



**Universidad Nacional Mayor de San Marcos**

**Universidad del Perú. Decana de América**

**Facultad de Ciencias Biológicas**

**Escuela Profesional de Ciencias Biológicas**

**Obtención del genoma nuclear de *Eugenia uniflora* L.  
(Myrtaceae) para la identificación de terpeno sintasas**

**TESIS**

Para optar el Título Profesional de Biólogo con mención en  
Biología Celular y Genética

**AUTOR**

Frank Lino GUZMAN ESCUDERO

**ASESORES**

Mabel ALVARADO GUTIERREZ

Rogelio MARGIS (Coasesor)

Lima, Perú

2022

## Referencia bibliográfica

---

Guzman, F. (2022). *Obtención del genoma nuclear de Eugenia uniflora L. (Myrtaceae) para la identificación de terpeno sintasas*. [Tesis de pregrado, Universidad Nacional Mayor de San Marcos, Facultad de Ciencias Biológicas, Escuela Profesional de Ciencias Biológicas]. Repositorio institucional Cybertesis UNMSM.

---



Reconocimiento - No Comercial - Compartir Igual - Sin restricciones adicionales

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Usted puede distribuir, remezclar, retocar, y crear a partir del documento original de modo no comercial, siempre y cuando se dé crédito al autor del documento y se licencien las nuevas creaciones bajo las mismas condiciones. No se permite aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros a hacer cualquier cosa que permita esta licencia.

## Metadatos complementarios

<b>Datos de autor</b>	
Nombres y apellidos	Frank Lino Guzman Escudero
Tipo de documento de identidad	DNI
Número de documento de identidad	40959413
URL de ORCID	<a href="https://orcid.org/0000-0002-5048-4213">https://orcid.org/0000-0002-5048-4213</a>
<b>Datos de asesor</b>	
Nombres y apellidos	Mabel Alvarado Gutiérrez
Tipo de documento de identidad	DNI
Número de documento de identidad	41353553
URL de ORCID	<a href="https://orcid.org/0000-0001-8135-9223">https://orcid.org/0000-0001-8135-9223</a>
<b>Datos de coasesor</b>	
Nombres y apellidos	Rogelio Margis
Tipo de documento de identidad	Cédula de identidad
Número de documento de identidad	BR / 43103839049
URL de ORCID	<a href="https://orcid.org/0000-0002-2871-4473">https://orcid.org/0000-0002-2871-4473</a>
<b>Datos del jurado</b>	
<b>Presidente del jurado</b>	
Nombres y apellidos	Monica Arakaki Makishi
Tipo de documento	DNI
Número de documento de identidad	08136445
<b>Miembro del jurado 1</b>	
Nombres y apellidos	Gustavo Sandoval Peña

Tipo de documento	DNI
Número de documento de identidad	41020762
<b>Miembro del jurado 2</b>	
Nombres y apellidos	Claudia Barletta Carrillo
Tipo de documento	DNI
Número de documento de identidad	40846770
<b>Datos de investigación</b>	
Línea de investigación	A.1.1.3. Diversidad Genética
Grupo de investigación	Laboratorio de Genómica de Plantas y Poblaciones - LGPP
Agencia de financiamiento	Brasil. Coordinación de la formación del personal de nivel superior (CAPES).
Ubicación geográfica de la investigación	Edificio: Centro de Biotecnología País: Brasil Estado: Rio Grande del Sur Ciudad: Porto Alegre Barrio: Agronomia
Año o rango de años en que se realizó la investigación	Enero 2020 – enero 2022
URL de disciplinas OCDE	Bioquímica, Biología molecular <a href="https://purl.org/pe-repo/ocde/ford#1.06.03">https://purl.org/pe-repo/ocde/ford#1.06.03</a> Genética, Herencia <a href="https://purl.org/pe-repo/ocde/ford#1.06.07">https://purl.org/pe-repo/ocde/ford#1.06.07</a> Biotecnología agrícola, Biotecnología alimentaria <a href="https://purl.org/pe-repo/ocde/ford#4.04.01">https://purl.org/pe-repo/ocde/ford#4.04.01</a>



**Universidad Nacional Mayor de San Marcos**  
(Universidad del Perú, Decana de América)

**FACULTAD DE CIENCIAS BIOLÓGICAS**

**ACTA DE SESIÓN PARA OPTAR AL TÍTULO PROFESIONAL DE  
BIÓLOGO CON MENCIÓN EN BIOLOGÍA CELULAR Y GENÉTICA  
(MODALIDAD: SUSTENTACIÓN VIRTUAL DE TESIS)**

Siendo las 10:02 horas del 21 de abril de 2022, en el Salón de Grados Virtual de la Facultad de Ciencias Biológicas cuya dirección electrónica fue <https://zoom.us/j/92523453080?pwd=SXZhSlJoNEtyRDhWMmpBMFcvNGM2QT09>, y en presencia del Jurado formado por los profesores que suscriben, se inició la sesión para optar al **Título Profesional de Biólogo con mención en Biología Celular y Genética** de **FRANK LINO GUZMAN ESCUDERO**.

Luego de dar lectura y conformidad al expediente N°, el titulado expuso su tesis: **“OBTENCIÓN DEL GENOMA NUCLEAR DE *Eugenia uniflora* L. (MYRTACEAE) PARA LA IDENTIFICACIÓN DE TERPENOS SINTASAS”**, y el Jurado efectuó las preguntas del caso calificando la exposición con la nota **19**, calificativo: **Aprobado con máximos honores**.

Finalmente, el expediente será enviado a la Escuela Profesional de Ciencias Biológicas y al Consejo de Facultad para que se apruebe otorgar el **Título Profesional de Biólogo con mención en Biología Celular y Genética** a **FRANK LINO GUZMAN ESCUDERO** y se eleve lo actuado al Rectorado para conferir el respectivo título, conforme a ley.

Siendo las 11:50 horas se levantó la sesión.

Ciudad Universitaria, 21 de abril de 2022.

---

**Dra. MONICA ARAKAKI MAKISHI**  
(PRESIDENTA)

---

**Dra. MABEL ALVARADO GUTIERREZ**  
(ASESORA)

---

**Mg. GUSTAVO SANDOVAL PEÑA**  
(MIEMBRO)

---

**Mg. CLAUDIA BARLETTA CARRILLO**  
(MIEMBRO)

A mi familia

## **AGRADECIMIENTOS**

Al Dr. Rogerio Margis por la oportunidad de pertenecer a su grupo de investigación, y así permitir mi desarrollo académico y profesional, gracias por todo el apoyo brindado en este periodo de mi vida en tierras brasileras.

A los miembros del Laboratorio de Genomas y Poblaciones de Plantas, en especial a Nureyev Ferreira, Guilherme Cordenonsi, Igor Paim, Natalia Balbinott y Pabulo Rampelotto por los aprendizajes, conversaciones filosóficas, momentos de relajación y por el apoyo brindado.

De manera muy especial, quiero dedicar el presente trabajo a Edilena Sperb. Gracias por el apoyo que me brindaste en todo momento, y por enseñarme la importancia de la palabra amistad y familia, siempre te recordare y tendré presente mi gran amiga.

A la Coordinación de la formación del personal de nivel superior (CAPES) y al Consejo Nacional de Desarrollo Científico y Tecnológico (CNPq) por el financiamiento de la presente tesis.



## ABREVIATURAS

BAM	Binary alignment map
Bp	Pares de bases
CAPES	Coordinación de la formación del personal de nivel superior
CNPq	Consejo Nacional de Desarrollo Científico y Tecnológico
CTAB	Bromuro de hexadeciltrimetilamonio
DNA	Ácido desoxirribonucleico
DMAPP	Difosfato de dimetilalilo
EDTA	Ácido etilendiaminotetraacético
FDP	Farnesil difosfato
GO	Ontología genica
GPP	Geranil pirofosfato
GGPP	Geranilgeranil pirofosfato
HTML	HyperText markup language
IPP	Isopentenil pirofosfato
Kbp	Mil pares de bases
miRNA	Micro RNA
MP	Mate-paired
MPE	Vía del metileritritolo fosfato
Mbp	Un millón de pares de bases
NaCl	Cloruro de sodio
NCBI	National Center for Biotechnology Information
NGS	Next Generation Sequencing
LINEs	Secuencias repetitivas dispersas largas
LTRs	Repeticiones de terminal larga
PE	Paired-end
pH	Potencial de hidrogeno

pre-miRNA	Precursor del miRNA
PT	Preniltransferasa
RNA	Ácido ribonucleico
SAM	Sequence alignment map
SINEs	Secuencias repetitivas dispersas cortas
TE	Tris EDTA
TE	Elementos transponibles
TPS	Terpeno sintasa
UFRGS	Universidade Federal do Rio Grande do Sul

## INDICE GENERAL

1.	INTRODUCCIÓN.....	1
2.	MARCO TEÓRICO.....	3
2.1	Generalidades.....	3
2.1.1	Posición taxonómica.....	3
2.1.2	La familia Myrtaceae.....	3
2.1.3	<i>Eugenia uniflora</i> .....	4
2.2	Los terpenos.....	5
2.3	Las terpeno sintasas.....	6
2.4	Secuenciamiento de genomas en plantas.....	6
2.5	Genómica de Myrtaceae.....	7
2.6	Recursos genómicos en <i>E. uniflora</i> .....	7
3.	HIPÓTESIS.....	9
4.	OBJETIVOS.....	9
5.	MATERIALES Y MÉTODOS.....	10
5.1	MATERIALES.....	10
5.1.1	Material biológico.....	10
5.2	MÉTODOS.....	11
5.2.1	Extracción de DNA genómico.....	11
5.2.2	Secuenciamiento del DNA.....	11
5.2.3	Calidad y limpieza de las bibliotecas.....	12
5.2.4	Estimación del tamaño del genoma y heterocigosidad.....	12
5.2.5	Ensamblaje del genoma.....	13
5.2.6	Identificación de secuencias repetitivas en el genoma.....	13
5.2.7	Anotación y validación del genoma.....	14
5.2.8	Análisis de genómica comparativa.....	16
5.2.9	Identificación de las terpeno sintasas.....	16
6.	RESULTADOS.....	18
6.1.1	Extracción y secuenciamiento de DNA genómico.....	18
6.1.2	Calidad y limpieza de las bibliotecas.....	19
6.1.3	Estimación del tamaño del genoma y heterocigosidad.....	20
6.1.4	Ensamblaje del genoma.....	21
6.1.5	Identificación de secuencias repetitivas en el genoma.....	21
6.1.6	Anotación y validación del genoma.....	22
6.1.7	Análisis de genómica comparativa.....	23

6.1.8	Identificación de las terpeno sintasas.....	28
7.	DISCUSIÓN.....	31
8.	CONCLUSIONES.....	37
9.	RECOMENDACIONES.....	38
10.	REFERENCIAS BIBLIOGRÁFICAS.....	39
	ANEXOS.....	52

## ÍNDICE DE FIGURAS

	Pág.
<b>Figura 1.</b> Hojas del árbol de pitanga que fueron colectadas para la extracción del DNA.	10
<b>Figura 2.</b> Calidad del DNA de tres diferentes hojas colectadas del mismo individuo de pitanga en gel de agarosa al 1%.	18
<b>Figura 3.</b> Estimación del tamaño del genoma de pitanga.	20
<b>Figura 4.</b> Comparación de los valores obtenidos, en la validación con BUSCO, del genoma ensamblado y de la predicción génica de pitanga.	23
<b>Figura 5.</b> Diagrama de Venn mostrando los clústeres de genes ortólogos identificados entre las especies de plantas del estudio.	24
<b>Figura 6.</b> Clústeres de ortólogos y número de genes codificadores de proteínas compartidos entre las especies de plantas del estudio.	25
<b>Figura 7.</b> Categorías y términos de Ontología de Genes (GO) identificados en los genes exclusivos de <i>E. uniflora</i> .	27
<b>Figura 8.</b> Análisis filogenético de las TPS completas de <i>E. uniflora</i> y <i>E. grandis</i> .	30
<b>Figura 9.</b> Output del programa FastQC de la biblioteca PE de 250 bp antes de la limpieza con el programa Trim galore!.	54
<b>Figura 10.</b> Output del programa FastQC de la biblioteca PE de 250 bp después de la limpieza con el programa Trim galore!.	56
<b>Figura 11.</b> Output del programa FastQC de la biblioteca PE de 550 bp antes de la limpieza con el programa Trim galore!.	58
<b>Figura 12.</b> Output del programa FastQC de la biblioteca PE de 550 bp después de la limpieza con el programa Trim galore!.	60
<b>Figura 13.</b> Output del programa FastQC de la biblioteca MP de 8 Kbp	62

antes de la clasificación con el programa Nextclip.

**Figura 14.** Output del programa FastQC de la biblioteca MP-usable después de la limpieza con el programa Trim galore!. 64

**Figura 15.** Reporte del secuenciamiento Illumina realizado por la empresa Macrogen. 77

**Figura 16.** Diagrama de Venn mostrando los clústeres de genes ortólogos identificados entre *E.uniflora* y *E. grandis*. 83

## ÍNDICE DE TABLAS

	Pág.
<b>Tabla 1.</b> Clasificación de los pares de lecturas de la biblioteca MP después de eliminar los adaptadores.	19
<b>Tabla 2.</b> Número total de pares de lecturas antes y después de la limpieza.	20
<b>Tabla 3.</b> Estadísticas del ensamblaje del genoma de <i>E. uniflora</i> .	21
<b>Tabla 4.</b> Clasificación de las secuencias repetitivas identificadas en el genoma de <i>E. uniflora</i> .	22
<b>Tabla 5.</b> Clústeres y singletons identificados en las especies de plantas del estudio.	23
<b>Tabla 6.</b> Categorías de Ontología de Genes (GO) de los genes exclusivos de pitanga.	26
<b>Tabla 7.</b> Resultados de BLAST y clasificación de las TPS candidatas de <i>E. uniflora</i> .	28
<b>Tabla 8.</b> Número de TPS identificadas en <i>E. uniflora</i> y <i>E. grandis</i> .	30
<b>Tabla 9.</b> Repeticiones que fueron excluidas por tener similitud por BLAST con secuencias codificadoras de proteínas en Viridiplantae.	78
<b>Tabla 10.</b> Anotación de los genes exclusivos de <i>E. uniflora</i> .	84
<b>Tabla 11.</b> Descripción de los términos de Ontología de Genes (GO) de los genes exclusivos de <i>E. uniflora</i> .	119

## RESUMEN

Pitanga (*Eugenia uniflora* L.) es un miembro de la familia Myrtaceae que se distribuye en la Mata Atlántica y se caracteriza por la síntesis de metabolitos secundarios con importantes actividades biológicas, así como por tener un amplio espectro ecológico en ambientes que son climáticamente contrastantes. Sin embargo, la disponibilidad de recursos genómicos que permitan estudiar esta especie aún es limitada y escasa en comparación con otras especies de plantas. Por ese motivo, el presente trabajo de investigación tiene como objetivo la obtención de un genoma de referencia para *E. uniflora* que permita identificar los miembros de la familia de las terpeno sintasas (TPS) que son responsables de la síntesis de terpenos de esta especie. El ensamblaje *de novo* del genoma, a partir de lecturas secuenciadas utilizando la plataforma Illumina, permitió obtener un tamaño total de 385.1 Mbp con un valor de N50 igual a 26 199 bp. Asimismo, la predicción génica utilizando como referencia proteínas de Viridiplantae permitió obtener 30 663 genes codificadores de proteínas de pitanga. Un análisis de genómica comparativa con otras especies de plantas permitió identificar 2 219 genes específicos de *E. uniflora*, de los cuales el 40% fueron anotados funcionalmente con 1 772 términos ontológicos. Por otro lado, utilizando la información de la anotación de los genes predichos, se identificaron 20 genes de TPS que representan a las subfamilias TPS-b, TPS-c, TPS-f y TPS-g. Los datos generados en este trabajo permitirán comprender y conocer los mecanismos moleculares involucrados en la adaptabilidad a ambientes naturales y en la síntesis de compuestos con importancia farmacológica en pitanga.

Palabras clave: Myrtaceae, secuenciamiento de DNA, ensamblaje de DNA, predicción génica, genómica comparativa, terpeno sintasas.



## ABSTRACT

Pitanga (*Eugenia uniflora* L.), a member of the Myrtaceae family, is distributed in the Atlantic Forest and is characterized by the synthesis of secondary metabolites with important biological activities, as well as by having a wide ecological spectrum in environments that are climatically contrasting. However, the availability of genomic resources to study this species is still limited and scarce compared to other plant species. For this reason, this research aims to obtain a reference genome for *E. uniflora* and identify the members of the terpene synthase (TPS) family that are responsible for the synthesis of terpenes. *De novo* assembly of the genome, from reads sequenced using the Illumina platform, allowed to obtain a total size of 385.1 Mbp with an N50 value equal to 26 199 bp. Furthermore, using Viridiplantae proteins as a reference, was possible to obtain 30,663 pitanga protein coding genes. A comparative genomics analysis with other plant species identified 2 219 genes specific to *E. uniflora*, 40% of which were functionally annotated with 1 772 ontological terms. On the other hand, using the predicted gene annotation information, 20 TPS genes representing the TPS-b, TPS-c, TPS-f, and TPS-g subfamilies were identified. The data generated in this work may allow us to understand and know the molecular mechanisms involved in the adaptability to natural environments and in the synthesis of compounds with pharmacological importance in pitanga.

Keywords: Myrtaceae, DNA sequencing, DNA assembly, gene prediction, comparative genomics, terpene synthases.

## 1. INTRODUCCIÓN

*Eugenia uniflora* (Myrtaceae), conocida comúnmente como pitanga o nangapiri, es una especie nativa de la Mata Atlántica y se distribuye naturalmente en Brasil (desde el estado de Rio Grande do Sul hasta el estado de Pernambuco), Argentina, Uruguay y Paraguay (Consolini *et al.*, 2002; Bicas *et al.*, 2011). La pitanga produce frutos semejantes a una cereza que son consumidos *in natura* o son utilizados en la fabricación de jugos, helados y licores (Lim, 2012). Otro aspecto importante de esta especie reside en el punto de vista medicinal y farmacéutico, debido a la presencia de diversos metabolitos secundarios en las hojas, incluyendo muchos aceites volátiles como los terpenos, flavonoides, taninos y esteroides (Amorin *et al.*, 2009). En el caso de su zona de distribución, la pitanga tiene un amplio espectro ecológico y puede encontrarse en ambientes que son climáticamente contrastantes. En ese contexto, esta especie puede ser encontrada como arbusto o un árbol pequeño en la vegetación de la planicie costera arenosa o restinga del sudoeste y sur de Brasil, o con un porte arbóreo cuando se encuentra en la vegetación ribereña de la región sur de Brasil, Paraguay y Argentina (Salgueiro *et al.*, 2004). Debido a las propiedades biológicas de sus metabolitos secundarios, así como a su capacidad de desarrollarse en diferentes tipos de hábitats y de alterar la composición de sus aceites esenciales por causa de los factores ambientales (Costa *et al.*, 2009), *E. uniflora* puede considerarse como una importante especie modelo para futuros estudios de variación adaptativa en plantas en respuesta a factores abióticos (estrés salino e hídrico).

A nivel genómico, la disponibilidad de secuencias nucleotídicas de *Eugenia uniflora* es escasa. El número de secuencias disponibles de esta especie es muy pequeño en comparación con las secuencias de *Eucalyptus grandis*, la cual es otra especie perteneciente a la familia Myrtaceae y con un genoma ya secuenciado (Myburg

*et al.*, 2014). Algunas especies de la familia Myrtaceae son buenas candidatas como modelos en el estudio y comprensión de diferentes procesos biológicos que ocurren en las plantas por ser diploides y tener un genoma pequeño (Da Costa *et al.*, 2008). Por esos motivos, el presente proyecto tiene como objetivo secuenciar y anotar el genoma nuclear de *E. uniflora* para obtener el catálogo de genes característicos de esta especie y así poder identificar las terpeno sintasas, enzimas responsables de la síntesis y variedad de terpenos presentes en esta especie. Asimismo, este trabajo permitirá generar recursos genómicos de esta especie que podrán ser utilizados en futuros estudios.

## 2. MARCO TEÓRICO

### 2.1 Generalidades

#### 2.1.1 Posición taxonómica

Reino: Viridiplantae

Filo: Streptophyta

Subfilo: Streptophytina

Clase: Magnoliopsida

Orden: Myrtales

Familia: Myrtaceae

Subfamilia: Myrtoideae

Tribu: Myrteae

Género: *Eugenia*

#### 2.1.2 La familia Myrtaceae

La familia Myrtaceae está constituida de géneros con importancia económica y ecológica que se encuentran distribuidos en las regiones neotropical y australiana (Lucas *et al.*, 2015). A nivel mundial presenta 5 671 especies agrupadas en 132 géneros (Govaerts *et al.*, 2008); mientras que en Brasil representa una de las principales familias de flora con 26 géneros y aproximadamente 1 000 especies (Souza y Lorenzi, 2005). Algunos géneros se caracterizan por tener un alto número de especies, como en el caso de *Syzygium*, *Eugenia* y *Eucalyptus* que contienen 1,500, 1,050 y 700 especies, respectivamente (Brooker 2000; Craven y Biffin, 2010). Mientras que, el género *Eugenia*, integrante de la subfamilia Myrtoideae, está constituida por cerca de 400 especies en

Brasil (Henriques *et al.*, 1993); la mayoría de estas especies son utilizadas principalmente como plantas medicinales por la presencia de metabolitos secundarios en los aceites esenciales de sus hojas y frutos (Adebajo *et al.*, 1989). A nivel morfológico, las especies de *Eugenia* se caracterizan por presentar un porte que varía de árbol a arbusto, con flores tetrámeras o pentámeras, solitarias o en racimos, cáliz abierto o cerrado, ovario bilocular con una o dos semillas y embriones con 2 cotiledones (Braga, 1985).

### **2.1.3 *Eugenia uniflora***

*Eugenia uniflora*, pitanga o pitangueira, es una especie nativa de la Mata Atlántica y crece en regiones de clima tropical y subtropical, pudiendo ser encontrada desde el centro de Brasil hasta el norte de Argentina (Bezerra *et al.*, 2000). Consecuencia de su adaptabilidad a diferentes tipos de ambientes, ha sido diseminada ampliamente y actualmente puede ser encontrada en diferentes partes del mundo como un árbol pequeño o un arbusto en el arenal, así como árbol en la vegetación de la ribera (Salgueiro *et al.*, 2004). Por esta plasticidad ambiental, esta especie se presenta como un excelente modelo en el aspecto genético y ecológico en la comprensión de la adaptabilidad al medio ambiente e interrelaciones entre las plantas y los factores abióticos (como el estrés hídrico y salino). En el aspecto económico, además de que el fruto de la pitanga es consumido *in natura* o utilizado en la fabricación de jugos, helados y licores (Lim, 2012), los extractos de hojas son usados en la medicina folclórica para el tratamiento de algunas enfermedades, como enfermedades intestinales, presión sanguínea alta, fiebre y diarrea, debido a la presencia de diferentes tipos de metabolitos secundarios en sus aceites esenciales (Consolini *et al.*, 1999). En este contexto, *E. uniflora* es el foco de diversos estudios fitoquímicos que describen los diferentes tipos de propiedades farmacológicas en sus aceites esenciales (Lim, 2012). Otro dato importante obtenido de estos estudios es que la concentración de los metabolitos

secundarios, especialmente de los terpenos, pueden variar entre los individuos de diferentes regiones como consecuencia de la composición del suelo, altitud, temporada de colecta y otros factores (Costa *et al.*, 2009).

## 2.2 Los terpenos

Los terpenos, también conocidos como terpenoides o isoprenoides, constituyen una gran clase de productos naturales de origen vegetal con una alta diversidad funcional y más de 20 000 miembros (Schwab, 2003). Todos los terpenos están basados en unidades de 5 carbonos (C5), como el isopentenil pirofosfato (IPP) o difosfato de dimetilalilo (DMAPP), y sus esqueletos carbonados son elaborados a partir de la unión de dos o más de estas unidades (Liu *et al.*, 2014). Los terpenos están divididos en los volátiles monoterpenos (C10) y sesquiterpenos (C15), en los menos volátiles diterpenos (C20), y los no volátiles triterpenos y esteroides (C30), pigmentos carotenoides (C40) y poliisoprenos (Cn) (Harborne, 1984). En plantas, el IPP o DMAPP son sintetizados por medio de dos vías compartimentalizadas: la vía del mevalonato (en el retículo endoplasmático y los peroxisomas), y la vía del metileritritolo fosfato (MPE) (en los plastidios). La condensación del IPP y DMAPP es catalizada por las preniltransferasas (PTs) y tienen como resultado tres principales intermediarios de la vía de los terpenos: farnesil difosfato (FDP), geranil pirofosfato (GPP) y geranylgeranil pirofosfato (GGPP) (Ogura *et al.*, 1997). Las centenas de esqueletos básicos típicos de los terpenos de las plantas son formadas a partir del DMAPP, GPP, FDP o GGPP por las terpeno sintasas o TPS (Davis y Croteau, 2000). Las TPS forman el hemiterpeno isopreno (C5) a partir del DMAPP (Miller *et al.*, 2001), monoterpenos (C10) a partir del GPP (Wise y Croteau, 1999), sesquiterpenos (C15) a partir del FDP (Cane, 1999), y diterpenos (C20) a partir del GGPP (Macmillan y Beale, 1999).

### **2.3 Las terpeno sintasas**

En la actualidad, el aislamiento de los genes de las TPS y su expresión heteróloga han permitido una mejor evidencia en la caracterización de la función y proceso catalítico de las enzimas de esta familia. A partir de estos estudios, se observó que las TPS poseen la capacidad de generar múltiples productos diferentes a partir del mismo sustrato consecuencia de la flexibilidad conformacional de su centro activo (Degenhardt *et al.*, 2009). Recientemente, los estudios a nivel genómico permitieron la identificación de los genes de las TPS en plantas con genomas secuenciados: 32, 69, 113 y 106 potenciales genes de TPS fueron reportados en *Arabidopsis thaliana*, *Vitis vinífera*, *Eucalyptus grandis* y *Eucalyptus globulus*, respectivamente (Aubourg *et al.*, 2002; Martin *et al.*, 2010; Külheim *et al.*, 2015).

### **2.4 Secuenciamiento de genomas en plantas**

El secuenciamiento de un genoma de referencia puede proveer el acceso al catálogo completo de genes de una especie, a los elementos regulatorios que controlan sus funciones y a un marco referencial para comprender la variación genómica a nivel poblacional (Feuillet *et al.*, 2011). Al mismo tiempo, es un prerrequisito para una mejor comprensión de las funciones de los genes en el desarrollo, generar abordajes a nivel de sistemas biológicos y explorar eficientemente la diversidad genética natural e inducida de un organismo. Los primeros proyectos de secuenciamiento de genoma fueron muy costosos y llevaron años para ser concluidos, pero el desarrollo de las tecnologías de secuenciamiento de la próxima generación (NGS) aumentó significativamente la producción de corridas de secuenciamiento a un costo mucho menor (Quail *et al.*, 2012). En ese sentido, la aplicación del NGS revolucionó la velocidad y resolución de los proyectos de secuenciamiento en plantas, ya sea utilizando solo esta tecnología o en combinación con el secuenciamiento clásico de Sanger (Bolger *et al.*,

2014). En los últimos años, el secuenciamiento por Illumina fue considerado como la plataforma de secuenciamiento dominante debido al alto número de secuencias generadas por corrida y por la alta calidad de las secuencias, siendo usado en la generación de datos genómicos en diversas especies de plantas: papa (PGS Consortium, 2011), plátano (D'Hont *et al.*, 2012), naranja (Xu *et al.*, 2013), sandía (Guo *et al.*, 2013), entre otros.

## 2.5 Genómica de Myrtaceae

En los últimos años se han reportado el secuenciamiento y ensamblaje de genomas de diferentes especies de la familia Myrtaceae utilizando NGS. A nivel de tribus se han reportado los siguientes genomas: Eucalypteae (*Eucalyptus grandis* con 640 Mbp y *Eucalyptus pauciflora* con 594.9 Mbp), Leptospermeae (*Leptospermum scoparium* con 297 Mbp), Myrteae (*Eugenia uniflora* con 3.15 Mbp y *Psidium guajava* con 303.8 Mbp), Metrosidereae (*Metrosideros polymorpha* con 304 Mbp), y Melaleuceae (*Melaleuca alternifolia* con 362 Mbp) (Myburg *et al.*, 2014; Izuno *et al.* 2016; Stefenon *et al.*, 2019; Thrimawithana *et al.*, 2019; Wang *et al.*, 2020; Thakur *et al.*, 2021; Voelker *et al.*, 2021; Zhang *et al.*, 2022).

## 2.6 Recursos genómicos en *E. uniflora*

A nivel genómico, la disponibilidad de secuencias nucleotídicas de *Eugenia uniflora* es limitada. Hasta la actualidad, solo 245 secuencias están disponibles en el banco de datos GenBank de la *National Center for Biotechnology Information* (NCBI), siendo la mayoría secuencias de marcadores moleculares del tipo microsatélites (Ferreira-Ramos *et al.*, 2008) y secuencias de genes obtenidos a partir de estudios filogenéticos (Clausing y Renner, 2001; Wilson *et al.*, 2001; Van der Merwe *et al.*, 2005;



Biffin *et al.*, 2006; Lucas *et al.*, 2007; Rutschmann *et al.*, 2007; Soh y Parnell, 2011; Da Cruz *et al.*, 2013; Kitson *et al.*, 2013). Por otro lado, recientemente fueron identificadas 42 secuencias de pre-miRNAs y sus respectivos blancos (Guzman *et al.*, 2012), así como el transcriptoma de hoja (Guzman *et al.*, 2014), el genoma de cloroplasto (Eguiluz *et al.*, 2017), y 2,601 *contigs* de tamaño pequeño y no anotados del genoma nuclear (Stefenon *et al.*, 2019) utilizando las tecnologías de secuenciación de la próxima generación (NGS).

### 3. HIPÓTESIS

H1: Es posible identificar los genes de las terpeno sintasas de *E. uniflora* utilizando el genoma ensamblado y anotado.

Ho: No es posible identificar los genes de las terpeno sintasas de *E. uniflora* utilizando el genoma ensamblado y anotado.

### 4. OBJETIVOS

#### **Objetivo general:**

- Obtener el genoma de *E. uniflora* para identificar los miembros de la familia de terpeno sintasas.

#### **Objetivos específicos:**

- Ensamblar el genoma de *E. uniflora* utilizando lecturas secuenciadas en la plataforma Illumina.
- Anotar el genoma de *E. uniflora* utilizando secuencias proteicas de otras especies de Viridiplantae y herramientas bioinformáticas.
- Validar el ensamblaje y la predicción génica utilizando secuencias de proteínas ortólogas de otras especies de la familia Viridiplantae.
- Realizar un análisis de genómica comparativa de *E. uniflora* con *E. grandis* y otras especies representativas de plantas.
- Identificar las terpeno sintasas presentes en el genoma de *E. uniflora*.

## 5. MATERIALES Y MÉTODOS

### 5.1 MATERIALES

#### 5.1.1 Material biológico

Las hojas jóvenes de un árbol de pitanga fueron colectadas en el Campus do Vale de la *Universidade Federal do Rio Grande do Sul* (UFRGS), localizado en la ciudad de Porto Alegre, estado de Rio Grande del Sur, Brasil (latitud S: 30°4'2.71"; longitud W: 51°7'11.88") (Fig. 1). Una muestra de este individuo fue depositada en el Herbario del *Instituto de Ciencias Naturais da UFRGS*. Las hojas colectadas fueron colocadas en tubos de Falcon de 50 ml (Eppendorf) y almacenadas en nitrógeno líquido para así preservar la integridad del DNA hasta la etapa de la extracción.



**Figura 1.** Hojas del árbol de pitanga que fueron colectadas para la extracción del DNA.

## 5.2 MÉTODOS

### 5.2.1 Extracción de DNA genómico

El DNA fue extraído a partir de las hojas colectadas y almacenadas en nitrógeno líquido utilizando el método de CTAB (Doyle y Doyle, 1990) en el Laboratorio de Genética de Poblaciones y Plantas (LGPP) del Centro de Biotecnología de la UFRGS. En este método, se agregó 1000 µl de buffer de extracción de bromuro de hexadeciltrimetilamonio (2% CTAB, 1.4 M NaCl, 0.2% 2-mercaptoetanol, 20 mM EDTA, 100 mM Tris-HCl, pH 8.0) a 100 mg de tejido foliar macerado en nitrógeno líquido. Después de una incubación de 30 minutos a 60°C, se adicionó cloroformo-alcohol isoamílico (24:1) y mediante una centrifugación se separaron las fases de la solución. La fase acuosa fue removida y transferida a un tubo de centrifuga y se agregaron 2/3 de volumen de isopropanol helado para precipitar los ácidos nucleicos. Posteriormente, los ácidos nucleicos fueron transferidos a otro tubo conteniendo el buffer de lavado (76% de etanol y 10 mM de acetato de amonio) para 1 o 2 lavados y finalmente resuspendidos en 1 ml de TE (10 mM Tris-Ha, 1 mM EDTA, pH 7.4). Adicionalmente, se agregó RNAasa A (concentración final de 10 µg/ml) y se incubó por 30 minutos a 37°C para así eliminar cualquier contaminación de RNA presente en la extracción. La calidad del DNA extraído fue evaluada en un gel de 1% de agarosa y su cuantificación fue realizada en un espectrofotómetro Nanodrop (Nanodrop Technologies, USA) a 260 nm y la pureza fue comprobada con la proporción A260/A280.

### 5.2.2 Secuenciamiento del DNA

El DNA total extraído y cuantificado fue enviado a la empresa Macrogen (Corea) para el servicio de preparación y secuenciamiento de una biblioteca *paired-end* (PE) y otra biblioteca *mate-paired* (MP). En el caso de la biblioteca PE, 10 µg de DNA fueron

fragmentados a un tamaño máximo de 500 bp por sonicación y el *TruSeq DNA Sample Preparation Kit* de Illumina fue usado para construir la biblioteca PE con un tamaño de inserto de 550 bp. Por otro lado, en el caso de la biblioteca MP, 10 µg de DNA fueron fragmentados a un tamaño máximo de 8 Kbp y el *Illumina Mate Pair v2 kit* fue usado para la construcción de la biblioteca con un tamaño de inserto de 8 Kbp. Después de evaluar la calidad de las bibliotecas, se procedió con el secuenciamiento utilizando el secuenciador Illumina HiSeq2000.

### 5.2.3 Calidad y limpieza de las bibliotecas

Además de las dos bibliotecas que fueron secuenciadas, en el ensamblaje del genoma se utilizó una biblioteca *paired-end* (con un tamaño de inserto de 250 bp) secuenciada a partir del DNA del mismo individuo que fue utilizada por Eguiluz *et al.* (2017) para la obtención del genoma de cloroplasto de pitanga. La calidad de las lecturas y la presencia de adaptadores en todas las bibliotecas fueron visualizadas utilizando el programa FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Este programa es una aplicación JAVA de 11 módulos que importa datos en formato BAM, SAM y FASTQ y genera un reporte en HTML que contiene gráficos y estadísticas de los datos importados (Anexo 1). Posteriormente, las secuencias adaptadoras contaminantes aun presentes, así como las lecturas de baja calidad con un valor de *phred* menor de 30, fueron removidas utilizando los programas Nextclip (Leggett *et al.*, 2014) y Trim galore! ([http://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore/](http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/)) en las bibliotecas MP y PE, respectivamente. Ambos programas reconocen y procesan lecturas en formato FASTA y FASTQ utilizando diferentes herramientas que son ejecutadas mediante líneas de comandos.

### 5.2.4 Estimación del tamaño del genoma y heterocigosidad

Un análisis de conteo de k-mers fue realizado con las lecturas limpias de las dos bibliotecas PE utilizando el programa Jellyfish (Marçais y Kingsford, 2011) utilizando un tamaño de  $k = 21$ . Después de obtener las frecuencias de 21 k-mer en todas las lecturas, estas fueron convertidas en un histograma que posteriormente fue analizado con el programa GenomeScope (Vurture *et al.*, 2017) para obtener estimaciones del tamaño de genoma, heterocigosidad, y secuencias únicas y repetidas de *E. uniflora*.

### 5.2.5 Ensamblaje del genoma

El genoma de pitanga fue ensamblado *de novo* utilizando el programa Platanus (Kajitani *et al.*, 2014) utilizando parámetros por defecto. En un inicio, este programa ensambla *contigs* utilizando las lecturas de las bibliotecas PE mediante extensiones de *kmers* que permitieron la construcción de grafos de Bruijn. Posteriormente, el ordenamiento de los *contigs* que generaron los *scaffolds* se determinó utilizando la información de las bibliotecas PE y MP, donde el tamaño del inserto fue estimado utilizando los pares de lecturas que mapean en el mismo *contig*. Finalmente, los *gaps* presentes fueron cerrados utilizando las lecturas que mapean en los *scaffolds*. Las métricas de calidad del ensamblaje fueron obtenidas utilizando el programa QUAST (Gurevich *et al.*, 2013) considerando un tamaño mínimo de *scaffold* de 800 bp. Este programa utiliza el alineador MUMmer para evaluar la calidad de un genoma ensamblado en formato FASTA y producir métricas basadas o independientes de un genoma de referencia (número de *scaffolds*, N50, L50, etc).

### 5.2.6 Identificación de secuencias repetitivas en el genoma

Antes de realizar la predicción de los genes codificadores de proteínas, se procedió con la identificación *de novo* de las secuencias repetitivas en el genoma de pitanga utilizando el programa RepeatModeler con parámetros por defecto

(<https://github.com/rmhubble/RepeatModeler>). Este programa realiza una búsqueda de las secuencias repetitivas presentes en el genoma que se desea enmascarar utilizando una búsqueda por BLAST. Posteriormente, se utilizó el programa RepeatScout con parámetros por defecto (<http://www.repeatmasker.org/>) para construir una librería de repeticiones utilizando los resultados obtenidos previamente. Este programa cuenta cada secuencia con un tamaño de 12 bp y luego las extiende para así obtener secuencias consenso. A continuación, se filtraron las repeticiones simples y en tándem de la librería de repeticiones preliminar. Esta librería preliminar fue usada en el programa RepeatMasker (<http://www.repeatmasker.org/>) para realizar una búsqueda preliminar en el genoma y así eliminar las secuencias repetitivas que estén presentes al menos de diez veces en todo el genoma. Después de este nuevo filtro, para garantizar que las familias de proteínas conservadas no se cataloguen erróneamente como repeticiones, se utilizó el programa BLAST2GO (Conesa *et al.*, 2005) para determinar la similitud entre las repeticiones putativas y cualquier proteína conocida utilizando BLAST. Por otro lado, las repeticiones fueron clasificadas utilizando el programa en línea TEclass (Abrusán *et al.*, 2009). Este programa clasifica los elementos transponibles (TE) desconocidos en cuatro categorías de acuerdo a su mecanismo de transposición: transposones de DNA, LTRs (repeticiones de terminal larga), LINEs (secuencias repetitivas dispersas largas) y SINEs (secuencias repetitivas dispersas cortas). Finalmente, utilizando la librería de repeticiones filtrada y clasificada, se procedió con el enmascaramiento suave del genoma de pitanga utilizando nuevamente el programa RepeatMasker.

