



UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS
(*Universidad del Perú, DÉCANA DE AMÉRICA*)

ESCUELA DE POSGRADO
FACULTAD DE CIENCIAS BIOLÓGICAS
UNIDAD DE POSGRADO

TRANSCRIPTOMA DE LA RESPUESTA A LA SEQUÍA EN
Solanum tuberosum* subsp. *andigena

TESIS
Para optar al Grado de Magister en Biología Molecular

Bach. Yerisf Carla Torres Ascurra

LIMA-PERU

2014

AGRADECIMIENTOS

A la Unidad de Genómica de la Universidad Cayetano Heredia, por el financiamiento y todas las facilidades para el desarrollo de esta tesis, y al Consejo de Ciencia y Tecnología (CONCYTEC) por el financiamiento de los estudios de maestría.

A la Dra. Gisella Orjeda, por la oportunidad que me brindó para la realización de este trabajo, a Roberto Lozano, por su constante apoyo durante todo este proceso. A Patricia Ponce, por la amistad y los buenos momentos en el lab, a Diana Martinez, por su compañía durante ese año de clases nocturnas en la maestría.

A mi asesora, Giovanna Sofil, por todo el apoyo brindado, las valiosas correcciones y sugerencias que hicieron posible culminar con este trabajo.

A la Dra.Noemí Zúñiga de la Estación Experimental de Huancayo del INIA por todo su apoyo. A la Sra. Norma y el Sr. Juan también por su ayuda.

A mi familia, a mi adorada madre Zoraya, a mis queridas hermanas Yozara y Nela, a mis dos bellos ángeles Anel y Terry, y a mi padre. Gracias por su compañía, por los momentos únicos e irremplazables, por el apoyo y por ser siempre mi puerto seguro.

A mis amigos de siempre, a Aling por su compañía y apoyo incondicional, a Ingrid y Mildred, por los divertidos momentos que pasamos juntas, aunque pocos sean.

DEDICATORIA

A mi madre, por ser mi fuerza y motivación

CONTENIDO

AGRADECIMIENTOS.....	ii
DEDICATORIA.....	iii
CONTENIDO.....	iv
INDICE DE TABLAS.....	vi
ÍNDICE DE FIGURAS.....	viii
LISTA DE ABREVIATURAS.....	xi
RESUMEN.....	xii
ABSTRACT.....	xiii
I. INTRODUCCIÓN.....	1
II. ANTECEDENTES	6
1. LA SEQUÍA	6
2. EFECTOS DE LA SEQUÍA EN LAS PLANTAS	7
a. Crecimiento y rendimiento del cultivo.....	7
b. Crecimiento de las raíces.....	8
c. Relaciones hídricas	8
d. Fotosíntesis.....	9
3. LA RESISTENCIA A LA SEQUÍA	10
a. Evitamiento	10
b. Tolerancia	11
4. TRANSDUCCIÓN DE SEÑALES Y EXPRESIÓN GÉNICA EN RESPUESTA A LA SEQUÍA....	12
a. El ácido abscísico (ABA).....	12
b. Percepción celular del déficit hídrico	12
c. Segundos mensajeros y moléculas señal	13
d. Control transcripcional.....	14
e. Proteínas Funcionales.....	15
5. LA PAPA	17
a. Las papas cultivadas.....	17
b. La especie <i>Solanum tuberosum</i>	18
6. LA PAPA Y LA SEQUÍA.....	19
a. Fisiología de la papa frente a la sequía.....	19
b. Expresión génica de plantas de papa en respuesta a la sequía	21
7. RNA-SEQ: SECUENCIAMIENTO MASIVO DE LECTURAS DE RNA.....	22
a. Metodología del RNA-Seq.....	23

