



UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
DEPARTAMENTO DE AGRONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM AGRONOMIA -
MELHORAMENTO GENÉTICO DE PLANTAS

FLÁVIA GOMES DA SILVA

FILOGENIA MOLECULAR E ANÁLISE *IN SILICO* DO INIBIDOR DE
PROTEASE BOWMAN-BIRK DE ESPÉCIES DA SUBFAMÍLIA
PAPILIONOIDEAE

RECIFE – PE
JANEIRO - 2023

FLÁVIA GOMES DA SILVA

**Filogenia molecular e análise *in silico* do inibidor de protease
Bowman-Birk de espécies da subfamília Papilionoideae**

Tese apresentada ao Programa de Pós- Graduação em Agronomia - Melhoramento Genético de Plantas (PPGAMGP), da Universidade Federal Rural de Pernambuco, como parte dos requisitos para obtenção do título de Doutora em Agronomia - Melhoramento Genético de Plantas.

Orientadora: Profa. DSc. Gheysa
Coelho Silva

Coorientador: DSc. Rômulo Maciel
de Moraes Filho

**RECIFE – PE
JANEIRO - 2023**

Dados Internacionais de Catalogação na Publicação
Universidade Federal Rural de Pernambuco
Sistema Integrado de Bibliotecas
Gerada automaticamente, mediante os dados fornecidos pelo(a) autor(a)

S586f

Silva, Flávia Gomes da
Filogenia molecular e análise in silico do inibidor de protease Bowman-Birk de espécies da subfamília
Papilionoideae / Flávia Gomes da Silva. - 2023.
114 f. : il.

Orientadora: Gheysa Coelho Silva.
Coorientador: Romulo Maciel de Moraes Filho.
Inclui referências e anexo(s).

Tese (Doutorado) - Universidade Federal Rural de Pernambuco, Programa de Pós-Graduação em
Agronomia - Melhoramento Genético de Plantas, Recife, 2023.

1. Leguminosae. 2. Bioinformática. 3. Análise filogenética. 4. Tempo de divergência, . 5. Modelagem por
homologia. I. Silva, Gheysa Coelho, orient. II. Filho, Romulo Maciel de Moraes, coorient. III. Título

CDD 581.15

FLÁVIA GOMES DA SILVA**Filogenia molecular e análise *in silico* do inibidor de protease Bowman-Birk de espécies da subfamília Papilionoideae**

Tese defendida e aprovada pela banca examinadora em: 31/01/2023

ORIENTADORA

Profa. DSc. Gheysa Coelho Silva
Área de Genética / Departamento de Biologia-UFRPE

EXAMINADORES

Prof. DSc. Horace José Jimenez (Membro Titular Externo)
Secretaria de Educação do Estado de Pernambuco

Prof. DSc. Pierre Teodosio Felix (Membro Titular Externo)
Centro Universitário da Vitória de Santo Antão - UNIVISA

Prof. DSc. Rafael Trindade Maia (Membro Titular Externo)
Universidade Federal de Campina Grande - UFCG

Prof. DSc. Reginaldo de Carvalho (Membro Titular Interno)
Universidade Federal Rural de Pernambuco - UFRPE

RECIFE – PE
JANEIRO - 2023

Hey Pai

Eu sei que nunca me deixou
Eu sei que sempre esteve aqui comigo
Livrando o meu coração do que não é Seu
E eu sei que, todas as noites, Tu vens
Pra acalmar a minha mente
E pondo no meu coração que eu sou Teu

Hey Pai, olhando pra Você, me lembro bem
Das noites que passei e que lutei aqui
Pensando estar sozinho, sem ninguém por perto
E eu, mesmo tão sozinho, não desanimei
Pois sabia que cuidarias de mim
Mesmo sem eu merecer, me daria Sua mão

Eu sei que nunca me deixou
Eu sei que sempre esteve aqui comigo
Livrando o meu coração do que não é Seu
E eu sei que, todas as noites, Tu vens
Pra acalmar a minha mente
E pondo no meu coração que eu sou Teu

Eu sei que nunca me deixou
Eu sei que sempre esteve aqui comigo
Livrando o meu coração do que não é Seu
E eu sei que, todas as noites, Tu vens
Pra acalmar a minha mente
E pondo no meu coração que eu sou Teu
E Tu és meu, Tu és meu.

ISADORA POMPEO

OFEREÇO

À Deus, pela sua infinita misericórdia e proteção.

DEDICO

À minha família, sem vocês nada disso seria possível.

AGRADECIMENTOS

À Deus por estar sempre presente em todos os momentos da minha vida, me concedendo saúde, abençoando em todas as decisões, me fortalecendo em cada obstáculo superado.

Ao meus pais por estarem sempre ao meu lado, confiando, torcendo, incentivando e comemorando a cada objetivo alcançado.

Às minhas irmãs, por me apoiarem e pelos momentos de descontração.

À minha querida Marília Damasceno pela compreensão, incentivo e amor.

Ao pinguinho de gente chamada Mallu, que me fez conhecer o significado do amor incondicional, que tornou os meus meus dias mais alegres e me motiva a correr atrás dos meus objetivos.

À minha orientadora, Gheysa Coelho Silva, pela confiança, ensinamentos e apoio em todas as situações.

Ao meu coorientador, Rômulo Maciel Moraes Filho pelo ensinamentos, empenho e dedicação durante as análises e obtenção dos resultados da tese.

Ao colega, Horace José Jimenez, pela colaboração nas análises e contribuição na redação do segundo artigo da tese.

Aos professores da Universidade Federal Rural de Pernambuco (UFRPE), pelos conhecimentos compartilhados, que auxiliou no desenvolvimento desta tese.

À colega de doutorado, Ana Kelly pelos momentos de descontração e companheirismo nesta jornada.

À UFRPE, pela oportunidade de qualificação profissional no doutorado de Agronomia (Melhoramento Genético de Plantas).

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa de estudo ao longo dos quatro anos de doutorado.

E por fim, a todos que de forma direta ou indireta contribuíram para conclusão desta pesquisa.

A todos vocês, o meu muito obrigada!

LISTA DE FIGURAS

CAPÍTULO I

Figura 1. Flor papilionácea (*Centrosema platycarpum* Benth).....19

CAPÍTULO II

Figura 1. Alinhamento de sequências completas de plastomas de 28 espécies da subfamília Papilionoideae.....76

Figura 2. Árvore filogenética baseada em quatro marcadores plastidiais (*matK*, *psaA*, *psbA*, *psbD* *rbcL* e *rpoC2*) e tempo de divergência entre espécies da subfamília Papilionoideae, utilizando seis pontos de calibração.....77

CAPÍTULO III

Figura 1. Alinhamento de sequências do inibidor de protease Bowman-Birk avaliados.....84

Figura 2. Previsão dos efeitos funcionais das mutações em aminoácidos pelo servidor SNAP2 e Domínios funcionais do inibidor de protease Bowman-Birk, observados pelo servidor ProDom.....85

Figura 3. Árvore filogenética das espécies da tribo Phaseoleae (Fabaceae - Papilionoideae).....86

Figura 4. Estrutura tridimensional predita pelo servidor GalaxyTBM para o inibidor Bowman-Birk das espécies *P. Vulgaris* e *C. cajan*. Superfície eletrostática e Hidrofobicidade representada como um gradiente de cor.....88

Figura 5. Gráfico de Ramachandran para BBI de *P. vulgaris* e *C. cajan*. Valor do Z-score pelo ProSA-web para *P. vulgaris* e *C. cajan*. Gráfico de energia de escores de resíduos da estrutura de proteína nativa para *P. vulgaris* e *C. cajan*.....89

LISTA DE TABELA

CAPÍTULO II

Tabela 1. Características das sequências do genoma plastidial de espécies da subfamília <i>Papilionoideae</i> e <i>outgroup</i> utilizadas na análise.....	72
---	----

CAPÍTULO III

Tabela 1. Análise da estrutura primária e localização subcelular das sequências do inibidor Bowman-Birk de espécies da tribo <i>Phaseoleae</i> (Fabaceae – <i>Papilionoideae</i>).....	83
--	----

SUMÁRIO

RESUMO.....	xi
ABSTRACT.....	xii

CAPÍTULO I: REVISÃO DE LITERATURA

1. Introdução Geral.....	16
2. Revisão de Literatura	18
2.1. Classificação, Distribuição Geográfica e Aspectos Gerais	18
2.2. Importância ecológica e econômica	20
2.3. Melhoramento Genético	23
2.4. Bioinformática.....	25
2.5. Genoma Plastidial	27
2.6. Filogenia Molecular	31
2.7. Sistemática Filogenética	34
2.8. Inibidor de Protease do Tipo Bowman-Birk	37
3. Referências	40

CAPÍTULO II: Filogenia com base em marcadores plastidiais revela *insights* sobre as relações de espécies de Papilionoideae

1. Resumo.....	59
2. Abstract	59
3. Introdução.....	60
4. Material e Métodos	60
5. Resultados.....	61
6. Discussão.....	63
7. Conclusões	67
8. Financiamento	67
9. Referências.....	67

CAPÍTULO III: Caracterização *in silico* e análise filogenética do inibidor de protease do tipo Bowman-Birk em espécies da tribo Phaseoleae (Fabaceae - Papilionoideae)

1. Resumo.....	80
2. Abstract	81
3. Introdução.....	81
4. Materiais e Métodos.....	82
5. Resultados e Discussão.....	83
6. Conclusões.....	90
7. Financiamento	90
8. Referências.....	90

ANEXOS

ANEXO I – NORMAS DE REDAÇÃO DE DISSERTAÇÃO OU TESE.....	96
ANEXO II – INSTRUÇÕES AOS AUTORES PARA PUBLICAÇÃO NA REVISTA CROP BREEDING AND APPLIED BIOTECHNOLOGY (CBAB).....	98
ANEXO III – INSTRUÇÕES AOS AUTORES PARA PUBLICAÇÃO NA REVISTA GENETIC RESOURCES AND CROP EVOLUTION.....	101
ANEXO IV – INSTRUÇÕES AOS AUTORES PARA PUBLICAÇÃO NA REVISTA BRAZILIAN ARCHIVES OF BIOLOGY AND TECHNOLOGY (BAPT).....	108

RESUMO

Fabaceae (Leguminosae) é a terceira maior família de dicotiledôneas. Subdividida em seis subfamílias, constitui a terceira maior família de angiospermas, ficando atrás de Asteraceae e Orchidaceae. Dentre essas subfamílias Papilionoideae se destaca por representar dois terços de todos os gêneros e espécies da família, com aproximadamente 14.000 espécies, 501 gêneros e 32 tribos. Essa subfamília é reconhecida por sua grande importância alimentar, forrageira e biológica, com ampla distribuição nos diversos tipos de biomas. Nesta subfamília também é relatado na literatura a existência do inibidor de protease do tipo Bowman-Birk (BBI) em diversos representantes de Papilionoideae, sendo reconhecido por sua ação inseticida, dentre outras funções importantes nas plantas. Com o surgimento de novas tecnologias de sequenciamento de DNA, bem como os avanços das ferramentas de bioinformática, favoreceu a utilização de inúmeras formas de estabelecer correlações evolutivas entre as espécies desta subfamília. Atualmente, em razão da baixa taxa mutacional dos genomas de cloroplasto (cpDNA), as análises filogenéticas baseadas em cpDNA, tem resolvido as relações entre os principais táxons de Papilionoideae. Nesse sentido, os transcritos de proteínas importantes também podem ser ferramentas para estabelecer correlações entre espécies, uma vez que partem do mesmo princípio de baixa taxa de mutação, bem como identificar, caracterizar e prever domínios funcionais de proteínas por meio da abordagem *in silico*. Com o objetivo de realizar uma reconstrução filogenética de espécies da subfamília Papilionoideae baseada em seis genes plastidiais, foram obtidas sequências completas dos plastomas de espécies da subfamília Papilionoideae e dos genes isolados do banco de dados GenBank. O alinhamento das sequências foram realizadas e a reconstrução filogenética foram realizadas com uso do *software* MEGA 7. Para reconstrução filogenética foi utilizado o método de Máxima Verossimilhança com suporte estatístico calculado pelo método de *bootstrap* (BS), com 1000 replicatas. As correlações filogenéticas para subfamília Papilionoideae resultaram em uma árvore monofilética, confirmando a divisão da subfamília em quatro principais clados, NPAAA, ADA, Genistoides e Dalbergioides. O clado ADA foi relacionado como grupo irmão do clado Genistoides, com alto suporte. Foi evidenciado o parafiletismo das tribos Phaseoleae e Millettiae, dentro do clado NPAAA. Também observamos, que apenas as espécies do clado ADA não possuem nódulos rizobianos, o que pode configurar uma possível sinapomorfia para sustentar as relações desse grupo. Da mesma forma, a análise do tempo de divergência sugere

que os principais clados de Papilionoideae divergiram a partir do Paleoceno. As sequências da BBI foram recuperadas do banco de dados GenBank, por meio da ferramenta BlastP. Os parâmetros físico-químicos foram avaliados através da ferramenta ProtParam. A identificação de domínios funcionais foi avaliada por meio do servidor Prodom e a estimativa de efeitos mutacionais foi feita por meio do servidor SNAP2. O alinhamento de sequência das proteínas foi realizado por meio do algoritmo ClustalW e a árvore filogenética produzida no *software* MEGA 7, pelo método de Máxima Verossimilhança. O servidor GalaxyTBM foi utilizado para predição de estrutura das proteínas em modelos 3D. Os resultados obtidos entre as sequências de BBI revelaram dois domínios funcionais, caráter hidrofílico e ponto isoelétrico ligeiramente ácido para todas as espécies analisadas. A análise de agrupamento por Máxima Verossimilhança das espécies da tribo Phaseolea (Papilionoideae) construídas por meio das sequências da proteína BBI, apresentam consistência com a classificação mais recentes para as espécies analisadas. A partir do gráfico de Ramachandran, foi possível verificar que os modelos 3D gerados, evidenciaram excelente qualidade estereoquímica e boa qualidade estrutural. Estas informações em conjunto, podem auxiliar os programas de melhoramento genético dessas culturas, por meio de métodos convencionais ou engenharia genética.

Palavras-chave: Leguminosae, bioinformática, análise filogenética, tempo de divergência, modelagem por homologia.

ABSTRACT

Fabaceae (Leguminosae) is the third largest family of dicots. Subdivided into six subfamilies, it constitutes the third largest family of angiosperms, after Asteraceae and Orchidaceae. Among these subfamilies Papilionoideae stands out for representing two thirds of all genera and species in the family, with approximately 14,000 species, 501 genera and 32 tribes. This subfamily is recognized for its great food, forage and biological importance, with a wide distribution in different types of biomes. In this subfamily, the existence of the Bowman-Birk-type protease inhibitor (BBI) in several representatives of Papilionoideae is also reported in the literature, being recognized for its insecticidal action, among other important functions in plants. The emergence of new DNA sequencing technologies, as well as advances in bioinformatics tools, favored the use of numerous ways to establish evolutionary correlations between the species of this subfamily. Currently, due to the low mutational rate of chloroplast

genomes (cpDNA), cpDNA-based phylogenetic analyzes have resolved the relationships between the main taxa of Papilionoideae. In this sense, transcripts of important proteins can also be tools to establish correlations between species, since they depart from the same principle of low mutation rate, as well as to identify, characterize and predict functional domains of proteins through the *in silico* approach. In order to carry out a phylogenetic reconstruction of species of the subfamily Papilionoideae based on six plastid genes, complete sequences of the plastomes of species of the subfamily Papilionoideae and of the genes isolated from the GenBank database were obtained. Sequence alignment was performed and phylogenetic reconstruction was performed using the MEGA 7 software. For phylogenetic reconstruction, the Maximum Likelihood method was used with statistical support calculated by the bootstrap method (BS), with 1000 replicates. Phylogenetic correlations for the subfamily Papilionoideae resulted in a monophyletic tree, confirming the division of the subfamily into four main clades, NPAAA, ADA, Genistoides and Dalbergioides. The ADA clade was related as a sister group to the Genistoides clade, with high support. The paraphyly of the Phaseoleae and Millettiae tribes within the NPAAA clade was evidenced. We also observed that only the species of the ADA clade do not have rhizobian nodules, which may configure a possible synapomorphy to support the relationships of this group. Likewise, analysis of the time of divergence suggests that the main clades of Papilionoideae diverged from the Paleocene onwards. The BBI sequences were retrieved from the GenBank database using the BlastP tool. The physical-chemical parameters were evaluated using the ProtParam tool. The identification of functional domains was evaluated using the Prodom server and the estimation of mutational effects was performed using the SNAP2 server. The protein sequence alignment was performed using the ClustalW algorithm and the phylogenetic tree produced in the MEGA 7 software, using the Maximum Likelihood method. The GalaxyTBM server was used for protein structure prediction in 3D models. The results obtained between the BBI sequences revealed two functional domains, hydrophilic character and slightly acidic isoelectric point for all analyzed species. Clustering analysis by Maximum Likelihood of the species of the Phaseolea tribe (Papilionoideae) constructed using BBI protein sequences is consistent with the most recent classification for the analyzed species. From the Ramachandran graph, it was possible to verify that the generated 3D models showed excellent stereochemical quality and good structural quality. Together, this information can help genetic improvement programs for these crops, through conventional

methods or genetic engineering.

Keywords: Leguminosae, bioinformatics, phylogenetic analysis, divergence time, homology modeling.

REVISÃO DE LITERATURA
CAPÍTULO I

1. INTRODUÇÃO GERAL

A família Fabaceae também conhecida como Leguminosae, é a terceira maior família botânica em diversidade de espécies com flores, com aproximadamente 770 gêneros, mais de 19.500 espécies distribuídas em seis subfamílias, que podem ser encontradas em todos os biomas do mundo, mesmo em ambientes adversos (Lewis et al. 2005, LPWG, 2013, 2017).

Dentre as seis subfamílias de Leguminosae, Papilionoideae é a maior subfamília com aproximadamente dois terços de todos os gêneros e espécies da família, compreendendo cerca de 501 gêneros e 14.000 espécies (LPWG 2021), agrupadas em 32 tribos (Polhill e Raven 1981). Suas espécies também são bem conhecidas por sua grande importância alimentar, econômica e biológica incluindo, soja (*Glycine max* L.), feijão comum (*Phaseolus vulgaris* L.), feijão caupi (*Vigna unguiculata* L.), feijão fava (*P. lunatus* L.), fava (*Vicia faba* L.), amendoim (*Arachis hypogaea* L.), ervilha (*Pisum sativum* L.), dentre outras. Esta subfamília é bem suportada nos trabalhos filogenéticos, sendo considerada monofilética (Tozzi 2016).

Essa monofilia foi amplamente elucidada em análises filogenéticas moleculares em toda a subfamília (Käss e Wink 1995, Doyle et al. 1997, Kajita et al. 2001, Wojciechowski et al. 2004, Lavin et al. 2005, McMahon e Sanderson 2006, Cardoso et al. 2012, Cardoso et al. 2013, LPWG 2013, Koenen et al. 2020), mas ainda existem lacunas a serem preenchidas na filogenia e evolução dessa subfamília, como por exemplo, espécies aninhadas fora de sua tribo de origem, ou clados mal suportados. Essas incongruências poderão ser melhor esclarecidas pela adição de um maior número de *loci*, haja vista que a maioria dos estudos são baseados apenas em um único gene (Cardoso et al. 2013).

Os estudos filogenéticos são fundamentais para a sistemática e evolução, fornecendo conhecimento sobre os padrões de diversificação, bem como a origem das características desejáveis das espécies (Machado et al. 2017).

Nesse sentido, os genomas de cloroplastos são marcadores úteis para realizar análises comparativas associadas a estudos filogenéticos e evolutivos. Tendo como as principais características tamanho, sua natureza conservada e herança materna (Soltis e Soltis 2020). Por conseguinte, sequências de genes de cloroplasto são adequadas para estudos filogenéticos, genética de populações, filogeografia e identificação de espécies (Santos e Almeida 2019, Song et al. 2019, Yang et al. 2019).

Os cloroplastos também são responsáveis por várias vias metabólicas essenciais para planta, possuindo genes essenciais para a fotossíntese, biossíntese de ácidos graxos, importação de proteínas plastidiais e proteólise (Xu *et al.* 2015). Dessa forma, análises comparativas do genoma completo de cloroplasto de espécies contrastantes também são úteis para um melhor entendimento da evolução estrutural dos genomas (Yang *et al.* 2019).

O surgimento de tecnologias de Sequenciamento de DNA de Nova Geração e os avanços das ferramentas de bioinformática, vem propiciando um aumento significativo no número de estudos de genomas de cloroplasto (Mardis 2008, Metzker 2010, Liu *et al.* 2019, Jimenez *et al.* 2019, Khan *et al.* 2019, Jimenez *et al.* 2020, Koenen *et al.* 2020, Zha *et al.* 2020). Para tanto, diversas metodologias e programas para análise *in silico* têm sido empregadas com sucesso na identificação, caracterização e comparação de regiões do genoma do cloroplasto para análises filogenéticas de diversos grupos de plantas, refletindo com precisão a história evolutiva das espécies estudadas (Ouzounis 2002, Hand *et al.* 2013, LPWG 2013, Morton e Telmer 2014, Song *et al.* 2015, Erguiluz *et al.* 2017, LPWG 2017, Vu *et al.* 2018). A partir de ferramentas da bioinformática também é possível identificar diferenças entre os aminoácidos e suas estruturas, por meio do alinhamento de estruturas de proteínas (Diniz e Canduri 2017), possibilitando compreender o papel das proteínas dentro de um grupo de espécies (Kumar *et al.* 2004).

Assim, estudos filogenéticos evolutivos também podem ser executados por meio de dados de sequências proteicas. Uma proteína de grande importância em espécies de leguminosas é o inibidor de protease Bowman-Birk (BBI), caracterizado por apresentarem dois domínios, que são responsáveis pela inibição autônoma e simultânea de duas serinoproteases distintas (Hellinger e Gruber, 2019), onde o primeiro sítio inibi a tripsina e o outro sítio inibi a quimotripsina (Lioi *et al.* 2010).

Na literatura são evidenciadas diversas funções defensivas dos BBIs nas plantas, sendo a ação inseticida a mais relatada em insetos de diversas culturas (Oppert *et al.* 1993, Franca *et al.* 2003, Rahbé *et al.* 2003, Aguirre *et al.* 2004, Azzouz *et al.* 2005, Pereira *et al.* 2007, Kuhar *et al.* 2013, Dantzger *et al.* 2015). Alguns BBIs também são associados ao efeito antimicrobianos em patógenos (Kim *et al.* 2009), ou até mesmo no desenvolvimento de tolerância à estresses abióticos, como salino (Shan *et al.* 2008) e hídrico (Dramé *et al.* 2013). Dessa forma, essas funções possibilitam a produção de plantas transgênicas resistentes a insetos (Clemente *et al.* 2019), patógenos e determinados estresses abióticos, o que

representa um potencial ganho econômico na produção agrícola.

Nesse sentido, o presente estudo teve como objetivo esclarecer a filogenia da subfamília Papilionoideae e estimar o tempo de divergência dentro da subfamília, por meio de ferramentas computacionais, utilizando como base marcadores plastidiais. O estudo também objetivou caracterizar, comparar e identificar domínios conservados em sequências de aminoácidos do BBI de espécies de Papilionoideae disponíveis em bancos de dados públicos; desenvolver modelos tridimensionais do BBI de representantes da subfamília Papilionoideae, com base na metodologia de modelagem por homologia, para uma melhor compreensão de suas estruturas e funções moleculares.

2. REVISÃO DE LITERATURA

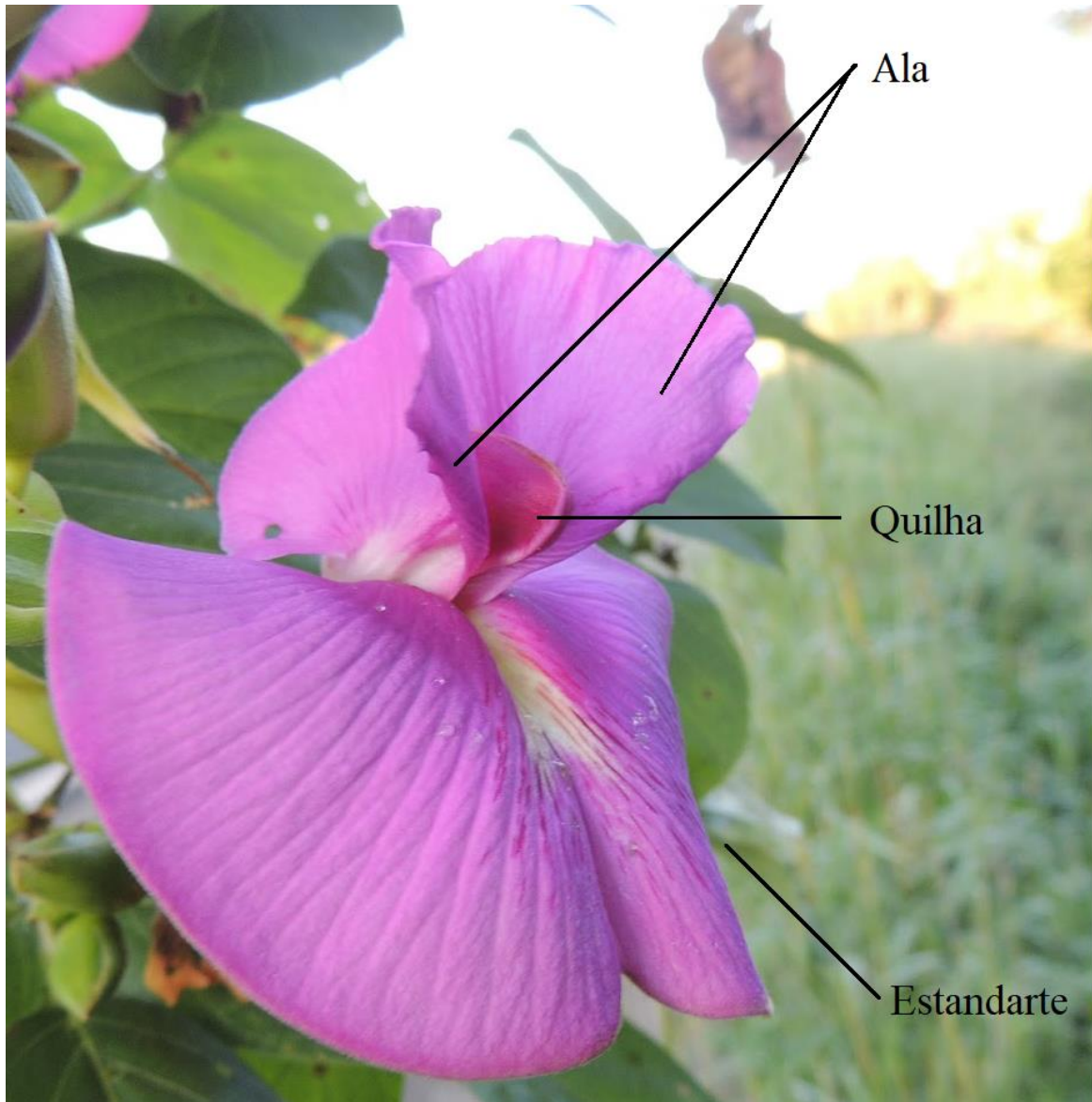
2.1. CLASSIFICAÇÃO, DISTRIBUIÇÃO GEOGRÁFICA E ASPECTOS GERAIS

Na última publicação do *Angiosperm Phylogeny Group IV* (APG IV 2016), a ordem Fabales é classificada como pertencente ao clado Fabidae, classe Eudicotiledôneas e a subclasse Rosidae. Essa ordem possui quatro famílias, sendo elas: Leguminosae (Fabaceae), Quillajaceae, Polygalaceae e Surianaceae. Dentre essas famílias, Leguminosae é uma das linhagens de plantas com flores de maior sucesso evolutivo (LPWG 2013). São a terceira maior família de angiospermas com maior diversidade de espécie, ficando apenas atrás de Asteraceae e Orchidaceae (LPWG 2017).

Leguminosae conta com aproximadamente 770 gêneros e mais de 19.500 espécies distribuídas em seis subfamílias, Cercidoideae, Detarioideae, Duparquetioideae, Dialioideae, Caesalpinioideae e Papilionoideae. Dessas subfamílias Papilionoideae é a maior, com aproximadamente dois terços de todos os gêneros e espécies da família Leguminosae, compreendendo cerca de 501 gêneros e 14.000 espécies (LPWG 2021), distribuídas nas 32 tribos reconhecidas (Polhill e Raven 1981). É uma subfamília homogênea, sendo mais difundida, comumente com distribuição global, estendendo-se ainda mais em regiões de clima temperado e subtropicais quando comparado com as demais subfamílias (Gurcharan 2004, Duane e Paul 2012).

As espécies da subfamília Papilionoideae podem ser anuais, bienais, trienais ou perenes, do tipo herbáceos, arbustivos, cipós, arbóreos, ervas e trepadeiras entrelaçadas ou com gavinhas (Polhill *et al.* 1981, Lewis *et al.* 2005, LPWG 2017).

Essa subfamília é normalmente caracterizada por flores pentâmeras papilionáceas (semelhantes a borboletas) de simetria zigomorfa (Tucker 2003, ILDIS 2005, Cullen *et al.* 2011). Que são pétalas que se diferenciam em duas alas, onde duas pétalas unidas formam a quilha e uma única pétala que forma o estandarte (Arroyo 1981, Tucker 2003).



Fonte da Foto: <http://leguminosae.blogspot.com/>

Figura 1. Flor papilionácea típica (*Centrosema platycarpum* Benth).

No entanto, sua morfologia floral pode ser bem mais diversa, desde simetrias actinomorfas com cinco pétalas indiferenciadas, até zigomorfas com poucas pétalas diferenciadas, que também podem ser ausentes ou restritas, somente a pétala adaxial padrão com estames livres e frequentemente numerosos (Pennington *et al.* 2000, Cardoso *et al.* 2012a).

Seus frutos variam muito, incluindo vagens com poucas ou muitas sementes, deiscentes ou indeiscentes, lomento, sâmara ou drupa. As sementes variam em

morfologia e coloração, possuindo normalmente uma testa dura, eventualmente crescida, raramente com uma sarcotesta, com uma válvula hilar complexa, hilo alongado e cristalino frequentemente presente, pleurograma ausente, podendo ter um embrião curvo ou raramente reto (LPWG 2017).

Esta subfamília também é bastante reconhecida por sua importância ecológica mais difundida, que é a sua capacidade de desenvolver nódulos rizobianos (indeterminados ou determinados) simbióticos com bactérias fixadoras de nitrogênio (Doyle 2016, LPWG 2017), permitindo a fixação do nitrogênio atmosférico. Na literatura foi observado que dentre as espécies da subfamília Papilionoideae as integrantes do clado ADA não desenvolvem nódulos rizobianos (Cannon *et al.* 2015, Sprent 2009, Zhao *et al.* 2021). Em estudos mais recentes foi observado uma possível associação entre a poliploidia e a nodulação, que consistiu no evento de duplicação do genoma inteiro no ancestral comum mais recente de Papilionoideae, com a poliploidia influenciando de forma direta para a evolução do sistema de nodulação rizobial simbiótico moderno (Cannon *et al.* 2015, Koenen *et al.* 2020, Leebens-Mack *et al.* 2019, Stai *et al.* 2019, Zhao *et al.* 2021).

As leguminosas Papilionoideae também são caracterizadas por produzir compostos secundários de natureza irritante ou venenosa que podem estar associadas ao sistema de defesa da planta ao ataque de insetos. Podendo serem encontrados nas folhas e partes frutíferas, incluindo isoflavonóides, flavonóides prenilados, indolizidina ou alcalóides quinolizidínicos. Também pode ser encontrado aminoácidos não proteicos, como a canavanina, identificados exclusivamente em Papilionoideae (Cardoso *et al.* 2013, LPWG 2017), podendo compreender até 5% do peso seco das sementes (Berry *et al.* 2018). Bem como a substância orgânica de origem natural rotenona, que é tóxico para vários organismos, encontrada nos extratos de raízes, caules e folhas de muitas espécies de Papilionoideae dos gêneros *Derris*, *Lonchocarpus* e *Tephrosia*, que os povos primitivos costumavam usar essas plantas para envenenar peixes (Berry *et al.* 2018), que posteriormente identificado por suas propriedades inseticidas com efeito para uma grande maioria de insetos (Zubairi *et al.* 2016).

2.2. IMPORTÂNCIA ECOLÓGICA E ECONÔMICA

As leguminosas Papilionoideae são ecologicamente e economicamente importantes (Lewis *et al.* 2005, Yahara *et al.* 2013). Sendo a característica ecológica mais conhecida na subfamília a fixação biológica de nitrogênio (FBN) atmosférico em

simbiose com bactérias do gênero *Rhizobium* (LPWG 2017), que é o segundo processo biológico natural mais relevante, ficando apenas atrás da fotossíntese (Unkovich 2013). O nitrogênio é um elemento essencial no metabolismo vegetal e animal, sendo um componente importante localizado em todos os aminoácidos presentes nas proteínas. Ainda que o nitrogênio seja o elemento mais abundante na atmosfera, com aproximadamente 80% da atmosfera composta de gás nitrogênio (N_2), esse reservatório abundante existe em uma forma que não está diretamente disponível para os seres vivos, para que o nitrogênio possa ser metabolizado pelos organismos vivos é necessário que esteja na forma de nitratos ou compostos de amônia (Berry *et al.* 2018).

Para que o nitrogênio atmosférico seja assimilável pelos organismos vivos, é necessária uma associação mutua exclusiva entre bactérias diazotróficas do solo e as leguminosas (principalmente Papilionoideae), que induzem a formação de nódulos nas raízes das plantas hospedeiras. Que por meio de uma simbiose entre leguminosas e bactérias *Rhizobium*, os rizóbios fixam o dinitrogênio atmosférico (N_2), sendo disponibilizados às plantas na forma de amônio, que pode ser facilmente metabolizado pela planta (Püschel *et al.* 2017). Para tanto, ocorre uma troca entre carbono produzida pela planta e do amônio fixado pela bactéria nos nódulos rizobianos, em uma complexa interação biossintética entre a planta hospedeira (leguminosas) e a bactéria, onde são formados compostos de nitrogênio a serem utilizados pelas plantas (Berry *et al.* 2018), que possibilita que as leguminosas possam se desenvolver naturalmente em solos pobres em nitrogênio, que são aparentemente favorecidas em condições áridas (Pellegrini *et al.* 2016). Podendo a nodulação estar associada com uma maior eficiência do uso da água em áreas determinadas, se tornando uma característica significativa na adaptação as mudanças climáticas (Adams *et al.* 2016).

Dessa maneira, a simbiose leguminosa-rizóbio traz benefícios para a agricultura, seja pela economia de fertilizantes nitrogenados (Signorelli *et al.* , 2020), redução das emissões dos gases de efeito estufa (Puozaa, Jaiswal e Dakora, 2019), restringindo seu impacto negativo ao meio ambiente, sendo imprescindível para se aproximar cada vez mais de uma agricultura sustentável (Signorelli *et al.* 2020). Também foi observado que essas leguminosas apresentam um maior conteúdo de proteínas que o observado em outras culturas, esse fato se deve a presença das bactérias fixadoras de nitrogênio presentes nos nódulos rizobianos nas raízes das leguminosas (Kouris-Blazos e Belski 2016).

Todos esses benefícios tem proporcionado um aumento nas pesquisas voltadas para entender os mecanismos da FBN, a partir dessas informações os melhoristas têm se empenhado para conseguir introduzir a capacidade de realizar FBN em plantas não leguminosas, por meio de engenharia genética pela introdução da enzima nitrogenase nessas plantas (Oldroyd e Dixon 2014) ou utilizando a bactéria fixadora de nitrogênio endofítica *Gluconacetobacter diazotrophicus*, uma bactéria fixadora de nitrogênio não nodulante (Cocking *et al.* 2006, Dent 2018).

Além de todos esses benefícios advindos do uso de leguminosas Papilionoideae para manutenção do nitrogênio do solo. Essas leguminosas geralmente são utilizadas para revitalizar solos pobres em nutrientes, em particular terras agricultáveis e pastagens abandonadas ou degradadas. Como também por ser utilizado no controle de plantas daninhas a partir da rotação de culturas, por manter a área sempre coberta, permitindo uma baixa infestação de plantas daninhas. Essas espécies tem por característica produzir uma grande quantidade de biomassa que posteriormente ser incorporado ao solo como adubo verde, fornecendo nutrientes e melhorando a qualidade do solo e fixação de nitrogênio para as próximas safras (LPWG 2017, Berry *et al.* 2018).

As leguminosas alimentares pertencentes à subfamília Papilionoideae, economicamente são a segunda em importância, ficando atrás apenas das espécies de Poaceae (LPWG 2017). Em sua maioria as leguminosas usadas para alimentação humana são plantas multivalentes, também sendo utilizadas para alimentação animal (forragem) e melhoramento das propriedades físicas e químicas solo (Berry *et al.* 2018). Sendo as espécies agrícolas representantes de Papilionoideae mais importantes, a soja (*G. max*), feijão (*P. vulgaris*), feijão-caupi (*V. unguiculata*), feijão guandu (*Cajanus cajan*), grão de bico (*Cicer arietinum*), lentilha (*Lens culinaris*), ervilha (*P. sativum*) e amendoim (*A. hypogaea*). Além da enorme importância alimentar dessas leguminosas Papilionoideae, a ervilha foi o a espécie chave utilizada no trabalho pioneiro de Mendel que estabelece a base da hereditariedade (Smýkal 2014).

Entre as leguminosas papilionoideae forrageiras temperadas estão trevos (*Trifolium* ssp.), alfafa (*Medicago sativa*), trifólio de pés de ave (*Lotus corniculatus*) e ervilhaca (*Vicia craca*). Já nos trópicos ou regiões semiáridas e áridas, algumas espécies importantes dessas regiões são Glycine (soja), *Stylosanthes* e *Desmodium* (carrapato trevo). Além de todas essas finalidades essas plantas são utilizadas como madeira, taninos, óleos e resinas, na fabricação de vernizes, tintas, medicamentos e

no paisagismo (LPWG 2017).

Em 2019, a produção mundial de leguminosas (leguminosas colhidas para sementes secas) alcançou mais de 90 milhões de toneladas, com um incremento de 34 milhões de toneladas em relação a 2001 (FAOSTAT 2021). Sendo o 818º produto mais comercializado no mundo, com uma negociação total de US \$ 1,48 bilhão, representando 0,0082% do comércio mundial total. Neste ano, os principais exportadores de leguminosas foram Marrocos (\$ 272 milhões), Guatemala (\$ 161 milhões), Holanda (\$ 130 milhões), México (\$ 122 milhões) e Espanha (\$ 104 milhões). E os principais importadores foram Estados Unidos (\$ 227 milhões), Espanha (\$ 183 milhões), Holanda (\$ 172 milhões), Reino Unido (\$ 153 milhões) e Canadá (\$ 116 milhões) (BACI 2021). No Brasil, a produção total das cinco principais culturas de leguminosas (amendoim, ervilha, fava, feijão e soja) em grão foi de 125.503.865 toneladas para o ano de 2020 (IBGE 2021).

