



UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR

SABRINA DELGADO SOARES

**MONTAGEM E ANÁLISE DO TRANSCRITOMA DO MOGNO AFRICANO (*Khaya
grandifoliola* C. DC.)**

Goiânia - Goiás

2019



UFG

UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO (TECA) PARA DISPONIBILIZAR VERSÕES ELETRÔNICAS DE TESES

E DISSERTAÇÕES NA BIBLIOTECA DIGITAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a [Lei 9.610/98](#), o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou download, a título de divulgação da produção científica brasileira, a partir desta data.

O conteúdo das Teses e Dissertações disponibilizado na BDTD/UFG é de responsabilidade exclusiva do autor. Ao encaminhar o produto final, o autor(a) e o(a) orientador(a) firmam o compromisso de que o trabalho não contém nenhuma violação de quaisquer direitos autorais ou outro direito de terceiros.

1. Identificação do material bibliográfico

Dissertação Tese Outro*: _____

*No caso de mestrado/doutorado profissional, indique o formato do Trabalho de Conclusão de Curso, permitido no documento de área, correspondente ao programa de pós-graduação, orientado pela legislação vigente da CAPES.

Exemplos: Estudo de caso ou Revisão sistemática ou outros formatos.

2. Nome completo do autor

SABRINA DELGADO SOARES

3. Título do trabalho

MONTAGEM E ANÁLISE DO TRANSCRITOMA DO MOGNO AFRICANO (*Khaya grandifoliola* C. DC.)

4. Informações de acesso ao documento (este campo deve ser preenchido pelo orientador)

Concorda com a liberação total do documento SIM NÃO¹

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. Após esse período, a possível disponibilização ocorrerá apenas mediante:

- a)** consulta ao(à) autor(a) e ao(à) orientador(a);
- b)** novo Termo de Ciência e de Autorização (TECA) assinado e inserido no arquivo da tese ou dissertação.

O documento não será disponibilizado durante o período de embargo.

Casos de embargo:

- Solicitação de registro de patente;
- Submissão de artigo em revista científica;

- Publicação como capítulo de livro;
- Publicação da dissertação/tese em livro.

Obs. Este termo deverá ser assinado no SEI pelo orientador e pelo autor.



Documento assinado eletronicamente por **SABRINA DELGADO SOARES, Discente**, em 07/06/2023, às 17:35, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Evandro Novaes, Usuário Externo**, em 11/06/2023, às 08:08, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **3808170** e o código CRC **942150F9**.

Referência: Processo nº 23070.032519/2023-23

SEI nº 3808170

SABRINA DELGADO SOARES

MONTAGEM E ANÁLISE DO TRANSCRITOMA DO MOGNO AFRICANO (*Khaya grandifoliola* C. DC.)

Tese apresentada ao Programa de Pós-graduação em Genética e Biologia Molecular, da Universidade Federal de Goiás para obtenção do título de Doutora em Genética e Biologia Molecular.

Orientador: Prof. Dr. Evandro Novaes

Goiânia – Goiás

2019

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UFG.

Soares, Sabrina Delgado
MONTAGEM E ANÁLISE DO TRANSCRITOMA DO MOGNO
AFRICANO (*Khaya grandifoliola* C. DC.) [manuscrito] / Sabrina
Delgado Soares. - 2019.
50 f.: il.

Orientador: Prof. Dr. Evandro Novaes.
Tese (Doutorado) - Universidade Federal de Goiás, Instituto de
Ciências Biológicas (ICB), Programa de Pós-Graduação em Genética e
Biologia Molecular, Goiânia, 2019.
Bibliografia.

1. Montagem de novo. 2. RNA-seq. 3. Bioinformática. 4. Madeira
nobre. I. Novaes, Evandro, orient. II. Título.

CDU 575



SERVIÇO PÚBLICO FEDERAL
UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR

SESSÃO PÚBLICA DE DEFESA DE TESE Nº 020

TÍTULO DO TRABALHO: "MONTAGEM E ANÁLISE DO TRANSCRITOMA DO
MOGNO AFRICANO (*Khaya grandifoliola* C. DC.)".

AUTOR (A): Sabrina Delgado Soares

TESE Nº 020 DO PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR,
DEFENDIDA E APROVADA EM SESSÃO PÚBLICA, NA DATA DE 25/07/2019, ÀS 13hs30min, no
Auditório ICB I, UFG, CUJA BANCA EXAMINADORA ESTEVE CONSTITUÍDA DOS SEGUINTE
PROFESSORES:

 Prof. Dr. Evandro Novaes Universidade Federal de Lavras	
 Profa. Dra. Mariana Pires de Campos Telles Universidade Federal de Goiás	 Prof. Dr. Alexandre Siqueira Guedes Coelho Universidade Federal de Goiás
 Profa. Dra. Thannya Nascimento Soares Universidade Federal de Goiás	 Profa. Dra. Rosana Pereira Vianello EMBRAPA

HOUVE MUDANÇA NO TÍTULO? SIM () NÃO (X)

CASO HAJA MUDANÇA, ESCREVER ABAIXO O NOVO TÍTULO DA DISSERTAÇÃO:

.....

OBS: O título na dissertação impressa deve corresponder exatamente ao que se encontra na ata.

GOIÂNIA (GO) 25/07/19

SUMÁRIO

RESUMO	10
ABSTRACT	11
1 INTRODUÇÃO	12
2 MATERIAL E MÉTODOS	13
2.1 COLETA DO MATERIAL BIOLÓGICO, EXTRAÇÃO E QUANTIFICAÇÃO DO RNA.....	13
2.4 CONTROLE DE QUALIDADE DO SEQUENCIAMENTO	14
2.5 MONTAGEM <i>de novo</i> DO TRANSCRITOMA.....	14
2.6 ANOTAÇÃO FUNCIONAL.....	15
2.7 ANÁLISE DE EXPRESSÃO GÊNICA E ENRIQUECIMENTO FUNCIONAL	16
3 RESULTADOS	17
3.1 MONTAGEM E ANOTAÇÃO FUNCIONAL DO TRANSCRITOMA DE <i>K.</i> <i>grandifoliola</i>	17
3.2 EXPRESSÃO GÊNICA DIFERENCIAL E ENRIQUECIMENTO FUNCIONAL.....	23
3.3 ANÁLISE DE VIAS METABÓLICAS PARA <i>K. grandifoliola</i>	28
4 DISCUSSÃO	32
4.1 MONTAGEM E ANOTAÇÃO FUNCIONAL.....	32
4.2 EXPRESSÃO GÊNICA DIFERENCIAL E ENRIQUECIMENTO FUNCIONAL.....	34
5 CONCLUSÃO	44
6 REFERÊNCIAS	45

AGRADECIMENTOS

Ao programa de Pós-graduação em Genética e Biologia Molecular (PGBM), Instituto de Ciências Biológicas da UFG e todos os docentes do programa pela oportunidade de realizar o doutorado.

Ao Laboratório de Genética e Genômica de Plantas da Escola de Agronomia da Universidade Federal de Goiás (EA/UFG) e a equipe que o compõe. Em especial à Ludmila Bandeira que fez a extração do RNA enquanto eu estava participando do processo de seleção do PGBM.

À professora Mariana Pires de Campos Telles e ao professor Alexandre Siqueira Guedes Coelho pelas contribuições na etapa de qualificação.

À Universidade Federal de Goiás pela estrutura e apoio no desenvolvimento da pesquisa.

À Fundação de Amparo à Pesquisa do Estado de Goiás (FAPEG - Proc. 201201272750053) e ao viveiro Mudas Nobres pelo apoio financeiro.

À Professora Sybelle Barreira pela permissão de coleta do material vegetal em sua área experimental.

Agradeço a Deus e a Nossa Senhora por me dar a oportunidade durante estes anos de conviver com pessoas, desde a moça da limpeza, das secretárias quanto aos mestres, que de alguma forma contribuíram para mais uma etapa da minha formação acadêmica.

À minha mãe (Maria) e meu irmão (Bruno) que sempre me incentivaram nos estudos e sempre me deram forças para almejar esse sonho.

Ao meu esposo, Divino Henrique, pela compreensão e paciência nos períodos de ausência.

À minha filha Maria Clara que desde o início de sua vida (dentro do meu ventre) foi companheira da mamãe durante as viagens, análises, escrita e agora... a defesa.

Aos meus avós maternos (Maria e Geraldo), aos meus tios (Zeca, Antônio e José Oscar) e tias (Marizete e Eliana) que de alguma forma me ajudaram durante esses anos.

Ao meu orientador, professor Evandro Novaes, que com sua humildade e enorme coração sempre esteve pronto para orientar. Seja puxando a orelha ou reconhecendo o esforço de seus alunos (as). E, principalmente pela enorme compreensão e por sempre estar preocupado com a nossa saúde mental.

Às minhas amigas Amanda Cândido e Lucilla Ávila pelo companheirismo e apoio nesta jornada.

Aos meus amigos e amigas que a Pós-graduação me deu, especialmente aos “meus” Stela Ribeiro, Mariane Brom, Rhewter Nunes, Ivone de Bem e Lucas Vieira que me ajudaram e apoiaram com uma estadia, com uma conversa, com abraços e com os afazeres da Pós.

Aos meus colegas de trabalho que sempre estavam dispostos a me substituir quando era necessário me ausentar, especialmente à minha amiga Livia Vieira. E, aos meus alunos que sempre compreenderam a minha ausência.

RESUMO

SOARES, S. D. **Montagem e análise do transcrito do mogno africano (*Khaya grandifoliola* C. DC.)** 2019. Tese (Doutorado em Genética e Biologia Molecular) – Instituto de Ciências Biológicas, Universidade Federal de Goiás¹.

O *Khaya grandifoliola* C. DC. é uma espécie tropical com madeira de alto valor no comércio internacional. Apesar do interesse por reflorestadores e produtores de madeira, praticamente nada se sabe sobre os aspectos genéticos e genômicos da espécie. Com o desenvolvimento das plataformas de sequenciamento de nova geração (NGS), a caracterização de transcritos pode ser realizada em larga escala e com custo acessível. Nesse contexto, o objetivo do trabalho foi o de identificar os genes de *K. grandifoliola* por meio da técnica de RNA-Seq e identificar aqueles diferencialmente expressos entre os tecidos da folha e do xilema *K. grandifoliola*. O RNA total foi extraído de amostras de folhas e xilemas de quatro árvores plantadas em área experimental da UFG. As amostras de mRNA foram sequenciadas em plataforma Illumina HiSeq 2000, no modo “*paired-end*” 2×100 pb. O sequenciamento resultou em um total de 857 milhões de sequências que, após controle de qualidade, diminuiu para 808 milhões com um tamanho médio de 24 a 85 pb. O programa Trinity foi utilizado para montagem *de novo* do transcrito obtendo-se 116.289 transcritos para um total de 36.271 genes. Os transcritos variaram de 524 a 12.031 pb com média de 1.366 pb e N50 de 1.657 pb. A análise de expressão gênica, identificou a presença de 13.626 genes diferencialmente expressos (GDEs) (FDR <0,05) entre os dois tecidos. Desses, 5.438 foram mais ativados em folha e 8.188 em xilema. Uma análise de enriquecimento funcional de categorias do Gene Ontology entre os GDEs foi realizada pelo programa Blast2GO. Essa análise identificou 26 termos GO enriquecidos entre os GDEs mais ativos em xilema. Em folha foram encontrados 48 termos enriquecidos. Como esperado, as categorias identificadas como enriquecidas entre os GDE apontam, para processos biológicos e componente celular mais específicos de cada tecido. Por exemplo, entre os genes mais ativos no tecido foliar foram identificados termos associados a tilacoides (GO:0009579), plastídeos (GO:0009536) e fotossíntese (GO:0015979). Já para o xilema foram identificados termos relacionados ao metabolismo de carboidratos (GO: 0005975), parede celular (GO:0071554) e de proteínas do citoesqueleto (GO:0008092). Esse estudo apresenta, de forma inédita, as sequências dos genes do *K. grandifoliola*. A análise de expressão diferencial permitiu identificar quais genes são mais ativos em folha e xilema. Análises das vias metabólicas no iPath3 possibilitou identificar as rotas mais ativas em folhas e em xilema. Um Blast contra o proteoma de laranja (*Citrus sinensis*) revelou a presença de 95 GDEs anotados para as enzimas da via de biossíntese dos limonóides, um importante metabólito secundário com propriedades medicinais.

Palavras-chave: Montagem *de novo*; RNA-seq; Bioinformática; madeira nobre.

¹ Orientador: Evandro Novaes, PhD. Escola de Agronomia –UFG.

ABSTRACT

SOARES, S. D. **Assembly and analysis of the transcriptome of the African mahogany (*Khaya grandifoliola* C. DC).** 2019. Thesis (PhD in Genetics and Molecular Biology) - Institute of Biological Sciences, Federal University of Goiás.

Khaya grandifoliola C. DC. is a tropical species with high value wood in international markets. Despite the high interest for reforestation and timber production, virtually nothing is known about the genetic and genomic aspects of the species. With the development of next-generation sequencing platforms (NGS), the characterization of transcriptomes can be performed on a large scale and at affordable cost. In this context, the objective of the work was to discover *K. grandifoliola* genes by using RNA-Seq technique and to identify those differentially expressed between the leaf and the xylem *K. grandifoliola* tissues. Total RNA was extracted from leaf and xylem samples of four trees planted in the experimental area of the UFG. mRNA samples were sequenced on the Illumina HiSeq 2000 platform, in paired-end mode 2×100 bp. Sequencing resulted in a total of 857 million sequences which, after quality control, decreased to 808 million with an average size of 24 to 85 bp. Trinity program was used to reassemble the transcriptome, obtaining 116,289 transcripts for a total of 36,271 genes. Transcripts ranged from 524 to 12,031 bp with a mean of 1,366 bp and N50 of 1,657 bp. Analyses of gene expression identified 13,626 differentially expressed genes (DEGs) (FDR <0.05) between the two tissues. Of these, 5,438 had higher expression in leaves and 8,188 in xylem. A functional enrichment analysis of Gene Ontology categories among the DEGs was performed by the Blast2GO program. This analysis identified 26 GO terms enriched among the most active DEGs in xylem. Among genes up-regulated in leaves, 48 enriched terms were found. As expected, the categories identified as enriched between DGEs point to biological processes and cellular components specific of each tissue. For example, among the most active genes in leaf tissue, terms associated with thylakoids (GO:0009579), plastids (GO:0009536) and photosynthesis (GO:0015979) were identified. On the other hand, for the xylem, terms related to carbohydrate metabolism (GO:0005975), cell wall (GO:0071554) and cytoskeletal proteins (GO:0008092) were significantly enriched. To the best of our knowledge, this is the first report of large-scale gene sequences for *K. grandifoliola*. Differential expression analysis allowed us to identify which genes are most active in leaf and xylem. Analysis in iPath3 identified the metabolic routes activated in leaf and xylem tissues. A BlastX search against the orange proteome (*Citrus sinensis*) revealed the presence of 95 DEGs annotated for the enzymes of the limonoid, a secondary metabolite with medicinal properties.

Keywords: *de novo* assembly; RNA-seq; Bioinformatics; high-value wood.

1 INTRODUÇÃO

A *Khaya grandifoliola* C. DC. é uma espécie tropical com madeira de alto valor no comércio internacional. Esse valor se deve à excelente trabalhabilidade e durabilidade de sua madeira. Além disso, a madeira é muito apreciada pela sua coloração, com gradiente de tonalidades que vai do amarelo ao vermelho, bem como pelas suas propriedades físicas que permitem um polimento fino, sendo utilizada para uma gama de indústrias na fabricação de móveis, navios, carpintaria e pavimentos. Tais características do mogno africano são similares à espécie de mogno brasileiro *Swietenia macrophylla* King [1].

Assim, o mogno africano é um excelente substituto para a *S. macrophylla* que não pode sofrer cortes e ser comercializada devido a uma instrução normativa do IBAMA [2]. Essa proibição foi instaurada por causa de sua inclusão no catálogo de espécies com risco de extinção [3]. Além disso, no Brasil os plantios de mogno brasileiro são devastados pela mariposa *Hypsipyla grandella* Zeller (1948) (Lepidoptera: Pyralidae), tornando seu cultivo inviável. A *K. grandifoliola* é susceptível a outra espécie de mariposa, a *H. robusta*. No entanto, esta é uma espécie africana não encontrada no Brasil. Além disso, o mogno africano se adapta bem ao clima e ao solo do país e ainda é pouco atacado pela *H. grandella*, que prefere as Meliaceae nativas do Brasil [1, 4–8], o que o torna potencialmente interessante para reflorestadores e produtores de madeira.

