 18 a 21 meses de plantio. Tal característica a coloca como uma alternativa para a redução de custo do plantio do amendoim forrageiro pelos produtores (ASSIS et al., 2013).

Apesar disso, ainda existe a necessidade de desenvolvimento de novas cultivares. Por esse motivo, o programa de melhoramento de amendoim forrageiro tem o objetivo de desenvolver novas cultivares adaptadas as diversas condições edafoclimáticas do território brasileiro, as quais poderão reduzir o custo de implementação da cultura e torná-la mais competitiva. Com a evolução das técnicas da genética molecular, o conhecimento sobre a

espécie tem sido ampliado. Por esse motivo, os programas de melhoramento têm investido cada vez mais na genética molecular como um suporte.

3.2 Biologia do amendoim forrageiro

O gênero *Arachis* foi descrito por Linnaeus (1753), a partir do amendoim comum, *Arachis hypogaea*, que é o seu representante com maior importância econômica. O gênero é formado por espécies diplóides (*A. pintoi*: 2n = 2x = 20) e tetraplóides (*A. hypogaea*: 2n = 4x = 40), pertencendo à família Fabaceae (LEWIS et al. 2005; JUDD et al. 2009).

Atualmente, o gênero *Arachis* está dividido em nove secções (*Arachis*, *Erectoides*, *Caulorrhizae*, *Extranervosae*, *Heteranthae*, *Procumbentes*, *Rhizomatosae*, *Trierectoides* e *Triseminatae*), de acordo com seu modo de reprodução, morfologia e modo de dispersão ao longo da América do Sul (VALLS; SIMPSON, 2005; KRAPOVICKAS; GREGORY, 2007).

As flores possuem variação de cor (branca, creme, amarela e laranja), sendo formadas por um estandarte que possui quilha pontiaguda, curvada e aberta ventralmente na base, muito delgada (ARGEL; PIZARRO, 1992; SIMPSON et al., 1994; ASSIS et al., 2010). São hermafroditas com os órgãos reprodutivos protegidos por duas pétalas em formato de asas (Figura 2). Suas flores são frágeis, murcham e perdem a coloração, aproximadamente, seis horas após a antese, que ocorre por volta de 6h da manhã, com pico da floração entre 8h15 às 12h00, nos meses de novembro a meados de abril (COSTA, 2012). A receptividade do estigma pode variar ao longo do dia entre os genótipos, com pico de 90% até 7h30 da manhã e queda de 50% a partir das 9h30 (CAPISTRANO, 2015). Apesar disso, os grãos de pólen já são viáveis seis a oito horas antes da antese da flor, o que favorece a autofecundação (SIMPSON et al., 1994).

Figura 2. Flor de *A. pintoi*. *: Estandarte; seta: asa.

Por outro lado, a presença de flores chamativas pode indicar a evolução de caracteres florais visando favorecer a visita de insetos (COSTA, 2012). Os principais visitantes das flores de *A. pintoi* são himenópteros e coleópteros, como *Paratrigona lineata*, *Apis mellifera*, *Diabrotica* sp, *Cerotoma* sp, *Pheidole* sp, *Camponotus* sp, *Trigona spinipes*, *Acyloscelis* sp e *Centris* sp.

Apesar de algumas peculiaridades, o comportamento dos insetos que visitam as flores é parecido. Eles pousam nas pétalas das asas da flor, empurrando-as para baixo afim de ter acesso ao pólen e, após coletado, o transferem para o terceiro par de patas, voando em seguida para outra flor e repetindo o processo (Figura 3) (NIGAM et al., 1990; COSTA, 2012).

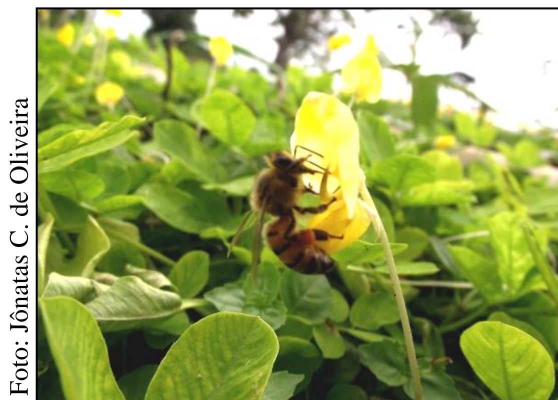


Figura 3. *Apis mellifera* coletando pólen em flor de *Arachis pintoi*.

Segundo Nigam et al. (1990), a espécie *A. hypogaea* é autógama, porém cerca de 10% das flores têm polinização cruzada, em decorrência da visita de insetos como *Apis melífera* e membros do gênero *Bombus*, *Megachiline*, *Nomia*, *Pithitis* e *Lasioglossum*.

Diversos autores confirmam a presença de polinizadores nas flores de *Arachis* (HAMMONS, 1963; NIGAM et al., 1990; KRAPOVICKAS; GREGORY, 2007; DRUMOND; CARDOSO, 2010; COSTA, 2012). Entretanto, em sua busca por pólen, esses insetos podem tanto predar as flores impedindo a produção dos frutos e do “peg”, como também contribuir para a polinização, mesmo que indiretamente, proporcionando tanto a polinização cruzada como autopolinização da flor.

3.3 Estudos genéticos com amendoim forrageiro

3.3.1 Marcadores moleculares

O uso de marcadores moleculares em estudos com plantas tem sido cada vez maior, especialmente, por permitir a identificação precisa de diferenças (polimorfismo) entre genótipos. Os marcadores moleculares baseados no DNA permitem que o polimorfismo seja visualizado sem a interferência do ambiente, o que não ocorre com marcadores morfológicos (RAMALHO et al., 2012). Dentre os marcadores moleculares, os microssatélites, também conhecidos como SSRs (*Simple Sequence Repeats*), são sequências curtas de DNA repetitivo compostas de um a seis nucleotídeos que ocorrem em todos os organismos (KALIA et al., 2011; RAMALHO et al., 2012). Sua amplificação é realizada *in vitro* por meio de PCR (*Polimerase Chain Reaction*), o que permite que pequenas quantidades de DNA possam ser analisadas (KALIA et al., 2011).

As principais vantagens desses marcadores são: natureza codominante, que possibilita a identificação de indivíduos heterozigotos; seu alto grau de polimorfismo, permite detectar um grande número de alelos por loco; regiões flanqueadoras aos microssatélites, em geral, são altamente conservadas dentro de cada gênero, o que permite a transferência entre espécies (HODEL et al., 2016). A principal desvantagem é o custo envolvido em seu desenvolvimento.

No gênero *Arachis* existem, aproximadamente, 4,5 mil marcadores microssatélites (HE et al., 2003; FERGUSON et al., 2004; MORETZSOHN et al., 2005; MARTINS et al., 2006; GIMENES et al., 2007; PROITE et al., 2007; CUC et al., 2008; GUO et al., 2008; NAITO et al., 2008; LIANG et al., 2009; QIN et al., 2012; SHIRASAWA et al., 2012, 2013; TANG et al., 2012; WANG et al., 2012; ZHANG et al., 2012; HUANG et al., 2016; ZHOU et al., 2016). A maioria foi desenvolvida para o amendoim comum (*A. hypogaea*), em virtude de sua importância econômica.

Para o amendoim forrageiro (*A. pintoi*) existem atualmente 25 locos (PALMIERI et al., 2002; PALMIERI et al., 2005; PALMIERI et al., 2010). Estes locos apresentaram elevados índices de polimorfismo e transferibilidade dentro do gênero *Arachis* (BRAVO et al., 2006; GIMENES et al., 2007; ANGELICI et al., 2008; PALMIERI et al., 2010; AZÊVEDO et al., 2016). Um protocolo foi desenvolvido para identificar híbridos em cruzamentos controlados de amendoim forrageiro, indicando os melhores locos (CAMPOS et al., 2016). Essa é uma etapa essencial para o desenvolvimento de novas cultivares, pois permite a identificação entre

genótipos que, na maior parte dos casos, não podem ser identificados através de descritores morfológicos.

Quando se considera características ideais para genotipagem, como ausência de bandas inespecíficas e elevado polimorfismo, esse número se restringe a dez locos com adequado padrão (AZÊVEDO et al., 2016). Dessa forma, há a necessidade de ampliar a quantidade de marcadores, e novas metodologias têm sido utilizadas. Dentre essas, o RNA-Seq (*RNA sequencing*) é uma técnica robusta e tem permitido o estudo do genoma funcional de espécies que ainda não possuem sequências genômicas disponíveis (WIT et al., 2012).

Com o uso do RNA-Seq é possível desenvolver milhares de SSRs com menor custo e esforço do que as técnicas tradicionais (TAHERI et al., 2018). No gênero *Arachis* essa técnica tem sido empregada, principalmente, no amendoim comum (*A. hypogaea*) e nas espécies silvestres mais próximas filogeneticamente (*A. ipaensis* e *A. duranensis*) (CHOPRA et al., 2014). Assim, para estas espécies existem cerca de 250 mil SSRs desenvolvidos a partir de sequências genômicas e de transcriptomas depositadas em bancos de dados públicos (ZHANG et al., 2012; PENG et al., 2016; LOU et al., 2017; WANG et al., 2018).

3.3.2 Estudos moleculares associados ao programa de melhoramento

3.3.2.1 Caracterização da diversidade genética molecular de amendoim forrageiro

Uma das formas de conservação de espécies é *ex situ*, na qual as espécies são conservadas fora do seu ambiente natural. A conservação de uma espécie em banco ativo de germoplasma (BAG) tem por objetivo preservar a variabilidade genética, evitando a perda de alelos. Para este fim, são coletadas amostras (acessos) silvestres e domesticadas, visando obter a representatividade da máxima variabilidade genética possível da espécie conservada (BORÉM; MIRANDA, 2009). No entanto, os bancos de germoplasma podem ter uma grande quantidade de acessos e, em muitos casos, pode haver duplicatas, o que aumenta o custo e trabalho na manutenção (MOURA et al., 2013). Por esse motivo, estudos de caracterização molecular dos BAGs são importantes, pois permitem a identificação de acessos redundantes, além de fornecer maiores informações sobre a divergência genética que podem auxiliar no trabalho dos pesquisadores.

Azêvedo (2014) avaliou a diversidade genética dos 145 acessos do BAG de amendoim forrageiro das espécies *A. pintoi, A. repens, A. glabrata, A. helodes* e híbridos intra e interespecíficos localizados na Embrapa Acre, com marcadores microssatélites. Foram

encontrados valores elevados de diversidade genética e ausência de duplicatas, indicando que os acessos possuem uma boa representatividade da base genética. Não houve a detecção de grupos definidos diferenciando *A. pintoi* e *A. repens*, confirmando a similaridade filogenética compartilhada entre essas espécies (AZÊVEDO et al., 2016).

Os dados de genotipagem também forneceram informações para a estimativa de uma coleção nuclear com base molecular. Foram identificados 15 acessos representativos da diversidade do BAG existente para *A. pintoi* e *A. repens* (AZÊVEDO et al., 2015). Dentre os acessos da coleção nuclear, houve também a representação de características morfológicas únicas no germoplasma: a maior média para o comprimento do folíolo basal, único acesso com flor branca, presença de florescimento precoce. Assim, a análise molecular conseguiu discriminar genótipos divergentes. Conclui-se que houve a detecção de um número mínimo de acessos para ser usado como coleção de trabalho e referência da diversidade da espécie.

Os resultados da diversidade molecular de amendoim forrageiro forneceram informações inéditas e valiosas para o conhecimento e o avanço do melhoramento na espécie. A divergência molecular também poderá ser usada juntamente à caracterização morfológica para escolha de genitores do programa de melhoramento.

3.3.2.2 Estimativa da reprodução cruzada em amendoim forrageiro

O sistema reprodutivo define como os genes são transmitidos para a próxima geração. Por isso, seu conhecimento é fundamental para estudos de conservação *ex situ* e melhoramento de plantas, pois permite delinear estratégias que aperfeiçoem a amostragem da variabilidade genética, além de nortear as melhores formas de multiplicação de sementes e os modelos mais adequados de melhoramento. Apesar disso, informações sobre a taxa de cruzamento no gênero *Arachis*, são limitadas a espécie *A. hypogaea*, o amendoim comum (COFFELT, 1989; KNAUFT et al., 1992) e são baseadas em marcadores morfológicos.

Em relação ao sistema reprodutivo, as plantas podem ser classificadas em autógamas, alógamas e mistas. Os métodos tradicionais de determinação do sistema de reprodução consistem na observação dos cruzamentos, comportamento dos polinizadores, exame da morfologia floral e resultados de experimentos controlados de polinização (MORAES; MONTEIRO, 2002).

A determinação do sistema reprodutivo pode ser feita por meio de marcadores moleculares, que possuem muitos locos com alelos codominantes e segregantes, e são

encontrados frequentemente nas populações, possibilitando a obtenção de estimativas mais acuradas da taxa de cruzamento (RITLAND; JAIN, 1981).

Para avaliar a reprodução cruzada em *A. pintoi*, Oliveira (2015) analisou 14 acessos do BAG localizado na Embrapa Acre. A taxa de cruzamento variou entre os acessos analisados (2% a 91%), com valor médio de 36% para a espécie (Tabela 1).

Tabela 1. Estimativas da taxa de cruzamento para os acessos de *A. pintoi*. t_m: taxa de cruzamento multilocos.

Acessos	t_m
V 6727	0,704 ±0,071
V 6784	0,625 ±0,154
W 34	0,173 ±0,074
W 647	0,564 ±0,062
V 5895	0,168 ±0,065
V 6740	0,792 ±0,096
V 13196	0,118 ±0,040
V 13198	0,425 ±0,074
V 6791wf	0,309 ±0,064
Belomonte	0,916 ±0,001
W 1000	0,265 ±0,094
BRS Mandobi	0,864 ±0,193
Amarillo MG-100	0,019 ±0,000
V 13888	0,266 ±0,062
A. pintoi	0,367 ±0,076

Esses valores indicaram que a espécie possui um sistema de cruzamento misto com predominância de autogamia, e que os polinizadores podem influenciar significativamente a taxa de cruzamento. Até a realização desse estudo, acreditava-se que as espécies do gênero *Arachis* eram autógamas restritas devido aos estudos com marcadores morfológicos em amendoim comum que detectaram valores de fecundação cruzada entre 1,5 e 8% (COFFELT, 1989; KNAUFT et al., 1992) e devido a anatomia do sistema floral (COSTA, 2012). O resultado da estimativa molecular conseguiu detectar com maior acurácia o fluxo gênico entre os acessos e comprovou uma mudança conceitual na perspectiva da biologia reprodutiva da espécie.

A partir dessas informações, novas estratégias de conservação do BAG devem ser adotadas para evitar a atuação dos polinizadores, o que pode resultar no efetivo cruzamento entre acessos da coleção. Além disso, as estratégias para melhoramento de *A. pintoi* também devem ser alteradas, o que interfere na forma como são realizados os cruzamentos, o avanço de gerações e produção de sementes, por exemplo.

4. REFERÊNCIAS

ANGELICI, C. M. L. C. D.; HOSHINO, A. A.; NÓBILE, P. M.; PALMIERI, D. A.; VALLS, J. F. M.; GIMENES, M. A.; LOPES, C. R. Genetic diversity in section *Rhizomatosae* of the genus *Arachis* (Fabaceae) based on microsatellite markers. **Genetics and Molecular Biology**, v. 31, n. 1, p. 79-88, 2008.

ARGEL, P. J.; PIZARRO, E. A. Germoplasm case study: *Arachis pintoi*. In.: **Pastures for the tropical lowlands – CIAT's Contribution**. Cali: CIAT, 1992. pp. 57-76.

ASSIS, G. M. L. **Produção de sementes de *Arachis pintoi* cv. BRS BRS Mandobi no Acre**. 2011. Disponível em: <http://sistemasdeproducao.cnptia.embrapa.br/FontesHTML/Amendoim/ProducaoSementesArachisAcre/index.htm>. Acesso em: 13 mai. 2020.

ASSIS, G. M. L.; VALLS, J. F. M.; CARVALHO, M. A.; VALENTIM, J. F.; ANDRADE,C. M. S. **Descritores morfológicos para condução de ensaios de distinguibilidade, homogeneidade e estabilidade em *Arachis pintoi* Krapov. & W.C. Greg**. Rio Branco: Embrapa Acre, 2010.

ASSIS, G. M. L.; VALENTIM, J. F. Forage peanut breeding program in Brazil. **Proceedings of the 2nd International Symposium of Forage Breeding**, 2009.

ASSIS, G. M. L.; VALENTIM, J. F.; ANDRADE, C. M. S. BRS Mandobi: a new forage peanut cultivar propagated by seeds for the tropics. **Tropical Grasslands-Forrajes Tropicales**, v. 1, n. 1, p. 39-41, 2013.

AZÊVEDO, H. S. F. S. **Caracterização da diversidade genética de amendoim forrageiro com marcadores microssatélites**. 2014. Dissertação (Mestrado em Ciência, Inovação e Tecnologia para Amazônia), Universidade Federal do Acre, Rio Branco, 2014.

AZÊVEDO, H. S. F. S.; ALVES FILHO, J.; SOUSA, A. C. B.; ASSIS, G. M. L.; CAMPOS, T. A coleção nuclear de germoplasma de amendoim forrageiro. **I Congresso Regional de Pesquisa do Estado do Acre e XXIV Seminário de Iniciação Científica da UFAC**, 2015.

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 Paranã River valleys of Brazil. **Genetics and Molecular Research**, v.15, n. 3, p. 1-11, 2016.

BARCELLOS, A. O.; ANDRADE, R. P.; KARIA, C. T.; VILELA, L. Potencial e uso de leguminosas forrageiras dos gêneros *Stylosanthes*, *Arachis* e *Leucena*. **Simpósio sobre Manejo da Pastagem**, 2000.

BARRETO, F. Z.; ROSA, J. R. B. F.; BALSALOBRE, T. W. A.; PASTINA, M. M.; SILVA, R. R.; HOFFMANN, H. P.; SOUZA, A. P.; GARCIA, A. A. F.; CARNEIRO, M. S. A genome-wide association study identified loci for yield component traits in sugarcane (*Saccharum* spp.). **PloS One**, v. 14, n. 7, p. e0219843, 2019.

BORÉM, A.; CAIXETA, E. T. **Marcadores moleculares**. 2ª ed. Editora UFV, 2009.

BRAVO, J. P.; HOSHINO, A. A.; ANGELICI, C. M. L. C.; LOPES, C. R.; GIMENES, M. A. Transferability and use of microsatellite markers for the genetic analysis of the germplasm of some *Arachis* section species of the genus *Arachis*. **Genetics and Molecular Biology**, v. 29, p. 516-524, 2006.

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.** 2016. Disponível em: <https://ainfo.cnptia.embrapa.br/digital/bitstream/item/156276/1/26263.pdf>. Acesso em: 22 jun. 2020.

CAPISTRANO, M. da C. **Fatores determinantes na fertilidade dos gametas e conservação de polens do amendoim forrageiro (*Arachis pintoi*) no Acre**. Dissertação (Mestrado em Produção Vegetal), Faculdade de Agronomia, Universidade Federal do Acre, Rio Branco, 2015.

CHOPRA, R.; BUROW, G.; FARMER, A.; MUDGE, J.; SIMPSON, C. E.; BUROW, M. D. Comparisons of *de novo* transcriptome assemblers in diploid and polyploid species using peanut (*Arachis* spp.) RNA-seq data. **PloS One**, v.9, n.12, p. 1-16, 2014.

COFFELT, T. A. Natural crossing of peanut in Virginia. **Peanut Science**. v. 16, p. 46-48, 1989.

CONTRERAS-SOTO, R. I.; MORA, F.; DE OLIVEIRA, M. A.; HIGASHI, W.; SCAPIM, C. A.; SCHUSTER, I. A Genome-Wide Association Study for Agronomic Traits in Soybean Using SNP Markers and SNP-Based Haplotype Analysis. **PloS One**, v. 12, n. 2, p. e0171105, 2017.

COSTA, L. C. **Biologia floral de espécies do gênero Arachis L. (Fabaceae-Papilonoideae), com ênfase em aspectos da morfologia floral e na anatomia de ovários**. Tese (Doutorado em Botânica), Universidade de Brasília, Brasília, 2012.

CUC, L. M.; MACE, E. S.; CROUCH, J. H.; QUANG, V. D.; LONG, T. D.; VARSHNEY, R. K. Isolation and characterization of novel microsatellite markers and their application for diversity assessment in cultivated groundnut (*Arachis hypogaea* L.). **BMC Plant Biology**, v. 8, n. 1, p. 1-11, 2008.

DRUMOND, P. M.; CARDOSO, G. A. **As abelhas e a produção de sementes do amendoim forrageiro**. Rio Branco: Embrapa Acre, 2010. 8 p.

EMBRAPA. Empresa Brasileira de Pesquisa Agropecuária. **Balanço Social 2019**. 2020. Disponível em: <https://bs.sede.embrapa.br/2019/balancosocialembrapa2019print.pdf> Acesso em: 14 ago. 2020.

FERGUSON, M. E.; BUROW, M. D.; SCHULZE, S. R.; BRAMEL, P. J.; PATERSON, A. H.; KRESOVICH, S.; MITCHELL, S. Microsatellite identification and characterization in peanut (*A. hypogaea* L.). **Theoretical and Applied Genetics**, v. 108, n. 6, p. 1064-1070, 2004.

GIMENES, M. A.; HOSHINO, A. A.; BARBOSA, A. V. G.; PALMIERI, D. A.; LOPES, C. R. Characterization and transferability of microsatellite markers of the cultivated peanut (*Arachis hypogaea*). **BMC Plant Biology**, v. 7, n. 1, p. 1-13, 2007.

GONZALES, M. S.; NEURKVAN, L. M.; ROMERO, F.; PEZO, D. A.; ARGEL, P. J. Producion de leche en pasturas de estrella africana (*Cynodon nlemfluensis*) solo y associado com *Arachis pintoi* o *Desmodium ovalifolium*. **Pasturas tropicales**, v. 18, n. 1, p. 2-12, 1996.

GUO, B. Z.; CHEN, X.; DANG, P.; SCULLY, B. T.; LIANG, X.; HOLBROOK, C. C.; YU, J.; CULBREATH, A. K. Peanut gene expression profiling in developing seeds at diferente reproduction stages during *Aspergillus parasiticus* infection. **BMC Developmental Biology**, v. 8, n. 1, p. 1-16, 2008.

HAMMONS, R. O. Artficial cross-pollination of the peanut with bee-collected pollen. **Crop Science,** v. 3, p. 562-563, 1963.

HE, G. M.; NEWMAN, M. G. G.; PITTMAN, R. N.; PRAKASH, C. S. Microsatellites as DNA markers in cultivated peanut (*A. hypogaea* L.). **BMC Plant Biology**, v. 3, n. 1, p. 1-6, 2003.

HODEL, R. G. J.; SEGOVIA-SALCEDO, M. C.; LANDIS, J. B.; CROWL, A. A.; SUN, M.; LIU, X.; GITZENDANNER, M. A.; DOUGLAS, N. A.; GERMAIN-AUBREY, C. C.; CHEN, S.; SOLTIS, D. E.; SOLTIS, P. S. The report of my death was an exaggeration: a review for researchers using microsatellites in the 21st century. **Applications in Plant Sciences**, v. 4, n. 6, p. 1600025, 2016.

HUANG, L.; WU, B.; ZHAO, J.; LI, H.; CHEN, W.; ZHENG, Y.; REN, Y.; ZHOU, X.; LEI, Y.; LIAO, B.; JIANG, H. Characterization and transferable utility of microsatellite markers in the wild and cultivated *Arachis* species. **PloS One**, v. 11, n. 5, p. 1-15, 2016.

JUDD, W. S., CAMPBELL, C. S., KELLOGG, E. A., STEVENS, P. F., DONOGHUE, M. J. **Sistemática Vegetal: Um Enfoque Filogenético**. 3º ed. Porto Alegre: Artmed, 2009. 612p.

KALIA, R.K.; RAI, M.K.; KALIA, S.; SINGH, R.; DHAWAN, A.K. Microsatellite markers: na overview of the recent progress in plants. **Euphytica**, v. 177, n. 3, p. 309-334, 2011.

KNAUFT, D. A.; CHIYEMBEKEZA, A. J.; CORBET, D. W. Possible Reproductive Factors Contributing to Outcrossing in Peanut (*Arachis hypogaea* L.). **Peanut Science,** v. 19, p. 29-31, 1992.

KRAPOVICKAS, A.; GREGORY, W. C. Taxonomy of the genus *Arachis* (Leguminosae). (Translated by D.E. Williams and C.E. Simpson.) **Bonplandia**, v. 16, p. 1–205, 2007.

