

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
DEPARTAMENTO DE BIOLOGIA

Fernanda Alves de Freitas Guedes

**PIROSSEQUENCIAMENTO DO TRANSCRIPTOMA
DE FOLHA DE *Lippia alba* POR MEIO DA
PLATAFORMA 454 GS FLX (ROCHE)**

Juiz de Fora-MG
2011

2011

PGCBIO

Fernanda Alves de Freitas Guedes

ICB/UFJF

FERNANDA ALVES DE FREITAS GUEDES

**PIROSSEQUENCIAMENTO DO TRANSCRIPTOMA DE FOLHA
DE *Lippia alba* POR MEIO DA PLATAFORMA 454 GS FLX (ROCHE)**

Dissertação de Mestrado do Curso de Pós-Graduação em Ciências Biológicas, para obtenção do Título de Mestre em Ciências Biológicas na área de Genética e Biotecnologia.

Orientador: Prof. Marcelo de Oliveira Santos

Juiz de Fora-MG
2011

Guedes, Fernanda Alves de Freitas

Pirossequenciamento do transcriptoma de folha de *Lippia alba* por meio da plataforma 454 GS FLX (Roche)./ Fernanda Alves de Freitas Guedes. — 2011.

63 f. : il.

Dissertação (Mestrado em Ciências Biológicas – Imunologia e DIP/Genética e Biotecnologia)—Universidade Federal de Juiz de Fora, Juiz de Fora, 2011.

1. Plantas medicinais. 2. Ervas. I. Título.

CDU 615.32:633.88

FERNANDA ALVES DE FREITAS GUEDES

**Pirossequenciamento do transcriptoma de folha
de *Lippia alba* por meio da plataforma 454 GS FLX (Roche)**

Dissertação defendida no Programa de Pós-Graduação em Ciências Biológicas da
Universidade Federal de Juiz de Fora para obtenção do título de Mestre em
Genética e Biotecnologia, aprovada em 28 de Fevereiro de 2011, pela Banca
Examinadora constituída pelos professores:

Prof. Marcelo de Oliveira Santos
Departamento de Biologia – UFJF
Presidente da Banca

Prof. Lyderson Facio Viccini
Universidade Federal de Juiz de Fora

Profa. Marisa Fabiana Nicolás
Laboratório Nacional de Computação Científica

Dedico aos meus pais, pelo amor e confiança.

AGRADECIMENTOS

Agradeço ao meu orientador, professor Dr. Marcelo de Oliveira Santos, pela oportunidade de trabalho tão valiosa; aos meus pais, Sonia e Célio, por sua presença e apoio incondicional; aos meus irmãos Marcela, Vinícius e Paulo, pelas conversas e momentos felizes partilhados; ao Filipe, companheiro paciente mesmo nas horas mais difíceis; à amiga Letícia, por sua disponibilidade em ajudar sempre; aos demais amigos do laboratório capazes de tornar a convivência mais agradável; à Universidade Federal de Juiz de Fora, ao Programa de Pós-Graduação em Imunologia, Genética e Biotecnologia da UFJF; à CAPES e FAPEMIG pelo apoio financeiro.

RESUMO

Lippia alba, também conhecida popularmente como erva cidreira, é uma espécie amplamente distribuída pelas Américas podendo ser encontrada por praticamente todo o Brasil. Muito usada pela medicina popular para o tratamento de problemas gastrointestinais e respiratórios, a folha desta espécie produz um óleo essencial rico em terpenos, principalmente mono e sesquiterpenos. Estes compostos não são apenas de interesse farmacológico, como também industrial. A composição deste óleo pode variar em função de fatores abióticos e também de variações genótípicas. Diante da complexidade da síntese destes compostos a proposta deste trabalho foi uma ampla caracterização do transcriptoma de folha de *Lippia alba*, além da identificação de prováveis enzimas envolvidas na síntese de terpenos. Para isso, foi feito um sequenciamento deste transcriptoma usando a plataforma 454 (Roche) seguido de uma montagem *de novo*. Esta plataforma tem sido cada vez mais utilizada para o sequenciamento de transcriptomas numa abordagem conhecida como RNA-Seq. O sequenciamento de biblioteca preparada a partir de RNA total em 1/8 de placa gerou 104.631 leituras com comprimento médio de 184,48bp num total de 19.302.161 bases. Foram feitas montagens das leituras usando 2 diferentes *assemblers* a fim de compará-las. Com o *Newbler 2.5* foi possível montar 2.686 *contigs* com comprimento médio de 349bp, enquanto o *SeqMan2.2* gerou 13.448 *contigs* com média de 284bp. Em seguida, foi feita a anotação funcional com o Blast2GO para os *contigs* obtidos nas duas montagens, tendo sido anotados 51,49% e 30,88%, respectivamente, dos *contigs* do *Newbler* e do *SeqMan*. Por fim, a análise das sequências anotadas revelou algumas enzimas potencialmente envolvidas com a síntese de terpenos. Os resultados obtidos neste estudo pioneiro sobre a espécie comprovam que as tecnologias NGS podem ser uma ferramenta bastante eficiente para o sequenciamento de transcriptomas e servirão como referência para o preparo mais específico de novas bibliotecas. Futuros sequenciamentos devem contribuir para uma melhor cobertura do transcriptoma permitindo a descoberta inclusive de transcritos raros.

Palavras-chave

Lippia alba, terpenos, transcriptoma, sequenciamento, plataformas NGS, 454, *contig*, anotação funcional.

ABSTRACT

Lippia alba, popularly known as erva cidreira, is widely distributed throughout the Americas and can be found through almost whole Brazil. This species is largely used in folk medicine to treat gastrointestinal and respiratory problems, especially leaves, which produce an essential oil rich in terpenes, mainly mono- and sesquiterpenes. These compounds are not only of pharmacological interest, as well as industrial. Composition of essential oil can vary depending on the developmental stage, the plant part and other abiotic factors. However, genotypic variations also contribute to oil composition variation. Given the complexity of terpenes synthesis, including diversity of enzymes involved in these metabolic pathways, the purpose of this work was a *L. alba* leaf transcriptome characterization, in addition to identifying some enzymes probably involved in terpene synthesis. For that, it was made a transcriptome sequencing using 454 platform (Roche) followed by a de novo assembly. This platform, along with other NGS technologies, has been increasingly used for transcriptome sequencing in an approach known as RNA-Seq. Sequencing of a library prepared from total RNA in 1/8 plate generated 104,631 reads with average length of 184.48 bp and a total of 19,302,161 bases. Read assemblies were made using two different assemblers in order to compare them. While Newbler 2.5, proprietary software platform, assembled 2686 contigs with average length of 349bp, SeqMan2.2 generated 13,448 contigs with an average of 284bp. Then, functional annotation was performed with Blast2GO for all contigs from both assemblies; 51.49% and 30.88% of contigs, respectively, from Newbler and SeqMan were annotated. Finally, analysis of annotated sequences revealed some enzymes potentially involved in terpene synthesis. Results obtained from this pioneering study on the species show that NGS technology can be a very efficient tool for transcriptome sequencing and they will serve as reference for preparation of other more specific libraries. New sequencings should contribute to a better coverage of this transcriptome, allowing discovery of even rare transcripts.

Keywords

Lippia alba, terpenes, transcriptome, sequencing, NGS platforms, 454, contig, functional annotation.

LISTA DE FIGURAS

- Figura1.1** Biossíntese de terpenóides nos diferentes compartimentos celulares (BOHLMANN, MEYER-GAUEN e CROTEAU, 1998). 20
- Figura1.2** A Síntese de IPP: rota do mevalonato (A) e rota do MEP (B) (DUBEY, BHALLA e LUTHRA, 2003). 21
- Figura1.3** Esquema da síntese de terpenos pela via do Ácido Mevalônico (McGARVEY e CROTEAU, 1995). 22
- Figura1.4** Painel comparativo das plataformas next-generation atualmente disponíveis no mercado, com seus respectivos tipos de preparo de biblioteca e química de sequenciamento. 27
- Figura1.5** Continuação da figura 1.4. 28
- Figura1.6** Técnica de pirosequenciamento empregada no 454 (DROEGE e HILL, 2008). 30
- Figura3.1** Softwares usados na montagem das leituras: SeqMan2.2 (A) e Newbler2.5 (B). 39
- Figura3.2** Blast2GO, usado para anotação funcional das sequências. 40
- Figura4.1** Eletroforese de cDNA dupla fita e RNA total do acesso CL16 de *L. alba* em gel de agarose 1%. 42
- Figura4.2** Análise do RNA Total no Agilent 2100 Bioanalyser destacando os picos 18S e 28S; nos eixos do gráfico temos tempo (s) e fluorescência (FU). 42

Figura4.3 Análise da otimização do número de ciclos de PCR; foram aplicados 5 µl em cada lane.	42
Figura4.4 Distribuição de frequência do comprimento das leituras obtidas com o 454 GS FLX; o comprimento médio foi de 184,48bp.	44
Figura4.5 Distribuição de tamanho dos contigs montados com o Newbler2.5.	45
Figura4.6 Distribuição de tamanho dos contigs da montagem final com o SeqMan2.2, usada para anotação funcional.	47
Figura4.7 Percentual de similaridade das sequências encontradas pelo Blast2GO para os contigs das montagens do Newbler2.5 e SeqMan2.2.	51
Figura4.8 Percentual de anotação funcional dos contigs gerados com o Newbler2.5 (2.686) e SeqMan2.2 (13.448).	52
Figura4.9 Representação hierárquica parcial de categorias de processos biológicos com seus respectivos termos GO; as categorias do esquema estão relacionadas ao metabolismo de terpenos.	52
Figura4.10 Histograma comparativo do percentual de sequências em categorias de processos biológicos prováveis com os quais os transcritos estão envolvidos. As sequências anotadas de cada montagem foram agrupadas de acordo com a classificação GO de processo biológico, dentre as quais algumas relacionadas ao metabolismo secundário.	53
Figura4.11 Representação da via de síntese de terpenóides e algumas enzimas anotadas em azul (KEGG map 00900). Usando o Blast2GO, foram obtidos números EC para as enzimas e estes	57

números foram usados para a obtenção deste mapa da rota de biossíntese de terpenóides.

LISTA DE TABELAS

Tabela1.1 Características das leituras geradas por diferentes plataformas.	26
Tabela4.1 Resultados obtidos com sequenciamento e montagem com o Newbler2.5.	43
Tabela4.2 Comparação das montagens feitas com o SeqMan2.2 variando os parâmetros básicos. Os valores definidos na primeira linha da tabela são valores default, usados na montagem inicial; a montagem que usou os valores definidos na última linha da tabela foi aquela escolhida para a anotação funcional.	46
Tabela4.3 Resultados obtidos na montagem final com o SeqMan2.2, utilizada para a anotação funcional.	47
Tabela4.4 Percentual de contigs gerados pelo Newbler2.5 e SeqMan2.2 distribuídos em categorias de tamanho.	49
Tabela4.5 Sequências anotadas do Newbler (N) e do SeqMan (S), com seus respectivos números EC, potencialmente envolvidas com a síntese de terpenos.	55
Tabela4.6 Potenciais terpeno sintases identificadas; Newbler (N) e do SeqMan (S).	56

LISTA DE ABREVIATURAS

CG-MS - Cromatografia gasosa acoplada à espectrometria de massa

CIM - Concentração inibitória mínima

EA - Extrato aquoso

mL - mililitro

µg - micrograma

Acetil-coA - acetil coenzima A

MEP - metileritrol fosfato/ 2-C-metil-D-eritritol-4-fosfato

DOXP - 1-deoxi-D-xilulose-5-fosfato

MECP - 2-C-metil-D-eritritol 2,4 ciclodifosfato

HMBPP - 1-hidroxi-2-metil-2(E)-butenil-4-difosfato

CDP - copalil difosfato

DMAPP - dimetilalil difosfato

EST - “Expressed Sequence Tag”

FPP - farnesil difosfato

GPP - geranil difosfato

GGPP - geranilgeranil difosfato

IPP - isopentenil difosfato

LPP - linalil difosfato

kDa - quilodáltons

NGS - “Next-generation Sequencing”

GS - “Genome Sequencer”

EST - “Expressed Sequence Tag”

sst-DNA - “single stranded DNA” (DNA fita simples)

emPCR - PCR em emulsão

OLC - “Overlapping-Layout-Consensus”

bp - “base pair” (par de bases)

pg - picogramas

GO - “Gene Ontology”

BLAST - “Basic Local Alignment Search Tool”

PTP - “Pico Titter Plate”

EC - “Enzyme Code”

KEGG - “Kyoto Encyclopedia of Genes and Genomes”

SUMÁRIO

1 INTRODUÇÃO	14
1.1 O gênero <i>Lippia sp.</i>	14
1.2 <i>Lippia alba</i>: importância e composição do óleo essencial	15
1.3 O metabolismo secundário e as terpeno sintases	19
1.4 Sequenciadores da nova geração	26
1.5 Pirosequenciamento na análise de transcriptomas	33
2 OBJETIVOS	36
2.1 Objetivo geral	36
2.2 Objetivos específicos	36
3 MATERIAL e MÉTODOS	37
3.1 Material Vegetal	37
3.2 Extração de RNA total e síntese do cDNA	37
3.3 Sequenciamento com o 454 GS FLX (Roche)	38
3.4 Montagem dos ESTs	38
3.5 Anotação funcional com o Blast2GO	40
3.6 Mapas KEEG	40
4 RESULTADOS e DISCUSSÃO	41
4.1 Sequenciamento e montagem dos ESTs	41
4.2 Anotação funcional e Mapas KEGG para vias metabólicas	50
5 CONCLUSÕES	58
6 REFERÊNCIAS BIBLIOGRÁFICAS	59

1 INTRODUÇÃO

1.1 O GÊNERO *Lippia* sp

O gênero *Lippia*, pertencente à família Verbenaceae, ordem Lamiales, inclui cerca de 200 espécies de ervas, arbustos e árvores, distribuídas principalmente pela América do Sul e Central, e territórios da África Tropical (PASCUAL *et al.*, 2001). O Brasil é um dos principais centros de diversidade deste gênero, possuindo aproximadamente 70-75% das espécies conhecidas. Muitas delas estão concentradas e são endêmicas da “Cadeia do Espinhaço”, nos estados de Minas Gerais e Goiás (VICCINI *et al.*, 2006). Membros da ordem Lamiales podem ser importantes fontes de óleos essenciais cujos componentes podem ser utilizados tanto para fins medicinais quanto na indústria de cosméticos e perfumaria (ALAGNA *et al.*, 2009). São exemplos de plantas desta ordem a oliveira (*Olea europaea*, família Oleaceae), hortelã e alecrim (*Mentha* sp e *Rosmarinus officinalis*, família Lamiaceae).

As espécies do gênero *Lippia* são popularmente usadas principalmente para o tratamento de problemas gastrointestinais e respiratórios sendo utilizadas partes aéreas, como folhas ou flores, na maioria dos casos. Entretanto, são atribuídas a estas plantas outras propriedades como: analgésica, sedativa, anti-inflamatória e antipirética. Foram relatados ainda o uso de algumas espécies para tratar lesões cutâneas (PASCUAL *et al.*, 2001).

Apesar do amplo e diverso uso tradicional de diferentes espécies mencionado acima, ainda são poucos os estudos sobre composição química e aspectos farmacológicos. Desta forma, podemos encontrar diferentes referências no que diz respeito a este uso popular. *Lippia chevalieri*, por exemplo, é tomada como chá estimulante no Senegal enquanto na Nigéria a mesma infusão é bebida como sedativa e relaxante (PASCUAL *et al.*, 2001). Dentre as espécies mais estudadas do gênero, podemos citar *L. multiflora* e *L. alba*.

Viccini *et al.* (2006) e Campos *et al.* (2010) determinaram o número cromossômico de diversas espécies de *Lippia*, número este que variou de n=10 a

n=28 nas espécies estudadas.

1.2 *Lippia alba*: IMPORTÂNCIA E COMPOSIÇÃO DO ÓLEO ESSENCIAL

A espécie *Lippia alba* é popularmente conhecida como erva cidreira, melissa, salvia morada, chá de tabuleiro, cidrila, alecrim selvagem, cidreira brava, falsa melissa, carmelitana, salva do Brasil, salva, salva limão, alecrim do campo e salva brava (SIANI *et al.*, 2002; AGUIAR e COSTA, 2005). Trata-se de uma planta arbustiva que pode ser encontrada em praticamente todas as regiões do Brasil, sendo muito empregada pela medicina popular para o tratamento de distúrbios estomacais como cólicas, indigestão, dores, náuseas e espasmos; nos casos de tosse, gripe, resfriado, bronquite; como tranqüilizante ou calmante, sedativo, analgésico, bem como nos casos de distúrbios hepáticos, sífilis, diarreia, disenteria, como carminativo, no tratamento de dores de cabeça, no combate à hipertensão e malária (JULIÃO *et al.*, 2003; HEINZMANN e BARROS, 2007). Tais propriedades aliadas à sua baixa toxicidade (JULIÃO *et al.*, 2003) contribuem para o uso popular de *L. alba*, cujas folhas são a parte da planta mais utilizada, principalmente na forma de chá (AGUIAR e COSTA, 2005). Entretanto, raízes também podem ser utilizadas (HEINZMANN e BARROS, 2007).