Devido à alta exploração de sua madeira visando abastecer o comércio internacional e sua baixa regeneração nos locais afetados, a espécie foi incluída na lista vermelha de espécies ameaçadas da IUCN (União Internacional para a Conservação da Natureza e dos Recursos Naturais) [9].

Apesar do potencial econômico e importância ecológica da *K. grandifoliola*, praticamente nada se sabe sobre os aspectos genéticos e genômicos da espécie. Com o advento dos métodos de sequenciamento de nova geração (*Next Generation Sequencing*) (NGS), tornou a análise do perfil genético e funcional de qualquer organismo mais rápido e acessível [10–13]. O sequenciamento de transcritomas com plataformas NGS é denominado de RNA-seq. Esta técnica permite caracterizar o conjunto de transcritos, possibilitando a quantificação dos elementos funcionais do genoma em diferentes tecidos e condições. As aplicações da técnica são variadas, podendo ser útil para estudos que visam identificar marcadores moleculares como microssatélites e polimorfismos de nucleotídeos únicos (SNPs), descoberta de genes funcionais

e de isoformas (produto de *splicing* alternativo), e quantificar a expressão gênica em diferentes tecidos e condições fisiológicas de um determinado organismo [14–17].

Contudo, para acessar as informações biológicas contidas nas milhões de sequências geralmente obtidas são necessários métodos computacionais, como, por exemplo, ferramentas para montagem do transcrito [18]. Na literatura já existe um amplo arcabouço de ferramentas bioinformáticas desenvolvidas para dados obtidos em plataformas NGS [13].

Devido ao alto valor de sua madeira e à necessidade de conservar o vulnerável germoplasma de *K. grandifoliola*, é importante que ferramentas moleculares sejam criadas para catalisar o melhoramento e os estudos de genética de populações para conservação da espécie. Como virtualmente nada se sabe sobre os genes e o genoma dessa espécie, os objetivos desse trabalho foram: 1) sequenciar e montar o transcrito do mogno africano *Khaya grandifoliola*; 2) anotar funcionalmente os genes de *K. grandifoliola*; 3) avaliar a expressão gênica diferencial entre os tecidos da folha e do xilema da espécie.

2 MATERIAL E MÉTODOS

2.1 COLETA DO MATERIAL BIOLÓGICO, EXTRAÇÃO E QUANTIFICAÇÃO DO RNA

Foram coletados folhas e xilema de quatro árvores adultas de mogno-africano (*K. grandifoliola*), com idade de dois anos, plantadas em área experimental da Escola da Agronomia da Universidade Federal de Goiás. A coleta do xilema foi feita com raspagem do tronco após remoção da casca. Os tecidos foram congelados imediatamente em nitrogênio líquido e armazenados em ultra-freezer (-80°C) para preservação do RNA.

Para a extração do RNA, as amostras de folhas e xilema congeladas foram maceradas em cadinho com nitrogênio líquido. Em seguida, procedeu-se a extração de RNA de acordo com o protocolo de [19], que se baseia na precipitação das moléculas de RNA com LiCl.

A integridade do RNA extraído de cada uma das oito amostras (4 árvores x 2 tecidos) foi avaliada em gel de agarose 2% e a concentração foi aferida no instrumento Qubit Fluorometer (Life Technologies) com o kit Qubit RNA Assay (Life Technologies). As amostras de RNA das quatro plantas foram combinadas de maneira equimolar em dois pools, um com as amostras de folha e outro com as de xilema. Esses dois pools foram enviados à empresa BGI

Co. Ltd. (Hong Kong) em colunas estabilizadoras (GenTegra-RNA, GenTegra, Pleasanton, USA) para sequenciamento na plataforma HiSeq 2000 (Illumina). As amostras de RNA das folhas e xilemas foram preparadas com o kit TruSeq® Stranded mRNA Library Prep (Illumina) e sequenciadas em duas canaletas, uma para cada tecido. O sequenciamento foi feito nas duas pontas dos fragmentos (*paired-end*), obtendo-se 100 pares de bases (pb) de cada lado.

2.4 CONTROLE DE QUALIDADE DO SEQUENCIAMENTO

O sequenciamento gera arquivos grandes em formato fastq, os quais foram compactados para transferência. A qualidade das sequências obtidas foi avaliada com o programa FastQC [20].

No sequenciamento Illumina são utilizados adaptadores que são ligados nos flancos dos fragmentos sequenciados. Como esses adaptadores podem ser sequenciados, esses foram removidos das sequências com o programa Trimmomatic [21], utilizando a opção ILLUMINACLIP:adapters.fa:2:30:10. Além disso, o Trimmomatic também foi utilizado para proceder à limpeza das sequências com baixa qualidade. Para isso foram utilizadas as opções SLIDINGWINDOW:4:20 para remoção das sequências das pontas quando uma janela de quatro bases tinha média de qualidade inferior a 20 (*Phred Quality Score* $Q \geq 20$); LEADING:20 e TRAILING:20 para o corte de bases com baixa qualidade no início e final de cada sequência, respectivamente; HEADCROP:15 para o corte das quinze primeiras bases de cada sequência. Essa opção foi utilizada porque as primeiras bases possuíam conteúdo GC distorcido; MINLEN:24 para remoção de sequências que, após a filtragem, tinham tamanho inferior a 24pb. Após a remoção dos adaptadores e das sequências de baixa qualidade, as sequências foram avaliadas novamente no programa FastQC, confirmando a melhoria nas métricas de qualidade.

2.5 MONTAGEM *de novo* DO TRANSCRITOMA

Como não há sequências de referência para o mogno africano *K. grandifoliola*, a montagem (*assembly*) das sequências dos transcritos foi realizada *de novo* com o programa Trinity [22]. Este utiliza três ferramentas (*Inchworm*, *Chrysalis* e *Butterfly*), de forma

sequencial, para montagem de grande volume de sequências de RNA-seq com base em Grafos de Brujn [22]. A montagem dos transcritos foi realizada com a opção de reportar somente *contigs* com tamanho mínimo de 500 pb. Também foi informado o tipo de orientação das *reads* na biblioteca: *forward* e *reverse* (FR).

Possíveis contaminantes nas sequências montadas foram investigados através de uma busca, com a ferramenta BlastN, contra sequências do genoma de *Escherichia coli* e dos bancos de rRNA do SILVA [23], UniVec (sequências de oligonucleotídeos e vetores de bactéria, fagos, vírus, leveduras e sequências sintéticas), de mitocôndria (mtDNA) e cloroplasto (cpDNA) do *National Center for Biotechnology Information* (NCBI). Sequências similares, com um *e-value* de 10^{-5} , foram eliminadas da montagem final com o programa Pyfasta (<http://github.com/brentp/pyfasta/>).

2.6 ANOTAÇÃO FUNCIONAL

O programa Blastx foi utilizado para encontrar similaridade entre os transcritos montados de mogno africano (*K. grandifoliola*), traduzidos nas seis fases de leitura, contra o banco de dados de proteínas do sub-reino de Embriófitas no banco de dados nr (*non-redundant*) do NCBI. O Blastx foi utilizado com a opção “max_target_seqs” de 20 e *e-value* 10^{-5} .

Como o Trinity monta as diferentes isoformas de um mesmo loco (gene), primeiramente foi feita a seleção no *output* do Blastx (.txt) de uma única isoforma, representativa, por gene. Essa seleção de uma única isoforma por gene foi feita para evitar a super-representação de categorias funcionais de genes com muitas isoformas. A partir do identificador atribuído pelo Trinity a cada transcrito, foi possível investigar o número de isoformas por gene e o tamanho dos transcritos através de um script *Shell*. Na plataforma R, a seleção dos transcritos mais representativos ocorreu através do melhor hit (*Blast hit*) com base no menor valor do *e-value*. Em caso de empate, o segundo critério utilizado foi o *BitScore* do Blastx, seguido do maior comprimento (*length*). Assim, se houvesse empate nos resultados do Blastx para duas ou mais isoformas, a mais representativa seria aquela de maior tamanho.

Com as isoformas selecionadas, foi criado um novo arquivo fasta com estas sequências utilizando o PyFasta. Um novo Blastx foi realizado com este arquivo das isoformas “mais representativas” contra o banco de dados nr do NCBI. O arquivo xml resultante do Blast foi

utilizado como *input* para anotação com categorias do Gene Ontology (GO) pelo programa Blast2GO [24]. Com isso, os transcritos foram classificados com diferentes funções biológicas nos três domínios definidas pelo sistema de ontologias (“Cellular Component”, “Molecular Function” e “Biological Process”) do *Gene Ontology*.

Os transcritos de *K. grandifoliola* foram comparados também com os peptídeos de *Citrus*, *Theobroma*, *Populus* e *Vitis*, obtidos do Phytozome [24]. A similaridade foi identificada com a ferramenta BlastX, utilizando um limiar de E-value 10^{-5} e de 70% de identidade. O número de transcritos com homólogos nessas espécies foi visualizado através de um diagrama de Venn, por meio da plataforma online Venny 2.1 [25].

2.7 ANÁLISE DE EXPRESSÃO GÊNICA E ENRIQUECIMENTO FUNCIONAL

Com o pipeline do programa Trinity foi realizado uma avaliação dos níveis de expressão gênica em ambos tecidos. Com isso, foi possível identificar genes específicos ou preferencialmente expressos em cada tecido. A análise foi feita em nível de genes e não de isoformas.

Seguindo o pipeline do Trinity, descrito em <http://trinityrnaseq.github.io>, primeiro procedeu-se ao mapeamento e contagem das *reads* nos genes com o script `align_and_estimate_abundance.pl`. O mapeamento das *reads* foi realizado com o programa `bowtie2` [26] e a contagem com o RSEM [27]. A matriz de contagem do número de sequências alinhados em cada gene e em cada tecido (folha e xilema) foi então utilizada na análise de expressão diferencial. Como o conjunto de dados deste trabalho não possui réplicas biológicas para os tratamentos (pools de folhas e xilemas), o pacote `edgeR` [28] foi utilizado com a opção “*-dispersion 0.4*” como valor empírico para a dispersão aleatória dos dados de expressão. Segundo o manual do `edgeR`, esse valor é razoável quando utilizado para conjunto de dados advindo de experimento controlado. O limite de significância adotado foi um p-valor ajustado menor que 0,05 ($FDR < 0,05$) e uma diferença de expressão mínima de $4\times$ (ou $|\log_2(\text{folha/xilema})| > 2$).

O número de genes expressos de forma específica ou em ambos tecidos (folhas e xilema) foi visualizado através de um diagrama de Venn, construído com o Venny 2.1 [25]. A análise de enriquecimento funcional foi realizada com os GDEs divididos em uma lista com aqueles mais ativados no xilema e outra para os genes mais ativados na folha. Essa análise foi realizada

no programa Blast2GO por meio do Teste Exato de Fisher considerando o valor de FDR < 0,05, para identificar categorias funcionais enriquecidas entre os genes expressos em cada tecido.

2.8 IDENTIFICAÇÃO DE VIAS METABÓLICAS

A identificação das vias metabólicas para os genes de *K. grandifoliola* foi realizada com os identificadores do KEGG Orthology (Kos) [29]. Para isso, primeiramente foi feito o *download* das sequências e do arquivo de anotação das proteínas das espécies de *Citrus sinensis*, *Populus trichocarpa* e *Arabidopsis thaliana*. Com o programa Blastx foi feito o alinhamento do transcrito de mogno com as sequências de proteínas das espécies descritas acima.

Por meio do resultado do Blastx, foram obtidos os melhores alinhamentos (Best Hits) utilizando um limiar de identidade maior que 70% com e-value 10^{-5} . A partir desses melhores alinhamentos foram obtidos os identificadores das respectivas vias metabólicas (KO e/ou EC). Para visualização das vias metabólicas representadas, dentre os transcritos montados, foi utilizado a ferramenta iPath3 [30, 31].

3 RESULTADOS

3.1 MONTAGEM E ANOTAÇÃO FUNCIONAL DO TRANSCRITOMA DE *K. grandifoliola*

No total, foram geradas 228.537.681 sequências pareadas 2x100 pb a partir do RNA extraído das folhas de *K. grandifoliola*. Para a biblioteca de xilema produziu-se um total de 200.247.822 sequências neste mesmo formato.

Tanto a biblioteca de folha (Figura 5) quanto a de xilema apresentaram sequências com média de qualidade de *Phred* acima de 30 (coloração verde), o que demonstra que o sequenciamento teve uma ótima qualidade e que as amostras de RNA não foram degradadas durante o armazenamento e transporte. Os gráficos do FastQC mostram que as primeiras bases das sequências apresentaram uma pequena queda de qualidade e uma distorção no conteúdo

GC. Na etapa de limpeza (filtragem) foi feito a remoção dos adaptadores Illumina, bem como a remoção de sequências de baixa qualidade.

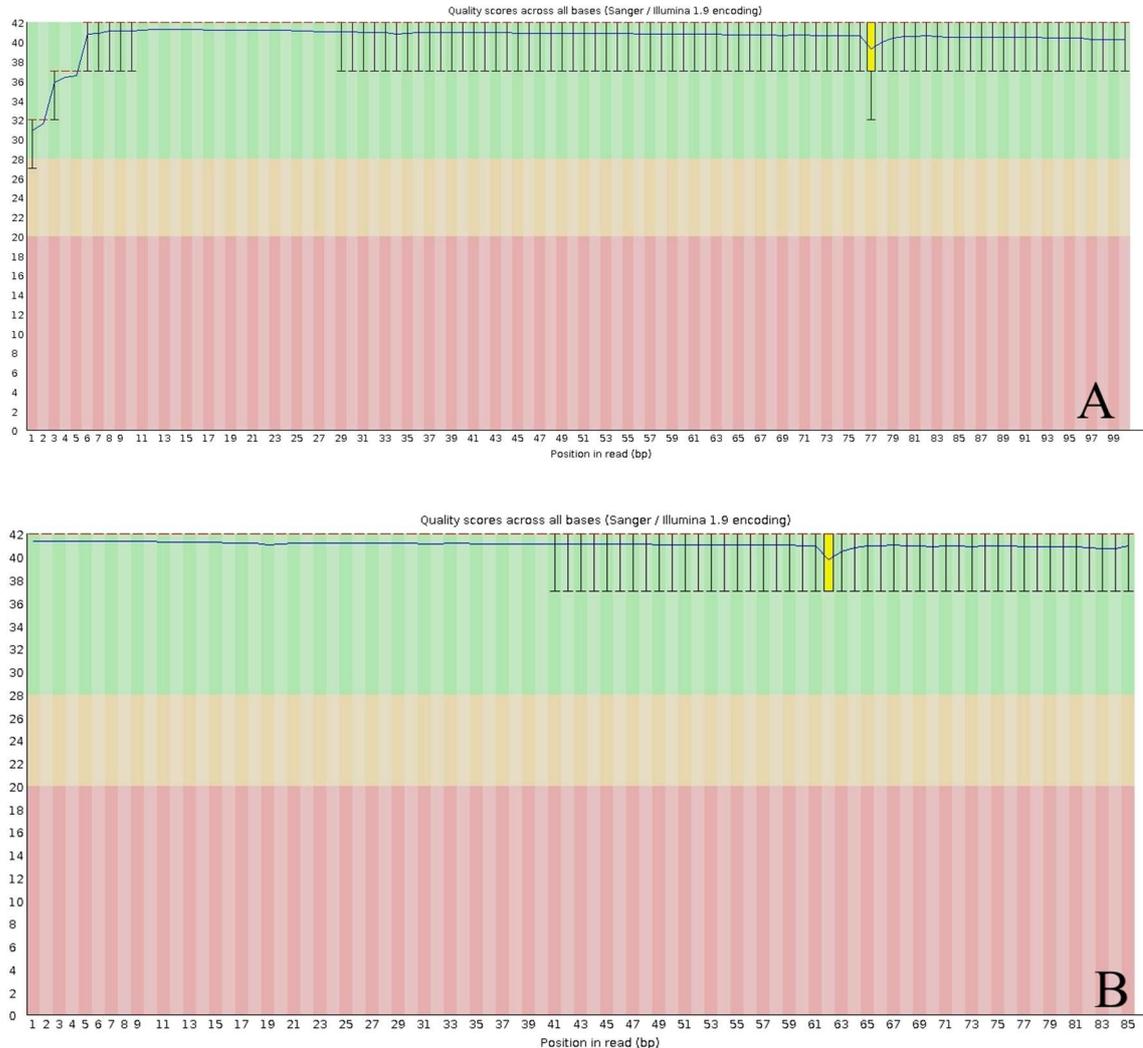


Figura 1. Resultado do programa FastQC para a biblioteca de folha. No eixo x estão a posição das sequências (pb) e no eixo y os valores de Phred. A: Primeira avaliação da qualidade das bases com detecção de ruído no início da sequência. B: Melhora da qualidade das sequências após o corte das 15 primeiras bases realizada com o programa Trimmomatic.

