

RESSALVA

Atendendo solicitação do autor, o texto completo desta dissertação será disponibilizado somente a partir de 29/11/2024.





PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS
(BIOLOGIA VEGETAL)

DESENVOLVIMENTO DE RECURSOS GENÔMICOS E TRANSCRICIONAIS EM

Quillaja brasiliensis

GIAN CARLOS COSTA BARROS

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS
(BIOLOGIA VEGETAL)

DESENVOLVIMENTO DE RECURSOS GENÔMICOS E TRANSCRICIONAIS EM
Quillaja brasiliensis

GIAN CARLOS COSTA BARROS

Dissertação apresentada ao Instituto de Biociências do Câmpus de Rio Claro, Universidade Estadual Paulista, como parte dos requisitos para obtenção do título de Mestre em Ciências Biológicas (Biologia Vegetal).

Orientador: Douglas Silva Domingues

Rio Claro – SP
2022

B277d Barros, Gian Carlos Costa
Desenvolvimento de recursos genômicos e transcricionais
em *Quillaja brasiliensis* / Gian Carlos Costa Barros. -- Rio
Claro, 2022
66 p.

Dissertação (mestrado) - Universidade Estadual Paulista
(Unesp), Instituto de Biociências, Rio Claro
Orientadora: Douglas Silva Domingues

1. *Quillaja brasiliensis*. 2. saponinas triterpênicas. 3.
biotecnologia. 4. recursos moleculares. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do
Instituto de Biociências, Rio Claro. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

CERTIFICADO DE APROVAÇÃO

TÍTULO DA DISSERTAÇÃO: Desenvolvimento de recursos genômicos e transcricionais em *Quillaja brasiliensis*

AUTOR: GIAN CARLOS COSTA BARROS

ORIENTADOR: DOUGLAS SILVA DOMINGUES

Aprovado como parte das exigências para obtenção do Título de Mestre em Ciências Biológicas (Biologia Vegetal), área: Biologia Vegetal pela Comissão Examinadora:



Pesquisador Dr. DOUGLAS SILVA DOMINGUES (Participação Virtual)
Departamento de Botânica / IB UNESP Rio Claro

Prof. Dr. ARTHUR GERMANO FETT-NETO (Participação Virtual)
Botânica - Instituto de Biociências / Universidade Federal do Rio Grande do Sul

Profa. Dra. MARINA ALVES GAVASSI (Participação Virtual)
Instituto de Biociências de Rio Claro-IB-UNESP / Rio Claro/SP

Rio Claro, 29 de novembro de 2022

Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001. Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), processo 2019/15477-3. Para o projeto “Moléculas sinalizadoras na modulação do transcriptoma, fluxo metabólico e produção de saponinas triterpênicas imunoadjuvantes em *Quillaja brasiliensis* Mart. (Quillajaceae)”. Ao Instituto de Biociências de Rio Claro pelo suporte durante o desenvolvimento da minha dissertação; agradecimentos e carinho especial ao grupo de Genômica e Transcricção de plantas do departamento de Botânica, especialmente pela orientação do Prof. Douglas Silva Domingues; aos meus amigos e familiares, por todo o apoio e carinho durante os anos de vivência.

Resumo

A Família Quillajaceae pertence à ordem Fabales e é composta por duas espécies: *Quillaja saponaria* e *Quillaja brasiliensis*, ambas nativas da América do Sul. Esse gênero é conhecido, do ponto de vista biotecnológico, pela produção de saponinas triterpênicas com uso na indústria farmacêutica como imunoestimulante. Apesar do potencial biotecnológico da espécie, *Q. brasiliensis* não possui dados genômicos ou transcricionais públicos que permitiriam maiores avanços no desenvolvimento de ferramentas biotecnológicas. Nesse contexto, o presente estudo teve por objetivo o desenvolvimento de recursos genômicos e transcricionais para *Quillaja brasiliensis*. Foram desenvolvidos recursos genômicos por meio da plataforma Illumina de sequenciamento, e recursos transcricionais por meio das plataformas PacBio e Illumina, com avaliação de folhas e culturas celulares. A partir dos dados genômicos, foi montado o genoma cloroplastidial de *Q. brasiliensis*, que apresenta 160528 pares de base. Com ele, foram realizados estudos filogenéticos comparativos, incluindo-se dados plastidiais de espécies das famílias Quillajaceae, Surianaceae e Polygalaceae. Com base nos recursos transcricionais PacBio, foi montado um transcriptoma de referência para *Q. brasiliensis*. Com base nessa referência, identificamos por sequenciamento Illumina 588 genes com estabilidade potencial para uso como normalizadores, dos quais 13 foram selecionados para testes iniciais por PCR. Três deles foram selecionados para teste em RT-qPCR: *Heat shock protein Hsp90* (HSP), *Cyclophilin-like domain* (CYP) e *Ubiquitin* (UBQ). Estes genes tiveram sua estabilidade em folhas, caule e raízes. HSP foi o gene que apresentou os resultados mais consistentes de estabilidade considerando-se os três tecidos avaliados. Nas análises individuais em cada tecido, CYP foi o gene normalizador mais bem ranqueado para folha e raiz e HSP o mais bem ranqueado em caule. Com isso, o presente estudo estabelece recursos moleculares iniciais em *Q. brasiliensis*, que podem ser utilizados tanto para o estudo do genoma e análises filogenéticas bem como para a descoberta de genes candidatos à produção de saponinas triterpênicas.

Palavras-chave: *Quillaja brasiliensis*, saponinas triterpênicas, biotecnologia, recursos moleculares.

Abstract

The Quillajaceae family belongs to the order Fabales and it is composed by two species: *Quillaja saponaria* and *Quillaja brasiliensis*, both native to South America. This genus is known, from a biotechnological point of view, by the production of triterpene saponins used in the pharmaceutical industry as an immunostimulant. Despite the species biotechnological potential, *Q. brasiliensis* does not have public genomic or transcriptional data that would allow further advances in the development of biotechnological tools. In this context, the present study aimed to develop genomic and transcriptional resources for *Quillaja brasiliensis*. Genomic resources were developed through the Illumina sequencing platform, and transcriptional resources through the PacBio and Illumina platforms, with evaluation of leaves and cell cultures. From the genomic data, the chloroplastidial genome of *Q. brasiliensis* was assembled, which has 160,528 base pairs. We then carried out comparative phylogenetic studies, including plastid data from species of the Quillajaceae, Surianaceae, and Polygalaceae families. Based on PacBio cDNA sequencing, a reference transcriptome for *Q. brasiliensis* was assembled. Based on this reference, we identified by Illumina sequencing 588 genes with potential stability for use as normalizers, of which 13 were selected for PCR testing. Three of them were selected for tests in RT-qPCR: Heat shock protein Hsp90 (*HSP*), Cyclophilin-like domain (*CYP*) and Ubiquitin (*UBQ*). These genes had their stability evaluated in leaves, stems and roots. *HSP* was the gene that presented the most consistent stability results considering the three evaluated tissues. In the individual analyzes in each tissue, *CYP* was the highest ranked normalizing gene for leaf and root and *HSP* was the highest ranked in stem. Thus, the present study establishes initial molecular resources in *Q. brasiliensis*, which can be used both for the study of the genome and phylogenetic analyzes as well as for the discovery of candidate genes for the production of triterpene saponins.