Em estudo realizado por Aguiar *et al.* (2008) foi testada a atividade antimicrobiana *in vitro* de diferentes extratos brutos de *L. alba*. Os resultados obtidos mostraram que os extratos hexânico, etanólico e metanólico das folhas inibiram o crescimento de *Staphylococcus*, *Micrococcus luteus*, *Bacillus subtilis*, *Mycobacterium smegmatis* e *Monilia sitophila*. Extratos da raiz foram ativos frente a esses microrganismos e também a *Candida albicans*. Foi também demonstrada atividade antiulcerogênica de *L.alba* (PASCUAL *et al.*, 2001), mostrando que a infusão de suas folhas foram capazes de prevenir a ulceração induzida, sem modificar o pH gástrico nem a acidez total em ratos machos Wistar.

O óleo essencial de três quimiotipos de *L.alba* foi usado em testes de efeito comportamental. Os principais componentes dos óleos eram: citral, b-mirceno e

limoneno no quimiotipo I, citral e limoneno no quimiotipo II, carvona e limoneno para o quimiotipo III. Sua aplicação intraperitoneal em ratos mostrou efeitos ansiolítico, hipotérmico, sedativo e miorelaxante (VALE *et al.*, 1999).

O linalol, que pode ser componente do óleo essencial de *L.alba*, mostrou efeito depressor sobre o sistema nervoso central, resultando em propriedades hipnóticas, hipotérmicas e anticonvulsivantes (JULIÃO *et al.*, 2003). Efeitos sedativo e miorelaxante do percolato etanólico de folhas de *L.alba* a 80% (v/v) administrado via oral, cujo conteúdo de flavonoides foi elevado, também foram demonstrados em ratos (ZÉTOLA *et al.*, 2000).

A potencialidade de plantas nativas para o desenvolvimento de fármacos e matérias-primas farmacêuticas, fundamenta-se no tripé biodiversidade, aceitabilidade e mercado econômico (HEINZMANN e BARROS, 2007). Em função de seu uso popular, sustentado pelos crescentes estudos farmacológicos, *Lippia alba* revela seu potencial para o desenvolvimento de novos fármacos.

Além do amplo uso medicinal, compostos do óleo essencial de *L. alba* podem ser de interesse industrial. O linalol, por exemplo, é um importante composto para a indústria de aromas, sendo usado como substituto para o óleo da lavanda francesa por causa da similaridade de odor de sua forma levorrotatória. Este composto ocorre naturalmente como os dois isômeros ópticos, que possuem odores distintos. Também podem ser fontes de linalol espécies brasileiras do gênero *Aniba* ou a espécie chinesa *Cinnamomum camphora* (Lauraceae). Entretanto, a primeira corre risco de extinção e seu óleo possui ambos os isômeros, e a segunda produz um óleo essencial de qualidade inferior. Assim, na busca por fontes alternativas de linalol para suprir a demanda do mercado a espécie *L. alba* aparece como uma fonte potencial (SIANI *et al.*, 2002).

Um documento, intitulado “*Market Brief in the European Union for selected natural ingredients derived from native species*”, divulgado em 2005 após Conferência de Comércio e Desenvolvimento das Nações Unidas, apresentou um resumo do mercado da União Europeia para produtos naturais vegetais, dando ênfase à espécie *Lippia alba*. O tipo e a qualidade do óleo essencial determinam que tipo de produto pode ser gerado a partir dele, ou seja, se ele é destinado à indústria de cosméticos, de alimentos ou farmacêutica. Segundo dados apresentados, em 2003 o total de importações em óleos essenciais chegou a 490 milhões de euros, com maior interesse em óleos essenciais não derivados de frutos cítricos, dentre os

quais o de *Lippia alba*. Países como a França, conhecida por seus perfumes, e o Reino Unido lideram as importações de óleos essenciais na União Europeia, respondendo por quase metade do valor total. O documento ressalta ainda que a exportação de óleo essencial de *L. alba* para países da União Europeia deve estar de acordo com requisitos governamentais e da indústria.

O óleo essencial de *L. alba* apresenta, entre outras substâncias, uma grande concentração de compostos terpênicos, principalmente monoterpenos e sesquiterpenos (PASCUAL *et al.*, 2001). Dentre os compostos comumente encontrados no óleo essencial de *L. alba* podemos citar os monoterpenóides: neral, geranial, carvona, 1,8-cineol, linalol, borneol, citronelol, limoneno, mirceno, citral, cânfora; e os sesquiterpenóides: b-cariofileno, a-muuroleno, b-cubebeno, b-elemenol, cadineno, alo-aromadendreno e óxido de cariofileno (PASCUAL *et al.*, 2001; SIANI *et al.*, 2002; JULIÃO *et al.*, 2003; AGUIAR e COSTA, 2005).

A composição do óleo essencial, entretanto, pode ser afetada por diversos fatores. Dentre eles podemos citar o estágio de desenvolvimento, a parte da planta usada (a composição varia entre folhas e demais partes aéreas, na presença ou ausência de flores, em material fresco ou seco), a localização geográfica de coleta, características do solo, do clima e outros fatores ecológicos. A influência destes fatores no acúmulo de terpenos nas espécies de *Lippia* determinam o perfil químico de seu óleo e seus componentes majoritários que, por sua vez, determinam o quimiotipo (SIANI *et al.*, 2002).

Entretanto, não apenas variações químicas, mas também variações morfológicas são encontradas nesta espécie (MATOS *et al.*, 1996). Quimiotipos podem apresentar diversidade morfológica, especialmente no tamanho e textura das folhas, coloração e tamanho das flores (MATOS, 1996).

Vale (2010) estudou variações morfológicas encontradas em 32 acessos de *L. alba* coletados em diferentes regiões do Brasil e mantidos em casa de vegetação, sob as mesmas condições durante 14 meses. Foram encontradas variações no hábito, filotaxia, folha, inflorescência, bráctea, cálice e corola. Apesar dos órgãos vegetativos serem bastante plásticos, ao contrário dos reprodutivos, neste trabalho foram encontradas variações em ambos e estas não se devem a alterações ambientais, uma vez que os espécimes foram mantidos sob as mesmas condições; ao contrário, elas devem refletir principalmente diferenças genotípicas. Além dos caracteres morfológicos, este trabalho avaliou a quantidade de DNA por citometria

de fluxo e no número cromossômico para alguns dos acessos. As estimativas de quantidade de DNA variaram de 2,31-2,44pg a 6,70pg, tendo sido este último valor obtido apenas para 1 acesso que apresentou também características morfológicas e número cromossômico discrepantes dos demais. Para o número cromossômico, foram encontrados $2n=30$, $2n=34$, $2n=51$, $2n=58$, $2n=60$ e $2n=93$. Considerando *L. alba* uma espécie diploide ($2n=30$), os valores encontrados podem representar eventos de poliploidização. As variações genéticas observadas nesses acessos, juntamente com as variações morfológicas, sugerem que os eventos de poliploidização sejam relevantes na evolução da espécie.

Atti-Serafini *et al.* (2002) investigaram a variação da composição dos óleos essenciais de *L. alba* nos diferentes meses do ano. Através de análise por CG-MS, linalol (50,0% - 79,2%) e 1,8-cineol (7,3% - 14,1%) foram identificados como componentes majoritários sendo o melhor período de colheita entre os meses de dezembro e março, quando verificou-se o maior teor de linalol e maior rendimento de óleo essencial.

O óleo essencial extraído de folhas dos quimiotipos citral, carvona e linalol de *L. alba*, provenientes de diferentes regiões do Brasil, mas cultivados em condições semelhantes, foi analisado por cromatografia gasosa acoplada a espectrometria de massa. Os resultados mostraram que a composição do óleo se manteve inalterada para os três quimiotipos mesmo após o cultivo em condições semelhantes, reforçando a ideia de que as diferenças na composição do óleo essencial dos quimiotipos refletem principalmente variações genotípicas entre as plantas. Este trabalho mostrou ainda que o rendimento de óleo essencial é maior no período vegetativo do que no de floração (TAVARES *et al.*, 2005).

Cabe ressaltar que variações quantitativas e qualitativas do óleo essencial de *L. alba* podem resultar em atividades farmacológicas distintas (TAVARES *et al.*, 2005). Diante do amplo uso medicinal de *Lippia alba* e das muitas variáveis que podem afetar composição e rendimento do óleo essencial produzido nas folhas, cuja variação pode resultar em diferentes propriedades medicinais, fica evidente a necessidade de trabalhos que aumentem nossa compreensão sobre a síntese destes compostos de interesse.

1.3 O METABOLISMO SECUNDÁRIO E AS TERPENOSINTASES

O metabolismo secundário dos vegetais inclui os processos que resultam na produção de compostos que, em sua maioria, não tem função direta no crescimento ou desenvolvimento da planta, os chamados metabólitos secundários. Entretanto, alguns compostos como as giberelinas e o ácido abscísico, importantes hormônios vegetais, ou ainda os carotenóides, são considerados terpenos (TAIZ e ZEIGER, 2004).

Existem diferentes tipos de metabólitos secundários, dentre os quais podemos destacar os terpenos, ou terpenóides, que constituem uma família de produtos naturais com mais de 30.000 membros. Os terpenóides têm diferentes papéis, não só para as plantas, como também na saúde humana e comércio. Aqueles que constituem óleos essenciais, como mono e sesquiterpenos, são importantes agentes de sabor e fragrâncias, em alimentos, bebidas, cosméticos, perfumes, sabonetes, além de possuírem atividades biológicas, farmacêuticas e terapêuticas (DUBEY, BHALLA e LUTHRA, 2003).

Todos os terpenos são derivados da união de unidades de cinco carbonos chamadas isopreno ou unidade isoprênica (TAIZ e ZEIGER, 2004). Considerando o número de carbonos que possuem, os terpenos podem ser classificados em hemiterpenos (5 carbonos), monoterpenos (10 carbonos), sesquiterpenos (15 carbonos), diterpenos (20 carbonos), triterpenos (30 carbonos), tetraterpenos (40 carbonos) ou ainda, politerpenos (DUBEY, BHALLA e LUTHRA, 2003).

Os diferentes tipos de terpenos são sintetizados em diferentes compartimentos celulares. A síntese dos sesquiterpenos bem como dos triterpenos, ocorre no citosol; já a síntese de monoterpenos, diterpenos e tetraterpenos ocorre nos plastídios (Figura 1.1). Assim, não só a separação física das enzimas envolvidas como também a disponibilidade de precursores nos diferentes compartimentos celulares controlam a síntese dos terpenóides (McGARVEY e CROTEAU, 1995).

O IPP (isopentenil difosfato) é a unidade ativa básica na formação de terpenos. Em plantas, esta unidade pode ser produzida a partir de duas vias (Figura 1.2): a via do acetato-mevalonato (Ac-MVA) e via metileritrol fosfato (MEP) (DUBEY, BHALLA e LUTHRA, 2003; TAIZ e ZEIGER, 2004). A via do mevalonato ocorre no

citoplasma e mitocôndria para sintetizar principalmente esteróis; já nos plastídeos, os terpenos são formados exclusivamente pela via do MEP. As duas vias operam simultaneamente e a separação nos compartimentos não é absoluta, podendo ocorrer trocas de metabólitos (DUBEY, BHALLA e LUTHRA, 2003).

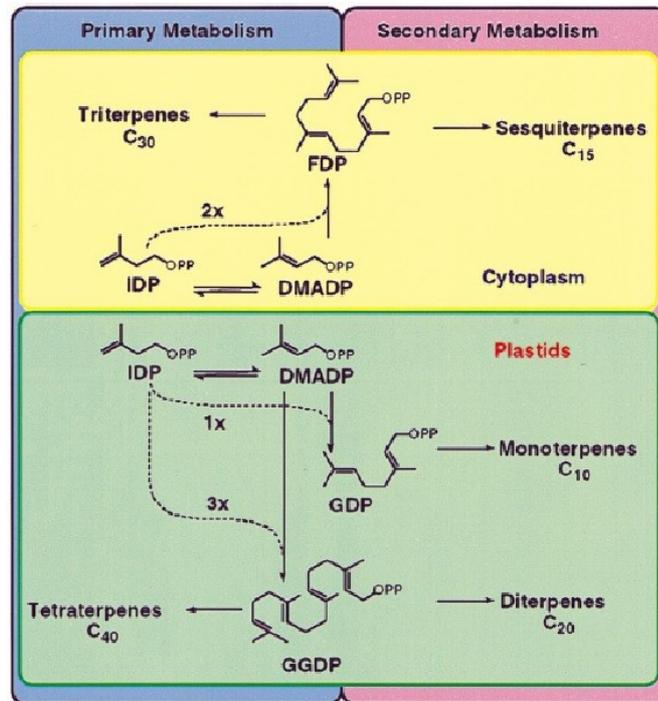


Figura 1.1: Biossíntese de terpenóides nos diferentes compartimentos celulares (BOHLMANN, MEYER-GAUEN e CROTEAU, 1998).

Inicialmente, pela via do ácido mevalônico (Figura 1.3), ocorre a síntese do IPP a partir de três moléculas de acetil-coA. Entretanto, o IPP precisa sofrer isomerização formando o DMAPP (dimetilalil difosfato) para que possa, então, ocorrer a condensação em terpenóides maiores. Esta isomerização é realizada pela **IPP isomerase** (EC:5.3.3.2), uma enzima que requer um íon metálico divalente que pode ser Mg^{2+} , Mn^{2+} ou Ca^{2+} . O DMAPP sofre, então, uma reação catalizada pela **isopreno sintase** (EC:4.2.3.27), uma enzima que requer Mg^{2+} ou Mn^{2+} , para formar o isopreno, o mais simples dos terpenóides (McGARVEY e CROTEAU, 1995).

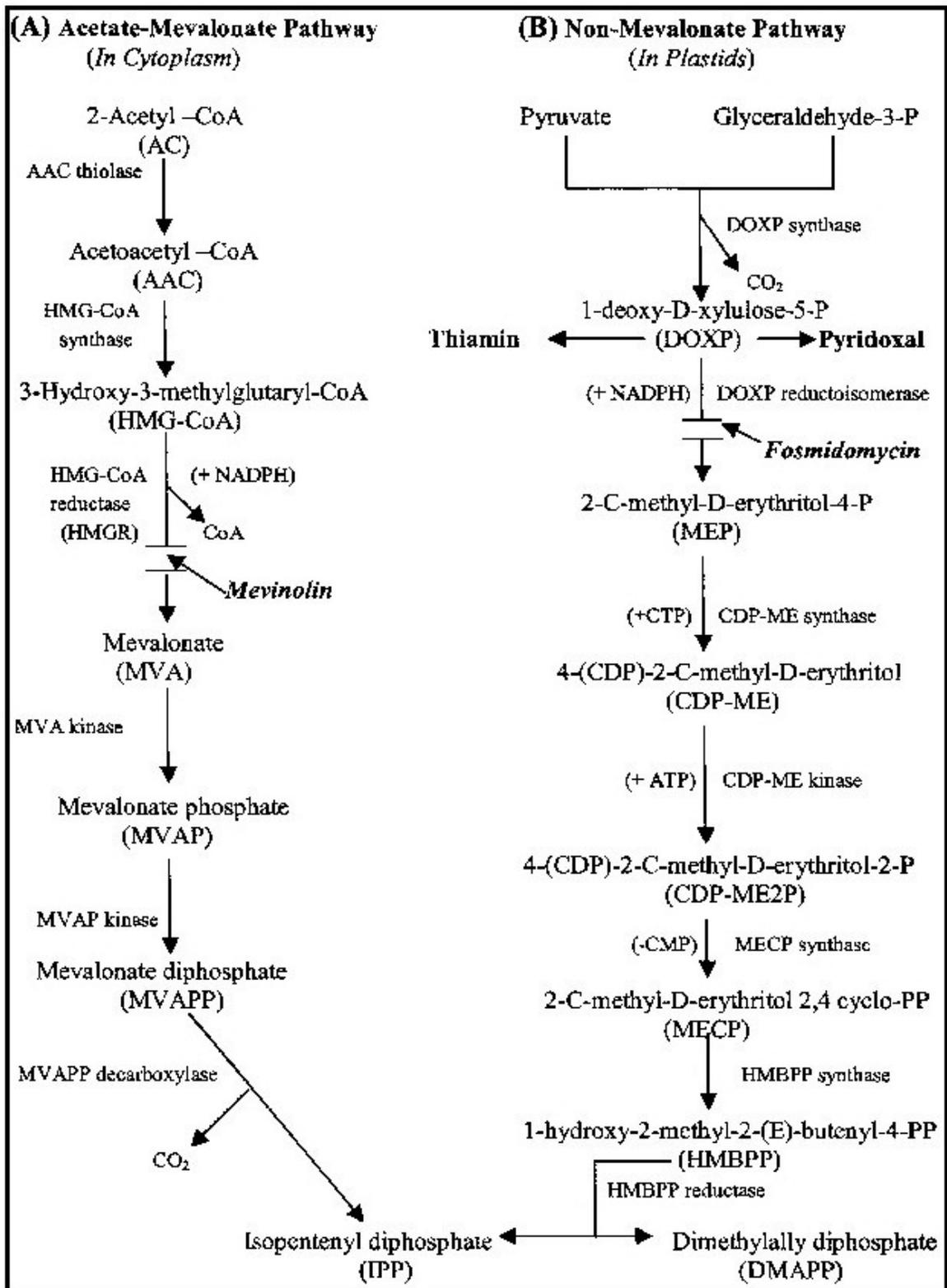


Figura 1.2: A Síntese de IPP: rota do mevalonato (A) e rota do MEP (B) (DUBEY, BHALLA e LUTHRA, 2003).