LASCANO, C.E. Nutritive value and animal production of forage *Arachis*. In: KERRIDGE, P.C.; HARDY, B. **Biology and agronomy of forage Arachis**. CIAT, 1994.

LEWIS, G. P; MACKINDER, B.; LOCK, M. **Legumes of the World**. Kew: Royal Botanic Gardens, 2005. 578p.

LIANG X.; CHEN, X.; HONG, Y.; LIU, H.; ZHOU, G.; LI, S.; GUO, B. Utility of EST-derived SSR in cultivated peanut (*Arachis hypogaea* L.) and *Arachis* wild species. **BMC Plant Biology**, v. 9, n. 1, p. 1-9, 2009.

LUO, H.; XU, Z.; LI, Z.; LI, X.; LV, J.; REN, X.; et al. Development of SSR markers and identification of major quantitative trait loci controlling shelling percentage in cultivated peanut (*Arachis hypogaea* L.). **Theoretical and Applied Genetics**, v. 130, n. 8, p. 1635-1648, 2017.

MACHADO, M. L. C. **Desempenho de novilhos Nelore castrados e não-castrados sob pastejo em *Brachiaria humidicola* pura e em consórcio com amendoim forrageiro**. Dissertação (Mestrado em Sanidade e Produção Animal Sustentável na Amazônia Ocidental), Universidade Federal do Acre, Rio Branco, 2017.

MAIA, G. F. N. **Desempenho produtivo de dois grupos genéticos de bovinos de corte em pastos puros e consorciados na Amazônia Ocidental**. Dissertação (Mestrado em Sanidade e Produção Animal Sustentável na Amazônia Ocidental), Universidade Federal do Acre, Rio Branco, 2018.

MAPA. MINISTÉRIO DA AGRICULTURA, PECUÁRIA E ABASTECIMENTO. **CultivarWeb**. 2020. Disponível em: <http://sistemas.agricultura.gov.br/snpc/cultivarweb/cultivares_registradas.php>. Acesso em: 24 set. 2020.

MARTINS, W. S. D.; PROITE, K. P.; MORETZSOHN, M.; BERTIOLI, D. J. New *softwares* for automated microsatellite marker development. **Nucleic Acids Research**, v. 34, n. 4, p. 1-4, 2006.

MORAES, P. L. R.; MONTEIRO, R. Taxas de cruzamento em uma população natural de *Cryptocarya moschata* Nees (Lauraceae). **Biota Neotropica**, v. 2, n. 2, p. 1-10, 2002.

MORETZSOHN, M. C.; LEOI, L.; PROITE, K.; GUIMARÃES, P. M.; LEAL-BERTIOLI, S. C. M.; GIMENES, M. A.; et al. A microsatellite based, gene-rich linkage map for the AA genome of *Arachis* (Fabaceae). **Theoretical and Applied Genetics**, v. 111, n. 6, p. 1060-1071, 2005.

MORRIS, G. P.; RAMU, P.; DESHPANDE, S. P.; HASH, C. T.; SHAH, T.; UPADHYAYA, H. D.; RIERA-LIZARAZU, O.; BROWN, P. J.; ACHARYA, C. B.; MITCHELL, S. E.; HARRIMAN, J.; GLAUBITZ, J. C.; BUCKLER, E. S.; KRESOVICH, S. Population genomic and genome-wide association studies of agroclimatic traits in sorghum. **Proceedings of the National Academy of Sciences**, v. 110, p. 453-458, 2013.

NAITO, Y.; SUZUKI, S.; IWATA, Y.; KUBOYAMA, T. Genetic diversity and relationship analysis of peanut germplasm using SSR markers. **Breeding Science**, v. 58, n. 3, p. 293-300, 2008.

NIE, X.; HUANG, C.; YOU, C.; LI, W.; ZHAO, W.; SHEN, C.; ZHANG, B.; WANG, H.; YAN, Z.; DAI, B.; WANG, M.; ZHANG, X.; LIN, Z. Genome-wide SSR-based association mapping for fiber quality in nation-wide upland cotton inbreed cultivars in China. **BMC Genomics**, v. 17, p. 352-367, 2016.

NIGAM, S. N.; RAO, M. J. V.; GIBBONS, R. W. **Artificial hybridization in groundnut**. Patancheru: International Crops Research Institute for the Semi-Arid Tropics, 1990. 27p.

OLIVEIRA, J. C. **Estimativa da taxa de cruzamento e diversidade genética em *Arachis pintoi* com marcadores microssatélites**. Dissertação (Mestrado em Ciência, Inovação e Tecnologia para a Amazônia), Universidade Federal do Acre, Rio Branco, 2015.

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, 2019.

PAGANELLA, M. B.; VALLS, J. F. M. Caracterização morfológica de cultivares e acessos selecionados de *Arachis pintoi* Krapov. & Gregory. **Pasturas Tropicales**, v. 24, n. 2, p. 23-30, 2002.

PALMIERI, D. A.; BECHARA, M. D.; CURI, R. A.; GIMENES, M. A.; LOPES, C. R. Novel polymorphic microsatellite markers in section *Caulorrhizae* (*Arachis*, Fabaceae). **Molecular Ecology Notes**, v. 5, n. 1, p. 77-79, 2005.

PALMIERI, D. A.; BECHARA, M. D.; CURI, R. A.; MONTEIRO, J. P.; VALENTE, S. E. S.; GIMENES, M. A.; LOPES, C. R. Genetic diversity analysis in the section *Caulorrhizae* (genus *Arachi*s) using microsatellite markers. **Genetics and Molecular Biology**, v. 33, n. 1, p. 109-118, 2010.

PALMIERI, D. A.; HOSHINO, A. A.; BRAVO, J. P.; LOPES, C. R.; GIMENES, M. A. Isolation and characterization of microsatellite loci from the forage species Arachis pintoi (Genus *Arachis*). **Molecular Ecology Notes**, v. 2, n. 4, p. 551-553, 2002.

PENG, Z.; GALLO, M.; TILLMAN, B. L.; ROWLAND, D.; WANG, J. Molecular marker development from transcript sequences and germplasm evaluation for cultivated peanut (*Arachis hypogaea* L.). **Molecular Genetics and Genomics**, v. 291, n. 1, p. 363-381, 2016.

PEREIRA, J. M.; RESENDE, C. P.; SANTANA, J. R. **Amendoim forrageiro cv. Belomonte (*Arachis pintoi* Krapov & Gregory): Uma nova opção de leguminosa forrageira**. 1999. Disponível em: <http://www.ceplac.gov.br/radar/amendoim%20forrageiro.htm> Acesso em: 20 fev. 2020.

POCZAI, P.; VARGA, I.; LAOS, M.; CSEH, A.; BELL, N.; VALKONEN, J. P. T.; HYVÖNEN, J. Advances in plant gene-targeted and functional markers: a review. **Plant Methods**, v. 9, n. 6, p. 1-31, 2013.

PROITE, K.; LEAL-BERTIOLI, S. C. M.; BERTIOLI, D. J.; MORETZSOHN, M. C.; SILVA, F. R.; MARTINS, N. F.; GUIMARÃES, P. M. ESTs from a wild *Arachis* species for gene Discovery and marker development. **BMC Plant Biology**, v. 7, n. 1, p. 1-10, 2007.

QIN, H.; FENG, S.; CHEN, C.; GUO, Y.; KNAPP, S.; CULBREATH, A.; HE, G.; WANG, M. L.; ZHANG, X.; HOLBROOK, C. C.; OZIAS-AKINS, P.; GUO, B. An integrated genetic linkage map of cultivated peanut (*Arachis hypogaea* L.) constructed from two RIL populations. **Theoretical and Applied Genetics**, v. 124, n. 4, p. 653-664, 2012.

RAFALSKI, A. Applications of single nucleotide polymorphisms in crop genetics. **Current Opinion in Plant Biology,** v. 5, p. 94-100, 2002.

RAMALHO, M.A.P.; SANTOS, J.B.; PINTO, C.A.B.; SOUZA, E.A.; GONÇALVES, F.M.A.; SOUZA, J.C. **Genética na Agropecuária**. 5ª ed. Editora UFLA, 2012.

RESENDE, F. V.; VIDAL, M. C.; SOUZA, R. B. **Plantio direto de tomate sobre coberturas vivas em sistema orgânico de produção**. Distrito Federal: Embrapa Hortaliças, 2020.

RITLAND, K.; JAIN, S. K. A model for the estimation of outcrossing rate and gene frequencies using independent loci. **Heredity**, v. 47, p. 35-52, 1981.

ROSE, T. J.; KEARNEY, L. J.; MORRIS, S.; VAN ZWIETEN, L.; ERLER, D. V. Pinto peanut cover crop nitrogen contributions and potential to mitigate nitrous oxide emissions in subtropical coffee plantations. **Science of the Total Environment**, v. 656, p. 108-117, 2019.

SANTOS, J. C. F.; CUNHA, A. J.; MELO, B. Soil cover and weed control on coffee intercropping perennial legume. **International Journal of Applied Science and Technology**, v. 4, n. 4, p. 149-157, 2014.

SHIRASAWA, K.; BERTIOLI, D. J.; VARSHNEY, R. K.; MORETZSOHN, M. C.; LEALBERTIOLI, S. C. M; THUDI, M.; et al. Integrated consensus map of cultivated peanut and wild relatives reveals structures of the A and B genomes of *Arachis* and divergence of the legume genomes. **DNA Research**, v. 20, n. 2, p. 173-184, 2013.

SHIRASAWA, K.; KOILKONDA, P.; AOKI, K.; HIRAKAWA, H.; TABATA, S.; WATANABE, M.; et al. In silico polymorphism analysis for the development of simple sequence repeat and transposon markers and construction of linkage map in cultivated peanut. **BMC Plant Biology**, v. 12, n. 1, p. 1-13, 2012.

SILVA, G. P. P.; RESENDE, F. V.; PEREIRA, T. S.; SOUZA, R. B.; ALBUQUERQUE, J. O.; VIDAL, M. C.; SOUSA, J. M. M. **Desempenho agronômico de cultivares de pepino em ambiente protegido cultivado em solo com cobertura viva de amendoim forrageiro em sistema orgânico de produção**. 2012. Disponível em: <https://www.alice.cnptia.embrapa.br/bitstream/doc/930040/1/IIJCEHGeanny.pdf>. 2012. Acesso em: 08 out. 2020.

SIMPSON, C. E.; VALLS, J. F. M.; MILES, J. W. Reproductive biology and potencial for recombination in *Arachis*. In: P. C. KERRIDGE e B. HARDY, eds., **Biology and Agronomy of Forage *Arachis***. Cali: Centro Internacional de Agricultura Tropical (CIAT), 1994. pp. 43-52.

SWAIN, M.; BLOMQVIST, L.; McNAMARA, J.; RIPPLE, W. J. Reducing the environmental impact of global diets. **Science of the Total Environment**, v. 610, p. 1207–1209, 2018.

TAHERI, S.; LEE ABDULLAH, T.; YUSOP, M. R.; HANAFI, M. M.; SAHEBI, M.; AZIZI, P.; SHAMSHIRI, R. R. Mining and development of novel SSR markers using next generation sequencing (NGS) data in plants. **Molecules**, v. 23, n. 2, p. 399-419, 2018.

TANG, M.; CHEN, Y.; REN, X.; HUANG, L.; ZHOU, X.; YAN, H.; JIANG, H. Genetic diversity of *Arachis* accessions by EST-SSR from cultivated peanut (*Arachis hypogaea* L.). **Acta Agronomica Sinica**, v. 38, p. 1221-1231, 2012.

URBANSKI, A. S. **Consórcio de pastagens como ferramenta para aumento de produtividade animal na Amazônia Ocidental**. Dissertação (Mestrado em Sanidade e Produção Animal Sustentável na Amazônia Ocidental), Universidade Federal do Acre, Rio Branco, 2016.

VALENTIM, J. F.; ANDRADE, C. M. S. **Tropical kudzu (*Pueraria phaseoloides*): Successful adoption in sustainable cattle production systems in the western Brazilian Amazon.** 2005. Disponível em: <https://www.tropicalgrasslands.asn.au/Tropical%20Grasslands%20Journal%20archive/PDFs/Vol_39_2005/Vol_39_04_2005_pp221_221.pdf>. Acesso em: 25 fev. 2020.

VALENTIM, J. F.; CARNEIRO, J. C.; SALES, M. F. L. **Amendoim forrageiro cv. Belomonte para a diversificação das pastagens e conservação do solo no Acre**. Rio Branco: Embrapa-Acre, 2001. 18p.

VALENTIM, J. F.; CARNEIRO, J. C.; VAZ, F. A.; SALES, M. F. L. **Produção de mudas de *Arachis pintoi*.** Rio Branco: Embrapa Acre, 2000. 4 p.

VALLS, J. F. M. Origem do germoplasma de *Arachis pintoi* disponível no Brasil. **Red International de Evaluación de Pastos Tropicales – RIEPT**, 1992.

VALLS, J.F.M.; SIMPSON. New species of *Arachis* (Leguminosae) from Brazil, Paraguay and Bolivia. **Bonplandia**, v.14, p.35-65, 2005.

WANG, H.; LEI, Y.; YAN, L.; WAN, L.; CAI, Y.; YANG, Z.; LV, J.; ZHANG, X.; XU, C.; LIAO, B. Development and validation of simple sequence repeat markers from *Arachis hypogaea* transcript sequences. **The Crop Journal**, v. 6, n. 2, p. 172-180, 2018.

WANG, H.; PENMETSA, R. V.; YUAN, M.; GONG, L.; ZHAO, Y.; GUO, B.; et al. Development and characterization of BAC-end sequence derived SSRs, and their incorporation into a new higher density genetic map for cultivated peanut (*Arachis hypogaea* L.). **BMC Plant Biology**, v. 12, n. 1, p. 1-11, 2012.

WANG, Y. X.; WENG, B. Q.; YE, J.; ZHONG, Z. M.; HUANG, Y. B. Carbon sequestration in a nectarine orchard as affected by green manure in China. **European Journal of Horticultural Science**, v. 80, n. 5, p. 208-215, 2015.

WIT, P.; PESPENI, M. H.; LADNER, J. T.; BARSHIS, D. J.; SENECA, F.; JARIS, H.; et al. The simple fool's guide to population genomics via RNA-Seq: an introduction to high-throughput sequencing data analysis. **Molecular Ecology Resources**, v. 12, n. 6, p. 1058-1067, 2012.

ZHANG, J.; LIANG, S.; DUAN, J.; WANG, J.; CHEN, S.; CHENG, Z.; et al. *De novo* assembly and characterization of the transcriptome during seed development, and generation of genic-SSR markers in peanut (*Arachis hypogaea* L.). **BMC Genomics**, v. 13, n. 1, p. 90-96, 2012.

ZHANG, X.; WARBURTON, M. L.; SETTER, T.; LIU, H.; XUE, Y.; YANG, N.; YAN, J.; XIAO, Y. Genome-wide association studies of drought-related metabolic changes in maize using an enlarged SNP panel. **Theoretical and Applied Genetics**, v. 129, n. 8, p. 1449-1463, 2016.

ZHOU, X.; DONG, Y.; ZHAO, J.; HUANG, L.; REN, X.; CHEN, Y.; et al. Genomic survey sequencing for development and validation of single-locus SSR markers in peanut (*Arachis hypogaea* L.). **BMC Genomics**, v. 17, n. 1, p. 1-14, 2016.

CAPÍTULO I

A MODIFIED PROTOCOL FOR TOTAL RNA ISOLATION FROM FORAGE PEANUTS

Este manuscrito foi publicado na revista South American Journal of Basic Education, Technical and Technological.

SOUTH AMERICAN JOURNAL
of Basic Education, Technical and Technological

SAJEBTT, Rio Branco, UFAC
v. 7 n. 1 (2020): Edição: jan/abr p. 683-687
ISSN: 2446-4821

A MODIFIED PROTOCOL FOR TOTAL RNA ISOLATION FROM FORAGE PEANUTS

UM PROTOCOLO MODIFICADO PARA ISOLAMENTO DE RNA TOTAL DE AMENDOIM FORRAGEIRO

[1]Jônatas Chagas de Oliveira, [2]Carla Cristina da Silva, [2]Anete Pereira de Souza, [*3]Tatiana de Campos
[1]Universidade Federal do Acre – UFAC,
[2]Universidade Estadual de Campinas – Unicamp,
[3]Empresa Brasileira de Pesquisa Agropecuária – Embrapa Acre

*Autor correspondente: tatiana.campos@embrapa.br

ABSTRACT

The development of specific protocols for extraction of high-quality RNA are key advances in molecular techniques for elucidating biological processes and mechanisms. The forage peanut has significant economic value for use in mixed pastures. However, it is difficult to extract high-quality RNA from forage peanuts due to high contents of tannins, polysaccharides, and other secondary metabolites. Here, it is described an efficient method for obtaining high-quality, high-yield total RNA that is usable in downstream transcriptome analysis. These modifications helped eliminate the problems associated with the presence of high concentrations of secondary metabolites in forage peanut leaves, and lacked signs of degradation.
Keywords: Total RNA extraction, high yield, high quality, secondary metabolites.

RESUMO

O desenvolvimento de protocolos específicos para extração de RNA de alta qualidade são avanços chave em técnicas moleculares para elucidar mecanismos e processos biológicos. O amendoim forrageiro tem significante valor econômico para uso em pastagens consorciadas. Entretanto, é difícil extrair RNA de alta qualidade de amendoim forrageiro devido aos altos conteúdos de taninos, polissacarídeos e outros metabólitos secundários. Aqui, descreve-se um eficiente método para obtenção de RNA total de alta qualidade e alto rendimento que é necessário para análises posteriores do transcriptoma. Essas modificações ajudaram a eliminar os problemas associados com a presença de altas concentrações de metabólitos secundários em folhas de amendoim forrageiro e sem sinais de degradação.
Palavras-chave: Extração de RNA total, alto rendimento, alta qualidade, metabólitos secundários.

The forage peanut (*Arachis pintoi* Krapov. and W.C. Greg.) is a tropical herbaceous legume with attributes facilitating use in mixed pastures due to its desirable dry matter and high protein contents; it can improve weight gain in cattle and provide biological nitrogen fixation, thus contributing to soil fertility and recovery of degraded areas [1,2]. In 2018, it was reported that forage peanut had a positive economic impact of approximately USD 20 million in Acre State, located in the Brazilian Amazon [3]. Despite its economic importance, molecular studies of *A. pintoi* are still lacking, particularly due to difficulty in the extraction of high-quality RNA.

Enviado em: 07/04/2020 - Aprovado em: 03/06/2020

SOUTH AMERICAN JOURNAL
of Basic Education, Technical and Technological
SAJEBTT, Rio Branco, UFAC
v. 7 n. 1 (2020): Edição: jan/abr p. 683-687
ISSN: 2446-4821

A. pintoi contains a large quantity of polysaccharides and secondary metabolites such as phenolic compounds and tannins [4], which directly interfere with the quality of extracted RNA [5]. Thus, a protocol to obtain RNA of high purity and quality is desirable.

Herein, we developed a method modified from the lithium chloride (LiCl) precipitation-based method described by [6] for use in *Arabidopsis thaliana*, *Avicennia schaueriana*, *Theobroma cacao*, *Paspalum notatum*, and *Sorghum bicolor*. LiCl was effective at solving problems related to the large number of secondary metabolites present in *A. pintoi* leaves.

Incompletely expanded young leaves were harvested from healthy forage peanut plants. Approximately two leaflets (100 mg of fresh weight) were sampled into 2.0 mL microfuge tubes and placed in liquid nitrogen. After cooling the mortar and pestle with liquid N_2, add leaf materials, frozen in liquid N_2 and grind. Add 2 mL of cold extraction buffer (0.465 M Tris-HCl (pH 8.0), 0.23 M LiCl, 0.0215 M EDTA (pH 8.0), 2.33% SDS, 4.7% β-mercaptoethanol, H_2O DEPC) into the mortar, then homogenize, and transfer 1 mL to 2 mL microfuge tubes, totaling two tubes per sample which will be processed separately until 12 M LiCl precipitation step. Add 600 µL of cold acid phenol, vortex for 1 minute and centrifuge for 15 minutes at 4 °C and 12,000 rpm. Transfer the supernatant (450 – 500 µL) to a 2 mL RNAse-free microfuge tube and add an equal volume of cold acid phenol. Then vortex for 1 minute and centrifuge for 15 minutes at 4 °C and 12,000 rpm. Repeat twice.

Transfer the supernatant (approximately 200 to 400 µL) to a 2 mL tube and add the same volume of chloroform. Vortex for 1 minute and centrifuge for 15 minutes at 4 °C and 12,000 rpm. Repeat 4 times. Transfer the supernatant (approximately 400 µL) in the two 2 mL tubes of each sample to a 1.5 mL RNAse free microfuge tube. Add half a volume of 12 M LiCl and gently mix. Incubate for 2 h at -80 °C. Centrifuge for 50 min at 4 °C and 12,000 rpm. Discard the supernatant and wash the pellet with 500 µL of 6 M LiCl. Do not vortex. Centrifuge for 10 min at 4 °C and 12,000 rpm. Repeat twice.

Discard the supernatant and resuspend the pellet in 300 µL DEPC H_2O. Vortex for 1 minute. Add 100 µL of 3 M sodium acetate (pH 5.2) and 800 µL ice-cold absolute ethanol. Homogenize by inversion. Incubate overnight at -20 °C or for 2 h at -80 °C. Centrifuge for 50 min at 4 °C and 12,000 rpm. Discard the supernatant and add 500 µL of ice-cold 70% ethanol. Centrifuge for 10 min at 4 °C and 12,000 rpm. Repeat twice. The pellet is dried for approximately 20 min

Enviado em: 07/04/2020 - Aprovado em: 03/06/2020

SOUTH AMERICAN JOURNAL
of Basic Education, Technical and Technological
SAJEBTT, Rio Branco, UFAC
v. 7 n. 1 (2020): Edição: jan/abr p. 683-687
ISSN: 2446-4821

at room temperature in a laminar flow cabinet, then resuspended in 30 – 50 µL DEPC H_2O and stored at -20 °C.

A spectrophotometric analysis was performed to assess the purity of the extracted total RNA. Ratios for UV absorption at A_{260}/A_{280} were recorded using a Nano-Drop ND-2000 (ThermoScientific, USA). Total RNA degradation was assessed by 1% agarose gel electrophoresis and the RNA integrity number (RIN) was obtained using an Agilent BioAnalyzer 2100 (Agilent Technologies Inc., USA). RIN ranges from one to ten, with one representing degraded RNA and ten intact RNA.

The protocol described here overcame the problem caused by severe contamination with high concentrations of secondary metabolites present in *A. pintoi* leaves, thus avoiding a loss of quality in extracted RNA [5]. Although an extraction test with the protocol developed by [6] was able to obtain RNA from *A. pintoi*, the quality and yield were lower than those obtained using our modified protocol (Table 1), demonstrating the efficiency of the modifications.

Table 1. Yield, purity, and integrity of RNA extracted from *A. pintoi* leaves (± SD, n = 10).

Protocol	RNA yield (µg)	A260/280	RIN
Original protocol [6]	1.09±0.53	1.8±0.14	2.37±0.35
Modified	21.55±3.85	1.92±0.06	6.04±2.17

The amount of RNA obtained in our study was similar to that found in *A. thaliana* [6] but lower than that obtained in *Arachis hypogaea* by other methods [5,7]. However, the protocols developed for *A. hypogaea* require the addition of proteinase K or DNase to obtain contaminantfree RNA. Our particular modifications are important because they reduce the cost and work time of researchers in extracting RNA.

The integrity of the RNA extracted from the samples using our protocol was higher than that reported by [6], with no apparent signs of degradation (Figure 1). RIN values ranged from 1.70 to 7.60 (Table 1). Samples with RIN less than six are not recommended for analysis because high quality is important for analysis and saves time and costs of researchers. However, low to medium RIN scores or other RNA quality metrics do not necessarily indicate that the samples are not suitable for analysis [8]. We observed a small quantity of DNA in the samples, but this didn’t interfere with downstream analysis [6].

Enviado em: 07/04/2020 - Aprovado em: 03/06/2020

SAJEBTT, Rio Branco, UFAC
v. 7 n. 1 (2020): Edição: jan/abr p. 683-687
ISSN: 2446-4821

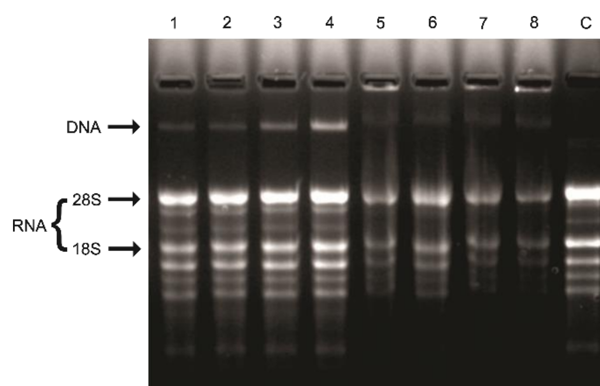


Figure 1. Integrity of RNA extracted from *A. pintoi* leaves on 1% agarose denaturing gel. Lanes 1 to 4: total RNA extracted using the modified protocol. Lanes 5 to 8: total RNA extracted using the original protocol [6]. Lane C: positive control (total RNA extracted from *Arabdopsis* sp. using the original protocol [6]).