Key words: *Quillaja brasiliensis*, triterpenic saponins, biotechnology, molecular resources.

Sumário

Resumo	4
Abstract	5
1. Introdução	8
1.1. Genoma dos cloroplastos.....	12
1.2. Sequenciamento Illumina	13
1.3. Sequenciamento Iso-seq.....	14
1.4. Genes Normalizadores.....	15
2. Objetivo	16
The complete chloroplast genome of <i>Quillaja brasiliensis</i> and comparative analysis in Quillajaceae and Fabales.....	19
Introduction.....	19
Material and methods	19
Plant material, DNA extraction and sequencing.....	19
Genome assembling, annotation and SSR analysis	20
Sequence divergence analysis	20
Phylogenetic analysis	21
Results.....	21
<i>Q. brasiliensis</i> and <i>Q. saponaria</i> assembly	21
Overall organization and general features of the plastomes	21
Discussion	31
Changes in the content of plastomes: gene gain/loss	31
Variable regions were identified based on plastome sequences.....	32
Phylogenomic analysis and relationship among species.....	32
Conclusion	33
Reference	34
Análises transcricionais em <i>Quillaja brasiliensis</i> : identificação de genes normalizadores.....	38

Introdução	38
Material e metodos.....	40
Material vegetal e extração de RNA	39
RNA-seq: sequenciamento Iso-seq e Illumina	40
Montagem dos transcritos e quantificação de expressão	41
Seleção de genes normalizadores e desenho de <i>primers</i>	42
Síntese de cDNA e PCR convencional	42
RT-qPCR	43
Estabilidade de genes de referência.....	43
Resultados e discussão.....	44
Sequenciamento Iso-seq: estatísticas gerais.....	44
Normalizadores e análises transcricionais	47
Identificação de candidatos a genes normalizadores para RT-qPCR....	47
Considerações finais.....	56
Referências	56

1. Introdução

A Família Quillajaceae pertence à ordem Fabales e é composta apenas por duas espécies: *Quillaja brasiliensis* (St. Hil. et Tul.) (Atualmente, *Q. lancifolia*) e *Quillaja saponaria*. A espécie *Quillaja brasiliensis* (Figura 1) é nativa do nordeste da Argentina, sul do Brasil e Uruguai. *Q. saponaria* tem sua ocorrência descrita no Chile



Figura 1: Espécie *Quillaja brasiliensis* (Elaboração Própria).

central (MARTICORENA & MARCHESI 2008, LUEBERT 2013).

Economicamente, esse gênero está relacionado à produção de saponinas triterpênicas. Em espécies angiospermas dicotiledôneas, estas moléculas são usualmente compostas de um triterpeno - terpenos de 30 carbonos, derivados de seis moléculas de isopreno, provenientes do metabolismo citoplasmático de terpenoides (LIAO et al., 2016) - hidrofóbico, muitas vezes chamado de sapogenina.

Além disso, as saponinas contêm de quatro a cinco anéis de agliconas hidrofóbicas. A partir de reações enzimáticas mediadas por enzimas do tipo glicosiltransferases, grupos de açúcar são adicionados em posições específicas dos anéis, bem como podem também ocorrer oxidações mediadas por enzimas P450 monooxigenases (SEKI et al., 2015). Essas agliconas apresentam aspecto apolar e açúcares, que em contato com solução aquosa apresentam um comportamento próximo ao sabão (BISWAS & DWIVEDI 2019; MAGEDANS et al., 2019). Dessa forma, as sapogeninas são agliconas usualmente “decoradas” com sacarídeos que

vão compor a parte hidrofílica da molécula (RIBEIRO et al., 2020; MAGEDANS et al., 2021).

Do ponto de vista ecológico, estes compostos fornecem proteção devido ao seu sabor amargo e adstringente, que impede animais herbívoros de consumirem a planta. As saponinas também servem como substâncias de proteção devido à sua toxicidade para animais monogástricos e insetos herbívoros (RIBEIRO et al., 2020).

As saponinas são moléculas altamente versáteis e apresentam aplicações em diversas atividades industriais. No setor farmacêutico, são relevantes por suas atividades como imunoadjuvantes (promovendo respostas imunes em formulações de vacinas veterinárias e humanas), entre outras propriedades farmacêuticas. Elas também são produtos relevantes na indústria alimentícia. Alguns de seus principais usos são como agentes espumantes em bebidas e emulsificantes em alimentos processados, incluindo encapsulamento e estabilização de compostos bioativos (MAGEDANS et al., 2021).

As saponinas são base de medicamentos expectorante, anti-inflamatório, hipocolesterolêmico e hemolítico. O extrato patenteado Q-Naturale® foi o composto, oriundo de *Q. saponária*, mais estudado no que diz respeito a sua ação imunoadjuvante e emulsificante (DE FARIA et al., 2017; ALIMENTARIUS 2018; FLECK et al., 2019.). Outros compostos de saponinas, como a frações de extratos da casca de *Quillaja saponaria*, exemplo da QS-21, foram testados como imunoestimulante (MARCIANI, 2018).

As saponinas de *Q. brasiliensis*, assim como saponinas de *Q. saponaria* formaram complexos imunoestimulantes (ISCOMs) (CIBULSKI et al., 2016), garantindo para as saponinas de *Q. brasiliensis* uma capacidade de estímulo do sistema imunológico

A ação imunoadjuvante de saponinas de *Quillaja* ocorre de duas maneiras. Em células T, o grupo aldeído se une a QS-21 e desencadeia uma cascata molecular auxílio de quinase ativada por mitógeno (MAP) (ERK2) e em conjunto com transportadores de K⁺ e Na⁺ resultado na secreção de citocinas Th1. Já em células dendríticas as saponinas agem como facilitadoras da liberação da vesícula endossomal, garantindo uma resposta rápida dessas células e a comunicação imediata com células T medidas por apresentação cruzada do antígeno (MARCIANI 2018). Estes eventos estão descritos na figura 3.