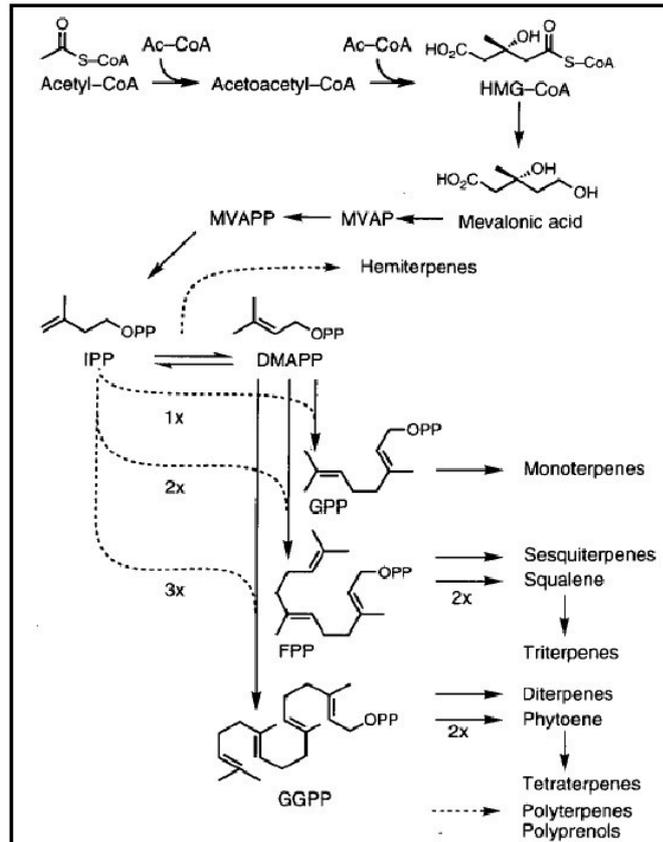


Figura 1.3: Esquema da síntese de terpenos pela via do Ácido Mevalônico (McGARVEY e CROTEAU, 1995).

A rota do MEP se inicia com a síntese do 1-deoxi-D-xilulose-5-fosfato (DOXP) por ação da enzima DOXP sintase (EC:2.2.1.7). Em seguida, outra enzima, a DOXP redutoisomerase (EC:1.1.1.267) transforma o DOXP em MEP (metileritrol fosfato), através de rearranjo e reação de redução. O próximo passo desta via envolve a ação da enzima 4-(difosfocitidil)-2-C-metil-D-eritritol (CDP-ME) sintase, também chamada 2-C-metil-D-eritritol 4-fosfato citidililtransferase (EC:2.7.7.60), que realiza a conversão do MEP em CDP-ME. Este é, em seguida, fosforilado por uma quinase e transformado em um intermediário cíclico por outra enzima, a 2-C-metil-D-eritritol 2,4 ciclodifosfato (MECP) sintase (EC:4.6.1.12). O passo final é a conversão do MECP em IPP e DMAPP. Estudos em *E. coli* e também com *A. thaliana* mostram que esta conversão ocorre através de um intermediário, o hidróxi-2-metil-2(E)-butenil-4-difosfato (HMBPP), envolvendo reações catalizadas pelas enzimas HMBPP sintase (EC:1.17.4.3) e redutase (EC:1.17.1.2) (DUBEY, BHALLA e LUTHRA, 2003; SEEMANN *et al.*, 2005).

Os terpenos maiores são formados por enzimas genericamente denominadas **preniltransferases**. As chamadas **GPP sintases** formam o GPP (geranyl difosfato), um precursor de 10 carbonos que origina os monoterpenos; as **FPP sintases**

formam o FPP (farnesil difosfato) com 15 carbonos e precursor de sesquiterpenos; a **GGPP sintase** forma o GGPP (geranilgeranil difosfato) de 20 carbonos, que dá origem aos diterpenos (McGARVEY e CROTEAU, 1995).

As preniltransferases são homodímeros de 70 a 80 kDa e que necessitam de Mg^{2+} ou Mn^{2+} para sua atividade, sendo necessários dois íons por sítio ativo. Estas enzimas são conservadas em plantas, animais, fungos e bactérias. Dois domínios encontrados nas preniltransferases com a sequência DDX(X)D altamente conservada parecem estar envolvidos na ligação do complexo íon-pirofosfato liberado do substrato (McGARVEY e CROTEAU, 1995).

Outro grupo de enzimas envolvidas na síntese dos terpenóides, que são compostos cíclicos em sua maioria, são as chamadas **sintases** ou **ciclases**. Estas enzimas catalisam reações de ciclização, que podem ser acompanhadas de reações redox, resultando na formação dos diversos terpenos que compõem óleos essenciais e resinas vegetais a partir de seus respectivos precursores. Cabe ressaltar que são conhecidos exemplos de sintases que produzem compostos acíclicos, como as monoterpreno-sintases que produzem mirceno e linalol (BOHLMANN, MEYERGAUEN e CROTEAU, 1998).

As reações catalisadas pelas monoterpreno sintases se iniciam com a isomerização do precursor GPP (geranil difosfato), passo preliminar necessário por causa da dupla ligação existente entre C_2 e C_3 que impede ciclização. GPP sofre então ionização, processo que envolve a participação de íons metálicos divalentes, para formar o LPP (linalil difosfato) a partir do qual forma-se o cátion α -terpinil, altamente reativo e intermediário universal das subseqüentes reações (BOHLMANN, MEYER-GAUEN e CROTEAU, 1998). Uma característica peculiar das monoterpensosintases é sua capacidade de produzir diferentes compostos. A limoneno sintase, por exemplo, é capaz de produzir limoneno, mirceno e α e β -pineno (McGARVEY e CROTEAU, 1995).

Monoterpreno sintases clonadas e caracterizadas de diferentes organismos mostram características semelhantes como: peso molecular de 50 a 100 kDa, podem ser monômeros ou homodímeros, exigência de íon metálico divalente (Mg^{2+} ou Mn^{2+}), pI próximo de 5 e pH ótimo próximo do neutro. Entretanto, as terpeno-sintases de gimnospermas exigem um cátion monovalente (K^+) e tem pH ótimo mais elevado (BOHLMANN, MEYER-GAUEN e CROTEAU, 1998).

Comparado aos monoterpensos, o número de sesquiterpenos existentes é

bem maior, uma vez que seu precursor conta com 5 carbonos a mais possibilitando mais ciclizações. A síntese dos sesquiterpenos se dá a partir do precursor de 15 carbonos FPP, que também precisa sofrer isomerização preliminar, originando o nerolidil difosfato, seguida de ciclização. Segue-se, então, uma série de reações que podem envolver as duplas ligações, os grupos metil e rearranjos, dando origem aos diferentes sesquiterpenos, dentre os quais também podemos encontrar compostos acíclicos como o b-farneseno. Assim como os mecanismos usados, as propriedades das sintases responsáveis pela produção de monoterpenos e sesquiterpenos também são similares (BOHLMANN, MEYER-GAUEN e CROTEAU, 1998).

As ciclizações de diterpenos ocorrem de forma similar, embora a isomerização preliminar não seja necessária em precursores maiores e mais flexíveis. O uso do mesmo mecanismo pelas prenilttransferases e terpeno sintases sugere uma origem evolutiva comum (McGARVEY e CROTEAU, 1995).

A diversidade funcional observada entre as terpeno sintases está relacionada ao número de cópias do gene e à tolerância a mutações. Aquelas enzimas envolvidas no metabolismo secundário são codificadas por várias cópias de genes, originadas por duplicação e, além disso, como não são essenciais ao desenvolvimento, toleram mais mutações levando a maior diversidade de produtos gerados. Já as terpeno sintases envolvidas no metabolismo primário, como as CDPsintases, geralmente são codificadas por um único gene e por sua função essencial, não toleram mutações que resultem em produtos diferentes (BOHLMANN, MEYER-GAUEN e CROTEAU, 1998).

A comparação de sequências e reconstrução filogenética de terpeno sintases de gimnospermas revelam que estas são muito mais relacionadas entre si do que com suas correspondentes em angiospermas (BOHLMANN, MEYER-GAUEN e CROTEAU, 1998).

As terpeno sintases apresentam elementos estruturais característicos. As enzimas pertencentes às famílias Tpsa, Tpsb e Tpsc apresentam um domínio C-terminal com o sítio ativo e outro N-terminal. No caso das monoterpeno sintases, podemos encontrar na porção N-terminal o peptídeo de endereçamento ao plastídio, fazendo com estas enzimas sejam um pouco maiores que as sesquiterpeno sintases. Este peptídeo parece não ser fundamental na atividade enzimática, uma vez que sua deleção não prejudica tal atividade. O conservado motivo RR parece delimitar o peptídeo de endereçamento, sendo que downstream as sequências são

similares (BOHLMANN, MEYER-GAUEN e CROTEAU, 1998).

As diterpeno sintases são ainda um pouco maiores que as monoterpeno sintases por causa de um elemento interno conservado em posição e sequência entre todas as enzimas desta classe. Este elemento não parece estar envolvido na catálise ((BOHLMANN, MEYER-GAUEN e CROTEAU, 1998).

Outro motivo altamente conservado é o DDxxD. Rico nos resíduos de aspartato, ele coordena os íons Mg^{2+} , que interagem com grupo difosfato do precursor no sítio ativo. Este motivo pode ser encontrado não só em terpeno sintases como também em preniltransferases, não só em plantas como também em microrganismos, sempre envolvido com a coordenação do íon metálico divalente. Uma exceção é a (-) CDP sintase, ela não possui o motivo DDxxD, mas possui outro muito similar, o DxxDD; esta enzima realiza inicialmente a protonação do substrato ao invés da clivagem do grupo difosfato (BOHLMANN, MEYER-GAUEN e CROTEAU, 1998).

Além da grande diversidade de compostos terpênicos existentes, as funções exercidas por eles também variam. Dentre estas funções podemos citar a atração de polinizadores, defesa contra patógenos (KESSLER e BALDWIN, 2002) e comunicação, podendo variar de acordo com o tecido vegetal, o estágio de desenvolvimento ou ainda com a condição fisiológica.

Dudareva *et al.* (2003) clonaram e caracterizaram três monoterpeno sintases de flor de *Antirrhinum majus* e foi também analisada a produção dos monoterpenos mirceno e (E)- β -ocimeno nessas flores. Enquanto nas partes vegetativas os monoterpenos tem função principalmente protetora, atuando nos mecanismos de defesa direta ou indireta, nas flores eles atuam principalmente na atração de polinizadores. Entretanto, pouco se sabe sobre genes e enzimas envolvidos na síntese dos componentes de fragrâncias florais.

Em angiospermas que produzem elevados níveis de terpenos, a biossíntese costuma ocorrer em estruturas glandulares especializadas, como os tricomas (McGARVEY e CROTEAU, 1995).

A complexidade inerente à biossíntese dos metabólitos secundários juntamente com a importância econômica e/ou medicinal dos mesmos, justificam um amplo estudo sobre o assunto.

1.4 SEQUENCIADORES DA NOVA GERAÇÃO

Desde o surgimento do método Sanger na década de 70 (MOROZOVA e MARRA, 2008) e sua aplicação no sequenciamento automatizado de DNA, avanços como a introdução da eletroforese capilar vem sendo feitos para tornar o sequenciamento mais rápido, com menores custos e maior acurácia. Recentemente, surgiram os chamados sequenciadores next-generation, second generation, de ultra-high throughput ou ainda, de sequenciamento massivo paralelo (MARDIS, 2008; MOROZOVA e MARRA, 2008). As características destas plataformas tem permitido mais do que a simples determinação da ordem de bases. Os microarrays, por exemplo, vem sendo substituídos pelos métodos seq-based que podem identificar e quantificar transcritos, inclusive raros, sem conhecimento do prévio do gene ou ainda detectar variantes de splicing (METZKER, 2010).

Dentre as plataformas de sequenciamento next-generation (Figuras 1.4 e 1.5) atualmente disponíveis no mercado podemos citar: Illumina/Solexa, ABI/SOLiD (Life/APG), Helicos BioSciences, Polonator G.007, 454 FLX/Roche. Além destas, chegou ao mercado em 2010 o PacBio RS, da Pacific Biosciences (MOROZOVA e MARRA, 2008; KUMAR e BLAXTER, 2010; METZKER, 2010). Apesar da redução relativa dos custos, a maior parte das leituras geradas por estas plataformas são menores que aquelas geradas pelo método Sanger (Tabela 1.1) (MOROZOVA e MARRA, 2008; METZKER, 2010).

Tabela 1.1: Características das leituras geradas por diferentes plataformas.

Plataforma	Comprimento da leitura (bp)	Bp/corrída
Sanger automatizado ABI370	Até 900	96Kb
454 FLX/Roche	330*	450Mb
Illumina/Solexa	75 ou 100	18Gb
ABI/SOLiD	50	30Gb
Polonator G.007	26	(-)
Helicos Bioscience	32*	37Gb
PacBioRS	964*	(-)

(*)Comprimento médio de leitura; os valores bp/corrída se referem ao sequenciamento de bibliotecas de fragmentos que não são mate-pair; (-), valor indisponível.

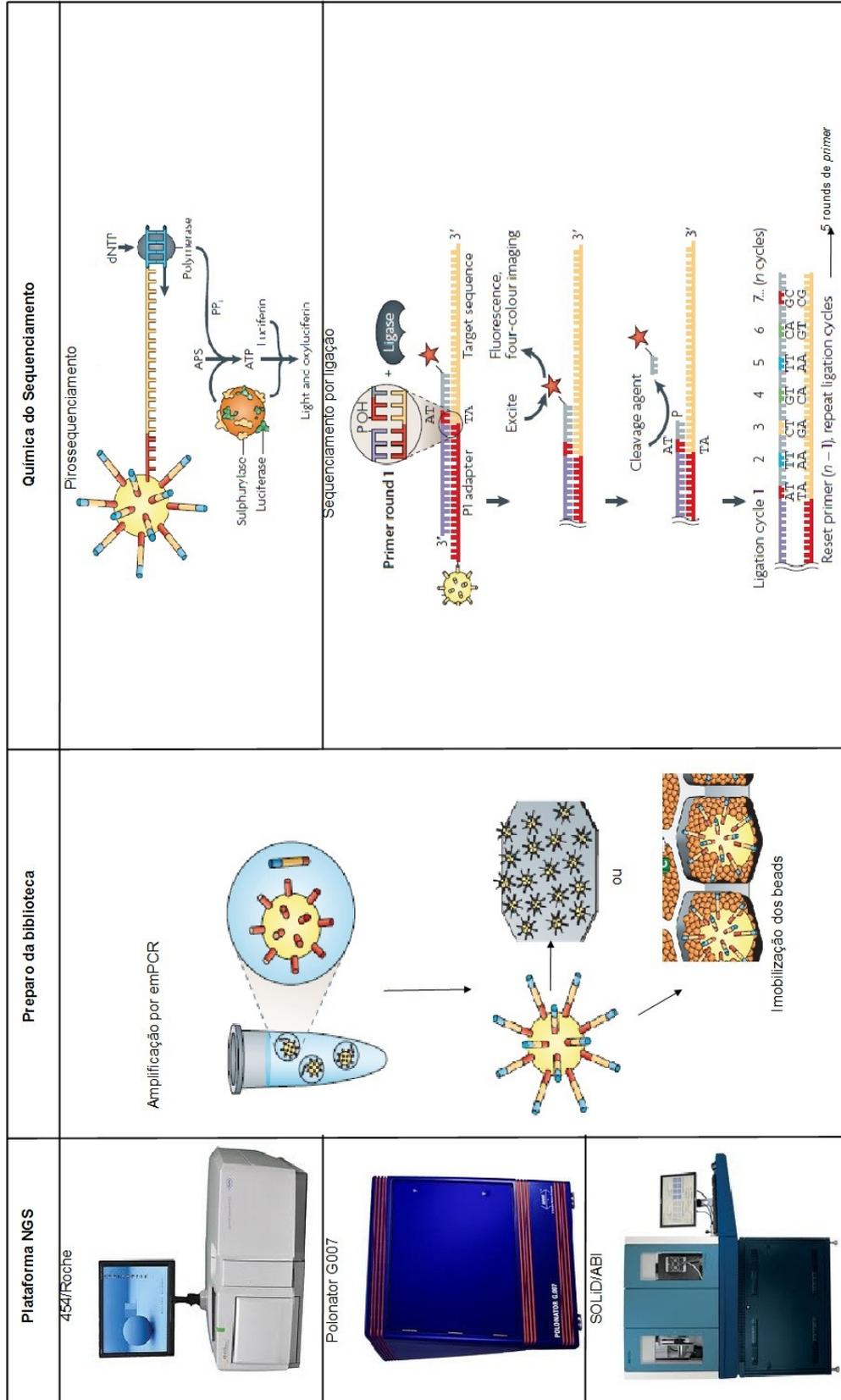


Figura 1.4: Painel comparativo das plataformas next-generation atualmente disponíveis no mercado, com seus respectivos tipos de preparo de biblioteca e química de sequenciamento.