### 5.2.7 Anotación y validación del genoma

Para la predicción de genes en el genoma enmascarado suavemente, se combinaron métodos de predicción *ab initio* y basados en homología utilizando el pipeline BRAKER2 (Brůna *et al.*, 2021). Este *pipeline* permitió el entrenamiento completamente automático de las herramientas de predicción génica Augustus (Stanke

*et al.*, 2006) y GeneMark-ES/ET/EP (Lomsadze *et al.*, 2005) utilizando datos de RNA-seq o de homología de proteínas, y luego se integró la evidencia extrínseca de estos mismos datos en la predicción. La predicción *ab initio* se llevó a cabo utilizando Augustus, y en el caso de la predicción basada en homología se generaron indicios para el programa GeneMark-ES/ET/EP con el programa ProHint (Bruna *et al.*, 2020) utilizando como referencia las proteínas de plantas del banco de datos de secuencias OrthoDB ([https://v100.orthodb.org/download/odb10\\_plants\\_fasta.tar.gz](https://v100.orthodb.org/download/odb10_plants_fasta.tar.gz)) así como también las secuencias de proteínas del genoma de *Eucalyptus grandis* que fueron obtenidos del banco de datos Phytozome (<https://phytozome-next.jgi.doe.gov/>). ProHint es un programa que permite predecir y puntuar indicios (en forma de intrones, y codones de inicio y terminación) en un determinado genoma por medio del mapeo y alineamiento de los genes predichos a un banco de datos de secuencias de proteínas de referencia. A continuación, el programa Augustus fue entrenado utilizando los indicios generados previamente con parámetros por defecto y posteriormente fue usado para realizar la predicción. A esto le siguió otra ronda de entrenamiento y de predicción basado en los resultados de la primera iteración y los indicios de GeneMark-ES/ET/EP. A continuación, utilizando el archivo de anotación en formato GTF que fue generado por el programa BRAKER2, se filtraron las secuencias codificadoras que no tenían codones de inicio y terminación, así como aquellos genes que tenían exones sobrepuestos a otros, utilizando el programa gFACs (Caballero y Wegrzyn, 2019). Este programa permitió filtrar, analizar y convertir las anotaciones de los genomas por medio de alineamientos y marcos de predicción génica. Asimismo, fueron excluidos los genes que tenían un 95% de su tamaño sobrepuestos a las regiones enmascaradas del genoma utilizando *scripts in house* y el programa PRINSEQ (Schmieder y Edwards, 2011). Este programa está compuesto por un conjunto de herramientas que permiten filtrar, reformatear o editar datos de secuencias genómicas y metagenómicas, integrando un sumario de estadísticas a las secuencias en formato gráfico y tabular. Finalmente, la calidad del ensamblaje y la predicción de genes fue validado con el programa BUSCO (Seppey *et*



*al.*, 2019) utilizando como referencia el banco de secuencias de Viridiplantae. Este programa permite cuantificar la integridad de los genomas y transcriptomas ensamblados, así como las predicciones de genes codificadores de proteínas, identificando coincidencias en el conjunto de genes que se esperan estén presentes como ortólogos de copia única en un determinado grupo taxonómico.

### 5.2.8 Análisis de genómica comparativa

La plataforma web OrthoVenn2 (Xu *et al.*, 2019) fue usada de manera local con parámetros por defecto para generar clústeres de ortólogos utilizando las proteínas predichas del genoma de *E. uniflora*, así como las proteínas de los genomas de *Arabidopsis thaliana* (L.) Heynh, *Eucalyptus grandis* W. Mill ex Maiden, *Populus trichocarpa* Torr. & A. Gray y *Vitis vinifera* L. que fueron descargadas del banco de datos Phytozome (<https://phytozome-next.jgi.doe.gov/>). Las proteínas que pertenecen a los clústeres de ortólogos específicos de pitanga fueron anotadas funcionalmente utilizando el programa BLAST2GO como fue descrito previamente.

### 5.2.9 Identificación de las terpeno sintasas

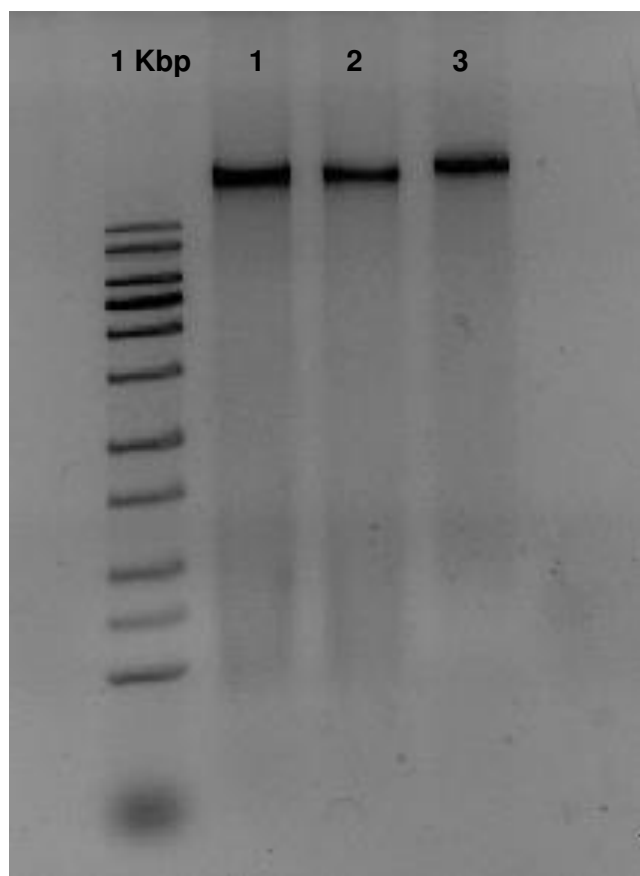
Las secuencias de las TPS fueron identificadas en los genes predichos en el genoma de *E. uniflora* buscando las familias Pfam PF01397 (dominio N terminal de las TPS) y PF03936 (dominio C terminal de las TPS) utilizando el programa BLAST2GO. Las secuencias de las TPS identificadas fueron utilizadas conjuntamente con las 113 TPS identificadas de *E. grandis* para realizar un análisis filogenético y así inferir por analogía la subfamilia de TPS a la que pertenecen. Todos los alineamientos de las secuencias de nucleótidos y proteínas de las TPS fueron realizados utilizando parámetros por defecto en el programa MEGA (Tamura *et al.*, 2011). Posteriormente, un análisis bayesiano fue realizado utilizando el programa BEAST (Drummond y Rambaut,

2007) usando los alineamientos realizados previamente. Los modelos de evolución proteica y de sustitución de nucleótidos fueron seleccionados usando los programas ProTest3 (Abascal *et al.*, 2005) y jModelTest2 (Darriba *et al.*, 2012), respectivamente. ProTest3 es una herramienta escrita en Java que permite seleccionar el mejor modelo de reemplazo de aminoácidos para un determinado conjunto de datos utilizando el programa PhyML (Guindon y Gascuel, 2003) para la estimación de máxima verosimilitud de los árboles filogenéticos y parámetros de los modelos. Por otro lado, jModelTest2 es un programa que permite realizar la selección estadística de los mejores modelos de sustitución de nucleótidos implementando diferentes estrategias de selección de modelos diferentes, además proporciona estimaciones de incertidumbre en la selección del modelo y calcula los parámetros promedio de los modelos. Posteriormente, el proceso Yule fue seleccionado antes del análisis bayesiano y fueron realizadas 10 000 000 de generaciones usando algoritmos de MCMC. Finalmente, los árboles filogenéticos fueron visualizados usando el programa FIGTREE (<http://tree.bio.ed.ac.uk/software/figtree/>). Este programa es un visualizador gráfico de árboles filogenéticos que permite editar y generar figuras para las publicaciones científicas.

## 6. RESULTADOS

### 6.1.1 Extracción y secuenciamiento de DNA genómico

Utilizando el método de extracción a partir de tres diferentes hojas colectadas del mismo individuo, así como un posterior tratamiento con RNAasa A, se obtuvo DNA genómico de buena calidad y concentración (100 a 300 ng/ $\mu$ l) (Fig. 2). Por otro lado, la lectura de 1.80 obtenida en la proporción A260/A280 nos indicó que la pureza del DNA extraído es óptima. Posteriormente, este DNA fue liofilizado y enviado por vía aérea a la empresa Macrogen (Corea) para un servicio de preparación y secuenciamiento de las bibliotecas PE y MP.



**Figura 2.** Calidad del DNA de tres diferentes hojas colectadas del mismo individuo de pitanga en gel de agarosa al 1%. Carril 1Kbp: marcador de peso de 1Kbp de Promega); carril 1-3: muestras de hojas.

### 6.1.2 Calidad y limpieza de las bibliotecas

Los datos del secuenciamiento de las bibliotecas PE de 550 bp de inserto y MP fueron descargados del sitio FTP de la empresa Macrogen. El reporte del secuenciamiento de estas dos bibliotecas está disponible en el Anexo 2. El tamaño de las lecturas de las bibliotecas PE de 250 bp y 550 bp fueron de 100 bp y 150 bp, respectivamente. En el caso de la biblioteca MP, el programa Nextclip clasificó los pares de lecturas de 150 bp de tamaño en cuatro categorías dependiendo de la presencia o no de los adaptadores utilizados en la construcción de las bibliotecas. Después de eliminar los adaptadores, 46 512 167 (40.94%) pares de lecturas de la biblioteca MP son usables para el ensamblaje del genoma y están constituidos por las categorías A, B y C (Tabla 1). Estos pares de lecturas fueron agrupados en dos nuevos archivos FASTQ denominados MP-usable-1 y MP-usable-2, y están constituidos por lecturas no duplicadas y un tamaño mínimo de 25 bp.

**Tabla 1.** Clasificación de los pares de lecturas de la biblioteca MP después de eliminar los adaptadores. Solo se están considerando lecturas que no son duplicadas y tienen un tamaño mínimo de 25 bp.

<b>Categoría</b>	<b>Número de pares de lecturas</b>	<b>Porcentaje</b>
A	6 213 250	5.47
B	19 735 955	17.37
C	20 562 962	18.10
D	26 031 518	22.92

La presencia de adaptadores, la composición nucleotídica y la calidad de las lecturas de las bibliotecas MP-usable, PE de 250 bp y PE de 550 bp antes y después de la limpieza se muestran en el Anexo 1. En el caso de las bibliotecas PE de 250 bp y 550 bp se consideró un tamaño mínimo de 100 bp y 150 bp de las lecturas limpias, respectivamente. En el caso de la biblioteca MP-usable, el tamaño mínimo de las lecturas limpias fue de 25 bp. El número total de las lecturas de cada biblioteca antes y

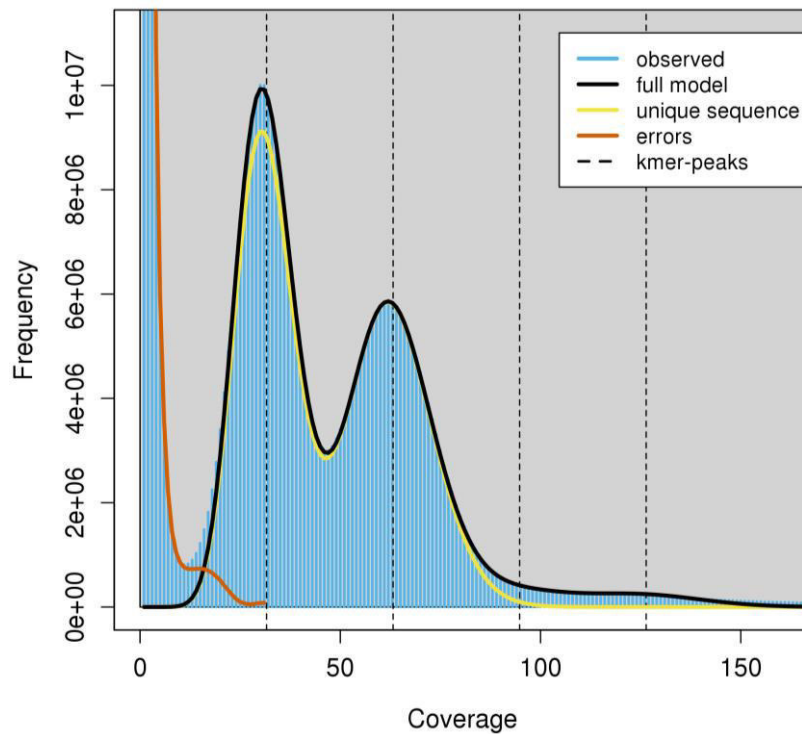
después de la limpieza con el programa Trim galore! se muestra en la Tabla 2.

**Tabla 2.** Número total de pares de lecturas antes y después de la limpieza.

Biblioteca	Número de pares de lecturas secuenciadas	Número de pares de lecturas limpias	Porcentaje
PE-250 bp	94 140 895	79 984 528	84.96
PE-550 bp	182 465 435	169 757 171	93.04
MP-usable	113 599 977	45 718 168	40.24

### 6.1.3 Estimación del tamaño del genoma y heterocigosidad

El tamaño estimado del genoma haploide de *E. uniflora* fue de 311.6 Mbp con un ajuste del modelo del 97.96 % utilizando un análisis de frecuencia por k-mers (k = 21). Por otro lado, se determinó que el genoma tiene un valor de 2.09% de heterocigosidad y el 25.03% de su contenido es repetitivo (Fig. 3).



**Figura 3.** Estimación del tamaño del genoma de pitanga. El primer pico localizado a una cobertura de 32X corresponde al pico heterocigoto, mientras que el segundo pico de 63X corresponde al pico homocigoto.

#### 6.1.4 Ensamblaje del genoma

El genoma ensamblado de *E. uniflora* tiene 385.1 Mbp y está fragmentando en 34 716 *scaffolds*, siendo el de mayor tamaño de 361 660 bp. Las secuencias del genoma fueron depositadas en el banco de datos del NCBI y tiene el número de acceso JAJNMO000000000. La Tabla 3 se muestran las métricas de calidad del ensamblaje.

**Tabla 3.** Estadísticas del ensamblaje del genoma de *E. uniflora*.

Métrica	Genoma de pitanga
Tamaño total	385 104 457
Número total de <i>scaffolds</i>	34 716
Número de <i>scaffolds</i> >= 5000 bp	16 582
Número de <i>scaffolds</i> >= 10000 bp	10 974
Número de <i>scaffolds</i> >= 25000 bp	4 360
Número de <i>scaffolds</i> >= 50000 bp	1 224
<i>Scaffold</i> de mayor tamaño	361 660
GC (%)	39.81
N50	26 199
N75	12 057
L50	4 080
L75	9 489
Número de N's por 100 Kbp	568.72

#### 6.1.5 Identificación de secuencias repetitivas en el genoma

Inicialmente se identificaron 9 888 secuencias repetitivas con una presencia mayor de 10 veces en el genoma ensamblado de pitanga. De este total, 181 secuencias repetitivas fueron excluidas por tener similitud con secuencias codificadoras de proteínas de Viridiplantae (Anexo 3). Utilizando la librería final de secuencias repetitivas,

el programa RepeatMasker identificó que 194.4 Mbp del genoma (50.47%) está conformado por secuencias repetitivas. En la Tabla 4 se muestra la clasificación de las secuencias repetitivas identificadas y enmascaradas en el genoma de pitanga.

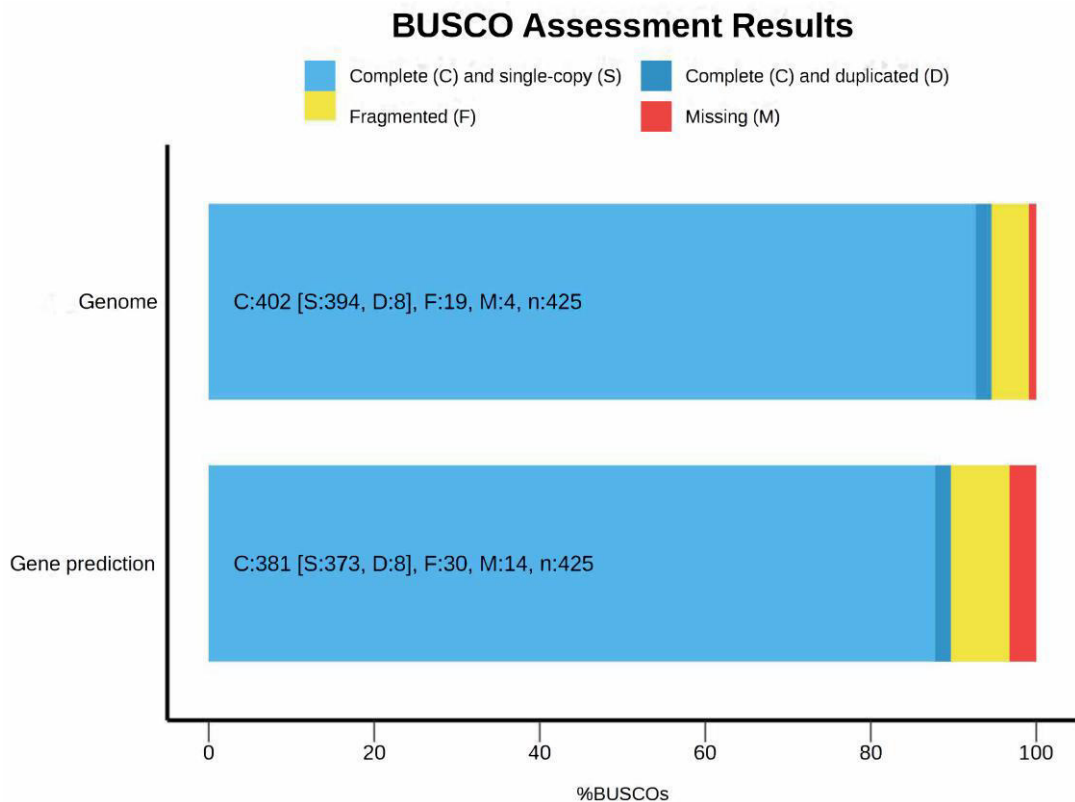
**Tabla 4.** Clasificación de las secuencias repetitivas identificadas en el genoma de *E. uniflora*.

Clase de repetición	Número de elementos	Tamaño ocupado en el genoma	Porcentaje
SINEs	16 586	1 814 643	0.47
LINEs	117 561	23 436 901	6.09
Elementos LTR	124 214	43 706 929	11.35
Elementos de DNA	313 852	56 620 676	14.70
No clasificado	181 781	68 791 662	17.86

### 6.1.6 Anotación y validación del genoma

Utilizando los indicios generados por el programa ProtHint en el genoma de pitanga, fueron predichos 32 201 genes codificadores de proteínas usando el programa BRAKER. Posteriormente, fueron filtrados 95 genes por tener un 95% de su tamaño sobrepuestos a las regiones enmascaradas del genoma. Asimismo, fueron filtrados 414, 245 y 784 genes por no tener codón de inicio, no tener codón de terminación y tener exones sobrepuestos a otros genes, respectivamente. El conjunto final de genes únicos codificadores de proteínas del genoma de *E. uniflora* está constituido por 30 663 genes.

La validación de la calidad del genoma ensamblado por BUSCO dio como resultado un total de 402 (94.6%) genes completos y 19 (4.5%) genes parciales (Fig. 4). Por otro lado, en la validación de los genes predichos se obtuvo un valor de 381 (89.7%) y 30 (7.1%) en los genes completos y parciales, respectivamente.



**Figura 4.** Comparación de los valores obtenidos, en la validación con BUSCO, del genoma ensamblado y de la predicción génica de pitanga.

### 6.1.7 Análisis de genómica comparativa

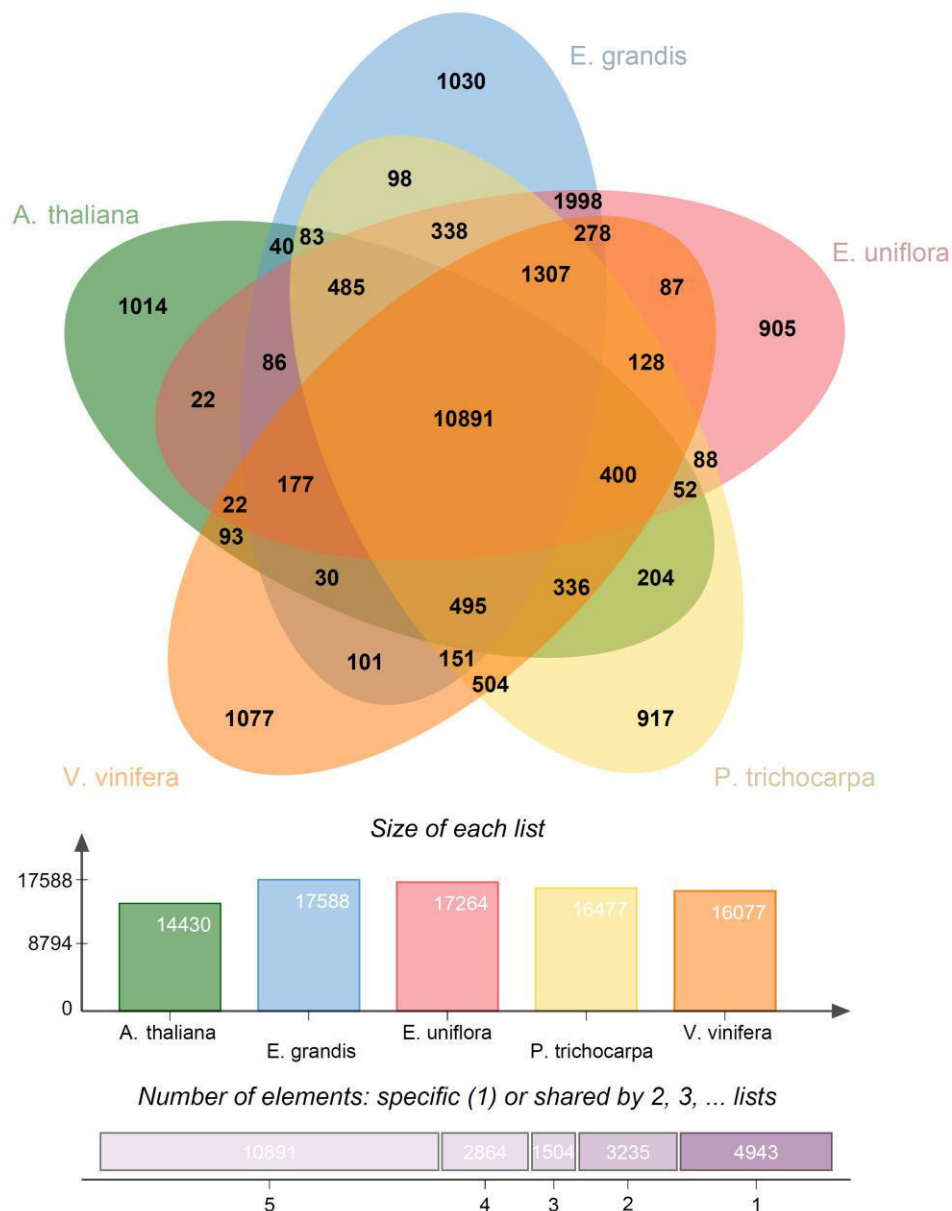
El análisis realizado con el programa OrthoVenn2 identificó que los 30 663 genes predichos en el genoma de *E. uniflora* fueron agrupados en 17 264 *clústeres* y 9 787 *singletons*. En la Tabla 5 se reportan los *clústeres* y *singletons* identificados en las proteínas de los genomas de *A. thaliana*, *E. grandis*, *P. trichocarpa* y *V. vinifera*.

**Tabla 5.** *Clústeres* y *singletons* identificados en las especies de plantas del estudio.

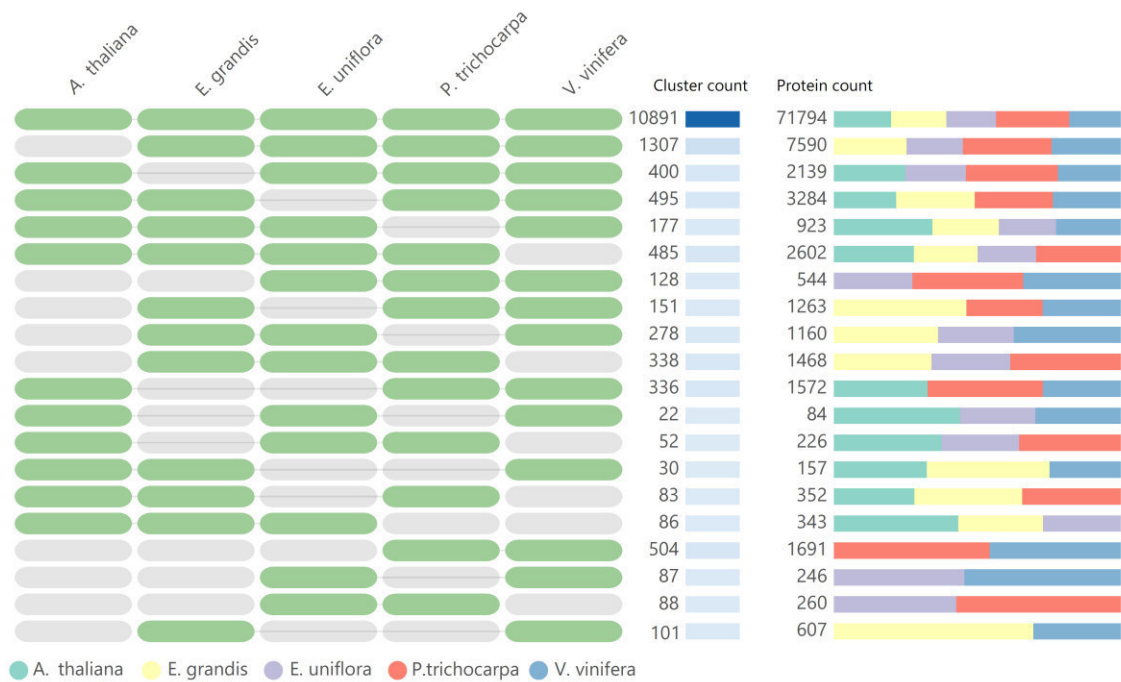
<b>Especie</b>	<b>Proteínas</b>	<b>Clústeres</b>	<b>Singletons</b>
<i>E. uniflora</i>	30 663	17 264	9 787
<i>A. thaliana</i>	27 416	14 430	5 173
<i>E. grandis</i>	34 349	17 588	9 305
<i>P. trichocarpa</i>	34 699	16 477	4 957
<i>V. vinifera</i>	31 845	16 077	9 460



El análisis comparativo realizado permitió identificar 10 891 *clústeres* ortólogos en común entre todas las cinco especies (Fig. 5). Asimismo, se identificaron 1 030 *clústeres* (constituido por 2 219 genes) que son específicos de pitanga. Por otro lado, se encontraron 1 014, 1 030, 917 y 1 077 *clústeres* específicos de *A. thaliana*, *E. grandis*, *P. trichocarpa* y *V. vinífera*, respectivamente. Los valores de *clústeres* y genes de las diferentes comparaciones entre especies se pueden visualizar en la Fig. 6.



**Figura 5.** Diagrama de Venn mostrando los clústeres de genes ortólogos identificados entre las especies de plantas del estudio.



**Figura 6.** Clústeres de ortólogos y número de genes codificadores de proteínas compartidos entre las especies de plantas del estudio.

Por otro lado, en la comparación entre *E. uniflora* y *E. grandis*, se identificaron 15 560 clústeres ortólogos en común entre estas dos especies de Myrtaceae (Anexo 4). Asimismo, 1 704 y 1 382 clústeres son específicos para *E. uniflora* y *E. grandis*, respectivamente.

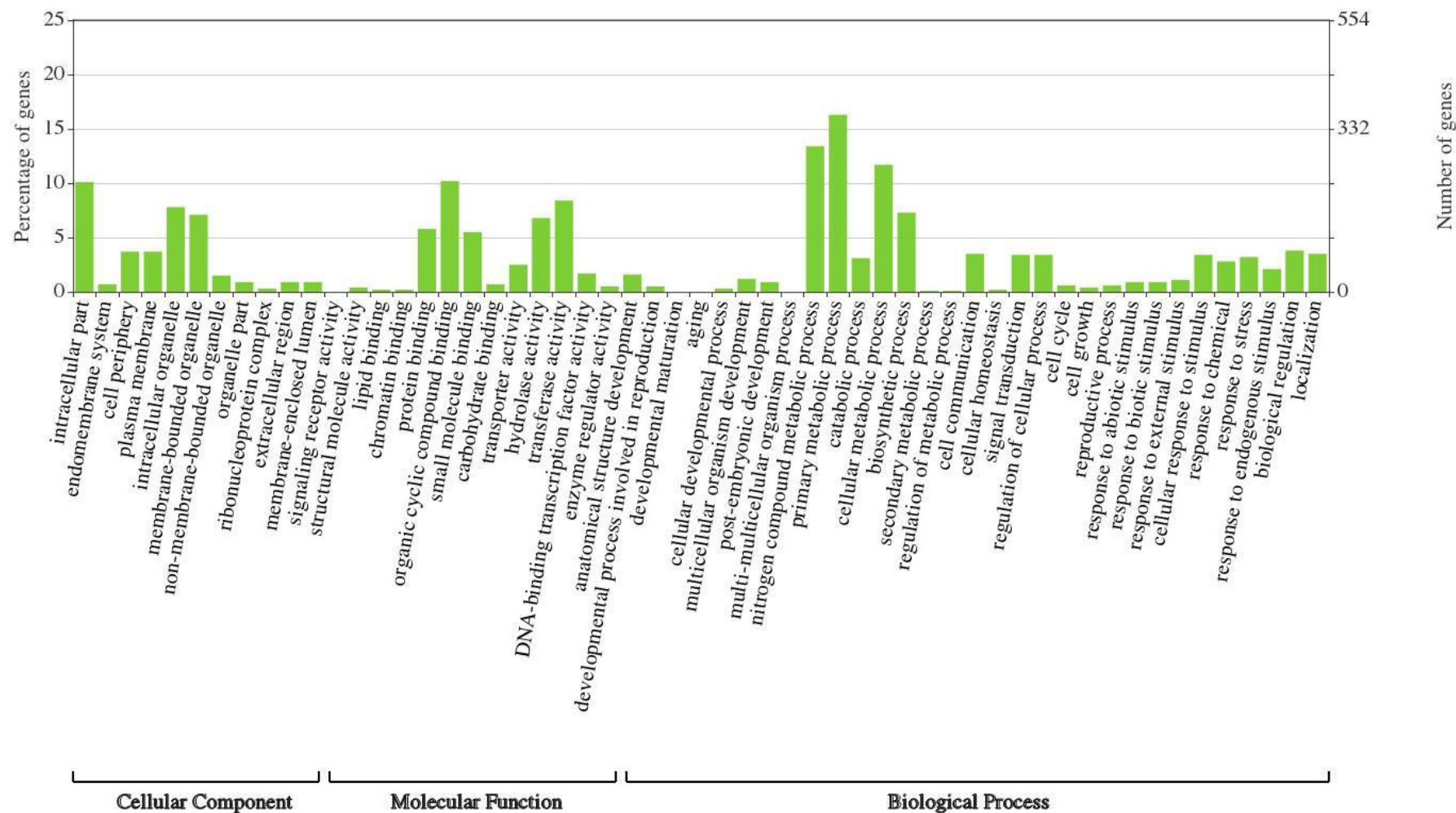
Se realizó la anotación de los 2 219 genes específicos de *E. uniflora* obtenidos en el análisis de genómica comparativa con otras especies de Viridiplantae. Del total, 1 512 genes (68.12%) tuvieron similitud con genes ya reportados en otras especies de plantas (Anexo 5). A nivel de ontología de genes (GO), 888 genes (40.02%) fueron anotados funcionalmente con 1 772 términos ontológicos divididos en las categorías de componente celular, función molecular y proceso biológico (Tabla 6).

**Tabla 6.** Categorías de Ontología de Genes (GO) de los genes exclusivos de pitanga.

<b>Categoría</b>	<b>Numero de términos de GO</b>
Componente celular	460
Función molecular	723
Proceso biológico	589
Total	1 772

A nivel de componente celular, los términos más abundantes fueron los de parte intracelular (GO:0044424), organelo intracelular (GO:0043229) y organelo delimitado por membrana (GO:0043227) con 224, 172 y 157 genes, respectivamente (Fig. 7). Por otro lado, en la categoría de función molecular, los términos que contienen el mayor número de genes fueron los de unión de compuestos orgánicos cíclicos (GO:0097159), actividad transferasa (GO:0016740) y actividad hidrolasa (GO:0016787) con 227, 186 y 150 genes, respectivamente. Asimismo, a nivel de proceso biológico, los términos más abundantes fueron los de procesos metabólicos primarios (GO:0044238), procesos metabólicos de compuestos de nitrógeno (GO:0006807) y procesos metabólicos celulares (GO:0044237) con 361, 298 y 259 genes, respectivamente.

Debido al potencial de pitanga como un modelo biológico en adaptación a diferentes ambientes, es importante señalar que se identificaron los términos de respuesta celular al estímulo (GO:0051716), respuesta al estrés (GO:0006950), respuesta a químicos (GO:0042221), respuesta a estímulos externos (GO:0009605) y respuesta a estímulos abióticos (GO:0009628) con 75, 70, 62, 24 y 19 genes, respectivamente (Anexo 6).



**Figure 7.** Categorías y términos de Ontología de Genes (GO) identificados en los genes exclusivos de *E. uniflora*.

### 6.1.8 Identificación de las terpeno sintasas

Después de realizar la búsqueda de los dominios Pfam PF01397 (dominio N terminal de las TPS) y PF03936 (dominio C terminal de las TPS) en los resultados obtenidos con el programa BLAST2GO, se identificaron 33 genes candidatos a TPS. De este total, fueron eliminados 13 genes que solo tenían unos de los dos dominios PF01397 (4) y PF03936 (9). La Tabla 7 muestra los resultados de BLAST de los 20 TPS candidatos de *E. uniflora* al ser comparados con el banco de secuencias de proteínas no redundante de Viridiplantae del NCBI.

**Tabla 7.** Resultados de BLAST y clasificación de las TPS candidatas de *E. uniflora*.

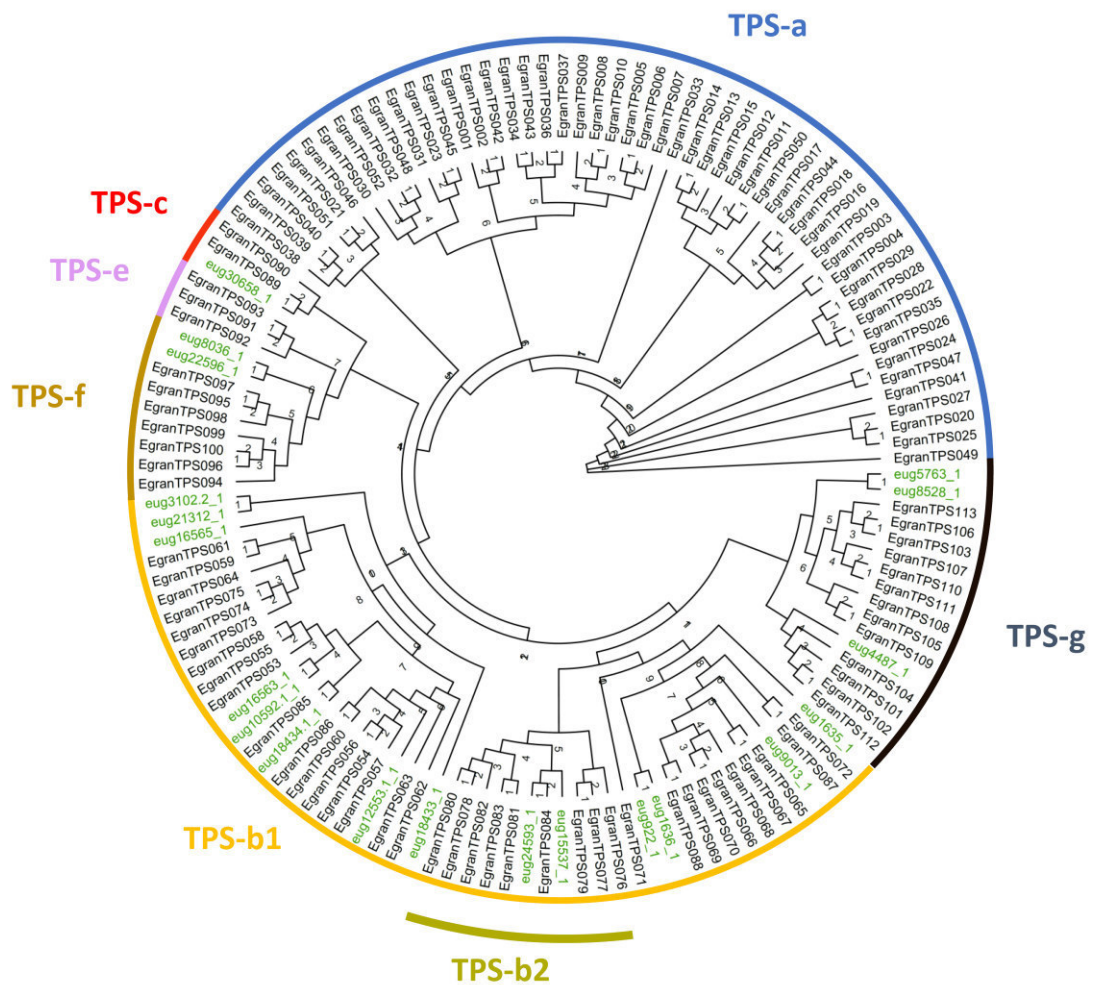
Código	Tamaño (aa)	Descripción	Valor E	Subfamilia de TPS
eug10592.1	593	Linalool synthase [ <i>Backhousia citriodora</i> ]	0	b1
eug12553.1	594	Beta-bisabolene synthase-like [ <i>Eucalyptus grandis</i> ]	0	b1
eug15537	582	Isoprene synthase, chloroplastic-like isoform X2 [ <i>Syzygium oleosum</i> ]	0	b2
eug1635	549	Alpha-farnesene synthase-like [ <i>Rhodamnia argentea</i> ]	0	b1
eug1636	568	Alpha-farnesene synthase-like [ <i>Syzygium oleosum</i> ]	0	b1
eug16563	593	Linalool synthase [ <i>Backhousia citriodora</i> ]	0	b1
eug16565	592	Beta-bisabolene synthase-like [ <i>Rhodamnia argentea</i> ]	0	b1
eug18433	577	Beta-bisabolene synthase-like [ <i>Rhodamnia argentea</i> ]	0	b1
eug18434.1	550	Linalool synthase [ <i>Backhousia citriodora</i> ]	0	b1

eug21312	373	Beta-bisabolene synthase-like [ <i>Rhodamnia argentea</i> ]	0	b1
eug22596	841	S-linalool synthase-like [ <i>Rhodamnia argentea</i> ]	0	f
eug24593	582	Terpene synthase 1 [ <i>Rhodomyrtus tomentosa</i> ]	0	b2
eug30658	639	(-)-Kolavenyl diphosphate synthase TPS28, chloroplastic-like [ <i>Rhodamnia argentea</i> ]	0	c
eug3102.2	416	Beta-bisabolene synthase-like [ <i>Rhodamnia argentea</i> ]	7.00E-143	b1
eug4487	519	(3S,6E)-nerolidol synthase 1-like [ <i>Rhodamnia argentea</i> ]	0	g
eug5763	533	(3S,6E)-nerolidol synthase 1-like [ <i>Syzygium oleosum</i> ]	0	g
eug8036	519	S-linalool synthase-like [ <i>Rhodamnia argentea</i> ]	0	f
eug8528	449	(3S,6E)-nerolidol synthase 1-like [ <i>Syzygium oleosum</i> ]	0	g
eug9013	568	Alpha-farnesene synthase-like [ <i>Rhodamnia argentea</i> ]	0	b1
eug922	568	Alpha-farnesene synthase-like [ <i>Rhodamnia argentea</i> ]	0	b1

El análisis filogenético de los 20 TPS candidatos de pitanga con las TPS de *E. grandis* permitió clasificarlos como miembros de las siguientes subfamilias de TPS: TPS-b (14 genes), TPS-c (1 gen), TPS-f (2 genes) y TPS-g (3 genes) (Tabla 8 y Fig. 8). En el caso de la subfamilia TPS-b, se encontraron 12 miembros del grupo TPS-b1 y 2 miembros del grupo TPS-b2. Por otro lado, la subfamilia TPS-a no fue identificada en las TPS candidatas de *E. uniflora*.

**Tabla 8.** Número de TPS identificadas en *E. uniflora* y *E. grandis*.

Especie	Total de TPS identificadas	TPS completos	a	b	c	d	e	f	g	h
<i>Eugenia uniflora</i>	33	20	0	14	1	0	0	2	3	0
<i>Eucalyptus grandis</i>	172	113	52	36	2	0	3	7	13	0



**Figura 8.** Análisis filogenético de las TPS completas de *E. uniflora* y *E. grandis*. Con letras verdes se indican las TPS de pitanga.