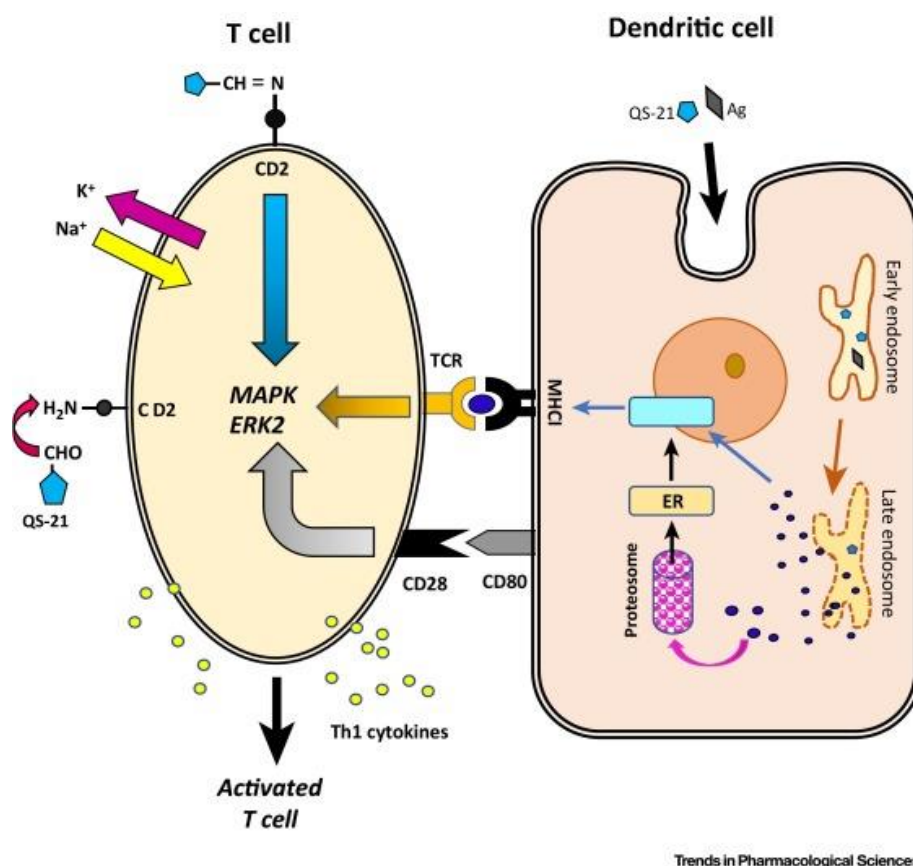


Figura 3: Ação das saponinas em células T e células dendríticas para resposta imunológica. Fonte: MARCIANI 2018.

Nos últimos anos, diversos estudos caracterizaram bioquimicamente as saponinas oriundas da representante brasileira *Quillaja brasiliensis* com potencial imunológico e identificaram características bioquímicas semelhantes as saponinas presentes em *Q. saponaria* (FLECK et al., 2019).

Em estudos de avaliação da capacidade imunoadjuvante de *Q. brasiliensis*, a fração QB-80 foi utilizada para síntese do neoadjuvante IMXQB-80 usado em vacinas experimentais inativadas contra o vírus Zika. Estudos com ratos submetidos a esse neoadjuvante indicaram um aumento significativo de imunoglobulinas IgG e outros subtipos relacionados ao Zika vírus (IgG1, IgG2b, IgG2c), assim como anticorpos na comparação com testes de vacinas sem adjuvante (CIBULSKI et al., 2021).

A espécie *Q. saponaria* é a mais explorada comercialmente para aplicabilidade na indústria alimentícia, de cosméticos e farmacêutica (GONZALEZ &

SÖRENSEN 2020; MA et al., 2020; DAI et al., 2020). As saponinas são extraídas da casca dessa árvore, demandando uma grande quantidade de material vegetal. Assim, as pesquisas recentes investigam formas alternativas de obtenção deste composto (REICHERT et al., 2019).

Dessa forma, a obtenção de saponinas de *Q. brasiliensis* tem maior potencial de sustentabilidade que as de *Q. saponaria*, uma vez que as saponinas estão em maior concentração na folha. Em função desse maior potencial de uso sustentável, o processo de indução da produção de saponinas em *Quillaja brasiliensis* tem sido estudado. Por exemplo, folhas sob efeito de luz UV-C produzem mais saponinas bioativas (YENDO et al., 2015). Outro caminho para o aumento da produção de saponinas bioativas seria por meio de extração destes compostos em culturas celulares (BISWAS E DWIVEDI, 2019).

Como o conhecimento de agentes e condições que estimulam o acúmulo de saponinas é de grande interesse e vários estudos têm sido realizados com esse objetivo. A elucidação das etapas de sua biossíntese em nível genômico e transcricional é um passo crucial nesse sentido.

Uma drástica diminuição dos custos de sequenciamento nos últimos anos tem possibilitado um maior número de projetos de sequenciamento de larga escala (KUMAR et al., 2019). Novos projetos surgiram para a geração de dados genômicos - como o The Earth Biogenome Project (EBP) (<https://www.earthbiogenome.org/>) que pretende contribuir para as diversas áreas do conhecimento como a conservação de espécies ameaçadas.

Em função da proliferação de dados de sequenciamento, quase 4000 genomas cloroplastidiais estão depositados no NCBI (DOBROGOJSKI et al., 2020). Análises do genoma dos cloroplastos obtidos a partir do sequenciamento de nova geração trouxeram maior entendimento sobre relações evolutivas dentro de grupos vegetais além de possibilitar aplicações biotecnológicas como desenvolvimento de vacinas e biofármacos (DANIELL et al., 2016).

Os avanços nas técnicas de sequenciamento também possibilitaram um desenvolvimento acelerado nas análises de RNA. Hoje o RNAseq se transformou em rotina nos trabalhos de biologia molecular. Com ele conseguimos identificar aspectos biológicos do RNA, compreender funcionalmente genes e genomas além de investigar a regulação de genes em diversos organismos (STARK et al., 2019).

No momento, a despeito de seu grande potencial biotecnológico, o desenvolvimento de recursos moleculares para a descoberta de vias metabólicas e compreensão de relações filogenéticas em Quillajaceae é relativamente incipiente. O único estudo público de transcriptoma em Quillajaceae foi o realizado na iniciativa “1kp transcriptomes” (One Thousand Plant Transcriptomes Initiative, 2019), no qual foi sequenciado o transcriptoma foliar de *Q. saponaria* pela técnica de sequenciamento Illumina. Com esta mesma técnica, foi obtido o sequenciamento do genoma de cloroplasto de *Q. saponaria* (Vizoso et al., 2019). Já para *Q. brasiliensis*, não há no momento dados públicos de sequenciamento.