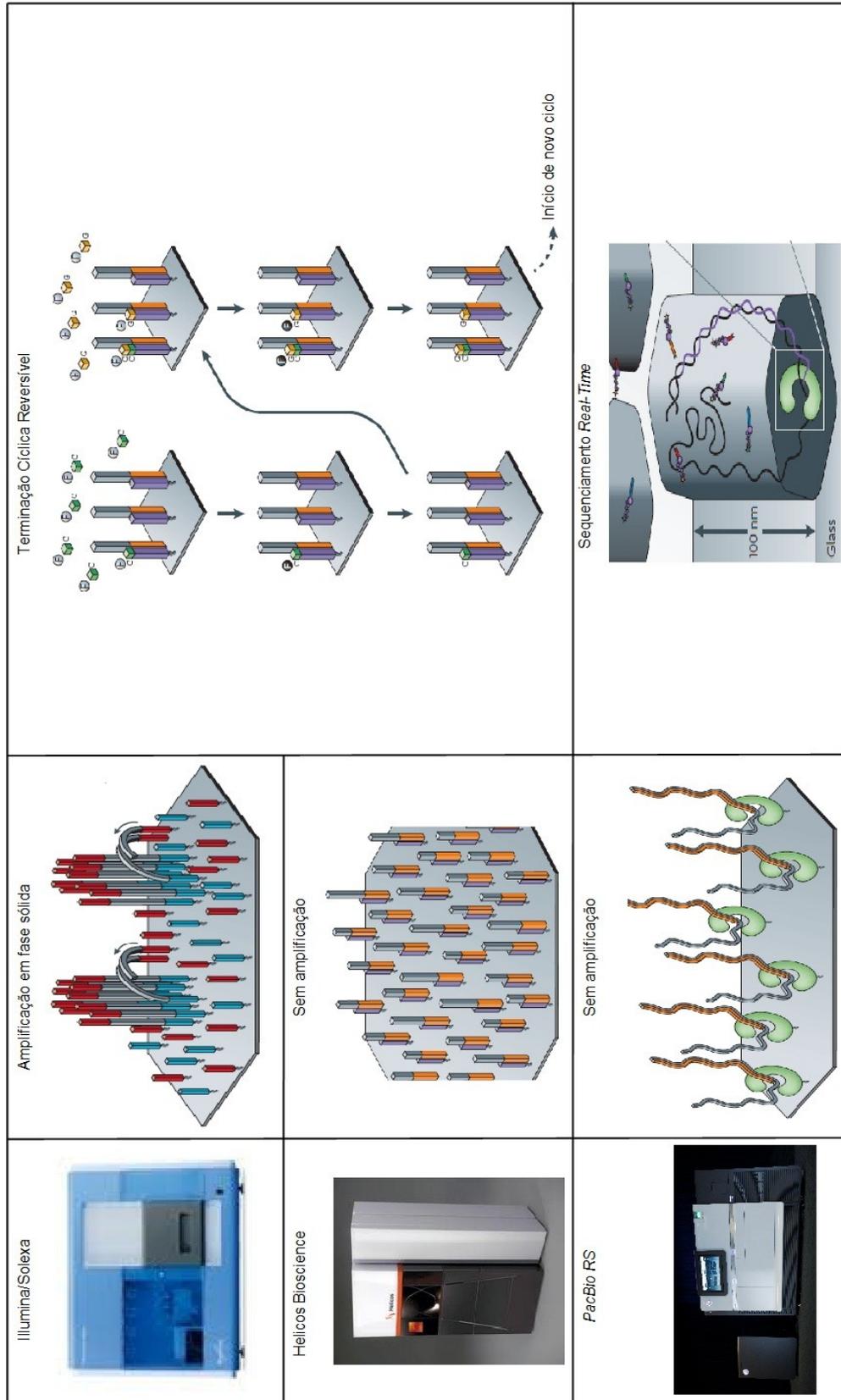


Figura 1.5: Continuação da figura 1.4.

De forma geral, o sequenciamento com estas novas tecnologias envolve as etapas de preparo da biblioteca, sequenciamento e obtenção de imagens, e análise dos dados. O que as diferencia umas das outras são os protocolos específicos para cada tecnologia resultando em tipos de dados diferentes, o que dificulta a comparação entre elas (METZKER, 2010).

No que diz respeito a etapa de preparo, primeiramente o material é randomicamente fragmentado e estes fragmentos podem ser usados para gerar uma biblioteca shotgun de fragmentos ou uma biblioteca mate-pair. A diferença entre elas é que, no preparo da biblioteca mate-pair, os fragmentos devem ser selecionados pelo tamanho, circularizados e por fim linearizados, mantendo unidas as extremidades que estavam distantes por um certo número conhecido de bases (METZKER, 2010).

Uma característica comum a estas novas tecnologias é a imobilização dos fragmentos molde separadamente sobre uma superfície, o que permite a execução simultânea de até bilhões de reações de sequenciamento (METZKER, 2010). A alta processividade, e conseqüentemente a redução de custos, proporcionadas pelos sistemas next-generation reforça a diferença entre estes instrumentos e aqueles baseados em eletroforese capilar (HERT, FREDLAKE e BARRON, 2008).

As tecnologias next-generation possuem outra característica bastante vantajosa. Durante o preparo de bibliotecas não é necessária a custosa etapa de clonagem dos fragmentos em bactérias. Ao invés disso, são empregadas outras técnicas podendo se trabalhar com fragmentos clonalmente amplificados ou não. Esta amplificação pode ser obtida pela técnica de PCR em emulsão ou por amplificação de fase sólida. No PCR em emulsão, adaptadores são ligados aos fragmentos servindo como sítios para primers que possibilitam a amplificação como num PCR comum. Entretanto, no emPCR, os fragmentos já ligados aos adaptadores são capturados por beads nos quais a amplificação acontece. Ao final, cada um dos beads deverá estar repleto de cópias de um mesmo fragmento e eles poderão ser imobilizados sobre diferentes dispositivos de acordo com a plataforma next-generation. Para a Life/APG (SOLiD) e a Polonator eles são quimicamente ligados a uma superfície de vidro; já para a 454/Roche eles são depositados em poços do dispositivo PTP (Pico Titter Plate). Na amplificação em fase sólida, usada pela plataforma Illumina/Solexa, os fragmentos presos por uma extremidade sobre uma lâmina de vidro são amplificados e a extremidade livre destes serve de sítio para o

início do sequenciamento (METZKER, 2010).

O preparo da biblioteca para plataformas como Helicos Bioscience e Pacific Bioscience, entretanto, não envolve a amplificação dos fragmentos numa abordagem chamada “molde de molécula única” o que pode ser particularmente vantajoso em algumas situações como RNA-seq (METZKER, 2010).

O GS 20 foi o primeiro sistema de sequenciamento next-generation lançado no mercado em 2005 pela 454 Life Science (Roche) (DROEGE e HILL, 2008). Esta plataforma emprega uma técnica de sequenciamento por síntese chamada pirosequenciamento (Figura 1.6). Esta técnica se baseia no fato de que, à medida que um nucleotídeo é incorporado à fita crescente de DNA, um pirofosfato é liberado e enzimaticamente convertido em ATP que, por sua vez, é usado pela enzima luciferase para gerar um sinal luminoso que é detectado pelo sistema (RONAGHI, UHLÉN e NYRÉN, 1998; HERT, FREDLAKE e BARRON, 2008; MARDIS, 2008).

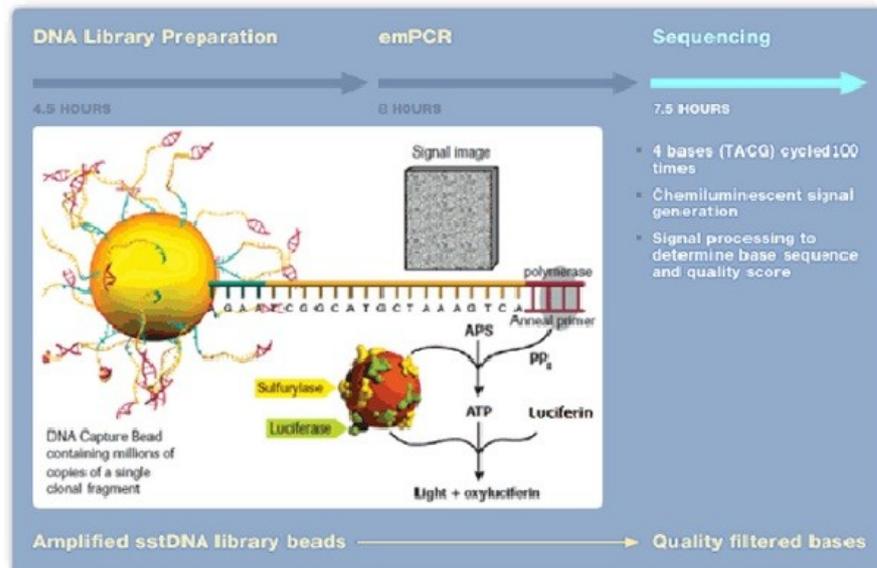


Figura 1.6: Técnica de pirosequenciamento empregada no 454 (DROEGE e HILL, 2008).

O GS FLX, um modelo melhorado em relação ao GS 20, é um sequenciador que combina alta processividade, maior comprimento médio da leitura, maior acurácia e a possibilidade de bibliotecas paired-end (DROEGE e HILL, 2008; HERT, FREDLAKE e BARRON, 2008). O dispositivo PTP recebeu uma cobertura de titânio, o que permitiu o aumento do comprimento de leituras e melhorias na qualidade dos dados evitando a comunicação entre pocinhos próximos contendo beads com fragmentos diferentes (METZKER, 2010).

Além de não ser baseada em eletroforese, conforme mencionado anteriormente a tecnologia empregada no 454 permite a eliminação da trabalhosa e demorada etapa de clonagem em bactérias. Assim, a construção da biblioteca a ser sequenciada nesta plataforma, bem como para outras tecnologias next-generation, representa menores custos sem a necessidade de produzir e armazenar clones (MOROZOVA e MARRA, 2008).

O sequenciamento com o 454 segue basicamente as etapas já mencionadas. Após a etapa inicial de nebulização, adaptadores curtos A e B são ligados aos fragmentos usando técnicas moleculares; estes adaptadores serão usados posteriormente na purificação, amplificação e sequenciamento. Na etapa seguinte, o chamado PCR em emulsão, a biblioteca de sst-DNA (single-stranded DNA) é misturada com beads que carregam oligonucleotídeos complementares a sequência do adaptador B, de modo que cada bead carregará um dos fragmentos da biblioteca. Após a adição dos reagentes de amplificação numa mistura water-in-oil, cada bead é capturado em um micro-reator onde a amplificação do fragmento acontece. A etapa final, o sequenciamento, requer ainda um mix de incubação contendo DNA polimerase e beads ainda menores recobertos das enzimas sulfúrilase e luciferase. Os beads carregando os fragmentos da biblioteca, juntamente com os beads de enzimas, são então depositados sobre um dispositivo óptico chamado PTP que será levado a plataforma. A imobilização dos beads sobre o dispositivo durante o sequenciamento otimiza a reação (DROEGE e HILL, 2008).

Durante o sequenciamento por síntese na plataforma 454, a incorporação de nucleotídeos complementares à fita molde, através do fluxo sequencial dos mesmos, emite um sinal luminoso que é detectado por uma câmera CCD. Isto ocorre porque o pirofosfato liberado é convertido pela sulfúrilase em ATP que, por sua vez, é usado por outra enzima, a luciferase, para converter luciferina em oxiluciferina, emitindo o sinal luminoso. Desta forma, o pirosequenciamento permitiu reduzir o volume da reação para o mínimo necessário à detecção. Além da miniaturização, este princípio permitiu o sequenciamento em paralelo de milhares de sequências diferentes (ROTHBERG e LEAMON, 2008).

Esta etapa de sequenciamento e obtenção de imagens também varia nas diferentes tecnologias next-generation. O método descrito acima para o sequenciamento com o 454 é referido como pirosequenciamento, ou adição de nucleotídeo único. Já as plataformas Illumina/Solexa e Helicos Bioscience, por

exemplo, usam o método chamado “cyclic reversible termination” ou terminação reversível cíclica, que consiste em realizar ciclos de adição de nucleotídeos modificados e marcados com fluorescência. Esses ciclos são intercalados por lavagens que removem nucleotídeos não incorporados. Após a incorporação do nucleotídeo, a imagem é obtida, o fluoróforo e o grupo terminador são removidos e lavados e um novo nucleotídeo pode ser incorporado. Enquanto a Illumina usa 4 cores de fluorescência, a Helicos usa apenas 1 cor (METZKER, 2010).

As plataformas Polonator e Life/APG (SOLiD) usam o método de sequenciamento por ligação, em que sondas marcadas por fluorescência hibridizam a sequências complementares e são unidas a um primer adjacente por uma DNA ligase. Neste método, a adição cíclica das sondas bem como a remoção do excedente por lavagem também são fundamentais entre a obtenção de imagens. Por fim, a plataforma da Pacific Bioscience combina uma biblioteca que não precisa de amplificação com um método conhecido como sequenciamento real-time. Neste, os nucleotídeos usados são marcados com 4 cores diferentes de fluorescência mas não param o processo de síntese e a incorporação contínua de nucleotídeos é acompanhada pela obtenção contínua de imagens. Assim, moléculas individuais de DNA polimerase são presas à superfície de detectores especiais, responsáveis por obter a sequência à medida que nucleotídeos são incorporados à fita crescente (METZKER, 2010).

Apesar da diversidade de técnicas empregadas e dos avanços conseguidos com as tecnologias next-generation, o sequenciamento com elas ainda está sujeito a erros. Os erros mais comuns são substituições nas plataformas Illumina e SOLiD, e inserções e deleções para o 454 (METZKER, 2010). Algumas desvantagens desta última tecnologia são: a necessidade de lavagem rigorosa da amostra e adição de enzimas entre cada adição de nucleotídeos, e o problema em sequenciar regiões homopoliméricas com mais de três bases de comprimento, uma vez que a intensidade do sinal luminoso está relacionada ao número de bases incorporadas (HERT, FREDLAKE e BARRON, 2008). De maneira limitada, a intensidade do sinal emitido é proporcional ao número de nucleotídeos incorporados e, conseqüentemente, podem ocorrer indels falsos (DROEGE e HILL, 2008).

1.5 PIROSSEQUENCIAMENTO NA ANÁLISE DE TRANSCRIPTOMAS

As tecnologias NGS tem sido aplicadas em estudos com diversos organismos e objetivos, inclusive em estudos de genômica funcional. As plataformas de sequenciamento next-generation estão mudando a maneira como a expressão gênica é estudada. Misturas complexas de RNA podem ser analisadas numa abordagem chamada “RNA-Seq”. Vários estudos de RNA-Seq abrem caminho para uma cobertura completa de transcriptomas com excelente resolução para a análise de nível e estrutura de transcritos, processados e não processados em diferentes condições (MARGUERAT, WILHELM e BÄHLER, 2008). Tais métodos modernos de sequenciamento tem se mostrado eficientes para a análise de transcriptomas tanto de espécies modelo quanto não modelo (MOROZOVA e MARRA, 2008).

O 454 GS FLX é uma poderosa plataforma que tem sido empregada para gerar milhões de bases de sequências de cDNA usadas na análise do transcriptoma (BARAKAT *et al.*, 2009). Suas leituras longas e acuradas permitem uma série de aplicações como anotação de genoma, identificação de novos transcritos, detecção de variantes de splicing, análise de expressão, montagem de genes inteiros e descoberta de SNP (JARVIE e HARKINS, 2008). Para sequenciamento de transcriptomas, esta plataforma tem sido a mais amplamente usada (MOROZOVA e MARRA, 2008), apesar de outras tecnologias como o Illumina também poderem ser utilizadas. O sequenciamento next-generation tem grande potencial para a caracterização de transcriptomas por causa da grande quantidade de dados gerados a custos relativamente menores quando comparados aos métodos de sequenciamento tradicionais (WALL *et al.*, 2009).