Essas altas cifras envolvendo as leguminosas Papilionoideae são determinadas pelo aumento global da demanda por leguminosas por serem importantes componentes alimentares, visto que as leguminosas podem ser um complemento aos cereais nas necessidades proteicas e por possuírem uma alta taxa de proteína digestível (Varshney *et al.* 2018), que fazem parte de uma dieta que fornece ao corpo os nutrientes essenciais para uma boa nutrição humana, fazendo parte da alimentação de quase todas as populações. Que provém uma diversidade de nutrientes essenciais, como proteínas, carboidratos (baixo índice glicêmico), fibra, minerais e vitaminas (Çakir *et al.* 2018). O que propicia benefícios a saúde principalmente em regiões de alta densidade populacional, e como cultura de subsistência para pequenos agricultores principalmente nos países subdesenvolvidos.

2.3. MELHORAMENTO GENÉTICO DA SUBFAMÍLIA PAPILIONOIDEAE

Para iniciar ou executar os programas de melhoramento é indispensável que o melhorista conheça a história evolutiva da cultura, o que possibilita a compreensão da base genética dos fatores que podem limitar a adaptação e a produção das espécies a serem melhoradas (Abbo *et al.* 2019).

Nesse sentido, a reconstrução e compreensão detalhada das relações filogenéticas da subfamília Papilionoideae é fundamental para o entendimento da origem e diversificação dessas leguminosas de grande importância alimentar, econômica e ecológica (Smýkal *et al.* 2015). Para tanto, a inferência da datação

molecular de divergências é uma etapa indispensável na maioria das análises de filogenia, uma vez, que permite adicionar uma dimensão temporal à árvore analisada (Tamura *et al.* 2012). A inclusão dos dados de tempo de divergência permite averiguar as correlações entre os eventos evolutivos. Deixando que os eventos evolutivos não sejam comparados somente com outros eventos evolutivos, mas também como outros eventos históricos e geológicos (Crosby e Williams 2017).

Desse modo, a genômica comparativa de genomas de cloroplasto torna possível a compreensão das diferenças evolucionárias e moleculares entre espécies de um grupo, por meio de análises filogenéticas de aplicabilidade no pré-melhoramento genético visando identificar características úteis de interesse aos programas de melhoramento genético das espécies dessa subfamília.

Dessa forma, as informações das sequências completas dos genomas, que por meio de ferramentas da bioinformática estão disponíveis nos bancos públicos, como o tamanho do genoma, número de genes, RNA transportador (tRNA), RNA ribossômico (rRNA) e número cromossômico, podem servir de embasamento para futuros trabalhos associados ao melhoramento genético das culturas. Para isso, é imprescindível o conhecimento dos aspectos evolutivos dessa subfamília, tendo em vista, que mesmo não havendo variação no número cromossômico, pode existir alteração no tamanho do genoma, sendo uma informação bastante significativa para estudos de sequenciamento de genomas e caracterização de novos marcadores moleculares (Loureiro *et al.* 2010), reforçando a relevância das estimativas do conteúdo de DNA para a compreensão da evolução cariotípica de espécies homoplóides, auxiliando na preparação de cruzamentos interespecíficos para fins de melhoramento (Loureiro *et al.* 2010, Prieto 2020).

Os dados de rRNA e tRNA também podem servir de base para estudos de modificações de RNA, tendo potencial para estimular as respostas das plantas ao estresse salino, de altas temperaturas e infecções virais (Anderson *et al.* 2018, Martínez-Pérez *et al.* 2017). Dessa forma, diversos pequenos RNAs atuam na tolerância ao estresse das plantas por meio do sistema de silenciamento do gene pós-transcricional (Shukla *et al.* 2008). Conseqüentemente, os componentes essenciais das pequenas vias de biogênese de RNA sugerem possíveis alvos para a edição do genoma (Yu e Bao 2020).

Outro aspecto a ser explorado nas Papilionoideae é a presença de alcalóides quinolizidínicos sintetizados pelas espécies da tribo Genisteeae, que são compostos de sabor amargo associados a defesa dessas plantas ao ataque patógenos e insetos

predadores (Bunsupa *et al.* 2013, Wink e Mohamed 2003). Em vista disso, essas substâncias podem ser exploradas nos programas de melhoramento das espécies que não sintetizam essas substâncias, a partir de cruzamentos das espécies compatíveis ou por meio de engenharia genética em espécies distantes.

2.4. BIOINFORMÁTICA

A bioinformática é um ramo interdisciplinar da ciência que se baseia em análise de dados, que abrange principalmente a ciência da computação, biologia molecular, genética, matemática, estatística e afins, destinada ao agrupamento, interpretação e compreensão de eventos biológicos a partir da grande quantidade de informações geradas pela biologia molecular (Can 2014). Dessa forma, várias ferramentas computacionais foram criadas visando processar o grande volume de informação disponíveis nos bancos de dados moleculares (Edwards e Batley 2004).

Ao passo que as ferramentas de sequenciamento se tornaram mais acessíveis, ágeis e precisas, observou-se um elevado crescimento no volume de dados armazenados nos bancos de dados genômicos, que resultou em sequências de nucleotídeos, aminoácidos e proteínas disponíveis para uso em várias pesquisas a partir dos recursos computacionais de bioinformática (Morais 2003, Bayat 2002). Essas sequências de nucleotídeos (em torno de milhões de pares de bases) ao serem inseridas nos bancos de dados públicos recebem uma identificação (código) única chamada de acesso, o que dificulta a possibilidade de duplicação de dados a serem disponibilizados aos pesquisadores e a comunidade acadêmica (Morais 2003, Reis *et al.* 2011).

O banco de dados público mais importante e abrangente é o *GenBank*, que distribui todas as sequências disponíveis de DNA, RNA e proteínas de múltiplos organismos, construídas e disponibilizadas pelo *National Center for Biotechnology Information* (NCBI), que é uma divisão da *National Library of Medicine* (NLM), situado no campus do Instituto Nacional de Saúde dos EUA (NIH) em Bethesda, MD, EUA (Sayers *et al.* 2020). Além de todas essas informações obtidas no *GenBank*, também é possível localizar dados acerca de genomas completos, mapas gênicos, taxonomia e estruturação proteica (Benson *et al.* 2000, Reis *et al.* 2011). Ferramentas da bioinformática também tem sido utilizadas em estudos de taxonomia, filogenia e na caracterização da estrutura tridimensional de proteínas de diferentes espécies (Liu *et al.* 2016, Moraes Filho e Martins 2016, Jimenez *et al.* 2019, Khan *et al.* 2019, Liu *et al.* 2019, Jimenez *et al.* 2020, Koenen *et al.* 2020, Zha *et al.* 2020).

O NCBI disponibiliza várias ferramentas que podem ser utilizadas pelos pesquisadores no momento de busca por sequências de interesse, dentre elas o *Basic Local Alignment Search Tool* (BLAST) que pode ser utilizado para gerar comparações entre as diferentes sequências disponíveis no banco de dados, permitindo agrupar por ordem de prioridade com base na similaridade máxima das sequências analisadas, sendo considerada homóloga àquela sequência com maior pontuação no banco de dados (Bayat 2002). Posteriormente a esse alinhamento, pode-se fazer *download* das sequências que foram selecionadas em formato FASTA ou Genbank.

Diferentemente do BLAST o *software* de alinhamento *Multiple Sequence Comparison by Log-Expectation* (MUSCLE), tem como função a execução de alinhamento de múltiplas sequências, sendo caracterizado pelo alinhamento aos pares para determinar os níveis de similaridades das sequências com complexidade de tempo e espaço reduzida (Edgar 2004).

A partir da grande disponibilidade de dados inseridos nos bancos públicos, se fez necessário o desenvolvimento de novos programas de bioinformática que possibilitassem a análise dessas informações de forma mais eficientes. Nesse sentido ao longo dos anos vários programas foram desenvolvidos, dentre eles o programa *Molecular Evolutionary Genetics Analysis* (MEGA), que possibilitou que o número de estudos relacionados com a análise comparativa dos acessos disponíveis nos bancos públicos se intensificassem, buscando analisar as similaridades entre as sequências de nucleotídeos e proteínas, propiciando a reconstrução da história evolutiva das espécies a partir dessas sequências (Kumar *et al.* 2004, Tamura *et al.* 2007, Kumar *et al.* 2016). No *software* MEGA alguns algoritmos podem ser utilizados, como *Unweighted Pair Group Method with Arithmetic Means* (UPGMA), Máxima Verossimilhança, Máxima Parcimônia e Neighbor-Joinning, sendo a Máxima Parcimônia e Máxima Verossimilhança os métodos cladísticos mais utilizados (Lesk 2008).

Nas últimas décadas, o aumento do número de pesquisas na área da genômica comparativa foi propiciado pelo surgimento de tecnologias de sequenciamento e genotipagem de próxima geração (Chawla *et al.* 2016, Kagale e Close 2021) e o desenvolvimento de ferramentas da bioinformática. Como consequência a utilização do sequenciamento de DNA e RNA está permitindo obter informações significativas sobre variabilidades genéticas, epigenéticas e estruturais ocultas que podem auxiliar no melhoramento de caracteres complexos de leguminosas (Kagale e Close 2021).

Por meio desses avanços a genômica das leguminosas vem se desenvolvendo

rapidamente. A primeira espécie de leguminosa pertencente a subfamília Papilionoideae a ser sequenciada foi a soja (*G. max* L.), há mais de duas décadas (Schmutz *et al.* 2010).

Posteriormente foram geradas as sequências do genoma de referência para as principais culturas de leguminosas, incluindo feijão guandu (*C. cajan* L.) (Varshney *et al.* 2012), grão de bico (*C. arietinum* L.) (Varshney *et al.* 2013), feijão comum (*P. vulgaris* L.) (Schmutz *et al.* 2014), ervilha (*P. sativum* L.) (Kreplak *et al.* 2019), feijão-caupi (*V. unguiculata* L.) (Lonardi *et al.* 2019). E recentemente os genomas do Lablab (*Lablab purpureus* L.) (Chang *et al.* 2018) e de duas espécies de tremoço (*Lupinus angustifolius* L. e *Lupinus albus* L.) (Kamphuis *et al.* 2021) foram sequenciadas.

2.5. GENOMA PLASTIDIAL

Os plastídios são caracterizados por possuírem estruturas subcelulares (organelas) que desempenham funções essenciais nas células vegetais, como a fotossíntese e geração e armazenamento de metabólitos (Choi *et al.* 2021). Desenvolveram-se inicialmente por meio de um evento endossimbiótico entre os procariontes fotossintéticos e os ancestrais eucarióticos das algas (Keeling 2013, Ševčíková *et al.* 2015), que em decorrência desse acontecimento desenvolveram membrana dupla (Cavalier-Smith 2000, Raven *et al.* 2001). Os plastídios podem ser classificados em indiferenciados e diferenciados. Os indiferenciados são pequenos e de estrutura clara chamados de proplastídeos, detectadas nos tecidos meristemáticos e reprodutivos. Em contrapartida os diferenciados são subdivididos em leucoplastos (branco), cloroplastos (verde) e cromoplastos (amarelo, laranja ou vermelho) (Jarvis e López-Juez 2013, Choi *et al.* 2021).

Os cloroplastos são organelas vegetais especializadas em fotossíntese presentes em algas e plantas, sendo a fotossíntese a sua principal função. Além da importância da fotossíntese na sustentação, desenvolvimento vegetal e na produtividade das culturas, os cloroplastos também são responsáveis por diversas funções essenciais, como síntese de aminoácidos e ácidos graxos, importação de proteínas plastidiais, proteólise e sinalização de estresse nas plantas (Xu *et al.* 2015, Chan *et al.* 2016).

O genoma do cloroplasto é constituído do seu próprio DNA (cpDNA), conformação circular e fita dupla, sendo independente do genoma nuclear com uma

estrutura altamente conservada e herança exclusivamente materna (Raubeson e Jansen 2005, Bock 2007, Xiong *et al.* 2009, Wicke *et al.* 2011).

Geralmente, esse genoma é composto por uma grande região de cópia única (LSC), uma pequena área de cópia única (SSC) e por duas regiões invertidas (IR) (Bock 2007, Wicke *et al.* 2011), com 120 a 130 genes (Rogalski *et al.* 2015), codificando rRNA, tRNA, peptídeos e três genes para a subunidade da RNA polimerase, genes para o fotossistema I, fotossistema II, citocromo e síntese de ATP com aproximadamente 80 proteínas (Green *et al.* 2011, Huang *et al.* 2013). Seu tamanho pode variar de 107 kb a 218 kb, a depender da espécie (Daniell *et al.* 2016, Menezes *et al.* 2018).

Dentre todos os genes encontrados nos genomas plastidiais alguns ganham grande destaque nas análises filogenéticas, como os genes *matK*, *psaA*, *psbA*, *psbD*, *rbcL* e *rpoC2*, por estarem ligadas a funções fisiológicas de grande importância nas plantas. O gene *matK* possui aproximadamente 1570 pb e codifica para a organela intron maturase, uma proteína que *splicing* íntrons RNA do tipo II (Neuhaus e Link 1987). Nesse sentido, o *matK* tem uma função imprescindível no cloroplasto, como um fator de *splicing* pós-transcricional, haja vista que os transcritos de RNA são necessários para o pleno funcionamento do cloroplasto, como no processo de fotossíntese (Michelle *et al.* 2007). Por possuir uma maior taxa evolutiva (se comparado a outros genes plastidiais), o gene *matK* é amplamente utilizado na em construções filogenéticas à nível de ordem ou família (Lahaye *et al.* 2008), bem como à nível de gênero ou espécie (Wilson *et al.* 2005), sendo também bastante relatado em estudos de DNA *Barcode* de plantas (Gao *et al.* 2008).

O gene *rbcL* codifica a maior subunidade da enzima RuBisCO (ribulose-1,5-bifosfato carboxilase oxidase), que é uma enzima reguladora que atua na catalisação e fixação de CO₂ e a oxigenase, determinante na fotossíntese líquida (Andersson e Backlund 2008). Esse gene é bastante utilizado em estudos que envolvem a sistemática de plantas, por apresentar uma taxa de evolução mais conservadora, que propicia uma significativa resolução das relações à nível de família botânica (Chase *et al.* 1993, Salamin *et al.* 2005, Pinto 2015). Esse gene também tem por vantagem a combinação do íntron *trnL* e os espaçadores intergênicos *trnL-trnF* que tem sido bastante empregada nas análises filogenéticas à nível de táxon, tornando-se eficiente em identificar variações inter e intraespecífica em diferentes espécies (Tsai *et al.* 2006). Nesse sentido, os genes *rbcL* têm sido amplamente utilizados em estudos filogenéticos à nível de subclasses e famílias, inclusive entre diversos grupos de

gimnospermas e angiospermas (Hollingsworth *et al.* 2016), bem como nos estudos de *barcoding*, por ser universal, facilmente amplificado e por sua comparabilidade (Chase *et al.* 2007).

O gene *psbA* codifica a proteína D1 envolvida no centro da reação do fotossistema II (PSII) (Mulo *et al.* 2009, Wegener *et al.* 2015). Constantemente a proteína D1 está sujeita a fotodano devido à natureza da fotoquímica do PSII, reduzindo assim a atividade fotossintética. Nesse sentido, após a desmontagem parcial do complexo PSII, as proteínas D1 que sofreram fotodanos são substituídas por novas proteínas D1 (Takahashi e Badger 2011, Lu *et al.* 2013). Dessa forma, a eficiência de transcrição e tradução de *psbA* durante a fotoinibição desempenham um papel importante nas respostas ao estresse de alta luz e temperatura (Pfannschmidt *et al.* 1999, Danilova *et al.* 2018). Huo *et al.* (2016) analisando o efeito do gene *psbA* do milho (*Zm_{psbA}*) na tolerância à seca do tabaco (*Nicotiana tabacum* L.), constataram que a superexpressão de *Zm_{psbA}* melhora a tolerância à seca em tabaco, atenuando a redução fotossintética, diminuindo o acúmulo de espécies reativas de oxigênio e danos na membrana, acioando o sistema de defesa da planta contra o estresse abiótico. Nesse sentido o *Zm_{psbA}* pode ser utilizado no melhoramento molecular de plantas para tolerância à seca.

O gene *psbD* codifica a proteína D2 do centro de reação da PSII, sendo controlado por promotor único responsivo à luz azul (BLRP) que por sua vez, é transcrito pelo plastídio de RNA polimerase (Yuichi *et al.* 2004). A transcrição de *psbD* BLRP além de ser regulada pelo fator sigma 5 localizado no cloroplasto (Nagashima *et al.* 2004, Sunoyama *et al.* 2004), também é induzida por estresses abióticos, tais como salino, osmótico e baixa temperatura (Nagashima *et al.* 2004). A transcrição de *psbD* BLRP também é intermediado pelo processo de sinais de fitocromos em resposta às proporções relativas de luz vermelha e vermelha distante (Belbin *et al.* 2017). Dessa forma, a transcrição do *psbD* BLRP é permeada pela resposta da planta ao estresse abiótico e a percepção da luz (Zhang *et al.* 2020). Por conseguinte, a redução dos efeitos negativos ocasionados pela luz intensa ao PSII, pode ser obtido pela alta atividade de *psbD* BLRP que favorece a síntese de D2 (Tsunoyama *et al.* 2002).

O gene *psaA* codifica a clorofila α -apoproteína P700 do fotossistema I, que é dividido em três éxons separados que se apresentam de forma dispersa em *loci* distante do genoma cpDNA (Kück *et al.* 1987). Pelos éxons serem ladeados por sequências de íntrons do grupo II, são necessárias duas reações de *trans-splicing*

para gerar um RNA mensageiro maduro (Glanz e Kück 2009). Duas proteínas de repetição octatricopeptídica (*RAT2* e *RAA1*) são indispensáveis para o *trans-splicing* do gene *psaA* (Balczun *et al.* 2005, Merendino *et al.* 2006, Eberhard *et al.* 2011). Dessa forma, mutantes com falha no *RAT2* ou *RAA1*, além da deficiência no *trans-splicing*, também têm uma baixa eficiência fotossintética, uma vez que não possuem a clorofila α -apoproteína P700 (Kück *et al.* 1987 Goldschmidt-Clermont *et al.* 1990, Jacobs *et al.* 2010).

O gene *rpoC2* desempenha uma importante função, codificar a subunidade β da RNA polimerase (RNA polimerase tipo I) (Androsiuk *et al.* 2020). Sendo a RNA polimerase uma enzima fundamental no controle da transcrição genômica e expressão gênica (Han *et al.* 2022). O gene *rpoC2* também é considerado uma sequência de cloroplasto de rápida evolução (Logacheva *et al.* 2007), que passa por seleção positiva em diversos grupos de espécies de plantas (Androsiuk *et al.* 2020).

Além da diversidade dos genes e suas funções, também é observado uma variação no tamanho do plastoma, que em leguminosas se deve a um fenômeno evolutivo, ocasionando a expansão, contração ou perda das regiões invertidas (IR) (Davis e Soreng, 2010). Sendo os plastomas menores caracterizados pela perda da IR, agrupadas no clado sem repetição invertida (IRLC) (Wojciechowski *et al.* 2000). Em contrapartida os plastomas maiores tem por característica as IRs em expansão, típico das espécies do clado de expansão repetida invertida (IREC) (Dugas *et al.* 2015, Wang *et al.* 2017). Nas espécies da subfamília Papilionoideae comumente são observadas inversões do plastoma (Doyle *et al.* 1996). Por apresentarem desvios da estrutura típica do plastoma, as leguminosas (Fabaceae) denotam vários rearranjos de relevância filogenética (Wojciechowski *et al.* 2000). Desse modo, análises comparativas do cpDNA de espécies diversas também propiciam uma melhor compreensão da evolução estrutural dos genomas (Yang *et al.* 2019).

O genoma plastidial tornou-se uma ferramenta bastante importante em estudos envolvendo a filogenia e evolução de espécies, visto que oferece diversas vantagens, incluindo tamanho altamente conservado, estrutura, ordem linear dos genes e conteúdo genético (Chumley *et al.* 2006, Shaw *et al.* 2007, Avise 2009, Green *et al.* 2011), bem como natureza não recombinante, herança uniparental e estar distribuído abundantemente na célula (Nishikawa *et al.* 2002, Dhingra e Folta 2005, Provan 2008).

Por todas essas vantagens e com o advento do Sequenciamento de Nova Geração e os avanços das ferramentas de bioinformática, nos últimos anos observou-

se um aumento significativo no número de estudos a partir dos genomas plastidiais, incluindo análises filogenéticas (Penjor *et al.* 2013, Morton e Telmer 2014, Fu *et al.* 2017, Khan *et al.* 2019, Liu *et al.* 2019, Koenen *et al.* 2020, Jimenez *et al.* 2020, Zha *et al.* 2020), caracterização *in silico* de proteína (Jimenez *et al.* 2019), estudos filogeográficos (Pfeil *et al.* 2008, Scarcelli *et al.* 2011, Apeelhans *et al.* 2018, Wu *et al.* 2018), geralmente obtidos a partir de genes plastidiais.

Atualmente no banco de dados NCBI, estão disponíveis as sequências de nucleotídeos do genoma completo de cloroplasto de diversas espécies vegetais. Para as espécies da subfamília Papilionoideae é possível encontrar as sequências dos plastomas de 394 espécies (NCBI 2022). As sequências dos genomas plastidiais das leguminosas Papilionoideae depositadas nos bancos de dados propiciará novos avanços nas diversas áreas da biologia molecular.

2.6. FILOGENIA MOLECULAR

Entende-se por filogenia o estudo das relações evolutivas entre grupos taxonômicos, que é representada por uma árvore filogenética (representação gráfica) que associa essas relações evolutivas. Geralmente essas árvores apresentam nós que são ligados por ramos, onde esses ramos representam a estabilidade de um ancestral genético ao longo do tempo, e cada nó representa o surgimento de uma nova linhagem a partir desse ancestral comum (Ridley 2006, Yang e Rannala 2012). Desse modo, as relações filogenéticas entre um grupo específico de espécies podem ser inferidas a partir de caracteres análogos partilhados por esse grupo de espécies. Esses caracteres podem ser morfológicos ou moleculares, contudo, atualmente nos estudos filogenéticos dão-se preferência por associar um maior número de caracteres moleculares, por meio de marcadores moleculares de diversas espécies (Ridley 2006, Aleixo *et al.* 2006, Gompert *et al.* 2006, Maira *et al.* 2013).

Neste sentido, a filogenia molecular se diferencia da filogenia de caracteres morfológicos pela utilização de marcadores moleculares, que podem ser sequências de nucleotídeos, RNA e proteína, que auxiliam no esclarecimento das relações evolutivas de espécies pertencentes a um determinado táxon (Graur e Li 2000). O estudo de caracteres moleculares é muito semelhante ao da inferência com caracteres morfológicos, entretanto, não existe tanta variação na sequência de DNA e aminoácidos, por esse motivo os marcadores moleculares se sobressaem por se basearem apenas na expressão fenotípica, bem como moderada interação epistática e pleiotrópica, sendo indiferente ao estágio de desenvolvimento do indivíduo (Aleixo

et al. 2006, Ridley 2006). Por esses motivos os dados moleculares tem sido correlacionados com os dados morfológicos, afim de confirmar a ligação evolutiva de espécies pertencentes ao mesmo táxon (Aleixo *et al.* 2006). Neste sentido, os dados moleculares revolucionaram os estudos evolutivos, aumentando de forma exponencial a resolução das árvores filogenéticas antes não bem esclarecidas (Patil *et al.* 2015).

Dessa forma, vários genes plastidiais são utilizados em estudos que relacionam espécies dentro de família ou gênero, incluindo, *accD*, *matK*, *ycf1*, *ycf2*, *rpoC2*, *ndhF*, em contrapartida os genes *psbC*, *psbB*, *rbcL*, *ndhB*, *psbA*, *psbD*, *psaB* e *psaA*, permite fazer inferências filogenéticas a grupos taxonômicos mais distantes, como entre angiospermas e gimnospermas (Kim *et al.* 2009).

Diante disso, os estudos filogenéticos da subfamília Papilionoideae baseado em genes de cloroplastos vem revelando informações esclarecedoras para compreensão das relações evolutivas dessa subfamília, entretanto a maioria dessas análises utilizam poucos genes, como *matK* ou *rbcL*, sendo necessário a inserção de um maior conjunto de genes e táxons para um estudo evolutivo mais esclarecedor (Cardoso *et al.* 2013).

Para tanto, os tipos de análises utilizados na construção de árvores filogenéticas por meio de marcadores moleculares são as fenéticas e cladísticas (Eymygdio *et al.* 2003). Nas análises fenéticas, são consideradas apenas medidas de de similaridade fenotípica, utilizando o máximo de caracteres possíveis, não levando em consideração as questões de ancestralidade, sendo o UPGMA o algoritmo mais utilizado. Por outro lado, no método cladístico se baseia na história evolutiva do ancestral comum entre espécies do mesmo táxon (Lesk 2008, Marouelli 2009).

Os métodos cladísticos mais utilizados na reconstrução filogenética são os de Máxima Parcimônia (Farris 1983), Máxima verossimilhança (Edwards e Cavalli-Sforza 1964) e Análise Bayesiana (Huelsenbeck *et al.* 2001).

O método da Máxima Parcimônia baseia-se na escolha da hipótese que implica no menor número de mudanças ao longo de todos os ramos para gerar reconstruções filogenéticas (Ridley 2006, Lesk 2008). Está análise geralmente geram mais de uma árvore, dessa forma o pesquisador deverá escolher a árvore de maior consenso entre as árvores geradas para a análise filogenética em questão (Marouelli 2009). Esse método tem como desvantagem a não garantia de que todas as informações contidas nas sequências serão recuperadas, uma vez que nem toda variabilidade será utilizada como sequência de sítios de DNA elucidativo, em virtude de o princípio da parcimônia não conseguir assegurar que esses sítios não poderiam ser alterados ao longo das

mudanças evolutivas na espécie atual e o ancestral mais próximo (Nei e Kumar 2000, Ridley 2011). A Máxima Parcimônia não especifica um modelo de substituição determinante. Dessa forma tem como suposição uma taxa de evolução lenta (Xia 2020).

Diferentemente da Máxima Parcimônia, na Máxima Verossimilhança sua análise leva em consideração todos os sítios informativos (até eventos mutacionais) sem diferenciação (Schneider 2007, Lesk 2008, Lemey *et al.* 2009). No qual o princípio da Máxima Verossimilhança baseia-se em uma construção filogenética por meio de probabilidades quantitativas complexas, que a partir das sequências averiguadas, um modelo estatístico é disponibilizado, sendo levado em consideração sua topologia e comprimento dos ramos baseado nos valores de probabilidade, para construção da ancestralidade das espécies taxonômicas analisadas. Dessa forma, a topologia de maior probabilidade será a mais correta, sendo a árvore filogenética mais precisa (Schneider 2007, Lesk 2008, Lemey *et al.* 2009). Atualmente, com a elevada disponibilização de sequências nos bancos públicos, atrelado ao surgimento de novas ferramentas de bioinformática e uma maior capacidade computacional, a Máxima Verossimilhança se tornou um dos métodos mais comumente usados por sua maior precisão para inferência filogenética (Ridley 2006).

A análise bayesiana é um método estatístico que consiste na avaliação de hipóteses pela Máxima Verossimilhança. Essa análise é fundamentada em três tipos de distribuições de probabilidades, a priori, verossimilhança e posterior. No cálculo da inferência bayesiana a probabilidade posterior de uma hipótese (uma árvore) é advinda pela verossimilhança multiplicada pela probabilidade a priori desta hipótese, estando esse valor associado a probabilidade da árvore estar correta por meio da probabilidade, modelo e dados selecionados anteriormente. Para tanto a hipótese perfeita é a que exhibe a máxima probabilidade posterior, para essa estimativa é necessário selecionar um modelo evolutivo, onde este valor pode ser analisado como sendo a probabilidade da árvore estar correta (Felsenstein 2004, Huelsenbeck *et al.* 2001).

Nos métodos cladísticos a melhor árvore é configurada como sendo aquela que apresenta uma menor variação ao longo dos ramos, que geralmente expressam a confiabilidade (teste de confiança) das árvores filogenéticas por meio do método do *bootstrap* (Yang e Rannala 2012).

2.7. SISTEMÁTICA FILOGENÉTICA

A sistemática vegetal é a área da botânica responsável pela construção de hipóteses filogenéticas que tem por intuito estudar a diversidade e as relações evolutivas das plantas. O papel da sistemática é reunir as espécies dentro de um sistema que leva em consideração suas características morfológicas, relações genéticas e afinidades. A sistemática muitas vezes pode ser confundida com a taxonomia, entretanto essa ciência atua na elaboração das leis que norteiam a classificação, diferentemente da sistemática que se baseia na morfologia para classificar os seres vivos (Barroso 2004).

Nesse sentido, a sistemática filogenética permite reconstruir a história evolutiva dos organismos por meio de hipóteses apoiada na ancestralidade comum entre esses organismos (Marques e Amorim 1996, Lopes *et al.* 2014). Para tanto, é necessário entender que uma determinada característica possivelmente poderá sofrer variações na descendência, passando a se apresentar com modificações, que serão posteriormente herdadas nas gerações seguintes (Lopes *et al.* 2014). Sendo a apomorfia o termo designado para definir a característica (variação) mais recente, oriunda de um caráter derivado de uma espécie ancestral, e que são determinantes para definir um novo grupo. Os caracteres homólogos apomórficos podem ser exclusivos de apenas um táxon (autapomorfia) ou compartilhada por dois ou mais táxons originados de um mesmo ancestral (sinapomorfia) (Amorim 2002, Lopes *et al.* 2014).

Para o entendimento de uma filogenia é necessário compreender a distribuição dos táxons ao longo da árvore filogenética, onde os organismos são posicionados em árvores do tipo cladograma baseado no seu parentesco, em que os grupos unidos por apomorfias são denominados de monofiléticos, por possuírem um ancestral comum mais recente para todos os seus descendentes (Santos 2008). No entanto, agrupamentos que excluem uma das linhagens descendentes do mesmo ancestral comum, são chamados parafiléticos (Amorim 2002, Lopes *et al.* 2014). Nos estudos de filogenia molecular, comumente as sinapomorfias são utilizadas para classificar as espécies mais próximas (Lesk 2008).

Atualmente, cada vez mais as análises moleculares são associadas a caracteres morfológicos, anatômicos, embriológicos e bioquímicos para solucionar problemas na filogenia de vários grupos de espécies. Para tanto se faz necessário compreender como foi realizado a classificação de cada grupo em estudo.

Nesse sentido, a família Leguminosae anteriormente era distribuída em três subfamílias, Papilionoideae, Mimosoideae e Caesalpinioideae, tomando como base em um pequeno conjunto de caracteres florais conspícuos, que foram recentemente reclassificadas por meio de uma filogenia mais abrangente de táxons associado a evidências morfológicas e moleculares (gene de cloroplasto *matK*), o que possibilitou a reclassificação das três subfamílias tradicionais em seis subfamílias, sendo elas Caesalpinioideae, Cercidoideae, Detarioideae, Dialioideae, Duparquetioideae e Papilionoideae (LPWG 2017).

A origem das leguminosas está diretamente relacionado com o limite Cretáceo-Paleógeno (K-Pg) (Koenen *et al.* 2020, Zhao *et al.* 2021), sendo a idade da coroa estimada em mais de 67,28 milhões de anos atrás (*mya*), (Zhao *et al.* 2021). Posteriormente a esta época, sucedeu uma rápida diversificação entre as seis subfamílias dentro de 0,82 *mya*, próximo de 67 *mya* (Zhao *et al.* 2021). Que proporcionou a essas espécies uma dominância ecológica nos mais diversos tipos de vegetação, que por meio de registro fósseis sugere que elas alcançaram tal destaque após o K-Pg em paralelo com diversos grupos de animais bem estudados, incluindo Placentalia e Neoaves (Koenen *et al.* 2021).

O tempo de diversificação para a maioria das tribos de Papilionoideae é estimado em 64 a 50 milhões de anos (final do Paleoceno ao início do Eoceno). Dentro de Papilionoideae, NPAAA (acúmulo de aminoácidos não proteicos) é o maior clado, que inclui a maioria das leguminosas de importância alimentar e econômica, divergiu da tribo Baphieae em 59,49 *mya*. Os representantes de algumas tribos de Papilionoideae, como Desmodieae e Phaseoleae, possivelmente originaram-se a 50-35 *mya* (durante o Eoceno médio e tardio).

Li *et al.* (2013), estudando a diversidade das leguminosas phaseolóides (Papilionoideae) utilizando três marcadores de cloroplasto (*rbcL*, região *trnL-F* e região *trnK/matK*), obtiveram um tempo de divergência estimado de 39,5 *mya* (35,7–43,2 *mya*) para idade de caule e 28,6 *mya* (25,8–31,2 *mya*) para idade de coroa.

Na literatura não existe um consenso acerca do número cromossômico base de Papilionoideae, sendo observado estudos a nível de clados ou tribos. Para Zhao *et al.* (2021) o número base de Papilionoideae é $x = 28$, originário no ancestral dos Genistóides, Dalbergioides e em outros papilionóides, ou até antes no ancestral de Papilionoideae. Para as tribos Dalbergieae e Phaseoleae seguiu um padrão de números cromossômicos de base $x = 10$ (Moraes *et al.* 2020) e 11 (Lackey 1980, Bruneau *et al.* 1995, Espert *et al.* 2008), respectivamente.

Outros estudos também observaram espécies que apresentam comportamento de $x = 20$ ou 40 (Glycines) sendo estes poliplóides estabilizados (Doyle 2012, Lackey 1980, Singh 2019). Doyle (2012) analisando quase 400 gêneros com base no número cromossômico para poliploidia, observou que em um quarto desses gêneros tinha táxons poliplóides, sendo muitos membros de Genisteae, Pterocarpus ($2n = 22, 24, 44$), numerosos Mirbelioides, Caragana ($2n = 16, 24, 32, 48$) e Glycine ($2n = 38, 40$). Para os representantes do gênero *Abrus* (tribo Abreae) apontam para um status tetraplóide com base em $2n = 22$ para *A. preclatorius* e *A. pulchellus*, concordando com um número cromossômico básico de 11 (Agbagwa e Okoli 2020).

A subfamília Papilionoideae é bem suportada nos trabalhos filogenéticos, sendo considerada monofilética (Doyle 1995, Käss e Wink 1997, Doyle *et al.* 2000, Kajita *et al.* 2001, Wojciechowski *et al.* 2000, Kajita *et al.* 2001; Wojciechowski *et al.* 2004, Lavin *et al.* 2005, McMahon e Sanderson 2006, Cardoso *et al.* 2012, Cardoso *et al.* 2013, LPWG 2013, Tozzi 2016). Entretanto, as relações entre vários subclados não estão elucidados dentro do clado de inversão de 50 kb, que inclui a tribo Exostyleae, os clados Andira e vataireoid, os genistoids sl, o dalbergioids sl, e o clado NPAAA (Zhao *et al.* 2021). Dessa forma, várias tribos ou clados em Papilionoideae não são monofiléticos, como observado por Egan *et al.* (2016) para as tribos Millettieae e Phaseoleae.

Na mais recente filogenia o clado NPAAA foi considerado monofilético, com as representantes de Andira e Baphieae como irmãs desse clado (Zhao *et al.* 2021). Dentro deste mesmo clado foi observado a tribo Hypocalypteae como sendo irmã dos demais grupos, ao passo que, sendo os clados Mirbelioides e Hologalegina irmãos de Indigoferae/milettioids. Para o clado Hologalegina foi observado que as tribos Robinieae e Sesbanieae + Loteae são irmãs do clado sem clado sem repetição invertida (Zhang *et al.* 2020, Zhao *et al.* 2021).

Zhao *et al.* (2021) no seu trabalho de filotranscriptômica nuclear e filogenômica comprovam que o clado ADA (representado pelas tribos monofiléticas Angylocalyceae, Dipterygeae e Amburaneae), Swartzieae e Cladrastideae, são irmãs dos papilionóides do clado de inversão de 50 kb. Assim como as linhagens Exostyleae e vataireoid classificadas como irmãs do clado de inversão de 50 kb, que originam dois grandes clados. Sendo um desses dois clados representados por nove tribos divididas em dois subclados de dalbergioides sl (duas tribos) e genistóides sl (sete tribos), com o outro grande clado dividido no clado do gênero *Andira*, tribo Baphieae e no clado NPAAA.

A tribo Indigofereae é suportada como monofilética e irmã do clado Millettoid/Phaseoloid (Hu *et al.* 2000, Kajita *et al.* 2001, Wojciechowski 2003, LPWG 2017, Zhang *et al.* 2020, Zhao *et al.* 2021). Sendo os representantes do clado Millettoid/Phaseoloid os gêneros *Aganope*, *Austrorostenia* e *Dalbergiella* irmãos dos demais Millettoid/Phaseoloid (Zhao *et al.* 2021). Estudos suportam a tribo Abreae como irmã de Diocleae+Millettieae (Queiroz *et al.* 2015, LPWG 2017, Zhao *et al.* 2021).

Dentro do clado NPAAA, a tribo Phaseoleae é parafilético, com as tribos Desmodieae e Psoraleeae aninhadas a ela (Kajita *et al.* 2001, Egan *et al.* 2016, LPWG 2017, Zhao *et al.* 2021). Contudo, as relações entre os representantes do clado Millettoid/Phaseoloid, especialmente a tribo Phaseoleae, ainda possuem uma sucessão de dúvidas (Stefanovic *et al.* 2009, LPWG 2017, Zhao *et al.* 2021).

2.8. INIBIDOR DE PROTEASE DO TIPO BOWMAN-BIRK

Os inibidores de protease (IPs) são reconhecidos como peptídeos ou pequenas proteínas que impedem a atividade proteolítica de enzimas, possibilitando a formação de complexos estequiométricos de elevada afinidade (Bode e Huber 2000). Sendo encontrados de forma abundante em plantas, animais e microrganismos (Lone *et al.* 2021). São produzidos no sistema de defesa endógeno das plantas que se expressam sob condições de estresses bióticos, como ataque de patógenos e insetos; ou abióticos como estresse salino, metais pesados e por variações na temperatura; como também por ferimentos na planta (Krishnan e Murugan 2015).

Esses inibidores são distribuídos em famílias, apoiados por suas estruturas primárias e terciária, e seus mecanismos de inibição (Laskowski e Qasim 2000). Informações disponíveis no banco de dados MEROPS, indicam que os IPs são divididos em 38 grupos e subdivididos em 78 famílias, dependendo de sua especificidade inibitória (Rawlings *et al.* 2018). Dentre essas, uma família bem estudada é a dos inibidores do tipo Bowman-Birk (BBI). Nas plantas, os BBIs são mais abundantemente encontrados em sementes e tubérculos de leguminosas (Fabaceae), Solanaceae, Cucurbitaceae, Poaceae (Connors *et al.* 2002) e Astaraceae (girassol) (Korsinczky *et al.* 2001).

O primeiro BBI foi isolado a partir de sementes de soja (*G. max*) por Donald E. Bowman em 1946 (Bowman 1946), e posteriormente caracterizado por Birk em 1963 (Birk *et al.* 1963). Na literatura os BBIs são evidenciados como uma das famílias de IPs naturais mais bem reconhecidas e caracterizadas (Kennedy 1998, Losso 2008,

Clemente *et al.* 2011, Clemente e Arques 2014).