b.	Estudios de estrés en plantas empleando RNA-Seq	21
III.	HIPÓTESIS	26
IV.	OBJETIVOS	26
1.	OBJETIVO GENERAL.....	26
2.	OBJETIVOS ESPECÍFICOS	26
V.	MATERIALES Y MÉTODOS.....	27
1.	LOCALIZACIÓN Y CONDICIONES CLIMÁTICAS	27
2.	MATERIAL VEGETAL.....	27
3.	INDUCCIÓN DEL ESTRÉS POR SEQUÍA	30
4.	EVALUACIONES FISIOLÓGICAS	30
5.	ANÁLISIS DE LA EXPRESIÓN GÉNICA	31
a.	Extracción de RNA	31
b.	Secuenciación de RNA	32
c.	Análisis bioinformático.....	32
c.1.	Control de calidad.....	32
c.2.	Alineamiento de secuencias al genoma de referencia.....	33
c.3.	Ensamblaje de transcritos y cuantificación de su expresión	33
c.4.	Clasificación funcional de los transcritos	34
c.5.	Redes de co-expresión génica.....	35
VI.	RESULTADOS.....	37
1.	EVALUACIÓN DE LA TASA FOTOSINTÉTICA	37
a.	Evaluación visual durante la exposición a sequía	39
2.	EXTRACCIÓN DE RNA	40
3.	ANÁLISIS BIOINFORMÁTICO DE LOS DATOS GENERADA POR RNA-SEQ	41
a.	Control de calidad de secuencias	41
b.	Mapeo de las lecturas	45
c.	Ensamblaje de transcritos	47
d.	Expresión diferencial.....	48
e.	Clasificación funcional de los transcritos.....	60
f.	Redes de co-expresión génica	69
VII.	DISCUSIONES.....	82
VIII.	CONCLUSIONES.....	90
IX.	RECOMENDACIONES	91
X.	REFERENCIAS BIBLIOGRÁFICAS	92
XI.	ANEXOS	104

INDICE DE TABLAS

	Pág.
Tabla 1. Descripción de las variedades de <i>S. tuberosum</i> subsp. <i>andigena</i> empleadas en este estudio.	27
Tabla 2. Composición del medio utilizado para la propagación de plántulas de <i>Solanum tuberosum</i> subsp. <i>andigena</i> .	28
Tabla 3. Composición de la solución nutritiva empleada para el cultivo aeropónico (Otazú, 2010).	29
Tabla 4. Códigos de las muestras colectadas de <i>S. tuberosum</i> subsp. <i>andigena</i> de las variedades tolerante y susceptible para la extracción de RNA.	31
Tabla 5. Tiempos de muestreo de hojas y raíces calculados a partir de la variación de la tasa de fotosíntesis (Pn) durante la respuesta temprana (T1), tardía (T2) y recuperación del estrés (T3).	38
Tabla 6. Cuantificación (ng/ μ L) y calidad (índices de absorbancia 260nm/280nm, 260nm/230nm) de las extracciones de RNA utilizando un equipo NanoDrop.	41
Tabla 7. Lista de bibliotecas correspondiente a cada variedad y tratamiento y número total de lecturas generadas.	42
Tabla 8. Porcentajes del tipo de lecturas según su mapeo al genoma de referencia de DM1-3 516R44.	46
Tabla 9. Número de transcriptos identificados en las 16 librerías.	47

Tabla 10. Niveles de expresión (en fpkm) de genes inducidos específicamente en la respuesta tardía, manteniendo o disminuyendo luego en la recuperación.	49
Tabla 11. Algunos de los genes inducidos en raíces de la variedad 703671 durante la recuperación	51
Tabla 12. Algunos de los genes reprimidos en hojas de la variedad 703671 durante la recuperación.	52
Tabla 13. Algunos genes específicos de hojas de la variedad 703671 durante la etapa temprana.	53
Tabla 14. Niveles de expresión de transcritos de pequeñas proteínas de choque térmico en cloroplastos en hojas de la variedad tolerante 703671 y la variedad susceptible 703248 durante la recuperación.	54
Tabla 15. Número de genes en los tejidos diferencialmente expresados en respuesta a la sequía. Se muestra el número de genes específicos y comunes a los tejidos.	59
Tabla 16. Algunos genes del módulo de co-expresión B1 de hojas de 703671.	77

