

**UNIVERSIDADE FEDERAL DE JUIZ DE FORA**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS - IMUNOLO-**  
**GIA / GENÉTICA E BIOTECNOLOGIA**

**VINICIUS CARIUS DE SOUZA**

**ANÁLISE DO TRANSCRIPTOMA DE *Lippia alba* (MILL.) N.E.BR. (Verbenaceae)**  
**POR RNASEQ VISANDO A IDENTIFICAÇÃO DE ENZIMAS TERPENO SINTASES**

**Juiz de Fora**

**2016**

**2016 PGCBIO**

**Vinicius Carius de Souza**

**ICB/UFJF**

**VINICIUS CARIUS DE SOUZA**

**ANÁLISE DO TRANSCRIPTOMA DE *Lippia alba* (MILL.) N.E.BR. (Verbenaceae) POR  
RNASEQ VISANDO A IDENTIFICAÇÃO DE ENZIMAS TERPENO SINTASES**

Dissertação apresentada ao Curso de Pós-Graduação em Ciências Biológicas, da Universidade Federal de Juiz de Fora como requisito parcial para obtenção do grau de Mestre em Ciências Biológicas na área de Genética e Biotecnologia.

Orientador: Prof. Dr. Marcelo de Oliveira Santos

Co-Orientadora: Prof<sup>ª</sup>. Dr<sup>ª</sup>. Priscila Vanessa Zabala Capriles Goliatt

**Juiz de Fora**

**2016**

Ficha catalográfica elaborada através do programa de geração automática da Biblioteca Universitária da UFJF, com os dados fornecidos pelo(a) autor(a)

Carius de Souza, Vinicius.

Análise do Transcriptoma de *Lippia alba* (Mill.) N.E.Br. (Verbenaceae) por RNAseq visando a Identificação de Enzimas Terpeno Sintases / Vinicius Carius de Souza. -- 2016.  
79 p.

Orientador: Marcelo de Oliveira Santos

Coorientadora: Priscila Vanessa Zabala Capriles Goliatt

Dissertação (mestrado acadêmico) - Universidade Federal de Juiz de Fora, Instituto de Ciências Biológicas. Programa de Pós-Graduação em Ciências Biológicas: Imunologia e Genética, 2016.

1. *Lippia alba*. 2. Transcriptoma. 3. Sequenciamento. 4. Anotação funcional. 5. PCR em tempo real. I. de Oliveira Santos, Marcelo, orient. II. Vanessa Zabala Capriles Goliatt, Priscila, coorient. III. Título.

**Vinicius Carius de Souza**

**Análise do Transcriptoma de *Lippia alba* (Mill.) N.E.Br. (Verbenaceae) por  
RNAseq Visando a Identificação de Enzimas Terpeno Sintases**

Dissertação apresentada ao Curso de Pós-Graduação em Ciências Biológicas, da Universidade Federal de Juiz de Fora como requisito parcial para obtenção do grau de Mestre em Ciências Biológicas na área de Genética e Biotecnologia:

Aprovada em \_\_\_\_ de \_\_\_\_\_ de \_\_\_\_\_

**BANCA EXAMINADORA**

---

Prof. Dr. Marcelo de Oliveira Santos – Orientador  
Universidade Federal de Juiz de Fora

---

Prof<sup>a</sup>. Dr<sup>a</sup>. Priscila Vanessa Zabala Capriles Goliatt – Co-orientadora  
Universidade Federal de Juiz de Fora

---

Dr<sup>a</sup>. Ana Luisa Sousa Azevedo  
Empresa Brasileira de Pesquisa Agropecuária

---

Prof. Dr. Lyderson Facio Viccini  
Universidade Federal de Juiz de Fora

“ Você certamente encontrará algo se procurar, mas nem sempre é o que você procura! ”

The Hobbit, J.R.R. Tolkien

Dedico este trabalho aos meus pais pelo amor e confiança. E acima de tudo pelos conselhos e incentivos.

## **AGRADECIMENTOS**

Agradeço aos meus pais Valter e Imaculada, meus irmãos Vitor e Igor e demais membros da minha família pelo amor incondicional e suporte em todos os momentos, participando e torcendo por cada passo meu; ao meu orientador, professor Dr. Marcelo de Oliveira Santos, pela oportunidade de trabalho tão valiosa e também por acreditar em mim; à minha co-orientadora, professora Dra. Priscila Capriles Goliatt, por sua disponibilidade e ajuda na realização deste trabalho; à professora Dra. Cristiane Zorzatto e ao professor Dr. Humberto Henrique de Carvalho, os quais contribuíram com este trabalho; à professora Dra. Letícia Stephan por sua disponibilidade em sempre ajudar; aos amigos Danielle, Francisco, Gustavo, Juliana, Leonardo, Rafaella, Sirlei e Vinicius Schmitz pelo companheirismo, ajuda e discussões enriquecedoras; aos demais professores e alunos do laboratório de genética por tornarem a convivência mais agradável; à EMBRAPA por ceder o material biológico usado neste trabalho; à UFJF, ao PPGCBIO-UFJF, ao PGMC-UFJF, CAPES, CNPq e FAPEMIG pelo apoio financeiro.

## RESUMO

*Lippia alba*, popularmente conhecida por erva-cidreira, é uma espécie vegetal amplamente distribuída pelas Américas e encontrada praticamente em todo o território brasileiro. Esta espécie possui importante uso na medicina tradicional para o tratamento de cólicas, indigestão, náuseas, espasmos, diarreia, disenteria, doenças respiratórias, problemas hepáticos e no tratamento de sífilis e gonorreia. As folhas de *L. alba*, as quais são preparadas sob a forma de infusão ou decocção e ingeridas por via oral, produzem um óleo essencial rico em moléculas isoprenóides denominadas terpenóides. Estes compostos não são apenas de interesse farmacológico, mas também industrial já que são usados na confecção de fragrâncias. A composição dos óleos essenciais pode variar em função de diferentes fatores abióticos e genotípicos, como por exemplo nível de ploidia. Neste contexto, os objetivos deste trabalho foram caracterizar o transcriptoma de folha da espécie *L. alba* e buscar sequências putativas de enzimas envolvidas na produção de metabólitos secundários. O transcriptoma foi sequenciado pela plataforma *Mi-seq* (Illumina) com bibliotecas *paired-end* de 300 bp. O sequenciamento resultou em um total de 47.498.310 *reads paired-end* (23.749.155 *reads* para cada *end* sequenciado) de 35-308 bp, compreendendo 12.148.327.567 nucleotídeos (~12 Gb). A montagem *de novo* dos transcritos foi processada a partir do software *Trinity* que gerou 193.532 transcritos, sendo 128.209 uni-genes, com o valor de N50 igual a 1.187 bp. Um total de 86.122 ORF (Open Read Frame) foi obtido e a seguir submetido ao algoritmo de alinhamentos *BlastP*, o qual encontrou 75.533 sequências com referência no banco de dados NR (*Non-Redundant*) de proteínas. Aproximadamente, 78,4% dessas sequências foram anotadas funcionalmente a partir do *pipeline* utilizado pelo software *Blast2GO*. As análises das sequências anotadas revelaram prováveis enzimas para síntese de terpenóides como geraniol e linalol/nerolidol. Para validação da montagem e anotação, foram realizados ensaios de qPCR para amplificação de sequências de 13 genes para controles endógenos e 4 genes de terpeno sintases. Os resultados obtidos aqui corroboram outros estudos de transcriptoma de espécies não modelo usando tecnologias de sequenciamento de alto-desempenho.

### Palavras-chave

*Lippia alba*, terpenóides, terpeno sintases, transcriptoma, sequenciamento, montagem *de novo*, anotação funcional, PCR em tempo real.

## ABSTRACT

*Lippia alba*, popularly known as erva-cidreira, is a widely distributed specie in Americas and it is found throughout Brazil. This specie has important using in popular medicine for cramping, indigestion, nausea, diarrhea, dysentery, respiratory diseases, liver disorders treatment and infectious diseases such as syphilis and gonorrhea. The leaves of *L. alba*, which are prepared by infusion or decoction and orally ingested, producing an essential oil rich in terpene compounds. These compounds are of pharmacological and industrial interest, due to their use in fragrance preparation. Interestingly, the composition of essential oils change according to different abiotic factors and genetic variations such as ploidy level. In this context, the aims of this work were to characterize the transcriptome of leaves of *L. alba* (linalool chemotype) and to search putative enzymes sequences involved in production of secondary metabolites. The transcriptome was sequenced by Miseq platform (Illumina) running pair-end libraries 300 bp. The sequencing resulted in 47,498,310 reads (23,749,155 reads for each end sequenced) of 35-308 bp, comprising 12,148,327,567 nucleotides (~12 Gb). The *de novo* assembly of transcripts was processed by Trinity software and generated 193,532 transcripts, in 128,209 uni-genes, with N50 equal to 1,187 bp. 86,122 ORFs (Open Read Frame) were obtained and submitted to BlastP algorithm, finding 75,533 sequences included in NR (Non-Redundant) protein database. Approximately 78.4% of these sequences were functionally annotated using Blast2Go pipeline. Analysis of annotated sequences revealed putative enzymes for synthesis of terpenoids such as geraniol and linalool/nerolidol. For assembly and annotation validation, qPCR assay were realized by amplification of 13 endogenous control genes and 4 terpene synthases genes. The results found here corroborate transcriptome studies in non-model organisms using high-performance sequencing technologies.

### **Keywords:**

*Lippia alba*, terpenoids, terpene synthases, transcriptome, sequencing, *De novo* assembly, functional annotation, real-time PCR.

## ÍNDICE DE FIGURAS

Figura 1.1: Espécime de <i>Lippia alba</i> cultivados na Estação Experimental de Cultivo e Manutenção de Plantas da Universidade Federal de Juiz de Fora.....	18
Figura 1.2: Organização das vias biossintéticas de terpenóides e seus compartimentos subcelulares em plantas.....	21
Figura 1.3: Compartimentalização das vias do MVA e MEP em células vegetais.....	23
Figura 3. 1: <i>Workflow</i> adotado para anotação de sequências de transcritos encontradas, após montagem de novo. ....	36
Figura 4.1: <i>Workflow</i> de procedimentos desde a extração do RNA total até a anotação de sequências de transcritos encontrados após montagem das reads geradas por sequenciamento NGS do transcriptoma de <i>L. alba</i> quimiotipo BGEN-04. ....	40
Figura 4.2: Eletroforese de RNA total extraído do quimiotipo BGEN-04 de <i>L. alba</i> em gel de agarose 1%. ....	42
Figura 4.3: Análise do RNA total de <i>Lippia alba</i> no <i>Agilent 2100 System bioanalyzer</i> destacando os picos 18S e 28S.....	42
Figura 4.4: Qualidade em <i>phred score</i> das <i>reads paired-end</i> gerados pelo sequenciamento do transcriptoma de <i>Lippia alba</i> usando a plataforma <i>Illumina Miseq</i> .....	43
Figura 4.5: Qualidade <i>phred score</i> por base dos <i>reads paired-end</i> gerados pelo sequenciamento do transcriptoma de <i>Lippia alba</i> usando a plataforma <i>Illumina Miseq</i> .....	44
Figura 4.6: Percentual de similaridade das sequências, obtidas a partir da montagem do transcriptoma de <i>Lippia alba</i> , alinhadas contra o banco de dados NR de proteínas usando o algoritmo <i>BlastP</i> local. ....	47
Figura 4.7: Classificação de sequências, obtidas pela montagem e anotação do transcriptoma de <i>Lippia alba</i> , anotadas pelo <i>Gene Ontology</i> .....	49
Figura 4.8: Distribuição das sequências, obtidas a partir da montagem e anotação do transcriptoma de <i>Lippia alba</i> , de acordo com as classes enzimáticas. ....	50
Figura 4.9: Representação das vias do ácido mevalônico (MVA) e do 2-C-metil-D-eritriol-4-fosfato (MEP). ....	55
Figura 4.10: Representação da via de síntese dos monoterpénóides.....	56
Figura 4.11: Representação da via de síntese dos diterpenóides.....	57
Figura 4.12: Representação da via dos carotenóides.....	58
Figura 4.13: Representação da via de síntese dos sesqui- e triterpenóides. ....	59

Figura 4. 14: qPCR de quatro genes de terpeno sintase para três acessos de *Lippia alba*. .....65

## ÍNDICE DE TABELAS

Tabela 4.1: Resultados obtidos com sequenciamento e montagem do transcriptoma de <i>Lippia alba</i> . .....	46
Tabela 4.2: Sequências de enzimas anotadas para a síntese de terpenóides obtidas a partir da montagem de novo do transcriptoma de <i>Lippia alba</i> usando o programa BlastP. ....	53
Tabela 4.3: Sequências usadas para desenhar primers para qPCR, afim de validar o sequenciamento e montagem do transcriptoma de <i>Lippia alba</i> . ....	61
Tabela 4.4: Sequência de primers usados e validados no presente estudo sobre anotação do transcriptoma de <i>Lippia alba</i> . ....	63

## ABREVIATURAS

**(Z,Z)-FPP** - Z,Z farnesil difosfato  
**AACT** - acetoacetil-CoA tiolase  
**ATP** - adenosina trifosfato  
**bp** - par de base (do inglês, *base-pair*)  
**cDNA** - DNA complementar  
**CDS** - sequência de DNA codificante (do inglês, *Coding DNA Sequence*)  
**CMK** - 4-(citidina 5'-difosfo)-2-C-metil-D-eritritol quinase  
**CoA** - acetil-Coezima A  
**CPP** - copalil difosfato  
**DF** - Distrito Federal  
**DHDDS** - dehidrodolicol difosfato sintase  
**DMAPP** - do inglês, *dimethylallyl pyrophosphate*  
**DNA** - ácido desoxirribonucleico (do inglês, *deoxyribonucleic acid*)  
**DOXP** - 1-deoxi-D-xilulose-5-fosfato  
**dsDNA – fitas duplas de DNA (do inglês, *double-strand DNA*)**  
**DXR** - DOXP redutoisomerase  
**DXS** - DOXP sintase  
**EC** - do inglês, *Enzyme Commission*  
**emPCR** - PCR em emulsão  
**EST** - marcador de sequência expressa (do inglês, *Expressed Sequence Tag*)  
**FPP** - farnesil difosfato  
**FPS** - farnesil difosfato sintase  
**FU** - fluorescência  
**GAP** - gliceraldeído-3-fosfato  
**Gb** - gigabase  
**GGPP** - geranylgeranyl difosfato  
**GGPS** - geranylgeranyl difosfato sintase  
**GO** - do inglês, *Gene Ontology*  
**GPP** - geranyl difosfato  
**GPS** - geranyl difosfato sintase  
**HDR** - (E)-4-hidroxi-3-metil-but-2-enil difosfato redutase  
**HDS** - (E)-4-hidroxi-3-metil-but-2-enil difosfato sintase  
**HMBPP** - 1-hidroxi-2-metil-2-(E)-butenil-4-difosfato  
**HMG-CoA** - 3-hidroxi-3-metilglutaril-CoA  
**HMGR** - 3-hidroxi-3-metilglutaril-CoA redutase  
**HT-NGS** - do inglês, *high-throughput next generation sequencing*  
**HT-Seq** - sequenciamento de alto desempenho (do inglês, *high-throughput sequencing*)  
**IDS** - isoprenil difosfato sintase  
**IPP** - isopentenil pirofosfatase  
**Kb** - kilobases  
**MCT** - 2-C-metil-D-eritritol-4-fosfato citidilil-transferase  
**MDS** - 2-C-metil-D-eritritol-2,4-ciclodifosfato sintase  
**MEP** - 2-C-metil-D-eritritol-4-fosfato  
**mg** - miligramas  
**MK** - MVA quinase  
**MPDC** - difosfato-MVA descarboxilase  
**mRNA** - RNA mensageiro  
**MVA** - ácido mevalônico

**NAD** - Nicotinamida-Adenina Dinucleotídeo  
**NDPS1** - neril difosfato sintase 1  
**ng** - nanogramas  
**NGS** - sequenciamento de nova geração (do inglês, *Next Generation Sequencing*)  
**NR** - não redundante (do inglês, *Non-Redundant*)  
**nt** - nucleotídeo  
**ORF** - quadro aberto de leitura (do inglês, *Open Read Frame*)  
**OSC** - oxidosqualeno ciclases  
**PCR** - reação em cadeia da polimerase (do inglês, *polymerase chain reaction*)  
**PHS1** - felandreno sintase 1  
**PMK** - Fosfo-MVA quinase  
**PTP** - PicoTiterPlate  
**qPCR** - **PCR quantitativo**  
**RAM** - memória de acesso aleatório (do inglês, *Random Access Memory*)  
**RNA** - ácido ribonucleico (do inglês, *ribonucleic acid*)  
**RNA-Seq** - sequenciamento do transcriptoma/RNA  
**SHC** - esqualeno-epene ciclase  
**ShSBS** - santaleno/bergamoteno sintase  
**ShZIS** - 7-epizingibereno sintase  
**SNP** - polimorfismos de nucleotídeos únicos  
**SQE** - esqualeno epoxidase  
**SQS** - esqualeno sintase  
**TPS** - terpeno sintase

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	<b>16</b>
1.1 <i>Lippia alba</i> .....	17
1.2 PRODUÇÃO DE TERPENÓIDES.....	20
1.3 SEQUENCIAMENTO DE NOVA GERAÇÃO .....	28
1.4 PCR EM TEMPO REAL (qPCR – <i>Quantitative Polymerase Chain Reaction</i> ) .....	30
<b>2 OBJETIVOS</b> .....	<b>33</b>
2.1 OBJETIVO GERAL .....	33
2.2 OBJETIVOS ESPECÍFICOS .....	33
<b>3 MATERIAL E MÉTODOS</b> .....	<b>34</b>
3.1 COLETA DAS AMOSTRAS PARA SEQUENCIAMENTO.....	34
3.2 EXTRAÇÃO DE RNA TOTAL PARA SEQUENCIAMENTO .....	34
3.3 SEQUENCIAMENTO NGS DO RNA TOTAL: RNAseq .....	34
3.4 VERIFICAÇÃO DE QUALIDADE E <i>TRIMMING</i> DE <i>READS</i> .....	35
3.5 MONTAGEM DOS TRANSCRITOS .....	35
3.6 ANOTAÇÃO FUNCIONAL DE SEQUÊNCIAS GÊNICAS .....	36
3.7 ESCOLHA DE SEQUÊNCIAS E DESENHO DE <i>PRIMERS</i> PARA ANÁLISE DE PCR QUANTITATIVO (qPCR) .....	37
3.8 EXTRAÇÃO DE RNA E SÍNTESE DE CDNA PARA qPCR .....	37
3.9 PCR QUANTITATIVO (qPCR) E ANÁLISE DOS DADOS.....	38
<b>4 RESULTADOS E DISCUSSÃO</b> .....	<b>40</b>
4.1 EXTRAÇÃO E QUALIDADE DO RNA TOTAL PARA SEQUENCIAMENTO...	41
4.2 SEQUENCIAMENTO E MONTAGEM DOS DADOS DE RNAseq.....	43
4.3 ANOTAÇÃO FUNCIONAL.....	47
4.3.1 Obtenção de vias metabólicas .....	50
4.3.2 Terpeno sintases .....	51
4.4 OBTENÇÃO DE <i>PRIMERS</i> PARA ESTUDOS DE qPCR E VALIDAÇÃO DA MONTAGEM .....	60
4.5 AVALIAÇÃO DO PERFIL DE EXPRESSÃO DE TERPENO SINTASES.....	64
<b>5 CONCLUSÃO</b> .....	<b>66</b>
<b>6 REFERÊNCIAS BIBLIOGRÁFICAS</b> .....	<b>67</b>
<b>ANEXO: CURVAS DE DISSOCIAÇÃO DOS <i>PRIMERS</i></b> .....	<b>77</b>

# 1 INTRODUÇÃO

As plantas de um modo geral sintetizam compostos químicos a partir dos nutrientes, água e luz obtidos do meio em que se localizam. Sua rede metabólica possui maior número de vias e algumas mais complexas em relação a de outros organismos, como por exemplo os animais, uma vez que as plantas são autotróficas. Além da produção de metabólitos primários que estão envolvidos nos processos vitais de seu ciclo (crescimento e desenvolvimento), as plantas produzem também metabólitos secundários relacionados à interação com o ambiente, desde a atração de polinizadores até a defesa contra predadores e patógenos (Aharoni e Galili, 2011). Portanto, os metabólitos secundários possuem uma importante função ecológica.

Farmacologicamente, estes metabólitos possuem um importante potencial medicinal, sendo considerados princípios ativos de várias plantas utilizadas por culturas tradicionais no tratamento de inflamações, espasmos, dores, e problemas diuréticos (Savitharamma, Rao e Suhrulatha, 2011). Além disso, são de grande importância econômica, já que são usados na produção de perfumes, sabores para alimentos e outras propriedades industriais (Fisher e Phillips, 2008).

No entanto, as aplicações comerciais dos metabólitos especializados possuem limitações, devido à baixa quantidade produzida pelos vegetais, tempo requerido para o crescimento da planta, problemas durante a colheita e extração dos compostos. Portanto, a identificação e caracterização de genes de enzimas relacionadas ao metabolismo de compostos especializados, bem como o uso de técnicas de clonagem e biologia sintética, são estratégias interessantes que possibilitariam a comercialização dessas moléculas (Facchini *et al.*, 2012).

A exploração do transcriptoma, a partir das tecnologias de sequenciamento de nova geração (*Next Generation Sequencing* - NGS), tornou-se um importante método para a descoberta de genes codificantes de enzimas responsáveis pela síntese de compostos secundários (Facchini *et al.*, 2012; Wang *et al.*, 2013), especialmente sustentável para espécies não modelo, as quais apresentam sequências genômicas desconhecidas (Wang *et al.*, 2013).

## 1.1 *Lippia alba*

O gênero *Lippia* é um dos 41 gêneros de ervas, arbustos e árvores pertencentes à família Verbenaceae. Este gênero compreende aproximadamente 200 espécies distribuídas principalmente na África tropical, Américas do Sul e Central, Índia e Austrália (Pascual *et al.*, 2001; Ombito *et al.*, 2014). O Brasil, uma das regiões de maior biodiversidade do gênero, possui aproximadamente 75% das espécies conhecidas que, em sua maioria, são endêmicas da “Cadeia do Espinhaço”, nos estados de Minas Gerais e Goiás (Viccini *et al.*, 2005).

O uso de plantas do gênero *Lippia* na medicina popular, geralmente envolvem o uso de folhas, flores e partes aéreas preparadas sob a forma de infusão ou decocção e administrados por via oral. As principais indicações são para o tratamento de doenças dos sistemas respiratório e digestivo, além do tratamento de infecções (Silva *et al.*, 2015).

Dentre as plantas do gênero *Lippia*, a espécie *L. alba* (Figura 1.1) é uma das mais estudadas. A espécie é popularmente conhecida por diversos nomes, como erva cidreira, falsa melissa, carmelitana, salva do Brasil, salva, salva limão, melissa, alecrim do campo e salva brava (Siani *et al.*, 2002; Aguiar *et al.*, 2008; Guedes, 2011).

Do ponto de vista farmacológico, alguns trabalhos apresentam a importância medicinal de *L. alba* (Hennebelle *et al.*, 2008; López, Stashenko e Fuentes, 2011; Blanco *et al.*, 2013). O uso desta espécie na medicina tradicional é indicado para tratamento de cólicas, indigestão, náuseas, espasmos, diarreia, disenteria, doenças respiratórias, problemas hepáticos e no tratamento de sífilis e gonorreia (Heinzman e Barros, 2007; Tavares Momenté e Nascimento, 2011).



Figura 1.1: Espécime de *Lippia alba* cultivado na Estação Experimental de Cultivo e Manutenção de Plantas da Universidade Federal de Juiz de Fora.

Machado *et al.* (2014) demonstrou que os óleos essenciais de *L. alba* possuem importante atividade antimicrobiana sobre bactérias Gram positivas e Gram negativas, sob diferentes concentrações. Estudos mostram que compostos produzidos a partir do metabolismo secundário de *L. alba* apresentam efeitos sedativos, dependentes da dose, no sistema nervoso central, incluindo hipotermia, propriedades anticonvulsivas e atividade antioxidante (Neto *et al.*, 2009; Chies *et al.*, 2013). A partir da análise das frações hexanólica e etanólica dos metabólitos voláteis de 16 quimiotipos de *Lippia* spp, Singulani *et al.* (2012) concluiu-se que estas plantas representam uma importante fonte de substâncias antioxidantes, as quais desempenham um papel chave na prevenção de doenças neurodegenerativas, cardiovasculares e câncer.

De acordo com o componente volátil majoritário do óleo essencial, essas plantas podem ser separadas em “variedades químicas”, denominadas quimiotipos (Pascual *et al.*, 2001; Tavares *et al.*, 2005; Teles *et al.*, 2012). Os principais quimiotipos encontrados na espécie são o citral, citral-mirceno, citral-limoneno, carvona-limoneno e linalol (Jannuzzi *et al.*, 2011). Os quimiotipos citral-mirceno e citral-limoneno são bastante utilizados no Brasil na forma de chá de folhas frescas, as quais possuem ações calmante, espasmódica, analgésica, sedativa e ansiolítica (Jannuzzi *et al.*, 2011).

Interessantemente, no Brasil *L. alba* tem sido indicada como uma importante fonte alternativa de linalol, o qual é muito utilizado nas indústrias de cosméticos e alimentícia. Esta fonte alternativa é importante, uma vez que o pau-rosa (*Aniba roseodora*), a maior fonte deste composto está ameaçada de extinção (Vieira, 1999; Siani *et al.*, 2002). A boa produtividade (89±9%) e a qualidade do linalol produzido por *L. alba* são pontos a serem considerados e explorados, uma vez que os quimiotipos que produzem linalol como principal componente do óleo essencial possuem baixos níveis de outros terpenóides (Forero-Peñuela *et al.*, 2013).

Além dos compostos citados anteriormente, os diferentes quimiotipos de *L. alba* possuem vários outros constituintes em seus óleos essenciais, variando entre 29 e 42 tipos de compostos. Alguns são encontrados apenas em determinados quimiotipos, constituindo um total aproximado de 90% da composição do óleo destas plantas (Tavares *et al.*, 2005; Hennebelle *et al.*, 2008).

Análises por cromatografia de fase gasosa acoplada a espectrometria de massa dos óleos essenciais dos quimiotipos citral, carvona e linalol oriundos de diferentes regiões, cultivados sob mesmas condições, indicaram que a composição do mesmo manteve-se inalterada, sugerindo que os diferentes perfis de óleos essenciais são devidos principalmente à variação genética dos diferentes quimiotipos (Tavares *et al.*, 2005). A variação genética entre plantas de *L. alba* tem sido confirmada a partir de análises cromossômicas, estudos de triagem molecular e conteúdo de DNA (Pierre *et al.*, 2011; Reis *et al.*, 2014).