A plataforma de sequenciamento 454 GS FLX permite trabalhar com diferentes materiais de partida, como DNA genômico, cDNA, BACs e produtos de PCR. A nebulização, que gera fragmentos de 300-800bp que serão usados no preparo da biblioteca shotgun, é necessária apenas para alguns tipos de amostra (DROEGE e HILL, 2008). A nebulização do cDNA, por exemplo, pode ou não ser feita devendo-se adequar a metodologia ao tamanho dos cDNAs desejados: cDNAs menores que 800bp não devem ser nebulizados; já aqueles menores que 2kb podem não fragmentar uniformemente, o que pode se manifestar por uma cobertura

não uniforme do transcrito (JARVIE e HARKINS, 2008).

O comprimento médio das leituras obtidas com o GS FLX em experimentos com transcriptoma geralmente fica em torno dos 200bp, enquanto que para DNA genômico esta média fica entre 250-300bp (JARVIE e HARKINS, 2008). O comprimento das leituras também é afetado pelo primer usado (oligo-dT ou random primer) na síntese da primeira fita de cDNA. O uso de um primer do tipo oligo-dT para a síntese da primeira fita de cDNA, bem como purificações de RNAm após a extração do RNA Total, deve reduzir a presença dos RNAr na biblioteca a ser sequenciada (WALL *et al.*, 2009). Além disso, após a síntese do cDNA dupla fita, esta biblioteca pode ou não ser normalizada. Para pequenos sequenciamentos, uma biblioteca não normalizada deve produzir um maior número de sequências gênicas completas quando comparadas às normalizadas. À medida que o sequenciamento aumenta, isto se inverte e as bibliotecas normalizadas tendem a capturar mais sequências completas de cDNA (WALL *et al.*, 2009).

A tecnologia atualmente empregada no 454 tem se mostrado adequada ao sequenciamento de ESTs podendo gerar cerca de 400.000 leituras em uma única corrida, além de uma representação não-tendenciosa de todas as regiões de um transcrito independente de seu comprimento ou nível de expressão. Desta forma, uma única corrida pode gerar um perfil representativo de uma população de RNAm. Além disso, diferentemente das leituras muito curtas geradas em sequenciadores como o Illumina e o ABI, o comprimento das leituras originadas pelo 454 permitem a interpretação de sequências provenientes de espécies que não possuem genoma de referência ou mesmo sequências de transcriptoma para comparação (MOROZOVA e MARRA, 2008). Neste caso, as leituras devem ser montadas e usadas para identificar o espaço gênico do novo genoma. Quando sequenciamos um organismo desconhecido, pode-se ainda mapear as leituras contra um genoma/transcriptoma de organismo próximo disponível para identificar homologia entre eles (JARVIE e HARKINS, 2008).

As leituras geradas pelas tecnologias next-generation podem ser alinhadas a um genoma de referência conhecido ou montadas *de novo*. A escolha por uma das duas abordagens deve levar em conta o objetivo pretendido, os custos e o tempo demandado. Mesmo quando a opção é pelo alinhamento de leituras a um genoma de referência disponível pode-se enfrentar dificuldades com o alinhamento a regiões repetitivas ou regiões que faltam neste genoma de referência, resultantes de gaps

ou variantes estruturais. Bibliotecas mate-pair podem ajudar a resolver alguns destes problemas (METZKER, 2010).

2 OBJETIVOS

2.1 OBJETIVO GERAL

O objetivo principal deste trabalho foi caracterizar o transcriptoma de folha do acesso CL16 de *Lippia alba*, contribuindo assim para ampliar os conhecimentos sobre esta espécie.

2.2 OBJETIVOS ESPECÍFICOS

O presente trabalho teve ainda os seguintes objetivos:

- Sequenciar o transcriptoma de folha do acesso CL16 de *Lippia alba* utilizando a plataforma 454;
- Comparar a performance de dois assemblers, o Newbler2.5 e SeqMan2.2, na montagem das leituras geradas com 454;
- *In silico*, identificar transcritos expressos em folhas de *Lippia alba* e determinar a provável função destes transcritos;
- Buscar os genes/transcritos envolvidos no metabolismo secundário, mais especificamente aqueles que participam da síntese de terpenos.

3 MATERIAL e MÉTODOS

3.1 MATERIAL VEGETAL

Foram coletadas folhas, em diferentes estágios de desenvolvimento, do espécime denominado acesso CL16 de *Lippia alba*, cultivado na Estação Experimental de Plantas da Universidade Federal de Juiz de Fora, Minas Gerais. O acesso tem exsicata depositada no Herbário Leopoldo Krieger (CESJ) com número de registro 48.395. A coleta foi realizada no mês de Julho de 2010, período em que o espécime não se encontrava em floração. As amostras coletadas foram imediatamente congeladas em nitrogênio líquido até a extração do RNA total.

3.2 EXTRAÇÃO DE RNA TOTAL E SÍNTESE DE CDNA

Inicialmente foi feito um pool das folhas coletadas. O RNA total foi extraído com o Rneasy Midi Kit (Qiagen) a partir de 400 miligramas deste pool de folhas de acordo com as informações do fabricante. O RNA resultante foi eluído em 250 μ L de água Rnase-Free. A qualidade do RNA total obtido foi analisada em gel de agarose 1% e a quantificação foi feita em 260nm no UV Spectrometer Biomate3 (Thermo, USA). Além disso, o RNA total foi também analisado no Agilent 2100 Bioanalyser (Agilent Technologies).

O cDNA foi sintetizado usando o SMARTer PCR cDNA Synthesis kit (Clontech) e o Advantage PCR kit (Clontech). Com o objetivo de conseguir bom rendimento e um cDNA de qualidade, alguns testes foram feitos para otimização de parâmetros como o número e duração de ciclos.

A primeira fita de cDNA foi feita usando o SMARTer PCR cDNA Synthesis kit e um primer do tipo oligo-dT de acordo com as instruções do fabricante. A reação foi feita a partir de 1 μ g do RNA total e, ao final, eluída em 40 μ L de buffer TE (10 mM

Tris [pH 8.0], 0,1 mM EDTA). O método empregado pelo kit favorece a amplificação de cDNAs inteiros.

Para a síntese da segunda fita, foi feito o LD-PCR usando o Advantage PCR kit (Clontech). Conforme recomendado pelo fabricante, foi usado 1 µL da reação de primeira fita. O programa de amplificação executado foi: desnaturação inicial por 1 min a 95°C, seguido por 18 ciclos de: desnaturação por 15 seg a 95°C, anelamento por 30 seg a 65°C, extensão por 3 min a 68°C. O cDNA dupla fita foi quantificado no UV Spectrometer Biomate3 (Thermo, USA) e sua qualidade analisada em gel de agarose 1%.

3.3 SEQUENCIAMENTO COM O 454 GS FLX (ROCHE)

Uma amostra de cDNA dupla fita foi enviada para sequenciamento no Interdisciplinary Center for Biotechnology Research (ICBR/USA). O cDNA foi inicialmente fragmentado por nebulização e a biblioteca preparada a partir dele foi sequenciada com a plataforma 454 GS FLX (Roche). O sequenciamento foi feito em 1/8 de placa usando reagentes da série Titanium.

3.4 MONTAGEM DOS ESTs

A montagem *de novo* das sequências foi feita com dois assemblers que usam o mesmo tipo de algoritmo (Overlapping-Layout-Consensus) (Figura 3.1): o Newbler 2.5 (Roche 454) e SeqMan Ngen 2.2 (DNASStar). Para o software SeqMan foi solicitada e obtida uma licença de teste. Em ambos os softwares o arquivo “.sff” resultante do sequenciamento foi usado como entrada. A montagem com o Newbler 2.5 foi feita usando os valores default, ou seja, aqueles determinados pelo próprio programa.

Para o SeqMan, foi feita uma montagem inicial com os valores default. A partir

3.5 ANOTAÇÃO FUNCIONAL COM O Blast2GO

Usando o Blast2GO (Figura 3.2), os contigs resultantes da montagem do Newbler2.5 e da montagem escolhida do SeqMan2.2, foram anotados contra as sequências do banco de dados nr (non-redundant) do NCBI usando o algoritmo Blastx e evalue de corte de 1.0E-3. Diante da inexistência de trabalhos similares para esta espécie, o evalue aqui adotado foi baseado no trabalho feito por Alagna *et al.* (2009). O algoritmo utilizado compara sequências de nucleotídeos contra bancos de dados de proteínas e é usado para encontrar potenciais produtos de tradução para a sequência de nucleotídeos buscadas (CONESA e GÖTZ, 2009).

nr	sequence name	seq description	length	hits	min. eValue	sim mean	#GOs	GO IDs	Enzyme	InterPro
1	contig00793	photosystem II cp47 chlorophyll a/b apoprotein	2615	20	0,0E0	97,35%	8	C:photosystem II; P:transport; P:photosynthetic electron transport chain; P:protein-chromophore linkage; F:chlorophyll binding; C:integral to membrane; C:photosystem I; C:chloroplast thylakoid membrane	EC:2.1.1.14	IPR000932; IPR017486; SignalP (SIGNALP); SSF81558 (SUPERFAMILY)
2	contig00794	methionine synthase	2348	20	0,0E0	92,5%	4	C:cytoplasm; P:methionine biosynthetic process; F:zinc ion binding; F:5-methyltetrahydropteroyltylglutamate-homocysteine S-methyltransferase activity	EC:2.1.1.14	IPR002926; G3DSA:3.20.20.210 (GENE3D); SSF51726 (SUPERFAMILY)
3	contig00795	atp synthase beta subunit	2207	20	0,0E0	91,25%	8	C:proton-transporting ATP synthase complex, catalytic core F1; P:plasma membrane ATP synthesis coupled proton transport; P:atp biosynthetic process; F:hydrogen-exporting ATPase activity; phosphorylative mechanism; F:hydrogen ion transporting ATP synthase activity, rotational mechanism; F:ATP binding; F:proton-transporting ATPase activity, rotational mechanism;	EC:3.6.3.6; EC:3.6.3.14	IPR001944; IPR000793; IPR005722; IPR020003; G3DSA:1.10.1140.10 (GENE3D); G3DSA:3.40.50.300 (GENE3D); PTHR15184

Figura 3.2: Blast2GO, usado para anotação funcional das sequências.

3.6 MAPAS KEEG

Usando o Blast2GO, números EC (Enzyme Commission) foram atribuídos a sequências anotadas e estes números foram usados para o mapeamento KEGG. Foram obtidos mapas KEGG da via de biossíntese de terpenóides para as enzimas potencialmente envolvidas na síntese de terpenos.

4 RESULTADOS e DISCUSSÃO

4.1 SEQUENCIAMENTO E MONTAGEM DOS ESTs

O material de partida usado para a extração do RNA total foi um pool de folhas coletadas de um espécime de *L. alba* denominado acesso CL16. Este acesso de *Lippia alba* é proveniente do Distrito Federal e o componente majoritário de seu óleo essencial é o linalol (dados não mostrados). Juntamente com outros acessos da mesma espécie, ele vem sendo cultivado na Estação Experimental de Plantas da UFJF. Em estudo comparativo Vale (2010) mostrou que este acesso possui número cromossômico $2n=34$ e quantidade de DNA estimada em 3,11 pg por citometria de fluxo. Após análise dos resultados e agrupamento dos acessos em função de suas similaridades, este acesso foi colocado em grupo próprio.

A qualidade do RNA total foi analisada por eletroforese (Figura 4.1) e também por meio do Agilent 2100 Bioanalyser (Agilent Technologies), cujo gráfico gerado (Figura 4.2) mostra claramente os picos referentes ao 18S e 28S, o que garante a integridade do RNA total. Uma revisão das metodologias atuais de trabalhos de sequenciamento de transcriptomas com o 454 revela que tanto RNA total (ALAGNA *et al.*, 2009; VEGA-ARREGUÍN *et al.*, 2009; SUN *et al.*, 2010), quanto RNA poli-A (BARAKAT *et al.*, 2009; PARCHMAN *et al.*, 2010) são utilizados como material de partida.

A síntese do cDNA foi feita de acordo com as instruções do fabricante. A primeira fita foi sintetizada a partir da quantidade máxima sugerida (1 µg) de RNA total usando um oligo-dT, com o objetivo de reduzir a presença de sequências ribossomais nas bibliotecas a serem sequenciadas (WALL *et al.*, 2009).

Para a síntese da segunda fita de cDNA foi feita a otimização do número de ciclos. Foram testados 15, 18, 21, 24, 27 e 30 ciclos para o LD-PCR (Figura 4.3). Analisando o gel, observa-se um aumento dos produtos de PCR de 15 para 18 ciclos; a partir de 18 ciclos, no entanto, a reação atinge o plateau. Assim, o número ótimo determinado foi 18 ciclos. Após a síntese do cDNA, a maioria dos produtos obtidos variou de 500 bp a 5 Kb (Figura 4.1).

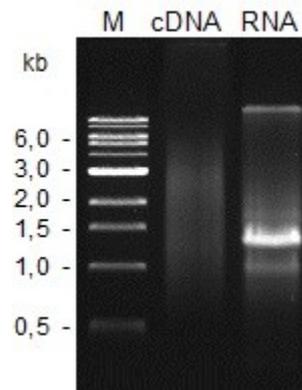


Figura 4.1: Eletroforese de cDNA dupla fita e RNA total do acesso CL16 de *L. alba* em gel de agarose 1%.

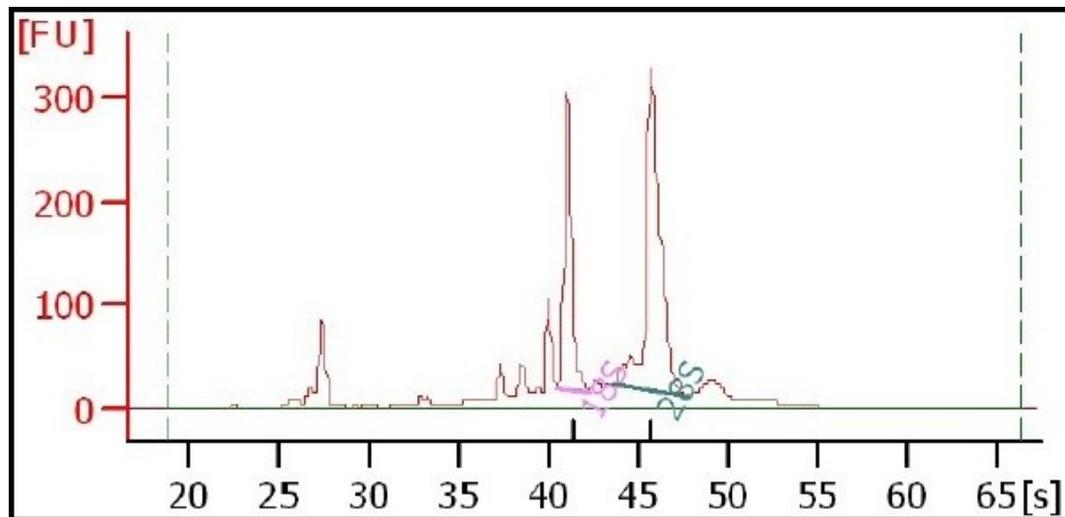


Figura 4.2: Análise do RNA total no Agilent 2100 Bioanalyser destacando os picos 18S e 28S; nos eixos do gráfico temos tempo (s) e fluorescência (FU).

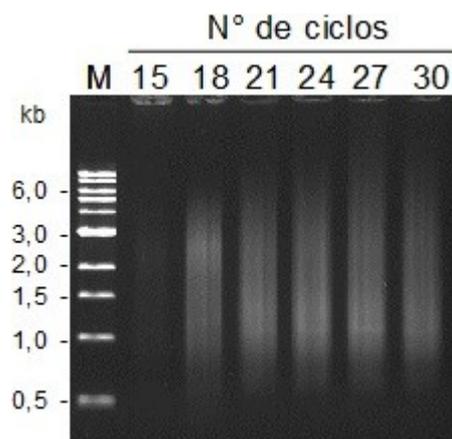


Figura 4.3: Análise da otimização do número de ciclos de PCR; foram aplicados 5 μ l em cada lane.

A biblioteca enriquecida de ds-cDNAs (double-stranded cDNA) completos foi sequenciada em corrida feita em 1/8 de placa gerando um total de 104.631 leituras e 19.302.161bp (Tabela 4.1). O tamanho médio das leituras (Figura 4.4) foi de 184,48bp, próximo a 200bp, o valor esperado em experimentos com transcriptoma realizados com o sequenciador GS FLX (Jarvie e Harkins, 2008). Este valor reflete o fato dos reads serem aparados por causa da baixa qualidade dos homopolímeros dT. Quando comparado ao de algumas outras plataformas NGS, o maior comprimento médio das leituras geradas pela 454 GS FLX facilitam a montagem *de novo* e anotação funcional das sequências em trabalhos com organismo não-modelo (WALL *et al.*, 2009; KUMAR e BLAXTER, 2010).