In conclusion, adjustments in RNA extraction protocols are common in plant species because of variations in the concentrations of phenolic compounds, polysaccharides, and other secondary metabolites between species. This method allowed us to obtain pure and intact nucleic acid, which is necessary for most transcriptomics studies. In addition, this optimized protocol may be useful for other species that have similar secondary metabolite profiles and concentrations.

REFERENCES

[1] OLIVEIRA, J.C.; CAMPOS, T. Tecnologias em genética molecular para intensificação do uso de amendoim forrageiro em pastagens na Amazônia, In: MENEGUETTI, D.U.O; CARVALHO, C.M.; ZAN, R.A.; SILVA, R.P.M. (Orgs.). Ciência, Inovação e Tecnologia na Amazônia. Rio Branco: Stricto Sensu, 2019, pp. 105-119. Available at: https://sseditora.com.br/wp-content/uploads/Ci%C3%AAncia-Inova%C3%A7%C3%A3o-eTecnologia-na-Amaz%C3%B4nia.pdf. Accessed on: 01 dec. 2019.

[2] TERRA, A.B.C.; FLORENTINO, L.A.; REZENDE, A.V.; et al. Forage legumes in pasture recovery in Brazil. Revista de Ciências Agrárias, v.42, n.2, p.305-313, 2019. https://doi.org/10.19084/rca.16016.

Enviado em: 07/04/2020 - Aprovado em: 03/06/2020

SAJEBTT, Rio Branco, UFAC
v. 7 n. 1 (2020): Edição: jan/abr p. 683-687
ISSN: 2446-4821

[3] EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. Balanço Social 2018. http://bs.sede.embrapa.br/2018/balancosocialembrapa2018web.pdf, 2019. Accessed on: 01 dec. 2019.

[4] PAULINO, V.T.; BUENO, M.S.; ABDALLA, A.L. Composição química e compostos fenólicos em *Arachis pintoi* 'Belmonte'. Archivos de zootecnia, v.61, n.236, p.611-614, 2012. http://dx.doi.org/10.4321/S0004-05922012000400013.

[5] YIN, D.; LIU, H.; ZHANG, X.; et al. A protocol for extraction of high-quality RNA and DNA from peanut plant tissues. Molecular biotechnology, v.49, p.187-191, 2011. https://doi.org/10.1007/s12033-011-9391-9.

[6] OLIVEIRA, R.R.; VIANA, A.J.C.; REÁTEGUI, A.C.E.; et al. An efficient method for simultaneous extraction of high-quality RNA and DNA from various plant tissues. Genetics and Molecular Research, v.14, n.4, p.18828-18838, 2015. http://dx.doi.org/10.4238/2015.December.28.32.

[7] HUANG, C.; PICIMBON, J.F.; LI, H.Q.; et al. An efficient method for total RNA extraction from peanut seeds. Russian journal of plant physiology, v.59, n.1, p.129-133, 2012. https://doi.org/10.1134/S1021443712010074.

[8] JOHNSON, M.T.J.; CARPENTER, E.J.; TIAN, Z.; et al. Evaluating Methods for Isolating Total RNA and Predicting the Success of Sequencing Phylogenetically Diverse Plant Transcriptomes. PloS one, v.7, n.11, p.e50226, 2012. https://doi.org/10.1371/journal.pone.0050226.

Enviado em: 07/04/2020 - Aprovado em: 03/06/2020

CAPÍTULO II

De novo TRANSCRIPTOME ASSEMBLY AND DEVELOPMENT OF SNP MARKERS FROM *Arachis pintoi* LEAVES

Este manuscrito foi formatado para submissão à revista Molecular Breeding.

De novo transcriptome assembly and development of SNP markers from *Arachis pintoi* leaves

Jônatas Chagas de Oliveira[1], Eduardo Fernandes Formighieri[2], Carla Cristina da Silva[3], Anete Pereira de Souza[3], Ana Letycia Basso[4], Gabriel Rodrigues Alves Margarido[4] and Tatiana de Campos[5]

[1]Universidade Federal do Acre (UFAC), Rio Branco, AC, Brazil.
[2]Empresa Brasileira de Pesquisa Agropecuária (Embrapa Agroenergia), Brasília, DF, Brazil.
[3]Universidade Estadual de Campinas (Unicamp), Campinas, SP, Brazil.
[4] Escola Superior de Agricultura Luiz de Queiróz (ESALQ), Piracicaba, SP, Brazil.
[5]Empresa Brasileira de Pesquisa Agropecuária (Embrapa Acre), Rio Branco, AC, Brazil.
Corresponding author: tatiana.campos@embrapa.br;

Abstract Forage peanut (*Arachis pintoi*) has received prominence as an alternative fertilizer source, such as green manure, and to increase forage productivity with reduction of livestock deforestation. Next generation sequencing (NGS) allows the use of new molecular approaches with lower cost, which will contribute to accelerate the genetic gains in *Arachis* breeding programs. Here we report the use of NGS technologies, and *de novo* transcriptome assembly developed from the leaf transcriptome. The sequencing output generated approximately 223 million high-quality paired-end reads with Trinity program. We assembled a *de novo* transcriptome with a total of 98,432 transcripts and a transcript average length of 1,102.2 bp. The N50 value and complete alignment rate were 1,924 bp and 95 %, respectively. Sequence similarity search against the non-redundant National Center of Biotechnology Information (NCBI) protein database returned 69% hits. The molecular function (36.0 %) and biological process (35.8 %) presented most of the Gene Ontology terms assigned, whereas in the cellular component (28.2 %) was assigned the smallest number. A search for functional single nucleotide polymorphisms (SNPs) identified 374,385 markers, of which 39.3 % and 8.1 % were exclusively identified with GATK and SAMtools SNP callers, respectively. These results are the first report of *A. pintoi* leaf transcriptome and functional SNP identification, providing an important information source for advances in *Arachis* breeding programs.

Keywords: Forage peanut, Functional SNPs, Functional annotation, RNA-Seq.

34 ## Introduction

35 The use of legume as green cover intercropped with commercial crops is an important strategy
36 for maintaining, conserving, and improving soil quality, as it contributes to maintaining soil
37 moisture in addition to biological nitrogen fixation. Among the legumes used, forage peanut
38 (*Arachis pintoi* Krapov. and W.C. Greg) has stood out for its positive results in coffee consortia,
39 providing N and low nitrous oxide emissions, and helping in the control of weeds (Santos et al.
40 2014; Rose et al. 2019). In mixed pastures, forage peanut is an alternative to increase
41 productivity with reduction of livestock deforestation in the Brazilian Amazon (Oliveira and
42 Campos 2019). In 2019, the use of this crop had a positive impact of approximately USD 17
43 million per year on intercropped pastures in southwestern Amazon (Embrapa 2020). However,
44 despite its importance, genetic knowledge is still small when compared to cultivated peanuts
45 (*A. hypogaea*).

46 Single nucleotide polymorfism (SNPs) have been shown to be useful tools. They have co-
47 dominant nature, distributed throughout the genome, highly polymorphic, and are biallelic
48 (Rafalski 2002). With the advent of RNA sequencing (RNA-Seq), the cost of developing SNPs
49 has reduced and, at the same time, allowed obtaining information that contributes to the
50 association of the phenotype to alternative splice events or gene expression changes (Wit et al.
51 2015). SNP-based genome-wide association studies (GWAS) have been successfully applied
52 in *Sorghum bicolor* (Morris et al. 2013), *Zea mays* (Zhang et al. 2016), and *Glycine max*
53 (Contreras-Soto et al. 2017) demonstrating the great potential of these approach to discover
54 marker-trait associations (MTAs).

55 In *Arachis* genus, RNA-Seq has been used mainly in cultivated peanut and their wild ancestors
56 (*A. ipaënsis* Krapov. and W.C. Gregory and *A. duranensis* Krapov. and W.C. Gregory). *A.*
57 *pintoi* has few genomic sequences available, of which 25 microsatellite markers were developed
58 (Palmieri et al. 2002; Palmieri et al. 2005; Palmieri et al. 2010) and there are no SNPs developed
59 for *A. pintoi* to date. In this study, we presented the first trancriptome assembly of *A. pintoi*
60 leaves from two divergent genotypes in terms of seeds production. In addition, we have
61 identified potential functional SNPs, which will play a relevant role in advancing of the *Arachis*
62 breeding.

63

64 ## Material and methods

65 Plant material, RNA extraction and sequencing

66

67 Two divergent genotypes of *A. pintoi* were sampled for RNA sequencing: the cultivars Amarillo
68 and Belomonte. Amarillo has a good production of seeds and it is the most used forage peanut
69 cultivar around the world. In despite to produce high quantity of seeds has a crossing rate about
70 2% (Oliveira et al., 2019). Belomonte has low seed production and propagate mainly by
71 vegetative way. Both genotypes also differ in terms of forage production during the drought
72 period.
73 In October 2017, four stolons from each genotype were collected in the Germplasm Bank
74 maintained by Brazilian Agricultural Research Corporation (Embrapa Acre) and planted in
75 pots. The samples were transported to the Multiuser Laboratory of Genotyping and Sequencing
76 (LMGS) at the Center of Molecular Biology and Genetic Engineering (CBMEG/Unicamp, SP,
77 Brazil). Young leaves from each genotype were collected and used for total RNA isolation. The
78 RNA isolation, the integrity and quantity evaluation were performed as previously described
79 (Oliveira et al. 2020) and RNA integrity confirmed using a 2100 Bioanalyzer (Agilent
80 Technologies). The RNA was then used for cDNA library preparation of each sample using
81 Illumina TruSeq RNA Sample Preparation Kit (Illumina Inc., USA) according to the
82 manufacturer's instructions. The sequencing of each cDNA library was carried out on an
83 Illumina HiSeq 2500 platform (Illumina Inc., USA) to get pair-end sequences reads of 100 bp
84 length.
85
86 Data filtering and *de novo* assembly
87
88 We assessed the quality of raw reads with FastQC
89 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Next, we removed low quality
90 reads (Phred quality score < 33), adaptors and the first 12 bp using trimommatic (trimommatic-
91 0.39.jar version) (Bolger et al. 2014). . Additionally, we attempted for removing residual rRNA
92 using the SortMeRNA (Kopylova et al. 2012). Then, a *de novo* transcriptome was assembled
93 using Trinity (Trinity-v2.8.5 version) with *k-mer* size of 25 and default values for other
94 parameters (Grabherr et al. 2011). To access the quality of the *de novo* assembly we aligned the
95 clean reads to the assembled transcriptome using Benchmarking Universal Single-Copy
96 Orthologs (BUSCO) program (Simão et al. 2015) with *viridiplantae* database
97 (https://busco.ezlab.org/datasets/prerelease/viridiplantae_odb10.tar.gz).
98
99 Functional annotation

100

101 We compared the transcriptome against the NCBI non-redundant protein (nr) database. A
102 homology search against these databanks were performed using the BlastX option from the
103 BLAST+ suite (Camacho et al. 2009). The BLAST+ results against the nr database were
104 imported to Blast2GO suite (Gotz et al. 2008) for mapping and retrieving Gene Ontology (GO)
105 and unique enzyme code (EC) annotations of assembled unigenes. The retrieved GO terms were
106 allocated to query sequences and the genes present in the transcriptome were classified into
107 cellular component, molecular function and biological process categories. Besides the
108 Blast2GO functional annotation, we also annotated the *de novo* transcriptome against the
109 Uniprot/SwissProt data based, using the Trinotate pipeline
110 (https://github.com/Trinotate/Trinotate.github.io/wiki), which also retrieve confident results,
111 but in a shorter time. We used the Blast2GO results to complement the lack of annotation for
112 some unigenes with Trinotate. The assembled transcriptome was further annotated against the
113 Kyoto Encyclopedia of Genes and Genomes (KEGG) metabolic pathways database (Kanehisa
114 and Goto, 2000).

115

116 Differential expression analysis and Functional enrichment

117

118 We estimated gene counts with Salmon (Patro et al. 2017) and evaluated the expression at gene
119 level, as we grouped transcripts to genes using tximport package (Soneson et al. 2015). After
120 that, we filtered the counts matrix to maintain only genes with three or more samples with at
121 least counts per million (CPM) ≥ 2. Next, we normalized the libraries by the Trimmed Mean of
122 M-values (TMM). The resulting matrix was used to perform differential expression (DE)
123 analysis using to contrast the cultivars Amarillo and Belomonte. The downstream analysis were
124 performed applying Quasi-Likelihood method, implemented in the edgeR package (Robinson
125 et al. 2010). The differentially expresses genes was used to perform enrichment analysis with
126 GOseq package (Young et al. 2010) of GO terms.

127

128 Putative SNP calling and filtering

129

130 To identify putative SNP markers positions, the STAR aligner (STAR-2.7.3a version) was used
131 to align the reads to the *de novo* assembled transcriptome to obtain the sorted transcripts (BAM
132 files) for each cultivar (Dobin et al. 2013). Subsequently, the sequence variant discovery for

133 SNP calling was done in Genome Analysis Tool Kit (GATK) pipeline version 4.1.4.0 (Poplin
134 et al. 2017) and SAMtools pipeline (Li 2011) using default settings. The *de novo* assembly was
135 used as a reference for SNP calling. A position was called a putative SNP if any cultivar had a
136 different allele against the reference. The putative SNP found by GATK and SAMtools were
137 filtered for QD < 2 and QUAL <20, respectively. To determine which SNPs were found by
138 each pipeline and by both, we used the BCFtools isec command, and these results were used to
139 construct a Venn diagram using the venndiagram R package (Chen and Boutros 2011).
140
141 **Results and discussion**
142 Sequencing, filtering and *de novo* assembly
143
144 Before the extraction of the RNA, two stolons of the cultivar Amarillo and one of Belomonte
145 died. Thus, the RNA extraction and subsequent analyses were made with two subsamples of
146 Amarillo and three of Belomonte. Each genotype was sequenced independently, and the
147 libraries (100 bp) generated a total of 267,234,958 (~ 26.7 Gb) raw pair-end reads from both
148 libraries combined (Table 1). Sequencing of the Amarillo produced a slightly lower number of
149 reads (43.2 % of total reads) compared with Belomonte (56.8 % of total reads). However, this
150 difference did not interfere with the assembly or characterization of the *A. pintoi* leaf
151 transcriptome. After filtering, 223,242,941 (~22.3 Gb) high quality pair-end read were obtained
152 from both genotypes. Amarillo retained more high-quality reads than Belomonte (87.8 and 80.3
153 %, respectively). Despite the reduction in absolute numbers in the filtered data, the quality of
154 the sequences obtained is an important factor to be considered, as it contributes to the correction
155 of sequencing errors and improves the accuracy of the new assembly, which is desirable
156 especially for SNP development (MacManes 2014).
157 **Table 1** Summary of Illumina sequencing outputs statistics

		Amarillo	Belomonte	Total
Sequenced	Number of reads	115,367,890	151,867,068	267,234,958
	Total bases (Gb)	11.5	15.2	26.7
Filtered	Number of reads	101,336,876	121,906,065	223,242,941
	Total bases (Gb)	10.1	12.2	22.3
	High quality data	87.8 %	80.3 %	83.5 %

158
159 The high-quality reads were used for *de novo* assembly with Trinity assembler, applying
160 established criteria. A total of 108,493,213 bases were assembled, generating 98,432 transcripts
161 (Table 2). The length of 42,782 contigs was <500 bp whereas 17,623 where longer than 500 bp
162 size. A total of 21,137 were over 1000 bp and 985 were over 5000 bp.

163 **Table 2** Statistics of *de novo* assembly of *A. pintoi* leaf transcriptome

Characterísticas	Details
Total number of contigs	98,432
Total assembled bases	108,493,213
Min length (bp)	180
Max length (bp)	16,665
Average length (bp)	1,102.2
Median length (bp)	642
Number of contigs < 500 bp	42,782
Number of contigs ≥ 500 bp	17,623
Number of contigs ≥ 1000 bp	21,137
Number of contigs ≥ 2000 bp	15,859
Number of contigs ≥ 5000 bp	985
Number of contigs ≥ 10000 bp	46
N50 (bp)	1924
GC content (%)	40.8

164

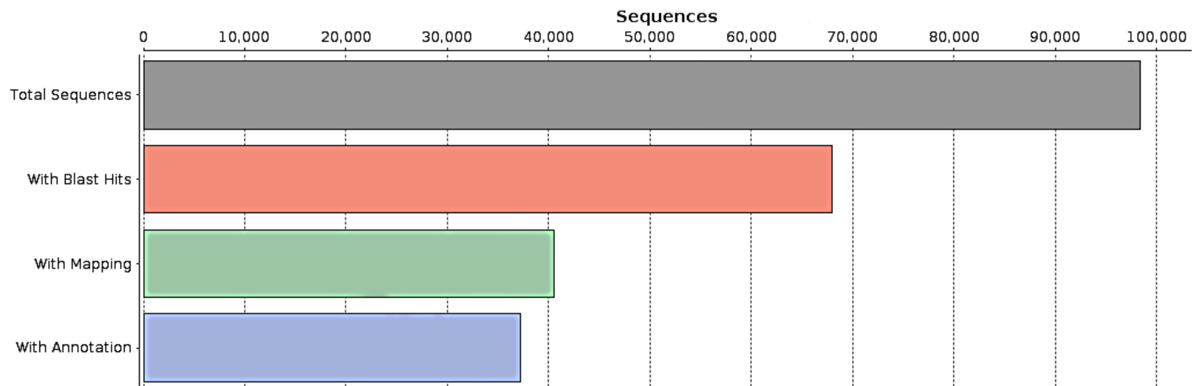
165 The average contig length found in this study was higher than that of *Pueraria montana var.*
166 *lobate* and *Neustanthus phaseoloides* (Haysen et al. 2018), *Gastrodia elata* (Wang et al. 2019)
167 and similar to *A. hypogaea* (Chopra et al. 2014). In addition, the N50 value was higher than that
168 of *A. ipaënsis*, *A. duranensis*, and *A. hypogaea* (Chopra et al. 2014), *Onobrychis viciifolia*
169 (Mora-Ortiz et al. 2016), *Urochloa humidicola* (Vigna et al. 2016), *G. elata* (Wang et al. 2019),
170 which suggest that the *de novo* assembly from *A. pintoi* leaves was satisfactory as that of the
171 other *de novo* assemblies.
172 The clean reads were aligned to the assembled transcriptome to assess the quality of assembly
173 using BUSCO program and *viridiplantae* database. The complete alignment rate was 95 %, of
174 which 52.3 % reads could map single-copy to transcriptome, while 43.3 % reads could map
175 duplicated-copy on transcriptome. These results were superior to that found in *P. montana var.*
176 *lobate* and *N. phaseoloides de novo* transcriptome assembly (Haysen et al. 2018), and *A.*
177 *monticola de novo* genome assembly (Yin et al. 2018), indicating the high completeness of the
178 assembly.
179

180 Functional Annotation

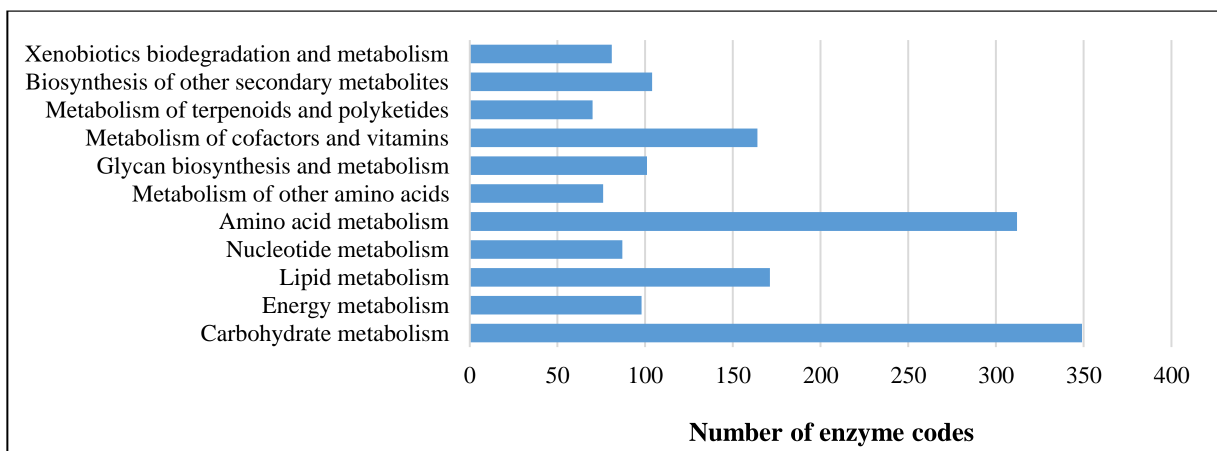
181

182 Functional annotations were done by comparison of the transcriptome sequences with the public
183 databases. Among the 98,432 transcripts, 67,939 (69.0 %) displayed homology to sequences in
184 the NCBI nr database (Figure 2). The annotation resulted in 37,219 transcripts that were
185 assigned with 146,844 GO terms, and these terms were grouped into three main classes of
186 onthologies: molecular function, biological process, and cellular component. The molecular
187 function (46,082 terms, 36.0 %) and biological process (45,800 terms, 35.8 %) presented the

188 majority of the GO terms assigned, whereas in the cellular component (36,131 terms, 28.2 %)
189 were assigned the smaller number.
190



191
192 **Fig. 1** Functional annotation of *A. pintoi* leaf transcriptome.
193
194 To correlate *A. pintoi* transcriptome with known metabolic pathways, we perform annotation
195 against KEGG database to assign sequences with KO terms and their respective KEGG maps.
196 A total of 845 assembled transcripts were associated with 1,671 enzymes codes and 153
197 pathways. These annotated enzymes were categorized into five major pathways in KEGG
198 database – "metabolism" (1,618), "genetic information processing" (23), "environmental
199 information processing" (25), "organismal systems" (3), and "human diseases" (2). The top five
200 pathways in "metabolism" category were "carbohydrate metabolism" (349), "amino acid
201 metabolism" (312), "lipid metabolism" (171), "metabolism of cofactors and vitamins" (164),
202 "biosynthesis of other secondary metabolites" (104) (Figure 4). The KEGG results indicate that
203 the assembled transcriptome has genes distributed among several metabolic pathways.
204



205
206 **Fig. 2** Classification of enzyme codes in KEGG "metabolism" category.

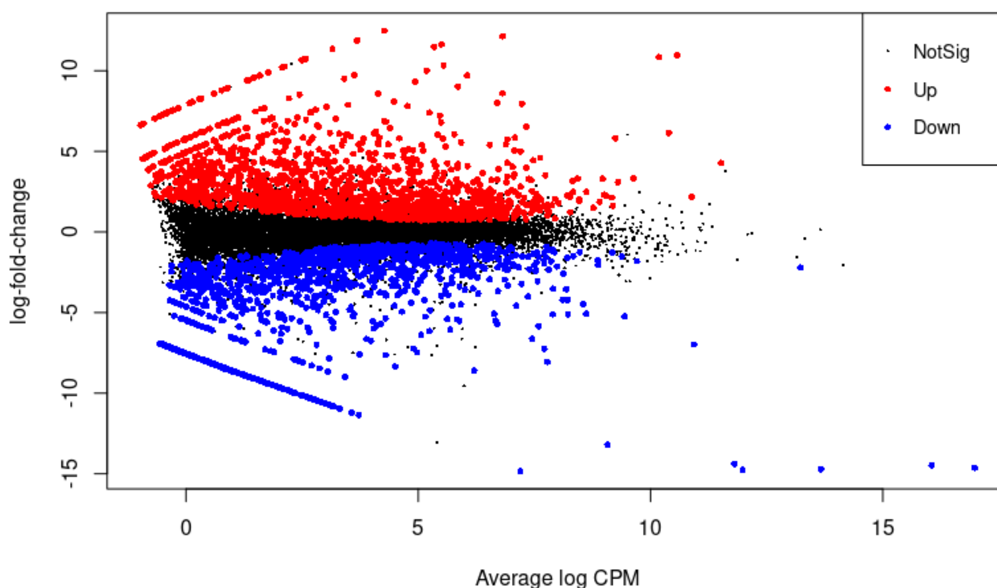
207

208 Differential expression analysis and functional enrichment

209

210 The cultivars Amarillo and Belomonte showed 85.74 % of similar gene expression in leaf
211 transcriptome. The differential expression analysis found 1,550 genes with higher expression
212 level in Amarillo and 1,357 genes with higher expression level in Belomonte (Figure 3). Further
213 studies are needed to investigate the functions of the genes specifically and commonly involved
214 in various leaf characteristics.