Após filtragem dos dados, a biblioteca de folha passou a ter 215.003.241 sequências pareadas, enquanto que a de xilema 189.247.018. Esta etapa resultou em sequências com tamanho que variaram de 24 a 85 pb para ambas as bibliotecas.

A montagem *de novo* com o Trinity resultou em 116.289 contigs ou transcritos, de 36.271 locos ou genes. Com isso, a montagem gerou em média 3,2 isoformas por gene. O

tamanho médio dos transcritos foi de 1366 pb e o N50 de 1657 pb (Tabela 1). Não foram montados contigs com tamanho inferior a 500 pb.

Tabela 1. Estatísticas da montagem *de novo* do transcrito de *K. grandifoliola* realizada no programa Trinity.

Parâmetros montagem	Valor
Número de sequências antes limpeza (folha e xilema)	428.785.503
Número de sequências após limpeza (folha e xilema)	404.250.259
Número de contigs (transcritos e isoformas)	116.289
Número de genes	36.271
Tamanho do menor contig	524
Tamanho do maior contig	12.031
% de contigs maiores que 1Kb	56,3
Número de contigs > 10 Kb	12
Média dos tamanhos dos contigs	1.366
Mediana dos tamanhos dos contigs	1.112
N50	1.657
Conteúdo GC (%)	41,58

A análise de similaridade das sequências de *K. grandifoliola* contra o banco de peptídeos de Embriófitas, realizada pelo Blastx, demonstrou que 88,14% dos 116,286 transcritos tiveram *hits* no Blastx. A porcentagem média de identidade do alinhamento das sequências de mogno contra as outras espécies foi relativamente alta (66,6%) (Tabela 2).

Tabela 2. Estimativas dos parâmetros de anotação no Blastx.

Parâmetros Blastx	Valor
Número de <i>hits</i>	102.503
Número de <i>no hits</i>	13.786
Comprimento médio dos <i>hits</i>	254.036
% de identidade média	66,62
Número de transcritos com identidade acima de 70%	20.141
Transcritos com 100% de identidade	326

Após a seleção dos *besthit* com porcentagem de identidade >70%, a análise pelo Blastx revelou que o transcrito do mogno africano possui maior similaridade com o gênero *Citrus*, seguido do *Theobroma*, *Populus* e *Vitis*. Cerca de 70% dos genes possuem *besthit* em laranja

(*Citrus*), seguido de Cacaú (*Theobroma*) com 20%. Foram identificados ainda 1.292, 469, 440 e 373 sequências com Blast hit somente em cada uma dessas espécies (*Citrus*, *Theobroma*, *Populus* e *Vitis*), respectivamente (Figura 2).

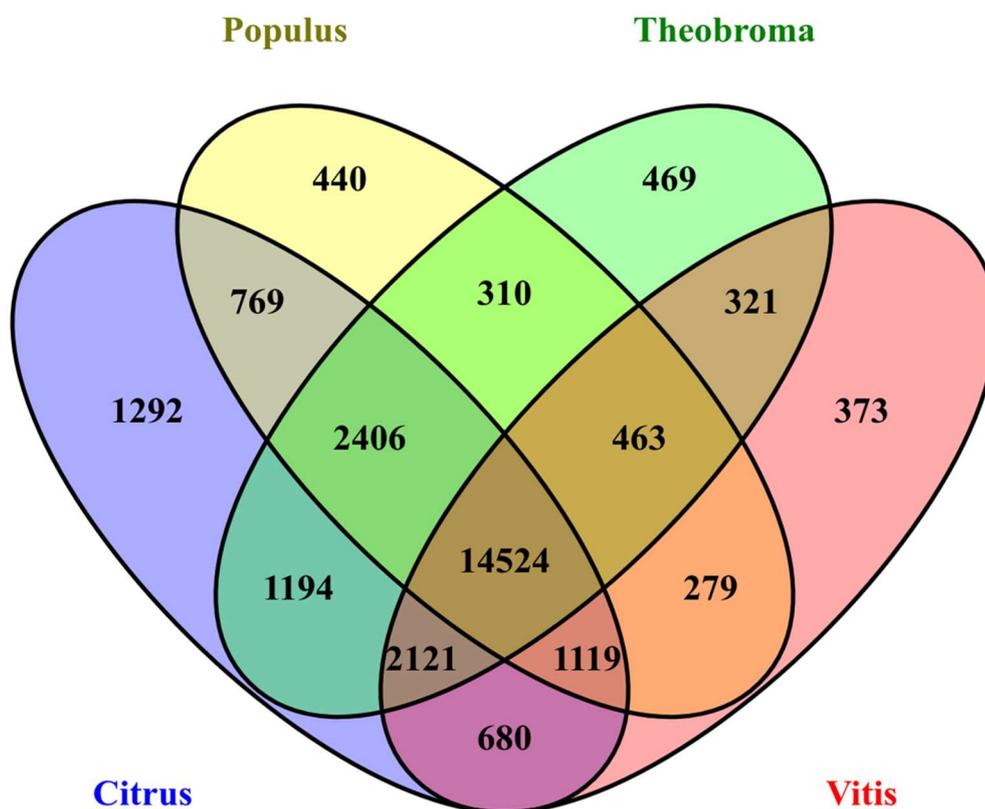


Figura 2. Diagrama de Venn demonstrando o número de transcritos de *K. grandifoliola* similares aos peptídeos de *Citrus*, *Theobroma*, *Populus* e *Vitis*. A similaridade foi identificada com a ferramenta BlastX (E-value < 10^{-5} e identidade > 70%).

Para anotação funcional dos genes preditos, foram selecionadas uma isoforma representativa para cada loco. Dos 36.271 genes, 27.575 obtiveram Blast hit e, portanto, tiveram uma isoforma selecionada. Destes, foram obtidas 83.477 categorias funcionais do Gene Ontology (GO) pelo Blast2Go. Do total, a maioria das anotações estão relacionadas com os domínios Processo Biológico (PB) com 37.936 anotações (45,4%), seguido de Função Molecular (FM) com 25.361 anotações (30,7%) e Componente Celular (CC) com 20.180 anotações (24,17%) (Figura 3).

Para o domínio CC, 7,5% das sequências foram anotadas para a região nuclear (GO:0005634), 7,4% como componente integrante da membrana (GO:0016021), 6,63%

membrana plasmática (GO:0005886), 5,44% para o citoplasma (GO:0005829), dentre outras categorias obtidas. A categoria mais representada para FM foi a de ligações a ATP (GO:0005524), com 9,02% das sequências anotadas (interação seletiva e não covalente com ATP) e 3,98% para ligação a íons de zinco (GO:0008270). Por fim, no domínio PB destaca-se o maior número de sequências anotadas (4,04%) para as reações de oxidação/redução (GO:0055114), 3,26% para processos metabólicos (GO:0008152), 2,25% para metabolismo de serina, dentre outras. A anotação funcional indica que, devido ao grande número de genes e transcritos obtidos, existe uma grande diversidade de processos biológicos e funções moleculares nas sequências montadas.

Como para *K. grandifoliola* não existem trabalhos prévios de descoberta de genes, as sequências aqui anotadas oferecem uma oportunidade importante para futuros estudos com a espécie.

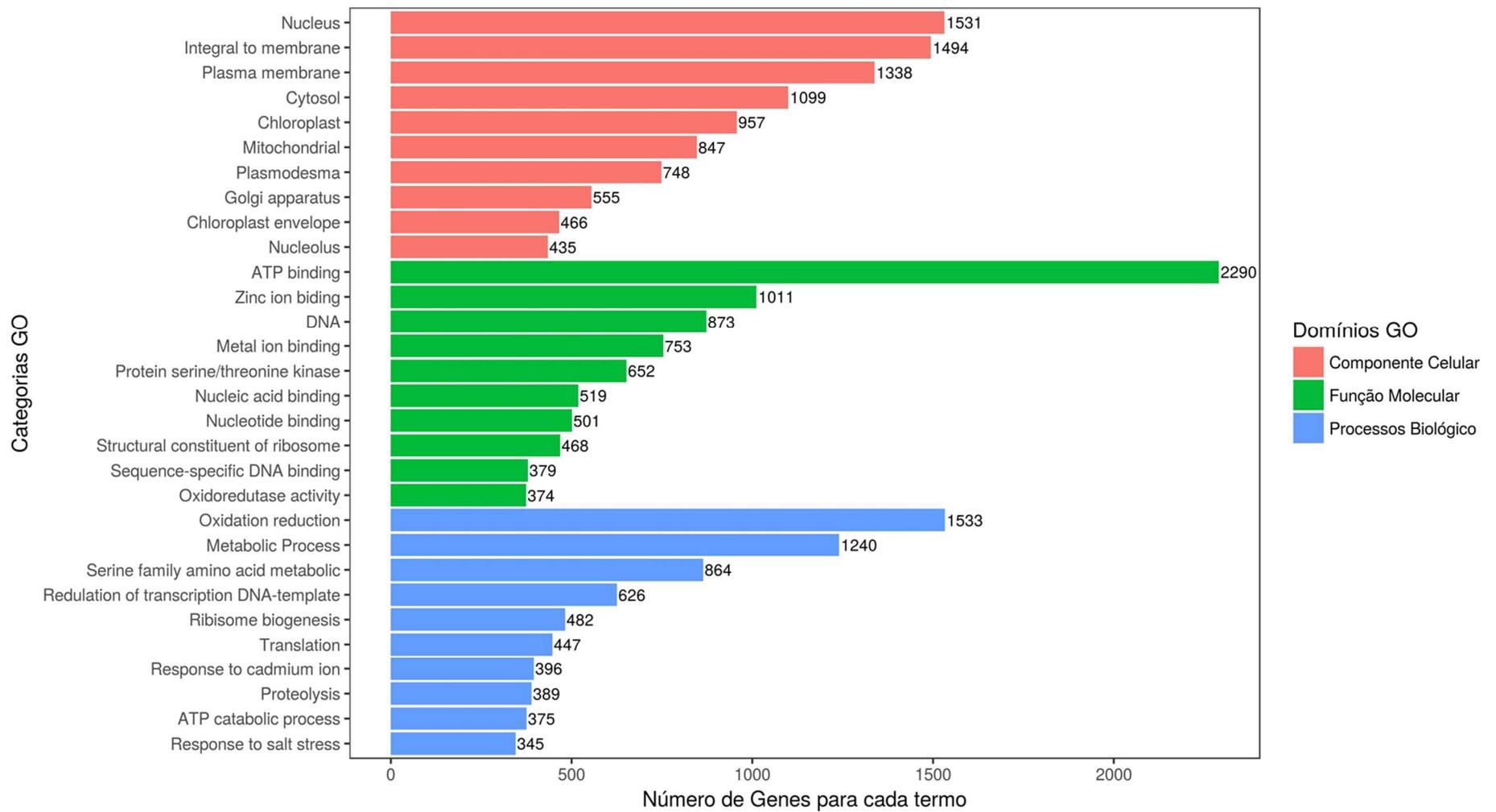


Figura 3. Dez primeiras categorias mais representadas, nas sequências de *K. grandifoliola* anotadas, para cada um dos três domínios funcionais do GO: componente celular (vermelho), função molecular (verde) e processos biológicos (azul).

3.2 EXPRESSÃO GÊNICA DIFERENCIAL E ENRIQUECIMENTO FUNCIONAL

A análise de expressão gênica identificou um total de 13.626 genes diferencialmente expressos (GDEs) entre os tecidos das folhas e do xilema de *K. grandifoliola* (FDR<0,05). Destes, 5.438 foram mais expressos em folhas e 8.188 em xilema.

No gráfico de dispersão da média de contagem (logCounts) versus log₂ do *fold change* (logFC) (Figura 4A), os pontos em vermelho demonstram os transcritos diferencialmente expressos entre os tecidos. No eixo y, os valores acima de zero (positivo) identificam transcritos mais expressos em xilema enquanto que aqueles abaixo de zero (negativo) os mais expressos nas folhas.

As duas linhas pontilhadas em diagonal, nos extremos da distribuição, representam os transcritos específicos de cada tecido. Ou seja, esses pontos representam os transcritos identificados somente em um e não no outro tecido. Dentre os GDEs, foram detectados 3.566 GDEs exclusivos de folhas e 6.129 exclusivos de xilema. Os GDEs que foram detectados como transcricionalmente ativos em ambos tecidos totalizaram 3.931 (Figura 4B).

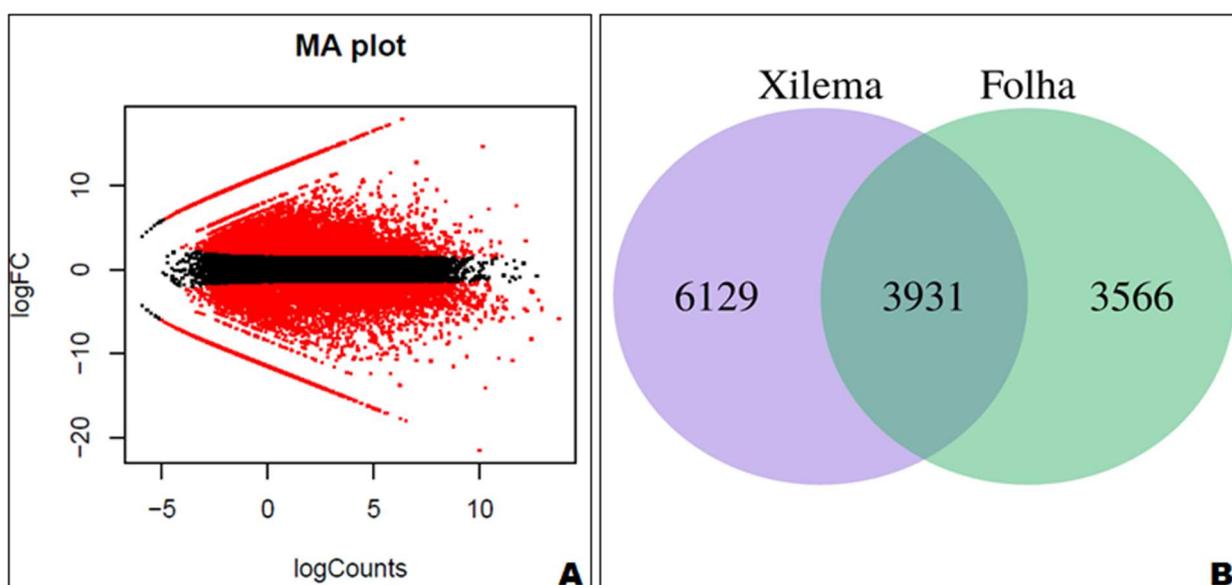


Figura 4. Resultado da análise de expressão gênica de mogno africano *K. grandifoliola*. **A:** Diagrama de dispersão da média de contagem (logCounts – eixo x) versus diferença de expressão entre os tecidos (log₂ do *fold change* ou logFC – eixo y). **B:** Diagrama de Venn representando o número de GDEs específicos de cada tecido e compartilhados entre ambos.

Um Blast contra o proteoma de laranja revelou a presença de 95 GDEs anotados para as enzimas da via de biossíntese dos limonóides (Tabela 3). Os limonóides são metabólitos secundários com propriedades medicinais contra doenças como a malária, atividade citotóxica contra células tumorais, atividade antifúngica, antimicrobiana e imunossupressora. Dentre esses 95 GDEs estão as enzimas limonene sintase, limonoide glicosiltransferase, oxidoesqualeno ciclase e a CYP714A1 corresponderam a um único gene no transcrito de mogno. Esses genes são o TR10120|c0_g1, TR2028|c0_g1, TR506|c0_g1 e TR15994|c0_g1, respectivamente. Para a enzima aciltransferase foram anotados oito genes, enquanto que para a enzima CYP90D1 foram anotados três genes, e por fim, para as UDPs-glicosiltransferases foram encontrados 81 genes de mogno. Dentre as UDPs-glicosiltransferases, quatro genes foram anotados para a enzima vomilenina glicosiltransferase, 34 genes para hidroquinona glicosiltransferase e 42 genes para poligalacturonato 4-alfa-galacturonosiltransferase.