1.1. Genoma dos cloroplastos

No aspecto metabólico, o cloroplasto pode ser considerado a organela mais importante presente nas plantas superiores. O mecanismo bioquímico e biofísico realizado por essa organela está diretamente relacionada a sua função no uso da radiação solar para gerar precursores na geração de energia (Helmut Kirchhoff 2019).

A função dos cloroplastos não se resume ao processo de aproveitamento da radiação solar para a síntese de açúcares para o metabolismo vegetal, mas também estão relacionados a sintetizam fitohormônios, ácidos graxos, aminoácidos além de outros metabólitos secundários, assim atuando diretamente nos aspectos do desenvolvimento vegetal, respostas ao ambiente e a imunidade vegetal (LITTLEJOHN et al., 2021).

As características funcionais e dos cloroplastos estão relacionadas ao seu genoma. Resultado de um processo ancestral de endossimbiose entre organismos eucariotos e uma cianobactéria. Os cloroplastos possuem genoma próprio e altamente diversificado.

A organização genômica dos cloroplastos é bastante variada, podendo apresentar conformações circulares ou linear de DNA, variando de 15.553 pares de bases (pb) em *Asarum minus* a 521.168 pb em *Floydiella terrestris* (BROUARD et al., 2010). Proteínas presentes nos cloroplastos nem sempre são sintetizadas pelo núcleo das células, podendo variar de 0 a 315 genes codificados pelo DNA cloroplastidial. No geral, o genoma cloroplastidial (cpDNA) está organizado em dois fragmentos idênticos chamados repetições invertidas (IR) separadas por uma seção

longa de cópia única (LSC) e uma seção curta de cópia única (SSC) (DOBROGOJSKI et al., 2020).

O sequenciamento de cloroplastos trouxe uma maior compreensão de diversos aspectos biológicos, como a o maquinário transcricional celular e ainda aspectos evolutivos (SANITÁ LIMA et al., 2016). Mesmo com os avanços nas técnicas de sequenciamento de nova geração os cloroplastos têm ficado em segundo plano na comparação com o sequenciamento de mitocôndrias por exemplo (SMITH 2018).

Devido às suas características únicas como a presença de genes altamente conservados e a herança uniparental, o genoma cloroplastidial é fundamental para a reconstrução filogenética de plantas (FANG et al 2017). As filogenias que utilizam os cloroplastos sustentam a monofilia de grandes grupos, como por exemplo a de briófitas (ZHONG et al., 2014) Ou ainda de grupos de importância econômica como o grupo das fabales (AYGOREN ULUER et al., 2020). Mas para que as análises filogenéticas que utilizam dados genômicos de cloroplasto ocorra, é preciso conhecer estes genomas por completo nas diferentes linhagens. Um exemplo é a família foco deste estudo, Quillajaceae, que possui apenas um genoma cloroplastidial reportado de *Q. saponaria* (VIZOSO et al., 2019), mas não existem até o momento trabalhos de sequenciamento com a espécie *Quillaja brasiliensis*.

1.2. Sequenciamento Illumina

A plataforma Illumina foi de extrema importância para o desenvolvimento de estudos relacionados ao sequenciamento de genomas e transcriptomas por trazer rapidez e diminuir custos da técnica de sequenciamento. Este tipo de sequenciamento é baseado na hibridização do DNA a oligonucleotídeos presentes em uma plataforma sólida (lâmina de vidro). diferente da técnica de sanger baseada na amplificação por PCR(VAN DIJK et al., 2018).

O DNA é fragmentado e adaptadores são adicionados em suas extremidades para que elas sejam hibridizadas na lâmina de vidro, em seguida o fragmento se dobra formando pontes. Rodadas sucessivas de síntese de fitas complementares são realizadas nas “pontes” formadas. A adição de nucleotídeos marcados faz com que a fluorescência seja detectada e assim as sequências são elucidadas (KUMAR et al., 2019). Os dados de saída do sequenciamento illumina variam de 1,2GB a 360GB dependendo do número de leituras realizadas para uma dada amostra. Na

plataforma Illumina, as leituras apresentam tamanho entre 150 e 300 bp (SLATKO et al 2018).

O método de sequenciamento utilizado pela plataforma Illumina tem como maior limitação o tamanho de sequência gerado, que vem sendo superado por novas técnicas como a plataforma PacBio e Oxford Nanopore. No sequenciamento illumina, assim como sequenciamento de DNA é possível realizar o sequenciamento do RNA a partir da síntese de um DNA complementar cDNA. O conjunto total de RNA expresso pelas células um organismo é chamado de transcriptoma. Em resumo, a técnica é baseada na fragmentação do RNA gerando leituras curtas assim essas leituras são montadas a partir de ferramentas de Bioinformática (STARK et al., 2019).

O desenvolvimento dessa tecnologia trouxe uma expansão aos estudos funcionais de aspectos biológicos como a regulação e descoberta de genes (STARK et al., 2019). O desenvolvimento dos estudos relacionados ao sequenciamento de RNA trouxe, especificamente para as plantas, possibilidades aos estudos evolutivos e de melhoramento (LE NGUYEN et al., 2019).

Com estes dados, podem ser descobertos genes relacionados à síntese de compostos de interesse em plantas, como genes relacionados à produção de terpenoides em cúrcuma (ANNANDURAI et al., 2013) e pitanga (GUZMAN et al., 2014), bem como a coloração de frutos em aração (VETÖ et al., 2020).

1.3. Sequenciamento Iso-seq

Como uma forma de melhorar o rendimento dos métodos de sequenciamento, um novo método surgiu com a possibilidade de leitura de uma molécula inteira de RNA sem a obrigatoriedade da fragmentação das sequências polipeptídicas para a leitura. Isso possibilitou a identificação de isoformas resultantes dos *splicing* alternativo além de reduzir os erros de montagem (STARK et al., 2019). Este método foi lançado em 2011 pela empresa Pacific Biosciences (PacBio). Nele, a polimerização é realizada em moléculas circulares, diferente dos *reads* fragmentados e fixados em plataforma sólida no Illumina. As moléculas circulares ficam dispostas em poços com uma única polimerase que realiza a amplificação e consecutiva leitura por fluoróforos (KUMAR et al., 2019).

Para os trabalhos de RNA-seq, a tecnologia *long-read* possibilita a identificação clara de RNAs e suas isoformas, possibilitando um maior entendimento

do perfil transcricional da planta uma vez que os erros relacionados a montagem das sequências são menos frequentes (DIJK et al., 2018). Além disso, a tecnologia *long-read* pode contribuir em estudos voltados à descoberta de genes relacionados à síntese de metabólitos especializados, como em espécies utilizadas na medicina chinesa (CUI et al., 2021).