215

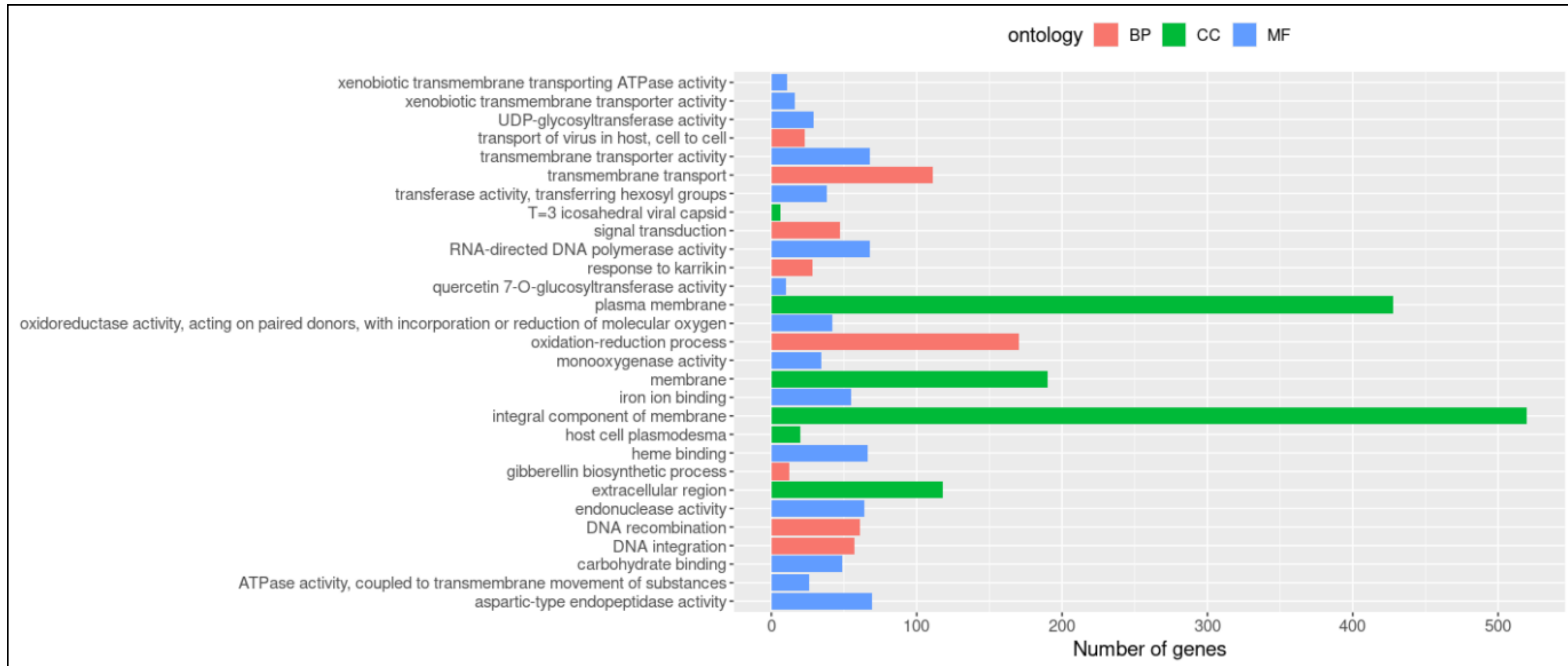


216

217 **Fig. 1** MA plot of differential expression analysis of cultivar Amarillo and Belomonte of *A.*
218 *pintoi*. UP: genes with higher expression level in Amarillo. Down: genes with higher expression
219 level in Belomonte. NotSig: genes with differential expression not significative.

220

221 The enrichment analysis of the annotated GO terms showed that cellular component (1,282
222 genes, 55.63 %) category was the most abundant, followed by molecular function (645 genes,
223 26.48 %) and biological process (509 genes, 20.89 %) (Figure 4). The GO terms "integral
224 component of membrane" (520 genes) and "plasma membrane" (428 genes) were the largest
225 groups in class cellular component. In molecular function class, "aspartic-type endopeptidase
226 activity" (69 genes) and "RNA-directed DNA polymerase activity" (68 genes) were the largest,
227 and "oxidation-reduction process" (170 genes) and "transmembrane transport" (111 genes)
228 were the first and second largest groups biological process.



229

230 **Fig. 3** GO-level distributions in *A. pintoi* leaf transcriptome. BP: biological process. CC: cellular component. MF: molecular function.

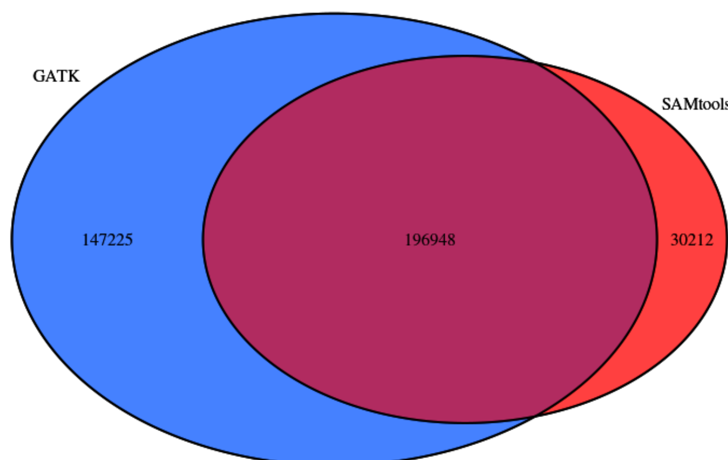
231

232 Putative SNPs calling and filtering

233

234 In this study, we identified 358,519 and 303,996 SNPs by GATK and SAMtools, respectively.
235 After filtering, 344,173 (96.0 %) and 227,160 (74.7 %) SNPs remaining from GATK and
236 SAMtools, respectively. A total of 196,948 (48.3 %) SNPs were found by both, of which
237 147,225 (39.3 %) were exclusive for GATK and 30,212 (8.1 %) for SAMtools (Figure 5). The
238 number of transition (Ts) SNPs were more abundant than transversion (Tv) SNPs, resulting in
239 a Ts/Tv ratio of 1.75 for both SNP callers (Table 3). Among the transition variation, C ↔ T
240 was the most highly represented, with 114,622 and 74,149 SNPs for GATK and SAMtools,
241 respectively, and A ↔ T was the transversion SNPs most common with 36,277 SNPs (GATK)
242 and 23,727 (SAMtools). Variations in the number of SNPs identified between variant callers
243 have been observed and may be the result of different approaches in each pipeline (Cornish &
244 Guda, 2015; Hwang, Kim, Lee, & Marcotte, 2015). To avoid bias in the SNP call the use of
245 filtering is essential. Hwang et al. (2015) observed ~ 92 % agreement between the variant callers
246 when they filtered out low confident variants below the quality score (QUAL) threshold of 20.
247 The variants identified by both variant callers may be more accurate.

248



249

250 **Fig. 5** Venn diagram showing SNPs found by GATK and SAMtools pipeline, after filtering.
251 Numbers in the circles indicate the numbers of SNPs found by single or both pipeline.

252

253 **Table 3** Sumary of the putative SNPs identified using GATK and SAMtools.

	GATK	SAMtools
Number of SNPs	344,173	227,160
Transitions	220,355	145,458
A ↔ G	105,733	71,309
C ↔ T	114,622	74,149
Transversions	126,029	82,928
A ↔ C	29,707	20,479
A ↔ T	36,277	23,727
C ↔ G	26,884	18,198
G ↔ T	33,161	20,524

254

255 Here we presented the first report of transcriptome analysis and SNPs detection in *A. pintoi*.
256 Further studies are needed to validate and characterize the identified SNPs, and relate them to
257 phenotypic traits of agronomic interest. The results of differential expression may play an
258 important role in SNP selection for validation. In addition, these set of information will open
259 new opportunities for genomic selection, linkage mapping, and GWAS on *Arachis* breeding
260 programs.
261

262 **Conclusions**

263 RNA-Seq approach and *de novo* assembly provided the first transcriptome in *A. pintoi*. New
264 SNP markers were identified and are valuable tools to breeding programs. Thus, this study
265 provides the largest genetic resource in *A. pintoi*, which will allow expression analysis,
266 marker-assisted breeding, and association studies.
267

268 **Acknowlegments**

269 We are gratefully Governo Federal/ CNPq/ Governo do Estado do Acre/ FAPAC TO 024/2018,
270 Embrapa to financial support. We thank the staff of the Life Sciences Core Facility (LaCTAD)
271 from State University of Campinas (UNICAMP), for the Genomics analysis.
272

273 **References**

274 Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina
275 sequence data. Bioinformatics 30: 2114-2120
276 Camacho C, Coulouris, G, Avagyan, V, Ma, N, Papadopoulos, J, Bealer, K, Madden, T. L.
277 2009. BLAST+: architecture and applications. BMC Bioinformatics, 10:421
278 Chopra R, Burow G, Farmer A, Mudge J, Simpson CE, Burow MD (2014) Comparisons of *de*
279 *novo* transcriptome assemblers in diploid and polyploid species using peanut (*Arachis* spp.)
280 RNA-Seq data. PloS One 9: e115055

281 Contreras-Soto RI, Mora F, de Oliveira MA, Higashi W, Scapim CA, Schuster I (2017) A
282 Genome-Wide Association Study for Agronomic Traits in Soybean Using SNP Markers and
283 SNP-Based Haplotype Analysis. PloS One 12:e0171105
284 Cornish A, Guda C. (2015) A Comparison of Variant Calling Pipelines Using Genome in a
285 Bottle as a Reference. Biomed Res Int 2015:1-11
286 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M,
287 Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15-21
288 Embrapa (2020) Balanço Social 2019.
289 https://bs.sede.embrapa.br/2019/balancosocialembrapa2019print.pdf. Accessed 28 september
290 2020
291 Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talon
292 M, Dopazo J, Conesa A (2008) High-throughput functional annotation and data mining with
293 the Blast2GO suite. Nucleic Acids Res 36: 3420-3435
294 Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L,
295 Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F,
296 Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length
297 transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29:
298 644-652
299 Haynsen MS, Vatanparast M, Mahadwar G, Zhu D, Moger-Reischer RZ, Doyle JJ, Crandall
300 KA, Egan AN (2018) *De novo* transcriptome assembly of *Pueraria montana var. lobata* and
301 *Neustanthus phaseoloides* for the development of eSSR and SNP markers: narrowing the US
302 origin(s) of the invasive kudzu. BMC Genomics 19:439-456
303 Hwang, S, Kim, E, Lee, I, Marcotte, E. M. 2015. Systematic comparison of variant calling
304 pipelines using gold standard personal exome variants. Sci Rep 5:17875
305 Kanehisa M, Goto S. (2000) KEGG: kyoto encyclopedia of genes and genomes. Nucleic
306 Acids Res. 28:27-30
307 Kopylova E, Noe L, Touzet H (2012) SortMeRNA: fast and accurate filtering of ribosomal
308 RNAs in metatranscriptomic data. Bioinformatics 28: 3211-3217
309 Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping
310 and population genetical parameter estimation from sequencing data. Bioinformatics 27:
311 2987-2993
312 Macmanes MD (2014) On the optimal trimming of high-throughput mRNA sequence data.
313 Front Genet 5:1-13

314 Mora-Ortiz M, Swain MT, Vickers MJ, Hegarty MJ, Kelly R, Smith LM, Skot L (2016) *De-*
315 *novo* transcriptome assembly for gene identification, analysis, annotation, and molecular
316 marker discovery in *Onobrychis viciifolia*. BMC Genomics 17(1): 756-769
317 Morris GP, Ramu P, Deshpande SP, Hash CT, Shah T, Upadhyaya HD, Riera-Lizarazu O,
318 Brown PJ, Acharya CB, Mitchell SE, Harriman J, Glaubitz JC, Buckler ES, Kresovich S
319 (2013) Population genomic and genome-wide association studies of agroclimatic traits in
320 sorghum. Proc Natl Acad Sci USA 110:453-458
321 Oliveira JC, Campos T (2019) Tecnologias em genética molecular para intensificação do uso
322 de amendoim forrageiro em pastagens na Amazônia. In: Meneguetti DUO, Carvalho CM, Zan
323 RA, Silva RPM (Orgs) Ciência, Inovação e Tecnologia na Amazônia, Stricto Sensu, Rio
324 Branco, pp 105-119. https://sseditora.com.br/wp-content/uploads/Ci%C3%AAncia-
325 Inova%C3%A7%C3%A3o-eTecnologia-na-Amaz%C3%B4nia.pdf. Accessed 01 september
326 2020.
327 Oliveira JC, Rufino PB, Azêvedo HSFS, Sousa ACB, Assis GML, Silva LM, Sebbenn AM,
328 Campos T (2019) Inferring mating system parameters in forage peanut, *Arachis pintoi*, for
329 Brazilian Amazon conditions. Acta Amazonica 49:277-282
330 Oliveira JC, Silva CC, Souza AP, Campos T (2020) A modified protocol for total RNA
331 isolation from forage peanuts. South American Journal of Basic Education, Technical and
332 technological 7:683-687
333 Palmieri DA, Bechara MD, Curi RA, Gimenes MA, Lopes CR (2005) Novel polymorphic
334 microsatellite markers in section *Caulorrhizae* (*Arachis*, Fabaceae). Molecular Ecology Notes
335 5:77-79
336 Palmieri DA, Bechara MD, Curi RA, Monteiro JP, Valente SES, Gimenes MA, Lopes CR
337 (2010) Genetic diversity analysis in the section *Caulorrhizae* (genus *Arachi*s) using
338 microsatellite markers. Genetics and Molecular Biology 33:109-118
339 Palmieri DA, Hoshino AA, Bravo JP, Lopes CR, Gimenes MA (2002) Isolation and
340 characterization of microsatellite loci from the forage species *Arachis pintoi* (Genus *Arachis*).
341 Molecular Ecology Notes 2:551-553
342 Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C (2017) Salmon provides fast and bias-
343 aware quantification of transcript expression. Nat Methods 14:417-419
344 Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Van der Auwera GA,
345 Kling DE, Gauthier LD, Levy-Moonshine A, Roazen D, Shakir K, Thibault J, Chandran S,
346 Whelan C, Lek M, Gabriel S, Daly MJ, Neale B, MacArthur DG, Banks E (2017) Scaling

347 accurate genetic variant discovery to tens of thousands of samples. bioRxiv.
348 https://doi.org/10.1101/201178
349 Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. Current
350 Opinion in Plant Biology 5:94-100
351 Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for
352 differential expression analysis of digital gene expression data, Bioinformatics 26(1):139-140.
353 https://doi.org/10.1093/bioinformatics/btp616
354 Rose TJ, Kearney LJ, Morris S, Van Zwieten L, Erler DV (2019) Pinto peanut cover crop
355 nitrogen contributions and potential to mitigate nitrous oxide emissions in subtropical coffee
356 plantations. Sci Total Environ 656:108-117
357 Santos JCF, Cunha AJ, Melo B (2014) Soil cover and weed control on coffee intercropping
358 perennial legume. International Journal of Applied Science and Technology 4:149-157
359 Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO:
360 assessing genome assembly and annotation completeness with single-copy orthologs.
361 Bioinformatics 31:3210-3212
362 Soneson C, Love MI, Robinson MD (2015) Differential analyses for RNA-seq: transcript-
363 level estimates improve gene-level inferences. F1000Research, 4. doi:
364 10.12688/f1000research.7563.1.
365 Vigna BB, de Oliveira FA, de Toledo-Silva G, da Silva CC, do Valle CB, de Souza AP
366 (2016) Leaf transcriptome of two highly divergent genotypes of *Urochloa humidicola*
367 (Poaceae), a tropical polyploid forage grass adapted to acidic soils and temporary flooding
368 areas. BMC Genomics 17: 910-929
369 Wang Y, Shahid MQ, Ghouri F, Ercisli S, Baloch FS (2019) Development of EST-based SSR
370 and SNP markers in *Gastrodia elata* (herbal medicine) by sequencing, de novo assembly and
371 annotation of the transcriptome. 3 Biotech 9: 292-302
372 Wit P, Pespeni MH, Palumbi SR (2015) SNP genotyping and population genomics from
373 expressed sequences – current advances and future possibilities. Molecular Ecology 24:2310-
374 2323
375 Yin, D, Ji, C, Ma, X, Li, H, Zhang, W, Li, S, Liu, F, Zhao, K, Li, F, Li, K, Ning, L, He, J,
376 Wang, Y, Zhao, F, Xie, Y, Zheng, H, Zhang, X, Zhang, Y, Zhang, J (2018) Genome of an
377 allotetraploid wild peanut *Arachis monticola*: a de novo assembly. Gigascience 7:1-9

378 Young MD, Wakefield MJ, Smyth GK, Oshlack A (2010) Gene ontology analysis for RNA-
379 seq: accounting for selection bias. Genome Biol 11:R14. https://doi.org/10.1186/gb-2010-11-
380 2-r14
381 Zhang X, Warburton ML, Setter T, Liu H, Xue Y, Yang N, Yan J, Xiao Y (2016) Genome-
382 wide association studies of drought-related metabolic changes in maize using an enlarged
383 SNP panel. Theor Appl Genet 129:1449-1463
384
385

CAPÍTULO III

DEVELOPMENT OF MICROSATELLITE MARKERS FROM AN *Arachis pintoi* TRANSCRIPTOME

Este manuscrito foi formatado para submissão à revista Crop Science.

1 **Development of microsatellite markers from an *Arachis pintoi* transcriptome**

2

3 Jônatas Chagas de Oliveira[1], André Lucas Domingos da Silva[1], Luciélio Manoel da Silva[2],
4 Eduardo Fernandes Formighieri[3], Carla Cristina da Silva[4], Anete Pereira de Souza[4] and Tatiana
5 de Campos[2]

6

7 [1]Universidade Federal do Acre (UFAC), Rio Branco, AC, Brazil.
8 [2]Empresa Brasileira de Pesquisa Agropecuária (Embrapa Acre), Rio Branco, AC, Brazil.
9 [3]Empresa Brasileira de Pesquisa Agropecuária (Embrapa Agroenergia), Brasília, DF, Brazil.
10 [4]Universidade Estadual de Campinas, Cidade Universitária Zeferino Vaz, Campinas, SP,
11 Brazil.
12 Corresponding author: tatiana.campos@embrapa.br; phone: +55 (68) 32123256.

13

14 **ABSTRACT**

15 The forage peanut (*Arachis pintoi*) has increased the livestock productivity due to its high
16 protein content, in addition the species contributes to soil fertility and recovery of degraded
17 areas. Nevertheless, there are a small number of molecular markers in *A. pintoi*. The objective
18 of this study was to develop and to characterize microsatellites from leaf transcriptome of *A.*
19 *pintoi* and to test their transferability to peanut (*A. hypogaea*) and other six wild species of
20 *Arachis* genus. We identified 4,461 putative simple sequence repeats (SSR) in 4,046 unigenes.
21 After filtering out primers with same annealing site and search for sequences related to open
22 reading frames (ORFs), PCR primer pairs were designed for 999 SSRs regions. The
23 dinucleotide motif was the most abundant (628; 62.86%), followed by trinucleotide (336;
24 33.63%). New 186 SSRs were characterized, of which 63 (33.87%) were polymorphic with
25 average of 7.37 alleles per locus. The markers presented high average for polymorphic
26 information content (PIC = 0.70) and discriminatory power (*D* = 0.80). One hundred and twenty
27 (64.52%) sequences presented similarity with proteins that are involved in growth, development
28 and cell differentiation, pathogen resistance and defense mechanism. Thirty-three SSRs were
29 evaluated for the transferability to peanut and other six wild species from *Arachis* genus, which
30 resulted in variable cross-species amplification (63.64 to 100%). These new informative
31 microsatellites, possibly linked to genes of agronomic importance, can be used to accelerate
32 *Arachis* breeding programs.

33 **Index terms**: forage peanut, *Arachis*, peanut, RNASeq.

34

35 **INTRODUCTION**

36 The genus *Arachis* is composed by 81 species, from which the cultivated peanut (*Arachis*
37 *hypogaea* L.) stands out for being one of the main food crops of the world, especially as a source
38 of vegetable oil. Among the wild species of *Arachis* genus, some have potential for forage use.
39 This is the case of forage peanut (*Arachis pintoi* Krapov. & WC Greg.), which in 2019 had a
40 positive economic impact of approximately US$ 17 million for livestock production in
41 Occidental Brazilian Amazon (Embrapa, 2020). The use of forage peanut increased the weight
42 gain in cattle due to the high quantity of protein content, reducing the slaughter time in up to
43 six months, and contributed to soil fertility and recovery of degraded areas through biological
44 nitrogen fixation (Maia, 2018; Terra, Florentino, Rezende, & Silva, 2019; zu Ermgassen et al.,
45 2018).

46 Among the tools for characterizing genome variation, microsatellite markers (Simple Sequence
47 Repeats – SSR) have been shown to be ideal tools. They are distributed throughout the genome
48 and present high polymorphism. In addition, they are multi-allelic and have co-dominant nature
49 (Hodel, Segovia-Salcedo et al., 2016). With the use of RNA-Seq (RNA sequencing) it is
50 possible to develop thousands of SSRs with less cost and effort than traditional techniques
51 (Taheri et al., 2018). The greatest advantage of RNA-based markers is that they are derived
52 from expressed regions of the genome, and they generate fragments that can be easily related
53 to phenotypic traits, which is extremely important for association studies (Poczai et al., 2013).
54 SSR-based genome-wide association studies (GWAS) have been successfully applied in
55 *Panicum maximum* (Simeão, Souza, Jank & Souza, 2015), *Gossypium* spp. (Nie et al., 2016)
56 and *Saccharum* spp. (Barreto et al., 2019), demonstrating the great potential of these approach
57 to discover marker-trait associations (MTAs).

58 In *Arachis* genus, RNA-Seq has been used mainly in cultivated peanut and their wild ancestors,
59 *A. ipaensis* and *A. duranensis*. However, the number of microsatellites available for *A. pintoi* is
60 restricted to 25 loci (Palmieri, Hoshino, Bravo, Lopes, & Gimenes, 2002; Palmieri, Bechara,
61 Curi, Gimenes, & Lopes, 2005; Palmieri et al., 2010), which were developed from few genomic
62 sequences. If we consider just the loci with adequate quality to PCR amplification to automated
63 genotyping, this number is reduced to ten loci (Azêvedo et al., 2016).

64 Heterologous microsatellites have been a useful and informative tool in *A. pintoi* to germplasm
65 evaluation, hybrid identification and mating system evaluation (Azêvedo et al., 2016; Campos
66 et al., 2016; Oliveira et al., 2019). Therefore, the development of new microsatellites markers,
67 especially from expressed genes, is crucial to advance of breeding programs, to develop genetic
68 maps, to map quantitative trait loci (QTLs) and to perform GWAS (Huang et al., 2016).

Thus, the objective of this study was to develop and to characterize transcript-derived microsatellite markers from *A. pintoi* transcriptome, and to test the transferability of these markers to peanuts (*A. hypogaea*) and other six wild species of *Arachis* genus.

MATERIAL AND METHODS

Plant material, DNA and RNA extraction, and RNA sequencing

Leaves from two accessions of *A. pintoi* were sampled for RNA sequencing: the cultivars Amarillo and Belomonte. These genotypes belong to Germplasm Bank of forage peanut maintained by Brazilian Agricultural Research Corporation (Embrapa Acre), Rio Branco, AC, Brazil. These cultivars are divergent in terms of seed production. The RNA extraction was performed according to Oliveira et al., 2020. The RNA was then used to perform cDNA library and sequencing, through Illumina TruSeq RNA Sample Preparation Kit and Illumina HiSeq 2500 platform (Illumina, USA), according to the manufacturer's instructions.

We used 19 accessions of *A. pintoi,* also belong to Germplasm Bank maintained by Embrapa Acre, to characterize the SSRs (Table 1). Four *A. pintoi* cultivars (Alqueire 1, BRS Mandobi, Amarillo MG-100, and Belomonte) were used for varietal identification. In order to test the transferability, we used seven species of the *Arachis* genus: *A. glabrata*, *A. valsii*, *A. stenosperma*, *A. duranensis*, *A. ipaensis*, *A. hypogaea*, *A. repens*. The genomic DNA was extracted from fresh leaves using a modified CTAB method (Campos et al., 2016) and their quantity were detected by a fluorimeter Qubit (Thermo Fisher Scientific Inc., USA).