INDICE DE FIGURAS

	Pág.
Figura 1. Posibles mecanismos responsables de la reducción del crecimiento en plantas bajo condiciones de estrés	8
Figura 2. Componentes de la resistencia a la sequía en plantas	11
Figura 3. Representación gráfica del sistema aeropónico de cultivo de papa (Otazú, 2010)	29
Figura 4. Variación de la tasa de fotosíntesis en la variedad 703671 (tolerante) durante las condiciones control (riego inicial: R1), sequía (sin riego) y recuperación (reinicio del riego: R2).	37
Figura 5. Variación de la tasa de fotosíntesis en la variedad 703248 (susceptible) durante las condiciones control (riego inicial: R1), sequía (sin riego) y recuperación (reinicio del riego: R2).	38
Figura 6. Plantas de la variedad 703671 luego de 60 minutos de exposición a sequía.	39
Figura 7. Plantas de la variedad 703248 luego de 60 minutos de exposición a sequía.	39
Figura 8. Evaluación de calidad de RNA extraído de hojas y raíces de <i>S. tuberosum</i> subsp. <i>andigena</i> empleando geles de agarosa al 1%.	40
Figura 9. Valores Phred de calidad para cada base de las lecturas.	44
Figura 10. Número de genes diferencialmente expresados en hojas y raíces de las variedades 703248 (susceptible) y 703671 (tolerante).	48

Figura 11. Distribución del número de genes suprimidos e inducidos por sequía.	50
Figura 12. Número total de genes diferencialmente expresados por sequía (respuesta temprana y tardía) y en la recuperación.	55
Figura 13. Número de genes inducidos por sequía (respuesta temprana y tardía) y en la recuperación	56
Figura 14. Número de genes reprimidos por sequía (respuesta temprana y tardía) y en la recuperación.	57
Figura 15. Número de asignaciones a las tres categorías de Gene Ontology.	60
Figura 16. Categorías GO de los 200 transcritos más inducidos por sequía durante la respuesta temprana en hojas de 703248 y 703671.	63
Figura 17. Categorías GO de los 200 transcritos más inducidos por sequía durante la respuesta tardía en hojas de 703248 y 703671.	64
Figura 18. Categorías GO de los 200 transcritos más inducidos por sequía durante la recuperación en hojas de 703248 y 703671.	65
Figura 19. Categorías GO de los 200 transcritos más inducidos por sequía durante la respuesta temprana en raíces de 703248 y 703671.	66
Figura 20. Categorías GO de los 200 transcritos más inducidos por sequía durante la respuesta tardía en raíces de 703248 y 703671.	67
Figura 21. Categorías GO de los 200 transcritos más inducidos por sequía durante la recuperación en raíces de 703248 y	68

703671.

Figura 22. Heat Map de hojas de 703248 y 703671. El color indica el grado de correlación.	70
Figura 23. Heat Map de raíces de 703248 y 703671. El color indica el grado de correlación.	71
Figura 24. Gráficos de la tendencia de expresión de los módulos de genes formados en Hojas de 703248.	73
Figura 25. Heatmap de los eigengenes de los 6 módulos de co-expresión génica de hojas de 703248.	74
Figura 26. Gráficos de la tendencia de expresión de los módulos de genes formados en Hojas de 703671.	75
Figura 27. HeatMap de los eigengenes de los 6 módulos de coexpresión génica de hojas de 703671.	76
Figura 28. Gráficos de la tendencia de expresión de los módulos de genes formados en raíces de 703248.	78
Figura 29. HeatMap de los eigengenes de los 6 módulos de coexpresión génica de raíces de 703248.	79
Figura 30. Gráficos de la tendencia de expresión de los módulos de genes formados en raíces de 703671.	80
Figura 31. HeatMap de los eigengenes de los 6 módulos de coexpresión génica de raíces de 703671.	81