Tabela 3. Enzimas da via biossintética dos limonóides e os genes correspondentes do transcrito da *K. grandifoliola*.

Citrus (gene)	Enzima	KO/EC	Gene <i>K. grandifoliola</i>
orange1.1g040856m, orange1.1g043404m	(4R)-limonene sintase	EC:4.2.3.20	TR10120 c0_g1
ciclev10010416m	Oxidoesqualene ciclase	K01853 EC:5.4.99.8	TR506 c0_g1
ciclev10031065m	CYP714A1	K10717	TR15994 c0_g1
ciclev10025448m	CYP90D1	K12638 EC:1.14.13.112	TR22323 c0_g1, TR4033 c0_g1, TR5956 c0_g1
ciclev10001521m	Aciltransferases	K13510 EC:2.3.1.23	TR20382 c1_g2, TR5906 c0_g1, TR9597 c2_g1, TR13146 c0_g1, TR16527 c0_g1, TR18300 c1_g1, TR543 c0_g1, TR543 c0_g2
orange1.1g032826m	UDP glicosiltransferases (limonoide glicosiltransferase)	EC:2.4.1.210	TR2028 c0_g1_i3
ciclev10020010m	UDP glicosiltransferases (hidroquinona glicosiltransferase)	K08237 EC:2.4.1.218	TR7889 c0_g2, TR12333 c0_g1, TR12870 c0_g1, TR13046 c0_g1, TR14581 c0_g1, TR16159 c0_g1, TR17138 c0_g1, TR17138 c0_g2, TR17138 c1_g1, TR1823 c0_g1, TR22480 c0_g1, TR22480 c1_g1, TR24873 c0_g1, TR5950 c0_g2, TR6700 c0_g1, TR6719 c0_g1, TR7889 c0_g2, TR6700 c0_g1, TR12333 c0_g1, TR12870 c0_g1, TR12870 c0_g2, TR14581 c0_g1, TR16159 c0_g1, TR17138 c0_g1, TR17138 c0_g2, TR17138 c1_g1, TR1823 c0_g1, TR22480 c0_g1, TR22480 c1_g1, TR24873 c0_g1, TR5950 c0_g2, TR6700 c0_g1, TR6719 c0_g1, TR7889 c0_g2
ciclev10015042m	UDP glicosiltransferases (vomilenina glicosiltransferase)	K08238 EC:2.4.1.219	TR12259 c1_g1, TR19143 c0_g1, TR25544 c0_g2, TR9407 c0_g1
ciclev10001658m	UDP glicosiltransferases (poligalacturonato 4-alpha- galacturonosiltransferase)	K13648 EC:2.4.1.43	TR12897 c0_g1, TR15183 c0_g1, TR15264 c0_g1, TR15264 c1_g1, TR15264 c1_g1, TR19272 c0_g2, TR20384 c0_g1, TR21617 c0_g1, TR21617 c0_g2, TR26217 c1_g1, TR7219 c0_g1, TR7219 c0_g2, TR9196 c2_g1, TR9311 c1_g1, TR12897 c0_g1, TR13392 c0_g1, TR13392 c0_g2, TR15183 c0_g1, TR15264 c0_g1, TR15264 c1_g1, TR19272 c0_g2, TR20384 c0_g1, TR21617 c0_g1, TR21617 c0_g2, TR21710 c0_g1, TR25984 c0_g1, TR26217 c1_g1, TR3173 c0_g1, TR467 c0_g1, TR5172 c0_g1, TR5527 c0_g1, TR5527 c1_g1, TR5527 c2_g1, TR5527 c3_g1, TR6545 c0_g1, TR7219 c0_g1, TR7219 c0_g2, TR9095 c0_g1, TR9098 c0_g1, TR9196 c0_g1, TR9196 c2_g1, TR9311 c1_g1

Para identificar as categorias funcionais super-representadas em cada tecido foi feita uma análise de enriquecimento funcional das categorias GO entre os GDEs (Tabela 4).

A análise de enriquecimento para GDEs ativados em folha encontrou 40 categorias super-representadas. Estas refletem a função das folhas no metabolismo energético e do nitrogênio, como por exemplo, categorias associadas ao endereçamento para os tilacoides (GO:0009579), plastídeos (GO:0009536) e parede celular (GO:0005618), geração de metabólitos precursores e energia (GO:0006091), processo metabólico do amido (GO:0043603), processo metabólico do ciclo do nitrogênio (GO:0071941), fotossíntese (GO:0015979).

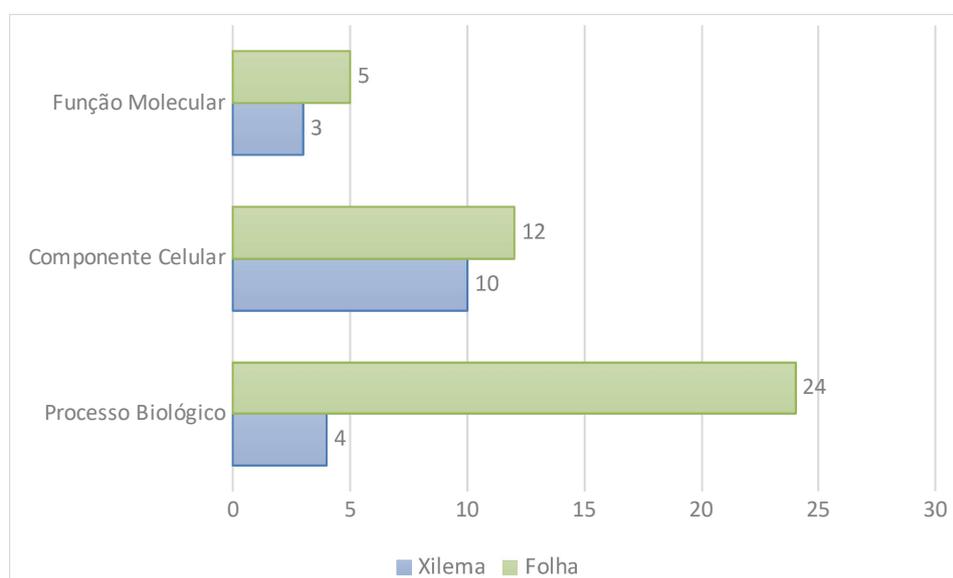


Figura 5. Distribuição do número de termos enriquecidos dentre os GDEs pela análise do Blast2GO. O número de categorias enriquecidas está subdividido em seus respectivos domínios GO (Função Molecular, Componente Celular e Processos Biológicos) e nos tecidos em que os genes analisados estão mais expressos (xilema em barras azul e folha em verde).

A análise identificou também 17 termos GO enriquecidos entre os GDEs mais ativos no xilema. Muitas dessas categorias, como esperado, estão associadas à função do tecido e a formação de parede celular secundária. Dentre as categorias enriquecidas em xilema encontram-se aquelas relacionadas ao metabolismo de carboidratos (GO:0005975), biogênese de macromoléculas da parede celular (GO:0071554) e de proteínas do citoesqueleto (GO:0008092) e processo metabólico secundário (GO:0019748).

Tabela 4. Categorias do Gene Ontology (GO) enriquecidas entre os GDEs mais ativos nos tecidos da folha e do xilema.

	FOLHA		XILEMA	
Função Molecular	Structural constituent of ribosome (GO:0003735)		Hydrolase activity, acting on glycosyl bonds (GO:0016798)	
	Structural molecule activity (GO:0005198)		Cytoskeletal protein binding (GO:0008092)	
	Transmembrane transporter activity (GO:0022857)		Oxidoreductase activity (GO:0016491)	
	Transporter activity (GO:0005215)			
Componente Celular	Thylakoid (GO:0009579)	Cytoplasmic part (GO:0044444)	Cell wall (GO:0005618)	Intracellular ribonucleoprotein complex (GO:0030529)
	Plastid (GO:0009536)	Cytosol (GO:0005829)	Extracellular region (GO:0005576)	Ribosome (GO:0005840)
	Ribonucleoprotein complex (GO:1990904)	Nucleolus (GO:0005730)	External encapsulating structure (GO:0030312)	Cell periphery (GO:0071944)
	Intracellular ribonucleoprotein complex (GO:0030529)	External encapsulating structure (GO:0030312)	Cytoskeleton (GO:0005856)	Non-membrane-bounded organelle (GO:0043228)
	Ribosome (GO:0005840)	Cell wall (GO:0005618)	Ribonucleoprotein complex (GO:1990904)	Intracellular non-membrane-bounded organelle (GO:0043232)
	Cytoplasm (GO:0005737)	Vacuole (GO:0005773)		
Processo Biológico	Photosynthesis (GO:0015979)	Cellular amide metabolic process (GO:0043603)	Cell wall organization or biogenesis (GO:0071554)	
	Generation of precursor metabolites and energy (GO:0006091)	Macromolecule biosynthetic process (GO:0009059)	Carbohydrate metabolic process (GO:0005975)	
	Ribosome biogenesis (GO:0042254)	Cellular nitrogen compound biosynthetic process (GO:0044271)	Secondary metabolic process (GO:0019748)	
	Ribonucleoprotein complex biogenesis (GO:0022613)	Transmembrane transport (GO:0055085)	Generation of precursor metabolites and energy (GO:0006091)	
	Peptide metabolic process (GO:0006518)	Cellular component biogenesis (GO:0044085)		
	Translation (GO:0006412)	Gene expression (GO:0010467)		
	Organonitrogen compound biosynthetic process (GO:1901566)	Localization (GO:0051179)		
	Cellular macromolecule biosynthetic process (GO:0034645)	Transport (GO:0006810)		
	Organic substance biosynthetic process (GO:1901576)	Establishment of localization (GO:0051234)		
	Cellular biosynthetic process (GO:0044249)	Response to stress (GO:0006950)		
	Peptide biosynthetic process (GO:0043043)	Nitrogen cycle metabolic process (GO:0071941)		
	Amide biosynthetic process (GO:0043604)	Cell death (GO:0008219)		

3.3 ANÁLISE DE VIAS METABÓLICAS PARA *K. grandifoliola*

A análise pelo iPath3 com os GDEs demonstrou que das 11 rotas metabólicas representadas no mapa, seis apresentaram maior expressão no tecido xilema, enquanto que folha apresentaram cinco rotas metabólicas com destaque (Figura 6).

Os níveis de expressão diferencial são identificados pela espessura das linhas. Quanto mais espessa, maior é a diferença de expressão entre os tecidos para aquela via, enquanto que as linhas mais finas apresentam menor nível de expressão diferencial, de acordo com a média do logFC dos genes que contém a categoria KO. Nas linhas destacadas, é possível identificar as enzimas que foram ativadas para cada tecido (Figura 7).

Os transcritos de *K. grandifoliola* tiveram enzimas diferencialmente expressos em praticamente todas as rotas metabólicas. Foram descritas aqui, apenas as que apresentaram linhas espessas identificadas no mapa (Figura 7).

A rota de metabolismo de terpenóides (7A) mostrou maior ativação para xilema, mais especificamente para a via de biossíntese de monoterpénóides. Para folha, houve destaque para a rota de metabolismo de nucleotídeos (7J) com ativação da via de metabolismo de purina. Na rota do metabolismo de lipídeos (7B e 7G), ambos evidenciaram a ativação para as vias de metabolismo de glicerolipídeos e glicerofosfolipídeos. No entanto, houve a ativação de vias diferentes relacionadas a esta rota. Em xilema, houve especificidade para a via de ácido alfa-linolênico. Enquanto que folha, a via de biossíntese de ácidos graxos teve genes ativados. Quanto a rota de metabolismo de aminoácidos (7C), houveram genes ativos em xilema nas vias do metabolismo de tirosina e fenilalanina. Além disso, também houve ativação na via de biossíntese de fenilalanina, tirosina e triptofano para ambos tecidos.

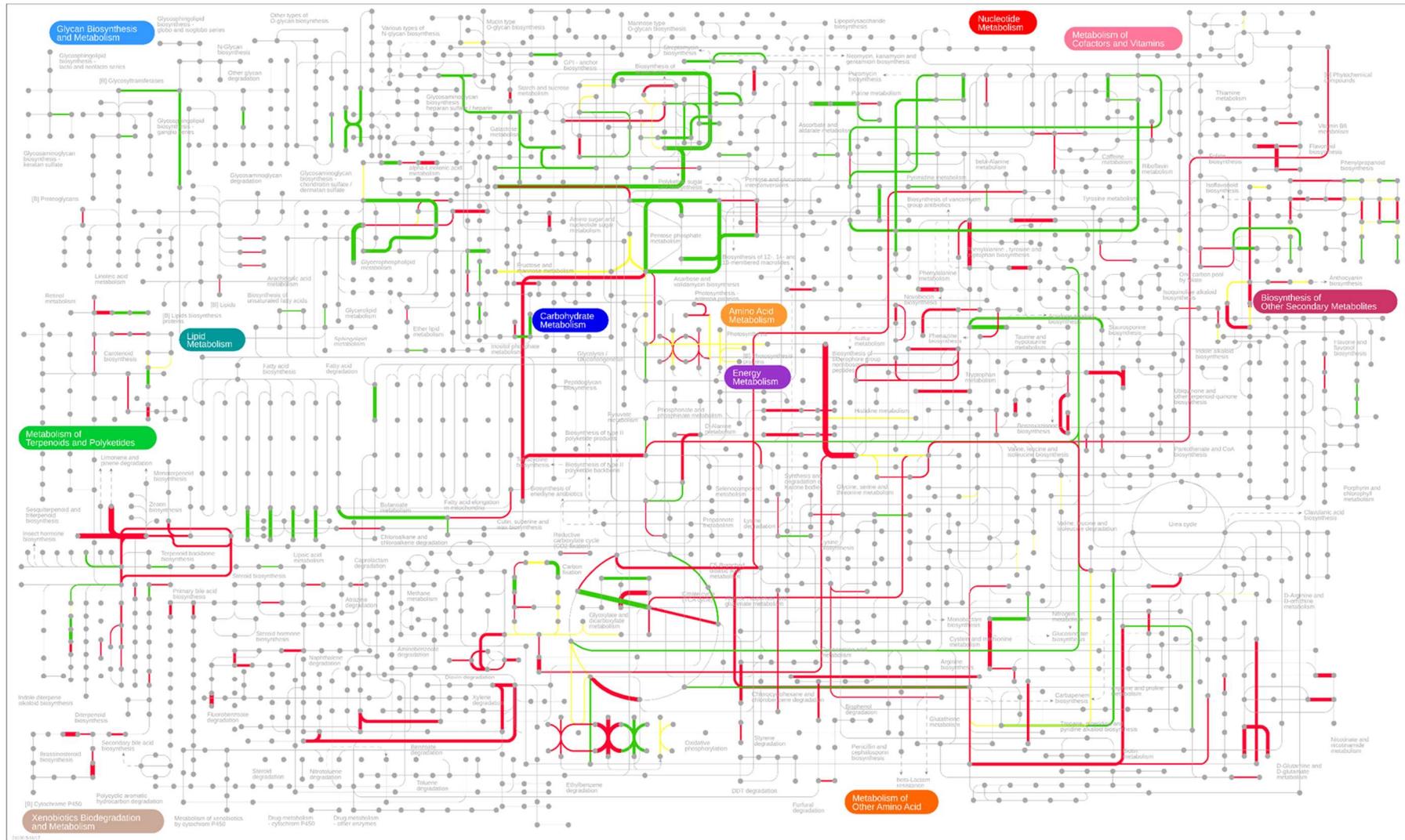


Figura 6. Visão geral das vias metabólicas com GDEs entre os tecidos de *K. grandifoliola*. As cores das linhas representam enzimas ativadas em folha (verde), xilema (vermelho) ou vias com a presença de enzimas ativadas em ambos (amarelo). A figura foi gerada pelo programa iPath3 [31].