Table 1. Genotypes of *A. pintoi* and seven species of the genus *Arachis* used for characterization and transferability of microsatellite markers.

Genotypes *A. pintoi*		Other species of *Arachis*
V14951	V14966	*A. glabrata*
V6741	Amarillo	*A. valsii*
V6784	V15062	*A. stenosperma*
W225	W944	*A. duranensis*
W647	V13288	*A. ipaensis*
V5895	V13298	*A. hypogaea*
V13196	V13294	*A. repens*
V13211-1	W34 (B)	
V6791wf	V13372	
Belomonte		

SSR development and selection

The *de novo* assembly was performed using Trinity (Trinity-v2.8.5 version) pipeline, with default parameters, until supertranscripts were obtained. The SSR's searches were performed using supertranscript sequences. The microsatellites were identified using the MISA platform

96 (MIcroSAtellite identification tool), with a minimum of eight copies of the repetitive motif for
97 di to hexanucleotides. Primers were designed using Primer 3
98 (https://sourceforge.net/projects/primer3/) with an ideal size of 18 base pairs, an ideal
99 temperature of 60 °C (min.: 57 °C; max.: 62 °C), percentage of GC between 20 and 80, 'GC
100 end' equal to zero, and product size range was from 100 to 300 bp. To avoid primers with same
101 site of annealing and search loci near to open reading frames (ORFs) we used the pipeline
102 developed by Hodel, Gitzendanner, et al. (2016)
103 (https://github.com/soltislab/transcriptome_microsats/tree/master/Tutorial).

104 **SSR amplification and analysis**

105 The amplification reactions to evaluate the ideal annealing temperature and the alleles' size
106 were performed according to Azêvedo et al. (2017). The SSR markers were evaluated for
107 amplification of the loci, their sharpness and the absence of nonspecific bands. The
108 polymorphism information content (PIC), expected (H_E) and observed (H_O) heterozygosity
109 were obtained using the Tools For Population Genetic Analyses (TFPGA) software (Miller,
110 1997). The PIC of each SSR was calculated using the equation of Botstein et al. (1980):

111
$$PIC = 1 - \sum_{i=1}^{n} f_i^2 - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 2\,fi^2 fj^2$$

112 where f_i is the frequency of the ith allele and f_j is the frequency of the jth allele, summing over
113 alleles.

114 We estimate the power of discrimination (*D*) to compare the efficiency of the markers to
115 identify the core collection genotypes using the equation described by Tessier, David, This,
116 Boursiquot, & Charrier (1999):

117
$$D_k = 1 - \sum_{j-1}^{l} p_j \frac{Np_j - 1}{N - 1}$$

118 Genetic distance among the accessions were calculated using the modified Rogers distance
119 method (Wright, 1978) in TFPGA, and a dendrogram was constructed using the unweighted
120 pair group method with arithmetic mean (UPGMA) clustering criterion using MegaX
121 (https://megasoftware.net/). The Mojena (1977) method was used to determinate the optimal
122 number of groups formed in the final stage of the UPGMA method.

123 Functional annotation of the loci associated with traits was performed using the sequences that
124 gave rise to the SSR marker using the nonredundant NCBI database through BLASTX with
125 Blast2GO software and applying standard parameters.

126 The transferability rate analysis was performed through evaluation of the presence or absence
127 of products in 3% agarose gel electrophoresis, and applying the formula:
128 $\text{Transferability}(\%) = \frac{\text{number of transferible loci}}{\text{number of tested loci}} \times 100.$

129 **RESULTS**

130 **Microsatellites search analysis, primer design and selection for validation**

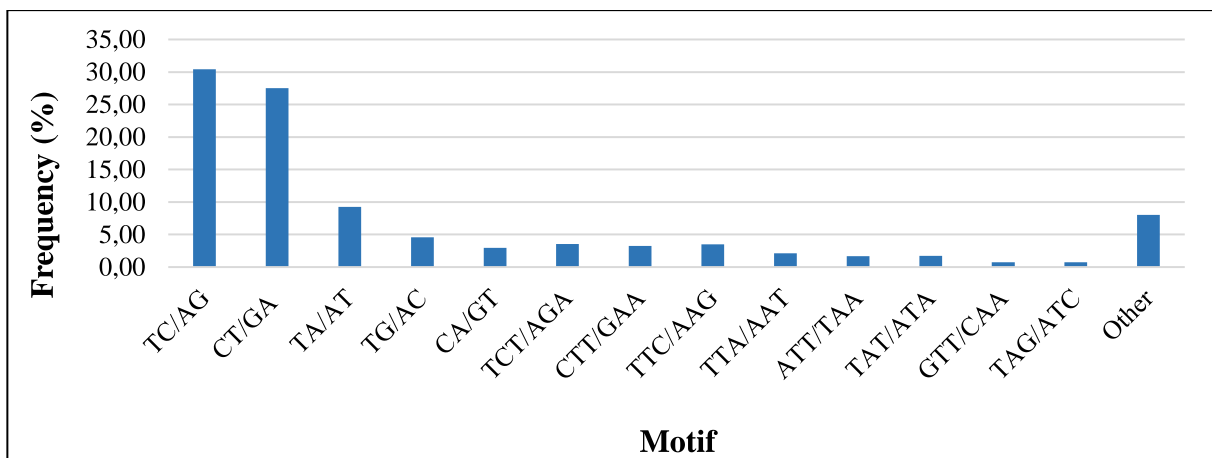
131 A total of 4,461 potential SSRs were identified in 4,046 transcript sequences, and 366 sequences
132 contained more than one SSR. The number of SSRs motifs varied. Dinucleotide repeat motif
133 was the most abundant (3,331; 74.67%), followed by trinucleotide (1,105; 24.77%) (Table 2).
134 Additionally, 164 SSRs presented compound motifs. The number of repeat units ranged from
135 eight to 31. The most common repeat was eight (1,423; 31.90%), followed by nine (590;
136 13.23%). The most common motif was TC/AG (1,356; 30.40%), followed by CT/GA (1,227;
137 27.50%) (Figure 1). The remaining motifs have lower frequencies than those.

138

139 **Table 2.** Summary of the SSRs search results.

Repeat type	Repeat number					Total
	8	9	10	11	≥12	
Dinucleotide	716	490	333	404	1.388	3.331
Trinucleotide	694	94	106	81	130	1.105
Tetranucleotide	7	4	2	2	0	15
Pentanucleotide	0	1	0	0	0	1
Hexanucleotide	6	1	2	0	0	9

140



141

142 **Figure 1.** Frequency distribution of SSRs based on the motif sequence types.

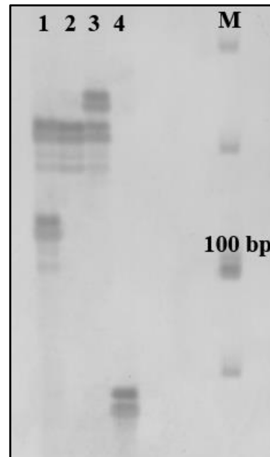
143

144 A total of 2,856 SSRs had primer pairs designed. The flanking sequences of the others 1,605
145 SSRs did not fulfill the primer design criteria, which did not provide suitable PCR primers pair
146 for them. Three alternative primer pairs were designed for each SSR locus. After filtering, 999
147 SSRs were found associated with some ORFs. Dinucleotides were the most common motif

148 (628; 62.86%), followed by trinucleotide (336; 33.63%), compound SSRs (31; 3.10%),
149 hexanucleotide (two; 0.20%), tetra and pentanucleotide (one; 0.10%).
150 One hundred eighty-six primer pairs were used for validation (Supplemental Table S1), from
151 which 179 were dinucleotide and seven were trinucleotide.

152 **Polymorphism analysis and cross-species transferability**

153 From the 186 tested loci, 148 (79.57%) presented amplification products, from which 87 loci
154 (58.78%) produced good amplification products, with good intensity, absence of secondary
155 products or nonspecific bands.
156 Sixty-four were analyzed for polymorphism. Only one (1,56%) was monomorphic among a set
157 of 19 *A. pintoi* accessions (Supplemental Table S2). A total of 464 alleles were found, with
158 fragments ranging from 100 to 300 bp in length. The number of alleles per locus varied from
159 two to 18, with an average of 7.37 alleles per locus.
160 The expected (H_E) and observed (H_O) heterozygosities were 0.72 and 0.31, respectively. Only
161 the locus Ap(AG)87 had an H_O higher than H_E. The PIC values ranged between 0.05 and 0.92.
162 Using the formulation proposed by Botstein et al. (1980) 53 loci were highly informative (PIC
163 > 0.5), nine were reasonably informative ($0.5 > PIC > 0.25$), and one slightly informative (PIC
164 < 0.25).
165 The *D* values ranged from 0.06 to 0.99, with a mean of 0.80. The Ap(TC)88 marker presented
166 the highest *D* value (0.99), and provide the identification of 15 unique profiles, with exclusive
167 alleles for each accession. The remaining four undifferentiated accessions (V 14951 and V
168 15062, V 6741 and V 6791wf) were distinguished by Ap(CT)54 profile. This marker subset
169 allowed the successful molecular differentiation of all core collection genotypes analyzed.
170 The dendrogram obtained by the UPGMA and with Mojena (1977) method did not represent
171 defined groups in clustering. Redundant genotypes weren't detected.
172 To perform the fingerprinting of the four cultivar, we used the loci that presented higher values
173 of PIC and *D*. The loci Ap(CT)75 was the most effective for discriminating among the cultivars
174 analyzed in this study, because showed a unique profile for all genotypes (Figure 2).



175

176 **Figure 2.** Amplification profile of the four cultivars of *A. pintoi* with the locus Ap(CT)75
177 analyzed by polyacrylamide gel. 1–4: Belomonte, Amarillo MG-100, Alqueire 1, and BRS
178 Mandobi, respectively. M: 10 bp ladder.

179

180 **Functional annotation analysis**

181 The sequences that gave rise to the SSR markers were blasted against the nonredundant protein
182 (nr) NCBI database using BLASTX and similarity was found for 120 (64.52%) sequences
183 (Supplemental Table S1). The Ap(AT)40, Ap(CT)54, Ap(CT)134, Ap(TC)90, and Ap(TC)99
184 sequences showed similarity to genes FMO1, At1g01540, OXI1, ACD6, and IQD1,
185 respectively, which are related with plant pathogen resistance and defense response.
186 The Ap(CT)15, Ap(CT)33, Ap(GA)39, Ap(AG)66, Ap(GAA)106, Ap(AG)89, Ap(GA)29, and
187 Ap(GA)111 presented similarity with genes RMS3, LHT, COG, UBP3, CFM3, ATHB-13,
188 *APETALA* (AP), and *CAULIFLOWER A* (CAL-A), respectively. These genes are related to
189 plant growth, development and differentiation of cells and tissues, as well as correct formation
190 of flowers, pollen, and seeds.

191 **Cross-species transferability analysis**

192 Of the 63 polymorphic markers, 33 SSRs were used for transferability analysis in the *Arachis*
193 genus. The loci presented a variable transferability rate. The highest percentage was observed
194 in the species *A. repens* (100%), followed by *A. glabrata* and *A. hypogaea* (87.88% and 84.85%,
195 respectively), *A. duranensis* (72.73%), *A. valsii* (69.70%) and, finally, *A. ipaensis* and *A.*
196 *stenosperma*, with the lowest transferability rate (63.64%). Sixteen loci (45.71%) amplified in
197 all analyzed species. The loci Ap(GA)48, Ap(CT)72, Ap(TCT)118 and Ap(AG)124 amplified
198 only in *A. repens*. The Ap(TCT)118 marker, showed similarity with *replication protein A 70*
199 *kDa DNA-binding subunit B-like isoform X4* (RPA1B), which may functions mostly in DNA
200 repair (Ishibashi et al., 2005). All loci showed an adequate profile in *A. repens*, the closest

201 species of *A. pintoi*, with absence of secondary products or nonspecific bands. For the other
202 species, 22 (66,66%) loci had the desirable amplifying characteristics (Supplemental Table S3).

203 **DISCUSSION**

204 **SSR validation and polymorphism analysis**

205 The amount of SSR markers that presented amplification products (79.57%) indicates the
206 efficiency of the design and filtering pipeline. The failure of amplification in the remaining
207 20.43% SSR markers may be due to primer mismatch, the extension of primers across a splice
208 site or the presence of large introns in the genomic DNA fragment to be amplified (Varshney
209 et al., 2006). In addition, the presence of multiples sites complementary to the primers may
210 have caused more than one amplicon in some loci (Tanwar, Pruthi & Randhawa, 2017).
211 Changes in the designs of the primers may be a solution to these problems.

212 The values of H_E and H_O found in our study were similar to those found by Azêvedo et al.
213 (2016), indicating that genic markers were also polymorphic and effective in accessing the
214 genetic diversity. The values of H_O higher than H_E observed in Ap(AG)87 could be due to the
215 mixed reproductive system of *A. pintoi*, which may provide crossings between different
216 accessions before the collection and the heterozygosis could be preserved by vegetative
217 propagation (Azêvedo et al., 2016; Oliveira et al., 2019).

218 The PIC value was variable, but the majority (84.13 %) were highly informative. The average
219 PIC value (0.70) was higher than those observed in three previous reports in *A. hypogaea*
220 (Bosamia, Mishra, Thankappan, & Dobaria, 2015; Wang et al., 2018), suggesting that these
221 new markers are highly informative even though they are derived from expressed regions.
222 Markers with higher *D* values are the most recommended to use in fingerprinting studies due
223 to the high capacity of discriminating genotypes. However, the selection of markers for cultivar
224 identification should also consider the quality of amplification products and the minimal
225 number of markers that could discriminate varieties (Rosa et al., 2010). The loci Ap(TC)88 and
226 Ap(CT)54 fulfill these requirements and are purposed to discriminate efficiently all the
227 genotypes. In addition, Ap(CT)75 marker was able to differentiate the four cultivars analyzed.
228 Therefore, the set of markers developed in this study would be a molecular tool for cultivar
229 protection in breeding programs.

230 **Functional annotations analysis**

231 To verify if the new developed SSR markers could possibly be associated with genes of
232 agronomic importance, we annotated the sequences from which the markers were originated.
233 The FMO1, ACD6, At1g01540, OXI1, and IQD1 genes and they have an important role on the
234 pathogen resistance and defense response mechanisms (Rentel et al. 2004; Levy, Wang, Kaspi,

235 Parrella, & Abel, 2005; Bartsch et al., 2006; Zhang, Shrestha, Tateda, & Greenberg, 2014;
236 Tateda, Zhang, & Greenberg, 2015; Shoala, Edwards, Knight, & Gatehouse, 2018). The study
237 of these genes may help to identify genotypes with improved resistance to pathogens that affect
238 the correct development of the *Arachis*.
239 The LHT, COG, UBP3, CFM3, *APETALA* (AP), and *CAULIFLOWER A* (CAL-A) are related
240 with the correct formation of flowers, pollen, and seeds pathways (Fang, Shen, Liu, Tang, &
241 He, 2000; Lee & Tegeder, 2004; Hirner et al., 2006; Doelling et al., 2007; Liu et al., 2007
242 Asakura, Bayraktar, & Barkan, 2008; Foster, Lee, & Tegeder, 2008; Ishikawa et al., 2008; Tan
243 et al., 2016; Rui, Wang, Li, Tan, & Bao, 2020). They can affect the reproduction of the species.
244 Thus, studies evaluating the expression of these genes may help to identify genotypes with
245 reproductive problems.
246 **Cross-species transferability**
247 The total success transferability observed in *A. repens* demonstrates its monophyletic nature
248 with *A. pintoi.* Both species belong to the same taxonomic section and occur in the same region,
249 and genetic studies detected similar genomes from a common ancestor (Friend, Quandt,
250 Tallury, & Stalker, 2010; Azêvedo et al., 2016). RNA-based markers are also expected to be
251 transferable between related species and genera as primers may be designed from conserved
252 coding regions of genome. However, the success of transfer may be related to genomic
253 complexity, taxonomic distance, and the function/evolution of the gene from which the
254 expressed sequence tags (EST) primers are derived (Poczai et al., 2013). The lower percentages
255 observed in the other species may be related to molecular divergence between the studied
256 species and *A. pintoi.* They belong to sections that are more distant phylogenetically. Loci that
257 amplified in all species indicates were highly conserved among all of them, and some were
258 related to housekeeping genes as correct growth, development, and reproduction (Ap(GA)39,
259 Ap(TC)109, and Ap(AG)121), and to pathogen resistance (Ap(CT)54). Huang et al. (2016)
260 tested the cross amplification of SSR loci, developed from RNA-Seq, in different species of the
261 genus *Arachis* and found a transferability percentage similar to that observed in this study. On
262 the other hand, the loci (Ap(GA)48, Ap(CT)72, Ap(TCT)118 and Ap(AG)124) that amplified
263 only in *A. repens* may suggest that the flanking site of these loci has changed over time. As the
264 Ap(TCT)118 is related to housekeeping gene responsible for DNA repair, so the flanking
265 primers may be extending across a splice site with a large intron or chimeric cDNA contigs
266 (Varshney et al., 2006; Tanwar, Pruthi & Randhawa, 2017), which caused primers annealing
267 problems. Therefore, changes on primer designs may help to solve this problem.

268 SSRs-based genome-wide association studies in *Panicum maximum* (Simeão, Souza, Jank & 269 Souza, 2015), *Gossypium* spp. (Nie et al., 2016) and *Saccharum* spp. (Barreto et al., 2019) has 270 demonstrated the great potential of these approach to discover MTAs. This information has 271 giving important insights about the gene networks of expressed traits, and assisting breeding 272 programs to increase the genetic gain rate of target traits, which is very important to improve 273 the speed and efficiency of the genome selection of superior genotypes and development of 274 new cultivars. Further studies are necessary to validate the putative loci found and characterize 275 them with GWAS to discover MTAs, which may play an important role to accelerate genetic 276 gains in *Arachis* breeding programs.

277 **ACKNOWLEDGMENTS**

278 We are gratefully to Governo Federal/ CNPq/ Governo do Estado do Acre/ FAPAC TO 279 024/2018, Embrapa to financial support. We thank the staff of the Life Sciences Core Facility 280 (LaCTAD) from State University of Campinas (UNICAMP), for the Genomics analysis.

281 **REFERENCES**

282 Asakura, Y., Bayraktar, O. A., & Barkan, A. (2008). Two CRM protein subfamilies cooperate 283 in the splicing of group IIB introns in chloroplasts. *RNA, 14*:2319–2332.

284 Azêvedo, H. S. F. S. (2014). *Caracterização da diversidade genética de amendoim forrageiro* 285 *com marcadores microssatélites* (Master's thesis, Universidade Federal do Acre). 286 Retrieved from UFAC Library (http://www2.ufac.br/cita/dissertacoes/2014/hellen-287 sandra-freires-da-silva-288 azevedo.pdf/@@download/file/Hellen%20Sandra%20Freires%20da%20Silva%20Aze 289 %CC%82vedo.pdf).

290 Azêvedo, H. S. F. S., Benvindo, F. D., Cavalcante, L. N., Haverroth, M., Wadt, L. H. O., & 291 Campos, T. (2017). Transferability of heterologous microsatellite loci between species of 292 *Euterpe* genus. *Genetics and Molecular Research, 16*:1-7.

293 Azêvedo, H. S. F. S., Sousa, A. C. B., Martins, K., Oliveira, J. C., Yomura, R. B. T., Silva, L. 294 M., Valls, J. F. M., Assis, G. M. L., & Campos, T. (2016). Genetic diversity of the forage 295 peanut in the Jequitinhonha, São Francisco, and Paraná River valleys of Brazil. *Genetics* 296 *and Molecular Research, 15*(3):1-11.

297 Barreto, F. Z., Rosa, J. R. B. F., Balsalobre, T. W. A., Pastina, M. M., Silva, R. R., Hoffmann, 298 H. P., Souza, A. P., Garcia, A. A. F., Carneiro, M. S. (2019). A genome-wide association 299 study identified loci for yield component traits in sugarcane (*Saccharum* spp.). *PLoS One,* 300 *14*(7): e0219843. https://doi.org/10.1371/journal.pone.0219843

301 Bartsch, M., Gobbato, E., Bednarek, P., Debey, S., Schultze, J. L., Bautor, J., & Parker, J. E.
302 (2006). Salicylic Acid–Independent ENHANCED DISEASE SUSCEPTIBILITY1
303 Signaling in Arabidopsis Immunity and Cell Death Is Regulated by the Monooxygenase
304 FMO1 and the Nudix Hydrolase NUDT7. *Plant Cell, 18*:1038-1051.
305 https://doi.org/10.1105/tpc.105.039982.
306 Bosamia, T. C., Mishra, G. P., Thankappan, R., & Dobaria, J.R. 2015. Novel and stress relevant
307 EST derived SSR markers developed and validated in peanut. *PLoS One, 10*:e0129127.
308 Brewera, P. B., Koltaib, H., & Beveridge, C. A. (2013). Diverse Roles of Strigolactones in Plant
309 Development. *Molecular Plant, 6*(1):18-28.
310 Campos, T., Azêvedo, H. S. F. S., Oliveira, J. C., Ferreira Filho, J. A., Yomura, R. B. T., &
311 Silva, L. M. (2016) Protocolo para identificação de híbridos de amendoim forrageiro
312 utilizando marcador molecular microssatélite. Retrieved from
313 https://ainfo.cnptia.embrapa.br/digital/bitstream/item/156276/1/26263.pdf.
314 Doelling, J. H., Phillips, A. R., Soyler-Ogretim, G., Wise, J., Chandler, J., Callis, J., Otegui, M.
315 S., & Vierstra, R. D. (2007). The Ubiquitin-specific protease subfamily UBP3/UBP4 is
316 essential for pollen development and transmission in *Arabidopsis*. *Plant Physiology,*
317 *145*:801–813.
318 EMBRAPA. (2020). Balanço Social 2019. Retrieved from
319 https://bs.sede.embrapa.br/2019/balancosocialembrapa2019print.pdf.
320 Fang, L. X., Shen, R. J., Liu, P. L., Tang, Z. C., & He, Y. K. (2000). Molecular characters and
321 morphological genetics of CAL gene in Chinese cabbage. *Cell Research, 10*:29-38.
322 Foster, J., Lee, Y. H., & Tegeder, M. (2008). Distinct expression of members of the LHT amino
323 acid transporter family in flowers indicates specific roles in plant reproduction. *Sexual*
324 *Plant Reproduction, 21*(2):143–152.
325 Friend, S. A. D., Quandt, S. P., Tallury, H. T., & Stalker, K. W. H. (2010). Species, genomes,
326 and section relationships in the genus *Arachis* (Fabaceae): a molecular phylogeny. *Plant*
327 *Systematics and Evolution, 290*, 185-199.
328 Hanson, J., Johannesson, H., & Engström, P. (2001). Sugar-dependent alterations in cotyledon
329 and leaf development in transgenic plants expressing the HDZhdip gene ATHB13. *Plant*
330 *Molecular Biology, 45*:247–262.
331 Henriksson, E., Olsson, A. S. B., Johannesson, H., Johansson, H., Hanson, J., Engström, P., &
332 Söderman, E. (2005). Homeodomain leucine zipper class I genes in *Arabidopsis*.
333 Expression patterns and phylogenetic relationships. *Plant Physiology, 139*:509–518.