Os BBIs de dicotiledônea são caracterizados por apresentarem dois domínios homólogos, em que cada um possui uma alça de inserção, uma estrutura simétrica de “duas cabeças”, onde cada domínio contém um sítio reativo de ligação de inibição independente e simultânea de duas serinoproteases de especificidade diferentes (Chen *et al.* 1992, Lin *et al.* 1993, Rao e Suresh 2007, Hellinger e Gruber 2019). Via de regra, o primeiro sítio é ativado para inibir a tripsina, ao passo que o segundo sítio inibi a quimotripsina ou elastase (Qi *et al.* 2005, Lioi *et al.* 2010). Onde sua estrutura tridimensional é estabilizada, principalmente pela presença de sete pontes dissulfeto, associado a um pequeno auxílio do núcleo hidrofóbico da molécula (Chen *et al.* 1992), que aumentam sua estabilidade, com uma estrutura sólida e consistente que se mantém estáveis à pH drásticos e condições de temperaturas extremas (Rao e Suresh 2007, Hellinger e Gruber 2019).

Existe diferença entres os BBIs de dicotiledônea e monocotiledônea. Mello *et al.* (2003), evidenciou que plantas dicotiledônea de diferentes gêneros são constituídos por 14 resíduos de cisteína e sete ligações dissulfeto, ao passo que os BBIs monocotiledôneas possam apresentar de dez a 26 cisteínas, que formam de 5 a 13 ligações com resíduos de cisteína conservados. Esses mesmos autores sugerem que a estrutura de duas cabeças possa ter surgido de um inibidor ancestral de uma cabeça, por meio de uma duplicação genética interna. Análises filogenéticas, também evidenciaram que os BBIs de plantas dicotiledôneas são relativamente próximos.

Na literatura é relatada diversas funções defensivas dos BBIs nas plantas, sendo sua maior refletida na ação inseticida, essa função está intimamente relacionada com sua atividade nas proteases intestinais, que reduz a assimilação de aminoácidos, que possibilita um retardo no desenvolvimento dos insetos que se alimentam dessas plantas (Deepika e Nath 2010). Além disso, alguns BBIs também podem inibir proteases produzidas por patógenos, que possibilita seu efeito antimicrobianos (Kim *et al.* 2009); bem como no desenvolvimento de tolerância ao estresse abiótico, como salino (Shan *et al.* 2008) e hídrico (Dramé *et al.* 2013).

Diversas análises voltadas para função anti-inseto de BBI têm sido relatadas, como o efeito inseticida de dois BBIs isolados a partir de sementes de feijão guandu (*C. cajan*) (Prasad *et al.* 2010a) e *Vigna mungo* (Prasad et a. 2010b) contra diferentes insetos lepidópteros. No amendoim, o BBI (PnBBI) isolado de suas sementes apresentou-se com um efeito inibitório contra uma das pragas mais destrutivas dessa cultura, *Helicoverpa armigera* (Lokya *et al.* 2020).

O isolado de BBI de *Clitoria fairchildiana* favoreceu um efeito inseticida considerável em análise *in vitro* contra enzimas similares a tripsina do intestino médio larval de traça da farinha (*Anagasta kuehniella*), broca da cana-de-açúcar (*Diatraea saccharalis*) e a lagarta das maçãs (*Heliothis virescens*) (Dantzger *et al.* 2015). Também foi descrito o efeito inseticida de BBIs contra pulgões, como o pulgão-da-ervilha (*Acyrtosiphon pisum*) (Rahbé *et al.* 2003) e pulgão verde escuro (*Macrosiphum euphorbiae*) (Azzouz *et al.* 2005). BBIs também exibem atividade inseticida contra broca do café (*Hypothenemus hampei*) (Pereira *et al.* 2007), algodão com o bicudo-do-algodoeiro (*Anthonomus grandis*) (Francoa *et al.* 2003) e *H. armigera* (Kuhar *et al.* 2013), milho com a broca maior do grão de milho (*Prostephanus truncatus*) (Aguirre *et al.* 2004) e o besouro castanho (*Tribolium castaneum*) (Oppert *et al.* 1993).

Assim, o uso dos BBIs pode evidenciar genes para a engenharia genética de plantas com importância econômica e que não produzem esse inibidor, o que configuraria como uma estratégia promissora para produzir plantas transgênicas resistentes a insetos (Clemente *et al.* 2019), como também a patógenos e alguns estresses abióticos. Dessa forma, possibilitaria a redução dos custos, do tempo e os esforços demandados na proteção das culturas contra ao ataque de insetos e patógenos, contribuindo para o meio ambiente mais sustentável.

Além disso, estudos revelam que os BBIs possuem propriedades anti-inflamatórias (Ware *et al.* 1999; Qi *et al.* 2005), propiciada pela inibição ou modulação da atividade de serinoproteases que participam da produção de citocinas inflamatórias (Gitlin-Domagalska *et al.* 2020). Como também, propriedades anticancerígenas (Armstrong *et al.* 2003; Kennedy *et al.* 2002), que está intimamente associada à sua capacidade específica de inibir serina proteases (Gitlin-Domagalska *et al.* 2020). Dessa forma, na literatura é demonstrado que os dois sítios de ligação do BBI, que exibe atividades antitripsina e antiqumotripsina, está ligado ao efeito antiproliferativo nas células cancerígenas do cólon (Clemente *et al.* 2010).

Jacobsen *et al.* (1998) relataram que populações que consomem quantidades mais elevadas de sementes de leguminosas a incidência de câncer de cólon, próstata, mama e pele é moderadamente mais baixo. Essas funções fazem com que os BBIs ofereçam alto potencial farmacêutico, como novos agentes no tratamento do câncer (Lioi *et al.* 2010).

3. REFERÊNCIAS

Adams MA, Turnbull TL, Sprent JI e Buchmann N (2016) Legumes are different: Leaf nitrogen, photosynthesis, and water use efficiency. **Proceedings of the National Academy of Sciences** **113**: 4098 - 4103.

Agbagwa IO e Okoli BE (2020) Chromosome number and cytomorphological characterization of a polyploid *Abrus*. **Frontiers of Agriculture and Food Technology** **10**: 1-4.

Aguirre C, Valdés-Rodríguez S, Mendoza-Hernández G, Rojo-Domínguez A e Blanco-Labra A (2004) A novel 8.7 kDa protease inhibitor from chan seeds (*Hyptis suaveolens* L.) inhibits proteases from the larger grain borer *Prostephanus truncatus* (Coleoptera: Bostrichidae). **Comparative Biochemistry and Physiology B** **138**: 81–89.

Aleixo A, Burlamaqui T, Goncalves E e Schneider P (2006) Molecular systematics of the Ocellated Woodcreeper complex (Dendrocolaptidae) in tropical South America: Implications for taxonomy, conservation, and historical biogeography. **Journal of Ornithology** **147**: 125-126.

Anderson SJ, Kramer MC, Gosai SJ, Yu X, Vandivier LE, Nelson ADL, Anderson ZD, Beilstein MA, Fray RG, Lyons E e *et al.* (2018) N6-Methyladenosine Inhibits Local Ribonucleolytic Cleavage to Stabilize mRNAs in Arabidopsis. **Cell Reports** **25**: 1146–1157.

Armstrong WB, Wan XS, Kennedy AR e Taylor TH (2003) Development of the Bowman-Birk inhibitor for oral cancer chemoprevention and analysis of Neu immunohistochemical staining intensity with Bowman-Birk inhibitor concentrate treatment. **Laryngoscope** **113**: 1687–1702.

Arroyo MTK (1981) Breeding systems and pollination biology in Leguminosae. In: Polhill RM e Raven PH (eds) **Advances in legume systematics**. Kew: Royal Botanic Gardens, United Kingdom, p. 723-769.

Avise JC (2009) Phylogeography: retrospect and prospect. **Journal of Biogeography** **36**: 3-15.

Azzouz H, Cherqui A, Campan EDM, Rahbé Y, Duport G, Jouanin L, Kaiser L e Giordanengo P (2005) Effects of plant protease inhibitors, oryzacystatin I and soybean

SILVA, F. G. Filogenia molecular e análise *in silico* do inibidor de protease Bowman-Birk de espécies da subfamília Papilionoideae

Bowman-Birk inhibitor, on the aphid *Macrosiphum euphorbiae* (Homoptera, Aphididae) and its parasitoid *Aphelinus abdominalis* (Hymenoptera, Aphelinidae). **Journal of Insect Physiology** **51**: 75–86.

Barroso LR (2004) **Sistemática de Angiospermas do Brasil**. UFV, Minas Gerais, 310p.

Bayat A (2002) Science, medicine, and the future Bioinformatics. **The BMJ** **324**: 1018-1022.

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Rapp BA e Wheeler DL (2000) GenBank. **Nucleic Acids Research** **28**:15-18.

Berry PE, Duane I e Turner BL (2018) "**Fabales**". Encyclopedia Britannica, <https://www.britannica.com/plant/Fabales>. Accessed 10 November 2021.

Birk Y, Gertler A e Khalef S (1963) A pure trypsin inhibitor from soya beans. **Biochemical Journal** **87**: 281–284.

Bock R (2007) Structure, function, and inheritance of plastid genomes. In: Bock R (ed) **Cell and Molecular Biology of Plastids**. Göteborg University, Göteborg, p. 29-64.

Bode W e Huber R (2000) Structural basis of the endoproteinase–protein inhibitor interaction. *Biochimica et biophysica acta*. **Protein structure and molecular enzymology** **1477**: 241–252.

Bowman DE (1946) Differentiation of Soy Bean Antitryptic Factors. **Experimental Biology and Medicine** **63**: 547–550.

Bunsupa S, Saito K e Yamazaki M (2013) Molecular biology and biotechnology of quinolizidine alkaloid biosynthesis in Leguminosae plants. In: Chandra S, Lata H e Varma A (eds) **Biotechnology for Medicinal Plants**. Springer, Berlin, Heidelberg, p. 263–273.

Çakir Ö, Uçarlı C, Tarhan C, Pekmez M e Turgut-Kara N (2018) Nutritional and health benefits of legumes and their distinctive genomic properties. **Food Science and Technology** **39**: 1-12.

Can T (2014) Introduction to Bioinformatics. In: Yousef M e Allmer J (eds) **miRNomics: MicroRNA Biology and Computational Analysis**, Methods in Molecular Biology.

Springer, New York, p. 51-71.

Cannon SB, Michael R, McKain MR, Harkess A, Nelson MN, Dash S e *et al.* (2015) Multiple Polyploidy Events in the Early Radiation of Nodulating and Nonnodulating Legumes. **Molecular Biology and Evolution** **32**: 193–210.

Cardoso D, Pennington RT, Queiroz LP, Boatwright JS, Van Wyk BE, Wojciechowski MF e Lavin M (2013) Reconstructing the deep-branching relationships of the papilionoid legumes. **South African Journal of Botany** **89**: 58–75.

Cardoso D, Queiroz LP, Pennington RT, Lima HC, Fonty E, Wojciechowski MF e Lavin M. 2012. Revisiting the phylogeny of papilionoid legumes: New insights from comprehensively sampled early-branching lineages. **American Journal of Botany** **99**: 1991–2013.

Cavalier-Smith T (2000) Membrane heredity and early chloroplast evolution. **Trends in Plant Science** **5**: 174-182.

Chan KX, Phua SY, Crisp P, McQuinn R e Pogson BJ (2016) Learning the languages of the chloroplast: retrograde signaling and beyond. **Annual Review of Plant Biology** **67**: 25–53.

Chang Y, Liu H, Liu M, Liao X, Sahu SK e *et al.* (2018). The draft genomes of five agriculturally importante African orphan crops. **GigaScience** **8**: 1–16.

Chawla S, Sangwan S, Sharma R (2016) Statistical Significance of Biological DataMarkets. **International Journal of Engineering Science and Computing** **6**: 7585-7588.

Chen P, Rose P, Love R, Wei CH, Wang BC (1992) Reactive sites of na anticarcinogenic Bowman–Birk proteinase inhibitor are similar to Other trypsin inhibitors. **Journal of Biological Chemistry** **267**: 1990–1994.

Choi H, Yi T e Ha S-H (2021) Diversity of Plastid Types and Their Interconversions. **Frontiers in Plant Science** **12**: 692024.

Chumley TW, Palmer JD e Mower JP (2006) The complete chloroplast genome sequence of *Pelargonium x hortorum*: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. **Molecular Biology and**

Evolution 23: 2175-2190.

Clemente A e Arques MDC (2014) Bowman-Birk inhibitors from legumes as colorectal chemopreventive agentes. **World Journal of Gastroenterology 20:** 10305–10315.

Clemente A, Moreno FJ, Marín-Manzano MDC, Jiménez E e Domoney C (2010) The cytotoxic effect of Bowman-Birk isoinhibitors, IBB1 and IBBD2, from soybean (*Glycine max*) on HT29 human colorectal câncer cells is related to their intrinsic ability to inhibit serine proteases. **Molecular Nutrition & Food Research 54:** 396–405.

Clemente A, Sonnante G e Domoney C (2011) Bowman-Birk inhibitors from legumes and human gastrointestinal health: Current status and perspectives. **Current Protein & Peptide Science 12:** 358–373.

Clemente M, Corigliano MG, Pariani SA, Sánchez-López EF, Sander VA e Ramos-Duarte VA (2019) Plant serine protease inhibitors: Biotechnology application in agriculture and molecular farming. **International Journal of Molecular Sciences 20:** 1345.

Cocking EC, Stone PJ e Davey, M. R. (2006). Intracellular colonization of roots of *Arabidopsis* and crop plants by *Gluconacetobacter diazotrophicus*. **In Vitro Cellular & Developmental Biology 42:** 74–82.

Connors BJ, Laun NP, Maynard CA e Powell WA (2002) Molecular characterization of a gene encoding a cystatin expressed in the stems of American chestnut (*Castaneadentata*). **Planta 215:** 510–514.

Crosby R e Williams T (2017) Fast algorithms for computing phylogenetic divergence time. **BMC Bioinformatics 18:** 514.

Cullen J, Knees SG e Cubey HS (2011). **The european garden flora**. Cambridge University Press, Cambridge, 652p.

Daniell H, Lin CS, Yu M e Chang WJ (2016) Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. **Genome Biol 17:** 1-29.

Dantzger M, Vasconcelos IM, Scorsato V, Aparicio R, Marangoni S e Macedo MLR (2015) Bowman-Birk proteinase inhibitor from *Clitoria fairchildiana* seeds: Isolation, biochemical properties and insecticidal potential. **Phytochemistry 118:** 224–235.

SILVA, F. G. Filogenia molecular e análise *in silico* do inibidor de protease Bowman-Birk de espécies da subfamília Papilionoideae

Davis JI e Soreng RJ (2010) Migration of endpoints of two genes relative to boundaries between regions of the plastid genome in the grass family (Poaceae). **American Journal of Botany** **97**: 874-892.

Deepika R e Nath AK (2010) Studies on Trypsin Inhibitor in Bean (*Phaseolus vulgaris* L) Cultivars of Himalayan Region. **Journal of Plant Biochemistry & Biotechnology** **19**: 103-105.

Dent DR (2018) Non-nodular Endophytic Bacterial Symbiosis and the Nitrogen Fixation of *Gluconacetobacter diazotrophicus*. In: Rigobelo EC (ed) **Symbiosis**. IntechOpen, London, p. 53–81.

Dhingra A e Folta KM (2005) ASAP: Amplification, sequencing & annotation of plastomes. **BMC Genomics** **6**: 1-13.

Dilworth MJ, James EK, Sprent JI e Newton WE (2008) **Nitrogen Fixation: Origins, Applications and Research Progress**. Springer, Berlin, 403p.

Doyle JJ (2012) Polyploidy in legumes. In: Soltis PS e Soltis DE (eds) **Polyploidy and Genome Evolution**. Springer, Berlin, p. 147–180.

Doyle JJ (2016) Chasing unicorns: nodulation origins and the paradox of novelty. **American journal of botany** **103**: 1865–1868.

Doyle, JJ, Doyle, JL, Ballenger, JA e Palmer, JD (1996). A distribuição e significado filogenético de uma inversão de DNA de cloroplasto de 50 kb na família de plantas com flores Leguminosae. **Molecular Phylogenetics and Evolution** **5**: 429-438.

Dramé KN, Passaquet C, Repellin A e Zuily-Fodil Y (2013) Cloning, characterization and differential expression. of a Bowman-Birk inhibitor during progressive water deficit and subsequent recovery in peanut (*Arachis hypogaea*) leaves. **Journal of Plant Physiology** **170**: 225–229.

Duane I e Paul EB (2012). Origin of Fabales and its relationship with other plant families. Encyclopaedia Britannica Inc. Retrieved on 2012 Oct 12 from <http://www.britannica.com/EBchecked/topic/199654/Fabales/72896/Classification-of-Fabaceae>.

Dugas DV, Hernandez D, Koenen EJM, Schwarz E, Straub S, Hughes CE e *et al.*

SILVA, F. G. Filogenia molecular e análise *in silico* do inibidor de protease Bowman-Birk de espécies da subfamília Papilionoideae

(2015). Mimosoid legume plastome evolution: IR expansion, tandem repeat expansions, and accelerated rate of evolution in *clpP*. **Scientific Reports** 5:16958.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. **Nucleic Acids Research** 32: 1792-7.

Edwards AWF e Cavalli-Sforza LL (1964) Reconstruction of evolutionary trees. In: Heywood V e McNeill J (eds) **Phenetic and phylogenetic classification**. Systematic Association Publication, London, p. 67-76.

Edwards D e Batley J (2004) Plant bioinformatics: from genome to phenome. **Trends in Biotechnology** 22: 232-237.

Egan AN, Vatanparast M e Cagle W (2016). Parsing polyphyletic *Pueraria*: delimiting distinct evolutionary lineages through phylogeny. **Molecular Phylogenetics and Evolution** 104: 44–59.

Emygdio BM, Antunes IF, Choer E e Nedel JL (2003) Eficiência de coeficientes de similaridade em genótipos de feijão mediante marcadores RAPD. **Pesquisa Agropecuária Brasileira** 38: 243-250.

Farris JS (1983) The logical basis of phylogenetic analysis. In Platnick N e Funk V (eds) **Advances in cladistics**. Columbia University Press: New York, p. 7-36.

Felsenstein J (2004) **Inferring phylogenies**. Sinauer Associates Incorporated, Sunderland, 664p.

Francoa OL, Santos RC, Batista JAN, Mendes ACM, Araújo MAM, Monnerat RG, Fátima Grossi-de-Sá, M e Freitas SM (2003) Effects of black-eyed pea trypsin/chymotrypsin inhibitor on proteolytic activity and on development of *Anthonomus grandis*. **Phytochemistry** 63: 343–349.

Fu CN, Li HT, Milne R, Zhang T, Ma PF, Yang J, Li DZ e Gao LM. 2017. Comparative analyses of plastid genomes from fourteen *Cornales* species: inferences for phylogenetic relationships and genome evolution. **BMC Genomics** 18: 1-14.

Gitlin-Domagalska A, Maciejewska A e Dębowski D (2020) Bowman-Birk Inhibitors: Insights into Family of Multifunctional Proteins and Peptides with Potential Therapeutical Applications. **Pharmaceuticals** 13: 421.

SILVA, F. G. Filogenia molecular e análise *in silico* do inibidor de protease Bowman-Birk de espécies da subfamília Papilionoideae

Gompert Z, Nice CC, Fordyce JA, Forister ML e Shapiro AM (2006) Identifying units for conservation using molecular systematics: the cautionary tale of the Karner blue butterfly. **Molecular Ecology** **15**: 1759-1768.

Graur D e Li W (2000) **Fundamentals of Molecular Evolution**. Sinauer Associates, Sunderland, 479p.

Green BR (2011) Chloroplast genomes of photosynthetic eukaryotes. **The Plant Journal: for cell and molecular biology** **66**: 34-44.

Gurcharan S (2004) **Plant systematics: An integrated approach**. Science Publishers, New Hampshire, 702p.

Hellinger R e Gruber CW (2019) Peptide-based protease inhibitors from plants. **Drug Discovery Today** **24**: 1877–1889.

Hu JM, Lavin M, Wojciechowski MF e Sanderson MJ (2000). Phylogenetic systematics of the tribe Millettieae (Leguminosae) based on chloroplast *trnK/matK* sequences and its implications for evolutionary patterns in Papilionoideae. **American Journal of Botany** **87**: 418–430.

Huang YY, Matzke AJ e Matzke M (2013) Complete sequence and comparative analysis of the chloroplast genome of coconut palm (*Cocos nucifera*). **PLoS One** **8**: 1-12.

Huelsenbeck JP, Ronquist F, Nielsen R e Bollback JP (2001) Bayesian inference of phylogeny and its impact on evolutionary biology. **Science** **294**: 2310-2314.

Huo Y, Wang M, Wei Y, Xia Z (2016) Over expression of the Maize *psbA* Gene Enhances Drought Tolerance Through Regulating Antioxidant System, Photosynthetic Capability, and Stress Defense Gene Expression in Tobacco. **Frontiers in Plant Science** **6**:1223.

International Legume Database & Information Service (ILDIS) (2005). **Sub-family Papilionoideae**. Version 10.01, Retrieved on 2014 Sept 2 from <http://www.ildis.org/>.

Jacobsen BK, Knutsen SF e Fraser GE (1998) Does high soy milk intake reduce prostate cancer incidence? The Adventist Health Study (United States). **Cancer Causes Control** **9**: 553-7.

SILVA, F. G. Filogenia molecular e análise *in silico* do inibidor de protease Bowman-Birk de espécies da subfamília Papilionoideae

Jansen RK, Cai Z, Raubeson LA, Daniell H, Pamphilis CW, Leebensmack J, Muller KF, Guisinger-Bellian M, Haberle RC, Hansen AK, Chumley TW, Lee SB, Peery R, Mcneal JR, Kuehl JVE e Janzantti NS, Machado TV e Monteiro M (2011) Sensory acceptance of juice from FCOJ processing steps. **Journal of Sensory Studies** **26**: 322-330.

Jarvis P e López-Juez E (2014). Biogenesis and homeostasis of chloroplasts and other plastids. **Nature Reviews Molecular Cell Biology** **15**: 147–147.

Jimenez HJ, Martins LSS, Carvalho R, Montarroyos AVV e Moraes Filho RM (2019) In silico characterization and phylogenetic analysis of a mannose-specific lectin in *Allium* species. **Genetics and Molecular Research** **18**: 1-15.

Jimenez HJ, Silva ADF, Martins LSS, Carvalho R e Moraes Filho RM (2020) Comparative genomics plastomes of the Amaryllidaceae family species. **Scientia Plena** **16**: 1-11.

Judd W, Campbell CS, Kellogg EA, Stevens PF e Donoghue MJ (2009) **Sistemática vegetal: Um enfoque Filogenético**. Artmed, Porto Alegre, 632p.

Judd WS, Campbell CS, Kellogg EA, Stevens PFE e Donoghue MJ (2002) **Plant systematics: a phylogenetic approach**. Sinauer Associates, Sunderland, 576p.

Kagale S e Close TJ (2021) Legumes: Embracing the genome era. **Legume Science** **3**: e113.

Kajita T, Ohashi H, Tateishi Y, Bailey CD e Doyle JJ (2001) *rbcL* and legume phylogeny, with particular reference to Phaseoleae, Millettieae, and allies. **Systematic Botany** **26**: 515–536.

Kamphuis LG, Garg G, Foley R e Singh KB (2021) Genomic resources for lupins are coming of age. **Legume Science** **3**: 1–15.

Käss E e Wink M (1996) Molecular evolution of the Leguminosae: phylogeny of the three subfamilies based on *rbcL*-sequences. **Biochemical Systematics and Ecology** **24**: 365–378.

Keeling PJ (2013). The number, speed, and impact of plastid endosymbioses in eukaryotic evolution. **Annual Review of Plant Biology** **64**: 583–607.

Kennedy AR (1998) Chemopreventive agents: Protease inhibitors. **Pharmacology & Therapeutics 78**: 167–209.

Kennedy AR, Billings PC, Wan XS e Wan PM (2002) Effects of Bowman-Birk inhibitor on rat colon carcinogenesis. **Nutrition and Cancer 43**: 174–186.

Khan A, Asaf S, Khan AL, Al-Harrasi A, Al-Sudairy O, AbdulKareem NM e *et al.* (2019) First complete chloroplast genomics and comparative phylogenetic analysis of *Commiphora gileadensis* and *C. foliacea*: Myrrh producing trees. **PLoS ONE 1**: 14.

Kim YK, Park CW e Kim KJ (2009) Complete chloroplast DNA sequence from a Korean endemic genus, *Megaleranthus saniculifolia*, and its evolutionary implications. **Molecules and Cells 27**: 365–381.

Koenen EJM, Ojeda DI, Bakker FT, Wieringa JJ, Kidner C, Hardy OJ, Pennington RT, Herendeen PS, Bruneau A e Hughes CE (2021) The origin of the legumes is a complex paleopolyploid phylogenomic tangle closely associated with the Cretaceous-Paleogene (K-Pg) mass extinction event. **Systematic Biology 70**: 508–526.

Koenen EJM, Ojeda DI, Steeves R, Migliore J, Bakker FT e *et al.* (2020) Large-scale genomic sequence data resolve the deepest divergences in the legume phylogeny and support a nearsimultaneous evolutionary origin of all six subfamilies. **New Phytologist 225**: 1355–1369.

Kouris-Blazos A e Belski R (2016) Health benefits of legumes and pulses with a focus on Australian sweet lupins. **Asia Pacific Journal of Clinical Nutrition 25**: 1-17.

Kreplak J, Madoui M-A, Cápál P, Novák P, Labadie K e *et al.* (2019) A reference genome for pea provides insight into legume genome evolution. **Nature Genetics 51**: 1411–1422.

Krishnan VGM e Murugan K (2015) Purification, characterization and kinetics of protease inhibitor from fruits of *Solanum aculeatissimum* Jacq. **Food Science and Human Wellness 4**: 97–107.

Kuhar K, Kansal R, Subrahmanyam B, Koundal KR, Miglani K e Gupta VK (2013) A Bowman-Birk protease inhibitor with antifeedant and antifungal activity from *Dolichos biflorus*. **Springer 35**: 1887–1903.

SILVA, F. G. Filogenia molecular e análise *in silico* do inibidor de protease Bowman-Birk de espécies da subfamília Papilionoideae

Kumar S, Stecher G e Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. **Molecular Biology and Evolution** **33**: 1870-1874.

Kumar S, Tamura K e Nei M (2004) MEGA3: an integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. **Briefings In Bioinformatics** **5**: 150-163.

Lackey JA (1980) Chromosome numbers in the phaseoleae (fabaceae:faboideae) and their relation to taxonomy. **American Journal of Botany** **4**: 595-602. 1980.

Laskowski M e Qasim MA (2000) What can the structures of enzyme-inhibitor complexes tell us about the structures of enzyme substrate complexes?. **Biochimica et Biophysica Acta** **1477**: 324–337.

Lavin M, Herendeen PS e Wojciechowski MF (2005) Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the Tertiary. **Systematic Biology** **54**: 575–594.

Leebens-Mack JH, Barker MS, Carpenter EJ, Deyholos MK, Gitzendanner MA, Graham SW, Grosse I, Li Z, Melkonian M, Mirarab S e *et al.* (2019) One thousand plant transcriptomes and the phylogenomics of green plants. **Nature** **574**: 679–685.

Lemey P, Salemi M, Vandamme A-M (2009) **The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing**. Cambridge University Press, Cambridge, 750p.

Lesk AM (2008) **Introduction to Bioinformatics**. Oxford University Press, New York, 432p.

Lewis G, Schrire B, Mackinder B, Lock M (2005) **Legumes of the World**. Kew, Royal Botanic Gardens, London, 577p.

Li Q-G, Zhang L, Li C, Dunwell JM e Zhang Y-M (2013) Comparative genomics suggests that an ancestral polyploidy event leads to enhanced root nodule symbiosis in the Papilionoideae. **Molecular Biology and Evolution** **30**: 2602–2611.

Lioi L, Galasso I, Daminati MG e Piergiovanni AR (2010) Inhibitory properties and binding loop polymorphism in Bowman-Birk inhibitors from Phaseolus species.

Genetic Resources and Crop Evolution 57: 533–542.

Liu Y, Zhang J, Li W, Guo C e Shu Y (2016) In silico identification, phylogeny and expression analysis of expansin superfamily in *Medicago truncatula*. **Biotechnology & Biotechnological Equipment 30**: 197-203.

Liu E, Yang C, Liu J, Jin S, Harijati N, Hu Z, Diao Y e Zhao L (2019) Comparative analysis of complete chloroplast genome sequences of four major *Amorphophallus* species. **Scientific Reports 9**: 809.

Lokya V, Swathi M, Mallikarjuna N e Padmasree K (2020) Response of Midgut Trypsin and Chymotrypsin Like Proteases of *Helicoverpa armigera* Larvae Upon Feeding With Peanut BBI: Biochemical and Biophysical Characterization of PnBBI. **Frontiers in Plant Science 11**: 266.

Lonardi S, Muñoz-Amatriaín M, Liang Q, Shu S, Wanamaker SI e *et al.* (2019). The genome of cowpea (*Vigna unguiculata* [L.] Walp.). **The Plant Journal 98**: 767–782.

Lone JK, Lekha MA, Bharadwaj RP, Ali F, Pillai MA, Wani SH, Yasin JK e Chandrashekharaiiah KS (2021) Multimeric Association of Purified Novel Bowman-Birk Inhibitor From the Medicinal Forage Legume *Mucuna pruriens* (L.) DC. **Frontiers in Plant Science 12**: 772046.

Losso JN (2008) The biochemical and functional food properties of the Bowman-Birk inhibitor. **Critical Reviews in Food Science and Nutrition 48**: 94–118.

Loureiro J, Trávníček P, Rauchová J, Urfus T, Vít P, Štech M, Castro S e Suda J (2010) The use of flow cytometry in the biosystematics, ecology and population biology of homoploid plants. **Preslia 82**: 3–21.

LPWG (Legume Phylogeny Working Group) (2013) Legume phylogeny and classification in the 21st century: progress, prospects and lessons for other species-rich clades. **Taxon 62**: 217–248.

LPWG (Legume Phylogeny Working Group) (2017) A new subfamily classification of the Leguminosae based on a taxonomically comprehensive phylogeny. **Taxon 66**: 44–77.

Maira N, Torres TM, Oliveira AL, Medeiros SRB, Agnez LF, Lima JPMS e Scortecci

KC (2013) Identification, characterisation and molecular modelling of two AP endonucleases from base excision repairs pathway in sugarcane provide insights on the early evolution of green plants. **Plant biology** **16**: 622-631.

Marouelli LP (2009) **Análise filogenética de acessos do gênero *Heliconia* L. (Heliconiaceae) utilizando marcadores moleculares.** 2009. 88f. Dissertação (Mestrado em Botânica) - Universidade de Brasília, Brasília.

Martínez-Pérez M, Aparicio F, López-Gresa MP, Bellés JM, Sánchez-Navarro JA e Pallás V (2017) Arabidopsis m6A demethylase activity modulates viral infection of a plant virus and the m6A abundance in its genomic RNAs. **Proceedings of the National Academy of Sciences USA** **114**: 10755–10760.

McMahon MM e Sanderson MJ (2006) Phylogenetic supermatrix analysis of GenBank Sequences from 2228 Papilionoid Legumes. **Systematic Biology** **55**: 818–836.

Menezes APA, Resende-Moreira LC, Buzatti RSO, Nazareno AG, Carlsen M, Lobo FP, *et al.* (2018) Chloroplast genomes of *Byrsonima* species (Malpighiaceae): Comparative analysis and screening of high divergence sequences. **Scientific Reports** **8**: 1-12.

Moraes Filho R e Martins LSS (2016) In silico comparative analysis of tylenchid nematodes pectate lyases. **Genetics and Molecular Research** **15**:1-14.

Moraes AP, Vatanparast M, Polido C, Marques A, Souza G, Fortuna-Perez AP e Forni-Martins ER (2020) Chromosome number evolution in dalbergioid legumes (Papilionoideae, Leguminosae). **Brazilian Journal of Botany** **43**: 575–587.

Morais DAL (2003) **Análise bioinformática de genes de resistência à patógenos no genoma expresso da cana-de-açúcar.** 2003. 217f. Tese (Doutorado em Biotecnologia/Biologia Celular e Molecular) Universidade Federal de Pernambuco, Recife.

Morton CM e Telmer C (2014) New subfamily classification for Rutaceae. **Annals of the Missouri Botanical Garden** **99**: 620–641.

Nei M e Kumar S (2000) **Molecular evolution and phylogenetics.** Oxford University Press, New York, 333p.

Nishikawa T, Salomon B, Komatsuda T, Von Bothmer R e Kadowaki K (2002) Molecular phylogeny of the genus *Hordeum* using three chloroplast DNA sequences. **Genome** **45**: 1157-1166.

Oppert B, Morgan TD, Culbertson C e Kramer KJ (1993) Dietary mixtures of cysteine and serine proteinase inhibitors exhibit synergistic toxicity toward the red flour beetle, *Tribolium castaneum*. *Comparative Biochemistry and Physiology Part C: Pharmacology* **105**: 379–385.

Patil RG, Jadhao KR, Samal KC e Rout GR (2015) Molecular phylogeny of Indian indigenous aromatic rice based on sequence diversity of the chloroplast-encoded *matK* gene. **Rice Genomics and Genetics** **6**: 1-8.

Pellegrini AFA, Staver CA, Hedin LO, Charles-Dominique T e Tourgee A (2016) Aridity, not fire, favors nitrogen- fixing plants across tropical savanna and forest biomes. **Ecology** **97**: 2177–2183.

Penjor T, Yamamoto M, Uehara M, Ide M, Matsumoto N, Matsumoto R e Nagano Y (2013) Phylogenetic relationships of *Citrus* and its relatives based on *matK* gene sequences. **PloS one** **8**: 1-13.

Pereira RDA, Valencia-Jiménez A, Magalhães CP, Prates MV Melo JAT, Lima LM, Sales MP, Nakasu EYT, *et al.* (2007) Effect of a Bowman-Birk proteinase inhibitor from *Phaseolus coccineus* on *Hypothenemus hampei* gut proteinases *in vitro*. **Journal of Agricultural and Food Chemistry** **55**: 10714–10719.

Pfeil BE e Crisp MD (2008) The age and biogeography of *Citrus* and the orange subfamily (Rutaceae: Aurantioideae) in Australasia and New Caledonia. **American Journal of Botany** **95**:1621-1631.

Polhill RM, Raven PH e Stirton CH (1981) Evolution and systematics of the Leguminosae. In Polhill RM e Raven PH (eds) **Advances in legume systematics**. Royal Botanic Gardens, Kew, London, p. 1–26.

Prasad ER, Dutta-Gupta A e Padmasree K (2010b) Purification and characterization of a Bowman-Birk proteinase inhibitor from the seeds of black gram (*Vigna mungo*). **Phytochemistry** **71**: 363–372.

Prasad ER, Merzendorfer H, Madhurarekha C, Dutta-Gupta A, Padmasree K (2010a)

Bowman-birk proteinase inhibitor from *Cajanus cajan* seeds: Purification, characterization, and insecticidal properties. **Journal of Agricultural and Food Chemistry** **58**: 2838–2847.

Prieto P (2020) Chromosome Manipulation for Plant Breeding Purposes. **Agronomy** **10**: 1695.

Provan J e Bennett K (2008) Phylogeographic insights into cryptic glacial refugia. **Trends in Ecology and Evolution** **23**: 564–571.

Puozaa DK, Jaiswal SK e Dakora FD (2019) Phylogeny and distribution of Bradyrhizobium symbionts nodulating cowpea (*Vigna unguiculata* L. Walp) and their association with the physicochemical properties of acidic African soils. **Systematic and Applied Microbiology** **42**: 403-414.

Püschel D, Janoušková M, Voříšková A, Gryndlerová H, Vosátka M e Jansa J (2017) Arbuscular Mycorrhiza Stimulates Biological Nitrogen Fixation in Two *Medicago* spp. through Improved Phosphorus Acquisition. **Frontiers in Plant Science** **8**: 390.

Qi RF, Song ZW, Chi CW (2005) Structural features and molecular evolution of Bowman-Birk protease inhibitors and their potential application. **Acta Biochimica et Biophysica Sinica** **7**:283-92.

Queiroz LP, Pastore JF, Cardoso D, Snak C, Lima ALC, Gagnon E, Vatanparast M, Holland AE e Egan AN (2015) A multilocus phylogenetic analysis reveals the monophyly of a recircumscribed papilionoid legume tribe Diocleae with wellsupported generic relationships. **Molecular Phylogenetics and Evolution** **90**:1–19.

Rahbé Y, Ferrasson E, Rabesona H e Quillien L (2003) Toxicity to the pea aphid *Acyrtosiphon pisum* of anti-chymotrypsin isoforms and fragments of Bowman-Birk protease inhibitors from pea seeds. **Insect Biochemistry and Molecular Biology** **33**: 299–306.

Raven P, Evert RF e Eichhorn SE (2001) **Biologia Vegetal**. Guanabara-Koogan, Rio de Janeiro, 906p.

Rawlings ND, Barrett AJ, Thomas PD, Huang X, Bateman A e Finn RD (2018) The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. **Nucleic Acids Research**

46: D624–D632.

Ridley, M (2006). **Evolução**. Artmed, Porto Alegre, 752p.

Rogalski M, Vieira LN, Fraga HP e Guerra MP (2015) Plastid genomics in horticultural species: importance and applications for plant population genetics, evolution, and biotechnology. **Frontiers in Plant Science** 6: 586.

Santos CMD (2008) Os dinossauros de Hennig: sobre a importância do monofiletismo para a sistemática biológica. **Scientia Studia** 6: 179-200.

Sayers EW, Cavanaugh M, Clark K, Pruitt KD, Schoch CL, Sherry ST e Karsch-Mizrachi I (2020) GenBank. **Nucleic Acids Research** 50: 161–164.

Scarcelli N, Barnaud A, Eiserhardt W, Treier UA, Seveno M, D'Anfray A, Vigouroux Y e Pintaud J-C (2011) A set of 100 chloroplast DNA primer pairs to study population genetics and phylogeny in monocotyledons. **PLoS ONE** 6:1-8.

Schmutz J, McClean PE, Mamidi S, Wu GA, Cannon SB e *et al.* (2014) A reference genome for common bean and genome-wide analysis of dual domestications. **Nature Genetics** 46: 707–713.

Schneider H (2007) **Métodos de Análise Filogenética: Um Guia Prático**. 3ª ed, Holos Editora Sociedade Brasileira de Genética, Ribeirão Preto, 200p.

Ševčíková T, Horák A, Klimeš V, Zbránková V, Demir-Hilton E, Sudek S e *et al.* (2015). Updating algal evolutionary relationships through plastid genome sequencing: did alveolate plastids emerge through endosymbiosis of an ochrophyte?. **Scientific Reports** 5: 10134.

Shaw J, Lickey E, Schilling E e Small R (2007) Comparison of whole chloroplast genome sequences to choose non-coding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. **American Journal of Botany** 94: 275-288.

