



UNIVERSIDADE  
ESTADUAL DE LONDRINA

---

DANILO RIBEIRO DE BRITO

**MONTAGEM DO TRANSCRIPTOMA DE *Coffea eugenioides*  
E IDENTIFICAÇÃO DOS GENES DIFERENCIALMENTE  
EXPRESSOS EM FOLHAS E FRUTOS**

---

DANILO RIBEIRO DE BRITO

**MONTAGEM DO TRANSCRIPTOMA DE *Coffea eugenioides*  
E IDENTIFICAÇÃO DOS GENES DIFERENCIALMENTE  
EXPRESSOS EM FOLHAS E FRUTOS**

DANILO RIBEIRO DE BRITO

**MONTAGEM DO TRANSCRIPTOMA DE *Coffea eugenioides*  
E IDENTIFICAÇÃO DOS GENES DIFERENCIALMENTE  
EXPRESSOS EM FOLHAS E FRUTOS**

Dissertação apresentada ao Programa de Pós-Graduação em Genética e Biologia Molecular, da Universidade Estadual de Londrina, como requisito parcial para a obtenção do título de Mestre.

Orientador: Dr. Paulo Maurício Ruas  
Co-orientador: Dr. Luiz Filipe Protasio Pereira

Londrina  
2019

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

DE BRITO, DANILO RIBEIRO .

MONTAGEM DO TRANSCRIPTOMA DE *Coffea eugenioides* E IDENTIFICAÇÃO DOS GENES DIFERENCIALMENTE EXPRESSOS EM FOLHAS E FRUTOS / DANILO RIBEIRO DE BRITO. - Londrina, 2019.  
90 f. : il.

Orientador: Dr. PAULO MAURÍCIO RUAS.

Coorientador: Dr. LUIZ FILIPE PROTASIO PEREIRA.

Dissertação (Mestrado em Genética e Biologia Molecular) - Universidade Estadual de Londrina, Centro de Ciências Biológicas, Programa de Pós-Graduação em Genética e Biologia Molecular, 2019.

Inclui bibliografia.

1. CAFÉ - Tese. 2. RNA-SEQ - Tese. 3. ANOTAÇÃO FUNCIONAL - Tese. I. RUAS, Dr. PAULO MAURÍCIO. II. PEREIRA, Dr. LUIZ FILIPE PROTASIO. III. Universidade Estadual de Londrina. Centro de Ciências Biológicas. Programa de Pós-Graduação em Genética e Biologia Molecular. IV. Título.

DANILO RIBEIRO DE BRITO

**MONTAGEM DO TRANSCRIPTOMA DE *Coffea eugenioides* E  
IDENTIFICAÇÃO DOS GENES DIFERENCIALMENTE EXPRESSOS  
EM FOLHAS E FRUTOS**

Dissertação apresentada ao Programa de Pós-Graduação em Genética e Biologia Molecular, da Universidade Estadual de Londrina, como requisito parcial para a obtenção do título de Mestre.

**BANCA EXAMINADORA**

---

Dr. Luiz Filipe Protasio Pereira  
Universidade Estadual de Londrina - UEL

---

Dra. Suzana Tiemi Ivamoto  
Universidade Estadual Paulista - UNESP

---

Dr. Rogério Fernandes de Souza  
Universidade Estadual de Londrina - UEL

Londrina, 22 de fevereiro de 2019.

Dedico este trabalho aos meus pais Nestor e Aparecida, à minha esposa Juliana, ao meu filho Pedro, à minha família e aos meus amigos. Todos me deram forças e incentivo, sempre me amparando com muito amor e carinho para vencer as batalhas de estudo.

## **AGRADECIMENTOS**

Agradeço a Deus, o autor da vida, por ter me contemplado e amparado com a sua luz divina e sabedoria em todo o decorrer de minha vida;

Agradeço infinitamente à minha eterna companheira e amada esposa Juliana, por tanto amor, carinho, paciência, pelo apoio e compreensão, por acreditar em mim me incentivando em meus estudos, além da confiança depositada nas minhas escolhas e também por fazer-me mais feliz me ajudando a tornar uma pessoa melhor a cada dia. Sem contar sua imensa ajuda dentro desta pesquisa, sem você meu entendimento seria infinitamente menor, obrigada por ser minha bióloga;

Agradeço muito pelo presente maravilhoso que recebi este ano, meu filho Pedro, minha riqueza, minha maior felicidade;

Aos meus amados pais, Nestor e Aparecida que sempre me apoiaram e torceram por minhas vitórias. A meu sogro por toda a ajuda nessa caminhada;

Obrigada à UEL e ao IAPAR pelo espaço cedido e toda a colaboração que permitiu a execução desta pesquisa;

À CAPES pelo apoio financeiro, o presente trabalho foi realizado com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nivel Superior – Brasil (Capes) – Código de Financiamento 001.

Muito obrigado a minha orientadora Dra. Claudete de Fátima Ruas, por ter dado continuidade na minha orientação, pelos ensinamentos científicos e toda a atenção, ao Dr. Paulo Ruas pela oportunidade e disposto seu tempo para dar as primeiras orientações nessa caminhada;

Agradeço imensamente a meu co-orientador Dr. Luiz Filipe Protasio Pereira, por acreditar em mim, por incentivar sempre e pelas grandes oportunidades que me ofereceu, sem sua ajuda esse trabalho não existiria;

Aos meus amigos: Martina, Luciano, Rafaela, Jean e Renato por tanto apoio. Vocês foram fundamentais para minha persistência nesta etapa da vida e também

por terem me ajudado e contribuído de alguma forma nesse trabalho, principalmente;

Aos Professores Dr. Rogério Fernandes de Souza e Dr. Laurival Villas-Boas, por me apresentarem essa área interessantíssima que é a Bioinformática e por me incentivarem a seguir nesse Programa de Genética e Biologia Molecular, muito obrigado;

Ao Programa de Genética e Biologia Molecular pela oportunidade concedida, juntamente agradeço à professora Silvia Helena Sofia pela dedicação e esforço que emprega para ajudar todos os alunos nos motivando dia após dia nessa caminhada;

Obrigada Dra. Suzana Tiemi Ivamoto pelas grandes contribuições, incentivos e amizade depositados a mim neste percurso, você fez muita diferença;

Agradeço também a todas as pessoas que me ajudaram e me apoiaram de alguma maneira na execução desta dissertação e conclusão do mestrado.

*“Em tudo o que você fizer, seja sempre humilde, guardando zelosamente a pureza de seu coração e a pureza de seu corpo. Se tudo o que somos e temos recebemos de Deus, do que podemos gloriar-nos? É preciso calar e ter silêncio ao seu redor para poder ouvir a voz de Deus.”*

*São Padre Pio de Pietrelcina*

DE BRITO, Danilo Ribeiro. **Montagem do Transcriptoma de *Coffea eugenoides* e Identificação dos Genes Diferencialmente Expressos em Folhas e Frutos**. 2019. 90 f. Dissertação (Mestrado em Genética e Biologia Molecular) – Universidade Estadual de Londrina, 2019.

## RESUMO

O café encontra-se entre as principais *commodities* agrícolas do Brasil, o qual ocupa o posto de maior produtor e exportador de café no mundo. Dentre as 124 espécies do gênero, *Coffea arabica*, *Coffea canephora* e *Coffea eugenoides* são as de maior destaque e as duas primeiras representam a maior parte da produção de café. Com exceção do *C. arabica*, todas as demais espécies do gênero são diplóides. *C. arabica* é um alotetraplóide, formado a partir de uma hibridação natural entre *C. canephora* e *C. eugenoides*. O parental *C. eugenoides* ainda foi pouco estudado, porém muitas características genéticas dos cafés Árabicas comerciais foram herdadas desta espécie. Desta forma, é justificada-se aumentar o conhecimento sobre os genes funcionalmente ativos no transcriptoma de *C. eugenoides*. A tecnologia de sequenciamento de RNAs (RNA-Seq) tem sido muito utilizada em trabalhos de identificação e anotação funcional de genes para diversas espécies de plantas, assim como para verificar os níveis de atividade transcricional e identificação de genes diferencialmente expressos em determinadas condições e/ou tecidos específicos. O presente projeto teve como objetivo realizar a montagem, anotação funcional dos transcritos de folhas e frutos de *C. eugenoides* utilizando o seu genoma como referência para o alinhamento das sequências, e comparar os resultados obtidos com os disponíveis até o presente momento em bancos de dados públicos. Foram utilizadas duas bibliotecas de RNA-Seq (folhas e frutos) de *C. eugenoides* sequenciadas via plataforma Illumina HiSeq2000. Primeiramente foi realizado o alinhamento das bibliotecas contra o genoma de referência utilizando o software HISAT2. Para montagem dos *reads* alinhado em transcritos foi utilizado o software StringTie e na sequência o software Kallisto para a quantificação dos transcritos. O software DESeq2 foi utilizado para identificação de genes diferencialmente expressos em frutos e folhas. Para anotação utilizamos o BlastX contra o banco de dados de sequências não redundantes (NCBI-nr). Foi realizada uma análise comparativa com a ferramenta BlastN contra o banco de dados de sequências expressas (EST) de *C. arabica* e *C. canephora* pré-existentes. O resultado da montagem de transcritos identificou 16.743 *contigs* únicos para *C. eugenoides*, estas sequências apresentaram similaridade de 82,06% com ESTs de *C. arabica*, 91,38% em CDS de *C. canephora*, 94,87% com o RNA-Seq de *C. arabica* e 98,07% com o RNA-Seq de *C. eugenoides*. Em relação ao trabalho original, foram identificados 322 novos transcritos, das quais 36 não foram descritos nas bases de dados da *Coffea spp*. Além disso, foram identificados 923 genes diferencialmente expressos em folhas e frutos, sendo 414 mais expressos em folhas e 509 mais expressos em frutos. Seis genes diferencialmente expressos em frutos tiveram sua atividade confirmada por RT-qPCR. Três deles tiveram expressão exclusiva em frutos, apresentando potencial biotecnológico para a produção de compostos específicos ou para identificação e análise de promotores.

**Palavras-chaves:** Café. RNA-Seq. Anotação funcional.

DE BRITO, Danilo Ribeiro. **Assembly the Transcript of *Coffea eugenioides* and Identification of Differentially Expressed Genes in Leaves and Fruits**. 2019. 90 pp. Dissertation (Master's in Genetics and Molecular Biology) – Universidade Estadual de Londrina ,Londrina, 2019.

## ABSTRACT

Coffee is one of the main agricultural commodities in Brazil, the largest producer and exporter of coffee worldwide. Among 124 coffee species, *Coffea arabica*, *Coffea canephora* and *Coffea eugenioides* are the most important and the first two represent the major part of coffee production. With the exception of *C. arabica*, all other species of the genus are diploid. *C. arabica* is an allotetraploid, formed from a natural hybridization between *C. canephora* and *C. eugenioides*. The parental *C. eugenioides* is still poorly studied, but many genetic characteristics of commercial Arabica coffees are inherited from this species. Thus, it is important to increase knowledge about the functionally active genes in the transcriptome of *C. eugenioides*. Sequencing of RNAs (RNA-Seq) has been widely used in the identification and functional annotation of genes for several plant species, as well as to verify the levels of transcriptional activity and identification of genes differentially expressed under certain conditions and / or specific tissues. The present project had the objective of to assemble and perform functional annotation of the transcriptome genes of leaves and fruits of *C. eugenioides* using its genome as a reference for sequences alignment. Also to compare our results with the public data banks available. Two sequential RNA-Seq (leaf and fruit) libraries of *C. eugenioides* were used via the Illumina HiSeq2000 platform. First, the alignment of the libraries against the reference genome was performed using the HISAT2 software. For assembly of the transcripts, the software StringTie was used and in sequence the Kallisto software for the quantification of the transcripts. DESeq2 was used to identify genes differentially expressed in fruits and leaves. For annotation we used BlastX against the non-redundant sequence database (NCBI-nr). A comparative analysis was performed with the BlastN tool against the preexisting *C. arabica* and *C. canephora* Expressed Sequence Database (EST). The result of the assembly of transcripts identified 16,743 unique contigs for the species, these sequences presented similarity of 82.06% with ESTs of *C. arabica*, 91.38% in CDS of *C. canephora*, 94.87% with RNA-Seq of *C. arabica* and 98.07% with *C. eugenioides* RNA-Seq. In relation to the original work, 322 new transcripts were identified, 36 of which were not described in the *Coffea spp* databases. In addition, 923 differentially expressed genes in leaves and fruits were identified, 414 being more expressed in leaves and 509 being more expressed in fruits. Six differentially expressed genes in fruits had their activity confirmed by RT-qPCR. Three of these had exclusive expression in fruits, presenting biotechnological potential for the production of specific compounds or for identification and analysis of promoters.

**Keywords:** Coffe. RNA-Seq. transcripts.

## LISTA DE FIGURAS

<b>Figura 1</b> – Principais países importadores de cafés brasileiros em 2018. ....	16
<b>Figura 2</b> – Produção bienal de café Arabica e Conilon entre os anos de 2003 e 2019 (Previsão). ....	17
<b>Figura 3</b> – Distribuição dos <i>contigs</i> em relação ao seu tamanho. ....	33
<b>Figura 4</b> – Distribuição de termos GO encontrados no transcriptoma de <i>C. eugenioides</i> nas categorias processo biológico, função molecular e componente celular. ....	36
<b>Figura 5</b> – Os termos GO mais abundantes para cada categoria. ....	37
<b>Figura 6</b> – Caracterização dos vinte domínios/famílias de proteínas conservadas mais abundantes no transcriptoma de <i>C. eugenioides</i> . ....	38
<b>Figura 7</b> – Principais categorias de vias metabólicas obtidas pelo KEGG. ....	39
<b>Figura 8</b> – Resultado da análise de quantificação relativa de transcritos via RT-qPCR para 6 genes de <i>C. eugenioides</i> . ....	42

## LISTA DE TABELAS

<b>Tabela 1</b> – Informações sobre os <i>primers</i> dos seis genes para validação do perfil de atividade transcricional observado <i>in silico</i> via RT-qPCR. ....	30
<b>Tabela 2</b> – Resultado do alinhamento das bibliotecas contra o genoma de referência. ....	32
<b>Tabela 3</b> – Total de contigs montados pelo software StringTie individualmente para cada biblioteca e o total após o Merge. ....	33
<b>Tabela 4</b> – Análise de similaridade comparativa entre os unigenes de <i>C. eugenioides</i> e unigenes de trabalhos anteriores de <i>Coffea sp.</i> ....	35
<b>Tabela 5</b> – Caracterização dos genes candidatos escolhidos para validação do perfil de expressão diferencial entre folhas e frutos de cafeeiros via RT-qPCR. ....	41

## LISTA DE ABREVIATURAS E SIGLAS

BLAST	<i>Basic Local Alignment Search Tool</i>
Des	Diferencialmente Expressos
DNA	<i>Deoxyribonucleic Acid</i> – Ácido Desoxirribonucleico
EMBL-EBI	<i>The European Bioinformatics Institute</i>
GO	<i>Gene Ontology</i>
KEGG	<i>Kyoto Encyclopedia of Genes and Genomes</i>
NCBI	<i>National Center for Biotechnology Information</i>
P450	Citocromos P450
RNA	<i>Ribonucleic Acid</i> – Ácido Ribonucleico
RNA-Seq	Sequenciamento de alto rendimento de RNAs
RT-qPCR	PCR quantitativo em Tempo Real
TPM	<i>Transcripts Per Kilobase Million</i>

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>14</b>
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	<b>15</b>
2.1	A HISTÓRIA DO CAFÉ	15
2.2	A IMPORTÂNCIA ECONÔMICA DO CAFÉ	16
2.3	AS CARACTERÍSTICAS DO CAFÉ	18
2.4	O SEQUENCIAMENTO DE MRNA PARA A DESCOBERTA DE GENES	21
<b>3</b>	<b>OBJETIVOS</b>	<b>26</b>
3.1	OBJETIVO GERAL	26
3.2	OBJETIVOS ESPECÍFICOS	26
<b>4</b>	<b>MATERIAL E MÉTODOS</b>	<b>27</b>
4.1	OBTENÇÃO E PROCESSAMENTO DOS DADOS DE RNA-SEQ	27
4.2	ALINHAMENTO DAS SEQUÊNCIAS DE RNA-SEQ NO GENOMA DE REFERÊNCIA	27
4.3	MONTAGEM DO TRANSCRIPTOMA E QUANTIFICAÇÃO DE SEQUÊNCIAS ALINHADAS POR TRANSCRITO	28
4.4	ANOTAÇÃO E CLASSIFICAÇÃO DOS TRANSCRITOS	28
4.5	IDENTIFICAÇÃO DOS GENES DIFERENCIALMENTE EXPRESSOS	29
4.6	VALIDAÇÃO DA EXPRESSÃO OBSERVADA <i>IN SILICO</i> VIA RT-qPCR	30
<b>5</b>	<b>RESULTADOS E DISCUSSÃO</b>	<b>32</b>
5.1	ALINHAMENTO CONTRA O GENOMA DE REFERÊNCIA E MONTAGEM	32
5.2	ANOTAÇÃO E CARACTERIZAÇÃO FUNCIONAL DOS UNIGENES DE <i>C. EUGENIOIDES</i>	34
5.3	IDENTIFICAÇÃO DOS GENES DIFERENCIALMENTE EXPRESSOS	39
5.4	SELEÇÃO DOS GENES CANDIDATOS PARA A VIA DE BIOSÍNTESE DE DITERPENOS	40
5.5	VALIDAÇÃO DOS GENES DIFERENCIALMENTE EXPRESSOS POR RT-qPCR	42
<b>6</b>	<b>CONSIDERAÇÕES FINAIS</b>	<b>45</b>
	REFERÊNCIAS	46
	ANEXOS	55

## 1 INTRODUÇÃO

A cafeicultura brasileira apresenta números expressivos que traduzem a grande importância econômica e social que a atividade representa para o país (ABIC, 2019). O Brasil é o maior produtor e exportador de café do mundo e o segundo maior consumidor (CECAFE, 2018). O aumento do consumo da bebida do café nos últimos anos tem sido associado, dentre outros fatores, sua maior qualidade aliada a bons preços de mercado, assim como às suas propriedades funcionais e nutracêuticas (ABIC, 2019).

Em função de sua importância, é preciso aumentar o conhecimento sobre as bases genéticas que influenciam a composição bioquímica dos cafés. A maioria dos estudos genéticos de cafeeiros possuem foco na espécie de maior importância econômica, a alotetraplóide *Coffea arabica* (CARDOSO *et al.*, 2014; MOFFATO *et al.*, 2016; IVAMOTO *et al.*, 2017a), alguns no seu parental diplóide *Coffea canephora* (DENOUEDE *et al.*, 2014; DEREPPER *et al.*, 2014) e apenas um trabalho com dados de RNA-Seq para seu outro parental diplóide *Coffea eugenioides* (YUYAMA *et al.*, 2015).

Desta forma, é preciso aumentar o conhecimento sobre a espécie *C. eugenioides*, como por exemplo, identificando e caracterizando os genes diferencialmente expressos através de estudos de transcriptômica. Uma das formas de se alcançar este objetivo, é realizar uma análise do transcriptoma em larga escala através do sequenciamento de RNAs mensageiros (RNAm), também conhecido como RNA-Seq. Essa técnica tem proporcionado resultados relevantes com relação à identificação, anotação e caracterização de genes de espécies de plantas (WANG *et al.*, 2009).

A presente pesquisa teve como objetivo, fazer uma comparação dos resultados de uma montagem baseada no genoma de referência com a montagem *de novo* disponível até o momento (YUYAMA *et al.*, 2015) e aumentar o conhecimento básico sobre a espécie *C. eugenioides* a partir da identificação de genes diferencialmente expressos em folhas e, especialmente em frutos. Esta informação será utilizada para identificar e selecionar genes candidatos para as vias metabólicas de interesse relacionadas à composição bioquímica de grãos de café e servirá como base para novos projetos de melhoramento genético de cafeeiros visando o aumento da qualidade da bebida.

## 2 FUNDAMENTAÇÃO TEÓRICA

### 2.1 A HISTÓRIA DO CAFÉ

A descoberta do café é envolta por lendas, dentre elas, a mais aceita e divulgada é a do pastor Kaldi que viveu na Absínia, região onde hoje é a Etiópia, cerca de mil anos passados (NEVES, 1974). Segundo a lenda, Kaldi, observando suas cabras, notou que elas ficavam mais alegres e com energia renovada sempre que mastigavam alguns frutos de coloração amarelo avermelhada dos arbustos que existiam nos campos onde elas pastavam. Com isto, o pastor resolveu experimentar os frutos em forma de infusão e percebeu que a bebida de fato apresentava características estimulantes e energéticas, logo a descoberta se espalhou e o consumo da bebida se difundiu (NEVES, 1974).

A planta de café, originária da Etiópia (LASHERMES *et al.*, 1999), foi propagada culturalmente pela Arábia e seu nome é derivado da palavra árabe “qahwa”, que significa vinho. O café tornou-se de extrema importância para os árabes, que detinham o controle sobre o cultivo e preparação da bebida protegendo as mudas de café com a própria vida (ABIC, 2019). Por volta do ano de 1615, o café começou a ser saboreado na Europa oriundo de viajantes que traziam do oriente (ABIC, 2019). Contudo, somente os árabes sabiam produzir o café, tendo os alemães, franceses e italianos tentado desenvolver o plantio em suas colônias. Porém, foram os holandeses, em suas colônias na Indonésia, que conseguiram realizar o cultivo e contribuíram para a disseminação do consumo da bebida no velho continente (ABIC, 2019).

No Brasil, o café chegou ao norte do país, mais precisamente em Belém-PA, em 1727, trazido da Guiana Francesa. Devido às condições climáticas favoráveis, logo seu cultivo se espalhou pelo Brasil, passando de uma produção voltada para o mercado doméstico para a posição de produto base da economia brasileira (CECAFE, 2018).

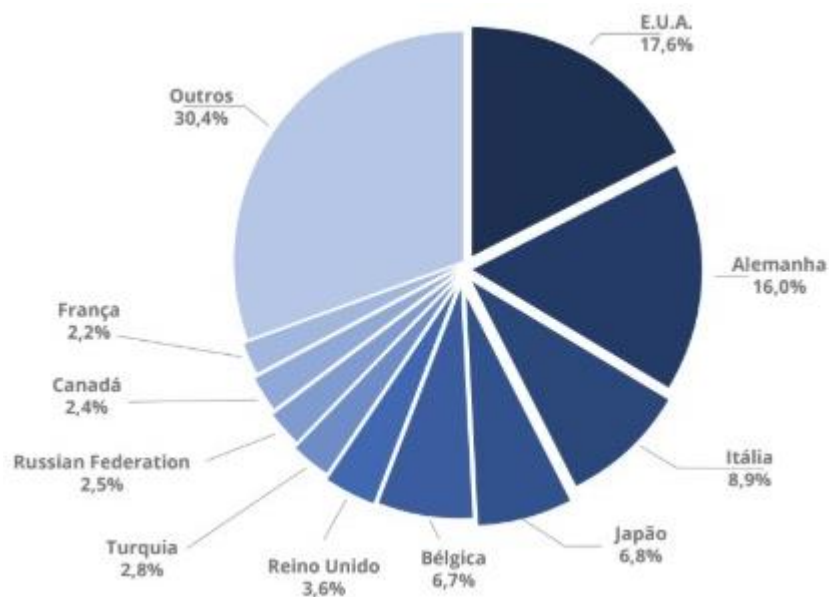
A produção de café estabeleceu-se como fonte de riqueza brasileira por quase um século, e as divisas proporcionadas pela economia do cultivo do café contribuíram para o desenvolvimento do Brasil (ABIC, 2018).

## 2.2 A IMPORTÂNCIA ECONÔMICA DO CAFÉ

O Brasil é o maior produtor e exportador mundial de café, responsável por 35% do mercado internacional (ICO, 2019). O país também possui o segundo maior mercado consumidor, ficando apenas atrás dos EUA (ABIC, 2019). Juntamente com o Brasil, outros três países possuem grandes parcelas na produção mundial: Vietnã, Colômbia e Indonésia com 18%, 9% e 6% respectivamente (CONAB, 2019; ICO, 2019).

O café figura entre os dez principais setores exportadores do Brasil, situado na 5º posição. Em 2018, um total de 35,2 milhões de sacas de café foram exportadas, o que já era previsto pelo mercado, esses valores representam um aumento de 13,9% em relação ao ano anterior, que registrou 30,9 milhões de sacas (CECAFE, 2019). Essas exportações geraram uma receita cambial de US\$ 5,1 bilhões, correspondendo a 5,4% de participação no total de exportações do agronegócio brasileiro (CONAB, 2019). Os principais destinos do café brasileiro em 2018 foram os EUA, seguido pela Alemanha, Itália, Japão e Bélgica (Figura 1) (CECAFE, 2019).

**Figura 1** – Principais países importadores de cafés brasileiros em 2018.

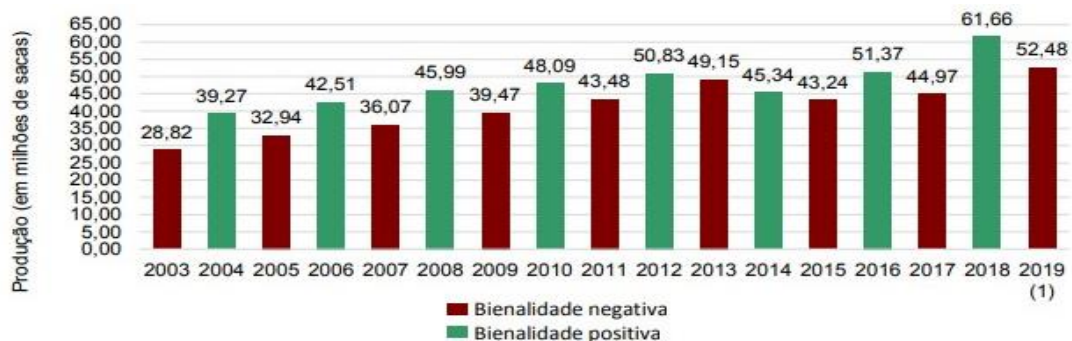


Fonte: CECAFE, 2018.

Além disso, os cafés do Brasil ocupam o quinto lugar no ranking de faturamento de lavouras. O faturamento bruto total das lavouras brasileiras em 2018 atingiu o montante de R\$ 383,87 bilhões, o qual foi calculado com base no volume produzido e nos preços médios recebidos pelos agricultores dos 21 principais produtos agrícolas. Desta forma, o ranking dos cinco produtos que apresentaram maior faturamento ficou com soja, em primeiro lugar, com R\$ 142,36 bilhões, que equivale a 37% do faturamento total; em segundo, cana-de-açúcar – R\$ 61,08 bilhões (16%); milho, em terceiro – R\$ 47,12 (12%). Na sequência, em quarto lugar, algodão herbáceo, com R\$ 33,99 bilhões (9%); e, por fim, o café, que figura em quinto lugar nesse ranking, com R\$ 24,92 bilhões, cujo faturamento representa 6,5% do total (MAPA, 2019).

Em 2018 o Brasil teve uma produção de 61,66 milhões de sacas, o que representou em um aumento de 37% em relação à safra de 2017 (CONAB, 2019) (Figura 2). A previsão da área total cultivada para 2019 é de 2,16 milhões de hectares (316,6 mil ha em formação e 1,84 mi ha em produção), com produção estimada de 50,48 a 54,48 milhões de sacas de café beneficiado, representando uma redução de 18,1% a 11% em relação à safra passada (CONAB, 2019) (Figura 2). O estado de Minas Gerais é o maior produtor nacional com aproximadamente 60% do total produzido no país, seguido por Espírito Santo (17%) e São Paulo (11%). O café Arábica (*C. arabica*) representa 77% do total produzido nacionalmente e Conilon (*C. canephora*) participa com 23% da produção nacional, tendo como maior produtor o estado de Espírito Santo com aproximadamente 63% da produção em 2018 (CONAB, 2019; ABIC, 2019).

**Figura 2** – Produção de café Arábica e Conilon entre os anos de 2003 e 2019 (Previsão).



Fonte: CONAB, 2019.

Além da importância econômica e cultural, o café ainda possui uma grande importância social, pois a maior parte da sua produção é feita no sistema de agricultura familiar gerando renda para milhares de empregados diretos e indiretos (VEZZALI, 2006; FAO, 2017).

O consumo de café segue em crescimento, tanto no Brasil quanto no exterior, desde o tradicional café torrado e moído até as cápsulas e bebidas geladas, proporcionando um mercado com cenário promissor e encorajador para o investimento das indústrias em inovação (BICC, 2017). No Brasil, o consumo de café em 2018 teve um crescimento de 4% (22,9 milhões de sacas de 60 kg) em relação ao ano anterior (ABIC, 2018). Ainda segundo a ABIC, o consumo deve continuar crescendo em 2019, mais 4%, acima da média mundial, impulsionando o consumo brasileiro a alcançar os EUA talvez, nos próximos anos.

### 2.3 AS CARACTERÍSTICAS DO CAFÉ

O cafeeiro é uma planta pertencente à família Rubiaceae e ao gênero *Coffea* (GUERREIRO FILHO *et al.*, 2008). Existem 124 espécies catalogadas, incluindo espécies do gênero *Psilanthus sp* (DAVIS *et al.*, 2011) que, atualmente, após análises de morfologia e genéticas, e também devido à baixa diversidade genética, encontra-se no mesmo gênero de *Coffea* (LOMBELLO; PINTO-MAGLIO, 2003, 2004; MAURIN *et al.*, 2007; DAVIS *et al.*, 2011).

*C. arabica* é uma espécie autógama, alotetraplóide ( $2n = 4x = 44$ ), resultado da combinação de dois conjuntos de genomas distintos que se formou a partir de uma hibridação natural seguida de duplicação (LASHERMES *et al.*, 1999). Sua origem é recente, estimada em 100 e 500 mil anos (ANTHONY, 2010). Contudo, alguns autores acreditam que a sua origem é mais recente, entre 10 e 50 mil anos (CENCI *et al.*, 2012). A hibridação que resultou em sua origem ocorreu entre duas espécies diploides e alógamas, *C. canephora* e *C. eugenioides* (LASHERMES *et al.*, 1999).