Considerando a rota de carboidratos (7D e 7H), ambos tecidos tiveram destaque para ativação da via de fosforilação oxidativa, ciclo do ácido cítrico (Ciclo de Krebs) e metabolismo de piruvato. No entanto, xilema teve ativação da via de metabolismo de propionato e de butanoato. Em folha, foram ativadas outras vias, como a de metabolismo de galactose, amido e sacarose, glioxilato e dicarboxilato, ascorbato e aldarato e vias de pentose fosfato. A rota de energia, mostrou que o xilema teve ativação para a via de metabolismo de metano (7E) e folha, para a via de fixação de carbono (7I). Por fim, a rota de metabólitos secundários, ativou no tecido xilema a via de biossíntese de flavonoides (7F), enquanto folha ativou a via de biossíntese de fenazina (7K).

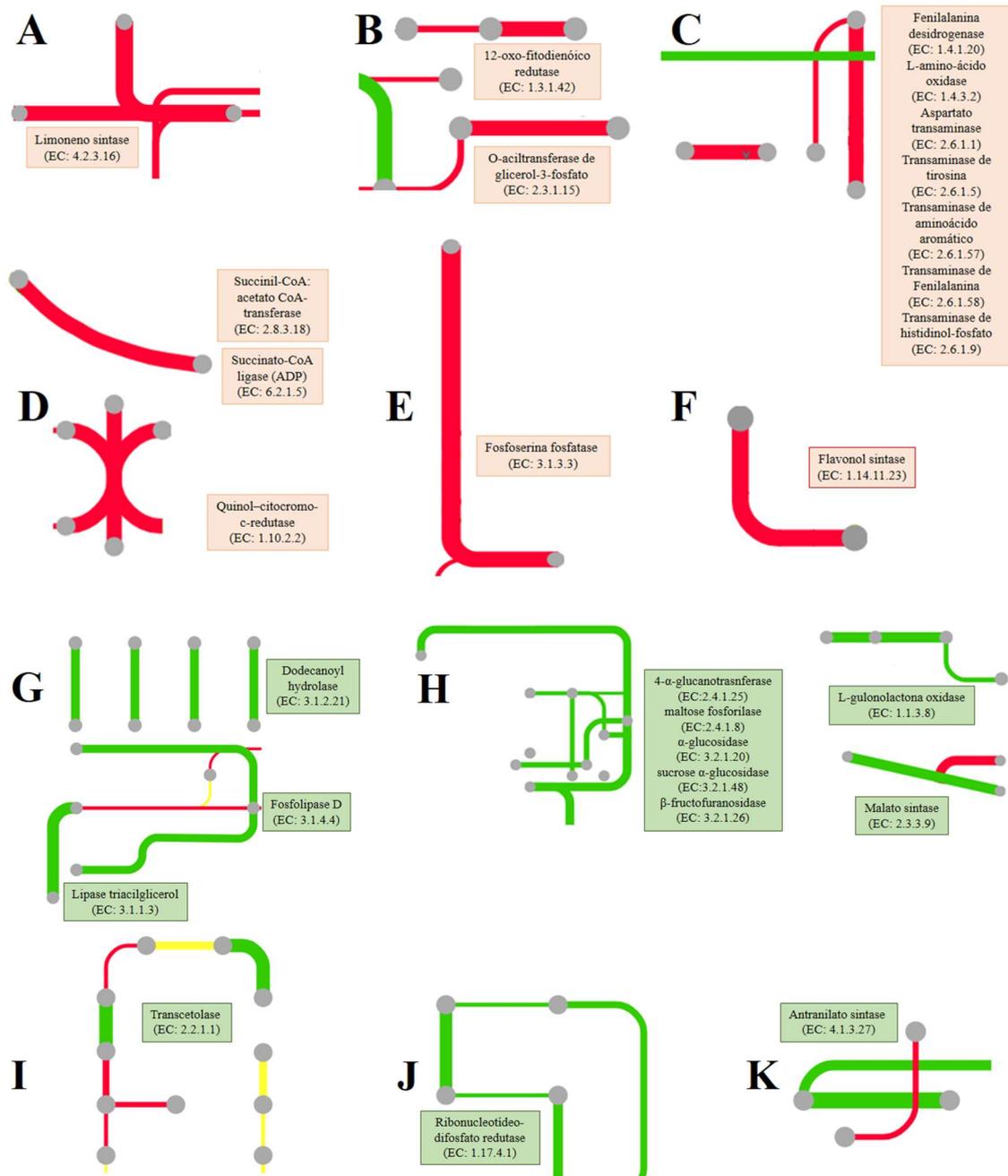


Figura 7. Representação das vias metabólicas mais ativadas em xilema (vermelho) e folha (verde) com seus respectivos ECs. (A) metabolismo de terpenoides, (B e G) metabolismo de lipídeos, (C) metabolismo de aminoácidos, (D e H) metabolismo de carboidratos, (E e I) metabolismo de energia, e (F e K) metabolismo de metabólitos secundários, (J) metabolismo de nucleotídeos.

4 DISCUSSÃO

4.1 MONTAGEM E ANOTAÇÃO FUNCIONAL

Em *K. grandifoliola* o número de genes foi maior (36.271 genes) que o encontrado para uma outra espécie da mesma família, a *Azadirachta indica* (20 mil genes) [32]. Cabe destacar que esse estudo diferiu do nosso, pois tratou-se da montagem e anotação do genoma de *Azadirachta indica* e não somente do transcrito. Por outro lado, um resultado próximo ao do presente estudo foi encontrado para o transcrito de *Cedrela balansae*, em que a montagem *de novo* apresentou um total de 149.572 contigs, com tamanho médio de 976 pb, com valor de N50 igual a 1.033 [33].

Os transcritos de mogno-africano, montados nesse estudo, variaram de 524 a 12.031 pb, com média de 1.366 pb e N50 de 1.657 pb. Os contigs menores do que 500 pb foram desconsiderados a fim de evitar RNAs parcialmente montados ou RNAs não codificantes (ncRNAs) que geralmente possuem tamanhos menores que este [34]. O presente estudo tem como foco analisar no transcrito apenas os genes codantes do mogno africano.

Apesar das sequências curtas geradas pelo sequenciamento Illumina, quando comparadas com outras metodologias, estas podem ser montadas em grandes *contigs*, gerando informações importantes para o descobrimento de genes em organismos não-modelos [35–37]. Além disso, o transcrito de espécies cujos recursos genômicos são escassos, como é o caso da *K. grandifoliola*, pode ser utilizado não só para a descoberta de genes, mas também para a identificação de marcadores microssatélites (SSR) ou de polimorfismo de nucleotídeo único (SNP) [38].

O resultado do Blastx demonstrou que as sequências de mogno africano são semelhantes às de *Citrus*. Esse resultado já era esperado, uma vez que essas espécies estão inseridas na ordem Sapindales que incluem as famílias Rutaceae (onde o gênero *Citrus* está inserido) e a Meliaceae (do gênero *Khaya*). Em estudo realizado com *nim* [32], também da família Meliaceae, a análise filogenética também verificou a proximidade taxonômica da espécie com *Citrus* spp. Assim, os resultados de *K. grandifoliola* corroboram aqueles obtidos com o transcrito de *nim*.

O baixo número de Blast hits contra espécies do gênero *Khaya* se dá pelo fato do banco de dados *nr* utilizado ainda não ter sequências depositadas desse gênero. Mesmo para a família

dos mognos, Meliaceae, a quantidade de sequências no banco *nr* ainda é muito pequena: apenas sete atualmente. Apesar disso, a porcentagem média de identidade (~66%) possibilitou a realização da anotação putativa dos transcritos. Quanto maior a porcentagem de identidade, maior a probabilidade das proteínas terem a mesma estrutura e função nas espécies analisadas.

A proximidade filogenética entre o nim e os mognos africanos explica o fato dessas espécies armazenarem metabólitos secundários bioativos, como os limonoides que tem amplo valor econômico nas indústrias da agricultura e saúde. O mogno africano *K. senegalensis* também apresenta importância na área da saúde, sendo utilizado pela população africana e cultivada na China para fins medicinais. Seu extrato possui funções citotóxicas, atividades antimaláricas e anti-inflamatórias derivados dos limonóides melíacos (metabólitos da família Meliaceae) [32, 39]. Além disso, esses metabólitos são considerados os fitoquímicos mais importantes de sua classe, estando também presentes em outras famílias, como Cneoraceae e Simaroubaceae [40].

Na etapa de anotação funcional a seleção das isoformas representativas para cada gene foi realizada para evitar estimativas viesadas das proporções das diferentes categorias funcionais. Genes com grande número de isoformas teriam maior influência na estimativa dessas proporções, caso essa seleção de uma isoforma por gene não fosse realizada. Os resultados da anotação se tornam importantes do ponto de vista biológico, pois permitem inferir a função biológica putativa da proteína codificada por um determinado gene no organismo estudado [41]. As sequências de *K. grandifoliola* anotadas com domínios do Gene Ontology (GO) corroboram com a literatura.

Dentro do domínio componente celular (CC), que representa onde o produto gênico é ativado, a espécie *K. grandifoliola* teve sequências anotadas como endereçadas para a região nuclear, componentes da membrana plasmática e citoplasma. A anotação em um estudo com a espécie *Magnolia champaca* (champak), árvore conhecida pela utilização de sua madeira e flores para a indústria alimentar e de fragrâncias, também identificou genes anotados para componente integral da membrana (51,6%), membrana plasmática (50%) e citoplasma (25,6%) [42]. Em *Pinus monticola*, 16,9% das anotações estavam relacionadas a membrana plasmática [43]. Para *A. indica* e *Melia azedarach* foram anotados genes relacionados às partes celulares e organelas [44].

Para o domínio função molecular (FM), que representa a atividade bioquímica do produto gênico, as categorias representativas em *K. grandifoliola* foram similares com os resultados de anotação encontrados para o transcrito da casca da espécie arbórea *Hevea brasiliensis* (seringueira), como por exemplo, as categorias de ligação (0,46%) e atividade

catalítica (0,39%) [36]. Em *Lonicera japônica*, uma planta medicinal chinesa, foram avaliados nove tecidos da planta [45], com anotação de Função Molecular para as categorias ligação (10,9%), atividade catalítica (9,3%) e ligação DNA (4,9%). Já para *M. champaca*, [42] identificaram as categorias de ligação (32,5%), ATP (50,5%), íons metálicos (23,7%), DNA (20,2%) e zinco (17,8%). Em *P. monticola* também foram detectadas categorias relacionadas à ligação (42,6%), atividade catalítica (41,7%) e atividade de transporte (5,5%) [43]. Em *A. indica* e *M. azedarach* [44] os genes estavam associados à atividade e ligação catalítica. Esses trabalhos também apresentaram a maioria das sequências anotadas para ligação a ATP, íons metálicos, DNA e a íons de zinco.

O domínio processo biológico (PB) define a atividade biológica em que o produto gênico está contribuindo para o funcionamento do organismo como um todo (células, tecidos e órgãos). Para outras espécies, como *H. brasiliensis* [36], *L. japonica* (RAI et al., 2017), *P. monticola* [43], espécies da família Meliaceae *A. indica* e *M. azedarach* [44], também se encontrou a maioria dos genes como envolvidos com a categoria de processos metabólicos.

Dessa maneira, a anotação funcional com as categorias GO permite conhecer, pelo menos de forma putativa, as vias metabólicas e os comportamentos biológicos dos genes identificados no organismo em estudo [36].

4.2 EXPRESSÃO GÊNICA DIFERENCIAL E ENRIQUECIMENTO FUNCIONAL

A análise do transcrito por meio da expressão gênica diferencial permite mensurar precisamente, ainda que de forma relativa, a quantidade de transcritos presentes em um determinado tecido ou organismo de acordo com o estágio específico do desenvolvimento ou da condição fisiológica em que estes foram submetidos (WANG et al., 2009). Em *K. grandifoliola*, foram identificados 3.566 genes diferencialmente expressos (GDEs) exclusivo do tecido foliar. Já em xilema, foram 6.129 GDE identificados de forma também exclusiva. Uma maior diversidade de transcritos em xilema já foi identificada em outros estudos, como em Eucalipto [46] e *Populus* [47]. Este mesmo resultado pode ser observado em outro estudo com a espécie *Eucalyptus grandis* [48] em que foram identificados GDEs anotados em maior quantidade para xilema (1.897) do que para folha (1.531).

Os órgãos fotossintetizantes da planta estão relacionados principalmente com a produção de ATP por meio da energia solar, produzindo carboidratos pela redução do carbono envolvido no processo. Dessa forma, estes carboidratos são utilizados pelos órgãos não fotossintetizantes para que através da respiração celular seja liberada a energia para o metabolismo da planta e, conseqüentemente, resultando em precursores que participarão da biossíntese de biomoléculas [46].

O xilema é responsável por diversas funções, dentre elas armazenamento de nutrientes, distribuição de água, minerais, aminoácidos, açúcares e o suporte físico para que a copa atinja enormes alturas para a planta captar luz [49–51]. O transporte dessas substâncias pelo xilema, demonstra a importância desse tecido no sistema de comunicação de longa distância para o desenvolvimento e manutenção dos processos fisiológicos e de defesa da planta. Esse fornecimento de informações, incluindo os hormônios peptídicos, proteínas e RNA permite que o sistema vascular se envolva efetivamente na coordenação destes processos da planta como um todo [51]. Isso sugere que moléculas como RNAs podem estar presentes no xilema, mas não necessariamente são específicos do próprio tecido, conseqüentemente causando diferença na quantidade de genes expressos entre os tecidos avaliados (folha e xilema).

Para os genes de *K. grandifoliola* diferencialmente expressos e anotados para a categoria parede celular, 21 apresentaram maior ativação em xilema do que em folha. Dentre essas categorias, cinco apresentaram uma indução maior considerando o valor do *foldchange* bem alto em relação aos demais, sendo eles, o TR12084|c0_g1 (*fold change* de 11,07), TR20307|c0_g1 (*fold change* de 11,22), TR21864|c0_g1 (*fold change* de 11,64), TR22442|c0_g1 (*fold change* de 11,01), TR24975|c0_g1 (*fold change* de 15,14). As categorias enriquecidas para xilema em *K. grandifoliola* estão intimamente relacionadas com a formação de parede celular secundária, assim como em outros estudos que realizaram a análise de enriquecimento para o mesmo tecido. As categorias GO representadas estão relacionadas ao desenvolvimento de biomassa lenhosa, como por exemplo, a biogênese e organização da parede celular (GO:0071554), cujos genes estão envolvidos na biossíntese de macromoléculas para construção da parede celular, parede celular (GO:0005618), e ligação de proteínas do citoesqueleto (GO:0008092) para entrega de biopolímeros à parede.

Em xilema, o perfil transcricional está especificamente direcionado à formação de madeira, a qual é basicamente composta de parede celular secundária [52, 53]. A identificação de GDEs entre tecidos é importante, pois indicam a relação dos transcritos com os eventos moleculares que levam ao desenvolvimento de um importante tecido ou órgão em estudo. Em

trabalho realizado com algodão, por exemplo, foram identificados GDEs relacionados ao desenvolvimento das fibras que são utilizadas na indústria têxtil [36].

Em *P. trichocarpa* (álamo), a análise de enriquecimento dos GDEs em xilema também identificou termos relacionados ao processo metabólico secundário, como biossíntese de lignina e morte celular [53]. Essas categorias estão associadas ao desenvolvimento da parede secundária e aos estágios terminais do desenvolvimento do tecido que resultam na morte celular programada, componentes importantes para o desenvolvimento da biomassa [54, 55].

Categorias relacionadas a biogênese da parede celular, organização de proteínas do citoesqueleto e processos metabólicos estão relacionados ao crescimento da planta [17]. As fibras da madeira são constituídas principalmente por parede celular secundária, contendo celulose, hemiceluloses e ligninas. Estas, em conjunto, determinam as propriedades da madeira [56]. A formação do xilema ocorre por diferentes fases de diferenciação, como as sucessivas divisões celulares, a expansão celular (alongamento e aumento radial), o espessamento da parede celular secundária (em que são depositados celulose, hemicelulose, lignina e proteínas da parede celular), e pela morte de células que atingem a maturidade (morte celular programada). Além disso, esse conjunto de alterações anatômicas, estruturais e químicas resultam na propriedade e qualidade da madeira [49, 57–59].

Em um estudo realizado com *Populus tremula* (álamo), por meio de análise proteômica foi possível observar os perfis de expressão de proteínas no tecido xilema durante as fases de desenvolvimento [60]. Nesse estudo foi observado um grande número de proteínas diferencialmente expressas em cada fase de diferenciação. A formação de madeira envolve padrões específicos de formação do citoesqueleto, formação e organização da parede celular e mudanças nos processos biológicos. É evidente que a maquinaria celular trabalha constantemente para o desenvolvimento da planta através da regulação gênica no tecido em desenvolvimento. Contudo, para mensurar e correlacionar o nível de expressão gênica de um determinado tecido é interessante relacionar estudos que abrangem tanto a transcritômica quanto a proteômica [59], como por exemplo, os que foram realizados em *P. tremula* e *Picea abies* (pinheiro) [50, 60–62].