Viccini *et al.* (2014) demonstraram que as diferenças entre os quimiotipos de *L. alba* podem estar, de fato, relacionadas às variações genéticas encontradas, a respeito das distintas ploidias. Considerando plantas com número cromossômico  $2n=30$  como indivíduos diploides, foram observados espécimes triploides ( $2n=45$ ) e tetraploides ( $2n=60$ ). Notadamente, todas as plantas diploides e tetraploides possuem como principal componente de seus óleos essenciais o citral. Por outro lado, triploides produzem linalol como principal terpenóide, sugerindo uma possível relação entre este composto e o conteúdo de DNA (Viccini *et al.*, 2014)

Diante do amplo uso medicinal de *L. alba* e das lacunas que ainda existem com relação aos fatores determinantes na produção de óleos essenciais, fica evidente a necessidade de estudos genéticos adicionais.

## 1.2 PRODUÇÃO DE TERPENÓIDES

Os óleos essenciais compreendem misturas complexas constituídas por moléculas voláteis, produzidas a partir do metabolismo das plantas. Estas moléculas voláteis denominadas de terpenóides (também conhecidas como isoprenóides) constituem uma extensa família de produtos naturais que inclui mais de 40.000 compostos, oriundos tanto do metabolismo primário quanto do metabolismo secundário (Goto *et al.*, 2010). Estas moléculas podem ser obtidas de raízes e rizomas (*e.g.*, gengibre), folhas (*e.g.*, menta, erva-cidreira e orégano), casca e ramos (*e.g.*, capim-cidreira, cânfora), flores (*e.g.*, jasmim, rosa, violeta e lavanda) e de frutos e sementes (*e.g.*, laranja, limão, pimenta, noz-moscada) (Fornari *et al.*, 2012).

Os terpenóides são todos biossintetizados a partir de unidades básicas de isopreno, mais precisamente, isopentenil difosfato (IPP) e dimetilalil difosfato (DMAPP), os quais são originados de dois caminhos biossintéticos alternativos localizados em compartimentos subcelulares diferentes. O caminho citosólico do ácido mevalônico (MVA) produz o IPP a partir de acetil-Coenzima A (CoA), enquanto que a via plastidial do 2-C-metil-D-eritriol-4-fosfato (MEP) leva a formação de IPP e DMAPP a partir de piruvato e gliceraldeído-3-fosfato (GAP) (Figura 1.2) (Nagegowda, 2010). Estas duas vias operam de maneira simultânea e a separação dentro de compartimentos celulares distintos não é absoluta, o que permite trocas de metabólitos (Hemmerlin *et al.*, 2003).

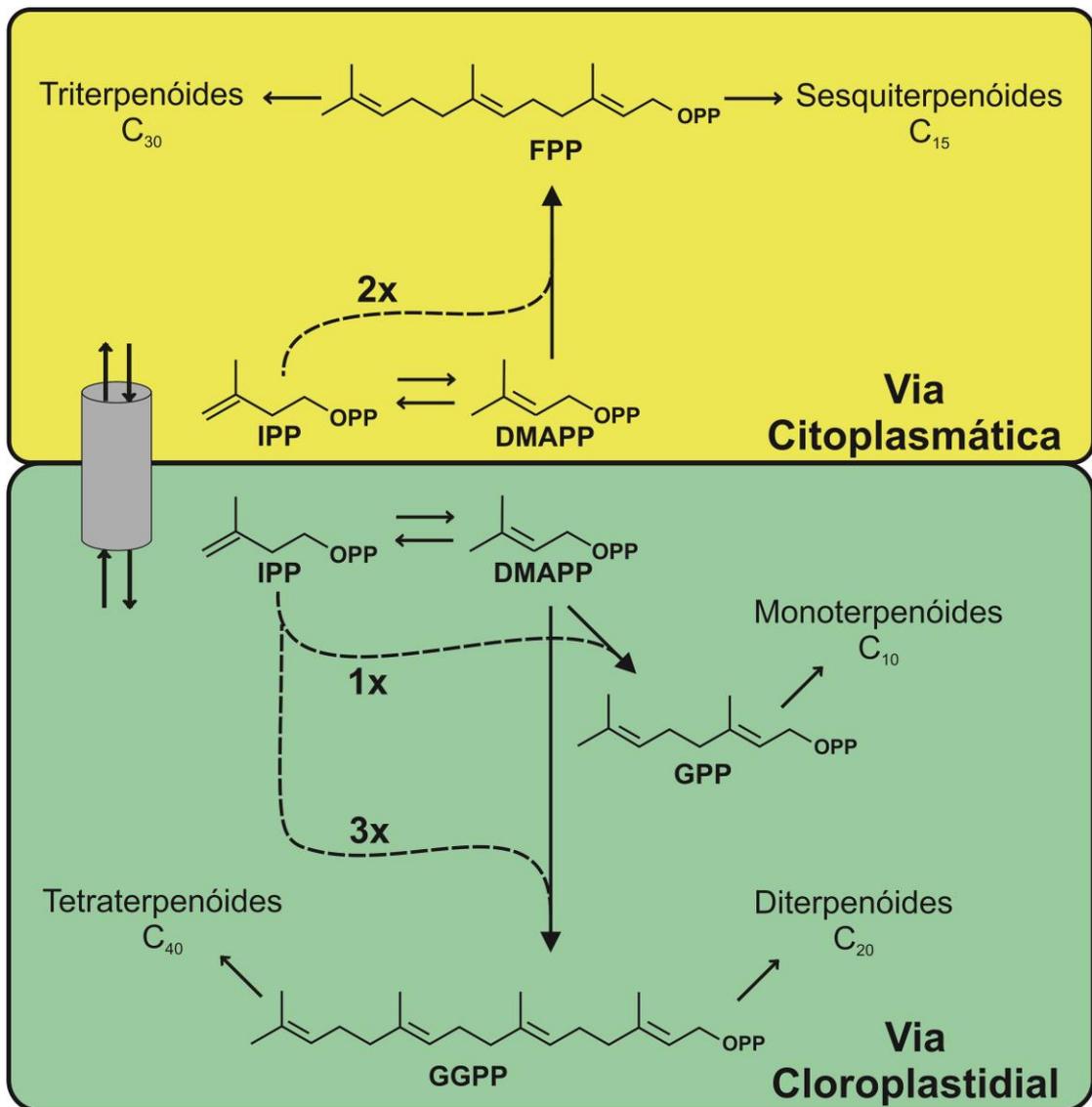


Figura 1.2: Organização das vias biossintéticas de terpenóides e seus compartimentos subcelulares em plantas. DMAPP, dimetilalil difosfato. IPP, isopentenil difosfato. (Adaptado de Bohlmann, Meyer-Gauen, Croteau 1998).

A reação inicial da via do MVA (Figura 1.3.A) é caracterizada pela enzima acetil-CoA-C-acetil-transferase, também conhecida como acetoacetil-CoA tiolase (AACT; EC 2.3.1.9) (Vranová, Coman e Grussem, 2013). A AACT é uma enzima responsável pela condensação de duas CoA para acetoacetil-CoA em uma reação reversível. No segundo passo, este produto é convertido para 3-hidroxi-3-metilglutaril-CoA (HMG-CoA), reação promovida pela enzima HMG-sintase (HMGS; EC 2.3.3.10).

O HMG-CoA é convertido para MVA em dois passos de redução, cada um destes requer a presença da molécula de Nicotinamida-Adenina Dinucleotídeo (NAD) em sua forma

reduzida NADH, a qual servirá como o equivalente redutor. Essas reações de oxido-redução são catalisadas pela enzima 3-hidroxi-3-metilglutaril-CoA redutase (HMGR; EC 1.1.1.34). Duas reações sucessivas de fosforilação são catalisadas pelas enzimas MVA quinase (MK; EC 2.7.1.36) e Fosfo-MVA quinase (PMK; EC 2.7.4.2), sendo o MVA fosforilado para MVA 5-difosfato.

O último passo da biossíntese do IPP é uma reação de descarboxilação dependente de adenosina trifosfato (ATP) do MVA 5-difosfato, reação catalisada pela difosfato-MVA descarboxilase (MPDC; 4.1.1.33). Após sua produção, o IPP sofre isomerização formando o DMAPP para que, então, ocorra a condensação de terpenóides maiores (Vranová, Coman e Gruissem, 2013).

A rota do MEP (Figura 1.3.B), descrita como alternativa a via do MVA, consiste de sete passos enzimáticos envolvendo a formação de IPP e DMAPP a partir das moléculas de piruvato e gliceraldeído-3-fosfato (GAP). O primeiro passo desta rota é a condensação de piruvato e GAP para formar 1-deoxi-D-xilulose-5-fosfato (DOXP), reação catalisada pela enzima DOXP sintase (DXS; EC 2.2.1.7). Em seguida, o DOXP é transformado em MEP por uma reação de redução/isomerização realizada pela enzima DOXP redutoisomerase (DXR; EC 1.1.1.267), também denominada de MEP sintase.

O MEP é convertido em 1-hidroxi-2-metil-2-(E)-butenil-4-difosfato (HMBPP) pela atividade consecutiva das enzimas 2-C-metil-D-eritritol-4-fosfato citidilil-transferase (MCT; EC 2.7.7.60), 4-(citidina 5'-difosfo)-2-C-metil-D-eritritol quinase (CMK; EC 2.7.1.148), 2-C-metil-D-eritritol 2,4-ciclodifosfato sintase (MDS; EC 4.6.1.12) e (E)-4-hidroxi-3-metil-but-2-enil difosfato sintase (HDS; EC 1.17.7.1). O passo final é a conversão do HMBPP para IPP e DMAPP pela atividade da enzima (E)-4-hidroxi-3-metil-but-2-enil difosfato redutase (HDR; EC 1.17.1.2) (Dudareva, Klempien e Kaplan, 2013).

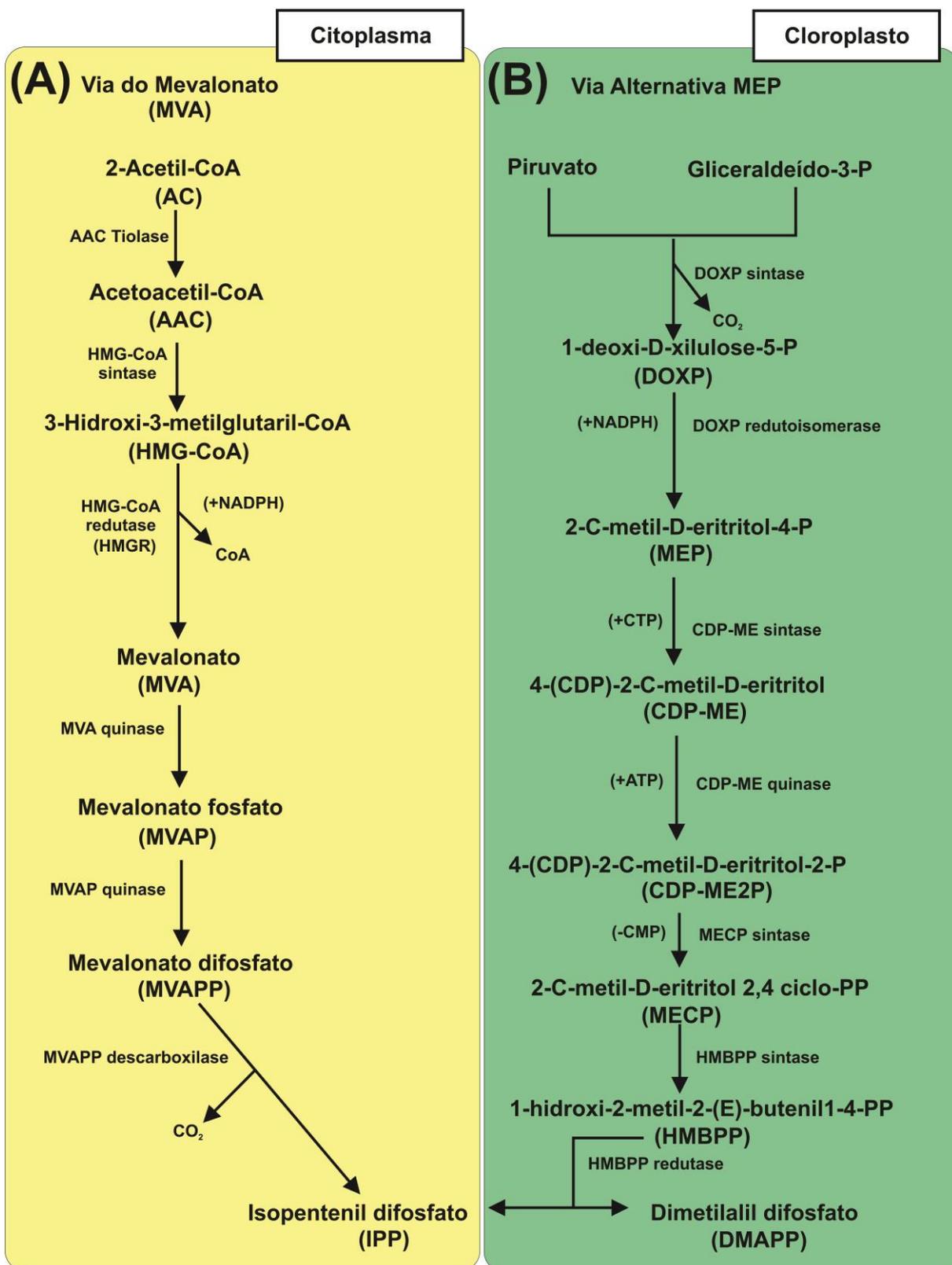


Figura 1.3: Compartimentalização das vias do MVA e MEP em células vegetais. A produção de IPP e DMAPP resultam de um conjunto de reações desencadeadas por enzimas compartmentalizadas e específicas de cada organela (adaptado de Vranová, Coman e Gruissem, 2013).

As unidades básicas de isoprenos, descritas anteriormente, podem se associar de diferentes formas gerando esqueletos carbônicos múltiplos de cinco, por uma reação de condensação. Desta maneira, podemos classificar os terpenóides de acordo com suas cadeias em (Rodríguez-Concepción, 2014): (i) monoterpenóides, os quais apresentam 10 carbonos e derivam da combinação entre duas unidades ou moléculas de isopreno; (ii) sesquiterpenos, moléculas com 15 átomos de carbono em sua estrutura, formados pela associação de três moléculas de isopreno; (iii) diterpenóides, derivados da combinação entre quatro unidades de isopreno, apresentando 20 átomos de carbono em sua cadeia; (iv) triterpenóides, os quais possuem 30 átomos de carbono gerados a partir da união de duas cadeias com 15 carbonos, sendo incluídos nesta classe os brassinosteroides, fitoesteroides de membrana e o ácido oleonólico; (v) tetraterpenóides, derivados de oito moléculas de isopreno, são moléculas com 40 átomos de carbono, constituem os pigmentos acessórios carotenóides do sistema fotossintético; (vi) politerpenóides, que apresentam mais de oito unidades de isopreno.

Os isoprenóides maiores, originados das associações anteriores, são produzidos a partir da atividade de enzimas genericamente denominadas de preniltransferases. Estas enzimas, também conhecidas como isoprenil difosfato sintase (IDS; EC 2.5.1.1), catalisam a condensação de moléculas de IPP com DMAPP (Muhlemann, Klempien e Dudareva, 2014). Elas são divididas em duas classes: *Cis* ou (*Z*)-preniltransferases, as quais geram moléculas contendo apenas duplas ligações do tipo *cis*, e *trans* ou (*E*)-preniltransferases que produzem moléculas com duplas ligações do tipo *trans* em cada unidade de IPP adicionada à cadeia.

Apenas um número limitado de membros da família das *cis*-preniltransferases tem sido estudado em detalhes. Estas enzimas geram produtos que são usados para a síntese de terpenóides de cadeia longa, como os dolicois, que estão ligados em conformação *cis*, exemplo de uma enzima desta família é a dehidrodolicol difosfato sintase (DHDDS). Interessantemente, uma (*Z,Z*)-FPP sintase foi identificada em tomate selvagem (*Solanum habrochaites*), a qual produz (*Z,Z*)-FPP a partir de IPP e DMAPP juntamente com outras duas enzimas, santaleno/bergamoteno sintase (ShSBS) e 7-epizingibereno sintase (ShZIS) que utilizam o substrato *cis*óide para a formação de sesquiterpenos (Gutensohn *et al.*, 2014).

Do mesmo modo, outra *cis*-preniltransferase, neril difosfato sintase 1 (NDPS1), a qual catalisa a formação de neril difosfato (NPP) utilizando como substrato o IPP e o DMAPP, também foi isolada de tomate (*Solanum lycopersicum*) (Gutensohn *et al.*, 2014). Em paralelo, Gutensohn e colaboradores demonstraram que a enzima felandreno sintase 1 (PHS1) também

é capaz de produzir monoterpénóides. Estes dados indicam que o papel das (Z)-prenil difosfato sintase tem sido subestimado (Gutensohn *et al.*, 2014).

As *trans*-preniltransferases constituem uma extensa família de proteínas que inclui vários membros de enzimas, as quais estão envolvidas com a produção dos principais percussores para as enzimas especializadas na produção de terpenóides, dentro desta classe estão (Tholl e Lee, 2011): (i) geranyl difosfato sintase (GPS; EC 2.5.1.1) que sintetiza o geranyl difosfato (GPP), o qual é um precursor de 10 carbonos para a geração de monoterpénóides; (ii) geranylgeranyl difosfato sintase (GGPS; EC 2.5.1.29), enzimas responsáveis pela formação do geranylgeranyl difosfato (GGPP), sendo este uma molécula de 20 carbonos precursora para síntese de diterpenóides; (iii) farnesil difosfato sintase (FPS; EC 2.5.1.10) que catalisam a produção farnesil difosfato (FPP), uma com 15 átomos de carbono e primordial na síntese de sesquiterpenos.

A biossíntese dos terpenóides, os quais são compostos cíclicos em sua maioria, envolve ainda outro grupo de enzimas denominadas terpeno sintases (TPS) ou ciclases. Estas enzimas catalisam reações de ciclização, que podem ocorrer acompanhadas de reações de oxido-redução, resultando na formação de terpenos que compõem óleos essenciais (Bohlmann, Meyer-Gauen e Croteau, 1998). Algumas exceções de terpeno sintases produzem compostos acíclicos, como as monoterpeno sintases que sintetizam micerno, geraniol e linalol (Bohlmann, Meyer-Gauen e Croteau, 1998).

De um modo geral, as terpeno sintases apresentam um tamanho entre 550 até 850 resíduos de aminoácidos, o que está de acordo com o peso molecular observado experimentalmente de 50-100 kDa (quilodaltons) (Bohlmann, Meyer-Gauen e Croteau, 1998). Elas são encontradas em compartimentos celulares específicos, por exemplo, terpeno sintases para monoterpénóides (C10), diterpenóides (C20) e tetraterpenóides/carotenóides (C40) estão localizadas nos plastídios, enquanto que terpeno sintases para sesquiterpenos (C15) e triterpenos (C30) estão presentes no citoplasma (Lange, 2014).

Todas as monoterpeno sintases apresentam em comum o mecanismo de reação que catalisa o produto terpenóide final (Mcgarvey e Croteau, 1995; Kampranis *et al.*, 2007). A catálise inicia com o processo de isomerização do GPP, promovendo a quebra da dupla ligação entre os átomos de carbono C<sub>2</sub> e C<sub>3</sub> que impede a ciclização da molécula. A seguir, o GPP sofre ionização dependente de íons metálicos divalentes, formando produtos catiônicos intermediários, por exemplo, o cátion geranyl, linalil difosfato (nas formas *trans* e *cis*), cátion linalil, e o cátion  $\alpha$ -terpenil. Estes produtos intermediários sofrem uma série de ciclizações,

deslocamento de hidretos ou outros rearranjos moleculares até a formação de um composto estável. Os monoterpenóides acíclicos como linalol, mirceno e (*E*)- $\beta$ -ocimene podem ser catalisados tanto pela via do cátion geranil quanto pela via do cátion linalil, enquanto que o geraniol provém da via do cátion geranil por uma reação de hidratação (Degenhardt, Köllner e Gershenzon, 2009).

Os sesquiterpenos compreendem um extenso grupo de moléculas isoprenóides, apresentando mais de 7.000 compostos, sendo catalisados a partir da atividade de enzimas denominadas de sesquiterpenos sintases. O número maior de sesquiterpenos deve-se ao fato do precursor destas moléculas apresentarem 5 carbonos a mais, o que possibilita maior número de ciclizações (Bohlmann, Meyer-Gauen e Croteau, 1998).

A reação catalisada pelas enzimas sesquiterpenos sintases é conservada, iniciando com processo de ionização do FPP que resulta no cátion farnesil. Esta molécula sofre um conjunto de ciclizações, sendo algumas delas precedidas por um mecanismo de isomerização, formando um composto intermediário denominado nerolidil. Seguidos passos de ciclização, deprotonação, mudanças de hidretos e grupamentos metil, além de outros rearranjos, são necessários para a formação de diferentes produtos estáveis (Degenhardt, Köllner e Gershenzon, 2009).

Os diterpenóides são moléculas derivadas do (*E,E,E*)-geranilgeranil difosfato (GGPP), o qual forma um grande grupo de mais de 12.000 produtos naturais. Notavelmente, uma significativa fração das moléculas diterpenóides são policíclicas (aproximadamente 7.000 compostos), originando de duas ciclizações e/ou reações de rearranjo molecular independentemente do número de anéis carbônicos finais (Peters, 2010). As diterpeno sintase podem ser divididas em dois grupos: (i) diterpeno sintases bifuncionais, as quais convertem o GGPP para uma molécula intermediária denominada copalil difosfato (CPP) e consecutivamente a conversão deste para diterpenóides estáveis; (ii) diterpeno sintases monofuncionais que catalisam diterpenóides a partir do CPP ou diretamente do GGPP, sem a produção de moléculas intermediárias (Chen *et al.*, 2011).

Os triterpenóides e esteróides são isoprenóides sintetizados na via do MVA, a partir de um composto intermediário denominado de 2,3-oxidosqualeno que é produzido pela ação consecutiva de duas enzimas. A esqualeno sintase (SQS; EC 2.5.1.21) converte duas moléculas de FPP para esqualeno, consumindo NADPH. A seguir, o esqualeno resultante da primeira catalise é usado como substrato para as enzimas esqualeno monooxigenase, também chamada de esqualeno epoxidase (SQE; EC 1.14.13.132), gerando 2,3-oxidoesqualeno a partir

do consumo de NADPH e oxigênio molecular. A ciclização de 2,3-oxidosqualeno é realizada por enzimas conhecidas como oxidosqualeno ciclases (OSC), gerando esqueletos carbônicos que serão utilizados para a geração de esteróides e triterpenóides.

A variação na ciclização, nos rearranjos moleculares e no dobramento do substrato inicial, são passos críticos para a diversidade de esqueletos carbônicos originados. Estes podem apresentar a conformação Chair-Chair-Chair (C-C-C), a qual dá origem aos triterpenóides, e a conformação Chair-Boat-Chair (CBC) que dá origem aos esteróides (Thimmappa *et al.*, 2014). Em bactérias, o esqualeno é ciclizado para o anel pentacíclico por uma enzima denominada esqualeno-anel ciclase (SHC; EC 5.4.99.17), o qual sofrerá rearranjos moleculares complexos até originar um esteróide estável (Abe, 2007).

Os carotenóides e tetraterpenos (C<sub>40</sub>) são moléculas isoprenóides produzidas pela união de dois GGPP, produzindo um composto denominado fitoeno, de maneira similar a formação do esqualeno. A reação de produção de fitoeno é catalisada pela enzima fitoeno sintase (PSY; EC 2.5.1.32), a qual desempenha um mecanismo muito similar ao desempenhado pela SQS. O fitoeno é dessaturado para formar uma cadeia de anéis-cromóforos que compõe o backbone dos carotenóides das plantas e determina as propriedades físicas e biológicas. A dessaturação e isomerização do fitoeno resultam no licopeno, o qual é um carotenóide vermelho. A ciclização desta molécula resulta na produção de carotenóides com diferentes anéis, os quais poderão sofrer modificações gerando a grande diversidade de carotenos encontrada em plantas (Ruiz-Sola e Rodríguez-Concepción, 2012).

Baseado no mecanismo de catalise e produtos formados, as terpeno sintases podem ser classificadas em dois grupos: classe I e classe II (Tholl, 2006). A reação enzimática catalisada pelas TPS classe I inicia com a ionização do substrato prenil difosfato pela remoção do grupamento difosfato da molécula de forma dependente de um cátion divalente. Em contraste, as TPS classe II catalisam a ionização de seu substrato a partir da adição de um próton ao anel epóxido ou via protonação da dupla ligação entre os carbonos C14 e C15 do GGPP (Gao, Honzatko e Peters, 2012; Tholl, 2015).

As TPS classe I possuem um domínio  $\alpha$  funcionalmente ativo, o qual possui um motivo altamente conservado rico em aspartato, **DDXXD**, e outro motivo menos conservado, **NSE/DTE**, que estão localizados de maneira oposta na entrada do sítio catalítico (Christianson, 2006). Estes motivos são responsáveis por auxiliar no posicionamento do substrato difosfato, coordenação de íons divalentes e moléculas de água, além de estabilizarem o sítio ativo (Christianson, 2006).

As TPS classe II dispõem de dois domínios, um domínio  $\beta$  e um domínio  $\gamma$ , sendo que o domínio  $\beta$  funcionalmente ativo carrega o motivo conservado **DXDD**, similar ao **DDXXD** de TPS classe I, o qual é indispensável para o processo de protonação do substrato ao invés da clivagem como ocorre com algumas de classe I (Christianson, 2006; Cao *et al.*, 2010). Algumas monoterpene sintases e diterpene sintases apresentam tipicamente um motivo **RR(X)<sub>8</sub>W**, o qual está relacionado com o processo de ciclização do substrato isoprenil. Este domínio delimita o peptídeo de endereçamento plastidial na região N-terminal (Martin *et al.*, 2010).

As relações entre as sequências de aminoácidos das enzimas terpeno sintases de plantas, permite subdividi-las em oito subfamílias, de TPSa até TPSh, cada uma apresentando um mínimo de 40% de identidade entre seus membros. O estudo filogenético entre as subfamílias de TPS, demonstrou que as TPSa (subdividida em TPSa-1, tipicamente de dicotiledôneas, e TPSa-2, presente em monocotiledôneas), TPSb e TPSg formam clados específicos de angiospermas, sendo que as TPSg estão mais próximas evolutivamente das TPSb. As TPSc, TPSe/f e TPSh formam um novo clado entre elas, sendo que as TPSf derivam das TPSe. Enquanto que as TPSd, as quais estão subdivididas em TPSd-1, TPSd-2 e TPSd-3, formam um clado isolado de enzimas tipicamente de gimnospermas (Chen *et al.*, 2011).

A complexidade na biossíntese dos terpenóides, seu potencial informativo sobre a filogenia, além da importância econômica e medicinal, justificam um amplo estudo sobre estas vias metabólicas.