## 7. DISCUSIÓN

En la actualidad, el análisis de los datos de secuenciamiento de DNA pueden ofrecer diversas estadísticas sobre el genoma de una especie antes de realizar un ensamblaje. En ese sentido, el conteo de k-mers realizado en las lecturas limpias de las bibliotecas PE permitió obtener un valor de heterocigosidad de 2.09% en *Eugenia uniflora*. Considerando que cuando este valor es igual o mayor a 0.8% indica que el genoma de esa especie es altamente heterocigótico (Wu *et al.*, 2014), se puede sugerir que el genoma de pitanga tiene altos niveles de heterocigosidad en su genoma. Este valor identificado en pitanga es mucho mayor de los ya reportados para otras especies de la familia Myrtaceae con genoma secuenciado como *Melaleuca alternifolia* (0.8%), *Eucalyptus grandis* (1.0%) y *Eucalyptus pauciflora* (1.5%) (Myburg *et al.*, 2014; Wang *et al.*, 2020; Zhang *et al.*, 2022), y podría ser consecuencia de que pitanga presenta flores hermafroditas que son polinizadas por insectos y que generalmente son entrecruzadas entre si (Silva y Pinheiro, 2009). Por otro lado, este alto nivel de heterocigosidad observado en esta especie podría complicar el ensamblaje del genoma utilizando NGS (Wei *et al.*, 2018), ya que las diferencias a nivel de nucleótidos que puedan existir en la misma región de dos cromosomas homólogos podrían ocasionar complicaciones en el algoritmo del programa de ensamblaje, y así generar *contigs* y *scaffolds* con un tamaño pequeño (Michael y VanBuren, 2020).

Las lecturas generadas en el secuenciamiento de las bibliotecas PE y MP, utilizando la plataforma Illumina, permitió ensamblar un genoma de *Eugenia uniflora* con un tamaño total de 385.1 Mbp. El tamaño obtenido es cercano al obtenido en el análisis de conteo de k-mers (311.6 Mbp) y mucho mayor al valor estimado por Da Costa *et al.* en el 2008 (245 Mbp) utilizando citometría de flujo en muestras de hoja provenientes de una pitanga del estado de Rio de Janeiro. Esta diferencia observada entre el tamaño de un genoma obtenido por citometría de flujo y secuenciamiento ya ha sido observada en



otras especies de plantas (Pellicer y Leitch, 2019) y hongos (Kooij y Pellicer, 2020), y podría ser consecuencia de la misma naturaleza del genoma (presencia de poliploidía, abundancia y composición de DNA repetitivo, y nivel de heterocigosidad). En nuestro caso, esta diferencia de valores podría ser consecuencia del grado de heterocigosidad que estaría relacionada con la plasticidad adaptativa a diferentes ambientes geográficos observada en pitanga (Salgueiro *et al.*, 2004), ya que el análisis por citometría de flujo se realizó con un arbusto colectado muy cerca de la playa en Rio de Janeiro (Restinga de Maricá) mientras que el individuo secuenciado procede de un árbol de la ciudad de Porto Alegre (estado de Rio Grande del Sur). Esta observación es similar a la realizada por Salgueiro *et al.* (2004), en donde utilizando marcadores AFLPs en diferentes poblaciones de pitanga de Brasil reportaron que los individuos de Rio Grande del Sur tienen mayor heterocigosidad que los individuos procedentes del estado de Rio de Janeiro. Por otro lado, se observa un caso similar en *Psidium guajava* (otra especie de la familia Myrtaceae), donde el tamaño del genoma obtenido en un cultivar de la India fue de aproximadamente 304 Mbp (Thakur *et al.*, 2021), mientras que la estimación del tamaño por citometría de flujo en los cultivares blanco y rojo de Brasil fueron de 247.9 y 269.4 Mbp, respectivamente (Da Costa *et al.*, 2008).

El tamaño del genoma de 385.1 Mbp obtenido en *E. uniflora* es más completo y representativo que el genoma de 3.15 Mbp obtenido para esta misma especie por Stefenon *et al.*, (2019) utilizando la plataforma Ion torrent PGM. Este tamaño de genoma obtenido en pitanga es muy parecido al obtenido en *P. guajava* (303.8 Mbp), siendo ambas especies pertenecientes a la tribu Myrteae (Thakur *et al.*, 2021). Asimismo, las especies *Metrosideros polymorpha* (304 Mbp) de la tribu Metrosidereae (Izuno *et al.* 2016), y *Melaleuca alternifolia* (362 Mbp) de la tribu Melaleuceae (Voelker *et al.*, 2021), tienen genomas secuenciados con un tamaño similar al de pitanga y pertenecen a tribus filogenéticamente cercanas a la tribu Myrteae (Maurin *et al.*, 2021). Por otro lado, ya existen reportes de secuenciamiento del genoma de especies de las tribus Eucalypteae

(*Eucalyptus grandis* con 640 Mbp y *Eucalyptus pauciflora* con 594.9 Mbp), y Leptospermeae (*Leptospermum scoparium* con 297 Mbp) (Myburg *et al.*, 2014; Thrimawithana *et al.*, 2019; Wang *et al.*, 2020; Zhang *et al.*, 2022). Los resultados obtenidos indican que el genoma secuenciado de *E. uniflora* en este estudio es más completo y de un tamaño moderadamente pequeño en comparación a otras especies de la familia Myrtaceae.

A nivel de secuencias repetitivas, el 50.47% (194.4 Mbp) del genoma de *E. uniflora* está compuesto por este tipo de secuencias. Este valor es mayor al reportado en otras especies de la familia Myrtaceae con genoma secuenciado y analizaron este tipo de secuencias, como *Eucalyptus pauciflora* (44.77%), *Eucalyptus grandis* (41.22%), *P. guajava* (37.95%) y *L. scoparium* (32.4%) (Myburg *et al.*, 2014; Thrimawithana *et al.*, 2019; Wang *et al.*, 2020; Thakur *et al.*, 2021). La presencia de secuencias repetitivas en la mitad del genoma de pitanga podría ser una de las principales causas del alto nivel de heterocigosidad mencionado previamente, y ya ha sido reportado en otras especies de Viridiplantae (Kyriakidou *et al.*, 2018). Asimismo, la obtención de un genoma fragmentado en 34 716 *scaffolds* es consecuencia de la presencia de estas regiones repetitivas, las cuales hacen imposible obtener genomas completos utilizando secuenciamiento Sanger e Illumina (Li *et al.*, 2018).

La integridad del genoma ensamblado de pitanga evaluado por el análisis de BUSCO dio como resultado un valor del 94.6%. Este valor es el segundo más alto obtenido a nivel de todas las especies de la familia Myrtaceae con genoma secuenciado hasta la fecha, siendo el mayor y menor valor los reportados en *Melaleuca alternifolia* (98.1%) y *P. guajava* (89.9%), respectivamente (Voelker *et al.*, 2021; Thakur *et al.*, 2021). Consecuencia de que el valor de BUSCO es mucho mayor al obtenido en *P. guajava*, el genoma ensamblado de *E. uniflora* estaría más íntegro y representaría mejor el conjunto de genes propios de la tribu Myrteae. Esto se vería también reforzado por la

estadística N50, donde el genoma de *E. uniflora* y *P. guajava* tienen los valores de 26 199 y 17 992 bp, respectivamente. Esta relación directamente proporcional entre los valores de BUSCO y N50 ya ha sido observado en diferentes especies de plantas, hongos y animales con genomas secuenciados (Jauhal y Newcomb, 2021).

La predicción génica realizada en el genoma enmascarado de pitanga permitió identificar un total de 30 663 genes codificadores de proteínas. Para estimar la sensibilidad del método de predicción de genes utilizado, se realizó un análisis BUSCO en las proteínas predichas y se obtuvo un valor de 89.7%, el cual fue menor al obtenido utilizando el genoma ensamblado (94.6%). Esta diferencia también fue reportada en otras especies de la familia Myrtaceae como *L. scoparium* (Thrimawithana *et al.*, 2019) y *E. grandis* (Myburg *et al.*, 2014). En nuestro caso, esta diferencia de valores puede ser consecuencia de utilizar proteínas de otras especies de Viridiplantae en la generación de los indicios y no bibliotecas de secuenciamiento de RNA-seq de diferentes tejidos de pitanga, lo que generaría la pérdida de predicción de genes o isoformas específicas de esta especie. Otra posible causa de esta diferencia es el enmascaramiento inexacto de las repeticiones, lo cual generaría que no se realice la predicción génica en determinadas regiones del genoma (Bruna *et al.*, 2021).

El análisis comparativo génico con otras especies de Viridiplantae permitió identificar 2 219 genes (agrupado en 1 030 clústeres) que son específicos de *E. uniflora*. La anotación funcional de los genes predichos es un análisis que permite asociar cada secuencia con una red de moléculas que interactúan en la célula para así proporcionar información acerca de los procesos biológicos, componentes celulares y funciones moleculares (Kanehisa *et al.*, 2014). Del total de genes específicos de pitanga, el 40% fueron anotados funcionalmente con 1 772 términos ontológicos, lo que sugiere que estos genes están involucrados en una amplia diversidad de funciones biológicas de pitanga que necesitan ser explorados en futuros trabajos. Interesantemente, dentro de

este grupo de genes específicos de pitanga, fueron identificados genes que podrían estar involucrados en la adaptación de esta especie a diferentes ambientes (Bezerra *et al.*, 2000). Un ejemplo de este tipo de genes identificados en pitanga son las que pertenecen a la familia de proteínas de choque térmico (hsps), cuyos niveles de expresión son inducidos por estreses que ocurren naturalmente o en condiciones anormales, y que pueden servir como indicadores de estrés y respuestas al estrés en diversos organismos y circunstancias (Chen *et al.*, 2018). Por otro lado, existen 707 genes específicos de pitanga que no tuvieron similaridad con ningún gen de Viridiplantae. El estudio más detallado de estos genes podría generar un nuevo conocimiento sobre los mecanismos moleculares de diversos procesos biológicos de pitanga.

Utilizando la anotación del genoma ensamblado de *E. uniflora*, fueron identificadas 20 TPS completas que presentan los dos dominios Pfam característicos de esta familia de genes. El bajo número de TPS identificadas en pitanga en comparación con *E. grandis* podría indicar que la predicción génica realizada fue inexacta en este grupo de genes debido a los *scaffolds* fragmentados del genoma ensamblado. Por otro lado, el bajo número de TPS identificadas también podría estar relacionado al hecho de que pitanga posiblemente no tenga una gran diversidad de miembros para este grupo de genes. Las especies del género *Eucalyptus*, las cuales crecen en bosques abiertos y abundantes, necesitan defender sus hojas con una mayor intensidad por lo que necesitan producir una mayor diversidad de terpenos volátiles a través de las TPS (Külheim *et al.*, 2015; Grubb *et al.*, 1998).

La subfamilia TPS-a, la cual codifican únicamente TPS que producen sesquiterpenos en plantas dicotiledóneas y monocotiledóneas (Martin *et al.*, 2010), no fueron identificadas en pitanga. En un estudio realizado por Guzman *et al.*, (2014) fueron identificadas 3 TPS que sintetizan sesquiterpenos, por lo que es más probable que los

miembros de la subfamilia TPS-a existen en pitanga pero no fueron predichos en la versión actual del genoma ensamblado. Por otro lado, la subfamilia monofilética TPS-b es dividida en dos grupos, un clado conteniendo monoterpenos sintasas cíclicas putativas descritas como TPS-b1 que generalmente están localizados en los plastidios (Külheim *et al.*, 2015) y otro clado que contiene sintasas que sintetizan isopreno/ocimeno denominado TPS-b2 (Calvert *et al.*, 2018). El grupo TPS-b1 tiene la mayor cantidad de TPS identificadas en el genoma de pitanga (60%), siendo un indicativo de una rápida evolución y expansión de este grupo de sintasas en respuesta las rápidas fluctuaciones de temperaturas en un hábitat subtropical cálido (Singsaas y Sharkey, 2000; Sharkey y Yeh, 2001), o de que algunos monoterpenos cíclicos puedan actuar en la defensa antimicrobiana (Calvert *et al.*, 2018). En el caso de la subfamilia TPS-g, la cual está dentro del clado de la subfamilia TPS-b, se identificaron 3 genes en pitanga y están involucrados en la síntesis de monoterpenos acíclicos en otras plantas (Chen *et al.*, 2011). Por otro lado, las subfamilias TPS-c y TPS-f, que conforman un clado conjuntamente con la subfamilia TPS-e, presentan 1 y 2 identificados en pitanga, respectivamente. Estas dos subfamilias están involucradas en la formación de monoterpenos, diterpenos y sesquiterpenos en otras plantas (Külheim *et al.*, 2015).

Una posterior caracterización bioquímica de las TPS identificadas en el genoma de *E. uniflora* podrá determinar y confirmar de manera específica que tipo de terpenos son sintetizados en determinadas condiciones ambientales. Asimismo, será necesaria la caracterización de la expresión génica de estas TPS en hojas y flores utilizando secuenciamiento de bibliotecas de RNA-seq o PCR en tiempo real. Finalmente, debido al alto grado de heterocigosidad observado en el genoma, sería importante identificar si existen SNPs en las secuencias de los genes de las TPS identificadas en individuos de diferentes poblaciones y verificar si existe una relación de los mismos con la adaptabilidad de pitanga a diferentes tipos de ambientes.

## 8. CONCLUSIONES

- El genoma ensamblado de *E. uniflora* tiene un tamaño de 385.1 Mbp y 50.47% de su contenido está constituido por secuencias repetitivas.
- Los valores de integridad de genoma y N50 confirman que el genoma obtenido en este estudio es de mejor calidad y tamaño en comparación al genoma reportado en el 2019. Por este motivo, este nuevo genoma de *E. uniflora* podría ser utilizado como referencia en futuros estudios comparativos con otras especies de la tribu Myrteae.
- La predicción génica realizada en el genoma permitió identificar 30 663 genes codificadores de proteínas de pitanga, de las cuales 2 219 genes son específicos de *E. uniflora*.
- Se identificaron 20 TPS en el genoma de *E. uniflora* y que representan a las subfamilias TPS-b, TPS-c, TPS-f y TPS-g.

## 9. RECOMENDACIONES

- El ensamblaje del genoma de pitanga se puede mejorar utilizando el secuenciamiento SMRT (Pacific Biosciences, USA) conjuntamente con los datos de secuenciamiento Illumina obtenidos en este trabajo de investigación.
- El genoma secuenciado y los genes predichos en este trabajo servirán de base para conocer y comprender los procesos biológicos que ocurren en la especie *E. uniflora*, especialmente los relacionados a la adaptabilidad ambientes y síntesis de terpenos.

## 10. REFERENCIAS BIBLIOGRÁFICAS

1. ABASCAL, F.; ZARDOYA, R.; POSADA, D. ProtTest: selection of best-fit models of proteinevolution. *Bioinformatics*, 2005, vol. 21, no 9, p. 2104-2105.
2. ABRUSÁN, G., et al. TEclass—a tool for automated classification of unknown eukaryotic transposable elements. *Bioinformatics*, 2009, vol. 25, no 10, p. 1329-1330.
3. ADEBAJO, A. C.; OLOKE, K. J.; ALADESANMI, A. J. Antimicrobial activities and microbial transformation of volatile oils of *Eugenia uniflora*. *Fitoterapia*, 1989, vol. 60, no 5, p. 451-455.
4. AMORIM, A.C.L., et al. Antinociceptive and hypothermic evaluation of the leaf essential oil and isolated terpenoids from *Eugenia uniflora* L. (Brazilian Pitanga). *Phytomedicine*, 2009, vol. 16, no 10, p. 923-928.
5. AUBOURG, S.; LECHARNY, A.; BOHLMANN, J. Genomic analysis of the terpenoid synthase (AtTPS) gene family of *Arabidopsis thaliana*. *Molecular Genetics and Genomics*, 2002, vol. 267, no 6, p. 730-745.
6. BEZERRA, J.E.F.; SILVA J.R., J.F.D.; LEDERMAN, I. E. Pitanga (*Eugenia uniflora* L.), Jaboticabal: FUNEP, 2000. 30p. Série Frutas Nativas, vol. 1.
7. BICAS, J.L., et al. Volatile constituents of exotic fruits from Brazil. *Food Research International*, 2011, vol. 44, no 7, p. 1843-1855.
8. BIFFIN, E., et al. Molecular systematics of *Syzygium* and allied genera (Myrtaceae): evidence from the chloroplast genome. *Taxon*, 2006, vol. 55, no 1, p. 79-94.



9. BOLGER, M. E., et al. Plant genome sequencing—applications for crop improvement. *Current opinion in biotechnology*, 2014, vol. 26, p. 31-37.
10. BRAGA, R. Plantas do Nordeste, especialmente do Ceará. *Colecao Mossoroense*, V., 1976, vol. 315, p. 540.
11. BROOKER, M.I.H. A new classification of the genus *Eucalyptus* L'Her.(Myrtaceae). *Australian Systematic Botany*, 2000, vol. 13, no 1, p. 79-148.
12. BRUNA, T.; LOMSADZE, A.; BORODOVSKY, M. GeneMark-EP and-EP+: automatic eukaryotic gene prediction supported by spliced aligned proteins. *bioRxiv*, 2020, p. 2019.12. 31.891218.
13. BRŮNA, T., et al. BRAKER2: Automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR genomics and bioinformatics*, 2021, vol. 3, no 1, p. lqaa108.
14. CABALLERO, M.; WEGRZYN, J. gFACs: Gene filtering, analysis, and conversion to unify genome annotations across alignment and gene prediction frameworks. *Genomics, proteomics& bioinformatics*, 2019, vol. 17, no 3, p. 305-310.
15. CANE, D. E. Sesquiterpene biosynthesis. *comprehensive, natural products chemistry*, 1999, vol. 2, p. 155-200.
16. CALVERT, Jed, et al. Terpene synthase genes in *Melaleuca alternifolia*: comparative analysis of lineage-specific subfamily variation within Myrtaceae. *Plant Systematics and Evolution*, 2018, vol. 304, no 1, p. 111-121.

17. CHEN, Feng, et al. The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. *The Plant Journal*, 2011, vol. 66, no 1, p. 212-229.
18. CHEN, Bing; FEDER, Martin E.; KANG, Le. Evolution of heat-shock protein expression underlying adaptive responses to environmental stress. *Molecular ecology*, 2018, vol. 27, no 15, p. 3040-3054.
19. CLAUSING, G.; RENNER, S.S. Molecular phylogenetics of Melastomataceae and Memecylaceae: implications for character evolution. *American Journal of Botany*, 2001, vol. 88, no 3, p. 486-498.
20. CONESA, A., et al. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, 2005, vol. 21, no 18, p. 3674-3676.
21. CONSOLINI, A.E.; BALDINI, O.A.N.; AMAT, A.G. Pharmacological basis for the empirical use of *Eugenia uniflora* L.(Myrtaceae) as antihypertensive. *Journal of ethnopharmacology*, 1999, vol. 66, no 1, p. 33-39.
22. CONSOLINI, A. E.; SARUBBIO, M. G. Pharmacological effects of *Eugenia uniflora* (Myrtaceae) aqueous crude extract on rat's heart. *Journal of ethnopharmacology*, 2002, vol. 81, no 1, p. 57-63.
23. COSTA, D. P., et al. Seasonal variability of essential oils of *Eugenia uniflora* leaves. *Journal of the Brazilian Chemical Society*, 2009, vol. 20, p. 1287-1293.
24. CRAVEN, L. A.; BIFFIN, E. An infrageneric classification of *Syzygium* (Myrtaceae).

Blumea-Biodiversity, Evolution and Biogeography of Plants, 2010, vol. 55, no 1, p. 94-99.

25. D'HONT, A., et al. The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature*, 2012, vol. 488, no 7410, p. 213-217.
26. DA COSTA, I. R.; DORNELAS, M. C.; FORNI-MARTINS, E. R. Nuclear genome size variation in fleshy-fruited Neotropical Myrtaceae. *Plant systematics and evolution*, 2008, vol. 276, no 3, p. 209-217.
27. DA CRUZ, F., et al. Phylogenetic analysis of the genus *Hexachlamys* (Myrtaceae) based on plastid and nuclear DNA sequences and their taxonomic implications. *Botanical Journal of the Linnean Society*, 2013, vol. 172, no 4, p. 532-543.
28. DARRIBA, D., et al. jModelTest 2: more models, new heuristics and parallel computing. *Nature methods*, 2012, vol. 9, no 8, p. 772-772.
29. DAVIS, E. M.; CROTEAU, R. Cyclization enzymes in the biosynthesis of monoterpenes, sesquiterpenes, and diterpenes. *Biosynthesis*, 2000, p. 53-95.
30. DEGENHARDT, J.; KÖLLNER, T. G.; GERSHENZON, J. Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. *Phytochemistry*, 2009, vol. 70, no 15-16, p. 1621-1637.
31. DOYLE, J. J. and DOYLE, J. L. Isolation of plant DNA from fresh tissue. *Focus*, 1990, vol. 12, p. 13-15.
32. DRUMMOND, A. J.; RAMBAUT, A. BEAST: Bayesian evolutionary analysis by

- samplingtrees. BMC evolutionary biology, 2007, vol. 7, no 1, p. 1-8.
33. EGUILUZ, M., et al. The chloroplast genome sequence from *Eugenia uniflora*, a Myrtaceae from Neotropics. Plant Systematics and Evolution, 2017, vol. 303, no 9, p. 1199-1212.
  34. FERREIRA-RAMOS, R., et al. Genetic analysis of forest species *Eugenia uniflora* L. throughof newly developed SSR markers. Conservation Genetics, 2008, vol. 9, no 5, p. 1281-1285.
  35. FEUILLET, C., et al. Crop genome sequencing: lessons and rationales. Trends in plant science, 2011, vol. 16, no 2, p. 77-88.
  36. GOVAERTS, R., et al. World checklist of Myrtaceae. Royal Botanic Gardens, 2008.
  37. GRATTAPAGLIA, D., et al. Progress in Myrtaceae genetics and genomics: *Eucalyptus* as the pivotal genus. Tree Genetics & Genomes, 2012, vol. 8, no 3, p. 463-508.
  38. GRUBB, Peter J., et al. Nitrogen-richness and protection of seeds in Australian tropical rainforest: a test of plant defence theory. Oikos, 1998, p. 467-482.
  39. GUO, S., et al. The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. Nature genetics, 2013, vol. 45, no 1, p. 51-58.
  40. GUINDON, Stéphane; GASCUEL, Olivier. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic biology, 2003, vol. 52, no 5, p. 696-704.

41. GUREVICH, A., et al. QUASt: quality assessment tool for genome assemblies. *Bioinformatics*, 2013, vol. 29, no 8, p. 1072-1075.
42. GUZMAN, F., et al. Identification of microRNAs from *Eugenia uniflora* by high-throughput sequencing and bioinformatics analysis. *PLoS One*, 2012, vol. 7, no 11, p. e49811.
43. GUZMAN, F., et al. De novo assembly of *Eugenia uniflora* L. transcriptome and identification of genes from the terpenoid biosynthesis pathway. *Plant Science*, 2014, vol. 229, p. 238-246.
44. HARBORNE, A. J. *Phytochemical methods a guide to modern techniques of plant analysis*. Springer science & business media, 1998.
45. HENRIQUES, A. T., et al. Aromatic plants from Brazil. II. The chemical composition of some *Eugenia* essential oils. *Journal of Essential Oil Research*, 1993, vol. 5, no 5, p. 501-505.
46. IZUNO, Ayako, et al. Genome sequencing of *Metrosideros polymorpha* (Myrtaceae), a dominant species in various habitats in the Hawaiian Islands with remarkable phenotypic variations. *Journal of plant research*, 2016, vol. 129, no 4, p. 727-736.
47. JAUHAL, April A.; NEWCOMB, Richard D. Assessing genome assembly quality prior to downstream analysis: N50 versus BUSCO. *Molecular Ecology Resources*, 2021, vol. 21, no 5, p. 1416-1421.
48. KAJITANI, Rei, et al. Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome research*, 2014, vol. 24, no 8, p. 1384-1395.

49. KANEHISA, Minoru, et al. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic acids research*, 2014, vol. 42, no D1, p. D199-D205.
50. KYRIAKIDOU, Maria, et al. Current strategies of polyploid plant genome sequence assembly. *Frontiers in Plant Science*, 2018, vol. 9, p. 1660.
51. KITSON, J. J. N., et al. Molecular characterization of trophic ecology within an island radiation of insect herbivores (Curculionidae: Entiminae: *Cratopus*). *Molecular ecology*, 2013, vol. 22, no 21, p. 5441-5455.
52. KOOIJ, Pepijn W.; PELLICER, Jaume. Genome size versus genome assemblies: are the genomes truly expanded in polyploid fungal symbionts?. *Genome biology and evolution*, 2020, vol. 12, no 12, p. 2384-2390.
53. KÜLHEIM, Carsten, et al. The *Eucalyptus* terpene synthase gene family. *BMC genomics*, 2015, vol. 16, no 1, p. 1-18.
54. MACMILLAN, J.; BEALE, M. H. Diterpene biosynthesis. In "Comprehensive Natural Products Chemistry No. 2, Isoprenoids Including Carotenoids and Steroids," eds. Barton, SD, and Nakanishi, K. 1999.
55. MARÇAIS, Guillaume; KINGSFORD, Carl. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*, 2011, vol. 27, no 6, p. 764-770.
56. MARTIN, D. M., et al. Functional annotation, genome organization and phylogeny of the grapevine (*Vitis vinifera*) terpene synthase gene family based on genome assembly, FLcDNAcloning, and enzyme assays. *BMC plant biology*, 2010, vol. 10, no 1, p. 1-22.

57. MAURIN, Olivier, et al. A nuclear phylogenomic study of the angiosperm order Myrtales, exploring the potential and limitations of the universal Angiosperms353 probe set. *American Journal of Botany*, 2021, vol. 108, no 7, p. 1087-1111.
58. MICHAEL, Todd P.; VANBUREN, Robert. Building near-complete plant genomes. *Current Opinion in Plant Biology*, 2020, vol. 54, p. 26-33.
59. MILLER, B.; OSCHINSKI, C.; ZIMMER, W. First isolation of an isoprene synthase gene from poplar and successful expression of the gene in *Escherichia coli*. *Planta*, 2001, vol. 213, no 3, p. 483-487.
60. MYBURG, A. A., et al. The genome of *Eucalyptus grandis*. *Nature*, 2014, vol. 510, no 7505, p. 356-362.
61. LAGO, J. H. G., et al. Chemical and biological evaluation of essential oils from two species of Myrtaceae—*Eugenia uniflora* L. and *Plinia trunciflora* (O. Berg) Kausel. *Molecules*, 2011, vol. 16, no 12, p. 9827-9837.
62. LEGGETT, R. M., et al. NextClip: an analysis and read preparation tool for Nextera Long MatePair libraries. *Bioinformatics*, 2014, vol. 30, no 4, p. 566-568.
63. LI, Changsheng, et al. Genome sequencing and assembly by long reads in plants. *Genes*, 2018, vol. 9, no 1, p. 6.
64. LIM, T. K. Edible medicinal and non-medicinal plants. Dordrecht, The Netherlands: Springer, 2012.

65. LIU, J., et al. Genome-wide analysis of terpene synthases in soybean: functional characterization of GmTPS3. *Gene*, 2014, vol. 544, no 1, p. 83-92.
66. LOMSADZE, A., et al. Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic acids research*, 2005, vol. 33, no 20, p. 6494-6506.
67. LUCAS, E. J., et al. Suprageneric phylogenetics of Myrteae, the generically richest tribe in Myrtaceae (Myrtales). *Taxon*, 2007, vol. 56, no 4, p. 1105-1128.
68. LUCAS, E. J.; BÜNGER, M. O. Myrtaceae in the Atlantic forest: their role as a 'model' group. *Biodiversity and Conservation*, 2015, vol. 24, no 9, p. 2165-2180.
69. OGURA, K.; KOYAMA, T.; SAGAMI, H. Polyprenyl diphosphate synthases. *Cholesterol*, 1997, p. 57-87.
70. PGS CONSORTIUM, et al. Genome sequence and analysis of the tuber crop potato. *Nature*, 2011, vol. 475, no 7355, p. 189-195.
71. PELLICER, Jaume; LEITCH, Ilia J. The Plant DNA C-values database (release 7.1): an updated online repository of plant genome size data for comparative studies. [*" New Phytologist"*], 2019.
72. QUAIL, M. A., et al. A tale of three next generation sequencing platforms: comparison of IonTorrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC genomics*, 2012, vol. 13, no 1, p. 1-13.
73. RUTSCHMANN, F., et al. Assessing calibration uncertainty in molecular dating: the



- assignment of fossils to alternative calibration points. *Systematic Biology*, 2007, vol. 56, no 4, p. 591-608.
74. SALGUEIRO, F., et al. Even population differentiation for maternal and biparental gene markers in *Eugenia uniflora*, a widely distributed species from the Brazilian coastal Atlantic rain forest. *Diversity and Distributions*, 2004, vol. 10, no 3, p. 201-210.
75. SCHMIEDER, R.; EDWARDS, R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, 2011, vol. 27, no 6, p. 863-864.
76. SCHWAB, W. Metabolome diversity: too few genes, too many metabolites?. *Phytochemistry*, 2003, vol. 62, no 6, p. 837-849.
77. SEPPEY, M.; MANNI, M.; ZDOBNOV, E. M. BUSCO: assessing genome assembly and annotation completeness. En *Gene prediction*. Humana, New York, NY, 2019. p. 227-245.
78. SHARKEY, Thomas D.; YEH, Sansun. Isoprene emission from plants. *Annual review of plant biology*, 2001, vol. 52, no 1, p. 407-436.
79. SINGSAAS, E. L.; SHARKEY, T. D. The effects of high temperature on isoprene synthesis in oak leaves. *Plant, Cell & Environment*, 2000, vol. 23, no 7, p. 751-757.
80. SILVA, André Luiz Gomes da; PINHEIRO, Maria Célia Bezerra. Reproductive success of four species of *Eugenia* L. (Myrtaceae). *Acta Botanica Brasilica*, 2009, vol. 23, p. 526-534.
81. SOH, W. K.; PARNELL, J. Comparative leaf anatomy and phylogeny of *Syzygium*

Gaertn. Plant Systematics and Evolution, 2011, vol. 297, no 1, p. 1-32.

82. SOUZA, V. C.; LORENZI, H. Botânica sistemática: guia ilustrado para identificação das famílias de Angiospermas da flora brasileira, baseado em APG II. Instituto Plantarum, 2005.
83. STANKE, M., et al. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC bioinformatics, 2006, vol. 7, no 1, p. 1-11.
84. STEFENON, V. M.; SARZI, D. S.; ROESCH, L. F. W. High-throughput sequencing analysis of *Eugenia uniflora*: insights into repetitive DNA, gene content and potential biotechnological applications. 3 Biotech, 2019, vol. 9, no 5, p. 1-7.
85. TAMURA, K., et al. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular biology and evolution, 2011, vol. 28, no 10, p. 2731-2739.
86. THAKUR, Sujata, et al. Development of Genome-Wide Functional Markers Using Draft Genome Assembly of Guava (*Psidium guajava* L.) cv. Allahabad Safeda to Expedite Molecular Breeding. Frontiers in plant science, 2021, vol. 12.
87. THRIMAWITHANA, Amali H., et al. A whole genome assembly of *Leptospermum scoparium* (Myrtaceae) for mānuka research. New Zealand Journal of Crop and Horticultural Science, 2019, vol. 47, no 4, p. 233-260.
88. VAN DER MERWE, M. M.; VAN WYK, A. E.; BOTHA, A. M. Molecular phylogenetic analysis of *Eugenia* L. (Myrtaceae), with emphasis on southern African taxa. Plant

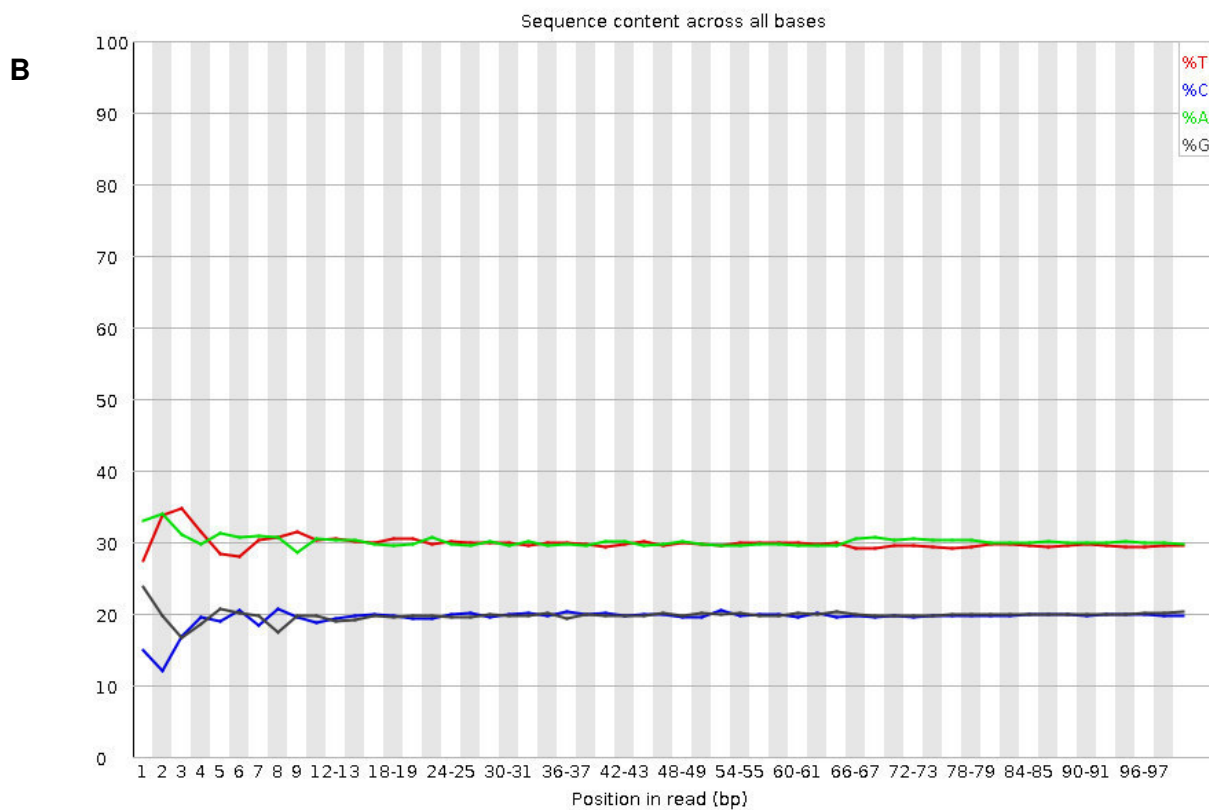
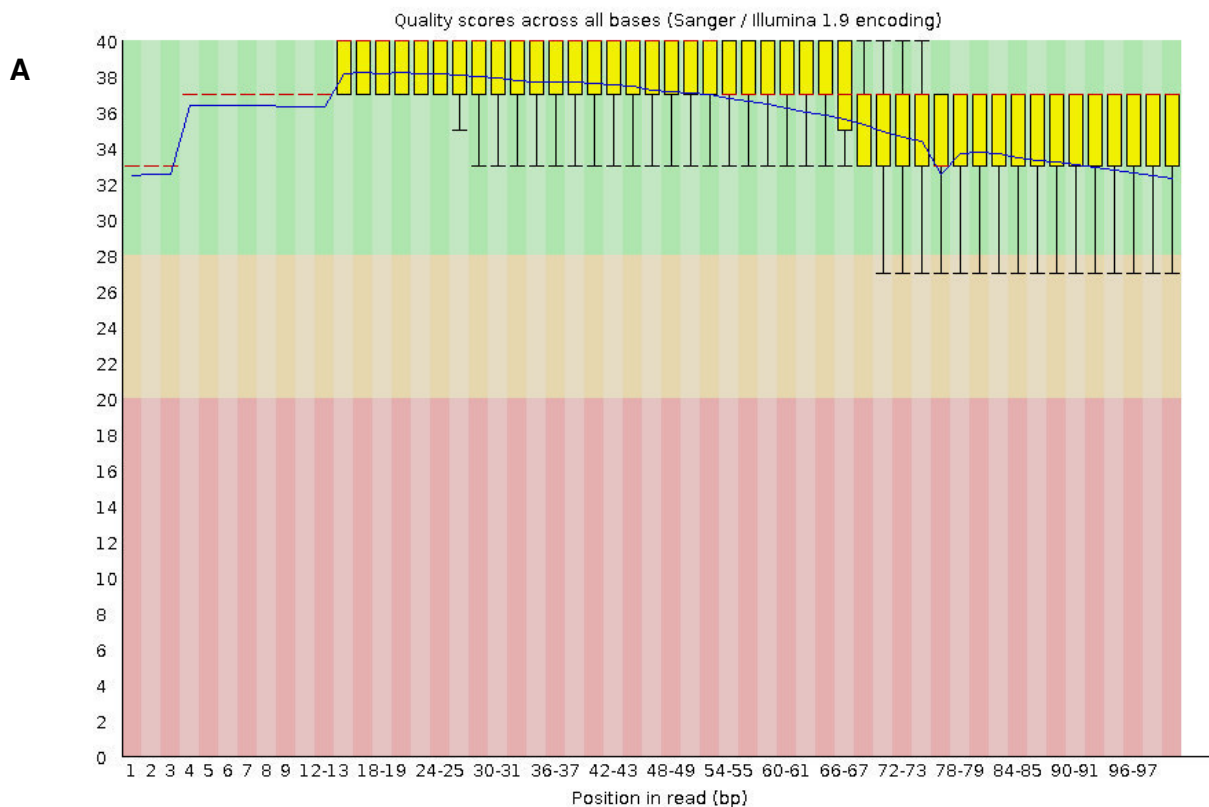
Systematics and Evolution, 2005, vol. 251, no 1, p. 21-34.