Shukla L, Chinnusamy V e Sunkar R (2008) The role of microRNAs and other endogenous small RNAs in plant stress responses. **Biochimica et Biophysica Acta - Gene Regulatory Mechanisms** 1779: 743–748.

Signorelli S Sainz M, Rosa ST e Monza J (2020) The Role of Nitric Oxide in Nitrogen Fixation by Legumes. **Frontiers Plant Science** 11: 521.

SILVA, F. G. Filogenia molecular e análise *in silico* do inibidor de protease Bowman-Birk de espécies da subfamília Papilionoideae

Singh RM (2019) Cytogenetics and genetic introgression from wild relatives in soybean. **Nucleus** **62**: 3–14.

Smýkal P e Konečná E (2014) Advances in pea genomics. In Gupta S, Nadarajan N e Gupta D (eds) **Legumes in the Omic Era**. Springer, New York.

Smýkal P, Coyne CJ, Ambrose MJ, Maxted N, Schaefer H e *et al.* (2015) Legume Crops Phylogeny and Genetic Diversity for Science and Breeding. **Plant Sciences** **34**:1-3, 43-104.

Sprent JI (2009) Legume Nodulation: A Global Perspective. Wiley-Blackwell, Oxford, UK, 200p.

Stai JS, Yadav A, Sinou C, Bruneau A, Doyle JJ, Fernández-Baca D e Cannon SB (2019) Cercis: a non-polyploid genomic relic within the generally polyploid legume family. **Frontiers in Plant Science** **10**: 345.

Stefanović, S., Pfeil, B.E., Palmer, J.D., and Doyle, J.J. (2009). Relationships among Phaseoloid legumes based on sequences from eight chloroplast regions. **Systematic Botany** **34**: 115–128.

Tamura K, Dudley J, Nei M e Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. **Molecular Biology and Evolution** **24**: 1596-9.

Tamura K, Battistuzzi FU, Billing-Ross P, Murillo O, Filipowski A e Kumar S (2012) Estimating Divergence Times in Large Molecular Phylogenies. **Proceedings of the National Academy of Sciences** **109**: 19333-19338.

Tozzi AMGA (2016) Papilionoideae. In Tozzi AMGA, Melhem TS, Forero E, Fortuna-Perez AP, Wanderley MGL, Martins SE, Romanini RP, Pirani JR, Fiuza de Melo MMR, Kirizawa M, Yano O, Cordeiro I (eds) **Flora Fanerogâmica do Estado de São Paulo**. Instituto de Botânica, SP, p. 167-397.

Tucker S C (2003) Floral development in legumes. **Plant Physiology** **131**: 911-926.

Unkovich M (2013) Isotope discrimination provides new insight into biological nitrogen fixation. **New Phytologist** **198**: 643–646.

Varshney RK, Chen W, Li Y, Bharti AK, Saxena RK e *et al.* (2012). Draft genome

SILVA, F. G. Filogenia molecular e análise *in silico* do inibidor de protease Bowman-Birk de espécies da subfamília Papilionoideae

sequence of pigeonpea (*Cajanus cajan*), na orphan legume crop of resource-poor farmers. **Nature Biotechnology** **30**: 83–89.

Varshney RK, Song C, Saxena RK, Azam S, Yu S e *et al.* (2013). Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. **Nature Biotechnology** **31**: 240–246.

Varshney RK, Thudi M, Pandey MK, Tardieu F e *et al.* (2018) Accelerating Genetic Gains in Legumes for the Development of Prosperous Smallholder Agriculture: Integrating Genomics, Phenotyping, Systems Modelling and Agronomy. **Journal of Experimental Botany** **69**: 3293-3312.

Wang YH, Qu XJ, Chen SY, Li DZ e Yi TS (2017) Plastomes of Mimosoideae: structural and size variation, sequence divergence, and phylogenetic implication. **Tree Genetics and Genomes** **13**: 41.

Ware JH, Wan XS, Newberne P e Kennedy AR (1999) Bowman-Birk inhibitor concentrate reduces colon inflammation in mice with dextran sulfate sodium-induced ulcerative colitis. **Digestive Diseases and Sciences** **44**: 986-990.

Wicke S, Schneeweiss GM, Pamphilis CW, Müller KF e Quandt D (2011) The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. **Plant Molecular Biology** **76**: 273–297.

Wink M e Mohamed GIA (2003) Evolution of chemical defense traits in the Leguminosae: mapping of distribution patterns of secondary metabolites on a molecular phylogeny inferred from nucleotide sequences of the *rbcL* gene. **Biochemical Systematics and Ecology** **31**: 897–917.

Wojciechowski MF, Lavin M e Sanderson MJ (2004) A phylogeny of legumes (Leguminosae) based on analysis of the plastid *matK* gene resolves many well-supported subclades within the family. **American Journal of Botany** **11**: 1846–1862.

Wojciechowski MF, Sanderson MJ, Steele KP e Liston A (2000) Molecular phylogeny of the “Temperate Herbaceous Tribes” of papilionoid legumes: a supertree approach. In Herendeen PS e Bruneau A (eds) **Advances in Legume Systematics**. Royal Botanic Gardens, Kew, London, p. 277–298.

Wojciechowski MF (2003) Reconstructing the phylogeny of legumes (Leguminosae):

an early 21st century perspective. In Klitgaard BB e Bruneau A (eds) **Higher Level Systematics, Advances in Legume Systematics**. Royal Botanic Gardens, Kew, London, p. 5–35.

Wu GA, Terol J, Ibanez V, García AL, Román EP, Borredá C, Domingo C, Tadeo FR, Caballero JC, Alonso R (2018) Genomics of the origin and evolution of *Citrus*. **Nature** **554**: 311-316.

Xia X (2020) Maximum parsimony method in phylogenetics. In *A Mathematical Primer of Molecular Phylogenetics*, Apple Academic Press, Flora, p. 185-208.

Xiong AS, Peng RH, Zhuang J, Gao F, Zhu B, Fu XB, Xue Y, Jin XF, Tian YS, Zhao W e Yao QH (2009) Gene duplication, transfer, and evolution in the chloroplast genome. **Biotechnology advances** **27**: 340–347.

Xu JH, Liu Q, Hu W, Wang T, Xue Q e Messing J (2015) Dinâmica de genomas de cloroplastos em plantas verdes. **Genomics** **106**: 221-231.

Yahara T, Javadi F, Onoda Y, Queiroz LP, Faith D e *et al.* (2013) Global legume diversity assessment: Concepts, key indicators, and strategies. **Taxon** **62**: 249–266.

Yang Z e Rannala B (2012) Molecular phylogenetics: principles and practice. *Nature Reviews*. **Genetics** **13**: 303-314.

Yang Z, Wang G, Ma Q, Ma W, Liang L e Zhao T (2019) The complete chloroplast genomes of three Betulaceae species: implications for molecular phylogeny and historical biogeography. **PeerJ** **7**: e6320.

Zha X, Wang X, Li J, Gao F e Zhou Y (2020) Complete chloroplast genome of *Sophora alopecuroides* (Papilionoideae): molecular structures, comparative genome analysis and phylogenetic analysis. **Journal of Genetics** **99**: 13.

Zhang R, Wang YH, Jin JJ, Stull GW, Bruneau A, Cardoso D e *et al.* (2020) Exploration of plastid phylogenomic conflict yields new insights into the deep relationships of Leguminosae. **Systematic Biology** **69**: 613–622.

Zhao Y, Zhang R, Jiang K-W, Qi J, Hu Y e *et al.* (2021) Nuclear phylotranscriptomics and phylogenomics support numerous polyploidization events and hypotheses for the evolution of rhizobial nitrogenfixing symbiosis in Fabaceae. **Molecular Plant** **14**: 1–26.

Filogenia com base em marcadores plastidiais revela *insights* sobre as relações de espécies de Papilionoideae

CAPÍTULO II

Artigo submetido à revista Genetic Resources and Crop Evolution (Qualis A2)

1 Filogenia com base em marcadores plastidiais revela *insights* sobre as relações de espécies 2 de Papilionoideae

3 Flávia Gomes da Silva^{1*}, Rômulo Maciel de Moraes Filho¹, Luiza Suely Semen Martins²,
4 Robson da Silva Ramos^{3,4}, Gheysa Coelho Silva¹

5 **Resumo** A subfamília Papilionoideae abrange mais de 14.000 espécies, 501 gêneros e 32 tribos,
6 representando dois terços de todos os gêneros e espécies da família Fabaceae (Leguminosae).
7 Os papilionóides são reconhecidos por sua importância alimentar e forrageira, ampla
8 distribuição nos diversos biomas e grande variação na arquitetura floral, bem como estrutura
9 do plastoma. Devido à elevada conservação dos genomas de cloroplasto, se comparados com
10 os genomas nucleares e mitocondriais, as análises filogenéticas baseadas em DNA de
11 cloroplasto (cpDNA) vem elucidando as relações entre os principais táxons de Papilionoideae,
12 entretanto a filogenia de alguns clados da subfamília permanece não solucionadas. Assim, este
13 estudo teve como objetivo realizar uma reconstrução filogenética das espécies de ramificação
14 profunda de Papilionoideae. Para análise de máxima verossimilhança foram utilizadas
15 sequências concatenadas de seis *loci* (*matK*, *psaA*, *psbA*, *psbD*, *rbcL* e *rpoC2*) de cpDNA de
16 117 espécies de Papilionoideae. Os plastomas de Papilionoideae evidenciaram baixa
17 conservação e similaridade. A análise filogenética resultou em uma árvore monofilética,
18 confirmando a divisão da subfamília em quatro principais clados, NPAAA, ADA, Genistoides
19 e Dalbergioides. Foi demonstrado a relação de grupo irmão, do clado ADA com o clado
20 Genistoides, com alto suporte. Dentro do clado NPAAA, foi evidenciado o parafiletismo das
21 tribos Phaseoleae e Millettiae, com filogenia do clado Genistoides não resolvida. Assim como,
22 observamos que apenas as espécies do clado ADA não possuem nódulos rizobianos, podendo
23 essa ser uma possível sinapomorfia para suportar as relações desse grupo. As análises também
24 sugerem que os principais clados de Papilionoideae divergiram a partir do Paleoceno.

25 **Palavras-chave** Leguminosae, Plastomas, Tempo de divergência, Análise *in silico*

26

27 ¹Programa de Pós-Graduação em Melhoramento Genético de Plantas, Departamento de Agronomia, Universidade
28 Federal Rural de Pernambuco, Recife – PE, Brasil

29 ²Programa de Pós-Graduação em Melhoramento Genético de Plantas, Departamento de Biologia, Universidade
30 Federal Rural de Pernambuco, Recife – PE, Brasil

31 ³Biofábrica Governador Miguel Arraes, Centro de Tecnologias Estratégicas do Nordeste, Recife – PE, Brasil

32 ⁴Departamento de Agronomia, Faculdades Nova Esperança, João Pessoa – PB, Brasil

33 *Autor para correspondência <flavia.gomess12@gmail.com>

34 **Abstract** The Papilionoideae subfamily comprises more than 14,000 species, 501 genera and
35 32 tribes, representing two-thirds of all genera and species in the Fabaceae family. Papilionoids
36 are recognized for their food and forage importance, wide distribution in different biomes and
37 variation in floral architecture as well as plastome structure. Due to the high-level conservation
38 of chloroplast genomes, when compared to nuclear and mitochondrial genomes, phylogenetic
39 analyzes based on chloroplast DNA (cpDNA) have been elucidating the relationships among
40 the main Papilionoideae's taxon. However, the phylogeny of some clades of the subfamily
41 remains unresolved. Aiming at the phylogenetic reconstruction of the deep branching species
42 of Papilionoideae, concatenated sequences of six loci (*matK*, *psaA*, *psbA*, *psbD*, *rbcL* and
43 *rpoC2*) of cpDNA from 117 species of Papilionoideae were analyzed using maximum
44 likelihood. The plastomes of Papilionoideae showed low conservation and similarity.
45 Phylogenetic analysis resulted in a monophyletic tree, confirming the division of the subfamily
46 into four main clades (NPAAA, ADA, Genistoids and Dalbergioids). The sibling group
47 relationship of the ADA clade with the Genistoids clade was demonstrated, with high support.
48 The paraphyly of the Phaseoleae and Millettiae tribes was evidenced within the NPAAA clade
49 with unresolved phylogeny of the Genistoids clade. As well, was observed that only species of

50 the ADA clade have no rhizobian nodules, which may be a possible synapomorphy to support
51 the relationships of this group. The analyzes also suggest that the main Papilionoideae clades
52 diverged from the Paleocene onwards.

53 **Keyword** Leguminosae, Plastomes, Divergence time, *In silico* analysis

54 **Introdução**

55 Papilionoideae é a maior subfamília de Fabaceae (Leguminosae), com cerca de dois
56 terços de todos os gêneros e espécies da família, representada por 501 gêneros e 14.000 espécies
57 (LPWG, 2021). Essa subfamília também é bem reconhecida por suas espécies economicamente
58 importantes, como soja (*Glycine max* L.), feijão comum (*Phaseolus vulgaris* L.), feijão fava (*P.*
59 *lunatus* L.), feijão caupi (*Vigna unguiculata* L.), amendoim (*Arachis hypogaea* L.), ervilha
60 (*Pisum sativum* L.), lentilha (*Lens culinaris*) *Clitoria ternatea* L., espécies do gênero
61 *Stylosanthes*, dentre outras.

62 A classificação filogenética da subfamília Papilionoideae vem sendo modificada ao
63 longo do tempo (Cardoso *et al.*, 2012, 2013; LPWG, 2013, 2017), mas ainda existem relações
64 não totalmente esclarecidas na filogenia dessa subfamília (Cardoso *et al.*, 2013; Choi *et al.*,
65 2022; Zhao *et al.*, 2019).

66 Os estudos filogenéticos são fundamentais para a sistemática e evolução, fornecendo
67 conhecimento sobre os padrões de diversificação, assim como a origem dos caracteres de
68 interesse das espécies (Machado *et al.*, 2017). Esses caracteres podem ser explorados no
69 melhoramento de plantas, como a resistência a pragas, doenças e tolerância ao estresse abiótico,
70 que possibilitam o desenvolvimento de cultivares mais produtivos (Hallauer, 2011).

71 O DNA de genomas plastidiais (cpDNA) são úteis para realizar análises comparativas
72 associadas a estudos filogenéticos e evolutivos, tendo como as principais características
73 tamanho e natureza conservada, herança materna e ser abundante nas células (Soltis e Soltis,
74 2020). Por conseguinte, sequências de genes de cpDNA são cruciais em estudos filogenéticos,
75 genética de populações, filogeografia e identificação de espécies (Santos e Almeida, 2019; Song
76 *et al.*, 2019; Yang *et al.*, 2019). Apesar dos diversos estudos de filogenia e diversificação de
77 Papilionoideae, os resultados evidenciam a necessidade de novos estudos filogenéticos, uma
78 vez que a maioria dos estudos são baseados apenas em um único gene, sendo necessária a adição
79 de mais *loci* para uma análise filogenética mais esclarecedora (Cardoso *et al.*, 2013).

80 Neste sentido, o objetivo desta pesquisa foi realizar a reconstrução filogenética das
81 espécies da subfamília Papilionoideae e estimar o tempo de divergência, por meio de sequências
82 de marcadores plastidiais, para entender a história evolutiva da subfamília.

83 **Materiais e métodos**

84 **Obtenção das sequências**

85 As sequências de genomas plastidiais completos de 121 espécies foram recuperadas por
86 meio da ferramenta BLAST do banco de dados Genbank no NCBI (National Center of
87 Biotechnology Information), sendo realizado o *download* em formato FASTA. Deste total de
88 espécies, 117 sequências são das espécies da subfamília Papilionoideae, quatro do grupo
89 externo (*outgroup*) da análise filogenética, sendo duas sequências de espécies da ordem
90 Malvales, algodoeiro (*Gossypium hirsutum*) e cacauero (*Theobroma cacao*), e duas espécies
91 da ordem Rosales, macieira (*Malus domestica*) e figueira (*Ficus carica*) (Tabela 1). Também
92 foram obtidas as sequências de seis genes do cpDNA (*matK*, *psaA*, *psbA*, *psbD*, *rbcL* e *rpoC2*)
93 de cada espécie.

94 **Obtenção dos dados das sequências dos genomas plastidiais**

95 Os valores de números cromossômicos das espécies utilizadas nesta pesquisa, foram
96 obtidos a partir de informações disponíveis na literatura e nos bancos de dados CCDB (The
97 Chromosome Counts Database, 2015) e IPCN (Index to Plant Chromosome Numbers, 2021).
98 Os dados de nodulação foram obtidos em Sprent (2009) e no banco de dados interativo ILDON
99 (Interactive Legume Database Of Nodulation: www.ildon.org). A partir da ferramenta GeSeq
100 do Chlorobox (Chlorobox, 2021) foram extraídas, o tamanho do genoma em pares de bases
101 (PB), número de genes (Genes), número de RNA ribossômico (rRNA) e número de RNA
102 transportador (tRNA). Para a obtenção da porcentagem de guanina e citosina (GC%) a
103 ferramenta Science buddies (Science buddies, 2021) foi utilizada (Tabela 1).

104 Alinhamento das sequências de genomas plastidiais

105 O alinhamento foi realizado no servidor mVISTA (Brudno 2003) utilizando o algoritmo
106 mLAGAN para 28 espécies de Papilionoideae, e a espécie *Stryphnodendron adstringens*,
107 pertencente a subfamília Caesalpinioideae como grupo externo. Os parâmetros padrões foram
108 aplicados, e a estrutura de anotação do genoma do cloroplasto do feijoeiro (NC_009259) foi
109 utilizado como referência, para verificar a ocorrência de inversões (Fig. 1). A representação
110 linear do plastoma foi obtida pelo servidor OGDRAW (Organelar Genome DRAW) (Lohse e
111 Oliver 2013). O gráfico VISTA (Frazer *et al.* 2004) permitirá verificar o percentual de
112 identidade de cada plastoma em relação ao feijoeiro.

113 Alinhamento dos genes e reconstrução filogenética

114 Os genes *matK*, *psaA*, *psbA*, *psbD*, *rbcL* e *rpoC2* foram selecionados previamente por
115 meio do gráfico vista (Fig. 1) e por sua importância fisiológica. As sequências dos genes de
116 cada genoma completo foram alinhadas com o algoritmo MUSCLE (Edgar 2004) no *software*
117 MEGA, versão 7.0 (Análise Genética Evolutiva Molecular) (Kumar *et al.* 2016).
118 Posteriormente, as regiões foram concatenadas em uma única supermatriz que também foi
119 alinhada no algoritmo MUSCLE (Edgar 2004), e a árvore filogenética foi criada com o *software*
120 MEGA 7.0 pelo método de Máxima Verossimilhança. O modelo de substituição foi o tipo GTR
121 (G + I) (General Time Reversible) obtido pelo JmodelTEST (Posada 2008). O suporte
122 estatístico foi calculado pelo método de *bootstrap* (BS) com 1000 replicatas. Como parâmetro
123 para verificação desta precisão, BS de 90-100 foram considerados fortemente suportados, de
124 70-89 moderadamente suportados e de 50-69 fracamente suportados.

125 Análise do tempo de divergência entre as espécies

126 Utilizou-se o servidor TimeTree (The timescale of life, 2021) para obtenção dos dados
127 do tempo de divergência estimado entre os pares de espécies, atribuindo idade aos nós dos
128 clados com variação mínima e máxima das idades em milhões de anos (Ma), utilizados como
129 pontos de calibração. Sendo selecionados seis pontos com a menor variação encontrada. O
130 *software* MEGA 7.0 foi utilizado para obtenção da árvore do tempo de divergência entre as
131 espécies de Papilionoideae, por meio da Máxima Verossimilhança e o modelo GTR (General
132 Time Reversible).

133 Resultados

134 Informações das sequências de genomas plastidiais

135 O tamanho do genoma variou de 120.289 (*Lathyrus odoratus*) a 179.362 pb (*Myroxylon*
136 *balsamum*), com uma variação no conteúdo GC de 33,8% (*Medicago hybrida* e *M. papillosa*)
137 a 36,8% (*Ammopiptanthus mongolicus*, *Genista tinctoria* e *Mirbelia oxylbioides*); nos genes

138 codificantes de proteínas foram observados valores com variação de 62 (*Centrosema*
139 *pubescens*) a 100 (*Trifolium pratense*); o rRNA variou de 4 a 8 para a maioria das espécies; o
140 tRNA variou de 26 (*M. Papillosa*) a 38 (várias espécies); a quantidade de genes variou de 96 a
141 139, para *M. papillosa* e *O. boluoensis*, respectivamente; e o número cromossômico variou de
142 6 (*Lotus corniculatus*, *Vicia fava* e *V. sativa*) a 40 (*G. dolichocarpa* e *G. tomentella*) (Tabela
143 1).

144

145 **Comparação genômica de plastomas**

146 Foi observado a ocorrência de várias regiões de baixa similaridade, o que pode indicar
147 possíveis eventos de inversão e rearranjo em trechos dos genomas das espécies de
148 Papilionoideae, em comparação com o feijoeiro, principalmente na região dos genes *ndhF* e
149 *yc1fB*. A espécie *S. adstringens*, utilizada como grupo externo, apresentou-se divergente em
150 relação às espécies de Papilionoideae (Fig. 1). Dentre os genes que exibiram pontos de
151 divergência, os genes *matK*, *psaA*, *psbA*, *psbD*, *rbcL* e *rpoC2*, foram selecionados para análise
152 filogenética.

153 **Filogenia da subfamília Papilionoideae**

154 A árvore filogenética baseada nos genes *matK*, *psaA*, *psbA*, *psbD*, *rbcL* e *rpoC2* resultou
155 em uma árvore monofilética, em que separa a subfamília Papilionoideae em quatro clados
156 principais, NPAAA, ADA, Genistoides e Dalbergioides, e o *outgroup* (Fig. 2). O *outgroup*
157 apresentou-se visivelmente separado das demais espécies, com alto valor de suporte (BS=
158 100%). O clado ADA é representado pelas tribos Angylocalyceae, Dipterygeae e Amburaneae
159 com relação bem suportada entre as espécies dessas tribos (BS= 100%). O clado Genistoides é
160 representado pelas tribos Genisteae, Crotalariaeae, Podalyrieae, Sophoreae, Brongniartieae e
161 Ormosieae, com um baixo valor de suporte (BS= 56%) entre as espécies dessas tribos. Para o
162 clado Dalbergioides, tem-se as tribos Amorpeae e Dalbergieae com suas relações bem
163 sustentadas (BS= 97%).

164 O clado NPAAA apresentou o maior número de espécies e tribos, sendo moderadamente
165 suportadas (BS= 72%), tendo a tribo Psoraleae e Desmodieae aninhada a tribo Phaseoleae.
166 Neste grande clado a espécie *C. ternatea* ficou isolada da sua tribo de origem (Phaseoleae), bem
167 como a espécie *Callerya nitida* (Millettieae). Também foi observado neste grande clado, que a
168 relação das tribos Mirbelieae e Bossiaeeae com os demais clados apresentaram suporte
169 moderado (BS= 72%). Dentro da tribo Fabeae, também foram observados agrupamentos
170 fracamente suportados, como nos gêneros *Vicia* e *Lens* (BS= 56%), e *Lathyrus* e *Vavilovia* (BS=
171 52%).

172 A relação das tribos de NPAAA com as tribos dos clados ADA e Genistoides foi
173 altamente suportada (BS= 91%). Dentro do clado Genistoides as relações entre as tribos
174 Ormosieae e Brongniartieae foram fracamente suportadas (BS= 41%), bem como sua relação
175 com as demais tribos desse clado (BS= 56%). Os clados ADA e Genistoides foram recuperados
176 como grupos irmãos, com alto valor de suporte (BS= 94%). No clado Dalbergioides, as relações
177 dentro da tribo Amorpeae (BS= 93%) e Dalbergieae foi altamente suportada (BS= 100%).

178 **Estimativa de tempo de divergência**

179 As estimativas de tempo de divergência indicam que a diferenciação de Papilionoideae
180 que dão origem aos clados NPAAA, ADA, Genistoides e Dalbergioides ocorreu há 62,13 Ma,
181 no início do Paleoceno (65,50-61,70 Ma) (Fig. 2). NPAAA se diferenciou há 58,85 Ma
182 (Paleoceno médio), separando as tribos Mirbelieae e Bossiaeeae dos demais clados de
183 Papilionoideae. O clado ADA se diversificou em 44,96 Ma, no Eoceno final. No clado
184 Genistoides a diversificação ocorreu em 58,50 Ma no período Paleoceno (65,50-55,80 Ma). No

185 clado Genistoides a separação de Ormosieae e Brongniartieae das tribos Genisteeae,
186 Crotalariaeae, Podalyrieae, Sophoreae ocorreu em 58,49 Ma, no período Paleoceno. A separação
187 de NPAAA dos clados ADA e Genistoides ocorreu em 62,13 Ma, no início do Paleoceno. Para
188 as tribos do clado ADA e do clado Genistoides a diversificação ocorreu em 61,83 Ma, no início
189 do Paleoceno. O clado Dalbergioides se diversificou em 56,96 Ma, no Paleoceno final (58,70-
190 55,80 Ma), com a diferenciação das tribos Amorpeae e Dalbergieae.

191 **Discussão**

192 **Genômica comparativa de Papilionoideae**

193 O tamanho do genoma (pb) das espécies analisadas assemelha-se aos encontrados por
194 outros autores em espécies de Papilionoideae (121,020 - 158,367 bp) (Choi e Choi 2017); *D.*
195 *alata* (158,647 pb) (Antunes *et al.* 2020); *Mucuna Macrocarpa* (153,826 pb) (Jin *et al.* 2019);
196 feijão caupi (152,415 pb) (Tian *et al.* 2021); *G. dolichocarpa* (152,804 pb) (Asaf *et al.* 2017).
197 Como observado, os resultados mostraram uma variação no tamanho do cpDNA (120.289 a
198 179.362 pb) dessas espécies. Essa constatação sugere que esses genomas podem ter passado
199 por diferentes processos evolutivos, dentre eles perda de gene e/ou intron, inserção e/ou deleção
200 (Wicke *et al.* 2011) e expansão e/ou contração de sua estrutura quadripartida (Davis e Soreng
201 2010; Wicke *et al.* 2011).

202 Os genomas de cloroplasto apresentaram valores de conteúdo de GC% semelhantes aos
203 encontrados na literatura para a maioria das espécies avaliadas (Choi e Choi 2017; Antunes *et*
204 *al.* 2020; Jin *et al.* 2019; Tian *et al.* 2021; Asaf *et al.* 2017). Segundo Moghaddam *et al.* (2022),
205 o elevado conteúdo de GC é favorável para a estabilidade do genoma e manutenção da
206 complexidade da sequência. Quanto ao número de genes que codificam proteínas, os valores
207 foram próximos aos encontrados para as Glycines (Asaf *et al.* 2017), bem como valores no
208 intervalo de 74-85 encontrado por Zha *et al.* (2020) para a maioria das espécies avaliadas neste
209 estudo. Valores semelhantes também foram relatados por Antunes *et al.* (2020) em *D. alata*,
210 onde 109 genes foram observados com 76 genes que codificam proteínas, 4 genes de rRNA e
211 29 genes de tRNA, e Tian *et al.* (2021) para feijão-fava e feijoeiro, observando 127 genes para
212 as duas espécies, incluindo 82 e 83 genes codificadores de proteínas, 37 e 36 genes de tRNA e
213 8 genes de rRNA, respectivamente.

214 Estudos recentes evidenciam a importância que modificações de RNA podem estimular
215 as respostas das plantas ao estresse salino e térmico, e infecções virais (Anderson *et al.* 2018;
216 Martínez-Pérez *et al.* 2017).

217 Na literatura não existe um consenso acerca do número cromossômico base de
218 Papilionoideae, sendo observado estudos a nível de clados ou tribos. Para Zhao *et al.* (2021) o
219 número base de Papilionoideae é $x = 28$, originário no ancestral dos Genistóides, Dalbergioides
220 e em outros papilionóides, ou até antes no ancestral de Papilionoideae. Para as tribos
221 Dalbergieae e Phaseoleae seguiu um padrão de números cromossômicos base $x = 10$ e 11
222 (Lackey 1980; Moraes *et al.* 2020), respectivamente.

223 Estudos observaram espécies que apresentam comportamento de $x = 20$ ou 40
224 (Glycines), sendo estas poliplóides estabilizadas (Doyle 2012; Lackey 1980; Singh 2019).
225 Doyle (2012) analisando aproximadamente 400 gêneros com base no número cromossômico
226 para poliploidia, observou que um quarto desses gêneros tinha táxons poliplóides, sendo muitos
227 membros de Genisteeae, Pterocarpus ($2n = 22, 24, 44$), numerosos Mirbelioides, Caragana ($2n$
228 $= 16, 24, 32, 48$) e Glycine ($2n = 38, 40$). Para os representantes do gênero *Abrus* (tribo Abreae),
229 estudos apontam para um status tetraplóide com base em $2n = 22$ para *A. precatorius* e *A.*
230 *pulchellus*, concordando com um número cromossômico básico de $x = 11$ (Agbagwa e Okoli
231 2020).

232 Através do número cromossômico, Hu (2000a) sugere que as espécies dos gêneros
233 *Callerya* são distintas de outras Millettieae, que aqui apresentou $x = 16$ para *Callerya* e $x = 8$

234 para as espécies da tribo Millettieae.

235 Mesmo quando os números cromossômicos são iguais, existe variação no tamanho dos
236 genomas, reforçando a relevância das estimativas do conteúdo de DNA para a compreensão da
237 evolução cariotípica de espécies homoplóides (Loureiro *et al.* 2010), auxiliando na escolha de
238 espécies para cruzamentos interespecíficos no melhoramento de plantas.

239 Com base no alinhamento do gráfico mVista (Fig. 1), as espécies com maior
240 similaridade ao genoma de referência (feijoeiro), foram feijão caupi e feijão-da-china (*V.*
241 *mungo*), indicando um elevado nível de conservação do genoma dessas espécies. Entretanto
242 comparando os plastomas das demais espécies de Papilionoideae pelo gráfico mVista, foi
243 possível observar pontos de divergência entre os genes *ndhF* e *ycf1B*, localizados em regiões
244 codificantes e não codificantes (menor proporção). A ocorrência de inversões e rearranjos nas
245 sequências gênicas dos genomas das espécies estudadas, sugerem um elevado nível de
246 diversidade entre as espécies de Papilionoideae. Essa grande diversidade apresentada pelas
247 espécies de Papilionoideae é de grande importância para a evolução das culturas e o sucesso
248 dos programas de melhoramento, que necessitam sobretudo da diversidade genética disponível
249 nos bancos de germoplasma (Baudoin 1988).

250 Na literatura alguns autores sugerem que os plastomas de vários membros de
251 Papilionoideae não são conservados em sua estrutura genômica, e que a elevada diversidade do
252 plastoma se deve a rearranjos e taxas de mutação aceleradas, como por exemplo a perda de
253 genes/íntrons (Jansen *et al.* 2008; Sabir *et al.* 2014; Schwarz *et al.* 2015). Ainda, existem
254 diversos outros motivos para a ocorrência de rearranjos no plastoma, como a falta de uma região
255 de repetições invertidas (IR), variação no tamanho da região IR e a presença de muitas
256 sequências repetidas em tandem (Keller *et al.* 2017).

257

258 **Análise filogenética**

259 A árvore filogenética baseada nos genes *matK*, *psaA*, *psbA*, *psbD*, *rbcL* e *rpoC2*,
260 evidenciou a formação de quatro clados principais, que corrobora com os resultados de Cardoso
261 *et al.* (2012, 2013).

262 As relações entre as espécies do clado ADA apresentaram um alto valor de suporte (BS=
263 100%), sendo consideradas papilionóides de ramificações potencialmente mais antigas, que
264 inclui as tribos monofiléticas Angylocalyceae, Dipterygeae e Amburanae (Antunes *et al.* 2020;
265 Cardoso *et al.* 2012, 2013; Zhang *et al.* 2020). Este clado é formado por um conjunto de gêneros
266 morfologicamente diferentes (Cardoso *et al.* 2013). A distribuição geográfica das espécies que
267 compõem as tribos do clado ADA, são bastante diversas, desde África e Madagascar
268 (*Angylocalyx* e *Xanthocercis*), Austrália (*Castanospermum*), até as florestas tropicais do norte
269 da América do Sul (*Alexa* e *Uleanthus*), para a tribo *Angylocalyceae* (Pennington *et al.* 2005).
270 Em quanto as espécies da tribo Dipterygeae e Amburanae são predominantemente neotropicais
271 (Cardoso *et al.*, 2013). Até então este clado permanece sem uma definição respaldada por nós,
272 assim como nenhuma sinapomorfia morfológica foi reconhecida.

273 Entretanto, observamos que a maioria dos clados de Papilionoideae apresentam
274 nodulação, sendo apenas as espécies representantes do clado ADA que não possuem nódulos
275 rizobianos, corroborando o observado por outros autores (Cannon *et al.* 2015; Sprent 2009;
276 Zhao *et al.* 2021), podendo essa ser uma possível sinapomorfia para associar as relações
277 filogenéticas desse clado.

278 Genistoides é considerado o clado de ramificação precoce mais rico em espécies
279 (Cardoso *et al.* 2013), que aqui foi representado pelas tribos Genisteeae, Crotalariaeae,
280 Podalyrieae, Sophoreae, Brongniartieae e Ormosieae, tribos predominantemente do hemisfério
281 sul (Cardoso *et al.* 2012, 2013). Essas leguminosas também apresentam simetria floral radial,
282 bem como pétalas indiferenciadas ou estames livres (Polhill 1981). Apesar do fraco valor de
283 suporte encontrado em nosso estudo para Genistoides (BS= 56%), é um clado monofilético com

284 sua filogenia bem estabelecida (Antunes *et al.* 2020; Cardoso *et al.* 2012, 2013; Zhang *et al.*
285 2020), determinado pelo ancestral comum mais recente de *Poecilanthe parviflora* e *Lupinus*
286 *argenteus* (Wojciechowski *et al.* 2004).

287 As leguminosas genistóides tem por sinapomorfia química o acúmulo de alcalóides do
288 grupo quinolizidina (Van Wyk 2003; Wink e Mohamed 2003; Kite *et al.* 2013), ainda que
289 existam alguns grupos que acumulam alcalóides pirrolizidínicos (Wink e Mohamed 2003) ou
290 por ausência total de alcalóides (Boatwright *et al.* 2008). Os alcalóides do tipo quinolizidínicos
291 são caracterizados pelo sabor amargo, que são frequentemente associados a defesa de plantas
292 ao ataque de patógenos e insetos (Bunsupa *et al.* 2013; Wink e Mohamed 2003). Dessa forma,
293 essa característica pode ser explorada nos programas de melhoramento de espécies que não
294 sintetizam esses alcalóides, por meio de hibridações interespecífica ou intraespecífica, e entre
295 espécies geneticamente distantes por meio da engenharia genética.

296 Este estudo apoiou fortemente (BS= 94%) o clado ADA como grupo irmão do clado
297 Genistoides. Cardoso *et al.* (2012, 2013) também observaram que o clado ADA, juntamente
298 com o clado Swartzioideae como irmãos sucessivos das demais papilionóides.

299 O clado Dalbergioideae incluiu representantes das tribos Dalbergieae e Amorphaeae. Este
300 clado apresentou-se altamente suportado (BS= 97%), sendo solucionado de forma consistente
301 como monofilético em análises filogenéticas moleculares (Cardoso *et al.* 2012, 2013; LPWG
302 2013; Wojciechowski 2013). Cardoso *et al.* (2013) sugerem que as vagens indeiscentes podem
303 ser uma boa sinapomorfia morfológica para este clado. A distribuição geográfica do clado se
304 dá a partir das duas tribos que o compõe, sendo a tribo Dalbergieae distribuída em qualquer
305 região dos trópicos (pantropical), principalmente nos Neotrópicos e na África Subsaariana
306 (Klitgaard e Lavin 2005), e a tribo Amorphaeae, distribuída do sudoeste dos EUA e México ao
307 norte da Argentina (Wojciechowski 2013).

308 O clado NPAAA inclui as espécies papilionóides que acumulam aminoácidos não
309 proteicos, que seria o acúmulo do aminoácido não proteínogênico canavanina nas sementes
310 dessas leguminosas (Bell 1981). Esse clado compreende muitos gêneros de importância
311 alimentar e econômica, como *Phaseolus*, *Glycine*, *Vigna*, *Vicia*, *Cajanus*, *Cicer*, *Pisum* e *Lens*.
312 Todas as papilionóides NPAAA são caracterizadas pela evolução de flores altamente
313 especializadas (flor papilionácea) (Lewis *et al.* 2005). Este grande clado é bem resolvido de
314 forma consistente em filogenias moleculares a partir do gene *matK*, (Cardoso *et al.* 2012, 2013;
315 Queiroz *et al.* 2015; Wojciechowski *et al.* 2004) e *rbcL* (Wojciechowski 2003).

316 Nossa análise filogenética sugere a parafilia das tribos Millettieae e Phaseoleae,
317 corroborada por estudos anteriores (Egan *et al.* 2016; Hu *et al.* 2000b; Queiroz *et al.* 2015;
318 Oyebanji *et al.* 2020; Vatanparast *et al.* 2018; Wojciechowski *et al.* 2004, Zhao *et al.* 2021).

319 Também destacamos as tribos Desmodieae e Psoraleae aninhadas dentro de
320 Phaseoleae, conforme relatado anteriormente em análise com marcadores de DNA de plastídios
321 (Kajita *et al.* 2001; LPWG 2017; Hu *et al.* 2000b; Queiroz *et al.* 2015; Oyebanji *et al.* 2020;
322 Stefanovic' *et al.* 2009; Vatanparast *et al.* 2018; Wojciechowski *et al.* 2004). Esse fato
323 possivelmente pode estar relacionado com a presença de nódulos de crescimento determinado
324 e exportação de ureídeos, característico das espécies dessas tribos (Sprent 2008).

325 Aqui, evidenciamos que a espécie *C. nitida* é distinta das demais espécies da tribo
326 Millettieae, diferença essa que também pode ser observada através do número cromossômico
327 dessas espécies 16 e 8, respectivamente. Além disso, demonstramos o agrupamento com alto
328 valor de suporte (BS= 100%) dos gêneros *Callerya* e *Wisteria*. Esse fato também foi observado
329 por Hu *et al.* (2000a), que analisou a filogenia da tribo Millettieae baseada em sequências de
330 cloroplasto *trnK/matK*, verificou que *Callerya* e *Wisteria* são parentes próximas e que são
331 bastante distintas de outros membros da tribo Millettieae. Essa característica também foi
332 confirmada por Compton *et al.* (2019).

333 Também observamos que *C. ternatea* não ficou aninhada na tribo Phaseoleae. Esse
334 resultado também foi encontrado em análises anteriores que mostraram que a subtribo

335 Clitoriinae, cuja a maior representante é *C. ternatea* tem uma relação distante da maioria das
336 espécies da tribo Phaseoleae (Schrire 2005; Cardoso *et al.* 2013; Queiroz *et al.* 2015).

