

I. ÍNDICE DE CONTENIDOS

II. ÍNDICE DE TABLAS	IV
III. ÍNDICE DE FIGURAS	VI
IV. RESUMEN	VII
V. ABSTRACT	IX
1. INTRODUCCIÓN	1
1.1 Papaya cultivada en Chile	1
1.2 Exportación de productos	3
1.3 Calidad frutal y maduración	4
1.4 Importancia de fitohormona Etileno	6
1.5 Ablandamiento de los frutos	7
1.6 Abordaje genómico computacional	12
2. DESCRIPCIÓN DEL PROBLEMA	16
3. HIPÓTESIS	18
4. OBJETIVOS	19
4.1 Objetivo General	19
4.2 Objetivos Específicos	19
5. METODOLOGÍA	20
5.1 Lenguajes y paquetes bioinformáticos utilizados en estudio	21
Perl	21
BioPerl	21
Awk	22
BLAST	22
MySql	24
5.2 Obtención del material genético	24
5.3 Secuenciación	25
5.4 Procesamiento y análisis de ESTs	26
5.4.1 Ensamble y aplicación de ESTs	26

5.4.2 Mapeo comparativo	30
5.4.3 Anotación funcional	32
5.4.3.1 Anotación mediante genomas y bases de datos	32
5.4.3.2 Mejores hits bidireccionales	34
5.4.3.3 Genes candidatos	35
5.5 Enriquecimiento del proceso de anotación	36
5.5.1 Clasificación por función biológica	36
5.5.2 Clasificación mediante dominios conservados	38
5.5.3 Reconstrucción de rutas metabólicas	38
5.6 Almacenamiento en base de datos	39
6. RESULTADOS	40
6.1 Procesamiento de las secuencias del transcriptoma	40
6.1.1 Información del ensamble	40
6.1.2 Identificación de variantes	42
6.2 Mapeo contra genoma de referencia	43
6.3 Anotación funcional de transcritos	45
6.3.1 Anotación mediante similitud entre secuencias	45
6.3.2 Mejores hits bidireccionales	56
6.4 Enriquecimiento del proceso de anotación	57
6.4.1 Análisis de clasificación biológica	57
6.4.2 Dominios conservados	59
6.4.3 Reconstrucción metabólica	61
6.5 Base de datos	62
7. DISCUSIÓN	63
7.1 Procesamiento del transcriptoma	63
7.1.1 Proceso de secuenciación	63
7.1.2 Identificación de variantes de splicing	64
7.2 Mapeo contra genoma de referencia	64
7.3 Anotación funcional de transcritos	65

7.3.1 Anotación mediante base de datos y genomas	65
7.3.2 Clasificación por categorías asociadas a maduración	66
7.3.3 Mejores hits bidireccionales	69
7.4 Enriquecimiento del proceso de anotación	69
7.4.1 Análisis mediante KOG	69
7.4.2 Análisis mediante CDD	70
7.4.3 Reconstrucción metabólica usando PRIAM	70
8. CONCLUSIONES	71
9. ANEXOS	73
10. REFERENCIAS	83

II. ÍNDICE DE TABLAS

Tabla 1	Rendimiento asociado a técnicas de secuenciación de segunda y tercera generación	13
Tabla 2	Distribución de estado de <i>reads</i> posterior a ensamble.	28
Tabla 3	Total de Singletons y secuencias de baja calidad.	29
Tabla 4	Distribución de isogrupos.	29
Tabla 5	Distribución de isotigs.	29
Tabla 6	Resumen taxonómico de genomas utilizados.	34
Tabla 7	Categorías de Cog y Kog.	37
Tabla 8	Resumen final de proceso de ensamble del transcriptoma.	43
Tabla 9	Resultados del mapeo contra genoma de referencia <i>C. papaya</i> con diferentes valores de alineamiento.	44
Tabla 10	Comparación de mapeo de <i>V. pubescens</i> contra genoma de <i>C. papaya</i> de diferentes bases de datos.	44
Tabla 11	Resumen de resultados de anotación de <i>V. pubescens</i> .	46
Tabla 12	Contigs anotados mediante genoma de referencia <i>C. papaya</i> .	46
Tabla 13	Clasificación de contigs según procesos de maduración frutal.	48
Tabla 14	Componentes clasificados en categoría Acumulación de azúcar.	49
Tabla 15	Componentes clasificados en categoría Biosíntesis de carotenoides.	49
Tabla 16	Componentes clasificados en categoría Cambios de color en cáscara.	49
Tabla 17	Componentes clasificados en categoría Crecimiento frutal.	49
Tabla 18	Componentes clasificados en categoría Desarrollo frutal.	50
Tabla 19	Componentes clasificados en categoría Enzimas asociadas a la degradación de pared.	50
Tabla 20	Componentes clasificados en categoría Expansinas.	50