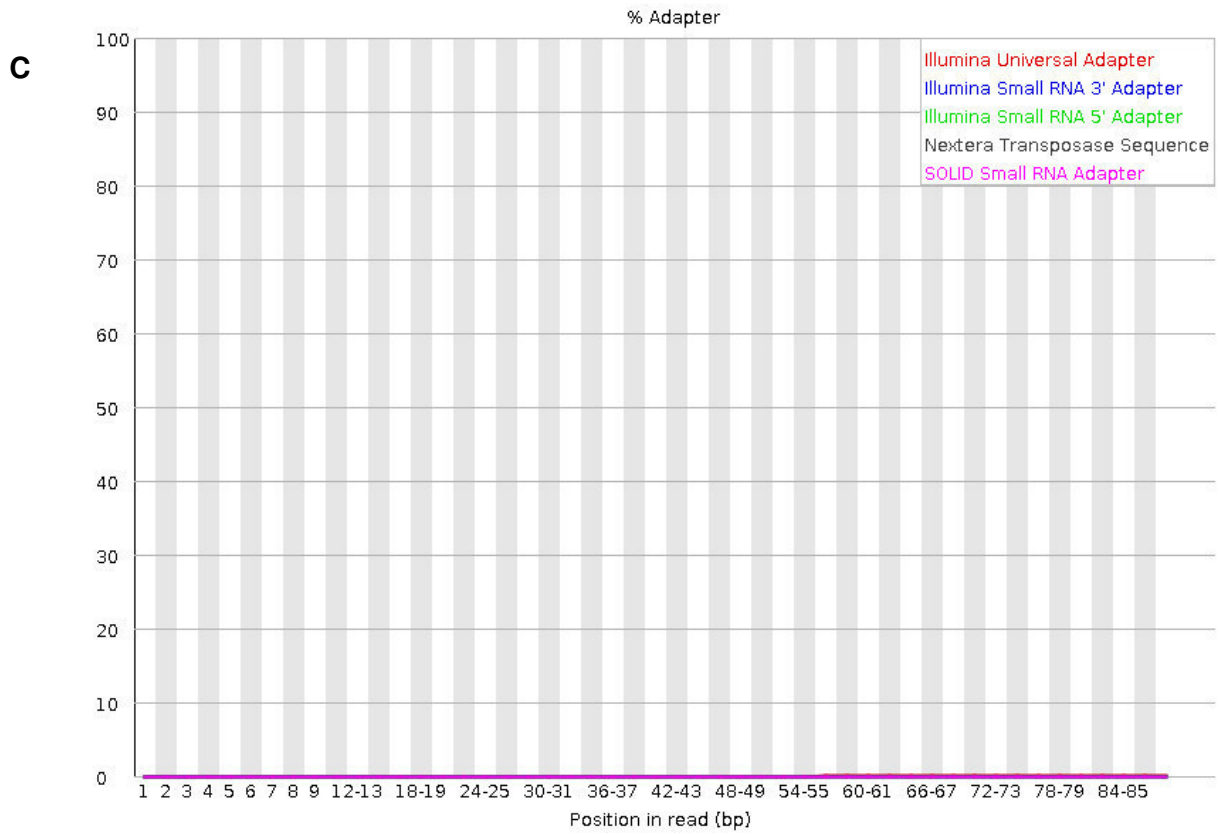
89. VICTORIA, F. N., et al. Essential oil of the leaves of *Eugenia uniflora* L.: antioxidant and antimicrobial properties. Food and chemical toxicology, 2012, vol. 50, no 8, p. 2668-2674.
90. VOELKER, Julia; SHEPHERD, Mervyn; MAULEON, Ramil. A high-quality draft genome for *Melaleuca alternifolia* (tea tree): a new platform for evolutionary genomics of myrtaceous terpene-rich species. Gigabyte, 2021, vol. 2021, p. 1-15.
91. VURTURE, Gregory W., et al. GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics, 2017, vol. 33, no 14, p. 2202-2204.
92. WANG, Weiwen, et al. The draft nuclear genome assembly of *Eucalyptus pauciflora*: a pipeline for comparing de novo assemblies. GigaScience, 2020, vol. 9, no 1, p. giz160.
93. WEI, Chaoling, et al. Draft genome sequence of *Camellia sinensis* var. *sinensis* provides insights into the evolution of the tea genome and tea quality. Proceedings of the National Academy of Sciences, 2018, vol. 115, no 18, p. E4151-E4158.
94. WILSON, P. G., et al. Myrtaceae revisited: a reassessment of infrafamilial groups. American Journal of Botany, 2001, vol. 88, no 11, p. 2013-2025.
95. WISE, M. L.; CROTEAU, R. Biosynthesis of monoterpenes. Comprehensive natural products chemistry, 1999, vol. 2, p. 97-153.
96. XU, Q., et al. The draft genome of sweet orange (*Citrus sinensis*). Nature genetics, 2013, vol. 45, no 1, p. 59-66.

97. XU, L., et al. OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic acids research*, 2019, vol. 47, no W1, p. W52-W58.
98. ZHANG, Xiaoning, et al. Draft Genome of the Medicinal Tea Tree *Melaleuca Alternifolia* (Maiden & Betche) Cheel (Myrtaceae). 2022.

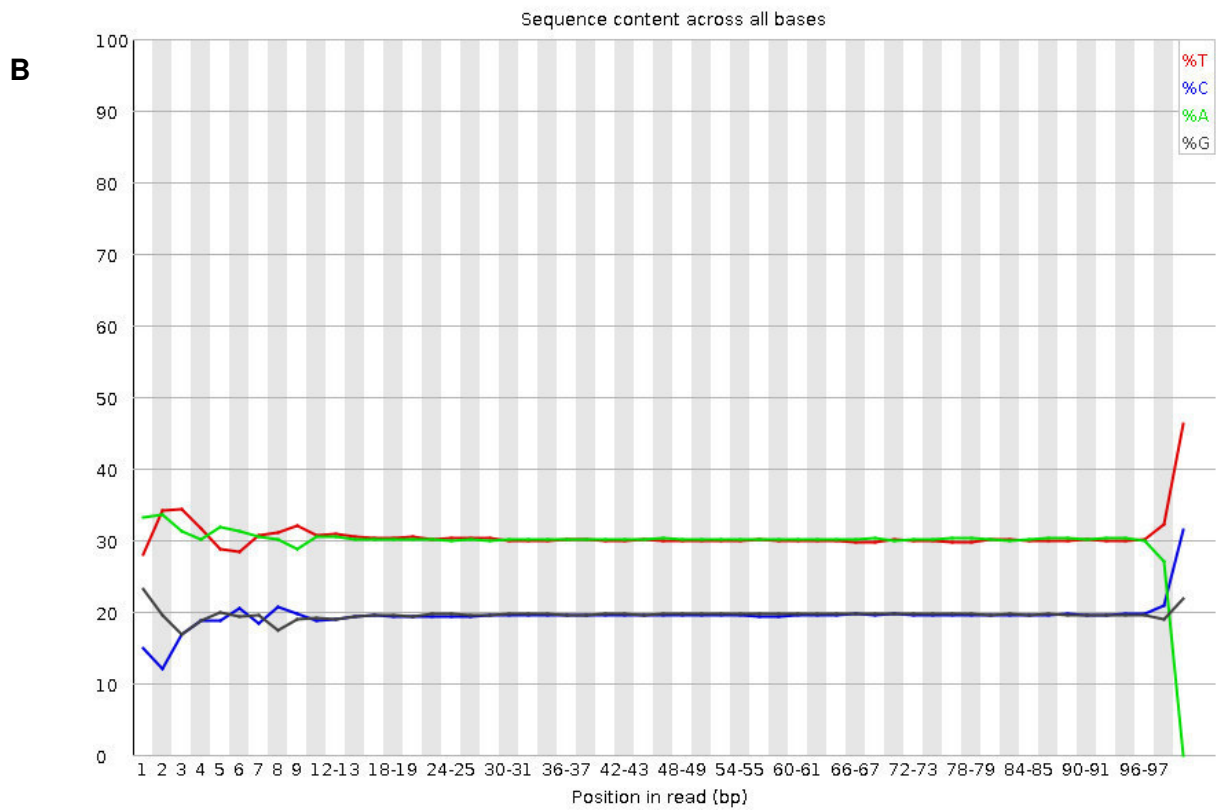
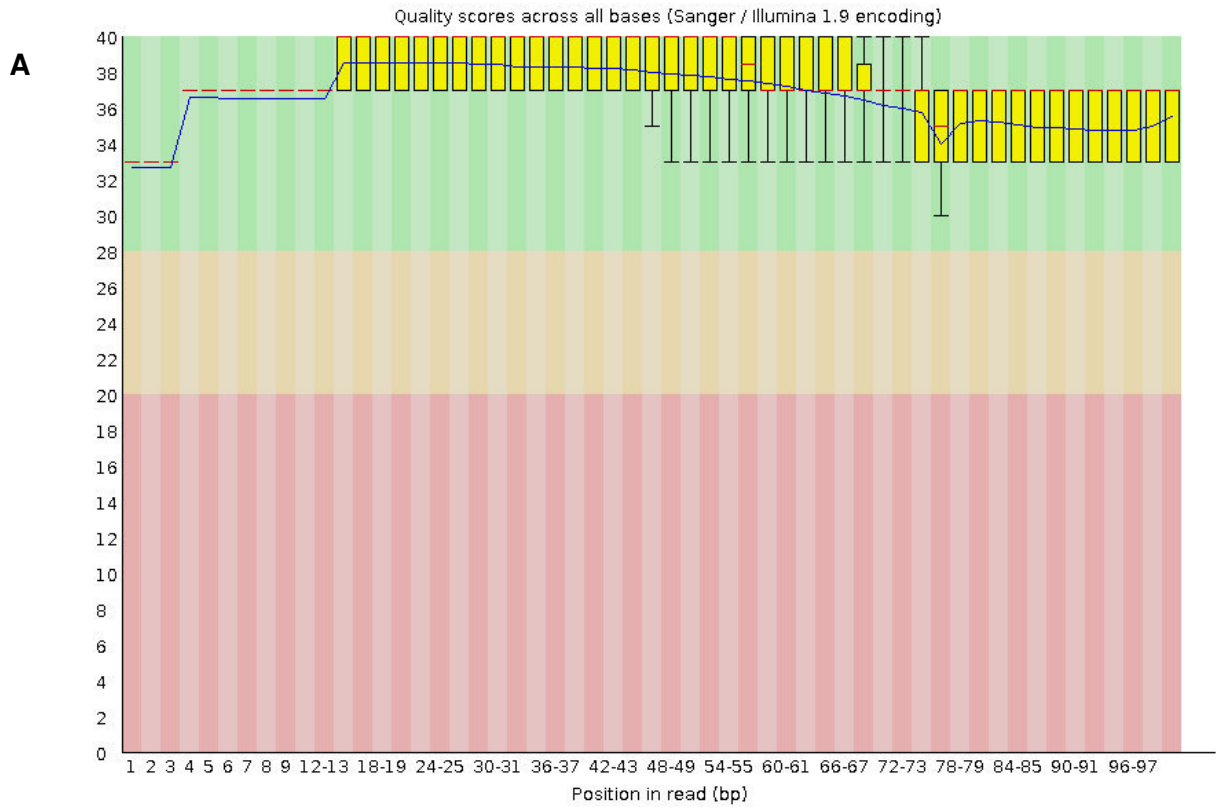
# **ANEXOS**

# ANEXO 1

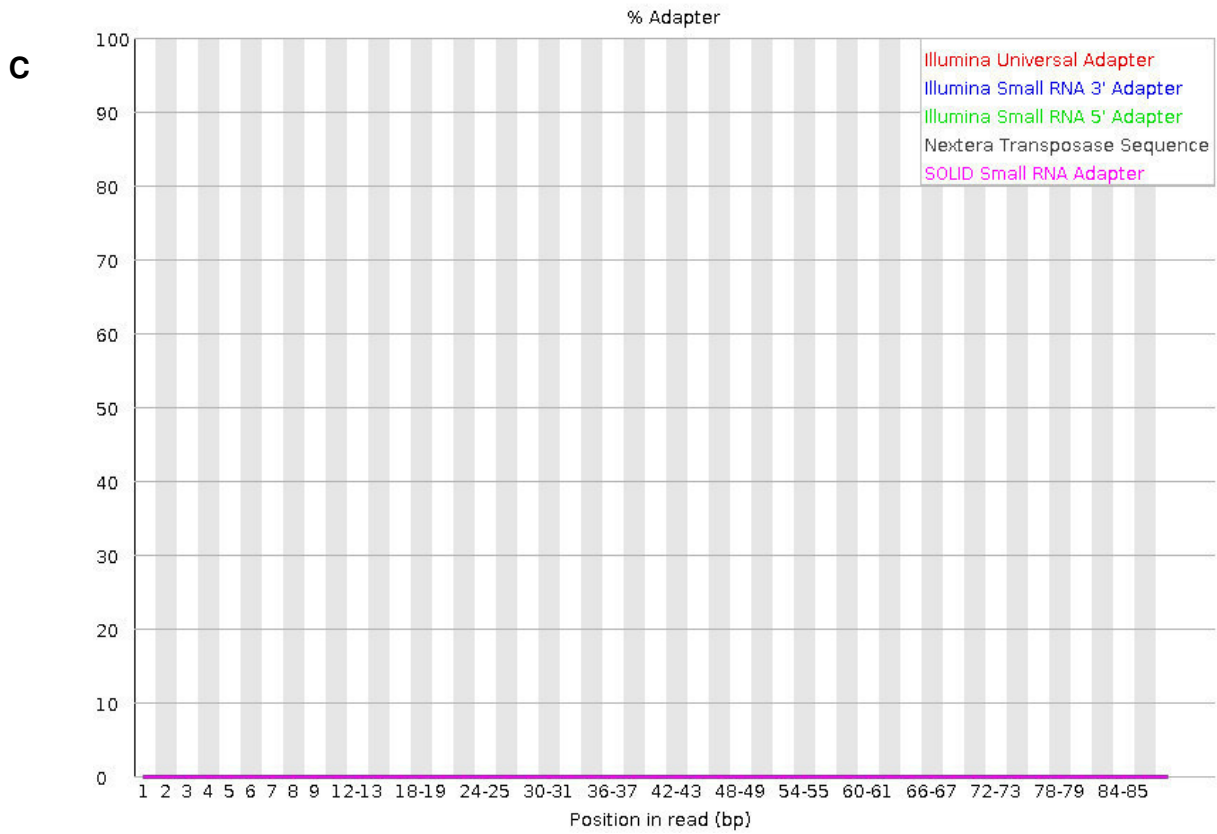




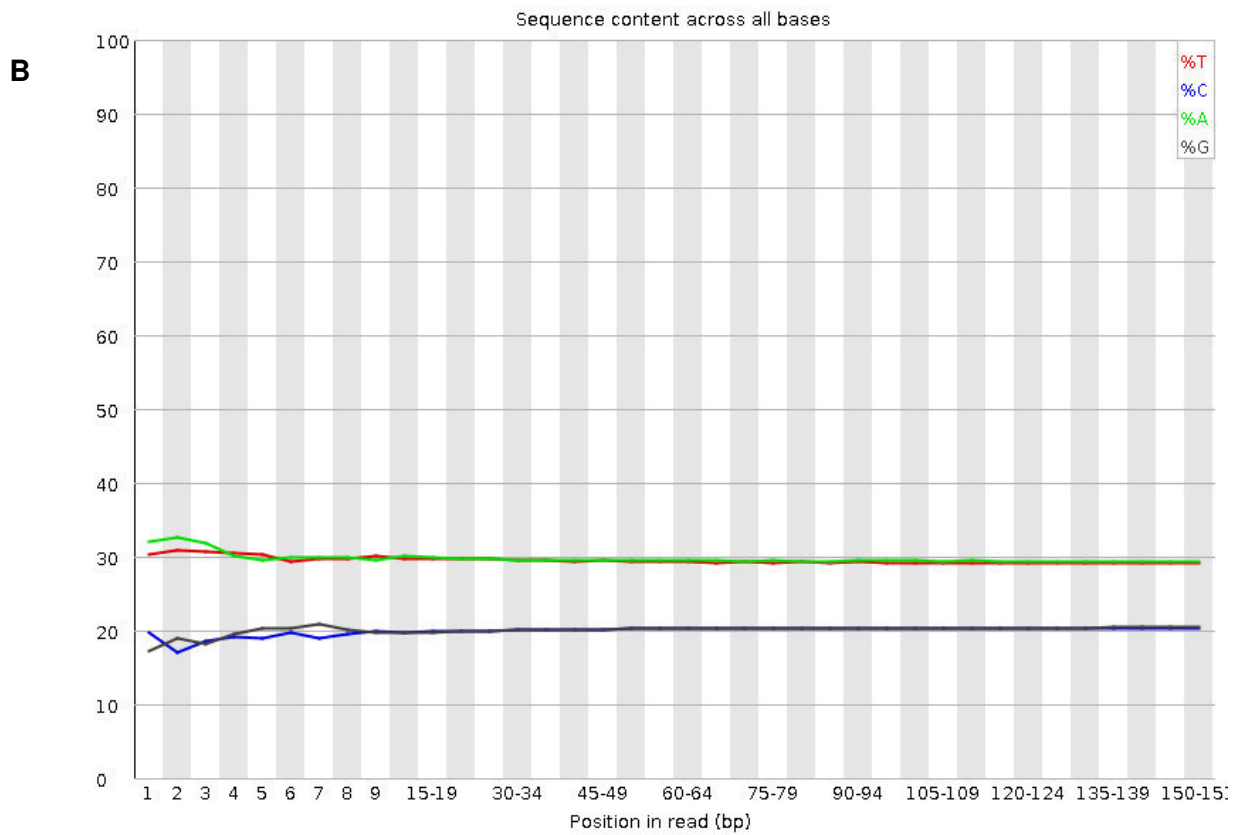
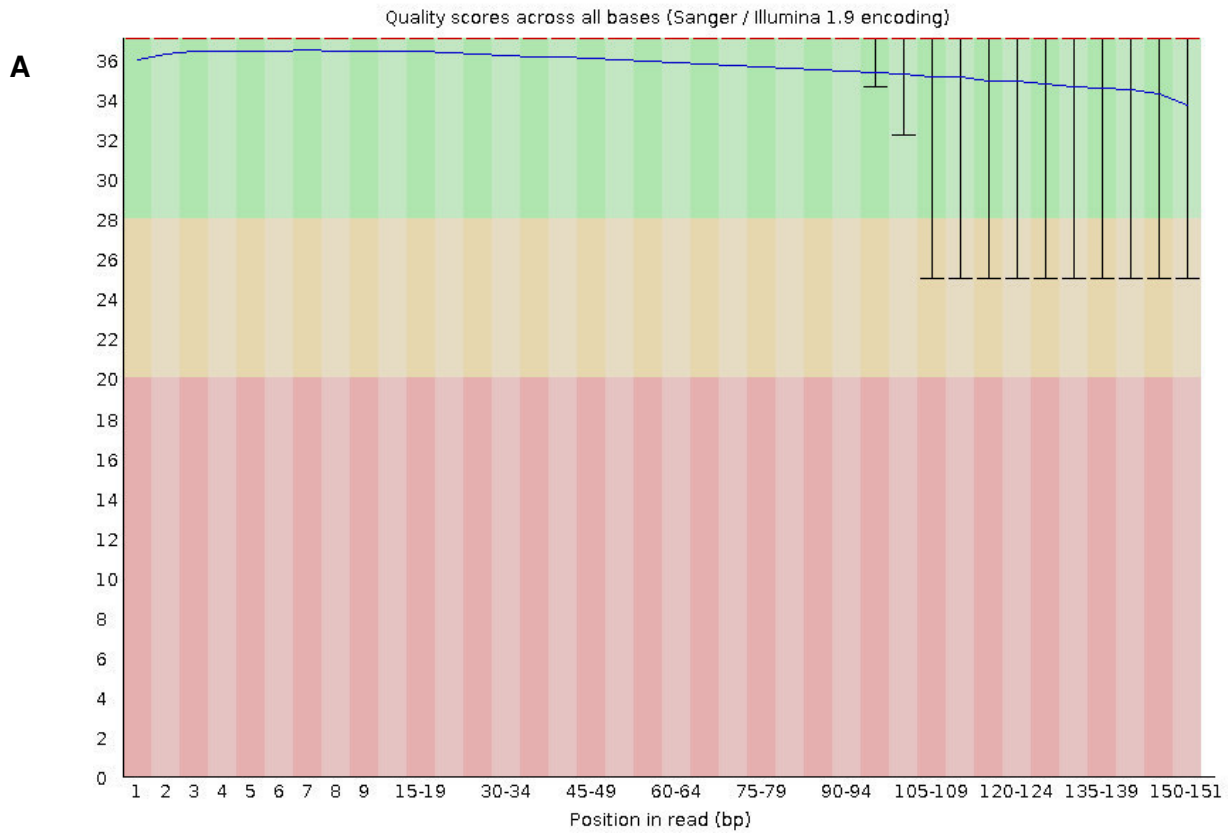
**Figura 9.** Output del programa FastQC donde se visualiza A) la calidad de las lecturas, B) la composición nucleotídica y C) la presencia de adaptadores del archivo R1 de la biblioteca PE de 250 bp antes de la limpieza con el programa Trim galore!.

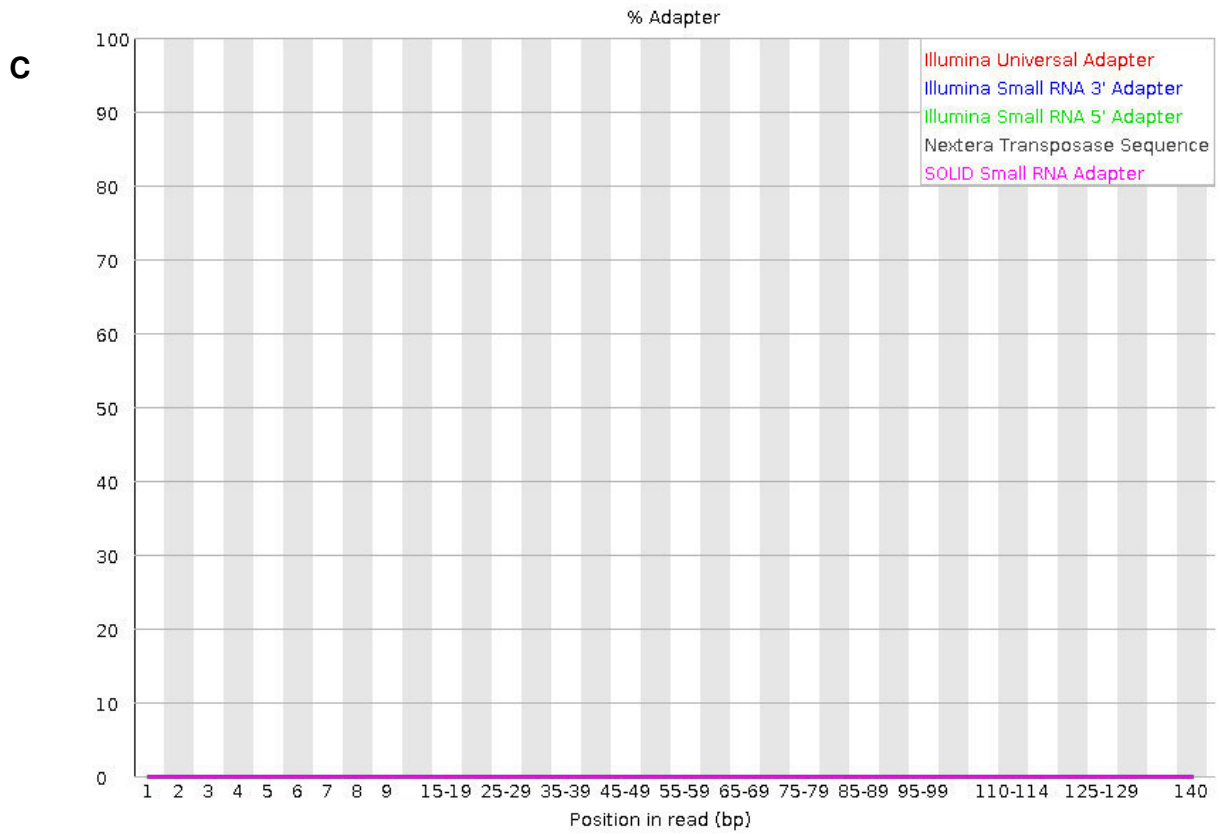




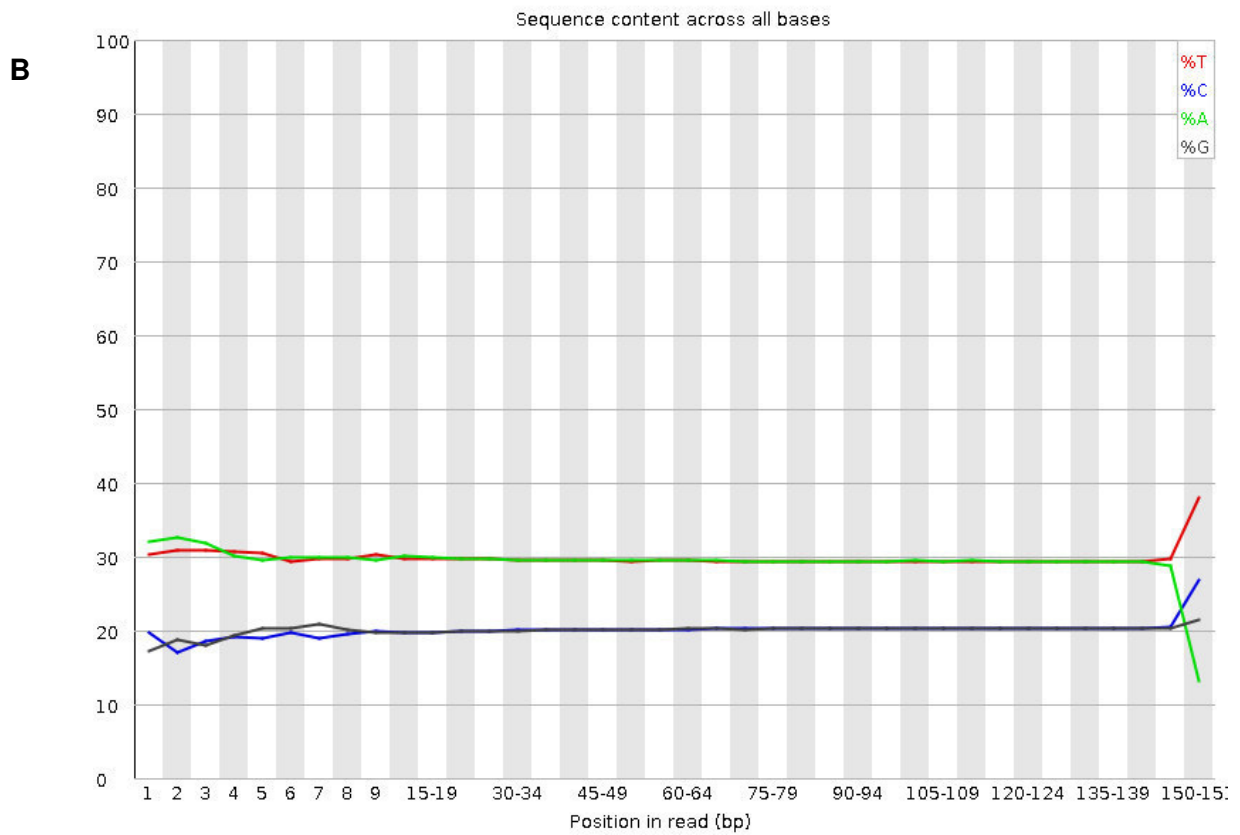
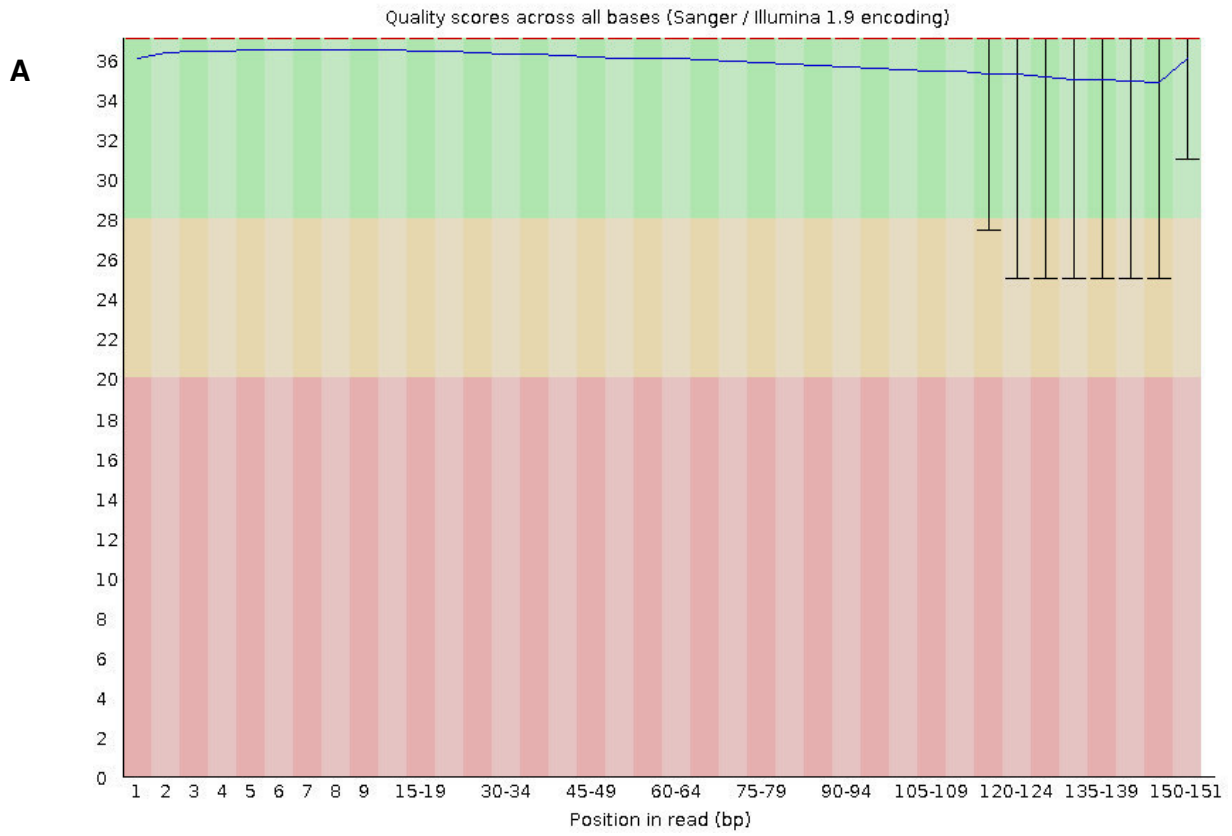


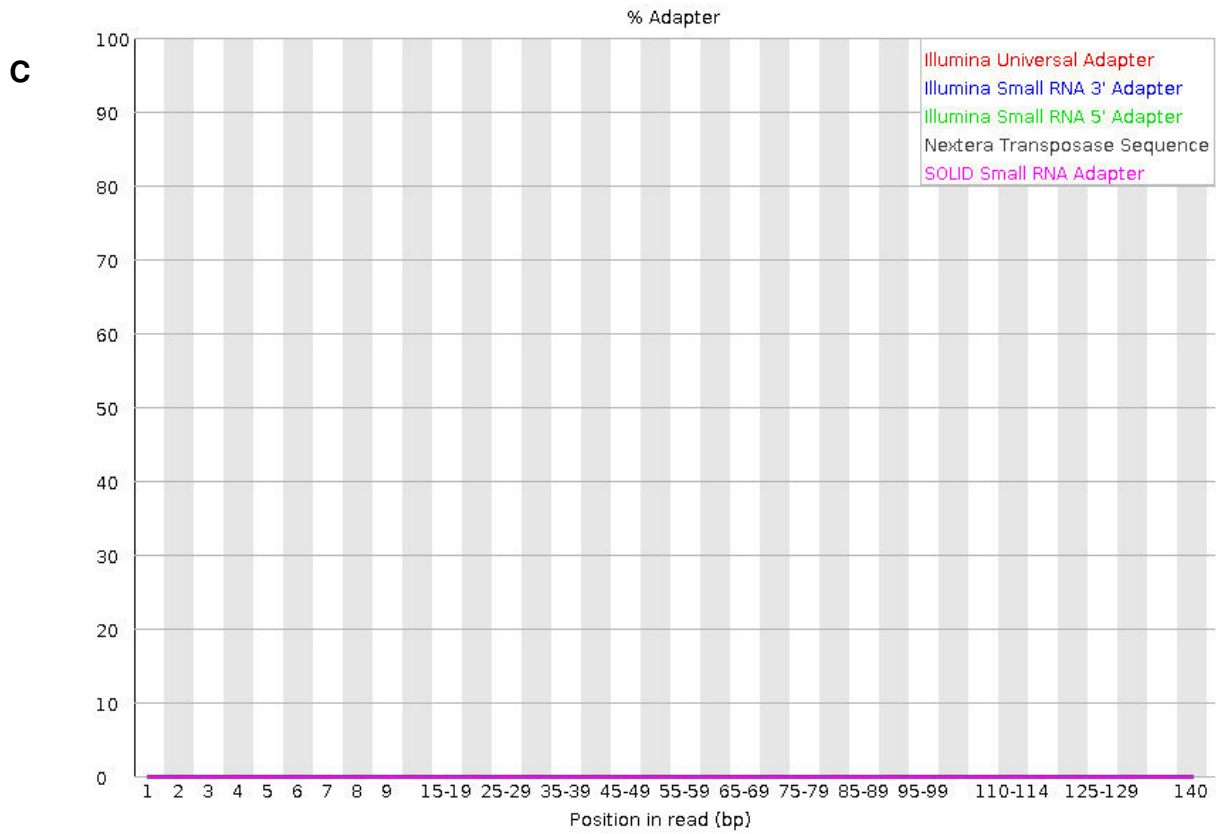
**Figura 10.** Output del programa FastQC donde se visualiza A) la calidad de las lecturas, B) la composición nucleotídica y C) la presencia de adaptadores del archivo R1 de la biblioteca PE de 250 bp después de la limpieza con el programa Trim galore!.



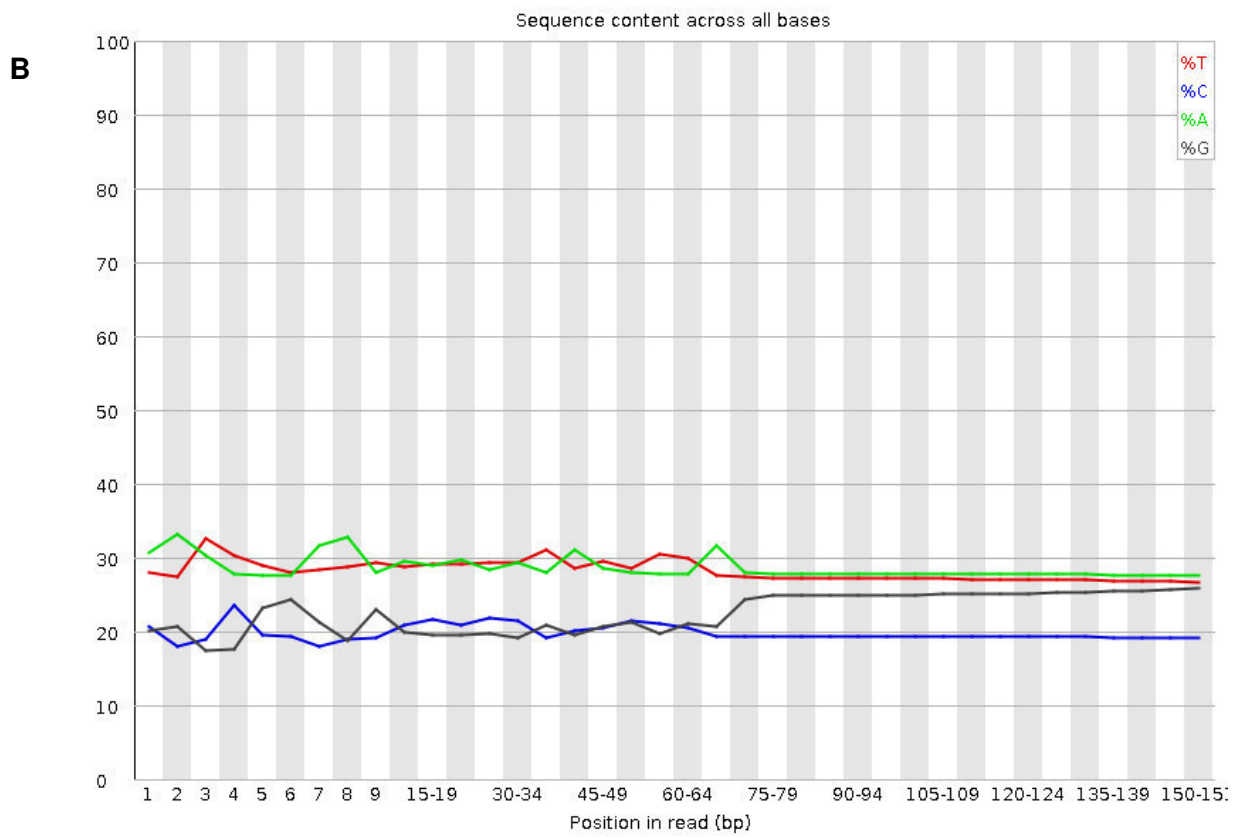
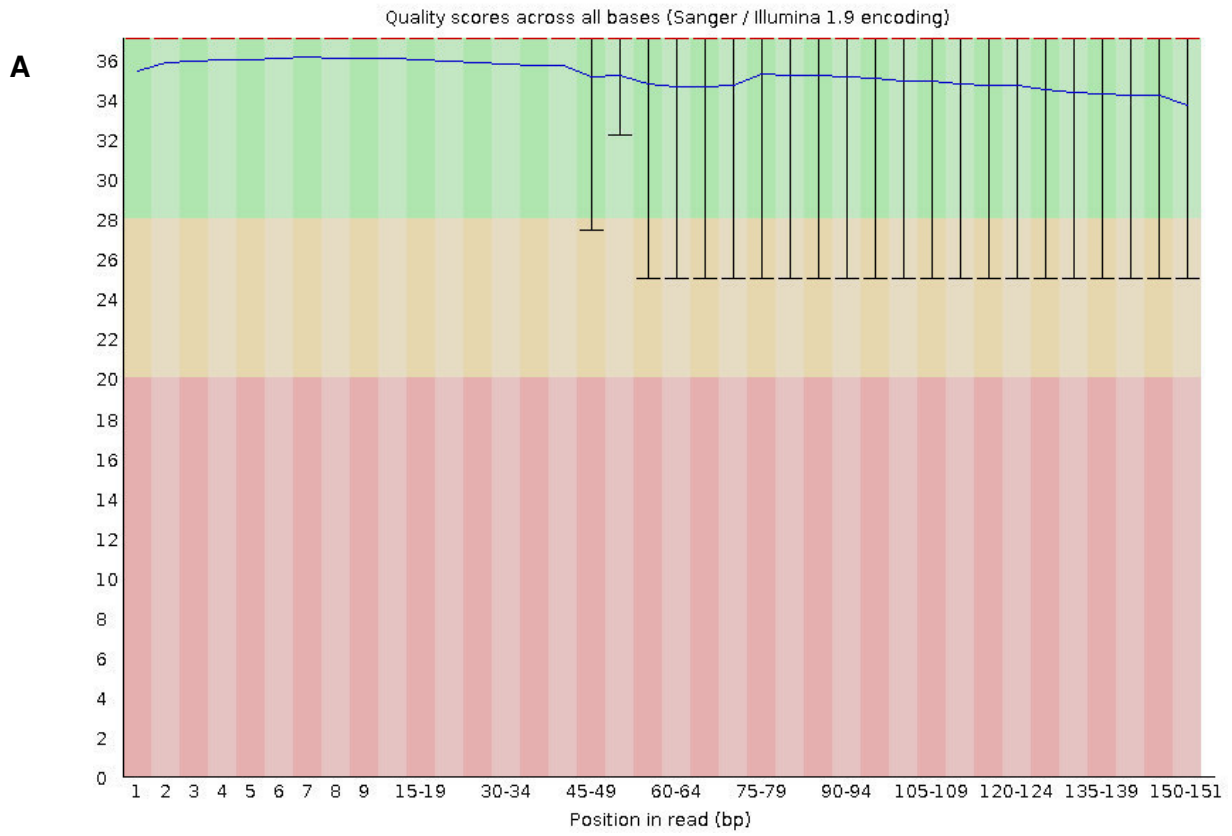


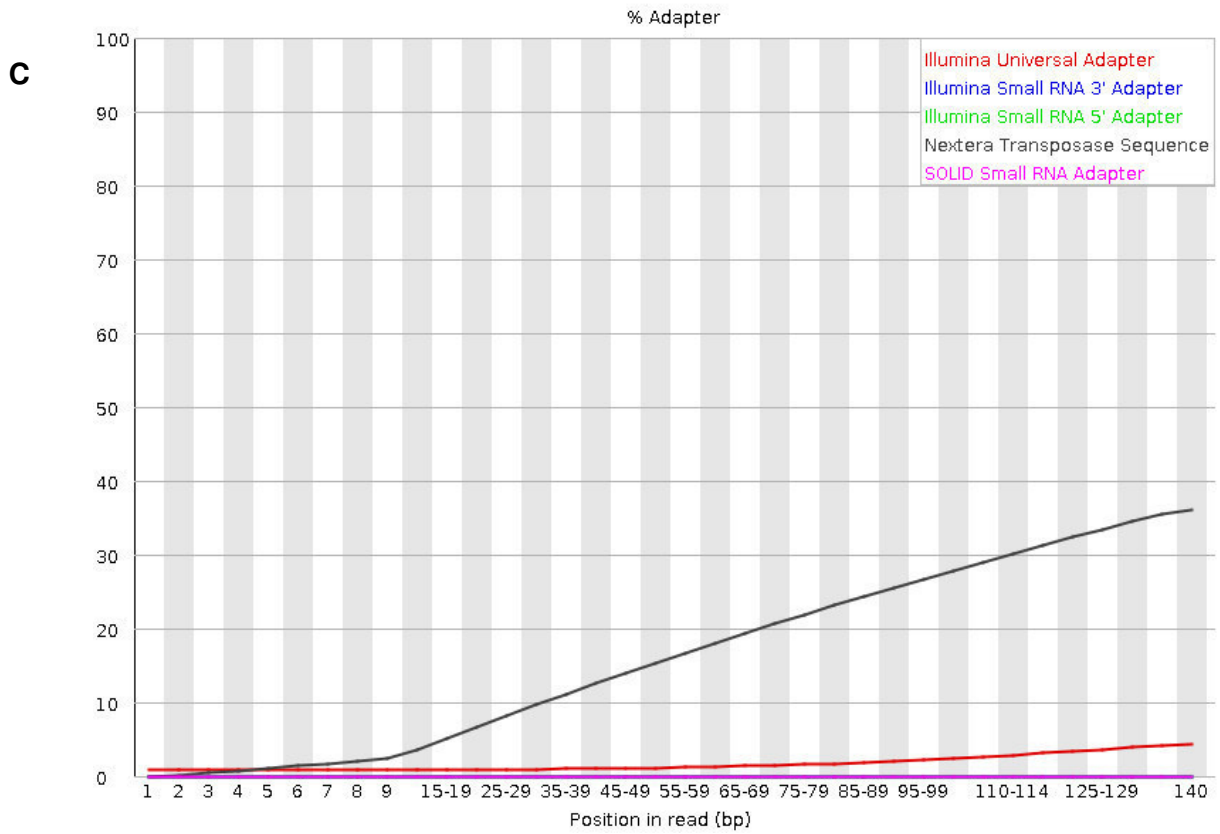
**Figura 11.** Output del programa FastQC donde se visualiza A) la calidad de las lecturas, B) la composición nucleotídica y C) la presencia de adaptadores del archivo R1 de la biblioteca PE de 550 bp antes de la limpieza con el programa Trim galore!.