A espécie *C. arabica* cresce em ambientes com variações de amplitudes térmicas médias de 20 a 24°C e, preferencialmente, cultivadas em altitudes acima de 800 metros (VIDAL, 2010). Sua bebida é considerada de maior qualidade do que o café proveniente de plantas de *C. canephora*, pois possui um aroma intenso,

sabores diversificados e inúmeras variações de corpo e acidez (CLARKE; MACRAE, 1985). Essas características sensoriais explicam o seu valor mais elevado no mercado (CLARKE; MACRAE, 1985; DAMATTA; RAMALHO, 2006; LEROY *et al.*, 2006; CAMPANHA, 2008). Porém ele é suscetível à doenças como a ferrugem e pragas como o bicho mineiro (CONCEIÇÃO; GUERREIRO-FILHO; GONÇALVES, 2005, VAN DER VOSSSEN; BERTRAND; CHARRIER, 2015).

Os cultivares de *C. arabica* mais utilizados na produção são: Caturra, Catuaí e Mundo Novo. Estes foram selecionados de duas populações tipicamente conhecidas como Típica e Bourbon (ANTHONY *et al.*, 2010). Caturra é um cultivar anão e mutante do grupo Bourbon, já Mundo Novo é resultado da hibridação entre Bourbon e Típica. Catuaí, por sua vez, é proveniente de um cruzamento entre Caturra e Mundo Novo (VIDAL, 2010).

*C. canephora* é uma espécie diploide ( $2n = 2x = 22$ ) com maior vigor e maior diversidade genética, adaptada à regiões de baixas altitudes com temperaturas médias em torno de 22 e 26°C (DAMATTA; RAMALHO, 2006). Possui como características a sua alta produtividade, resistência às secas prolongadas, tolerância à pragas e também por conter um alto teor de cafeína (LEROY *et al.*, 2005). Sua bebida é considerada inferior à do *C. arabica*, com menor acidez e mais encorpada, sendo muito utilizada na produção de café solúvel e em *blends* com *C. arabica* para o produto torrado e moído (CLARKE; MACRAE, 1985; DAMATTA; RAMALHO, 2006; SOUZA; BENASSI, 2012; FAO, 2018).

Por sua vez, o outro ancestral de *C. arabica*, o *C. eugenoides*, é uma espécie silvestre também diploide que cresce em regiões com altitudes elevadas próximas as bordas de florestas, em temperatura média de 18 a 23°C. Sua produção não é em escala comercial devido à baixa produtividade e do pequeno tamanho das sementes, que têm com baixo teor de cafeína (MAURIN *et al.*, 2007). Medina-Filho *et al.* (2007), em sua revisão acerca do melhoramento genético e a biodiversidade do café, propõe que a melhor qualidade da bebida do *C. arabica* está relacionada a seu ancestral materno *C. eugenoides*, pois este, em sua bebida, apresenta atributos superiores se comparados ao *C. canephora*, resultando em um licor fraco, suave e limpo, com um sabor muito mais agradável, baixa adstringência e um aroma floral.

A composição química do café é um dos fatores mais determinantes em relação à qualidade da bebida (SOUZA; BENASSI, 2012; FAO, 2017). Seu aroma e sabor são resultados da combinação de vários constituintes químicos voláteis e não

voláteis. Como principais compostos químicos que podem influenciar a qualidade do produto final, pode-se destacar os açúcares, cafeína, trigonelina, ácidos clorogênicos e lipídeos (FARAH *et al.*, 2006; KIM *et al.*, 2009).

O consumo de cafés está aumentando muito com o passar dos anos, e uma das principais razões é o aumento da qualidade da bebida via seleção assistida por marcadores, com o intuito de realizar melhoramento genético de cafeeiros (FARAH, 2012), além da expansão de mercado e criação de novas categorias do produto, como as capsulas e os cafés Gourmet ou Especiais (ABIC, 2019; BSCA, 2018). Esta relação entre aumento do consumo do café/melhor qualidade da bebida, demonstra a importância de aumentar os estudos sobre a espécie *C. eugenioies*, já que esta possui uma bebida com um sabor mais agradável e por ser um parental do *C. arabica*.

Outro fator para explicar o aumento do consumo de café é a relação do consumo moderado da bebida com o benefício na saúde humana. Estudos apontam que essa relação pode reduzir o risco de desenvolvimento de doenças do coração, Parkinson, diabetes e alguns tipos de cânceres (ALVES; CASAL; OLIVEIRA, 2009; BØHN *et al.*, 2012; O'KEEFE *et al.*, 2013; DI LORENZO *et al.*, 2017). Uma pesquisa realizada por Guercio *et al.* 2015, mostrou que a ingestão de 4 xícaras ou mais de café por dia reduziu, em pacientes com câncer de cólon em estágio III, a probabilidade de recorrência da doença e da morte dos pacientes. O café também possui propriedades anti-inflamatórias e antioxidantes que auxiliam na proteção do fígado contra o desenvolvimento de doenças hepáticas (ALVES; CASAL; OLIVEIRA, 2009).

Existem várias tecnologias desenvolvidas para analisar o transcriptoma e o funcionamento dos genes em plantas e em outros organismos, contribuindo assim para entender o funcionamento de determinado gene e sua consequente resposta no fenótipo do organismo estudado. Para que desta forma, no caso de plantas, possam ser realizados programas de melhoramento genético ou de transgenia, produzindo assim, por exemplo, cultivares mais produtivas ou resistentes à pragas e estresses. Como exemplo dessas tecnologias temos os *microarrays*, que foram amplamente utilizados para a análise de transcriptomas e descoberta de genes de forma eficaz, embora com algumas limitações (MORTAZAVI *et al.*, 2008).

Todavia, foram os sequenciamentos em larga escala de RNAs (via RNA-Seq) que revolucionaram os estudos de transcriptômica e proporcionaram avanços

significativos em estudos de várias espécies, inclusive espécies anteriormente pouco estudadas (WANG *et al.*, 2009; FU *et al.*, 2018).

#### **2.4 O SEQUENCIAMENTO DE MRNA PARA A DESCOBERTA DE GENES**

O sequenciamento de alta capacidade de DNA/RNA foi um grande avanço tecnológico para o estudo de genética e biologia molecular, possibilitando identificação e caracterização de genes. Até poucas décadas atrás, o método de sequenciamento de nucleotídeos disponível mais utilizado era aquele desenvolvido pelo cientista inglês Frederick Sanger (SANGER *et al.*, 1977). Conhecido como método de sequenciamento de Sanger ou método de terminação de cadeia, este serviu de base para o advento da genômica (KOLODNER, 1979; WANG; MESSING, 2011).

No sequenciamento de Sanger, são sequenciadas regiões do DNA com tamanho de até 900 pares de bases sem grande perda de qualidade, e o processamento desses dados eram realizados em períodos de semanas e meses, em um exausto trabalho em laboratório (MARTINS, 2014). Contudo, no início dos anos 2000, as automatizações das tecnologias de sequenciamento foram sendo aperfeiçoadas até o ponto em que equipamentos totalmente automatizados conseguiram gerar sequências em apenas 24 horas (MARTINS, 2014). Com estes avanços, um projeto de sequenciamento de bactérias, por exemplo, contendo 20 mil a 200 mil *reads*, podiam ser montados em um computador, porém, projetos maiores, como o genoma humano com cerca de 3 bilhões de bases, necessitavam de uma estrutura muito grande de recursos computacionais, envolvendo vários laboratórios e centros de pesquisa científica (MARTINS, 2014).

Esta época ficou caracterizada pelos grandes avanços técnicos, tendo como ponto principal a automatização dos equipamentos de sequenciamento e de análise de segmentos de DNA, permitindo o sequenciamento de vários genomas completos de diversos organismos (BORTIRI *et al.*, 2008). Ainda segundo Bortiri *et al.* (2008), apesar da larga utilização e do sucesso obtido pelo método de Sanger, a necessidade de uma metodologia de sequenciamento com um custo mais baixo, com maior rapidez e eficiência se tornou evidente. Assim, novas tecnologias de sequenciamento de DNA e RNA acabaram sendo desenvolvidas.

As técnicas de sequenciamento têm evoluído e vem apresentando a cada ano um menor custo e, o desenvolvimento de ferramentas computacionais para analisar o grande volume de informações geradas, também tem crescido. Isto tem proporcionando um ambiente favorável que auxilia no aumento do conhecimento genômico, possibilitando o sequenciamento de genomas de organismos ainda pouco estudados (RAI; SHEKHAWAT, 2014), como mamão (MING *et al.*, 2008), mandioca (BREDESON *et al.*, 2016) e pêssego (VERDE *et al.*, 2017), além de importantes commodities como soja (SCHMUTZ *et al.*, 2010), milho (SCHNABLE *et al.*, 2009), trigo (IWGSC, 2014) e algodão (PATERSON *et al.*, 2012).

Ao longo do tempo, diferentes metodologias foram descritas para realizar a identificação de novos transcritos em espécies de plantas não modelo (RAI; SHEKHAWAT, 2014), e atualmente uma das mais utilizadas é a de sequenciamento em larga escala de RNAs (RNA-Seq) (TJADEN, 2015).

O RNA-Seq é uma abordagem relativamente nova, desenvolvida para perfis de transcriptomas em que são utilizadas tecnologias de sequenciamento profundo (*deep-sequencing*) (WANG *et al.*, 2009; ZHANG *et al.*, 2015). O RNA-Seq proporciona uma medida mais precisa dos níveis de transcritos e de suas isoformas do que outros métodos existentes (WANG, 2009; ZHANG *et al.*, 2015). Surgiu como uma ferramenta poderosa para estudar transcriptomas, com o objetivo de fornecer uma lista abrangente de todos os transcritos e seus níveis de expressão de uma célula ou população de células sob determinada condição (MARTIN *et al.*, 2010).

Essa abordagem funciona basicamente da seguinte forma: certa quantidade de RNA é convertida em uma biblioteca contendo fragmentos de cDNA, em seguida esses fragmentos recebem adaptadores (bases de DNA) e assim passam por sequenciamento, gerando uma sequência curta, em torno de 30 a 500 pb, dependendo da tecnologia utilizada (CONESA *et al.*, 2016; HRDLICKOVA; TOLOUE; TIAN, 2017). Logo após, esses *reads* curtos são alinhados a um genoma de referência, ou alinhados contra um transcriptoma previamente anotado, quando disponível, e depois utiliza o resultado do alinhamento para montar, através de um gráfico de sobreposições dos *reads*, os possíveis transcritos (GARBER *et al.*, 2011; ZHAO *et al.*, 2011; FLOREA; SALZBERG, 2013; CONESA *et al.*, 2016). Todavia, quando não se têm disponíveis genoma ou transcriptoma de referência, os *reads* são remontados a fim de criar um mapa em escala genômica composto pela estrutura transcricional ou o nível de expressão de cada gene individualmente,

método chamado de montagem *de novo* (WANG *et al.*, 2009; GRABHERR *et al.*, 2011; SCHULZ *et al.*, 2012; FU *et al.*, 2018).

As duas estratégias apresentam suas vantagens, porém, é possível destacar na montagem guiada pelo genoma de referência: i) os erros de sequenciamento podem ser resolvidos pelo mapeamento dos *reads* contra o genoma; ii) os *reads* alinhados proximais podem ser agrupados em *exons* e os locais de *splice* podem ajudar a estender os limites do *exon*; iii) a montagem guiada pelo genoma é bastante sensível a transcritos pouco expressos, uma vez que eles podem ser mapeados para o genoma de referência (LU; ZENG; SHI, 2013). Já, para a montagem *de novo*, tem-se destaque: i) com uma cobertura de *read* suficiente, a montagem de novo pode montar eficientemente um transcriptoma completo; ii) é possível a identificação de transcritos com *introns* longos e detectar eventos especiais como *trans-splicing*; iii) pode ser utilizada em qualquer organismo independente de genoma de referência (LU; ZENG; SHI, 2013).

A melhor estratégia dependerá da disponibilidade de um genoma para o organismo que está sendo estudado. Alguns autores constatam em teste que a montagem guiada por referência se sobressai ao método *de novo* (ZHAO *et al.*, 2011; JAIN; KRISHNAN; PANDA, 2013; BOLEY *et al.*, 2014) por apresentarem transcritos mais completos, porém uma abordagem unindo os dois métodos apresentam os melhores resultados (GARBER *et al.*, 2011; JAIN; KRISHNAN; PANDA, 2013; LU; ZENG; SHI, 2013)

Sendo assim, o RNA-Seq tem sido amplamente aplicado tanto em organismos modelo bem estudados como em organismos não-modelo, para fornecer informações sobre o perfil de transcrição destes e para fornecer pontos de vista importantes em processos biológicos (ZHAO *et al.*, 2011; YANG; SMITH, 2013).

Existem várias tecnologias desenvolvidas para a dedução de transcritos e quantificação de transcriptomas, incluindo técnicas de hibridização e baseadas em sequências (CLARK, 2002). O RNA-Seq oferece várias vantagens em relação a essas tecnologias, incluindo os *microarrays*. Por exemplo, ele apresenta uma maior sensibilidade para transcritos com baixa abundância (CHENG; WANG; SHI, 2011; ZHAO *et al.*, 2011; HASS *et al.*, 2012). Além disso, ao contrário das abordagens baseadas em hibridização, o RNA-Seq não está limitado à detecção de transcritos correspondentes às sequências genômicas preexistentes, ou seja, para aquelas que possuem genomas de referência (WANG *et al.*, 2009). Neste último caso, como visto

anteriormente, quando a sequência genômica de referência não está disponível, é possível a construção de uma montagem *de novo* a partir de RNA-Seq (ZHAO *et al.*, 2011).

O RNA-Seq também mostrou ser altamente preciso para quantificar os níveis de expressão, bem como altos níveis de reprodutibilidade, tanto para réplicas técnicas quanto para as biológicas (ZHAO *et al.*, 2011). Ademais, o RNA-Seq requer uma quantidade muito menor de amostras de RNA para as detecções de variações nas sequências genômicas das regiões transcritas (DA SILVA, 2012). Levando em conta todas essas vantagens, o RNA-Seq é o primeiro método baseado em sequenciamento que permite que todo o transcriptoma seja pesquisado de maneira muito ampla e quantitativa (WANG *et al.*, 2009).

Mesmo apresentando todas essas vantagens, o método RNA-Seq ainda possui alguns desafios em sua aplicação (ZHAO *et al.*, 2011; WU *et al.*, 2013). Por exemplo, um problema muito comum está relacionado com a área de informática, em questões de armazenamento, leitura e processamento das informações geradas, isso pode resultar em erros no momento da análise e, mesmo com *reads* de boa qualidade, ainda é um desafio ter precisão nas etapas de alinhamento ou montagem destes (DA SILVA, 2012). Este método deveria ser capaz de identificar todo tipo de RNA de qualquer tamanho, porém, como o RNA-Seq possui muitas etapas de manipulação durante a produção das bibliotecas de cDNA, moléculas maiores de RNA acabam tendo que ser fragmentadas para se adequarem às tecnologias de sequenciamento de nova geração, influenciando diretamente o resultado produzido (WANG *et al.*, 2009).

Apesar dos desafios, o RNA-Seq tem sido considerado por muitos pesquisadores um marco na Genética, pois sua alta resolução e sensibilidade possibilitou que várias regiões de transcrição e isoformas de *splicing* de genes já conhecidos fossem descobertas, incluindo o mapeamento de seus limites 5'-3' (WANG *et al.*, 2009). Novas regiões de transcrição também foram descobertas utilizando RNA-Seq para cada genoma que este foi empregado (WANG *et al.*, 2009). O RNA-Seq também provou ser capaz de rastrear com precisão mudanças nos padrões de expressão em diferentes estágios do desenvolvimento de alguns organismos (WANG *et al.*, 2009).

Existem vários trabalhos publicados recentemente que utilizaram essa tecnologia para estudar transcriptoma de cafeeiro (CHENG; FURTADO; HENRY,

2017; FLOREZ *et al.*, 2017; IVAMOTO *et al.*, 2017a), com foco na identificação de genes relacionados a estresses bióticos (CARDOSO *et al.*, 2014), e relacionados ao desenvolvimento de marcadores moleculares e identificação de polimorfismos de base única em cafés Arábica (HERRERA *et al.*, 2014).

Um trabalho pioneiro de transcriptoma *de novo* de *C. eugenioides* (YUYAMA *et al.*, 2015) utilizou a tecnologia de RNA-Seq e identificou 79 genes diferencialmente expressos entre folhas e frutos da espécie. Os dados de sequenciamento deste trabalho foram utilizados no presente projeto de montagem de transcriptoma com a finalidade de melhorar e aumentar o número de genes identificados e, inclusive, identificar mais genes diferencialmente expressos entre folhas e frutos, pois será utilizado o genoma de referência, que anteriormente não estava disponível, para realizar um novo alinhamento.

### 3 OBJETIVOS

#### 3.1 OBJETIVO GERAL

Realizar a montagem do transcriptoma de *Coffea eugenoides* utilizando o genoma de referência como base para o alinhamento.

#### 3.2 OBJETIVOS ESPECÍFICOS

- ✓ Comparar os resultados da nova montagem baseada no genoma de referência com a montagem *de novo* disponível até o presente momento;
- ✓ Realizar a predição de genes e verificar se existem transcritos ainda não descritos para esta espécie;
- ✓ Identificar os genes diferencialmente expressos em *C. eugenoides*;
- ✓ Validar o perfil de atividade transcricional observado *in silico* para alguns genes através da tecnologia de RT-qPCR.

## 4 MATERIAL E MÉTODOS

### 4.1 OBTENÇÃO E PROCESSAMENTO DOS DADOS DE RNA-SEQ

O material genético utilizado para gerar os dados de RNA-Seq é proveniente de folhas jovens e frutos maduros no estádio 'cereja' que foram colhidos de quatro plantas de *C. eugenoides* dos mesmos pais de seis a oito anos mantidos no COCARI, Mandaguari, PR, Brasil [latitude (S): 23° 30' 52"; longitude (W): 51° 42' 86"]. A região tem uma altura de 650 m e uma temperatura média anual de 22-23°C. Essas plantas são de geração F<sub>3</sub> a partir de acessos originais de *C. eugenoides* no Instituto Agronômico de Campinas, Campinas, SP, Brasil. Todo o processo de extração do RNA dessas plantas e o envio para sequenciamento e construção das bibliotecas foram realizados por Yuyama *et al.* (2015).

Na montagem foram utilizadas duas bibliotecas (folhas e frutos) disponibilizadas pelo Instituto Agronômico do Paraná – IAPAR, no NCBI (número de acesso PRJNA273321). Foi utilizado o genoma de referência de *C. eugenoides* (versão 1.9) gentilmente cedido pelo ACGC (*Arabica Coffee Genome Consortium*) para o alinhamento das bibliotecas de RNA-Seq. A verificação da qualidade das sequências foi realizada com o software FastQC v.0.11.5 (ANDREWS, 2010). Em seguida, foi realizada a trimagem (limpeza) com o software Trimmomatic v.36 (BOLGER; LOHSE; USADEL, 2014) para a remoção dos adaptadores, sendo selecionados para as próximas análises apenas os fragmentos com *Phred score* igual ou maior que 20 e com tamanho mínimo de 36 pares de base.

### 4.2 ALINHAMENTO DAS SEQUÊNCIAS DE RNA-SEQ NO GENOMA DE REFERÊNCIA

Para alinhamento dos *reads* contra o genoma de referência do *C. eugenoides* foi utilizado o alinhador HISAT2 versão 2.1.0 com parâmetros *default* (KIM; LANGMEAD; SALZBERG, 2015). Além de ser uma das ferramentas de alinhamento mais utilizadas atualmente, o HISAT2 é o mais indicado para esse tipo de trabalho, por ser software de alinhamento mais rápido e com precisão igual ou melhor do que os outros softwares dessa categoria, como por exemplo, TopHat2 (KIM *et al.*, 2013)

e GSNAP (WU; NACU, 2010). Além de utilizar o esquema de indexação baseado na transformação Burrows-Wheeler e no índice Ferragina Manzini (FM).

### **4.3 MONTAGEM DO TRANSCRIPTOMA E QUANTIFICAÇÃO DE SEQUÊNCIAS ALINHADAS POR TRANSCRITO**

Para montagem dos transcritos, foi utilizado o software StringTie versão 1.3.4 (PERTEA *et al.*, 2015) com seus parâmetros *default*, sendo selecionados transcritos com no mínimo 200bp de tamanho para as próximas análises. Foi utilizado o software SamTools versão 1.6 (LI *et al.*, 2009) para conversão dos arquivos de saída do alinhamento (SAM) para o formato aceito pelo StringTie (BAM).

Num primeiro momento foi realizada a montagem das bibliotecas separadamente e, após, com sua opção *-MERGE*, o StringTie cria um conjunto único de transcritos para todas as amostras, facilitando assim, as próximas análises de identificação de genes diferencialmente expressos em ambos os tecidos utilizados.

Para a quantificação da abundância de transcritos foi utilizado o software Kallisto (versão 0.44) (BLAY *et al.*, 2016).

### **4.4 ANOTAÇÃO E CLASSIFICAÇÃO DOS TRANSCRITOS**

Todos os transcritos foram anotados através da análise comparativa realizada pela ferramenta BlastX contra o banco de dados de sequências não redundantes (NCBI nr) disponíveis na plataforma do *National Center for Biotechnology Information* (NCBI). O parâmetro de corte utilizado foi de *e-value* menor ou igual a  $1e^{-5}$ . Além do NCBI-nr, foi realizada uma segunda análise comparativa (BlastX) dos unigenes de *C. eugenoides* contra o banco de dados curados de proteínas curado (Swiss-Prot - *The UniProt Consortium*, 2014).

A anotação funcional dos *contigs* montados e de seus respectivos processos biológicos, funções moleculares e componentes celulares foi realizada utilizando o software Blast2GO Pro 5 (CONESA *et al.*, 2005). O banco de dados do InterProScan (MITCHELL *et al.*, 2019) também foi utilizado para identificar domínios de proteínas

conservados e depositados no banco de dados EMBL-EBI (*The European Bioinformatics Institute*) para melhorar a anotação funcional dos genes. O banco de dados do *Kyoto Encyclopedia of Genes and Genomes* (KEGG) foi usado para identificar as vias metabólicas dos genes putativamente encontrados no transcriptoma (KENEHISA *et al.*, 2018). Estas duas últimas análises foram realizadas via de ferramentas disponíveis no software Blast2GO (parâmetros *default*).

A análise comparativa entre o transcriptoma gerado por este trabalho e os previamente publicados foi realizada utilizando a ferramenta BlastN contra os bancos de dados existentes de sequências nucleotídicas e/ou proteínas de café: i) sequências ESTs (*Expressed Sequence Tags*) de *C. arabica* (MONDEGO *et al.*, 2011); ii) sequências codantes (CDS) de *C. canephora* (DENOEUDE *et al.*, 2014); iii) sequências expressas de *C. eugenioides* por (YUYAMA *et al.*, 2015) e iv) sequências expressas de *C. arabica* utilizando a tecnologia Illumina (IVAMOTO *et al.*, 2017a).

#### 4.5 IDENTIFICAÇÃO DOS GENES DIFERENCIALMENTE EXPRESSOS

A análise de expressão diferencial gênica entre as bibliotecas (folha e fruto) foi realizada pelo software DESeq2 (LOVE; HUBER; ANDERS, 2014), cuja finalidade foi quantificar o número total de sequências alinhadas em cada transcrito gerado pelo software Kallisto. Os parâmetros utilizados foram  $p\text{-value} < 0.05$  em combinação com  $\text{Log}_2\text{FC} \geq 1,58$  para transcritos positivamente (*up*) regulados e  $\leq -1,58$  para transcritos negativamente (*down*) regulados. Após a identificação, foram verificadas as anotações nos bancos NCBI-nr e Swiss-Prot e também a análise GO (*Gene Ontology*) via Blast2GO dos genes identificados como diferencialmente expressos. O software TransDecoder (<https://transdecoder.github.io/>) foi utilizado para identificar os unigenes que possuem sequências completas para a codificação de proteínas (*full-lengths*).

#### 4.6 VALIDAÇÃO DA EXPRESSÃO OBSERVADA *IN SILICO* VIA RT-qPCR

Pares de *primers* foram desenhados utilizando o software CLC Genomics Workbench v. 9.5.3.27 (<https://www.qiagenbioinformatics.com/>) e os *amplicons* variaram de 100 a 132 pb com Tm de 60°C (Tabela 1).

**Tabela 1** – Informações sobre os *primers* dos seis genes para validação do perfil de atividade transcricional observado *in silico* via RT-qPCR.

Nome do Gene	Anotação	ID Gene	N Acesso	Primer Forward	Primer Reverse
<i>CeCYP81E8</i>	Cytochrome P450	CE.2959	sp Q6WVQ8 C81E8_MEDTR	CGCAAGATGTGCAAAGAGAA	CTGCTAACGCTGGACCTAT
<i>CeCYP76AH1</i>	Ferruginol synthase	CE.12743	sp S4UX02 CYPH1_SALMI	ACTCTATTTTGTCTATCTGGGCA	GGAACCTCCCCTTCAACTTCA
<i>CeCYP734A1</i>	Cytochrome P450	CE.6367	sp O48786 C734A_ARATH	AATGACGGCAAGGTGGAA	AGCCCTTCCATCTTCATAGC
<i>CeGA20OX1</i>	Gibberellin 20 oxidase 1	CE.5367	sp Q39110 GAOX1_ARATH	AGTGAAGCCATGAGCAGT	AGGTTTTTGGCATGGAGGA
<i>CeCYP76B10</i>	Geraniol 8-hydroxylase	CE.12740	sp D1MI46 C76BA_SWEMU	CGCAAGATGTGCAAAGAGAA	CTGCTAACGCTGGACCTAT
<i>CeCYP71D55</i>	Premnaspirodien oxygenase	CE.12668	sp A6YIH8 C7D55_HYOMU	GGAACCTCCCCTTCAACTTCA	GGTGAAGCTTTTGGCC

**Fonte:** o próprio autor.

DNAs complementares (cDNAs) para todas as amostras foram sintetizados com o kit GoScript™ 3 Reverse Transcription System Kit (Promega), segundo o protocolo do fabricante, em um volume final de 20 µl e utilizando 1,50 µg de RNA total.

A expressão gênica foi analisada com o auxílio do equipamento Fast 7500 (Thermo Fisher Scientific). A reação consistiu em um volume final total de 15 µL com 7.5 µL de GoTaq® qPCR Master Mix (Promega), 0.5 µL de *primer forward* e *reverse* (10 µM), 1 µL de cDNA (50 ng/µL) e 5.5 µL de água livre de nucleases. As condições da ciclagem foram de 95°C por 20 segundos, seguido por 40 ciclos de 95°C por 30s, 60°C por 30s e 72°C por 30s. Todas as reações foram seguidas por uma análise de curva de dissociação para verificar a presença de produtos (banda) únicos de PCR. Todas as reações foram realizadas com três replicatas biológicas e três replicatas técnicas, e foram seguidas todas as recomendações do *MIQE guidelines* para experimentos de RT-qPCR (BUSTIN *et al.*, 2009).

Os dados foram analisados para determinar os valores do limiar do ciclo (Ct). A especificidade dos produtos de PCR gerados para cada conjunto de *primers* foi

verificada analisando a  $T_m$  (dissociação) dos produtos amplificados. Os níveis de expressão foram calculados aplicando a fórmula  $(1 + E) - Ct$ , onde  $\Delta C_{\text{target}} = \text{Gene } Ct_{\text{target}} - Ct_{\text{CeGAPDH}}$  e  $\Delta\Delta Ct = \Delta Ct_{\text{target}} - \Delta Ct_{\text{Amostra de referência}}$  (COTTA *et al.*, 2014). O gene com menor expressão entre os órgãos foi utilizado como amostra de referência para calcular a expressão relativa. Os níveis de expressão gênica foram normalizados utilizando o gene da gliceraldeído-3-fosfato desidrogenase (GAPDH) como referência. Este gene tem sido amplamente utilizado para calibrar a expressão em experimentos de RT-qPCR de café (CARVALHO *et al.*, 2013; YUYAMA *et al.*, 2015; IVAMOTO *et al.*, 2017b).

## 5 RESULTADOS E DISCUSSÃO

### 5.1 ALINHAMENTO CONTRA O GENOMA DE REFERÊNCIA E MONTAGEM

Do total de 3.618.185 de *reads* para a biblioteca de folhas e 4.657.475 para a biblioteca de frutos, foram alinhados com o software HISAT2 contra o genoma de referência 82,98% (3.002.213) e 79,42% (3.769.938) das sequências respectivamente (Tabela 2). Esses resultados para os alinhamentos estão de acordo com os resultados apresentados pelo protocolo de referência para este tipo de análise (PERTEA *et al.*, 2016)

**Tabela 2** – Resultado do alinhamento das bibliotecas contra o genoma de referência.

Informações do Alinhamento	Valores	
	Folhas	Frutos
<i>Reads</i> de alta qualidade	3.618.185	4.657.475
Porcentagem de <i>reads</i> alinhados 1 vez	82,98% (3.002.213)	81,14% (3.779.132)
Porcentagem de <i>reads</i> não alinhados	17,02% (615.972)	18,86% (878.343)
Porcentagem de <i>reads</i> alinhados mais de 1 vez	0% (0)	0% (0)

**Fonte:** o próprio autor.

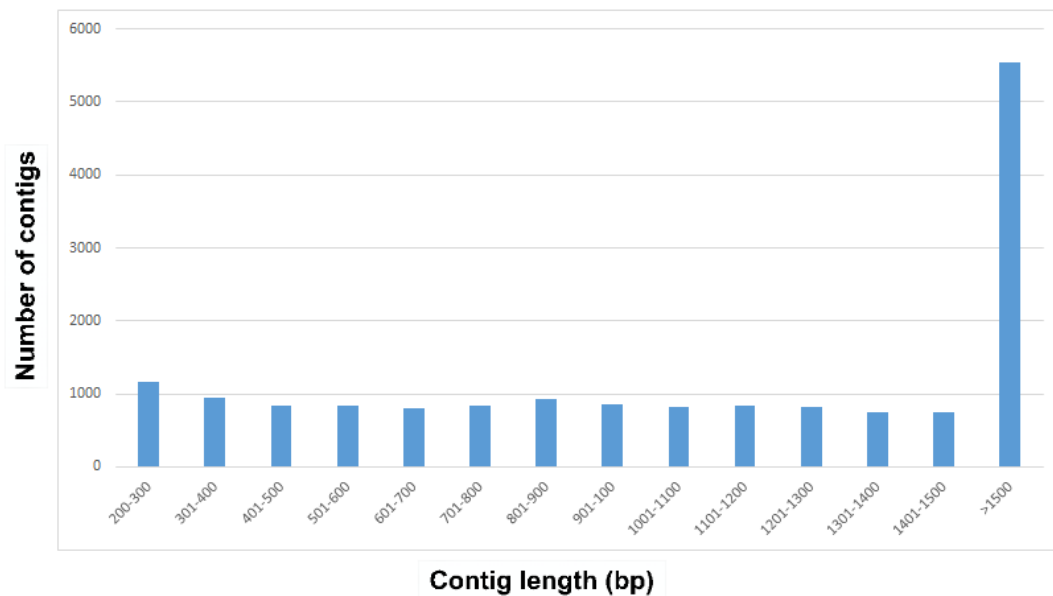
Submetidos os *reads* alinhados para a montagem dos *contigs*, obtivemos os seguintes resultados: a biblioteca de folhas produziu um total de 13.668 transcritos e a de frutos, 14.543 transcritos. Após a aplicação da opção de merge do StringTie, um total de 16.743 transcritos foram identificados com comprimento médio de 1.248 pb, deste total de transcritos, mais de 9.000 tiveram comprimentos maiores que 1.000 pb (Tabela 3). É possível observar a distribuição dos transcritos de acordo com seus tamanhos na Figura 3.