### 1.3 SEQUENCIAMENTO DE NOVA GERAÇÃO

O método de sequenciamento Sanger automatizado, originado a partir de avanços na eletroforese capilar, permitiu o término do primeiro genoma humano marcando o início da era do sequenciamento moderno. Os avanços posteriores resultaram na invenção e desenvolvimento de novas estratégias de sequenciamento de alto desempenho (do inglês, *high-throughput sequencing* – HT-Seq), sendo as novas plataformas de sequenciamento denominadas de *high-throughput next generation sequencing* (HT-NGS). As características desta nova estratégia têm permitido técnicas mais rápidas, com menores custos e com maior acurácia (Mardis, 2008; Pareek, Smoczynski e Tretyn, 2011).

O progresso dessas tecnologias tem permitido não apenas a análise das variações estruturais e sequenciais em genomas, mas também, variações epigenômicas, modificações da

cromatina, genômica do câncer e estudos de transcriptomas, como a detecção e quantificação de transcritos, bem como outros elementos da genômica funcional (Soon, Hariharan e Snyder, 2013). Interessantemente, técnicas como o *microarray* vêm sendo substituídas pelos sequenciamentos de nova geração, uma vez que estes podem detectar genes raros mesmo sem conhecimentos prévios e ainda detectar variantes de *splicing* alternativo (Metzker, 2010).

As plataformas de HT-NGS diferenciam-se entre si em configurações de engenharia e química de sequenciamento. No entanto todas compartilham uma característica, um paradigma técnico, que é o processamento paralelo massivo de fragmentos de DNA ou RNA em uma única *flow cell*. Essa é a principal diferença destas plataformas em relação ao sequenciamento Sanger, o qual é baseado na separação por eletroforese dos produtos da terminação de cadeia precoce por inserção de bases modificadas, em sequenciamentos individuais (Voelkerding, Dames e Durtschi, 2009).

Dentre as plataformas e métodos de sequenciamento HT-NGS estão (Liu *et al.*, 2012; Myllykangas, Buenrostro e Ji, 2012): (i) SOLiD desenvolvido pela Applied Biosystems (agora parte da *Life Technologies*); (ii) Solexa/Illumina; (iii) 454 FLX desenvolvido pela *Life Sciences* (Roche); (iv) Ion Torrent desenvolvido pela *Ion Torrent Systems* (agora parte da *Life technologies*); (v) Heliscope desenvolvido pela *Helicos Biosciences*; (vi) PacBio RS desenvolvido pela *Pacific BioSciences*. Além destas, outras tecnologias que parecem se destacar nos últimos anos são o Polonator G.007, desenvolvido por George M. Church, em Harvard, como uma tecnologia aberta (combina bibliotecas *paired-tag* com PCR em emulsão e o sequenciamento baseado em ligação da plataforma SOLiD proporcionando baixo custo e alta acurácia), e o Oxford Nanopore, desenvolvido pela *Oxford Nanopore Technologies Limited* (Pareek, Smoczynski e Tretyn, 2011)

De maneira geral, as plataformas HT-NGS envolvem três etapas primordiais e críticas, as quais são (Myllykangas, Buenrostro e Ji, 2012): (i) preparação das amostras (ou bibliotecas), etapa na qual as amostras são randomicamente fragmentadas em tamanhos específicos, envolvendo também a ligação de pequenas sequências definidas, denominadas de adaptadores, aos 5' e 3' terminais dos fragmentos gerados; (ii) imobilização, neste passo as bibliotecas preparadas na etapa anterior são ancoradas em um suporte sólido, no qual ocorrerão reações de sequenciamento de forma simultânea, proporcionando o processamento de bilhões de sequências; (iii) sequenciamento, nesta terceira etapa ocorre a amplificação massiva das bibliotecas por reações de PCR, formando *clusters* de cópias destas bibliotecas

que serão detectados a partir de sinais luminosos (fluorescentes) ou químicos (detecção por pH).

Comparada à montagem por referência, a montagem *de novo* também apresenta muitas vantagens, destacando-se três (Martin e Wang, 2011): (i) os dados de transcriptoma de organismos não modelo podem ser reconstruídos sem a necessidade de uma referência, permitindo a detecção de sequências desconhecidas; (ii) a montagem não depende do alinhamento perfeito dos *reads* em sítios de *splicing*, como na montagem baseada em referência, portanto os *introns* longos não são problemas nesta estratégia; (iii) além disso, transcritos similares providos de rearranjos cromossômicos ou transcritos oriundos de mecanismos de *splicing-trans* podem ser detectados.

Existem vários *softwares* para cada estratégia. Para a montagem por referência existe a plataforma *web Galaxy* (Goecks, Nekrutenko e Taylor, 2010), que permite a construção de um *workflow* interativo para montagem tanto de genomas quanto de transcritos. Além do *Galaxy*, os programas de alinhamento de sequências curtas Blat, TopHat, GSNAP, ApliceMap e MapSplice podem também serem usados para a estratégia de montagem por referência (Martin e Wang, 2011). Com relação a montagem *de novo*, os *softwares* mais utilizados para dados providos de *reads* curtos são SOAPdenovo-Trans, Velvet/Oases, Trans-ABYSS e Trinity (Chen *et al.*, 2011).

A escolha por uma das duas abordagens deve considerar o objetivo pretendido, os custos e o tempo demandado. Desta forma, é importante determinar o número e tipos de réplicas que serão necessários e como os dados serão processados e manipulados (Strickler, Bombarely e Mueller, 2012).

#### **1.4 PCR EM TEMPO REAL (qPCR – *Quantitative Polymerase Chain Reaction*)**

O estudo da mudança de expressão gênica promove o entendimento sobre a rede regulatória do gene. Neste contexto, a técnica de qPCR (também conhecida como qRT-PCR) é comumente usada para análises quantitativas, devido sua sensibilidade, especificidade e reprodutibilidade (Lee *et al.*, 2015). A qPCR pode ser considerada um refinamento da PCR original, o qual teve início na década de 90 a partir da exploração da atividade exonucleásica da *taq*-polimerase e da capacidade quantitativa da PCR utilizando o brometo de etídio (Holland *et al.*, 1991; Higuchi *et al.*, 1992).

De forma similar à PCR convencional, a qPCR usa *taq*-polimerase, tampão, dNTPs e *primers* para a amplificação de pequenos fragmentos (80-150 bp) de cDNA (DNA complementar). A diferença em relação ao método convencional é que para cada ciclo de amplificação um sinal fluorescente é emitido, por sondas e corantes, e monitorado por um termociclador computadorizado, permitindo análise em tempo real do produto amplificado (Thornton e Basu, 2011). Os principais produtos químicos para detecção do fragmento amplificado são baseados em corantes intercalantes ao DNA, exemplo o *SybrGreen*, e na hidrólise de sondas fluorescentes, exemplo as sondas *TaqMan* (Buh Gašparič *et al.*, 2010).

Um corante muito usado na técnica de qPCR é o *SybrGreen*, o qual se liga de forma inespecífica a fitas duplas de DNA (dsDNA – do inglês, *double-strand DNA*) geradas durante a amplificação, gerando fluorescência. A grande vantagem do *SybrGreen* é que pode ser usado com qualquer par de *primer* para qualquer alvo, sendo, seu uso menos custoso que sondas (Giulietti *et al.* 2001). No entanto, a especificidade é diminuída devido ao risco de amplificar produtos inespecíficos e dímeros de *primer*. Assim, uma forma de contornar esse problema é analisar as curvas de dissociação (*melting*) pela temperatura de fusão (*melting temperature-Tm*), que é definida como a temperatura onde 50% dos produtos amplificados estão dissociados em fita simples. De maneira geral, as curvas de dissociação devem apresentar um único pico, garantindo que uma amplificação inespecífica não ocorreu (Giulietti *et al.*, 2001; Buh Gašparič *et al.*, 2010).

Outra forma de gerar a fluorescência é pelo uso de sondas específicas para uma região interna da sequência amplificada, como no caso as sondas *TaqMan*. Em solução, o sinal fluorescente do fluoróforo é suprimido pelo *quencher*. Porém durante a fase de extensão a sonda ligada é degradada pela atividade 5'-3' exonucleásica da *taq* polimerase, gerando fluorescência pela separação do fluoróforo do *quencher*. Este processo é repetido a cada ciclo sem interferir com a síntese exponencial dos produtos de PCR (Navarro *et al.*, 2015).

Os termocicladores para estudos de qPCR consistem de uma fonte de luz integrada para excitação, um sistema de detecção de fluorescência e um *software*, o qual realiza a análise quantitativa do sinal detectado durante o ensaio (Navarro *et al.* 2015). Durante a análise de qPCR um valor importante para os cálculos de quantificação é Ct (*Cycle threshold*), que é definido como o número de ciclos requeridos para que o sinal de fluorescência cruze um valor limiar de detecção (*Threshold*). De forma geral o Ct está diretamente relacionado à quantidade de DNA amplificado (Valasek e Repa, 2005).

A partir da análise dos resultados de um ensaio de qPCR a partir dos valores de Ct obtidos, podemos realizar dois métodos de quantificação: a quantificação absoluta (também denominada de curva-padrão) e a quantificação relativa. A quantificação absoluta é baseada na comparação dos valores de Ct com uma curva-padrão gerada da amplificação de quantidades conhecidas do DNA (gene) alvo (Postollec *et al.*, 2011). A curva-padrão é geralmente usada para determinar a eficiência de amplificação (**E**) de um par de *primer*, uma alta eficiência está relacionada a uma inclinação (*slope*) de 3,32 da reta representada pelas diluições de 10 vezes do DNA alvo (Taylor *et al.*, 2010). Valores de eficiência de 100% indicam que a quantidade de DNA dobra a cada ciclo de amplificação (Bustin *et al.*, 2009).

Quanto a quantificação relativa, esta é usada para estimar mudanças na expressão de um gene específico, baseando-se na normalização da quantidade do gene alvo por um gene controle endógeno com expressão constante (*housekeeping gene*). De maneira geral, a expressão relativa é dada por  $2^{-\Delta Ct}$ , onde a variação de Ct ( $\Delta Ct$ ) é dada por  $Ct_{(amostra)} - Ct_{(controle\ end\acute{o}geno)}$  (Arya *et al.*, 2005; Postollec *et al.*, 2011).

O fato do qPCR envolver a amplificação de sequências curtas do DNA alvo, permite reduzir de forma significativa o tempo de extensão. Além disso, outra vantagem do qPCR em relação ao PCR convencional é que o passo de eletroforese em gel de agarose ou acrilamida não é requerido para visualizar os fragmentos amplificados (Levin, 2005).

## **2 OBJETIVOS**

### **2.1 OBJETIVO GERAL**

Caracterizar o transcriptoma de folhas do quimiotipo BGEN-04 de *Lippia alba*, contribuindo com novas informações sobre as vias metabólicas de terpenóides desta espécie.

### **2.2 OBJETIVOS ESPECÍFICOS**

O presente trabalho tem como objetivos específicos:

- Montar e anotar o transcriptoma de folhas de *Lippia alba*;
- Identificar por métodos *in silico* transcritos das enzimas terpeno sintase expressas em folhas de *L. alba*;
- Validar *primers* para controle endógeno e genes de terpeno sintases para trabalhos futuros de qPCR.

## **3 MATERIAL E MÉTODOS**

### **3.1 COLETA DAS AMOSTRAS PARA SEQUENCIAMENTO**

Para o sequenciamento do transcriptoma do acesso BGEN-04 de *Lippia alba*, foram coletadas folhas do quarto nó a partir do ápice que estavam em estágios iniciais de desenvolvimento, isto é, com o limbo foliar em abertura. O acesso BGEN-04 foi cultivado na Estação Experimental de Plantas da Universidade Federal de Juiz de Fora, Minas Gerais, Brasil. E o mesmo possui exsicata depositada no Herbário Leopoldo Krieger (CESJ), com número de *voucher* 48372.

Esta planta é caracterizada como triploide e seu principal composto terpenóide é o linalol. A fim de garantir que as folhas estivessem em fase de maior produção de óleo essencial, a coleta foi realizada no mês de março de 2014 durante o período em que as plantas não estavam em floração, uma vez que na fase de floração pode haver uma diminuição no número de transcritos de enzimas relacionadas com a síntese de terpenóides (Tavares *et al.*, 2005). As amostras coletadas foram imediatamente congeladas e armazenadas em nitrogênio líquido até a extração de RNA total.

### **3.2 EXTRAÇÃO DE RNA TOTAL PARA SEQUENCIAMENTO**

Um *pool* de folhas jovens expandidas de *L. alba* BGEN-04 foi coletado e macerado em nitrogênio líquido utilizando almofariz e pistilo estéreis RNase-free. O RNA total foi extraído utilizando método de Trizol juntamente com o Kit *PureLink®RNA Mini Kit* (Ambio®) da Life-Technologies, a partir de 100 mg de tecido foliar macerado, seguindo instruções do fabricante. Ao final do procedimento o RNA total foi ressuspensionado em água RNase-free. A quantificação e análise de integridade foram realizadas em gel de agarose 1%.

### **3.3 SEQUENCIAMENTO NGS DO RNA TOTAL: RNAseq**

Após a extração, quantificação e verificação da integridade do RNA total em gel de agarose 1%, a amostra extraída foi enviada para o Centro de Alto Desempenho de Sequenciamento de Genoma – Genômica-DF, na Universidade Católica de Brasília – Distrito

Federal, Brasil. O *workflow* da etapa de sequenciamento, após chegada ao laboratório Genômica-DF, envolveu três etapas sequenciais:

(i) Análise do RNA total via Bioanalyzer: A amostra de RNA total enviada foi analisada quantitativamente e qualitativamente, utilizando-se o equipamento Bioanalyzer modelo *Agilent 2100 System*;

(ii) Preparo das Bibliotecas para Sequenciamento: O preparo das bibliotecas foi realizado utilizando o Kit *TruSeq Sample Preparation* (Illumina®), seguindo as recomendações do fabricante;

(iii) Sequenciamento NGS propriamente dito: A plataforma de sequenciamento de nova geração utilizada foi o *MiSeq* da Illumina. Foi executada uma corrida de 2X300 bp, ou 600 ciclos, com bibliotecas *paired-end*.

### **3.4 VERIFICAÇÃO DE QUALIDADE E *TRIMMING* DE *READS***

Após o sequenciamento, foi realizada verificação da qualidade das *reads*, isto é, analisou-se o valor do *Phred score* (também denominado de *Phred Quality*), o qual indica o número de erros de base identificados durante os ciclos de sequenciamento. O *Trimming*, ou retirada dos adaptadores, não foi necessário, uma vez que as sequências enviadas após o sequenciamento já estavam tratadas neste sentido, portanto sem adaptadores. As análises de qualidade foram realizadas utilizando o *software* FastQC versão 0.11.2.

### **3.5 MONTAGEM DOS TRANSCRITOS**

A montagem *de novo* das sequências foi realizada utilizando o *software* Trinity versão 20140717. Foram submetidos ao software os dois arquivos “\*.fastq” referentes aos dois *ends* sequenciados, sendo a montagem processada usando os parâmetros: *--normalize\_reads*: realiza a normalização das *reads in silico*; *--seqType*: para definir se arquivo dos *reads* estava no formato fasta (fa) ou fastq (fq); *--JM*: denominado de *jellyfish memory*, usado para definir a quantidade de memória RAM utilizada, neste trabalho foi definido 10 *gigabytes*; *--CPU*: define o número de núcleos do processador que serão usados, foi usado para este estudo 10 núcleos. O valor de *k-mer* usado foi 25, o qual é definido por *default* no programa Trinity.

### 3.6 ANOTAÇÃO FUNCIONAL DE SEQUÊNCIAS GÊNICAS

Após a montagem dos transcritos, as sequências contínuas (*contigs*) obtidas foram submetidas ao *software Blast2Go*, a fim de identificar quais genes estão relacionados. Seguindo o *workflow* do *Blast2Go*, os *contigs* foram primeiramente alinhados contra sequências de proteínas do banco de dados NR (*Non-Redundant*), usando o algoritmo *BlastP* com *e-value* de 1.0E-6. Este passo é denominado de mapeamento (*mapping*) (Figura 3.1).

Após o mapeamento, foi realizada uma pesquisa funcional dos genes preditos, isto é, se tais sequências estavam relacionadas à alguma proteína funcional como, por exemplo, uma enzima de alguma via metabólica. Tal etapa é conhecida como anotação funcional. Neste passo, foi usado o banco de dados *Gene Ontology* e a ferramenta *InterProScan*, a qual realiza uma busca por famílias proteicas a partir de domínios e motivos conservados. Após o passo de anotação funcional, realizou-se uma busca pelo mapa metabólico das enzimas determinadas no banco de dados *KEGG*, de acordo com o número EC (*Enzyme Commission*) (Figura 3.1).

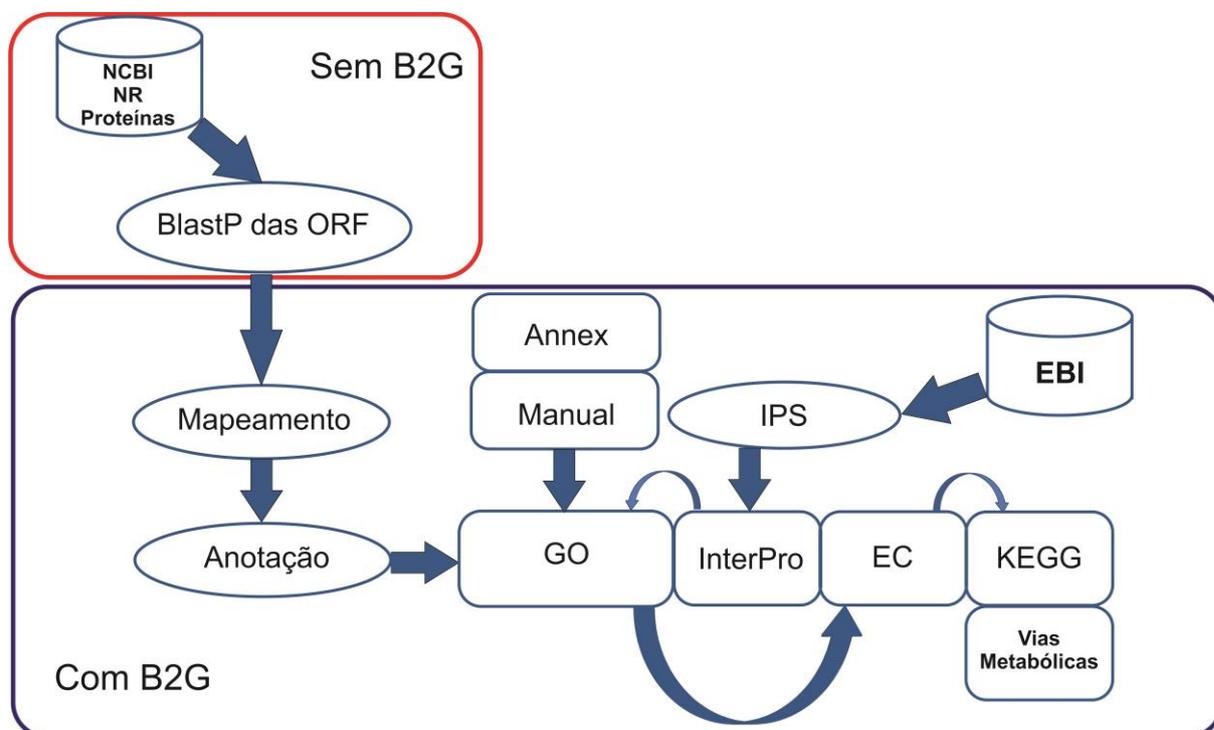


Figura 3.1: *Workflow* adotado para anotação de sequências de transcritos encontradas, após montagem *de novo* do transcriptoma de *Lippia alba*.

### 3.7 ESCOLHA DE SEQUÊNCIAS E DESENHO DE *PRIMERS* PARA ANÁLISE DE PCR QUANTITATIVO (qPCR)

Após montagem e anotação do transcriptoma de *Lippia alba*, foram selecionados um total de 17 *contigs* para estudos de qPCR, sendo 13 destes controles endógenos e 4 referentes a genes da biossíntese de terpenóides. Os genes controles foram escolhidos baseando-se em trabalhos previamente publicados (Reid *et al.*, 2006; Kozera e Rapacz, 2013; Liang *et al.*, 2014; Campos *et al.*, 2015; Castanera *et al.*, 2015; Wang *et al.*, 2015): actina, gliceraldeído-3-fosfato desidrogenase, tubulina alfa, NADH desidrogenase, citocromo b, fator de alongação alfa, RNA polimerase, DNA polimerase, proteína ribossomal L2, sacarose sintase, glicose-6-fosfato isomerase, malato desidrogenase e serina/treonina fosfatase. Quanto aos genes de terpeno sintase, foram escolhidos: farnesil difosfato sintase, geranylgeranyl difosfato redutase, geraniol sintase e linalol/nerolidol sintase. A escolha destes foi feita de acordo com o interesse nas vias de síntese de geraniol, linalol/nerolidol e de diterpenóides.

Dentre os vários *contigs* anotados para os genes escolhidos tanto para controle quanto para terpeno, foram selecionados para desenho dos *primers* aqueles com melhores valores de identidade, cobertura e *e-value*, obtidos durante a anotação funcional usando o *workflow* do programa *Blast2Go*. Os *primers* para PCR quantitativo (qPCR) foram projetados utilizando-se o *software* PrimerExpress V3.0 (Applied Biosystems). A temperatura de *melting* ( $T_m$ ) para cada *primer* foi 65-67 °C e o tamanho do fragmento amplificado (*amplicon*) de cada par de *primer* desenhado foi de 80 bp.

### 3.8 EXTRAÇÃO DE RNA E SÍNTESE DE CDNA PARA qPCR

Foi coletado separadamente um *pool* de folhas jovens expandidas a partir do quarto segmento nodal (a contar do ápice) dos acessos triploides de *L. alba* BGEN-04 (*voucher* 48372), BGEN-23 (*voucher* 48460) e BGEN-37 (*voucher* 48381), cultivados na Estação Experimental de Plantas da Universidade Federal de Juiz de Fora, Minas Gerais, Brasil. As folhas coletadas de cada acesso foram maceradas em nitrogênio líquido utilizando almofariz e pistilo estéreis RNase-free. O RNA total foi extraído utilizando o *Kit illustra™ RNAspin Mini RNA isolation* (GE Healthcare), a partir de 30 mg de tecido foliar macerado, seguindo instruções do fabricante. Ao final do procedimento o RNA total foi ressuspenso em água RNase-

*free*. A quantificação e análise de integridade foram realizadas por espectrofotometria no equipamento *nanodrop* 2000 UV-Vis (Thermo Scientific).

Para obtenção do cDNA, 2 µg do RNA total foram usados como molde, juntamente com 40 µM de oligo-dT *primer* e 2,5 mM de dNTP mix, em um volume final de 16 µL. A mistura foi aquecida a 70°C por 5 minutos e depois imediatamente incubada em gelo. Em seguida, foram adicionados para a mistura anterior 2µL do tampão 10X da enzima transcriptase reversa (500 mM Tris-HCL pH 8,3; 750 mM de KCL; 30 mM MgCl<sub>2</sub>; 100 mM DTT (ditio-treitol)), 1 µL de inibidor de RNase e 1 µL de M-MuLV RT Transcriptase Reversa (NEB, New England Biolabs) a mistura foi incubada a 42°C por 1 hora. A enzima foi inativada a 90°C por 10 minutos.

### 3.9 PCR QUANTITATIVO (qPCR) E ANÁLISE DOS DADOS

Primeiramente, os cDNAs preparados para cada um dos três acessos de *L. alba* foram unidos em um *pool*, o qual foi diluído para as concentrações de 800, 200, 50 e 12,5 ng/µL. Cada concentração foi testada em triplicata para cada par de *primer*, afim de obter o valor de eficiência de amplificação destes.

A análise via qPCR dos *primers* desenhados para cada gene selecionado foi realizada no equipamento *Rotor-gene* Q (Qiagen) em um volume de 11 µL. Os parâmetros usados na amplificação foram: 5 minutos a 95°C (afim de ativar a polimerase), 40 ciclos de desnaturação a 95°C por 5 segundos e *annealing*/extensão a 60°C por 10 segundos e ao final dos 40 ciclos foi realizado passo de dissociação (*melting*) de 72 a 95°C com incremento de 1°C a cada medida de fluorescência. As medidas dos produtos amplificados se deram pela incorporação do marcador fluorescente *SYBR Green* na dupla fita de cDNA para cada reação de amplificação.

As análises de eficiência para cada par de *primer* foram realizadas utilizando o programa *ROTOR-GENE SERIES Q Software* v2.3.1. O programa determina a partir dos valores de Ct (*cycle threshold*) a equação da reta que melhor se ajusta aos dados (Curva Padrão) dada por  $y = Mx + B$ , onde  $M$  é o coeficiente angular da reta denominado *slope* e  $B$  é o intercepto da reta. De posse do valor de *slope* o programa calcula os valores de eficiência segundo a fórmula  $E = 10^{(-1/slope)} - 1$ , onde  $10^{(-1/slope)}$  é denominado valor de amplificação. O programa calcula também os valores de coeficiente de correlação  $R^2$ .

Após validação dos pares de *primers* e respectivo cálculo de eficiência, foi realizado análise do nível de expressão dos genes referentes as enzimas linalol/nerolidol sintase, gerani-

ol sintase, farnesil difosfato sintase e geranylgeranyl redutase para cada um dos quimiotipos citados anteriormente. Nesta etapa foram usados separadamente os cDNA de cada planta e a reação seguiu conforme descrito no segundo parágrafo desta seção com uma concentração de cDNA de 100 ng/ $\mu$ l.

## 4 RESULTADOS E DISCUSSÃO

A figura 4.1 representa o *workflow* adotado no presente estudo para montagem e anotação de seqüências oriundas do sequenciamento do transcriptoma de *Lippia alba*. Os resultados obtidos para cada um dos passos serão apresentados e discutidos nas subseções desta seção de resultados.