Tabela 4.1: Resultados obtidos com sequenciamento e montagem com o Newbler2.5.

	Sequência (n)	Bases (bp)
Sequenciamento		
Leituras	104.631	19.302.161
Comprimento médio das leituras	-	184,48
Leituras usadas na montagem	103.245	16.809.359
Contigs		
Leituras montadas em contigs	48.623	8.058.681
Número de contigs	2.686	937.010
Comprimento médio dos contigs	-	348,85
Singletons		
Número de singletons	45.265	-

O sequenciamento de transcriptomas usando as tecnologias next-generation tem se tornado cada vez mais popular graças a alta processividade e redução do custo por base proporcionados por estas tecnologias. Além disso, se comparados aos de genomas, os sequenciamentos de transcriptomas custam menos, são computacionalmente menos exigentes e ainda assim podem fornecer informações suficientes a muitos estudos (KUMAR e BLAXTER, 2010).

O sequenciamento gerou três diferentes arquivos: .qual, .fna, .sff. O terceiro, conhecido como standard flowgram file, foi o arquivo usado na etapa de montagem das leituras, que define os prováveis transcritos. A montagem correta das milhares

de leituras geradas é importante para determinar as sequências do maior número de transcritos possíveis. Alguns problemas que podem ser enfrentados na montagem de transcriptomas são erros e polimorfismos, que complicam o reconhecimento de sobreposições. Além disso, em bibliotecas não-normalizadas, a variação na abundância dos transcritos pode afetar a cobertura (KUMAR e BLAXTER, 2010).

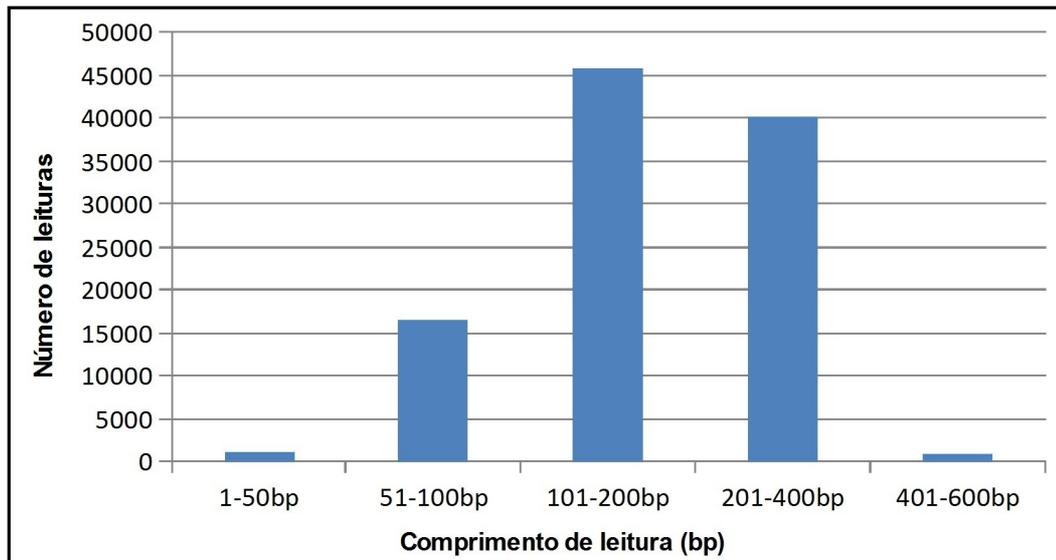


Figura 4.4: Distribuição de frequência do comprimento das leituras obtidas com o 454 GS FLX; o comprimento médio foi de 184,48bp.

Diante da inexistência de um genoma de referência que pudesse ser usado para o alinhamento, neste trabalho foram feitas montagens *de novo* das leituras usando dois diferentes assemblers. Tanto o Newbler2.5 quanto o SeqMan2.2 usam o mesmo tipo de algoritmo, Overlapping-Layout-Consensus (KUMAR e BLAXTER, 2010), apesar das diferenças nos parâmetros que podem ser ajustados para cada assembler. O algoritmo OLC envolve uma comparação inicial, em pares, de todas as leituras para a detecção de sequências únicas de bases, chamadas mers, que são usadas para determinar sobreposições entre elas. Mers comuns a duas ou mais leituras são alinhados para determinar um layout aproximado da sequência, sendo que apenas alguns mers são selecionados para isso. Por esta razão é importante escolher mers que não ocorram em regiões repetitivas, uma vez que o resultado poderia ser o alinhamento incorreto de leituras. O layout final é o resultado do alinhamento múltiplo de sequências. Três parâmetros afetam a descoberta de sobreposições e, conseqüentemente, a robustez da montagem: o tamanho do mer, o comprimento mínimo de sobreposição e a porcentagem mínima de sobreposição.

Quanto maior os valores desses parâmetros, mais acurados e menores os contigs (MILLER, KOREN e SUTTON, 2010).

Na montagem feita com o Newbler2.5, cerca de 47,09% das leituras (48.623) foram usadas gerando 2.686 contigs; 45.265 (43,84%) leituras foram consideradas singletons (Tabela 4.1). Singleton é uma leitura que não pode ser combinada com outras porque não possui sobreposição suficiente, e cujo percentual pode refletir a redundância da biblioteca. O percentual de singletons encontrado pode ser reduzido com novos sequenciamentos (WALL *et al.*, 2009). O tamanho médio dos contigs para esta montagem ficou em torno de 349bp e o maior contig gerado tem 2.815bp (Figura 4.5).

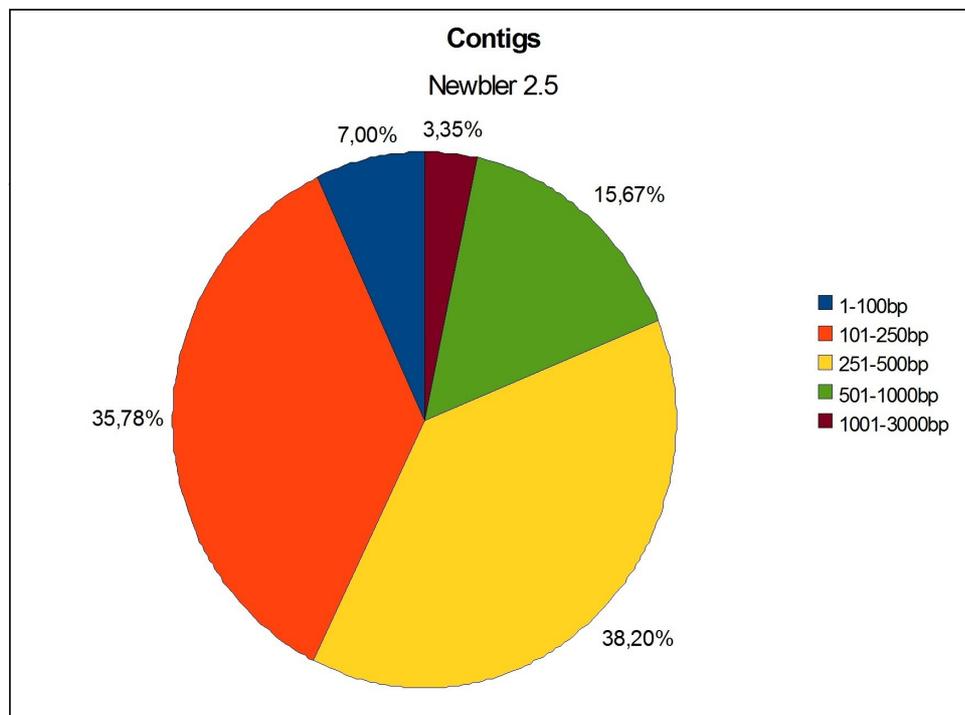


Figura 4.5: Distribuição de tamanho dos contigs montados com o Newbler2.5.

Para a montagem com o SeqMan, os parâmetros básicos do software foram variados. Os parâmetros básicos são aqueles que mais frequentemente precisam ser ajustados. São eles: match quality, match size e repeat handling. Para avaliar como a variação destes parâmetros afeta a montagem *de novo* das leituras foram feitas 9 montagens (Tabela 4.2) a partir de diferentes combinações de valores para estes parâmetros. Os resultados foram diferentes para o número de leituras usadas na montagem, o tamanho médio e o número de contigs gerados.

Tabela 4.2: Comparação das montagens feitas com o Seq-Man2.2 variando os parâmetros básicos. Os valores definidos na primeira linha da tabela são valores default, usados na montagem inicial; a montagem que usou os valores definidos na última linha da tabela foi aquela escolhida para a anotação funcional.

Match Size	Match Quality (%)	Repeat Handling/ coverage = 20*	Resultados			
			N° de contigs	Contigs >500bp	Comprimento médio (bp)	N° sequências montadas
19	85	Não	13.637	1.433	303	73.074
19	90	Não	13.717	1.409	301	67.966
19	85	Sim	13.589	1.115	290	66.004
19	90	Sim	13.666	1.113	289	61.904
21	90	Sim	13.488	1.043	287	61.205
21	95	Sim	12.705	999	286	47.382
21	100	Sim	7.173	767	300	16.295
25	85	Sim	13.270	949	286	63.812
25	90	Sim	13.448	961	284	60.131

(*) Quando a opção Repeat Handling está ativada, o valor expected coverage é solicitado; por default, este valor é 20.

O parâmetro “match size” especifica o número mínimo de bases correspondentes consecutivas necessárias para determinar a sobreposição de leituras. Já o parâmetro “match quality” especifica a porcentagem mínima de correspondência numa sobreposição para juntar duas sequências num único contig. Por fim, “repeat handling” é um conjunto de parâmetros que ajuda a identificar possíveis sequências repetitivas e que, por isso, não são boas para a detecção de sobreposições.

Ao selecionar o tipo de montagem a ser realizado e o tipo de leituras usado, o software automaticamente atribui os valores default aos diversos parâmetros. Para este trabalho, foi escolhida uma montagem *de novo* de transcriptoma e leituras obtidas com a tecnologia 454. Neste caso, os valores default atribuídos foram: 19 para match size, 85% para match quality e a opção repeat handling é desativada. Esta montagem inicial usou quase 70% das leituras (73.074) gerando 13.637 contigs com comprimento médio de 303bp.

Ainda usando o SeqMan2.2, outras combinações dos parâmetros básicos foram feitas (Tabela 4.2). Para uma dessas montagens, os valores atribuídos aos parâmetros foram: match size = 25 , match quality = 90% e a opção repeat handling

foi ativada, com expected coverage = 20 (default). Nesta montagem, cerca de 57,47% das leituras (60.131) foram montadas em 13.448 contigs (Tabela 4.3). O comprimento médio foi de 284bp e o maior contig gerado tem 2.088bp (Figura 4.6). Esta montagem, que será chamada montagem final, foi a usada na etapa de anotação funcional.

Tabela 4.3: Resultados obtidos na montagem final com o SeqMan 2.2, utilizada para a anotação funcional.

	Sequências(n)
Leituras montadas em contigs	60.131
Leituras não montadas	44.500
Número de contigs	13.448
Número de contigs >500bp	961
Comprimento médio dos contigs	284bp

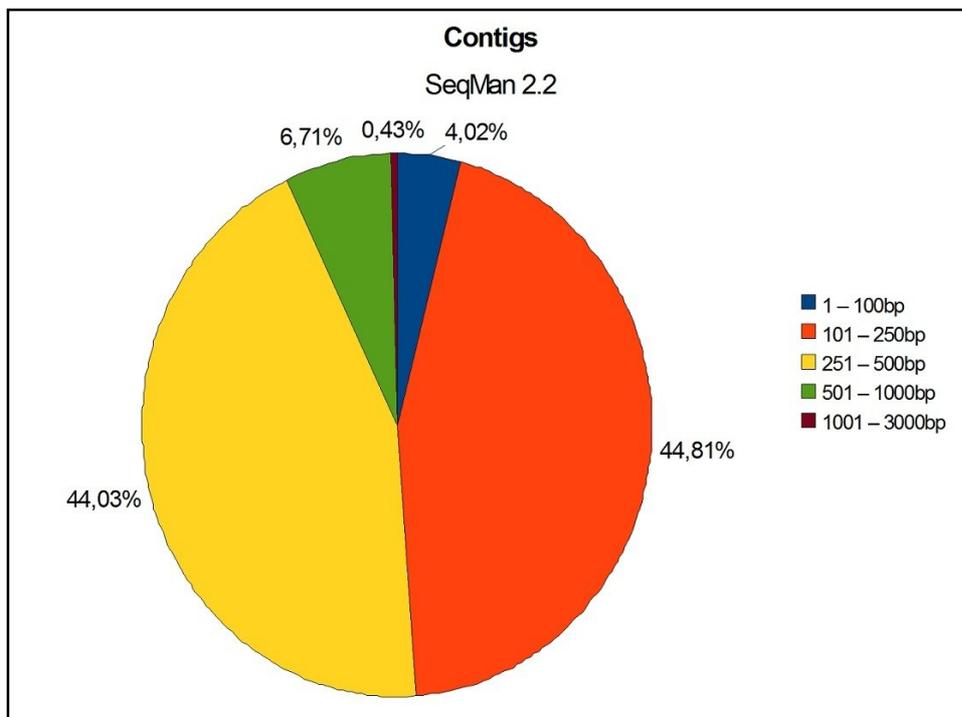


Figura 4.6: Distribuição de tamanho dos contigs da montagem final com o SeqMan2.2, usada para anotação funcional.

Comparando os resultados obtidos nas montagens inicial e final (Tabela 4.3), podemos notar que houve apenas uma pequena redução do número e do comprimento médio dos contigs.

Quando o *percentual match quality* é aumentado de 85 para 90%, o número de sequências montadas e o comprimento médio de contigs diminuiu. O número de contigs, entretanto, aumentou. Isto significa que, com 85%, um número maior de leituras foi montado num número menor de contigs. Este resultado sugere que, quando o *percentual* foi aumentado para 90%, alguns desses contigs devem ter sido desmontados originando mais de 1 contig. Resultados similares foram encontrados por Parchman *et al.* (2010).

Ainda comparando as montagens feitas com o SeqMan2.2, podemos notar que o uso da opção *repeat handling* também diminuiu o número de leituras montadas bem como o comprimento médio e número de contigs. Por fim, quanto maior o valor para *match size*, menor o número de contigs e de leituras montadas. Apesar da consequente redução no número de contigs, de leituras montadas e mesmo do tamanho médio, espera-se que montagens feitas com maiores valores para os parâmetros básicos sejam mais confiáveis, uma vez que eles estão envolvidos com a boa escolha dos *mers*. Um maior valor *match size*, por exemplo, aumenta a chance do *mer* encontrado ser único; o uso do *repeat handling* ajuda a eliminar *mers* cuja frequência seja maior que o limiar determinado e, conseqüentemente, diminui as montagens incorretas; o valor *match quality* aumentado também auxilia na solução de dificuldades provenientes de regiões repetidas.

Apesar de usar o mesmo tipo de algoritmo, os dois assemblers diferem na forma como as leituras são tratadas (KUMAR e BLAXTER, 2010), o que se refletiu nos resultados obtidos. Enquanto a montagem do Newbler, usando seus valores default, montou apenas 47,09% das leituras, o SeqMan foi capaz de montar cerca de 70% delas usando seus valores default. Mesmo aumentando os valores dos parâmetros básicos do SeqMan, este assembler montou 57,47% das leituras, *percentual* maior que o do Newbler. Outra diferença observada, que pode estar relacionada ao *percentual* de sequências montadas, foi no número de contigs gerados, aproximadamente 5 vezes maior para o SeqMan2.2.

Avaliando o critério tamanho dos contigs (Tabela 4.4), a montagem feita com o Newbler 2.5 teve cerca de 19% de seus contigs com tamanho entre 501 e 3000bp,

enquanto menos de 5% dos contigs montados pelo SeqMan se encontram nessa faixa de tamanho. A montagem com o Newbler2.5 gerou ainda maior tamanho médio de contigs bem como o contig de maior tamanho. Cabe ressaltar que uma montagem ótima não é necessariamente a maior (KUMAR e BLAXTER, 2010).

Tabela 4.4: Percentual de contigs gerados pelo Newbler2.5 e SeqMan2.2 distribuídos em categorias de tamanho.

	Newbler 2.5	SeqMan2.2
1-100bp	7,00%	4,02%
101-250bp	35,78%	44,81%
251-500bp	38,20%	44,03%
501-1000bp	15,67%	6,71%
1001-3000bp	3,35%	0,43%

Utilizando sequências obtidas com a plataforma 454, Kumar e Blaxter (2010) fizeram um estudo comparativo entre 5 assemblers atualmente usados para a montagem *de novo* de transcriptomas. Usando valores default para os parâmetros dos softwares avaliados, o Newbler2.5 e SeqMan2.2 tiveram os melhores desempenhos. Enquanto o primeiro foi capaz de gerar maiores contigs, o segundo gerou maior número de contigs, porém menores. Neste trabalho, observamos resultados similares.