**Tabela 3** – Total de *contigs* montados pelo software StringTie individualmente para cada biblioteca e o total após o Merge.

Total de <i>contigs</i> identificados		
Folhas	Frutos	Após MERGE
13.668	14.543	16.743
Tamanho médio dos <i>contigs</i>	Tamanho do menor <i>contig</i>	Tamanho do maior <i>contig</i>
1.248 pb	200 pb	5.884 pb

Fonte: o próprio autor.

**Figura 3** – Distribuição dos *contigs* em relação ao seu tamanho.



Fonte: o próprio autor.

Os tamanhos de *contigs* variaram de 200 pares de base (pb) a 5.884 pb, com tamanho médio de 1.248 pb. Este último foi semelhante ao encontrado em um estudo de transcriptoma de *C. arabica* (1.264 pb) (IVAMOTO *et al.*, 2017a), próximo ao descrito por Venturini *et al.* 2013, em *Vitis vinífera* (1.307 pb) e maior que o encontrado em *C. eugenioides* (701 pb) em uma análise *de novo* (YUYAMA *et al.*, 2015).

Foram obtidos um total de 16.743 transcritos, esse número foi menor do que a quantidade encontrada por Yuyama *et al.* (2015) (36.935 *contigs*), o que já era esperado por se tratar de um alinhamento contra um genoma de referência, onde parametrizamos para utilizar apenas um alinhamento por *read*, desta forma mais

conservadora, obtvemos uma quantidade de *contigs* menor em relação a montagem *de novo*. Lembrando que, a montagem *de novo* usando a técnica de RNA-Seq *single-end*, caso semelhante ao de Yuyama *et al.* (2015), normalmente gera um número alto de *contigs* (HAN *et al.*, 2013).

## 5.2 ANOTAÇÃO E CARACTERIZAÇÃO FUNCIONAL DOS UNIGENES DE *C. EUGENIOIDES*

Os 16.743 *contigs* de *C. eugenioides* foram analisados com base na similaridade de suas sequências com o BlastX contra sequências não redundantes de proteínas disponíveis no banco de dados NCBI (NCBI-nr). Sendo que, (91,49%) apresentaram similaridade com pelo menos uma sequência do NCBI-nr. Além disso, foi realizada uma análise de anotação funcional dos unigenes com a ferramenta BlastX contra o banco de dados de proteínas curadas (UniProt – Swiss-Prot), onde foram anotados 12.277 *contigs* (73,32%) nessa análise. Os resultados de similaridade indicaram que houve maiores semelhanças entre os unigenes obtidos por este trabalho com as espécies de planta *C. canephora* (12.702 *contigs*), seguido por *C. arabica* (14,38%), *C. eugenioides* (9,25%) e *Sesamum indicum* (0,74%).

Uma análise de comparativa de similaridade entre os *contigs* montados de *C. eugenioides* com 4 bancos de dados públicos de *Coffea sp.* através da ferramenta BlastN: i) *expressed sequence tags* (ESTs) de *C. arabica* contendo 35.113 *contigs* (MONDEGO *et al.*, 2011); ii) *coding region sequences* (CDS) do genoma de *C. canephora* (DENOEUDE *et al.*, 2014); iii) transcritos de *C. arabica* obtidos via RNA-seq contendo 65364 *unigenes* (IVAMOTO *et al.*, 2017); iv) 36.935 *unigenes* de um transcriptoma *ab initio* de folhas e frutos de *C. eugenioides* obtidos via RNA-seq (YUYAMA *et al.*, 2015), essas mesmas bibliotecas foram utilizadas para montar o transcriptoma do presente trabalho, porém, utilizamos o genoma de referência para esta nova montagem.

De um total de 16.743 *unigenes*, 82,06% (13740 *contigs*) apresentaram similaridade maior que 90% com os ESTs de *C. arabica* (MONDEGO *et al.*, 2011) (Tabela 4). Os mesmos *contigs* quando comparados com os CDS de *C. canephora* (DENOEUDE *et al.*, 2014) obtivemos 91,38% (15.300 *contigs*) de similaridade (Tabela 4). Em uma outra análise comparativa contra *unigenes* de sequências expressas de

*C. arabica* obtidos via RNA-seq (IVAMOTO *et al.*, 2017) 94,87% (15.885 *contigs*) das sequências apresentaram similaridade maior que 90% (Tabela 4).

**Tabela 4** – Análise de similaridade comparativa entre os unigenes de *C. eugenoides* e unigenes de trabalhos anteriores de *Coffea sp.*

Banco de dados	Contigs anotados	Porcentagem de anotação
EST <i>database</i>	13.740	82,06 %
<i>C. canephora</i> (CDS)	15.300	91,38%
<i>C. arabica</i> (RNA-seq)	15.885	94,87%
<i>C. eugenoides</i> (RNA-seq)	16.421	98,07%

**Fonte:** o próprio autor.

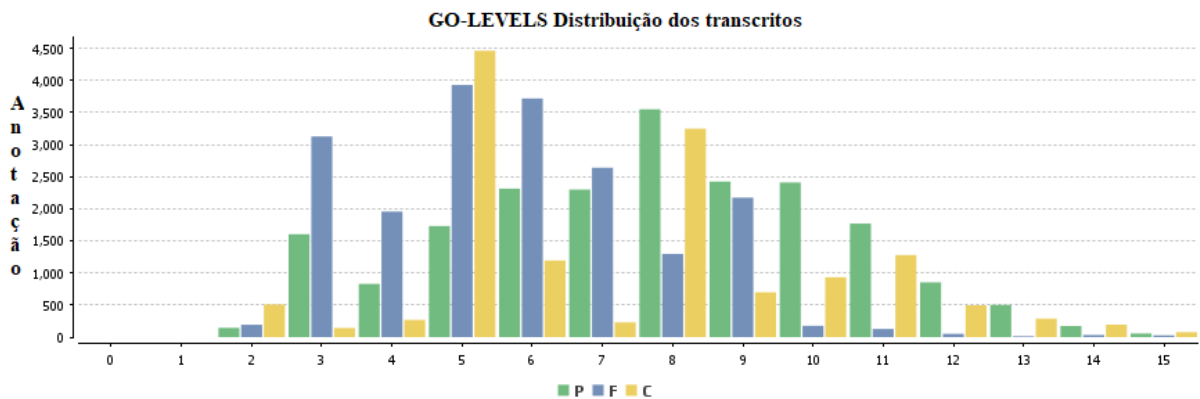
A última comparação realizada foi feita utilizando os unigenes do transcriptoma *ab initio* de *C. eugenoides* (YUYAMA *et al.*, 2015), geradas com as mesmas bibliotecas utilizadas por este trabalho, e o resultado foi e 98,07% (16.421 *contigs*) de similaridade entre as sequências, onde apenas 322 *contigs* não apresentaram similaridade (anexo A). Estes 322 *contigs* podem conter novos genes ainda não identificados para a espécie e configuram dados que podem ser explorados futuramente. Desses *contigs*, 286 também foram encontrados nos bancos de dados de *C. canephora* (DENOUEE *et al.*, 2014) e de *C. arabica* (MONDEGO *et al.*, 2011), restando 36 *contigs* não anotados em nenhum dos dados de sequência de café analisados neste trabalho.

As análises comparativas realizadas até o presente momento, evidenciam que a montagem do transcriptoma de *C. eugenoides* utilizando o genoma de referência para o alinhamento dos *reads* mostrou-se muito mais eficiente em identificar genes que codificam proteínas funcionais do que o encontrado pelo transcriptoma *ab initio* (YUYAMA *et al.*, 2015), o qual apresentou uma porcentagem menor de similaridade contra os mesmos bancos de dados utilizados neste trabalho. Apenas 45,8%, 63,1%, 50,2 e 65,1% dos unigenes apresentaram similaridade com os bancos de dados do NCBI-nr, Uniprot/Swiss-Prot, ESTs de café (MONDEGO *et al.*, 2011) e CDS de *C. canephora* (DENOUEE *et al.*, 2014), respectivamente.

Quanto a caracterização funcional dos *contigs* de *C. eugenoides*, um total de 13.716 *contigs* (81,92%) apresentaram atribuição de um ou mais termos GOs, sendo que estes variaram de 1 e 34 por transcrito. No total, 56.059 termos GOs foram

atribuídos ao longo dos 13.716 *contigs*, onde 38,39% estavam relacionados com processo biológico, 35,69% com função molecular e 25,91% com a categoria de componentes celulares. A classificação de GOs é distribuída em 15 níveis categóricos, sendo que os níveis mais informativos para este transcriptoma foram o cinco, seis e o oito, com 10.121, 7.124 e 8.093 termos GOs anotados, respectivamente (Figura 4). Resultados semelhantes, em relação a distribuição dos termos GOs entre os 15 níveis, foram identificados em um trabalho anterior de transcriptoma de *C. arabica* (IVAMOTO *et al.*, 2017a).

**Figura 4** – Distribuição de termos GO encontrados no transcriptoma de *C. eugenoides* nas categorias processo biológico, função molecular e componente celular.



\*P: processo biológico (barras verdes). F: função molecular (barras azuis). C: componente celular (barras amarelas).

**Fonte:** o próprio autor.

Os termos GO mais abundantes para categoria de função molecular foram *ATP binding* (GO:0005524), *protein binding* (GO:0005515), *metal ion binding* (GO:0046872) e *DNA binding* (GO:0003677). Na categoria de processo biológico, *oxidation-reduction process* (GO:0055114), *protein phosphorylation* (GO:0006468), *regulation of transcription, DNA-templated* (GO:0006355) e *serine family amino acid metabolic process* (GO:0009069) foram os termos atribuídos ao maior número de *contigs*. Por fim, na categoria de componente celular, os termos mais abundantes foram *integral component of membrane* (GO:0016021), *nucleus* (GO:0005634), *cytosol* (GO:0005829) e *cytoplasm* (GO:0005737) (Figura 5).

**Figura 5 – Os termos GO mais abundantes para cada categoria.**

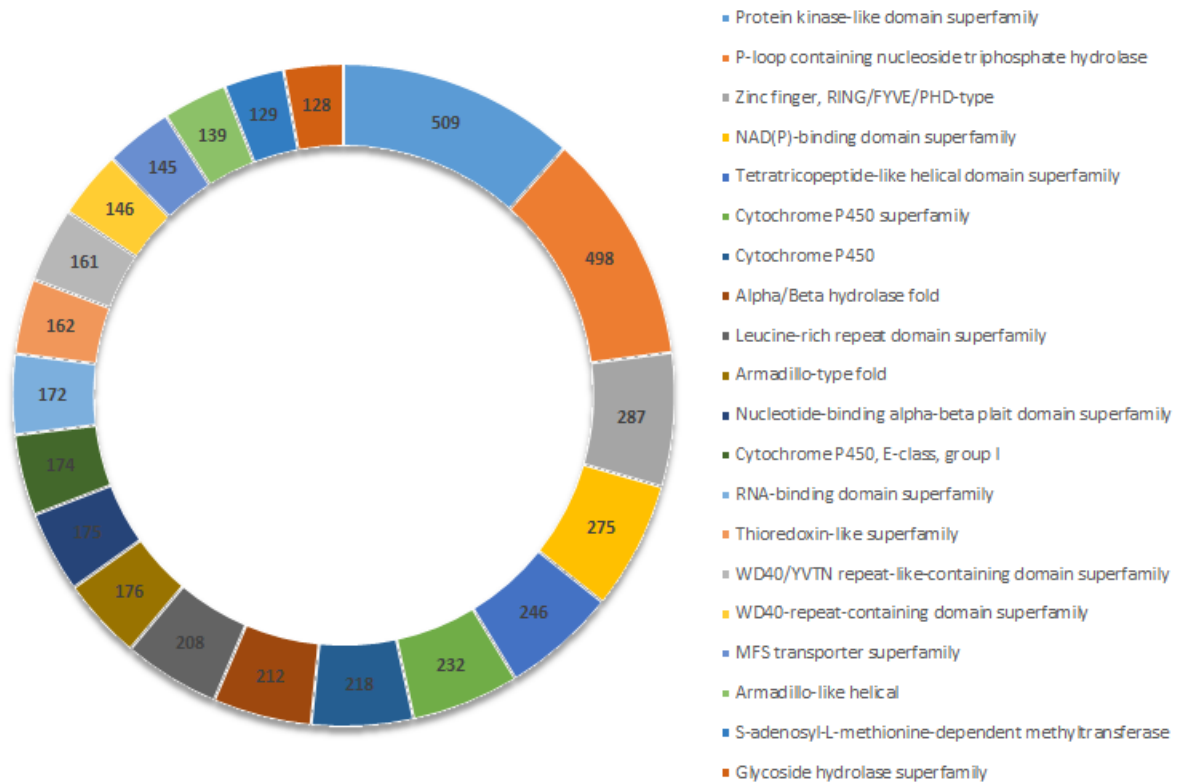


\*Função molecular (barras azuis). Processo biológico (barras laranjas). Componente celular (barras verdes).

**Fonte:** o próprio autor

Adicionalmente a análise de termos GOs, foram anotados os domínios conservados presentes nos transcritos montados de *C. eugenoides* utilizando o banco de dados InterProScan e a ferramenta presente no software Blast2GO. Um total de 14.782 *contigs* apresentaram 4.249 domínios/famílias conservados. Os 20 domínios/famílias conservados mais abundantes foram as proteínas quinases (509 *contigs*), P-loop (498 *contigs*), zinc finger (287 *contigs*), NAD(P)-binding (275 *contigs*), Tetratricopeptide-like helical (246 *contigs*) e citocromo P450 com 232 *contigs* (Figura 6).

**Figura 6** – Caracterização dos vinte domínios/famílias de proteínas conservadas mais abundantes no transcriptoma de *C. eugenioides*.

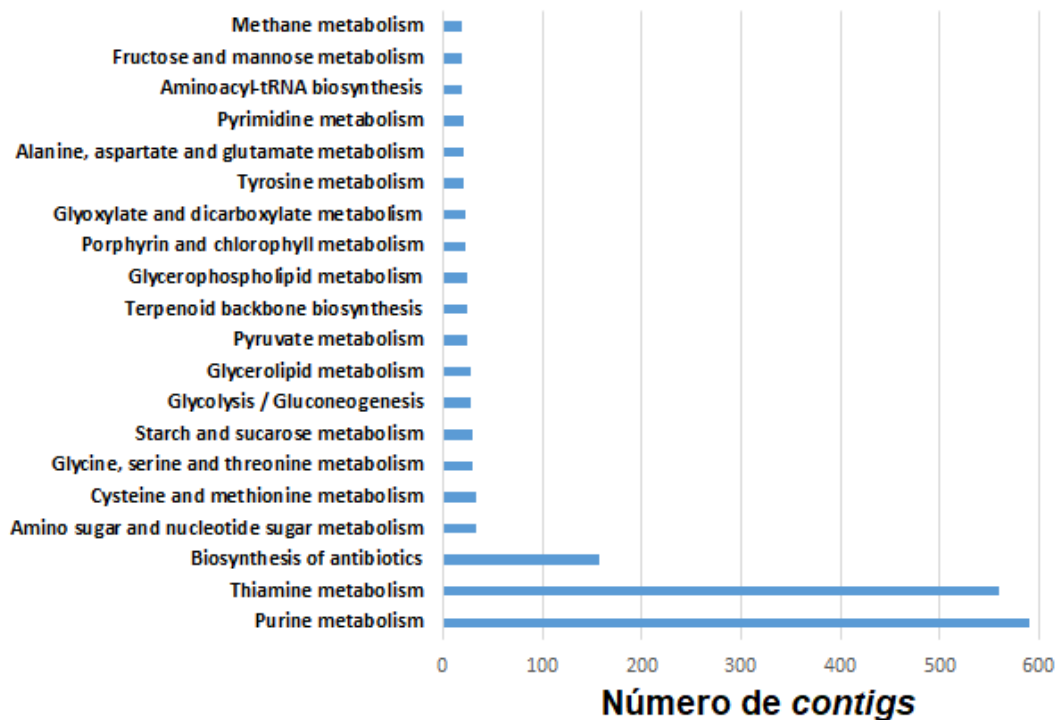


**Fonte:** o próprio autor.

Os resultados dos domínios conservados foram semelhantes aos encontrados em outras análises de transcriptômica de cafeeiros, onde as proteínas quinases, *zinc finger*, citocromos P450s e domínios de ligação foram as categorias mais frequentemente identificadas (LIN *et al.*, 2005; MONDEGO *et al.*, 2011; CARDOSO *et al.*, 2014; YUYAMA *et al.*, 2015; IVAMOTO *et al.*, 2017a).

Posteriormente, na identificação de mapas de vias metabólicas via banco de dados do KEGG, um total de 1.249 *contigs* (7,45%) foram distribuídos em 149 vias metabólicas e representam 777 enzimas diferentes. A categoria mais abundante foi o metabolismo de purina com 590 membros, seguido de metabolismo de tiamina (560 membros) e biossíntese de antibióticos (158 membros). Um gráfico com as 20 principais categorias é apresentado na Figura 7.

**Figura 7** – Principais categorias de vias metabólicas obtidas pelo KEGG.



Fonte: o próprio autor.

A análise de vias metabólicas pelo KEGG apresentou números superiores aos encontrados por Yuyama *et al.* (2015), sendo 1.249 contigs contra 802, além de 777 enzimas contra 374. Em relação as categorias mais abundantes, ambos os trabalhos apresentaram semelhanças, como por exemplo, o metabolismo de purina, metabolismo de amido e sacarose e glicólise/gliconeogênese. Resultados semelhantes obtendo como categorias mais abundantes o metabolismo de purina e metabolismo de tiamina foi encontrado por Tran *et al.* (2018), em *C. arabica*.

### 5.3 IDENTIFICAÇÃO DOS GENES DIFERENCIALMENTE EXPRESSOS

Foram identificados 923 possíveis transcritos diferencialmente expressos entre folhas e frutos de *C. eugenoides*, onde 414 foram preferencialmente expressos em folhas e 509 em frutos (Anexo B).

Dos 923 transcritos, 855 apresentaram sequências completas para a codificação de proteínas, 873 e 714 transcritos apresentaram similaridade com sequências previamente descritas em outras plantas nos bancos de dados NCBI-nr e UniProt/Swiss-Prot, respectivamente.

O nosso trabalho identificou 923 transcritos diferencialmente expressos e o trabalho anterior encontrou apenas 79 (YUYAMA *et al.*, 2015). Destes 79, nós identificamos 69 deles, o que evidencia uma correspondência entre os dois trabalhos em busca de genes diferencialmente expressos (DEs). Além disso, 9 *contigs* que estão entre os 322 *contigs* que foram identificados como novos em relação ao trabalho de Yuyama *et al.* (2015) (Anexo A), foram classificados com DEs.

Provavelmente, a melhora no processo de identificação desses genes é resultado da evolução das tecnologias e dos softwares que realizam essas análises (COSTA-SILVA; DOMINGUES; LOPES, 2017) e também pelo fato de termos utilizado um genoma de referência para realização dos alinhamentos das sequências provenientes do RNA-seq. Entre os 923 possíveis genes, 69 tiveram similaridade com as sequências dos 79 genes encontrados por Yuyama *et al.* (2015),

#### **5.4 SELEÇÃO E VALIDAÇÃO DOS GENES CANDIDATOS PARA A VIA DE BIOSÍNTESE DE DITERPENOS**

Dos 923 unigenes DEs, foram selecionados 5 genes da superfamília das citocromos P450 (*CeCYPs*) e um gene possivelmente relacionado a via metabólica dos diterpenos (*CeGA20OX1*). Utilizamos a tecnologia de quantificação de transcritos (RT-qPCR) para validar o padrão transcricional desses seis unigenes com expressão maior em frutos do que em folhas com base nos resultados *in silico* gerados pelo software DESeq2 (Tabela 5), e também, utilizando dados prévios de RNA-Seq (YUYAMA *et al.*, 2015). Os dados de RT-qPCR corroboraram com os perfis transcricionais observados *in silico* (Figura 8).

**Tabela 5** – Caracterização dos genes candidatos escolhidos para validação do perfil de expressão diferencial entre folhas e frutos de cafeeiros via RT-qPCR.

<b>Genes ID</b>	<b>Anotação (NCBI-nr)</b>	<b>Log2FC</b>	<b>p-valor</b>	<b>TPM em folha</b>	<b>TPM em fruto</b>
<i>CeGA20OX1</i>	Giberelina oxidase 1	3,9330	0,006391	0,78806	674,612
<i>CeCYP71D55</i>	Premnaspirodione oxigenase	2,4806	0,045615	13,6821	94,9079
<i>CeCYP734A1</i>	Citocromo P450 734A1	2,8952	0,043103	1,49748	29,1271
<i>CeCYP76B10</i>	Geraniol 8-hidroxilase	3,0892	0,032474	0,97598	233,423
<i>CeCYP76AH1</i>	Ferruginol sintase	4,3428	0,001596	3,74668	263,423
<i>CeCYP81E8</i>	Citocromo P450 81E8	3,0464	0,035220	1,13807	28,5038

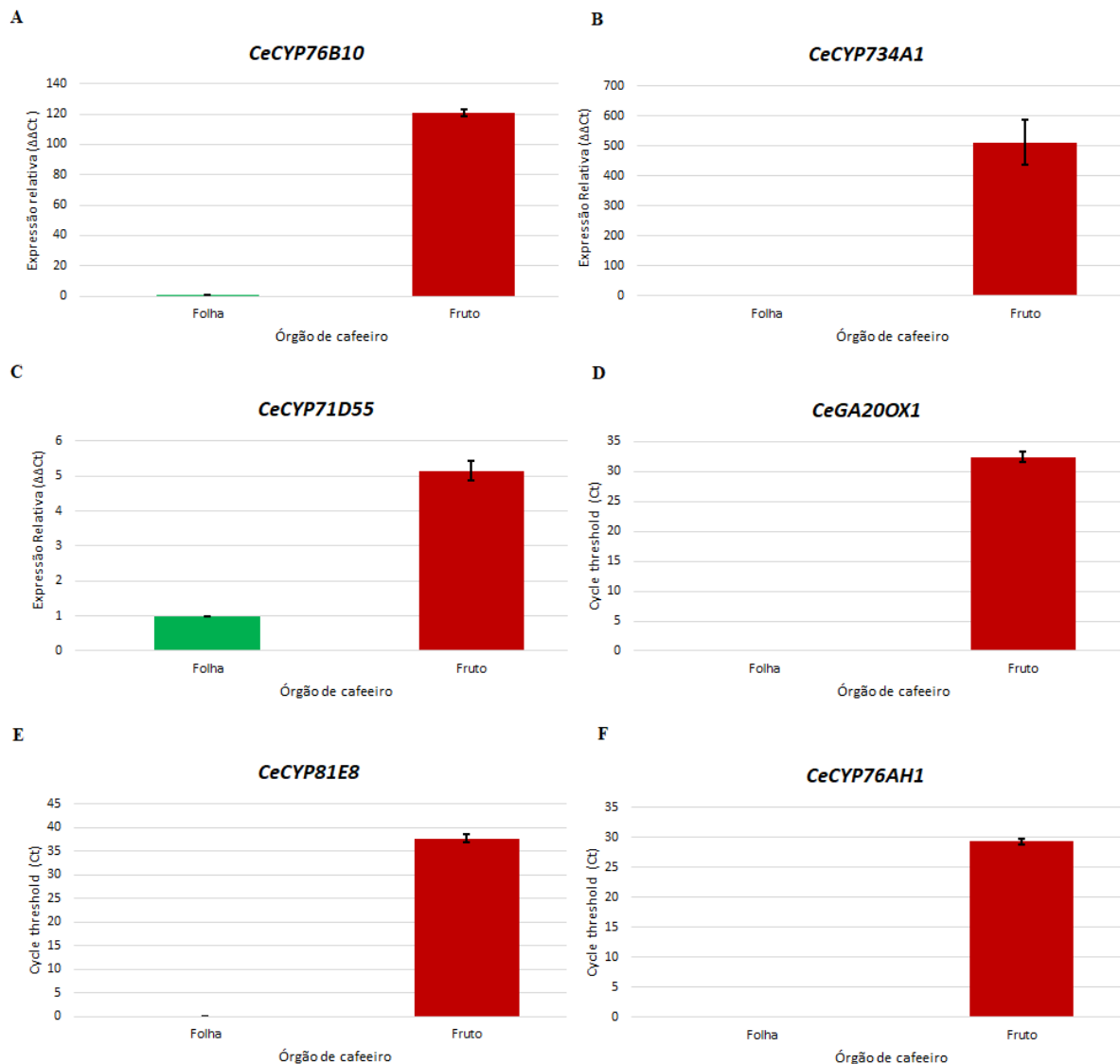
\*Log2FC: Log 2 fold change. metodologia de análise de expressão gênica TPM (transcritos por milhão).

**Fonte:** o próprio autor.

Três genes que codificam citocromos P450 monooxigenases (*CeCYP76B10*, *CeCYP734A1* e *CeCYP71D55*) evidenciaram uma maior atividade transcricional em frutos quando comparados com folhas. *CeCYP76B10*, *CeCYP734A1* e *CeCYP71D55* foram 120, 500 e 5 vezes mais expressos em frutos do que em folhas, respectivamente (Figuras 8A, 8B e 8C).

Dois genes que codificam P450 (*CeCYP81E8* e *CeCYP76AH1*) e um que codifica compostos diterpênicos (*CeGA20OX1*), apresentaram atividade transcricional específica em frutos, pois não foram detectados mRNAs em folhas (Figuras 8D, 8E e 8F). Estes três genes (8D, 8E e 8F) possuem níveis transcricionais menores do que os três primeiros (8A, 8B e 8C), pois seus produtos de PCR foram detectados mais tardiamente (*cycle threshold*).

**Figura 8** – Resultado da análise de quantificação relativa de transcritos via RT-qPCR para 6 genes de *C. eugenioides*.



**Fonte:** o próprio autor.

Dos unigenes selecionados, cinco (*CeCYP81E8*, *CeCYP734A1*, *CeCYP76B10*, *CeCYP76AH1* e *CeCYP71D55*) foram anotados como genes que codificam citocromo P450. Essa superfamília gênica é responsável pela produção de inúmeras proteínas enzimáticas, e os mesmos podem ser encontrados em quase todos os organismos. Em plantas, as citocromos estão envolvidas em inúmeras reações biossintéticas, como por exemplo, a produção de metabólitos primários (hormônios vegetais) e síntese de diversos metabólitos secundários (WERCK-REICHHART; FEYEREISEN, 2000; SCHULER; WERCK-REICHHART, 2003;

NELSON; WERCK-REICHHART, 2011; WANG *et al.*, 2012; PETERAKI *et al.*, 2015; IVAMOTO *et al.*, 2015). As P450 já foram descritas, em diversas plantas, como as responsáveis pela produção de compostos diterpênicos (WANG *et al.*, 2012; ZERBE *et al.*, 2013). Os diterpenos são um grupo de isoprenóides cíclicos, que no caso de cafés, estão presentes na fração lipídica da planta. Além disso, os diterpenos de café, cafestol e caveol, são compostos importantes, pois influenciam a qualidade final da bebida (SANT'ANA *et al.*, 2018). A quantidade de diterpenos encontrados nas plantas é variável inter e intra espécie (KITZBERGER *et al.*, 2013) e de acordo com o tecido/orgão analisado (IVAMOTO *et al.*, 2017b)

O unigenes *CeCYP71D55* pertence à família de genes CYP71, sendo que *CYP71D55* está descrito como uma oxigenase prenaspirodienne envolvida na biossíntese de solvetivona, uma potente fitoalexina antifúngica (TAKAHASHI *et al.*, 2007). Os genes que compõem a família CYP71, descrita como a maior família do P450, estão envolvidos em várias funções (NELSON; WERCK-REICHHART, 2011). Em tomate, genes CYP71 tiveram alta expressão durante os estágios iniciais do desenvolvimento do fruto do tomateiro (LI *et al.*, 2010). Genes da subfamília CYP71D também estão envolvidos em reações que levam na oxidação de rotundona em uvas (TAKASE *et al.*, 2015). O rotundona é um terpeno, sendo uma potente molécula odorante com um aroma picante característico existente em várias plantas, incluindo videiras (TAKASE *et al.*, 2015).

Os unigenes *CeCYP76B10* e *CeCYP76AH1* pertencem a mesma subfamília (CYP76) a qual está envolvida na biossíntese de diterpenoides em arroz (WANG *et al.*, 2012). O gene *CYP76B10* é descrito como uma hidroxilase envolvida na biossíntese de geraniol (WANG *et al.*, 2010). O geraniol é um álcool monoterpenoide presente em óleos essenciais de algumas plantas, como por exemplo, rosas e citronela, e sendo utilizado comercialmente como fragrância (CHEN; VILJOEN, 2010). Por sua vez, o gene *CYP76AH1* é descrito na literatura como produtor de ferruginol em diversos tipos de plantas (GUO *et al.*, 2013; MA *et al.*, 2016). O ferruginol é um diterpenoide abietano que tem um papel como agente antineoplásico, agente antibacteriano, agente protetor e metabólito vegetal (HO *et al.*, 2012). Alguns genes da subfamília CYP76 também podem estar envolvidos na produção de diterpenos cafestol em café (IVAMOTO *et al.*, 2017b). O cafestol é um dos componentes principais da fração lipídica em grãos de café verde e torrado

(SCHOLZ *et al.*, 2016) e também está relacionado à qualidade da bebida (DEL TERRA *et al.*, 2013).