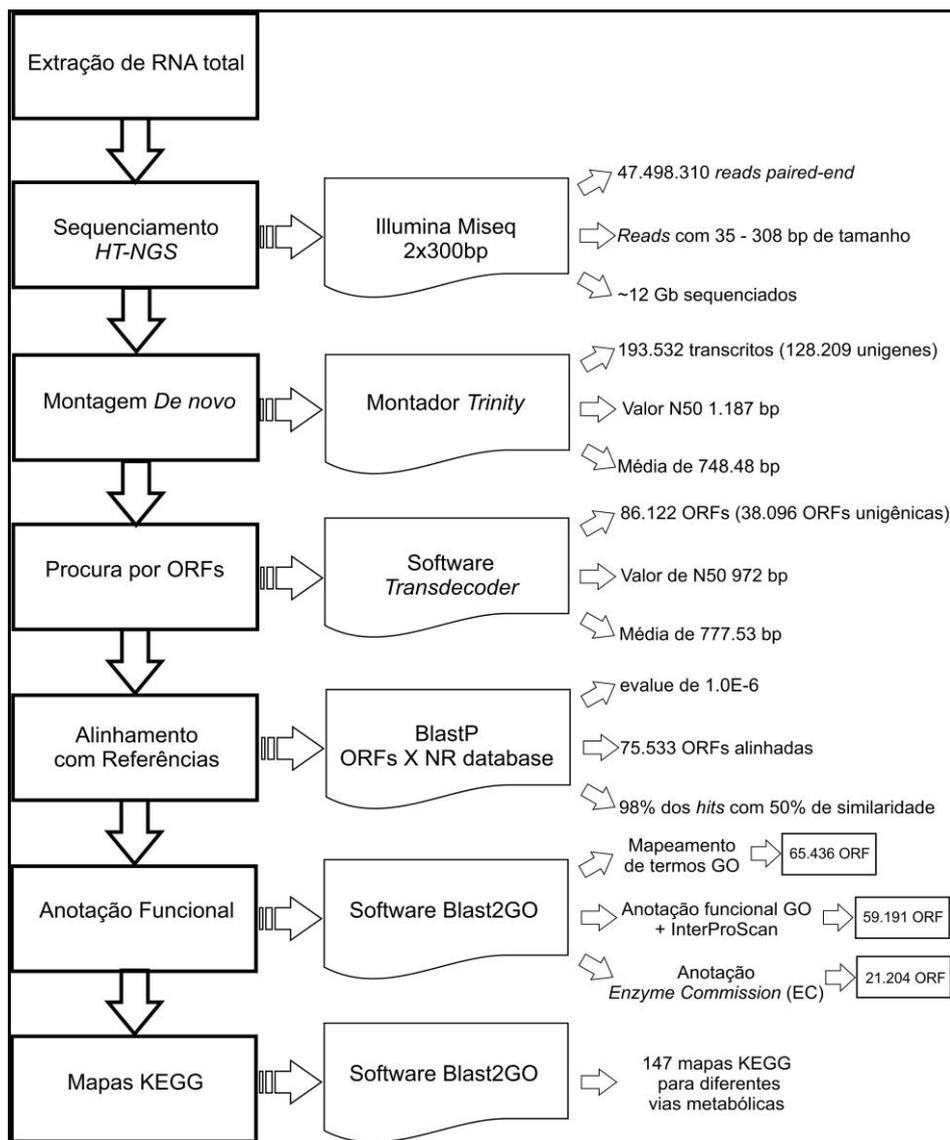


Figura 4.1: *Workflow* de procedimentos desde a extração do RNA total até a anotação de seqüências de transcritos encontrados após montagem das *reads* geradas por sequenciamento NGS do transcriptoma de *L. alba* quimiotipo BGEN-04. Para cada passo, estão indicadas as ferramentas usadas e os resultados obtidos.

## 4.1 EXTRAÇÃO E QUALIDADE DO RNA TOTAL PARA SEQUENCIAMENTO

Em *L. alba* a floração parece não influenciar na qualidade dos óleos essenciais, porém sua quantidade pode ser reduzida nesta fase (Tavares *et al.*, 2005). Portanto, a coleta foi realizada no período em que as plantas não estavam em floração, uma vez que um dos objetivos específicos deste trabalho foi a detecção de transcritos de enzimas terpenos sintases.

O segmento nodal e o estágio das folhas também influenciam a produção de terpenóides. De acordo com Pandeló e colaboradores (2012) folhas jovens do quarto segmento nodal apresentam maior expressão de genes relacionados às enzimas do metabolismo de terpenóides, seguida por folhas do segundo nó. Além disso, folhas maduras presentes no oitavo nó parecem ter menor taxa de expressão de genes e baixa produção de óleos, provavelmente, devido ao fato dos tricomas neste estágio estarem colapsados. Assim, para obtenção do RNA total de *L. alba* BGEN-04 para sequenciamento, foram coletadas folhas conforme pressuposto nos trabalhos de Tavares *et al.* (2005) e Pandeló *et al.* (2012).

O isolamento de RNA com boa qualidade e quantidade é um desafio em plantas que possuem altas taxas de metabólitos secundários, como por exemplo, polissacarídeos e compostos polifenólicos. Estes metabólitos podem se ligar ao RNA e precipitam, resultando em baixo rendimento na extração (Djami-Tchatchou e Straker, 2012). Além disso, a oxidação dos compostos fenólicos durante a extração pode ocasionar degradação e perda da amostra de RNA (Yu *et al.*, 2012).

A qualidade do RNA total extraído neste trabalho foi analisada por eletroforese em gel de agarose 1% (Figura 4.1) e também utilizando o equipamento *Agilent 2100 System bioanalyzer* (Agilent Technologies), o qual resultou em um gráfico (Figura 4.2) mostrando os picos referentes aos RNA ribossômicos 18S e 28S. Estes resultados revelam que apesar das dificuldades pertinentes a extração de RNA de plantas com altas taxas de metabólitos secundários, foi possível obter uma amostra para sequenciamento com boa qualidade.

O kit *Truseq* utilizado para preparo das amostras para sequenciamento, faz uso do método de ligação de oligo-dT para separação e enriquecimento das amostras de mRNA, eliminando o rRNA, mesmo que degradado. No estudo realizado por Adiconis *et al.* (2013), foram comparados diferentes métodos de enriquecimento para amostras de RNA com sinais de degradação e baixa quantidade, concluiu-se que metodologias baseadas na seleção por cauda poliA possuem uma maior fração de *reads* alinhadas em regiões exônicas. Interessantemente,

de acordo com o fabricante, o kit *Truseq* está otimizado para trabalhar com baixas quantidades de RNA total (0,1 a 1 µg).

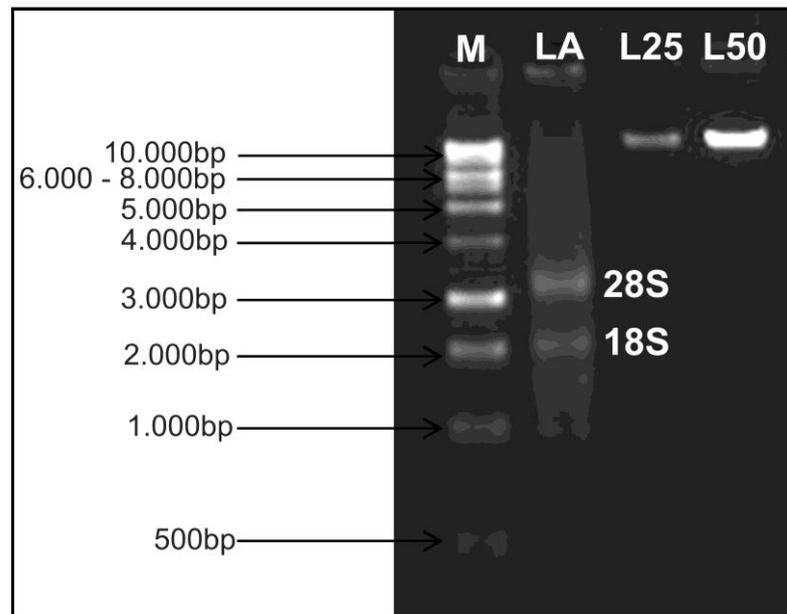


Figura 4.2: Eletroforese de RNA total extraído do quimiotipo BGEN-04 de *L. alba* em gel de agarose 1%. Em sequência são apresentados: M, marcador de 1Kb ladder; LA, RNA total de *L. alba*; L25, marcador lambda de 25 ng; L50, marcador lambda de 50 ng.

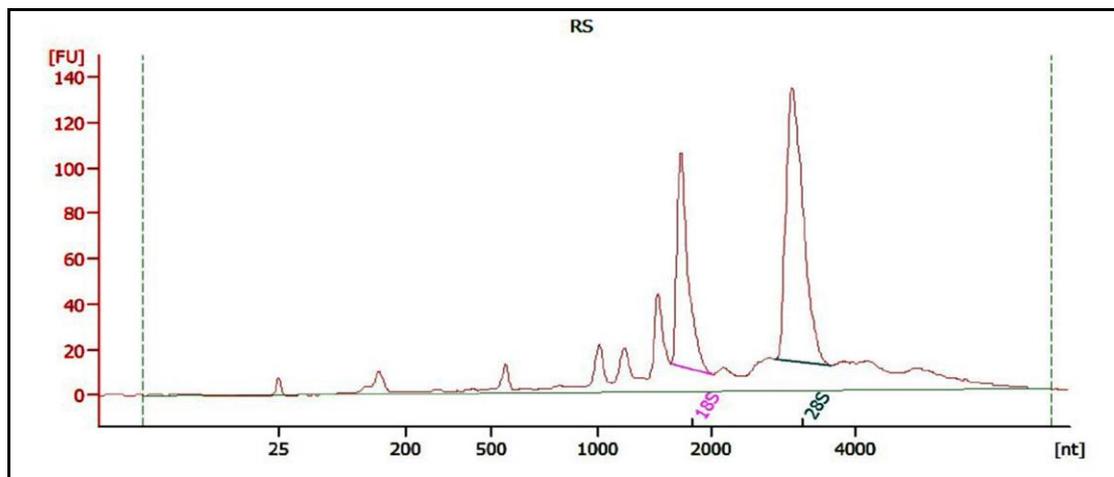


Figura 4.3: Análise do RNA total de *Lippia alba* no *Agilent 2100 System bioanalyzer* destacando os picos 18S e 28S. Nos eixos do gráfico estão descritos tamanho em nucleotídeos (nt) e fluorescência (FU).

## 4.2 SEQUENCIAMENTO E MONTAGEM DOS DADOS DE RNAseq

As bibliotecas *paired-end* geradas foram submetidas à plataforma de sequenciamento *Miseq* da Illumina. Foram obtidas um total de 47.498.310 *reads paired-end* (23.749.155 *reads* para cada *end* sequenciado) de 35-308 bp, compreendendo 12.148.327.567 nucleotídeos (~12 Gb) (Tabela 4.1). Aproximadamente 82% e 70% dos *reads* para o primeiro e segundo *end*, respectivamente, tiveram valores de *phred score* igual ou acima de Q30 (Figura 4.4).

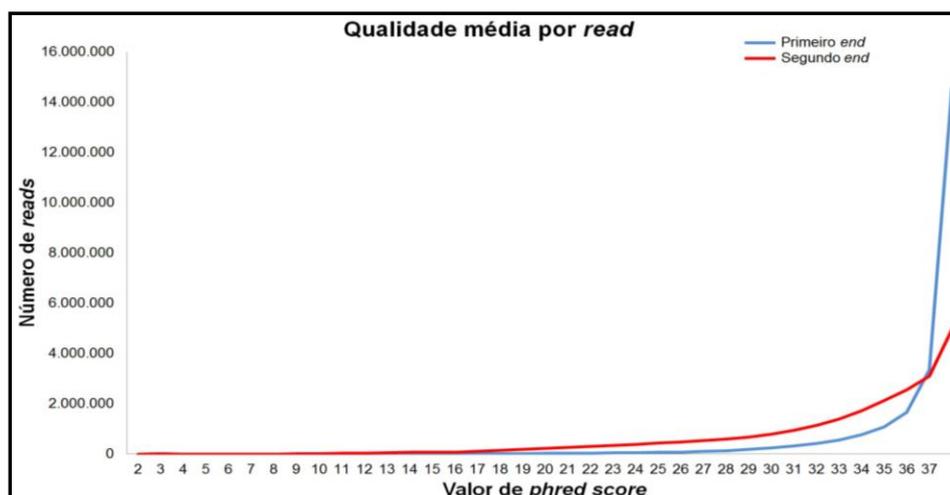


Figura 4.4: Qualidade em *phred score* das *reads paired-end* gerados pelo sequenciamento do transcriptoma de *Lippia alba* usando a plataforma *Illumina Miseq*. Em azul está representado as *reads* para a extremidade *forward* e vermelho as *reads* para a extremidade *reverse*.

Um problema comum, observado em sequenciamentos *paired-end*, é a menor qualidade dos *reads* sequenciados para o segundo *end* em relação ao primeiro (Figura 4.5). Este problema não é bem discutido na literatura, no entanto parece ser de fato uma limitação da tecnologia em relação à capacidade da enzima e das fluorescências que se desgastam ao longo do processo.

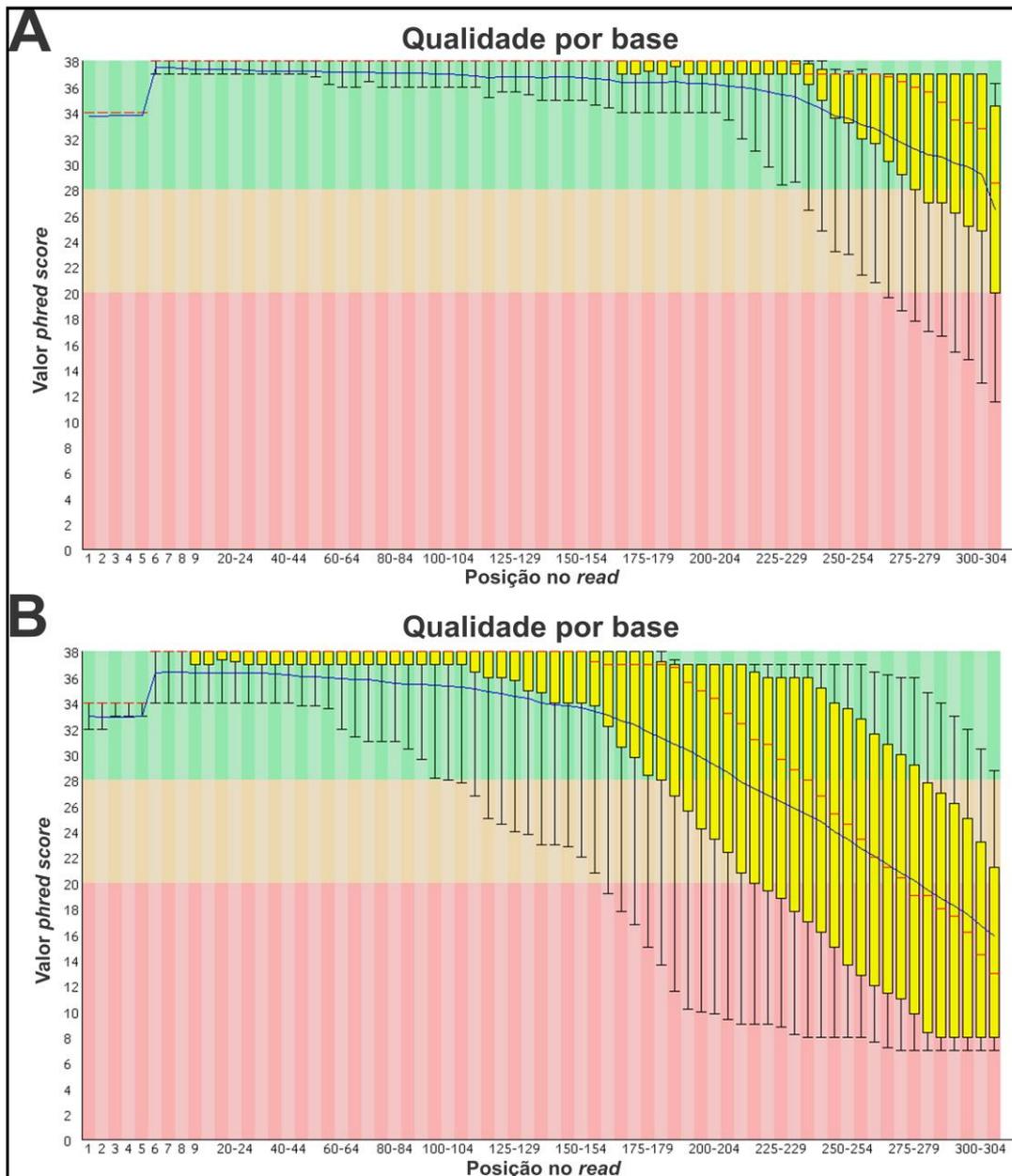


Figura 4.5: Qualidade *phred score* por base dos *reads paired-end* geradas pelo sequenciamento do transcriptoma de *Lippia alba* usando a plataforma *Illumina Miseq*. (A) qualidade por base das *reads* para a extremidade *forward* e (B) qualidade das bases das *reads* para a extremidade *reverse*.

Diante da ausência de um genoma de referência disponível, neste trabalho foi realizada a montagem *de novo* dos *reads* em sequências contínuas denominadas *contigs*. Para isso, foi usado o *software Trinity*, um montador (*assembler*) baseado em *k-mers* (fragmentos de sequências com um tamanho *k* determinado) com 25 bp. O *software* utiliza um algoritmo denominado de grafo *De Bruijn* que processa o alinhamento considerando uma subsequência com conservação de *k-1* bases, reduzindo o custo computacional. De maneira interessante, o pro-

grama *Trinity* consiste de três módulos independentes, porém usados sequencialmente para montagem dos transcritos: (i) *inchworm*, o qual monta os dados de RNAseq para sequências únicas de transcritos (*contigs*); (ii) *chrysalis*, agrupa os *contigs* gerados pelo *inchworm* em *clusters* e constrói grafos *De Bruijn* para cada *cluster*; (iii) *butterfly*, avalia e extrai as prováveis sequências de cada componente dos grafos, resultando portanto em sequências alternativas para um mesmo transcrito (Haas *et al.*, 2013). Trabalhos de montagem de transcritos tem demonstrado que o *Trinity* é altamente efetivo quando comparado a outros métodos, principalmente no que diz respeito ao uso de recurso computacional (Zhao *et al.*, 2011; Duan *et al.*, 2012).

A montagem do transcriptoma de *L. alba* BGEN-O4 resultou em 193.532 transcritos, dos quais 128.209 são unigenes, isto é, sequências que possivelmente representam *splicing* alternativo. O valor de N50 calculado, o qual designa que 50% do total de bases montadas estão contidas em *contigs* de tamanho N ou maiores (Farias, 2013), foi de 1.187 bp considerando todos os transcritos gerados. O tamanho médio e mediano dos *contigs* foram 748,48 bp e 447 bp, respectivamente. Lulin e colaboradores (2012) durante montagem e análise do transcriptoma de folhas de *Carthamus tinctorius*, também usando o programa *Trinity*, obtiveram 195.320 transcritos e 120.778 unigenes (Lulin *et al.*, 2012). Esses valores de transcritos montados e unigenes, foram similares aos obtidos para uma primeira análise do transcriptoma de *Lippia alba*.

Após a montagem, o arquivo “Trinity.fasta” gerado, foi submetido ao software Transdecoder, parte do pacote *Trinity assembler*, a fim de encontrar dentre os transcritos possíveis ORF (*Open Read Frame*) de proteínas. O *Transdecoder* realiza, primeiramente, uma busca por possíveis ORF dentro do arquivo “Trinity.fasta”. Dessa forma, subconjuntos de ORF longos são formados e utilizados como parâmetros em um modelo de Markov. As ORF são randomizadas para sequências não-codificante gerando probabilidades que são pontuadas pelo modelo de Markov, o qual define o melhor *frame* de leitura entre todos os possíveis. Ao final foram obtidos quatro arquivos: (i) “\*.pep”, o qual possui as sequências de aminoácidos para os candidatos finais de ORF; (ii) “\*.cds” que possui as sequências de nucleotídeos para as regiões codificantes; (iii) “\*.gff3” onde estão as posições das possíveis ORF dentro dos transcritos alvos; e (iv) “\*.bed”, o qual descreve as posições das ORF.

Um total de 86.122 sequências de ORF foram encontradas, dentre as quais 38.096 são referentes a unigenes. A análise do arquivo “\*.cds” revelou um valor de N50 de 972b considerando ORF curtas e longas. A mesma análise estatística realizada apenas para sequências lon-

gas por unigenes revelou um N50 de 1062 bp. Com relação às médias de *contigs* para ambos os conjuntos foram obtidos valores de 777,53 bp e 794,40 bp, respectivamente. Muitas das sequências montadas não representavam *frames* de leitura com *códons* de início e término, por isso foi observado a redução no número de sequências unigênicas após a procura por possíveis ORF usando o programa *Transdecoder*.

Tabela 4.1: Resultados obtidos com sequenciamento e montagem do transcriptoma de *Lippia alba*.

	Número de sequências	Número de bases (bp)
<b>Sequenciamento</b>		
Número de <i>reads</i>	47.498.310	12.148.327.567
Tamanho dos <i>reads</i>	-	35-308
<b>Montagem dos transcritos</b>		
Número de transcritos	193.532	144.937.310
Número de unigenes	128.209	-
Tamanho médio dos transcritos	-	749
Tamanho N50	-	1.187
Número N50	37.399	-
Menor transcrito	-	201
Maior transcrito	-	15.616
<b>ORF</b>		
Número de ORF	86.122	66.962.337
Número de ORF unigênicas	38.096	-
Tamanho médio das ORF	-	778
Tamanho N50	-	972
Número N50	21.473	-
Menor transcrito	-	297
Maior transcrito	-	15.213

Resultados obtidos usando o *software Trinity* para montagem e o programa *Transdecoder* para procura de ORF.

O valor de ORF unigênicas obtido no presente estudo com o processo de montagem foi próximo ao valor publicado por Xiao e colaboradores em 2013, os quais realizaram o sequenciamento do transcriptoma de *Lippia dulcis*, espécie próxima de *L. alba*, usando a plataforma *Illumina genome analyzer*. Em seu trabalho, Xiao obteve como resultados para o sequenciamento de *L. dulcis* 59.685.424 *reads*, as quais foram montadas em 41.220 sequências

contínuas unigênicas preditas como CDS (*Coding DNA Sequence*). Além disso, os valores de N50 e tamanho médio obtido para as ORF obtidos neste trabalho, foram próximos aos dados publicados para outras espécies como por exemplo *Pisum sativum* (Sudheesh *et al.*, 2015), *Camellia taliensis* (Zhang *et al.*, 2015), *Litsea cubeba* (Han *et al.*, 2013) e *Lens culinaris* (Verma, Shah e Bhatia, 2013). Uma vez que os valores de N50, tamanho médio dos *contigs* e número de unigenes estão próximos aos resultados descritos na literatura, podemos inferir que os resultados para montagem *de novo*, usando o *software Trinity*, para transcriptoma de *Lippia alba* BGEN-04 são válidos.

### 4.3 ANOTAÇÃO FUNCIONAL

O primeiro passo para anotação foi a realização de um alinhamento das sequências peptídicas presentes no arquivo “\*.pep” contra o banco de dados NR (Non-Redundant) de proteínas do NCBI. Tal etapa foi processada utilizando o algoritmo *BlastP* versão 2.2.28, instalado localmente, com *E-value* igual a 1.0E-6. Dentre as 86.122 sequências de ORF, 75.533 alinharam com pelo menos um *hit*, sendo que 98% dos alinhamentos tiveram valores de similaridade acima de 50% (Figura 4.6).

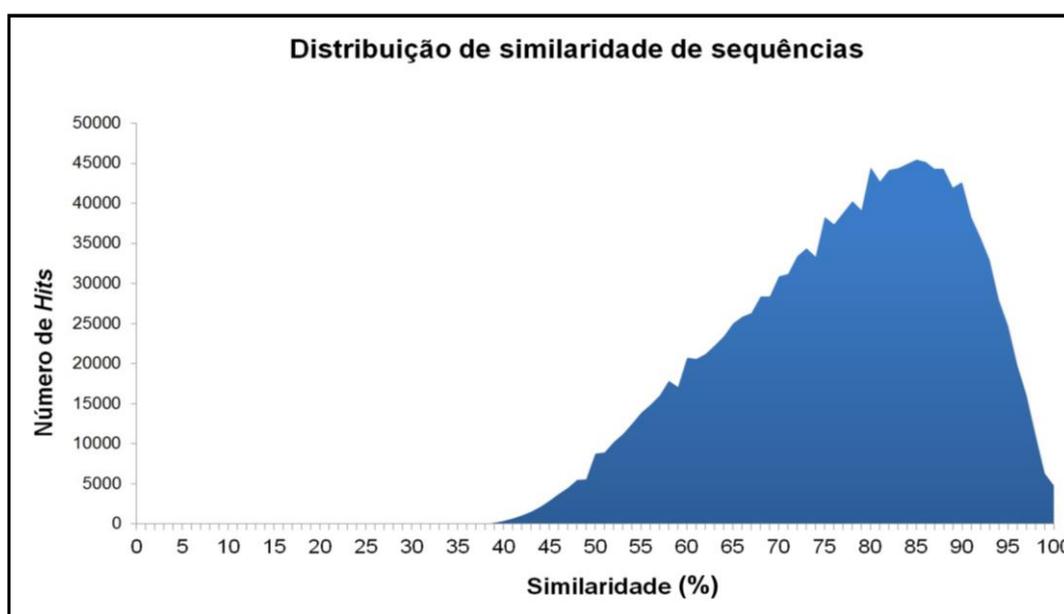


Figura 4.6: Percentual de similaridade das sequências, obtidas a partir da montagem do transcriptoma de *Lippia alba*, alinhadas contra o banco de dados NR de

proteínas usando o algoritmo *BlastP* local. O gráfico foi gerado utilizando o programa *Blast2GO*.

Após o alinhamento, o arquivo “\*.xml” contendo os resultados do *BlastP* foi submetido ao software *Blast2GO*, no qual foi realizado o passo de anotação *in silico*. Do total de sequências submetidas ao *workflow* de anotação, 59.191 sequências foram anotadas. Nesta anotação, um gene pode ser atribuído a mais do que uma categoria e, por conseguinte, o número total de transcritos categorizados em cada categoria excede o número total de transcritos observados.

A anotação usando o *gene ontology* (GO) para as três principais categorias biológicas revelou que: 45.633 sequências foram associadas a categoria de processos biológicos (GO:0008150), 51.297 para função molecular (GO:0003674) e 34.875 associadas a componentes celulares (GO:0005575). Dentro da categoria de processos biológicos 36.585 transcritos foram anotados para a classe processo metabólico (GO:0008052), representando cerca de 61,80% de todas as sequências anotadas. Na categoria função molecular, a maior parte dos transcritos foram anotados para a classe ligação (GO:0005488) com 35.060 sequências, seguido pela classe atividade catalítica (GO:0003824) com 30.916 sequências anotadas. Quanto a categoria componente celular, foi observado que 30.071 sequências representam partes celulares (GO:0044464). Tais informações podem ser observadas na figura 4.7.

Aproximadamente 27.163 transcritos anotados para processos metabólicos foram associados ao metabolismo primário (GO:0044238). Enquanto que a busca por termos GO para metabolismo de isoprenóides (GO:0006720) e terpenóides (GO:0006721) indicou que 807 (1,36%) e 530 (0,89%) das sequências foram anotadas para esses processos, respectivamente.

No estudo do transcriptoma de *Raphanus sativus* (popularmente conhecida como rabanete), realizado por Wang e colaboradores em 2013, foram obtidos 46.966 transcritos com assinatura no *gene ontology*, sendo que: (i) para a categoria processo biológico foram encontradas 27.531 (58,62%) sequências referentes a processos metabólicos; (ii) dentro de função molecular as classes mais abundantes foram ligação com 28.856 sequências e atividade catalítica com 21.742 sequências; (iii) na categoria componentes celulares a classe mais abundante foi partes celulares com 36.289 sequências (Wang *et al.*, 2013). Comparativamente, os resultados obtidos para *L. alba* foram similares aos descritos para *Raphanus sativus*.