Apesar de cada assembler possuir vantagens e desvantagens, diante da necessidade de optar por um único assembler, seja por tempo ou recursos limitados, Kumar e Blaxter (2010) recomendam o Newbler2.5 para montagens *de novo* de transcriptomas, uma vez que é fácil e rápido de usar, gera maiores contigs, além de possuir outra característica interessante e única, que é o agrupamento de transcritos alternativos e de isoformas. Para isso, este assembler incorpora os conceitos de isogrupo e isotig. Um isogrupo é uma coleção de contigs contendo leituras que implicam uma conexão entre eles. Após a tentativa de resolver estruturas ramificadas, o assembler agrupa os contigs cujas ramificações não puderam ser resolvidas em isogrupos. Dentro de cada isogrupo, o assembler produz então um conjunto de isotigs. Um isotig deve ser análogo a um transcrito individual e diferentes isotigs de um isogrupo podem ser variantes de splicing. Conexões entre contigs de um isogrupo são representadas por leituras que tem alinhamentos com dois ou mais contigs diferentes.

4.2 ANOTAÇÃO FUNCIONAL E MAPAS KEGG PARA VIAS METABÓLICAS

A anotação funcional, que consiste em atribuir uma função conhecida a uma sequência codificadora, foi feita com o Blast2GO. Os termos GO compõem vocabulários estruturados e descrevem produtos gênicos de acordo com sua participação em determinado processo biológico, como componente celular ou ainda por sua função molecular, independente da espécie. Esses termos já são usados por vários bancos de dados colaboradores do consórcio, e seu principal objetivo é gerar uniformidade facilitando as buscas nos mesmos.

Inicialmente, o Blast2GO busca no banco de dados selecionado sequências similares às sequências objeto. Em seguida, os termos GO associados às sequências similares são selecionados e anotados às sequências objeto. Além disso, o Blast2GO permite a associação do número EC (Enzyme Code) às sequências, o qual pode ser usado para a obtenção de mapas KEGG para rotas metabólicas.

Assim, através de comparação com o banco de dados nr (non-redundant) do NCBI usando o algoritmo Blastx e *eval* de corte $1.0E-3$, a anotação foi feita separadamente para os contigs da montagem com o Newbler2.5 (2.686) e para os contigs da montagem final feita com o SeqMan2.2 (13.448). O menor percentual de similaridade das sequências observado ficou em torno de 50% para ambas as montagens (Figura 4.7). Ao fim do processo de anotação, 1.383 e 4.153 contigs foram anotados representando, respectivamente, 51,49% da montagem do Newbler2.5 e 30,88% da montagem SeqMan2.2 (Figura 4.8). Esses resultados mostram que, apesar do SeqMan2.2 gerar um número bem maior de contigs, seu percentual de anotação funcional é menor quando comparado ao do Newbler2.5.

Depois de terminada a anotação funcional, as sequências anotadas foram analisadas para algumas categorias de processo biológico relacionadas ao metabolismo de terpenóides (Figura 4.9). Do total de sequências anotadas, 81,34% (1.125) e 76,55% (3.179), respectivamente, do Newbler e do SeqMan, foram relacionadas a algum processo biológico. Em cada montagem, cerca de 45% das sequências foram relacionadas ao metabolismo primário enquanto menos de 3% devem estar envolvidas com metabolismo secundário. As demais categorias

mostradas no gráfico (Figura 4.10) não ultrapassam 1% das sequências. Além disso, podemos observar que, nas duas montagens, o percentual de sequências em cada categoria avaliada foi bastante próximo.

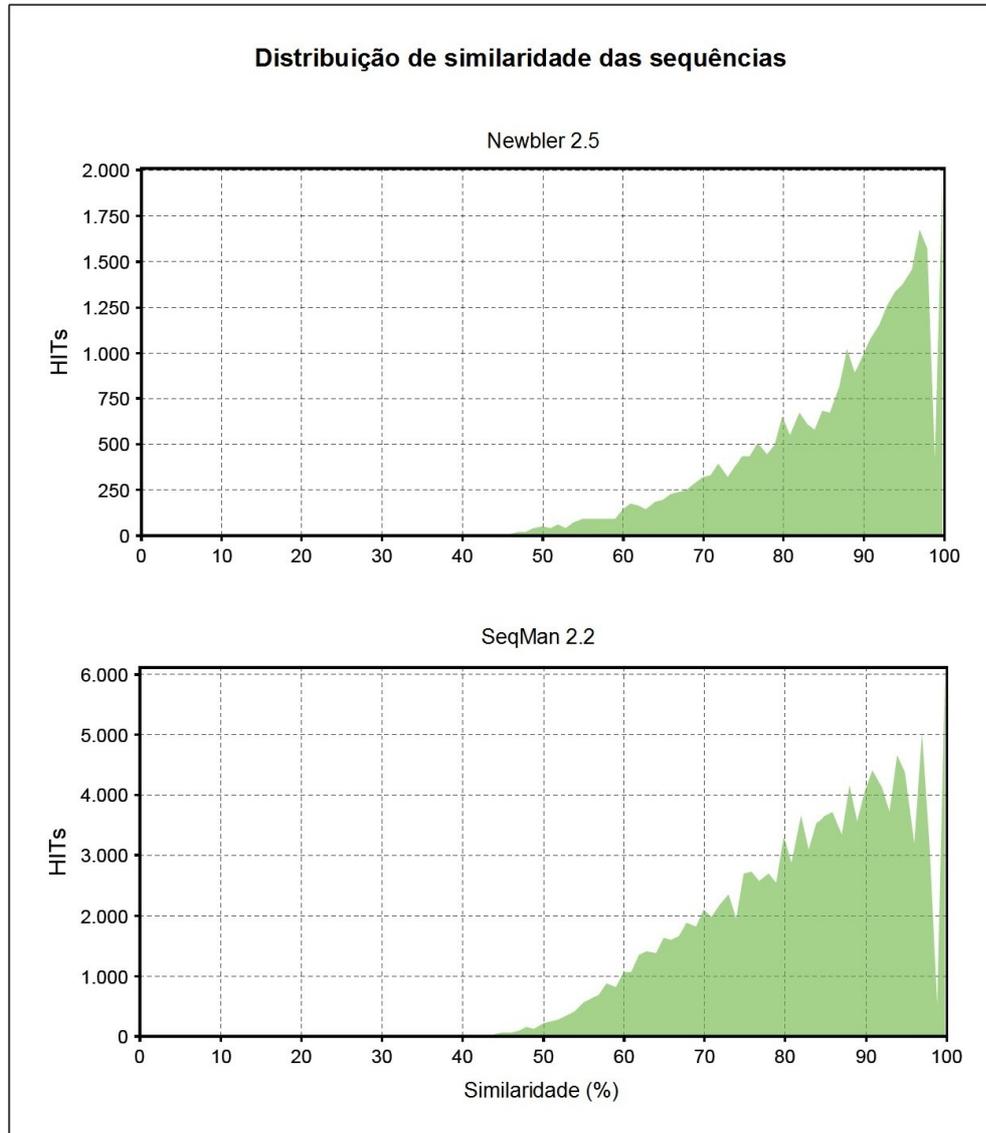


Figura 4.7: Percentual de similaridade das sequências encontradas pelo Blast2GO para os contigs das montagens do Newbler 2.5 e SeqMan 2.2.

Dentre todas as sequências anotadas para as duas montagens, algumas prováveis enzimas envolvidas na síntese de terpenos foram destacadas. Algumas destas enzimas foram identificadas em ambas as montagens, outras apenas na montagem feita com o SeqMan2.2 e uma delas apenas na montagem do Newbler2.5 (Tabela 4.5). A partir do número EC atribuído a estas sequências e ainda utilizando o BlastGO, foi obtido o mapa KEGG 00900, que mostra a via de biossíntese dos terpenóides. Estas enzimas foram manualmente identificadas e marcadas neste

mapa, a fim de posicioná-las nesta via (Figura 4.11).

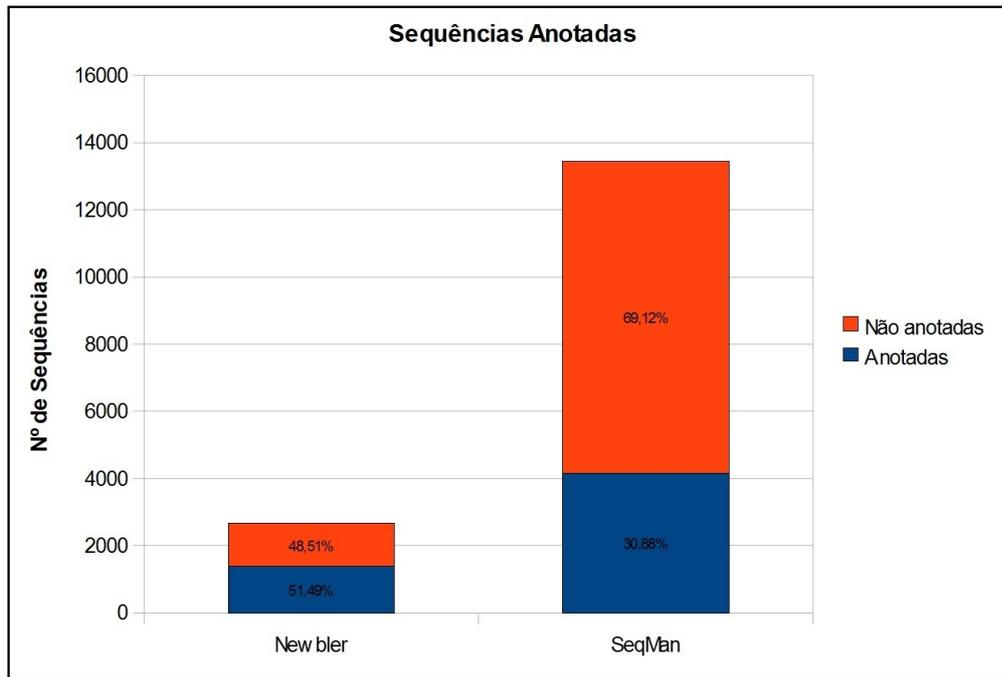


Figura 4.8: Percentual de anotação funcional dos contigs gerados com o Newbler2.5 (2.686) e SeqMan2.2 (13.448).

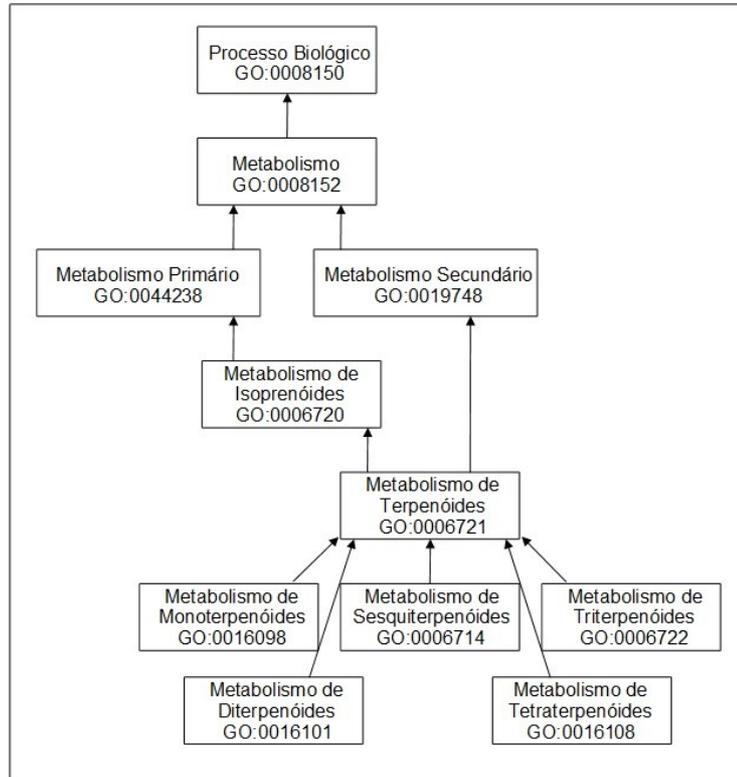


Figura 4.9: Representação hierárquica parcial de categorias de processos biológicos com seus respectivos termos GO; as categorias do esquema estão relacionadas ao metabolismo de terpenos.

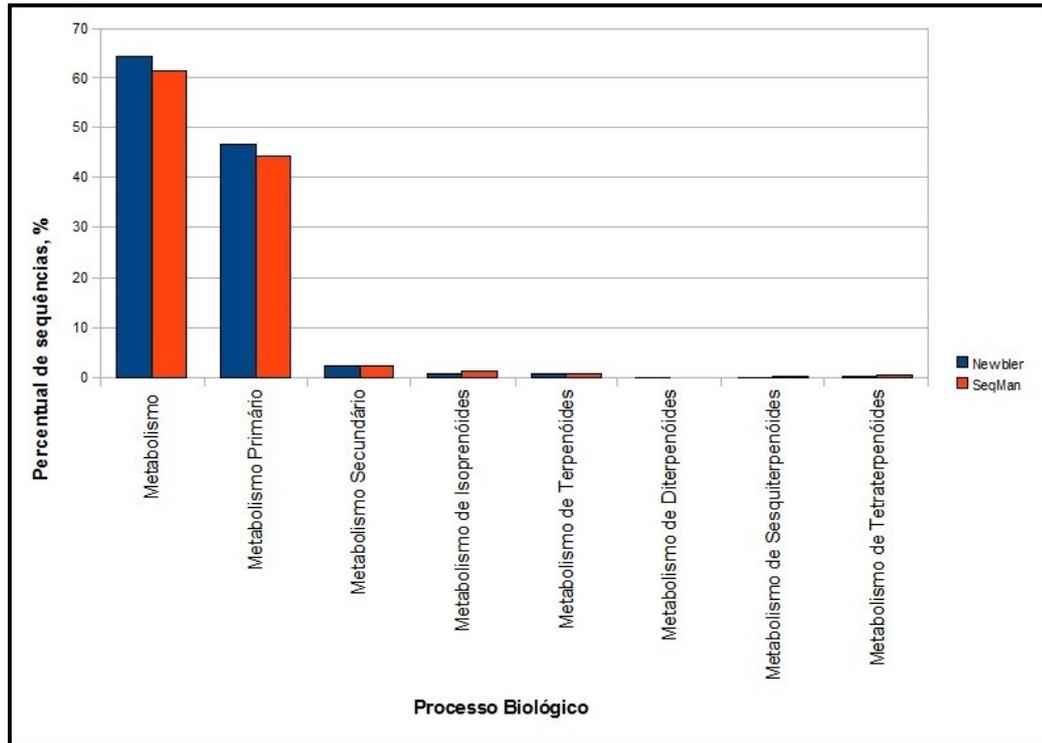


Figura 4.10: Histograma comparativo do percentual de sequências em categorias de processos biológicos prováveis com os quais os transcritos estão envolvidos. As sequências anotadas de cada montagem foram agrupadas de acordo com a classificação GO de processo biológico, dentre as quais algumas relacionadas ao metabolismo secundário.

A isopentenilpirofosfato (ou isopentenil difosfato) isomerase é uma enzima inicial da via do ácido mevalônico, responsável pela isomerização do substrato IPP em DMAPP. Esta etapa de isomerização é fundamental, uma vez que o substrato IPP não é suficientemente reativo para sofrer ionização e iniciar a condensação em terpenóides maiores (McGARVEY e CROTEAU, 1995). A rota do MEP, que também resulta na síntese do precursor IPP e ocorre nos plastídeos, também está representada por várias enzimas encontradas (Figura 4.11).

A farnesil pirofosfato sintase e a geranylgeranyl pirofosfato sintase são preniltransferases que formam os intermediários de 15 e 20 carbonos, FPP e GGPP, respectivamente (McGARVEY e CROTEAU, 1995).

A escaleno sintase cataliza uma reação de condensação de duas moléculas de farnesil difosfato gerando o escaleno (CHRISTIANSON, 2006). A fitoeno sintase, por sua vez, cataliza a conversão de duas moléculas de geranylgeranyl difosfato (GGPP) a fitoeno, ou seja, o segundo passo na biossíntese de carotenóides a partir de isopentenil difosfato. A fitoeno sintase é encontrada em todos os organismos que sintetizam carotenóides como as plantas, e está localizada nos cloroplastos. Além da

similaridade funcional, a escaleno sintase e a fitoeno sintase apresentam similaridades em sua estrutura primária.

Por fim, a análise *in silico* do transcriptoma de *L. alba* mostrou sequências parciais de duas prováveis sesquiterpeno sintases (valenceno sintase e patchoulol sintase) e duas monoterpene sintases (linalol sintase e -linalol-nerolidol ()-geraniol sintase) (Tabela 4.6).