337 Com base nos pontos observados nesta análise, talvez uma nova classificação
338 filogenética tribal de Papilionoideae devesse ser realizada para solucionar as incongruências
339 aqui verificadas na relação filogenética dessa subfamília. Como por exemplo, acomodar *C.*
340 *ternatea* em sua própria tribo (Clitoriinae), bem como as tribos Psoraleae e Desmodieae sendo
341 subtribos da tribo Phaseoleae. Contudo, estudos complementares que englobem uma
342 amostragem de táxons mais representativa da subfamília serão necessários antes de uma decisão
343 taxonômica mais definitiva.

344 **Tempo de divergência entre as espécies de Papilionoideae**

345

346 Esse estudo evidenciou que a provável divergência das tribos de Papilionoideae ocorreu
347 em 62,13 Ma (início do Paleoceno), resultados semelhantes foram obtidos por Lavin *et al.*
348 (2005) (~ 58,6 Ma) e por Zhao *et al.* (2021) (63,02 Ma). Nosso estudo também demonstra que
349 as espécies de Papilionoideae se diferenciaram em quatro principais clados, com início no
350 período Paleoceno (65,50-55,80 Ma). A diversificação do clado NPAAA ocorreu em 58,85 Ma
351 (Paleoceno médio), separando as tribos Mirbelieae e Bossiaeeae, que se diferenciaram em 58,85
352 Ma das demais Papilionoideae, que corrobora com resultados próximos obtidos por Zhao *et al.*
353 (2021), em que a diversificação de Mirbelieae e Bossiaeeae ocorreu em 51,76 Ma, no
354 Paleoceno.

355 Zhao *et al.* (2021), relatam que esse tempo coincide com o aumento da temperatura
356 global durante o máximo térmico Paleoceno-Eoceno e o ótimo climático Eoceno Inferior.
357 Hyland *et al.* (2015) também relatam que o Paleoceno é caracterizado por eventos hipertermais,
358 que propiciaram mudanças nas condições do ambiente, bem como na vegetação e na
359 diversificação do ecossistema.

360 A elevação da temperatura tem por função estimular a biodiversidade (Benton 2010;
361 Condamine *et al.* 2013; Erwin 2009), dessa forma acredita-se que com o aumento das
362 temperaturas as taxas de processos biológicos sejam intensificadas, assim como na geração de
363 tempos mais reduzidos e por consequência taxas mais elevadas de especiação (Zhao *et al.*
364 2021).

365 Ainda no clado NPAAA, as espécies da tribo Phaseoleae se diversificaram no período
366 Mioceno (23,03-5,33 Ma). Idades de divergência semelhantes foram relatadas por Egan *et al.*
367 2016 e Zhao *et al.* (2021), com 23,52 Ma e 23,96 Ma, respectivamente. Em contrapartida, Lavin
368 *et al.* (2005) e Li *et al.* (2013) sugerem que as leguminosas phaseolóides se diferenciaram no
369 Oligoceno (24,2–32,1 Ma). Vale salientar que ao decorrer do período Oligoceno até o Mioceno,
370 ocorreram mudanças marcantes nas condições ambientais globais, como fenômenos de
371 aquecimento e resfriamento do planeta, que propiciou grandes mudanças na formação da terra
372 e na configuração dos continentes, tal qual são reconhecidas nos dias de hoje (Zachos *et al.*
373 2001; Pälike *et al.* 2006; Wade e Pälike 2004).

374 Outro fator que pode ter atuado como uma revolução que sucedeu no aumento da taxa
375 de diversificação das leguminosas phaseolóides no período Mioceno, foi a transformação de
376 plantas lenhosas para herbáceas (Li *et al.* 2013). Isso se deve ao fato que as plantas herbáceas
377 têm um tempo de geração mais curto, podendo ter uma maior taxa de mutação anual, o que
378 possibilita ampliar a divergência genética, bem como a taxa de especiação, tornando as plantas
379 herbáceas mais diversas que as lenhosas (Eriksson e Bremer 1992; Dodd *et al.* 1999; Verdú
380 2002).

381 Para o clado ADA, nosso estudo revelou que a diversificação ocorreu durante o Eoceno
382 final (44,96 Ma), idade não muito distante da evidenciada por Zhao *et al.* (2021), com 61,05
383 Ma. Esse período foi marcado por um pico ótimo, com menores oscilações de amplitude
384 (Zachos *et al.* 2001), que propiciou condições ideais para a plena diversificação das espécies

385 desse clado.

386 As variações de tempo de divergência observadas no presente estudo para o clado ADA,
387 em relação as estimativas obtidas por Zhao *et al.* (2021), podem estar relacionadas as espécies
388 amostradas, número de fósseis e/ou calibrações utilizadas, bem como o método empregado.
389 Haja vista, que no nosso estudo foram utilizados seis pontos de calibração obtidos por meio da
390 plataforma TimeTree, ao contrário de Zhao *et al.* (2021) que utilizaram 23 restrições de idade
391 baseadas em fósseis reconhecidos.

392 A diversificação do clado Genistoides ocorreu no Paleoceno (58,50 Ma), corroborando
393 os resultados observados por Lavin *et al.* (2005), com idade de 56,4 Ma (~ 0,2 Ma) e Zhao *et al.*
394 *et al.* (2021), com 58,9 Ma. Já o clado Dalbergioides se diversificou no Paleoceno final (56,96
395 Ma). Essa estimativa de idade foi semelhante a observada por Zhao *et al.* (2021), com 57,15
396 Ma, e Lavin *et al.* (2005), com 55,3 Ma (~ 0,5 Ma).

397 **Conclusões**

398 As espécies de Papilionoideae apresentaram baixa conservação e similaridade entre os
399 plastomas.

400 A análise filogenética resultou em uma árvore monofilética, confirmando a divisão da
401 subfamília em quatro principais clados, NPAAA, ADA, Genistoides e Dalbergioides. Foi
402 demonstrado a relação de grupo irmão, do clado ADA com o clado Genistoides, com alto
403 suporte. No clado NPAAA, foi demonstrado o parafiletismo das tribos Phaseoleae e Millettiae.
404 Também foi evidenciado que apenas as espécies do clado ADA não possuem nódulos
405 rizobianos, podendo essa característica ser uma possível sinapomorfia para suportar as relações
406 desse grupo.

407 Os principais clados de Papilionoideae diversificaram entre o período Paleoceno e
408 Eoceno.

409

410 **Financiamento**

411 O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de
412 Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001. Silva FG foi
413 financiado por uma bolsa de doutorado CAPES, e Moraes Filho RM foi financiado por uma
414 bolsa de pós-doutorado CAPES.

415

416 **Referências**

417 Agbagwa IO, Okoli BE (2020) Chromosome number and cytomorphological characterization
418 of a polyploid *Abrus*. *Front Agric and Food Technol* 10:1-4

419 Anderson SJ, Kramer MC, Gosai SJ, Yu X, Vandivier LE, Nelson ADL, Anderson ZD,
420 Beilstein MA, Fray RG, Lyons E, Gregory BD (2018) N6-Methyladenosine Inhibits Local
421 Ribonucleolytic Cleavage to Stabilize mRNAs in Arabidopsis. *Cell Rep* 25:1146–1157.
422 <https://doi.org/10.1016/j.celrep.2018.10.020>

423 Antunes AM, Soares TN, Targueta CP, Novaes E, Coelho ASG, Telles MPC (2020) The
424 chloroplast genome sequence of *Dipteryx alata* Vog. (Fabaceae: Papilionoideae): genomic
425 features and comparative analysis with other legume genomes. *Braz J Biol* 43:271–282.
426 <https://doi.org/10.1007/s40415-020-00599-3>

427 Asaf S, Khan AL, Khan MA, Imran QM, Kang SM, Al-Hosni K, Jeong EJ, Lee KE, Lee I-J
428 (2017) Comparative analysis of complete plastid genomes from wild soybean (*Glycine soja*)
429 and nine other *Glycine* species. *PLoS ONE* 8: e0182281.
430 <https://doi.org/10.1371/journal.pone.0182281>

- 431 Baudoin JP (1988) Genetic resources, domestication and evolution of lima bean, *Phaseolus*
432 *lunatus*. J Emerg Med 39:253–60
- 433 Benton MJ (2010) The origins of modern biodiversity on land. Philos Trans R Soc Lond B Biol
434 Sci 365:3667–3679. <https://doi.org/10.1098/rstb.2010.0269>
- 435 Boatwright JS, Le Roux MM, Wink M, Morozova T, Van Wyk BE (2008) Phylogenetic
436 relationships of tribe Crotalarieae (Fabaceae) inferred from DNA sequences and morphology.
437 Syst Bot 33:752–761. <https://doi.org/10.1600/036364408786500271>
- 438 Bunsupa S, Saito K, Yamazaki M (2013) Molecular Biology and Biotechnology of
439 Quinolizidine Alkaloid Biosynthesis in Leguminosae Plants. In: Chandra S, Lata H, Varma A
440 (eds), Biotechnology for Medicinal Plants. Springer, Berlin, Heidelberg, pp 263–273.
- 441 Cannon SB, McKain MR, Harkess A, Nelson MN, Dash S, Deyholos MK, Peng Y, Joyce B,
442 Stewart CN, Jr, Rolf Jr M, Kutchan T, Tan X, Chen C, Zhang Y, Carpenter E, Wong GK-S,
443 Doyle JJ, Leebens-Mack J, (2015) Multiple Polyploidy Events in the Early Radiation of
444 Nodulating and Nonnodulating Legumes. Mol Biol Evol 32:193–210.
445 <https://doi.org/10.1093/molbev/msu296>
- 446 Cardoso D, Pennington RT, Queiroz LP, Boatwright JS, Van Wyk BE, Wojciechowski MF,
447 Lavin M (2013) Reconstructing the deep-branching relationships of the papilionoid legumes. S
448 African J Bot 89:58–75. <https://doi.org/10.1016/j.sajb.2013.05.001>
- 449 Cardoso D, Queiroz LP, Pennington RT, Lima HC, Fonty E, Wojciechowski MF, Lavin M
450 (2012) Revisiting the phylogeny of papilionoid legumes: New insights from comprehensively
451 sampled early-branching lineages. Am J Bot 99:1991–2013.
452 <https://doi.org/10.3732/ajb.1200380>
- 453 Choi I-S, Choi B-H (2017) The distinct plastid genome structure of *Maackia fauriei* (Fabaceae:
454 Papilionoideae) and its systematic implications for genistoids and tribe Sophoreae. PLoS ONE
455 4:e0173766. <https://doi.org/10.1371/journal.pone.0173766>
- 456 Choi I-S, Cardoso D, Queiroz LP, Lima HC, Lee C, Ruhlman TA, Jansen RK, Wojciechowski
457 MF (2022) Highly Resolved Papilionoid Legume Phylogeny Based on Plastid Phylogenomics.
458 Front. Plant Sci 13:823190. <https://doi.org/10.3389/fpls.2022.823190>
- 459 Compton JA, Schrire BD, Könyves K, Forest F, Malakasi P, Mattapha S, Sirichamorn Y (2019)
460 The Callerya Group redefined and Tribe Wisterieae (Fabaceae) emended based on morphology
461 and data from nuclear and chloroplast DNA sequences. PhytoKeys 124:1-112.
462 <https://doi.org/10.3897/phytokeys.125.34877>
- 463 Condamine FL, Rolland J, Morlon H (2013) Macroevolutionary perspectives to environmental
464 change. Ecology Letters 16:72–85. <https://doi.org/10.1111/ele.12062>
- 465 Davis JI, Soreng RJ (2010) Migration of endpoints of two genes relative to boundaries among
466 regions of the plastid genome in the grass family (Poaceae). Am J Bot 97:874-892.
467 <https://doi.org/10.3732/ajb.0900228>
- 468 Dodd ME, Silvertown J, Chase MW (1999) Phylogenetic analysis of trait Evolution and species
469 diversity variation among angiosperm families. Evolution 53:732–744.
470 <https://doi.org/10.2307/2640713>
- 471 Doyle JJ (2012) Polyploidy in legumes. In: Soltis PS, Soltis DE (eds) Polyploidy and Genome
472 Evolution. Springer, Heidelberg, Berlin, pp 147–180

- 473 Doyle JJ (2016) Chasing unicorns: nodulation origins and the paradox of novelty. *Am J Bot*
474 103:1865–1868. <https://doi.org/10.3732/ajb.1600260>
- 475 Egan AN, Vatanparast M, Cagle W (2016) Parsing polyphyletic Pueraria: delimiting distinct
476 evolutionary lineages through phylogeny. *Mol Phylogenet Evol* 104:44–59.
477 <https://doi.org/10.1016/j.ympev.2016.08.001>
- 478 Eriksson O, Bremer B (1992) Pollinationsystems dispersalmodos lifeforms and diversification
479 rates in angiosperm families. *Evolution* 46:258–266. <https://doi.org/10.2307/2409820>
- 480 Erwin DH (2009) Climate as a driver of evolutionary change. *Curr Biol* 19:575–583.
481 <https://doi.org/10.1016/j.cub.2009.05.047>
- 482 Hallauer A (2011) Evolution of plant breeding. *CBAB* 11:197–206.
483 <https://doi.org/10.1590/S1984-70332011000300001>
- 484 Hu J-M (2000a) Phylogenetic relationships of the tribe Millettieae and allies - the current status.
485 In: Herendeen PS, Bruneau A (eds) *Advances in Legume Systematics part 9*. Royal Botanic
486 Gardens, Kew, UK, pp 299–310
- 487 Hu J-M, Lavin M, Wojciechowski MF, Sanderson J (2000b) Phylogenetic Systematics of the
488 Tribe Millettieae (Leguminosae) Based on Chloroplast *trnK/matK* Sequences and its
489 Implications for Evolutionary Patterns in Papilionoideae. *Am J Bot* 87:418–
490 30. <https://doi.org/10.2307/2656638>
- 491 Hyland EG, Sheldon ND, Cotton JM (2015) Terrestrial evidence for a two-stage mid-Paleocene
492 biotic event. *Palaeogeogr Palaeoclimatol Palaeoecol* 417:371–378.
493 <https://doi.org/10.1016/j.palaeo.2014.09.031>
- 494 Jansen RK, Wojciechowski MF, Sanniyasi E, Lee SB, Daniell H (2008) Complete plastid
495 genome sequence of the chickpea (*Cicer arietinum*) and the phylogenetic distribution of *rps12*
496 and *clpP* intron losses among legumes (Leguminosae). *Mol Phylogenet Evol* 48:1204–17.
497 <https://doi.org/10.1016/j.ympev.2008.06.013>
- 498 Jin D-P, Choi I-S, Choi B-H (2019) Plastid genome evolution in tribe Desmodieae (Fabaceae:
499 Papilionoideae). *PLoS ONE* 6:e0218743. <https://doi.org/10.1371/journal.pone.0218743>
- 500 Kajita T, Ohashi H, Tateishi Y, Bailey CD, Doyle JJ (2001) *rbcL* and legume phylogeny, with
501 particular reference to Phaseoleae, Millettieae, and allies. *Syst Bot* 26:515–536.
- 502 Keller J, Rousseau-Gueutin M, Martin GE, Morice J, Boutte J, Coissac E, Ourari M, Ainouche
503 M, Salmon A, Cabello-Hurtado F, Ainouche A (2017) The evolutionary fate of the chloroplast
504 and nuclear *rps16* genes as revealed through the sequencing and comparative analyses of
505 novel legume chloroplast genomes from *Lupinus*. *DNA Res* 24:343–58.
506 <https://doi.org/10.1093/dnares/dsx006>
- 507 Kite GC, Cardoso D, Veitch NC, Lewis GP (2013) Quinolizidine alkaloid status of *Acosmium*
508 s.s., *Guianodendron* and *Leptolobium*, the segregate genera of *Acosmium* s.l. *South African J*
509 *Botany* 89:176–180. <http://dx.doi.org/10.1016/j.sajb.2013.05.009>
- 510 Klitgaard BB, Lavin M (2005) Dalbergieae sensu lato. In: Lewis GP, Schrire B, Mackinder B,
511 Lock M (eds) *Legumes of the world (mansc.)*, Royal Botanic Gardens, Kew, UK, 557p
- 512 Lackey JA (1980) Chromosome numbers in the phaseoleae (fabaceae:faboideae) and their
513 relation to taxonomy. *Am J Bot* 4:595–602. 1980. <https://doi.org/10.1002/j.1537-514.2197.1980.tb07689.x>

- 515 Lavin M, Herendeen PS, Wojciechowski MF (2005) Evolutionary rates analysis of
516 Leguminosae implicates a rapid diversification of lineages during the Tertiary. *Syst Biol*
517 54:575–594. <https://doi.org/10.1080/10635150590947131>
- 518 Lewis G, Schrire B, Mackinder B, Lock M (2005) *Legumes of the World*. Royal Botanic
519 Gardens, Kew, UK
- 520 Li H, Wang W, Lin L, Zhu X, Li J, Zhu X, Chen Z (2013) Diversification of the phaseoloid
521 legumes: effects of climate change, range expansion and habit shift. *Front Plant Sci* 4:386.
522 <https://doi.org/10.3389/fpls.2013.00386>
- 523 Loureiro J, Trávníček P, Rauchová J, Urfus T, Vít P, Štech M, Castro S, Suda J (2010) The use
524 of flow cytometry in the biosystematics, ecology and population biology of homoploid plants.
525 *Preslia* 82:3–21.
- 526 LPWG (Legume Phylogeny Working Group) (2013) Legume phylogeny and classification in
527 the 21st century: progress, prospects and lessons for other species-rich clades. *Taxon* 62:217–
528 248. <https://doi.org/10.12705/622.8>
- 529 LPWG (Legume Phylogeny Working Group) (2017) A new subfamily classification of the
530 Leguminosae based on a taxonomically comprehensive phylogeny. *Taxon* 66:44–77.
531 <https://doi.org/10.5061/dryad.61pd6>
- 532 LPWG (Legume Phylogeny Working Group) (2021) The World Checklist of Vascular Plants
533 (WCVP): Fabaceae, vers. June 2021. Govaerts, R. ed. Available online at:
534 http://sftp.kew.org/pub/data_collaborations/Fabaceae/DwCA/. Accessed in 13 December 2021
- 535 Machado LO, Vieira LN, Stefenon VM, Pedrosa FO, Souza EM, Guerra MP, Nodari RO (2017)
536 Phylogenomic relationship of feijoa (*Acca sellowiana* (O.Berg) Burret) with other Myrtaceae
537 based on complete chloroplast genome sequences. *Genetica* 145:163–174.
538 <https://doi.org/10.1007/s10709-017-9954-1>
- 539 Martínez-Pérez M, Aparicio F, López-Gresa MP, Bellés JM, Sánchez-Navarro JÁ, Pallás V
540 (2017) Arabidopsis m6A demethylase activity modulates viral infection of a plant virus and the
541 m6A abundance in its genomic RNAs. *PNAS* 114:10755–10760.
542 <https://doi.org/10.1073/pnas.1703139114>
- 543 Moghaddam M, Ohta A, Shimizu M, Terauchi R, Kazempour-Osaloo S (2022) The complete
544 chloroplast genome of *Onobrychis gaubae* (Fabaceae-Papilionoideae): comparative analysis
545 with related IR-lacking clade species. *BMC Plant Biol* 22:75. <https://doi.org/10.1186/s12870-546-022-03465-4>
- 547 Moraes AP, Vatanparast M, Polido C, Marques A, Souza G, Fortuna-Perez AP, Forni-Martins
548 ER (2020) Chromosome number evolution in dalbergioid legumes (Papilionoideae,
549 Leguminosae). *Braz J Biol* 43:575–587. <http://dx.doi.org/10.1007/s40415-020-00631-6>
- 550 Oyebanji O, Zhang R, Chen SY, Yi T-S (2020) New Insights Into the Plastome Evolution of
551 the Millettoid/Phaseoloid Clade (Papilionoideae, Leguminosae). *Front Plant Sci* 11:151.
552 <https://doi.org/10.3389/fpls.2020.00151>
- 553 Pälke H, Norris RD, Herrle JO, Wilson PA, Coxall HK, Lear CH (2006) The heartbeat of the
554 Oligocene climate system. *Science* 314:1894–1898. <https://doi.org/10.1126/science.113382>
- 555 Pennington RT, Stirton CH, Schrire BD (2005) Tribe Sophoreae. In: Lewis G, Schrire B,
556 Mackinder B, Lock M (eds) *Legumes of the World*. Royal Botanic Gardens, Kew, UK, pp 227–
557 249

- 558 Polhill RM, Raven PH (1981) *Advances in Legume Systematics, Part 1*. Royal Botanic
559 Gardens, Kew, UK
- 560 Queiroz LP, Pastore JFB, Cardoso D, Snak C, Lima ALC, Gagnon L, Vatanparast M, Holland
561 AE, Egan AN (2015) A multilocus phylogenetic analysis reveals the monophyly of a
562 recircumscribed papilionoid legume tribe Diocleae with well-supported generic relationships.
563 *Mol Phylogenet Evol* 90:1-19. <https://doi.org/10.1016/j.ympev.2015.04.016>
- 564 Sabir J, Schwarz E, Ellison N, Zhang J, Baeshen NZ, Mutwakil M, Jansen R, Ruhlman T (2014)
565 Evolutionary and biotechnology implications of plastid genome variation in the inverted-repeat-
566 lacking clade of legumes. *Plant Biotechnol J* 12:743–54. <https://doi.org/10.1111/pbi.12179>
- 567 Santos V, Almeida C (2019) The complete chloroplast genome sequences of three *Spondias*
568 species reveal close relationship among the species. *Genet Mol Biol* 42:132–138.
569 <https://doi.org/10.1590/1678-4685-GMB-2017-0265>
- 570 Schwarz EM, Ruhlman TA, Sabir JSM, Hajrah NH, Alharbi NS, Al-Malki AL, Bailey CD,
571 Jansen RK (2015) Plastid genome sequences of legumes reveal parallel inversions and multiple
572 losses of rps16 in papilionoids. *J Syst Evol* 53:458–68. <https://doi.org/10.1111/jse.12179>
- 573 Schrire, BD. 2005. Phaseoleae, In: Lewis G, Schrire B, Mackinder B, Lock M (eds) *Legumes*
574 of the world. Royal Botanic Gardens, Kew, UK, pp 393–431
- 575 Singh RM (2019) Cytogenetics and genetic introgression from wild relatives in soybean.
576 *Nucleus* 62:3–14. <https://doi.org/10.1007/s13237-019-00263-6>
- 577 Soltis PS, Soltis DE (2020) Plant genomes: markers of evolutionary history and drivers of
578 evolutionary change. *Plants People Planet* 3:74-82. <https://doi.org/10.1002/ppp3.10159>
- 579 Song Y, Chen Y, Lv J, Zhu S, Li M (2019) Comparative chloroplast genomes of *Sorghum*
580 species: sequence divergence and phylogenetic relationships. *Biomed Res Int* 1:11.
581 <https://doi.org/10.1155/2019/5046958>
- 582 Sprent JI (2009) *Legume Nodulation: A Global Perspective*. Wiley-Blackwell, Oxford, UK
- 583 Sprent JI (2008) Evolution and diversity of legumes symbiosis. In: Dilworth MJ, James EK,
584 Sprent JI, Newton WE (eds) *Nitrogen fixing Leguminous Symbioses*. Springer, The
585 Netherlands, pp 1-21
- 586 Stefanovic S, Pfeil BE, Palmer JD, Doyle JJ (2009) Relationships among phaseoloid legumes
587 based on sequences from eight chloroplast regions. *Syst Bot* 34:115-128.
588 <https://doi.org/10.1600/036364409787602221>
- 589 Tian S, Lu P, Zhang Z, Wu JQ, Zhang H, Shen H (2021) Chloroplast genome sequence of
590 Chongming lima bean (*Phaseolus lunatus* L.) and comparative analyses with other legume
591 chloroplast genomes. *BMC Genom* 22:194. <https://doi.org/10.1186/s12864-021-07467-8>
- 592 Van Wyk, B-E (2003) The value of chemosystematics in clarifying relationships in the
593 Genistoid tribes of papilionoid legumes. *Biochem Syst Ecol* 31:875–884.
594 [https://doi.org/10.1016/S0305-1978\(03\)00083-8](https://doi.org/10.1016/S0305-1978(03)00083-8)
- 595 Vatanparast M, Powell A, Doyle JJ, Egan AN (2018) Targeting legume loci: a comparison of
596 three methods for target enrichment bait design in Leguminosae phylogenomics. *APPS*
597 6:e1036. <https://doi.org/10.1002/aps3.1036>
- 598 Verdú M (2002) Age at maturity and diversification in woody angiosperms. *Evol* 56:1352–
599 1361. <https://doi.org/10.1111/j.0014-3820.2002.tb01449.x>

- 600 Wade BS, Pälke H (2004) Oligocene climate dynamics. *Paleoceanography* 19, PA4019.
601 <https://doi.org/10.1029/2004PA001042>
- 602 Wicke S, Schneeweiss GM, Pamphilis CW, Muller KF, Quandt D (2011) The evolution of the
603 plastid chromosome in land plants: Gene content, gene order, gene function. *Plant Mol Biol*
604 76:273–97. <https://doi.org/10.1007/s11103-011-9762-4>
- 605 Wink M, Mohamed GIA (2003) Evolution of chemical defense traits in the Leguminosae:
606 mapping of distribution patterns of secondary metabolites on a molecular phylogeny inferred
607 from nucleotide sequences of the *rbcL* gene. *Biochem Syst Ecol* 31:897–917.
608 [https://doi.org/10.1016/S0305-1978\(03\)00085-1](https://doi.org/10.1016/S0305-1978(03)00085-1)
- 609 Wojciechowski MF, Lavin M, Sanderson MJ (2004) A phylogeny of legumes (Leguminosae)
610 based on analysis of the plastid *matK* gene resolves many well-supported subclades within the
611 family. *Am J Bot* 11:1846–1862. <https://doi.org/10.3732/ajb.91.11.1846>
- 612 Wojciechowski MF (2013) Towards a new classification of Leguminosae: Naming clades using
613 non-Linnaean phylogenetic nomenclature. *S Afr J Bot* 89:85–93.
614 <https://doi.org/10.1016/j.sajb.2013.06.017>
- 615 Yang Z, Wang G, Ma Q, Liang L, Zhao T (2019) The complete chloroplast genomes of three
616 Betulaceae species: implications for molecular phylogeny and historical biogeography. *PeerJ*
617 7:e6320. <https://doi.org/10.7717/peerj.6320>
- 618 Zachos J, Pagani M, Sloan L, Thomas E, Billups K (2001) Trends, rhythms, and aberrations in
619 global climate 65 Ma to present. *Science* 292:686-93. <https://doi.org/10.1126/science.10594>
- 620 Zhang R, Wang, Y-H, Jin J-J, Stull GW, Bruneau A, Cardoso D, Queiroz LP, Moore MJ, Zhang
621 S-D, Chen S-Y, Wang J, Li D-Z, Yi T-S (2020) Exploration of Plastid Phylogenomic Conflict
622 Yields New Insights into the Deep Relationships of Leguminosae. *Syst Biol* 4:613–622.
623 <https://doi.org/10.1093/sysbio/syaa013>
- 624 Zhao Y, Zhang R, Jiang K-W, Qi J, Hu Y, Guo J, Zhu R, Zhang T, Egan AN, Yi T-S, Huang
625 C-H, Ma H (2021) Nuclear phylotranscriptomics and phylogenomics support numerous
626 polyploidization events and hypotheses for the evolution of rhizobial nitrogen fixing symbiosis
627 in Fabaceae. *Mol Plant* 14: 1–26. <https://doi.org/10.1016/j.molp.2021.02.006>

Tabela 1 Características das sequências do genoma plastidial de espécies da subfamília Papilionoideae e *outgroup* utilizadas nesta análise

¹ Espécies	Ref. Seq	PB	GC (%)	Prot.	rRNA	tRNA	Genes	Crom. (n)	Tribo
<i>Abrus precatorius</i>	NC_047402.1	156500	35,4	71	8	37	116	11	Abreae
<i>Abrus pulchellus</i>	MT328396.1	156497	35,4	72	8	37	117	11	Abreae
<i>Ammopiptanthus mongolicus</i>	KY034453.1	153935	36,8	71	8	37	116	9	Sophoreae
<i>Amorpha californica</i> var. <i>napensis</i>	OK274088.1	158294	36,0	66	8	37	111	10	Amorpheae
<i>Amorpha fruticosa</i>	NC_047310.1	158245	36,0	66	8	37	111	20	Amorpheae
<i>Amorpha roemeriana</i>	MW628937.1	158322	36,0	66	8	37	111	10*	Amorpheae
<i>Arachis duranensis</i>	NC_047337.1	156318	36,4	66	8	37	111	10	Dalbergieae
<i>Arachis hypogaea</i>	NC_037358.1	156391	36,4	73	8	37	118	20	Dalbergieae
<i>Arachis ipaensis</i>	MK144829.1	156453	36,3	71	8	37	116	10	Dalbergieae
<i>Arachis monticola</i>	MK144818.1	156395	36,3	72	8	37	117	20	Dalbergieae
<i>Angylocalyx braunii</i>	NC_047392.1	162989	35,5	73	8	37	118	13-14*	Angylocalyceae
<i>Astragalus gummifer</i>	NC_047251.1	123425	34,1	74	4	30	108	32	Galegeae

SILVA, F. G. Filogenia molecular e análise *in silico* do inibidor de protease Bowman-Birk de espécies da subfamília Papilionoideae

<i>Astragalus laxmannii</i>	NC_052923.1	122844	34,1	74	4	30	108	8*	Galegeae
<i>Astragalus strictus</i>	MT120746.1	122796	34,2	74	4	30	108	16	Galegeae
<i>Cajanus cajan</i>	NC_031429.1	153841	34,9	70	8	37	115	11	Phaseoleae
<i>Cajanus scarabaeoides</i>	KU729878.1	152201	35,0	71	8	37	116	11	Phaseoleae
<i>Callerya nitida</i>	MT120748.1	132319	33,9	74	4	30	108	16	Millettieae
<i>Caragana arborescens</i>	MT211962.1	129473	34,3	78	4	31	113	8	Caraganeae
<i>Caragana jubata</i>	MT211963.1	128132	34,4	79	4	30	113	8	Caraganeae
<i>Caragana korshinskii</i>	KX289923.1	129331	34,4	78	4	31	113	8	Caraganeae
<i>Carmichaelia australis</i>	MF597719.1	122805	34,3	74	4	30	108	15	Galegeae
<i>Centrosema pubescens</i>	NC_057278.1	154141	34,8	62	8	37	107	10	Phaseoleae
<i>Cicer arietinum</i>	NC_011163.1	125319	33,9	83	4	30	117	8	Cicereae
<i>Clitoria ternatea</i>	NC_047365.1	151673	34,5	67	8	37	112	8	Phaseoleae
<i>Crotalaria pallida</i>	MT920364.1	152658	36,7	73	8	37	118	8	Crotalariaeae
<i>Cullen corylifolium</i>	NC_042700.1	153114	35,3	71	8	37	116	10	Psoraleeae
<i>Dahlstedtia araripensis</i>	MN966622.1	153217	34,8	71	8	37	116	11*	Millettieae
<i>Dalbergia cultrata</i>	NC_044117.1	156385	36,1	70	8	38	116	10	Dalbergieae
<i>Dalbergia hainanensis</i>	NC_036961.1	156073	36,1	71	8	38	117	10	Dalbergieae
<i>Dalbergia odorifera</i>	NC_049008.1	156049	36,1	71	8	38	117	10	Dalbergieae
<i>Dipteryx alata</i>	MT119080.2	158647	35,0	76	8	36	120	8	Dipterygeae
<i>Dussia macropphyllata</i>	NC_047341.1	167738	35,0	65	8	38	111	13	Amburaneae
<i>Ficus carica</i>	NC_035237.1	160602	35,8	66	8	37	111	13	Ficeae
<i>Genista tinctoria</i>	NC_047375.1	152222	36,8	74	8	37	119	20	Genisteae
<i>Glycine canescens</i>	NC_021647.1	152518	35,3	68	8	37	113	20	Phaseoleae
<i>Glycine cyrtoloba</i>	NC_021645.1	152381	35,3	71	8	37	116	20	Phaseoleae
<i>Glycine dolichocarpa</i>	NC_021648.1	152804	35,3	71	8	37	116	40	Phaseoleae
<i>Glycine falcata</i>	NC_021649.1	153023	35,3	71	8	37	116	20	Phaseoleae
<i>Glycine gracilis</i>	NC_030329.1	152218	35,4	71	8	37	116	20	Phaseoleae
<i>Glycine max</i>	NC_007942.1	152218	35,4	71	8	37	116	20	Phaseoleae
<i>Glycine soja</i>	NC_022868.1	152217	35,4	71	8	38	116	20	Phaseoleae
<i>Glycine stenophita</i>	NC_021646.1	152618	35,3	71	8	37	116	20	Phaseoleae
<i>Glycine syndetika</i>	NC_021650.1	152783	35,3	71	8	37	116	20	Phaseoleae
<i>Glycine tomentella</i>	KC893633.1	152728	35,4	71	8	37	116	40	Phaseoleae
<i>Goodia macrocarpa</i>	NC_047331.1	154333	35,9	72	8	37	117	-	Bossiaeeae
<i>Gossypium hirsutum</i>	HQ901196.1	160256	37,2	65	8	37	110	26	Gossypiae
<i>Haymondia wallichii</i>	NC_052910.1	153668	35,4	71	8	37	116	-	Phaseoleae
<i>Hylodesmum podocarpum</i>	MG867568.1	149564	35,2	68	8	38	114	11	Desmodieae
<i>Indigofera linifolia</i>	NC_047353.1	160040	35,8	74	8	37	119	8	Indigofereae
<i>Indigofera tinctoria</i>	NC_026680.1	158367	35,8	73	8	37	118	8	Indigofereae
<i>Kummerowia striata</i>	NC_044114.1	148450	35,0	68	8	37	113	11	Desmodieae
<i>Lathyrus clymenum</i>	NC_027148.1	121263	34,9	77	4	30	111	7	Fabeae
<i>Lathyrus odoratus</i>	NC_027150.1	120289	35,2	72	4	30	106	7	Fabeae
<i>Lathyrus sativus</i>	NC_014063.1	121020	35,1	73	4	30	107	7	Fabeae
<i>Lens culinaris</i>	KF186232.1	123096	34,4	72	4	30	106	7	Fabeae
<i>Lens culinaris cult. Northfield</i>	NC_027152.1	122967	34,4	73	4	30	107	7	Fabeae
<i>Lespedeza bicolor</i>	NC_046836.1	148930	35,0	67	8	37	112	11	Desmodieae

SILVA, F. G. Filogenia molecular e análise *in silico* do inibidor de protease Bowman-Birk de espécies da subfamília Papilionoideae

<i>Lespedeza davurica</i>	NC_042748.1	147151	34,7	70	8	35	113	21	Desmodieae
<i>Lespedeza maritima</i>	NC_044115.1	149022	35,0	68	8	37	113	11*	Desmodieae
<i>Lotus corniculatus</i>	MT528596.1	150700	36,0	80	8	37	125	6	Loteae
<i>Lotus japonicus</i>	NC_002694.1	150519	36,0	72	8	37	117	12	Loteae
<i>Lupinus albus</i>	NC_026681.1	154140	36,5	75	8	37	120	25	Genisteae
<i>Lupinus atlanticus</i>	KU726827.1	152272	36,7	73	8	37	118	19	Genisteae
<i>Lupinus luteus</i>	NC_023090.1	151894	36,6	74	8	37	119	26	Genisteae
<i>Maackia floribunda</i>	NC_034774.1	154541	36,5	75	8	37	120	9	Sophoreae
<i>Malus domestica</i>	MK434916.1	159926	36,6	66	8	37	111	17	Maleae
<i>Medicago falcata</i>	NC_032066.1	124430	34,0	83	4	28	115	9	Trifolieae
<i>Medicago hybrida</i>	NC_027153.1	125208	33,8	75	4	30	109	8	Trifolieae
<i>Medicago papillosa</i>	NC_027154.1	125203	33,8	66	4	26	96	8	Trifolieae
<i>Medicago sativa</i>	NC_042841.1	125330	33,9	79	4	29	112	16	Trifolieae
<i>Millettia dura</i>	MN966639.1	152714	34,9	75	8	37	120	8	Millettieae
<i>Millettia pinnata</i>	JN673818.2	152968	34,8	74	8	37	119	8	Millettieae
<i>Mirbelia oxylobioides</i>	NC_047371.1	151897	36,8	68	8	35	111	8	Mirbelieae
<i>Mucuna macrocarpa</i>	NC_044116.1	153826	35,1	70	8	37	115	11	Phaseoleae
<i>Myroxylon balsamum</i>	MW628963.1	179362	36,4	75	8	38	121	13	Amburaneae
<i>Orbexilum onobrychis</i>	NC_047389.1	152378	35,5	71	8	37	116	11*	Psoraleeae
<i>Ormosia boluoensis</i>	MN886968.1	175760	35,5	94	8	37	139	8*	Ormosieae
<i>Ormosia formosana</i>	MT258921.1	173587	35,8	91	8	37	136	8	Ormosieae
<i>Ormosia hosiei</i>	NC_039418.1	171642	35,8	93	8	37	138	8	Ormosieae
<i>Pachyrhizus erosus</i>	NC_026682.1	151947	35,3	72	8	37	117	11	Phaseoleae
<i>Phaseolus lunatus</i>	MW423611.1	150902	35,4	69	8	38	115	11	Phaseoleae
<i>Phaseolus vulgaris</i>	NC_009259	150285	35,4	70	8	37	115	11	Phaseoleae
<i>Pisum abyssinicum</i>	NC_037830.1	122174	34,8	77	4	30	111	7	Fabeae
<i>Pisum fulvum</i>	NC_036828.1	120837	34,9	74	4	30	108	7	Fabeae
<i>Pisum sativum</i>	NC_014057.1	122169	34,8	76	4	30	110	7	Fabeae
<i>Platylobium obtusangulum</i>	NC_045516.1	150090	36,7	71	8	37	116	8	Bossiaeeae
<i>Podalyria calyptrata</i>	MN966651.1	152992	36,6	73	8	37	118	7	Podalyrieae
<i>Poecilanthe parviflora</i>	NC_047393.1	160576	35,6	73	8	37	118	9	Brongniartieae
<i>Pterocarpus indicus</i>	NC_049082.1	158107	36,4	73	8	37	118	10	Dalbergieae
<i>Pterocarpus santalinus</i>	NC_049084.1	158966	36,3	74	8	37	119	11	Dalbergieae
<i>Pterocarpus tinctorius</i>	MH033829.1	159116	36,2	72	8	38	118	11*	Dalbergieae
<i>Pueraria edulis</i>	MN515038.1	153434	35,4	72	8	37	117	11*	Phaseoleae
<i>Pueraria montana var. lobata</i>	MT818508.1	153442	35,4	72	8	37	117	11	Phaseoleae
<i>Robinia pseudoacacia</i>	NC_026684.1	154835	35,9	71	8	37	116	11	Robinieae
<i>Sesbania cannabina</i>	MN105118.1	153978	35,6	71	8	36	115	12	Sesbanieae
<i>Sesbania grandiflora</i>	NC_047382.1	157345	35,4	70	8	37	115	12	Sesbanieae
<i>Sophora alopecuroides</i>	NC_036102.1	154108	36,5	74	8	37	119	18	Sophoreae
<i>Sophora flavescens</i>	MH748034.1	154378	36,6	74	8	37	119	9	Sophoreae
<i>Sophora tonkinensis</i>	NC_042688.1	155640	36,4	75	8	37	120	9	Sophoreae
<i>Sphaerophysa salsula</i>	MW122834.1	123300	34,1	71	4	30	105	8	Galegeae
<i>Stryphnodendron adstringens</i>	NC_044627.1	162169	35,9	72	8	37	117	13	Mimoseae
<i>Stylosanthes hamata</i>	NC_039159.1	156502	36,6	71	8	37	116	10	Dalbergieae
<i>Stylosanthes scabra</i>	NC_039160.1	156502	36,6	71	8	37	116	10	Dalbergieae

SILVA, F. G. Filogenia molecular e análise *in silico* do inibidor de protease Bowman-Birk de espécies da subfamília Papilionoideae

<i>Stylosanthes viscosa</i>	NC_039161.1	156244	36,6	70	8	37	115	10	Dalbergieae
<i>Templetonia retusa</i>	NC_047372.1	162216	35,3	69	8	37	114	8	Brongniartieae
<i>Tipuana tipu</i>	NC_047321.1	158362	36,4	73	8	37	118	10	Dalbergieae
<i>Theobroma cacao</i>	NC_014676.2	160619	36,9	64	8	37	109	10	Theobromateae
<i>Trifolium pratense</i>	NC_047412.1	146573	34,3	100	4	31	135	7	Trifolieae
<i>Trifolium repens</i>	NC_024036.1	132120	34,3	82	4	30	116	16	Trifolieae
<i>Trigonella foenum-graecum</i>	NC_042857.1	125645	33,9	84	4	31	119	8	Trifolieae
<i>Vavilovia formosa</i>	NC_047213.1	122196	34,9	83	4	31	118	7	Fabeae
<i>Vicia bungei</i>	MT362055.1	130796	34,8	88	4	30	122	12	Fabeae
<i>Vicia faba</i>	KF042344.1	123722	34,6	72	4	30	106	6	Fabeae
<i>Vicia sativa</i>	NC_027155.1	122467	35,2	76	4	28	108	6	Fabeae
<i>Vigna angularis</i>	NC_021091.1	151683	35,2	71	8	37	116	11	Phaseoleae
<i>Vigna mungo</i>	NC_050260.1	151294	35,2	72	8	37	117	11	Phaseoleae
<i>Vigna radiata</i>	NC_013843.1	151271	35,2	71	8	37	116	11	Phaseoleae
<i>Vigna unguiculata</i>	NC_018051.1	152415	35,2	74	8	37	119	11	Phaseoleae
<i>Wisteria brachybotrys</i>	MT120814.1	131101	34,3	70	5	30	105	8	Wisterieae
<i>Wisteria floribunda</i>	KM103376.1	130960	34,3	70	5	30	105	8*	Wisterieae
<i>Wisteria sinensis</i>	KT200359.1	130561	34,4	70	5	30	105	8	Wisterieae

¹Nome científico; código da sequência no NCBI (Ref. Seq.); número de pares de bases (PB); conteúdo de guanina e citosina (GC); genes que codificam proteínas (Prot.); RNA ribossômico (rRNA); RNA transportador (tRNA); número de genes (Genes); Cromossomos (Crom.). *Número cromossômico base para o gênero.