O gene *CeCYP734A1* pertence a subfamília CYP734 de P450. A subfamília CYP734 é descrita em *Arabidopsis* como enzimas inativadoras de brassinosteroides importantes para a transdução de sinal mediada por esteroides (THORNTON *et al.*, 2010). Os brassinoesteroides são lactonas polihidroxi-esteroides que ocorrem naturalmente em plantas. Eles são hormônios que promovem o alongamento e expansão celular (ARAGÃO *et al.*, 2008). Em *Pyrus pyrifolia* (pera-nashi), foi identificado um gene *CYP734A1* altamente expresso em frutos jovens (JIANG *et al.*, 2018).

*CeCYP81E8*, que também é um citocromo P450, foi descrito na literatura primeiramente em *Medicago tribuloides* (LIU *et al.*, 2003). Genes *CYP81E* são isoflavona 2'-hidroxilase. Isoflavonas são fitoestrógenos encontradas predominantemente em leguminosas (THAM *et al.*, 1998). As isoformas de *CYP81E* também foram identificadas em *L. japonicus* (SHIMADA *et al.*, 2000) e em soja (USHIDA; AKASHI; AOKI, 2015) com funções de hidroxilação de isoflavonas.

Por sua vez o unigene restante, *CeGA20OX1*, foi anotado como uma proteína gibberelina 20 oxidase 1, responsável por biossíntese do fitohormônio gibberelina. As gibberelinas são fitohormônios diterpênicos que possuem uma grande importância no desenvolvimento das plantas (YAMAGUCHI, 2008). Genes da subfamília GA20OX estão envolvidos na regulação do florescimento e na frutificação em *V. vinifera*, apresentando, em sua maioria, maior expressão em flores. Porém, alguns destes genes são expressos no desenvolvimento dos frutos (GIACOMELLI *et al.*, 2013). Em tomates (*Solanum lycopersicum*), os genes *GA20OX* também estão envolvidos no desenvolvimento do fruto (SERRANI *et al.*, 2007).

## 6 CONSIDERAÇÕES FINAIS

No presente trabalho de reanálise de dados de RNA-Seq de *C. eugenoides*, foram montados 16.743 transcritos, sendo identificados 322 transcritos novos em relação ao trabalho original de Yuyama *et al.* (2015), o que levanta a possibilidade de novos genes descritos para a espécie *C. eugenoides*. Apesar do número de transcritos ter sido menor em comparação Yuyama *et al.* (2015), esta pesquisa apresentou uma porcentagem maior de anotação em todos os bancos de dados utilizados em ambos os trabalhos em relação ao número de transcritos. Também foram identificados 36 genes ainda não descritos. O número de genes diferencialmente expressos também foi maior em relação a Yuyama *et al.* (2015), sendo 923 genes contra 79.

Entre os genes diferencialmente expressos, foram validados 6 genes confirmando as análises *in silico*. A confirmação da preferência e especialidade de expressão de genes da família Citocromo P450 aqui identificados abre uma nova linha de estudo para a sua caracterização, algo importante para a análise dos genes envolvidos na biossíntese de diterpenóides em cafeeiros.

## REFERÊNCIAS

- ABIC. Associação Brasileira da Indústria de Café. 2018. Disponível em: <<http://www.abic.com.br>>. Acesso em 20 de janeiro de 2019.
- ALVES, R. C.; CASAL, S.; OLIVEIRA, B. Benefícios do café na saúde: mito ou realidade. **Quim Nova**, v. 32, n. 2169, p. e80, 2009.
- ANDREWS, S. FastQC: A Quality Control tool for High Throughput Sequence Data. <<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>> 2010.
- ANTHONY, F. *et al.* Adaptive radiation in *Coffea* subgenus *Coffea* L. (Rubiaceae) in Africa and Madagascar. **Plant Systematics and Evolution**, v. 285, n. 1-2, p. 51-64, 2010.
- ARAGÃO, P. H. *et al.* Brassinosteróide e substratos na aclimatização do abacaxizeiro 'Imperial'. **Acta Scientiarum. Agronomy**, v. 30, n. 3, 2008.
- BICC, Barea de Inteligência Competitiva de Café. 2017. Disponível em: <<http://www.icafebr.com.br/>>. Acesso em 20 de agosto de 2017.
- BOLGER, A. M.; LOHSE, M.; USADEL, B. Trimmomatic: A flexible trimmer for Illumina Sequence Data. **Bioinformatics**, v. 30, p. 2114–2120, 2014.
- BOLEY, N. *et al.* Genome-guided transcript assembly by integrative analysis of RNA sequence data. **Nature biotechnology**, v. 32, n. 4, p. 341, 2014.
- BØHN, S. K. *et al.* Effects of tea and coffee on cardiovascular disease risk. **Food & function**, v. 3, n. 6, p. 575-591, 2012.
- BORTIRI, E. *et al.* The complete chloroplast genome sequence of *Brachypodium distachyon*: sequence comparison and phylogenetic analysis of eight grass plastomes. **BMC Research Notes**, v. 1, n. 1, p. 61, 2008.
- BREDESON, J. V. *et al.* Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. **Nature biotechnology**, v. 34, n. 5, p. 562, 2016.
- BUSTIN, S. A. *et al.* The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. **Clin Chem**, v. 55, p. 611-622, 2009.
- CAMPANHA, F. G. Discriminação de espécies de café (*Coffea arabica* e *Coffea canephora*) pela composição de diterpenos. Tese de Doutorado. **Universidade Estadual de Londrina**, 2008.
- CARDOSO, D. C. *et al.* Large-scale analysis of differential gene expression in coffee genotypes resistant and susceptible to leaf miner—toward the identification of candidate genes for marker assisted-selection. **BMC genomics**, v. 15, n. 1, p. 66, 2014.

CARVALHO, K. *et al.* Nitrogen starvation, salt and heat stress in coffee (*Coffea arabica* L.): identification and validation of new genes for qPCR normalization. **Molecular biotechnology**, v. 53, n. 3, p. 315-325, 2013.

CECAFE. Conselho dos Exportadores de Café do Brasil. 2018. Disponível em: <<http://www.cecafe.com.br>>. Acesso em: 12 de agosto de 2018.

CENCI, A.; COMBES, M. C.; LASHERMES, P. Genome evolution in diploid and tetraploid *Coffea* species. **Plant Mol Biol.**, v. 78 p.135-145, 2012.

CHEN, W.; VILJOEN, A. M. Geraniol—a review of a commercially important fragrance material. **South African Journal of Botany**, v. 76, n. 4, p. 643-651, 2010.

CHEN, G.; WANG, C.; SHI, T. Overview of available methods for diverse RNA-Seq data analyses. **Science China Life Sciences**, v. 54, n. 12, p. 1121-1128, 2011.

CLARK, J. *et al.* Identification of amplified and expressed genes in breast cancer by comparative hybridization onto microarrays of randomly selected cDNA clones. **Genes, Chromosomes and Cancer**, v. 34(1), p. 104-114, 2002.

CLARKE, R. J.; MACRAE, R. Coffee chemistry (Vol. 1). **Essex, England: Elsevier Science**, v. 306, 1985.

CONAB, Companhia Nacional de Abastecimento 2018. Disponível em: <<https://www.conab.gov.br/info-agro/safras/cafe/boletim-da-safra-de-cafe>> Acesso em 20 de janeiro de 2019.

CONCEIÇÃO, C. H. C.; GUERREIRO-FILHO, O.; GONÇALVES, W. Flutuação populacional do bicho-mineiro em cultivares de café arábica resistentes à ferrugem. **Bragantia**, v. 64, n. 4, p. 625-631, 2005.

COSTA-SILVA, J.; DOMINGUES, D.; LOPES, F. M. RNA-Seq differential expression analysis: An extended review and a software tool. **PloS one**, v. 12, n. 12, p. e0190152, 2017.

COTTA, M. G. *et al.* Lipid transfer proteins in coffee: isolation of *Coffea* orthologs, *Coffea arabica* homeologs, expression during coffee fruit development and promoter analysis in transgenic tobacco plants. **Plant molecular biology**, v. 85, n. 1-2, p. 11-31, 2014.

DAMATTA, F. M.; RAMALHO, J. D. C. Impacts of drought and temperature stress on coffee physiology and production: a review. **Brazilian Journal of Plant Physiology**, v. 18, n. 1, p. 55-81, 2006.

DA SILVA, J. Genetic transcript analyzer-Ferramenta computacional para análise de transcrição gênica por RNA-Seq. Tese de Doutorado. Curitiba: Programa de Pós-Graduação em Bioinformática, **Universidade Federal do Paraná**. 2012.

DAVIS, A. P. *et al.* Growing coffee: *Psilanthus* (Rubiaceae) subsumed on the basis of molecular and morphological data; implications for the size, morphology, distribution

and evolutionary history of *Coffea*. **Botanical Journal of the Linnean Society**, v. 167(4), p. 357-377, 2011.

DEL TERRA, L. et al. Functional characterization of three *Coffea arabica* L. monoterpene synthases: Insights into the enzymatic machinery of coffee aroma. **Phytochemistry**, v. 89, p. 6-14, 2013.

DENOEUDE, F. et al. The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. **Science**, v. 345, n. 6201, p. 1181-1184, 2014.

DEREEPER, A. et al. The coffee genome hub: a resource for coffee genomes. **Nucleic acids research**, v. 43, n. D1, p. D1028-D1035, 2014.

DI LORENZO, A. et al. Effects of tea and coffee consumption on cardiovascular diseases and relative risk factors: An update. **Current pharmaceutical design**, v. 23, n. 17, p. 2474-2487, 2017.

FAO. (2017) Food and Agriculture Organization of the United Nations. Disponível em: [www.fao.org](http://www.fao.org). Acesso em 17 ago 2017.

FARAH, A. Coffee constituents. **Coffee: emerging health effects and disease prevention**, v. 1, p. 22-58, 2012.

FARAH, A. et al. Correlation between cup quality and chemical attributes of Brazilian coffee. **Food Chemistry**, v. 98, n. 2, p. 373-380, 2006.

FLOREA, L. D.; SALZBERG, S. L. Genome-guided transcriptome assembly in the age of next-generation sequencing. **IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)**, v. 10, n. 5, p. 1234-1240, 2013.

FU, S. et al. IDP-denovo: de novo transcriptome assembly and isoform annotation by hybrid sequencing. **Bioinformatics**, v. 1, p. 9, 2018.

GIACOMELLI, L. et al. Gibberellin metabolism in *Vitis vinifera* L. during bloom and fruit-set: functional characterization and evolution of grapevine gibberellin oxidases. **Journal of experimental botany**, v. 64, n. 14, p. 4403-4419, 2013.

GUERCIO, B. J. et al. Coffee intake, recurrence, and mortality in stage III colon cancer: results from CALGB 89803 (Alliance). **Journal of Clinical Oncology**, v. 33, n. 31, p. 3598, 2015.

GUERREEIRO FILHO, O. et al. Origem e classificação botânica do cafeeiro. Cultivares de café: origem, características e recomendações. Brasília: **Embrapa Café**, p. 27-33, 2008.

GUO, J. et al. CYP76AH1 catalyzes turnover of miltiradiene in tanshinones biosynthesis and enables heterologous production of ferruginol in yeasts. **Proceedings of the National Academy of Sciences**, v. 110, n. 29, p. 12108-12113, 2013.

GUTTMAN, M. *et al.* Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. **Nature biotechnology**, v. 28, n. 5, p. 503-510, 2010.

HAAS, B. J. *et al.* How deep is deep enough for RNA-Seq profiling of bacterial transcriptomes? **BMC genomics**, v. 13, n. 1, p. 734, 2012.

HAN, X. *et al.* Transcriptome sequencing and expression analysis of terpenoid biosynthesis genes in *Litsea cubeba*. **PLoS one**, v. 8, n. 10, p. e76890, 2013.

HERRERA, J. C. *et al.* Genomic relationships among different Timor hybrid (*Coffea* L.) accessions as revealed by SNP identification and RNA-Seq analysis. In: **Advances in Computational Biology**. Springer, Cham, p. 161-168, 2014.

HO, C. *et al.* Composition, antioxidant, antimicrobial and anti-wood-decay fungal activities of the twig essential oil of *Taiwania cryptomerioides* from Taiwan. **Natural product communications**, v. 7, n. 2, p. 261-264, 2012.

HRDLICKOVA, R.; TOLOUE, M.; TIAN, B. RNA-Seq methods for transcriptome analysis. **Wiley Interdisciplinary Reviews: RNA**, v. 8, n. 1, p. e1364, 2017.

ICO. International Coffee Organization. 2019. Disponível em: <<http://www.ico.org>>. Acesso em 20 de janeiro de 2019.

INTERNATIONAL WHEAT GENOME SEQUENCING CONSORTIUM *et al.* A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. **Science**, v. 345, n. 6194, p. 1251788, 2014.

IVAMOTO, S. T. *et al.* Identification of the transcriptionally active cytochrome P450 repertoire in *Coffea arabica*. **Gen. Mol. Res**, v. 14, p. 2399-2412, 2015.

IVAMOTO S. T. *et al.* Transcriptome analysis in leaves, flowers and during initial fruit development of *Coffea arabica* L. reveals differential expression in genes related to raffinose biosynthesis. **PLoS One**, 2017a.

IVAMOTO, S. T. *et al.* Diterpenes biochemical profile and transcriptional analysis of cytochrome P450s genes in leaves, roots, flowers, and during *Coffea arabica* L. fruit development. **Plant Physiology and Biochemistry**, v. 111, p. 340-347, 2017.

JAIN, P.; KRISHNAN, N. M.; PANDA, B. Augmenting transcriptome assembly by combining de novo and genome-guided tools. **PeerJ**, v. 1, p. e133, 2013.

JIANG, S. *et al.* Comparative Analysis of Transcriptomes to Identify Genes Associated with Fruit Size in the Early Stage of Fruit Development in *Pyrus pyrifolia*. **International journal of molecular sciences**, v. 19, n. 8, p. 2342, 2018.

KIM, D.; LANGMEAD, B.; SALZBERG, S. L. HISAT: a fast spliced aligner with low memory requirements. **Nature methods**, v. 12, n. 4, p. 357-360, 2015.

KIM, J. G., BEPPU, K., KATAOKA, I. Varietal differences in phenolic content and astringency in skin and flesh of hardy kiwifruit resources in Japan. **Scientia Horticulturae**, v. 120(4), p. 551-554, 2009.

KITZBERGER, C. S. et al. Diterpenes in green and roasted coffee of *Coffea arabica* cultivars growing in the same edapho-climatic conditions. **Journal of Food Composition and Analysis**, v. 30, n. 1, p. 52-57, 2013.

KOLODNER, R.; TEWARI, K. K. Inverted repeats in chloroplast DNA from higher plants. **Proceedings of the National Academy of Sciences**, v. 76, n. 1, p. 41-45, 1979.

LASHERMES, P. et al. Molecular characterisation and origin of the *Coffea arabica* L. genome. **Molecular and General Genetics MGG**, v. 261, n. 2, p. 259-266, 1999.

LANGMEAD, B.; SALZBERG, S. L. Fast gapped-read alignment with Bowtie 2. **Nature methods**, v. 9, n. 4, p. 357-359, 2012.

LEROY, T. et al. Construction and characterization of a *Coffea canephora* BAC library to study the organization of sucrose biosynthesis genes. **Theoretical and Applied Genetics**, v. 111, n. 6, p. 1032-1041, 2005.

LEROY, T. et al. Genetics of coffee quality. **Brazilian Journal of Plant Physiology**, v. 18, n. 1, p. 229-242, 2006.

LI, Z. et al. Molecular cloning and expression analysis of a cytochrome P450 gene in tomato. **Plant growth regulation**, v. 61, n. 3, p. 297-304, 2010.

LIN, C. et al. Coffee and tomato share common gene repertoires as revealed by deep sequencing of seed and cherry transcripts. **Theoretical and applied genetics**, v. 112, n. 1, p. 114-130, 2005.

LIU, C. et al. Regiospecific hydroxylation of isoflavones by cytochrome p450 81E enzymes from *Medicago truncatula*. **The Plant Journal**, v. 36, n. 4, p. 471-484, 2003.

LOMBELLO, R. A.; PINTO-MAGLIO, C. A. F. Cytogenetic studies in *Psilanthus ebracteolatus* Hiern., a wild diploid coffee species. **Cytologia**, v. 68, n. 4, p. 425-429, 2003.

LOMBELLO, R. A.; PINTO-MAGLIO, C. A. F. Cytogenetic studies in *Coffea* L. and *Psilanthus* Hook. f. using CMA/DAPI and FISH. **Cytologia**, v. 69, n. 1, p. 85-91, 2004.

LOVE, M. I.; HUBER, W.; ANDERS, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. **Genome biology**, v. 15, n. 12, p. 550, 2014.

LU, B.; ZENG, Z.; SHI, T. Comparative study of de novo assembly and genome-guided assembly strategies for transcriptome reconstruction based on RNA-Seq. **Science China Life Sciences**, v. 56, n. 2, p. 143-155, 2013.

- MA, Y. *et al.* RNA interference targeting CYP76AH1 in hairy roots of *Salvia miltiorrhiza* reveals its key role in the biosynthetic pathway of tanshinones. **Biochemical and biophysical research communications**, v. 477, n. 2, p. 155-160, 2016.
- MAPA, Ministério da Agricultura, Pecuária e Abastecimento. 2018. Disponível em: < <http://www.agricultura.gov.br/assuntos/politica-agricola/valor-bruto-da-producao-agropecuaria-vbp> > Acesso em 10 de fevereiro de 2019.
- MARTIN, J. *et al.* Rnnotator: an automated de novo transcriptome assembly pipeline from stranded RNA-Seq reads. **BMC genomics**, v. 11, n. 1, p. 663, 2010.
- MARTINS, A. M. Sequenciamento de DNA, montagem de novo do genoma e desenvolvimento de marcadores microssatélites, indels e SNPs para uso em análise genética de *Brachiaria ruziziensis*. Tese de Doutorado. **Universidade de Brasília**, 2014.
- MAURIN, O. *et al.* Towards a phylogeny for *Coffea* (Rubiaceae): identifying well-supported lineages based on nuclear and plastid DNA sequences. **Annals of Botany**, v. 100, n. 7, p. 1565-1583, 2007.
- MING, R. *et al.* The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). **Nature**, v. 452, n. 7190, p. 991, 2008.
- MOFATTO, L. S. *et al.* Identification of candidate genes for drought tolerance in coffee by high-throughput sequencing in the shoot apex of different *Coffea arabica* cultivars. **BMC plant biology**, v. 16(1), p. 1, 2016.
- MONDEGO, J. M. C. *et al.* An EST-based analysis identifies new genes and reveals distinctive gene expression features of *Coffea arabica* and *Coffea canephora*. **BMC plant biology**, v. 11, n. 1, p. 30, 2011.
- NELSON, D.; WERCK-REICHHART, D. A P450-centric view of plant evolution. **The Plant Journal**, v. 66, n. 1, p. 194-211, 2011.
- NEVES, C. A história do café. Rio de Janeiro, **Instituto Brasileiro do Café**, p. 52, 1974.
- NOIROT, M. *et al.* Genome size variations in diploid African *Coffea* species. **Annals of Botany**, v. 92, n. 5, p. 709-714, 2003.
- O'KEEFE, J. H. *et al.* Effects of habitual coffee consumption on cardiometabolic disease, cardiovascular health, and all-cause mortality. **Journal of the American College of Cardiology**, v. 62, n. 12, p. 1043-1051, 2013.
- PATERAKI, I. *et al.* Cytochromes P450 for terpene functionalisation and metabolic engineering. In: **Biotechnology of Isoprenoids**. Springer, Cham, 2015. p. 107-139.
- PATERSON, A. H. *et al.* Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. **Nature**, v. 492, n. 7429, p. 423, 2012.

- PERTEA, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. **Nature biotechnology**, v. 33, n. 3, p. 290-295, 2015.
- PINTO-MAGLIO, C. A. F.; DA CRUZ, N. D. Pachytene chromosome morphology in *Coffea* L. II. *C. arabica* L. complement. **Caryologia**, v. 51, n. 1, p. 19-35, 1998.
- RAI, M. K.; SHEKHAWAT, N. S. Genomic resources in fruit plants: an assessment of current status. **Critical reviews in biotechnology**, v. 35, n. 4, p. 438-447, 2014.
- SANGER, F.; NICKLEN, S.; COULSON, A. R. DNA sequencing with chain-terminating inhibitors. **Proceedings of the National Academy of Sciences**, v. 74, n. 12, p. 5463-5467, 1977.
- SCHOLZ, M. B. *et al.* Chemical composition in wild Ethiopian Arabica coffee accessions. **Euphytica**, v. 209, n. 2, p. 429-438, 2016.
- SCHULZ, M. H. *et al.* Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. **Bioinformatics**, v. 28, n. 8, p. 1086-1092, 2012.
- SCHMUTZ, J. *et al.* Genome sequence of the palaeopolyploid soybean. **Nature**, v. 463, n. 7278, p. 178, 2010.
- SCHNABLE, P. S. *et al.* The B73 maize genome: complexity, diversity, and dynamics. **Science**, v. 326, n. 5956, p. 1112-1115, 2009.
- SCHULER, M. A.; WERCK-REICHHART, D. Functional genomics of P450s. **Annual review of plant biology**, v. 54, n. 1, p. 629-667, 2003.
- SERRANI, J. C. *et al.* Gibberellin regulation of fruit set and growth in tomato. **Plant physiology**, v. 145, n. 1, p. 246-257, 2007.
- SHIMADA, N. *et al.* Induction of isoflavonoid pathway in the model legume *Lotus japonicus*: molecular characterization of enzymes involved in phytoalexin biosynthesis. **Plant Science**, v. 160, n. 1, p. 37-47, 2000.
- SOUZA, R.; BENASSI, M. T. Discrimination of commercial roasted and ground coffees according to chemical composition. **Journal of the Brazilian Chemical Society**, v. 23(7), p. 1347-1354, 2012.
- TAKAHASHI, S. *et al.* Functional characterization of prenaspirodiene oxygenase, a cytochrome P450 catalyzing regio- and stereo-specific hydroxylations of diverse sesquiterpene substrates. **Journal of Biological Chemistry**, 2007.
- TAKASE, H. *et al.* Cytochrome P450 CYP71BE5 in grapevine (*Vitis vinifera*) catalyzes the formation of the spicy aroma compound (-)-rotundone. **Journal of experimental botany**, v. 67, n. 3, p. 787-798, 2015.
- THAM, D. *et al.* Potential health benefits of dietary phytoestrogens: a review of the clinical, epidemiological, and mechanistic evidence. **The Journal of Clinical Endocrinology & Metabolism**, v. 83, n. 7, p. 2223-2235, 1998.

TJADEN, B. De novo assembly of bacterial transcriptomes from RNA-seq data. **Genome biology**, v. 16, n. 1, p. 1, 2015.

TRAN, H. T. M. *et al.* SNP in the *Coffea arabica* genome associated with coffee quality. **Tree Genetics & Genomes**, v. 14, n. 5, p. 72, 2018.

TRAPNELL, C. *et al.* Transcript assembly and abundance estimation from RNA-Seq reveals thousands of new transcripts and switching among isoforms. **Nature biotechnology**, v. 28, n. 5, p. 511, 2010.

UCHIDA, K.; AKASHI, T.; AOKI, T. Functional expression of cytochrome P450 in *Escherichia coli*: an approach to functional analysis of uncharacterized enzymes for flavonoid biosynthesis. **Plant Biotechnology**, v. 32, n. 3, p. 205-213, 2015.

UNIPROT CONSORTIUM *et al.* Activities at the universal protein resource (UniProt). **Nucleic acids research**, v. 42, n. D1, p. D191-D198, 2014.

VERDE, I. *et al.* The Peach v2.0 release: high-resolution linkage mapping and deep resequencing improve chromosome-scale assembly and contiguity. **BMC genomics**, v. 18, n. 1, p. 225, 2017.

VEZZALI, F. Agricultura familiar gera empregos, mas recebe pouco recurso. **Repórter Brasil, São Paulo**, v. 10, 2006.

VIDAL, R. O. *et al.* A high-throughput data mining of single nucleotide polymorphisms in *Coffea* species expressed sequence tags suggests differential homeologous gene expression in the allotetraploid *Coffea arabica*. **Plant Physiology**, v. 154, n. 3, p. 1053-1066, 2010.

VAN DER VOSEN, E; BERTRAND, B.; CHARRIER, A. Next generation variety development for sustainable production of arabica coffee (*Coffea arabica* L.): a review. **Euphytica**, v. 204(2), p. 243–256, 2015.

WANG, Z.; GERSTEIN, M.; SNYDER, M. RNA-Seq: a revolutionary tool for transcriptomics. **Nature Reviews**, v.10, p. 57-63, 2009.

WANG, J. *et al.* Cloning and functional analysis of geraniol 10-hydroxylase, a cytochrome P450 from *Swertia mussotii* Franch. **Bioscience, biotechnology, and biochemistry**, v. 74, n. 8, p. 1583-1590, 2010.

WANG, W.; MESSING, J. High-throughput sequencing of three Lemnoideae (duckweeds) chloroplast genomes from total DNA. **PloS one**, v. 6, n. 9, p. e24670, 2011.

WANG, Q. *et al.* Characterization of CYP76M5-8 indicates metabolic plasticity within a plant biosynthetic gene cluster. **Journal of Biological Chemistry**, p. jbc. M111. 305599, 2012.

WERCK-REICHHART, D.; FEYEREISEN, R. Cytochromes P450: a success story. **Genome biology**, v. 1, n. 6, p. reviews3003. 1, 2000.

- WU, A. R. *et al.* Quantitative assessment of single-cell RNA-sequencing methods. **Nature methods**, v. 11, n. 1, p. 41, 2013.
- YAMAGUCHI, S. Gibberellin metabolism and its regulation. **Annu. Rev. Plant Biol.**, v. 59, p. 225-251, 2008.
- YANG, Y.; SMITH, S. A. Optimizing de novo assembly of short-read RNA-seq data for phylogenomics. **BMC genomics**, v. 14, n. 1, p. 328, 2013.
- YUYAMA, P. M. *et al.* Transcriptome analysis in *Coffea eugenioides*, an Arabica coffee ancestor, reveals differentially expressed genes in leaves and fruits. **Molecular genetics and genomics**, v. 291, n. 1, p. 323-336, 2015.
- ZERBE, P. *et al.* Gene discovery of modular diterpene metabolism in non-model systems. **Plant physiology**, v. 162, p. 1073-1091, 2013.
- ZHANG, W. *et al.* Comparison of RNA-seq and microarray-based models for clinical endpoint prediction. **Genome biology**, v. 16, n. 1, p. 133, 2015.
- ZHAO, Q. *et al.* Optimizing de novo transcriptome assembly from short-read RNA-Seq data: a comparative study. **BMC bioinformatics**, v. 12, n. 14, p. S2, 2011.