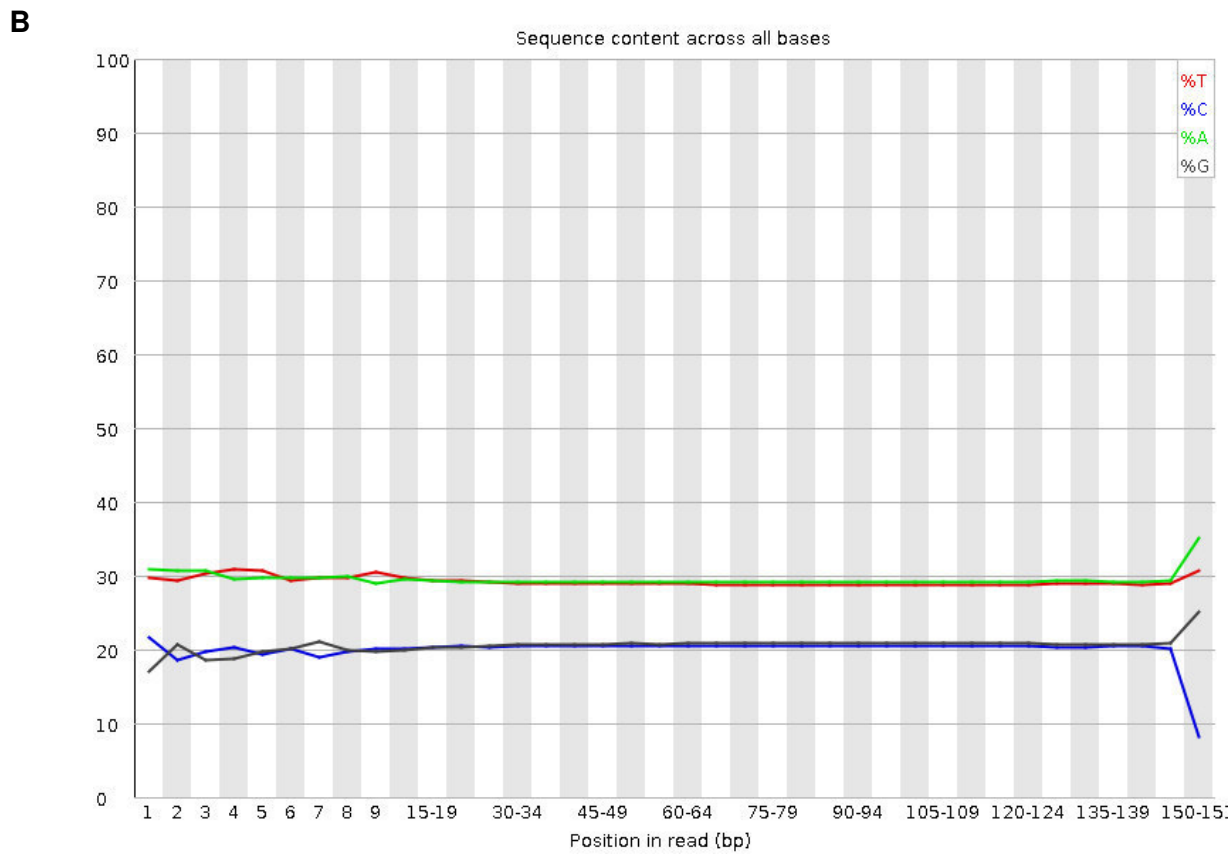
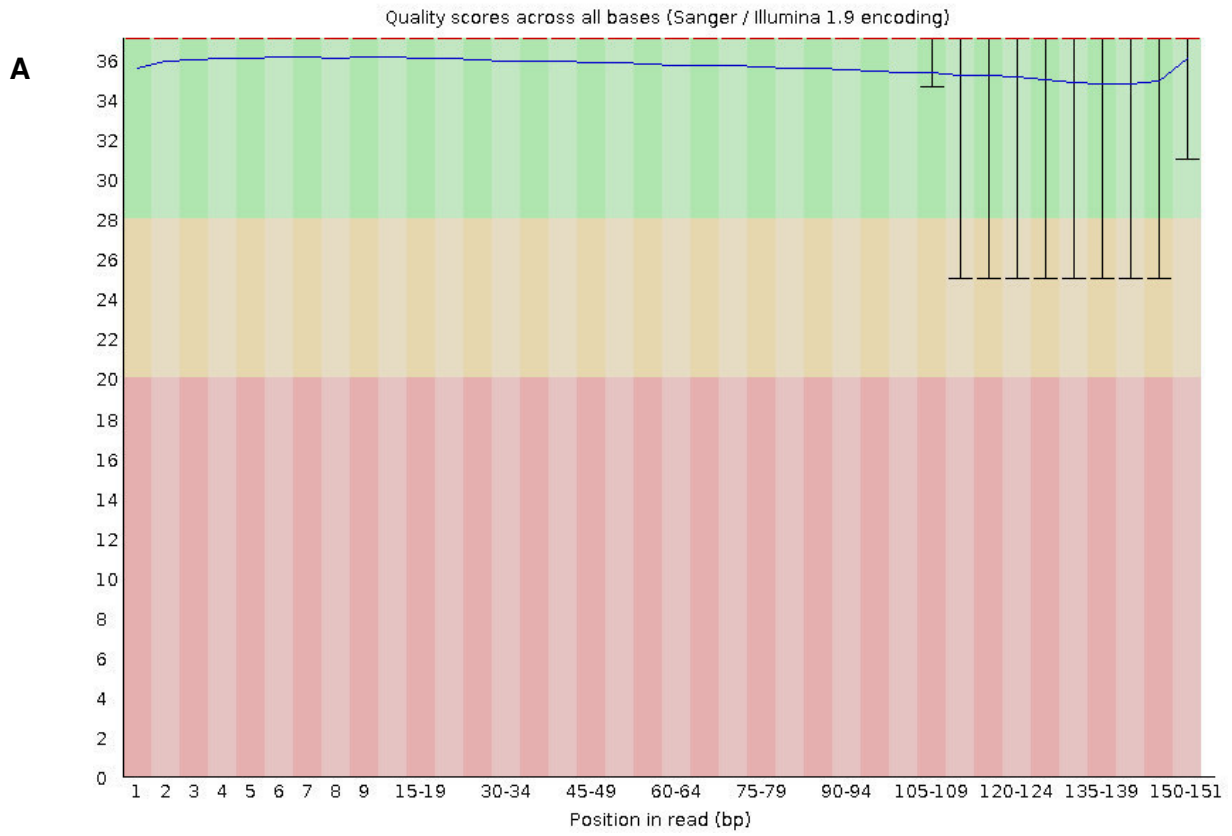


**Figura 12.** Output del programa FastQC donde se visualiza A) la calidad de las lecturas, B) la composición nucleotídica y C) la presencia de adaptadores del archivo R1 de la biblioteca PE de 550 bp después de la limpieza con el programa Trim galore!.

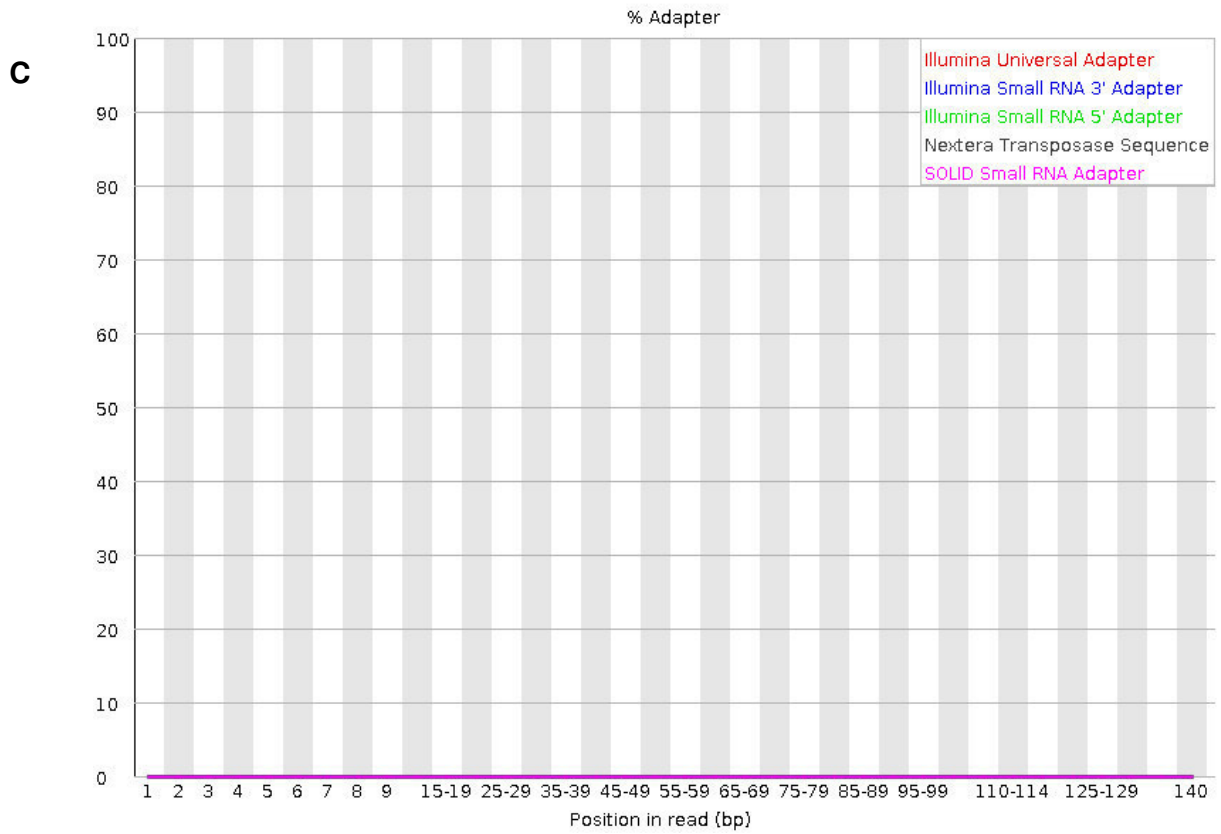




**Figura 13.** Output del programa FastQC donde se visualiza A) la calidad de las lecturas, B) la composición nucleotídica y C) la presencia de adaptadores del archivo R1 de la biblioteca MP de 8 Kbp antes de la clasificación con el programa Nextclip.







**Figura 14.** Output del programa FastQC donde se visualiza A) la calidad de las lecturas, B) la composición nucleotídica y C) la presencia de adaptadores del archivo R1 de la biblioteca MP-usable después de la limpieza con el programa Trim galore!.

# Raw Data Report

April 2019



COPYRIGHT © MACROGEN INC.



## Project Information

Client Name	Andreia Zolet
Company / Institution	Federal University of Rio Grande do Sul
Order Number	HN00104380
Type of Read	Paired-end
Read Length	151
Number of Samples	8
Library Kit	Nextera Mate Pair Sample Prep Kit, TruSeq Nano DNA Kit
Library Protocol	Nextera Mate Pair Library Prep Reference Guide 15035209 v02, TruSeq Nano DNA Sample Preparation Guide, Part # 15041110 Rev. D
Type of Sequencer	Illumina platform

# Table of Contents

---

<b>Project Information</b>	<b>2</b>
<b>1. Data Download Information</b>	<b>4</b>
1. 1. Raw Data and Analysis Results	4
<b>2. Experimental Methods and Workflow</b>	<b>5</b>
2. 1. Experiment Overview	5
2. 2. Generation of Raw Data	6
<b>3. Summary of Produced Data</b>	<b>7</b>
3. 1. Raw Data Statistics	7
3. 2. Total Read Bases	8
3. 3. Total Reads	9
3. 4. GC/AT Content	10
3. 5. Q20/Q30 (%)	11
<b>4. Appendix</b>	<b>12</b>
4. 1. FAQ	12
4. 2. FASTQ File	12
4. 3. Phred Quality Score Chart	12

# 1. Data Download Information

## 1.1. Raw Data and Analysis Results

Download link	File size	md5sum
<a href="#">EUNI_550bp_1.fastq.gz</a>	13.5G	57cd62c6030f40002915884e5c2e5738
<a href="#">EUNI_550bp_2.fastq.gz</a>	14.9G	af978e619a912b9f40d58c92f417693c
<a href="#">EUNI_8kb_1.fastq.gz</a>	8.2G	fc5e7c04215f3e458bb75fdbfd1381b
<a href="#">EUNI_8kb_2.fastq.gz</a>	8.5G	97060f8b5b3d5dce0853d7e41967ed5a
<a href="#">PSI-RD_3kb_1.fastq.gz</a>	7.3G	e26c681b138240efd95a7269c161c3db
<a href="#">PSI-RD_3kb_2.fastq.gz</a>	7.5G	333d862465b54fca83a99ae21a596b21
<a href="#">PSI-RD_550bp_1.fastq.gz</a>	14.8G	0ed58f446cbe4aaa23f902597064d90c
<a href="#">PSI-RD_550bp_2.fastq.gz</a>	16.5G	295cafb4008594a5cfa6b5a4d9d24a70
<a href="#">PSI-RD_8kb_1.fastq.gz</a>	8.3G	55659ec6c7a92134b7d8243056466bb9
<a href="#">PSI-RD_8kb_2.fastq.gz</a>	8.5G	b4442913c4b67b9caf895e5701736204
<a href="#">PSI-YW_3kb_1.fastq.gz</a>	6.8G	f01cf4b2ef7bdc34b02e2ff959e06de7
<a href="#">PSI-YW_3kb_2.fastq.gz</a>	7.1G	850d12f765173eef329aa72167c8eeb5
<a href="#">PSI-YW_550bp_1.fastq.gz</a>	11.6G	2c6635f3d9dccac806550b62a1334941
<a href="#">PSI-YW_550bp_2.fastq.gz</a>	12.8G	876b93dc116d59334b4f528ba11a47d0
<a href="#">PSI-YW_8kb_1.fastq.gz</a>	12.0G	03edd470b2d74dda77d16ac43ec70810
<a href="#">PSI-YW_8kb_2.fastq.gz</a>	12.6G	341ea9ebbf5fe78cbd54c7484eff772b

- fastq.gz : This is a zip file of raw data used in analysis.
- md5sum : In order to verify the integrity of files, md5sum is used. If the values of md5sum are the same, there is no forgery, modification or omission.

**Your data will be retained in our server for 3 months. Should you wish to extend the retention period, please email ([ngs@macrogen.com](mailto:ngs@macrogen.com)) or contact our sales team.**

## 2. Experimental Methods and Workflow

### 2.1. Experiment Overview

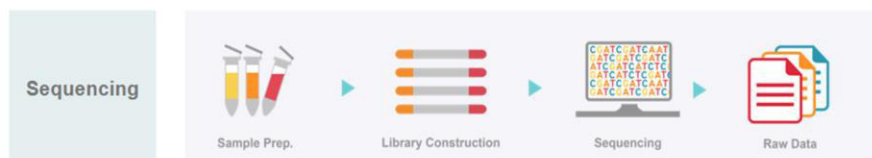


Fig1. Experiment overview

The Illumina NGS workflow includes 4 basic steps :

#### 1) Sample Preparation

For library construction, DNA/RNA is extracted from a sample. After performing quality control (QC), qualified samples proceed to library construction.

#### 2) Library Construction

The sequencing library is prepared by random fragmentation of the DNA or cDNA sample, followed by 5' and 3' adapter ligation. Alternatively, "tagmentation" combines the fragmentation and ligation reactions into a single step that greatly increases the efficiency of the library preparation process. Adapter-ligated fragments are then PCR amplified and gel purified.

#### 3) Sequencing

For cluster generation, the library is loaded into a flow cell where fragments are captured on a lawn of surface-bound oligos complementary to the library adapters. Each fragment is then amplified into distinct, clonal clusters through bridge amplification. When cluster generation is complete, the templates are ready for sequencing.

Illumina SBS technology utilizes a proprietary reversible terminator-based method that detects single bases as they are incorporated into DNA template strands. As all 4 reversible, terminator-bound dNTPs are present during each sequencing cycle, natural competition minimizes incorporation bias and greatly reduces raw error rates compared to other technologies. The result is highly accurate base-by-base sequencing that virtually eliminates sequence-context-specific errors, even within repetitive sequence regions and homopolymers.

#### 4) Raw data

Sequencing data is converted into raw data for the analysis.

## 2. 2. Generation of Raw Data

The Illumina sequencer generates raw images utilizing sequencing control software for system control and base calling through an integrated primary analysis software called RTA (Real Time Analysis). The BCL (base calls) binary is converted into FASTQ utilizing illumina package bcl2fastq. Adapters are not trimmed away from the reads.

### 3. Summary of Produced Data

#### 3.1. Raw Data Statistics

The total number of bases, reads, GC (%), Q20 (%), and Q30 (%) are calculated for the 8 samples. For example, in EUNI\_550bp, 364,930,870 reads are produced, and total read bases are 55.1G bp. The GC content (%) is 40.907% and Q30 is 87.182%.

Table 1. Raw data Stats (maximum 20 samples)

Sample ID	Total read bases (bp)	Total reads	GC(%)	AT(%)	Q20(%)	Q30(%)
EUNI_550bp	55,104,561,370	364,930,870	40.907	59.09	94.283	87.182
EUNI_8kb	34,307,193,054	227,199,954	43.400	56.6	94.236	87.292
PSI-RD_3kb	30,496,593,294	201,964,194	41.778	58.22	95.467	88.956
PSI-RD_550bp	60,493,011,470	400,615,970	39.999	60.0	94.081	86.812
PSI-RD_8kb	34,925,843,376	231,296,976	42.112	57.89	95.018	88.576
PSI-YW_3kb	28,311,217,406	187,491,506	41.760	58.24	94.918	88.043
PSI-YW_550bp	47,431,122,192	314,113,392	39.763	60.24	94.902	87.906
PSI-YW_8kb	50,361,802,672	333,521,872	41.863	58.14	94.794	87.981

- Sample ID : Sample name.
- Total read bases : Total number of bases sequenced.
- Total reads : Total number of reads. For Illumina paired-end sequencing, this value refers to the sum of read 1 and read 2.
- GC(%) : GC content.
- AT(%) : AT content.
- Q20(%) : Ratio of bases that have phred quality score of over 20.
- Q30(%) : Ratio of bases that have phred quality score of over 30.



### 3. 2. Total Read Bases

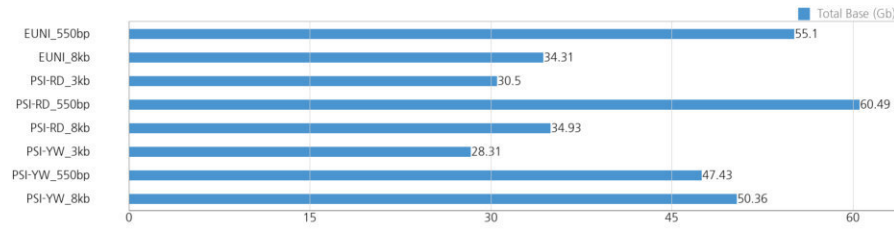


Figure 2.Throughput of Raw data

### 3. 3. Total Reads

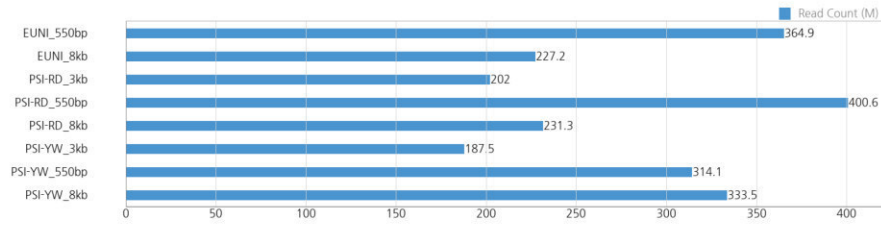


Figure 3. Total read count of Raw data

### 3. 4. GC/AT Content

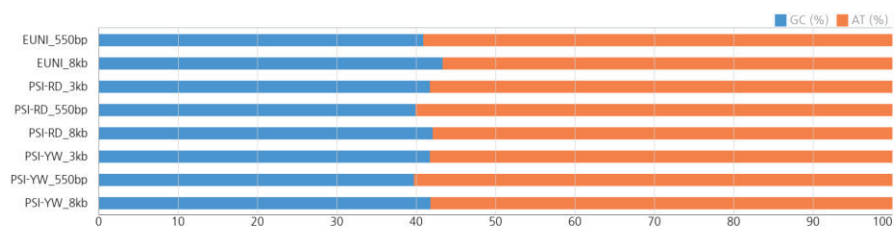


Figure 4. GC/AT Content of Raw data

### 3. 5. Q20/Q30 (%)

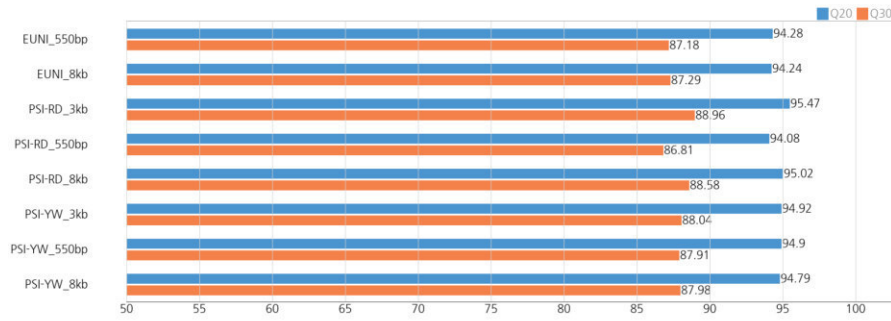


Figure 5. Q20/Q30 scores of Raw data

## 4. Appendix

### 4.1. FAQ

**Q:** I want to see the produced data. How can I open the files?

**A:** As the large size zip files provided by our company are hard to process in the Windows environment, we highly recommend using Linux environment for a smoother operation.

### 4.2. FASTQ File

Example of FASTQ

```
@HISEQ-MFG:501:HB0TFADXX:1:1101:1247:2183 1:N:0:
CTCAGCTAAACTTTGACACNGTANNANNNNNNNNNNTNNNNNNNNNNNN
+
@@@BDDDDHHHHFHIIIIIII#3AC#####
```

FASTQ file is composed of four lines.

Line 1 : ID line includes information such as flow cell lane information.

Line 2 : Sequences line.

Line 3 : Separator line (+ mark).

Line 4 : Quality values line about sequences.

### 4.3. Phred Quality Score Chart

Phred quality score numerically expresses the accuracy of each nucleotide. Higher Q number signifies higher accuracy. For example, if Phred assigns a quality score of 30 to a base, the chances of having base call error are 1 in 1000.

Phred Quality Score Q is calculated with  $-10\log_{10}P$ , where P is probability of erroneous base call.

Quality of phred score	Probability of incorrect base call	Base call accuracy	Characters
10	1 in 10	90%	!"#\$%&'()*+,-./012345
20	1 in 100	99%	,-./012345
30	1 in 1000	99.9%	6789;h=i?
40	1 in 10000	99.99%	@ABCDEFGHIJ

- Encoding : Sanger Quality (ASCII Character Code=Phred Quality Value + 33)



MacroGen Inc. | 08511 10F, 254 Beotkkot-ro, Geumcheon-gu, Seoul, Rep of Korea

**Contact us**  
Tel: +82-2-2180-7016  
Site: [www.macrogen.com](http://www.macrogen.com) | <http://dna.macrogen.com>

**Figura 15.** Reporte del secuenciamiento Illumina realizado por la empresa MacroGen. Las bibliotecas EUNI\_550bp y EUNI\_8kb corresponden a las bibliotecas PE de 550 bp de inserto y MP, respectivamente.

### ANEXO 3

**Tabla 9.** Repeticiones que fueron excluidas por tener similitud por BLAST con secuencias codificadoras de proteínas en Viridiplantae.

Repetición	Anotación	Tamaño	Valor E
R_102	Retrovirus-related Pol polyprotein from transposon RE1	1753	0
R_1026	UBN2 domain-containing protein	1929	8.74E-153
R_1056	benzyl alcohol O-benzoyltransferase	628	7.00E-39
R_1131	retrotransposon protein, putative, Ty1-copia subclass	2404	4.27E-43
R_1142	uncharacterized mitochondrial protein AtMg00810-like	2115	1.47E-116
R_1208	receptor-like protein 9DC3	1584	4.13E-08
R_1238	retrotransposon protein putative Ty1-copia subclass	1840	4.76E-119
R_1273	putative reverse transcriptase domain, Viral movement protein	1780	8.51E-160
R_1278	Retrovirus-related Pol polyprotein from transposon TNT 1-94	1256	1.33E-144
R_1329	probable LRR receptor-like serine/threonine-protein kinase At3g47570	3702	0
R_1378	serine/threonine-protein phosphatase 7 long form homolog	2906	6.98E-141
R_1486	PREDICTED: uncharacterized protein LOC104429064	1003	5.79E-103
R_1512	TMV resistance protein N isoform X1	1027	2.11E-98
R_1576	putative disease resistance protein At3g14460	2159	0
R_1591	coiled-coil domain-containing protein 186-like	580	2.44E-48
R_1659	amino acid permease 2-like	4135	2.98E-56
R_1697	putative nuclease HARBI1	1504	5.07E-164
R_1708	uncharacterized protein LOC115677553	2203	1.88E-24
R_1714	TMV resistance protein N-like isoform X1	848	1.29E-53
R_1791	Transposon Ty3-I Gag-Pol polyprotein	1580	0
R_1820	hypothetical protein EUGRSUZ_G03355	800	9.19E-06
R_1833	DNA mismatch repair protein MSH1, mitochondrial isoform X1	803	4.58E-32
R_1852	TMV resistance protein N-like	759	9.81E-77
R_1877	hypothetical protein CRG98_047229, partial	1757	1.20E-138
R_1881	disease resistance protein RPM1	1055	0
R_1890	Transposase, MuDR, plant	1762	1.08E-72
R_1916	disease resistance protein RPM1	294	2.81E-08
R_1974	receptor-like protein 7	1389	0
R_201	non-structural maintenance of chromosomes element 4 homolog A-like	913	2.35E-18
R_2054	protein FAR1-RELATED SEQUENCE 5-like	995	8.85E-42
R_2243	LEAF RUST 10 DISEASE-RESISTANCE LOCUS RECEPTOR-LIKE PROTEIN KINASE-like 2.2	843	2.02E-96
R_2255	MDIS1-interacting receptor like kinase 2-like	517	3.36E-41
R_2270	retroelement pol polyprotein-like	928	7.18E-76
R_2406	(-)-germacrene D synthase	799	2.74E-51
R_2663	L10-interacting MYB domain-containing protein-like	1051	1.15E-112
R_2740	uncharacterized protein LOC115686568	863	1.75E-58
R_2975	protein PLASTID REDOX INSENSITIVE 2, chloroplastic	147	5.22E-07
R_2980	probable LRR receptor-like serine/threonine-protein kinase At3g47570	1289	2.22E-40

R_2995	MDIS1-interacting receptor like kinase 2-like	1678	0
R_3110	uncharacterized protein LOC115666717	906	1.20E-43
R_3195	transcription termination factor MTERF6, chloroplastic/mitochondrial-like	1485	0
R_3297	ankyrin repeat-containing protein ITN1-like	1256	8.46E-114
R_3310	TMV resistance protein N-like	660	7.60E-78
R_3406	retrovirus-related Pol polyprotein from transposon TNT 1-94	1274	0
R_3493	lipid phosphate phosphatase 2	1263	7.05E-20
R_3663	TMV resistance protein N-like	847	1.09E-122
R_3753	zinc finger BED domain-containing protein RICESLEEPER 1-like	629	2.42E-56
R_3814	pentatricopeptide repeat-containing protein At3g05340	325	1.69E-18
R_3864	kinesin 1	775	3.81E-08
R_3898	DNA excision repair protein ERCC-1 isoform X1	640	1.86E-25
R_3904	PREDICTED: uncharacterized protein LOC107944841	629	8.65E-20
R_3953	B3 domain-containing protein REM16 isoform X4	680	1.51E-10
R_3965	uncharacterized protein LOC115661891	924	3.93E-39
R_3977	TMV resistance protein N-like	411	1.89E-35
R_4141	uncharacterized protein LOC115665076, partial	3326	2.63E-115
R_4174	integrator complex subunit 3	671	5.37E-139
R_4211	Retrovirus-related Pol polyprotein from transposon TNT 1-94	1662	0
R_4271	S-locus lectin protein kinase family protein, putative	1063	6.12E-56
R_4339	hypothetical protein CRG98_019469	1212	3.48E-115
R_4350	zf-CCHC domain-containing protein	688	1.17E-99
R_4395	receptor-like protein 12	690	2.54E-75
R_4428	Von Willebrand factor A domain-containing protein	170	1.52E-10
R_4706	rust resistance kinase Lr10-like	306	6.54E-10
R_4754	LEAF RUST 10 DISEASE-RESISTANCE LOCUS RECEPTOR-LIKE PROTEIN KINASE-like 1.1	947	9.38E-162
R_4847	RNA-directed DNA polymerase	1557	0
R_4958	L-type lectin-domain containing receptor kinase IX.1	783	1.56E-08
R_5003	Retrovirus-related Pol polyprotein from transposon TNT 1-94	410	6.85E-60
R_5019	zinc finger BED domain-containing protein RICESLEEPER 1-like	906	1.91E-127
R_5022	hypothetical protein EUGRSUZ_D024122, partial	948	3.24E-06
R_5065	putative disease resistance protein RGA3	776	3.31E-87
R_5189	uncharacterized protein LOC115663949	1464	7.95E-64
R_622	Gag-pol polyprotein-like protein	1288	1.08E-131
R_806	gag/pol protein	1822	4.25E-105
R_897	protein FAR1-RELATED SEQUENCE 5 isoform X1	2493	0
R_918	heat shock factor protein HSF30-like	697	7.89E-13
R_10058	putative RNA-directed DNA polymerase	775	1.68E-111
R_10095	Retrovirus-related Pol polyprotein from transposon RE2	279	1.33E-07
R_10447	cytochrome P450 714C2-like	258	3.90E-36
R_10844	SART-1 family protein DOT2	376	7.64E-45
R_10868	Retrovirus-related Pol polyprotein from transposon TNT 1-94	688	3.97E-36
R_11067	Gag-pol polyprotein, putative	1098	9.20E-137
R_11150	probable disease resistance protein At1g61310	1271	4.18E-139

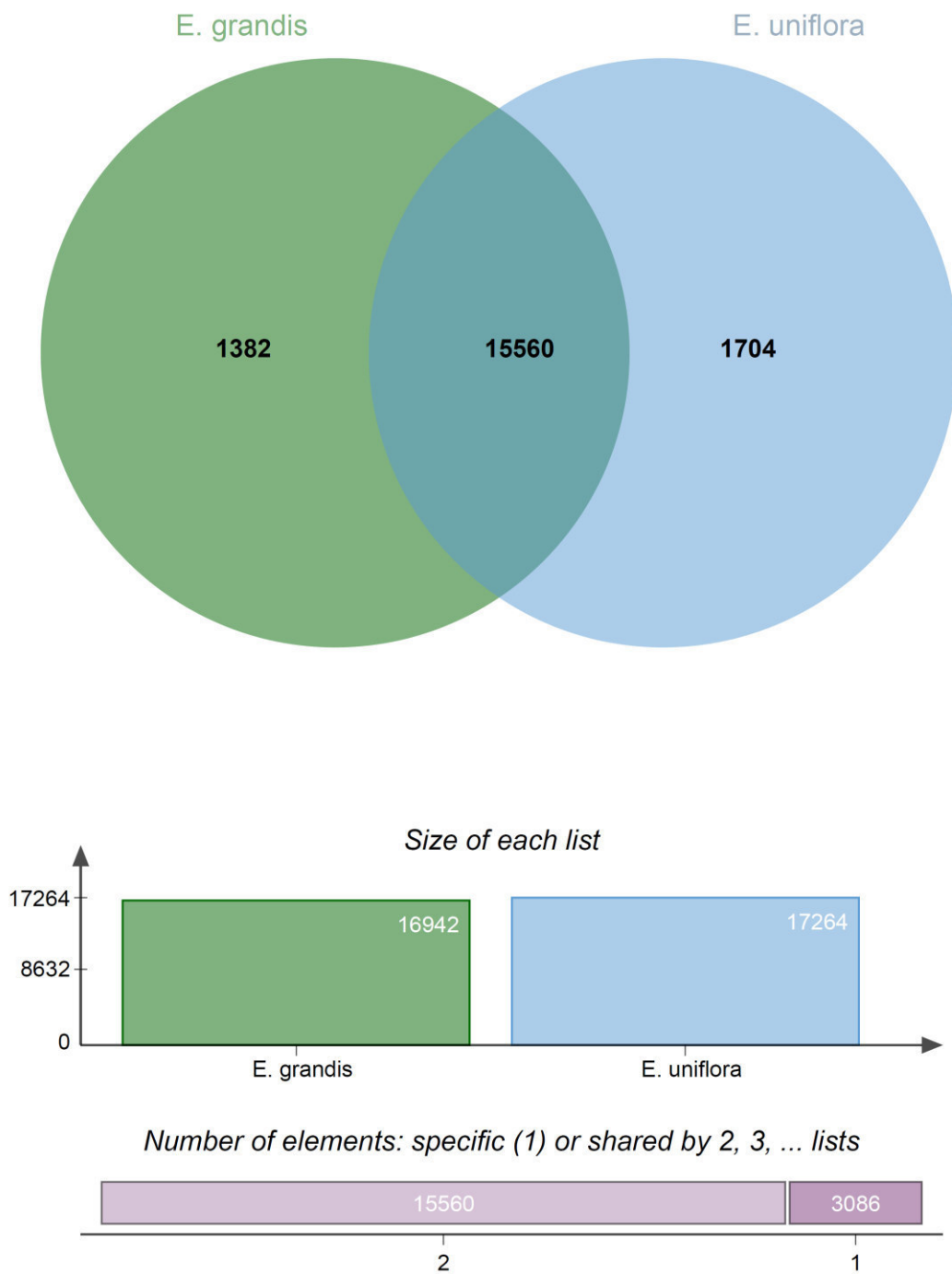


R_11201	patatin-like protein 2	85	7.65E-11
R_11209	G-type lectin S-receptor-like serine/threonine-protein kinase At2g19130	1851	0
R_11316	putative leucine-rich repeat receptor-like protein kinase At2g19210 isoform X2	402	3.88E-47
R_12054	putative RNA-directed DNA polymerase	838	2.35E-114
R_12223	beta-bisabolene synthase-like	392	1.20E-67
R_12692	Retrovirus-related Pol polyprotein from transposon TNT 1-94	1123	4.04E-112
R_13283	PREDICTED: uncharacterized protein LOC108955652	983	1.31E-170
R_13288	cellulose synthase-like protein H1	106	2.40E-15
R_13319	TMV resistance protein N-like isoform X1	2228	7.07E-174
R_13359	uncharacterized protein LOC115679525	870	7.38E-155
R_13846	F-box protein At5g03100-like isoform X3	787	1.63E-39
R_14205	uncharacterized protein LOC115680630	938	1.16E-10
R_14474	aspartic proteinase Asp1-like isoform X2	735	1.44E-15
R_14494	serine carboxypeptidase-like 18 isoform X1	202	3.23E-16
R_14550	MDIS1-interacting receptor like kinase 2-like	1549	0
R_14711	probable leucine-rich repeat receptor-like protein kinase At1g35710	1559	3.85E-155
R_14769	zf-CCHC domain-containing protein/UBN2_2 domain-containing protein	199	4.33E-11
R_15005	Ribosomal protein	110	1.34E-13
R_15590	protein ENHANCED DOWNY MILDEW 2 isoform X2	419	1.17E-06
R_15646	BEACH domain-containing protein C2	935	2.90E-09
R_15829	protein DETOXIFICATION 24 isoform X2	90	7.79E-13
R_16094	protein NRT1/ PTR FAMILY 8.1	907	0
R_16323	polyprotein	792	7.04E-136
R_16494	wall-associated receptor kinase-like 1	1044	0
R_16543	hypothetical protein EUGRSUZ_B02205	722	5.35E-06
R_16635	receptor-like protein kinase FERONIA	652	4.76E-123
R_16638	endo-1,3;1,4-beta-D-glucanase	320	6.44E-11
R_16647	ankyrin repeat-containing protein ITN1 isoform X1	311	5.28E-36
R_16728	Retrovirus-related Pol polyprotein from transposon RE1	655	1.80E-140
R_16730	Retrovirus-related Pol polyprotein from transposon TNT 1-94	1273	0
R_17813	probable serine/threonine-protein kinase PIX13 isoform X2	541	4.70E-17
R_18007	uncharacterized protein LOC115685074	324	5.43E-11
R_18449	probable leucine-rich repeat receptor-like protein kinase At1g35710	718	1.14E-109
R_18560	Serine carboxypeptidase 24	799	2.13E-10
R_19218	NAC domain-containing protein 83-like	297	2.70E-44
R_19507	lysosomal Pro-X carboxypeptidase-like	168	1.94E-24
R_19828	ankyrin repeat-containing protein At5g02620-like isoform X1	671	1.87E-18
R_20020	cytochrome P450 CYP736A12-like	644	9.73E-58
R_20630	uncharacterized protein LOC115692171	827	3.72E-10
R_20947	acyltransferase-like protein At3g26840, chloroplastic isoform X2	333	1.14E-23
R_21207	ankyrin repeat-containing protein BDA1-like	250	4.92E-07
R_21630	TMV resistance protein N-like	264	7.68E-10
R_21714	aspartyl protease family protein At5g10770	1029	3.09E-142
R_22035	rust resistance kinase Lr10 isoform X1	216	1.54E-16

R_22291	cullin-1 isoform X1	991	1.29E-90
R_23009	G-type lectin S-receptor-like serine/threonine-protein kinase At4g27290	780	1.22E-116
R_5238	cellulose synthase-like protein H1	1161	6.26E-76
R_5248	putative B3 domain-containing protein At1g78640	504	1.59E-13
R_5419	(-)-germacrene D synthase-like	1535	6.41E-129
R_5674	Ty3/gypsy retrotransposon protein	764	7.22E-55
R_5696	uncharacterized protein LOC115693178	3276	1.08E-149
R_5697	ankyrin repeat-containing protein At5g02620-like isoform X2	701	1.22E-44
R_5817	galactinol synthase 2	649	1.50E-151
R_5851	uncharacterized protein LOC115666035	1021	0
R_5867	zinc finger BED domain-containing protein RICESLEEPER 2-like	1251	5.36E-77
R_5980	zinc finger BED domain-containing protein RICESLEEPER 2-like	1255	0
R_6121	probable 2-oxoglutarate-dependent dioxygenase At5g05600	287	2.20E-51
R_6134	disease resistance protein At4g27190-like	687	7.04E-34
R_6150	RNA-directed DNA polymerase	847	2.81E-151
R_6200	Retrovirus-related Pol polyprotein from transposon TNT 1-94	181	1.95E-27
R_6208	rust resistance kinase Lr10 isoform X1	1237	0
R_6212	ankyrin repeat-containing protein ITN1-like	1266	6.35E-110
R_6379	probable leucine-rich repeat receptor-like protein kinase At1g35710	634	5.35E-115
R_6413	CCHC-type integrase	772	3.52E-75
R_6540	PREDICTED: uncharacterized protein LOC109166774	1203	3.01E-66
R_6946	receptor-like protein 12	646	1.48E-82
R_7067	Glutamine synthetase leaf isozyme, chloroplastic	228	8.72E-15
R_7134	protein ROOT HAIR DEFECTIVE 3-like	585	1.88E-10
R_7334	RNA-directed DNA polymerase, eukaryota, Reverse transcriptase zinc-binding domain protein	359	1.32E-28
R_7374	hypothetical protein CRG98_029300	266	4.25E-06
R_7466	putative ribosome biogenesis protein C8F11.04	206	2.07E-06
R_7578	neurofilament medium polypeptide	153	3.72E-08
R_7648	LEAF RUST 10 DISEASE-RESISTANCE LOCUS RECEPTOR-LIKE PROTEIN KINASE-like 1.1	1022	0
R_7780	rust resistance kinase Lr10 isoform X1	1313	0
R_7811	uncharacterized protein LOC115693178	2722	6.72E-153
R_7842	(-)-germacrene D synthase	970	6.78E-18
R_7844	UBN2_3 domain-containing protein	633	4.06E-79
R_7945	disease resistance protein TAO1-like	904	2.68E-96
R_7956	uncharacterized protein E5676_scaffold237G00650	666	2.26E-15
R_8085	lipid phosphate phosphatase 2	1507	6.63E-19
R_8360	TMV resistance protein N	1749	0
R_8425	G-type lectin S-receptor-like serine/threonine-protein kinase At1g61370 isoform X1	1188	1.89E-63
R_8429	G-type lectin S-receptor-like serine/threonine-protein kinase At2g19130	1129	0
R_8443	TMV resistance protein N-like	283	1.07E-41
R_8519	copla-like polyprotein	958	6.64E-136
R_8669	G-type lectin S-receptor-like serine/threonine-protein kinase At4g27290	760	1.97E-134
R_8790	uncharacterized protein LOC115679517	97	4.28E-12

R_8837	putative F-box/LRR-repeat protein At5g02700	177	4.39E-07
R_8918	UPF0481 protein At3g47200	728	6.05E-131
R_9064	serine/threonine-protein phosphatase 7 long form homolog	1112	7.12E-50
R_9066	interactor of constitutive active ROPs 2, chloroplastic-like	822	2.11E-121
R_9075	hypothetical protein EUGRSUZ_J01625	459	3.70E-09
R_9440	Retrovirus-related Pol polyprotein from transposon TNT 1-94	590	7.45E-87
R_9606	unnamed protein product, partial	242	1.84E-18
R_9666	hypersensitive-induced response protein 1-like	611	1.15E-11
R_9689	actin-related protein 9	732	1.32E-51
R_9709	glutamine synthetase leaf isozyme, chloroplastic isoform X2	221	6.85E-09
R_9822	Retrovirus-related Pol polyprotein from transposon TNT 1-94	920	6.02E-81

ANEXO 4



**Figura 16.** Diagrama de Venn mostrando los clústeres de genes ortólogos identificados entre *E.uniflora* y *E. grandis*.