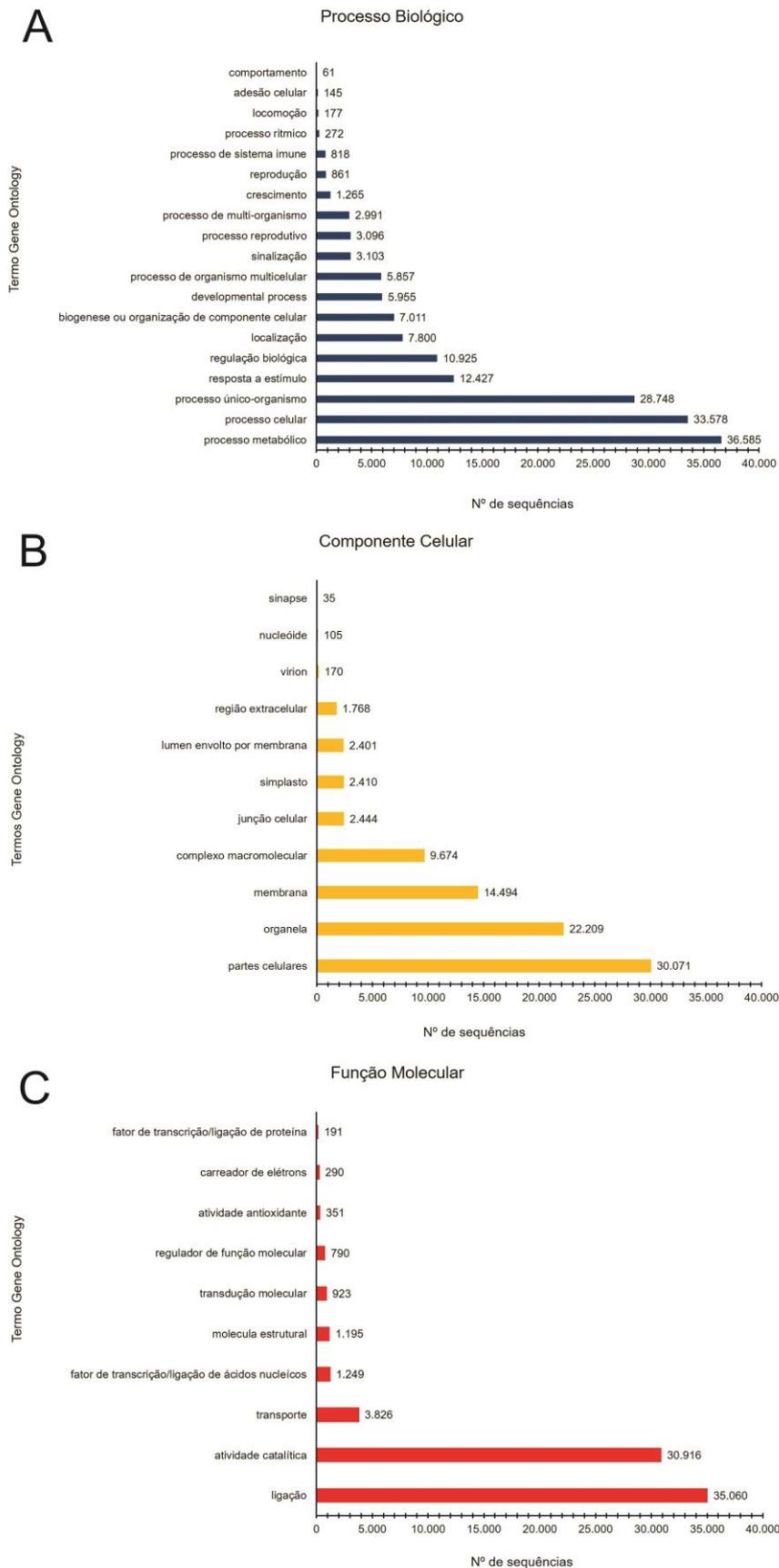


Figura 4.7: Classificação de seqüências, obtidas pela montagem e anotação do transcriptoma de *Lippia alba*, anotadas pelo *Gene Ontology*. Em **A** estão apresentados os números de seqüências para cada classe dentro da categoria **Processo Biológico**; **B** são apresentadas as distribuições das seqüências para cada uma das classes dentro da categoria **Componente Celular**; Em **C** são apresentados os números de seqüências para as classes dentro da categoria **Função Celular**.

### 4.3.1 Obtenção de vias metabólicas

Das 59.191 sequências anotadas utilizando o *workflow* do *Blast2GO*, 21.204 apresentaram código EC (*Enzyme commission*), agrupados em seis diferentes classes de enzimas, conforme apresentado na Figura 4.8. Um total de 1.265 EC distintos foram encontrados e a partir deles obteve-se 147 mapas de vias metabólicas (*KEGG maps*). Liu e colaboradores (2013) em um estudo de análise do transcriptoma de *Boehmeria nivea* (popularmente conhecida como rami) encontraram 19.846 sequências anotadas com EC distribuídas em 126 mapas KEGG, valores próximos aos relatados para *L. alba* BGEN-04 neste estudo.

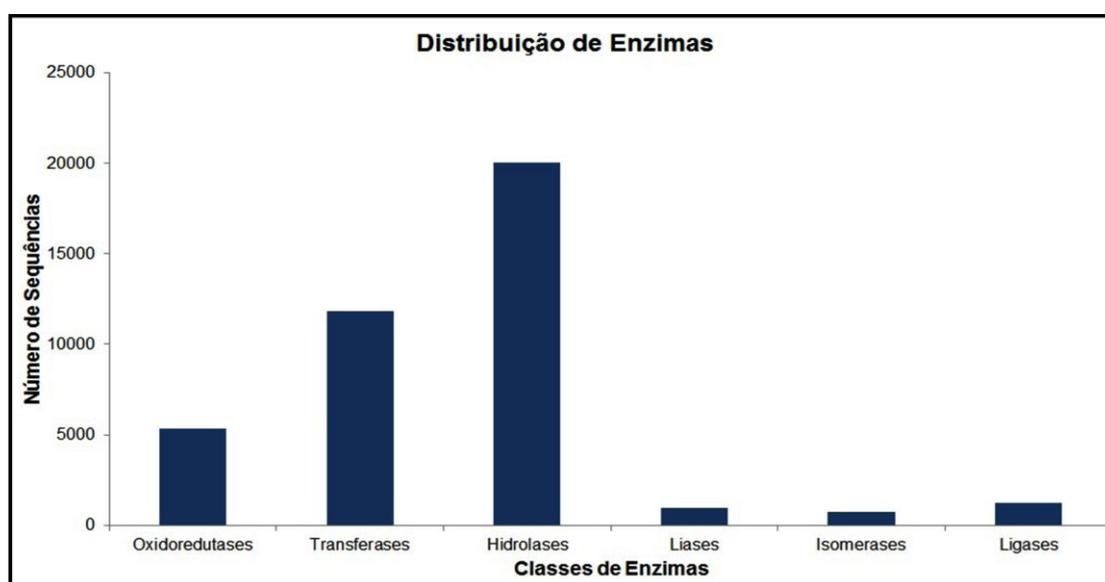


Figura 4.8: Distribuição das sequências, obtidas a partir da montagem e anotação do transcriptoma de *Lippia alba*, de acordo com as classes enzimáticas. A informação da classe de enzima foi obtida a partir do número EC (*Enzyme Commission*) obtido pela anotação no *software Blast2GO*. O número de sequências encontradas para cada classe foi: 5.363 para oxidoredutases; 11.830 para Transferases; 20.047 para hidrolases; 970 para liases; 712 para isomerasas; e 1.243 para ligases.

Como no presente estudo objetivamos o estudo de terpeno sintases, somente esses mapas KEGG (Figura 4.9 até figura 4.13) foram selecionados para apresentação na seção a seguir.

### 4.3.2 Terpeno sintases

A partir das anotações usando os bancos de dados *gene ontology* e Interpro, foi possível encontrar prováveis enzimas envolvidas na síntese dos terpenóides geraniol, nerolidol/linalol, farnesil difosfato e geranilgeranil difosfato (estes dois últimos precursores da via dos sesquiterpenos e diterpenos), as quais foram identificadas acuradas manualmente (Tabela 4.2). Estas sequências foram submetidas aos preditores de clivagem (ProP, SignalP e TatP) e de localização celular (ChloroP e TargetP). Os motivos encontrados foram determinados conforme descrito por Martin *et al* (2010) em seu estudo sobre anotação funcional de terpeno sintases de *Vitis vinifera*. De acordo com a presença destes motivos, as sequências foram marcadas como completas ou incompletas, conforme a presença de motivos importantes para a síntese de terpenóides.

Para a enzima geraniol sintase (EC: 3.1.7.11) foram encontradas duas sequências completas. O alinhamento a partir do *BlastP* contra o banco de dados do Uniprot/Swissprot resultou em valores de identidade de 65% e cobertura de 99% com geraniol sintase de *Ocimum basilicum* (Q6USK1). O alinhamento com a geraniol sintase de *Lippia dulcis* (E9JGT2) demonstrou uma alta homologia entre as sequências, isto é, uma identidade de aproximadamente 89% e cobertura de 99%. A análise usando os programas ChloroP e TargetP indicou que essas sequências anotadas possuem localização em cloroplasto, sendo localizada uma provável região de peptídeo sinal envolvendo o primeiro resíduo até o resíduo 43 para o contig c32759\_g1\_i2. Para o contig c32759\_g1\_i2 foi incluído do primeiro aminoácido até o resíduo 45, conforme indicado na tabela 4.2.

Com relação a enzima linalol/nerolidol sintase (EC: 4.2.3.25/EC: 4.2.3.48) em *L. alba* foram anotadas 4 sequências completas, as quais possuem valores médios de identidade 48% e cobertura 98% em relação a nerolidol sintases (EC: 4.2.3.48) de *Fragaria ananassa* (P0CV94/ P0CV95). De forma oposta às de geraniol sintases, a análise de localização subcelular para as sequências de linalol/nerolidol sintases indicou que as mesmas não possuem sinal de direcionamento a plastídios.

Para as enzimas terpeno sintases farnesil difosfato sintase, geraniol sintase e linalol/nerolidol sintase foi possível mapear os motivos **DDXXD** e **NSE/DTE**, os quais são responsáveis pela ligação de íons metálicos pelo processo de ionização do substrato. Porém o motivo **RR(X)<sub>8</sub>W**, o qual está envolvido com a ciclização do substrato, não foi encontrado. Este motivo nem sempre está presente em monoterpene sintases e sua ausência parece estar relacionada ao fato dessas enzimas produzirem monoterpeneóides acíclicos. O mesmo foi ob-

servado em trabalhos com monoterpene sintases de manjeriço e soja (Iijma *et al.*, 2004; Zhang *et al.*, 2013).

A ausência de peptídeo sinal e do motivo **RR(X)<sub>8</sub>W** são características comuns das terpeno sintases da subfamília TPSg, a qual inclui as enzimas nerolidol sintases (ZHANG *et al.*, 2013). Por outro lado, a subfamília das TPSb, que inclui as geraniol sintases, são caracterizadas por apresentarem um peptídeo de endereçamento. Devido as similaridades, sugere-se que a subfamília TPSb teve origem a partir da subfamília TPSg no início da evolução das angiospermas (Sharkey *et al.*, 2013).

Foi demonstrado para *V. vinifera* que a linalol e a nerolidol podem ser produzidas *in vitro* por uma enzima bifuncional, a qual utiliza tanto o FPP quanto o GPP como substratos. Todavia, a produção *in vivo* desses terpenóides é compartimentalizada em locais distintos das células, graças à presença ou ausência de um peptídeo de endereçamento para o cloroplasto no N-terminal das enzimas. Conforme sugerido para *Anthriscum majus*, dois genes distintos podem codificar enzimas semelhantes com função linalol/nerolidol sintase e com diferentes endereços dentro da célula (Nagegowda *et al.*, 2008).

A hipótese da presença de genes distintos que codificam enzimas linalol/nerolidol sintase com diferentes endereços, juntamente com a expressão diferencial de terpeno sintases, conforme sugerido por Pandeló *et al.* (2012) para *L. alba*, podem explicar o fato do mapeamento não encontrar sequências de linalol/nerolidol sintase com peptídeo sinal. Isto é, nerolidol sintase estaria mais expressa em estágios jovens das folhas, enquanto que linalol sintase mais expressa em folhas maduras. Desta forma são necessários estudos em *L. alba* que ajudem a responder esta questão e para isto foram desenhados *primers* para análise por qPCR desses genes.

Tabela 4.2: Sequências de enzimas anotadas para a síntese de terpenóides obtidas a partir da montagem *de novo* do transcriptoma de *Lippia alba* usando o programa *BlastP*.

Contig	Número de AA	Subfamília	DDXXD	NSE/DTE	SignalP		Clivagem (score)	Peptídeo Sinal(score)	Clivagem (score)	ProP	Peptídeo Sinal (score)	ChloroP	TargetP	Tipo
					Clivagem (score)	Peptídeo Sinal(score)								
c27834_g1_i1	468	?	-	-	Ausente	Ausente	Ausente	Ausente	Resíduo 324 (0,830)/resíduo 370 (0,958)/resíduo 433 (0,546)	Ausente	Ausente	Cloroplasto (0,562)/resíduo 44	Cloroplasto	Completa
c27834_g1_i2	466	?	-	-	Ausente	Ausente	Ausente	Ausente	Resíduo 322 (0,830)/resíduo 368 (0,958)/resíduo 431(0,546)	Ausente	Ausente	Cloroplasto (0,567)/resíduo 42	Cloroplasto	Completa
c28162_g3_i2	103	?	DDIFD	Ausente	Ausente	Ausente	Ausente	Ausente	Resíduo 90 (0,732)	Resíduo 26-27 (TLA-IL)	Ausente	?	?	Fragmento
c30822_g1_i1	410	TPSg	DDIFD	LVDDLGS AKDE	KEA-QD(0,606)	Resíduo 286(0,884)	Resíduo 286(0,884)	Resíduo 286(0,884)	Resíduo 47 (0,863)/resíduo 237 (0,853)/ resíduo 366 (0,814)	Ausente	Ausente	?	?	Completa
c30822_g1_i2	376	?	DDIFD	Ausente	Resíduo 248(0,408)	Resíduo 369 (0,790)	Resíduo 369 (0,790)	Resíduo 369 (0,790)	Resíduo 159 (0,863)/ resíduo 349 (0,959)	Ausente	Ausente	?	?	Fragmento
c30822_g1_i3	481	TPSg	DDIFD	LVDDLGS AKDE	KEA-QD(0,606)	Resíduo 357 (0,884)	Resíduo 357 (0,884)	Resíduo 357 (0,884)	Resíduo 118 (0,863)/ resíduo 308 (0,853)/ resíduo 437 (0,814)	Ausente	Ausente	?	?	Completa
c30822_g1_i4	322	?	DDIFD	Ausente	Resíduo 248 (0,408)	Ausente	Ausente	Ausente	Resíduo 159 (0,863)	Ausente	Ausente	?	?	Fragmento
c30822_g1_i6	522	TPSg	DDIFD	LVDDLGS AKDE	KEA-QD (0,606)	Resíduo 398 (0,884)	Resíduo 398 (0,884)	Resíduo 398 (0,884)	Resíduo 159 (0,863)/resíduo 349 (0,853)/resíduo 478 (0,814)	Ausente	Ausente	?	?	Completa
c30822_g1_i7	449	TPSg	DDIFD	LVDDLGS AKVI	KEA-QD (0,606)	Ausente	Ausente	Ausente	Resíduo 159 (0,863)/resíduo 349 (0,853)	Ausente	Ausente	?	?	Completa
c32759_g1_i2	587	TPSb	DDIFD	LWDDLGTAKEE	VTA-QN (0,976)	Resíduo 454(0,905)	Resíduo 454(0,905)	Resíduo 454(0,905)	Ausente	Ausente	Ausente	Cloroplasto (0,581)/resíduo 43	Cloroplasto	Completa
c32759_g1_i3	589	TPSb	DDIFD	LWDDLGTAKEE	VTA-QN (0,976)	Resíduo 456 (0,905)	Resíduo 456 (0,905)	Resíduo 456 (0,905)	Ausente	Ausente	Ausente	Cloroplasto (0,565)/resíduo 45	Cloroplasto	Completa
c32759_g1_i5	240	TPSb	Ausente	LWDDLGTAKEE	VTA-QN (0,976)	Resíduo 107 (0,905)	Resíduo 107 (0,905)	Resíduo 107 (0,905)	Ausente	Ausente	Ausente	?	?	Fragmento
c34343_g2_i1	327	?	DDIMD/DDLYD <sup>a</sup>	-	Ausente	Ausente	Ausente	Ausente	Ausente	Ausente	Ausente	?	?	Completa
c34343_g2_i2	359	?	DDIMD/DDLYD <sup>a</sup>	-	Ausente	Ausente	Ausente	Ausente	Ausente	Ausente	Ausente	?	?	Completa
c34343_g2_i6	299	?	DDIMD/DDLYD <sup>a</sup>	-	Ausente	Ausente	Ausente	Ausente	Ausente	Ausente	Ausente	?	?	Completa

<i>Contig</i>	Melhor <i>hit</i> SWISSPROT	AC UNIPROT/ SWISSPROT	Identidade SWISSPROT	Cobertura BlastP	Melhor <i>hit</i> TREMBL	AC UNIPROT/ TREMBL	IDENTITY TREMBL	Cobertura BlastP
c27834_g1_i1	Geranylgeranyl diphosphate reductase ( <i>Nicotiana tabacum</i> )	Q9ZS34	86,3%	99,14%	Geranylgeranyl reductase ( <i>Sesamum orientale</i> )	E2D5V0	90,60%	99,36%
c27834_g1_i2	Geranylgeranyl diphosphate reductase ( <i>Nicotiana tabacum</i> )	Q9ZS34	86,4%	99,57%	Geranylgeranyl reductase ( <i>Sesamum orientale</i> )	E2D5V0	90,7%	99,78%
c28162_g3_i2	Geraniol synthase, chloroplastic ( <i>Ocimum basilicum</i> )	Q6USK1	66,00%	90,00%	Putative geraniol synthase ( <i>Olea europaea</i> )	J9ZU63	71,00%	96,00%
c30822_g1_i1	(3S,6E)-nerolidol synthase 1 ( <i>Fragaria ananassa</i> )	P0CV94	50,00%	99,00%	Putative uncharacterized protein ( <i>Vitis vinifera</i> )	F6H895	53,00%	99,00%
c30822_g1_i2	(3S,6E)-nerolidol synthase 2, chloroplastic/mitochondrial ( <i>Fragaria ananassa</i> )	P0CV95	44,00%	92,00%	(3S)-linalool/(E)-nerolidol synthase ( <i>Vitis vinifera</i> )	E5GAH3	45,00%	87,00%
c30822_g1_i3	(3S,6E)-nerolidol synthase 1 ( <i>Fragaria ananassa</i> )	P0CV94	48,00%	99,00%	(3S)-linalool/(E)-nerolidol synthase ( <i>Vitis vinifera</i> )	E5GAH3	53,00%	99,00%
c30822_g1_i4	(3S,6E)-nerolidol synthase 1 ( <i>Fragaria ananassa</i> )	P0CV94	40,00%	95,00%	(3S)-linalool/(E)-nerolidol synthase ( <i>Vitis vinifera</i> )	E5GAH3	44,00%	92,00%
c30822_g1_i6	(3S,6E)-nerolidol synthase 1 ( <i>Fragaria ananassa</i> )	P0CV94	47,00%	98,00%	(3S)-linalool/(E)-nerolidol synthase ( <i>Vitis vinifera</i> )	E5GAH3	51,00%	96,00%
c30822_g1_i7	(3S,6E)-nerolidol synthase 2, chloroplastic/mitochondrial ( <i>Fragaria ananassa</i> )	P0CV95	48,00%	96,00%	(3S)-linalool/(E)-nerolidol synthase ( <i>Vitis vinifera</i> )	E5GAH3	48,00%	92,00%
c32759_g1_i2	Geraniol synthase, chloroplastic ( <i>Ocimum basilicum</i> )	Q6USK1	65,60%	99,00%	Geraniol synthase ( <i>Lippia dulcis</i> )	E9JGT2	89,60%	99,00%
c2759_g1_i3	Geraniol synthase, chloroplastic ( <i>Ocimum basilicum</i> )	Q6USK1	65,30%	99,00%	Geraniol synthase ( <i>Lippia dulcis</i> )	E9JGT2	89,20%	97,00%
c32759_g1_i5	Geraniol synthase, chloroplastic ( <i>Ocimum basilicum</i> )	Q6USK1	79,00%	98,00%	Geraniol synthase ( <i>Lippia dulcis</i> )	E9JGT2	93,00%	99,00%
c34343_g2_i1	Farnesyl pyrophosphate synthase 1 ( <i>Lupinus albus</i> )	P49351	88,80%	100,00%	Farnesyl pyrophosphate synthase ( <i>Bacopa monnieri</i> )	G0TA25	97,40%	100,00%
c34343_g2_i2	Farnesyl pyrophosphate synthase 1 ( <i>Lupinus albus</i> )	P49351	85,70%	95,26%	Farnesyl pyrophosphate synthase ( <i>Bacopa monnieri</i> )	G0TA25	93,1%	97,21%
c34343_g2_i6	Farnesyl pyrophosphate synthase 1 ( <i>Lupinus albus</i> )	P49351	87,30%	87,43%	Farnesyl pyrophosphate synthase ( <i>Bacopa monnieri</i> )	G0TA25	94,60%	85,67%

**(Continuação)**

Estão apresentadas aqui sequências de enzimas terpeno sintases usadas no ensaio de qPCR descrito na seção 4.4, a seguir. Em destaque estão as sequências completas para nerolidol/linalol sintase (**Amarelo**) e geraniol sintase (**Verde**), de acordo com a presença dos motivos **DDXXD** e **NSE/DTE**. Os sinais de “?” e “-” indicam, respectivamente, que não foi possível obter a informação para a coluna e que os domínios não estão presentes para a enzima anotada. O símbolo “a” indica que foram encontrados dois motivos **DDXXD** para farnesil difosfato sintase.



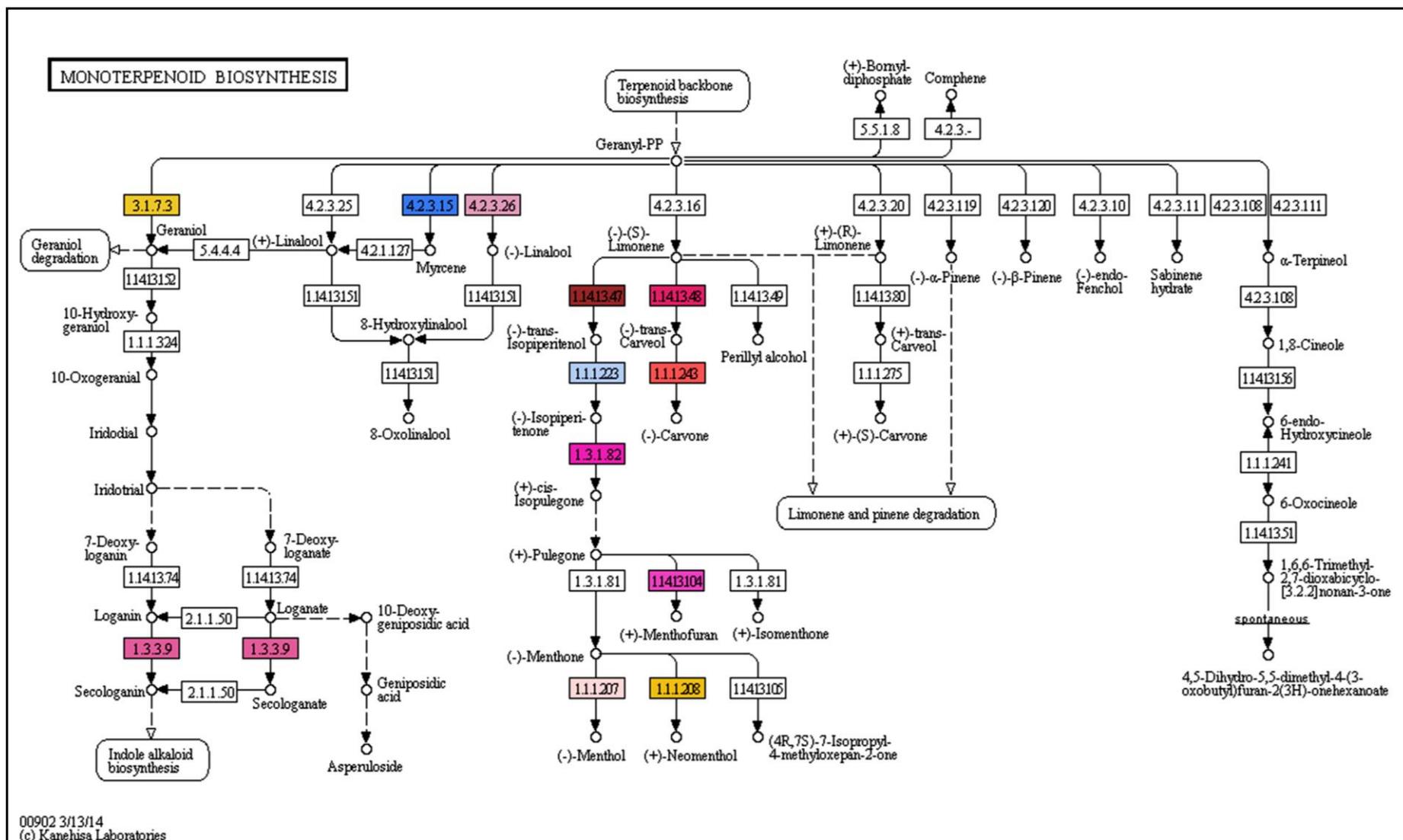


Figura 4.10: Representação da via de síntese dos monoterpenóides. As enzimas anotadas para *L. alba* e presentes nesta via estão coloridas de acordo com o número EC de sua categoria (Mapa KEGG 902).