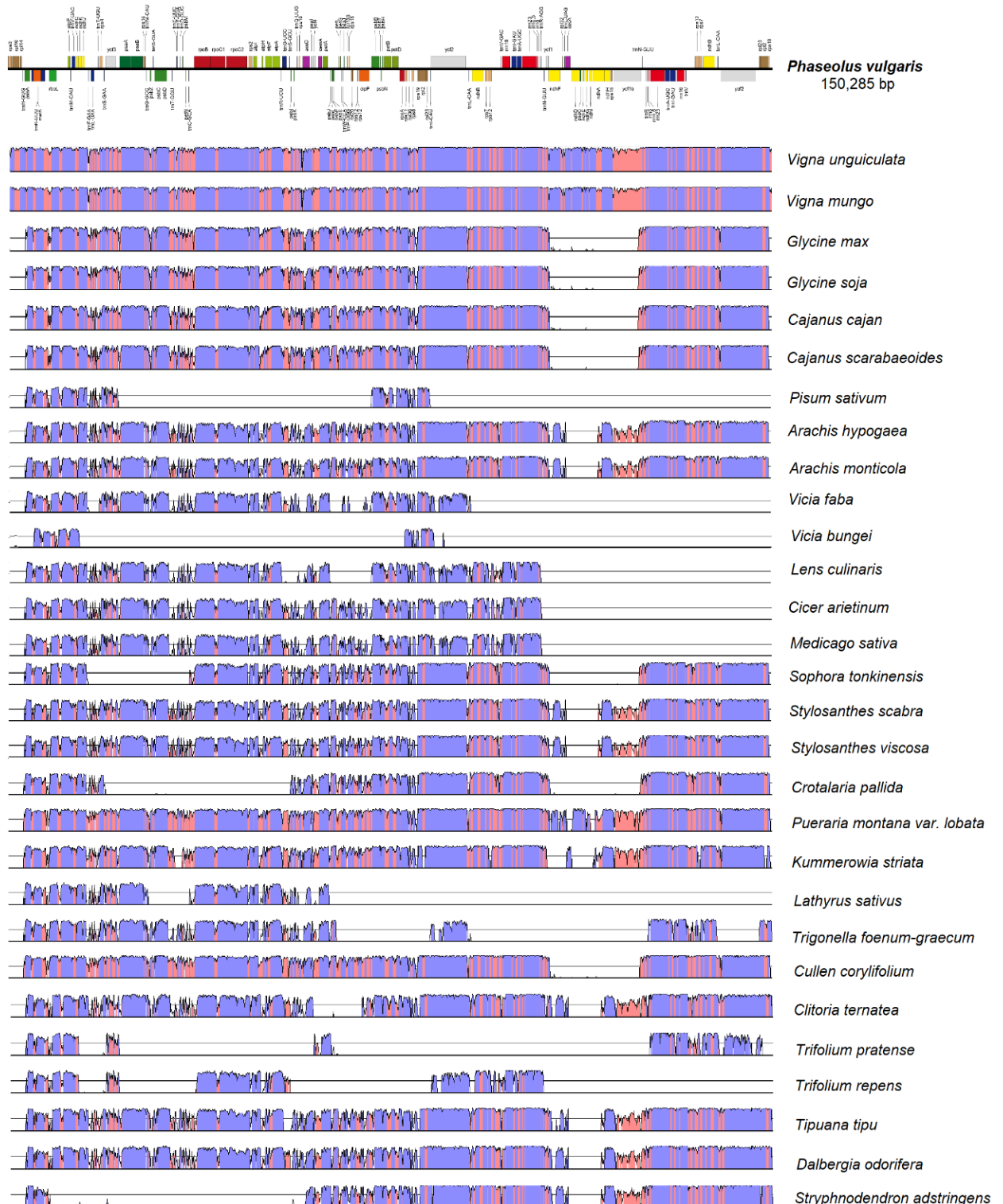


Fig. 1 Alinhamento de sequências completas de plastomas de 28 espécies da subfamília Papilionoideae. As regiões em azul representam as regiões codificantes, regiões na cor rosa indicam regiões não codificantes. Os gráficos da imagem foram obtidos pelo servidor mVISTA e pelo servidor OGDRAW. A espécie *S. adstringens* foi utilizada como grupo externo.

SILVA, F. G. Filogenia molecular e análise *in silico* do inibidor de protease Bowman-Birk de espécies da subfamília Papilionoideae

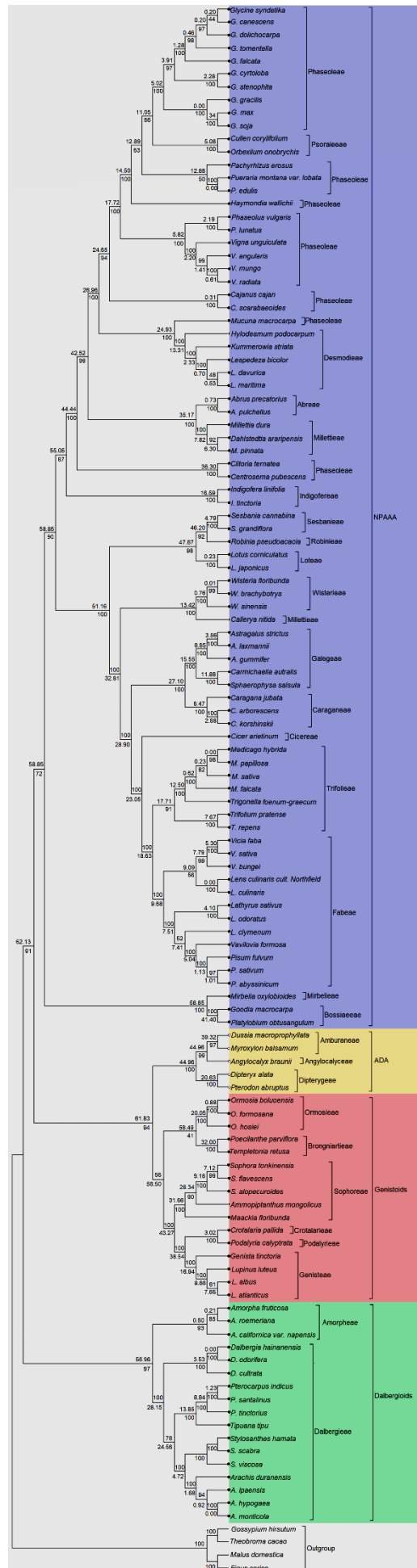


Fig. 2 Filogenia da subfamília Papilionoideae baseada em seis genes plastidiais (*matK*, *psaA*, *psbA*, *psbD*, *rbcL* e *rpoC2*) e tempo de divergência entre espécies da subfamília Papilionoideae utilizando seis pontos de calibração. As análises foram realizadas nos *softwares* TimeTree e MEGA7, através do método de máxima verossimilhança, usando suporte de *bootstrap* com 1000 replicatas. Os valores de suporte de bootstrap (%) são mostrados na interseção de cada ramo. O status de nodulação é mostrado com círculos: Preenchido, espécies nodulantes; vazio, espécies não nodulantes. x = Espécie sem relato para nodulação ou não nodulação.

Caracterização *in silico* e análise filogenética do inibidor de protease do tipo Bowman-Birk em espécies da tribo Phaseoleae (Fabaceae - Papilionoideae)

CAPÍTULO III

Artigo será submetido à Revista Brazilian Archives of Biology and Technology (Qualis A2)

1 Caracterização *in silico* e análise filogenética do 2 inibidor de protease do tipo Bowman-Birk em 3 espécies da tribo Phaseoleae (Fabaceae – 4 Papilionoideae)

5 Flávia Gomes da Silva^{1*}

6 <https://orcid.org/0000-0001-5413-5646>

Gheysa Coelho Silva¹

<https://orcid.org/0000-0002-47246441>

7 Rafael Trindade Maia²

8 <https://orcid.org/0000-0002-5870-1091>

9 Horace José Jimenez¹

10 <https://orcid.org/0000-0001-95568288>

11 Rômulo Maciel Moraes Filho¹

12 <https://orcid.org/0000-0003-39884379>

13 ¹ Universidade Federal Rural de Pernambuco, Departamento de Agronomia, Programa de Pós-
14 Graduação em Melhoramento Genético de Plantas, Recife, PE, Brasil.

15 ² Universidade Federal de Campina Grande, Centro de Desenvolvimento Sustentável do Semiárido,
16 Programa de Pós-Graduação em Ciências Naturais e Biotecnologia, Sumé, PB, Brasil.

17 *Correspondência: flavia.gomess12@gmail.com

18 DESTAQUES

- 19 • Foram utilizados métodos *in silico* para construção de modelos 3D do inibidor do tipo
20 Bowman-Birk (BBI) e análise filogenética das espécies da tribo Phaseoleae.
- 21
- 22 • Foi obtido o modelo 3D do inibidor BBI estável e dentro dos padrões.
- 23
- 24 • Os BBIs das espécies da tribo Phaseoleae possuem dois domínios funcionais.
- 25
- 26 • A filogenia está de acordo com a classificações taxonômicas mais recente para as
27 espécies da tribo Phaseoleae.

28 **Resumo:** A subfamília Papilionoideae (Fabaceae) possui aproximadamente 32 tribos, sendo
29 a tribo Phaseoleae a maior, com cerca de 90 gêneros e 1.600 espécies, distribuídas em sete
30 subtribos. Essa tribo possui uma grande diversidade de espécies de importância econômica
31 como fonte alimento humano e animal. Foi realizado uma caracterização estrutural e
32 funcional, bem como o desenvolvimento de modelos tridimensionais do inibidor de protease
33 do tipo Bowman-Birk (BBI), que é reconhecida por sua associação com funções de atividade
34 do inibidor de duas serinoproteases de especificidade diferentes, onde o primeiro sítio de ação
35 inibi a tripsina, e o segundo a quimotripsina. Foram examinadas 23 sequências de BBI de
36 espécies da tribo Phaseoleae, disponíveis em bancos de dados públicos. Uma característica
37 hidrofílica do BBI foi encontrada para todas as espécies, e o ponto isoelétrico demonstrou um
38 caráter ligeiramente ácido (5,18 a 6,80). Também foi identificado dois domínios funcionais nas
39 sequências BBI. A análise de agrupamento por Máxima Verossimilhança foi consistente com

40 as filogenias mais recentes das espécies analisadas.

41 **Palavras-chave:** Bioinformática, Análise computacional, Modelagem por homologia,
42 Serinoprotease

43 **Abstract:** The subfamily Papilionoideae (Fabaceae) has approximately 32 tribes, with the
44 Phaseoleae tribe being the largest, with about 90 genera and 1,600 species, distributed in
45 seven subtribes. This tribe has a great diversity of species of economic importance as a source
46 of human and animal food. A structural and functional characterization was performed, as well
47 as the development of three-dimensional models of the Bowman-Birk-type protease inhibitor
48 (BBI), which is recognized for its association with inhibitor activity functions of two serine
49 proteases of different specificity, where the first action site inhibits trypsin, and the second
50 chymotrypsin. We examined 23 BBI sequences from Phaseoleae species, available in public
51 databases. A hydrophilic characteristic of BBI was found for all species, and the isoelectric
52 point demonstrated a slightly acidic character (5.18 to 6.80). Two functional domains were also
53 identified in the BBI sequences. Cluster analysis by Maximum Likelihood was consistent with
54 the most recent phylogenies of the analyzed species.

55 **Keywords:** Bioinformatics; Computational analysis; Homology modeling; Serineproteases

56 INTRODUÇÃO

57 A subfamília Papilionoideae possui 32 tribos [67], sendo a tribo Phaseoleae s.l. a maior,
58 com aproximadamente 90 gêneros e 1.600 espécies, subdivididas em sete subtribos [72].
59 Além disso, essa tribo se destaca por sua grande importância econômica [64,78], por englobar
60 os principais grãos cultivados para alimentação humana e animal, como os feijões comum
61 (*Phaseolus vulgaris*), caupi (*Vigna unguiculata*) [6,72], guandu (*Cajanus cajan*) [72] e a soja
62 (*Glycine max*) [6,72], pertencentes as subtribos Phaseolinae, Cajaninae e Glycininae,
63 respectivamente.

64 Os integrantes dessa tribo possuem ampla distribuição global [81,86], sendo as espécies
65 da subtribo Phaseolinae distribuídas do Velho (continentes europeu, africano, asiático e os
66 quatro arquipélagos da Macaronésia) ao Novo Mundo (continente americano) [26-27]. Já a
67 maioria das espécies da subtribo Cajaninae são asiáticas [72], sendo a subtribo Glycininae
68 encontrada na Ásia, parte da Rússia a Austrália [75]. Além disso, são caracterizadas pela
69 presença de folhas primárias (eófilos) simples e opostas, folhas trifolioladas de base
70 assimétrica, corola papilionácea, inflorescências em pseudorracemos, grão de pólen
71 recoberto por uma espessa camada de endexina, presença de estipelas (pequena estípula),
72 número cromossômico básico de 11 [11,30,53].

73 Diversos representantes de leguminosas (Fabaceae) se sobressaem por conterem o
74 inibidor de protease do tipo Bowman-Birk (BBI) [20], como é o caso de espécies da tribo
75 Phaseoleae (Papilionoideae). BBI são moléculas proteicas com dois domínios inibitórios
76 diferentes, onde geralmente o primeiro sítio ativo é específico para tripsina, e o segundo inibe
77 a quimotripsina ou elastase [56,68]. Os BBIs foram assim nomeados porque em 1946, Donald
78 E. Bowman [10] isolou o primeiro BBI na soja (*G. max*), e posteriormente foi caracterizado em
79 1963 por [9]. Esses BBIs são encontrados abundantemente em sementes e tubérculos de
80 leguminosas (Fabaceae), Solanaceae, Cucurbitaceae e Poaceae [23].

81 Diversas funções dos BBIs nas plantas são mencionadas na literatura, sendo a ação
82 inseticida a mais relatada. Essa função está intimamente relacionada com sua atividade nas
83 proteases intestinais dos insetos, que possibilita a redução da assimilação de aminoácidos,
84 que pode retardar o desenvolvimento dos insetos que se alimentam dessas plantas [25].
85 Determinados BBIs também podem favorecer o efeito microbiano nas plantas [47], assim
86 como na tolerância a estresses abióticos, tais como salino [74] e hídrico [29].

87 Outros estudos evidenciam funções dos BBIs na saúde humana, tais como propriedades
88 anti-inflamatórias em doenças intestinais [42,54,71,83,87] e anticancerígenas [3,15,31,47,63].
89 Foi comprovado que os dois sítios de ligação do BBI (antitripsina e anti-quimotripsina) estão

90 intimamente ligados ao efeito antiproliferativo nas células cancerígenas do cólon [21].
91 Com o crescimento no número de sequências proteicas nos bancos de dados,
92 conseqüentemente o número de sequências funcionalmente relacionadas de um mesmo
93 grupo de proteína, vem crescendo de forma exponencial. Dessa maneira, se faz cada vez
94 mais necessário que as relações entre os membros de uma mesma família proteica sejam
95 analisadas [88]. Nesse sentido, análises filogenéticas podem auxiliar a entender as relações
96 de grupos de proteína nos diversos grupos de organismos, possibilitando a compreensão de
97 como essas proteínas estão relacionadas em diferentes espécies, e se elas evoluíram a partir
98 de um único ancestral comum [45]. Para ajudar neste entendimento a previsão da estrutura
99 tridimensional da proteína também é realizada, o que permite um melhor entendimento de
100 suas estruturas e funções moleculares.

101 Cada vez mais, estudos têm sido realizados para elucidar a estrutura [5,68-69], relações
102 evolutivas e filogenéticas do BBI [24,38,61], onde os resultados filogenéticos estão
103 continuamente sendo aperfeiçoados devido à crescente disponibilidade de uma grande
104 quantidade de dados biológicos e do surgimento de novos avanços nos métodos para analisá-
105 los [2,45].

106 Dessa forma os objetivos desta pesquisa foram: (1) Caracterizar, comparar e identificar
107 domínios conservados em sequências do inibidor de protease do tipo Bowman-Birk de
108 diferentes espécies da tribo Phaseoleae (Fabaceae - Papilionoideae) disponíveis em bancos
109 de dados públicos; (2) Realizar uma análise filogenética com base nas sequências
110 encontradas, para entender as suas relações. Desenvolver modelos tridimensionais de
111 representantes da tribo, baseados na metodologia de modelagem por homologia.

112 **MATERIAIS E MÉTODOS**

113 **Recuperação das sequências**

114 As sequências do BBI foram obtidas por meio do algoritmo de busca BLAST (*Basic*
115 *Local Alignment Search Tool*) disponível na plataforma NCBI (*National Center for*
116 *Biotechnology Information*), sendo realizado o *download* no formato FASTA. As sequências
117 de BBI em espécies da subfamília Papilionoideae foram recuperadas pela ferramenta
118 BLASTp, que procura as sequências homólogas no banco de dados, onde um total de 23
119 sequências foram selecionadas (Tabela 1).

120 **Análise das sequências**

121 O servidor ProtParam (<http://web.expasy.org/protparam>) [33] analisou os parâmetros
122 físico-químicos de BBI presentes em espécies da tribo Phaseoleae (subfamília
123 Papilionoideae). A presença de sítios de clivagem de peptídeos de sinal foi analisada por meio
124 do servidor TOPCONS (<http://topcons.cbr.su.se/pred>) [82]. Para identificação dos domínios
125 funcionais da proteína foi utilizado o servidor Prodom (<http://prodom.prabi.fr/prodom/>) [73]
126 para sua classificação e ontologia. Foi estimado os efeitos funcionais ocasionados por
127 mutações de sequências de aminoácidos por intermédio do servidor SNAP2
128 (<https://roslab.org/services/snap2web/>) [35].

129 **Alinhamento e Análise filogenética**

130 As sequências proteicas do BBI foram alinhadas utilizando o algoritmo ClustalW e a
131 árvore filogenética foi produzida no *software* MEGA 7.0 [52]. A árvore filogenética foi
132 desenvolvida utilizando o método de Máxima Verossimilhança (ML), com suporte estatístico
133 calculado pelo método de *bootstrap* (BS) com 1000 replicatas.

134 **Modelagem, avaliação e validação da estrutura terciária do modelo**

135 A predição da estrutura terciária dos BBI foi realizada pelo servidor GalaxyTBM

136 (<https://galaxy.seoklab.org/cgi-bin/submit.cgi?type=TBM>) [51] no modo Template Based
 137 Modelling. A visualização da estrutura 3D foi construída por meio do pacote UCSF Chimera
 138 [66]. Foi utilizado o servidor SAVES v6.0
 139 (<https://saves.mbi.ucla.edu/results?job=1294919&p=procheck>) para verificar a qualidade do
 140 modelo pela análise do gráfico de Ramachandran. O cálculo do Z-score foi executado no
 141 servidor interativo ProSA-web (<https://prosa.services.come.sbg.ac.at/prosa.php>), para
 142 identificar erros em estruturas tridimensionais [89].

143 RESULTADOS E DISCUSSÃO

144 Com base na análise da estrutura primária e localização subcelular das sequências do
 145 BBI de espécies da tribo Phaseoleae (Tabela 1), os locais conservados de peptídeos de sinal
 146 variaram de 22 a 30. O peptídeo de sinal é uma sequência curta e transitória que tem por
 147 função direcionar as proteínas para secreção ou para transferência para suas organelas
 148 características e para posterior processamento [19]. Após o direcionamento da proteína para
 149 a via de transporte correta, a mesma é clivada por peptidases de sinal [65].

150 Houve uma variação no tamanho dos aminoácidos de BBI de 93 aa (*Glycine microphylla*)
 151 a 121 aa (*Vigna subterranea*). O pequeno tamanho das sequências de BBI, normalmente
 152 entre 60 e 90 o diferencia de outras famílias de inibidores de protease [39]. Tiessen *et al.* [79],
 153 ressaltam que a composição de aminoácidos é influenciada pelo tamanho da proteína, dessa
 154 forma os aminoácidos das proteínas pequenas tendem a serem diferentes das proteínas
 155 maiores [79]. Para o peso molecular, obteve-se uma variação de 10.23kDa (*G. microphylla*) a
 156 13.19kDa (*P. microcarpus*), essa informação pode ser aplicada para isolar proteínas, a partir
 157 de técnicas de cromatografia e eletroforese [90].

158
 159 **Tabela 1.** Análise da estrutura primária e localização subcelular das sequências do inibidor
 160 Bowman-Birk de espécies da tribo Phaseoleae (Fabaceae - Papilionoideae).

Espécie	GB-ID	SCPS	Nº aa	PM (kDa)	pl	GRAVY
<i>Phaseolus microcarpus</i>	CAL64060.1	29-30	120	13.19	5.69	-0.222
<i>Phaseolus zimapanensis</i>	CAQ52360.1	29-30	120	13.11	5.61	-0.197
<i>Phaseolus filiformis</i>	CAL69282.1	28-29	120	13.09	5.69	-0.222
<i>Phaseolus grayanus</i>	CAQ52359.1	29-30	120	13.13	5.69	-0.233
<i>Phaseolus hintonii</i>	CAQ52357.1	29-30	119	13.02	5.22	-0.199
<i>Phaseolus oligospermus</i>	CAL51270.1	29-30	120	13.07	6.08	-0.166
<i>Phaseolus augusti</i>	CAL51269.1	28-29	120	13.03	5.61	-0.187
<i>Phaseolus glabellus</i>	CAL69238.1	29-30	120	13.04	5.18	-0.154
<i>Phaseolus lunatus</i>	CAL51268.1	28-29	120	13.09	5.61	-0.194
<i>Phaseolus coccineus subsp. polyanthus</i>	CAQ52820.1	29-30	120	13.05	5.86	-0.232
<i>Phaseolus vulgaris</i>	CAM34246.1	28-29	120	13.10	6.22	-0.232
<i>Phaseolus coccineus subsp. coccineus</i>	CAL69277.1	28-29	117	12.82	5.69	-0.195
<i>Phaseolus costaricensis</i>	CAL69280.1	28-29	120	13.04	5.44	-0.182
<i>Phaseolus maculatus</i>	CAL51271.1	28-29	120	13.03	5.86	-0.137
<i>Phaseolus acutifolius var. latifolius</i>	CAL49456.1	29-30	118	12.92	5.79	-0.204
<i>Vigna marina</i>	ABD97867.1	21-22	112	12.16	6.22	-0.072
<i>Vigna mungo</i>	AAK97766.1	21-22	103	11.31	6.80	-0.085
<i>Vigna subterranea</i>	AHY03234.1	29-30	121	13.15	6.50	-0.214
<i>Glycine soja</i>	XP_028181512.1	27-28	114	12.49	5.44	-0.108
<i>Glycine max</i>	NP_001236353.2	27-28	114	12.48	5.42	-0.108
<i>Glycine microphylla</i>	AAO89510.1	-	93	10.23	5.41	-0.567
<i>Rhynchosia sublobata</i>	ALA09300.2	22-23	112	12.41	5.44	-0.304
<i>Cajanus cajan</i>	KYP42282.1	21-22	101	11.22	6.00	-0.155

161 GB-ID: Identificação do Genbank, SCSP: Locais Conservados de Peptídeos de Sinal, Nº aa: Número
 162 de aminoácidos, PM: Peso Molecular, GRAVY: Grande Média de Hidropaticidade: pl: Ponto Isoelétrico.
 163

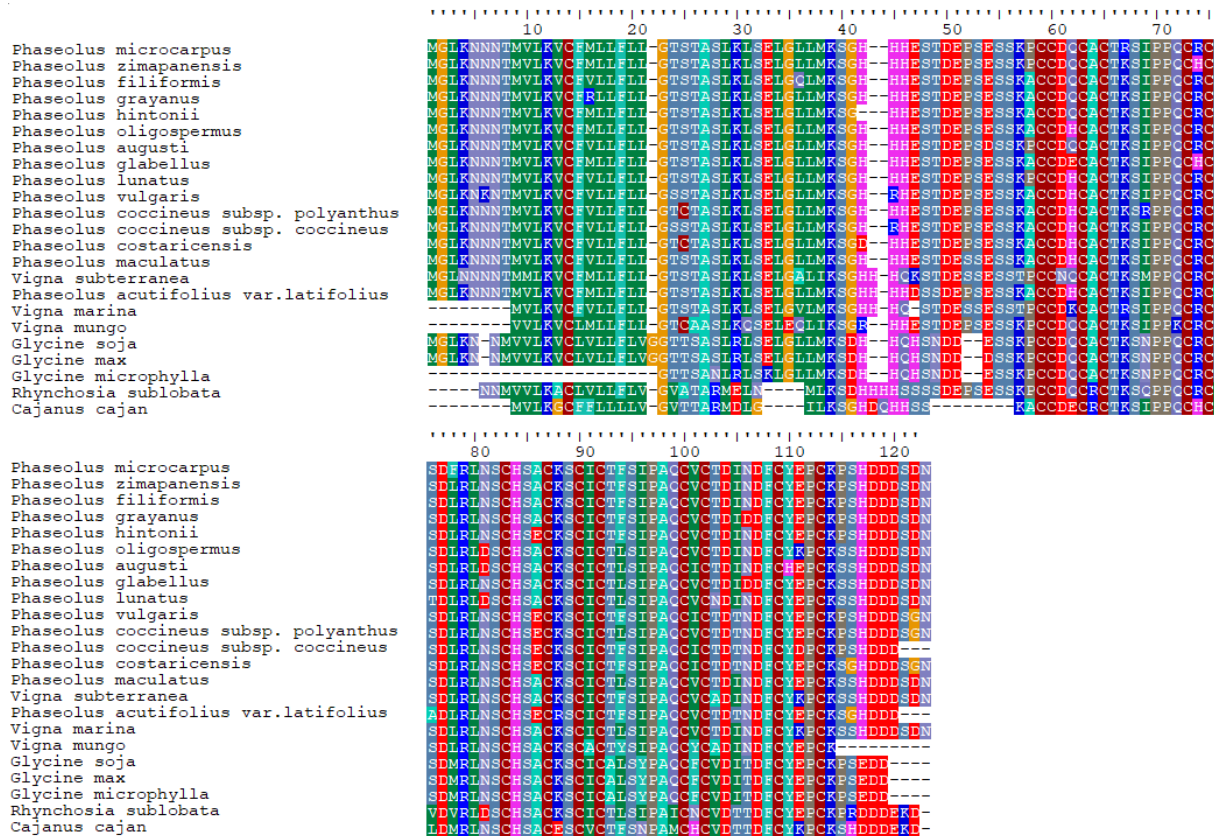
164 O ponto isoelétrico (pl) atestou um caráter ligeiramente ácido, que variou de 5,18 (*P.*

165 *glabellus*) a 6,80 (*Vigna mungo*). Na literatura são observados valores de pI de proteínas
 166 distintas, variando de valores altamente ácidos a demasiadamente alcalinos de 4,0 a 12,0
 167 [80]. Essa variação está associada a diferença entre os resíduos de aminoácidos de diversos
 168 BBI [57]. Kiraga *et al.* [48], também evidenciaram que o pI pode ser amplamente variado, a
 169 depender das inserções e deleções entre os ortólogos de proteínas e da ecologia do
 170 organismo.

171 Essa informação é muito significativa para se conhecer a dissolubilidade e a localização
 172 subcelular de proteínas, que indica ser um recurso importante para prever a sua carga líquida
 173 em um determinado pH e compreender as interações entre proteínas, proteínas e membranas
 174 (fosfolípidios), como também para indicar a existência de isoformas proteicas [77]. O
 175 conhecimento do pI de uma proteína é um parâmetro importante para muitas análises
 176 bioquímicas e proteômicas, como eletroforese em gel 2D-PAGE [8,28], focalização isoeletrica
 177 capilar [70], cromatografia líquida acoplada à espectrometria de massas [36] e cristalografia
 178 de raios-X [44,59].

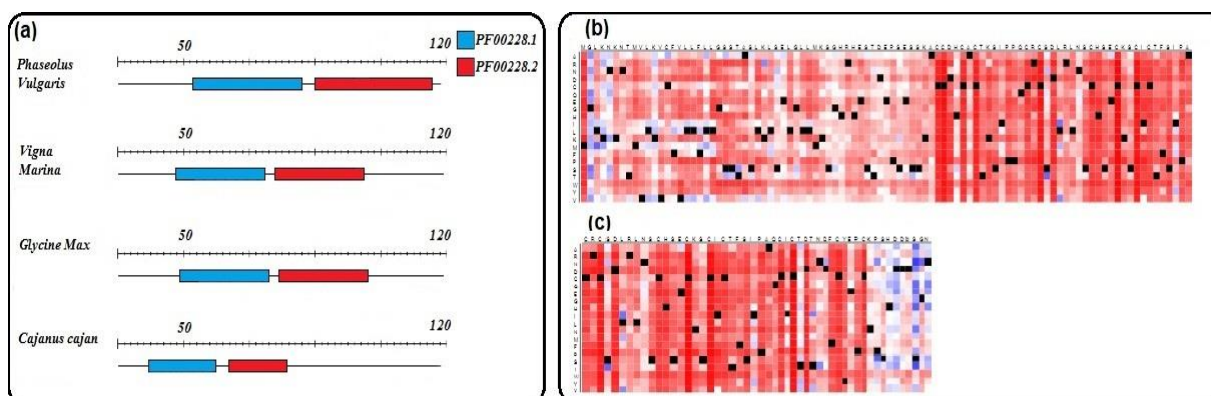
179 A média de hidropaticidade variou de -0.072 (*V. marina*) a -0.567 (*G. microphylla*). Essa
 180 média é utilizada para demonstrar o valor de hidrofobicidade da proteína, que é obtido a partir
 181 do cálculo da soma dos valores de hidropatia de todos os aminoácidos, dividido pelo tamanho
 182 da sequência [16], sendo as médias positivas para proteínas hidrofóbicas e negativas para
 183 proteínas hidrofílicas. Dessa forma, os valores de média de hidropaticidade indicaram que a
 184 proteína BBI é de caráter hidrofílica para todas as espécies. A média de hidropaticidade tem
 185 aplicações importantes, como ajudar na compreensão da estrutura 3D das proteínas, sua
 186 relação evolutiva e a localização de proteínas análogas e distantes [55].

187 Conforme o alinhamento múltiplo das sequências (Figura 1), foi possível verificar regiões
 188 de alta similaridade, onde a maior parte das posições apresentaram conservação entre as
 189 sequências de BBI das espécies analisadas. Clemente *et al.* [20], ao realizar o alinhamento
 190 de sequência de aminoácidos dos principais iso-inibidores de BBI de soja e de outras espécies
 191 representativas de leguminosas, também evidenciaram que existe uma alta identidade das
 192 sequências entre as isoformas de BBI.



193 **Figura 1.** Alinhamento das sequências do inibidor Bowman-Birk avaliadas. As sequências foram
 194 alinhadas por ClustalW, onde os resíduos idênticos e similares são exibidos na mesma cor.
 195

196 Para a caracterização funcional do BBI e previsão dos efeitos funcionais de mutações,
197 foram selecionadas quatro espécies (Figura 2). Foi identificada a existência de dois domínios
198 funcionais: PF00228.1 e PF00228.2 nas sequências de aminoácidos do BBI. Sendo esses
199 dois domínios funcionais encontrados em todas as sequências observadas, que estão
200 associados com funções de atividade do inibidor de duas serinoproteases de especificidade
201 diferentes [37]. Geralmente, o primeiro sítio inibi a tripsina, e o segundo a quimotripsina [56].
202



203 **Figura 2.** Previsão dos efeitos funcionais das mutações em aminoácidos pelo servidor SNAP2 e
204 Domínios funcionais do inibidor Bowman-Birk observados pelo servidor ProDom.

205 A partir da ação desses domínios, várias funções defensivas dos BBIs nas plantas são
206 relatadas, sendo a ação inseticida a mais estudada que são associadas à sua atividade nas
207 proteases intestinais, que reduz a assimilação de aminoácidos, o que retarda o
208 desenvolvimento dos insetos que se nutrem dessas plantas [25]. Kim *et al.* [47], também
209 evidenciaram que alguns BBIs podem inibir proteases produzidas por patógenos, que
210 possibilita seu efeito antimicrobiano. Além dessas funções, esses inibidores são associados
211 na tolerância a estresses abióticos, salino [74] e hídrico [29].

212 Dessa forma, o uso dos BBIs pode apontar genes para a engenharia genética de plantas
213 economicamente importantes, propiciando o desenvolvimento de plantas transgênicas
214 resistentes a insetos [22], assim como a patógenos e a maioria dos estresses abióticos. O que
215 possibilitaria a redução dos custos na mão de obra e na utilização de defensivos agrícolas
216 contra ao ataque de insetos e patógenos, para proporcionar um meio ambiente mais
217 sustentável.

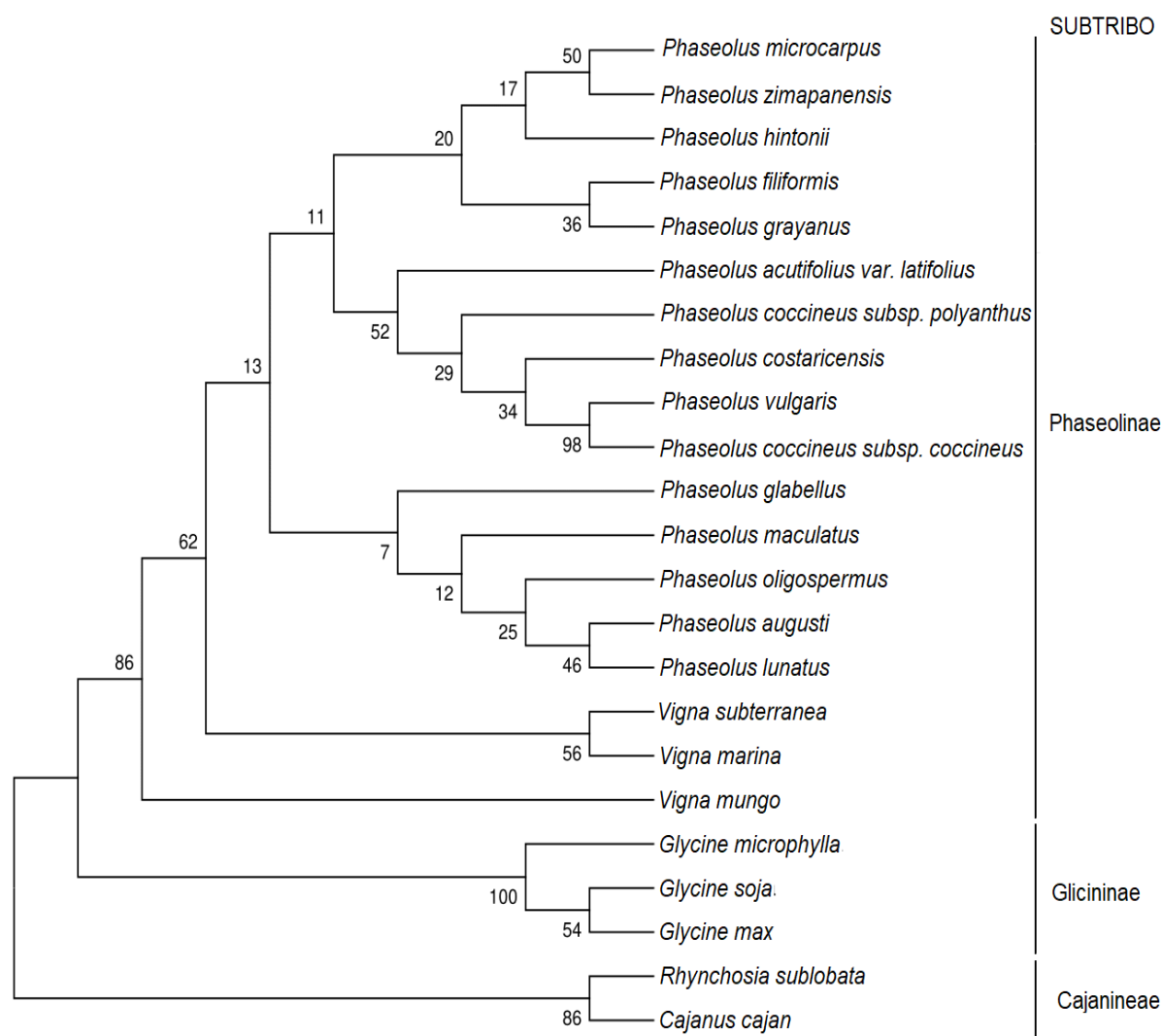
218 Além dessas funções, estudos também demonstram que os BBIs possuem propriedades
219 anti-inflamatórias [34,71,87], bem como propriedades anticancerígenas [4,34,46]. As
220 propriedades anticancerígenas estão ligadas aos dois sítios de ligação do BBI (atividades
221 antitripsina e anti-quimotripsina), propiciando um efeito antiproliferativo nas células
222 cancerígenas (principalmente do cólon) [21].