## ANEXOS

**Anexo A** – Tabela com os 322 genes identificados como novos para *C. eugenioides* (Destaque em vermelho para os 36 genes não descritos para *Coffea spp.*)

Gene ID	Swiss-Prot	CDS <i>C. canephora</i>	EST <i>C. arabica</i>	NCBI-nr
CE.62.1	Protochlorophyllide-dependent translocon component 52	Cc01_g00670	Contig1712	
CE.118.1	NO HIT	Cc01_g00900	Contig16720	
CE.120.1	NO HIT	NO HIT	Contig7237	
CE.134.1	NO HIT	Cc04_g05310	Contig12592	
CE.136.1	NO HIT	Cc00_g11080	NO HIT	
CE.161.1	NO HIT	NO HIT	NO HIT	NO HIT
CE.203.1	Retrovirus-related Pol polyprotein from transposon TNT 1-94	NO HIT	Contig13404	
CE.223.1	Serine/threonine-protein kinase SAPK2	Cc00_g27980	CA00-XX-AR1-015-D04-EB.F	
CE.247.1	Serine carboxypeptidase-like 2	Cc01_g02440	Contig2086	
CE.250.1	Serine carboxypeptidase-like 4	Cc00_g23530	CA00-XX-FB2-040-E04-AB.F	
CE.253.1	Glycine-rich RNA-binding protein 3	Cc00_g14830	Contig2789	
CE.277.1	NO HIT	NO HIT	Contig7237	
CE.278.1	NO HIT	NO HIT	Contig7237	
CE.282.1	NO HIT	Cc07_g17240	Contig6166	
CE.291.1	NO HIT	NO HIT	NO HIT	hypothetical protein PanWU01x14_013720
CE.320.1	Probable aldo-keto reductase 2	Cc00_g13690	Contig8284	
CE.352.1	NO HIT	Cc01_g04580	CA00-XX-PC1-003-F01-EC.F	
CE.419.1	NO HIT	NO HIT	CA00-XX-LV9-035-B02-UT.F	
CE.444.1	NO HIT	NO HIT	Contig6615	
CE.526.1	NO HIT	Cc11_g12800	Contig13903	
CE.566.1	NO HIT	NO HIT	NO HIT	NO HIT
CE.589.1	Cyclin-dependent kinase C-1	Cc01_g07380	Contig9843	
CE.605.1	NO HIT	NO HIT	CA00-XX-CS1-014-B07-EQ.F	
CE.614.1	NO HIT	Cc03_g14790	Contig13404	
CE.717.1	NO HIT	NO HIT	Contig11294	
CE.735.1	Major pollen allergen Ole e 10	Cc00_g27990	NO HIT	
CE.849.1	Probable glucuronoxylan glucuronosyltransferase IRX7	Cc01_g10400	Contig16448	
CE.874.1	Kiwelin	NO HIT	CA00-XX-LV8-080-G12-RF.F	
CE.893.1	CASP-like protein 5A2	Cc01_g11070	NO HIT	
CE.970.1	Peroxisome biogenesis protein 12	Cc06_g14670	Contig2579	
CE.1196.1	NO HIT	Cc00_g35020	CA00-XX-CS1-058-B01-SB.F	

CE.1234.1	Peroxisome biogenesis protein 12	Cc02_g08720	Contig2579	
CE.1358.1	NO HIT	NO HIT	NO HIT	NO HIT
CE.1469.1	Hsp70-Hsp90 organizing protein	Cc01_g18750	NO HIT	
CE.1474.1	Photosystem II reaction center W protein	Cc01_g18800	Contig4885	
CE.1497.1	Protein LIFEGUARD 4	Cc01_g19090	Contig491	
CE.1658.1	Probable galacturonosyltransferase-like 7	Cc01_g21200	CA00-XX-EA1-032-E06-EC.F	
CE.1668.1	NO HIT	Cc01_g21270	CA00-XX-CL2-003-F09-CB.F	
CE.2030.1	NO HIT	Cc00_g35020	NO HIT	
CE.2123.1	Mediator of RNA polymerase II transcription subunit 31	Cc00_g13060	Contig9903	
CE.2124.1	Mediator of RNA polymerase II transcription subunit 31	Cc00_g13060	Contig9903	
CE.2149.1	Probable RNA-binding protein EIF1AD	Cc02_g05210	Contig2045	
CE.2296.1	60S ribosomal protein L35a-1	Cc02_g07030	Contig8550	
CE.2297.1	60S ribosomal protein L35a-3	Cc02_g07030	Contig8550	
CE.2339.1	E3 ubiquitin-protein ligase SDIR1	Cc02_g07580	Contig15762	
CE.2361.1	NO HIT	Cc02_g07910	NO HIT	
CE.2433.1	Peroxisome biogenesis protein 12	Cc02_g08720	Contig11063	
CE.2451.1	Serine/threonine-protein phosphatase PP1	Cc02_g08990	Contig15304	
CE.2592.1	NO HIT	Cc11_g14620	CA00-XX-CB1-016-G08-AC.F	
CE.2939.1	Shaggy-related protein kinase eta	Cc02_g15390	Contig13995	
CE.2943.1	NO HIT	Cc02_g15490	Contig8401	
CE.2982.1	NO HIT	Cc02_g15490	NO HIT	
CE.3702.1	26S proteasome non-ATPase regulatory subunit 1 homolog A	Cc03_g07730	Contig3829	
CE.3703.1	26S proteasome non-ATPase regulatory subunit 1 homolog A	Cc03_g07730	Contig3829	
CE.3723.1	NO HIT	Cc02_g24600	Contig4686	
CE.3791.1	Probable indole-3-pyruvate monooxygenase YUCCA9	Cc01_g17530	NO HIT	
CE.3801.1	Sm-like protein LSM7	Cc02_g27590	Contig646	
CE.3851.1	tRNA-specific adenosine deaminase 2	Cc02_g28310	Contig12955	
CE.3883.1	Putative anthocyanidin reductase	Cc02_g28500	CA00-XX-SH2-037-G10-EM.F	
CE.3899.1	Serine carboxypeptidase-like 40	Cc02_g28910	Contig17538	
CE.3900.1	Serine carboxypeptidase-like 40	Cc02_g28910	Contig17538	
CE.3935.1	NO HIT	NO HIT	NO HIT	NO HIT
CE.4033.1	NO HIT	Cc02_g30290	Contig7005	
CE.4068.1	NO HIT	NO HIT	Contig13404	
CE.4087.1	E3 ubiquitin-protein ligase KEG	Cc00_g09560	NO HIT	
CE.4113.1	NO HIT	Cc04_g05310	Contig12592	
CE.4172.1	Cytochrome c	Cc02_g23400	Contig2813	
CE.4259.1	Polyubiquitin 14	Cc02_g31520	Contig3934	
CE.4302.1	NO HIT	NO HIT	NO HIT	NO HIT
CE.4358.1	NO HIT	Cc02_g33370	CA00-XX-FR1-034-H09-EP.F	

CE.4387.1	Probable mediator of RNA polymerase II transcription subunit 37b	Cc02_g33190	NO HIT	
CE.4422.1	F-box protein PP2-B15	Cc02_g34460	CA00-XX-SI3-082-A11-EM.F	
CE.4426.1	60S ribosomal protein L19-1	Cc02_g34490	Contig6864	
CE.4449.1	Probable protein disulfide-isomerase A6	Cc02_g34820	Contig16762	
CE.4522.1	Small RNA 2'-O-methyltransferase	Cc09_g07800	Contig14608	
CE.4798.1	NO HIT	Cc03_g01370	NO HIT	
CE.4812.1	NO HIT	Cc03_g01560	NO HIT	
CE.4966.1	NO HIT	Cc03_g03480	Contig1388	
CE.5053.1	NO HIT	Cc00_g35020	NO HIT	
CE.5270.1	Putative late blight resistance protein homolog R1B-13	Cc03_g08260	Contig9626	
CE.5277.1	Protein NLP2	Cc06_g21940	Contig15576	
CE.5293.1	Dynamin-2A	Cc10_g09200	CA00-XX-SI3-083-F05-EM.F	
CE.5343.1	Serine carboxypeptidase-like 40	Cc02_g28890	Contig16286	
CE.5344.1	Serine carboxypeptidase-like 40	Cc02_g28910	Contig17538	
CE.5345.1	Serine carboxypeptidase-like 40	Cc02_g28910	Contig17538	
CE.5406.1	KH domain-containing protein At3g08620	Cc06_g00210	CA00-XX-SH2-065-B07-EM.F	
CE.5451.1	Probable aldo-keto reductase 2	Cc03_g09870	Contig6269	
CE.5452.1	Probable aldo-keto reductase 3	Cc03_g09870	Contig6269	
CE.5453.1	Auxin-induced protein PCNT115	Cc03_g09870	Contig15398	
CE.5454.1	Probable aldo-keto reductase 2	Cc03_g09900	Contig8284	
CE.5455.1	NO HIT	NO HIT	Contig1163	
CE.5456.1	Auxin-induced protein PCNT115	Cc03_g09950	Contig8284	
CE.5457.1	Auxin-induced protein PCNT115	Cc03_g09950	CA00-XX-EA1-001-D03-EC.F	
CE.5458.1	Probable aldo-keto reductase 2	Cc00_g32490	Contig15509	
CE.5459.1	Auxin-induced protein PCNT115	Cc00_g17680	Contig15509	
CE.5470.1	Probable aldo-keto reductase 2	Cc03_g09870	Contig6269	
CE.5474.1	Oil body-associated protein 1A	Cc03_g10230	CA00-XX-FR2-005-C12-AB.F	
CE.5476.1	Probable aldo-keto reductase 2	Cc03_g10240	Contig634	
CE.5480.1	NO HIT	NO HIT	NO HIT	NO HIT
CE.5613.1	NO HIT	NO HIT	NO HIT	putative PIFa transposase
CE.5632.1	Cellulose synthase-like protein G1	Cc00_g08520	Contig13058	
CE.5633.1	Cellulose synthase-like protein G3	Cc00_g08520	Contig13058	
CE.5642.1	Cellulose synthase-like protein G3	Cc00_g08520	Contig13058	
CE.5669.1	NO HIT	Cc09_g09650	CA00-XX-CB1-099-C08-JF.F	
CE.5696.1	NO HIT	NO HIT	CA00-XX-CL2-047-B03-QH.F	
CE.5697.1	Actin-related protein 2/3 complex subunit 4	Cc00_g31310	CA00-XX-CS1-008-H12-EZ.F	
CE.5699.1	Actin-related protein 2/3 complex subunit 4	NO HIT	CA00-XX-CL2-047-B03-QH.F	
CE.5718.1	Mitochondrial pyruvate carrier 1	Cc03_g13260	Contig14247	

CE.5722.1	Mitochondrial pyruvate carrier 1	Cc03_g13260	Contig14247	
CE.5746.1	50S ribosomal protein L15	Cc03_g13010	CA00-XX-LV4-021-B08-MC.F	
CE.5772.1	NO HIT	Cc11_g02160	NO HIT	
CE.5783.1	NO HIT	NO HIT	Contig2579	
CE.5822.1	NO HIT	Cc04_g05310	Contig12592	
CE.5877.1	L-type lectin-domain containing receptor kinase IV.1	Cc00_g01460	Contig13754	
CE.5881.1	NO HIT	Cc00_g01380	Contig7886	
CE.5906.1	NO HIT	Cc10_g10480	NO HIT	
CE.5913.1	ALA-interacting subunit 3	Cc03_g13970	Contig6390	
CE.5919.1	NO HIT	Cc02_g10860	Contig11486	
CE.5925.1	NO HIT	Cc00_g32640	Contig12928	
CE.5987.1	NO HIT	Cc03_g15330	CA00-XX-SI3-064-G08-EM.F	
CE.6087.1	NO HIT	Cc04_g00380	Contig10477	
CE.6102.1	ATPase 10, plasma membrane-type	Cc04_g00560	Contig13504	
CE.6103.1	NO HIT	NO HIT	CA00-XX-AR1-025-E11-EB.F	
CE.6135.1	NO HIT	Cc04_g00930	NO HIT	
CE.6163.1	Putative receptor-like protein kinase At1g72540	Cc04_g01240	NO HIT	
CE.6336.1	NO HIT	NO HIT	CA00-XX-SI3-012-H11-EM.F	
CE.6350.1	Amino acid transporter AVT1B	Cc01_g05610	NO HIT	
CE.6351.1	NO HIT	Cc01_g05610	NO HIT	
CE.6458.1	NO HIT	Cc04_g05110	Contig2679	
CE.6465.1	NO HIT	Cc04_g05310	Contig2044	
CE.6477.1	NO HIT	NO HIT	Contig10183	
CE.6697.1	NO HIT	NO HIT	NO HIT	NO HIT
CE.6698.1	NO HIT	Cc02_g08720	NO HIT	
CE.6994.1	Peroxisome biogenesis protein 12	Cc02_g26650	Contig11063	
CE.7152.1	NO HIT	Cc06_g11850	Contig3829	
CE.7153.1	NO HIT	NO HIT	CA00-XX-CS1-012-A06-EZ.F	
CE.7155.1	NO HIT	NO HIT	CA00-XX-IC1-003-H01-EC.F	
CE.7163.1	NO HIT	NO HIT	CA00-XX-CS1-014-B07-EQ.F	
CE.7167.1	NO HIT	NO HIT	CA00-XX-SI3-066-H09-EM.F	
CE.7189.1	NO HIT	NO HIT	NO HIT	unnamed protein product
CE.7214.1	NO HIT	Cc01_g05610	NO HIT	
CE.7280.1	Amino acid transporter AVT1B	Cc04_g15130	NO HIT	
CE.7312.1	Ribulose-phosphate 3-epimerase, chloroplastic (Fragment)	Cc05_g01370	Contig7913	
CE.7587.1	AP-1 complex subunit sigma-2	Cc10_g10040	Contig1636	
CE.7632.1	Transmembrane protein 205	Cc05_g01780	NO HIT	
CE.7685.1	Viral IAP-associated factor homolog	NO HIT	Contig14209	
CE.7691.1	NO HIT	NO HIT	NO HIT	unnamed protein product

CE.7789.1	NO HIT	Cc03_g08260	NO HIT	
CE.7824.1	Putative late blight resistance protein homolog R1B-14	Cc05_g05400	Contig16136	
CE.7915.1	Tricyclene synthase TPS4, chloroplastic	Cc05_g05760	Contig11121	
CE.7943.1	UDP-glycosyltransferase 71A16	Cc05_g05860	Contig10759	
CE.7945.1	UDP-glycosyltransferase 71A16	Cc00_g13910	CA00-XX-IA2-024-E12-EC.F	
CE.8055.1	NO HIT	Cc05_g10170	NO HIT	
CE.8144.1	Peroxisomal (S)-2-hydroxy-acid oxidase	NO HIT	Contig10052	
CE.8247.1	NO HIT	NO HIT	Contig10183	
CE.8262.1	NO HIT	Cc05_g13370	NO HIT	
CE.8393.1	Protein disulfide-isomerase 5-4	NO HIT	Contig2451	
CE.8483.1	NO HIT	Cc00_g29930	NO HIT	
CE.8540.1	NO HIT	Cc05_g16460	NO HIT	
CE.8634.1	NO HIT	Cc06_g01320	Contig17298	
CE.8744.1	Symplekin	NO HIT	Contig11195	
CE.8777.1	NO HIT	Cc06_g02520	NO HIT	
CE.8826.1	Sanguinarine reductase	Cc06_g04080	Contig8656	
CE.8950.1	ABC transporter G family member 28	Cc06_g04860	NO HIT	
CE.9003.1	NO HIT	NO HIT	Contig7355	
CE.9032.1	NO HIT	Cc06_g06310	NO HIT	
CE.9114.1	Ubiquitin-conjugating enzyme E2 variant 1C	Cc06_g08870	Contig14067	
CE.9309.1	NO HIT	NO HIT	Contig4407	
CE.9313.1	Acyl-protein thioesterase 2	Cc06_g09690	NO HIT	
CE.9386.1	Transcription repressor OFP8	Cc06_g11360	NO HIT	
CE.9505.1	NO HIT	Cc08_g07990	Contig5531	
CE.9607.1	NO HIT	Cc06_g13090	NO HIT	
CE.9670.1	Peroxidase 45	Cc06_g13990	Contig12806	
CE.9726.1	rRNA-processing protein UTP23 homolog	Cc06_g15050	Contig11910	
CE.9789.1	Eukaryotic translation initiation factor 5A-4	Cc06_g15350	Contig2627	
CE.9822.1	Actin-related protein 4	Cc06_g15350	Contig7835	
CE.9823.1	Actin-related protein 4	NO HIT	Contig7835	
CE.9827.1	NO HIT	NO HIT	NO HIT	NO HIT
CE.9828.1	NO HIT	Cc06_g13990	NO HIT	
CE.9850.1	UTP23_MOUSE rRNA-processing protein UTP23 homolog	Cc06_g15350	Contig8543	
CE.9856.1	Actin-related protein 4	NO HIT	Contig7835	
CE.9858.1	NO HIT	Cc06_g15470	NO HIT	
CE.9859.1	Class V chitinase	Cc06_g17620	NO HIT	
CE.9966.1	NO HIT	NO HIT	NO HIT	RRP12-like protein
CE.10102.1	NO HIT	Cc06_g19740	Contig13404	
CE.10117.1	NO HIT	NO HIT	NO HIT	probable disease resistance protein At1g12280
CE.10159.1	NO HIT	Cc02_g28740	NO HIT	
CE.10285.1	Putative disease resistance protein At3g14460	Cc00_g08520	Contig1236	

CE.10358.				
1	Cellulose synthase-like protein G1	NO HIT	Contig13058	
CE.10359.				
1	Cellulose synthase-like protein G3	NO HIT	Contig13058	
CE.10405.				
1	NO HIT	NO HIT	Contig13404	
CE.10416.				
1	Cellulose synthase-like protein G3	NO HIT	Contig13058	
CE.10417.				
1	Cellulose synthase-like protein G3	Cc00_g08520	Contig13058	
CE.10496.				
1	NO HIT	Cc06_g23260	NO HIT	
CE.10519.				
1	NO HIT	NO HIT	Contig10203	
CE.10587.				
1	60S ribosomal protein L7a-1	Cc00_g20240	Contig11310	
CE.10588.				
1	60S ribosomal protein L7a-1	Cc00_g20240	Contig11310	
CE.10641.				
1	NO HIT	NO HIT	CA00-XX-NS1-003-B09- EF.F	
CE.10701.				
1	NO HIT	NO HIT	NO HIT	NO HIT
CE.10781.				
1	Translation factor GUF1 homolog, mitochondrial	Cc07_g03130	Contig3404	
CE.10857.				
1	Alpha-amylase 3, chloroplastic	Cc07_g04160	NO HIT	
CE.10903.				
1	Amino acid transporter AVT6A	Cc07_g04660	Contig16483	
CE.11068.				
1	NO HIT	Cc07_g07090	NO HIT	
CE.11174.				
1	Non-specific lipid-transfer protein 2	Cc07_g08570	Contig3039	
CE.11189.				
1	NO HIT	Cc00_g35020	NO HIT	
CE.11317.				
1	NO HIT	Cc07_g10590	NO HIT	
CE.11321.				
1	Small nuclear ribonucleoprotein SmD3b	Cc00_g23360	CA00-XX-EA1-006-G03- EC.F	
CE.11332.				
1	Calcium-dependent protein kinase 28	Cc07_g10750	Contig17057	
CE.11468.				
1	NO HIT	NO HIT	CA00-XX-LV9-036-F06- BM.F	
CE.11521.				
1	Ethylene-responsive transcription factor 1B	Cc07_g13110	NO HIT	
CE.11524.				
1	Potassium channel SKOR	Cc07_g13170	CA00-XX-SS1-006-H02- EF.F	
CE.11634.				
1	Probable LRR receptor-like serine/threonine-protein kinase At5g10290	NO HIT	Contig10829	
CE.11644.				
1	Cyclin-dependent kinase C-1	Cc07_g15140	Contig10807	
CE.11702.				
1	Stress enhanced protein 2	Cc07_g16120	Contig15929	
CE.11763.				
1	NO HIT	Cc07_g16690	CA00-XX-SI3-028-F11- EM.F	
CE.11773.				
1	NO HIT	NO HIT	Contig13404	

CE.11774. 1	Thioredoxin O1	Cc09_g05440	Contig12385	
CE.11839. 1	NO HIT	NO HIT	Contig9383	
CE.11891. 1	Probable ribosome biogenesis protein RLP24	Cc01_g02050	NO HIT	
CE.11974. 1	Small nuclear ribonucleoprotein SmD3b	Cc07_g21080	Contig10422	
CE.12010. 1	NO HIT	NO HIT	NO HIT	NO HIT
CE.12020. 1	NO HIT	NO HIT	NO HIT	NO HIT
CE.12074. 1	NO HIT	NO HIT	NO HIT	small nuclear ribonucleoprotein SmD3b-like
CE.12100. 1	NO HIT	NO HIT	Contig1704	
CE.12133. 1	NO HIT	Cc00_g17650	Contig780	
CE.12341. 1	NO HIT	NO HIT	NO HIT	NO HIT
CE.12342. 1	NO HIT	NO HIT	NO HIT	NO HIT
CE.12357. 1	Cell division control protein 2 homolog A	Cc08_g09900	Contig16395	
CE.12385. 1	NO HIT	NO HIT	Contig12455	
CE.12427. 1	NO HIT	NO HIT	Contig11294	
CE.12428. 1	NO HIT	NO HIT	Contig11294	
CE.12453. 1	60S ribosomal protein L35a-3	Cc02_g06370	Contig16484	
CE.12458. 1	NO HIT	Cc02_g32150	CA00-XX-CB1-067-B04-BM.F	
CE.12517. 1	NO HIT	NO HIT	Contig4069	
CE.12525. 1	Serine/threonine-protein kinase SAPK2	Cc00_g07830	CA00-XX-RX1-020-G07-EB.F	
CE.12531. 1	Protein disulfide isomerase-like 2-2	Cc09_g08100	Contig9182	
CE.12580. 1	NO HIT	NO HIT	NO HIT	NO HIT
CE.12636. 1	NO HIT	Cc11_g14620	Contig8519	
CE.12675. 1	Premnaspirodiene oxygenase	Cc08_g07370	Contig8797	
CE.12676. 1	Premnaspirodiene oxygenase	Cc08_g07360	Contig7947	
CE.12680. 1	Putative glutathione peroxidase 7	Cc08_g07250	Contig10263	
CE.12711. 1	NO HIT	Cc00_g27080	Contig13131	
CE.12719. 1	NO HIT	Cc11_g14620	CA00-XX-CB1-016-G08-AC.F	
CE.12950. 1	Serine/threonine-protein kinase SAPK3	Cc08_g11200	Contig14386	

CE.12951. 1		NO HIT	Cc08_g11220	Contig8849
CE.13025. 1		NO HIT	NO HIT	NO HIT
CE.13211. 1		NO HIT	Cc08_g15380	NO HIT
CE.13699. 1		Protease Do-like 10, mitochondrial	Cc09_g04340	Contig2117
CE.13725. 1		Trans-resveratrol di-O-methyltransferase	Cc00_g28800	Contig12031
CE.13726. 1		NO HIT	NO HIT	NO HIT
CE.13727. 1		Probable O-methyltransferase 3	Cc05_g07890	NO HIT
CE.13728. 1		Trans-resveratrol di-O-methyltransferase	Cc05_g07890	Contig7087
CE.13763. 1		Aspartate aminotransferase, mitochondrial	Cc05_g08500	Contig2257
CE.13787. 1		11-oxo-beta-amyrin 30-oxidase	Cc05_g08840	NO HIT
CE.13791. 1		Cytochrome P450 CYP72A219	Cc05_g08840	NO HIT
CE.13831. 1		(+)-neomenthol dehydrogenase	Cc09_g05870	Contig13813
CE.13848. 1		(+)-neomenthol dehydrogenase	Cc09_g05870	CA00-XX-LV8-022-D03-QH.F
CE.13867. 1		Glutathionyl-hydroquinone reductase YqjG	Cc09_g06140	Contig10782
CE.13888. 1		NO HIT	Cc00_g31220	Contig9383
CE.13993. 1		NO HIT	Cc00_g13570	CA00-XX-LV8-024-F01-CB.F
CE.14005. 1		NO HIT	Cc00_g22400	Contig16539
CE.14018. 1		NO HIT	Cc09_g08590	NO HIT
CE.14046. 1		Protein-lysine methyltransferase METTL21E	Cc06_g21220	Contig11155
CE.14048. 1		Amino acid transporter AVT1C	Cc01_g05610	NO HIT
CE.14071. 1		NO HIT	Cc01_g05610	NO HIT
CE.14076. 1		NO HIT	Cc00_g15320	CA00-XX-FR2-014-B01-BF.F
CE.14119. 1		Trans-resveratrol di-O-methyltransferase	Cc05_g07890	Contig7087
CE.14132. 1		NO HIT	NO HIT	CA00-XX-S13-051-C02-EM.F
CE.14200. 1		NO HIT	NO HIT	NO HIT
CE.14225. 1		NO HIT	NO HIT	NO HIT
CE.14304. 1		Pentatricopeptide repeat-containing protein At2g03880	Cc00_g35320	NO HIT
CE.14414. 1		Enoyl-CoA hydratase 2, peroxisomal	Cc10_g00990	Contig16153

CE.14415. 1	Enoyl-CoA hydratase 2, peroxisomal	Cc10_g00990	Contig16153	
CE.14416. 1	Enoyl-CoA hydratase 2, peroxisomal	Cc10_g00990	Contig16153	
CE.14697. 1	NO HIT	NO HIT	CA00-XX-CS1-061-H10-AC.F	
CE.14773. 1	NO HIT	NO HIT	NO HIT	NO HIT
CE.14860. 1	NO HIT	NO HIT	NO HIT	NO HIT
CE.14874. 1	NO HIT	NO HIT	NO HIT	uncharacterized protein LOC113713716
CE.14935. 1	NO HIT	Cc03_g06690	CA00-XX-CS1-110-D03-EQ.F	
CE.15001. 1	NO HIT	Cc00_g19520	CA00-XX-CA1-033-B08-BM.F	
CE.15010. 1	Nuclear cap-binding protein subunit 2	Cc10_g09490	Contig2326	
CE.15078. 1	NO HIT	Cc00_g16170	Contig3788	
CE.15085. 1	Bifunctional epoxide hydrolase 2	Cc05_g03980	CA00-XX-SI3-095-C08-EM.F	
CE.15143. 1	NO HIT	NO HIT	Contig2079	
CE.15301. 1	Pentatricopeptide repeat-containing protein At5g46460	Cc10_g13020	Contig2512	
CE.15311. 1	NO HIT	NO HIT	CA00-XX-SI3-051-C02-EM.F	
CE.15354. 1	NO HIT	Cc00_g28460	CA00-XX-CB1-026-E05-AG.F	
CE.15356. 1	NO HIT	Cc00_g28460	CA00-XX-CB1-026-E05-AG.F	
CE.15357. 1	NO HIT	Cc00_g28460	CA00-XX-CB1-026-E05-AG.F	
CE.15358. 1	NO HIT	Cc00_g28460	CA00-XX-CB1-026-E05-AG.F	
CE.15367. 1	NO HIT	Cc03_g12780	Contig15733	
CE.15369. 1	NO HIT	NO HIT	NO HIT	NO HIT
CE.15381. 1	NO HIT	NO HIT	Contig16802	
CE.15390. 1	NO HIT	Cc10_g14940	CA00-XX-RX1-029-A12-EB.F	
CE.15422. 1	NO HIT	Cc10_g15440	Contig460	
CE.15502. 1	NO HIT	Cc05_g01000	CA00-XX-CL2-016-D04-AC.F	
CE.15536. 1	Peroxisome biogenesis protein 12	Cc06_g14670	Contig2579	
CE.15585. 1	NO HIT	NO HIT	NO HIT	NO HIT
CE.15589. 1	NO HIT	NO HIT	CA00-XX-CB1-007-B10-CE.F	
CE.15595. 1	NO HIT	NO HIT	NO HIT	MDIS1-interacting receptor like kinase 2

CE.15603.	1	Telomerase reverse transcriptase	NO HIT	CA00-XX-EA1-049-F02-EC.F	
CE.15633.	1	3-hexulose-6-phosphate isomerase	Cc03_g13850	Contig12253	
CE.15637.	1	Photosystem I P700 chlorophyll a apoprotein A2	NO HIT	NO HIT	photosystem I P700 apoprotein A2
CE.15654.	1	NO HIT	NO HIT	NO HIT	NO HIT
CE.15660.	1	TERT_ARATH Telomerase reverse transcriptase	NO HIT	CA00-XX-EA1-049-F02-EC.F	
CE.15696.	1	UDP-glycosyltransferase 71A15	Cc00_g07170	NO HIT	
CE.15833.	1	Ras-related protein RABE1c	Cc00_g22400	Contig16539	
CE.15838.	1	Rab escort protein 1	Cc11_g03220	CA00-XX-FB1-106-G08-AC.F	
CE.15859.	1	NO HIT	NO HIT	Contig9383	
CE.15888.	1	Anthranilate synthase beta subunit 1	NO HIT	Contig2118	
CE.15911.	1	RCD1_ARATH Inactive poly	Cc08_g03260	Contig6615	
CE.15919.	1	Dolichol-phosphate mannose synthase subunit 2	Cc08_g03400	CA00-XX-RX1-049-E08-EB.F	
CE.15930.	1	NO HIT	NO HIT	NO HIT	NO HIT
CE.15975.	1	NO HIT	NO HIT	Contig8519	
CE.15979.	1	NO HIT	NO HIT	Contig4156	
CE.16017.	1	Peroxisome biogenesis protein 12	Cc06_g14670	Contig2579	
CE.16023.	1	26S proteasome non-ATPase regulatory subunit 1 homolog A	Cc03_g07730	Contig3829	
CE.16160.	1	NO HIT	Cc11_g09710	Contig5862	
CE.16171.	1	NO HIT	Cc04_g13280	CA00-XX-NS1-001-G07-EF.F	
CE.16245.	1	NO HIT	NO HIT	NO HIT	unnamed protein product
CE.16403.	1	NO HIT	NO HIT	NO HIT	NO HIT
CE.16717.	1	Phosphatidylinositol transfer protein 3	NO HIT	Contig3606	

Fonte: o próprio autor.