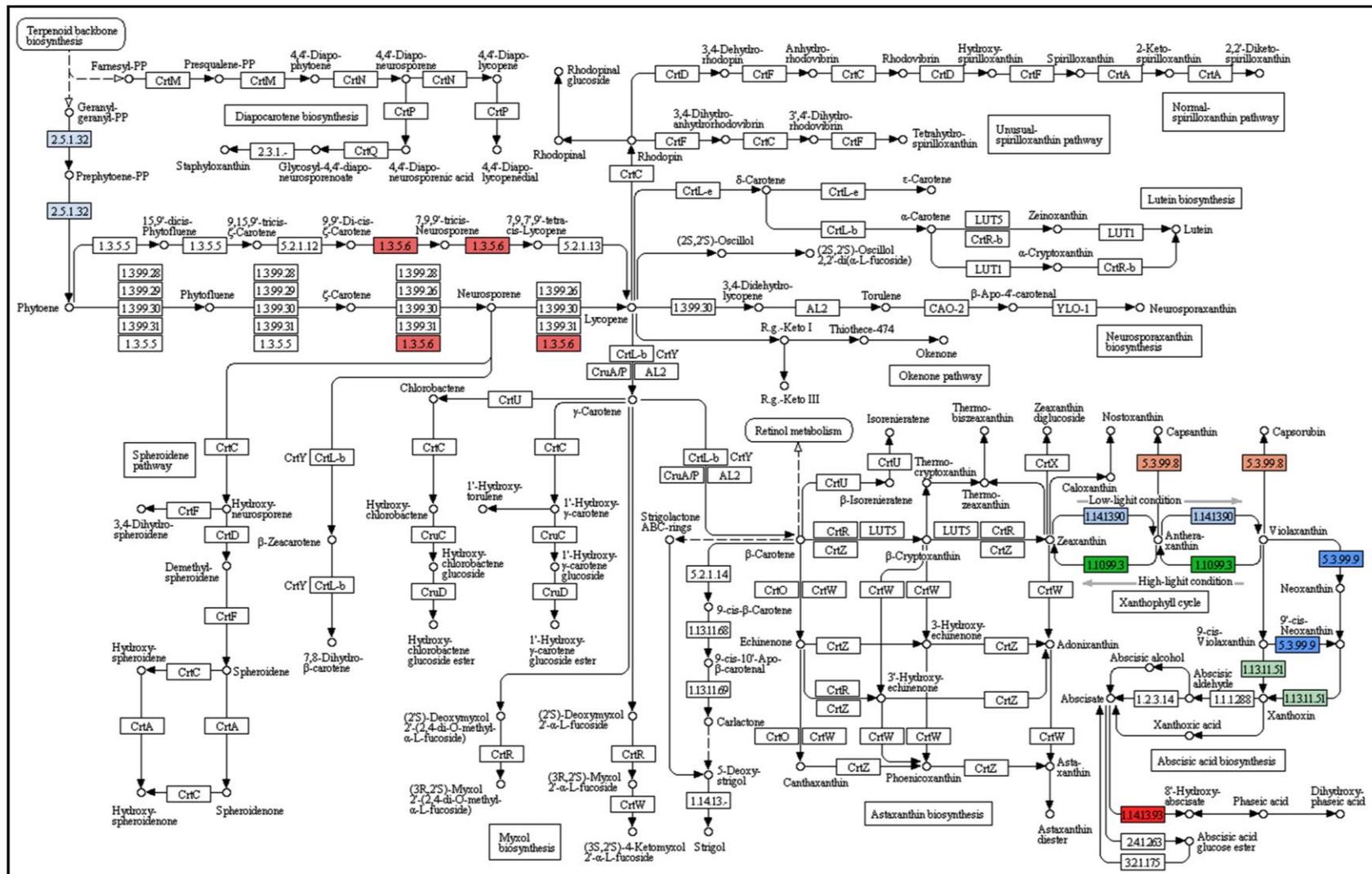


Figura 4.12: Representação da via dos carotenoides. As enzimas anotadas para *L. alba* e presentes nesta via estão coloridas de acordo com o número EC de sua categoria (Mapa KEGG 906).

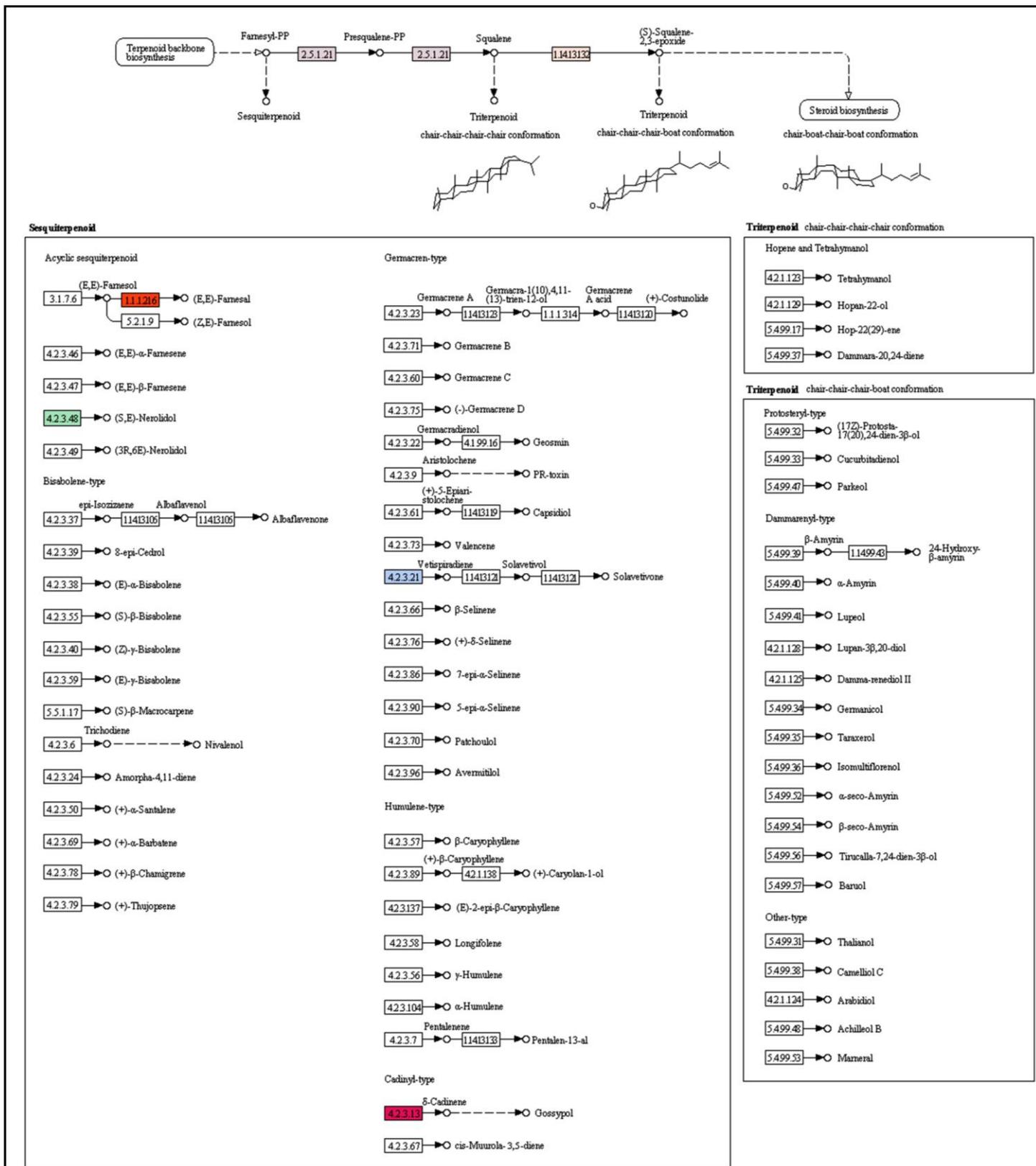


Figura 4.13: Representação da via de síntese dos sesqui- e triterpenóides. As enzimas anotadas para *L. alba* e presentes nesta via estão coloridas de acordo com o número EC de sua categoria (Mapa KEGG 909).

#### 4.4 OBTENÇÃO DE *PRIMERS* PARA ESTUDOS DE qPCR E VALIDAÇÃO DA MONTAGEM

A validação dos resultados obtidos com a montagem de transcriptomas pode ser feita utilizando a técnica de PCR quantitativo em tempo real (qPCR), uma vez que a mesma apresenta sensibilidade de detecção, especificidade e reprodutibilidade. Huang e colaboradores (2012) a fim de validar a montagem e anotação do transcriptoma do pinhão chinês (*Cunninghamia lanceolata*) selecionaram 18 genes relacionados com a biossíntese de lignina e celulose, os quais foram amplificados por qPCR confirmando suas sequências putativas. No estudo de análise do transcriptoma de folhas de banana (*Musa acuminata*) em busca de marcadores SSR (*Simple Sequence Repeat*) para genes relacionados com resistência ao patógeno *Mycosphaerella musicola*, realizado por Passos e colaboradores (2013), a técnica de qPCR foi também usada para a validação dos polimorfismos entre os acessos usados no trabalho. Em 2011, no estudo sobre a dinâmica do transcriptoma de algodão sob situações de defesa contra o fungo patógeno *Verticillium dahliae* foi realizada a validação dos resultados de RNA-seq a partir da análise dos principais genes diferencialmente expressos usando também a técnica de qPCR, encontrando correlação moderada e significativa entre as duas técnicas (Xu *et al.*, 2011).

O passo inicial para estudos de qPCR é a obtenção de *primers* e cálculo do valor da eficiência de amplificação destes. Este valor desempenha um papel fundamental na análise de quantificação absoluta, uma vez que com o seu decréscimo ocorre diminuição do poder de determinação do nível de expressão (TICHOPAD *et al.*, 2004). A eficiência ideal deve ter valor próximo a 1 (ou 100%), o que significa que durante a fase logarítmica da reação, o produto do PCR de interesse é o dobro do ciclo anterior. Contudo, a eficiência pode ser significativamente afetada pela qualidade dos *templates*, *primers* e da molécula fluorescente utilizada na detecção (ZHANG *et al.*, 2015). Desta forma, a determinação da eficiência de *primers* para realização do PCR quantitativo é um passo essencial na técnica.

A fim de validar experimentalmente a confiabilidade das sequências montadas e anotadas, a partir do sequenciamento do transcriptoma de *Lippia alba*, foram desenhados *primers* para análise por PCR quantitativo em tempo real para 13 controles endógenos: actina – *ACT*; gliceraldeído-3-fosfato desidrogenase – *GADPH*; tubulina alfa - *Tubul $\alpha$* ; NADH desidrogenase – *NADHD*; citocromo b – *CITb*; fator de alongação 1 alfa - *EF-1 $\alpha$* ; RNA polimerase –

*RNApol*; DNAPolimerase – *DNAPol*; proteína ribossomal L2 – *RPTL2*; sacarose sintase – *SUCRS*; glicose-6-fosfato isomerase – *G6ISO*; malato desidrogenase – *MALADH*; serina/treonina fosfatase – *SerThr*. Foram desenhados também *primers* para enzimas relacionadas com a síntese de terpenóides: farnesil difosfato sintase – *FRNDPS*; geraniol sintase – *GERS*; linalol/nerolidol sintase – *NERLIN*; geranilgeranil difosfato redutase – *GERGER*.

Para o desenho de *primers* para esses genes, foram selecionadas dentre as diversas sequências anotadas aquelas com melhores resultados de anotação funcional usando o programa *Blast2Go*, conforme descrito na seção 3.7. Após a seleção, foi realizado o alinhamento das CDS contra o banco de dados NR/NT usando o programa *BlastN*, afim de observar a conservação das sequências nucleotídica (Tabela 4.3).

Tabela 4.3: Sequências usadas para desenhar *primers* para qPCR, afim de validar o sequenciamento e montagem do transcriptoma de *Lippia alba*.

<i>Contig</i>	Símbolo	<i>Hit</i> NCBI NR/NT(Número de acesso)	Cobertura <sup>a</sup>	Identidade	<i>E-value</i>
c26477_g8_i3	ACT	Actina de <i>Sesamum indicum</i> (XM_011080520.1)	100%	90%	0
c30049_g1_i3	GADPH	Gliceraldeído-3-fosfato desidrogenase de <i>Sesamum indicum</i> (XM_009782054.1)	97%	85%	0
c30168_g1_i3	TUBUL	Cadeia alfa-3 tubulina de <i>Sesamum indicum</i> (XM_011091420.1)	100%	87%	0
c24835_g1_i1	NADHD	NADH desidrogenase de <i>Acacia mangium</i> (AB839205.1)	99%	98%	0
c9128_g1_i1	CITB	Apocitocromo b de <i>Nicotiana tabacum</i> (U92011.1)	99%	98%	0
c36689_g2_i1	EF-1 $\alpha$	Fator de alongação alfa de <i>Sesamum indicum</i> (XM_011091573.1)	100%	91%	0
c26987_g4_i4	RNApol	Subunidade II RNAPolimerase beta de <i>Lantana camara</i> (HM853922.1)	100%	99%	0
c30863_g2_i1	DNAPol	DNAPolimerase delta de <i>Sesamum indicum</i> (XM_011101402.1)	99%	91%	0
c33701_g2_i5	RPTL2	Proteína ribossomal L2 de <i>Davidia involucrata</i> (HM100382.1)	99%	98%	0
c35136_g1_i2	SUCRS	Sacarose sintase de <i>Sesamum indicum</i> (XM_011095588.1)	100%	85%	0
c29323_g1_i2	G6ISO	Glicose-6-fosfato isomerase 1 de <i>Sesamum indicum</i> (XM_011075926.1)	98%	90%	0
c5731_g1_i1	MALADH	Malato desidrogenase de <i>Sesamum indicum</i> (XM_011075227.1)	100%	93%	2E-150
c30533_g2_i3	SerThr	Serina/Treonina fosfatase de <i>Erythranthe guttatus</i> (XM_012975954.1)	100%	91%	0
c34343_g2_i6	FRNDPS	Farnesil difosfato sintase de <i>Sesamum indicum</i> (XM_011097585.1)	99%	90%	0
c32759_g1_i2	GERSF	Geraniol sintase de <i>Lippia dulcis</i> (GU136162.1)	100%	92%	0
c30822_g1_i6	NERLIN	(3S,6E) -nerolidol sintase de <i>Cucumis sativus</i> (XM_011650129.1)	5%	83%	7E-12
c27834_g1_i1	GERGER	Geranil geranil difosfato redutase de <i>Sesamum indicum</i> (XM_011073402.1)	98%	86%	0

Estão apresentados os valores de cobertura, identidade e *e-value*, obtidos durante o alinhamento das CDS de *L. alba* contra o banco de dados NR/NT do NCBI. <sup>a</sup> Cobertura referente a porcentagem de nucleotídeos da sequência de busca alinhada contra a sequência alvo no banco de dados NR/NT.

Na tabela 4.3, observa-se que as sequências nucleotídicas dos genes controles endógenos apresentaram valores de identidade e cobertura próximos a 100% e 90%, respectivamente, corroborando a eficácia da montagem e anotação do sequenciamento. Resultados similares

foram obtidos para as sequências referente aos genes das enzimas terpeno sintases, com exceção apenas da linalol/nerolidol sintase.

Para linalol/nerolidol sintase, o alinhamento da sequência CDS foi restrito às regiões dos motivos **DDXXD** e **NSE/DTE** e proximidades destes, sendo o valor de identidade igual a 83% e cobertura de 5%. Este resultado foi diferente do observado para o alinhamento proteico realizado durante a etapa de anotação, onde foram obtidos valores de identidade e cobertura iguais a 47% e 98%, respectivamente (Tabela 4.2). O valor de cobertura maior encontrado para o alinhamento proteico sugere que a sequência proteica de linalol/nerolidol sintase é mais conservada que a sequência nucleotídica, possivelmente devido o código genético ser degenerado. Apesar de um maior valor de cobertura para o alinhamento proteico, a identidade continuou sendo maior na região dos motivos conservados. Desta forma, foram desenhados três pares de *primers* para diferentes regiões da CDS de linalol/nerolidol sintase, tanto próximas aos motivos conservados quanto na região inicial da sequência.

Uma vez obtidos os *primers* para os controles e as terpeno sintases, foram realizados ensaios de qPCR para o cálculo da eficiência desses e validação da montagem e anotação do transcriptoma de *L. alba*. A análise da curva de amplificação de cada par de *primer* em diferentes concentrações de cDNA mostrou que todos foram eficientes com valores entre 0,9 a 1,07 de eficiência (*E*) com coeficiente de regressão  $R^2$  acima de 0,98 (tabela 4.4).

Enquanto a curva de amplificação fornece dados sobre a cinética de amplificação da sequência alvo, a curva de dissociação (*melting*) revela as características quanto a pureza e especificidade do produto final amplificado (Navarro *et al.*, 2015). A fim de assegurar a especificidade da reação, a curva de dissociação deve ser analisada para determinar o ponto de *melting*. A presença de dois ou mais picos sugere que mais do que uma sequência amplificada foi obtida, sendo assim, os *primers* pouco específicos. A ocorrência de dímeros também pode ser verificada a partir de picos de dissociação menores que os produtos específicos esperados (Valasek e Repa, 2005; Navarro *et al.*, 2015).

Todos os pares de *primers* desenhados neste trabalho, além de apresentarem valores de eficiência acima de 0,9 (90%), foram também capazes de amplificar sequências específicas para as 4 diluições de cDNA testadas, o que pode ser confirmado pela presença de apenas um pico nas curvas de dissociação (ANEXO). Este resultado juntamente com os valores de eficiência validam a montagem e anotação do transcriptoma de *Lippia alba*, mesmo na ausência de um genoma de referência.

Tabela 4.4: Sequência de *primers* usados e validados no presente estudo sobre anotação do transcriptoma de *Lippia alba*.

<i>Primer</i>	Sequência (F - <i>Forward</i> / R - <i>Reverse</i> )	Eficiência ( <i>E</i> )	$R^2$
ACT	<b>F:</b> CTGGAATGGTCAAGGCTGGAT <b>R:</b> GTGACGAGGGCGACCAACTA	1	0,99383
GADPH	<b>F:</b> CGAGGGTCTGATGACCACTGT <b>R:</b> CTCTCCAGTCCTTGGCTGATG	1,06	0,99462
TUBUL	<b>F:</b> AGGTTTGCAAGGGTTCTTGGT <b>R:</b> ACCGTTCCAAGAGCAACGAA	1,07	0,99149
NADH	<b>F:</b> TTGGTTACCCGAAGCTCATGT <b>R:</b> AAAAACCCGTAGGTTCCCAATT	1,07	0,99583
CITB	<b>F:</b> CTCATATTGTGCCGGAATGGT <b>R:</b> GGCTACACCTCCCGATTTGTC	1	0,98867
EF-Tu/EF-1 $\alpha$	<b>F:</b> TTGTCCCCATTTCTGGATTTG <b>R:</b> GAGGGTTGGTCCCTTGTACCA	0,97	0,99179
RNApol	<b>F:</b> AAAAGAGGGTTGACGGTTGGA <b>R:</b> TAGCTCAGCACCAATCAAGAATG	0,91	0,99068
DNApol	<b>F:</b> TGGATCCGAACAGGGAACCTTA <b>R:</b> GTGTGCGGATAGGCTTCTCAT	0,9	0,98
RPTL2	<b>F:</b> GGGAGGTCCGTTTGATATCCA <b>R:</b> CGGCTCTACCCAAACTTTTCTG	0,91	0,99029
SUCRS	<b>F:</b> TCGACATCACACCACGCATT <b>R:</b> CTCGAGCCGTTGACCACAA	0,98	0,99025
G6ISO	<b>F:</b> GGGTGGGTGGCAGAACATC <b>R:</b> CCCCAGCGAGCATTTCTTTA	0,92	0,99018
MALADH	<b>F:</b> CGAGGTGCTGCTATAATCAAAGC <b>R:</b> TCACGGATATGGTCACATGCA	0,9	0,98639
SerThr	<b>F:</b> GACCGTGTCCAAGAAGTTCCA <b>R:</b> CCCCACCAACGATCATC	0,95	0,99011
FRNDP	<b>F:</b> AACCTTCAAGGCGTCTTTGC <b>R:</b> TTGCTGGGATGAGCTTCGAT	0,97	0,99053
GERS	<b>F:</b> ACTGGCAACGATGCCCTACT <b>R:</b> CTCTCCTTGTCCCTCCTCTTC	1	0,99012
NERLIN-1	<b>F:</b> ACGAAAACCATAGAGCAAGTAAGGA <b>R:</b> AGCCGTTGGATCGTATCGAT	1,09	0,99612
NERLIN-2	<b>F:</b> TGGGAAGTGCAAAGGATGAG <b>R:</b> CATACGAGGAGATGTCGTGTTC	1,03	0,99277
NERLIN-3	<b>F:</b> GGTACACTTGAACAGCTTACTATCT <b>R:</b> GACTGATTCGTGGGTTCGATC	1,01	0,99042
GERGER	<b>F:</b> GGGAACCCGATTGAGGACTT <b>R:</b> ATCTCCTTCTCAGCGCATTAG	0,96	0,99034

Estão apresentados nas colunas os valores de eficiência (*E*) e coeficientes de regressão ( $R^2$ ) para cada par de *primer*. A eficiência deve variar entre 0,9 até 1,1, enquanto que quanto maior o valor  $R^2$  maior a significância do dado encontrado.

## 4.5 AVALIAÇÃO DO PERFIL DE EXPRESSÃO DE TERPENOSINTASES

Neste estudo, além do desenho e validação de *primers* para estudos de expressão por qPCR, foi realizado também um ensaio para avaliar o perfil de expressão de enzimas terpeno sintases encontradas *in silico*. Para isso, foram obtidas amostras de RNA total de folhas do quarto entrenó de três acessos triploides de *L. alba* conforme descrito na seção 3.8 em material e métodos. Os genes terpenóides avaliados foram referentes às enzimas: geraniol sintase (GERS), responsável pela síntese do monoterpénio geraniol; linalol/nerolidol sintase (NERLIN), a qual sintetiza linalol e nerolidol dependendo de sua localização; farnesil difosfato sintase (FRNDPS) que sintetiza o farnesil difosfato; e geranilgeranil difosfato redutase (GERGER), a qual sintetiza fitil difosfato um diterpenóide precursor de tocoferol e outros pigmentos fotossintetizantes.

Conforme apresentado na figura 4.14, os genes referentes às enzimas farnesil difosfato sintase e geranilgeranil redutase apresentaram valores de expressão maiores que os genes das outras duas enzimas avaliadas, isto pode ser explicado pelo fato dessas enzimas sintetizarem precursores da via de sesquiterpenóides e tocoferóis. A super expressão do gene de GERGER (figura 4.14.D), com *fold change* acima de 20 vezes, pode ser explicado pela importância do metabolismo de pigmentos fotossintetizantes para o desenvolvimento das plantas (Tanaka *et al.*, 1999). Observou-se ainda que a expressão do gene de linalol/nerolidol sintase foi relativamente maior que geraniol sintase, possivelmente devido ao fato de acessos triploides produzirem como principal terpenóide o linalol, conforme mostrado por Viccini e colaboradores (2014) durante estudo de avaliação dos compostos do óleo essencial de acessos de *L. alba* com diferentes ploidias.

Conforme apresentado na Tabela 4.4 na seção anterior, os *primers* NERLIN-1, NERLIN-2 e NERLIN-3 possuíram valores de eficiência acima de 1 e  $R^2$  de 99%. Entretanto, para a análise da expressão de linalol/nerolidol sintase foi selecionado apenas o par de *primer* NERLIN-1, pois este apresentou menor variação no valor de Ct. Pelo mesmo motivo, foi selecionado como controle endógeno o gene EF-Tu/EF-T $\alpha$  para normalização dos valores de expressão das enzimas terpeno sintases para os acessos triploides avaliados neste estudo.

Apesar das plantas avaliadas apresentarem perfil similar de expressão, existem variações nos níveis de expressão de cada terpeno sintase para os quimiotipos testados (figura 4.14). Isto pode estar relacionado a mudanças e inconstâncias do microambiente que as plan-

tas se encontravam, uma vez que as mesmas foram coletadas diretamente do campo e não estavam em condições totalmente homogêneas. Contudo, os resultados obtidos para os três acessos triploides de *L. alba* testados, sugerem que a produção de um óleo essencial rico no composto linalol pode estar relacionado ao genótipo da planta. Enquanto que a quantidade de óleo produzido seria influenciada por fatores ambientais. Além disso, a amplificação das sequências dos genes terpenóides, obtidos neste trabalho, em três diferentes acessos de *L. alba* permitiu validar tanto os *primers* quanto o sequenciamento e montagem do transcriptoma.

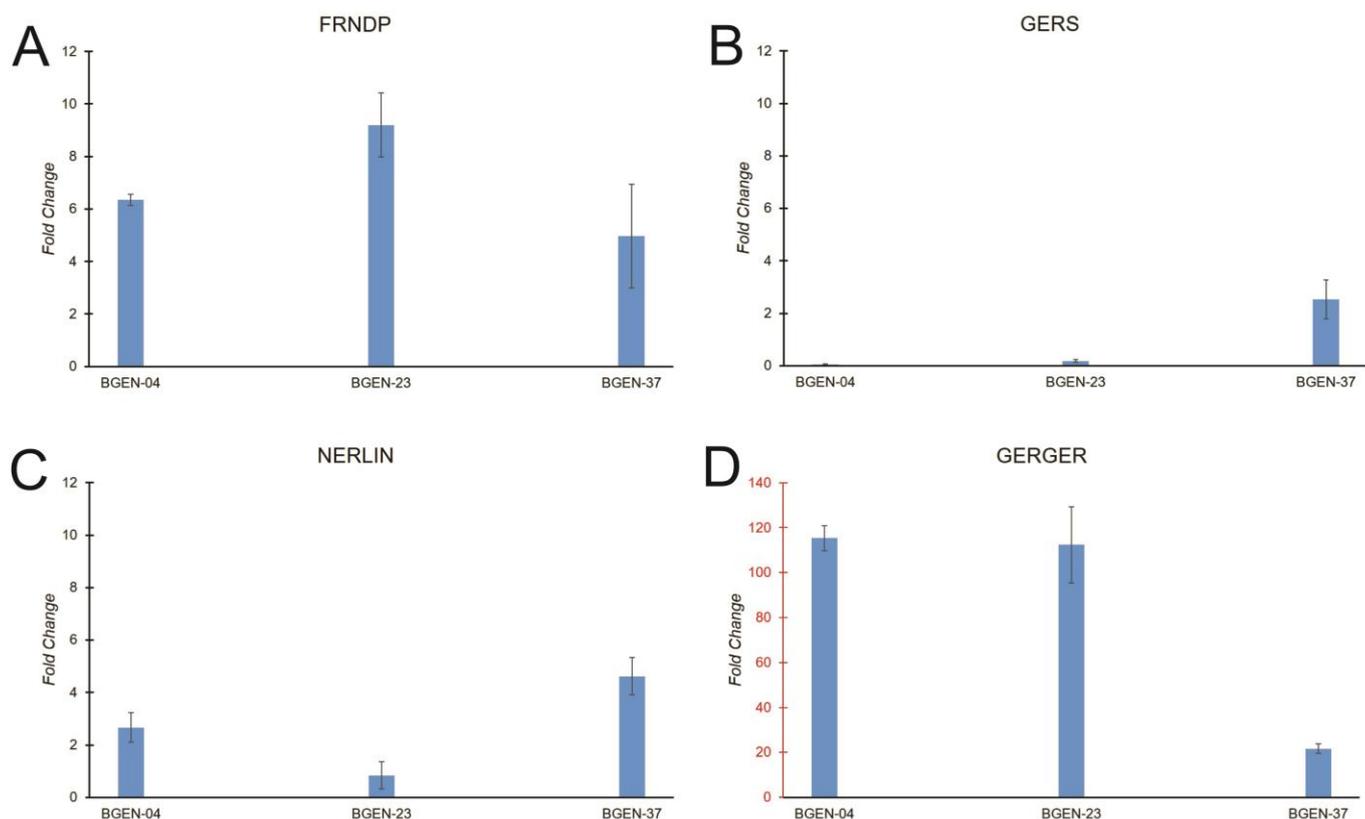


Figura 4.14: qPCR de quatro genes de terpeno sintase para três acessos de *Lippia alba*. O desvio-padrão foi calculado em três replicatas e o gene EF-Tu/EF-1 $\alpha$  foi usado como controle endógeno para a normalização dos dados. Para avaliar a expressão de nerolidol/linalol sintase foi usado o *primer* NERLIN-1. O *fold change* foi calculado a partir da fórmula  $2^{(-\Delta Ct)}$ , onde  $\Delta Ct = Ct_{(gene\ alvo)} - Ct_{(gene\ controle)}$ .