## ANEXO 5

**Tabla 10.** Anotación de los genes exclusivos de *E. uniflora*.

Código	Tamaño	Anotación	Valor E
eug14020	154	17.6 kDa class II heat shock protein-like	6.41E-95
eug3657	154	17.6 kDa class II heat shock protein-like	1.16E-90
eug13075	159	18.1 kDa class I heat shock protein-like	2.56E-95
eug16834	157	18.1 kDa class I heat shock protein-like	1.28E-32
eug17261	141	18.1 kDa class I heat shock protein-like	4.67E-31
eug17262	123	18.1 kDa class I heat shock protein-like	1.61E-39
eug21437	139	18.1 kDa class I heat shock protein-like	4.47E-36
eug21509	159	18.1 kDa class I heat shock protein-like	9.05E-95
eug3395	126	18.1 kDa class I heat shock protein-like	3.25E-32
eug3592	139	18.1 kDa class I heat shock protein-like	4.68E-10
eug20389	73	1-aminocyclopropane-1-carboxylate oxidase	2.95E-36
eug20797	73	1-aminocyclopropane-1-carboxylate oxidase	1.09E-36
eug4855	111	1-aminocyclopropane-1-carboxylate oxidase homolog 1-like isoform X1	7.82E-47
eug8502	114	1-aminocyclopropane-1-carboxylate oxidase homolog 1-like isoform X1	1.35E-48
eug123	98	2-alkenal reductase (NADP(+)-dependent)	2.51E-33
eug13879	98	2-alkenal reductase (NADP(+)-dependent)	4.08E-33
eug21402	97	2-alkenal reductase (NADP(+)-dependent)	2.18E-28
eug26777	91	2-alkenal reductase (NADP(+)-dependent)	1.40E-09
eug9743	93	2-alkenal reductase (NADP(+)-dependent)	3.60E-36
eug9744	98	2-alkenal reductase (NADP(+)-dependent)	6.56E-35
eug18099	85	2-alkenal reductase (NADP(+)-dependent)-like	5.16E-44
eug4629	85	2-alkenal reductase (NADP(+)-dependent)-like	2.56E-44
eug10698	96	2-alkenal reductase NADP+-dependent	7.41E-07
eug8013	112	2-oxoglutarate-Fe(II) type oxidoreductase hxnY-like isoform X2	1.48E-06
eug12572	133	33 kDa ribonucleoprotein, chloroplastic	2.07E-30
eug26176	126	33 kDa ribonucleoprotein, chloroplastic	2.24E-30
eug25111	178	3-oxoacyl-[acyl-carrier-protein] reductase FabG-like	1.90E-112
eug25621	177	3-oxoacyl-[acyl-carrier-protein] reductase FabG-like	1.19E-111
eug10945.2	156	40S ribosomal protein S3-3	3.08E-82
eug4027	129	40S ribosomal protein S3-3	6.48E-85
eug4412	262	40S ribosomal protein S3a	0
eug7491	262	40S ribosomal protein S3a	0
eug18771	154	40S ribosomal protein S6	2.08E-14
eug13046	109	AAA-ATPase ASD, mitochondrial	1.19E-07
eug20932	99	AAA-ATPase ASD, mitochondrial-like	4.32E-35
eug23441	99	AAA-ATPase ASD, mitochondrial-like	3.98E-35
eug6725	212	AAA-ATPase ASD, mitochondrial-like	1.32E-07
eug28119	338	ABC transporter B family member 19-like	3.49E-85
eug29861	186	ABC transporter B family member 19-like	3.63E-61
eug25397	117	ABC transporter C family member 12	2.73E-67

eug28514	70	ABC transporter C family member 12	1.17E-39
eug28837	70	ABC transporter C family member 12	7.91E-39
eug7921	117	ABC transporter C family member 12-like isoform X1	4.24E-61
eug14008	127	ACT domain-containing protein ACR4 isoform X2	3.70E-17
eug20705	169	Actin cross-linking protein	7.79E-14
eug19249	357	Actin cytoskeleton-regulatory complex protein pan1, putative isoform 2	0
eug9575	366	Actin cytoskeleton-regulatory complex protein pan1, putative isoform 2	0
eug23825	265	acyl-CoA N-acyltransferase	5.93E-29
eug10393	262	acyl-CoA-binding domain-containing protein 4	1.83E-168
eug16326	278	acyl-CoA-binding domain-containing protein 4	2.97E-167
eug29852	176	acyl-coenzyme A thioesterase 9, mitochondrial-like isoform X1	4.44E-59
eug6002	136	acyl-coenzyme A thioesterase 9, mitochondrial-like isoform X1	3.11E-58
eug311	108	adenosine monophosphate-protein transferase FICD homolog	3.61E-16
eug9541	157	agamous-like MADS-box protein AGL29	1.39E-18
eug26708	67	agamous-like MADS-box protein AGL61	1.46E-11
eug589	92	agamous-like MADS-box protein AGL61	6.28E-22
eug21498	148	alcohol dehydrogenase 7	2.98E-14
eug12721	165	aldehyde oxidase GLOX1-like	3.73E-71
eug21325	188	aldehyde oxidase GLOX1-like	1.61E-71
eug24176	70	alpha/beta hydrolase domain-containing protein 17B	4.11E-18
eug30563	70	alpha/beta hydrolase domain-containing protein 17B	7.38E-20
eug29705	191	Alpha/beta hydrolase fold	9.42E-65
eug12927	173	amino acid permease 3	5.56E-57
eug12928	181	amino acid permease 3	3.04E-70
eug12929	173	amino acid permease 3	4.30E-58
eug12930	124	amino acid permease 3	2.89E-69
eug24602	124	amino acid permease 3	2.91E-65
eug29759	67	amino acid permease 3	8.61E-10
eug674	124	amino acid permease 3	8.29E-62
eug675.2	74	amino acid permease 3	2.46E-17
eug676	124	Amino acid permease 3	2.72E-61
eug12931	80	amino acid permease 3-like	3.11E-11
eug16495	87	aminopeptidase M1-like	3.04E-40
eug17712	98	aminopeptidase M1-like	3.34E-40
eug24927	211	aminopeptidase M1-like	2.45E-46
eug19701	358	ammonium transporter 1 member 1-like	1.50E-51
eug20637	195	ammonium transporter 1 member 4-like	6.50E-43
eug23055	74	AMN1 homolog	5.96E-13
eug10179	99	anaphase-promoting complex subunit 2-like	8.09E-32
eug14469	204	ankyrin repeat-containing protein BDA1-like	7.64E-62
eug17986	107	ankyrin repeat-containing protein BDA1-like	1.66E-41
eug17990	101	ankyrin repeat-containing protein BDA1-like	3.12E-33
eug7649	231	ankyrin repeat-containing protein BDA1-like	2.58E-73
eug25662	203	ankyrin repeat-containing protein ITN1	5.75E-106
eug7667	252	ankyrin repeat-containing protein ITN1	3.15E-133

eug12669	360	ankyrin repeat-containing protein NPR4-like	1.76E-96
eug17306	187	ankyrin repeat-containing protein NPR4-like	2.23E-99
eug18134	130	ankyrin repeat-containing protein NPR4-like	4.63E-21
eug6929	172	anthocyanidin 3-O-glucosyltransferase 2	6.50E-78
eug20368	377	anthocyanidin 3-O-glucosyltransferase 2-like	1.09E-70
eug10112	127	antifungal protein ginkbilobin-like protein	9.79E-18
eug15144	134	antifungal protein ginkbilobin-like protein	9.77E-36
eug20948	113	antifungal protein ginkbilobin-like protein	5.21E-31
eug29026	136	antifungal protein ginkbilobin-like protein	1.31E-44
eug5167	135	antifungal protein ginkbilobin-like protein	1.24E-40
eug21837	109	AP-1 complex subunit gamma-2-like	5.05E-36
eug6127	73	AP-1 complex subunit gamma-2-like	2.74E-32
eug19379	85	aquaporin NIP2-1	1.10E-33
eug23576	116	aquaporin NIP2-1	5.04E-29
eug28719	182	armadillo repeat-containing protein 8	1.10E-90
eug29330	200	armadillo repeat-containing protein 8	8.00E-86
eug16064	180	aspartic proteinase Asp1-like	6.79E-110
eug24207	212	aspartic proteinase Asp1-like	1.17E-100
eug10989	109	aspartic proteinase CDR1-like	1.18E-38
eug27882	109	aspartic proteinase CDR1-like	4.97E-39
eug14023	208	Aspartyl protease family protein 1	3.18E-34
eug28131	208	aspartyl protease family protein 1	7.19E-27
eug11776	200	AT3G14810-like protein	8.26E-27
eug13672	197	AT-hook motif nuclear-localized protein 10-like isoform X1	3.38E-94
eug13673	318	AT-hook motif nuclear-localized protein 10-like isoform X1	0
eug2524	262	AT-hook motif nuclear-localized protein 10-like isoform X1	4.34E-89
eug8432	78	AT-hook motif nuclear-localized protein 20-like	9.49E-11
eug3652	129	AT-hook motif nuclear-localized protein 22-like	2.62E-10
eug8000	354	AT-hook motif nuclear-localized protein 6-like	8.01E-98
eug20126	88	ATP synthase subunit delta', mitochondrial	2.18E-36
eug20127	88	ATP synthase subunit delta', mitochondrial	2.66E-34
eug20718	88	ATP synthase subunit delta', mitochondrial	1.32E-39
eug3870	96	ATP-dependent Clp protease proteolytic subunit-related protein 4, chloroplastic	5.56E-42
eug8411	96	ATP-dependent Clp protease proteolytic subunit-related protein 4, chloroplastic	4.37E-40
eug26667	91	AT-rich interactive domain-containing protein 1-like	7.30E-36
eug7247	215	AT-rich interactive domain-containing protein 1-like	4.65E-53
eug12081	232	auxin efflux carrier component 5	1.76E-139
eug12082	89	auxin efflux carrier component 5	5.53E-43
eug22695	89	auxin efflux carrier component 5	8.31E-43
eug22696	264	auxin efflux carrier component 5	1.29E-118
eug11041	105	auxin-induced protein 15A-like	2.59E-08
eug17532	105	auxin-induced protein 15A-like	7.77E-09
eug20103	129	auxin-responsive protein SAUR32-like	3.44E-66
eug23568	128	auxin-responsive protein SAUR32-like	1.31E-67

eug23570	131	auxin-responsive protein SAUR32-like	1.66E-61
eug10941	118	auxin-responsive protein SAUR50-like	4.60E-52
eug237	217	B box-type domain-containing protein	6.76E-128
eug22252	192	B3 domain-containing protein	1.83E-07
eug4732	281	B3 domain-containing protein	8.97E-163
eug16838	442	B3 domain-containing protein At2g24670	6.90E-114
eug15132	106	B3 domain-containing protein Os01g0723500-like	1.30E-36
eug21771	106	B3 domain-containing protein Os01g0723500-like	3.44E-34
eug22330	168	B3 domain-containing protein Os01g0723500-like	4.52E-62
eug7581	106	B3 domain-containing protein Os01g0723500-like	4.40E-36
eug5189	160	B3 domain-containing protein Os06g0194400-like	2.58E-64
eug7890	153	B3 domain-containing protein Os06g0194400-like	6.17E-64
eug11439	265	B3 domain-containing protein REM14-like isoform X2	2.28E-148
eug1686	111	BAG family molecular chaperone regulator 5, mitochondrial	1.04E-12
eug20604	73	BAG family molecular chaperone regulator 5, mitochondrial	1.22E-17
eug22599	169	basic leucine zipper 34-like	4.50E-56
eug6005	168	basic leucine zipper 34-like	3.75E-62
eug21312	372	beta-bisabolene synthase-like	0
eug3102.2	415	beta-bisabolene synthase-like	4.80E-146
eug5160	283	beta-bisabolene synthase-like	0
eug1083	109	beta-glucosidase 11-like isoform X2	9.20E-62
eug19306	109	beta-glucosidase 11-like isoform X2	1.36E-61
eug29881	87	beta-glucosidase 46-like isoform X2	1.54E-37
eug6935	87	beta-glucosidase 46-like isoform X2	5.24E-38
eug16334	170	beta-hexosaminidase 1	9.56E-26
eug14050	119	beta-hexosaminidase 1-like	2.05E-41
eug19357	153	BI1-like protein	2.57E-16
eug19637	179	BI1-like protein	6.03E-35
eug6389	78	BI1-like protein	4.78E-17
eug6390	99	BI1-like protein	3.75E-15
eug17656	420	bifunctional aspartate aminotransferase and glutamate/aspartate-prephenate aminotransferase	0
eug18665	390	bifunctional aspartate aminotransferase and glutamate/aspartate-prephenate aminotransferase	2.87E-147
eug21548	212	bifunctional dihydrofolate reductase-thymidylate synthase-like	5.69E-56
eug6709	337	bifunctional dihydrofolate reductase-thymidylate synthase-like	1.91E-48
eug15657	159	bifunctional epoxide hydrolase 2-like isoform X1	8.61E-39
eug18600	123	bifunctional epoxide hydrolase 2-like isoform X1	1.67E-45
eug15087	332	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1-like	0
eug19531	328	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1-like	0
eug28685.1	343	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1-like	1.69E-172
eug5937	333	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1-like	1.88E-169
eug9709	250	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1-like	2.59E-16
eug6917	697	Bromodomain and PHD finger-containing 3	7.24E-65
eug27551	697	bromodomain-containing protein 3-like	3.51E-67
eug20662	117	BTB/POZ domain-containing protein At3g09030	3.75E-42



eug24974	114	BTB/POZ domain-containing protein At3g09030	1.50E-41
eug12133	84	BTB/POZ domain-containing protein At5g41330	2.55E-25
eug17050	91	BTB/POZ domain-containing protein At5g41330	2.16E-23
eug17809	150	BTB/POZ domain-containing protein At5g41330	8.11E-23
eug1655	101	Calmodulin	2.04E-33
eug3338	118	calmodulin	1.64E-53
eug26923	499	calmodulin-binding protein 60 D-like	0
eug29856	588	calmodulin-binding protein 60 D-like isoform X2	0
eug4827	146	calmodulin-like protein 8	1.24E-82
eug487	153	calmodulin-like protein 8	3.08E-102
eug10548	118	carbon catabolite repressor protein 4 homolog 5	5.47E-53
eug25626	118	carbon catabolite repressor protein 4 homolog 5	7.81E-52
eug11674	70	CASP-like protein 1C1	1.94E-38
eug28269	83	CASP-like protein 1C1	7.75E-42
eug12911	192	cation/H(+) antiporter 18-like	1.35E-22
eug6494	151	cation/H(+) antiporter 18-like	5.80E-22
eug3948	80	Cationic amino acid transporter 6, chloroplastic	6.29E-31
eug17107	93	cationic amino acid transporter 7, chloroplastic-like	6.09E-25
eug16294	75	CBL-interacting serine/threonine-protein kinase 14-like	6.42E-12
eug19241	147	CBL-interacting serine/threonine-protein kinase 14-like	1.76E-10
eug12591	336	CCR4-NOT transcription complex subunit 9-like isoform X1	0
eug26356	91	CDK-activating kinase assembly factor MAT1	6.83E-54
eug3020	91	CDK-activating kinase assembly factor MAT1	2.38E-53
eug6767	165	cell differentiation protein rcd1-like isoform X2	4.84E-40
eug29944	152	cell wall / vacuolar inhibitor of fructosidase 2-like	1.56E-71
eug29945	122	cell wall / vacuolar inhibitor of fructosidase 2-like	1.11E-52
eug12855	122	cellulose synthase A catalytic subunit 3 [UDP-forming]	2.13E-52
eug21011	122	cellulose synthase A catalytic subunit 3 [UDP-forming]	1.20E-45
eug16610	111	cellulose synthase-like protein E6	7.77E-64
eug26322	111	cellulose synthase-like protein E6	1.54E-61
eug14276	483	chaperone protein ClpB1	5.88E-119
eug13808	88	chaperone protein dnaJ C76, chloroplastic isoform X1	6.79E-34
eug8837	112	chaperone protein dnaJ C76, chloroplastic isoform X1	1.45E-53
eug23244	228	chlorophyllase-2, chloroplastic-like	4.59E-149
eug23625	215	chlorophyllase-2, chloroplastic-like	6.10E-121
eug23627	244	chlorophyllase-2, chloroplastic-like	1.01E-145
eug17130	116	chromatin modification-related protein eaf-1-like	5.15E-45
eug210	82	clathrin light chain 1-like	5.98E-18
eug23932	153	clathrin light chain 1-like	1.55E-21
eug11238	164	CLP protease regulatory subunit CLPX2, mitochondrial isoform X1	5.66E-36
eug11747	283	CLP protease regulatory subunit CLPX2, mitochondrial isoform X1	1.79E-76
eug27467	261	CLP protease regulatory subunit CLPX2, mitochondrial isoform X1	7.40E-73
eug29389	179	CLP protease regulatory subunit CLPX2, mitochondrial isoform X1	8.32E-59
eug19823	192	collagen alpha-1(XXV) chain-like	3.12E-87
eug19824	204	collagen alpha-1(XXV) chain-like	3.52E-66

eug28290	207	collagen alpha-1(XXV) chain-like	2.26E-90
eug7285	197	collagen alpha-1(XXV) chain-like	8.30E-61
eug20066	86	collagen alpha-1(XXVII) chain A-like isoform X1	9.75E-13
eug20067	255	collagen alpha-1(XXVII) chain A-like isoform X1	7.40E-27
eug20068	131	collagen alpha-1(XXVII) chain A-like isoform X1	3.42E-24
eug20069	168	collagen alpha-1(XXVII) chain A-like isoform X1	1.40E-19
eug8881	111	conserved hypothetical protein	5.41E-06
eug1674	143	cotton fiber protein	1.69E-55
eug18367	240	cotton fiber protein	2.16E-46
eug10234	553	coumaroyl-CoA:anthocyanidin 3-O-glucoside-6''-O-coumaroyltransferase 1-like	0
eug30006	467	coumaroyl-CoA:anthocyanidin 3-O-glucoside-6''-O-coumaroyltransferase 1-like	0
eug23339	128	cucumis-in-like	1.10E-46
eug15194	452	cyanidin 3-O-galactoside 2''-O-xylosyltransferase FGGT1-like	0
eug15197	636	cyanidin 3-O-galactoside 2''-O-xylosyltransferase FGGT1-like	0
eug17034	457	cyanidin 3-O-galactoside 2''-O-xylosyltransferase FGGT1-like	0
eug1419	387	Cysteine-rich domain-containing protein	7.24E-111
eug17510	323	Cysteine-rich domain-containing protein	2.54E-76
eug25407.2	409	cysteine-rich receptor-like protein kinase 10	3.77E-117
eug3290	476	cysteine-rich receptor-like protein kinase 25	6.71E-91
eug13264	331	cysteine-rich receptor-like protein kinase 29	1.16E-155
eug18677	311	cysteine-rich receptor-like protein kinase 29	1.91E-150
eug6388	76	cytochrome b5	3.62E-09
eug21503	168	cytochrome b561 and DOMON domain-containing protein At5g47530-like	3.26E-96
eug52	177	cytochrome b561 and DOMON domain-containing protein At5g47530-like	1.91E-100
eug12788	166	cytochrome b561 and DOMON domain-containing protein At5g48750-like	1.19E-96
eug51	166	cytochrome b561 and DOMON domain-containing protein At5g48750-like	1.19E-96
eug18342	183	cytochrome P450 81E8-like	5.31E-60
eug7437	193	cytochrome P450 81E8-like	4.56E-66
eug4237	96	cytochrome P450 94A1	3.00E-15
eug1836	130	cytochrome P450 94A2-like	1.55E-13
eug16678	289	cytochrome P450 CYP749A22-like	0
eug8869	289	cytochrome P450 CYP749A22-like	0
eug24051.1	219	DEAD-box ATP-dependent RNA helicase 24	4.31E-125
eug9899.1	218	DEAD-box ATP-dependent RNA helicase 24	4.13E-124
eug3854	204	DEAD-box ATP-dependent RNA helicase 28	9.65E-123
eug9774	204	DEAD-box ATP-dependent RNA helicase 28	9.65E-123
eug690	91	dehydrin 1	1.14E-25
eug12024	163	dehydrin DHN1	1.02E-49
eug7666	104	DELLA protein GAI	7.18E-34
eug24215	158	DELLA protein GAI-like	5.15E-41
eug12732	86	denticleless protein homolog	1.31E-18
eug19581	111	denticleless protein homolog	1.14E-25

eug30255	328	disease resistance protein At4g27190-like	1.29E-132
eug18849	105	disease resistance protein RPM1-like	7.20E-11
eug20704	107	disease resistance protein RPM1-like	1.76E-17
eug20706	243	disease resistance protein RPM1-like	4.57E-66
eug213	533	disease resistance protein RPM1-like	1.26E-112
eug17701	222	disease resistance protein RPV1-like	2.23E-50
eug1546	93	D-lactate dehydrogenase [cytochrome], mitochondrial isoform X1	1.84E-15
eug20925	96	D-lactate dehydrogenase [cytochrome], mitochondrial isoform X2	9.24E-18
eug1964	594	DNA polymerase V	1.72E-128
eug7103	117	DNA polymerase zeta catalytic subunit	8.93E-56
eug19406	209	DNA polymerase zeta catalytic subunit isoform X1	7.97E-105
eug2494	165	DNA polymerase zeta catalytic subunit isoform X1	2.01E-56
eug20537	199	DNA polymerase zeta catalytic subunit isoform X2	2.01E-96
eug10392	133	DNA topoisomerase 3-beta isoform X2	2.87E-59
eug9710	133	DNA topoisomerase 3-beta isoform X2	2.87E-59
eug28254	133	DNA-directed RNA polymerase I subunit rpa49-like	7.72E-76
eug4638	133	DNA-directed RNA polymerase I subunit rpa49-like	1.79E-74
eug17669	90	dnaJ homolog subfamily B member 9-like	2.68E-21
eug5634	76	dnaJ homolog subfamily B member 9-like	4.49E-17
eug22706	162	DUF1685 family protein	4.80E-54
eug7040	123	DUF1685 family protein	2.69E-17
eug22254	448	DUF313 domain-containing protein	1.42E-30
eug22256	410	DUF313 domain-containing protein	6.51E-21
eug13696	289	DUF4283 domain-containing protein/zf-CCHC_4 domain-containing protein	2.47E-12
eug10700	186	DUF581 domain-containing protein	6.62E-46
eug13838	156	DUF581 domain-containing protein	4.20E-12
eug18207	191	DUF581 domain-containing protein	4.00E-43
eug28605	184	DUF581 domain-containing protein	2.13E-45
eug9588	246	DUF674 domain-containing protein	1.73E-140
eug13014	73	dynamamin-related protein 4C-like	6.27E-26
eug20479.1	286	dynamamin-related protein 4C-like	3.88E-165
eug2183	291	dynamamin-related protein 4C-like	2.58E-175
eug7823	75	dynamamin-related protein 4C-like	3.59E-26
eug15541	96	E3 ubiquitin ligase BIG BROTHER-related	4.80E-25
eug5871	83	E3 ubiquitin ligase BIG BROTHER-related	4.50E-06
eug30364	266	E3 ubiquitin-protein ligase ATL23-like	7.06E-13
eug3150	70	E3 ubiquitin-protein ligase ATL23-like	3.06E-15
eug11264	185	E3 ubiquitin-protein ligase ATL31-like	1.41E-14
eug29045	186	E3 ubiquitin-protein ligase ATL31-like	1.26E-22
eug29901	264	E3 ubiquitin-protein ligase ATL6-like	4.35E-29
eug20060	145	E3 ubiquitin-protein ligase BRE1-like 1	7.36E-78
eug27387	145	E3 ubiquitin-protein ligase BRE1-like 1	4.14E-77
eug2461	88	E3 ubiquitin-protein ligase Praja-1	3.80E-12
eug30564	88	E3 ubiquitin-protein ligase Praja-1	5.60E-12

eug12535	241	E3 ubiquitin-protein ligase Praja-1-like	1.92E-126
eug6789	218	E3 ubiquitin-protein ligase Praja-1-like	2.69E-113
eug23561	118	E3 ubiquitin-protein ligase RNF12-B-like	4.08E-13
eug28848	123	E3 ubiquitin-protein ligase RNF12-B-like	1.97E-23
eug1370	109	early nodulin-75-like	9.85E-29
eug24163	170	early nodulin-75-like	4.84E-47
eug20306	232	ecdysone-induced protein 75B	8.09E-101
eug24317	83	ecdysone-induced protein 75B	6.10E-31
eug24489	120	Egg cell-secreted protein 1.4	3.07E-52
eug25395	120	Egg cell-secreted protein 1.4	7.00E-46
eug4219	120	Egg cell-secreted protein 1.4	1.94E-54
eug3350	127	ELKS/Rab6-interacting/CAST family	4.84E-35
eug8705	310	endoglucanase 11	0
eug16672	231	endoglucanase 11-like	7.68E-145
eug16673	223	endoglucanase 11-like	3.38E-142
eug3423	314	endoglucanase 11-like	0
eug20855	130	ethylene-responsive transcription factor 11-like	5.35E-10
eug18664	101	ethylene-responsive transcription factor 4	3.31E-24
eug19643	163	ethylene-responsive transcription factor 4	2.72E-32
eug7384	134	ethylene-responsive transcription factor 4	1.12E-48
eug27750	230	ethylene-responsive transcription factor CRF4-like	6.28E-85
eug27751	162	ethylene-responsive transcription factor CRF4-like	5.12E-62
eug27752.2	243	ethylene-responsive transcription factor CRF4-like	3.21E-90
eug27838	289	ethylene-responsive transcription factor CRF4-like	5.60E-126
eug11985	119	ethylene-responsive transcription factor ERF084	8.79E-07
eug7620	110	ethylene-responsive transcription factor ERF084	8.69E-12
eug23421	170	ethylene-responsive transcription factor ESR2-like	2.86E-42
eug24281	144	ethylene-responsive transcription factor ESR2-like	1.20E-42
eug23950	106	ethylene-responsive transcription factor RAP2-12	1.66E-28
eug639	106	ethylene-responsive transcription factor RAP2-12	1.07E-26
eug12099	172	FAR1-RELATED SEQUENCE 10	1.21E-48
eug12102	172	FAR1-RELATED SEQUENCE 10	4.26E-64
eug26795	145	fasciclin-like arabinogalactan protein 10	1.30E-34
eug8783	122	fasciclin-like arabinogalactan protein 10	1.02E-27
eug16207	164	F-box protein At2g02240-like isoform X1	8.59E-69
eug20865.1	197	F-box protein PP2-B10 isoform X1	2.15E-83
eug30581	206	F-box protein PP2-B10 isoform X1	8.46E-87
eug10598	127	F-box protein PP2-B11-like isoform X4	3.42E-42
eug7030	282	F-box protein SKIP1	6.11E-146
eug2501	302	F-box protein SKIP1-like	4.09E-155
eug2512	377	F-box protein SKIP22-like	6.03E-112
eug29463	343	F-box protein SKIP22-like	7.87E-74
eug6777	440	F-box/FBD/LRR-repeat protein At1g13570-like	0
eug999	434	F-box/FBD/LRR-repeat protein At1g13570-like	0
eug16498	312	F-box/kelch-repeat protein At1g57790-like	4.24E-72

eug29524	320	F-box/kelch-repeat protein At1g57790-like	1.27E-78
eug20863	175	FCS-Like Zinc finger 2	7.60E-11
eug3883	156	FCS-Like Zinc finger 2 isoform X1	5.87E-14
eug25532	443	flavin-containing monooxygenase FMO GS-OX-like 2 isoform X1	0
eug9853	448	flavin-containing monooxygenase FMO GS-OX-like 2 isoform X1	0
eug17824	194	flotillin-like protein 2	3.95E-52
eug16039	103	flowering-promoting factor 1-like	4.02E-51
eug16040	103	flowering-promoting factor 1-like protein 2	5.24E-58
eug15925	124	formin-like protein 14	5.86E-27
eug29329	154	formin-like protein 14	8.49E-27
eug14614	116	fringe-related protein	1.41E-18
eug23525	119	fringe-related protein	1.89E-19
eug17263	216	fructose-bisphosphate aldolase 1, cytoplasmic-like	1.02E-91
eug3396	175	fructose-bisphosphate aldolase 1, cytoplasmic-like	6.99E-81
eug3397	218	fructose-bisphosphate aldolase 6, cytosolic	1.37E-87
eug625	135	galactinol synthase 2	9.73E-23
eug29633	83	GATA transcription factor 5-like	1.07E-16
eug18944	104	GATA transcription factor 6	1.28E-14
eug13707	217	GDSL esterase/lipase At1g29670-like	1.35E-93
eug27165.1	280	GDSL esterase/lipase At3g27950-like	8.47E-123
eug29453	376	GDSL esterase/lipase At3g27950-like	0
eug5352	373	GDSL esterase/lipase At3g27950-like	9.01E-174
eug11152	361	GDSL esterase/lipase At5g45670-like	4.22E-156
eug24333	360	GDSL esterase/lipase At5g45670-like	1.17E-167
eug27481	344	GDSL esterase/lipase At5g45670-like	3.96E-157
eug11495	297	GDSL esterase/lipase EXL3-like	0
eug9839	278	GDSL esterase/lipase EXL3-like	2.49E-175
eug25039	129	GEM-like protein 5	5.70E-30
eug30507	212	GEM-like protein 5	6.40E-36
eug390	129	GEM-like protein 5	6.02E-29
eug25159	103	geranylgeranyl diphosphate reductase, chloroplastic-like	5.59E-06
eug7863	118	geranylgeranyl diphosphate reductase, chloroplastic-like	2.97E-13
eug9618	155	geranylgeranyl diphosphate reductase, chloroplastic-like	1.71E-41
eug10517	192	geranylgeranyl pyrophosphate synthase 7, chloroplastic-like	4.18E-58
eug10665	113	gibberellin 20 oxidase 2-like	1.64E-43
eug19642	90	gibberellin 20 oxidase 2-like	1.50E-49
eug20466	104	gibberellin 20 oxidase 2-like	8.17E-49
eug2834	90	gibberellin 20 oxidase 2-like	4.24E-47
eug29427	195	gibberellin 2-beta-dioxygenase 6	4.85E-80
eug29429.1	140	gibberellin 2-beta-dioxygenase 6	3.01E-75
eug20260	169	glutamate dehydrogenase 1	4.78E-20
eug26556	299	glutamate dehydrogenase 1	3.42E-10
eug4254	173	glutamate dehydrogenase 1	1.74E-21
eug16460	106	glutamate-1-semialdehyde 2,1-aminomutase, chloroplastic-like	3.48E-47
eug4782	108	glutamate-1-semialdehyde 2,1-aminomutase, chloroplastic-like	1.30E-38

eug29568	112	glutaredoxin domain-containing cysteine-rich protein 1	7.44E-52
eug808	189	glutaredoxin domain-containing cysteine-rich protein CG12206	2.54E-65
eug13603	227	glutathione S-transferase U8-like	2.19E-146
eug15205	214	glutathione S-transferase U8-like	2.06E-138
eug25228	225	glutathione S-transferase U8-like	2.51E-147
eug20518	202	glycine, alanine and asparagine-rich protein-like	5.14E-82
eug20519	200	glycine, alanine and asparagine-rich protein-like	1.37E-87
eug9307	245	glycine-rich cell wall structural protein 1.0-like	3.65E-22
eug28475	92	glycine-rich cell wall structural protein 1-like	3.74E-07
eug13759	148	glycine-rich cell wall structural protein-like	5.15E-37
eug13760	339	glycine-rich cell wall structural protein-like	1.48E-31
eug28852	131	glycine-rich protein 23-like	7.66E-26
eug28853	131	glycine-rich protein 23-like	3.58E-21
eug28854	137	glycine-rich protein 23-like	1.24E-19
eug28855	138	glycine-rich protein 23-like	2.66E-25
eug8040	83	glycine-rich protein 23-like	3.37E-15
eug16917	159	glycine-rich protein 5-like	1.77E-15
eug23203	129	glycine-rich protein 5-like	2.27E-16
eug29478	165	glycine-rich protein 5-like	2.83E-18
eug5399	122	glycine-rich protein 5-like	5.03E-09
eug23058	172	glycine-rich protein A3-like	2.06E-20
eug27472	177	glycine-rich protein A3-like	2.64E-36
eug403	107	glycine-rich protein DOT1-like	3.38E-17
eug9374	107	glycine-rich protein DOT1-like	1.24E-17
eug14763.2	428	Glycine--tRNA ligase, mitochondrial 1	0
eug22472	315	glycine--tRNA ligase, mitochondrial 1	5.74E-152
eug11397	513	glycine--tRNA ligase, mitochondrial 1-like	0
eug9075	73	glyoxylate/hydroxypyruvate reductase HPR3	5.46E-12
eug12	112	glyoxylate/hydroxypyruvate reductase HPR3-like	4.12E-46
eug13	73	glyoxylate/hydroxypyruvate reductase HPR3-like	8.96E-12
eug9074	112	glyoxylate/hydroxypyruvate reductase HPR3-like	3.93E-42
eug23491	272	GRAS family protein RAM1-like	2.69E-44
eug22221	204	GRAS family protein RAM1-like isoform X1	2.08E-33
eug4340	148	GRF1-interacting factor 3	5.29E-07
eug2738	150	GTP_EFTU_D3 domain-containing protein/Abhydrolase_5 domain-containing protein	2.13E-62
eug15528	242	G-type lectin S-receptor-like serine/threonine-protein kinase At4g03230	1.93E-107
eug22401	203	G-type lectin S-receptor-like serine/threonine-protein kinase At4g03230	1.07E-83
eug29709	109	G-type lectin S-receptor-like serine/threonine-protein kinase CES101 isoform X1	5.57E-43
eug29643	130	G-type lectin S-receptor-like serine/threonine-protein kinase SD1-29	1.23E-21
eug7352	102	G-type lectin S-receptor-like serine/threonine-protein kinase SD1-29	9.77E-07
eug7756	117	G-type lectin S-receptor-like serine/threonine-protein kinase SD1-29	2.47E-09
eug27201	155	Guanosine nucleotide diphosphate dissociation inhibitor	1.77E-07
eug18043	88	heat shock 70 kDa protein	3.84E-33

eug27068	75	heat shock 70 kDa protein	1.16E-19
eug2725	88	Heat shock cognate protein 80	3.34E-51
eug6608	88	Heat shock cognate protein 80	3.34E-51
eug17434	89	heavy metal-associated isoprenylated plant protein 2	5.13E-10
eug73	105	heavy metal-associated isoprenylated plant protein 2	2.66E-13
eug74	149	heavy metal-associated isoprenylated plant protein 2	3.31E-12
eug17660	112	heavy metal-associated isoprenylated plant protein 34	1.82E-46
eug8936	112	heavy metal-associated isoprenylated plant protein 34	5.92E-45
eug9300	93	heavy metal-associated isoprenylated plant protein 34	4.19E-30
eug20445	247	heavy metal-associated isoprenylated plant protein 34-like	4.79E-20
eug981	111	heavy metal-associated isoprenylated plant protein 34-like	2.66E-42
eug13924.1	280	helicase sen1	1.18E-149
eug17870.1	305	helicase sen1	5.54E-140
eug6050	346	heparan-alpha-glucosaminide N-acetyltransferase	9.19E-111
eug24835	297	heparan-alpha-glucosaminide N-acetyltransferase-like	8.88E-64
eug14829	104	heterodimeric geranylgeranyl pyrophosphate synthase large subunit 1, chloroplastic-like	1.20E-38
eug23357	99	hexokinase 2	1.02E-36
eug6634	103	Hexokinase 2	3.67E-37
eug13283	275	hexokinase-1	0
eug29710	275	hexokinase-1-like isoform X1	0
eug15860	152	high mobility group B protein 3-like	8.18E-43
eug11715	146	histidine-containing phosphotransfer protein 1-like	2.88E-83
eug11716	90	histidine-containing phosphotransfer protein 1-like	7.17E-51
eug12906	146	histidine-containing phosphotransfer protein 1-like	1.26E-97
eug8124	121	histidine-containing phosphotransfer protein 2-like	4.15E-62
eug8125	70	histidine-containing phosphotransfer protein 2-like	1.01E-31
eug24165	127	histidine-rich glycoprotein-like	1.72E-36
eug24167	123	histidine-rich glycoprotein-like	1.01E-33
eug24169	127	histidine-rich glycoprotein-like	3.53E-37
eug10672	83	histone-lysine N-methyltransferase ASHR2	1.05E-38
eug20628	256	histone-lysine N-methyltransferase ASHR2	2.41E-38
eug29049	91	histone-lysine N-methyltransferase SETD2-like	1.69E-22
eug17389	291	histone-lysine N-methyltransferase, H3 lysine-9 specific SUVH6-like	2.32E-103
eug17391	314	histone-lysine N-methyltransferase, H3 lysine-9 specific SUVH6-like	3.27E-103
eug15694	288	HMG1/2-like protein	1.07E-76
eug15862	149	HMG1/2-like protein	1.52E-35
eug19947	79	HVA22-like protein e	2.33E-07
eug29220	130	HVA22-like protein e	4.07E-17
eug17979	132	hydroxyproline-rich glycoprotein family protein	1.27E-08
eug9654	269	hydroxyproline-rich glycoprotein family protein	6.75E-10
eug19053	292	Hypothetical predicted protein	2.02E-18
eug24402	324	Hypothetical predicted protein	0
eug27087	416	Hypothetical predicted protein	1.63E-30
eug27091	254	Hypothetical predicted protein	7.59E-41