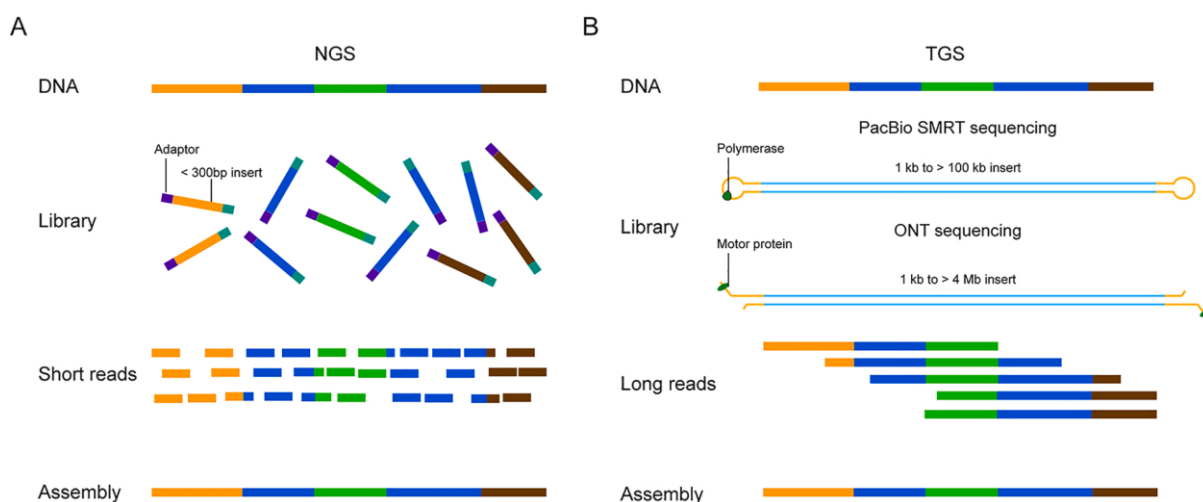


Figura 2: Métodos de leitura curta, equenciamento de alto rendimento(A) e leitura longa TGS (*Long reads*)(B). Destaque para menos sobreposições de reads durante a montagem utilizando *long reads* em comparação com *short reads*. Fonte: CHEN et al., (2021)

Assim, trabalhos que utilizam de métodos de RNA-seq combinados, Illumina e long-read, podem contribuir para a identificação de aspectos metabólicos e funcionais de compostos como os terpenos. Ou ainda possibilitar a construção de transcriptomas completos de forma rápida e econômica de espécies, sem um genoma de referência. Isso foi realizado na espécie de algodão selvagem *Gossypium australe* (FENG et al., 2019) e na espécie da medicina tradicional chinesa *Zanthoxylum armatum* (LIU et al., 2021).

1.4. Genes Normalizadores

Para a validação da expressão diferencial ou identificação de genes chaves para vias metabólicas de interesse, é preciso a identificação de genes considerados normalizadores, ou seja, genes que independente da condição experimental apresentam o mesmo padrão transcricional. Os experimentos de RT-qPCR dependem primariamente de uma extração de RNA, no entanto existe variação no conteúdo total de RNA presentes em uma amostra após a extração, assim o gene de

referência serviria como um calibrador e sua expressão será constante durante os ciclos de amplificação (ZHOU et al., 2017; KOZERA et al., 2013).

Para que o gene normalizador apresente essa característica durante as reações de RT-qPCR é preciso que ele seja previamente selecionado em estudos que caracterizam seu perfil transcricional e os dados obtidos sejam executados para identificar a estabilidade desse gene.

A estabilidade de um normalizador depende de algumas características, são elas; o gene deve ser expresso em diferentes tecidos; o gene precisa ser identificado em diferentes fases de crescimento do organismo de estudo e ser expresso em diferentes condições de tratamento (HELLEMANS et al., 2014). Assim para estudos comparativos, no qual mais de um tratamento é avaliado por RT-PCR quantitativo em tempo real (RT-qPCR), bem como em experimentos que avaliam a expressão de genes em condições diversas, o nível de mRNA do gene de referência é utilizado como calibrador para as condições da análise transcricional (JOSEPH et al., 2018).

Para tanto, é preciso identificar qual gene apresenta características específicas de um normalizador. Para espécies vegetais que não apresentam genes normalizadores conhecidos, é necessária uma avaliação da estabilidade de genes que possam ser utilizados como normalizadores, idealmente comparando tratamentos (JOSEPH et al., 2018). A identificação dos genes mais estáveis, no geral, ocorre por meio da comparação dos valores médios de CT exibidos ou variações desses dados. A plataforma RefFinder é um exemplo amplamente empregado em diversos trabalhos com normalizadores (AIMAS et al., 2018; LI et al., 2021; SARWAR et al., 2020; SHUKLA et al., 2019). Esta plataforma utiliza quatro algoritmos que individualmente indicam a partir dos valores de referência qual o gene mais estável para normalizadores, são elas BestKeeper, NormFinder, Genorm e o método comparativo delta-CT (VANDESOMPELE et al., 2002; ANDERSON et al. 2004; PFAFFL et al., 2004; SILVER et al., 2006).

2. Objetivo

O presente estudo tem como objetivo geral obter recursos genômicos e de transcriptoma da espécie *Quillaja brasiliensis*. Especificamente foram realizados:

- A montagem *de novo* e a anotação do genoma de cloroplasto da espécie;
- Uma análise comparativa do genoma de cloroplastos de espécies da família Quillajaceae;

- Sequenciamento do transcriptoma de folhas e culturas celulares de *Q. brasiliensis*;
- Identificação de genes normalizadores para folhas, caules e raízes de *Q. brasiliensis*.

Para apresentar o alcance desses objetivos, os conteúdos desta dissertação foram divididos em duas partes. No primeiro é apresentado o sequenciamento e análise do genoma plastidial de *Q. brasiliensis*. Na segunda parte apresentamos o desenvolvimento de um transcriptoma de referência para a espécie.

Considerações finais

O desenvolvimento das técnicas de RT-qPCR e sequenciamento genômico trouxeram muitos avanços para o entendimento da dinâmica metabólica dos organismos vegetais. Tais dinâmicas servem como mapa para o aperfeiçoamento de técnicas relacionadas a biologia molecular e biotecnologia. Além de servir como base nos estudos descritivos sobre evolução e fisiologia de plantas.

O presente trabalho buscou acrescentar a esses estudos subsídios genômicos e transcricionais da espécie *Quillaja brasiliensis*. Para isso executamos o sequenciamento, montagem e anotação do genoma plastidial desta espécie. Como resultado, pudemos realizar uma análise filogenética com outras espécies de Fabales bem como diferenciar caracteres evolutivos referente aos cloroplastos deste grupo.