A expressão gênica em *P. abies* [50] foi avaliada por meio de amostras que representassem os estágios de desenvolvimento do xilema, e como esperado a análise demonstrou que os GDEs apresentaram categorias enriquecidas específicas de cada fase, desde a divisão das células até a formação do lenho (madeira). As categorias apresentadas no estudo com *K. grandifoliola* são semelhantes com as encontradas neste estudo de pinheiro. Categorias variadas, por exemplo, para crescimento celular, processo metabólico de macromoléculas e

carboidratos, citoesqueleto, microtúbulos e atividade de hidrolase foram identificadas em diferentes *clusters* de genes envolvidos na formação da madeira.

O metabolismo é um conjunto de reações e vias químicas que podem ser específicos de um determinado tecido ou de diferentes tipos celulares, que no final gerarão os produtos necessários para o desenvolvimento e funcionamento de tecidos e órgãos do organismo [63]. As plantas por sua vez, sintetizam várias outras moléculas orgânicas que não atuarão diretamente nos processos do metabolismo primário (fotossíntese, respiração, nutrição e transporte de solutos). Essas moléculas são conhecidas metabólitos secundários, que poderão ser específicos de um determinado gênero, família ou uma espécie de planta. Exemplos de compostos secundários incluem as moléculas com funções direcionadas à pigmentação, odor, sabor e pesticidas naturais [64].

Neste estudo, uma das categorias enriquecidas dentre os genes mais ativos no xilema foi a de “processo metabólico secundário” (GO:0019748). Dentre os produtos do metabolismo secundário das espécies da família Meliaceae estão os limonoides. Estes, são metabólitos do grupo químico dos tetranoterpenoides, formados via oxidação a partir da via biossintética dos limonoides [65–67]. Esses compostos podem estar presentes em diferentes órgãos e tecidos vegetais, tais como caules, folhas, sementes e frutos [68–72].

A avaliação da atividade biológica dos limonoides de *K. ivorensis* revelou o potencial deste composto no tratamento de malária [39, 73], atividades citotóxicas contra células tumorais [65, 74], além de atividades antifúngicas, antimicrobianas [75], imunossupressoras [76] e pesticida [77].

O óleo das sementes de mogno africano *K. senegalensis* foi efetivo no controle do *Callosobruchus maculatus* (inseto conhecido como caruncho que degrada grãos e sementes) em feijão [78]. Segundo os autores, o controle provavelmente se deu pelo fato do mogno africano e de outras espécies deste gênero possuírem como agentes bioativos os metabólitos secundários como os óleos essenciais derivados dos monoterpenoides que apresentam alta toxicidade para insetos. O uso destes compostos alternativos, como os óleos das Meliaceae, é importante para saúde humana dado a elevada toxicidade dos pesticidas sintéticos convencionais [78, 79].

A via biossintética dos limonóides foi melhor estudada em *Citrus* e envolve a oxidação, isomerização, metilação, acetilação e hidrólise da nomilina [80, 81]. Ao avaliar diferentes tecidos de *Citrus grandis* (L.) Osbeck [40] encontraram 924 GDEs. Dentre esses genes, os autores identificaram 15 genes candidatos que participam da via de biossintese dos limonóides.

Os genes correspondem às enzimas oxidosqualeno ciclase (OSC), CYP450s, UDPs glicosiltransferases e aciltransferases.

Um transcrito da *K. grandifoliola* foi anotado para a enzima chave da biossíntese de terpenóides (Tabela 3), a oxidosqualeno ciclase (codificada pelo gene *OSC*) descrito em *C. grandis* [40]. Esta converte estruturas de carbono em precursores metabólitos de triterpenóides [82]. Outras enzimas são responsáveis por modificar estes precursores, como as CYP450s, glicosiltransferases, metiltransferases e aciltransferases, produzindo todos os tipos de agliconas e glicosídeos de triterpenos [83].

Dos seis genes anotados para CYP450 em *C. grandis* [40], quatro foram identificados para a *K. grandifoliola* (Tabela 3). Na via, estes são responsáveis por modificar os triterpenos em seus grupos hidroxila, cetona, aldeído e carboxílico. Para as glicosiltransferases, foram anotados três genes, enquanto que para *K. grandifoliola* foram 80 (Tabela 3). Estas, fazem a regulação da glicosilação dos triterpenos. A limonóide glicosiltransferase encontrada no banco de dados de *Citrus sinensis* também foi identificada em mogno africano, com um transcrito anotado. Esta enzima consegue converter os agliconas limonóides em glicosídeos limonóides que resultam na perda do amargor de frutos [84]. Para as aciltransferases foram encontrados oito transcritos de *K. grandifoliola* (Tabela 3) contra um de *C. grandis*. Também foi evidenciado em laranja, a expressão de três genes para a enzima hydroquinone glucosyltransferase que tiveram correlação com o conteúdo de limonóides em três estágios de desenvolvimento das folhas. Os 34 genes anotados no transcrito de mogno para esta enzima (Tabela 3) também foram expressos no tecido foliar.

Em *M. champaca*, uma espécie da família Magnoliaceae conhecida como “louro abacate”, foi identificado dentro do domínio PB termos relacionados com a transcrição, regulação transcricional, seguidos de resposta de defesa. Segundo os autores, esses termos sugerem que o tecido está ativo para o metabolismo secundário [42]. Essa planta possui compostos aromáticos em suas flores que são derivados de terpenoides e que são importantes para defesa da planta contra patógenos. Os terpenoides constituem o grupo mais abundante dos metabólitos secundários de plantas, possuindo uma enorme diversificação estrutural. Essas moléculas estão envolvidas em diferentes papéis nas interações planta-inseto, planta-patógeno e planta-planta. A regulação do metabolismo dos terpenoides deve ser elucidado. Para isso é preciso estudos que integrem diferentes áreas como a genética, transcritômica, proteômica e metabolômica [77].

As categorias enriquecidas para os GDEs em folha são plausíveis para o tecido estudado. Em *K. grandifoliola* apesar da quantidade de GDEs de folha ter sido menor do que xilema, a

análise de enriquecimento dos GDEs teve uma representatividade maior em relação ao número de categorias enriquecidas no xilema. O domínio com maior representação de categorias está associado especificamente ao componente celular, sendo que os termos cloroplastos, plastídeos e tilacóides foram identificados como enriquecidos nos genes ativados em folhas.

Um dos questionamentos feitos no estudo de *P. trichocarpa* [47] foi avaliar dentre os cinco tecidos amostrados o porquê da diferença da expressão gênica durante o desenvolvimento de uma planta lenhosa. Por meio da técnica de *microarray*, o caule apresentou um número maior de GDEs, mas com maior proporção de transcritos não caracterizados pelo fato da ausência de informações para espécies lenhosas, especialmente em 2008 quando o artigo foi publicado. Apesar da proporção maior de GDEs no xilema, o tecido foliar apresentou maior número de categorias na análise de enriquecimento funcional, o que se assemelha aos resultados do presente estudo com *K. grandifoliola*. A análise de enriquecimento identificou um número maior de categorias atribuídas a cloroplastos e plastídeos, grande parte envolvidos com a fixação de carbono. Ao comparar as categorias GO entre *A. thaliana* e *P. trichocarpa* (genes ortólogos) foi possível observar que se tratam de um conjunto de genes conservados em folhas jovens e maduras. Para estes órgãos, genes associados aos cloroplastos, plastídios, componentes intracelulares e citoplasmáticos foram super-representados. Esses estão ligados ao complexo fotossintético e ao metabolismo de carboidratos.

A maquinaria fotossintética promove a fixação de carbono pelo ciclo do ácido cítrico gerando energia a ser utilizada para o desenvolvimento da planta. Dentre os genes diferencialmente expressos em *K. grandifoliola* anotados para a categoria GO de fotossíntese, 40 genes apresentaram maior ativação no tecido foliar quando comparado com xilema, sendo que cinco tiveram uma indução maior considerando o valor do *foldchange*, como por exemplo, o TR14037|c0_g1 (*fold change* de -5,20), TR19495|c0_g1 (*fold change* de -9,32), TR5757|c0_g1 (*fold change* de -5,11), TR7042|c2_g1 (*fold change* de -5,22), TR8634|c0_g1 (*fold change* de -5,43). A fotossíntese pode ocorrer tanto por reações dependentes de luz, quanto por assimilação de carbono nos cloroplastos. Estes são organelas que, semelhantes às mitocôndrias, são formadas por duas membranas, compartimentalizado em vesículas que estão circundados por membranas (os tilacóides) organizados em pilhas (grana). Os pigmentos fotossintéticos e os complexos enzimáticos que realizarão as reações luminosas e a síntese de ATP estão presentes nas membranas dos tilacóides, denominadas lamelas. Isso justifica o enriquecimento da categoria GO:0009579 de localização das proteínas no tilacóide. Para as reações de assimilação de carbono, as enzimas estão contidas no estroma (local aquoso

delimitado pela membrana interna da organela). O resultado final das reações de assimilação são a formação de carboidratos como, as trioses e hexoses [85].

Outros termos enriquecidos importantes para o tecido foliar são aqueles associados ao metabolismo do nitrogênio (N) e biogênese de compostos nitrogenados, como por exemplo os aminoácidos e as bases nitrogenadas do DNA. Genes associados a essas categorias são essenciais para a formação de compostos orgânicos envolvidos em reações metabólicas (fotorrespiração), hidrólise de moléculas de transporte, armazenamento de nitrogênio, conversão de aminoácidos que visam o desenvolvimento da planta, etc [125]. No tecido foliar, 11 GDEs (TR18577|c1_g1, TR18577|c2_g1, TR18746|c0_g1, TR19335|c0_g1, TR19335|c0_g2, TR20548|c0_g1, TR21112|c0_g1, TR21112|c0_g2, TR542|c0_g1, TR9639|c1_g1 e TR12225|c0_g1) foram anotados para a proteína RuBisCo (EC: 4.1.1.39), enquanto que para xilema quatro genes foram anotados como tal (TR1748|c0_g1, TR4034|c0_g1, TR542|c0_g2 e TR18577|c2_g2). A RuBisCO é muito abundante em folhas, sendo responsável pela fixação de CO₂ no Ciclo de Calvin [86].

Outras categorias também enriquecidas em folhas, que corroboram com este estudo, foram demonstradas para os GDEs em *nim* (*A. indica*) e em *Apium graveolens* L. [87]. Entre as categorias apresentadas estão aquelas ligadas à homeostase de íons, fotossíntese, transporte de íons e atividade de ATPase em *nim*. Além disso, para processo metabólico foi encontrado enriquecimento funcional para as categorias de estrutura encapsuladora externa, tilacóides, resposta ao estresse, sugerindo que estas categorias estejam associadas com o crescimento da planta.

4.3 ANÁLISE DAS VIAS METABÓLICAS ATIVADAS EM *K. grandifoliola*

Em xilema, a rota metabólica dos terpenóides destaca a ativação da enzima limoneno sintase (EC: 4.2.3.16), enzima responsável pela produção dos monoterpenos [88]. Esse resultado, corrobora com a anotação GO, em que uma das categorias anotadas foi para processo metabólico secundário, que gera compostos da classe dos terpenóides, como é o caso dos limonóides.

Outra via relacionada a formação de metabólitos secundários, foi a via de biossíntese de flavonoides com a ativação da enzima flavonol sintase (FLS) (EC: 1.14.11.23). Esta é responsável por formar os flavonóis kaempferol e quercetina que possuem características

antifúngicas e antioxidantes [89, 90], formando metabólitos secundários, como os flavonóis glicosilados e a flavon-3-ol (catequina) [91].

A ativação das rotas metabólicas dos carboidratos em *K. grandifoliola* foram relacionadas às vias para produção energia, pelas vias do ciclo do ácido cítrico, fosforilação oxidativa, metabolismo de piruvato, propionato e de butanoato.

Em *K. grandifoliola*, a enzima quinol–citocromo-c-redutase (EC: 1.10.2.2) foi ativada na via de fosforilação oxidativa e está relacionada com a absorção de luz pelo citocromo C. A via do metabolismo de propionato, teve genes ativados para as enzimas succinil-CoA: Acetato CoA-transferase (EC: 2.8.3.18) e succinate-CoA ligase (ADP-forming) (EC: 6.2.1.5), que estão relacionadas ao metabolismo central (ciclo do ácido cítrico). A ativação destas enzimas, promove a conversão do succinil-CoA à propionil-CoA, pela via do metilmalonil-CoA [92, 93]. No metabolismo de butanoato (ácido butanoico ou butírico) um dos caminhos realizados na via é a degradação do ácido γ amino-butírico (GABA) pela enzima GABA transaminase para a síntese e transporte de succinato da matriz mitocondrial para o citosol. Em plantas, o GABA é uma enzima que se acumula no citosol das plantas em resposta a diferentes estresses, sinalização e crescimento da planta [94, 95].

Para as rotas de metabolismo de glicerolípídeos e glicerofosfolípídeos, a enzima O-aciltransferase de glicerol-3-fosfato (EC: 2.3.1.15), atua como componente estrutural das membranas biológicas. Essa enzima foi purificada em cloroplastos de *Pisum sativum* (ervilha) e *Spinacia oleracea* (espinafre) [96].

Outra via ativada no xilema de *K. grandifoliola* foi a do ácido α -linolênico com a ativação da enzima 12-oxo-fítodienóico redutase (EC: 1.3.1.42), responsável pela conversão de ácido linolênico (ω -3) em ácido jasmônico (metabólito secundário) nos cloroplastos. Essa enzima foi caracterizada a partir de plântulas de milho (*Zea mays* L.) [97]. O ácido jasmônico é considerado como um hormônio que atua como sinalizador para a ativação de genes em resposta às lesões causadas por herbívoros e situações de estresse, induzindo a rota de defesa vegetal [98, 99].

Em xilema, a enzima ativada para a via de metabolismo de metano fosfoserina fosfatase (EC: 3.1.3.3) (hidrolase) também participa do metabolismo de glicina, serina e treonina. Em estudos com *A. thaliana*, essa enzima foi isolada nos cloroplastos sendo responsável por catalisar a biossíntese de serina tanto em plastídeos quanto nas mitocôndrias. A enzima também é necessária para o desenvolvimento de embriões, pólen e raízes, além de biossíntese de serina em tecidos não fotossintéticos. O metabolismo de serina desempenha papel importante em vias como a glicólise, ciclo do ácido cítrico e na biossíntese de aminoácidos [100–102].

Para a rota de aminoácidos em xilema, foram ativadas as vias de metabolismo de tirosina e fenilalanina, e as vias de biossíntese de fenilalanina, tirosina e triptofano. Nessas vias houve a ativação das enzimas fenilalanina desidrogenase (EC: 1.4.1.20), L-amino-ácido oxidase (EC: 1.4.3.2), aspartato transaminase (EC: 2.6.1.1), transaminase de tirosina (EC: 2.6.1.5), transaminase de aminoácido aromático (EC: 2.6.1.57), transaminase de fenilalanina (EC: 2.6.1.58) e a transaminase de histidinol-fosfato (EC: 2.6.1.9). A biossíntese de triptofano ocorre por meio do composto intermediário corismato. O perfenato, outro composto intermediário derivado do corismato, participa da biossíntese de fenilalanina e tirosina. Estes três aminoácidos são precursores de compostos importantes para o tecido vegetal. Um dos polímeros mais importante que dá rigidez à planta, é a lignina, derivada dos aminoácidos fenilalanina e tirosina. [84]. O triptofano é responsável por ser precursor do ácido indolilacético (indol-3-acetato), conhecido como auxina, molécula importante para a regulação de diversos processos biológicos nas plantas, como por exemplo, o crescimento vegetal [84].