334 Hirner, A., Ladwig, F., Stransky, H., Okumoto, S., Keinath, M., Harms, A., Frommer, W. B.,
335 & Koch, W. (2006). *Arabidopsis* LHT1 is a high-affinity transporter for cellular amino
336 acid uptake in both root epidermis and leaf mesophyll. *The Plant Cell, 18*:1931–1946.
337 Hodel, R. G. J., Gitzendanner, M. A., Germain-Aubrey, C. C., Liu, X., Crowl, A. A., Sun, M.,
338 Landis, J. B., Segovia-Salcedo, M. C., Douglas, N. A., Chen, S., Soltis, D. E., & Soltis,
339 P. S. (2016). A new resource for the development of SSR markers: millions of loci from
340 a thousand plant transcriptomes. *Applications in Plant Sciences, 4*(6):1600024.
341 doi:10.3732/apps.1600024.
342 Hodel, R. G. J., Segovia-Salcedo, M. C., Landis, J. B., Crowl, A. A., Sun, M., Liu, X.,
343 Gitzendanner, M. A., Douglas, N. A., Germain-Aubrey, C. C., Chen, S., Soltis, D. E., &
344 Soltis, P. S. (2016). The report of my death was an exaggeration: a review for researchers
345 using microsatellites in the 21st century. *Applications in Plant Sciences, 4*(6):1600025.
346 doi:10.3732/apps.1600025.
347 Huang, L., Wu, B., Zhao, J., Li, H., Chen, W., Zheng, Y., Ren, X., Chen, Y., Zhou, X., Lei, Y.;
348 Liao, B., & JIANG, H. (2016). Characterization and transferable utility of microsatellite
349 markers in the wild and cultivated *Arachis* species. *PLoS One, 11*(5):e0156633.
350 doi:10.1371/journal.pone.0156633.
351 Ishikawa, T., Machida, C., Yoshioka, Y., Ueda, T., Nakano, A., & Machida, Y. (2008).
352 EMBRYO YELLOW gene, encoding a subunit of the conserved oligomeric Golgi
353 complex, is required for appropriate cell expansion and meristem organization in
354 *Arabidopsis thaliana*. *Genes to Cells, 13*:521–535.
355 Lee, Y. H. & Tegeder, M. (2004). Selective expression of a novel high-affinity transport system
356 for acidic and neutral amino acids in the tapetum cells of *Arabidopsis* flowers. *The Plant*
357 *Journal, 40*:60–74.
358 Levy, M., Wang, Q., Kaspi, R., Parrella, M. P., & Abel, S. (2005). *Arabidopsis* IQD1, a novel
359 calmodulin-binding nuclear protein, stimulates glucosinolate accumulation and plant
360 defense. *The Plant Journal, 43*:79–96.
361 Liu, C., Zhou, J., Bracha-Drori, K., Yalovsky, S., Ito, T., & Yu, H. (2007). Specification of
362 *Arabidopsis* floral meristem identity by repression of flowering time genes. *Development,*
363 *134*:1901-1910.
364 Luo, H., Xu, Z., Li, Z., Li, X., Lv, J., Ren, X., Huang, L., Zhou, X., Chen, Y., Yu, J., Chen, W.,
365 Lei, Y., Liao, B., & Jiang, H. (2017). Development of SSR markers and identification of
366 major quantitative trait loci controlling shelling percentage in cultivated peanut (*Arachis*
367 *hypogaea* L.). *Theoretical and Applied Genetics, 130*(8):1635-1648.

368 Maia, G. F. N. (2018). Desempenho produtivo de dois grupos genéticos de bovinos de corte em
369 pastos puros e consorciados na Amazônia Ocidental (Master's thesis, Universidade
370 Federal do Acre). Retrieved from UFAC Library
371 (http://www2.ufac.br/ppgespa_docs/dissertacoes/dissertacao_gerbson.pdf).
372 Marzec, M. (2016). Strigolactones as part of the plant defence system. *Trends in Plant Science,*
373 *21*(11):900-903.
374 Miller, M. P. (1997). Tools for population genetic analyses (TFPGA): A Windows program for
375 the analysis of allozyme and molecular population genetic data, version 1.3. Northern
376 Arizona University: Arizona, 33 p.
377 Mojena, R. (1977). Hierarchical grouping methods and stopping rules: an evaluation. *The*
378 *Computer Journal, 20*:359–363.
379 Nie, X., Huang, C., You, C., Li, W., Zhao, W., Shen, C., Zhang, B., Wang, H., Yan, Z., Dai,
380 B., Wang, M., Zhang, X., & Lin, Z. (2016). Genome-wide SSR-based association
381 mapping for fiber quality in nation-wide upland cotton inbreed cultivars in China. *BMC*
382 *Genomics, 17*:352-367.
383 Oliveira, J. C., Rufino, P. B., Azêvedo, H. S. F. S., Sousa, A. C. B., Assis, G. M. L., Silva, L.
384 M., Sebbenn, A. M., & Campos, T. (2019). Inferring mating system parameters in forage
385 peanut, *Arachis pintoi*, for Brazilian Amazon conditions. *Acta Amazonica, 49*(4):277-
386 282.
387 Palmieri, D. A., Bechara, M. D., Curi, R. A., Gimenes, M. A., & Lopes, C. R. (2005). Novel
388 polymorphic microsatellite markers in section *Caulorrhizae* (*Arachis*, Fabaceae).
389 *Molecular Ecology Notes, 5*(1):77-79.
390 Palmieri, D. A., Bechara, M. D., Curi, R. A., Monteiro, J. P., Valente, S. E. S., Gimenes, M. A.,
391 Lopes, C. R. (2010). Genetic diversity analysis in the section *Caulorrhizae* (genus
392 *Arachi*s) using microsatellite markers. *Genetics and Molecular Biology, 33*(1):109-118.
393 Palmieri, D.A., Hoshino, A.A., Bravo, J.P., Lopes, C.R., & Gimenes, M.A. (2002). Isolation
394 and characterization of microsatellite loci from the forage species *Arachis pintoi* (Genus
395 *Arachis*). *Molecular Ecology Notes, 2*(4):551-553.
396 Petersen, L. N., Ingle, R. A., Knight, M. R., & Denby, K. J. (2009). OXI1 protein kinase is
397 required for plant immunity against *Pseudomonas syringae* in *Arabidopsis*. *Journal of*
398 *Experimental Botany, 60*(13):3727–3735.
399 Poczai, P., Varga, I., Laos, M., Cseh, A., Bell, N., Valkonen, J. P. T., & Hyvönen, J. (2013).
400 Advances in plant gene-targeted and functional markers: a review. ***Plant Methods,***
401 *9*(6):1-31. doi:10.1186/1746-4811-9-6.

402 Rentel, M. C., Lecourieux, D., Ouaked, F., Usher, S. L., Petersen, L., Okamoto, H., Knight, H.,
403 Peck, S. C., Grierson, C. S., Hirt, H. et al. (2004). OXI1 kinase is necessary for oxidative
404 burst-mediated signalling in *Arabidopsis*. *Nature, 427*:858–861.
405 Rosa, P. M., Campos, T., Sousa, A. C. B., Sforça, D. A., Torres, G. A. M., & Souza, A. P.
406 (2010). Potato cultivar identifi cation using molecular markers. *Pesquisa Agropecuária*
407 *Brasileira, 45*(1):110-113.
408 Rui, Q., Wang, J., Li, Y., Tan, X., & Bao, Y. (2020). *Arabidopsis* COG6 is essential for pollen
409 tube growth and Golgi structure maintenance. *Biochemical and Biophysical Research*
410 *Communications, 528*:447-452.
411 Saint Germain, A., Clavé, G., Badet-Denisot, M. A., Pillot, J. P., Cornu, D., Le Caer, J. P.,
412 Burger, M., Pelissier, F., Retailleau, P., Turnbull, C., Bonhomme, S., Chory, J., Rameau,
413 C., & Boyer, F. D. (2016). An histidine covalent receptor/butenolide complex mediates
414 strigolactone perception. *Nature Chemical Biology*, *12*(10):787–794.
415 Shoala, T., Edwards, M. G., Knight, M. R., & Gatehouse, A. M. R. (2018). OXI1 kinase plays
416 a key role in resistance of *Arabidopsis* towards aphids (*Myzus persicae*). *Transgenic*
417 *Research, 27*:355–366.
418 Simeão, R. M., Sousa, A. C. B., Jank, L., Souza, A. P. (2015). Genome-wide selection for
419 apomixis in *Panicum maximum*. In: International Symposium of Forage Breeding, 5.,
420 2015, Buenos Aires. Proceedings… Buenos Aires: Facultad de Agronomía, 2015.
421 Taheri, S., Lee Abdullah, T., Yusop, M. R., Hanafi, M. M., Sahebi, M., Azizi, P., & Shamshiri,
422 R. R. (2018). Mining and development of novel SSR markers using next generation
423 sequencing (NGS) data in plants. *Molecules, 23*(2):399-419.
424 Tan, X., Cao, K., Liu, F., Li, Y., Li, P., Gao, C., Ding, Y., Lan, Z., Shi, Z., Rui, Q., Feng, Y.,
425 Liu, Y., Zhao, Y., Wu, C., Zhang, Q., Li, Y., Jiang, L., & Bao, Y. (2016). *Arabidopsis*
426 COG complex subunits COG3 and COG8 modulate golgi morphology, vesicle trafficking
427 homeostasis and are essential for pollen tube growth. *PLoS Genetics, 12*(7):e1006140.
428 https://doi.org/10.1371/journal.pgen.1006140.
429 Tanwar, U. K., Pruthi, V., & Randhawa, G. S. (2017). RNA-Seq of Guar (*Cyamopsis*
430 *tetragonoloba*, L. Taub.) leaves: *de novo* transcriptome assembly, functional annotation
431 and development of genomic resources. *Frontiers in Plant Science, 8*:91.
432 https://doi.org/10.3389/fpls.2017.00091.
433 Tateda, C., Zhang, Z., & Greenberg, J. T. (2015). Linking pattern recognition and salicylic acid
434 responses in *Arabidopsis* through ACCELERATED CELL DEATH6 and receptors. *Plant*

435 *Signaling & Behavior, 10*(10):e1010912.
436 https://doi.org/10.1080/15592324.2015.1010912.
437 Terra, A. B. C., Florentino, L. A., Rezende, A. V., & Silva, N. C. D. (2019). Forage legumes in
438 pasture recovery in Brazil. (In Portuguese, with English abstract.) *Revista de Ciências*
439 *Agrárias, 42*(2):305-313.
440 Tessier, C., David, J., This, P., Boursiquot, J. M., & Charrier, A. (1999). Optimization of the
441 choice of molecular markers for varietal identification in *Vitis vinifera* L. *Theorical*
442 *Applied Genetics, 98*:171-177.
443 Varshney, R. K., Grosse, I., Hähnel, U., Siefken, R., Prasad, M., Stein, N., Langridge, P.,
444 Altschmied, L., & Graner, A. (2006). Genetic mapping and BAC assignment of EST-
445 derived SSR markers shows non-uniform distribution of genes in the barley genome.
446 *Theoretical and Applied Genetics, 113*:239-250.
447 Wang, H., Lei, Y., Yan, L., Wan, L., Cai, Y., Yang, Z., Lv, J., Zhang, X., Xu, C., & Liao, B.
448 (2018). Development and validation of simple sequence repeat markers from *Arachis*
449 *hypogaea* transcript sequences. *The Crop Journal, 6*(2):172-180.
450 Wright, S. (1978). Evolution and the genetics of populations, vol 4: variability within and
451 among natural populations. Chicago: University of Chicago Press, 590 p.
452 Xiong, Y., DeFraia, C., Williams, D., Zhang, X., & Mou, Z. (2009). Deficiency in a cytosolic
453 ribose-5-phosphate isomerase causes chloroplast dysfunction, late flowering and
454 premature cell death in *Arabidopsis*. *Physiologia Plantarum, 137*:249–263.
455 Zhang, J., Liang, S., Duan, J., Wang, J., Chen, S., Cheng, Z., Zhang, Q., Liang, X., & Li, Y.
456 (2012). De novo assembly and characterisation of the transcriptome during seed
457 development, and generation of genic-SSR markers in peanut (*Arachis hypogaea* L.).
458 *BMC genomics, 13*(1):90-96.
459 Zhang, Z., Shrestha, J., Tateda, C., & Greenberg, J. T. (2014). Salicylic acid signaling controls
460 the maturation and localization of the *Arabidopsis* defense protein ACCELERATED
461 CELL DEATH6. *Molecular Plant, 7*:1365–1383.
462 zu Ermgassen, E., Alcântara, M., Balmford, A., Barioni, L., Neto, F., Bettarello, M., Brito, G.,
463 Carrero, G., Florence, E., Garcia, E., Gonçalves, E., Da Luz, C., Mallman, G., Strassburg,
464 B., Valentim, J., & Latawiec, A. (2018). Results from on-the-ground efforts to promote
465 sustainable cattle ranching in the Brazilian Amazon. *Sustainability, 10*:1301-1326.
466

467 **Supplemental Table S1.** Characterization of the 186 primer pairs used for validation, including the primer sequences, annealing temperature (Ta),
468 and expected size in base pair.

Primer	Motif	Forward primer sequence (5' - 3')	Reverse primer sequence (5' - 3')	Ta (°C)	Expected length (bp)	Amplification	BlastX hits
Ap(CT)01	(CT)16	CCTTCAGCACCCACATCTCA	GGTGGCCTTCAACACTCTCA	60.0	251	positive	lachrymatory-factor synthase
Ap(AG)02	(AG)8	ACGAGGGATTGAGGGGAGAA	CTTCTCCGTCGCAGTCTCTC	-	173	negative	no hit
Ap(GA)03	(GA)8	AAGAGGCGCTAGGAAGGAGA	AAGAGGCGCTAGGAAGGAGA	60.0	139	positive	no hit
Ap(CT)04	(CT)8	GGCTCCCTTGTGCCTATTGT	AGCCACGCTCTGCATTTTTG	-	178	negative	triose phosphate/phosphate translocator, non-green plastid, chloroplastic
Ap(TC)05	(TC)9	ACTCCCCTTCCTTCCTTCCT	TCCTTCTTCCTGCACGTCAC	60.0	167	positive	50S ribosomal protein L31, chloroplastic-like
Ap(TC)06	(TC)14	GCTCCTCCTAACTCCCAACG	GGCGTGGGAGTGTTTAGTGA	60.0	243	positive	Dihydrolipoyl dehydrogenase 2, chloroplastic
Ap(AT)07	(AT)8	ACCATTTTCTCGAGCCACCT	ACAGGGAAAGGTGGTGCATG	60.0	124	positive	no hit
Ap(AG)08	(AG)12	CAAATTGGCAAAGGGCGGAG	CCACCGAGTGTGTAACTGCT	-	151	negative	recQ-mediated genome instability protein 2
Ap(GA)09	(GA)8	CGTCGTCGCCAATGAGAAAC	CATGGTAGCGACGACGATGA	-	226	negative	no hit
Ap(AG)10	(AG)15	CCACACACTTCACCCACACT	TCGAGCCCTTTCTTAGCGTC	-	209	negative	no hit
Ap(AG)11	(AG)20	CAACCTTTGCAGCAGCCTTT	GCCATCATTCCACCACCTGT	45.0	246	positive	growth-regulating factor 1
Ap(CT)12	(CT)9	CGAGCTCACTTCCCTTCCTC	TGGCGAAGGAAAAGTGGGAT	45.8	268	positive	RNA-dependent RNA polymerase 6
Ap(TC)13	(TC)10	GTGCCAAAATGGAGGGTCCT	GCGGTCGAGACATAGCAGTT	60.0	231	positive	no hit
Ap(AG)14	(AG)9	CCAATTTCCCACTCCGACGA	GCTTCCGCCATTAATGCCTC	58.5	192	positive	no hit
Ap(CT)15	(CT)10	ACCCCAAACCTTCAAAACCA	ATACCGTGGGCCAAGACAAG	60.0	133	positive	strigolactone esterase RMS3
Ap(AG)16	(AG)9	ACTTCCGGTGTGCATGTCAA	CCACAACGTTGTTGGCATCA	60.0	176	positive	no hit
Ap(AG)17	(AG)15	AGAAGAGAGGAGGGGACGAC	TCCATCAACAGCAGCAGAGG	60.0	215	positive	uncharacterized protein DS421_3g81360
Ap(TC)18	(TC)8	ACCAACCCGACCTTATAGCT	TTCTCGGCAGCATCTTGGAG	55.4	219	positive	remorin

469 (-) = absence of amplification; (+) = observed amplification.

470 **Supplemental Table S1.** Characterization of the 186 primer pairs used for validation, including the primer sequences, annealing temperature
471 (Ta), and expected size in base pair. (continued).

Ap(CT)19	(CT)10	GGCTCCGCTCATCACTCTTT	CTCGCCATTCTCCACCACTT	45.0	229	positive	katanin p60 ATPase-containing subunit A-like 2
Ap(CA)20	(CA)12	CCTCTCACACTCAGCCTCAC	ACTCAGCACCAAGGATGCAA	50.2	216	positive	no hit
Ap(GA)21	(GA)8	CCACCACTTTTCCACCCCTT	CGCGTGTGACATCTTTGGTC	58.5	260	positive	no hit
Ap(TA)22	(TA)10	ACAAAGGCAACAACACGGTT	CACGTCGGGTTTCAGGTACA	60.0	162	positive	filaggrin-like isoform X1
Ap(CT)23	(CT)9	GTGGCGGTAACTAGTGAGCA	ATTGGTGGTGGAGCATGGAG	60.0	181	positive	pentatricopeptide repeat-containing protein At4g21065
Ap(AG)24	(AG)15	AAGAGGAAGGAAGCGCACTC	TGTTGCTCCACTCCTTGCAT	60.0	241	positive	ankyrin repeat domain-containing protein 30A-like
Ap(AG)25	(AG)8	GCGTGAACCAGAGGGAAGAA	CAGCGACGATGGTTTGACAC	-	201	negative	RNA-directed DNA polymerase
Ap(CT)26	(CT)21	GCCTCACCCATAACCTCCAC	TAGGAAGGCGAGGAAGGGAA	58.5	272	positive	no hit
Ap(TA)27	(TA)9	CATTAGCCTTTGCACGCACA	GAAGCCTTGGGGGAGCAATA	58.5	202	negative	inositol transporter 1 isoform X1
Ap(TC)28	(TC)9	CAACCCCGAAACCGAAACAC	GAGAGCAGCTTTGACTCCCA	55.4	130	positive	UPF0161 protein At3g09310
Ap(GA)29	(GA)10	AACCTTCCTCTGCAGCACTG	CCCCTTCAGCTGCAGTTACA	58.5	234	positive	Floral homeotic protein APETALA
Ap(GA)30	(GA)10	CGTCGCCGTCAAGCAAAAAT	ACGTCGCAAAACCTCAGCTA	60.0	278	positive	no hit
Ap(GA)31	(GA)8	CGAATCACTCGGTTGCTCAC	GAATGGAACCACGTCCCTGT	45.0	141	positive	WAT1-related protein At4g15540
Ap(TG)32	(TG)9	GCTACAGAAGGAGGCAGCAA	TCCAAATCCAGCCACAAACA	60.0	223	positive	probable ribosome biogenesis protein RLP24
Ap(CT)33	(CT)9	GCGGCATCTCTACCCAAGAA	ATCAGGTTCCGTTTGCGAGT	60.0	236	positive	Lysine histidine transporter
Ap(GA)34	(GA)12	CCCTTCTCATCGTTCAGGCA	AATTCAGGCCTCACTGGTCC	52.1	241	positive	GDSL esterase/lipase CPRD49
Ap(CT)35	(CT)10	CAGCTGCGCAACCAAAAGTA	CCGGCCTTCTTGGTTCTCTT	60.0	169	positive	60S ribosomal protein L37a
Ap(CT)36	(CT)12	GCCAATGCCGACATCACATC	TGTTTCCGGCGTTCATCTCA	52.1	198	positive	no hit

472 (-) = absence of amplification; (+) = observed amplification.

Supplemental Table S1. Characterization of the 186 primer pairs used for validation, including the primer sequences, annealing temperature (Ta), and expected size in base pair. (continued).

Ap(AG)37	(AG)9	TGTGCTCGAATTGTCCCCAA	GACTGCAGATTCACCGCTCT	-	219	negative	mitochondrial adenine nucleotide transporter ADNT1
Ap(TC)38	(TC)8	GCACTGCAGTTACACCTCCT	CGGCGTCTAAATTCCCCGTA	60.0	134	positive	26S proteasome non-ATPase regulatory subunit 12 homolog A
Ap(GA)39	(GA)10	TGGTTGTGATGAGTAGGGAGA	GACGATGATATGGCGTGGGT	58.5	164	positive	Conserved oligomeric Golgi complex subunit
Ap(AT)40	(AT)9	AGCTTGCTTAGGTTGTTGAGA	GGAGTCAGTGGCCTCAAACA	58.5	192	positive	probable flavin-containing monooxygenase 1
Ap(GA)41	(GA)9	GTGGATAAGGTCGCCACCAT	CAGCAGTGGCATGGAAAACC	60.0	270	positive	no hit
Ap(GA)42	(GA)15	CCGTCGCCATTGCTGTTAAA	GGTGACGGCAATAACAGGGA	58.5	153	positive	no hit
Ap(CT)43	(CT)8	TCAAACTCCGACCGCTTCTC	AGGAGTGTCGAGGTGTCCAT	60.0	223	positive	patellin-6
Ap(CT)44	(CT)15	GCTCGTCTTCGTCTCCTCTG	GAAGAAGAAGACGGTGGCCA	60.0	168	positive	no hit
Ap(GA)45	(GA)9	TTATCCTCACCCCTCGCTCA	CTCTCACCAACCCCAATCCC	60.0	275	positive	no hit
Ap(GA)46	(GA)9	AATCCCACACGACCACCATC	TTTTGCGCACACGACGAAAT	48.2	255	positive	no hit
Ap(GA)47	(GA)12	GTCCGTCGTCGTCTTCGTTA	GACAGAGGCAGAAGCTCTGG	48.2	172	positive	Transcription initiation factor TFIID subunit
Ap(GA)48	(GA)9	CTCTGCCTAACACCAGTGCA	TGGTGCTTCGAGTTTGGACC	60.0	107	positive	no hit
Ap(TC)49	(TC)12	TGCTCTCTGCGTCCTTTCTG	CTTGGGCACTAGGAGTGGTG	52.1	165	positive	ras GTPase-activating protein-binding protein 1
Ap(CT)50	(CT)8	GCGCATTGTTGGGTTCCATC	GAACACCAGACAAAGCTGCA	60.0	225	positive	mitochondrial import inner membrane translocase subunit TIM22-2
Ap(TA)51	(TA)8	CGCTCTCTGCTCTCTTCACC	TGCTTGGTTCTCAGCACACT	52.1	280	positive	GRF1-interacting factor 1
Ap(AG)52	(AG)9	TGTCAGCTTGTTGGGCTTCA	CACTCGCCGAGGAGAAGAAG	48.2	154	positive	Calcium-transporting ATPase 4, plasma membrane-type
Ap(CT)53	(CT)12	ACCCGAGAAGTAGAAGGGGA	TGGGGCAAGTCCAATCATCC	-	185	negative	ranBP2-type zinc finger protein At1g67325

(-) = absence of amplification; (+) = observed amplification.

476 **Supplemental Table S1.** Characterization of the 186 primer pairs used for validation, including the primer sequences, annealing temperature
477 (Ta), and expected size in base pair. (continued).

Ap(CT)54	(CT)8	GAAACTCTCACTGGCCGGAA	GAGGATTCCGATGACGACCC	58.5	268	positive	probable serine/threonine-protein kinase At1g01540
Ap(CT)55	(CT)8	GATACCGGATAGGGCGGTTG	CTGGGAGACGCCATCTTTGT	58.5	236	positive	arabinogalactan peptide 22
Ap(TC)56	(TC)8	CGGCTACTCGTGTTCCTTCA	AACCGTCTCTGCAGCTTTCA	58.5	198	positive	no hit
Ap(AG)57	(AG)12	AGAAGAGGGGAAGAGAGGGG	TCCCTCTCTGGCATCTTCCT	60.0	167	positive	no hit
Ap(AG)58	(AG)11	TGCCACCGAAGAAAGGAAGA	CCACTGTCGCTACCAACACT	58.5	140	positive	no hit
Ap(AG)59	(AG)8	CCCACGCTCATACACCATCA	TCAAACCAATGCTGCCTCCT	60.0	277	positive	cytokinin riboside 5'-monophosphate phosphoribohydrolase LOG1
Ap(AG)60	(AG)8	CATCACTGCCCTGTTTCCCT	CTCCTTCTTCGCGGTTCTGT	-	239	negative	uncharacterized protein DS421_12g364340
Ap(GA)61	(GA)9	CGGAGAGGGAGAAGAAGGGA	GTCTCCTTCTCCGCCTTTCC	60.0	176	positive	no hit
Ap(TC)62	(TC)9	GCAACGTCCTTGTTTCGGTC	ACGCGCGTAGAAGATAAGGG	60.0	250	positive	no hit
Ap(TA)63	(TA)8	GCTTTAACATGCCGGCACAA	AGCAGAGTTACACTGGACGT	58.5	268	positive	no hit
Ap(TC)64	(TC)8	GCACTCCGTGTCTCAGTCTC	AGAAGGTAGCTGCTAGGGCT	-	169	negative	no hit
Ap(TA)65	(TA)9	GCGAAGTAGTCCTTGGGTGT	AGGAAAGCCAAGGAACAGCA	58.5	266	positive	uncharacterized protein DS421_4g111450
Ap(AG)66	(AG)10	TCCCGCTTGGTTCTTCAGAC	AGCCTTCTCGAGCTTGGAAC	60.0	188	positive	Ubiquitin carboxyl-terminal hydrolase 3
Ap(GA)67	(GA)27	AGTGTGTCACTCTTTCGCCA	CCTCCACCCTTTCTTTCCCC	-	122	negative	Transcription initiation factor TFIID subunit
Ap(CT)68	(CT)9	ACATTGCAGGCACCCATCTT	AGCACCATCACCAACAACCC	55.4	192	positive	no hit
Ap(TC)69	(TC)20	ACACACACCCACACCACTTT	GGGAGATGAAGGTGGTGAGG	60.0	198	positive	nudix hydrolase 25
Ap(TC)70	(TC)9	AGTAGCCAGCATCAGCAGTC	GCAGCGCAAGTTAGCAAGTT	48.2	229	positive	double-stranded RNA-binding protein 2
Ap(TA)71	(TA)9	TCGTGCTCCCGATTCTTGAC	GTGTGAAGTAATGCGTGCAA	60.0	267	positive	no hit
Ap(CT)72	(CT)8	ACCGGTTGAATTGGGTCCAA	TTCCCAGCGTGAATGGCATA	60.0	108	positive	uncharacterized protein LOC107621065

478 (-) = absence of amplification; (+) = observed amplification.