Tabla 21	Componentes clasificados en categoría Factores de transcripción.	51
Tabla 22	Componentes clasificados en categoría Forma de fruto.	51
Tabla 23	Componentes clasificados en categoría Isotiocianatos.	51
Tabla 24	Componentes clasificados en categoría Laticíferos y Proteasas.	52
Tabla 25	Componentes clasificados en categoría Lignificación secundaria y pared celular.	52
Tabla 26	Componentes clasificados en categoría Polifenoloxidasas y Peroxidasas.	52
Tabla 27	Componentes clasificados en categoría Producción de volátiles.	53
Tabla 28	Componentes clasificados en categoría Regulación cambios de textura.	53
Tabla 29	Componentes clasificados en categoría Respiración.	53
Tabla 30	Componentes clasificados en categoría Respuesta a etileno.	53
Tabla 31	Componentes clasificados en categoría Síntesis de etileno.	54
Tabla 32	Componentes clasificados en categoría Síntesis de lignina.	54
Tabla 33	Selección de categorías interesantes en ablandamiento frutal.	55
Tabla 34	Total de hits bidireccionales.	57
Tabla 35	Clasificación de las secuencias de acuerdo a sub-categorías de KOG	58
Tabla 36	Clasificación de las secuencias de acuerdo a categorías de CDD.	59
Tabla 37	Total de componentes identificados en reconstrucción metabólica.	61

III. ÍNDICE DE FIGURAS

Figura 1	Comparación morfológica entre <i>V. pubescens</i> y <i>C. papaya</i> .	2
Figura 2	Secuencia de estadios de maduración de <i>V. pubescens</i> .	5
Figura 3	Cambios en tasa respiratoria de frutos climatéricos y no climatéricos.	6
Figura 4	Ciclo de biosíntesis de etileno.	7
Figura 5	Cambios en producción de etileno y pérdida de firmeza en <i>V. pubescens</i> .	8
Figura 6	Principales componentes de pared celular vegetal.	9
Figura 7	Etapas de ensamble de <i>reads</i> .	20
Figura 8	Relación entre exones, contigs e isotigs.	26
Figura 9	Distribución cromosómica de genomas utilizados.	27
Figura 10	Diagrama de metodología.	30
Figura 11	Distribución de largo de contigs totales.	34
Figura 12	Reciprocidad de “mejor hit bidireccional”.	35
Figura 13	Distribución de isotigs por isogrupos.	41
Figura 14	Distribución de núm. de contigs presentes en un isogrupo.	41
Figura 15	Proceso de ensamble para conformar un isotig.	42
Figura 16	Visualización de mapeo de <i>V. pubescens</i> y <i>C. papaya</i> en Gbrowse.	44
Figura 17	Resumen de anotación obtenida con y sin descripción.	46
Figura 18	Diagrama de Venn de número de aciertos específicos.	47
Figura 19	Distribución de hits obtenidos mediante anotación.	56
Figura 20	Porcentaje de secuencias significativas clasificadas por categorías funcionales de KOG.	57
Figura 21	Resumen final del proceso de anotación	60
Figura 22	Diagrama principal de base de datos	62