Para o tecido foliar, a rota mais ativada para lipídeos foi para o metabolismo de ácidos graxos, com a ativação da enzima dodecanoil hidrolase (EC 3.1.2.21). Esta é responsável por formar um dos ácidos graxos saturados mais comuns em plantas, o palmitato (produzido nos plastídeos) [103]. A biossíntese de ácidos graxos saturados e insaturados é importante, dado que estes compõem uma variedade de lipídeos envolvidos com funções estruturais das membranas celulares e de armazenamento da planta [104, 105]. Outra enzima ativada na rota, a fosfolipase D (EC: 3.1.4.4) é responsável por hidrolisar fosfolipídeos em diferentes posições nas ligações fosfodiéster entre glicerol e o grupamento fosfato. O resultado é o ácido fosfatídico [106] que atua em processos como transdução de sinal, respostas a condições de estresse biótico e abiótico, sinalização e regulação do crescimento da planta [107–109]. Também foram ativadas na mesma via a lipase triacilglicerol (EC: 3.1.1.3) que hidrolisa o triacilglicerol em diacilglicerol [110]. A enzima diacilglicerol O-aciltransferase (EC: 2.3.1.20) também foi encontrada no mogno-africano e catalisa a formação de triglicerídeos a partir de diacilglicerol e Acetil-CoA [103].

Em folha, na rota dos carboidratos, foram ativadas as vias de metabolismo de galactose, amido e sacarose. Todas as enzimas ativadas para estas vias estão intimamente relacionadas a produção de açúcares simples, fornecendo compostos necessários para a produção de energia e crescimento da planta. Entre as enzimas ativadas, a 4- α -glucanotransferase (EC: 2.4.1.25) foi identificada em *A. thaliana* para a degradação de amido [111]. A degradação de amido na folha é direcionada para a produção de sacarose a ser exportada a outros órgãos da planta ou para consumo pela própria folha através da respiração. Sendo assim, a sacarose exercerá papel

central no crescimento e no desenvolvimento da planta, sendo fonte primária para transporte de carbono e energia. Ou seja, a sacarose funciona como uma molécula importante para a distribuição de fotoassimilados, além de ser regulador da expressão gênica [112, 113].

Outra enzima ativada em folha foi a maltose fosforilase (EC: 2.4.1.8), que utiliza como substratos a maltose e o fosfato para gerar glicose-1-fosfato e glicose. Outras enzimas da classe das invertases também foram ativadas, como a sucrose α -glucosidase (EC: 3.2.1.48) e a β -fructofuranosidase (EC: 3.2.1.26), ambas enzimas hidrolíticas, estão relacionadas a degradação de sacabrose. As hexoses geradas nessa degradação são utilizadas para geração de substratos para armazenamento e produção de energia na planta [103]. A α -glucosidase (EC: 3.2.1.20) é responsável pela clivagem hidrolítica de polissacarídeos, como o glicogênio e o amido [114].

Ainda nas rotas de carboidratos do tecido foliar, houve ativação da enzima malato sintase (EC: 2.3.3.9). Essa enzima participa do Ciclo do Glioxilato nos glioxissomos, por meio da condensação de uma molécula de Acetil-CoA para formar o precursor malato. Este é oxidado a oxalacetato via gliconeogênese gerando glicose com a finalidade de gerar energia para a planta [85, 115]. Outra enzima ativada nesta rota, mas para a via de metabolismo de ascorbato e aldarato, foi a enzima L-gulonolactone oxidase (EC: 1.1.3.8). Essa enzima é responsável por catalisar a última reação de síntese do ácido ascórbico ou ascorbato, conhecido como vitamina C, a partir da glicose. O ácido ascórbico atua como co-fator enzimático permitindo o funcionamento correto das enzimas em plantas, animais, algas e outros organismos [116, 117]. Em plantas e algas, a vitamina C possui papel fotoprotetor contra espécies reativas de oxigênio (EROs), moléculas citotóxicas que são originadas em plantas em situações normais ou de estresse por organelas como as mitocôndrias, peroxissomos e cloroplastos [118]. Sendo assim, o metabolismo celular de organismos fotossintéticos evoluiu para a biossíntese de ascorbato para formar sistemas antioxidantes, como o ciclo ascorbato-glutationa, que visam evitar a toxicidade gerada pelas EROs [119, 120].

Outra enzima ativada em folha foi a transcetolase (EC: 2.2.1.1), responsável pela assimilação de CO₂, principalmente para a fixação de carbono. Esta é responsável por promover o rearranjo dos esqueletos de carbono das moléculas trioses-fosfato gerando as pentoses-fosfato, que são convertidas em outros compostos intermediários-chave do Ciclo de Calvin [85].

Em folha, a enzima ativada ribonucleotídeo-difosfato redutase (EC: 1.17.4.1) participa da via de metabolismo de nucleotídeos purina. Essa enzima catalisa a redução de ribonucleotídeos a desoxirribonucleotídeos, etapa crucial para a replicação e reparo do ácido desoxirribonucleico (DNA) [121].

A ativação da rota de metabólitos secundários no tecido foliar gerou a ativação da enzima antranilato sintase (EC: 4.1.3.27) para a via de biossíntese de fenazina. Esta enzima faz a conversão do corismato em antranilato. A via do chiquimato produz o composto corismato, que é utilizado para a produção de compostos aromáticos, como os aminoácidos triptofano, fenilalanina e tirosina, desempenhando papéis importantes para o crescimento, desenvolvimento e defesa de plantas. O corismato tem papel precursor para a produção de outros cofatores, como o folato, coenzimas benzóides e naftóideas, fenazinas e sideróforos [122–124].

De forma geral, a análise dos transcritos de *K. grandifoliola* demonstram uma ampla gama de vias metabólicas ativadas nos dois tecidos amostrados: folhas e xilema. Dessa maneira, esse estudo provê, de maneira inédita, informações sobre a genômica funcional da espécie, abrindo caminho para novos estudos sobre a biossíntese e regulação dos limonóides ou outras vias metabólicas.

5 CONCLUSÃO

O sequenciamento *de novo* possibilitou a montagem e análise do transcrito da espécie *K. grandifoliola*, com excelente cobertura dos genes e vias metabólicas ativadas em folhas e no xilema. Esse resultado é relevante pelo indeditismo, dado que não existem trabalhos publicados no campo da genômica com esta espécie. O transcrito revelou a presença de 36.271 genes e de acordo com a análise funcional das sequências de mogno africano, estas apresentaram maior similaridade com o gênero de *Citrus*, pertencente à ordem *Sapindales* onde também está inserido o gênero *Khaya*. Ambas análises de expressão gênica diferencial e das vias metabólicas demonstraram maior número de GDEs e de vias no tecido do xilema do que em folha, indicando maior diversidade de transcritos em xilema. Os termos GO enriquecidos e as vias metabólicas ativadas em cada tecido, demonstram suas funções e atividades metabólicas. A análise de enriquecimento e das vias metabólicas evidenciou que os GDEs do tecido do xilema estão relacionados à formação de parede celular e metabólitos secundários, enquanto que para folha, os GDEs foram enriquecidos para a maquinaria fotossintética.

6 REFERÊNCIAS

1. Pinheiro AL, Couto L, Pinheiro DT, Brunetta JMFC. Ecologia, silvicultura e tecnologia de utilização dos mognos-africanos (*Khaya* spp.). 2011.
2. Casara HN. INSTRUÇÃO NORMATIVA Nº 17, DE 19 DE OUTUBRO DE 2001. INSTITUTO BRASILEIRO DO MEIO AMBIENTE E DOS RECURSOS NATURAIS RENOVÁVEIS. 2001. http://www.mma.gov.br/estruturas/pnf/_arquivos/in_ibama_17_01.pdf.
3. CITES. CONVENTION INTERNATIONAL TRADE IN ENDANGERED. AMENDMENTS TO APPENDICES I AND II OF THE CONVENTION. 2013:89–100. https://www.cites.org/sites/default/files/eng/cop/09/prop/E09-Prop-100_Khaya.PDF. Acessado 8 de janeiro de 2019.
4. Newton AC, Leakey RRB, Mesén JF. Genetic variation in mahoganies: its importance, capture and utilization. *Biodivers Conserv*. 1993;2:114–26.
5. Maués MM. A review of *Hypsipyla grandella* Zeller research in Pará State, Brazil. *ACIAR Proc*. 2001;97:63–6.
6. Verzignassi JR, Poltronieri LS, Benchimol RL. Mancha-alvo em mogno-africano no Brasil. *Summa Phytopathol*. 2009;35:70–1.
7. Tremacoldi CR, Lunz AM, Coelho IL, Boari A de J. Cancro em mogno africano no estado do Pará. *Pesqui Florest Bras*. 2013;33:221–5.
8. Silva LF da, Ferreira GL, Santos ACA dos, Leite HG, Silva ML da. Equações Hipsométricas, Volumétricas e de Crescimento para *Khaya ivorensis* Plantada em Pirapora. *Floresta e Ambient*. 2016;23:362–8.
9. Hawthorne W. *Khaya grandifoliola*. The IUCN Red List of Threatened Species. 1998.
10. Ward JA, Ponnala L, Weber CA. Strategies for transcriptome analysis in nonmodel plants. *Am J Bot*. 2012;99:167–76.
11. Mardis ER. Next-Generation Sequencing Platforms. *Annu Rev Anal Chem*. 2013;6:287–303.
12. Dijk EL van, Auger H, Jaszczyszyn Y, Thermes C. Ten years of next-generation sequencing technology. *Trends Genet*. 2014;30:418–26.
13. Goodwin S, McPherson JD, McCombie WR. Coming of age: Ten years of next-generation sequencing technologies. *Nat Rev Genet*. 2016;17:333–51.
14. Mardis ER. Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet*. 2008;9:387–402.

15. Wang ET, Sandberg R, Luo S, Khrebtkova I, Zhang L, Mayr C, et al. Alternative isoform regulation in human tissue transcriptomes. *Nature*. 2008;456:470–6.
16. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*. 2009;10:57–63.
17. Mantello CC, Cardoso-Silva CB, Da Silva CC, De Souza LM, Scaloppi EJ, de Gonçalves PS, et al. De Novo assembly and transcriptome analysis of the rubber tree (*Hevea brasiliensis*) and SNP markers development for rubber biosynthesis pathways. *PLoS One*. 2014;9:1–14.
18. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc*. 2013;8:1494–512.
19. Chang S, Puryear J, Cairney J. A simple and efficient method for isolating RNA from pine trees. *Plant Mol Biol Report*. 1993;11:113–6.
20. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
21. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114–20.
22. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat Biotechnol*. 2011;29:644–52.
23. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, et al. SILVA: A comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res*. 2007;35:7188–96.
24. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, et al. Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Res*. 2012;40:1178–86.
25. Oliveros JC. Venny. An interactive tool for comparing lists with Venn’s diagrams. 2007. <http://bioinfogp.cnb.csic.es/tools/venny/index.html>.
26. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357–9.
27. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 2011;12:1–16.
28. Robinson MD, McCarthy DJ, Smyth GK. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2009;26:139–40.
29. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*. 2000;28:27–30.
30. Yamada T, Letunic I, Okuda S, Kanehisa M, Bork P. IPATH2.0: Interactive pathway explorer. *Nucleic Acids Res*. 2011;39 SUPPL. 2:412–5.

31. Darzi Y, Letunic I, Bork P, Yamada T. IPath3.0: Interactive pathways explorer v3. *Nucleic Acids Res.* 2018;46:510–3.
32. Krishnan NM, Pattnaik S, Jain P, Gaur P, Choudhary R, Vaidyanathan S, et al. A draft of the genome and four transcriptomes of a medicinal and pesticidal angiosperm *Azadirachta indica*. *BMC Genomics.* 2012;13:1.
33. Torales SL, Rivarola M, Gonzalez S, Inza MV, Pomponio MF, Fernández P, et al. De novo transcriptome sequencing and SSR markers development for *Cedrela balansae* C.DC., a native tree species of northwest Argentina. *PLoS One.* 2018;13:1–16.
34. Perkel JM. Visiting ‘noncodarnia’. *Biotechniques.* 2013;54:301–4.
35. Wang QQ, Liu F, Chen XS, Ma XJ, Zeng HQ, Yang ZM. Transcriptome profiling of early developing cotton fiber by deep-sequencing reveals significantly differential expression of genes in a fuzzless/lintless mutant. *Genomics.* 2010;96:369–76.
36. Xia Z, Xu H, Zhai J, Li D, Luo H, He C, et al. RNA-Seq analysis and de novo transcriptome assembly of *Hevea brasiliensis*. *Plant Mol Biol.* 2011;77:299–308.
37. Li D, Deng Z, Qin B, Liu X, Men Z. De novo assembly and characterization of bark transcriptome using Illumina sequencing and development of EST-SSR markers in rubber tree (*Hevea brasiliensis* Muell. Arg.). *BMC Genomics.* 2012;13.
38. Liu J, Wang Y, Li Q. Analysis of differentially expressed genes and adaptive mechanisms of *Prunus triloba* Lindl. under alkaline stress. *Hereditas.* 2017;154:10.
39. Agbedahunsi JM, Fakoya FA, Adesanya SA. Studies on the anti-inflammatory and toxic effects of the stem bark of *Khaya ivorensis* (Meliaceae) on rats. *Phytomedicine.* 2004;11:504–8.
40. Wang F, Wang M, Liu X, Xu Y, Zhu S, Shen W, et al. Identification of Putative Genes Involved in Limonoids Biosynthesis in Citrus by Comparative Transcriptomic Analysis. *Front Plant Sci.* 2017;8 May:1–12.
41. Stein L. Genome annotation: From sequence to biology. *Nat Rev Genet.* 2001;2:493–503.
42. Dhandapani S, Jin J, Sridhar V, Sarojam R, Chua NH, Jang IC. Integrated metabolome and transcriptome analysis of *Magnolia champaca* identifies biosynthetic pathways for floral volatile organic compounds. *BMC Genomics.* 2017;18:1–18.
43. Liu JJ, Sturrock RN, Benton R. Transcriptome analysis of *Pinus monticola* primary needles by RNA-seq provides novel insight into host resistance to *Cronartium ribicola*. *BMC Genomics.* 2013;14:1–16.
44. Wang Y, Chen X, Wang J, Xun H, Sun J, Tang F. Comparative analysis of the terpenoid biosynthesis pathway in *Azadirachta indica* and *Melia azedarach* by RNA-seq. *Springerplus.* 2016;5.

45. Rai A, Kamochi H, Suzuki H, Nakamura M, Takahashi H, Hatada T, et al. De novo transcriptome assembly and characterization of nine tissues of *Lonicera japonica* to identify potential candidate genes involved in chlorogenic acid, luteolosides, and secoiridoid biosynthesis pathways. *J Nat Med.* 2017;71:1–15.
46. Carvalho MC da CG de, Tese. Caracterização de perfis transcricionais de folhas e da região cambial de *Eucalyptus grandis* usando o SAGE. Universidade de São Paulo - Escola Superior de Agricultura Luiz de Queiroz; 2006.
47. Quesada T, Li Z, Dervinis C, Li Y, Bock P, Tuskan GA, et al. Comparative analysis of the transcriptomes of *Populus trichocarpa* and *Arabidopsis thaliana* suggests extensive evolution of gene expression regulation in angiosperms. *New Phytol.* 2008;180:408–20.
48. Mizrahi E, Joubert F, Myburg AA, Hefer CA, Ranik M. De novo assembled expressed gene catalog of a fast-growing *Eucalyptus* tree produced by Illumina mRNA-Seq. *BMC Genomics.* 2010;11:1–12.
49. Zhang J, Nieminen K, Serra JAA, Helariutta Y. The formation of wood and its control. *Curr Opin Plant Biol.* 2014;17:56–63.
50. Jokipii-Lukkari S, Sundell D, Nilsson O, Hvidsten TR, Street NR, Tuominen H. NorWood: a gene expression resource for evo-devo studies of conifer wood development. *New Phytol.* 2017;216:482–94.
51. Lucas WJ, Groover A, Lichtenberger R, Furuta K, Yadav SR, Helariutta Y, et al. The Plant Vascular System: Evolution, Development and Functions. *J Integr Plant Biol.* 2013;55:294–388.
52. Pavy N, Boyle B, Nelson C, Paule C, Giguère I, Caron S, et al. Identification of conserved core xylem gene sets: Conifer cDNA microarray development, transcript profiling and computational analyses. *New Phytol.* 2008;180:766–86.
53. Hefer CA, Mizrahi E, Myburg AA, Douglas CJ, Mansfield SD. Comparative interrogation of the developing xylem transcriptomes of two wood-forming species: *Populus trichocarpa* and *Eucalyptus grandis*. *New Phytol.* 2015;206:1391–405.
54. Dharmawardhana P, Brunner AM, Strauss SH. Genome-wide transcriptome analysis of the transition from primary to secondary stem development in *Populus trichocarpa*. 2010;11:1–19.
55. FUKUDA H. Signaling, transcriptional regulation, and asynchronous pattern formation governing plant xylem development. *Proc Japan Acad Ser B.* 2016;92:98–107.
56. Nairn CJ, Lennon DM, Wood-Jones A, Nairn A V., Dean JFD. Carbohydrate-related genes and cell wall biosynthesis in vascular tissues of loblolly pine (*Pinus taeda*). *Tree Physiol.* 2008;28:1099–110.
57. Hertzberg M, Aspeborg H, Schrader J, Andersson A, Erlandsson R, Blomqvist K, et al. A transcriptional roadmap to wood formation. *Proc Natl Acad Sci United States Am.* 2001;98:14732–14737.