223 O servidor SNAP2 mostrou que os domínios funcionais PF00228.1 e PF00228.2, são as
224 regiões da proteína mais sensíveis a mutações. Este servidor forneceu um mapa com a
225 presença de tonalidades diferentes com possível substituição em cada posição dessas
226 proteínas. A tonalidade vermelha sinaliza fortes sinais para mutação, coloração branca
227 apresenta indicador de baixo efeito e a coloração azul indica neutralidade (Figura 2). A
228 ocorrência de uma maior quantidade de sítios sensíveis a mutações dos domínios PF00228.1
229 e PF00228.2, pode ser explicado por sua conservação e incidência nas sequências aqui
230 analisadas. Outros trabalhos, a partir das análises para previsão dos efeitos funcionais das
231 mutações em aminoácidos pelo servidor SNAP2 e domínios funcionais da putrescina N-
232 metiltransferase em espécies de Solanaceae [32] e lectina manose-específica em espécies
233 de *Allium* [40] também evidenciaram uma maior incidência de sítios sensíveis a mutações dos
234 domínios.

235 A análise filogenética baseada no BBI, resultou em uma árvore monofilética (Figura 3),
236 onde foi observado a formação de três cladogramas principais que compreendem três subtribos da
237 tribo Phaseoleae (Fabaceae - Papilionoideae). O clado maior foi composto pelas espécies da

238 subtribo Phaseolinae (BS=86%), representada pelos gêneros *Phaseolus* e *Vigna*, o segundo
239 clado foi formado pelas espécies representantes da subtribo Gliciniinae (*Glycine microphylla*,
240 *G. soja* e *G. max*) (BS=100%) e o terceiro clado integrado pelas espécies da subtribo
241 Cajanineae (*Rhynchosia sublobata* e *C. cajan*) (BS=86%).

242 No clado da subtribo Phaseoleolinae, foi observado a formação de três clados dentro do
243 gênero *Phaseolus*, onde primeiro grupo (BS=20%) compreende as espécies *P. microcarpus*,
244 *P. zimapanensis*, *P. hintonii*, *P. filiformis* e *P. grayanus*, o segundo grupo (BS=52%) com *P.*
245 *acutifolius* var. *latifolius*, *P. coccineus* subsp. *polyanthus*, *P. costaricensis*, *P. vulgaris* e *P.*
246 *coccineus* subsp. *coccineus* e o terceiro (BS=7%) formado pelas espécies *P. glabellus*, *P.*
247 *maculatus*, *P. oligospermus*, *P. augusti* e *P. lunatus*. No clado do gênero *Vigna*, foi possível
248 observar que a espécie *V. mungo* apresentou um grupo isolado na árvore filogenética
249 (BS=86%), que por sua vez está próximo também das espécies *V. subterrânea* e *V. marina*
250 (BS=56%).



251 **Figura 3.** Árvore filogenética das espécies da tribo Phaseoleae (Fabaceae - Papilionoideae), geradas
252 com base nas sequências do inibidor Bowman-Birk pelo método Máxima Verossimilhança.
253

254 As espécies do gênero *Phaseolus* são todas endêmicas do Novo Mundo, sendo as
255 espécies do primeiro subclado encontradas principalmente do México a Califórnia (*P.*
256 *filiformis*) [26]. No segundo subclado, as espécies são distribuídas do sudeste do Canadá ao
257 sul do leste dos EUA e do sul dos EUA ao sudeste da Califórnia, por meio do México e da
258 América Central, como também na região andina da América do Sul [26]. Já no terceiro

259 subclado, a espécie *P. glabellus* ocorre do sul de Nuevo Leon e Tamaulipas até o centro de
260 Chiapas (México), *P. oligospermus*, *P. augusti* e *P. lunatus* do México a América Central e
261 partes da América do Sul, e *P. maculatus* do leste dos EUA até o leste do Texas e sul do
262 México [26]. Nesse sentido, a proximidade geográfica dessas espécies sustenta a formação
263 do clado *Phaseolus*.

264 Jin *et al.* [41], a partir do genoma plastidial de espécies da tribo Desmodieae e
265 Phaseoleae, evidenciaram que o feijão comum (*P. vulgaris*) e o feijão-caupi (*V. unguiculata*)
266 foram reconstruídos como um grupo monofilético. Em outros estudos, a partir DNA de
267 cloroplastos, foi possível verificar que as espécies do gênero *Vigna* do Novo Mundo estavam
268 mais intimamente relacionados com outros gêneros do Novo Mundo de Phaseolinae (como
269 *Phaseolus*) do que com *Vigna* do Velho Mundo [26-27], como pode-se observar na *V. mungo*
270 que ficou isolada das outras espécies de *Vigna*. As espécies de quase todos os gêneros
271 americanos têm por sinapomorfia a assimetria floral “destra” [27]. Essas espécies são todas
272 trepadeiras e localizadas em florestas tropicais de savana [60], embora seja possível localizar
273 alguns subgrupos que se diversificaram em florestas tropicais montanhosas, como é o caso
274 do gênero *Phaseolus* [26].

275 No clado Glycininae (*G. microphylla*, *G. soja* e *G. max*), o agrupamento das espécies *G.*
276 *max* e *G. soja* provavelmente está relacionada à sua proximidade genética, uma vez que
277 essas espécies pertencem ao subgênero soja, que inclui *G. max* e *G. soja* (parente selvagem
278 da soja comercial), que são nativos do leste da Ásia e parte da Rússia. Já a *G. microphylla*
279 pertence ao subgênero Glycine, que inclui cerca de 30 espécies perenes, cuja o centro de
280 diversidade se encontra na Austrália, distribuídas por habitats diversos, desde o deserto até
281 climas temperadas e subtropicais [75].

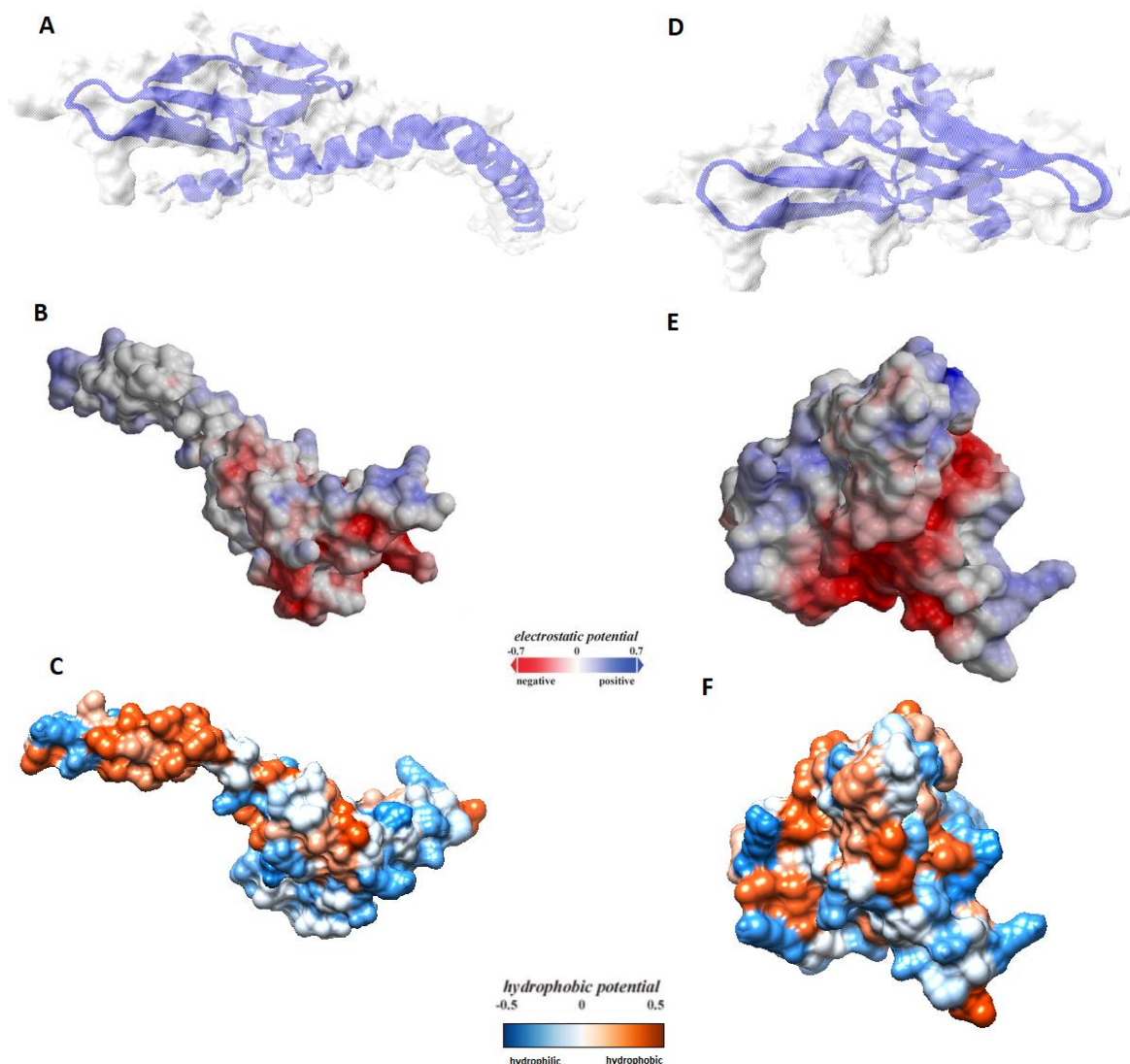
282 Análises filogenéticas a partir de sequências de genes plastidiais de espécies de
283 leguminosas phaseolóides, possibilitou a visualização de um grupo fortemente apoiado
284 (BS=98% e BS=85%) que incluía as subtribos Phaseolinae, Glycininae e Cajaninae [43,76],
285 onde foi evidenciado o grupamento monofilético de *Rhynchosia* e *Cajanus* [43], corroborando
286 os resultados aqui observados.

287 Na subtribo Cajaninae, os estudos de 12. Cândido *et al.* [12], baseados na análise
288 das regiões nucleares ITS e plastidiais *rpl32* e *trnQ*, também sustentaram o agrupamento de
289 *Cajanus* com *R. volubilis*. Características semelhantes são compartilhadas por esses dois
290 gêneros, como o hábito que pode variar de subarbuscivo a herbáceo, folhas trifolioladas a
291 unifolioladas (em algumas *Rhynchosia*), sendo a maior diferença observada no fruto, que
292 pode apresentar duas a três sementes (ou mais) em *Rhynchosia* e *Cajanus*, respectivamente.
293 A espécie *R. sublobata* geralmente, possui folhas trifolioladas, hábito trepador, flores amarelas
294 estriadas de vermelho e lobos do cálice transpassando a corola [85]. Outra característica
295 partilhada é a distribuição geográfica, onde as espécies de *Rhynchosia* são asiáticas, bem
296 como a maioria das espécies do gênero *Cajanus* [72].

297 As relações de baixo suporte apresentadas na árvore filogenética, podem ser justificadas
298 pela ineficiência do marcador (BBI) para separar ou organizar espécies geneticamente muito
299 próximas.

300 Análises filogenéticas com base em sequências proteicas têm sido bastante utilizadas
301 para esclarecer a função das proteínas dentro de um grupo de organismos [40].
302 Reconstruções filogenéticas podem explicar questionamentos sobre como as proteínas estão
303 relacionadas em diferentes espécies, e se elas podem ter evoluído a partir de um ancestral
304 comum [62,84]. A filogenia baseada em sequências do BBI, demonstrou similaridade com
305 outras filogenias de Papilionoideae a partir de genes plastidiais [14,58-59,91].

306 Para predição da estrutura terciária do BBI o servidor GalaxyTBM foi utilizado , foram
307 selecionadas as espécies *P. vulgaris* e *C. cajan*, representantes da tribo Phaseoleae
308 (Fabaceae - Papilionoideae), obtidas no banco de dados (Figura 4).

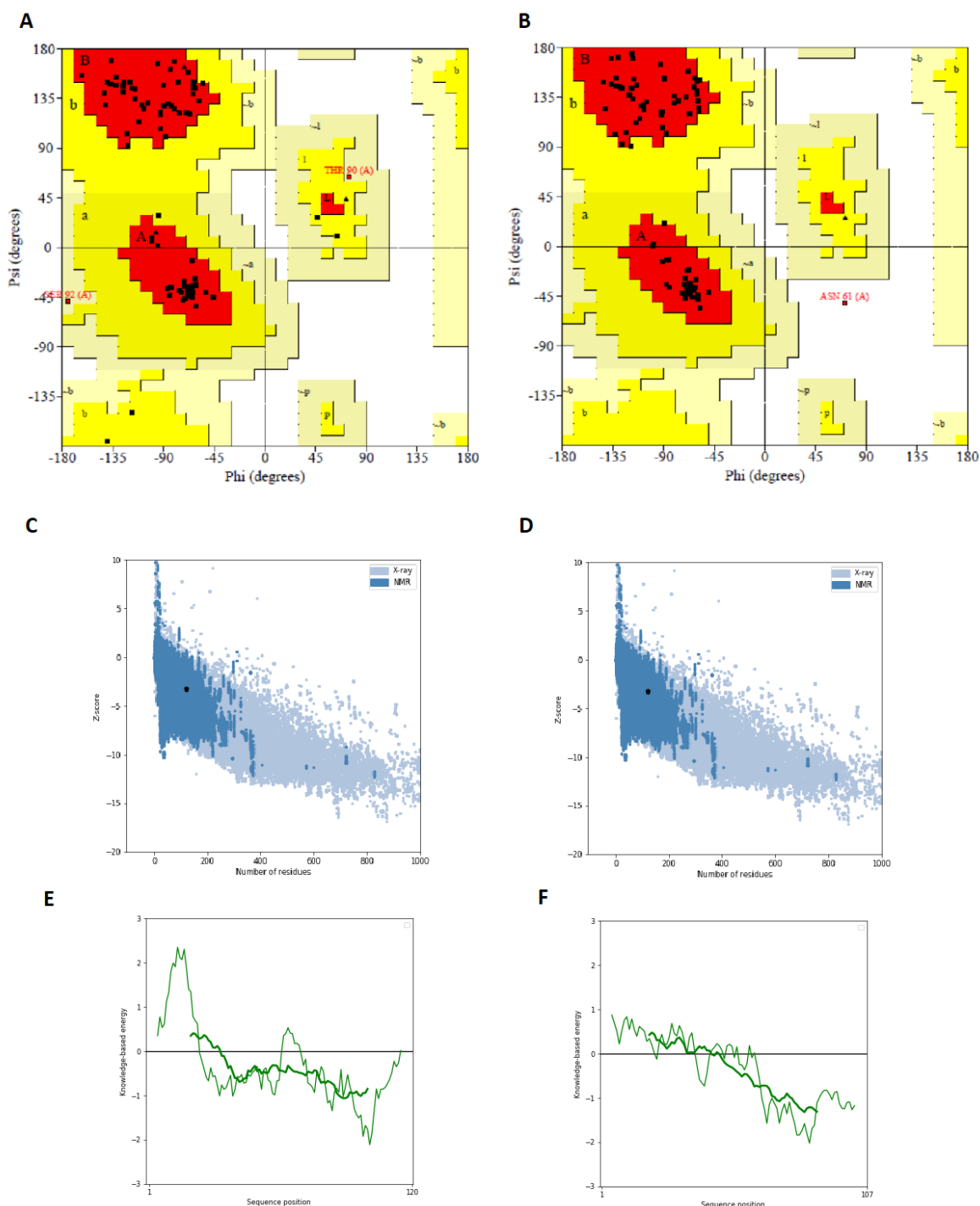


309
310 **Figura 4.** Estrutura tridimensional predita pelo servidor GalaxyTBM para o inibidor Bowman-Birk das
311 espécies *P. vulgaris* (A) e *C. cajan* (D). Superfície eletrostática representada como um gradiente de cor,
312 a partir do mais carregado negativamente (vermelho) para o mais positivamente carregado (azul) das
313 espécies *P. vulgaris* (B) e *C. cajan* (E). Hidrofobicidade representada como gradiente de cor, sendo o
314 azul o mais hidrofílico e vermelho alaranjado para os mais hidrofóbicos das espécies *P. vulgaris* (C) e
315 *C. cajan* (F).

316 Para validação do modelo, o gráfico de Ramachandran e o Z-score foram utilizados. Os
317 modelos refinados evidenciaram 93,5% e 96,9 % de resíduos de aminoácidos nas regiões
318 beneficiadas pela análise do gráfico de Ramachandran, para as espécies *P. vulgaris* e *C.*
319 *cajan*, respectivamente.

320 As estruturas tridimensionais são geradas pelas sequências de codificação, que por
321 conseguinte permitem que os efeitos biológicos possam ser avaliados com mais precisão.
322 Uma vez que, a previsão de estruturas 3D de proteínas geralmente resulta em aplicações
323 práticas de relevância no uso biotecnológico e terapêutico [1,13].

324 Por meio da análise no servidor ProSA, foram observados valores de Z-score de -1,61 e
325 -3,92 para *C. cajan* e *P. vulgaris*, respectivamente (Figura 5).



326
327
328
329
330

Figura 5. Gráfico de Ramachandran para BBI de *P. vulgaris* (A) e *C. cajan* (B). Valor do Z-score (ponto preto) pelo ProSA-web para *P. vulgaris* (C) e *C. cajan* (D). Uso do ProSA-web para o BBI, mostrando o gráfico de energia de escores de resíduos da estrutura de proteína nativa para *P. vulgaris* (E) e *C. cajan* (F).

331
332
333
334
335
336
337
338

A qualidade do modelo é obtida a partir do resultado do Z-score de estruturas proteicas resolvidas como referências [7], uma vez que esses valores podem ser significativos a luz da termodinâmica e qualidade da estrutura, com potencial de apresentar valores muito semelhantes ou até superiores quando comparados a proteínas que foram elucidadas por meio de experimento, o que confirma a confiabilidade do método [1]. Nesse sentido, os valores de Z-scores apontam que as estruturas são errôneas quando esse valor estiver fora de um intervalo característico para proteínas nativas [89]. Dessa forma, pode-se confirmar que os valores de Z-scores obtidos para a estrutura de BBI de *C. cajan* e *P. vulgaris* (-1,61 e -3,92)

339 estão dentro do intervalo de pontuações tipicamente identificada para proteínas nativas de
340 tamanhos semelhantes.

341 Os modelos 3D obtidos evidenciaram diferenças mínimas entre as estruturas do BBI de
342 *P. vulgaris* e *C. cajan* (Figura 4). Nesses modelos pode ser observado a presença de beta-
343 folhas e alfa-hélices nas regiões dos domínios PF00228.1 e PF00228.2 no modelo de *P.*
344 *vulgaris* e *C. cajan*. Os domínios PF00228.1 e PF00228.2 exibem uma conformação
345 semelhante, dessa forma estes domínios provavelmente apresentam a mesma funcionalidade
346 nas duas espécies analisadas, como também foi observado na estrutura terciária de BBI de
347 sementes de *V. unguiculata* [4]. Os dois domínios apresentam uma conformação conservada,
348 justificado por sua presença nas duas espécies analisadas, também observado por outros
349 autores em seus modelos 3D para outros tipos de proteínas [32,40]. Chen *et al.* [17], afirmam
350 que a estrutura tridimensional do BBI é estabilizada, sobretudo pela presença de sete pontes
351 dissulfeto, aliado a um pequeno auxílio do núcleo hidrofóbico da molécula.

352 A funcionalidade das proteínas está associada à sua configuração 3D, que por sua vez,
353 é determinada pela sequência de seus aminoácidos [40]. Ainda que seja complexo a análise
354 das sequências, métodos computacionais têm se tornado cada vez mais utilizados na
355 predição de proteínas. Dessa forma, a homologia ou modelagem comparativa é a mais
356 empregada por gerar bons resultados de forma efetiva, bem como a qualidade dos modelos
357 previstos, que se diferencia de outras técnicas por apresentarem uma relação evolutiva
358 aceitável e evidenciar uma maior precisão [1].

359 A modelagem tridimensional e a análise comparativa de BBI fornecerão informações
360 indispensáveis sobre suas funções moleculares e evolução nas espécies analisadas.

361 CONCLUSÕES

362 O presente estudo permitiu um melhor entendimento das funções do BBI na atividade
363 inibitoria das duas serinoproteases. Associado a isso, o agrupamento obtido a partir das
364 sequências da proteína BBI confirma as classificações taxonômicas recentes das espécies da
365 tribo Phaseoleae (Fabaceae - Papilionoideae).

366 O emprego de ferramentas *in silico* se mostrou válida e confiável na construção de
367 modelos 3D, na geração de modelos estáveis quanto ao seu nível energético e dentro dos
368 padrões, que é capaz ser considerado uma representação análoga da estrutura real da
369 proteína já estabelecida, o que proporciona a análise de suas estruturas terciárias bem como
370 suas funções moleculares.

371

372 **FINANCIAMENTO:** O presente trabalho foi realizado com apoio da Coordenação de
373 Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento
374 001. Silva FG foi financiado por uma bolsa de doutorado CAPES.

375 REFERÊNCIAS

- 376 1. Abreu RA, editor. Caracterização funcional e estrutural comparativa *in silico* de uma proteína pr-5
377 do tipo taumatina de *Poncirus trifoliata* [Internet]. Sumé: Universidade Federal de Campina Grande;
378 2015 [citado em 10 de novembro de 2022]. Disponível em:
379 <http://dspace.sti.ufcg.edu.br:8080/xmlui/handle/riufcg/5530?show=full>.
- 380 2. Andrade LF, Nahum LA, Avelar LGA. Eukaryotic protein kinases (ePKs) of the helminth parasite
381 *Schistosoma mansoni*. BMC Genomics. 2011, 12:215.
- 382 3. Armstrong WB, Taylor TH, Kennedy AR, Melrose RJ, Messadi DV, Gu M, et al. Bowman birk inhibitor
383 concentrate and oral leukoplakia: a randomized phase IIb trial. Cancer Prev Res (Phila).
384 2013;6(5):410-8.
- 385 4. Armstrong WB, Wan XS, Kennedy AR, Taylor TH. Development of the Bowman-Birk inhibitor for oral
386 cancer chemoprevention and analysis of Neu immunohistochemical staining intensity with Bowman-
387 Birk inhibitor concentrate treatment. Laryngoscope. 2003;113(10):1687-702.
- 388 5. Barbosa JARG, Silva LP, Teles RCL, Esteves GF, Azevedo RB, Ventura MM, et al. Crystal Structure
389 of the Bowman-Birk Inhibitor from *Vigna unguiculata* Seeds in Complex with b-Trypsin at 1.55 Å
390 Resolution and Its Structural Properties in Association with Proteinases. Biophys. J.
391 2007;92(5):1638-50.

- 392 6. Barreto KL, Fernandes MF, Queiroz LP. Flora da Bahia: Leguminosae – Centrosema
393 (Papilionoideae: Phaseoleae) Sitientibus série Ciências Biológicas. 2020; 20.
- 394 7. Benkert P, Biasini M, Schwede T (2011) Toward the Estimation of the Absolute Quality of Individual
395 Protein Structure Models. *Bioinformatics*. 2011;27(3):343–350.
- 396 8. Bezerra júnior RQ, Martins GR, Barroso IC, Marinho RC, Aguiar TDF, Teixeira MFS. Eletroforese
397 Bidimensional e espectrometria de massa como ferramentas proteômica aplicadas à definição de
398 marcadores proteicos associados à eficiência reprodutiva de caprinos. *Acta Vet Bras*. 2013;7(2):100-
399 112. Português.
- 400 9. Birk Y, Gertler A, Khalef, S. A pure trypsin inhibitor from soya beans. *Biochem J*. 1963;87(2):281–
401 284.
- 402 10. Bowman DE (1946) Differentiation of soy bean anti-tryptic factors. *Proc Soc Exp Biol Med*.
403 1946;63(3):547-50.
- 404 11. Bruneau A, Doyle JL, Doyle JJ. Phylogenetic relationships in Phaseoleae: evidence from chloroplast
405 restriction site characters. In: Crisp MD, Doyle JJ, editores. *Advances in Legume Systematics, Part*
406 *7: Phylogeny*. Richmond: Royal Botanic Gardens, Kew; 1995. p. 309-330.
- 407 12. Cândido ES, Vatanparast M, Vargas W, Bezerra LMPA, Lewis GP, Mansano VF, et al. Molecular
408 phylogenetic insights into the evolution of Eriosema (Fabaceae): a recent tropical savanna-adapted
409 genus. *Bot J Linn Soc*. 2020;194(4):439–459.
- 410 13. Capriles PVSZ, Trevizani R, Rocha GK, Dardenne LE, Custódio FL. Modelos Tridimensionais. In:
411 Verli H, organizador. *Bioinformática da Biologia à Flexibilidade Molecular*. São Paulo: Sociedade
412 Brasileira de Bioquímica e Biologia Molecular; 2014. p. 148-171.
- 413 14. Cardoso D, Pennington RT, Queiroz LP, Boatwright JS, Van Wyk BE, Wojciechowski MF, et al.
414 Reconstructing the deep-branching relationships of the papilionoid legumes. *S. African J Bot*.
415 2013;89:58–75.
- 416 15. Catalano M, Ragona L, Molinari H, Tava A, Zetta L. Anticarcinogenic Bowman Birk inhibitor isolated
417 from snail medic seeds (*Medicago scutellata*): Solution structure and analysis of self-association
418 behavior. *Biochemistry*. 2003;42(10):2836-46.
- 419 16. Chang KY, Yang J-R. Analysis and Prediction of Highly Effective Antiviral Peptides Based on
420 Random Forests. *PLoS ONE*. 2013;8(8):e70166.
- 421 17. Chen P, Rose P, Love R, Wei CH, Wang BC. Reactive sites of na anticarcinogenic Bowman–Birk
422 proteinase inhibitor are similar to Other trypsin inhibitors. *J. Biol. Chem*. 1992;267(3):1990–1994
- 423 18. Chen VB, Davis IW, Richardson DC. KING (Kinemage, Next Generation): a versatile interactive
424 molecular and scientific visualization program. *Protein Sci*. 2009;18(11):2403–2409.
- 425 19. Choo KH, Tan TW, Ranganathan S. SPdb – a signal peptide database. *BMC Bioinformatics*. 2005
426 6:249.
- 427 20. Clemente A, Arques MC. Bowman-Birk inhibitors from legumes as colorectal chemopreventive
428 agentes. *World J Gastroenterol*. 2014;20(30):10305–10315.
- 429 21. Clemente A, Moreno FJ, Marín-Manzano MDC, Jiménez E, Domoney C. The cytotoxic effect
430 of Bowman-Birk isoinhibitors, IBB1 and IBB2, from soybean (*Glycine max*) on HT29 human
431 colorectal cancer cells is related to their intrinsic ability to inhibit serine proteases. *Mol Nutr Food*
432 *Res*. 2010;54(3):396-405.
- 433 22. Clemente M, Corigliano MG, Pariani AS, Sánchez-López EF, Sander VA, Ramos-Duarte VA. Plant
434 serine protease inhibitors: Biotechnology application in agriculture and molecular farming. *Int J Mol*
435 *Sci*. 2019; 20(6):1345.
- 436 23. Connors, BJ, Laun, NP, Maynard, CA, Powell WA. Molecular characterization of a gene encoding a
437 cystatin expressed in the stems of American chestnut (*Castaneadentata*). *Planta*. 2002;215(3):510-
438 4.
- 439 24. De Paola D, Blanco E, Pierri CL, Sonnante G. Isolation and characterization of novel variants of BBI
440 coding genes from the legume *Lathyrus sativus*. *Plant Physiol Biochem*. 2012;57:45-53.
- 441 25. Deepika R, Nath AK. Studies on Trypsin Inhibitor in Bean (*Phaseolus vulgaris* L) Cultivars of
442 Himalayan Region. *J Plant Biochem Biotechnol*. 2010;19:103-105.
- 443 26. Delgado-Salinas A, Bibler R, Lavin M. Phylogeny of the genus *Phaseolus* (Fabaceae): A recent
444 diversification in na ancient landscape. *Syst Bot*. 2006;31(4):779-791.
- 445 27. Delgado-Salinas A, Thulin M, Pasquet R, Weeden N, Lavin M. *Vigna* (Fabaceae) sensu lato: the
446 names and identities of the american segregate genera. *Am J Bot*. 2011;98(10):1694-715.
- 447 28. Di Ciero L, Bellato CM Proteoma - Avanços Recentes em técnicas de eletroforese bidimensional e
448 espectrometria de massa. *Biociencia*. 2002;5(29):158-164. Português.
- 449 29. Dramé KN, Passaquet C, Repellin A, Zully-Fodil Y. Cloning characterization and differential
450 expression. of a Bowman-Birk inhibitor during progressive water deficit and subsequent recovery in
451 peanut (*Arachis hypogaea*) leaves. *J Plant Physiol*. 2013;170(2):225-229.

- 452 30. Espert, SM, Sede, SM, Ruiz LK, Fortunato RH, Poggio L. New chromosome reports in the subtribes
453 Diocleinae and Glycininae (Phaseoleae: Papilionoideae: Fabaceae). Bot J Linn Soc.
454 2008;158(2):336–341.
- 455 31. Fereidunian A, Sadeghalvad M, Oscoie MO, Mostafaie A. Soybean bowman-birk protease inhibitor
456 (BBI): Identification of the mechanisms of BBI suppressive effect on growth of two adenocarcinoma
457 cell lines: AGS and HT29. Arch Med Res. 2014;45(6):455-61.
- 458 32. Gama, BRA, Moraes Filho RM, Montarroyos AVV, Martins LSS. In silico characterization of
459 putrescine N-methyltransferase in Solanaceae species. Genetics and Molecular Research.
460 2022;21:gmr19076.
- 461 33. Gasteiger E, Hoogland C, Gattiker A and Duvaud S. Protein identification and analysis tools on the
462 ExPASy server. In: Walker JM, editor. The proteomics protocols handbook. Totowa: Humana Press.
463 2005. p. 571-607.2005.
- 464 34. Gitlin-Domagalska A, Maciejewska A, Debowski D. Bowman-Birk Inhibitors: Insights into Family of
465 Multifunctional Proteins and Peptides with Potential Therapeutical Applications. Pharmaceuticals.
466 2020; 13(12):421.
- 467 35. Hecht M, Bromberg Y, Rost B. Better prediction of functional effects for sequence variant. BMC
468 Genomics. 2015; 16:(Suppl 8)S1.
- 469 36. Heller M, Ye M, Michel PE, Morier P, Stalder D, Jünger MA, et al. Added value for tandem mass
470 spectrometry shotgun proteomics data validation through isoelectric focusing of peptides. J
471 Proteome Res. 2005;4(6):2273–2282.
- 472 37. Hellinger R, Gruber CW. Peptide-based protease inhibitors from plants. Drug Discov Today.
473 2019;24(9)1877-1889.
- 474 38. James AM, Jayasena AS, Zhang J, Berkowitz O, Secco D, Knott GJ, et al. Evidence for Ancient
475 Origins of Bowman-Birk Inhibitors from *Selaginella moellendorffii*. Plant Cell. 2017;29(3):461–473
- 476 39. Jaulent AM, Leatherbarrow RJ. Design, synthesis and analysis of novel bicyclic and bifunctional
477 protease inhibitors. Protein Eng Des Sel. 2004;17(9):681–687.
- 478 40. Jimenez HJ, Martins LSS, Carvalho R, Montarroyos AVV, Moraes Filho RM. In silico characterization
479 and phylogenetic analysis of a mannose-specific lectin in *Allium* species. Genet Mol Res.
480 2019;18:gmr18187.
- 481 41. Jin DP, Choi IS, Choi BH. Plastid genome evolution in tribe Desmodieae (Fabaceae: Papilionoideae).
482 PLoS ONE. 2019;14:e0218743.
- 483 42. Juritsch AF, Moreau R. Role of soybean-derived bioactive compounds in inflammatory bowel
484 disease. Nutr Rev. 2018;76(8):618-638.
- 485 43. Kajita T, Ohashi H, Tateishi Y, Bailey CD, Doyle JJ. rbcL and Legume Phylogeny, with particular
486 reference to Phaseoleae, Millettieae, and Allies. Syst Bot. 2001;26(3):515-536.
- 487 44. Kantardjieff KA, Rupp B. Protein isoelectric point as a predictor for increased crystallization screening
488 efficiency. Bioinformatics. 2004;20(14):2162-8.
- 489 45. Kasap S, Benkrid K. High performance phylogenetic analysis with maximum parsimony on
490 reconfigurable hardware. IEEE Transactions on VLSI Systems. 2010;19(5):796-808.
- 491 46. Kennedy AR, Billings PC, Wan XS, Wan PM. Effects of Bowman-Birk inhibitor on rat colon
492 carcinogenesis. Nutrition and Cancer. 2002;43(2):174-86.
- 493 47. Kim JY, Park SC, Hwang I, Cheong H, Nah JW, Hahm KS, Park Y. Protease inhibitors from plants
494 with antimicrobial activity. Int J Mol Sci. 2009;10:2860–2872.
- 495 48. Kiraga J, Mackiewicz P, Mackiewicz D, Kowalczyk M, Biecek P, Polak N, et al. The relationships
496 between the isoelectric point and: length of proteins, taxonomy and ecology of organisms. BMC
497 Genomics. 2007;8:163-10.
- 498 49. Kirkwood J, Hargreaves D, O’Keefe S, Wilson J. Using isoelectric point to determine the pH for initial
499 protein crystallization trials. Bioinformatics. 2015;31(9):1444–1451.
- 500 50. Krieger E, Joo K, Lee J, Raman S, Thompson J, Tyka M, et al. Improving physical realism,
501 stereochemistry and side-chain accuracy in homology modeling: four approaches that performed
502 well in CASP8. Proteins. 2009;77(Suppl 9):114–122.
- 503 51. Ko J, Park H, Seok C. GalaxyTBM: Template-based modeling by building a reliable core and refining
504 unreliable local regions. BMC Genomics. 2012;13(1):189.
- 505 52. Kumar S, Stecher G and Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0
506 for Bigger Datasets. Mol Biol Evol. 2016;33(7):1870-4.
- 507 53. Lackey JA. Phaseoleae. In: RM Polhill, PH Raven, editores. Papilionoideae. Advances in legume
508 systematics. Richmond: Royal Botanic Gardens, Kew. 1981. p. 301–335
- 509 54. Lichtenstein GR, Deren JJ, Katz S, Lewis JD, Kennedy AR, Ware JH. Bowman-Birk inhibitor
510 concentrate: a novel therapeutic agent for patients with active ulcerative colitis. Randomized control
511 trial. 2008;53(1):175-80.

- 512 55.Lolkema JS, Slotboom DJ. Estimation of structural similarity of membrane proteins by hydrophathy
513 profile alignment. *Mol Membr Biol.* 1998;15:33-42.
- 514 56.Lioi L, Galasso I, Daminati MG, Piergiovanni AR. Inhibitory properties and binding loop polymorphism
515 in Bowman-Birk inhibitors from Phaseolus species. *Genet Resour Crop Evol.* 2010;57:533–542.
- 516 57.Losso JN. The Biochemical and Functional Food Properties of the Bowman-Birk Inhibitor. *Nutr Food*
517 *Sci* 2008;48(1):94-118.
- 518 58.LPWG (Legume Phylogeny Working Group). Legume phylogeny and classification in the 21st
519 century: progress, prospects and lessons for other species-rich clades. *Taxon.* 2013;62(2):217–248.
- 520 59.LPWG (Legume Phylogeny Working Group). A new subfamily classification of the Fabaceae based
521 on a taxonomically comprehensive phylogeny. *Taxon.* 2017;66(1):44–77.
- 522 60.Maxted N, Mabuza-Dlamini P, Moss H, Padulosi S, Jarvis A, Guarino L. African Vigna: An
523 Ecogeographic Study. Rome: IPGRI; 2004.
- 524 61.Mello MO, Tanaka AS, Silva-Filho MC. Molecular evolution of Bowman–Birk type proteinase
525 inhibitors in flowering plants. *Mol Phylogenet Evol.* 2003;27(1):103–112.
- 526 62.Moraes Filho RM, Martins LSS. In silico comparative analysis of tylenchid nematodes pectate lyases.
527 *GMR.* 2016;15(3):gmr.15038402.
- 528 63.Oliás R, Becerra-Rodríguez C, Soliz-Rueda JR, Moreno FJ, Delgado-Andrade C, Clemente A.
529 Glycation affects differently the main soybean Bowman-Birk iso-inhibitors, IBB1 and IBB2, altering
530 their antiproliferative properties against HT29 colon cancer cells. *Food Funct.* 2019;10:6193–6202.
- 531 64.Oliveira ACS, Perez APF, Silva JS. Os gêneros *Eriosema* e *Rhynchosia* (Leguminosae -
532 Papilionoideae - Phaseoleae) no nordeste brasileiro. *Rodriguésia.* 2018;69(4):1825-1850.
533 Português.
- 534 65.Peng C, Shi C, Cao X, Li Y, Liu F, Lu F. Factors Influencing Recombinant Protein Secretion Efficiency
535 in Gram-Positive Bacteria: Signal Peptide and Beyond. *Front Bioeng Biotechnol.* 2019;7:139.
- 536 66.Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera-
537 a visualization system for exploratory research and analysis. *J Comput Chem.* 2004;25(13):1605-
538 12.
- 539 67.Pollhill RM, Raven PH e Stirton CH. Evolution and systematics of the Leguminosae. In Pollhill RM,
540 Raven PH, editores. *Advances in legume systematics.* Richmond: Royal Botanic Gardens, Kew.
541 1981. p. 1–26.
- 542 68.Qi RF, Song ZW, Chi CW. Structural features and molecular evolution of Bowman-Birk protease
543 inhibitors and their potential application. *Acta Biochim Biophys Sin.* 2005;7(5):283-92.
- 544 69.Rao KN, Suresh CG. Bowman–Birk protease inhibitor from the seeds of *Vigna unguiculata* forms a
545 highly stable dimeric structure. *Biochimica et Biophysica Acta.* 2007;1774(10):1264-73.
- 546 70.Righetti PG, Castagna A, Herbert B, Reymond F, Rossier JS. Prefractionation techniques in
547 proteome analysis. *Proteomics.* 2003;3(8):1397–1407.
- 548 71.Rui-Feng Q, Zhan-Wu S, Cheng-Wu C. Structural features and molecular evolution of Bowman-Birk
549 protease inhibitors and their potential application. *Acta Biochim Biophys.* 2005;37(5):283–292.
- 550 72.Schrire, BD. Tribe Phaseoleae. In: Lewis G, Schrire BD, Mackinder B, Lock M, editores. *Legumes of*
551 *the world.* Richmond: Royal Botanic Gardens, Kew. 2005. p. 393-431.
- 552 73.Servant F, Bru C, Carrère S, Courcelle E, Gouzy J, Peyruc D, et al. ProDom: Automated clustering
553 of homologous domains. *Brief. Bioinform.* 2002;3(3):246-51.
- 554 74.Shan L, Li C, Chen F, Zhao S, Xia GA. Bowman-Birk type protease inhibitor is involved in the
555 tolerance to salt stress in wheat. *Plant Cell Environ.* 2008;31(8):1128–37.
- 556 75.Sherman-Broyles S, Bombarely A, Powell AF, Doyle JL, Egan An, Coate JE, et al. (2014) The wild
557 side of a major crop: soybean’s perennial cousins from down under. *Am J Bot.* 101(10):1651-65.
- 558 76.Stefanovic S, Pfeil BE, Palmer JD, Doyle J. Relationships Among Phaseoloid Legumes Based on
559 Sequences from Eight Chloroplast Regions. *Systematic Botany* 2009;34(1):115–128.
- 560 77.Stekhoven SFM, Gorissen MH, Flik G. The isoelectric point, a key to understanding a variety of
561 biochemical problems: a minireview. *Fish Physiol Biochem.* 2008;34:1–8.
- 562 78.Teixeira JD, Ferreira JJS, Silva JS. A tribo Phaseoleae s.l. (Leguminosae-Papilionoideae) no
563 município de Caetité, Estado da Bahia, Brasil. *Hoehnea.* 2021;48:e392020. Português.
- 564 79.Tiessen A, Pérez-Rodríguez P, Delaye-Arredondo LJ. Mathematical modeling and comparison of
565 protein size distribution in different plant, animal, fungal and microbial species reveals a negative
566 correlation between protein size and protein number, thus providing insight into the evolution of
567 proteomes. *BMC Research Notes.* 2012;5:85.
- 568 80.Tokmakov A, Kurotani A, Sato K-I. Protein pI and Intracellular Localization. *Front Mol Biosci.*
569 2021;8:775736.
- 570 81.Tozzi AMGA. Papilionoideae. In: Tozzi AMGA, Melhem TS, Forero E, Fortuna-Perez AP, Wanderley
571 MGL, Martins SE, et al., editores. São Paulo: Instituto de Botânica. 2016. p. 167-397. Português.