Com os dados transcricionais sequenciamos e montamos o primeiro transcriptoma de referência para a o primeiro estudo sobre seleção de genes de referência. Estes genes podem servir como guia para estudos posteriores sobre o transcriptoma desta espécie em diversas condições experimentais.

Assim o presente trabalho permitiu contribui não só para biologia da espécie *Quillaja brasiliensis* e da família Quillajaceae, mas para estudos em Fabales e em organismos produtores de Saponinas.

Referências

ABBAS, Adeel et al. Selection and Validation of Reference Genes for RT-qPCR Analysis in *Aegilops tauschii* (Coss.) under Different Abiotic Stresses. **International journal of molecular sciences**, v. 22, n. 20, p. 11017, 2021.

ALMAS, D. Ebadi; KAMRODI, A. Rahmani. Validation of appropriate reference genes for real-time quantitative pcr gene expression analysis in rice plants exposed to metal stresses. **Russian Journal of Plant Physiology**, v. 65, n. 6, p. 890-897, 2018.

ANNADURAI, Ramasamy S. et al. De novo transcriptome assembly (NGS) of *Curcuma longa* L. rhizome reveals novel transcripts related to anticancer and antimalarial terpenoids. *PloS one*, v. 8, n. 2, p. e56217, 2013.

ANDERSEN, Claus Lindbjerg; JENSEN, Jens Ledet; ØRNTOFT, Torben Falck. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. **Cancer research**, v. 64, n. 15, p. 5245-5250, 2004.

BISWAS, Tanya; DWIVEDI, Upendra N. Plant triterpenoid saponins: biosynthesis, in vitro production, and pharmacological relevance. **Protoplasma**, v. 256, n. 6, p. 1463-1486, 2019.

BUSTIN, Stephen; NOLAN, Tania. Talking the talk, but not walking the walk: RT-qPCR as a paradigm for the lack of reproducibility in molecular research. **European Journal of Clinical Investigation**, v. 47, n. 10, p. 756-774, 2017.

BRAY, Nicolas L. et al. Near-optimal probabilistic RNA-seq quantification. **Nature biotechnology**, v. 34, n. 5, p. 525-527, 2016.

BUCCHINI, François et al. TRAPID 2.0: a web application for taxonomic and functional analysis of de novo transcriptomes. **Nucleic acids research**, v. 49, n. 17, p. e101-e101, 2021.

CIBULSKI, Samuel Paulo et al. Novel ISCOMs from *Quillaja brasiliensis* saponins induce mucosal and systemic antibody production, T-cell responses and improved antigen uptake. **Vaccine**, v. 34, n. 9, p. 1162-1171, 2016.

CHEN, Jingchao et al. Selection of relatively exact reference genes for gene expression studies in goosegrass (*Eleusine indica*) under herbicide stress. **Scientific reports**, v. 7, n. 1, p. 1-9, 2017.

CHIESA, Fernando Ferreira et al. *Quillaja brasiliensis* (A. St.-Hil. & Tul.) Mart. In: **Medicinal and Aromatic Plants of South America Vol. 2**. Springer, Cham, 2021. p. 447-459.

CHEN, Jiangfei et al. Genome-wide identification, classification and expression analysis of the HSP gene superfamily in tea plant (*Camellia sinensis*). **International Journal of Molecular Sciences**, v. 19, n. 9, p. 2633, 2018.

CHEN, Zhiao; HE, Xianghuo. Application of third-generation sequencing in cancer research. **Medical Review**, v. 1, n. 2, p. 150-171, 2021.

Codex Alimentarius International Food Standards; Food and Agriculture Organization of United Nations World Health Organization: Rome, Italy, 2018; Volume 49.

CUI, Yupeng et al. Full-length transcriptome analysis reveals candidate genes involved in terpenoid biosynthesis in *Artemisia argyi*. *Frontiers in genetics*, v. 12, 2021.

DA FONSECA, Rute R. et al. Next-generation biology: sequencing and data analysis approaches for non-model organisms. *Marine Genomics*, v. 30, p. 3-13, 201

DA SILVA MAGEDANS, Yve Veronica; PHILLIPS, Michael A.; FETT-NETO, Arthur Germano. Production of plant bioactive triterpenoid saponins: From metabolites to genes and back. **Phytochemistry Reviews**, v. 20, n. 2, p. 461-482, 2021.

DAI, Yu-Lin et al. Characterization and anti-tumor activity of saponin-rich fractions of South Korean sea cucumbers (*Apostichopus japonicus*). *Journal of Food Science and Technology*, p. 1-10, 2020.

DE FARIA, Janaína Teles et al. Performance of Quillaja bark saponin and β -lactoglobulin mixtures on emulsion formation and stability. **Food Hydrocolloids**, v. 67, p. 178-188, 2017

DE GROOT, Carolin; MÜLLER-GOYMANN, Christel C. Saponin interactions with model membrane systems—Langmuir monolayer studies, hemolysis and formation of ISCOMs. **Planta medica**, v. 82, n. 18, p. 1496-1512, 2016.

DAI, Yu-Lin et al. Characterization and anti-tumor activity of saponin-rich fractions of South Korean sea cucumbers (*Apostichopus japonicus*). *Journal of Food Science and Technology*, p. 1-10, 2020.

DONG, Wenpan et al. ycf1, the most promising plastid DNA barcode of land plants. **Scientific reports**, v. 5, n. 1, p. 1-5, 2015.

DUAN, Z. L. et al. Reference gene selections for real time quantitative PCR analysis of gene expression in different oat tissues and under salt stress. **Biol. Plant**, v. 64, p. 838-844, 2020.

EGAN, Ashley N.; SCHLUETER, Jessica; SPOONER, David M. Applications of next-generation sequencing in plant biology. *American journal of botany*, v. 99, n. 2, p. 175-185, 2012.

FLECK, Juliane Deise et al. Saponins from *Quillaja saponaria* and *Quillaja brasiliensis*: particular chemical characteristics and biological activities. **Molecules**, v. 24, n. 1, p. 171, 2019

FENG, Shouli et al. Reconstruction of the full-length transcriptome atlas using PacBio Iso-Seq provides insight into the alternative splicing in *Gossypium australe*. *BMC plant biology*, v. 19, n. 1, p. 1-16, 2019.

FIALLOS-JURADO, Jennifer et al. Saponin determination, expression analysis and functional characterization of saponin biosynthetic genes in *Chenopodium quinoa* leaves. *Plant Science*, v. 250, p. 188-197, 2016.

GEISLER, Markus; BAILLY, Aurélien; IVANCHENKO, Maria. Master and servant: Regulation of auxin transporters by FKBP and cyclophilins. ***Plant Science***, v. 245, p. 1-10, 2016.