LISTA DE ABREVIATURAS

ABA	Ácido abscísico
DNA	Ácido desoxiribonucleico
cDNA	DNA complementario
RNA	Ácido ribonucleico
mRNA	RNA mensajero
RNA-Seq	Secuenciamiento de RNA
EST	Expressed sequence tag
LEA	Proteínas abundantes de embriogénesis tardía
FAO	Food and Agriculture Organization
INIA	Instituto Nacional de Innovación Agraria
CIP	Centro Internacional de la papa
PGSC	Consortio de secuenciamiento del genoma de papa
TIGR	The Institute for Genomic Research
FPKM	Fragmentos por kilobase de exón por millón de fragmentos mapeados
GO	Gene ontology
AgriGO	Gene ontology for agriculture
Blast	Basic local alignment search tool
UNIREF	Uniprot reference cluster
WGCNA	Weighted Correlation Network Analysis

RESUMEN

El reciente desarrollo del RNA-Seq, un método de secuenciamiento masivo en paralelo para el análisis de transcriptomas permite conocer el perfil de expresión de las plantas en respuesta a estrés de tipo abiótico y biótico. En este estudio, se secuenció el mRNA proveniente de hojas y raíces de dos variedades de *Solanum tuberosum* subsp. andigena, una tolerante y otra susceptible, expuestas a diferentes niveles de sequía. Lecturas de 50 pares de bases provenientes de mRNA, se mapearon al genoma de papa: entre el 75 - 82% mapearon a posiciones únicas, 6 - 14% mapearon a múltiples posiciones y 9 - 12% no mapearon a posición alguna del genoma. Comparando los perfiles de expresión, se encontraron entre 887 a 1925 genes inducidos/reprimidos por sequía en la variedad susceptible y 998- 1995 en la tolerante. Se anotaron funcionalmente los 200 genes más inducidos por cada tratamiento encontrándose información primaria respecto a los procesos biológicos y moleculares involucrados durante la sequía. Finalmente, fue posible correlacionar los perfiles de expresión diferencial de los genes durante la sequía, formándose 31 módulos con aquellos genes altamente correlacionados. Este estudio generó información de gran valor que podrá ser utilizada en futuros estudios para comprender mejor los mecanismos moleculares de tolerancia a sequía en papa y especies cercanas.

Palabras clave: RNA-Seq, expresión génica, estrés hídrico, papa, mRNA.

ABSTRACT

The recent advent of RNA-Seq, a massively parallel sequencing method for transcriptome analysis, provides an opportunity to understand the expression profile of plants in response to biotic and abiotic stress. In this study, the mRNA was sequenced from leaves and roots of two native potato varieties of *Solanum tuberosum* subsp. *andigena* (tolerant and susceptible) at different levels of drought. Fifty-base-pair reads from whole mRNAs were mapped to the potato genomic sequence: 75 - 82% mapped uniquely, 6 - 14% mapped to several locations and 9 - 12% had no match in the genome. Comparing the expression profiles, 887 to 1925 genes were found to be induced/repressed by drought in the sensible variety and 998 to 1995 in the tolerant variety. It was obtained the functional annotation of the 200 genes most induced, which rendered primary information of the biological process and molecular functions involved in drought. Finally, it was possible to correlate the differential expression profile of genes during drought, and 31 modules were formed with those highly correlated genes. This research provides valuable information for future studies and deeper understanding of the molecular mechanism of drought tolerance in potato and related species.

Keywords: RNA-Seq, gene expression, drought stress, potato, mRNA.