479 **Supplemental Table S1.** Characterization of the 186 primer pairs used for validation, including the primer sequences, annealing temperature
480 (Ta), and expected size in base pair. (continued).

Ap(TC)73	(TC)10	TGATTCCCATGAAGCGCCAT	CCCATTCGTCGTCGGAGATT	58.5	235	positive	protein CHROMATIN REMODELING 19 isoform X1
Ap(CT)74	(CT)16	CAGCTTCTCCTCGACGGTTT	TCCCTCTCATGTTCAGCTGT	52.1	185	positive	NADPH-dependent diflavin oxidoreductase 1 isoform X1
Ap(CT)75	(CT)9	GGTTCATCGCCTTCCTCTGT	GAAGACCAGCTTCCATGGCT	58.5	270	positive	probable ribose-5-phosphate isomerase 2
Ap(TG)76	(TG)8	TGTGAACGGTGGTGATCAGT	AAGTAACGTGACCACCGCTT	58.5	235	positive	no hit
Ap(CT)77	(CT)8	CCACACCCTTCCCTTACCAC	ATAAGCATGCAGGGTGGTGG	-	204	negative	aspartic proteinase PCS1
Ap(GA)78	(GA)9	GTCCGTCGTCGTCATCGTTA	GAGCAAGTGAGGGCAGATGA	52.1	117	positive	no hit
Ap(AT)79	(AT)8	TGCCTCGCAACACTCAGATT	AGGAGGTGTAGGAAGAGCCA	-	154	negative	uncharacterized protein DS421_3g98260
Ap(AC)80	(AC)8	CAGTGGACTGATTCACCCCC	AAGCTCGTCTCTCCCTCCAT	48.2	267	positive	calpain-type cysteine protease DEK1
Ap(AG)81	(AG)22	AGGAAGGAGGAGAGGACGAC	CCTCCCGCGACCTCTACA	59.4	124	positive	no hit
Ap(AC)82	(AC)8	ACAGTCCCTGTCCTACAGCT	ATGCTGGGAAGTGGGAAGTG	58.5	198	positive	no hit
Ap(CT)83	(CT)22	ACGGCGATTTCTACAAGCCT	GAAAGTTTGCGGCCTGTGAG	60.0	228	positive	plastidial pyruvate kinase 2
Ap(CT)84	(CT)22	TCTTTCGCTTGCACCATCCT	ATCCCTGTGGAGTATGCCCT	48.2	271	positive	serine/threonine-protein kinase HT1
Ap(AG)85	(AG)12	GATTCTCCGCCCGTCCTTAG	ATGATGTCGGTCACCACACC	-	219	negative	60S ribosomal protein L8-3
Ap(AG)86	(AG)8	AGACCAATCATTGCCCATGT	GCTGGTCCCTCTTTCCCTTC	55.4	211	positive	no hit
Ap(AG)87	(AG)9	TGAGCCCAGCCCTCTCATAT	CAGAGCGAGACCGGAGAATC	55.4	205	positive	selenoprotein F
Ap(TC)88	(TC)8	TGCGCCTTCTGGAACATTCT	GGCGCCACTGAATTTCGATC	55.4	238	positive	protein CutA 1, chloroplastic isoform X2
Ap(AG)89	(AG)8	ATGACCAAACTGGCCTCTGG	GCTAGCACCCCTTTGAGTCA	60.0	164	positive	homeobox-leucine zipper protein ATHB-13
Ap(TC)90	(TC)8	CACCACCACTGTACCGGATC	CTCGCTCCATGATGGCTGAT	55.4	212	positive	protein ACCELERATED CELL DEATH 6
Ap(CT)91	(CT)8	AGCAGCAGTAGGAGTAGGCT	TGGAACTCAAAACCCTGCCT	60.0	192	positive	transcription factor UNE12-like

481 (-) = absence of amplification; (+) = observed amplification.

482 **Supplemental Table S1.** Characterization of the 186 primer pairs used for validation, including the primer sequences, annealing temperature
483 (Ta), and expected size in base pair. (continued).

Ap(CT)92	(CT)12	ACACACTGTTCCACACGACA	GGTGAAGGTGGAGGTTGGAG	58.5	115	positive	Peroxisomal membrane protein
Ap(TC)93	(TC)9	TCATCACCCTGACGGAACAC	AGGGGCTCCATCTCACAGAT	55.4	212	positive	no hit
Ap(CT)94	(CT)9	CCAGTCTCTGCTTCTGCCTC	TTCCAACTTGGGGCCATTGT	60.0	140	positive	no hit
Ap(AC)95	(AC)9	GCAACCATCCTCTTCACCCA	GATGACGACGACCCTAGCAG	58.5	143	positive	no hit
Ap(TC)96	(TC)11	TGCCTCTTCCCCTTCCTTCT	GAGGTTGGCGATGGCTTTTC	60.0	248	positive	Avr9/Cf-9 rapidly elicited protein
Ap(AT)97	(AT)8	GGCAGCATGGAGTCTTCACT	CCTCAAGTTCGCTTTTCCCC	-	252	negative	peroxidase 21
Ap(GA)98	(GA)9	GTGGCGCTCATGACTGATGA	GAGGCAGCAGAGTCTTCGTT	55.8	148	positive	TBC1 domain family member 22B
Ap(TC)99	(TC)11	ACTCACTCGCTCACTCTCTT	AGAGAACCAGCTTCCCCTCT	60.0	141	positive	protein IQ-DOMAIN 1
Ap(AG)100	(AG)18	CACAGCTCTGCAGGTCTCTC	GGGGCGTCGGAATCACTATC	45.0	214	positive	uncharacterized protein LOC112800625
Ap(CT)101	(CT)10	GCCGAAAATCCTCCGATTTCA	GACAGTGTCGCCTTCCTTCA	60.0	269	positive	10 kDa chaperonin, mitochondrial
Ap(AGG)102	(AGG)8	GGAAGAAGAAGGGAGTGGCT	TGCCTGTAGTGTGCTTGAGT	60.0	108	positive	no hit
Ap(TC)103	(TC)10	TCTCCACCTTCACCATACGC	CGACAGCAGAGGAAGCTCAT	-	106	negative	no hit
Ap(AG)104	(AG)12	TAGCTGCTGAACTGTCGCAG	GATGAGGCAGCAGCTTCTGA	50.2	226	positive	hypothetical protein Ahy_A10g049139 isoform A
Ap(AT)105	(AT)12	ACAACAACCAAAACCTTGACGA	CACCAATAATGCGCATTCACA	60.0	280	positive	plasma membrane-associated cation-binding protein 1
Ap(GAA)106	(GAA)13	CTCACCAACGACCCACTTCA	TCATCAGCAGCACCACCAAT	59.4	206	positive	CRM-domain containing factor CFM3, chloroplastic/mitochondrial
Ap(TG)107	(TG)9	ACAGAAGTAAAGCAGCAGAAGGA	GGCACCTAAACATTCCTCCCA	58.5	221	positive	no hit
Ap(CT)108	(CT)15	CTCACCCTAGTCGCATCACC	GGTCAGATCTCGCCGGAATT	58.5	257	positive	uncharacterized protein LOC110279909
Ap(TC)109	(TC)8	ATCGCTATCTGCCTCCGTTG	TCAGGTGGGTAGGCGTAGAA	60.0	274	positive	methionine-S-oxide reductase

484 (-) = absence of amplification; (+) = observed amplification.

485 **Supplemental Table S1.** Characterization of the 186 primer pairs used for validation, including the primer sequences, annealing temperature
486 (Ta), and expected size in base pair. (continued).

Ap(TC)110	(TC)8	TGCTTGCCATCACTGTTTGC	TGAGGCCCAGTCGTTGAATT	60.0	114	positive	no hit
Ap(GA)111	(GA)10	TCGCTGCTTCTTCGTCTTCT	ACTCTTCCCCTTCCCATCGA	55.4	161	positive	truncated transcription factor CAULIFLOWER A
Ap(CT)112	(CT)8	ACACCTTTTCCTCTCCAGCG	GAGCCGTGCCTTGGATGTT	55.4	189	positive	no hit
Ap(CT)113	(CT)13	GTGTCTCATTCTCCCGCTCC	CGGTTTTCTTGGTTTGTGTGGT	58.5	128	positive	no hit
Ap(CT)114	(CT)16	TGAGAAACTGAGAAGGGTGCC	GCGAGGAGACTGAAACGGTT	58.5	218	positive	no hit
Ap(CT)115	(CT)9	CCTATCCCTTTCCGCTTCCC	ACAGCACTAGCAGCACCATT	-	191	negative	Histone deacetylase complex subunit SAP30
Ap(TCT)116	(TCT)14	CCATTCTCCTCAACCGCCTT	GCTGCTACGAATGGGGCTAT	55.4	113	positive	Ribosomal RNA small subunit methyltransferase H
Ap(TTC)117	(TTC)8	GATTGCTTCCACACTGCCCA	TGGGGATCATCGGTTTGTTGT	62.0	159	positive	no hit
Ap(TCT)118	(TCT)8	AGGAGGTCAGGTAGCGAGAG	GGTTGGGCTTGAAGGGGAAT	58.5	206	positive	replication protein A 70 kDa DNA-binding subunit B-like isoform X4
Ap(CAA)119	(CAA)11	TCGCCCTCAACAACCTTCTC	GAAGACCGACAGAGCACACA	46.8	277	positive	TIR-NBS-LRR type disease resistance protein
Ap(TCT)120	(TCT)11	CCAAGCCTCATCAACCACCT	GCGTCCGAGTTGGTTTGTTC	58.5	267	positive	zinc finger protein ZAT9
Ap(AG)121	(AG)8	AAGCTCTTCTCTCCTTCCCA	ACCACCATTTTGCTCGTCCA	58.5	124	positive	WAT1-related protein At4g08300
Ap(TG)122	(TG)8	AGACCGAGGGCAACTTGAAG	TCCATCAGCCACATTCACCA	52.1	278	positive	RNA demethylase ALKBH10B isoform X1
Ap(TC)123	(TC)8	TCTCATTCACTCAAAACAAAACCT	CACGACGGGTAGCTCTCATC	-	247	negative	auxin-responsive protein SAUR78-like
Ap(AG)124	(AG)10	ATCGCCACCATTGAACCTGT	ACTTCCTCCTCTCTCGCCAT	58.5	234	positive	no hit
Ap(TC)125	(TC)10	CCTTCTCTTCTCTCACTCACTCA	CATTGTGGTGGTGGTTGCTG	52.1	149	positive	formin-like protein 1
Ap(CT)126	(CT)9	AGCAGTTGGTTGGTTAGTGT	GCGGTTTCGGAATTGGTTGA	58.5	232	positive	no hit
Ap(AG)127	(AG)9	GGGTAGAATCCAAGACAGCGT	CAACCTCGGAAACACCCTCA	58.5	108	positive	protein LURP-one-related 17
Ap(GT)128	(GT)11	CCCCTCCATCCCTCCTCTAC	GCGTTCTAACTCACCACATCC	59.4	193	positive	no hit

487 (-) = absence of amplification; (+) = observed amplification.

488 **Supplemental Table S1.** Characterization of the 186 primer pairs used for validation, including the primer sequences, annealing temperature
489 (Ta), and expected size in base pair. (continued).

Ap(CT)129	(CT)17	GGGAGGCTGATAAGATGTAAGAT	AAAGAACACCCCAACCCCAG	60.0	223	positive	no hit
Ap(TC)130	(TC)8	CATCACCCAACATCCCTCCC	TCTTCGTCGTCGTCGGATTC	-	209	negative	probable membrane-associated kinase regulator 5
Ap(AC)131	(AC)9	AAAACACACACACACTCACAC	ACCACCTCCTCCTTCTCCTC	-	174	negative	no hit
Ap(TC)132	(TC)9	CCCACTGCTTTCACTCTCTCA	CGGCTCTATAACCCGCCATT	58.5	116	positive	ras-related protein RABA1b
Ap(TA)133	(TA)8	TCCGATCATATCCCTCTCCCT	TGTGAACACTTCTTAACTGCTCT	-	225	negative	no hit
Ap(CT)134	(CT)9	TCAGCCACCAAAATAAAATCCCC	CTCGCTAGGAACACCACTCC	58.5	258	positive	serine/threonine-protein kinase OXI1
Ap(AG)135	(AG)9	TCTCTGCCACACATCACACC	GGCTTTGAATGAATGAGTGAAACG	-	248	negative	no hit
Ap(GA)136	(GA)8	TGAGAAGGATTAGAACTTAGGGGA	CCATGTAGGCACTCCTCTCC	-	118	negative	no hit
Ap(AG)137	(AG)10	GCATCCAACATTACCAACATATCA	GTACCGTGCTGCTCGATGTA	58.5	158	positive	transcription factor RAX3
Ap(TG)138	(TG)9	TGTCGCATAGTTGAGTCGCA	ACCCCTCATCCACAAGAATCA	-	270	negative	no hit
Ap(CT)139	(CT)10	GCCCTAAACCCACACACTCA	AGAAGGAGCCAAAGAGGAAAAGA	60.0	161	positive	uncharacterized protein DS421_15g491500
Ap(CT)140	(CT)8	TCTCCATTGATTCTTTCTGCGT	TGAGTGAGAGAGAGGGTCCG	-	173	negative	no hit
Ap(CT)141	(CT)12	CACTTCCCCTCAAACCCCTC	TCTCTAACAATGCTACGGCGT	53.6	123	positive	no hit
Ap(TC)142	(TC)8	TCAATTCACACTCCCCACCC	ACAGCGAAAGAGGAATAGCGA	58.5	162	positive	short-chain dehydrogenase TIC 32, chloroplastic
Ap(CA)143	(CA)13	GCAAAGATGGAGAGCGGAGT	AGTGCGGGAAAATGAGGAGG	52.1	227	positive	extra-large guanine nucleotide-binding protein 3
Ap(GA)144	(GA)8	CTCTCTCAACCCTCCCTCCA	AGACCCAAAAACACCAGCCA	45.0	261	positive	peroxisomal acyl-coenzyme A oxidase 1
Ap(CT)145	(CT)16	CACCAGCAACAGCAACAACA	GGGGAGAAAGAGAAAGGGGG	53.6	209	positive	protein LIGHT-DEPENDENT SHORT HYPOCOTYLS 4-like
Ap(TG)146	(TG)8	AGGTTTACGGGGTGTGTGTG	CTGGGGATTTGACCGCTGAT	60.0	229	positive	no hit

490 (-) = absence of amplification; (+) = observed amplification.

491 **Supplemental Table S1.** Characterization of the 186 primer pairs used for validation, including the primer sequences, annealing temperature
492 (Ta), and expected size in base pair. (continued).

Ap(TC)147	(TC)9	TTCCAGGCATTCCAACAGCA	GTTTGGTGGTTGTGGTGGTG	58.5	258	positive	nucleobase-ascorbate transporter 3
Ap(TC)148	(TC)8	GCTGAGCACAATAGGCGTTC	AGGACTTCTGATGCCCCGTA	60.0	181	positive	inositol-tetrakisphosphate 1-kinase 3 isoform X2
Ap(CT)149	(CT)9	TCTCTCTACCGAATACCGCA	GGAGGAGGAGGAGGAGGAAG	48.2	263	positive	no hit
Ap(CT)150	(CT)8	CTGCTCTCTGCTGCCTCTTT	TGTCGTAAGGGGGTGGGATA	58.5	222	positive	ABC transporter D family member 1
Ap(CT)151	(CT)12	TGTGTTGTTGGCTCCTTCGT	GCAGATGGGGAGAGAGTGTG	57.3	109	positive	zinc finger CCCH domain-containing protein 41
Ap(AT)152	(AT)8	CCTTCTTCCCCCACTTCCAC	GCGGTGTGCCATGTTTGTAA	57.3	203	positive	Dynein light chain LC6, flagellar outer arm
Ap(CT)153	(CT)16	GAGAACAGCAGGAACGGGAA	AGTGATGAGTGAGGGGGTGT	58.5	186	positive	receptor protein kinase TMK1
Ap(CT)154	(CT)11	TCCTCCTCATTGCTCCCTCA	CGCCTTCTTCTCCTCCGATC	58.5	244	positive	2-alkenal reductase (NADP(+)-dependent)-like
Ap(AC)155	(AC)9	GCCATCTTCGTAACACACACAC	CACGGCACAAATGAGCAACA	58.5	227	positive	probable methyltransferase PMT23
Ap(TC)156	(TC)17	GGGTTGTGTCTGCGTGTTTG	ACATCAATCGGCTGGCTCAA	-	262	negative	protein EXPORTIN 1A
Ap(AT)157	(AT)8	TCTCTCGCCTGTCACTTACT	AGGAGAAGAAGAAAGGATGTGAGA	-	259	negative	no hit
Ap(TC)158	(TC)19	TCAAGTCAGAAAGGGGTGGC	GTCTTTGGTGCTGTGCTTGG	58.5	237	positive	cytochrome b5
Ap(TC)159	(TC)8	CTTGCCCCTGCCCTAGTAAT	TGGTGGTTGGTGTTGCTGAA	48.2	190	positive	no hit
Ap(GA)160	(GA)8	TGAGAGGTAAAGCGGTTGGT	GTTGTGTGAGTCCCTGGCTT	-	273	negative	WUSCHEL-related homeobox 5
Ap(AC)161	(AC)9	ACCCACATTTTCCCCTTCCC	GCGAGTTTGACGATGATGGC	52.1	254	positive	uncharacterized protein LOC112728362 isoform X2
Ap(TC)162	(TC)21	CTGGGACGCCGATAACAAGA	ACGAGTGGTGCTTACATTTCT	50.2	257	positive	transmembrane ascorbate ferrireductase 1
Ap(CT)163	(CT)12	GCTTTTGGATTCTTGAGACGGC	TTTCTGTGTGGTGTACGGCA	-	127	negative	scarecrow-like protein 13
Ap(AG)164	(AG)12	ACCTCCTTCTTCTTCTTCTCCC	TCCCTTCTAACGGCATCTTCT	-	246	negative	transcription repressor KAN1

493 (-) = absence of amplification; (+) = observed amplification.

494 **Supplemental Table S1.** Characterization of the 186 primer pairs used for validation, including the primer sequences, annealing temperature
495 (Ta), and expected size in base pair. (continued).

Ap(CA)165	(CA)11	CTCACCTCACTCCTCACCCT	CCTCTGCTCTTCTCCTCCCT	52.1	188	positive	no hit
Ap(GA)166	(GA)11	CGAAAGCGGAAAGAAAGGCA	TCTGCGAAGTGTCCAACCTC	55.4	105	positive	Protein TIME FOR COFFEE
Ap(TC)167	(TC)11	ACACAGACACAAAACACACGG	CGAGCAGGAAGAGAGCAGAG	46.8	219	positive	AP-1 complex subunit mu-2
Ap(TC)168	(TC)12	GGGTTGCTCCGATTTCTACG	CTGGTGGTGATTGTGGTGGT	55.4	113	positive	membrin-11 isoform X1
Ap(CT)169	(CT)9	AGCGATGATAGCGTTGGGAG	AGGGGAGACGAGGTGGAATA	53.6	177	positive	no hit
Ap(AG)170	(AG)20	AGGAGGAGGCACATGAATGA	GGACACAAGGAGACTGGAGC	-	268	negative	zinc finger CCCH domain-containing protein 58
Ap(TC)171	(TC)15	TTTGAGAGACTTGGGGGCAC	AGAAGGAGACGAGACGGGAA	-	280	negative	no hit
Ap(CT)172	(CT)12	CTCACACTTCACTCCCCACC	AGAAGGAAGAAGCAGCAGCA	-	170	negative	no hit
Ap(TC)173	(TC)12	ACCGATCCCCGTTTTGTCTC	GCAGCGAGCGAATTGAAGAG	55.4	188	positive	TLD domain-containing protein 1
Ap(TC)174	(TC)8	ACCATTCCGATCCACCAACC	TGTGTAAAACCAGCAGAGCCA	-	232	negative	hexose transporter
Ap(CT)175	(CT)8	AGCCCCATTCAAAAACACAGA	CTAAACCCACCACCGCTCAT	60.0	276	positive	GBF-interacting protein 1-like
Ap(CT)176	(CT)11	CAGTACACACGACCACGACA	TGAAGCGGGAGAGGAGGTTA	60.0	255	positive	ras-related protein RABA1f
Ap(CT)177	(CT)11	TCCCATCGCTATCCACACAC	CTGGAAAGTGGGTTGGTGGT	58.5	178	positive	Abscisic acid receptor
Ap(AC)178	(AC)10	CCACCAATCCAACCCCATCA	AGACGAAAGCCTGCGAAGAA	-	192	negative	serine hydroxymethyltransferase, mitochondrial
Ap(AG)179	(AG)8	ACACCACCACCACCAACAAA	TCGAAAACCACCTCCGTCTG	60.0	153	positive	transcription initiation factor IIB
Ap(TC)180	(TC)8	TGCGTTTTCTTTCTCTCAGGT	GGTTCTGAAGGTGGAGGTGG	59.4	151	positive	no hit
Ap(TA)181	(TA)9	TTTTCCCTCGCCTTGCTACT	AAACTCCTTCGTCCGTCCTG	-	175	negative	no hit
Ap(CT)182	(CT)9	CCACCCTTTTTCCTTCGCCT	GGTAGGAGGAGCAAGTGGGA	58.5	148	positive	no hit
Ap(TC)183	(TC)12	ACCCTCTCTGACTTCCCTCC	CGGAATGGTGAACGGGAGAA	60.0	237	positive	polycomb group protein EMBRYONIC FLOWER 2-like
Ap(AT)184	(AT)8	CGGCATAGAGAGAAGAGAAGGG	AACCGAACTTTGAACCACTCT	50.2	154	positive	plant/T7H20-70 protein

496 (-) = absence of amplification; (+) = observed amplification.

497 **Supplemental Table S1.** Characterization of the 186 primer pairs used for validation, including the primer sequences, annealing temperature
498 (Ta), and expected size in base pair. (continued).

Ap(TG)185	(TG)10	AGCAAGCATCCTCAACCTCC	CCCTCGTTACCACACACACA	52.1	139	positive	calmodulin-binding protein 60 B
Ap(TC)186	(TC)10	AGCAGAGAGGAGAGAAGCCA	GGAACAGAGGAACATGCCCA	-	109	negative	U-box domain-containing protein 26

499 (-) = absence of amplification; (+) = observed amplification.

Supplemental Table S2. Characterization of the 64 developed SSR markers, including the annealing temperature (Ta), observed size in base pair, number of alleles, expected (H_E) and observed (H_O) heterozygosity, polymorphism information content (*PIC*), and discriminatory power (*D*).