GONZALEZ, Patricia Jurado; SÖRENSEN, Pia M. Characterization of saponin foam from *Saponaria officinalis* for food applications. *Food Hydrocolloids*, v. 101, p. 105541, 2020.

GUÉNIN, Stéphanie et al. Normalization of qRT-PCR data: the necessity of adopting a systematic, experimental conditions-specific, validation of references. *Journal of experimental botany*, v. 60, n. 2, p. 487-493, 2009.

GUZMAN, Frank et al. De novo assembly of *Eugenia uniflora* L. transcriptome and identification of genes from the terpenoid biosynthesis pathway. *Plant Science*, v. 229, p. 238-246, 2014.

HEALEY, Adam et al. Protocol: a simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species. *Plant methods*, v. 10, n. 1, p. 21, 2014.

HELLEMANS, Jan; VANDESOMPELE, Jo. Selection of reliable reference genes for RT-qPCR analysis. In: **Quantitative Real-Time PCR**. Humana Press, New York, NY, 2014. p. 19-26.

HUANG, Li-Min et al. Evolution of Terpene Synthases in Orchidaceae. *International journal of molecular sciences*, v. 22, n. 13, p. 6947, 2021.

JIN, Yuhuan et al. Identification of reliable reference genes for qRT-PCR in the ephemeral plant *Arabidopsis pumila* based on full-length transcriptome data. **Scientific reports**, v. 9, n. 1, p. 1-11, 2019.

JOSEPH, Joyous T.; POOLAKKALODY, Najya Jabeen; SHAH, Jasmine M. Plant reference genes for development and stress response studies. **Journal of biosciences**, v. 43, n. 1, p. 173-187, 2018.

JOZWIAK, Adam et al. Plant terpenoid metabolism co-opts a component of the cell wall biosynthesis machinery. *Nature Chemical Biology*, v. 16, n. 7, p. 740-748, 2020.

JUNG, Hyungtaek et al. Tools and strategies for long-read sequencing and de novo assembly of plant genomes. *Trends in plant science*, v. 24, n. 8, p. 700-724, 2019.

KARUNANITHI, Prema S.; ZERBE, Philipp. Terpene synthases as metabolic gatekeepers in the evolution of plant terpenoid chemical diversity. *Frontiers in Plant Science*, v. 10, p. 1166, 2019.

KIARASH, Jamshidi Goharrizi et al. Selection and validation of reference genes for normalization of qRT-PCR gene expression in wheat (*Triticum durum* L.) under drought and salt stresses. **Journal of Genetics**, v. 97, n. 5, p. 1433-1444, 2018.

KIRCHHOFF, Helmut. Chloroplast ultrastructure in plants. **New Phytologist**, v. 223, n. 2, p. 565-574, 2019.

KOZERA, Bartłomiej; RAPACZ, Marcin. Reference genes in real-time PCR. **Journal of applied genetics**, v. 54, n. 4, p. 391-406, 2013.

KUMAR, Arbind et al. Plant-derived immuno-adjuvants in vaccines formulation: a promising avenue for improving vaccines efficacy against SARS-CoV-2 virus. **Pharmacological Reports**, p. 1-17, 2022.

LIAO, Pan et al. The potential of the mevalonate pathway for enhanced isoprenoid production. **Biotechnology advances**, v. 34, n. 5, p. 697-713, 2016.

LI, Rong et al. Selection of the reference genes for quantitative gene expression by RT-qPCR in the desert plant *Stipagrostis pennata*. **Scientific reports**, v. 11, n. 1, p. 1-11, 2021.

LI, Yajuan et al. Systematic identification and validation of the reference genes from 60 RNA-Seq libraries in the scallop *Mizuhopecten yessoensis*. *BMC genomics*, v. 20, n. 1, p. 1-12, 2019.

Li, Yajuan, Lingling Zhang, Ruojiao Li, Meiwei Zhang, Yangping Li, Hao Wang, Shi Wang, and Zhenmin Bao. "Systematic Identification and Validation of the Reference Genes from 60 RNA-Seq Libraries in the Scallop *Mizuhopecten Yessoensis*." *BMC Genomics* 20 (April 11, 2019): 288.

LOGSDON, Glennis A.; VOLLGER, Mitchell R.; EICHLER, Evan E. Long-read human genome sequencing and its applications. *Nature Reviews Genetics*, p. 1-18, 2020.

LOVE, Michael I.; HUBER, Wolfgang; ANDERS, Simon. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, v. 15, n. 12, p. 1-21, 2014.

LUEBERT, Federico. Taxonomy and distribution of the genus *Quillaja* Molina (Quillajaceae). **Feddes Repertorium**, v. 124, n. 4, p. 157-162, 2013.

MA, Mengjie et al. Fabrication and characterization of zein/tea saponin composite nanoparticles as delivery vehicles of lutein. *LWT*, p. 109270, 2020.

MANNI, Mosè et al. BUSCO: assessing genomic data quality and beyond. **Current Protocols**, v. 1, n. 12, p. e323, 2021.

MANNI, Mosè et al. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. **Molecular Biology and Evolution**, v. 38, n. 10, p. 4647-4654, 2021.

MARCIANI, Dante J. Elucidating the mechanisms of action of saponin-derived adjuvants. **Trends in Pharmacological Sciences**, v. 39, n. 6, p. 573-585, 2018.

MARTICORENA, Clodomiro; MARCHESI, Eduardo. Catálogo de las plantas vasculares del Cono Sur (Argentina, Sur de Brasil, Chile, Paraguay y Uruguay).: Pteridophyta, Gymnospermae y Monocotyledoneae. **Missouri Botanical Garden Press, 2008.**

NAVARRO, E. et al. Real-time PCR detection chemistry. **Clinica chimica acta**, v. 439, p. 231-250, 2015.

NACHTIGALL, Pedro G.; KASHIWABARA, Andre Y.; DURHAM, Alan M. CodAn: predictive models for precise identification of coding regions in eukaryotic transcripts. **Briefings in bioinformatics**, v. 22, n. 3, p. bbaa045, 2021.

PRIYA, Piyush et al. Terzyme: a tool for identification and analysis of the plant terpenome. *Plant Methods*, v. 14, n. 1, p. 1-18, 2018.

PFAFFL, Michael W. et al. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper–Excel-based tool using pair-wise correlations. **Biotechnology letters**, v. 26, n. 6, p. 509-515, 2004.

RAMAKERS, Christian et al. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neuroscience letters*, v. 339, n. 1, p. 62-66, 2003.

RIBEIRO, Bianca et al. A Seed-Specific Regulator of Triterpene Saponin Biosynthesis in *Medicago truncatula*. *The Plant Cell*, v. 32, n. 6, 2020.