A maior parte destas enzimas identificadas estão envolvidas na síntese de precursores iniciais como o IPP e DMAPP, ou seja, cujo nível de expressão mais elevado deve ser compatível com a necessidade dos precursores sintetizados por elas. Assim, a maior abundância destes transcritos deve ter possibilitado sua detecção nesta análise, diferentemente de enzimas finais da via, cujo nível mais baixo de transcritos foi insuficiente para sua descoberta neste sequenciamento. Ainda assim, foram identificadas sequências parciais de terpeno sintases, como a linalol sintase e patchoulol sintase.

Tabela 4.5: Sequências anotadas do Newbler (N) e do SeqMan (S), com seus respectivos números EC, potencialmente envolvidas com a síntese de terpenos.

Identificação do contig	Comprimento do contig (bp)	Descrição da sequência	Número EC
00924(N); 04995(S)	801(N); 210(S)	isopentenilpirofosfato isomerase	5.3.3.2
00890(N); 01345(S)	874(N); 252(S)	fitoeno sintase	2.5.1.32
01001(N); 01811(S)	683(N); 276(S)	escaleno sintase	2.5.1.21
04302(S)	180(S)	farnesil pirofosfato sintase	2.5.1.10; 2.5.1.1
01509(N); 01040(S)	368(N); 255(S)	geranilgeranil pirofosfato sintase-related protein	2.5.1.29; 1.3.1.74
05085(S); 00877(S)	287(S); 273(S)	geranilgeranil difosfato sintase	2.5.1.29; 2.5.1.10; 2.5.1.1
01332(N)	432(N)	2-c-metil-d-eritritol 4-fosfato citidiltransferase	2.7.7.60
03901(S); 02049(N)	367(S);255(N)	2-c-metil-d-eritritol -ciclodifosfato sintase	4.6.1.12
01564(N); 00570(S); 02200(N)	353(N); 351(S); 225(N)	1-hidroxi-2-metil-2-butenil 4-difosfato redutase	1.17.1.2
05077(S); 00634(S); 05542(S)	368(S); 272(S); 259(S)	4-hidroxi-3-metilbut-2-enil difosfato redutase	1.17.1.2; 1.17.4.3
06994(S); 07309(S)	232(S); 199(S)	hidroximetilbutenil 4-difosfato sintase	1.17.4.3
00997(S); 07264(S)	266(S); 145(S)	zeaxantina epoxidase	1.14.13.90
09236(S); 09123(S)	228(S); 205(S)	zeta-caroteno desaturase	1.14.99.30
10937(S); 07032(S); 04213(S); 01225(S)	285(S); 240(S); 220(S); 184(S)	hidroximetilglutaril- sintase	2.3.3.10; 2.3.1.9
12295(S); 03850(S)	270(S); 227(S)	solanesil difosfato sintase	1.14.13.0; 2.4.1.14; 2.5.1.30; 2.5.1.11

Tabela 4.6: Potenciais terpeno sintases identificadas; Newbler(N) e do SeqMan (S).

Identificação do contig	Comprimento do contig (bp)	Descrição da sequência
01372(N)	420(N)	(+)-valenceno sintase
01307(N); 10096(S)	455(N); 311(S)	patchoulol sintase
04966(S); 08261(S)	386(S);386(S)	linalol sintase
13011(S); 08117(S)	279(S); 348(S)	-linalol-nerolidol ()-geranil linalol sintase

5 CONCLUSÕES

A diversidade de compostos de interesse que podem ser encontrados em *L. alba*, especialmente aqueles que compõem seu óleo essencial como os terpenos, deixam claro o potencial desta espécie. Assim, o conhecimento mais detalhado da síntese destes compostos, inclusive das enzimas envolvidas, pode resultar no desenvolvimento de fármacos e produção de substâncias de importância comercial. Diante de nossos objetivos, optamos pelo sequenciamento do transcriptoma da folha de *Lippia alba* com a plataforma 454 GS FLX (Roche).

A metodologia utilizada neste trabalho permitiu uma primeira e ampla caracterização do transcriptoma de folha deste acesso de *L. alba*, com a identificação dos transcritos e suas prováveis funções. O uso do RNA total juntamente com o oligo-dT para a síntese de cDNA pode produzir material de qualidade para um bom sequenciamento com o 454 GS FLX e montagem *de novo* das leituras geradas.

A exploração dos parâmetros básicos do assembler SeqMan2.2 se mostrou uma ferramenta útil para melhorar as montagens. Isto significa que, apesar dos valores default para esses parâmetros poderem ser utilizados, valores maiores podem ser uma forma de aumentar a acurácia dos contigs montados.

A comparação de montagens feitas com o Newbler2.5 e SeqMan2.2 deixou claro que as diferenças no modo como estes assemblers montam as leituras influenciam bastante nos resultados gerados. As vantagens e desvantagens exibidas por cada um permitem concluir que o uso de diferentes assemblers para montar um mesmo conjunto de leituras pode ser uma boa estratégia para a melhor descoberta de transcritos.

A anotação funcional e análise das sequências, feita para todos os contigs, permitiu a identificação de um grande número de transcritos de folha de *L. alba*, incluindo algumas enzimas provavelmente envolvidas na síntese de terpenos. Um novo sequenciamento pode melhorar a descoberta de transcritos menos abundantes, como as terpeno sintases.

Desta forma, este trabalho pioneiro deverá servir como referência para novos sequenciamentos do mesmo material, adequando futuras metodologias a objetivos mais específicos.

6 REFERÊNCIAS BIBLIOGRÁFICAS

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

AGUIAR, J.S.; COSTA, M.C.C.D. 2005. *Lippia alba* (Mill.) N. E. Brown (Verbenaceae): levantamento de publicações nas áreas química, agronômica e farmacológica, no período de 1979 a 2004. **Revista Brasileira de Plantas Mediciniais**, **8(1)**: 79-84.

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). **Revista Brasileira de Farmacognosia**, **18(3)**: 436-440.

ALAGNA, F.; D'AGOSTINO, N.; TORCHIA, L.; SERVILI, M.; RAO, R.; PIETRELLA, M.; GIULIANO, G.; CHIUSANO, M.L.; BALDONI, L.; PERROTA, G. 2009. Comparative 454 pyrosequencing of transcripts from two olive genotypes during fruit development. **BMC Genomics**, **10**: 399-414.

ATTI-SERAFINI, L.; PANSERA, M.R.; ATTI-SANTOS, A.C.; ROSSATO, M.; PAULETTI, G.F.; ROTA, L.D.; PAROUL, N.; MOYNA, P. 2002. Variation in essential oil yield and composition of *Lippia alba* (Mill.) N.E.Br. grown in southern Brazil. **Revista Brasileira de Plantas Mediciniais**, **4(2)**: 72-74.

BARAKAT, A.; DILORETO, D.S.; ZHANG, Y.; SMITH, C.; BAIER, K.; POWELL, W.A.; WHEELER, N.; SEDEROFF, R.; CARLSON, J.E. 2009. Comparison of the transcriptomes of American chestnut (*Castanea dentata*) and Chinese chestnut (*Castanea mollissima*) in response to the chestnut blight infection. **BMC Plant Biology**, **9**:51.

BOHLMANN, J.; MEYER-GAUEN, G.; CROTEAU, R. 1998. Plant terpenoid synthases: Molecular biology and phylogenetic analysis. **Proceedings of the National Academy of Sciences**, **95**: 4126-4133.

CAMPOS, J.M.S.; SOUSA, S.M.; SILVA, P.S.; PINHEIRO, L.C.; SAMPAIO, F.; VICCINI, L.F. 2010. Chromosome numbers and DNA C values in the genus *Lippia* (Verbenaceae). **Plant Systematics and Evolution**, **291**: 133-140.

CHRISTIANSON, D.W. 2006. Structural Biology and Chemistry of the Terpenoid Cyclases. **Chemical Reviews**, **106**: 3412-3442.

CONESA, A.; GÖTZ, S. 2009. **Blast2GO Tutorial, Bioinformatics and Genomics Department**. Prince Felipe Research Center, Spain, Valencia.

DROEGE, M.; HILL, B. 2008. The Genome Sequencer FLX(TM) System—Longer reads, more applications, straight forward bioinformatics and more complete data sets. **Journal of Biotechnology**, **136(1-2)**: 3-10.

DUBEY, V.S.; BHALLA, R.; LUTHRA, R. 2003. An overview of the non-mevalonate pathway for terpenoid biosynthesis in plants. **Journal of Biosciences**, **28**: 637-646.

DUDAREVA, N.; MARTIN, D.; KISH, C.M.; KOLOSOVA, N.; GORENSTEIN, N.; FÄLDT, J.; MILLER, B.; BOHLMANN, J. 2003. (E)-b-Ocimene and Myrcene synthase Genes of Floral Scent Biosynthesis in Snapdragon: Function and Expression of Three Terpene Synthase Genes of a New Terpene Synthase Subfamily. **The Plant Cell**, **15**: 1227-1241.

HEINZMANN, B.M.; BARROS, F.M.C. 2007). Potencial das plantas nativas brasileiras para o desenvolvimento de fitomedicamentos tendo como exemplo *Lippia alba* (Mill.) N.E.Brown (Verbenaceae). **Saúde, Santa Maria**, **33**: 43-48.

HERT, D.G.; FREDLAKE, C.P.; BARRON, A.E. 2008. Advantages and limitations of nextgeneration sequencing technologies: a comparison of electrophoresis and non-electrophoresis methods. **Electrophoresis**, **29**: 4618-4626.

JARVIE, T.; HARKINS, T. 2008. Transcriptome sequencing with the Genome Sequencer FLX system. **Nature Methods**, 6-8.

JULIÃO, L.S.; TAVARES, E.S.; LAGE, C.L.S.; LEITÃO, S.G. 2003. Cromatografia em camada fina de extratos de três quimiotipos de *Lippia alba* (Mill) N.E.Br. (erva-cidreira). **Revista Brasileira de Farmacognosia**, **13**: 36-38.

KESSLER, A.; BALDWIN, I.T. 2002. Plant responses to insect herbivory: the emerging molecular analysis. **Annual Review of Plant Biology**, **53**: 299–328.

KUMAR, S.; BLAXTER, M.L. 2010. Comparing de novo assemblers for 454 transcriptome data. **BMC Genomics**, **11**: 571.

LUCKER, J.; BOWEN, P.; BOHLMANN, J. 2004. *Vitis vinifera* terpenoid cyclases:

functional identification of two sesquiterpene synthase cDNAs encoding (+)-valencene synthase and (-)-germacrene D synthase and expression of mono- and sesquiterpene synthases in grapevine flowers and berries. **Phytochemistry**, **65(19)**: 2649-2659.

MARGUERAT, S.; WILHELM, B.T.; BÄHLER, J. 2008. Next-generation sequencing: applications beyond genomes. **Biochemical Society Transactions**, **36**: 1091-1096.

MARDIS, E.R. 2008. Next-Generation DNA Sequencing Methods. **Annual Review of Genomics and Human Genetics**, **9**: 387-402.

MATOS, F.J.A. 1996. As ervas cidreiras do Nordeste do Brasil. Estudo de três quimiotipos de *Lippia alba* (Mill.) N. E. Brown (Verbenaceae) Parte II - Farmacoquímica. **Revista Brasileira de Farmacognosia**, **77(4)**: 137-141.

MATOS, F.J.A.; MACHADO, M.I.L.; CRAVEIRO, A.A.; ALENCAR, J.W. 1996. The essential oil composition of two chemotypes of *Lippia alba* grown in Northeast Brazil. **Journal of Essential Oil Research**, **8**: 695-698.

McGARVEY, D.J.; CROTEAU, R. 1995. Terpenoid Metabolism. **The Plant Cell**, **7**: 1015-1026.

METZKER, M.L. 2010. Sequencing technologies: the next generation. **Nature Reviews Genetics**, **11(1)**: 31-46.

MILLER, J.R.; KOREN, S.; SUTTON, G. 2010. Assembly algorithms for next-generation sequencing data. **Genomics**, **95**: 315-327.

MOROZOVA, O.; MARRA, M.A. 2008. Applications of next-generation sequencing technologies in functional genomics. **Genomics**, **92(5)**: 255-264.

MUNCK, S.L.; CROTEAU, R. 1990. Purification and characterization of the sesquiterpene cyclase patchoulol synthase from *Pogostemon cablin*. **Archives of Biochemistry and Biophysics**, **282(1)**: 58-64.

PARCHMAN, T.L.; GEIST, K.S.; GRAHNEN, J.A.; BENKMAN, C.W.; BUERKLE, C.A. 2010. Transcriptome sequencing in an ecologically important tree species: assembly, annotation, and marker discovery. **BMC Genomics**, **11**: 180.

PASCUAL, M.E.; SLOWING, K.; CARRETERO, M.E.; Villar. 2001. Antiulcerogenic activity of *Lippia alba* (Mill.) N. E. Brown (Verbenaceae). **II Farmaco**, **56**: 501-504.

PASCUAL, M.E.; SLOWING, K.; CARRETERO, E.; MATA D.S.; Villar, A. 2001. *Lippia*: traditional uses, chemistry and pharmacology: a review. **Journal of Ethnopharmacology**, **76**: 201-214.

RONAGHI, M.; UHLÉN, M.; NYRÉN, P. 1998. DNA Sequencing: A sequencing method based on real-time pyrophosphate. **Science**, **281**: 363-365.

SEEMAN, M.; WEGNER, P.; SCHÜNEMANN, V.; BUI, B.T.S.; WOLF, M.; MARQUET, A.; TRAUTWEIN, A.X.; ROHMER, M. 2005. Isoprenoid biosynthesis in chloroplasts via the methylerythritol phosphate pathway: the (E)-4-hydroxy-3-methylbut-2-enyl diphosphate synthase (GcpE) from *Arabidopsis thaliana* is a [4Fe-4S] protein. **Journal of Biological Inorganic Chemistry**, **10(2)**: 131-137.

SIANI, A.C.; TAPPIN, M.R.R.; RAMOS, M.F.S.; MAZZEI, J.L.; RAMOS, M.C.K.V.; NETO, F.R.A.; FRIGGHETTO, 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**: 3518-3521.

SUN, C.; LI, Y.; WU, Q.; LUO, H.; SUN, Y.; SONG, J.; LUI, E.M.K.; CHEN, S. 2010. *De novo* sequencing and analysis of the American ginseng root transcriptome using a GS FLX Titanium platform to discover putative genes involved in ginsenoside biosynthesis. **BMC Genomics**, **11**: 262.

TAIZ, L.; ZEIGER, E. 2004. **Fisiologia Vegetal**. Artmed, Porto Alegre.

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) cultivados em condições semelhantes. **Brazilian Journal of Pharmacognosy**, **15(1)**: 1-5.

VALE, A.A. 2010. **Plasticidade fenotípica em acessos de *Lippia alba* (Mill.) N.E.BR EX Britton & Wilson (Verbenaceae)**. Dissertação de mestrado, Universidade Federal de Juiz de Fora, Brasil, Juiz de Fora, 75p.

VALE, T.G.; MATOS, F.J.A.; LIMA, T.C.M.; VIANA, G.S.B. 1999. Behavioral effects of essential oils from *Lippia alba* (Mill.) N.E. Brown chemotypes. **Journal of Ethnopharmacology**, **167**: 127-133.

VICCINI, L.F.; PIERRE, P.M.O.; PRAÇA, M.M.; COSTA, D.C.S.; ROMANEL, E.C.; SOUSA, S.M.; PEIXOTO, P.H.P.; SALIMENA, F.R.G. 2006. Chromosome numbers in the genus *Lippia* (Verbenaceae). **Plant Systematics and Evolution**, **256**: 171-178.

VEGA-ARREGUÍN, J.C.; IBARRA-LACLETTE, E.; JIMÉNEZ-MORAILA, B.; MARTÍNEZ, O.; VIELLE-CALZADA, J.P.; HERRERA-ESTRELLA, L.; HERRERA-ESTRELLA, A. 2009. Deep sampling of the Palomero maize transcriptome by a high throughput strategy of pyrosequencing. **BMC Genomics**, **10**: 299.

WALL, P.K.; LEEBENS-MACK, J.; CHANDERBALI, A.S.; BARAKAT, A.; WOLCOTT, E.; LIANG, H.; LANDHERR, L.; TOMSHO, L.P.; HU, Y.; CARLSON, J.E.; MA, H.; SCHUSTER, S.C.; SOLTIS, D.E.; SOLTIS, P.S.; ALTMAN, N.; PAMPHILIS, C.W. 2009. Comparison of next generation sequencing technologies for transcriptome characterization. **BMC Genomics**, **10**: 347.

ZÉTOLA, M.; LIMA, T.M.C.; SONAGLIO, D.; GONZÁLEZ-ORTEGA, G.; LIMBERGER, R.P.; PETROVICK, P.R.; BASSANI, V.L. 2002. CNS activities of liquid and spray-dried extracts from *Lippia alba* Verbenaceae (Brazilian false melissa). **Journal of Ethnopharmacology**, **82**: 207-215.