## Anexo B – Genes Diferencialmente expressos em folhas e frutos de *C. eugenioides*.

Gene ID	log2FoldChange	pvalue	Expressão em Folha	Expressão em Fruto	NCBI nr
CE.7407	-4,8087	0,00059773	684.01	4.67684	BURP domain-containing protein 16-like
CE.7551	-4,7044	0,00075236	863.552	6.59771	acidic endochitinase-like
CE.384	-4,6980	0,00077526	466.731	3.68934	BURP domain-containing protein 16-like
CE.15798	4,7077	0,00079990	2.26143	291.271	serine carboxypeptidase II-2
CE.11611	-4,6284	0,00080237	4564.97	43.936	NO HIT
CE.7884	-4,6242	0,00082196	934.589	8.94054	glucose-6-phosphate/phosphate translocator 2. chloroplastic-like
CE.14025	-4,5883	0,00088449	3148.25	31.7736	cysteine proteinase inhibitor 4-like
CE.387	-4,6444	0,00100256	232.533	1.69383	BURP domain-containing protein 16-like
CE.9588	-4,5158	0,00106166	810.674	8.86006	auxin-binding protein ABP19a-like
CE.12594	-4,6360	0,00106192	200.336	1.51506	glucose-1-phosphate adenyltransferase small subunit. chloroplastic/amyloplastic
CE.10685	4,8622	0,00107244	0	896.853	oleosin 5-like
CE.7991	-4,8597	0,00110299	139.545	0	photosystem II protein D1 (chloroplast)
CE.12310	4,8168	0,00120252	0	688.451	cytochrome P450 71A1-like
CE.2927	4,3880	0,00128891	24.7574	1690.12	protein DOG1-like 4
CE.10997	-4,5480	0,00130073	107.124	0.853545	putative lipase
CE.341	4,4042	0,00153219	0.325947	955.033	NO HIT
CE.12743	4,3428	0,00159554	3.74668	263.423	ferruginol synthase-like
CE.6645	4,6986	0,00161865	0.000421736	384.425	unnamed protein product
CE.11941	4,2398	0,00182536	9.81142	567.986	alcohol dehydrogenase-like 7
CE.15182	-4,3275	0,00183607	263.081	3.41373	kunitz trypsin inhibitor 2
CE.2701	-4,4185	0,00185299	61.212	0.541853	alpha-farnesene synthase-like isoform X1
CE.5671	-4,6303	0,00195830	71.2304	4.3431	cytochrome P450 71A4-like
CE.12213	4,4031	0,00196375	1.93271	209.502	putative cytochrome c oxidase subunit 6b-like
CE.10144	-4,2459	0,00196489	342.54	5.31482	BURP domain-containing protein 16-like
CE.12765	4,2598	0,00201758	9.34708	592.851	ferruginol synthase-like
CE.2967	-4,1534	0,00203149	5363.63	125.472	unnamed protein product
CE.6839	4,1831	0,00211357	39.1724	2151.81	BURP domain-containing protein 16-like
CE.8728	4,5901	0,00211827	0	233.815	protein SLE1
CE.14850	-4,5976	0,00212304	470.435	0	7-deoxyloganetin glucosyltransferase-like
CE.12822	-4,3803	0,00218548	107.379	0.893661	BURP domain protein RD22-like
CE.9608	4,3987	0,00220317	1.13157	129.979	probable 6-phosphogluconolactonase 4. chloroplastic
CE.10222	4,3971	0,00221088	0.860572	110.061	2-hydroxyisoflavanone dehydratase-like
CE.4071	-4,5803	0,00221502	58.1425	0	ribulose 1.5-bisphosphate carboxylase/oxygenase large subunit (chloroplast)
CE.5686	-4,5581	0,00233832	148.279	0	kirola-like
CE.7696	4,4369	0,00235815	0.22626	453.563	pectinesterase 2
CE.14510	-4,2547	0,00242715	89.1891	1.09902	protein WALLS ARE THIN 1

CE.16534	-4,0586	0,00243200	3760.94	79.6126	galactinol synthase 1
CE.2292	-4,0991	0,00254724	249.233	4.67414	sedoheptulose-1.7-bisphosphatase. chloroplastic
CE.14654	-4,2297	0,00256804	100.576	1.17145	BRI1 kinase inhibitor 1-like
CE.6929	-4,2633	0,00264865	64.2277	0.707831	Cytochrome P450 71A6
CE.2107	4,0317	0,00274858	13.7411	610.886	oleosin 1
CE.6828	4,2156	0,00279408	1.12258	86.8087	protein SHOOT GRAVITROPISM 5-like
CE.661	-3,9990	0,00279782	3537.08	79.5744	cysteine proteinase inhibitor CPI-4
CE.12872	-4,2703	0,00281825	128.581	1.23401	aquaporin TIP4-1
CE.9878	4,0339	0,00286995	13.8561	633.38	18 kDa seed maturation protein-like
CE.9799	-4,0274	0,00288124	236.138	4.85053	CBL-interacting serine/threonine-protein kinase 6-like
CE.8790	4,1053	0,00309305	6.90585	404.551	17.5 kDa class I heat shock protein
CE.15016	-3,9468	0,00309585	2591.84	62.122	cysteine proteinase inhibitor CPI-4
CE.16038	3,9869	0,00311687	6.72625	288.389	endoglucanase 8
CE.2838	3,9984	0,00315327	8.05001	357.23	ABA-inducible protein PHV A1-like
CE.11670	-4,1165	0,00317360	107.037	1.13221	unnamed protein product
CE.7900	-4,4300	0,00318347	101.943	0	B-box domain protein 30
CE.16060	3,9118	0,00321815	44.198	2701.17	kirola-like
CE.11093	-4,1711	0,00326852	593.202	0.716193	squalene monooxygenase
CE.11197	-3,9948	0,00327201	158.979	3.30888	3-ketoacyl-CoA synthase 6
CE.6313	-3,9153	0,00331823	1316.88	32.5906	acid phosphatase 1-like
CE.8580	-3,9599	0,00334483	617.922	13.805	Photosystem I reaction center subunit V. chloroplastic
CE.5731	-4,4013	0,00340722	913.075	0	cytochrome P450 71A4-like
CE.13338	-3,8895	0,00342260	655.246	16.9	granule-bound starch synthase 1. chloroplastic/amyloplastic
CE.9409	4,1225	0,00342353	0	231.239	bark storage protein A
CE.9763	3,9674	0,00345292	9.51364	343.928	secoisolariciresinol dehydrogenase-like
CE.4679	-3,9857	0,00345352	113.18	2.32965	cytochrome P450 78A3-like
CE.4570	-3,9221	0,00346076	214.275	5.12126	ferric reduction oxidase 6-like
CE.10023	4,1916	0,00356263	0.809419	841.476	probable F-box protein At5g04010
CE.2108	3,8686	0,00358539	78.125	2768.32	oleosin 1
CE.9025	4,0455	0,00367391	4.04936	228.99	unnamed protein product
CE.15628	3,8972	0,00375069	9.13323	355.772	17.8 kDa class I heat shock protein-like
CE.2587	-4,0600	0,00375093	100.448	1.56436	bark storage protein A-like
CE.8031	3,8926	0,00375375	6.71557	256.935	zinc finger CCCH domain-containing protein 2-like
CE.9983	3,9154	0,00379190	4.28663	173.962	zeatin O-glucosyltransferase-like
CE.3084	-3,8976	0,00384846	951.602	22.6968	protein RADIALIS-like 4
CE.4228	3,9318	0,00386083	16.0636	685.914	pathogenesis-related protein PR-4-like
CE.3480	3,8334	0,00388277	26.6077	910.513	glutelin type-A 3 isoform X1
CE.11222	3,8203	0,00391400	21.6132	722.445	vicilin-like seed storage protein At2g18540
CE.13285	-3,8761	0,00392902	229.141	5.61277	glutaredoxin domain-containing cysteine-rich protein 1
CE.4344	3,8823	0,00394508	10.4362	400.94	NADPH-dependent aldehyde reductase 1. chloroplastic-like

CE.7487	4,3305	0,00394618	0	116.939	em-like protein GEA6
CE.11845	4,1453	0,00396591	4.50482	421.183	NO HIT
CE.11138	3,9224	0,00399285	10.4859	447.097	cold-regulated protein
CE.7621	3,9687	0,00402268	6.9537	340.316	NO HIT
CE.376	-4,2296	0,00403325	125.721	0.686025	putative chitinase
CE.4338	3,8361	0,00404670	11.5567	405.992	basic 7S globulin-like
CE.15449	-3,8407	0,00417280	293.208	7.60811	MADS-box protein SVP
CE.9245	3,8433	0,00417793	8.83629	273.154	36.4 kDa proline-rich protein-like
CE.11336	-3,8207	0,00421831	349.894	9.43089	ferredoxin--NADP reductase. leaf-type isozyme. chloroplastic
CE.12311	3,9554	0,00428067	4.01778	197.406	cytochrome P450 71A1-like
CE.15212	-3,7842	0,00430270	745.842	16.2951	photosystem II 22 kDa protein. chloroplastic
CE.323	-3,8092	0,00435318	141.073	3.8363	linoleate 13S-lipoxygenase 2-1. chloroplastic-like
CE.1650	3,9094	0,00444719	2.52771	112.259	ethylene-responsive transcription factor ERF071-like
CE.13932	4,0883	0,00451143	0.425782	363.832	cannabidiolic acid synthase-like 2
CE.16059	3,7326	0,00457145	140.919	5723.37	kirola-like
CE.5377	3,7237	0,00460019	4.78059	19.2351	unnamed protein product
CE.5156	3,7880	0,00462828	15.7325	533.475	peroxygenase-like
CE.6366	3,8882	0,00464559	2.2981	99.5309	unnamed protein product
CE.12699	3,7363	0,00468279	10.171	665.078	late embryogenesis abundant protein D-34-like
CE.7490	3,7512	0,00471383	39.1506	1229.81	em-like protein GEA6
CE.13437	-3,8495	0,00477773	59.044	1.38663	oligopeptide transporter 3
CE.5475	3,9760	0,00477892	5.11276	298.038	oil body-associated protein 1A
CE.15629	3,7260	0,00478418	45.2379	1365.29	17.8 kDa class I heat shock protein-like
CE.9507	-3,9126	0,00479024	90.6544	1.79345	putative GATA transcription factor 22
CE.725	3,9275	0,00483422	0.526445	749.412	secoisolariciresinol dehydrogenase-like
CE.4772	3,8158	0,00490789	6.71873	193.791	outer envelope pore protein 16-2. chloroplastic
CE.11734	4,1127	0,00503517	0.275117	369.021	ethylene-responsive transcription factor ABR1-like
CE.15855	3,8771	0,00507525	1.14259	51.3651	protein indeterminate-domain 5. chloroplastic-like
CE.4457	3,7241	0,00507905	21.9989	682.096	universal stress protein A-like protein isoform X2
CE.2111	4,2189	0,00510104	0	408.692	phenylalanine N-monooxygenase
CE.4717	3,7961	0,00511172	4.52317	209.575	1-Cys peroxiredoxin
CE.10796	3,7044	0,00512741	45.5197	1361.19	non-specific lipid-transfer protein A-like
CE.11528	-3,8130	0,00515905	96.2179	2.35422	aspartic proteinase PCS1-like
CE.3124	3,7129	0,00516623	27.236	831.871	non-symbiotic hemoglobin 1
CE.15988	3,6627	0,00516678	65.4864	372.494	thaumatin-like protein
CE.15588	-4,0025	0,00519675	419.939	6.40344	acidic endochitinase-like
CE.7917	4,2079	0,00522957	0	605.511	isoprene synthase
CE.5115	3,6527	0,00524443	70.8636	1054.3	legumin B-like
CE.4601	-3,9573	0,00529395	424.672	0.660061	ferruginol synthase-like
CE.6659	-3,7354	0,00531029	167.465	4.79658	protein ASPARTIC PROTEASE IN GUARD CELL 1-like

CE.4227	3,6768	0,00536269	63.0205	1820.12	pathogenesis-related protein PR-4-like
CE.8702	3,7234	0,00537362	6.6804	100.314	aldose reductase
CE.12813	3,6318	0,00544109	229.195	6100.9	BURP domain protein RD22
CE.11554	3,7730	0,00564850	1.79747	66.1743	probable rhamnogalacturonate lyase B
CE.1500	-3,7772	0,00565079	64.8254	1.63344	sulfate transporter 3.1-like
CE.4502	-3,7438	0,00570765	124.498	3.37867	chloroplast stem-loop binding protein of 41 kDa b. chloroplastic
CE.1071	3,9808	0,00574539	1.63958	126.543	cell wall integrity and stress response component 1
CE.4704	3,6125	0,00575216	60.1999	1580.64	NO HIT
CE.15183	-3,6899	0,00575308	457.354	14.0223	kunitz trypsin inhibitor 2
CE.15750	-4,1682	0,00581810	413.365	0	chaperone protein dnaJ C76. chloroplastic-like
CE.10965	3,6741	0,00582052	1.16059	151.721	reticulon-like protein B13
CE.12944	3,6570	0,00582518	15.2077	441.186	17.1 kDa class II heat shock protein-like
CE.11278	3,5931	0,00584694	357.447	9108.16	dehydrin DH1a
CE.9023	3,9226	0,00585172	1.25069	72.2679	unnamed protein product
CE.2686	-3,6027	0,00585980	1366.33	49.1702	thiamine thiazole synthase 2. chloroplastic
CE.4044	-3,7660	0,00589878	84.775	2.13476	G-type lectin S-receptor-like serine/threonine-protein kinase RLK1
CE.13649	3,7864	0,00593533	2.02524	79.952	seipin-1
CE.15485	3,7708	0,00595883	2.8942	109.931	GDSL esterase/lipase At1g71250
CE.1958	3,5852	0,00598402	144.788	3666.08	unnamed protein product
CE.2345	3,5894	0,00604119	23.3113	596.073	expansin-A4
CE.9766	3,6337	0,00607566	6.16002	172.781	heat shock 70 kDa protein
CE.1873	3,7170	0,00610832	2.48325	83.8671	probable trehalose-phosphate phosphatase 2
CE.9173	3,6398	0,00612670	8.83587	242.354	universal stress protein YxiE-like isoform X1
CE.4396	3,8621	0,00614628	1.26552	64.9965	NO HIT
CE.5396	3,8060	0,00615412	1.783	765.978	crocetin glucosyltransferase. chloroplastic-like
CE.6500	3,6244	0,00617048	23.2058	643.461	unnamed protein product
CE.7787	-3,6109	0,00618962	333.761	11.618	aspartyl protease family protein At5g10770-like
CE.6972	3,5876	0,00629116	118.084	3070.92	aldehyde dehydrogenase family 2 member C4
CE.10947	3,6312	0,00633611	13.2668	377.899	PREDICTED: uncharacterized protein LOC104116127
CE.11421	-3,5935	0,00636241	277.342	9.80975	catalase isozyme 1
CE.9246	3,5982	0,00638134	8.93498	236.99	36.4 kDa proline-rich protein-like
CE.5793	3,9330	0,00639120	0.788065	674.612	gibberellin 20 oxidase 1-D-like
CE.8917	-3,9573	0,00642336	37.283	0.423983	Knob-associated histidine-rich protein
CE.8385	3,6433	0,00642687	19.243	566.353	2S albumin
CE.3829	3,7762	0,00642704	5.03746	108.547	unnamed protein product
CE.1199	3,6637	0,00644680	6.11084	188.409	expansin
CE.8304	3,5471	0,00648724	76.4573	1864.99	stem-specific protein TSJT1-like
CE.4869	-3,8631	0,00651209	144.884	2.48387	monothiol glutaredoxin-S2
CE.10431	3,8348	0,00651604	2.00479	93.7747	cysteine-rich repeat secretory protein 55
CE.6515	3,6504	0,00652409	0	154.814	ethylene-responsive transcription factor ERF114-like

CE.7887	3,5739	0,00656255	9.98629	257.614	Isocitrate lyase
CE.4458	3,6801	0,00658396	1.12055	36.325	histidine kinase CK11-like
CE.16159	-3,8543	0,00664101	185.317	3.20674	non-specific lipid-transfer protein 2-like
CE.11590	3,5208	0,00664474	124.107	3.20674	NO HIT
CE.15868	-3,7978	0,00665642	693.567	13.6801	isoflavone reductase homolog A622-like
CE.6894	3,5331	0,00670973	38.3895	924.933	probable pectate lyase 18
CE.10024	-3,5403	0,00673795	1904.94	72.8641	lipid transfer protein precursor
CE.10881	3,9832	0,00673912	1.18679	137.465	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit like
CE.13627	3,5291	0,00680150	28.8935	411.15	cilia- and flagella-associated protein 251 isoform X1
CE.12704	3,7542	0,00687688	1.38109	56.0573	dual specificity protein kinase splA
CE.5872	-3,8025	0,00696050	49.5128	1.0202	Cytochrome P450 71A6
CE.6780	3,5376	0,00696871	78.6417	1941.43	metallothionein-like protein 4B
CE.10784	3,6953	0,00697056	1.59936	55.8562	polyvinylalcohol dehydrogenase-like
CE.14363	-3,5637	0,00697131	246.769	8.88246	glyceraldehyde-3-phosphate dehydrogenase B. chloroplastic
CE.5364	-3,7764	0,00697191	526.125	11.2718	unnamed protein product
CE.7947	3,6208	0,00703303	3.81773	112.329	aromatic-L-amino-acid decarboxylase-like
CE.9064	-3,9132	0,00709138	672.685	0.801226	protein BIC1
CE.7544	3,8334	0,00711153	1.36237	75.157	probable 2-oxoglutarate-dependent dioxygenase AOP1
CE.12365	3,7623	0,00712227	3.72771	160.069	oleosin 1
CE.11357	3,5601	0,00715672	17.6969	462.165	heavy metal-associated isoprenylated plant protein 26
CE.4685	-3,6358	0,00720664	64.6201	1.96084	transcription factor MYC2-like
CE.15940	3,7307	0,00722779	6.65213	263.173	NO HIT
CE.12734	3,6181	0,00724580	3.83747	115.356	geraniol 8-hydroxylase-like
CE.3570	3,6507	0,00727017	2.35787	79.4139	pectinesterase 2
CE.7418	3,7063	0,00728601	7.83871	283.876	endochitinase EP3-like
CE.16039	3,4996	0,00731568	36.2032	848.902	21 kDa protein-like
CE.15305	3,5123	0,00738067	15.9278	384.013	alcohol dehydrogenase
CE.2188	3,7700	0,00749371	5.68509	263.614	HIG1 domain family member 1B like
CE.1530	3,5506	0,00750477	4.99864	131.038	basic 7S globulin-like
CE.5521	3,4802	0,00751107	22.1595	506.635	O-acyltransferase WSD1-like
CE.10995	-3,4717	0,00752012	346.634	4.86429	putative lipase
CE.15786	-3,5670	0,00752113	68.4877	2.36056	protein CHLOROPLAST IMPORT APPARATUS 2-like
CE.15959	3,7662	0,00755989	2.14683	99.1913	NO HIT
CE.4389	3,4971	0,00757403	24.1734	572.458	probable xyloglucan endotransglucosylase/hydrolase protein
CE.8038	3,4645	0,00759858	0.0407564	148.699	protein E6-like
CE.16616	3,6529	0,00761248	1.36598	46.7979	cytochrome P450 71A1-like
CE.8922	-3,7831	0,00775938	43.642	0.815315	NAC domain-containing protein 2-like
CE.11547	3,6956	0,00778016	5.22915	187.216	protein RSI-1
CE.9723	3,6171	0,00778369	8.78704	271.457	EG45-like domain containing protein
CE.14996	3,5628	0,00780836	6.37928	175.118	11-beta-hydroxysteroid dehydrogenase-like 5

CE.3904	-3,8699	0,00781543	427.375	0.531822	NO HIT
CE.7632	3,6137	0,00784819	2.06002	65.6685	unnamed protein product
CE.2168	3,4386	0,00786314	107.43	2318.7	expansin
CE.7955	3,7860	0,00787459	3.38269	176.712	Wound-responsive family protein
CE.3069	3,5157	0,00794563	13.8856	348.55	embryogenic cell protein 40
CE.5920	3,7098	0,00797994	1.97492	74.2813	putative receptor-like protein kinase At1g72540
CE.4041	-4,0227	0,00799733	336.624	0	annexin D4-like
CE.9024	3,7087	0,00801010	2.79831	117.315	unnamed protein product
CE.8029	-3,9238	0,00804311	438.561	0.336416	squamosa promoter-binding protein 1-like
CE.12818	-3,5058	0,00805322	248.78	9.39728	BURP domain protein RD22-like
CE.11194	-3,4208	0,00805637	5938.3	262.991	ribulose biphosphate carboxylase small chain. chloroplastic
CE.4129	3,9002	0,00809175	0.343573	363.172	putative B3 domain-containing protein Os03g0621600
CE.4744	3,5703	0,00810150	5.9809	170.377	cell wall / vacuolar inhibitor of fructosidase 1-like
CE.5474	3,5954	0,00813991	11.7699	356.463	oil body-associated protein 1A
CE.2648	3,4655	0,00822675	6.4997	150.21	transketolase. chloroplastic-like
CE.11217	-3,8451	0,00823001	250.597	0.321506	anthocyanidin 3-O-glucosyltransferase 5-like
CE.6378	3,4106	0,00823173	281.911	5890.49	late embryogenesis abundant protein Dc3-like
CE.13615	3,6002	0,00826408	3.97605	122.994	dehydration-responsive element-binding protein 2D-like
CE.16037	3,7571	0,00839207	2.88884	146.505	protein PLANT CADMIUM RESISTANCE 2-like
CE.3338	3,5112	0,00839809	3.93894	100.699	(-)-alpha-terpineol synthase
CE.3299	-3,4972	0,00842167	172.851	6.50315	malate dehydrogenase. glyoxysomal
CE.8403	3,6549	0,00850937	5.6951	208.067	NO HIT
CE.9174	3,4191	0,00855654	34.1359	735.268	universal stress protein A-like protein
CE.15108	3,4162	0,00857550	49.0273	993.199	NO HIT
CE.5336	-3,5541	0,00864107	204.27	6.70827	glycine cleavage system H protein. mitochondrial
CE.6046	3,4661	0,00874373	9.52278	226.509	NO HIT
CE.5867	-3,6963	0,00875622	68.8188	1.51975	BURP domain-containing protein 16-like
CE.3852	3,3870	0,00876025	49.5602	1016.75	DUF642 domain-containing protein
CE.357	3,5772	0,00889725	4.23068	126.267	endochitinase EP3-like
CE.2589	-3,9691	0,00896455	279.045	0	bark storage protein A-like
CE.14852	-3,6812	0,00903782	63.7273	1.42556	7-deoxyloganetin glucosyltransferase-like
CE.3086	-3,6335	0,00904118	186.592	4.84808	protein RADIALIS-like 1
CE.7351	3,7204	0,00904278	0.3977	206.585	NO HIT
CE.6594	3,3638	0,00906824	11.5433	52.4334	beta-glucosidase 44-like
CE.14241	3,8470	0,00908589	1.06224	106.151	hypothetical protein CDL12_10429
CE.9877	3,3594	0,00908605	189.312	3750.55	18 kDa seed maturation protein-like
CE.9609	3,9544	0,00911033	0	594.666	probable 6-phosphogluconolactonase 4. chloroplastic
CE.11668	-3,9576	0,00918326	316.838	0	annexin D2-like
CE.358	3,5339	0,00923381	5.03191	148.818	endochitinase EP3-like
CE.10684	3,3864	0,00927025	0	699.617	oleosin 5-like

CE.12035	3,6119	0,00928283	1.2091	42.0886	anthocyanidin 3-O-glucosyltransferase
CE.16443	3,3557	0,00932969	26.5024	526.815	protein NRT1/ PTR FAMILY 2.13
CE.16634	3,7067	0,00934999	0.654586	314.538	Ankyrin repeat
CE.214	-3,6912	0,00947020	584.435	0	protein YLS3
CE.12742	3,6230	0,00958262	1.08706	37.7841	ferruginol synthase-like
CE.11830	3,9284	0,00962357	0	434.835	unnamed protein product
CE.15004	-3,9341	0,00964670	184.806	0	alkane hydroxylase MAH1-like
CE.16720	-3,4982	0,00968633	47.3809	1.65036	protein NRT1/ PTR FAMILY 6.1
CE.15579	3,5170	0,00977444	0	126.378	quinone oxidoreductase PIG3
CE.15731	-3,3497	0,00984549	394.641	18.3511	glyceraldehyde-3-phosphate dehydrogenase A. chloroplastic
CE.3583	-3,4960	0,00988573	126.58	4.37716	SPX domain-containing protein 3
CE.13663	-3,6691	0,00992552	368.252	0.772732	flavonol sulfotransferase-like
CE.16486	3,5091	0,00995552	1.91348	53.118	UDP-glycosyltransferase 74B1
CE.43	3,6347	0,01000530	1.89824	78.8559	probable caffeine synthase 4 isoform X2
CE.1152	3,4088	0,01010156	3.83027	87.0668	inactive receptor-like serine/threonine-protein kinase At2g40270
CE.7786	-3,3922	0,01017388	134.79	5.75763	aspartyl protease family protein At5g10770-like
CE.2972	3,4636	0,01020330	5.02303	128.91	oil body-associated protein 2A-like
CE.5012	-3,5948	0,01026110	41.2641	1.07206	VAN3-binding protein isoform X1
CE.747	-3,6516	0,01028419	65.0886	1.39744	NAD(P)H-quinone oxidoreductase subunit N. chloroplastic
CE.14517	-3,4399	0,01030038	60.9665	2.33077	jacalin-related lectin 3
CE.10510	-3,6150	0,01042067	66.5497	1.59824	NO HIT
CE.4853	-3,4120	0,01043643	268.843	0.943177	probable protein phosphatase 2C 40
CE.3083	3,4421	0,01047497	1.32783	14.5856	transcription factor bHLH63-like
CE.3902	3,5807	0,01048661	1.69951	64.3075	aldose 1-epimerase-like
CE.11750	3,3847	0,01058092	5.07308	110.712	cytochrome P450 71A1-like
CE.14062	3,8826	0,01058710	0	145.739	7-deoxyloganetic acid glucosyltransferase-like
CE.10193	-3,2857	0,01064438	1194.93	60.2952	chlorophyll a-b binding protein of LHCII type 1
CE.7988	3,5483	0,01064970	0	397.717	cytochrome P450 71D9
CE.305	-3,8848	0,01068617	360.182	0	rust resistance kinase Lr10-like
CE.14047	3,6422	0,01072868	0.676496	303.967	probable pectinesterase/pectinesterase inhibitor 17
CE.6840	3,2735	0,01073258	39.64	719322	malate synthase. glyoxysomal
CE.6879	3,4775	0,01091479	2.7128	72.1221	cytochrome P450 82C4-like
CE.8111	-3,2975	0,01098102	466.457	22.7085	glucose-1-phosphate adenylyltransferase small subunit 2. chloroplastic
CE.4927	-3,3230	0,01098988	170.571	7.96952	glutamine synthetase leaf isozyme. chloroplastic
CE.16026	3,2620	0,01109645	277.869	4986.11	defensin-like protein P322
CE.6980	3,3641	0,01110175	17.5206	381.326	oleosin 18.2 kDa-like
CE.14671	3,6790	0,01111184	1.20141	67.1959	protein RALF-like 34
CE.11309	-3,4259	0,01119942	160.917	6.0486	lactoylglutathione lyase-like
CE.9721	3,2645	0,01127190	86.2338	1569.67	EG45-like domain containing protein
CE.7255	-3,8588	0,01127388	231.959	0	L-type lectin-domain containing receptor kinase IV.1-like

CE.11568	3,3956	0,01130438	3.29669	77.3101	putative UDP-rhamnose:rhamnosyltransferase 1
CE.9084	3,5395	0,01140785	0.909828	308.424	beta-fructofuranosidase. insoluble isoenzyme 1-like
CE.7899	3,8436	0,01147189	0	178.285	thaumatin-like protein
CE.9896	3,3253	0,01168835	0.253633	613.693	transcription factor HBP-1b(c38)-like isoform X2
CE.12314	3,8336	0,01171116	0	167.607	flavonoid 3'-monooxygenase-like
CE.1592	-3,3704	0,01181331	201.872	8.30224	monothiol glutaredoxin-S11
CE.6281	3,3895	0,01181965	6.70611	159.168	probable Xaa-Pro aminopeptidase P
CE.13921	-3,4681	0,01183636	45.5599	1.4931	berberine bridge enzyme-like 13
CE.11241	3,2212	0,01196825	53.9431	796.267	Submandibular gland secretory Glx-rich protein like
CE.5150	3,3455	0,01197878	2.17677	47.406	Protein transport protein like
CE.9574	3,2456	0,01197923	8.3498	150.443	subtilisin-like protease SBT1.8
CE.13128	3,2238	0,01199323	70.9802	1233.3	thaumatin-like protein
CE.3605	-3,4199	0,01200584	88.9381	3.24958	probable 2-oxoglutarate-dependent dioxygenase AOP1
CE.8557	-3,4427	0,01207042	557.808	0.143958	fructose-bisphosphate aldolase. cytoplasmic isozyme 1
CE.935	3,4808	0,01220537	1.58751	48.2439	NO HIT
CE.10207	3,5799	0,01221198	1.21884	51.2471	NO HIT
CE.13259	3,3949	0,01221430	14.5274	358.304	probable protein phosphatase 2C 51
CE.9798	3,2372	0,01223261	12.3317	213.245	CBL-interacting serine/threonine-protein kinase 6-like
CE.12827	-3,2257	0,01224943	513.805	27.4178	BURP domain protein RD22-like
CE.15871	-3,4505	0,01225576	240.654	7.60183	isoflavone reductase-like protein
CE.1677	-3,2057	0,01233747	1841.16	101.232	calvin cycle protein CP12-2. chloroplastic-like
CE.9899	3,3395	0,01242810	7.18699	157.611	polygalacturonase inhibitor-like
CE.7057	3,2597	0,01260096	8.44684	159.386	probable inactive poly [ADP-ribose] polymerase SRO5
CE.702	3,3648	0,01267865	1.53401	35.9224	subtilisin-like protease SBT1.5
CE.11598	-3,4323	0,01273123	43.8515	1.48994	protein DETOXIFICATION 27-like
CE.14013	3,7917	0,01275026	0	201.365	IQ motif. EF-hand binding site
CE.16605	-3,2467	0,01275137	414.872	20.9539	photosystem II core complex proteins psbY. chloroplastic
CE.15046	-3,3683	0,01281205	49.2007	1.94417	MADS-box protein TM3 subfamily
CE.3631	3,3702	0,01283122	14.0601	335.777	non-specific lipid-transfer protein A-like
CE.15799	3,6099	0,01283789	0.431891	224.679	PREDICTED: uncharacterized protein LOC104242831
CE.14718	3,4548	0,01284832	3.85023	109.173	F-box-like/WD repeat-containing protein TBL1XR1
CE.9308	3,2357	0,01285219	52.0266	947.205	early nodulin-93-like
CE.6973	3,2487	0,01293373	20.8153	389.196	aldehyde dehydrogenase family 2 member C4
CE.2335	-3,1594	0,01294659	802.708	46.896	ribulose bisphosphate carboxylase/oxygenase activase. chloroplastic isoform X1
CE.11991	3,2715	0,01295056	8.13884	124.419	probable aquaporin TIP3-2
CE.2787	3,1975	0,01297125	11.7142	200.034	polyphenol oxidase. chloroplastic-like
CE.5665	-3,7889	0,01298584	272.307	0	cytochrome P450 71A3-like
CE.12054	-3,4590	0,01298690	206.411	6.58836	protein ASPARTIC PROTEASE IN GUARD CELL 1-like
CE.6113	3,2090	0,01300721	10.0003	174.433	pectinesterase/pectinesterase inhibitor U1
CE.10801	-3,5035	0,01313578	327.138	0.88237	vacuolar amino acid transporter 1