eug8381	108	Hypothetical predicted protein	2.36E-07
eug5139	98	hypothetical protein	5.93E-08
eug9740	90	hypothetical protein	1.25E-07
eug4790	204	hypothetical protein ACMD2_05090, partial	9.91E-14
eug17175	154	hypothetical protein BHE74_00050784	9.48E-07
eug8459	140	hypothetical protein BT93_A0105	2.91E-32
eug17297	150	hypothetical protein BT93_A0944	1.17E-31
eug12367	96	hypothetical protein BT93_A1606	5.02E-06
eug19879	128	hypothetical protein BT93_A1617	1.94E-19
eug20253	136	hypothetical protein BT93_A1617	1.94E-12
eug16373	137	hypothetical protein BT93_A1661	1.28E-16
eug18174	123	hypothetical protein BT93_A1661	7.34E-13
eug13489	79	hypothetical protein BT93_A2191	1.16E-07
eug16423	156	hypothetical protein BT93_B0589	4.79E-61
eug3388	189	hypothetical protein BT93_B0589	3.12E-63
eug26801	113	hypothetical protein BT93_B1397	1.44E-10
eug2062	88	hypothetical protein BT93_C1557	6.08E-20
eug26897	88	hypothetical protein BT93_C1557	6.02E-20
eug590	123	hypothetical protein BT93_C1557	9.45E-11
eug13698	77	hypothetical protein BT93_C1995	5.61E-21
eug4821.1	666	hypothetical protein BT93_C2030	0
eug3720	180	hypothetical protein BT93_C2324	7.38E-55
eug22684	203	hypothetical protein BT93_D0722	1.26E-76
eug9305	311	hypothetical protein BT93_D1465	2.47E-23
eug2552	125	hypothetical protein BT93_D1885	7.24E-06
eug4220	128	hypothetical protein BT93_E0220	3.56E-50
eug13247	125	hypothetical protein BT93_E1405	4.70E-43
eug26702	99	hypothetical protein BT93_E1405	7.67E-22
eug13509	109	hypothetical protein BT93_E1502	6.73E-35
eug3155	111	hypothetical protein BT93_E1502	9.09E-40
eug13778	77	hypothetical protein BT93_F1784	5.67E-23
eug21345	81	hypothetical protein BT93_F1784	7.54E-28
eug29937	216	hypothetical protein BT93_G0211	8.29E-55
eug14972	74	hypothetical protein BT93_H0975	1.96E-27
eug14973	74	hypothetical protein BT93_H0975	1.72E-27
eug8830	72	hypothetical protein BT93_H2607	3.63E-22
eug8831	91	hypothetical protein BT93_H2607	9.21E-27
eug8833	77	hypothetical protein BT93_H2607	4.85E-22
eug24349	94	hypothetical protein BT93_H2982	2.00E-16
eug23164	174	hypothetical protein BT93_H3713	8.37E-21
eug4708	206	hypothetical protein BT93_H3713	2.29E-35
eug29440	101	hypothetical protein BT93_H3773	6.94E-06
eug13995	103	hypothetical protein BT93_I0196	1.35E-30
eug15839	103	hypothetical protein BT93_I0196	2.20E-29
eug19876	94	hypothetical protein BT93_I0346	1.32E-27



eug8575	94	hypothetical protein BT93_I0346	5.71E-29
eug18657	94	hypothetical protein BT93_I0347	2.78E-29
eug18659	94	hypothetical protein BT93_I0347	6.08E-31
eug11131	123	hypothetical protein BT93_I0776	1.68E-30
eug14389	98	hypothetical protein BT93_I0778	2.19E-17
eug9082	453	hypothetical protein BT93_J0793	2.06E-78
eug25429	70	hypothetical protein BT93_J1502	5.05E-07
eug772	175	hypothetical protein BT93_J1502	7.97E-07
eug24750	184	hypothetical protein BT93_K0453	2.55E-18
eug14210	108	hypothetical protein BT93_K0644	5.99E-57
eug14211	161	hypothetical protein BT93_K0644	5.11E-10
eug25879	138	hypothetical protein BT93_K1807	6.02E-56
eug25882	130	hypothetical protein BT93_K1807	1.24E-34
eug19392	92	hypothetical protein BT93_K1994	6.25E-20
eug10768	82	hypothetical protein BT93_L0939	1.33E-10
eug23590	251	hypothetical protein BT93_L1479	1.35E-133
eug16408	114	hypothetical protein BT93_L2608	2.27E-13
eug16137	86	hypothetical protein CDL15_Pgr002347	5.76E-07
eug13168	407	hypothetical protein CDL15_Pgr009339	6.60E-159
eug2294	407	hypothetical protein CDL15_Pgr009339	2.19E-159
eug22080	154	hypothetical protein CK203_073061	4.81E-10
eug23163	111	hypothetical protein CRG98_005410	2.67E-09
eug5798	208	hypothetical protein EUGRSUZ_C02042	2.21E-11
eug21093	77	hypothetical protein EUGRSUZ_G02791	1.72E-13
eug21094	78	hypothetical protein EUGRSUZ_G02791	2.16E-14
eug21097	74	hypothetical protein EUGRSUZ_G02791	9.25E-14
eug21098	74	hypothetical protein EUGRSUZ_G02791	8.19E-11
eug21100	75	hypothetical protein EUGRSUZ_G02791	5.31E-09
eug1452	153	hypothetical protein F0562_012874	1.63E-12
eug26956	99	hypothetical protein I3760_15G027200	4.73E-07
eug9494	98	hypothetical protein I3760_15G027200	5.52E-09
eug1106	117	hypothetical protein L484_005913	9.54E-12
eug19553	130	hypothetical protein L484_005913	1.94E-12
eug306	67	hypothetical protein L484_013392	1.35E-09
eug841	84	hypothetical protein RHGRI_028034	2.01E-21
eug8963	134	hypothetical protein RHGRI_028034	1.55E-21
eug21292	118	hypothetical protein SADUNF_Sadunf16G0033700	5.50E-07
eug24172	67	hypothetical protein TorRG33x02_273390	2.93E-06
eug3717	526	hypothetical protein Vadar_033820	2.04E-44
eug3718	212	hypothetical protein Vadar_033820	1.40E-29
eug415	90	increased DNA methylation 3-like	2.04E-31
eug417	127	increased DNA methylation 3-like	1.38E-35
eug18193	78	indole-3-acetic acid-amido synthetase GH3.10	5.10E-09
eug30673	102	indole-3-acetic acid-amido synthetase GH3.10	6.44E-07
eug26847	78	indole-3-acetic acid-amido synthetase GH3.6	5.90E-12

eug5741	78	indole-3-acetic acid-amido synthetase GH3.6	1.04E-11
eug15324	298	inner centromere protein A isoform X2	3.48E-13
eug23815	297	inner centromere protein A isoform X2	4.15E-13
eug20201	281	inorganic pyrophosphatase 1	0
eug20202	276	inorganic pyrophosphatase 1	0
eug8967	292	inorganic pyrophosphatase 2-like	8.52E-172
eug19640	95	inositol hexakisphosphate anddiphosphoinositol-pentakisphosphate kinase 1	5.31E-12
eug20534	77	inositol hexakisphosphate anddiphosphoinositol-pentakisphosphate kinase 1	3.49E-18
eug16526	130	integrator complex subunit 3 homolog	2.38E-63
eug16527	167	integrator complex subunit 3 homolog	2.34E-86
eug21202	170	integrator complex subunit 3 homolog	8.86E-95
eug5122	133	integrator complex subunit 3 homolog isoform X1	8.36E-65
eug10015	72	isocitrate dehydrogenase [NAD] regulatory subunit 1, mitochondrial-like	7.51E-09
eug14684	72	isocitrate dehydrogenase [NAD] regulatory subunit 1, mitochondrial-like	5.20E-10
eug3813	166	isoform 2 of histone-lysine n-methyltransferase trx1	7.96E-66
eug24160	117	isopentenyl-diphosphate Delta-isomerase I	4.33E-48
eug28234	117	isopentenyl-diphosphate Delta-isomerase I	2.72E-51
eug17198.2	802	jacalin-related lectin 4-like	4.10E-138
eug23430.1	176	L-ascorbate oxidase-like	2.80E-51
eug26910	104	L-ascorbate oxidase-like	1.83E-48
eug3768	104	L-ascorbate oxidase-like	5.99E-46
eug671	153	L-ascorbate oxidase-like	1.91E-58
eug12789	149	Late embryogenesis abundant	5.48E-34
eug2549	105	Late embryogenesis abundant	2.35E-45
eug8395	96	Late embryogenesis abundant	1.45E-40
eug13909	146	Late embryogenesis abundant protein	4.13E-26
eug10748	268	late embryogenesis abundant protein D-29	9.96E-108
eug3180	264	late embryogenesis abundant protein D-29	2.12E-107
eug12355	287	LEAF RUST 10 DISEASE-RESISTANCE LOCUS RECEPTOR-LIKE PROTEIN KINASE-like 1.2	1.63E-138
eug18319	291	LEAF RUST 10 DISEASE-RESISTANCE LOCUS RECEPTOR-LIKE PROTEIN KINASE-like 1.2	1.45E-134
eug22721.2	218	LEAF RUST 10 DISEASE-RESISTANCE LOCUS RECEPTOR-LIKE PROTEIN KINASE-like 1.2	2.70E-64
eug4601	119	LEAF RUST 10 DISEASE-RESISTANCE LOCUS RECEPTOR-LIKE PROTEIN KINASE-like 1.2	3.77E-41
eug15286	179	lecithin retinol acyltransferase-like	2.32E-83
eug23734	181	lecithin retinol acyltransferase-like	8.40E-80
eug11151	83	lectin-like	9.36E-27
eug5833	137	lectin-like	4.91E-62
eug13615	475	leucine-rich repeat extensin-like protein 3	2.21E-95
eug25927	240	leucine-rich repeat extensin-like protein 3	1.85E-64
eug27711	272	leucine-rich repeat extensin-like protein 3	4.02E-83
eug16828	113	leucine-rich repeat extensin-like protein 4	5.20E-56
eug27630	109	leucine-rich repeat extensin-like protein 4	2.19E-55

eug13852	642	leucine-rich repeat receptor-like tyrosine-protein kinase PXC3	0
eug13853	570	leucine-rich repeat receptor-like tyrosine-protein kinase PXC3	0
eug10093	71	L-galactose dehydrogenase	1.15E-27
eug19368	66	L-galactose dehydrogenase	5.02E-24
eug21602	370	LOC110651873 isoform	0
eug9873	470	LOC110651873 isoform	0
eug11484	219	LOC111013677 isoform	3.59E-62
eug148	248	LOC111013677 isoform	2.06E-108
eug2985	175	LOC111013677 isoform	1.69E-40
eug2986	286	LOC111013677 isoform	4.84E-84
eug29695	142	loganic acid O-methyltransferase-like	9.14E-63
eug28686	243	LRR receptor kinase SERK2-like isoform X3	4.42E-11
eug20291.1	644	LRR receptor-like serine/threonine-protein kinase GSO1	0
eug9590	229	L-type lectin-domain containing receptor kinase S.4-like	1.22E-28
eug26908	496	lysine histidine transporter-like 8	0
eug8944	461	lysine histidine transporter-like 8	0
eug17208	136	lysine-specific demethylase JMJ25	1.54E-69
eug28511	135	lysine-specific demethylase JMJ25-like	1.67E-63
eug3252	159	lysine-specific demethylase JMJ25-like	5.34E-94
eug29223	156	lysine-specific demethylase JMJ25-like isoform X1	1.86E-26
eug5523	143	lysine-specific demethylase JMJ25-like isoform X1	5.95E-36
eug16165	224	lysine-specific demethylase JMJ25-like isoform X2	3.04E-130
eug17207	313	lysine-specific demethylase JMJ25-like isoform X2	0
eug9248	280	lysine-specific demethylase JMJ25-like isoform X2	1.73E-98
eug14475	110	malate dehydrogenase-like	5.09E-09
eug16511	156	MDIS1-interacting receptor like kinase 2-like	1.00E-32
eug18056.2	404	MDIS1-interacting receptor like kinase 2-like	0
eug12809	70	mechanosensitive ion channel protein 6 isoform X1	3.75E-13
eug11777	154	mechanosensitive ion channel protein 6-like	4.48E-22
eug14337	73	mechanosensitive ion channel protein 6-like	2.34E-08
eug461	73	mechanosensitive ion channel protein 6-like	1.89E-08
eug1987	138	mediator of RNA polymerase II transcription subunit 7a-like	6.85E-28
eug8346	121	mediator of RNA polymerase II transcription subunit 7a-like	5.93E-42
eug13627	93	MFS transporter	2.68E-54
eug22136	93	MFS transporter	1.09E-53
eug23114	82	microtubule-associated protein RP/EB family member 1C	4.67E-12
eug23116	122	microtubule-associated protein RP/EB family member 1C	3.75E-15
eug11115	162	microtubule-associated protein TORTIFOLIA1	2.55E-18
eug19010	127	microtubule-associated protein TORTIFOLIA1	6.80E-22
eug15907	239	mitochondrial glycoprotein family protein	8.12E-122
eug25249	248	mitochondrial glycoprotein family protein	3.38E-128
eug3754	248	mitochondrial glycoprotein family protein	3.20E-121
eug25016.2	218	mitochondrial uncoupling protein 5	9.85E-111
eug2726	244	mitochondrial uncoupling protein 5	1.87E-128
eug9263	233	mitochondrial uncoupling protein 5	1.93E-114

eug21790.2	439	mitogen-activated protein kinase kinase 5	7.56E-141
eug10513	346	mitogen-activated protein kinase kinase 5-like	0
eug2486	513	MLO-like protein 15	0
eug2671	508	MLO-like protein 15	0
eug27615	509	MLO-like protein 15	0
eug16286	561	MLO-like protein 2	0
eug6876	538	MLO-like protein 6	0
eug25168.2	248	monooxygenase 2-like	2.01E-161
eug25169	259	monooxygenase 2-like	2.01E-158
eug20393	168	monooxygenase 2-like isoform X2	5.67E-105
eug23288	175	monooxygenase 2-like isoform X2	4.75E-110
eug14379	243	monoterpene synthase-like	1.75E-149
eug15514	258	monoterpene synthase-like	6.38E-148
eug19563	80	multisubstrate pseudouridine synthase 7	9.86E-22
eug27162	174	multisubstrate pseudouridine synthase 7	6.78E-17
eug5594	78	multisubstrate pseudouridine synthase 7	5.01E-18
eug19049	237	Myb_DNA-bind_6 domain-containing protein	1.52E-62
eug856	196	Myb_DNA-bind_6 domain-containing protein	3.77E-68
eug25553	312	Myosin-3 like	3.29E-153
eug17432	147	NAC domain-containing protein 2-like	1.24E-80
eug18706	283	NAC domain-containing protein 68-like	1.16E-147
eug23552	300	NAC domain-containing protein 68-like	6.70E-160
eug9362	300	NAC domain-containing protein 68-like	9.59E-159
eug9289	814	NAC domain-containing protein 91-like isoform X1	3.05E-29
eug12098	120	NADPH-dependent pterin aldehyde reductase-like	1.91E-34
eug21313	138	NADPH-dependent pterin aldehyde reductase-like	8.89E-37
eug12922	262	NDR1/HIN1-like protein 10	1.41E-84
eug12923	247	NDR1/HIN1-like protein 10	1.43E-84
eug18903	215	NDR1/HIN1-like protein 10	3.60E-50
eug3336	262	NDR1/HIN1-like protein 10	4.58E-81
eug18904	273	NDR1/HIN1-like protein 12	2.05E-49
eug24914	114	neutral ceramidase 2-like	5.75E-57
eug27300	167	neutral ceramidase 2-like	1.15E-56
eug3829	100	neutral ceramidase 2-like	3.46E-54
eug25784	92	nitrate regulatory gene2 protein-like	1.09E-32
eug27776	91	nitrate regulatory gene2 protein-like	3.60E-34
eug6414	77	nitrate regulatory gene2 protein-like	3.30E-21
eug23505.2	369	non-structural maintenance of chromosomes element 4 homolog A-like	0
eug2465.2	367	non-structural maintenance of chromosomes element 4 homolog A-like	0
eug29814.1	461	non-structural maintenance of chromosomes element 4 homolog A-like	0
eug26811	144	Nop14, putative isoform 1	3.12E-18
eug30674	122	Nop14, putative isoform 1	7.28E-21
eug26709	298	N-terminal acetyltransferase A complex catalytic subunit NAA10	1.53E-27
eug5445	120	N-terminal acetyltransferase A complex catalytic subunit NAA10-like	8.56E-06
eug591	218	N-terminal acetyltransferase A complex catalytic subunit NAA10-like	1.79E-12

eug14451	384	Nuclear factor related to kappa-B-binding protein	7.03E-147
eug15477.1	506	nuclear factor related to kappa-B-binding protein	5.17E-140
eug29681	145	nuclear transport factor 2B	9.59E-49
eug6644	130	nuclear transport factor 2B	5.88E-50
eug18581	394	nucleolar protein 58-like	6.29E-97
eug6660	182	nucleolar protein 58-like	4.72E-17
eug20434	126	Nucleoredoxin 3	7.42E-33
eug4221	352	O-acyltransferase WSD1-like	7.42E-139
eug4222	417	O-acyltransferase WSD1-like	0
eug11072	168	Octicosapeptide/Phox/Bem1p family protein	1.22E-43
eug869	261	O-fucosyltransferase 20	2.42E-17
eug16795	439	oligopeptide transporter 1-like	0
eug2782	439	oligopeptide transporter 1-like	0
eug2783	426	oligopeptide transporter 1-like	2.78E-180
eug15422	727	oligopeptide transporter 5-like	0
eug27986	760	oligopeptide transporter 5-like	0
eug13275	97	omega-hydroxypalmitate O-feruloyl transferase	1.18E-24
eug15413	123	omega-hydroxypalmitate O-feruloyl transferase	1.06E-27
eug19578	117	organelle RRM domain-containing protein 2, mitochondrial	1.53E-55
eug803	117	organelle RRM domain-containing protein 2, mitochondrial	2.19E-55
eug20914	171	organic cation/carnitine transporter 4	9.68E-86
eug21678	180	organic cation/carnitine transporter 4	3.52E-81
eug8474	183	palmitoyl-acyl carrier protein thioesterase, chloroplastic	1.92E-59
eug26864	427	palmitoyl-acyl carrier protein thioesterase, chloroplastic-like	0
eug7717	243	palmitoyl-acyl carrier protein thioesterase, chloroplastic-like	1.45E-119
eug8475	180	palmitoyl-acyl carrier protein thioesterase, chloroplastic-like	1.23E-97
eug15390	186	PAR1 protein	1.53E-109
eug19300	186	PAR1 protein	1.91E-110
eug23659	182	PAR1 protein	4.03E-103
eug29869	149	pathogenesis-related protein 4	1.29E-12
eug14484	103	pectate lyase-like	1.04E-63
eug26523	103	pectate lyase-like	1.04E-63
eug26875	515	pectinesterase 2-like	0
eug26876	515	pectinesterase 2-like	0
eug4215	209	pentatricopeptide repeat-containing protein	3.59E-60
eug11663	91	pentatricopeptide repeat-containing protein At2g15690, mitochondrial-like	5.50E-23
eug12990	125	pentatricopeptide repeat-containing protein At2g15690, mitochondrial-like	7.57E-13
eug16074	112	pentatricopeptide repeat-containing protein At2g15690, mitochondrial-like	8.21E-35
eug2367	122	pentatricopeptide repeat-containing protein At2g15690, mitochondrial-like	6.15E-34
eug14007	174	pentatricopeptide repeat-containing protein At2g26790, mitochondrial	1.32E-18
eug21370	97	pentatricopeptide repeat-containing protein At2g26790, mitochondrial	7.66E-17
eug18440	202	pentatricopeptide repeat-containing protein At2g33760-like	1.63E-134
eug21557	206	pentatricopeptide repeat-containing protein At2g33760-like	2.17E-117

eug16182	118	pentatricopeptide repeat-containing protein At3g25210, mitochondrial	1.66E-28
eug21323	155	pentatricopeptide repeat-containing protein At3g25210, mitochondrial	1.37E-27
eug15357	127	pentatricopeptide repeat-containing protein At3g54980, mitochondrial-like	2.21E-43
eug16141	197	pentatricopeptide repeat-containing protein At3g54980, mitochondrial-like	5.15E-83
eug50	153	pentatricopeptide repeat-containing protein At3g61360	1.98E-64
eug680	174	pentatricopeptide repeat-containing protein At3g61360	7.83E-59
eug7089	192	pentatricopeptide repeat-containing protein At3g61360-like	1.18E-52
eug5929	165	pentatricopeptide repeat-containing protein At4g02750-like	7.13E-81
eug7377	216	pentatricopeptide repeat-containing protein At4g02750-like	2.84E-114
eug12444	158	pentatricopeptide repeat-containing protein At4g32430, mitochondrial	1.16E-96
eug27901	158	pentatricopeptide repeat-containing protein At4g32430, mitochondrial	2.38E-97
eug21590.1	374	pentatricopeptide repeat-containing protein At5g08305	0
eug3755.2	236	pentatricopeptide repeat-containing protein At5g08305 isoform X1	1.28E-153
eug13092	189	pentatricopeptide repeat-containing protein At5g66520-like isoform X1	3.19E-67
eug27434	141	pentatricopeptide repeat-containing protein At5g66520-like isoform X1	1.73E-47
eug1243	406	peptidyl-prolyl cis-trans isomerase FKBP43-like	4.75E-179
eug13611	439	peptidyl-prolyl cis-trans isomerase FKBP43-like	0
eug13717	322	peroxidase 4-like	0
eug13718	322	peroxidase 4-like	0
eug16719	322	peroxidase 4-like	0
eug15046	69	peroxisomal and mitochondrial division factor 2	4.79E-16
eug24583	276	peroxisomal and mitochondrial division factor 2	2.05E-135
eug20621	271	peroxisomal and mitochondrial division factor 2-like	7.85E-101
eug10689	109	PHD finger protein MALE MEIOCYTE DEATH 1	1.17E-57
eug23254	109	PHD finger protein MALE MEIOCYTE DEATH 1	2.83E-57
eug499	109	PHD finger protein MALE MEIOCYTE DEATH 1	2.05E-55
eug1665	283	phosphatidate phosphatase PAH2-like isoform X1	0
eug21337	272	phosphatidate phosphatase PAH2-like isoform X1	9.07E-141
eug22836	109	phosphoinositide phospholipase C 2	2.93E-55
eug24086	72	phosphoinositide phospholipase C 2	3.37E-33
eug24264	109	phosphoinositide phospholipase C 2	8.80E-55
eug25084	109	phosphoinositide phospholipase C 2	8.07E-60
eug23275	109	phosphoinositide phospholipase C 2-like	5.70E-56
eug5936	69	phosphoinositide phospholipase C 2-like	2.49E-30
eug24896	156	phospholipase A and acyltransferase 3-like	2.50E-16
eug10142	107	phosphoribosylformylglycinamide cyclo-ligase, chloroplastic-like	7.90E-55
eug15321	107	phosphoribosylformylglycinamide cyclo-ligase, chloroplastic-like	4.08E-53
eug5933	117	phosphoribosylformylglycinamide cyclo-ligase, chloroplastic-like	5.44E-61
eug10795	149	phytosulfokine receptor 1	1.03E-27
eug16956	131	phytosulfokine receptor 1-like	1.98E-30
eug29940	355	phytosulfokine receptor 1-like	9.64E-82
eug3847	175	phytosulfokine receptor 1-like	7.09E-76
eug462	188	phytosulfokine receptor 1-like	4.64E-75
eug24525	151	PII uridylyltransferase	1.66E-55

eug15059	489	Plant basic secretory protein (BSP) family protein	1.08E-114
eug21747	166	Plastid transcriptionally active 15	2.50E-24
eug24665	148	Plastid transcriptionally active 15	2.67E-44
eug6196	78	PLATZ transcription factor family protein	3.47E-29
eug13966	322	pollen-specific leucine-rich repeat extensin-like protein 3	0
eug1591	169	pollen-specific leucine-rich repeat extensin-like protein 3	1.00E-81
eug163	203	pollen-specific leucine-rich repeat extensin-like protein 3	7.37E-78
eug28261	562	pollen-specific leucine-rich repeat extensin-like protein 3	0
eug3457	241	pollen-specific leucine-rich repeat extensin-like protein 3	1.85E-101
eug17351	160	Polyadenylate-binding protein	4.10E-95
eug10406	338	polyadenylate-binding protein 1-B-binding protein	2.78E-169
eug9814	340	polyadenylate-binding protein 1-B-binding protein	1.26E-164
eug9815	328	polyadenylate-binding protein 1-B-binding protein	1.48E-172
eug25208	353	polygalacturonase-1 non-catalytic subunit beta-like	0
eug8305	657	polygalacturonase-1 non-catalytic subunit beta-like	5.28E-143
eug9806	240	polyol transporter 5-like	1.06E-09
eug23767	209	polyubiquitin-B-like isoform X2	1.75E-51
eug2840	109	polyubiquitin-like isoform X1	5.71E-37
eug17342	95	precursor of CEP14	1.21E-33
eug22548	101	precursor of CEP14	4.47E-56
eug18140	197	precursor of CEP9-like	8.84E-63
eug30355	154	precursor of CEP9-like	7.75E-67
eug4331	128	PREDICTED: uncharacterized protein LOC108212167	1.96E-10
eug10400	560	primary amine oxidase	0
eug18986	126	primary amine oxidase	4.49E-54
eug18992	217	primary amine oxidase	8.73E-141
eug24550	217	primary amine oxidase	1.06E-147
eug4016	126	primary amine oxidase	4.37E-58
eug18816	242	primary amine oxidase-like	1.25E-54
eug18989	593	primary amine oxidase-like	0
eug3949	183	primary amine oxidase-like	1.28E-45
eug14018	314	probable 28S rRNA (cytosine-C(5))-methyltransferase isoform X1	0
eug19684	314	probable 28S rRNA (cytosine-C(5))-methyltransferase isoform X1	0
eug13494	89	probable 2-oxoglutarate-dependent dioxygenase AOP1	8.69E-17
eug25163	101	probable 2-oxoglutarate-dependent dioxygenase AOP1	6.29E-21
eug25092	130	probable 3-hydroxyisobutyrate dehydrogenase-like 1, mitochondrial	1.52E-12
eug16642	189	probable aldo-keto reductase 2	9.20E-81
eug2253	160	probable aldo-keto reductase 2	3.38E-93
eug8198	171	probable aldo-keto reductase 2	8.91E-89
eug10793	363	probable caffeine synthase 2	0
eug10794	361	probable caffeine synthase 2	0
eug3136.1	341	probable caffeine synthase 2	0
eug15049	117	probable cytosolic oligopeptidase A	3.23E-11
eug19485	98	probable cytosolic oligopeptidase A	2.59E-22
eug18161.2	122	probable disease resistance protein At4g27220 isoform X1	1.49E-49

eug20647	164	probable disease resistance protein At4g27220 isoform X1	6.49E-74
eug2325	334	probable disease resistance protein At4g27220 isoform X1	1.93E-165
eug17729	175	probable disease resistance protein At5g66900	4.40E-34
eug19222	124	probable disease resistance protein At5g66900	3.27E-27
eug20021	193	probable disease resistance protein At5g66900	5.63E-77
eug20650	153	probable disease resistance protein At5g66900	6.53E-25
eug2337	149	probable disease resistance protein At5g66900	4.53E-60
eug3921	154	probable disease resistance protein At5g66900	1.25E-27
eug6093	156	probable disease resistance protein At5g66900	1.62E-31
eug11405	130	probable glycosyltransferase At5g03795	1.84E-48
eug25795	128	probable glycosyltransferase At5g03795 isoform X2	6.09E-49
eug12687	83	probable H/ACA ribonucleoprotein complex subunit 1	2.57E-11
eug12761	83	probable H/ACA ribonucleoprotein complex subunit 1	1.30E-11
eug27425	98	probable H/ACA ribonucleoprotein complex subunit 1	2.04E-15
eug12920	119	probable inactive poly [ADP-ribose] polymerase SRO5	2.94E-40
eug15705	223	probable inactive poly [ADP-ribose] polymerase SRO5	4.42E-71
eug16076	400	probable inactive receptor kinase At1g48480	8.29E-137
eug18931	453	probable inactive receptor kinase At1g48480	5.02E-135
eug26838	358	probable inactive receptor kinase At1g48480	1.63E-134
eug11989	150	probable inactive receptor kinase At4g23740	4.80E-67
eug18781	249	probable inactive receptor kinase At4g23740	3.83E-125
eug12033	111	probable leucine-rich repeat receptor-like protein kinase At1g35710	2.00E-24
eug5633	90	probable leucine-rich repeat receptor-like protein kinase At1g35710	2.36E-28
eug16025	362	probable LRR receptor-like serine/threonine-protein kinase At1g56130	0
eug18148	316	probable LRR receptor-like serine/threonine-protein kinase At1g56130	0
eug19297.1	369	probable LRR receptor-like serine/threonine-protein kinase At1g56130	0
eug4522	219	probable LRR receptor-like serine/threonine-protein kinase At1g56130	2.86E-128
eug6568	171	probable LRR receptor-like serine/threonine-protein kinase At1g56130	8.36E-79
eug28482	107	probable LRR receptor-like serine/threonine-protein kinase At4g37250	4.42E-14
eug30253	116	probable LRR receptor-like serine/threonine-protein kinase At4g37250	1.81E-17
eug8279	106	probable LRR receptor-like serine/threonine-protein kinase At4g37250	4.32E-12
eug12817.2	198	probable mediator of RNA polymerase II transcription subunit 26b	2.63E-66
eug2108	298	probable mediator of RNA polymerase II transcription subunit 26b	4.07E-108
eug4356	71	probable nitronate monooxygenase	4.93E-14
eug23652	180	probable pectinesterase 29	2.79E-102
eug7785	202	probable pectinesterase 29	2.49E-122
eug15320	73	probable pectinesterase 55	1.82E-24
eug16836	66	probable pectinesterase 55	2.06E-23
eug18655	79	probable pectinesterase/pectinesterase inhibitor 51	9.68E-29
eug27341	106	probable pectinesterase/pectinesterase inhibitor 51	1.04E-49
eug24924	80	probable protein phosphatase 2C 27	1.39E-28
eug766	88	probable protein phosphatase 2C 27	4.74E-34
eug26436	90	probable pyridoxal 5'-phosphate synthase subunit PDX1	8.26E-11
eug9190	83	probable pyridoxal 5'-phosphate synthase subunit PDX1	6.13E-10
eug501	346	probable receptor-like protein kinase At5g15080	2.45E-145



eug11925	111	probable S-adenosylmethionine-dependent methyltransferase At5g38100	1.81E-51
eug13964	105	probable S-adenosylmethionine-dependent methyltransferase At5g38100	2.19E-49
eug13965	118	probable S-adenosylmethionine-dependent methyltransferase At5g38100	3.42E-66
eug29694	105	probable S-adenosylmethionine-dependent methyltransferase At5g38100	2.81E-50
eug30506	206	probable serine/threonine-protein kinase PBL11	2.56E-108
eug4745	144	probable serine/threonine-protein kinase PBL11	5.04E-73
eug4563	83	probable serine/threonine-protein kinase PBL19	2.18E-43
eug9278	83	probable serine/threonine-protein kinase PBL19	1.15E-42
eug25487.1	341	probable serine/threonine-protein kinase PIX13 isoform X2	1.70E-141
eug27466.1	464	probable serine/threonine-protein kinase PIX7	0
eug20307	215	probable serine/threonine-protein kinase yakA	2.47E-78
eug5588	187	probable serine/threonine-protein kinase yakA	2.41E-13
eug5897	347	probable serine/threonine-protein kinase yakA	1.87E-154
eug703	391	probable transcription factor At1g61730	8.62E-130
eug7573	463	probable transcription factor At1g61730	7.61E-120
eug27382	133	probable tRNA N6-adenosine threonylcarbamoyltransferase, mitochondrial	8.26E-28
eug4931	133	probable tRNA N6-adenosine threonylcarbamoyltransferase, mitochondrial isoform X1	5.03E-28
eug17635	79	probable UDP-N-acetylglucosamine--peptide N-acetylglucosaminyltransferase SPINDLY isoform X1	1.76E-41
eug27527	79	probable UDP-N-acetylglucosamine--peptide N-acetylglucosaminyltransferase SPINDLY isoform X1	6.98E-41
eug5606	377	probable WRKY transcription factor 30	1.99E-72
eug21669	244	probable WRKY transcription factor 31 isoform X1	2.38E-106
eug3334	279	probable WRKY transcription factor 31 isoform X2	2.44E-127
eug6880	96	Prolactin receptor like	8.44E-22
eug29028	95	protein ABCI7, chloroplastic	7.11E-35
eug4386	136	protein ABCI7, chloroplastic	1.07E-38
eug16904	124	protein ACCELERATED CELL DEATH 6 isoform X1	8.31E-49
eug28311	123	protein ACCELERATED CELL DEATH 6 isoform X1	1.31E-39
eug1199	157	protein AGENET DOMAIN (AGD)-CONTAINING P1	3.38E-78
eug21424	161	protein AGENET DOMAIN (AGD)-CONTAINING P1	5.32E-79
eug20755	528	protein ALP1-like	0
eug25174	491	protein ALP1-like	0
eug19402	162	protein ALWAYS EARLY	1.20E-23
eug20516	141	protein ALWAYS EARLY	1.12E-23
eug12726	159	protein AMN1 homolog	5.69E-12
eug20452	534	protein CHUP1, chloroplastic-like	0
eug22178	266	protein CHUP1, chloroplastic-like	1.03E-109
eug13453	489	protein DA1-related 1	0
eug21780	563	protein DA1-related 1	0
eug19143	563	protein DA1-related 1-like	0
eug19924	110	protein DETOXIFICATION 53-like	1.03E-47
eug28108	116	protein DETOXIFICATION 53-like	4.36E-22

eug29382	89	protein DETOXIFICATION 53-like	6.29E-25
eug17533	77	protein ENHANCED DOWNY MILDEW 2 isoform X1	2.23E-13
eug22480	86	protein ENHANCED DOWNY MILDEW 2 isoform X1	1.27E-13
eug11038	77	protein ENHANCED DOWNY MILDEW 2-like	1.42E-15
eug13280	172	protein ENHANCED DOWNY MILDEW 2-like	4.70E-70
eug21815	460	protein ENHANCED DOWNY MILDEW 2-like	0
eug5172	573	protein ENHANCED DOWNY MILDEW 2-like	0
eug11040	154	protein ENHANCED DOWNY MILDEW 2-like isoform X2	1.51E-45
eug18273	82	protein FLOURY 1-like	6.17E-18
eug22212	98	protein FLOURY 1-like isoform X2	7.55E-14
eug7128	80	protein FLOURY 1-like isoform X2	6.58E-11
eug15326	81	protein kinase PINOID	3.58E-17
eug3241	78	protein kinase PINOID	2.17E-17
eug19161	94	protein LIKE COV 2-like	4.93E-15
eug25202	91	protein LIKE COV 2-like	4.27E-16
eug26710	91	protein LIKE COV 2-like	5.05E-16
eug13139	198	protein LURP-one-related 10-like	1.41E-117
eug13135	220	protein LURP-one-related 15-like	3.21E-116
eug11829	103	protein MID1-COMPLEMENTING ACTIVITY 1-like	2.34E-26
eug4148	106	protein MID1-COMPLEMENTING ACTIVITY 1-like	7.79E-44
eug14843	370	protein NEN1-like	1.90E-174
eug2203	370	protein NEN1-like	2.88E-170
eug150	251	protein NETWORKED 2D	9.13E-76
eug151	188	protein NETWORKED 2D	1.19E-78
eug5951	239	protein NETWORKED 2D	1.13E-71
eug22411	91	protein N-lysine methyltransferase METTL21A	2.24E-23
eug28272	97	protein N-lysine methyltransferase METTL21A	7.64E-22
eug10628	168	protein NTM1-like 9	5.57E-79
eug11185	209	protein NTM1-like 9	3.32E-36
eug11187.1	467	protein NTM1-like 9	5.79E-43
eug16323	248	protein NTM1-like 9	1.89E-136
eug17819	220	protein NTM1-like 9	1.38E-55
eug2468	248	protein NTM1-like 9	3.19E-135
eug26496	206	protein NTM1-like 9	3.54E-50
eug3521	327	protein NTM1-like 9	0
eug9284	295	protein NTM1-like 9	4.06E-171
eug9285	279	protein NTM1-like 9	1.13E-179
eug9287	199	protein NTM1-like 9	4.36E-93
eug9473	327	protein NTM1-like 9	0
eug11953	151	protein NUCLEAR FUSION DEFECTIVE 4	1.18E-69
eug6379	151	protein NUCLEAR FUSION DEFECTIVE 4	2.58E-70
eug7220.2	168	protein NYNRIN-like	4.53E-08
eug16435	123	protein OCTOPUS-like	9.77E-38
eug3359	113	protein OCTOPUS-like	5.80E-30
eug15303	228	protein PHLOEM PROTEIN 2-LIKE A10-like	1.95E-91

eug22856	198	protein PHLOEM PROTEIN 2-LIKE A10-like	1.25E-88
eug16075	91	protein PHYTOCHROME KINASE SUBSTRATE 3-like	1.76E-44
eug26184	91	protein PHYTOCHROME KINASE SUBSTRATE 3-like	3.98E-42
eug12586	128	protein PLASTID MOVEMENT IMPAIRED 2	1.54E-54
eug1843	128	protein PLASTID MOVEMENT IMPAIRED 2	1.42E-55
eug22617	376	protein prune-like protein	4.92E-78
eug27500	327	protein prune-like protein	1.74E-76
eug17644	136	protein RMD5	2.36E-74
eug19189	136	protein RMD5	1.92E-75
eug14478	401	protein SHORTAGE IN CHIASMATA 1	1.75E-108
eug3950	467	protein SHORTAGE IN CHIASMATA 1	0
eug13882	635	protein SIEVE ELEMENT OCCLUSION B-like	0
eug14234	620	protein SIEVE ELEMENT OCCLUSION B-like	0
eug17653	464	protein SIEVE ELEMENT OCCLUSION B-like	9.09E-118
eug23733.2	492	protein SIEVE ELEMENT OCCLUSION B-like	3.21E-138
eug16184	236	protein spotted leaf 11-like	1.03E-125
eug10018	212	protein SRC2 homolog	5.77E-35
eug7416.1	164	protein SRC2 homolog	2.46E-28
eug15446	178	protein SRC2-like	1.18E-20
eug22785	140	protein SRC2-like	1.95E-21
eug25673	315	protein SRC2-like	1.27E-45
eug26583	98	protein SRC2-like	2.49E-16
eug26889	121	protein SRC2-like	1.33E-14
eug30155	214	protein SRC2-like	1.51E-26
eug4311	203	protein SRC2-like	4.47E-29
eug8057	121	protein SRC2-like	2.35E-16
eug22533	77	protein SRG1-like	5.45E-25
eug26865	107	protein SRG1-like	4.10E-26
eug5611	107	protein SRG1-like	9.02E-28
eug1018	245	protein STICHEL-like 4	4.96E-104
eug11735	505	protein STICHEL-like 4	0
eug4290	569	protein STICHEL-like 4	0
eug5143	152	protein tesmin/TSO1-like CXC 4 isoform X1	3.38E-63
eug9674	152	protein tesmin/TSO1-like CXC 4 isoform X1	3.93E-63
eug16387	153	protein TIFY 10b-like	6.73E-75
eug20301	109	protein TIFY 10b-like	1.79E-30
eug20302	118	protein TIFY 10b-like	1.54E-62
eug23028	117	protein TIFY 10b-like	1.45E-62
eug20976	173	protein trichome birefringence-like 19	1.93E-102
eug20980	280	protein trichome birefringence-like 19	3.59E-112
eug22561	113	protein ULTRAPETALA 1-like	1.29E-53
eug22560	113	protein ULTRAPETALA 2-like	2.17E-56
eug17450	96	proteoglycan 4-like	2.97E-34
eug17452	98	proteoglycan 4-like	1.77E-19
eug17453	82	proteoglycan 4-like	3.27E-10