## 5 CONCLUSÃO

Neste trabalho, foi realizado a caracterização do transcriptoma do acesso triploide BGEN-04 de *Lippia alba*, a partir do sequenciamento de RNA total usando a plataforma *Illumina Miseq*. As estatísticas referentes aos passos de montagem *de novo* e procura de ORF foram considerados satisfatórios, apresentando valores próximos aos relatados na literatura. Isto mostra que o pacote *Trinity* foi eficaz na montagem do transcriptoma de *L. alba*, mesmo na ausência de um genoma de referência.

Os métodos de montagem *de novo* e anotação funcional usados aqui permitiram obter transcritos de diferentes categorias biológicas em *L. alba*. Das diferentes vias identificadas, pode-se citar algumas vias de interesse como as de metabolismo fotossintético, fito hormônios, embriogênese somática e, com maior destaque e nesse trabalho, as vias de síntese de terpenóides.

Com relação a análise *in silico* dos transcritos de terpeno sintase, foi possível obter sequências completas dessas enzimas (*e.g.*, geraniol sintase e a linalol/nerolidol sintase), isto é, sequências que apresentam os motivos conservados **DDXXD** e **NSE/DTE**. Além disso, foi possível desenhar *primers* para estudos de qPCR para genes controles endógenos e de terpeno sintases, os quais foram validados e tiveram valor de eficiência satisfatório. A partir deste estudo foi possível realizar a análise do perfil de expressão de quatro genes terpeno sintases para os acessos triploides BGEN-04, BGEN-23 e BGEN-37 de *L. alba*.

Os perfis de expressão das terpeno sintases avaliadas nos três acessos foram similares, porém foi observado variações na expressão individual para cada enzima entre as plantas avaliadas. Essa variação, possivelmente está relacionada ao fato dos acessos não estarem em condições uniformes durante o cultivo e a coleta.

Através da comparação entre os perfis de expressão das enzimas geraniol sintase (GERS) e linalol/nerolidol sintase (NERLIN) foi possível verificar que a NERLIN teve uma maior expressão em relação a GERS o que corrobora os dados da literatura relacionados a maior porcentagem do composto linalol em acessos triploides. Os dados gerados com esse trabalho podem auxiliar em novos estudos de expressão gênica em acessos de *L. alba* com diferentes ploidias, colaborando para a verificação da diferença de composição dos óleos essenciais produzidos por cada quimiotipo.

## 6 REFERÊNCIAS BIBLIOGRÁFICAS

Abe, I. 2007. Enzymatic synthesis of cyclic triterpenes. *Natural Product Reports*, 24 (6): 1311–1331. Doi:10.1039/b616857b

Adiconis, X., Borges-Rivera, D., Satija, R., Deluca, D. S., Busby, M. A., Berlin, A. M., Sivachenki, A., Thompson, D.A., Wysoker, A., Fennell, T., Gnirke, A., Pochet, N., Regev, A., Levin, J. Z., 2013. Comparative analysis of RNA sequencing methods for degraded or low-input samples. *Nature methods*, 10 (7): 623–9. Doi:10.1038/nmeth.2483

Aguiar, J. S., Costa, M. C. C. D., Nascimento, S. C., Sena, K. X. F. R., 2008. Atividade antimicrobiana de *Lippia alba* (Mill.) N. E. Brown (Verbenaceae). *Brazilian Journal of Pharmacognosy*, 18 (3): 436–440. Doi:10.1590/S0102-695X2008000300018

Aharoni, A., Galili, G., 2011. Metabolic engineering of the plant primary-secondary metabolism interface. *Current Opinion in Biotechnology*, 22 (2): 239–44. Doi:10.1016/j.copbio.2010.11.004

Arya, M., Shergill, I. S., Williamson, M., Gommersall, L., Arya, N., Patel, H. R. 2005. Basic principles of real-time quantitative PCR. *Expert Rev. Mol. Diagn*, 5 (2): 209–219. Doi:10.1586/14737159.5.2.209

Blanco, M. A., Colareda, G. A., Van Baren, C., Bandoni, A. L., Ringuélet, J., Consolini, A. E., 2013. Antispasmodic effects and composition of the essential oils from two South American chemotypes of *Lippia alba*. *Journal of Ethnopharmacology*, 149 (3): 803–809. Doi:10.1016/j.jep.2013.08.007

Bohlmann, J., Meyer-Gauen, G., Croteau, R., 1998. Plant terpenoid synthases: molecular biology and phylogenetic analysis. *Proceedings of the National Academy of Sciences of the United States of America*, 95 (8): 4126–4133. Doi:10.1073/pnas.95.8.4126

Buh Gašparič, M., Tengs, T., La Paz, J. L., Holst-Jensen, A., Pla, M., Esteve, T., Žel, J., Gruden, K., 2010. Comparison of nine different real-time PCR chemistries for qualitative and quantitative applications in GMO detection. *Analytical and Bioanalytical Chemistry*, 396 (6): 2023–2029. Doi:10.1007/s00216-009-3418-0

Bustin, S. A., Benes, V., Garson, J. A., Hellemans, J., Huggett, J., Kubista, M., Muller, R., Nolan, T., Pfaffl, M. W., Shipley, G. L., Vandesompele, J., Wittwer, C. T., 2009. The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. *Clinical Chemistry*, 55 (4): 611–622. Doi:10.1373/clinchem.2008.112797

Campos, M. D., Frederico, A. M., Nothnagel, T., Arnholdt-Schmitt, B., Cardoso, H., 2015. Selection of suitable reference genes for reverse transcription quantitative real-time PCR studies on different experimental systems from carrot (*Daucus carota* L.). *Scientia Horticulturae*, 186: 115-123.

Cao, R., Zhang, Y., Mann, F. M., Huang, C., Mukkamala, D., Hudock, M. P., Mead, M. E., Wang, K., Fu-Yang, L., Reuben, J. P., Oldfield, E., 2010. Diterpene cyclases and the nature of

the isoprene fold. *Proteins: Structure, Function and Bioinformatics*, 78 (11): 2417–2432.  
Doi:10.1002/prot.22751

Castanera, R., López-Varas, L., Pisabarro, A. G., Ramírez, L., 2015. Validation of reference genes for transcriptional analyses in *Pleurotus ostreatus* by using reverse transcription-quantitative PCR. *Applied and environmental microbiology*, 81(12): 4120-4129.

Chen, F., Tholl, D., Bohlmann, J., Pichersky, E., 2011. The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. *The Plant Journal : For Cell and Molecular Biology*, 66 (1): 212–29.  
Doi:10.1111/j.1365-313X.2011.04520.x

Chen, G., Yin, K. P., Wang, C., Shi, T. L., 2011. De novo transcriptome assembly of RNA-Seq reads with different strategies. *Science China Life Sciences*, 54 (12): 1129–1133.  
Doi:10.1007/s11427-011-4256-9

Chies, C., Branco, C., Scola, G., Agostini, F., Gower, A., Salvador, M., 2013. Antioxidant Effect of *Lippia alba* (Miller) N. E. Brown. *Antioxidants*, 2 (4): 194–205.  
Doi:10.3390/antiox2040194

Christianson, D. W., 2006. Structural biology and chemistry of the terpenoid cyclases. *Chemical Reviews*, 106 (8): 3412–3442. Doi:10.1021/cr050286w

Coman, D., Rütimann, P., Gruissem, W., 2014. Plant Isoprenoids. (Manuel Rodríguez-Concepción, Ed.) *Methods in molecular biology* (Clifton, N.J.) (Vol. 1153). Doi:10.1007/978-1-4939-0606-2

Degenhardt, J., Köllner, T. G., Gershenzon, J., 2009. Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. *Phytochemistry*, 70 (15-16): 1621–1637. Doi:10.1016/j.phytochem.2009.07.030

Djami-Tchatchou, a. T., Straker, C. J., 2012. The isolation of high quality RNA from the fruit of avocado (*Persea americana* Mill.). *South African Journal of Botany*, 78 : 44–46.  
Doi:10.1016/j.sajb.2011.04.009

Duan, J., Xia, C., Zhao, G., Jia, J., Kong, X., 2012. Optimizing de novo common wheat transcriptome assembly using short-read RNA-Seq data. *BMC Genomics*, 13 (1): 392.  
Doi:10.1186/1471-2164-13-392

Dudareva, N., Klempien, A., Muhlemann, J. K., Kaplan, I., 2013. Biosynthesis, function and metabolic engineering of plant volatile organic compounds. *New Phytologist*.  
Doi:10.1111/nph.12145

Facchini, P. J., Bohlmann, J., Covello, P. S., De Luca, V., Mahadevan, R., Page, J. E., Dae-Kyun, R., Sensen, C. W., Storms, R., Martin, V. J. J., 2012. Synthetic biosystems for the production of high-value plant metabolites. *Trends in Biotechnology*, 30 (3): 127–131.  
Doi:10.1016/j.tibtech.2011.10.001

Farias, D. R., 2013. Desenvolvimento de Ferramentas de Bioinformática para Montagem e Prospecção de Genomas Selvagens Visando o Melhoramento de Arroz. Tese de Doutorado, Universidade Federal de Pelotas.

- Fisher, K., Phillips, C., 2008. Potential antimicrobial uses of essential oils in food: is citrus the answer? *Trends in Food Science and Technology*, 19 (3): 156–164.  
Doi:10.1016/j.tifs.2007.11.006
- Forero-Peñuela, L. Y., Biasi, L. A., Bizzo, H. R., Souza, M. S. De, Deschamps, C., 2013. Potential of *Lippia alba* (Mill.) N.E. Br. Ex Britt. & P. Wilson, as available source of linalool in southern Brazil. *Journal of Essential Oil Research*, 25 (6): 464–467.  
Doi:10.1080/10412905.2013.796495
- Fornari, T., Vicente, G., Vázquez, E., García-Risco, M. R., Reglero, G., 2012. Isolation of essential oil from different plants and herbs by supercritical fluid extraction. *Journal of Chromatography A*, 1250 : 34–48. Doi:10.1016/j.chroma.2012.04.051
- Gao, Y., Honzatko, R. B., Peters, R. J., 2012. Terpenoid synthase structures: a so far incomplete view of complex catalysis. *Natural Product Reports*, 29 (10): 1153.  
Doi:10.1039/c2np20059g
- Giulietti, A., Overbergh, L., Valckx, D., Decallonne, B., Bouillon, R., Mathieu, C., 2001. An overview of real-time quantitative PCR: applications to quantify cytokine gene expression. *Methods (San Diego, Calif.)*, 25 (4): 386–401. Doi:10.1006/meth.2001.1261
- Goecks, J., Nekrutenko, A., Taylor, J., 2010. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biology*, 11 (8): R86. Doi:10.1186/gb-2010-11-8-r86
- Guedes, F. A. de F., 2011. Pirosequenciamento do Transcriptoma de Folhas de *Lippia alba* por Meio da Plataforma 454 GS FLX (ROCHE). Dissertação de Mestrado, Universidade Federal de Juiz de Fora.
- Gutensohn, M., Nguyen, T. T. H., McMahon, R. D., Kaplan, I., Pichersky, E., Dudareva, N., 2014. Metabolic engineering of monoterpene biosynthesis in tomato fruits via introduction of the non-canonical substrate neryl diphosphate. *Metabolic Engineering*, 24 : 107–116.  
Doi:10.1016/j.ymben.2014.05.008
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., MacManes, M. D., Ott, M., Orvis, J., Porchet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., Henschel, R., LeDuc, R. D., Friedman, N., Regev, A., 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, 8 : 1494–512. Doi:10.1038/nprot.2013.084
- Han, X. J., Wang, Y. D., Chen, Y. C., Lin, L. Y., Wu, Q. K., 2013. Transcriptome Sequencing and Expression Analysis of Terpenoid Biosynthesis Genes in *Litsea cubeba*. *Plos ONE*, 8 (10): 1–14. Doi:10.1371/journal.pone.0076890
- Heinzman, B. M., Barros, F. M. C. de B., 2007. Potencial das plantas nativas brasileiras para o desenvolvimento de fitomedicamentos tendo como exemplo *Lippia alba* (MILL.) N.E. Brown (verbenaceae). *Saúde*, 33 (1): 43–48.

- Hemmerlin, A., Hoeffler, J. F., Meyer, O., Tritsch, D., Kagan, I. A., Grosdemange-Billiard, C., Rohmer, M., Bach, T. J., 2003. Cross-talk between the cytosolic mevalonate and the plastidial methylerythritol phosphate pathways in tobacco bright yellow-2 cells. *Journal of Biological Chemistry*, 278 (29): 26666–26676. Doi:10.1074/jbc.M302526200
- Hennebelle, T., Sahpaz, S., Joseph, H., Bailleul, F., 2008. Ethnopharmacology of *Lippia alba*. *Journal of Ethnopharmacology*, 116 : 211–222. Doi:10.1016/j.jep.2007.11.044
- Higuchi, R., Dollinger, G., Walsh, P. S., Griffith, R., 1992. Simultaneous amplification and detection of specific DNA sequences. *Bio/technology*, 10 (4), 413-417.
- Holland, P. M., Abramson, R. D., Watson, R., Will, S., Saiki, R. K., Gelfand, D. H., 1992. Detection of specific polymerase chain-reaction product by utilizing the 5'- 3' exonuclease activity of *thermus-aquaticus* DNA-Polymerase. *Clinical Chemistry*, 38 (3), 462-463.
- Huang, H. H., Xu, L. L., Tong, Z. K., Lin, E. P., Liu, Q. P., Cheng, L. J., Zhu, M. Y., 2012. De novo characterization of the Chinese fir (*Cunninghamia lanceolata*) transcriptome and analysis of candidate genes involved in cellulose and lignin biosynthesis. *BMC Genomics*, 13 (1): 648. Doi:10.1186/1471-2164-13-648
- Iijima, Y., Gang, D. R., Fridman, E., Lewinsohn, E., Pichersky, E., 2004. Characterization of geraniol synthase from the peltate glands of sweet basil. *Plant Physiology*, 134 (1): 370–379. Doi:10.1104/pp.103.032946
- Jannuzzi, H., Mattos, J. K. A, Silva, D. B., Gracindo, L. A M., Vieira, R. F., 2011. Avaliação agronômica e química de dezessete acessos de erva-cidreira [*Lippia alba* (Mill.) N.E.Brown] - quimiotipo citral, cultivados no Distrito Federa. *Revista Brasileira de Plantas Mediciniais*, 13 (3): 258–264.
- Kampranis, S. C., Ioannidis, D., Purvis, A., Mahrez, W., Ninga, E., Katerelos, N. A., Anssour, S., Dunwell, J. M., Degenhardt, J., Markis, A. M., Goodenough, J. W., Johnson, C. B., 2007. Rational conversion of substrate and product specificity in a *Salvia* monoterpene synthase: structural insights into the evolution of terpene synthase function. *The Plant Cell*, 19 (6): 1994–2005. Doi:10.1105/tpc.106.047779
- Kawada, T., Goto, T., Takahashi, N., Hirai, S., 2010. Various Terpenoids derived from herbal and dietary plants function as PPAR modulators and regulate carbohydrate and lipid metabolism. *PPAR Research*, 2010. Doi:10.1155/2010/483958
- Kozera, B., Rapacz, M., 2013. Reference genes in real-time PCR. *Journal of applied genetics*, 54(4): 391-406.
- Lange, B. M., 2014. The Evolution of Plant Secretory Structures and the Emergence of Terpenoid Chemical Diversity. *Annual Review of Plant Biology*, 66 (1): 150126112146002. Doi:10.1146/annurev-arplant-043014-114639
- Lee, W. J., Jang, S. J., Lee, S. C., Park, J. S., Jeon, R. H., Subbarao, R. B., Bharti, D., Shin, J. K., Park, B. W., Rho, G. J., 2015. Selection of reference genes for quantitative real-time polymerase chain reaction in porcine embryos. *Reproduction, Fertility and Development*. Doi:10.1071/RD14393

- Levin, R. E., 2005. The Application of Real-Time PCR to Food and Agricultural Systems. A Review. *Food Biotechnology*, 18 (1): 97–133. Doi:10.1081/FBT-120030386
- Liang, P., Guo, Y., Zhou, X., Gao, X., 2014. Expression profiling in *Bemisia tabaci* under insecticide treatment: indicating the necessity for custom reference gene selection. *PloS one*, 9(1): e87514.
- Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., Law, M., 2012. Comparison of next-generation sequencing systems. *Journal of Biomedicine and Biotechnology*, 2012 . Doi:10.1155/2012/251364
- Liu, T., Zhu, S., Tang, Q., Chen, P., Yu, Y., Tang, S., 2013. De novo assembly and characterization of transcriptome using Illumina paired-end sequencing and identification of cesa gene in ramie (*Boehmeria nivea* L. Gaud). *BMC Genomics*, 14 (1): 125. Doi:10.1186/1471-2164-14-125
- López, M. A., Stashenko, E. E., Fuentes, J. L., 2011. Chemical composition and antigenotoxic properties of *Lippia alba* essential oils. *Genetics and Molecular Biology*, 34 (3): 479–488. Doi:10.1590/S1415-47572011005000030
- Lulin, H., Xiao, Y., Pei, S., Wen, T., Shangqin, H., 2012. The First Illumina-Based De Novo Transcriptome Sequencing and Analysis of Safflower Flowers. *Plos one*. Doi:10.1371/journal.pone.0038653
- Machado, T. F., Pereira, R. C. A., Batista, V. C. V., 2014. Seasonal variability of the antimicrobial activity of the essential oil of *Lippia alba*. *Revista ciência agrônômica*, 45 : 515–519.
- Mardis, E. R., 2008. Next-generation DNA sequencing methods. *Annual Review of Genomics and Human Genetics*, 9: 387–402. Doi:10.1146/annurev.genom.9.081307.164359
- Martin, D. M., Aubourg, S., Schouwey, M. B., Daviet, L., Schalk, M., Toub, O., Lund, S. T., Bohlmann, J., 2010. Functional annotation, genome organization and phylogeny of the grapevine (*Vitis vinifera*) terpene synthase gene family based on genome assembly, flcDNA cloning, and enzyme assays. *BMC Plant Biology*, 10 (1): 226. Doi:10.1186/1471-2229-10-226
- Martin, J. A., Wang, Z., 2011. Next-generation transcriptome assembly. *Nature Reviews. Genetics*, 12 (10): 671–682. Doi:10.1038/nrg3068
- Mcgarvey, D. J., Croteau, R., 1995. Terpenoid metabolism. *The Plant Cell*, 7 (7): 1015–1026. Doi:10.1105/tpc.7.7.1015
- Metzker, M. L., 2010. Sequencing technologies - the next generation. *Nature Reviews Genetics*, 11 (1): 31–46. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/19997069>
- Muhlemann, J. K., Klempien, A., Dudareva, N., 2014. Floral volatiles: From biosynthesis to function. *Plant, Cell and Environment*, 37 (8): 1936–1949. Doi:10.1111/pce.12314
- Mylykangas, S., Buenrostro, J., Ji, H. P., 2012. Bioinformatics for High Throughput Sequencing. *Bioinformatics for High Throughput Sequencing*, 11–25. Doi:10.1007/978-1-4614-0782-9

- Nagegowda, D. A., 2010. Plant volatile terpenoid metabolism: Biosynthetic genes, transcriptional regulation and subcellular compartmentation. *FEBS Letters*.  
Doi:10.1016/j.febslet.2010.05.045
- Nagegowda, D. A., Gutensohn, M., Wilkerson, C. G., Dudareva, N., 2008. Two nearly identical terpene synthases catalyze the formation of nerolidol and linalool in snapdragon flowers. *Plant Journal*, 55 : 224–239. Doi:10.1111/j.1365-313X.2008.03496.x
- Navarro, E., Serrano-Heras, G., Castaño, M. J., Solera, J., 2015. Real-time PCR detection chemistry. *Clinica Chimica Acta*, 439 : 231–250. Doi:10.1016/j.cca.2014.10.017
- Neto, A. C., Netto, J. C., Pereira, P. S., Pereira, A. M. S., Taleb-Contini, S. H., França, S. C., Marques, M. O., Belebony, R. O., 2009. The role of polar phytocomplexes on anticonvulsant effects of leaf extracts of *Lippia alba* (Mill.) N.E. Brown chemotypes. *The Journal of Pharmacy and Pharmacology*, 61 (7): 933–939. Doi:10.1211/jpp/61.07.0013
- Ombito, J. O., Salano, E. N., Yegon, P. K., Ngetich, K., Mwangi, E. M., 2014. A review on the chemistry of some species of genus *Lippia* ( Verbenaceae family ). *Journal of Scientific and Innovative Research*, 3 (4): 460–466.
- Pandeló, D., Melo, T. D., Singulani, J. L., Guedes, F. A. de F., Machado, M. A., Coelho, C. M., Viccini, L. F., Santos, M. O., 2012. Oil production at different stages of leaf development in *Lippia alba*. *Revista Brasileira de Brazilian Journal of Pharmacognosy*, 22 (3): 497–501. Doi:10.1590/S0102- 695X
- Pareek, C. S., Smoczynski, R., Tretyn, A., 2011. Sequencing technologies and genome sequencing. *Journal of Applied Genetics*, 52 (4): 413–435. Doi:10.1007/s13353-011-0057-x
- Pascual, M. E., Slowing, K., Carretero, M. E., Villar, Á., 2001. Antiulcerogenic activity of *Lippia alba* (Mill.) N. E. Brown (Verbenaceae). *Farmaco*, 56 (5-7): 501–504. Doi:10.1016/S0014-827X(01)01086-2
- Passos, M. A, de Cruz, V. O., Emediato, F. L., de Teixeira, C. C., Azevedo, V. C., Brasileiro, A. C., Amorim, E. P., Ferreira, C. F., Martins, N. F., Togawa, R. C., Pappas Jr, G. J., Silva Jr, O. B., Miller, R. N., 2013. Analysis of the leaf transcriptome of *Musa acuminata* during interaction with *Mycosphaerella musicola*: gene assembly, annotation and marker development. *BMC Genomics*, 14 (1): 78. Doi:10.1186/1471-2164-14-78
- Peters, R. J., 2010. Two rings in them all: the labdane-related diterpenoids. *Natural Product Reports*, 27 (11): 1521–1530. Doi:10.1039/c0np00019a
- Pierre, P. M. O., Sousa, S. M., Davide, L. C., Machado, M. A., Viccini, L. F., 2011. Karyotype analysis, DNA content and molecular screening in *Lippia alba* (Verbenaceae). *Anais Da Academia Brasileira de Ciencias*, 83 (3): 993–1005. Doi:10.1590/S0001-37652011005000012
- Postollec, F., Falentin, H., Pavan, S., Combrisson, J., Sohier, D., 2011. Recent advances in quantitative PCR (qpcr) applications in food microbiology. *Food Microbiology*, 28 (5): 848–861. Doi:10.1016/j.fm.2011.02.008

- Reid, K. E., Olsson, N., Schlosser, J., Peng, F., Lund, S. T., 2006. An optimized grapevine RNA isolation procedure and statistical determination of reference genes for real-time RT-PCR during berry development. *BMC plant biology*, 6(1), 1.
- Reis, A. C., Sousa, S. M., Vale, A. A., Pierre, P. M. O., Franco, A. L., Campos, J. M. S., Vieira, R. F., Viccini, L. F., 2014. *Lippia alba* (Verbenaceae): A new tropical autoployploid complex? *American Journal of Botany*, 101 (6): 1002–1012. Doi:10.3732/ajb.1400149
- Ruiz-Sola., 2011. Carotenoid Biosynthesis in Arabidopsis: A Colorful Pathway. *The Arabidopsis Book*, 1–28. Doi:10.1199/tab.0158
- Savithramma, N., Rao, M. L., Suhulatha, D., 2011. Screening of traditional medicinal plants for secondary metabolites, 2 (4): 643–647.
- Sharkey, T. D., Gray, D. W., Pell, H. K., Breneman, S. R., Topper, L., 2013. Isoprene synthase genes form a monophyletic clade of acyclic terpene synthases in the Tps-b terpene synthase family. *Evolution*, 67 (4): 1026–1040. Doi:10.1111/evo.12013
- Siani, A. C., Tappin, M. R. R., Ramos, M. F. S., Mazzei, J. L., Ramos, M. C. K. V, De Aquino Neto, F. R., Frighetto, N., 2002. Linalool from *Lippia alba*: Study of the reproducibility of the essential oil profile and the enantiomeric purity. *Journal of Agricultural and Food Chemistry*, 50 (12): 3518–3521. Doi:10.1021/jf011592n
- Silva, F. S., Menezes, P. M. N., Sá, P. G. S. De, Oliveira, A. L. De S., Souza, E. A. A., Bamberg, V. M., Oliveira, H. R., Oliveira, S. A., Araújo, R. E., Uetanabaro, A. P. T., Silva, T. R. S., Almeida, J. R. G. S., Lucchese, A. M., 2015. Pharmacological Basis for Traditional Use of the *Lippia thymoides*. *Evidence-Based Complementary and Alternative Medicine*, 2015: 1–10. Doi:10.1155/2015/463248
- Singulani, J. L., Silva, P. S., Raposo, N. R. B., Siqueira, E. P., Zani, C. L., Alves, T. M. A., Viccini, L. F., 2012. Chemical composition and antioxidant activity of lippia species. *Journal of medicinal plants research*, 6 (27) 4416–4422. Doi:10.5897/jmpr12.533
- Soon, W. W., Hariharan, M., Snyder, M. P., 2013. High-throughput sequencing for biology and medicine. *Molecular Systems Biology*, 9 (640): 640. Doi:10.1038/msb.2012.61
- Strickler, S. R., Bombarely, A., Mueller, L. A., 2012. Designing a transcriptome next-generation sequencing project for a nonmodel plant species. *American Journal of Botany*, 99 (June 2011): 257–266. Doi:10.3732/ajb.1100292
- Sudheesh, S., Sawbridge, T. I., Cogan, N. O., Kennedy, P., Forster, J. W., Kaur, S., 2015. De novo assembly and characterisation of the field pea transcriptome using RNA-Seq. *BMC Genomics*, 16 (1): 611. Doi:10.1186/s12864-015-1815-7
- Tanaka, R., Oster, U., Kruse, E., Rüdiger, W., Grimm, B., 1999. Reduced Activity of Geranylgeranyl Reductase Leads to Loss of Chlorophyll and Tocopherol and to Partially Geranylgeranylated Chlorophyll in Transgenic Tobacco Plants Expressing Antisense RNA for Geranylgeranyl Reductase1. *Plant Physiology*, 120 (3): 695–704. Doi:10.1104/pp.120.3.695
- Tavares, E. S., Julião, L. S., Lopes, D., Bizzo, H. R., Lage, C. L. S., Leitão, S. G., 2005. Análise do óleo essencial de folhas de três quimiotipos de *Lippia alba* (Mill.) N. E. Br. (Verbenaceae).

ceae) cultivados em condições semelhantes. *Revista Brasileira de Farmacognosia*.  
Doi:10.1590/S0102-695X2005000100002

Tavares, I. B., Momenté, V. G., Nascimento, I. R., 2011. *Lippia alba*: estudos químicos, etnofarmacológicos e agrônômicos. *Revista Brasileira de Tecnologia Aplicada Nas Ciências Agrárias*, 4 (1): 204–220. Doi:10.5777/paet.V4.N1.12

Taylor, S., Wakem, M., Dijkman, G., Alsarraj, M., Nguyen, M., 2010. A practical approach to RT-qpcr—Publishing data that conform to the MIQE guidelines. *Methods*, 50 (4): S1–S5.  
Doi:10.1016/j.ymeth.2010.01.005

Teles, S., Pereira, J. A., Santos, C. H. B., Menezes, R. V., Malheiro, R., Lucchese, A. M., Silva, F., 2012. Geographical origin and drying methodology may affect the essential oil of *Lippia alba* (Mill) N.E. Brown. *Industrial Crops and Products*, 37 (1): 247–252.  
Doi:10.1016/j.indcrop.2011.12.029

Thimmappa, R., Geisler, K., Louveau, T., O'Maille, P., Osbourn, A., 2014. Triterpene biosynthesis in plants. *Annual Review of Plant Biology*, 65 (January): 225–57.  
Doi:10.1146/annurev-arplant-050312-120229

Tholl, D., 2011. Terpene Specialized Metabolism in *Arabidopsis thaliana*. *The Arabidopsis Book*, (9). Doi:10.1199/tab.0143

Tholl, D., 2006. Terpene synthases and the regulation, diversity and biological roles of terpene metabolism. *Current Opinion in Plant Biology*, 9 (3): 297–304. Doi:10.1016/j.pbi.2006.03.014

Tholl, D., 2015. Biosynthesis and Biological Functions of Terpenoids in Plants. *Advances in Biochemical Engineering/biotechnology*, 123 : 127–141. Doi:10.1007/10-2014-295

Thornton, B., Basu, C., 2011. Real-time PCR (qpcr) primer design using free online software. *Biochemistry and Molecular Biology Education*, 39 (2): 145–154. Doi:10.1002/bmb.20461

Tichopad, A., Didier, A., Pfaffl, M. W., 2004. Inhibition of real-time RT-PCR quantification due to tissue-specific contaminants. *Molecular and Cellular Probes*, 18 : 45–50.  
Doi:10.1016/j.mcp.2003.09.001

Valasek, M. A., Repa, J. J., 2005. The power of real-time PCR, 151–159.  
Doi:10.1152/advan.00019.2005.