CE.9061	3,2995	0,01314772	3.49849	3.84729	protein abscisic acid-insensitive 5
CE.14428	-3,2328	0,01322056	207.994	10.6127	thioredoxin-like protein CDSP32. chloroplastic
CE.14438	3,4687	0,01322401	2.23515	70.5272	expansin-A11-like
CE.16677	3,3539	0,01325673	3.59795	84.4876	Transmembrane protein
CE.7378	3,2237	0,01332045	15.1951	274.779	germin-like protein subfamily T member 2
CE.15162	3,2372	0,01332946	12.7027	235.551	Protein DMP4
CE.10191	-3,3565	0,01335076	36.3029	3.81473	chlorophyll a-b binding protein of LHCII type 1
CE.5632	-3,4960	0,01335701	96.5188	2.80077	Cellulose synthase-like protein G3
CE.5236	3,1953	0,01344099	10.4232	179.766	low-temperature-induced 65 kDa protein-like
CE.12109	3,3441	0,01349535	5.51893	127.124	probable F-box protein At5g04010
CE.14725	3,3728	0,01357795	1.4387	4.17142	agamous-like MADS-box protein AGL104 isoform X1
CE.12329	-3,1651	0,01364081	351.822	19.9843	phosphoribulokinase. chloroplastic
CE.13194	3,1271	0,01369337	53.5019	827.366	heat shock protein 83
CE.12714	3,5257	0,01369611	3.7868	163.802	probable 6-phosphogluconolactonase 4. chloroplastic
CE.8720	3,3782	0,01379808	10.0978	262.789	pyruvate decarboxylase 1
CE.15952	-3,5126	0,01381293	78.9758	1.94766	leucine-rich repeat receptor-like protein kinase PXC2
CE.2209	3,3574	0,01392175	2.52721	64.7196	dehydration-responsive element-binding protein 2A-like
CE.16574	-3,1241	0,01402258	1655.7	99.7476	unnamed protein product
CE.6710	-3,2006	0,01404346	529.837	27.7753	NO HIT
CE.7272	3,7425	0,01406689	0.073023	358.028	probable E3 ubiquitin-protein ligase XERICO
CE.11388	-3,2932	0,01412519	49.0139	2.19182	xyloglucan galactosyltransferase XLT2-like
CE.11047	-3,2675	0,01412668	42.8834	1.98507	subtilisin-like protease SBT1.7
CE.10341	3,4087	0,01412782	2.1531	60.7041	plant cysteine oxidase 2-like
CE.2620	3,1480	0,01414206	14.4143	297.525	probable cysteine protease RD19D
CE.1531	3,2791	0,01420275	2.53177	52.5884	basic 7S globulin-like
CE.8164	-3,1828	0,01421939	381.637	20.6842	heme-binding protein 2
CE.3426	-3,1722	0,01423068	175.988	9.73719	chlorophyll a-b binding protein CP24 10A. chloroplastic
CE.6823	3,3822	0,01423169	2.93623	77.5726	HVA22-like protein e
CE.7499	3,2957	0,01424002	19.8678	428.088	unnamed protein product
CE.2138	3,3295	0,01424430	0.00000312115	82.9843	NO HIT
CE.11127	-3,1654	0,01424605	157.622	8.80262	serine hydroxymethyltransferase. mitochondrial
CE.16016	3,1348	0,01435201	60.0732	956.561	defensin Ec-AMP-D2-like
CE.1278	3,4612	0,01435644	2.06146	69.174	cysteine-rich repeat secretory protein 55
CE.13724	-3,3591	0,01440620	4.78041	0	probable O-methyltransferase 3
CE.9832	3,2384	0,01441427	5.4522	105.424	acidic mammalian chitinase-like
CE.5741	-3,2168	0,01449403	194.755	9.83356	unnamed protein product
CE.9520	3,1387	0,01454636	16.1045	259.482	heat stress transcription factor A-2-like
CE.12924	3,7242	0,01459029	0	23.247	NO HIT
CE.8218	3,7242	0,01459029	0	153.044	EP1-like glycoprotein 2
CE.4877	3,6208	0,01460822	0.494492	390.085	NO HIT

CE.6707	-3,1431	0,01466029	795.939	45.4878	unnamed protein product
CE.3903	-3,7276	0,01466413	171.988	0	NO HIT
CE.15638	-3,7276	0,01466413	309.364	0	ATP synthase CF1 beta subunit (chloroplast)
CE.5338	-3,7276	0,01466413	141.534	0	pentatricopeptide repeat-containing protein At5g46580. chloroplastic
CE.5713	3,2469	0,01466503	4.87584	96.4147	purine permease 3
CE.14593	-3,5201	0,01472828	235.498	0.514193	BEL1-like homeodomain protein 4
CE.9058	3,2753	0,01481732	3.377	71.2518	probable purple acid phosphatase 20
CE.408	-3,1542	0,01483245	360.741	20.2245	chlorophyll a-b binding protein 36. chloroplastic
CE.6396	3,3806	0,01496353	3.13524	85.9378	26.5 kda heat shock protein. mitochondrial
CE.15256	-3,1934	0,01504463	137.989	7.16008	photosystem I reaction center subunit N. chloroplastic
CE.13108	3,2480	0,01505661	5.84491	117.471	bidirectional sugar transporter SWEET1
CE.7907	3,4363	0,01511653	0.959565	314.147	formin-like protein 3
CE.16268	3,4166	0,01518047	0.235707	278.011	zinc finger CCCH domain-containing protein 14
CE.1036	3,1224	0,01535716	17.2385	275.7	expansin-A4-like
CE.3993	3,3937	0,01539176	1.18951	34.6935	viridiflorene synthase
CE.15306	3,2903	0,01542781	2.97373	66.4876	alcohol dehydrogenase 1
CE.10426	3,4674	0,01544436	2.7638	113.627	cysteine-rich repeat secretory protein 55
CE.3137	3,2887	0,01547481	6.1224	136.64	NO HIT
CE.6343	3,5921	0,01550795	0.538579	414.069	auxin-induced protein AUX22
CE.5718	3,3862	0,01564733	3.77319	101.769	mitochondrial pyruvate carrier 1-like
CE.6377	3,2111	0,01573370	10.9109	208.139	unnamed protein product
CE.6639	3,1771	0,01574778	7.76554	138.635	NO HIT
CE.8798	3,4127	0,01587516	0.811992	259.828	rop guanine nucleotide exchange factor 7-like
CE.5981	3,3497	0,01592174	1.78196	51.0633	tetrahydrocannabinolic acid synthase-like
CE.16733	-3,1114	0,01596777	643.51	37.4982	Uncharacterised protein
CE.3489	-3,3808	0,01598766	31.8632	1.02882	potassium channel AKT2/3-like
CE.14761	3,3452	0,01606652	1.71288	45.2835	Histidine-rich glycoprotein
CE.2980	3,0623	0,01613657	32.7412	481.597	floral homeotic protein AGAMOUS isoform X2
CE.15858	-3,6781	0,01615128	207.863	0	isoflavone reductase-like protein
CE.14520	3,0642	0,01620792	5.30538	81.5268	cytokinin dehydrogenase 5
CE.8098	3,1168	0,01624806	2.24638	33.4151	G-type lectin S-receptor-like serine/threonine-protein kinase SD2-5
CE.11842	3,1671	0,01633189	6.78318	120.863	unnamed protein product
CE.4873	3,2934	0,01634989	3.91226	91.6855	unnamed protein product
CE.3037	-3,0851	0,01636025	228.282	13.8915	fructose-bisphosphate aldolase 1. chloroplastic
CE.9841	-3,5855	0,01637864	241.911	0.261177	zinc finger protein 8
CE.13575	3,2447	0,01643827	3.68319	77.3259	probable RNA-binding protein ARP1 isoform X3
CE.11131	3,0254	0,01645389	129.368	1804.09	transcription factor MYB90
CE.11044	-3,2136	0,01650278	119.301	5.67962	probable esterase KAI2
CE.5239	-3,3901	0,01658531	39.8949	1.31976	pollen-specific leucine-rich repeat extensin-like protein 1
CE.13735	-3,1860	0,01665522	53.0356	1.47007	transcription factor TCP4-like

CE.5732	-3,6609	0,01669826	33.981	0	cytochrome P450 71A3-like
CE.5079	3,1944	0,01673238	2.13469	31.5265	cellulose synthase-like protein G2
CE.14720	3,3500	0,01679635	3.13866	87.5702	agamous-like MADS-box protein AGL104 isoform X2
CE.13024	3,2778	0,01687466	3.29542	73.199	Heat shock 22 kDa protein. mitochondrial
CE.8990	3,4233	0,01691173	0.652583	234.919	UDP-glycosyltransferase 73D1
CE.13235	-3,2375	0,01691793	220.634	9.83203	unnamed protein product
CE.8089	3,6469	0,01696394	0.0000000492783	19.2601	NO HIT
CE.16671	3,0858	0,01701329	13.3156	208.007	agamous-like MADS-box protein AGL9 homolog
CE.14229	3,0907	0,01702351	6.88276	50.7942	Thebaine 6-O-demethylase
CE.11387	-3,1815	0,01708120	49.2727	2.41812	xyloglucan galactosyltransferase XLT2-like
CE.13738	3,0605	0,01708150	11.2235	168.073	seed biotin-containing protein SBP65-like isoform X1
CE.16596	3,0855	0,01714270	13.6599	213.925	vignain-like
CE.8123	-3,0228	0,01716369	571.702	37.9376	chlorophyll a-b binding protein 8. chloroplastic
CE.3905	3,1796	0,01722995	5.38379	98.988	aldose 1-epimerase-like
CE.5909	-3,6435	0,01726339	173.162	0	putative nuclease HARBI1
CE.12055	3,0392	0,01728520	16.6806	242.967	cinnamoyl-CoA reductase 2-like
CE.14551	-3,0684	0,01729296	183.872	11.274	S-adenosyl-L-methionine-dependent methyltransferase
CE.6418	-3,1624	0,01729444	34.1034	1.72549	protein TSS isoform X1
CE.691	-3,1885	0,01731148	47.1004	2.30078	protein FAF-like. chloroplastic
CE.3497	-2,9946	0,01734060	1146.02	79.2978	ferredoxin-1. chloroplastic
CE.15955	-3,2898	0,01756511	134.527	5.2263	unnamed protein product
CE.15465	3,0269	0,01763868	15.8682	227.217	Endoglucanase 9
CE.8161	3,2979	0,01764489	1.94432	48.9672	NO HIT
CE.12271	-3,3914	0,01766396	232.186	0.648611	NAC domain-containing protein 86
CE.269	-2,9855	0,01770863	2144.01	146.643	acidic endochitinase-like
CE.8743	-3,1538	0,01779354	58.3979	3.01714	phosphoinositide phospholipase C 2
CE.4343	3,3978	0,01779887	3.96401	139.045	NO HIT
CE.1303	3,0578	0,01790377	17.8752	271.339	chaperone protein dnaJ 20. chloroplastic-like
CE.6314	-2,9694	0,01790529	3904.36	278.255	acid phosphatase 1-like
CE.11055	-3,0324	0,01795035	181.324	11.6551	peroxidase 51-like
CE.16653	-2,9955	0,01797381	700.713	47.8439	chlorophyll a-b binding protein 7. chloroplastic
CE.2481	-3,4725	0,01801572	20.6248	0.357887	3-ketoacyl-CoA synthase 19-like
CE.6328	-3,2433	0,01802342	42.3304	1.784	protein NRT1/ PTR FAMILY 1.2
CE.6438	-3,0601	0,01810987	124.589	7.60868	putative UPF0481 protein At3g02645
CE.8546	3,1613	0,01813706	2.21396	41.1524	E3 ubiquitin-protein ligase At4g11680-like isoform X1
CE.15580	3,3453	0,01817842	1.23442	36.8962	quinone oxidoreductase PIG3
CE.15113	3,4341	0,01829993	0	310.507	putative 12-oxophytodienoate reductase 11
CE.11912	-3,0668	0,01830274	259.895	15.6187	probable calcium-binding protein CML29
CE.12110	-2,9715	0,01847179	1758.64	123.606	lipid transfer protein precursor
CE.11424	-3,1786	0,01849545	67.4136	2.05437	dehydrodolichyl diphosphate synthase 6-like

CE.2733	-3,0435	0,01854526	129.298	8.05783	Actin cytoskeleton-regulatory complex protein
CE.7512	3,1969	0,01854953	1.92916	39.2656	cytochrome P450 CYP82D47-like
CE.8712	-3,2623	0,01856832	46.3544	1.82158	mitoferrin-like
CE.9453	3,1212	0,01868813	25.8741	450.524	programmed cell death protein 4-like
CE.13630	3,0901	0,01885764	17.1583	282.544	DUF538 domain-containing protein
CE.1334	-3,3595	0,01886062	282.611	0.812355	DedA family
CE.5182	3,0284	0,01888903	13.3704	196.996	membrane protein PM19L
CE.9265	3,0764	0,01889825	26.2253	421.238	NO HIT
CE.5899	-3,1161	0,01890889	124.464	6.71878	basic endochitinase-like
CE.2903	2,9819	0,01892769	23.468	319.919	protein trichome birefringence-like 38
CE.12994	3,0329	0,01895920	31.0683	458.637	oleosin 1-like
CE.12265	3,2887	0,01902443	1.24436	32.7224	Basic-leucine zipper transcription factor family protein. putative isoform 2
CE.14822	3,0846	0,01909004	5.42348	89.0891	purine permease 3-like
CE.2734	3,4130	0,01921548	0.873551	372.584	equilibrative nucleotide transporter 3-like
CE.2498	3,5812	0,01923583	0	342.248	methyltransferase-like 2
CE.13318	3,0929	0,01923732	4.62336	77.2715	PREDICTED: uncharacterized protein LOC107863712
CE.8148	2,9473	0,01928723	23.3843	303.389	9-cis-epoxycarotenoid dioxygenase NCED1. chloroplastic-like
CE.5108	-2,9717	0,01936553	241.261	16.6886	inorganic phosphate transporter 1-4-like
CE.16689	-3,0631	0,01944841	148.998	8.77939	auxin-responsive protein SAUR71-like
CE.11719	3,4070	0,01945621	0.484474	206.195	alkane hydroxylase MAH1-like
CE.10593	3,1055	0,01953165	3.29164	56.8306	Transmembrane protein
CE.13050	2,9935	0,01964584	10.6242	149.678	vacuolar-processing enzyme
CE.1363	3,1057	0,01984205	2.45252	42.7213	B3 domain-containing transcription factor ABI3-like
CE.16253	-2,9901	0,01993659	66.5008	4.4135	protein CHUP1. chloroplastic
CE.13258	2,9282	0,01998821	3.77882	23.3636	endoglucanase 6-like
CE.13623	2,9527	0,02001895	26.1117	346.121	F-box protein PP2-B10-like
CE.13905	-2,9949	0,02008824	198.306	8.33148	photosystem I reaction center subunit III. chloroplastic-like
CE.12687	3,5516	0,02033991	0	85.293	unnamed protein product
CE.9977	3,5516	0,02033991	0	151.878	zeatin O-glucosyltransferase-like
CE.12255	-3,3176	0,02050302	50.2029	1.39957	NAC domain-containing protein 45
CE.3959	-3,3167	0,02053655	208.628	0.625544	cytokinin dehydrogenase 6-like
CE.4976	3,1584	0,02058305	3.25472	64.9907	Pyrimidine 5-nucleotidase protein
CE.5206	2,9452	0,02061835	11.0821	146.633	probable pectate lyase 8
CE.7911	3,3741	0,02078347	0.601863	248.372	transcription factor bHLH117
CE.5342	-3,2774	0,02082843	2.45252	42.7213	DNA-directed RNA polymerase subunit beta
CE.4806	2,8884	0,02087286	108.031	1498.7	NO HIT
CE.4042	3,3688	0,02096334	1.15948	41.3236	NO HIT
CE.10097	-3,3946	0,02101508	52.869	1.06711	putative UPF0481 protein At3g02645
CE.5070	3,1324	0,02101771	21.5747	411.596	inhibitor of trypsin and hageman factor-like protein
CE.8811	3,3658	0,02110549	1.76301	72.0282	S-crystallin like

CE.10799	3,2341	0,02121611	1.35081	33.6792	myb-related protein 306-like
CE.14504	3,0786	0,02125685	3.43157	48.8777	B-box zinc finger protein 21-like
CE.10162	-3,4526	0,02128189	173.866	0.213465	36.4 kDa proline-rich protein
CE.51	3,3549	0,02158876	2.4803	94.1377	probable caffeine synthase 4 isoform X2
CE.941	2,9157	0,02159640	29.2201	374.537	LOB domain-containing protein 1-like
CE.10649	3,1518	0,02162253	2.71604	55.2831	BAG family molecular chaperone regulator 2-like
CE.16308	-3,0002	0,02188651	489.712	30.7182	hypothetical protein CDL15_Pgr006384
CE.9612	3,2166	0,02192614	1.84419	45.1315	protein LAZ1 homolog 2
CE.15660	3,3462	0,02193687	1.53537	61.5465	telomerase reverse transcriptase isoform X2
CE.13515	-2,9635	0,02194027	111.075	7.43149	ATP synthase gamma chain. chloroplastic
CE.7732	-3,0496	0,02198246	70.0896	4.08039	probable tyrosine-protein phosphatase At1g05000
CE.6583	-3,0303	0,02198413	67.5282	3.99455	tubulin beta-4 chain
CE.5322	-3,2485	0,02202026	287.061	10.0092	unnamed protein product
CE.14724	2,9883	0,02204965	8.53151	126.214	F-box-like/WD repeat-containing protein TBL1XR1 isoform X1
CE.13521	-2,9354	0,02211941	167.843	11.7455	Immunoglobulin-like and fibronectin type III domain-containing protein
CE.6880	3,0204	0,02214014	1.78529	35.9768	cytochrome P450 82C4-like
CE.6736	3,1327	0,02242423	4.83277	96.492	kunitz trypsin inhibitor 2
CE.15677	3,2389	0,02245312	1.93604	52.0831	methylesterase 3-like
CE.13503	3,1759	0,02247213	1.78946	39.9584	putative Myb family transcription factor At1g14600
CE.10121	2,8874	0,02251490	23.0223	285.987	3-oxoacyl-[acyl-carrier-protein] reductase FabG-like
CE.15735	-2,8331	0,02252033	2887.49	255.597	NADP-dependent D-sorbitol-6-phosphate dehydrogenase-like
CE.5522	-2,9704	0,02252919	38.1268	2.51311	protein TSS
CE.3360	-3,0509	0,02256746	41.5197	41.5197	ABC transporter G family member 11-like
CE.2560	3,2366	0,02257085	1.11527	29.959	beta-D-glucosyl crocetin beta-1.6-glucosyltransferase-like
CE.8721	3,0533	0,02273821	6.86117	111.625	pyruvate decarboxylase 2
CE.7028	3,4905	0,02277778	0	342.605	Exo_endo_phos domain-containing protein/DUF4283 domain-containing protein
CE.8240	-3,1091	0,02278487	23.4107	1.14952	transcription factor TCP4-like
CE.11494	2,8761	0,02279697	54.5323	668.077	glutathione S-transferase F11-like
CE.4394	-2,9490	0,02290501	116.395	5.7739	transcription factor bHLH18-like
CE.14601	-2,8602	0,02299525	292.83	18.875	Pre-mRNA-splicing factor ATP-dependent RNA helicase
CE.5420	-2,8584	0,02302352	1061.49	76.1846	unnamed protein product
CE.9331	-2,9826	0,02307770	255.741	0	sodium/pyruvate cotransporter BASS2. chloroplastic
CE.532	3,1615	0,02312210	1.50751	33.2304	cytochrome P450 87A3-like
CE.10494	3,1894	0,02312462	5.69492	124.979	unnamed protein product
CE.8885	2,9664	0,02314970	13.2426	190.419	glutathione S-transferase L3-like
CE.7010	-3,3457	0,02315428	235.583	0.497005	plant UBX domain-containing protein 10-like
CE.16313	2,9145	0,02318172	7.32405	136.038	MADS-box protein CMB1-like
CE.6254	3,0415	0,02325796	9.41277	156.967	suppressor protein SRP40-like isoform X1
CE.7683	3,0303	0,02326959	9.80825	159.9	unnamed protein product
CE.6897	3,2628	0,02328244	0.812647	24.9797	vignain-like

CE.3229	-2,9327	0,02330553	86.3106	5.94432	Epstein-Barr nuclear antigen 2 like
CE.4520	-3,4039	0,02338335	90.9407	1.0178	TUBB3
CE.5888	3,4745	0,02345536	0	521.018	FBD-associated F-box protein
CE.3372	2,9387	0,02347290	6.76062	93.0539	7-deoxyloganetin glucosyltransferase-like
CE.14238	3,1073	0,02354745	2.2236	43.2918	SNF1-related protein kinase regulatory subunit beta-1
CE.14425	-2,9045	0,02372812	449.498	32.5233	plastocyanin
CE.490	-3,0156	0,02375280	136.699	8.10243	protein CURVATURE THYLAKOID 1B. chloroplastic-like
CE.16714	2,9827	0,02388233	4.57322	44.0109	ethylene-responsive transcription factor ERF113-like
CE.5425	3,0064	0,02395252	14.9164	235.384	cysteine proteinase inhibitor B-like
CE.14505	2,9637	0,02395507	11.695	169.977	MADS box transcription factor
CE.14540	2,8412	0,02397348	48.3637	570.246	Tat-binding like
CE.11108	-2,9361	0,02400428	33.2665	2.24935	LRR receptor-like serine/threonine-protein kinase HSL2
CE.11448	2,8890	0,02414262	6.13224	78.3509	monosaccharide-sensing protein 2-like
CE.7325	3,0218	0,02414869	8.44759	138.008	thioredoxin-like 1-2. chloroplastic
CE.10984	2,8311	0,02431101	4.08086	25.0283	protein DETOXIFICATION 48-like
CE.15937	3,2399	0,02432993	1.70469	51.3183	polygalacturonase-like
CE.3099	2,8955	0,02444759	24.2201	314.436	organ-specific protein S2-like
CE.7932	-2,8034	0,02445889	420.869	35.1489	Protein phosphatase 1 regulatory subunit. putative
CE.1254	-2,8356	0,02449655	503.473	39.9824	photosystem I reaction center subunit XI. chloroplastic
CE.9266	2,9062	0,02462988	20.7693	275.489	blue copper protein
CE.2478	2,9917	0,02463063	7.63658	118.86	zinc finger protein ZAT11-like
CE.15896	2,8627	0,02475896	10.8794	134.32	beta-D-glucosyl crocetin beta-1.6-glucosyltransferase-like
CE.3901	-2,8972	0,02486166	258.014	18.4969	germin-like protein subfamily T member 2
CE.10884	3,0078	0,02486811	0	409.196	developmental protein SEPALLATA 1-like isoform X2
CE.13860	3,0780	0,02488446	1.75386	34.517	NO HIT
CE.8157	-3,0502	0,02489648	34.5601	1.87614	catechol oxidase
CE.1660	2,9358	0,02494425	7.01871	98.7283	GATA transcription factor-like protein
CE.10145	-3,1578	0,02494518	42.4808	1.72766	BURP domain-containing protein 16-like
CE.13912	-3,3066	0,02499655	33.973	0.742418	agamous-like MADS-box protein AGL19
CE.6716	-3,3053	0,02505421	290.046	0.63474	tropinone reductase homolog At5g06060
CE.16259	-2,9789	0,02512228	31.0221	1.89489	phosphoenolpyruvate carboxylase 4
CE.7874	-3,4407	0,02514460	900.475	0.468876	Apoptotic ATPase
CE.7550	-2,8638	0,02522300	185.655	13.9371	acidic endochitinase-like
CE.3765	2,9193	0,02523636	26.8226	368.048	Transcription termination factor like
CE.2861	2,7681	0,02529119	164.707	1293.05	probable serine/threonine-protein kinase clkA
CE.4408	2,8071	0,02535415	82.9115	943.986	TRNA (guanine-N(7)-)-methyltransferase non-catalytic subunit like
CE.8725	2,8328	0,02536181	10.5624	124.566	pyruvate decarboxylase 2
CE.11277	2,9007	0,02547877	8.86799	118.187	bZIP transcription factor 53-like
CE.12238	2,9817	0,02551522	5.41196	84.1226	21 kDa protein-like
CE.1805	2,8528	0,02555212	0.0000000224847	48.6875	growth-regulating factor 3-like isoform X2

CE.9615	-2,9990	0,02557682	712.809	0	photosystem I chlorophyll a/b-binding protein 6. chloroplastic
CE.1669	3,3367	0,02561362	0.573606	258.893	unnamed protein product
CE.6415	-3,2939	0,02562488	220.339	0.486846	haloalkane dehalogenase
CE.7545	3,4254	0,02564031	0	710.274	probable 2-oxoglutarate-dependent dioxygenase AOP1.2
CE.12262	2,9202	0,02564726	6.72605	103.205	unnamed protein product
CE.13264	2,9316	0,02566103	8.91367	126.064	probable protein phosphatase 2C 51
CE.1068	3,2107	0,02580379	1.06061	31.0648	NO HIT
CE.4604	-3,1659	0,02587811	172.995	0.650664	receptor-like serine/threonine-protein kinase At2g45590
CE.7981	-3,2387	0,02592057	181.811	0.519649	jasmonic acid-amido synthetase JAR1-like
CE.5194	3,0227	0,02595727	3.75923	64.3966	E3 ubiquitin-protein ligase BRE1A-like
CE.13263	3,2097	0,02598412	3.40278	97.8155	probable protein phosphatase 2C 51
CE.9620	2,9783	0,02621374	6.03472	98.2187	ethylene-responsive transcription factor ERF016-like
CE.3529	3,0504	0,02624993	4.32026	79.5233	22.0 kDa class IV heat shock protein-like
CE.13686	2,9317	0,02631335	4.43736	63.5376	NO HIT
CE.6613	3,3210	0,02637684	0.311832	18.436	NO HIT
CE.9971	2,8905	0,02638388	8.71032	115.741	heat stress transcription factor B-3-like
CE.1920	2,8454	0,02640572	5.79504	71.237	heat shock protein 83
CE.13541	-2,8006	0,02657681	341.312	27.8639	chlorophyll a-b binding protein 6. chloroplastic-like
CE.8688	-3,2738	0,02661665	358.192	0.807272	heat shock protein 83
CE.3756	-2,9434	0,02685605	64.829	4.09922	protein YLS3-like
CE.14594	-3,0233	0,02690035	96.365	5.14044	BEL1-like homeodomain protein 2
CE.2006	2,7884	0,02691196	19.2895	217.572	Spore coat assembly protein like
CE.13933	3,1447	0,02696404	0.813447	204.238	cannabidiolic acid synthase-like 2
CE.11671	-2,7406	0,02697020	544.436	48.5297	inositol-3-phosphate synthase
CE.10112	3,1090	0,02709002	4.53617	100.173	anthocyanidin 3-O-glucosyltransferase 2-like
CE.12866	-3,0335	0,02712201	33.9654	1.75798	protein NRT1/ PTR FAMILY 1.2-like
CE.5757	-3,1137	0,02718548	55.4352	2.32068	tetrapyrrole-binding protein. chloroplastic
CE.8069	-3,2626	0,02720492	44.3122	1.00776	large ribosomal RNA subunit accumulation protein YCED homolog 2. chloroplastic isoform X1
CE.14869	-2,8333	0,02721928	59.0089	4.52738	cryptochrome-1-like isoform X1
CE.2175	-2,8664	0,02724699	141.056	10.2189	unnamed protein product
CE.11186	-3,0466	0,02731039	41.3676	2.06462	unnamed protein product
CE.15599	-3,3940	0,02732761	156.085	0	flavonol 3-sulfotransferase-like
CE.5132	-3,3940	0,02732761	275.141	0	BURP domain-containing protein 16-like
CE.5361	-3,3940	0,02732761	216.077	0	unnamed protein product
CE.3443	3,3024	0,02734233	0.386292	225.041	WAT1-related protein At1g25270-like
CE.3692	-2,8059	0,02734992	109.403	8.76582	protein ECERIFERUM 1-like
CE.13861	3,0257	0,02750620	1.90752	32.9737	NO HIT
CE.9160	-2,8721	0,02750682	36.115	2.56406	ABC transporter G family member 22 isoform X1
CE.11532	2,7264	0,02759170	68.1626	704.463	glutamine synthetase
CE.5352	2,8047	0,02764556	15.1457	541.921	cysteine proteinase inhibitor 5-like

CE.11983	-2,7579	0,02767240	83.0124	1.40562	protein OPAQUE10-like
CE.3075	-3,1314	0,02769969	23.057	0.895123	AAA-ATPase At3g50940-like
CE.8481	-2,8894	0,02799005	66.4484	4.56214	probable long-chain-alcohol O-fatty-acyltransferase 5
CE.3599	-3,1970	0,02802920	184.475	0.549964	transcription factor UNE10
CE.3746	-2,7205	0,02827814	392.896	35.5781	Outer capsid glycoprotein VP7 precursor
CE.3131	2,8088	0,02848412	4.64107	30.8792	zinc finger CCCH domain-containing protein 14
CE.9857	-3,3696	0,02852505	264.613	0	protein NRT1/ PTR FAMILY 3.1-like
CE.16260	-2,7216	0,02857349	2611.35	237.205	NO HIT
CE.7408	3,1560	0,02864126	3.38996	88.4622	endochitinase EP3-like
CE.811	2,8002	0,02885302	5.13763	60.5973	glucomannan 4-beta-mannosyltransferase 2-like
CE.12220	2,8974	0,02886779	28.9386	406.667	seed maturation protein PM41
CE.15464	2,7449	0,02898304	16.3492	176.592	mannan endo-1.4-beta-mannosidase 5-like
CE.4324	-2,9265	0,02903523	71.2419	4.46411	UPF0603 protein At1g54780. chloroplastic
CE.3361	-2,6960	0,02904264	539.662	50.2227	ABC transporter G family member 11-like
CE.11807	-2,9160	0,02909174	52.9581	3.38899	probable inorganic phosphate transporter 1-9
CE.5898	-3,2875	0,02909417	616.315	0.880345	basic endochitinase-like
CE.2731	-3,0127	0,02912613	21.1795	1.09254	disease resistance RPP13-like protein 4
CE.7934	-2,7268	0,02916278	696.875	61.9207	B-box domain protein 30
CE.6370	-2,8224	0,02917281	95.6218	7.26751	acyl-protein thioesterase 1 homolog 1-like
CE.3260	2,8076	0,02925698	28.9858	348.107	probable beta-D-xylosidase 5 isoform X2
CE.13737	3,1961	0,02934333	6.94943	242.01	unnamed protein product
CE.12330	-2,8095	0,02939699	74.3706	5.76145	plastidal glycolate/glycerate translocator 1. chloroplastic
CE.4673	-2,7909	0,02944991	208.516	16.6556	calcium-binding protein kic
CE.7451	3,1416	0,02945737	0.643572	176.393	HIPL1 protein-like
CE.11195	2,8939	0,02953944	5.05789	71.4663	LOB domain-containing protein 41-like
CE.7740	-2,9170	0,02954362	63.0229	4.04559	probable tyrosine-protein phosphatase At1g05000 isoform X1
CE.14664	-2,6879	0,02964192	596.774	13.3645	patatin-like protein 2
CE.16242	-2,8525	0,02968042	130.494	9.32708	allene oxide cyclase
CE.14109	3,0599	0,02970367	2.75354	57.9018	mitochondrial fission 1 protein A-like
CE.12608	-3,3455	0,02974959	138.121	0	pentatricopeptide repeat-containing protein At2g42920. chloroplastic
CE.15444	2,7795	0,02979923	5.53279	63.8363	AT-hook motif nuclear-localized protein 17
CE.2239	2,7773	0,02999913	4.36819	23.0059	PREDICTED: uncharacterized protein LOC109217179
CE.559	3,3347	0,03010797	0	303.896	transmembrane protein
CE.9362	-2,9594	0,03032955	41.9464	2.32912	strigolactone esterase D14
CE.12846	3,0467	0,03043281	2.92226	55.9574	aspartate aminotransferase. cytoplasmic
CE.5301	-2,6833	0,03052592	1632.86	152.744	unnamed protein product
CE.15456	-2,8224	0,03056711	71.0577	5.29142	protein ECERIFERUM 1
CE.6969	3,2416	0,03060500	0.348678	191.719	C2 and GRAM domain-containing protein At5g50170 isoform X2
CE.9629	-2,9168	0,03064341	81.3188	5.0353	peroxiredoxin Q. chloroplastic
CE.5081	2,7008	0,03071039	5.31811	56.3203	ethylene receptor 2-like