- 572 82.Tsirigos KD, Peters C, Shu N, Käll L. The TOPCONS web server for combined membrane protein
573 topology and signal peptide prediction. *Nucleic Acids Res.* 2015;43(W1):W401-7.
574 83.Utrilla MP, Peinado MJ, Ruiz R, Rodriguez-Nogales A, Algieri F, Rodriguez-Cabezas ME, et al. Pea
575 (*Pisum sativum* L.) seed albumin extracts show anti-inflammatory effect in the DSS model of mouse
576 colitis. *Mol Nutr Food Res.* 2015;59(4):807-19.
577 84.Van Holle S, De Schutter K, Eggermont L and Tsaneva M. Comparative Study of Lectin Domains in
578 Model Species: New Insights into Evolutionary Dynamics *Int J Mol Sci.* 2017;18(6):1136.
579 85.Vargas W, Fortuna-Perez AP, Lewis GP, Piva TC, Vatanparast M, Machado SR. Ultrastructure and
580 secretion of glandular trichomes in species of subtribe Cajaninae Benth. (Fabaceae, Phaseoleae).
581 *Protoplasma.* 2019;56(2):431-445.
582 86.Viviani T, Conte L, Cristofolini G, Speranza M. Sero-systematic and taxonomic studies on the
583 Phaseoleae (Fabaceae) and related tribes. *Bot J Linn Soc.* 1991;105(2):113-136.
584 87.Ware JH, Wan XS, Newberne P, Kennedy AR. Bowman-Birk inhibitor concentrate reduces colon
585 inflammation in mice with dextran sulfate sodium-induced ulcerative colitis. *Dig Dis Sci.*
586 1999;44(5):986-90.
587 88.Watson JD, Laskowski RA, Thornton JM. Predicting protein function from sequence and structural
588 data. *Curr Opin Struct Biol.* 2005;15(3):275–284.
589 89.Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in
590 threedimensional structures of proteins. *Nucleic Acids Res.* 2007;35(Web Server issue):W407-10.
591 90.Wu Y-T, Huang W-Y, Lin T-C, Sheu S-J. Determination of moutan tannins by high-performance liquid
592 chromatography and capillary electrophoresis. *J Sep Science.* 2003;26(18):1629-1634.
593 91.Zhao Y, Zhang R, Jiang K.-W, Qi J, Hu Y, Guo J, et al. Nuclear phylotranscriptomics and
594 phylogenomics support numerous polyploidization events and hypotheses for the evolution of
595 rhizobial nitrogen fixing symbiosis in Fabaceae. *Molecular Plant.* 2021; 14(5):748-773.

ANEXOS

ANEXO I - NORMAS DE REDAÇÃO DE DISSERTAÇÃO OU TESE

1. Normas Gerais

1.1. Dissertação constitui o produto final de pesquisas desenvolvidas em cursos de Mestrado e a Tese constitui o produto final de pesquisas desenvolvidas em cursos de Doutorado. Exigem investigações próprias à área de especialização e métodos específicos.

1.2. A Dissertação ou Tese é de responsabilidade do discente, da Comissão Orientadora e da Banca Examinadora.

2. Estrutura

2.1. A Dissertação ou Tese deverá ser composta de: (i) capa, (ii) páginas pré-textuais, (iii) corpo propriamente dito e, (iv) anexo (páginas pós-textuais).

2.2. A capa deverá conter a autoria, título, local e ano da aprovação. As capas encadernadas em mais de um volume deverão conter as mesmas informações acrescidas da identificação do respectivo volume. Dois (2) exemplares devem ser de capas duras de cor preta e letras amarelas.

2.3. As páginas pré-textuais serão compostas:

2.3.1. Primeira folha interna (página de rosto), contendo: (i) autoria, (ii) título; (iii) nota explicativa de que se trata de um trabalho de Dissertação ou Tese, mencionando o Programa de Pós-Graduação, a Universidade e o grau pretendido (Mestrado ou Doutorado); (iv) comitê de orientação e (v) local e ano de aprovação. Conterá, no verso desta folha, a ficha catalográfica.

2.3.2. Segunda folha interna deve conter, o título, o nome do pós-graduando(a), a data de aprovação, os nomes e as assinaturas do orientador e dos participantes da Banca Examinadora, local e data.

2.3.3. Opcionalmente, poderão ser incluídas páginas adicionais contendo: (i) agradecimento (ii) oferecimento, (iii) dedicatória e (iv) biografia do autor, obrigatoriamente, deve conter (v) lista de símbolos, figuras, tabelas e sumário.

2.3.4. Folha (s) em que conste (m) o resumo em português, palavras-chave, o abstract em inglês e key words. O resumo com no máximo 800 palavras deve destacar: o local da pesquisa, delineamento estatístico, caracterização do problema, focalizar o(s) objetivo(s), síntese da metodologia, resultados obtidos e conclusões.

2.4. O corpo da Dissertação ou Tese conterá todo o trabalho impresso, avaliado e aprovado pela Banca Examinadora. O corpo poderá ser organizado na forma de capítulos.

2.5. O corpo em capítulos será composto das seções:

Capítulo I: Introdução e Referencial Teórico; Capítulos (I ou mais a depender do número de artigos científicos); e Considerações Finais (opcional). As referências bibliográficas e citações seguirão as normas da Crop Breeding and Applied Biotechnology. As referências bibliográficas deverão aparecer ao final de cada capítulo.

2.6. O anexo (páginas pós-textuais) conterá material pertinente e suplementar.

2.7. Inserir cabeçalho com citação do autor e nome da dissertação ou tese, sendo a fonte tipo arial e tamanho 10, a partir do Capítulo I até a página inicial da folha anexos.

3. Editoração

3.1. Composição tipográfica. As dissertações ou teses deverão ser impressas em forma permanente e legível, com caracteres de alta definição e de cor preta no tipo Arial tamanho 12, com espaçamento 1,5.

3.2. Notação científica e medidas. A nomenclatura científica deverá ser diferenciada contextualmente, de acordo com as normas internacionais. As unidades métricas deverão seguir o padrão do Sistema Internacional de Unidades.

3.3. Papel. Utilizar papel A-4 (210 x 297 mm) branco, e suficientemente opaco para leitura normal.

3.4. Margens. A margem esquerda deve ser de 3 cm e as outras margens de 2 cm.

3.5. Paginação. Todas as páginas textuais e pós-textuais deverão ser numeradas em sequência contínua, isto é, desde a página do Capítulo I (texto corrido), até a última página, em algarismos arábicos. A sequência deverá incluir tudo que estiver como mapas, diagramas, páginas em branco e outros. As páginas pré-textuais deverão ser numeradas, sequencialmente, como algarismos romanos minúsculos.

3.6. Ilustrações. Fotografias e outras ilustrações deverão ser montadas de forma definitiva e incluídas no corpo da Dissertação ou Tese. É admitido o uso de cores nas figuras e ilustrações. Em nenhuma circunstância dever-se-á empregar fita adesiva ou material similar para afixação de ilustrações no corpo da Dissertação ou Tese. Folhas de tamanho superior a A4 serão aceitáveis, desde que dobradas, de forma a resultar em dimensões inferiores ao tamanho do papel adotado.

ANEXO II - INSTRUÇÕES AOS AUTORES PARA PUBLICAÇÃO NA REVISTA CROP BREEDING AND APPLIED BIOTECHNOLOGY (CBAB)

Forma e preparação de manuscritos

O CBAB publica exclusivamente em inglês. O autor pode acompanhar as etapas de tramitação do manuscrito por meio de seu e-mail e senha pessoal. Revisores especialistas ad hoc avaliam os manuscritos para auxiliar o Conselho Editorial na decisão final de aprovação, modificação ou reprovação.

O manuscrito completo deve seguir a seguinte sequência: título, resumo, palavras-chave, introdução, material e métodos, resultados e discussão, agradecimentos, referências, tabelas e figuras. O manuscrito deve ser digitado em Word for Windows, em fonte times new roman 12, espaçamento duplo, formato A4, com margens de 20 mm e numeração consecutiva superior direita. O texto em espaço duplo não deve exceder 18 páginas, incluindo tabelas e figuras colocadas separadamente (uma por página) no final do texto. Todas as equações, modelos e símbolos devem ser feitos no Microsoft Equation. O título deve ser claro, conciso e expressar a essência do artigo. Não deve ultrapassar 15 palavras, deve ser digitado em negrito, à esquerda, com letras iniciais maiúsculas. Os nomes completos dos autores e seus endereços institucionais devem constar na prova lida. O resumo não deve conter mais de 150 palavras. São permitidas no máximo 5 palavras-chave, diferentes do título. A introdução deve incluir uma breve revisão da literatura sobre o assunto e os objetivos do estudo. Material e Métodos devem permitir que outros pesquisadores repitam a experiência. Preferencialmente, Resultados e Discussão devem ser apresentados juntos para facilitar a leitura. Os agradecimentos devem ser sucintos e limitados a colaboradores efetivos e agências financiadoras.

Cuidado com as referências. Nunca cite resumos de eventos e teses, ou qualquer outra literatura inédita. Essas medidas ajudarão a moldar um manuscrito que será um crédito tanto para o seu artigo quanto para o periódico. Citações citadas no texto pelo sobrenome do autor e ano (por exemplo, Liu 1998, Pereira e Amaral Júnior 2001, William et al. 1990) devem ser listadas alfabeticamente no item Referências, conforme os exemplos a seguir:

1) Artigos em revistas:

Pereira MG and Amaral Júnior AT (2001) Estimation of genetic components in popcorn based on the nested design. *Crop Breeding and Applied Biotechnology* 1: 3-10.

Knapp SJ, Stroup WW and Ross WM (1985) Exact confidence intervals for heritability on a progeny mean basis. *Crop Science* 25: 192-194.

2) Livro

Hallauer AR, Carena MJ and Miranda Filho JB (2010) *Quantitative genetics in maize breeding*. Springer, New York, 664p.

3) Capítulo de livro

Morais PPP e Borem A (2017) Cultivares GM. In Silva FL, Borem A, Sedyama T e Ludke WH (eds) **Melhoramento da soja**. Springer, Nova York, p. 174-189.

4) Congresso:

Frey KJ (1992) Plant breeding perspectives for the 1990s. In: Stalker HT and Murphy JP (eds.) *Proceedings of the symposium on plant breeding in the 1990s*. CAB, Wallingford, p. 1-13.

O CBAB publica, além de artigos, outras formas de texto, igualmente sujeitas ao critério de pareceristas ad hoc.

Resenha

Aos principais autores de determinados temas será solicitada uma Revisão pelo Conselho Editorial (também restrita a 18 páginas), que deverá lançar luz especificamente sobre temas instigantes que mereçam uma análise mais aprofundada em seu estágio de desenvolvimento.

Notas

As notas são limitadas a 12 páginas e destinam-se a informar sobre novos estudos ou observações, portanto, as ferramentas analíticas não são necessárias. Eles podem se concentrar em um assunto de interesse amplo; descrever brevemente um estudo original; relatório sobre pesquisa participativa; expressar observações de especial interesse nas áreas de pesquisa, ensino e ciências aplicadas; ou comentar sobre o

lançamento de um novo software em uma área relacionada ao melhoramento de plantas.

Programas de melhoramento de plantas

Os programas de melhoramento que se destacarem quanto à inovação, eficiência, impacto e/ou continuidade podem ser retratados no CBAB, limitado a 18 páginas.

Lançamento de cultivares

Novas cultivares merecem atenção especial pelo seu papel fundamental no melhoramento vegetal e, conseqüentemente, na agricultura nacional. Uma contribuição para esta seção deve incluir um resumo de no máximo 50 palavras, palavras-chave, uma introdução, mencionar os métodos de melhoria aplicados, características de desempenho, produção de sementes básicas e conter um mínimo de referências (seguir exemplos de referências de artigos), tabelas e figuras. O texto inteiro não deve exceder 12 páginas.

Versão do software/dispositivo

Novos softwares e dispositivos serão publicados na seção Especial pela importância que representam para o aperfeiçoamento atual e, portanto, para a agricultura. Esta nova seção deve conter Resumos limitados a 50 palavras, Palavras-chave, introdução, métodos, histórico utilizado para o desenvolvimento do manuscrito, características de desempenho, aplicações, formas de acesso e um mínimo de referências, tabelas e figuras. Todo o manuscrito será limitado a 12 páginas.

Revisão do livro

Esta nova seção foi criada para anunciar novos livros relacionados ao melhoramento de plantas. A contribuição para esta seção deve incluir dois exemplares do livro enviado pelo Autor. O livro será revisado por um parecerista especializado escolhido pelo Conselho Editorial para editar um resumo.

Cartas

Cartas curtas de interesse geral também são bem-vindas para publicação, sujeitas a alterações pelo Conselho Editorial por questões de espaço ou clareza de expressão.

ANEXO III - INSTRUÇÕES AOS AUTORES PARA PUBLICAÇÃO NA REVISTA GENETIC RESOURCES AND CROP EVOLUTION

Submissão de manuscrito

A submissão de um manuscrito implica: que a obra descrita não tenha sido publicada anteriormente; que não está sendo considerado para publicação em nenhum outro lugar; que sua publicação foi aprovada por todos os coautores, se houver, bem como pelas autoridades responsáveis – tácita ou expressamente – no instituto onde o trabalho foi realizado. A editora não será considerada legalmente responsável caso haja algum pedido de indenização.

Permissões

Os autores que desejam incluir figuras, tabelas ou passagens de texto que já foram publicadas em outro lugar devem obter permissão do(s) proprietário(s) dos direitos autorais para o formato impresso e online e incluir evidências de que tal permissão foi concedida ao enviar seus trabalhos. Qualquer material recebido sem tais evidências será considerado como originário dos autores.

Submissão on-line

Por favor, siga o hiperlink “Enviar manuscrito” e carregue todos os seus arquivos de manuscrito seguindo as instruções dadas na tela.

Arquivos Fonte

Certifique-se de fornecer todos os arquivos de origem editáveis relevantes em cada envio e revisão. Deixar de enviar um conjunto completo de arquivos de origem editáveis fará com que seu artigo não seja considerado para revisão. Para o texto do seu manuscrito, sempre envie em formatos comuns de processamento de texto, como .docx ou LaTeX.

Artigo de revisão

As revisões são pesquisas oficiais, equilibradas e acadêmicas dentro do escopo da revista. Uma revisão deve se concentrar em um tópico que seja oportuno e relevante para o campo. A exigência de equilíbrio não precisa impedir os autores de propor pontos de vista específicos, mas se houver controvérsias no campo, os autores devem tratá-las de maneira imparcial.

Trabalho de pesquisa

Espera-se que artigos de pesquisa originais relatem avanços científicos e contribuam com novos insights sobre questões de pesquisa claramente identificadas no escopo da revista. Eles devem conter uma hipótese nova e bem formulada, uma abordagem experimental sólida e resultados que confirmem ou rejeitem a hipótese. Use a seguinte ordem: Resumo, Introdução, Materiais e Métodos, Resultados, Discussão, Conclusões, Agradecimentos, Referências.

Comunicação curta

Esses artigos comunicam breves relatos de dados de pesquisas originais. As Comunicações Breves típicas incluem táxons de culturas recém-descritos, notas nomenclaturais, relatórios de missões de coleta e resultados de avaliação de material de banco de genes.

Notas sobre Culturas Negligenciadas e Subutilizadas

Este tipo de artigo é semelhante a uma comunicação curta e se concentra em culturas raramente usadas e parentes selvagens de culturas.

Revisão do livro

Resenhas de Livros são artigos curtos (cerca de 400 palavras) que discutem livros importantes publicados recentemente. O objetivo de uma resenha de livro é fornecer um breve resumo dos pontos fortes e fracos do livro e avaliar a utilidade geral do livro para o público a que se destina.

Folha de rosto

Certifique-se de que sua página de título contém as seguintes informações.

Título

O título deve ser conciso e informativo.

Informação sobre o autor

- O(s) nome(s) do(s) autor(es)
- A(s) afiliação(ões) do(s) autor(es), ou seja, instituição, (departamento), cidade, (estado), país
- Uma indicação clara e um endereço de e-mail ativo do autor correspondente
- Se disponível, o ORCID de 16 dígitos do(s) autor(es)

Se as informações de endereço forem fornecidas com a(s) afiliação(ões), elas também serão publicadas.

Para autores que são (temporariamente) não afiliados, iremos capturar apenas sua cidade e país de residência, não seu endereço de e-mail, a menos que especificamente solicitado.

Resumo

Forneça um resumo de 150 a 250 palavras. O resumo não deve conter abreviações indefinidas ou referências não especificadas.

Apenas para revistas de ciências da vida (quando aplicável)

- Número de registro do ensaio e data de registro para ensaios registrados prospectivamente
- Número de registro do ensaio e data de registro, seguidos de “registrado retrospectivamente”, para ensaios registrados retroativamente

Palavras-chave

Forneça de 4 a 6 palavras-chave que podem ser usadas para fins de indexação.

Texto

Formatação de texto

Os manuscritos devem ser submetidos em Word.

- Use uma fonte normal e simples (por exemplo, Times Roman tamanho 10) para o texto.
- Use itálico para dar ênfase.
- Use a função de numeração automática de páginas para numerar as páginas.
- Não use funções de campo.
- Use paradas de tabulação ou outros comandos para recuos, não a barra de espaço.
- Use a função de tabela, não planilhas, para fazer tabelas.
- Use o editor de equações ou MathType para equações.
- Salve seu arquivo no formato docx (Word 2007 ou superior) ou no formato doc (versões mais antigas do Word).

Manuscritos com conteúdo matemático também podem ser enviados em LaTeX. Recomendamos o uso do modelo LaTeX da Springer Nature.

Títulos

Use no máximo três níveis de cabeçalhos exibidos.

Abreviaturas

As abreviaturas devem ser definidas na primeira menção e usadas consistentemente depois disso.

Notas de rodapé

As notas de rodapé podem ser usadas para fornecer informações adicionais, que podem incluir a citação de uma referência incluída na lista de referências. Eles não devem consistir apenas em uma citação de referência e nunca devem incluir os detalhes bibliográficos de uma referência. Também não devem conter figuras ou tabelas.

As notas de rodapé do texto são numeradas consecutivamente; aqueles às tabelas devem ser indicados por letras minúsculas sobrescritas (ou asteriscos para valores de significância e outros dados estatísticos). As notas de rodapé do título ou dos autores do artigo não recebem símbolos de referência.

Sempre use notas de rodapé em vez de notas de fim.

Agradecimentos

Agradecimentos de pessoas, doações, fundos, etc. devem ser colocados em uma seção separada na página de título. Os nomes das organizações financiadoras devem ser escritos por extenso.

Observe:

- Use a função de numeração automática de linhas para numerar as linhas do manuscrito.

Estilo científico

- Sempre use sinais e símbolos aceitos internacionalmente para unidades (unidades SI).
- Nomenclatura: Na medida do possível, os autores devem usar nomes sistemáticos semelhantes aos usados pelo Chemical Abstract Service ou IUPAC.

Nomes científicos de plantas

Os autores devem usar o nome de planta latino aceito, incluindo a citação do autor (por exemplo, *Zea mays* L.) na primeira vez que uma espécie é mencionada, tanto no resumo quanto no texto principal. Em caso de dúvida, os autores podem consultar os seguintes sites:

<http://www.theplantlist.org/>

<https://npgsweb.ars-grin.gov/gringlobal/taxon/taxonomysearch.aspx>

A primeira menção de um nome de gênero deve ser escrita por extenso, subsequentemente os nomes de gênero devem ser abreviados para sua letra inicial (por exemplo, *Z. mays*).

Nomes científicos de plantas devem estar em itálico no nível de gênero e abaixo.

Referências

Citação

Cite as referências no texto por nome e ano entre parênteses. Alguns exemplos:

- A pesquisa em negociação abrange muitas disciplinas (Thompson 1990).
- Este resultado foi posteriormente contestado por Becker e Seligman (1996).
- Este efeito foi amplamente estudado (Abbott 1991; Barakat et al. 1995a, b; Kelso e Smith 1998; Medvec et al. 1999, 2000).

Lista de referência

A lista de referências deve incluir apenas os trabalhos citados no texto e que tenham sido publicados ou aceitos para publicação. Comunicações pessoais e trabalhos inéditos devem ser mencionados apenas no texto.

As entradas da lista de referências devem ser colocadas em ordem alfabética pelos sobrenomes do primeiro autor de cada trabalho. Por favor, coloque em ordem alfabética de acordo com as seguintes regras: 1) Para um autor, pelo nome do autor, depois cronologicamente; 2) Para dois autores, por nome do autor, depois nome do co-autor, depois cronologicamente; 3) Para mais de dois autores, pelo nome do primeiro autor, depois cronologicamente.

Se disponível, sempre inclua DOIs como links DOI completos em sua lista de referência (por exemplo, “<https://doi.org/abc>”).

- **Artigo de jornal**

Gamelin FX, Baquet G, Berthoin S, Thevenet D, Nourry C, Nottin S, Bosquet L (2009) Effect of high intensity intermittent training on heart rate variability in prepubescent children. *Eur J Appl Physiol* 105:731-738. <https://doi.org/10.1007/s00421-008-0955-8>

Idealmente, os nomes de todos os autores devem ser fornecidos, mas o uso de “et al” em listas longas de autores também será aceito:

Smith J, Jones M Jr, Houghton L et al (1999) Future of health insurance. *N Engl J Med* 341:325–329

- **Artigo por DOI**

Slifka MK, Whitton JL (2000) Clinical implications of dysregulated cytokine production. *J Mol Med*. <https://doi.org/10.1007/s001090000086>

- **Livro**

South J, Blass B (2001) *The future of modern genomics*. Blackwell, London

- **Capítulo de livro**

Brown B, Aaron M (2001) The politics of nature. In: Smith J (ed) *The rise of modern genomics*, 3rd edn. Wiley, New York, pp 230-257

- **Documento on-line**

Cartwright J (2007) Big stars have weather too. IOP Publishing PhysicsWeb. <http://physicsweb.org/articles/news/11/6/16/1>. Accessed 26 June 2007

- **Dissertação**

Trent JW (1975) Experimental acute renal failure. Dissertation, University of California

ANEXO IV - INSTRUÇÕES AOS AUTORES PARA PUBLICAÇÃO NA REVISTA BRAZILIAN ARCHIVES OF BIOLOGY AND TECHNOLOGY – BAPT

Instruções aos autores

Breve guia para submissão

Este guia descreve pontos-chaves para a preparação de manuscritos para submissão ao BAPT.

Para facilitar a elaboração do manuscrito, os modelos abaixo foram criados para uso obrigatório.

Carta de Apresentação (Obrigatório)

Ao enviar o artigo, os autores devem incluir uma Carta de Apresentação ao editor, resumindo os achados de seu manuscrito, evidenciando o que é de interesse do Brazilian Archives of Biology and Technology publicá-lo, sua originalidade e que não foi submetido ou publicado anteriormente em outros lugares. Além disso, a Carta de Apresentação deve abordar claramente a Área do Conhecimento a que se destina, que não há conflitos de interesse associados ao manuscrito e que atendam a todas as normas aplicáveis sobre a ética da experimentação e da pesquisa, e que não possui publicações duplicadas, fraude, plágio ou preocupações com experimentos animais ou humanos. A Carta de Apresentação também deve informar se o autor aceita ou não que seu artigo seja utilizado para gerar conteúdo para as redes sociais de divulgação do periódico.

Se o artigo já foi compartilhado como um preprint, informe em "Comentários aos Editores" na Carta de Apresentação.

A Carta de Apresentação deve seguir o formulário de modelo na extensão docx anexado aqui.

Por favor atenção ao anexar a carta de apresentação no momento da submissão. Arquivos anexados em locais incorretos não serão aceitos. Preferencialmente, anexe o arquivo na extensão .docx

Manuscrito Principal

O manuscrito deve ser escrito apenas em inglês. Se os autores não tiverem experiência em escrever nesta língua, sugerimos que o manuscrito seja revisado por

um especialista da língua inglesa antes de sua submissão, de preferência com experiência científica na Área do Conhecimento.

A revista BAPT é rigorosa ao formato de submissões iniciais; deve estar no modelo anexado aqui.

Importante: Os manuscritos que não cumprirem as normas estabelecidas para submissão serão devolvidos aos autores para adequações até que estejam corretos para serem enviados à apreciação dos revisores.

Todas as informações devem ser apresentadas em um único arquivo, formato .docx, incluindo as informações dos autores e ORCID ID. Todas as figuras e tabelas devem ser apresentadas ao longo do texto, e o manuscrito deve ter no máximo 15 páginas (incluindo texto, figuras e tabelas).

Ao usar o modelo, use a galeria de estilos para formatar seu manuscrito.

Um guia para os estilos contidos no modelo é mostrado abaixo:

Título

Deve ter no máximo 3 linhas

Fonte: Arial, 20 pt, negrito, justificado, espaçamento da linha: 1,0 (Simples), Espaço

Depois: 12 pt,

Deve ser preciso e refletir claramente o conteúdo do manuscrito. (Coloque cada palavra na parte superior).

Nome do autor

Sem títulos ou graduações, preferencialmente sem abreviaturas;

Fonte: Arial, 11 pt, Negrito, Justificado Espaçamento da Linha: 1,0 (Simples), Espaço

Depois: 0 pt,

O autor para correspondência deve ser identificado com um asterisco

ORCID

Formato: <https://orcid.org/0000-0000-0000-0000>

Fonte: Arial, 10 pt, esquerda, espaçamento da linha: 1,0 (simples), espaço depois: 12 pt,

É obrigatório para todos os autores.

IMPORTANTE: Se houver mais de 4 Autores, você pode colocar a lista em duas colunas.

Filiação

Os nomes das instituições e programas deverão ser apresentados por extenso e em inglês. Para instituições de países de escrita latina os nomes podem ser escritos na língua original.

Formato: Universidade, Faculdade, Departamento, Cidade, Estado, País;

Fonte: Arial, 10 pt, esquerda, espaçamento da linha: 1,0 (simples), espaço antes e depois: 0 pt

É obrigatório para todos os autores.

Destaques

Deve resumir concisamente os principais achados e fornecer aos leitores uma visão geral do artigo. Devem incluir de três a cinco pontos, descrevendo as principais realizações da pesquisa (por exemplo, resultados ou conclusões) e destacam o que é único sobre o estudo

Inclua 3 a 5 destaques;

No máximo 85 caracteres em cada destaque, incluindo espaços;

Apenas os resultados principais do manuscrito devem ser destacados;

Escreva o destaque da pesquisa no tempo presente;

Seja conciso e específico;

Fonte: Arial, 11 pt, justificado, espaçamento da linha: 1,5

Abstract

Fonte: Arial, 11 pt, justificado, espaçamento da linha: 1,0 (simples), espaço antes: 12 pt.

A palavra “abstract” antes é em negrito.

Um único parágrafo de cerca de 250 palavras, no máximo. Para artigos de pesquisa, os resumos devem dar uma visão geral pertinente do trabalho. Encorajamos fortemente os autores a utilizar o seguinte estilo de resumos estruturados, mas sem cabeçalhos: Antecedentes: Colocar a questão abordada em um contexto amplo e destacar o objetivo do estudo; Métodos: Descrever sucintamente os principais métodos ou tratamentos aplicados; Resultados: Resumir os principais resultados do artigo; e Conclusão: Indicar as principais conclusões ou interpretações. O resumo deve ser uma representação objetiva do artigo, não deve conter resultados que não sejam apresentados e fundamentados no texto principal e não deve exagerar a conclusão principal.

Keywords

palavra-chave 1; palavra-chave 2; palavra-chave 3. Listar de três a seis palavras-chave pertinentes específicas do artigo; ainda razoavelmente comum dentro da disciplina do assunto.

Fonte: Arial, 11 pt, justificado, espaçamento da Linha: 1,0 (Simples), espaço Antes 12 pt.

Resumo Gráfico (Opcional)

Um Resumo Gráfico é um resumo único, conciso, pictórico e visual dos principais achados do artigo. Pode ser a figura final do artigo ou uma figura especialmente projetada para o efeito, que captura o conteúdo do artigo para os leitores em uma única imagem.

Seções principais

Introdução, Material e Métodos, Resultados, Discussão, Conclusão, Patentes, Referências.

Fonte: Arial, 11 pt., negrito, justificado, espaçamento de linha: 1,0 (simples), espaço antes: 12 pt., espaço Depois: 6 pt

Subseções

Subseção

Fonte: Arial, 11 pt. negrito, justificado, espaçamento de linha: 1,0 (Simples), espaço antes: 12 pt. espaço depois: 6 pt.

Subsubseção

Fonte: Arial, 11 pt. itálico, justificado, espaçamento de linha: 1,0 (Simples), espaço antes: 12 pt. espaço depois: 6 pt.

Subsubsubseção

Fonte: Arial, 11 pt., justificado, espaçamento de linha: 1,0 (Simples), espaço antes: 12 pt. espaço depois: 6 pt.

Introdução

Fonte: Arial, 11 pt., justificado, espaçamento de linha: 1,0 (simples), espaço antes: 0 pt., espaço depois: 0 pt., primeira linha recuo de 0,75cm.

A introdução deve conter brevemente o estudo em um contexto amplo e destacar porque é importante; definir o propósito da obra e seu significado. O estado atual do

campo de pesquisa deve ser revisto cuidadosamente e as principais publicações citadas. Por favor, destaque hipóteses controversas e divergentes quando necessário. Por fim, mencione brevemente o objetivo principal do trabalho. As referências devem ser numeradas na ordem em que aparecem e indicadas numericamente entre colchetes, por exemplo, [1] ou [2,3], ou [4-6]. Consulte o final do documento para obter mais detalhes sobre as referências.

Materiais e Métodos

Fonte: Arial, 11 pt., justificado, espaçamento de linha: 1,0 (simples), espaço antes: 0 pt., espaço depois: 0 pt., primeira linha recuo de 0,75cm.

Materiais e métodos devem ser descritos com detalhes suficientes para permitir que outros repliquem e construam os resultados publicados. Por favor, note que a publicação do seu manuscrito implica que você deve disponibilizar todos os materiais, dados, códigos de computador e protocolos associados à publicação. Revele na fase de submissão quaisquer restrições à disponibilidade de materiais ou informações na carta de apresentação. Novos métodos e protocolos devem ser descritos em detalhes, enquanto métodos bem estabelecidos podem ser brevemente descritos e apropriadamente citados. Esta seção pode ser dividida por subseções como mostrado acima.

Resultados

Fonte: Arial, 11 pt., justificado, espaçamento de linha: 1,0 (simples), espaço antes: 0 pt., espaço depois: 0 pt., recuos de primeira linha de 0,75cm.

Deve conter uma descrição concisa e precisa dos resultados experimentais, sua interpretação bem como a conclusão experimental que pode ser traçada.

Tabelas e Esquemas de Figuras

Coloque todas as tabelas e esquemas de figuras conforme são citadas no texto.

Todas as figuras e tabelas devem ser citadas no texto principal numeradas consecutivamente em dígitos arábicos como Figura 1, Tabela 1, etc.

Tabelas

Fonte da tabela: Arial, 10 pt., justificado, espaçamento de linha: 1,0 (simples), espaço antes: 12 pt., espaço depois: 0 pt.

Cada tabela deve ser acompanhada por uma frase de título curta, descrevendo o que a tabela mostra. Mais detalhes podem ser incluídos como notas de rodapé.

Fonte do corpo da tabela: Arial, 10 pt alinhada à esquerda, espaçamento de linha: 1,0 (simples), espaço antes: 0 pt., espaço depois: 0 pt.

Linhas verticais e diagonais e preenchimento não devem ser usadas em tabelas.

Fonte do rodapé: Arial, 10 pt., centralizado, espaçamento de linha: 1,0 (simples), espaço antes: 0 pt., espaço depois: 12 pt.

IMPORTANTE: As tabelas devem ser elaboradas e editadas em células, utilizando os recursos do editor de texto.

Observação: Tabelas em formato de figura não são aceitas ou que contenham linhas desenhadas, ou fundo preenchido.

Tabelas muito grandes para páginas com orientação retrato devem ser colocadas em uma página separada na sequência que aparecem no texto, no modo paisagem, como na página de exemplo contida no modelo.

Figuras e Esquemas

Legenda da figura Fonte: Arial, 10 pt., Justificado, Espaçamento de Linha: 1,0 (Simples), Espaço Antes: 0 pt., Espaço Depois: 12 pt.;

Se houver vários painéis, eles devem ser listados como: (a) descrição do que está contido no primeiro painel; (b) descrição do que está contido no segundo painel;

As imagens devem ter uma resolução mínima de 300 dpi.;

A legenda deve ter uma única linha e estar centralizada;

As figuras devem ser submetidas preferencialmente em cores, os textos dentro das figuras devem ser legíveis;

Quando as figuras apresentam diferentes partes, cada uma deve ser indicada por maiúscula (Figure 1 A, B, C etc.).

Equações e Teoremas

Equações não devem estar em formato de imagem;

Elas devem ser numeradas;

Equação: Arial, 11 pt., centralizado, espaçamento de linha: 1,0 (simples), espaço antes: 6 pt., espaço depois: 6 pt.;

Teoremas: Arial, 11 pt., itálico justificado, espaçamento de linha: 1,0 (simples), espaço antes: 12 pt., espaço depois: 8 pt. O título “Teorema n^o” no início da frase é em negrito;

Prova de Teoremas: Arial, 11 pt., justificado, espaçamento de linha: 1,0 (simples), espaço antes: 12 pt., espaço depois: 8 pt. O título “Prova de Teorema nº” no início da frase é em negrito.

Discussão

Fonte: Arial, 11 pt., justificado, espaçamento de linha: 1,0 (simples), espaço antes: 0 pt., espaço depois: 0 pt., recuos de primeira linha de 0,75cm.

Os autores devem discutir os resultados e como podem ser interpretados na perspectiva de estudos anteriores e das hipóteses de trabalho. Os achados e suas implicações devem ser discutidos no contexto mais amplo possível. Futuras direções de pesquisa também podem ser destacadas.

Conclusões

Fonte: Arial, 11 pt., justificado, espaçamento de linha: 1,0 (simples), espaço antes: 0 pt., espaço depois: 0 pt., recuos de primeira linha de 0,75cm.

Esta seção não é obrigatória, mas pode ser adicionada ao manuscrito se a discussão for extraordinariamente longa ou complexa.

Patentes

Fonte: Arial, 11 pt., justificado, espaçamento de linha: 1,0 (simples), espaço antes: 0 pt., espaço depois: 0 pt., recuos de primeira linha de 0,75cm.

Esta seção não é obrigatória, mas pode ser adicionada se houver patentes resultantes do trabalho relatado neste manuscrito.

Financiamento

Fonte: Arial, 10 pt., justificado, espaçamento de linha: 1,0 (simples), espaço antes: 12 pt., espaço depois: 0 pt. A palavra “Financiamento” no início da frase é em negrito.

Por favor, adicione: "Esta pesquisa não recebeu nenhum financiamento externo" ou "Esta pesquisa foi financiada pelo NOME DO FINANCIADOR", número de subvenção “XXX”. Verifique cuidadosamente se os dados são precisos.

Agradecimentos

Fonte: Arial, 10 pt., justificado, espaçamento de linha: 1,0 (simples), espaço antes: 0 pt., espaço depois: 0 pt. A palavra Agradecimentos no início da frase deve ser em negrito.

Nesta seção você pode reconhecer qualquer apoio dado que não esteja coberto pelas seções de contribuição ou financiamento do autor. Isso pode incluir suporte administrativo e técnico, ou doações em espécie (por exemplo, materiais usados para experimentos).

Referências

As referências devem ser numeradas conforme citadas no texto (incluindo citações em tabelas e legendas) e listadas individualmente no final do manuscrito. Recomendamos preparar as referências com um pacote de software de bibliografia, como EndNote, Reference Manager ou Zotero para evitar erros de digitação e referências duplicadas. No texto, os números de referência devem ser colocados em colchetes [], e colocados antes da pontuação; por exemplo [1], [1-3] ou [1,3].

Fonte: Arial, 10 pt., alinhada esquerda, espaçamento de linha: 1,0 (simples), espaço antes: 0 pt., espaço depois: 0 pt., numeradas.

A lista completa de referências, no final do artigo, deve seguir o estilo Vancouver:

Guia de Referências

Os manuscritos podem ser submetidos seguindo os “Uniform Requirements for Manuscripts Submitted to Biomedical Journals” produzidos pelo Committee of Medical Journal Editors também conhecido como o Estilo Vancouver. Neste caso, os autores devem seguir as diretrizes no link abaixo e serão responsáveis por modificar o manuscrito onde ele difere das instruções dadas aqui.

http://www.nlm.nih.gov/bsd/uniform_requirements.html

IMPORTANTE: Não haverá restrições na língua de origem do artigo de referência, apesar da recomendação de estar em inglês, porém o título deve ser traduzido e apresentado em inglês, entre colchetes. – Demais informações deverão ser preferencialmente na língua inglesa.

Importante: Salve o arquivo em formato .doc ou .docx.