Primer	Ta (°C)	Observed length (range, bp)	Nº of alleles	H_E	H_O	*PIC*	*D*
Ap(CT)01	60.0	250-300	10	0.83	0.37	0.81	0.91
Ap(TC)05	60.0	250-300	3	0.57	0.25	0.55	0.79
Ap(TC)06	60.0	200-300	2	0.48	0.00	0.47	0.46
Ap(AT)07	60.0	100-150	8	0.79	0.26	0.77	0.88
Ap(AG)14	58.5	150-250	13	0.86	0.26	0.83	0.90
Ap(CT)15	60.0	100-150	3	0.49	0.21	0.48	0.61
Ap(TC)18	55.4	200-250	6	0.82	0.26	0.80	0.90
Ap(GA)21	58.5	250-300	9	0.83	0.19	0.81	0.90
Ap(CT)26	58.5	200-250	11	0.86	0.32	0.84	0.93
Ap(TC)28	55.4	100-200	10	0.91	0.37	0.88	0.94
Ap(GA)29	58.5	200-250	5	0.55	0.11	0.54	0.62
Ap(GA)30	60.0	250-300	7	0.71	0.35	0.69	0.85
Ap(TG)32	60.0	200-250	11	0.85	0.54	0.82	0.90
Ap(CT)33	60.0	200-250	4	0.47	0.16	0.46	0.57
Ap(CT)35	60.0	150-250	15	0.91	0.47	0.88	0.98
Ap(GA)39	58.5	150-200	3	0.54	0.21	0.52	0.68
Ap(AT)40	58.5	150-200	7	0.83	0.32	0.81	0.93
Ap(GA)42	58.5	150-200	7	0.84	0.37	0.82	0.93
Ap(GA)45	60.0	250-300	5	0.64	0.26	0.62	0.77
Ap(GA)48	60.0	100-150	2	0.49	0.11	0.48	0.62
Ap(CT)54	58.5	250-300	14	0.89	0.42	0.87	0.94
Ap(CT)55	58.5	200-300	12	0.88	0.16	0.85	0.91
Ap(TC)56	58.5	150-250	13	0.90	0.39	0.88	0.98
Ap(AG)58	58.5	100-200	4	0.60	0.25	0.56	0.62
Ap(GA)61	60.0	150-200	5	0.78	0.38	0.76	0.92
Ap(TA)65	58.5	250-300	1	-	-	-	-
Ap(AG)66	60.0	150-250	12	0.91	0.37	0.89	0.96
Ap(CT)68	55.4	150-300	18	0.93	0.47	0.91	0.98
Ap(CT)72	60.0	100-200	15	0.94	0.82	0.92	0.98
Ap(TC)73	58.5	250-300	4	0.59	0.26	0.57	0.73
Ap(CT)75	58.5	200-300	11	0.87	0.21	0.85	0.92
Ap(TG)76	58.5	250-300	4	0.44	0.11	0.43	0.51
Ap(AC)82	58.5	200-250	2	0.05	0.05	0.05	0.06
Ap(AG)86	55.4	200-250	3	0.44	0.14	0.43	0.70
Ap(AG)87	55.4	200-250	8	0.79	1.00	0.77	0.62
Ap(TC)88	55.4	200-300	15	0.93	0.37	0.90	0.99
Ap(AG)89	60.0	150-200	3	0.60	0.16	0.59	0.72
Ap(TC)90	55.4	200-250	3	0.56	0.21	0.54	0.70
Ap(CT)92	58.5	100-150	7	0.77	0.33	0.75	0.89
Ap(TC)93	55.4	200-250	7	0.71	0.26	0.69	0.83
Ap(AC)95	58.5	100-200	8	0.75	0.26	0.73	0.80
Ap(GA)98	55.8	150-250	12	0.88	0.26	0.86	0.93
Ap(TC)99	60.0	100-200	8	0.84	0.50	0.81	0.89
Ap(AGG)102	58.5	100-150	6	0.57	0.47	0.56	0.73
Ap(GAA)106	59.4	150-250	12	0.88	0.37	0.86	0.95
Ap(TG)107	58.5	200-250	2	0.51	0.26	0.49	0.67
Ap(TC)109	60.0	250-300	5	0.66	0.00	0.64	0.70
Ap(TC)110	60.0	100-200	10	0.88	0.32	0.86	0.93
Ap(GA)111	55.4	150-200	4	0.51	0.37	0.49	0.68
Ap(CT)112	55.4	150-200	3	0.46	0.63	0.45	0.53
Ap(CT)113	58.5	100-150	10	0.85	0.32	0.83	0.91
Ap(CT)114	58.5	200-250	12	0.91	0.16	0.89	0.94

Supplemental Table S2. Characterization of the 64 developed SSR markers, including the annealing temperature (Ta), observed size in base pair, number of alleles, expected (H_E) and observed (H_O) heterozygosity, polymorphism information content (*PIC*), and discriminatory power (*D*). (continued).

Ap(TCT)116	55.4	100-150	8	0.86	0.47	0.84	0.91
Ap(TTC)117	62.0	150-200	6	0.78	0.33	0.76	0.87
Ap(TCT)118	58.5	200-300	8	0.76	0.00	0.74	0.77
Ap(TCT)120	58.5	250-300	5	0.73	0.32	0.71	0.80
Ap(AG)121	58.5	100-150	9	0.77	0.53	0.75	0.90
Ap(AG)124	58.5	200-250	7	0.67	0.21	0.65	0.76
Ap(CT)126	58.5	200-250	8	0.81	0.37	0.79	0.88
Ap(AG)127	58.5	100-150	3	0.61	0.42	0.59	0.77
Ap(GT)128	59.8	150-250	5	0.67	0.32	0.66	0.75
Ap(TC)132	58.5	100-150	4	0.55	0.11	0.54	0.62
Ap(CT)134	58.5	250-300	3	0.67	0.37	0.65	0.81
Ap(AG)137	58.5	150-200	5	0.74	0.42	0.72	0.88
Mean	-	-	7.37	0.72	0.31	0.70	0.80
Total	-	-	464	-	-	-	-

Supplemental Table S3. Results of cross amplification in the 33 loci analyzed for the seven species of *Arachis* in 3% agarose gel. The presence of secondary products in each locus are reported.

	Section/Species							
	Rhizomatosae	*Procumbentes*	*Arachis*				*Caulorrhizae*	
Loci	*A. glabrata*	*A. valsii*	*A. stenosperma*	*A. duranensis*	*A. ipaensis*	*A. hypogaea*	*A. repens*	Sec. Prod.
Ap(CT)01	+	+	+	+	+	+	+	No
Ap(AG)14	+	+	+	+	+	+	+	No
Ap(CT)15	+	-	+	-	+	+	+	Yes
Ap(TC)18	+	+	+	+	+	+	+	No
Ap(TC)28	+	+	+	+	-	+	+	Yes
Ap(GA)29	+	+	-	+	+	+	+	No
Ap(TG)32	+	+	+	+	+	+	+	No
Ap(CT)35	+	+	+	+	+	+	+	No
Ap(GA)39	+	+	+	+	+	+	+	No
Ap(AT)40	+	-	+	-	+	+	+	No
Ap(GA)42	-	-	-	-	-	-	+	No
Ap(CT)54	+	+	+	+	+	+	+	No
Ap(CT)55	+	+	+	+	-	+	+	No
Ap(AG)58	+	-	-	+	-	+	+	No
Ap(CT)68	+	+	+	+	+	+	+	No
Ap(CT)72	-	-	-	-	-	-	+	No
Ap(AG)87	+	+	+	+	+	+	+	Yes
Ap(TC)88	+	+	+	+	+	+	+	No
Ap(CT)92	+	-	-	-	-	+	+	No
Ap(TC)93	+	+	+	+	+	+	+	Yes
Ap(AC)95	+	+	-	+	-	+	+	Yes
Ap(GA)98	+	+	+	+	+	+	+	No
Ap(TC)99	+	-	-	-	-	+	+	No
Ap(TC)109	+	+	+	+	+	+	+	No
Ap(TC)110	+	-	-	-	+	-	+	No
Ap(GA)111	+	+	-	+	-	+	+	No
Ap(CT)112	+	+	+	+	+	+	+	Yes
Ap(CT)113	+	+	+	+	+	+	+	No
Ap(CT)114	+	+	-	+	+	+	+	No
Ap(TCT)116	+	+	+	+	-	+	+	Yes
Ap(TCT)118	-	-	-	-	-	-	+	No

T (%) = transferability rate; (-) = absence of amplification; (+) = observed amplification.

Supplemental Table S3. Results of cross amplification in the 33 loci analyzed for the seven species of *Arachis* in 3% agarose gel. The presence of secondary products in each locus are reported. (continued).

Ap(AG)121	+	+	+	+	+	+	+	No
Ap(AG)124	-	-	-	-	-	-	+	No
T (%)	87.88	69.70	63.64	72.73	63.64	84.85	100	-

T (%) = transferability rate; (-) = absence of amplification; (+) = observed amplification.

CAPÍTULO IV

MOLECULAR IDENTIFICATION OF VARIETIES AND INTERSPECIFIC HYBRIDS IN FORAGE PEANUTS

Este manuscrito foi formatado para submissão à revista South American Journal of Basic Education, Technical and Technological

Molecular identification of interspecific hybrids in forage peanuts

Identificação molecular de híbridos interespecíficos em amendoim forrageiro

Jônatas Chagas de Oliveira[1], Luciélio Manoel da Silva[2] and Tatiana de Campos[2*]
[1]Universidade Federal do Acre (UFAC), Rio Branco, AC, Brazil.
[2]Empresa Brasileira de Pesquisa Agropecuária (Embrapa Acre), Rio Branco, AC, Brazil.
*Corresponding author: tatiana.campos@embrapa.br

ABSTRACT

Arachis pintoi (forage peanut) has stood out in its use as a green cover in intercropping with commercial crops and in pastures due to its ability to control weeds, maintain soil moisture, biological nitrogen fixation and recovery of degraded areas, besides its high nutritional content. Therefore, the objective of this study was to perform the molecular identification of interspecific hybrids of crosses between *A. pintoi* and *A. hypogaea*. A set of four microsatellites were tested. The Ap(CT)68 locus allowed the certification among 229 crosses. The alleles polymorphism was detected on agarose gel, allowing the reduction of costs and time required for the hybrids identification process. The two species belong to different botanical sections and have different ploidy. In face of the biological challenge, new breeding strategies are proposed and should be tested in the breeding program.

Keywords: molecular certification, genes introgression, *A. pintoi*, *A. hypogaea*.

RESUMO

O *Arachis pintoi* (amendoim forrageiro) vem se destacando no uso como cobertura verde em cultivo intercalado com culturas comerciais e em pastagens devido à sua capacidade de controle de plantas daninhas, manutenção da umidade do solo, fixação biológica de nitrogênio e recuperação de áreas degradadas, além do seu alto teor nutritivo. Portanto, o objetivo desse estudo foi realizar a identificação molecular de híbridos interespecíficos de cruzamentos entre *A. pintoi* e *A. hypogaea*. O loco Ap(CT)68 permitiu a certificação entre os 229 cruzamentos. O polimorfismo dos alelos foi detectado em gel de agarose, permitindo a redução de custos e tempo necessários para o processo de identificação dos híbridos. As duas espécies pertencem a secções botânicas distintas e apresentam ploidias diferentes. Diante do desafio biológico, novas estratégias de cruzamento são propostas e deverão ser testadas no programa de melhoramento.

Palavras-chave: certificação molecular, introgressão de genes, *A. pintoi*, *A. hypogaea*.

1. INTRODUCTION

Forage peanut (*Arachis pintoi* Krapov. and WC Greg.) has received prominence for its use as forage in mixed pastures with grasses has improved livestock production, generating an annual positive impact of approximately USD 17 million per year in the state of Acre, located in the southwest of the Brazilian Amazon [1]. In addition, its use as green manure in intercropping with *Coffea arabica* [2,3], *Cucumis sativus* [4], and *Solanum lycopersicum* [5], contributing to soil protection, moisture maintenance, weed control, biological nitrogen fixation and nutrient cycling. However, there are still few cultivars available to producers and the high cost of commercialized seeds has made large-scale application less attractive than other forages [6].
Unlike the fruits of peanut (*A. hypogaea*), forage peanut has no hard PEG (ovarian stretch connecting the embryo to the mother plant) and seed are detached from the adult plant [7]. For the development of new cultivars, the introgression of genes between wild species and peanuts has been the main tool used in breeding [8]. Thus, the introgression of genes related to the stiffness and thickness of the PEG from common peanuts to forage peanuts is a goal of the breeding program to facilitate the fruit harvesting process and reduce cost.
In order to register a new cultivar, the breeder has to prove the characteristics of distinction, homogeneity and stability (DHE). The descriptors used must be more striking morphological, physiological or molecular characteristics [9]. However, crossings can be made between genotypes that have no obvious divergent phenotypic traits, making it impossible to use morphological markers [10]. In these cases, the use of microsatellite markers (SSR) has been an important alternative, as they are multi-allelic, have co-dominant nature, are distributed throughout the genome, and present high polymorphism [11]. Microsatellites have been used in early identification of hybrids [10,12]. Therefore, the objective of this study was to perform the molecular identification interspecific hybrids of crosses between *A. pintoi* and *A. hypogaea*.

2. MATERIAL AND METHODS

2.1 Plant material e DNA extaction

Two crosses were tested: 18.2001 (*A. hypogaea*) x W34 (*A. pintoi*) and between cultivar BRS 421 (*A. hypogaea*) x W34. The genotype W34 was the pollen donor and genotypes 18.2001 and

BRS 421 were the receptors. The crosses were carried out in a greenhouse at Embrapa Acre, in Rio Branco, Acre, Brazil, from February to May 2019 [13]. The produced seeds were collected and planted in germination cells and, later, young leaflets were collected for DNA extraction. The genomic DNA of each progeny was extracted with CTAB [14] and quantification was performed on a *Qubit* fluorimeter (Thermo Scientific Inc., MA, USA).

2.2 PCR amplification and genotyping

Four markers were used for hybrids identification (Table 1). The PCR reactions were made with a volume of 13 µL containing the following: buffer solution 1X (Ludwig Biotecnologia LTDA, RS, Brazil); 0.25 mg mL^{-1} BSA (Bovine Serum Albumin); 0.25 mM of each dNTP; 2.0 mM $MgCl_2$; 0.8 µM of each primer; 1 U of Taq DNA polymerase (Ludwig Biotecnologia LTDA, RS, Brazil); 7.5 ng of genomic DNA and sterile ultrapure water.

Table 1. Characteristics and source of the four microsatellite loci used for hybrid identification.

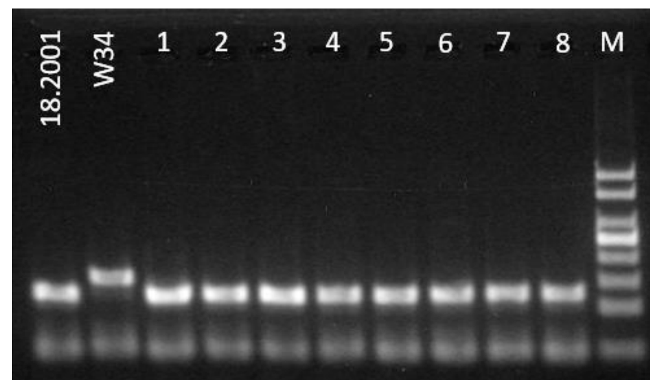
Loci	Primer sequence	Ta (°C)	Expected length (bp)	Source of primers
Ap(CT)68	F: ACATTGCAGGCACCCATCTT R: AGCACCATCACCAACAACCC	58.5	150-300	Oliveira et al., unplublished
Ap(GA)98	F: GTGGCGCTCATGACTGATGA R: GAGGCAGCAGAGTCTTCGTT	58.5	150-250	Oliveira et al., unplublished
Ap(CT)113	F: GTGTCTCATTCTCCCGCTCC R: CGGTTTTCTTGGTTTGTGTGGT	58.5	100-150	Oliveira et al., unplublished
Ap(AG)121	F: AAGCTCTTCTCTCCTTCCCA R: ACCACCATTTTGCTCGTCCA	58.5	100-150	Oliveira et al., unplublished

Amplifications were performed in a thermocycler (Analitik Jena) according to the following steps: 94°C for five minutes, followed by 30 cycles at 94°C for one minute, annealing temperature set for each primer for one minute and 72°C extension for one minute, followed by a final extension step at 72°C for five minutes. The PCR products were analyzed and genotyped on agarose gel electrophoresis 3%.

3. RESULTS AND DISCUSSION

One hundred seventy-seven seeds were obtained from the cross between 18,2001 x W34 and 52 between the cultivar BRS 421 x W34, totaling 229 samples. The germination period lasted, on average, 15 days, after which young leaves were observed in the size and quantity necessary for DNA extraction.

94 All analyzed markers showed amplification products and allowed differentiation in agarose gel
95 3% of the genotypes used in the crosses, which was not possible with the markers previously
96 used in hybrids identification [14]. This result is interesting, as it reduces costs and makes the
97 hybrid identification process faster and less laborious.
98 The Ap(CT)68 marker showed the best profile among the analyzed markers, allowing the
99 successful determination of the progeny status, whether hybrids or not. There was no positive
100 identification of hybrids in the analyzed progenies, which were identified as result of selfing
101 (Figure 2). This result corroborates with the literature, since the hybridization between *A. pintoi*
102 and *A. hypogaea* has not yet been achieved [15]. The great diversity of the genus *Arachis* has
103 made introgression a challenge because there are nine taxonomic sections and gene transfer of
104 these sections for *A. hypogaea* is complex [8]. Hybridization studies between *A. hypogaea* and
105 other wild species of the *Arachis* genus have been carried out, mainly, for introgression of
106 characteristics of interest such as resistance to abiotic stress and diseases, which interfere in
107 peanut productivity [8,12]. However, the success rate at these crosses has been low, ranging
108 from 0 to 18%, according to the genotypes used.
109



110
111 **Figure 2**. Visualization of PCR products of Ap(CT)68 marker on agarose gel 3%. The genotype
112 18.2001 and W34 are the female and male genitors, respectively. The numbers 1 to 8 are
113 putative hybrids. M: 50 bp latter marker.
114
115 In this study, a probable factor for the difficulty of crossing between *A. hypogaea* and *A. pintoi*
116 may be a difference in ploidy (tetraploid (4X) and diploid (2X), respectively) between the
117 genotypes. Therefore, other strategies can be addressed in future works, such as chromosomal
118 duplication of *A. pintoi* through treatment with colchicine [12]. In addition, it is possible to use
119 the crosses between *A. pintoi* and another diploid species in the section *Arachis*, which can act
120 as a bridge species in crossings with *A. hypogaea*.

121

122 **CONCLUSION**

123 The microsatellite markers used were able to carry out the identification of possible hybrids 124 analyzed. A marker (Ap(CT)68) was presented that allowed the certification of the absence of 125 hybrids in the crosses performed. There are biological barriers that may have interfered with 126 hybridization, however, it is important that new approaches be tested in future works in the face 127 of the promising scientific advance.

128

129 **REFERENCES**

130 [1] EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. **Balanço Social 2019**. 131 https://bs.sede.embrapa.br/2019/balancosocialembrapa2019print.pdf., 2020. Accessed on: 01 132 sep. 2020.

133 [2] SANTOS, J.C.F.; CUNHA, A.J.; MELO, B. Soil cover and weed control on coffee 134 intercropping perennial legume. **International Journal of Applied Science and Technology**, v.4, 135 n.4, p.149-157, 2014.

136 [3] ROSE, T.J.; KEARNEY, L.J.; MORRIS, S.; VAN ZWIETEN, L.; ERLER, D.V. Pinto 137 peanut cover crop nitrogen contributions and potential to mitigate nitrous oxide emissions in 138 subtropical coffee plantations. **Science of the Total Environment**, v.656, p. 108-117, 2019.

139 [4] SILVA, G.P.P.; RESENDE, F.V.; PEREIRA, T.S.; SOUZA, R.B.; ALBUQUERQUE, J.O.; 140 VIDAL, M.C.; SOUSA, J.M.M. **Desempenho agronômico de cultivares de pepino em 141 ambiente protegido cultivado em solo com cobertura viva de amendoim forrageiro em 142 sistema orgânico de produção**. 143 https://www.alice.cnptia.embrapa.br/bitstream/doc/930040/1/IIJCEHGeanny.pdf., 2012. 144 Accessed on: 08 oct. 2020.

145 [5] RESENDE, F.V.; VIDAL, M.C.; SOUZA, R.B. **Plantio direto de tomate sobre 146 coberturas vivas em sistema orgânico de produção**. Distrito Federal: Embrapa Hortaliças, 147 2020.

148 [6] VALENTIM, J.F.; ASSIS, G.M.L.; SÁ, C.P. Produção de sementes de amendoim forrageiro 149 (*Arachis pintoi*) no acre. **Amazônia: Ciência & Desenvolvimento**, v. 4, n. 8, p. 191-193, 2009.

150 [7] LIMA, J.A. de; PINTO, J.C.; EVANGELISTA, A.R.; SANTANA, R.A.V. **Amendoim 151 forrageiro (*Arachis pintoi* Krav. & Gregory)**. Lavras: Editora UFLA, 2003, p. 11-19.

152 [8] CASON, J.M.; SIMSPON, C.E.; ROONEY, W.L.; BRADY, J.A. First report of successful 153 hybridization between *Arachis vallsii* and *Arachis dardani*. **Agrosystems, Geosciences & 154 Environment**, v.2, n.1, p.1-3, 2019. https://doi.org/10.2134/age2019.03.0013.

155 [9] BRASIL. MINISTÉRIO DA AGRICULTURA, PECUÁRIA E ABASTECIMENTO.
156 **Proteção de Cultivares no Brasil**. Brasília: Mapa/ACS, 2011.
157 [10] CHU, Y.; WU, C.L.; HOLBROOK, C.C.; OZIAS-AKINS, P. Conditions that impact
158 artificial hybridization of *Arachis hypogaea* L. **Peanut Science**, v.43, n.2, p.106-115, 2016.
159 [11] HODEL, R.G.J.; SEGOVIA-SALCEDO, M.C.; LANDIS, J.B.; CROWL, A.A.; SUN, M.;
160 LIU, X.; GITZENDANNER, M.A.; DOUGLAS, N.A.; GERMAIN-AUBREY, C.C.; CHEN,
161 S.; SOLTIS, D.E.; SOLTIS, P.S. The report of my death was an exaggeration: a review for
162 researchers using microsatellites in the 21st century. **Applications in Plant Sciences**, v.4, n.6,
163 p.1600025, 2016. https://doi.org/10.3732/apps.1600025.
164 [12] FAVERO, A.P.; DOS SANTOS, R.F.; SIMPSON, C.E.; VALLS, J.F.; VELLO, N.A.
165 Successful crosses between fungal-resistant wild species of *Arachis* (section *Arachis*) and
166 *Arachis hypogaea*. **Genetics and Molecular Biology**, v.38, n.3, p.353-365, 2015.
167 https://doi.org/10.1590/S1415-475738320140376.
168 [13] NIGAM, S.N.; RAO, M.J.V.; GIBBONS, R.W. **Artificial hybridization in groundnut**.
169 Patancheru: International Crops Research Institute for the Semi-Arid Tropics, 1990. 27p.
170 [14] OLIVEIRA, J.C.; CAMPOS, T. Tecnologias em genética molecular para intensificação do
171 uso de amendoim forrageiro em pastagens na Amazônia, In: MENEGUETTI, D.U.O;
172 CARVALHO, C.M.; ZAN, R.A.; SILVA, R.P.M. (Orgs.). **Ciência, Inovação e Tecnologia na**
173 **Amazônia**. Rio Branco: Stricto Sensu, 2019, pp. 105-119. Available at:
174 https://sseditora.com.br/wp-content/uploads/Ci%C3%AAncia-Inova%C3%A7%C3%A3o-e-
175 Tecnologia-na-Amaz%C3%B4nia.pdf. Accessed on: 01 dec. 2019.
176 [15] GREGORY, M.P.; GREGORY, W.C. Exotic germ plasm of *Arachis* L. interspecific
177 hybrids. **Journal of Heredity**, v.70, p.185-193, 1979.
178

9. CONCLUSÕES

A preocupação com o desenvolvimento sustentável tem crescido cada vez mais, especialmente por consumidores do mercado externo. Dessa forma o uso de técnicas e soluções que associem o aumento da produtividade com a redução da degradação ambiental é altamente desejável. Nesse sentido, o uso de leguminosas forrageiras como cobertura vegetal do solo tanto em consórcios com culturas comerciais como para recuperação de áreas degradadas tem sido recomendado. Dentre as leguminosas, o amendoim forrageiro tem apresentado resultados importantes em consórcios com café, nectarina, pepino e tomate por contribuir na manutenção da umidade do solo, redução da erosão, ciclagem de nutrientes, fixação biológica de nitrogênio e baixas emissões de óxido nitroso. Além disso, seu uso em pastagens consorciadas tem contribuído tanto na melhoria da fertilidade dos solos de pastagens e redução da compactação como no aumento da produção de leite, ganho de peso no gado de corte e redução no tempo de abate em até nove meses. Com o objetivo de tornar o uso do amendoim forrageiro mais atrativo aos produtores, pesquisadores têm buscado desenvolver novas cultivares que possam atender as diversas necessidades. Estudos moleculares têm fornecido informações importantes sobre a biologia do amendoim forrageiro, o que tem contribuído no avanço do conhecimento da espécie para uso no melhoramento.

Neste trabalho foram apresentados resultados de extração de RNA, sequenciamento e montagem *de novo* do transcriptoma de *A. pintoi*, o qual permitiu o desenvolvimento dos primeiros marcadores SNP e novos marcadores SSR para a espécie. Os marcadores SSR validados apresentaram elevado polimorfismo e taxa variável de transferibilidade. Foi possível estabelecer o menor número necessário para identificação de variedades e híbridos interespecíficos de amendoim forrageiro. O conjunto desses resultados permitirão a aplicação de técnicas moleculares modernas como associação genômica, expressão diferencial e seleção genômica, os quais contribuirão significativamente no conhecimento da biologia de *A. pintoi* e avanço do programa de melhoramento do amendoim.