CE.727	3,1700	0,03081117	1.02922	32.6099	secoisolariciresinol dehydrogenase-like
CE.6963	-2,9660	0,03082266	89.9362	4.79453	Kynurenine 3-monooxygenase
CE.8782	-3,0231	0,03083992	31.9814	1.52894	L-type lectin-domain containing receptor kinase VIII.2-like
CE.3811	-2,8845	0,03088765	113.888	7.50211	thylakoid lumenal 19 kDa protein. chloroplastic
CE.11096	2,8675	0,03101514	3.00966	41.4447	1-aminocyclopropane-1-carboxylate synthase 3
CE.4682	-2,8891	0,03106749	49.624	3.22773	Rhamnosyl O-methyltransferase
CE.14770	-2,7365	0,03109146	241.174	19.0027	heavy metal-associated isoprenylated plant protein 24-like
CE.10192	-2,7200	0,03110901	458.912	40.3409	Chlorophyll a-b binding protein 3C. chloroplastic
CE.1496	-3,3195	0,03111590	195.137	0	chlorophyllase-1
CE.5133	-3,3195	0,03111590	194.088	0	BURP domain-containing protein 16-like
CE.5769	2,8856	0,03114669	2.40618	34.3277	unnamed protein product
CE.12188	-2,7515	0,03117571	393.97	32.8709	Photosystem II PsbX
CE.3917	-2,9593	0,03121891	23.5842	1.36189	L-type lectin-domain containing receptor kinase IV.1-like
CE.6278	-2,9757	0,03121975	336.304	0.669744	photosynthetic NDH subunit of subcomplex B 3. chloroplastic
CE.11576	2,8847	0,03122851	2.73576	39.1432	AP2-like ethylene-responsive transcription factor AIL6
CE.1328	-2,8113	0,03144760	49.5685	3.71765	glutaredoxin domain-containing cysteine-rich protein 1
CE.15042	2,8592	0,03147231	4.43678	60.5938	myb-related protein 330-like
CE.2002	-2,8314	0,03148452	66.6339	4.81397	unnamed protein product
CE.13851	3,2238	0,03161213	0.671219	362.757	wall-associated receptor kinase-like 1
CE.11238	-2,9878	0,03163629	39.4373	2.02323	cytochrome P450 90B1
CE.66	-3,2413	0,03165273	37.918	0.609278	protochlorophyllide-dependent translocon component 52. chloroplastic
CE.8423	3,1029	0,03166212	0.750582	198.017	eukaryotic translation initiation factor 5-like
CE.15396	-2,6311	0,03170028	4544.83	454.011	unnamed protein product
CE.7610	2,8363	0,03179650	2.42688	31.8779	ABC transporter G family member 25
CE.3140	-2,9091	0,03180266	95.2334	5.85696	Cytochrome P450 82C4. putative isoform 2
CE.10599	3,0993	0,03190721	0.607996	160.048	protein NUCLEAR FUSION DEFECTIVE 4 isoform X2
CE.8040	-2,8656	0,03193019	40.3975	2.71272	transcription activator GLK1-like
CE.1050	3,0976	0,03199925	2.34394	61.5873	Titin like
CE.6971	-2,8338	0,03202511	185.85	13.2536	Pathogenesis-related protein 1B
CE.1588	3,0963	0,03207029	1.04822	27.5027	bidirectional sugar transporter NEC1-like
CE.16115	2,9657	0,03211984	1.70958	29.9439	Protein involved in mRNA turnover and stability
CE.12074	3,2957	0,03221412	0	296.267	small nuclear ribonucleoprotein SmD3b-like
CE.15971	3,2957	0,03221412	0	14.4091	heat stress transcription factor A-8
CE.639	-2,8794	0,03223973	46.4562	3.01647	probable leucine-rich repeat receptor-like protein kinase At5g49770
CE.8932	2,9875	0,03228260	2.29692	42.7606	putative senescence regulator S40
CE.13188	-2,7102	0,03234516	66.2749	5.83255	homeobox protein BEL1 homolog
CE.5676	3,0149	0,03235532	4.05129	84.3028	cytochrome P450 71A4-like
CE.7922	-2,9562	0,03241225	170.495	0.926159	disease resistance RPP13-like protein 4
CE.12740	3,0892	0,03247433	0.975981	233.482	ferruginol synthase-like
CE.13918	3,0473	0,03247968	0.890182	199.277	unnamed protein product

CE.5380	-2,9547	0,03254013	416.19	23.339	unnamed protein product
CE.12769	-3,2929	0,03257398	178.374	0	unnamed protein product
CE.15692	-3,1180	0,03258765	51.8051	1.9006	NO HIT
CE.12131	-2,8949	0,03262619	24.3933	1.46531	photosystem II 44 kDa protein (chloroplast)
CE.12658	-2,6924	0,03273729	152.944	13.8612	aspartyl protease family protein At5g10770-like
CE.11990	3,2047	0,03276907	0.415478	22.143	unnamed protein product
CE.14665	-2,9043	0,03279304	39.8719	2.35234	patatin-like protein 2
CE.6637	2,7563	0,03282816	7.41067	85.4824	LOB domain-containing protein 41-like
CE.16057	2,6374	0,03298551	119.191	1139.82	kirola-like
CE.8994	-3,0705	0,03309650	15.2641	0.578113	purine-uracil permease NCS1
CE.2165	3,0783	0,03314209	0.765191	197.307	phenylalanine N-monooxygenase
CE.2830	-2,9421	0,03321126	31.9481	1.76557	pistil-specific extensin-like protein
CE.8422	2,6224	0,03328890	65.6501	615.138	NEP1-interacting protein-like 1
CE.2520	2,9460	0,03329726	3.08284	52.95	ferredoxin. root R-B2-like
CE.12688	3,0759	0,03332657	5.18955	133.689	unnamed protein product
CE.7186	-3,1538	0,03334650	130.915	0.329591	long-chain-alcohol oxidase FAO4A
CE.10058	3,2755	0,03334745	0	153.555	putative methyltransferase DDB_G0268948
CE.2467	2,9440	0,03344224	2.65918	46.5664	unnamed protein product
CE.16693	2,6743	0,03345012	15.9568	162.032	NAC domain-containing protein 83
CE.7733	-3,1510	0,03350607	130.312	0.329139	glucan endo-1.3-beta-glucosidase 12
CE.12974	-3,1502	0,03355688	207.324	0.52419	taxadien-5-alpha-ol O-acetyltransferase
CE.2562	2,9631	0,03378176	3.43392	62.4366	Nuclear valosin-containing protein
CE.4066	-2,7114	0,03395652	125.357	10.8748	probable carboxylesterase 120 isoform X1
CE.14026	-3,0216	0,03409700	91.1853	3.92995	cysteine proteinase inhibitor 5-like
CE.10509	-2,6787	0,03414587	97.8058	7.6088	Titin like
CE.3016	-3,0532	0,03417008	28.5345	1.09902	X intrinsic protein
CE.7501	-2,7377	0,03418182	23.1893	2.4545	transcription factor PIF4
CE.11107	3,2549	0,03453895	0	421.152	precursor of CEP16-like
CE.9241	2,9269	0,03454697	1.75821	29.6753	probable glycosyltransferase At3g42180
CE.10711	-2,6644	0,03467539	133.241	12.9503	carbonic anhydrase. chloroplastic
CE.13330	-2,8499	0,03468259	62.8401	4.15113	beta-carotene isomerase D27. chloroplastic
CE.8671	-2,7382	0,03478349	48.9315	3.95366	serine/threonine-protein kinase STN8. chloroplastic
CE.1397	-2,6806	0,03491490	23.7722	2.82502	Anaphase-promoting complex subunit like
CE.9572	-2,9345	0,03497669	100.004	5.39539	Nuclear transition protein like
CE.318	3,1683	0,03498972	0.314607	161.933	NO HIT
CE.4734	3,0064	0,03509311	1.09067	23.5081	transcription factor MYB108-like
CE.2959	3,0464	0,03522046	1.13807	28.5038	Cytochrome P450 81E8 like
CE.5667	-2,7962	0,03522201	74.6017	5.40242	desiccation-related protein PCC13-62-like
CE.4393	-3,1815	0,03527194	24.137	0.379044	transcription factor bHLH18-like
CE.12106	-2,5671	0,03528418	2277.04	242.013	hypothetical protein CTI12_AA323580

CE.10298	-2,8189	0,03538634	124.607	8.66802	unnamed protein product
CE.383	2,8737	0,03538676	8.41317	133.797	putative chitinase
CE.14833	2,5772	0,03561317	164.537	1470.4	unnamed protein product
CE.14574	-2,6940	0,03611250	40.9192	3.54567	J domain-containing protein required for chloroplast accumulation response 1
CE.4956	-2,7890	0,03622478	39.5542	2.8821	glucose-1-phosphate adenylyltransferase large subunit 3. chloroplastic/amyloplastic
CE.1892	-2,5876	0,03630676	95.0198	9.71912	ABC transporter C family member 10-like
CE.2277	2,6726	0,03632512	8.18799	85.4699	nematode resistance protein-like HSPRO2
CE.11306	-2,5776	0,03640768	273.851	28.4062	fructose-bisphosphate aldolase 1. chloroplastic
CE.14028	-2,7249	0,03643639	320.772	26.073	photosystem I reaction center subunit psaK. chloroplastic
CE.13864	2,9515	0,03643644	1.62498	30.8436	putative receptor-like protein kinase At3g47110
CE.3862	2,6261	0,03655454	15.5915	152.511	cinnamoyl-CoA reductase 2-like
CE.6660	-3,0159	0,03664631	90.9254	3.60621	protein ASPARTIC PROTEASE IN GUARD CELL 1-like
CE.16295	-2,6847	0,03672126	77.5364	6.77817	UDP-glycosyltransferase 89A2-like
CE.1692	-2,7653	0,03673772	71.4302	5.41132	long chain acyl-CoA synthetase 1-like
CE.4287	-2,6167	0,03684430	198.744	19.3618	Nodulation-signaling pathway protein isoform 2
CE.10991	-2,8399	0,03691855	28.9286	1.87484	unnamed protein product
CE.16	-2,6407	0,03696834	133.902	12.5472	protein MET1. chloroplastic-like
CE.5350	2,5344	0,03705699	7.43201	9.64408	cysteine proteinase inhibitor 5-like
CE.8332	2,7784	0,03718244	3.36675	42.9801	unnamed protein product
CE.14956	2,8015	0,03719808	9.6317	128.861	probable rhamnogalacturonate lyase B isoform X2
CE.359	2,7041	0,03719856	12.7861	139.416	putative chitinase
CE.10826	-2,6773	0,03731663	111.045	0	UPF0301 protein SCO2948
CE.8879	-2,9714	0,03743913	37.2766	1.71708	E3 ubiquitin-protein ligase SGR9. amyloplastic-like
CE.11250	2,7551	0,03745495	3.86607	47.3531	late embryogenesis abundant protein D-29
CE.592	2,5907	0,03747896	21.4451	199.345	alcohol dehydrogenase-like 7
CE.8100	-2,7976	0,03748182	70.5455	4.95451	Molecular chaperone (small heat-shock protein Hsp26/Hsp42)
CE.12712	-3,2079	0,03758289	172.212	0	NO HIT
CE.545	-2,8770	0,03761996	35.7469	2.096	glucan endo-1.3-beta-glucosidase
CE.5717	-2,6226	0,03766761	1408.25	134.896	putative 23S ribosomal RNA
CE.15228	2,6634	0,03771426	18.2679	187.635	putative germin-like protein 2-1
CE.8936	-2,5494	0,03778427	503.664	53.8405	29 kDa ribonucleoprotein A. chloroplastic-like
CE.7344	-3,1383	0,03809376	449.904	0.73319	Disulfide bond formation protein B 2. putative
CE.1095	-2,5565	0,03814157	3436.29	362.534	signal anchor. putative
CE.8287	-2,5128	0,03823519	39496.8	4434.4	metallothionein-like protein type 3
CE.7687	-2,6619	0,03840754	69.9929	6.26878	cytochrome P450 82C4-like
CE.5124	-2,6616	0,03842598	35.6038	3.17637	COP1-interacting protein 7
CE.14188	3,1899	0,03852826	0	284.779	subtilisin-like protease SBT3.9
CE.1075	2,5639	0,03853411	13.1602	281.961	17.6 kDa class I heat shock protein-like
CE.12981	2,5557	0,03858807	0	1.15534	oleosin 1-like
CE.13107	-2,9556	0,03860423	195.731	0.89579	beta-D-glucosyl crocetin beta-1.6-glucosyltransferase-like

CE.3429	-2,5850	0,03874012	391.491	0.0393797	sulfate transporter 1.3-like
CE.2311	2,5289	0,03884340	75.5079	645.989	RNA-binding protein like
CE.5439	2,6856	0,03884911	11.1376	122.216	UDP-glucose iridoid glucosyltransferase-like
CE.13199	2,7342	0,03888893	5.15422	61.8792	1-aminocyclopropane-1-carboxylate synthase-like
CE.13391	-2,5708	0,03893054	283.939	29.1335	ATP synthase subunit b'. chloroplastic
CE.4813	-2,6240	0,03919700	94.2521	9.91554	epoxide hydrolase
CE.5030	2,8210	0,03926606	1.34077	15.0131	major allergen Pru ar 1-like
CE.5419	-2,5149	0,03927497	7.09611	0	NO HIT
CE.1483	2,5814	0,03929003	10.5721	98.1898	Translation initiation factor IF-2 like
CE.13451	-2,6243	0,03940259	184.704	17.3544	inositol monophosphatase 3-like
CE.3907	-2,5531	0,03943129	174.027	18.264	Beta-glucosidase 11
CE.4443	2,9086	0,03944077	2.262	41.2326	GATA transcription factor 15-like
CE.5169	-2,7333	0,03946272	56.4484	4.37347	28 kDa ribonucleoprotein. chloroplastic
CE.7650	-2,6524	0,03947464	211.334	18.8602	unnamed protein product
CE.10960	-3,0147	0,03948866	31.8955	0.977454	salicylic acid-binding protein 2-like
CE.16657	-3,1778	0,03949669	124.544	0	Molybdenum cofactor sulfurase
CE.6463	2,6126	0,03957975	19.4875	190.633	chaperone protein dnaJ 8. chloroplastic-like
CE.7595	2,8495	0,03976998	3.14294	46.6061	NO HIT
CE.10014	2,5062	0,03983697	66.7587	560.067	glucan endo-1.3-beta-glucosidase. acidic-like
CE.3304	3,1673	0,03999117	0	303.165	unnamed protein product
CE.13286	-2,8825	0,04005848	196.931	11.2831	glutaredoxin domain-containing cysteine-rich protein 1
CE.500	-2,5147	0,04015813	393.315	43.4301	Serine--glyoxylate aminotransferase
CE.7954	2,7723	0,04020089	7.69547	101.688	Wound-responsive family protein
CE.16467	-2,5429	0,04025608	177.768	18.8193	transcription factor MYB1R1
CE.15143	2,6971	0,04025735	4.77037	54.3772	NO HIT
CE.4975	-2,5683	0,04028626	273.192	27.8203	photosystem I reaction center subunit II. chloroplastic
CE.3240	-2,6522	0,04030814	133.002	11.8164	protein TRI1-like
CE.11646	-3,1047	0,04037292	256.029	0.431216	BTB/POZ domain-containing protein DOT3 isoform X1
CE.9715	2,5277	0,04037481	7.88088	22.3592	probable alpha.alpha-trehalose-phosphate synthase [UDP-forming] 11
CE.9636	2,6432	0,04043153	13.9876	145.187	auxin-responsive protein SAUR72-like
CE.5683	2,8165	0,04067044	3.05495	46.1491	cytochrome P450 71A6-like
CE.6917	-2,6581	0,04072092	60.122	5.26361	cytochrome P450 71A1-like
CE.13939	2,8361	0,04080950	1.70586	25.0718	berberine bridge enzyme-like 8
CE.8094	-2,7845	0,04085597	42.8496	2.92724	unnamed protein product
CE.2916	2,5686	0,04091639	23.6757	218.614	non-specific lipid-transfer protein 1-like
CE.983	2,8603	0,04093348	1.7679	29.2104	glutelin type-B 2-like
CE.5674	2,8341	0,04094424	5.35668	84.95	cytochrome P450 71A4-like
CE.15403	2,5042	0,04095380	12.8522	61.939	ethylene-responsive transcription factor ABR1-like isoform X2
CE.8063	-2,9533	0,04108926	27.2729	1.26735	receptor-like protein kinase FERONIA
CE.8791	-2,5730	0,04109746	212.329	21.3177	17.6 kDa class I heat shock protein

CE.7553	-2,5281	0,04114907	342.154	36.7692	Oxygen-evolving enhancer protein 2. chloroplastic
CE.10016	-2,9196	0,04118958	19.7056	0.933787	protein IQ-DOMAIN 14
CE.12905	2,7195	0,04131153	4.66241	56.2309	NDR1/HIN1-like protein 6
CE.3652	3,1443	0,04153498	0	391.932	unnamed protein product
CE.355	3,1443	0,04153498	0	29.924	NO HIT
CE.4729	-2,6247	0,04173214	79.9369	7.3335	galactinol synthase 1
CE.2805	2,9538	0,04173521	0.959917	220.565	stem-specific protein TSJT1-like
CE.3614	2,5667	0,04178287	15.7054	145.671	Histidine phosphatase superfamily. clade-1
CE.6825	2,5157	0,04182286	27.6778	237.522	heavy metal-associated isoprenylated plant protein 7-like
CE.2243	-2,5471	0,04190900	107.693	11.2255	omega-3 fatty acid desaturase. chloroplastic
CE.14652	-2,8000	0,04193389	178.81	11.57	putative EF-hand domain pair protein
CE.7652	2,9507	0,04195216	1.91696	39.2343	germin-like protein subfamily 1 member 20
CE.6402	-2,5488	0,04203581	144.075	13.9703	Ribulose biphosphate carboxylase/oxygenase
CE.14398	-2,4732	0,04213922	288.209	33.2688	ATP-dependent zinc metalloprotease FTSH. chloroplastic
CE.15192	3,0632	0,04214124	0.591715	276.498	kunitz trypsin inhibitor 2
CE.12942	2,9969	0,04235766	1.39911	40.5851	17.9 kDa class II heat shock protein-like
CE.4880	3,1319	0,04239010	0.148284	375.924	RAB6A-GEF complex partner protein 1-like
CE.9554	2,7559	0,04255398	1.64511	21.795	phospholipid:diacylglycerol acyltransferase 1-like
CE.1021	-2,7242	0,04258511	8.64861	0	dof zinc finger protein DOF2.4-like
CE.7591	2,8995	0,04266871	3.53488	75.9427	protein RSI-1
CE.5080	2,9922	0,04269711	0.543585	156.828	transcriptional regulator SUPERMAN-like
CE.11024	2,6570	0,04275918	4.99725	54.5268	xyloglucan galactosyltransferase XLT2-like
CE.14080	-2,4548	0,04279290	3196.66	377.542	hypothetical protein DCAR_032037
CE.4677	-2,5844	0,04309285	68.7199	6.65492	vegetative cell wall protein gp1-like
CE.6367	2,8952	0,04310288	1.49748	29.1271	cytochrome P450 734A1-like
CE.7834	3,0501	0,04312622	0.256393	11.861	L-ascorbate oxidase homolog
CE.4593	-2,6913	0,04315717	43.0606	3.44596	fasciclin-like arabinogalactan protein 8
CE.14935	3,1206	0,04316537	0	325.631	NO HIT
CE.7848	-2,8937	0,04318782	49.7034	2.4115	Rubredoxin 4
CE.16044	-3,0097	0,04322165	41.5429	1.19396	probable NAD(P)H dehydrogenase subunit CRR3. chloroplastic
CE.10936	-2,5640	0,04322471	57.1686	5.71537	BEL1-like homeodomain protein 4
CE.8672	2,5413	0,04323523	6.72651	60.3535	mannan endo-1.4-beta-mannosidase 6 isoform X1
CE.6587	-2,8644	0,04334783	37.0541	1.96202	peptidyl-prolyl cis-trans isomerase FKBP17-2. chloroplastic
CE.14879	2,4462	0,04339585	134.852	1020.72	annexin D1
CE.10254	3,0462	0,04340442	0.857114	28.5662	NO HIT
CE.6126	-2,8146	0,04356166	162.855	0.984581	subtilisin-like protease SBT1.5
CE.15966	-3,0602	0,04361631	418.599	0.735068	receptor-like protein 12
CE.65	-2,7633	0,04362751	39.8611	2.72783	protochlorophyllide-dependent translocon component 52. chloroplastic
CE.10983	2,9285	0,04367248	1.43694	32.2458	unnamed protein product
CE.13931	2,6507	0,04373285	3.64888	39.6311	cannabidiolic acid synthase-like

CE.1516	-2,6756	0,04376007	78.336	6.41146	chloroplast stem-loop binding protein of 41 kDa a. chloroplastic
CE.11534	-3,1147	0,04376091	115.646	0	cysteine-rich repeat secretory protein 60-like
CE.12962	-3,1147	0,04376091	282.035	0	expansin-like B1
CE.16558	-3,1147	0,04376091	138.403	0	NDR1/HIN1-like protein 13
CE.620	-2,7448	0,04390162	40.1305	2.84594	sorbitol dehydrogenase
CE.8514	-2,7094	0,04397062	55.4506	4.23393	gem-associated protein 2
CE.15327	-2,4825	0,04416036	986.043	110.891	NO HIT
CE.4497	-2,5829	0,04432487	203.548	19.7716	unnamed protein product
CE.761	2,4384	0,04432547	92.3411	710.249	ABC transporter G family member 25
CE.14391	-2,6460	0,04442315	360.591	30.6404	unnamed protein product
CE.13658	-2,4812	0,04450394	450.723	50.6637	thioredoxin F-type. chloroplastic-like
CE.4844	-2,7242	0,04456486	19.4812	1.42982	translocase of chloroplast 159. chloroplastic
CE.11482	2,5734	0,04466266	10.5781	101.767	NDR1/HIN1-like protein 6
CE.16588	-2,6297	0,04477695	37.499	3.30363	E3 ubiquitin-protein ligase RGLG5-like
CE.11603	-2,8733	0,04486112	25.4926	1.25863	unnamed protein product
CE.3306	2,9132	0,04487488	2.1528	47.609	NO HIT
CE.5814	3,0964	0,04488852	0	390.492	NO HIT
CE.3167	2,7606	0,04493114	2.50495	34.9106	NO HIT
CE.13802	-2,6356	0,04525069	89.2962	7.58428	chlorophyll a-b binding protein P4. chloroplastic
CE.10731	2,9560	0,04549033	1.63066	45.4438	Forkhead box protein like
CE.9455	2,7768	0,04551032	4.34737	63.6975	osteocalcin 2-like
CE.14288	-2,5015	0,04555606	130.129	14.0797	heterodimeric geranylgeranyl pyrophosphate synthase small subunit. chloroplastic
CE.12668	2,4806	0,04561522	13.6821	94.9079	premnaspirodiene oxygenase
CE.322	-2,5763	0,04566718	45.5252	4.3161	L-type lectin-domain containing receptor kinase IV.1-like
CE.571	-2,5130	0,04576964	23.9514	3.31569	Outer capsid glycoprotein VP7 precursor
CE.15875	-2,5142	0,04596914	276.711	29.2585	isoflavone reductase-like protein
CE.11495	-2,5077	0,04608845	86.7285	9.25609	phytoene synthase 2. chloroplastic
CE.5457	3,0113	0,04611001	2.84594	126.92	aldo/keto reductase AKR1
CE.9843	-3,0811	0,04617217	161.438	0	dirigent protein 22-like
CE.15633	-3,0811	0,04617217	488.006	0	3-hexulose-6-phosphate isomerase-like
CE.9957	-2,4623	0,04627308	80.273	9.16564	magnesium-chelatase subunit ChlH. chloroplastic
CE.16246	-2,5614	0,04630801	55.4984	5.42176	serine/threonine-protein kinase STN7. chloroplastic
CE.12758	-2,7453	0,04631074	20.1751	1.37372	pleiotropic drug resistance protein 1-like
CE.1212	-2,5845	0,04634226	46.9263	4.40655	root phototropism protein 2
CE.1971	-2,8844	0,04648544	156.084	0.70298	WD repeat-containing protein RUP2
CE.5360	-2,9679	0,04651895	330.366	0.98516	unnamed protein product
CE.2009	3,0715	0,04671108	0	181.729	protein HOTHEAD-like isoform X2
CE.9936	2,9407	0,04673466	0.856875	235.505	Retrovirus-related Pol polyprotein from transposon TNT 1-94
CE.1458	2,7394	0,04681960	1.93	26.4292	cytochrome P450 94A2-like
CE.12050	2,8485	0,04686051	1.16279	21.5564	7-deoxyloganetic acid glucosyltransferase-like

CE.11900	-2,8481	0,04686879	24.5924	1.34787	urea-proton symporter DUR3
CE.8418	2,8888	0,04687403	2.58364	55.8858	Expression of the gene is downregulated in the presence of paraquat. putative
CE.10791	2,8133	0,04693796	3.28443	44.8379	NO HIT
CE.12479	-2,5438	0,04695501	17.3742	2.2023	pectin acetylesterase 8-like
CE.10332	-2,8197	0,04695525	49.147	2.72599	myb-related protein 306-like
CE.7680	-2,8788	0,04701683	176.759	0.83092	anthocyanidin 3-O-glucosyltransferase
CE.12721	-2,9165	0,04703811	65.5129	2.52186	unnamed protein product
CE.11686	2,7803	0,04729326	5.65755	86.6978	unnamed protein product
CE.14803	2,3859	0,04737182	203.294	1502.44	laccase-14-like
CE.3644	2,7532	0,04742831	2.29513	32.8689	protein DMR6-LIKE OXYGENASE 2-like
CE.3132	2,8063	0,04748370	1.43731	23.8158	allene oxide synthase 3-like
CE.2712	2,5337	0,04751707	13.572	125.393	probable NAD(P)H dehydrogenase (quinone) FQR1-like 1
CE.8597	-2,4257	0,04751959	620.891	74.1709	photosystem II 10 kDa polypeptide. chloroplastic
CE.6442	2,7766	0,04757726	2.89333	44.1698	ethylene-responsive transcription factor ERF003-like
CE.14407	2,5150	0,04773177	7.12405	63.9563	L-ascorbate oxidase homolog
CE.12891	-2,7272	0,04784982	47.5846	3.29461	transcription factor myb1r1
CE.13876	-2,5183	0,04785675	162.258	16.7914	protein PHLOEM PROTEIN 2-LIKE A1-like
CE.2702	-2,9062	0,04795183	33.1584	1.28586	alpha-farnesene synthase
CE.9010	2,4417	0,04797784	12.8452	116.972	embryonic protein DC-8-like isoform X2
CE.9933	2,6176	0,04803801	0	214.646	nuclear transcription factor Y subunit A-1-like
CE.7681	2,5437	0,04806440	4.92824	16.0063	uncharacterized membrane protein At3g27390
CE.1916	-2,9045	0,04808572	39.1321	1.52039	unnamed protein product
CE.7881	-2,4943	0,04808728	41.7711	4.48348	glutamate receptor 3.4-like
CE.5323	-2,7231	0,04824375	32.7108	2.37199	unnamed protein product
CE.8843	-2,5926	0,04826594	69.8437	6.34235	short-chain dehydrogenase TIC 32. chloroplastic-like
CE.16310	-2,7775	0,04837640	84.4429	5.09417	Cell migration-inducing and hyaluronan-binding protein
CE.10167	2,6822	0,04854310	5.1502	65.1742	cysteine inhibitor 1
CE.1459	2,8690	0,04858484	1.05292	22.3881	BEL1-like homeodomain protein 2
CE.15034	3,0460	0,04864034	0	106.904	transmembrane protein 205-like
CE.15406	3,0460	0,04864034	0	127.165	peroxidase 12
CE.7413	-3,0469	0,04873143	285.347	0	endochitinase EP3-like
CE.13125	2,4073	0,04873325	30.7814	206.971	bifunctional UDP-glucose 4-epimerase and UDP-xylose 4-epimerase 1
CE.2993	2,7159	0,04876726	2.48549	33.2622	amino acid permease 3-like
CE.7200	2,5866	0,04902533	2.48549	33.2622	cytochrome P450 71A1-like
CE.2034	-2,8210	0,04920871	22.0865	1.14558	unnamed protein product
CE.3645	2,7101	0,04928356	4.58566	61.0508	uclacyanin-3-like
CE.9829	2,5905	0,04937933	5.6957	59141	chitotriosidase-1-like
CE.15492	-2,4060	0,04945423	324.123	39.3573	chlorophyll a-b binding protein CP26. chloroplastic
CE.14354	-2,5506	0,04952690	50.8582	4.91462	psbP domain-containing protein 3. chloroplastic
CE.16064	-2,6134	0,04960311	90.2414	7.74607	MLP-like protein 43

CE.3918	-2,7644	0,04960906	31.1639	1.9	L-type lectin-domain containing receptor kinase IV.1
CE.11489	2,4786	0,04962655	17.564	151.007	Aspartate carbamoyltransferase
CE.6663	-2,8145	0,04973728	22.8175	1.1918	Transcription factor DIVARICATA
CE.13797	2,7806	0,04981264	3.34109	50.1495	unnamed protein product
CE.8290	-2,6290	0,04984122	18.729	1.55945	Cytidylate kinase
CE.4501	-2,9269	0,04996158	20.6565	0.638213	psbP domain-containing protein 1. chloroplastic
CE.7542	2,6842	0,04996323	9.82088	124.796	oxoglutarate/iron-dependent dioxygenase

---

**Fonte:** o próprio autor.