58. Paiva JAP, Alves A, Brach J, Frigerio J, Fevereiro P, Plomion C. Molecular and phenotypic profiling from the base to the crown in maritime pine wood-forming tissue. *New Phytol.* 2008;283–301.
59. Seyfferth C, Wessels B, Jokipii-Lukkari S, Sundberg B, Delhomme N, Felten J, et al. Ethylene-Related Gene Expression Networks in Wood Formation. *Front Plant Sci.* 2018;9:1–17.
60. Obudulu O, Bygdell J, Sundberg B, Moritz T, Hvidsten TR, Trygg J, et al. Quantitative proteomics reveals protein profiles underlying major transitions in aspen wood development. *BMC Genomics.* 2016;17:1–17.
61. Bygdell J, Srivastava V, Obudulu O, Srivastava MK, Nilsson R, Sundberg B, et al. Protein expression in tension wood formation monitored at high tissue resolution in *Populus*. *J Exp Bot.* 2017;68:3405–17.
62. Sundell D, Street NR, Kumar M, Mellerowicz EJ, Kucukoglu M, Johnsson C, et al. AspWood: High-Spatial-Resolution Transcriptome Profiles Reveal Uncharacterized Modularity of Wood Formation in *Populus tremula*. *Plant Cell.* 2017;29:1585–604.
63. The Gene Ontology Consortium. Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res.* 2017;45:331–338.
64. García AA, Carril EP-U. Metabolismo secundario de plantas. *Reduca (Biología).* 2009;2:119–45. <http://revistareduca.es/index.php/biologia/article/view/798>.
65. Kim J, Jayaprakasha GK, Patil BS. Limonoids and their anti-proliferative and anti-aromatase properties in human breast cancer cells. *Food Funct.* 2013;4:258–61.
66. Ji KL, Liao SG, Zheng XL, Na Z, Hu H Bin, Zhang P, et al. Limonoids from the Fruits of *Khaya ivorensis*. *Molecules.* 2014;19:3004–11.
67. Wang F, Yu X, Liu X, Shen W, Zhu S, Zhao X. Temporal and spatial variations on accumulation of nomilin and limonin in the pummelos. *Plant Physiol Biochem.* 2016;106:23–9.
68. McIntosh CA, Mansell RL, Rouseff RL. Distribution of Limonin in the Fruit Tissues of Nine Grapefruit Cultivars. *J Agric Food Chem.* 1982;30:689–92.
69. Rouseff RL, Nagy S. Distribution of limonoids in citrus seeds. *Phytochemistry.* 1982;21:85–90.
70. Herman Z, Hasegawa S, Fong CH, Ou P. Limonoids in *Citrus ichangensis*. *J Agric Food Chem.* 1989;37:850–1.
71. Li S, Wang Z, Ding F, Sun D, Ma Z, Cheng Y, et al. Content changes of bitter compounds in ‘Guoqing No.1’ Satsuma mandarin (*Citrus unshiu* Marc.) during fruit development of consecutive 3 seasons. *Food Chem.* 2014;145:963–9.

72. Chen J, Li S, Xu J, Ding F, Wang Z, Cheng Y, et al. Concentration and distribution of main bitter compounds in fruit tissues of 'Oroblanco' (*Citrus grandis* L. × *Citrus paradisi* Macf.). *Sci Hortic (Amsterdam)*. 2015;193:84–9.
73. Tepongning RN, Lucantoni L, Nasuti CC, Dori GU, Yerbanga SR, Lupidi G, et al. Potential of a *Khaya ivorensis* – *Alstonia boonei* extract combination as antimalarial prophylactic remedy. *J Ethnopharmacol*. 2011;137:743–51.
74. Kim H, Lee T, Park W, Lee JW, Kim J, Lee BY, et al. Peeling back the evolutionary layers of molecular mechanisms responsive to exercise-stress in the skeletal muscle of the racing horse. *DNA Res*. 2013;20:287–98.
75. Abdelgaleil SA, Hashinaga F, Nakatani M. Antifungal activity of limonoids from *Khaya ivorensis*. *Pest Manag Sci*. 2005;61:186–90.
76. Zhang B, Wang Y, Yang SP, Zhou Y, Wu W Bin, Tang W, et al. Ivorenolide a, an unprecedented immunosuppressive macrolide from *Khaya ivorensis*: Structural elucidation and bioinspired total synthesis. *J Am Chem Soc*. 2012;134:20605–8.
77. Cheng A-X, Lou Y-G, Mao Y-B, Lu S, Wang L-J, Chen X-Y. Plant Terpenoids: Biosynthesis and Ecological Functions. *J Integr Plant Biol*. 2007;49:179–186.
78. Bamaiyi LJ, Ndams IS, Toro WA, Odekina S. Effect of Mahogany *Khaya senegalensis* Seed Oil in the Control of *Callosobruchus maculatus* on Stored Cowpea. *Plant Prot Sci*. 2006;42:130–4.
79. Perez J, Eigenbrode S, Hilje L, Tripepi R, Aguilar ME, Mesén F. Leaves from grafted Meliaceae species affect survival and performance of *Hypsipyla grandella* (Zeller) (Lepidoptera: Pyralidae) larvae. *J Pest Sci (2004)*. 2010;83:95–104.
80. Hasegawa S, Bennett RD, Verdon CP. Metabolism of limonoids via a deoxylimonoid pathway in *Citrus*. *Phytochemistry*. 1980;19:1445–7.
81. Hasegawa S, Herman Z. Biosynthesis of obacunone from nomilin in *Citrus limon*. *Phytochemistry*. 1985;24:1973–4.
82. Phillips DR, Rasbery JM, Bartel B, Matsuda SP. Biosynthetic diversity in plant triterpene cyclization. *Curr Opin Plant Biol*. 2006;9:305–14.
83. Thimmappa R, Geisler K, Louveau T, O'Maille P, Osbourn A. Triterpene Biosynthesis in Plants. *Annu Rev Plant Biol*. 2014;65:225–57.
84. Hasegawa S, Suhayda CG, Hsu WJ, Robertson GH. Purification of limonoid glucosyltransferase from navel orange albedo tissues. *Phytochemistry*. 1997;46:33–7.
85. Nelson DL, Cox MM. *Principios de Bioquímica de Lehninger*. 6th edition. Porto Alegre: Artmed; 2014.
86. Irving LJ, Robinson D. A dynamic model of Rubisco turnover in cereal leaves. *New Phytol*. 2006;169:493–504.

87. Jia XL, Wang GL, Xiong F, Yu XR, Xu ZS, Wang F, et al. De novo assembly, transcriptome characterization, lignin accumulation, and anatomic characteristics: Novel insights into lignin biosynthesis during celery leaf development. *Sci Rep.* 2015;5:1–14.
88. Lücker J, El Tamer MK, Schwab W, Verstappen FWA, Van Der Plas LHW, Bouwmeester HJ, et al. Monoterpene biosynthesis in lemon (*Citrus limon*) cDNA isolation and functional analysis of four monoterpene synthases. *Eur J Biochem.* 2002;269:3160–71.
89. Costanzo CDG, Fernandes VC, Zingaretti S, Belebony RO, Pereira AMS, Marins M, et al. Isolation of flavonoids from *Anemopaegma arvense* (vell) stellf. ex de souza and their antifungal activity against *Trichophyton rubrum*. *Brazilian J Pharm Sci.* 2013;49:559–65.
90. Omezzine F, Bouaziz M, Daami-Remadi M, Simmonds MSJ, Haouala R. Chemical composition and antifungal activity of *Trigonella foenum-graecum* L. varied with plant ploidy level and developmental stage. *Arab J Chem.* 2014;10:3622–31.
91. Oliveira DM. ESTUDO FITOQUÍMICO E BIOLÓGICO DAS FOLHAS DE *Banisteriopsis argyrophylla* (A. Juss.) B. Gates (Malpighiaceae). UNIVERSIDADE FEDERAL DE UBERLÂNDIA; 2016.
92. Aldor IS, Kim SW, Jones Prather KL, Keasling JD. Metabolic engineering of a novel propionate-independent pathway for the production of poly(3-hydroxybutyrate-co-3-hydroxyvalerate) in recombinant *Salmonella enterica* serovar Typhimurium. *Appl Environ Microbiol.* 2002;68:3848–54.
93. Suvorova IA, Ravcheev DA, Gelfand MS. Regulation and evolution of malonate and propionate catabolism in proteobacteria. *J Bacteriol.* 2012;194:3234–40.
94. Clark SM, Di Leo R, Dhanoa PK, Van Cauwenberghe OR, Mullen RT, Shelp BJ. Biochemical characterization, mitochondrial localization, expression, and potential functions for an *Arabidopsis* γ -aminobutyrate transaminase that utilizes both pyruvate and glyoxylate. *J Exp Bot.* 2009;60:1743–57.
95. Michaeli S, Fait A, Lagor K, Nunes-Nesi A, Grillich N, Yellin A, et al. A mitochondrial GABA permease connects the GABA shunt and the TCA cycle, and is essential for normal carbon metabolism. *Plant J.* 2011;67:485–98.
96. Bertrams M, Heinz E. Positional Specificity and Fatty Acid Selectivity of Purified sn-Glycerol 3-Phosphate Acyltransferases from Chloroplasts. *Plant Physiol.* 2008;68:653–7.
97. Vick BA, Zimmerman DC. Characterization of 12-Oxo-Phytodienoic Acid Reductase in Corn: The Jasmonic Acid Pathway. *Plant Physiol.* 1986;80:202–5.
98. Soares AM dos S, Machado OLT. Defesa de plantas: Sinalização química e espécies reativas de oxigênio. *Rev Trópica – Ciência Agrárias e Biológicas.* 2007;1.
99. Deuner C, Borges C, Almeida A, Meneghello G, Tunes L. Ácido jasmônico como promotor de resistência em plantas. *Rev Ciências Agrárias.* 2015;38:275–81.

100. Ho C-L, Noji M, Saito K. Plastidic Pathway of Serine Biosynthesis. *J Biol Chem.* 1999;274:11007–12.
101. Cascales-Minana B, Munoz-Bertomeu J, Flores-Tornero M, Anoman AD, Pertusa J, Alaiz M, et al. The Phosphorylated Pathway of Serine Biosynthesis Is Essential Both for Male Gametophyte and Embryo Development and for Root Growth in *Arabidopsis*. *Plant Cell.* 2013;25:2084–101.
102. Baptista CG. Metabolismo de serina: Caracterização de Serina Hidroximetiltransferase de *Trypanosoma cruzi*. Universidade de São Paulo; 2017.
103. Caspi R, Altman T, Billington R, Dreher K, Foerster H, Fulcher CA, et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res.* 2014;42:459–71.
104. Dörmann P, Voelker TA, Ohlrogge JB. Accumulation of Palmitate in *Arabidopsis* Mediated by the Acyl-Acyl Carrier Protein Thioesterase FATB1. *Plant Physiol.* 2000;123:637–44.
105. Lopes ADS. Caracterização molecular e fisiológica de plantas transplastômicas de tabaco expressando dessaturases de cianobactéria. Universidade Federal de Viçosa; 2014.
106. Heller M. Phospholipase D. In: Paoletti R, Kritchevsky D, editors. *Advances in Lipid Research.* 1978. p. 267–326.
107. Guo L, Mishra G, Markham JE, Li M, Tawfall A, Welti R, et al. Connections between sphingosine kinase and phospholipase D in the abscisic acid signaling pathway in *Arabidopsis*. *J Biol Chem.* 2012;287:8286–96.
108. Pinosa F, Buhot N, Kwaaitaal M, Fahlberg P, Thordal-Christensen H, Ellerstrom M, et al. *Arabidopsis* Phospholipase D Is Involved in Basal Defense and Nonhost Resistance to Powdery Mildew Fungi. *Plant Physiol.* 2013;163:896–906.
109. Khatoon H, Mansfeld J, Schierhorn A, Ulbrich-Hofmann R. Purification, sequencing and characterization of phospholipase D from Indian mustard seeds. *Phytochemistry.* 2015;17:65–75.
110. Eastmond PJ. SUGAR-DEPENDENT1 Encodes a Patatin Domain Triacylglycerol Lipase That Initiates Storage Oil Breakdown in Germinating *Arabidopsis* Seeds. *Plant Cell Online.* 2006;18:665–75.
111. Critchley JH, Zeeman SC, Takaha T, Smith AM, Smith SM. A critical role for disproportionating enzyme in starch breakdown is revealed by a knock-out mutation in *Arabidopsis*. *Plant J.* 2001;26:89–100.
112. Koch K. Sucrose metabolism: Regulatory mechanisms and pivotal roles in sugar sensing and plant development. *Curr Opin Plant Biol.* 2004;7:235–46.

113. Peterson J, Gardin P. Desenvolvimento Reprodutivo Do Cafeeiro: Anatomia De Vasos Do Xilema E Dinâmica De Carboidratos. 2006.
114. Kumar S, Narwal S, Kumar V, Prakash O. α -glucosidase inhibitors from plants: A natural approach to treat diabetes. *Pharmacogn Rev.* 2011;5:19–29.
115. Dixon GH, Kornberg HL, Lund P. Purification and properties of malate synthetase. *Biochim Biophys Acta.* 1960;41:217–33.
116. Mandl J, Szarka A, Bánhegyi G. Vitamin C: Update on physiology and pharmacology. *Br J Pharmacol.* 2009;157:1097–110.
117. Blaschke K, Ebata KT, Karimi MM, Zepeda- J a, Goyal P, Mahapatra S, et al. Vitamin C induces Tet-dependent DNA demethylation in ESCs to promote a blastocyst-like state. *Nature.* 2013;500:222–6.
118. Costa MG de S. Alterações no proteoma caulinar de *Eucalyptus globulus* e *Eucalyptus grandis* em resposta a alterações no proteoma caulinar de *Eucalyptus*. Universidade Estadual Paulista; 2017.
119. Wheeler G, Ishikawa T, Pornsaksit V, Smirnov N. Evolution of alternative biosynthetic pathways for vitamin C following plastid acquisition in photosynthetic eukaryotes. *Elife.* 2015;4:1–25.
120. Noshi M, Hatanaka R, Tanabe N, Terai Y, Maruta T, Shigeoka S. Redox regulation of ascorbate and glutathione by a chloroplastic dehydroascorbate reductase is required for high-light stress tolerance in *Arabidopsis*. *Biosci Biotechnol Biochem.* 2016;80:870–7.
121. Yu Y, Rahmanto YS, Hawkins CL, Richardson DR. The Potent and Novel Thiosemicarbazone Chelators Di-2-pyridylketone-4,4-dimethyl-3-thiosemicarbazone and 2-Benzoylpyridine-4,4-dimethyl-3-thiosemicarbazone Affect Crucial Thiol Systems Required for Ribonucleotide Reductase Activity. *Mol Pharmacol.* 2011;79:921–31.
122. Dosselaere F, Vanderleyden J. A metabolic node in action: Chorismate-utilizing enzymes in microorganisms. *Crit Rev Microbiol.* 2001;27:75–131.
123. Guttenberger N, Blankenfeldt W, Breinbauer R. Recent developments in the isolation, biological function, biosynthesis, and synthesis of phenazine natural products. *Bioorganic Med Chem.* 2017;25:6149–66.
124. Parthasarathy A, Cross PJ, Dobson RCJ, Adams LE, Savka MA, Hudson AO. A Three-Ring Circus: Metabolism of the Three Proteogenic Aromatic Amino Acids and Their Role in the Health of Plants and Animals. *Front Mol Biosci.* 2018;5:1–30.
125. Ireland RJ, Lea PJ. The enzymes of glutamine, glutamate, asparagine, and aspartate metabolism. In BK Singh, ed, *Plant Amino Acids. Biochemistry and Biotechnology*, 1999: 49-109.