Verma, P., Shah, N., Bhatia, S., 2013. Development of an expressed gene catalogue and molecular markers from the de novo assembly of short sequence reads of the lentil (*Lens culinaris* Medik.) Transcriptome. *Plant Biotechnology Journal*, 11 (7): 894–905.  
Doi:10.1111/pbi.12082

Viccini, L. F., Pierre, P. M. O., Praça, M. M., Costa, D. C. S. D., Romanel, E. D. C., Sousa, S. M. D., Peixoto, P. H. P., Salimena, F. R. G., 2005. Chromosome numbers in the genus *Lippia* (Verbenaceae). *Plant Systematics and Evolution*, 256 (1-4): 171–178. Doi:10.1007/s00606-005-0351-3

Viccini, L. F., Silveira, R. S., Vale, A. A., Campos, J. M. S., Reis, A. C., Santos, M.O., Campos, V. R., Carpanez, A. G., Grazul, R. M., 2014. Citral and linalool content has been corre-

lated to DNA content in *Lippia alba* (Mill.) N.E. Brown (Verbenaceae). *Industrial Crops and Products*, 59 : 14–19. doi:10.1016/j.indcrop.2014.04.028

Vieira, R. F., 1999. Conservation of medicinal and aromatic plants in Brazil. *Perspectives on New Crops and New Uses*, 152 : 152–159.

Voelkerding, K. V., Dames, S. A., Durtschi, J. D., 2009. Next-generation sequencing: from basic research to diagnostics. *Clinical Chemistry*, 55 (4): 641–658. Doi:10.1373/clinchem.2008.112789

Vranová, E., Coman, D., Gruissem, W., 2013. Network analysis of the MVA and MEP pathways for isoprenoid synthesis. *Annual Review of Plant Biology*, 64 : 665–700. Doi:10.1146/annurev-arplant-050312-120116

Wang, H., Chen, S., Jiang, J., Zhang, F., Chen, F. 2015. Reference gene selection for cross-species and cross-ploidy level comparisons in *Chrysanthemum* spp. *Scientific reports*, 5.

Wang, Y., Pan, Y., Liu, Z., Zhu, X., Zhai, L., Xu, L., Yu, R., Gong, Y., Liu, L., 2013. De novo transcriptome sequencing of radish (*Raphanus sativus* L.) and analysis of major genes involved in glucosinolate metabolism. *BMC Genomics*, 14 (1): 836. Doi:10.1186/1471-2164-14-836

Wang, Y., Xu, L., Chen, Y., Shen, H., Gong, Y., Limera, C., Liu, L., 2013. Transcriptome profiling of radish (*Raphanus sativus* L.) Root and identification of genes involved in response to Lead (Pb) stress with next generation sequencing. *Plos One*, 8 (6): e66539. Doi:10.1371/journal.pone.0066539

Xiao, M., Zhang, Y., Chen, X., Lee, E. J., Barber, C. J. S., Chakrabarty, R., Desgagné-Penix, I., Haslam, T.M., Kim, Y., Liu, E., Macnevin, G., Masada-Atsumi, S., Reed, D.W., Stout, J.M., Zerbe, P., Zhang, Y., Bohlmann, J., Covello, P.S., De Luca, V., Page, J.E., Ro, D., Martin, V.J.J., Facchini, P.J., Sensen, C. W., 2013. Transcriptome analysis based on next-generation sequencing of non-model plants producing specialized metabolites of biotechnological interest. *Journal of Biotechnology*, 166 (3): 122–34. doi:10.1016/j.jbiotec.2013.04.004

Xu, L., Zhu, L., Tu, L., Liu, L., Yuan, D., Jin, L., Long, L., Zhang, X., 2011. Lignin metabolism has a central role in the resistance of cotton to the wilt fungus *Verticillium dahliae* as revealed by RNA-Seq-dependent transcriptional analysis and histochemistry. *Journal of Experimental Botany*, 62 (15): 5607–5621. Doi:10.1093/jxb/err245

Yu, D., Tang, H., Zhang, Y., Du, Z., Yu, H., Chen, Q., 2012. Comparison and Improvement of Different Methods of RNA Isolation from Strawberry (*Fragaria \* ananassa*). *Journal of Agricultural Science*, 4 (7): 51–56. Doi:10.5539/jas.v4n7p51

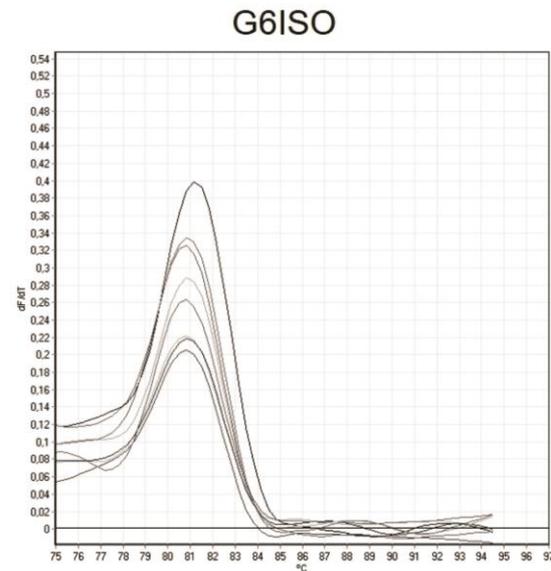
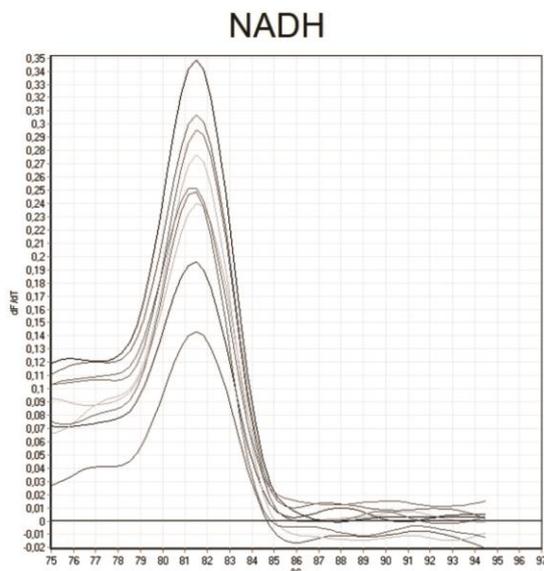
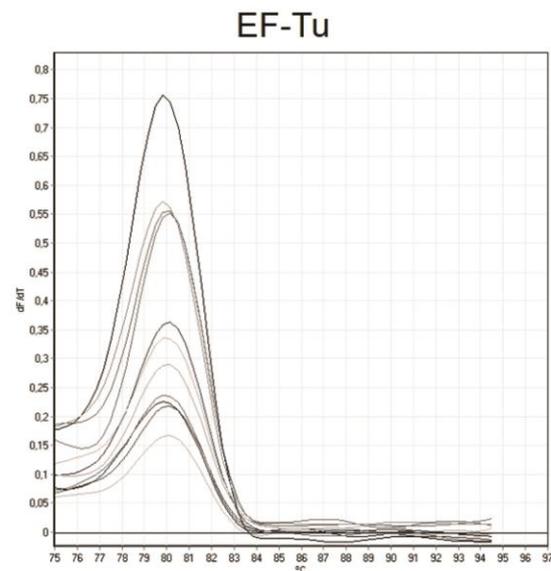
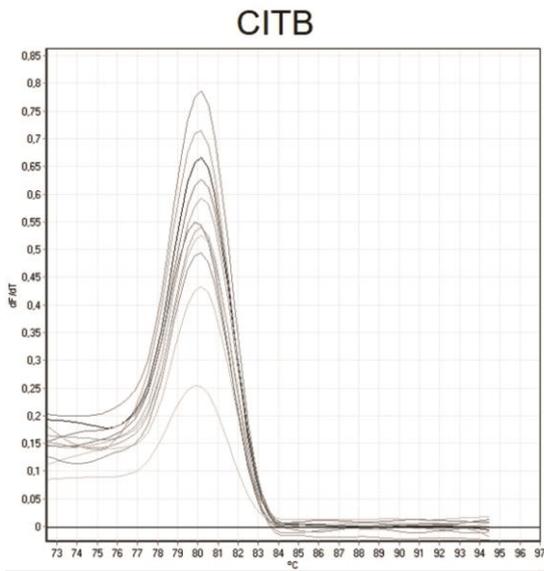
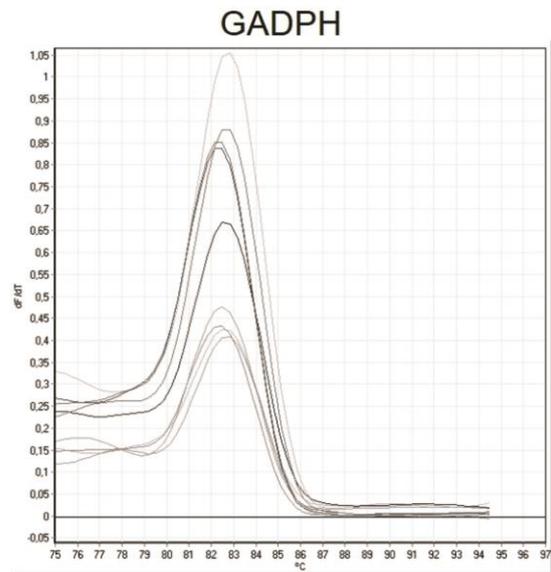
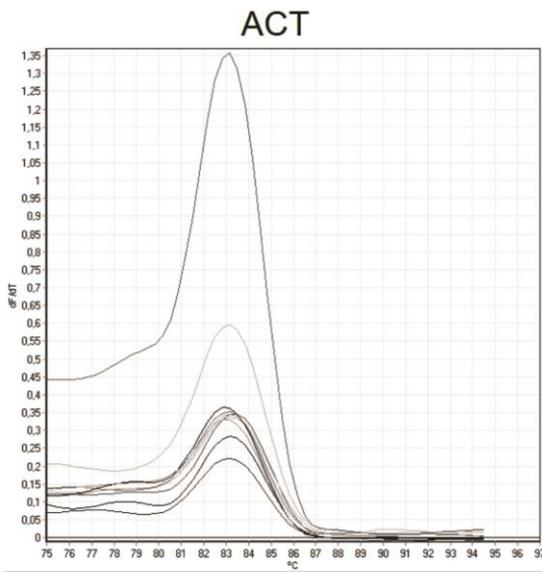
Zhang, H. B., Xia, E. H., Huang, H., Jiang, J. J., Liu, B. Y., Gao, L. Z., 2015. De novo transcriptome assembly of the wild relative of tea tree (*Camellia taliensis*) and comparative analysis with tea transcriptome identified putative genes associated with tea quality and stress response. *BMC Genomics*, 16 (1): 298. Doi:10.1186/s12864-015-1494-4

Zhang, M., Liu, J., Li, K., Yu, D., 2013., Identification and Characterization of a Novel Monoterpene Synthase from Soybean Restricted to Neryl Diphosphate Precursor. *Plos ONE*, 8 (10). Doi:10.1371/journal.pone.0075972

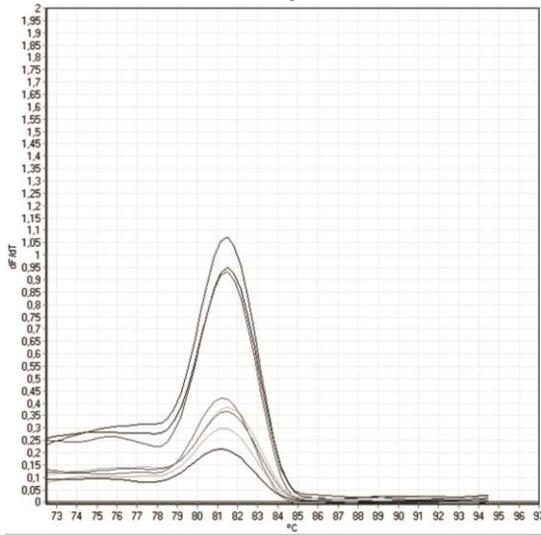
Zhang, Q., Wang, J., Deng, F., Yan, Z., Xia, Y., Wang, Z., Ye, J., Deng, Y., Zhang, Z., Qiao, M., Li, R., Denduluri, S. K., Wei, Q., Zhao, L., Lu, S., Wang, X., Tang, S., Liu, H., Luu, H. H., Haydon, R. C., He, T., Jiang, L., 2015. Tqpcr: A Touchdown qpcr Assay with Significantly Improved Detection Sensitivity and Amplification Efficiency of SYBR Green qpcr. *Plos One*, 10 (7): e0132666. Doi:10.1371/journal.pone.0132666

Zhao, Q.-Y., Wang, Y., Kong, Y.-M., Luo, D., Li, X., Hao, P. 2011. Optimizing de novo transcriptome assembly from short-read RNA-Seq data: a comparative study. *BMC Bioinformatics*. Biomed Central Ltd. Doi:10.1186/1471-2105-12-S14-S2

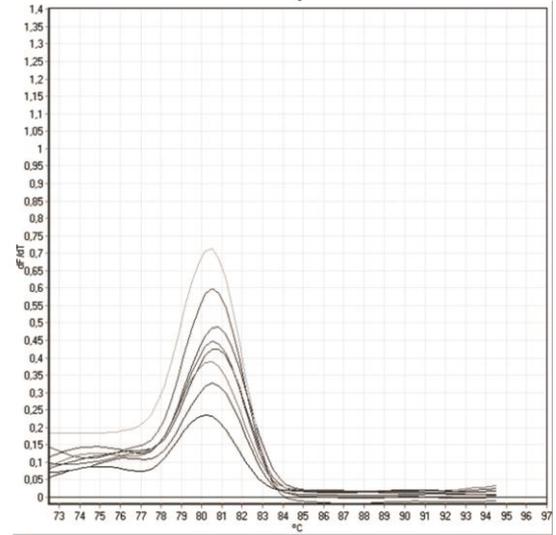
# ANEXO: CURVAS DE DISSOCIAÇÃO DOS *PRIMERS*



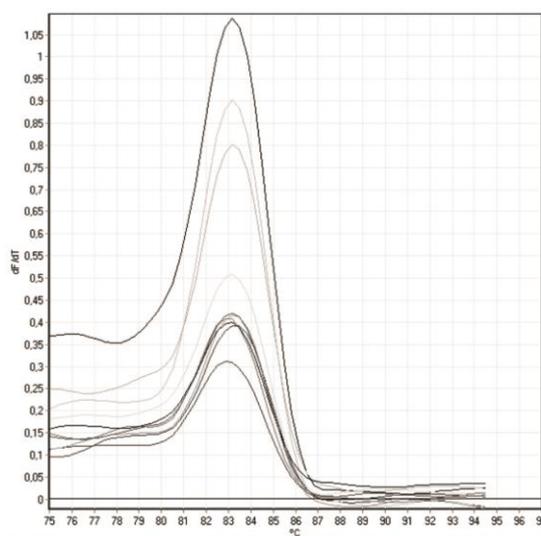
DNAPol



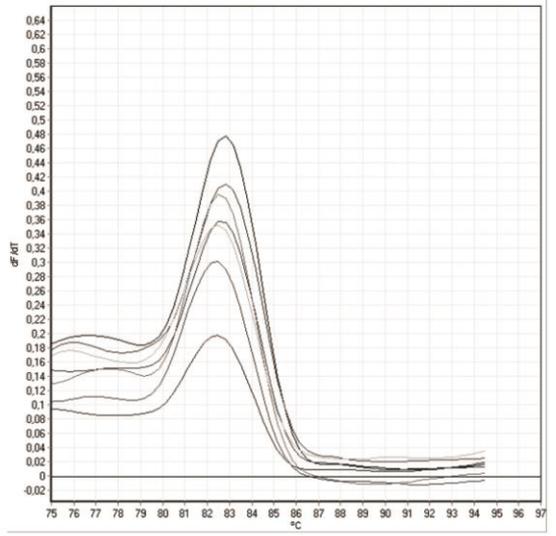
RNApol



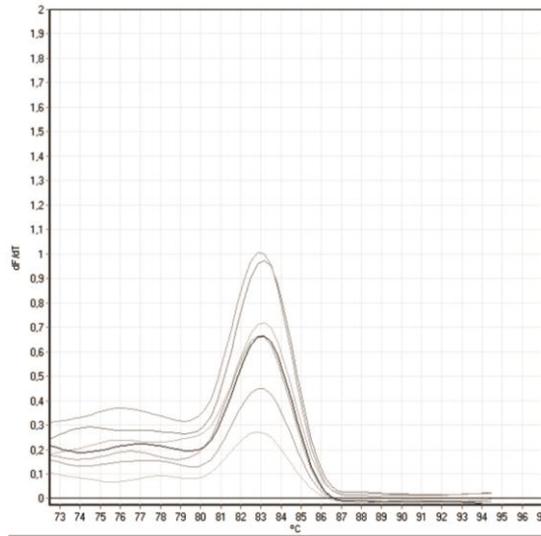
MALADH



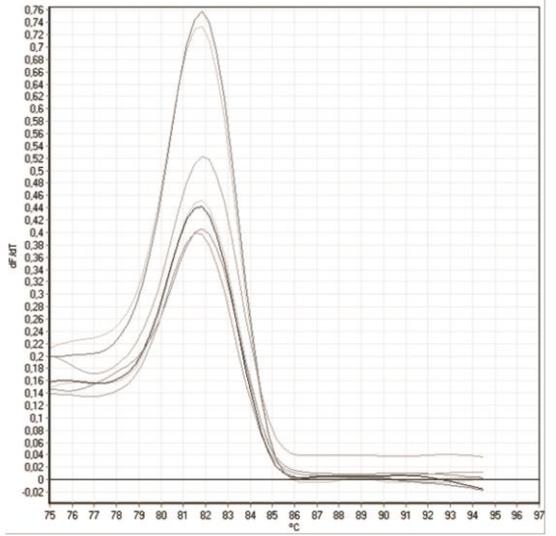
SUCRS



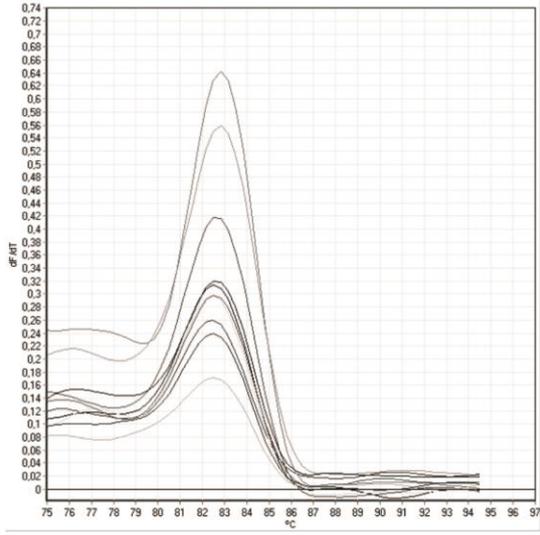
RPTL2



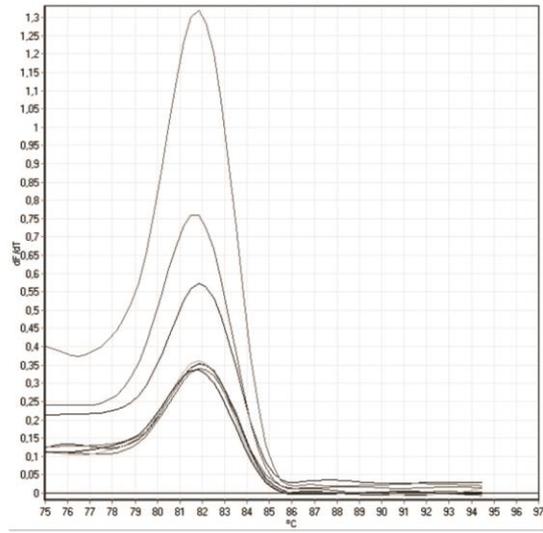
SerTre



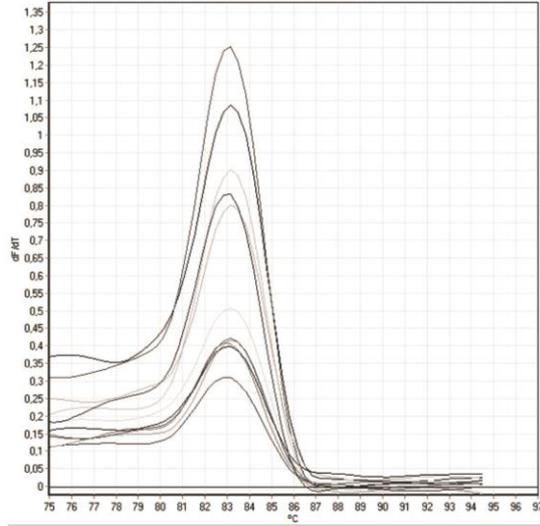
TUBUL



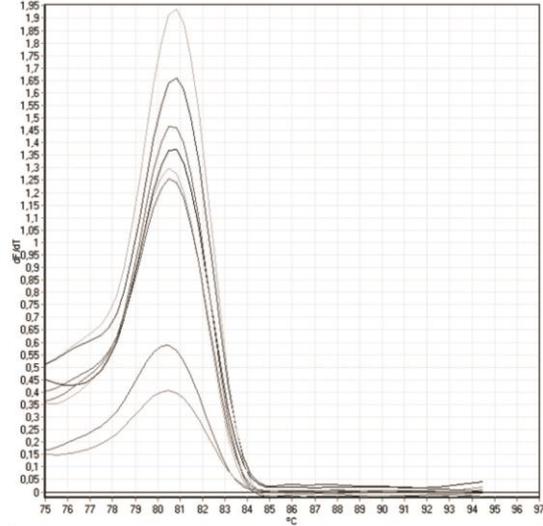
FRNDP



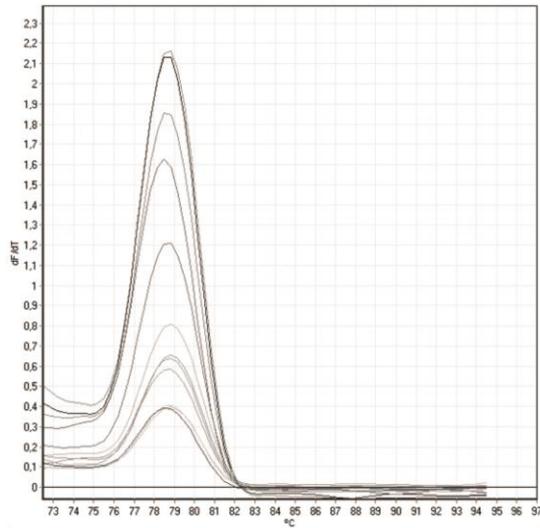
GERGER



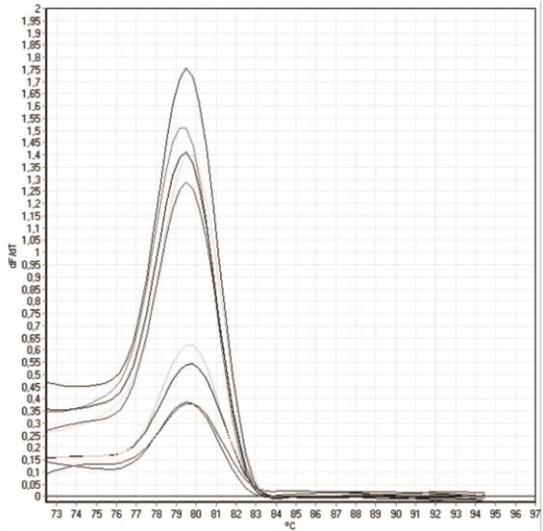
GERS



NERLIN-1



NERLIN-2



# NERLIN-3