eug26681	93	protodermal factor	2.46E-49
eug27499	93	protodermal factor	1.82E-46
eug5148	110	protodermal factor	3.15E-53
eug6018	98	protodermal factor	2.19E-55
eug14244	94	protoheme IX farnesyltransferase, mitochondrial	9.63E-53
eug22962	94	protoheme IX farnesyltransferase, mitochondrial	3.13E-51
eug20626	91	putative 30S ribosomal protein S2	4.96E-34
eug4552	445	putative ankyrin repeat protein RF_0381	0
eug4554.1	427	putative ankyrin repeat protein RF_0381	0
eug1493	110	putative B3 domain-containing protein Os03g0621600	6.44E-34
eug19214	213	putative B3 domain-containing protein Os03g0621600	4.19E-118
eug11237.1	173	putative BPI/LBP family protein At1g04970	2.42E-26
eug3116	103	putative BPI/LBP family protein At1g04970	1.13E-26
eug14876	100	putative calcium-binding protein CML10	2.75E-44
eug20759	100	putative calcium-binding protein CML10	1.06E-46
eug21338	100	putative calcium-binding protein CML10	2.26E-53
eug26324	95	putative calcium-binding protein CML10	4.11E-52
eug26325	138	putative calcium-binding protein CML10	1.87E-30
eug5579	109	putative calcium-binding protein CML10	2.31E-57
eug23549	638	putative calcium-transporting ATPase 13, plasma membrane-type	0
eug23550	761	putative calcium-transporting ATPase 13, plasma membrane-type	0
eug19729	81	putative cysteine-rich receptor-like protein kinase 35	1.99E-06
eug6103	81	putative cysteine-rich receptor-like protein kinase 35	1.85E-07
eug15840	108	putative DEAD-box ATP-dependent RNA helicase 1-like isoform X1	5.81E-09
eug20730	94	putative ETHYLENE INSENSITIVE 3-like 4 protein	5.97E-27
eug26175	91	putative ETHYLENE INSENSITIVE 3-like 4 protein	7.66E-26
eug17518	98	putative F-box protein PP2-B12 isoform X1	8.74E-41
eug10333	185	putative GAST1 protein precursor	1.70E-51
eug21506	164	putative GAST1 protein precursor	4.04E-55
eug26803	177	putative invertase inhibitor	6.69E-89
eug30590	106	putative Late embryogenesis abundant protein, LEA-14	4.49E-26
eug15959	87	putative nitronate monooxygenase	2.79E-19
eug18279.1	238	putative pentatricopeptide repeat-containing protein At1g19290	3.21E-66
eug12157	132	putative pentatricopeptide repeat-containing protein At3g23330	4.45E-36
eug12466	194	putative pentatricopeptide repeat-containing protein At5g08310, mitochondrial	1.37E-118
eug5782	173	putative pentatricopeptide repeat-containing protein At5g08310, mitochondrial	4.51E-98
eug9267	192	putative pentatricopeptide repeat-containing protein At5g08310, mitochondrial	1.27E-110
eug18850	89	putative P-loop containing nucleoside triphosphate hydrolase, leucine-rich repeat domain, L	3.07E-14
eug13564	567	putative receptor-like protein kinase At4g00960	0
eug19220	560	putative receptor-like protein kinase At4g00960	0
eug26751	138	putative ribosome biogenesis protein slx9-like	1.31E-44
eug7665	237	putative ribosome biogenesis protein slx9-like	3.02E-58
eug28497	106	putative RING-H2 finger protein ATL21B	7.14E-29

eug17611	199	putative serine/threonine-protein kinase roco5	1.67E-99
eug11710	304	putative two-component response regulator-like APRR6	2.74E-109
eug19074	489	putative two-component response regulator-like APRR6	2.15E-98
eug19075	114	putative two-component response regulator-like APRR6	2.73E-18
eug26652	512	putative WRKY transcription factor 19 -like protein	4.48E-121
eug25724.1	757	PYK10-binding protein 2-like	7.73E-136
eug3162.1	754	PYK10-binding protein 2-like	1.03E-134
eug366.1	911	PYK10-binding protein 2-like	3.04E-117
eug3163.2	680	pyrophosphate-energized vacuolar membrane proton pump	0
eug365	652	pyrophosphate-energized vacuolar membrane proton pump	0
eug17199.1	586	pyrophosphate-energized vacuolar membrane proton pump-like	4.28E-170
eug25725.1	655	pyrophosphate-energized vacuolar membrane proton pump-like	0
eug27627	91	RAB proteins geranylgeranyltransferase component	8.90E-09
eug13284	156	receptor-like protein kinase 5	6.09E-95
eug29713	156	receptor-like protein kinase 5	1.38E-91
eug15433	96	receptor-like protein kinase FERONIA	3.53E-53
eug20254	96	receptor-like protein kinase FERONIA	3.07E-53
eug17696	123	receptor-like protein kinase HSL1	1.26E-48
eug21291	228	receptor-like protein kinase HSL1	1.20E-106
eug23805	228	receptor-like protein kinase HSL1	2.72E-110
eug1453	228	receptor-like serine/threonine-protein kinase SD1-8	5.61E-103
eug7575	428	receptor-like serine/threonine-protein kinase SD1-8	0
eug9371	366	receptor-like serine/threonine-protein kinase SD1-8	1.45E-142
eug962	478	receptor-like serine/threonine-protein kinase SD1-8	0
eug15908	430	receptor-like serine/threonine-protein kinase SD1-8 isoform X1	0
eug26321	195	regulatory-associated protein of TOR 1	5.12E-14
eug27063	220	Retrovirus-related Pol polyprotein from transposon RE1	2.40E-42
eug28024	84	Retrovirus-related Pol polyprotein from transposon RE1	1.33E-20
eug786	117	retrovirus-related pol polyprotein from transposon RE1	4.40E-50
eug20146	97	Retrovirus-related Pol polyprotein from transposon TNT 1-94	1.18E-12
eug22049	217	retrovirus-related Pol polyprotein from transposon TNT 1-94	8.94E-33
eug25118	107	retrovirus-related Pol polyprotein from transposon TNT 1-94	1.24E-12
eug6525	120	retrovirus-related Pol polyprotein from transposon TNT 1-94	1.98E-17
eug29168.2	527	reverse transcriptase domain-containing protein	1.66E-90
eug17810	121	rho GTPase-activating protein gacO-like	6.90E-30
eug23205	170	rho GTPase-activating protein gacO-like	3.06E-30
eug3566	234	rho GTPase-activating protein gacO-like	3.64E-42
eug18062	190	rho GTPase-activating protein gacV-like	7.11E-47
eug30439	167	rho GTPase-activating protein gacV-like	5.20E-46
eug7110	169	rho GTPase-activating protein gacV-like	2.33E-45
eug23823	183	ribosomal protein S6e	8.83E-21
eug25260	127	ribosome biogenesis protein NSA2 homolog	3.59E-12
eug7776	75	ribosome biogenesis protein NSA2 homolog	1.03E-22
eug11790	398	RING finger protein B-like	0
eug11791	395	RING finger protein B-like	0

eug18096	393	RING finger protein B-like	0
eug19498	400	RING finger protein B-like	0
eug21816	394	RING finger protein B-like	0
eug18095	75	RING finger protein B-like isoform X3	5.18E-34
eug20417	75	RING finger protein B-like isoform X3	7.86E-35
eug26510	110	RING/U-box superfamily protein	8.72E-24
eug22803	82	RNA polymerase II C-terminal domain phosphatase-like 2 isoform X3	1.86E-31
eug3675	82	RNA polymerase II C-terminal domain phosphatase-like 2 isoform X3	3.38E-32
eug11589	93	RNA polymerase II C-terminal domain phosphatase-like 4	2.74E-43
eug14977.2	155	RNA polymerase II C-terminal domain phosphatase-like 4	5.98E-49
eug16561	76	RNA polymerase II C-terminal domain phosphatase-like 4	6.50E-30
eug18732	101	RNA polymerase II C-terminal domain phosphatase-like 4	7.02E-37
eug3713	324	RNA polymerase II C-terminal domain phosphatase-like 4	1.72E-167
eug3715	275	RNA polymerase II C-terminal domain phosphatase-like 4	6.59E-131
eug14515	132	RNA-binding protein 33-like	7.06E-63
eug27073	131	RNA-binding protein 33-like	8.25E-66
eug12856	111	RNA-dependent RNA polymerase 2	4.30E-35
eug20729	142	RNA-dependent RNA polymerase 2	6.88E-32
eug25270	277	RNA-dependent RNA polymerase 2	5.05E-29
eug2535	80	rop guanine nucleotide exchange factor 7-like	3.06E-37
eug3104	80	rop guanine nucleotide exchange factor 7-like	1.16E-37
eug4728	160	RPM1-interacting protein 4	5.40E-28
eug14702	137	RPM1-interacting protein 4-like	4.83E-50
eug15329	307	RPM1-interacting protein 4-like	9.99E-133
eug22464	80	rRNA biogenesis protein RRP5	1.23E-20
eug6451	80	rRNA biogenesis protein RRP5	8.95E-21
eug6723	137	rubisco accumulation factor 1.1, chloroplastic-like	2.76E-40
eug6724	79	rubisco accumulation factor 1.1, chloroplastic-like	3.97E-37
eug12364	261	rust resistance kinase Lr10-like	1.71E-90
eug16122	143	rust resistance kinase Lr10-like	6.76E-70
eug20822	147	rust resistance kinase Lr10-like	1.64E-61
eug5043	245	rust resistance kinase Lr10-like	7.95E-93
eug14061	216	scarecrow-like protein 21	8.10E-118
eug27379	216	scarecrow-like protein 21	1.73E-114
eug22100	277	secoisolariciresinol dehydrogenase	1.37E-158
eug9954	277	secoisolariciresinol dehydrogenase	1.17E-167
eug12255	107	secoisolariciresinol dehydrogenase-like	4.91E-40
eug12860	133	secoisolariciresinol dehydrogenase-like	1.19E-39
eug24866	131	self-incompatibility protein S1-like	3.67E-66
eug25617	94	self-incompatibility protein S1-like	6.15E-35
eug24843	524	senescence-associated carboxylesterase 101-like isoform X2	0
eug25167	649	senescence-associated carboxylesterase 101-like isoform X2	0
eug10007	84	serine carboxypeptidase 24-like	2.11E-17
eug26171	185	serine carboxypeptidase II-2	2.26E-111
eug28609	167	serine carboxypeptidase II-2	2.55E-110

eug9764.1	185	serine carboxypeptidase II-2	1.11E-109
eug23217	289	Serine protease inhibitor (SERPIN) family protein	1.25E-158
eug25356	280	Serine protease inhibitor (SERPIN) family protein	9.61E-156
eug26863	153	serine/arginine repetitive matrix protein 1	3.64E-62
eug3428	153	serine/arginine repetitive matrix protein 1	9.31E-57
eug19410	168	serine/arginine-rich splicing factor RSZ21 isoform X1	1.11E-16
eug28825	204	Serine/threonine protein kinase	1.42E-33
eug23052	114	serine/threonine-protein kinase SRPK	5.64E-25
eug22868	582	serine/threonine-protein kinase/endoribonuclease IRE1a-like isoform X2	3.39E-158
eug15306	459	serine/threonine-protein kinase/endoribonuclease IRE1b	1.81E-142
eug22735	227	serine/threonine-protein kinase/endoribonuclease IRE1b-like	1.34E-97
eug169	116	serine-rich adhesin for platelets	1.47E-50
eug23432	116	serine-rich adhesin for platelets	3.07E-53
eug19316	381	SIP1 domain-containing protein	0
eug29516	407	SIP1 domain-containing protein	0
eug14962	115	SKP1-like protein 11	1.07E-39
eug26343	150	SKP1-like protein 11	6.05E-72
eug19989	170	SKP1-like protein 1B	4.78E-33
eug24338	203	small heat shock protein, chloroplastic-like isoform X2	2.62E-65
eug24339.1	200	small heat shock protein, chloroplastic-like isoform X2	3.32E-65
eug10188	77	small ribosomal subunit biogenesis GTPase RsgA 1, mitochondrial	5.46E-23
eug21318	78	small ribosomal subunit biogenesis GTPase RsgA 1, mitochondrial-like	1.06E-21
eug29235	104	smr (Small MutS Related) domain-containing protein	5.78E-06
eug12200	148	SNF2 domain-containing protein CLASSY 4-like	9.43E-66
eug14213	613	SNF2 domain-containing protein CLASSY 4-like	2.26E-177
eug2916	84	spermidine synthase 1	6.35E-13
eug5548	95	spermidine synthase 1	7.22E-16
eug9015	120	spermidine synthase 1	3.81E-15
eug19482	162	splicing factor YJU2-like	1.98E-24
eug19483	378	splicing factor YJU2-like	1.67E-52
eug19484	202	splicing factor YJU2-like	1.07E-37
eug6017	145	splicing factor YJU2-like	1.50E-29
eug23065	97	S-protein homolog 18-like	6.06E-29
eug23101	135	S-protein homolog 18-like	1.48E-53
eug28758	151	SRSF protein kinase 2-like isoform X1	1.71E-27
eug14104	236	starch synthase	1.71E-39
eug21002	198	starch synthase 3, chloroplastic/amyloplastic isoform X1	4.98E-39
eug12414	191	stearoyl-[acyl-carrier-protein] 9-desaturase, chloroplastic	3.47E-120
eug23036	191	stearoyl-[acyl-carrier-protein] 9-desaturase, chloroplastic	2.70E-122
eug30560	191	stearoyl-[acyl-carrier-protein] 9-desaturase, chloroplastic	4.28E-125
eug13739	151	stilbene synthase	2.76E-18
eug13742	167	stilbene synthase	3.49E-19
eug25385	196	stress protein DDR48	4.98E-54
eug8951	97	stress protein DDR48	3.22E-43

eug11142	169	Subtilisin-like protease	1.14E-31
eug22331.1	191	subtilisin-like protease SBT1.2	6.80E-39
eug24319	127	subtilisin-like protease SBT1.2	9.07E-22
eug980	127	subtilisin-like protease SBT1.2	7.33E-22
eug3094	112	subtilisin-like protease SBT3	4.84E-41
eug8828	135	subtilisin-like protease SBT3	1.97E-44
eug21248	104	subtilisin-like protease SBT4.14	9.52E-56
eug23553	104	subtilisin-like protease SBT4.14	1.27E-55
eug1385	218	subtilisin-like protease SBT4.3	4.61E-89
eug6285	145	subtilisin-like protease SBT5.4	2.81E-57
eug821	128	subtilisin-like protease SBT5.4	1.00E-43
eug20328	77	sugar transport protein 7	7.01E-23
eug28707	109	sugar transport protein 7	4.12E-23
eug18097	191	sugar transport protein 8-like	1.96E-13
eug28113	108	sugar transport protein 8-like	3.38E-20
eug5144	130	synaptotagmin-4-like isoform X2	1.10E-09
eug18669	150	synaptotagmin-5-like	1.18E-09
eug12912	145	TGF-beta-activated kinase 1 and MAP3K7-binding protein 3-like	1.07E-30
eug310	99	TGF-beta-activated kinase 1 and MAP3K7-binding protein 3-like	2.04E-32
eug26317	305	thaumatin-like protein 1b	3.65E-160
eug982	301	thaumatin-like protein 1b	2.94E-156
eug12806	182	threonine dehydratase biosynthetic, chloroplastic	2.65E-78
eug3061	175	threonine dehydratase biosynthetic, chloroplastic	2.30E-73
eug408	154	thylakoid lumenal 19 kDa protein, chloroplastic	1.87E-42
eug8370	113	thylakoid lumenal 19 kDa protein, chloroplastic	3.06E-39
eug16520	119	TMV resistance protein N-like	1.41E-29
eug19002	380	TMV resistance protein N-like	0
eug22501	155	TMV resistance protein N-like	1.35E-31
eug24981	107	TMV resistance protein N-like	8.00E-30
eug26462	380	TMV resistance protein N-like	0
eug3092	176	TMV resistance protein N-like	2.93E-38
eug3117	127	TMV resistance protein N-like	1.27E-26
eug3182	247	TMV resistance protein N-like	4.75E-94
eug340	339	TMV resistance protein N-like	1.64E-131
eug341.1	313	TMV resistance protein N-like	1.26E-116
eug4800	74	TMV resistance protein N-like	6.63E-21
eug6119	138	TMV resistance protein N-like	7.97E-22
eug6768	124	TMV resistance protein N-like	6.95E-19
eug8425	214	TMV resistance protein N-like	9.22E-84
eug22981	174	TMV resistance protein N-like isoform X1	3.32E-14
eug22982	116	TMV resistance protein N-like isoform X1	2.35E-52
eug19790	165	TMV resistance protein N-like isoform X2	2.99E-22
eug8033.2	123	TMV resistance protein N-like isoform X2	2.19E-15
eug10336	114	toll/interleukin-1 receptor-like protein	8.76E-15
eug484	121	toll/interleukin-1 receptor-like protein	2.86E-26



eug8961	250	traB domain-containing protein-like isoform X1	1.10E-76
eug17250	103	transcription factor bHLH140 isoform X2	2.82E-32
eug3088	76	transcription factor bHLH140 isoform X2	2.14E-18
eug11275	114	transcription factor bHLH48	3.66E-46
eug19730	114	transcription factor bHLH48	9.87E-49
eug13113	226	transcription factor bHLH85-like	7.42E-127
eug28601	226	transcription factor bHLH85-like	4.74E-124
eug11918	189	transcription factor MYB113-like	2.69E-77
eug183	180	transcription factor MYB75-like	2.93E-88
eug20744	158	transcription factor MYB98	2.02E-21
eug757	485	transcription factor MYB98	2.64E-125
eug5688	320	transcription factor MYB98-like	6.31E-101
eug9875	115	transcription factor MYB98-like	1.50E-12
eug12406	167	Transcription factor TCP20	9.96E-15
eug14664	84	transcription factor UNE10	2.36E-13
eug5976	182	transcription factor UNE10	2.94E-12
eug7051	269	transcription factor VIP1-like	1.01E-101
eug8931	253	transcription factor VIP1-like	8.34E-98
eug8367	179	transcriptional repressor ILP1	3.80E-13
eug15213	296	translation initiation factor IF-2	4.76E-68
eug20029	541	translation initiation factor IF-2	7.31E-108
eug20208	132	Transmembrane 9 superfamily member 7	1.42E-13
eug13771	162	transmembrane 9 superfamily member 8-like	1.74E-12
eug16512	177	Transmembrane protein	3.54E-53
eug30126	178	Transmembrane protein	2.65E-51
eug4697	125	transposon Tf2-1 polyprotein	4.10E-47
eug6540	121	Transposon Ty3-G Gag-Pol polyprotein	5.97E-15
eug24318	163	trithorax group protein osa-like	1.35E-33
eug27287	138	tropinone reductase-like 1	1.83E-17
eug24650	104	tubulin-folding cofactor C	5.19E-26
eug26927	109	tubulin-folding cofactor C	4.16E-32
eug18980	104	Tudor domain-containing protein	2.67E-53
eug16448	205	two pore potassium channel c-like isoform X1	2.72E-26
eug13485	552	two-component response regulator ARR14-like	0
eug21372	188	two-component response regulator ARR14-like	1.98E-83
eug6856	464	two-component response regulator ARR14-like	0
eug6860	355	two-component response regulator ARR14-like	6.69E-138
eug19076	510	two-component response regulator ARR1-like	2.81E-177
eug8	382	two-component response regulator ARR2-like isoform X1	0
eug14326	100	two-pore potassium channel 3-like	2.26E-34
eug4822	101	two-pore potassium channel 3-like	2.15E-38
eug17551	108	type I inositol polyphosphate 5-phosphatase 5	6.62E-25
eug2223	110	type I inositol polyphosphate 5-phosphatase 5	3.64E-25
eug16786	162	ubiquitin carboxyl-terminal hydrolase 8-like	6.65E-37
eug5823	113	ubiquitin carboxyl-terminal hydrolase 8-like isoform X2	5.84E-16

eug14454	114	Ubiquitin system component Cue	2.89E-47
eug15406	108	Ubiquitin-conjugating enzyme E2 2	2.01E-33
eug16805	91	Ubiquitin-conjugating enzyme E2 2	9.75E-32
eug3584	86	ubiquitin-like protein 5	6.28E-09
eug14273	89	UBN2_3 domain-containing protein	3.67E-41
eug16057	215	U-box domain-containing protein 12-like	3.48E-138
eug26024	91	U-box domain-containing protein 26	2.28E-12
eug10504	67	U-box domain-containing protein 26-like	2.60E-15
eug11057	76	U-box domain-containing protein 26-like	1.40E-12
eug20693	67	U-box domain-containing protein 26-like	1.54E-14
eug25855	94	U-box domain-containing protein 26-like	1.03E-31
eug6286	76	U-box domain-containing protein 26-like	1.38E-13
eug10887	281	U-box domain-containing protein 35-like	4.23E-158
eug21843	91	U-box domain-containing protein 35-like	7.97E-49
eug3119	237	U-box domain-containing protein 35-like	6.78E-116
eug9165	91	U-box domain-containing protein 35-like	6.08E-50
eug20480	164	U-box domain-containing protein 51-like isoform X1	7.23E-89
eug25154	172	U-box domain-containing protein 52-like	1.49E-101
eug15186	146	UDP-D-apiose/UDP-D-xylose synthase 2	1.00E-49
eug15924.1	147	UDP-glycosyltransferase 73C3-like	3.61E-79
eug16823.1	147	UDP-glycosyltransferase 73C3-like	2.63E-84
eug14268.1	208	UDP-glycosyltransferase 91C1	3.24E-120
eug14271.1	209	UDP-glycosyltransferase 91C1	7.84E-101
eug19577	100	uncharacterized mitochondrial protein AtMg00860-like	2.73E-31
eug27658	173	uncharacterized protein LOC104422917	2.27E-11
eug1902	198	uncharacterized protein LOC104428113	5.65E-98
eug25516	93	uncharacterized protein LOC104439428	7.90E-29
eug6551	95	uncharacterized protein LOC104439428	2.35E-30
eug15258	103	uncharacterized protein LOC104449201	2.05E-26
eug17129	103	uncharacterized protein LOC104449201	3.18E-28
eug27643	239	uncharacterized protein LOC104453401	1.99E-09
eug29849	115	uncharacterized protein LOC104453401	1.59E-28
eug14961	122	uncharacterized protein LOC104455534	4.05E-16
eug24203.1	390	uncharacterized protein LOC104455534	4.50E-118
eug8469.1	349	uncharacterized protein LOC104455534	1.58E-97
eug14159	178	uncharacterized protein LOC104455547	6.53E-20
eug22442	143	uncharacterized protein LOC108955637	4.20E-09
eug29648	308	uncharacterized protein LOC108957620	3.61E-129
eug21041.2	439	uncharacterized protein LOC108958440 isoform X2	3.07E-151
eug2761	172	uncharacterized protein LOC108959267	1.41E-41
eug11487	70	uncharacterized protein LOC115661707	9.95E-18
eug18037	222	uncharacterized protein LOC115661707	2.01E-11
eug18410	78	uncharacterized protein LOC115661707	3.01E-16
eug23128	78	uncharacterized protein LOC115661707	2.86E-20
eug29129	78	uncharacterized protein LOC115661707	4.87E-19

eug19245	106	uncharacterized protein LOC115662167	5.00E-06
eug19064	124	uncharacterized protein LOC115663348	7.17E-59
eug20791	123	uncharacterized protein LOC115663348	6.22E-53
eug29826	230	uncharacterized protein LOC115664499	2.67E-49
eug29031	117	uncharacterized protein LOC115665593	1.18E-27
eug14787	259	uncharacterized protein LOC115667281 isoform X2	2.41E-33
eug2073	165	uncharacterized protein LOC115667281 isoform X2	1.91E-08
eug8924	205	uncharacterized protein LOC115668702	3.25E-42
eug11101	132	uncharacterized protein LOC115669680	8.19E-18
eug24710.1	300	uncharacterized protein LOC115671143	2.93E-157
eug11592	121	uncharacterized protein LOC115672919 isoform X2	1.99E-14
eug4830	118	uncharacterized protein LOC115674312	2.33E-32
eug30368	145	uncharacterized protein LOC115674840	3.55E-38
eug4454	124	uncharacterized protein LOC115674840	1.60E-47
eug2809	93	uncharacterized protein LOC115675023 isoform X1	1.53E-50
eug8029	96	uncharacterized protein LOC115675023 isoform X1	1.81E-51
eug10679	232	uncharacterized protein LOC115678941	2.08E-67
eug26882	152	uncharacterized protein LOC115679265	1.22E-51
eug23824	276	uncharacterized protein LOC115679404	6.19E-14
eug21727	94	uncharacterized protein LOC115679820	9.64E-17
eug22235	75	uncharacterized protein LOC115679973 isoform X2	1.81E-21
eug12525	198	uncharacterized protein LOC115680938	1.62E-65
eug29567	274	uncharacterized protein LOC115680938	1.92E-112
eug1077	384	uncharacterized protein LOC115681319	1.33E-154
eug20922	237	uncharacterized protein LOC115681319	2.61E-112
eug28442	114	uncharacterized protein LOC115682019	2.89E-09
eug28443	178	uncharacterized protein LOC115682019	6.56E-10
eug4355	148	uncharacterized protein LOC115682019	8.04E-11
eug615	148	uncharacterized protein LOC115682019	4.52E-11
eug27019	107	uncharacterized protein LOC115682466	5.42E-10
eug29162	106	uncharacterized protein LOC115682466	8.88E-16
eug1472.2	457	uncharacterized protein LOC115684511	2.46E-119
eug4307	142	uncharacterized protein LOC115685964 isoform X1	1.59E-20
eug14628.1	616	uncharacterized protein LOC115688176	0
eug9643	330	uncharacterized protein LOC115688215 isoform X1	9.60E-154
eug25402	136	uncharacterized protein LOC115691315	2.80E-12
eug1313	159	uncharacterized protein LOC115691316	4.53E-42
eug28060	125	uncharacterized protein LOC115691316	4.38E-28
eug29087	163	uncharacterized protein LOC115691316	3.07E-10
eug23518	176	uncharacterized protein LOC115691967	2.25E-12
eug25217	136	uncharacterized protein LOC115691967	3.47E-10
eug16781	129	uncharacterized protein LOC115692257	1.12E-23
eug2029	66	uncharacterized protein LOC115692257	1.59E-11
eug12997	354	uncharacterized protein LOC115692702	0
eug15123	522	uncharacterized protein LOC115692702	7.51E-167

eug15652	229	uncharacterized protein LOC115693306	1.69E-123
eug11742	133	uncharacterized protein LOC115725842	1.88E-45
eug19624	192	uncharacterized protein LOC115725842	2.59E-101
eug26809	232	uncharacterized protein LOC115725842	3.26E-122
eug7971	229	uncharacterized protein LOC115725884	8.69E-109
eug1306	384	uncharacterized protein LOC115726065	0
eug18671	217	uncharacterized protein LOC115726065	2.69E-113
eug5146	217	uncharacterized protein LOC115726065	7.10E-115
eug2531	124	uncharacterized protein LOC115726459	2.32E-39
eug19974	469	uncharacterized protein LOC115726631	0
eug5448	247	uncharacterized protein LOC115727112	1.68E-110
eug5736	186	uncharacterized protein LOC115727349	2.21E-31
eug7838	146	uncharacterized protein LOC115727349	2.64E-40
eug5947	377	uncharacterized protein LOC115728094	0
eug13287.1	225	uncharacterized protein LOC115728662	1.62E-100
eug16769	211	uncharacterized protein LOC115728662	6.93E-128
eug26222	211	uncharacterized protein LOC115728662	4.16E-125
eug29451.1	225	uncharacterized protein LOC115728662	3.30E-101
eug126	303	uncharacterized protein LOC115729390 isoform X1	8.72E-155
eug4824	328	uncharacterized protein LOC115729390 isoform X1	1.16E-168
eug8972	327	uncharacterized protein LOC115729390 isoform X1	0
eug30367	224	uncharacterized protein LOC115729390 isoform X2	2.58E-100
eug7668	256	uncharacterized protein LOC115729390 isoform X2	6.69E-135
eug25838.1	405	uncharacterized protein LOC115729396	7.72E-96
eug28070	463	uncharacterized protein LOC115729420	0
eug16516	128	uncharacterized protein LOC115729615	2.65E-61
eug19191	111	uncharacterized protein LOC115730567	5.78E-25
eug17749	159	uncharacterized protein LOC115730998	1.38E-75
eug6698	95	uncharacterized protein LOC115731028	4.78E-11
eug8666	129	uncharacterized protein LOC115731031	2.08E-09
eug16806	172	uncharacterized protein LOC115731239	8.40E-81
eug16807	207	uncharacterized protein LOC115731239	2.43E-80
eug21957	189	uncharacterized protein LOC115731239	1.13E-81
eug24850	128	uncharacterized protein LOC115731592	1.34E-58
eug8261	100	uncharacterized protein LOC115731592	1.75E-38
eug29307	127	uncharacterized protein LOC115731706	1.89E-07
eug12366	125	uncharacterized protein LOC115731936	5.96E-07
eug11278	148	uncharacterized protein LOC115735400	3.12E-24
eug20765	116	uncharacterized protein LOC115735518	1.21E-64
eug4506	116	uncharacterized protein LOC115735518	7.04E-64
eug22503	188	uncharacterized protein LOC115739970 isoform X1	2.06E-82
eug1845	283	uncharacterized protein LOC115739970 isoform X2	8.63E-83
eug9613	314	uncharacterized protein LOC115740933	7.83E-115
eug20204	150	uncharacterized protein LOC115741425	5.31E-60
eug2042	117	uncharacterized protein LOC115743494 isoform X1	9.21E-50

eug820	104	uncharacterized protein LOC115743494 isoform X1	1.15E-53
eug21652	104	uncharacterized protein LOC115743494 isoform X3	3.15E-47
eug12617	192	uncharacterized protein LOC115744539	8.58E-81
eug22801	172	uncharacterized protein LOC115744539	1.55E-85
eug13188	217	uncharacterized protein LOC115746823	7.97E-79
eug186	124	uncharacterized protein LOC115746823	7.50E-41
eug20672	487	uncharacterized protein LOC115747043	0
eug28198	165	uncharacterized protein LOC115747043	1.24E-72
eug633	547	uncharacterized protein LOC115747043	0
eug8477	138	uncharacterized protein LOC115750017 isoform X1	2.11E-83
eug24200	174	uncharacterized protein LOC115750017 isoform X5	1.31E-86
eug2556	253	uncharacterized protein LOC115750017 isoform X5	6.19E-127
eug8478	252	uncharacterized protein LOC115750017 isoform X7	7.16E-127
eug26813	106	uncharacterized protein LOC115750330	9.02E-19
eug16298	85	uncharacterized protein LOC115751743 isoform X5	3.91E-10
eug16796	151	uncharacterized protein LOC115751743 isoform X5	6.57E-22
eug2123	161	uncharacterized protein LOC115752611 isoform X1	1.35E-34
eug8524	159	uncharacterized protein LOC115752611 isoform X1	2.15E-37
eug2124	184	uncharacterized protein LOC115752611 isoform X2	2.08E-23
eug1935	185	uncharacterized protein LOC115752695	8.29E-57
eug23193	185	uncharacterized protein LOC115752695	5.77E-73
eug6484	186	uncharacterized protein LOC115752695	5.70E-74
eug25841	565	uncharacterized protein LOC115753693 isoform X1	3.40E-104
eug6804	269	uncharacterized protein LOC115753757 isoform X1	2.19E-153
eug9329	263	uncharacterized protein LOC115753757 isoform X1	1.29E-152
eug1509	149	uncharacterized protein LOC115753760 isoform X1	1.04E-79
eug25544	160	uncharacterized protein LOC115753760 isoform X1	1.38E-90
eug28151	149	uncharacterized protein LOC115753760 isoform X1	8.42E-77
eug14714	160	uncharacterized protein LOC115753760 isoform X2	1.32E-96
eug10120	166	uncharacterized protein LOC115754282	2.18E-49
eug25219	101	uncharacterized protein LOC115754361	1.85E-28
eug5740	182	uncharacterized protein LOC115755003 isoform X3	1.23E-80
eug7311	188	uncharacterized protein LOC115755003 isoform X3	1.21E-81
eug513	183	uncharacterized protein LOC115755041	4.32E-101
eug515	233	uncharacterized protein LOC115755041	1.04E-101
eug22433	276	uncharacterized protein LOC115756748	7.88E-152
eug18063	283	uncharacterized protein LOC115756749	7.52E-152
eug10163	176	uncharacterized protein LOC115756915	1.34E-49
eug10165	176	uncharacterized protein LOC115756915	1.27E-65
eug21726	190	uncharacterized protein LOC115757652 isoform X2	7.26E-83
eug8388	83	uncharacterized protein LOC116114809	1.56E-07
eug18144	277	uncharacterized protein LOC116207851	1.05E-85
eug21431	226	uncharacterized protein LOC116207851	1.08E-85
eug2874	246	uncharacterized protein LOC116207851	2.13E-98
eug28929	271	uncharacterized protein LOC116207851	2.32E-86

eug3114	147	uncharacterized protein LOC119330369	1.13E-32
eug9953	510	uncharacterized protein LOC120286529	0
eug26842	301	uncharacterized protein LOC120286880	1.60E-71
eug7591.1	226	uncharacterized protein LOC120286880	2.10E-35
eug29825	152	uncharacterized protein LOC120290446	5.04E-52
eug25794.2	551	uncharacterized protein LOC120290558	0
eug17562	286	uncharacterized protein LOC123150699	5.89E-26
eug27088.2	447	uncharacterized protein LOC8269817 isoform X3	6.54E-28
eug28165	83	uncharacterized protein LOC9329671	2.78E-06
eug12914	96	unnamed protein product	7.16E-17
eug17949	95	unnamed protein product	1.07E-08
eug25237	102	unnamed protein product	9.74E-10
eug3760	87	UPF0160 protein MYG1, mitochondrial	1.62E-54
eug5347	87	UPF0160 protein MYG1, mitochondrial	4.07E-55
eug10671	191	UPF0481 protein At3g47200-like	9.23E-21
eug11107	244	UPF0481 protein At3g47200-like	1.48E-15
eug12003	104	UPF0481 protein At3g47200-like	2.26E-09
eug18188	124	UPF0481 protein At3g47200-like	2.75E-06
eug20625	95	UPF0481 protein At3g47200-like	1.10E-20
eug30459	106	UPF0481 protein At3g47200-like	1.63E-21
eug5894	79	UPF0481 protein At3g47200-like	5.68E-17
eug11882	129	UPF0496 protein 4-like	2.38E-57
eug13445	95	UPF0496 protein 4-like	4.47E-24
eug14325	204	UPF0496 protein 4-like	6.59E-62
eug3967	108	UPF0496 protein 4-like	1.93E-35
eug4784	140	UPF0496 protein 4-like	2.94E-56
eug12457	494	vacuolar-processing enzyme-like	0
eug4413	484	vacuolar-processing enzyme-like	0
eug6034	469	vacuolar-processing enzyme-like	0
eug958	239	valine--tRNA ligase, mitochondrial 1-like isoform X1	7.34E-66
eug21575	129	vicilin-like seed storage protein At2g28490	2.35E-11
eug15717	195	VQ motif-containing protein	4.05E-102
eug29115	147	WAT1-related protein At5g07050-like	1.18E-76
eug7454	154	WAT1-related protein At5g07050-like	1.48E-76
eug10108	77	zeatin o-xylosyltransferase	3.12E-11
eug3597	77	zeatin o-xylosyltransferase	5.62E-11
eug10065	110	zf-GRF domain-containing protein	3.06E-40
eug24568	177	zinc finger CCCH domain-containing protein 29	2.78E-27
eug24892	190	zinc finger CCCH domain-containing protein 48-like	4.10E-96
eug27594	129	zinc finger CCCH domain-containing protein 48-like	3.06E-82
eug20070	139	zinc finger homeobox protein 4-like isoform X1	2.81E-35
eug12827	249	Zinc finger protein 1	3.16E-144
eug17853	378	zinc finger protein 197	0
eug13060	200	zinc finger protein GAI-ASSOCIATED FACTOR 1-like	4.23E-26
eug3518	246	zinc finger protein GAI-ASSOCIATED FACTOR 1-like	1.02E-25

eug12824	223	zinc finger protein GIS3-like	1.49E-109
eug12826	221	zinc finger protein GIS3-like	6.36E-100
eug12828	245	zinc finger protein GIS-like	4.65E-141
eug17304	89	zinc finger protein SHOOT GRAVITROPISM 5-like	5.58E-06
eug10619	93	zinc finger protein ZAT10-like	2.59E-26
eug13312	80	zinc finger protein ZAT10-like	1.06E-27
eug21072	93	zinc finger protein ZAT10-like	1.63E-27
eug9697	155	zinc finger protein ZAT10-like	6.61E-25
eug17854	425	zinc finger protein ZAT4-like	0
eug22899	95	zinc finger protein ZAT4-like	5.45E-13

## ANEXO 6

**Tabla 11.** Descripción de los términos de Ontología de Genes (GO) de los genes exclusivos de *E. uniflora*.

<b>Termino GO</b>	<b>Descripción</b>	<b>Número de genes</b>
GO:0044238	primary metabolic process	361
GO:0006807	nitrogen compound metabolic process	298
GO:0044237	cellular metabolic process	259
GO:0097159	organic cyclic compound binding	227
GO:0044424	intracellular part	224
GO:0016740	transferase activity	186
GO:0043229	intracellular organelle	172
GO:0009058	biosynthetic process	163
GO:0043227	membrane-bounded organelle	157
GO:0016787	hydrolase activity	150
GO:0005515	protein binding	128
GO:0036094	small molecule binding	123
GO:0065007	biological regulation	84
GO:0071944	cell periphery	83
GO:0005886	plasma membrane	82
GO:0051179	localization	78
GO:0007154	cell communication	77
GO:0007165	signal transduction	75
GO:0050794	regulation of cellular process	75
GO:0051716	cellular response to stimulus	75
GO:0006950	response to stress	70
GO:0009056	catabolic process	69
GO:0042221	response to chemical	62
GO:0005215	transporter activity	55
GO:0009719	response to endogenous stimulus	47
GO:0003700	DNA-binding transcription factor activity	37
GO:0048856	anatomical structure development	36
GO:0043228	non-membrane-bounded organelle	34
GO:0007275	multicellular organism development	27
GO:0009605	response to external stimulus	24
GO:0005576	extracellular region	21
GO:0009607	response to biotic stimulus	21
GO:0044422	organelle part	21
GO:0009628	response to abiotic stimulus	19
GO:0009791	post-embryonic development	19
GO:0031974	membrane-enclosed lumen	19
GO:0012505	endomembrane system	15
GO:0030246	carbohydrate binding	15
GO:0007049	cell cycle	14



GO:0022414	reproductive process	13
GO:0003006	developmental process involved in reproduction	12
GO:0030234	enzyme regulator activity	12
GO:0005198	structural molecule activity	9
GO:0016049	cell growth	9
GO:0048869	cellular developmental process	6
GO:1990904	ribonucleoprotein complex	6
GO:0003682	chromatin binding	5
GO:0008289	lipid binding	4
GO:0019725	cellular homeostasis	4
GO:0019222	regulation of metabolic process	3
GO:0019748	secondary metabolic process	2
GO:0007568	aging	1
GO:0021700	developmental maturation	1
GO:0038023	signaling receptor activity	1
GO:0044706	multi-multicellular organism process	1