REICHERT, Corina L.; SALMINEN, Hanna; WEISS, Jochen. Quillaja saponin characteristics and functional properties. *Annual review of food science and technology*, v. 10, p. 43-73, 2019.

RUIJTER, J. M. et al. Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. **Nucleic acids research**, v. 37, n. 6, p. e45-e45, 2009.

SARWAR, Muhammad Bilal et al. Identification and validation of superior housekeeping gene (s) for qRT-PCR data normalization in *Agave sisalana* (a CAM-plant) under abiotic stresses. **Physiology and Molecular Biology of Plants**, v. 26, n. 3, p. 567-584, 2020.

SEKI, Hikaru; TAMURA, Keita; MURANAKA, Toshiya. P450s and UGTs: key players in the structural diversity of triterpenoid saponins. *Plant and Cell Physiology*, v. 56, n. 8, p. 1463-1471, 2015.

SHUKLA, Pawan et al. Selection of suitable reference genes for quantitative real-time PCR gene expression analysis in Mulberry (*Morus alba* L.) under different abiotic stresses. **Molecular biology reports**, v. 46, n. 2, p. 1809-1817, 2019.

SANITÁ LIMA, Matheus et al. The (in) complete organelle genome: exploring the use and nonuse of available technologies for characterizing mitochondrial and plastid chromosomes. *Molecular Ecology Resources*, v. 16, n. 6, p. 1279-1286, 2016.

SILVA, Flávia Loyze Baldassarini et al. Validation of reference genes for real-time quantitative PCR in *Brachiaria* grass under salt stress. **Plant Gene**, v. 27, p. 100319, 2021.

SILVER, Nicholas et al. Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. **BMC molecular biology**, v. 7, n. 1, p. 1-9, 2006.

SINGH, Harpreet et al. Plant cyclophilins: Multifaceted proteins with versatile roles. **Frontiers in Plant Science**, v. 11, p. 585212, 2020.

SINGH, Khushwant et al. Cyclophilins: less studied proteins with critical roles in pathogenesis. **Phytopathology**, v. 108, n. 1, p. 6-14, 2018.

SINHA, Ragini; SHARMA, T. R.; SINGH, Anil Kumar. Validation of reference genes for qRT-PCR data normalisation in lentil (*Lens culinaris*) under leaf developmental stages and abiotic stresses. **Physiology and Molecular Biology of Plants**, v. 25, n. 1, p. 123-134, 2019.

SLATKO, Barton E.; GARDNER, Andrew F.; AUSUBEL, Frederick M. Overview of next-generation sequencing technologies. **Current protocols in molecular biology**, v. 122, n. 1, p. e59, 2018.

SONI, Praveen et al. Reference gene identification for gene expression analysis in rice under different metal stress. **Journal of Biotechnology**, v. 332, p. 83-93, 2021.

STARK, Rory; GRZELAK, Marta; HADFIELD, James. RNA sequencing: the teenage years. *Nature Reviews Genetics*, v. 20, n. 11, p. 631-656, 2019.

TAYLOR, Sean et al. A practical approach to RT-qPCR—publishing data that conform to the MIQE guidelines. *Methods*, v. 50, n. 4, p. S1-S5, 2010

UL HAQ, Saeed et al. Heat shock proteins: dynamic biomolecules to counter plant biotic and abiotic stresses. **International journal of molecular sciences**, v. 20, n. 21, p. 5321, 2019.

UNTERGASSER, Andreas et al. Primer3—new capabilities and interfaces. **Nucleic acids research**, v. 40, n. 15, p. e115-e115, 2012.

VAN DIJK, Erwin L. et al. The third revolution in sequencing technology. *Trends in Genetics*, v. 34, n. 9, p. 666-681, 2018.

VANDESOMPELE, Jo et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. **Genome biology**, v. 3, n. 7, p. 1-12, 2002.

VETÖ, Nicole M. et al. Transcriptomics analysis of *Psidium cattleianum* Sabine (Myrtaceae) unveil potential genes involved in fruit pigmentation. *Genetics and molecular biology*, v. 43, 2020.

WANG, Lijuan; WANG, Yancai; ZHOU, Ping. Validation of reference genes for quantitative real-time PCR during Chinese wolfberry fruit development. **Plant physiology and biochemistry**, v. 70, p. 304-310, 2013.

WHITE, Elizabeth J. et al. Modified Cetyltrimethylammonium bromide method improves robustness and versatility: The benchmark for plant RNA extraction. Weinheim: WILEY-VCH Verlag, 2008.

WANG, Bo et al. Systematic selection and validation of suitable reference genes for quantitative real-time PCR normalization studies of gene expression in *Nitraria tangutorum*. **Scientific reports**, v. 10, n. 1, p. 1-10, 2020.

XIE, Fuliang et al. miRDeepFinder: a miRNA analysis tool for deep sequencing of plant small RNAs. *Plant molecular biology*, v. 80, n. 1, p. 75-84, 2012.

YAMAMURA, Yoshimi; KUROSAKI, Fumiya; LEE, Jung-Bum. Elucidation of terpenoid metabolism in *Scoparia dulcis* by RNA-seq analysis. *Scientific reports*, v. 7,

YENDO, A. C. A. et al. Irradiance-based treatments of *Quillaja brasiliensis* leaves (A. St.-Hil. & Tul.) Mart. as means to improve immunoadjuvant saponin yield. *Industrial Crops and Products*, v. 74, p. 228-233, 2015.

YE, Jing et al. Selection of suitable reference genes for qRT-PCR normalisation under different experimental conditions in *Eucommia ulmoides* Oliv. **Scientific reports**, v. 8, n. 1, p. 1-11, 2018.

ZHANG, Zhaoping et al. Selection of the reference gene for expression normalization in *Papaver somniferum* L. under abiotic stress and hormone treatment. **Genes**, v. 11, n. 2, p. 124, 2020.

ZHAO, Yiyong et al. Nuclear phylotranscriptomics and phylogenomics support numerous polyploidization events and hypotheses for the evolution of rhizobial nitrogen-fixing symbiosis in Fabaceae. **Molecular Plant**, v. 14, n. 5, p. 748-773, 2021

ZHOU, Zhe et al. Using RNA-seq data to select reference genes for normalizing gene expression in apple roots. PloS one, v. 12, n. 9, p. e0185288, 2017.

ZHU, Jia-Hong et al. Identification, characterization and expression analysis of genes involved in steroidal saponin biosynthesis in *Dracaena cambodiana*. Journal of plant research, v. 131, n. 3, p. 555-562, 2018.

ZUO, Yang et al. The complete chloroplast genome of *Polygala japonica* Houtt.(Polygalaceae), a medicinal plant in China. Mitochondrial DNA Part B, v. 6, n. 1, p. 239-240, 2021