I. INTRODUCCIÓN

La papa es el tercer cultivo alimentario más importante del mundo (FAO STAT, 2008); en comparación a otros cultivos de importancia como el tomate, maíz o caña de azúcar, la papa es muy sensible a la sequía ya que necesita de riego frecuente (van Loon, 1981). La especie de papa más consumida e importante comercialmente *Solanum tuberosum* L. es altamente susceptible a la sequía, mientras que las especies nativas de los andes, cultivadas a altitudes tan altas como los 3500 msnm, están adaptadas a condiciones climáticas adversas (Vasquez-Robinet *et al.*, 2008). Esto hace a estas variedades candidatos ideales para estudiar la expresión de genes responsables de la tolerancia a la sequía, para que en un futuro puedan ser introgresados en *Solanum tuberosum* L., y lograr un incremento de la tolerancia que permitirían ahorrar agua de irrigación y garantizar el rendimiento y seguridad alimentaria.

Según una definición de campo, la sequía es un período sin lluvia que por su duración daña el cultivo y reduce significativamente los rendimientos económicos (Ekanayake, 1993). Es el estrés medio ambiental más importante en la agricultura, por lo que se han realizado muchos esfuerzos para mejorar la productividad de los cultivos bajo condiciones limitantes de agua. La ausencia de precipitación en el medio natural causa una disminución de la humedad de la atmósfera y del suelo, por lo que las plantas requieren de sistemas adecuados en raíces y hojas para percibir este déficit hídrico y desencadenar una transducción de señales que llevan posteriormente a la alteración del metabolismo celular, a cambios fisiológicos y de desarrollo de la planta (Yokota *et al.*, 2006).

Se han reportado diferencias entre cultivares en el mantenimiento del rendimiento bajo condiciones de estrés hídrico, las cuales son dependientes del genotipo (Lahlou *et al.*, 2003). Se dice que un genotipo es resistente a la sequía cuando produce un cultivo rentable dentro de los límites de su potencial de producción a pesar de la disponibilidad limitada de agua; esto se puede conseguir gracias a las siguientes estrategias: escape, tolerancia, evitamiento y

recuperación, las cuales no son mutuamente excluyentes y ofrecen al cultivo la capacidad de resistir la sequía en un período dado (Ekanayake, 1993).

La estrategia de escape permite a la planta completar su ciclo de vida durante el período de suficiente suministro de agua antes del inicio de la sequía. El evitamiento incluye estrategias que ayudan a la planta a mantener un alto estado hídrico durante periodos de estrés, mediante una eficiente absorción de agua por las raíces o reduciendo la evapotranspiración de las partes aéreas (Manavalan *et al.*, 2009). El mecanismo de tolerancia permite a la planta mantener el turgor y continuar el metabolismo incluso con un bajo potencial hídrico, por tolerancia protoplásmica o síntesis de osmoprotectores, osmolitos o solutos compatibles (Nguyen *et al.*, 1997). Finalmente, la recuperación es la capacidad de la planta para reestablecer su aparato fotosintético y sus principales funciones vitales una vez que la sequía termina (Manavalan *et al.*, 2009).

Las respuestas fisiológicas de la planta frente a la sequía incluyen la inhibición del crecimiento y metabolismo, especialmente la fotosíntesis y respiración. Las hojas cierran sus estomas inmediatamente para protegerse de la evaporación, y a medida que se pierde el agua de las hojas, la presión de turgencia de los tejidos disminuyen y las hojas se marchitan, para proteger su maquinaria fotosintética de los rayos solares (Yokota *et al.*, 2006). Aunque el estrés por sequía inhibe el crecimiento de los órganos aéreos, las raíces pueden continuar elongándose, como parte de un mecanismo adaptativo para permitir la obtención de agua de capas más profundas del suelo (Steudle, 2000).

Aún es incierto como este estrés es percibido o detectado por las plantas, pero varios estudios han encontrado moléculas que funcionan como osmosensores, es decir proteínas cuyo rol primario es monitorear fluctuaciones en la osmolaridad externa e iniciar una cascada de señales para una osmo-adaptación (Reisser *et al.*, 2003). En *Arabidopsis* se ha reportado la existencia de varios osmosensores candidatos, entre ellos ATHK1 (un homólogo de SLN1, identificado en levadura) que es sobreexpresado en raíces por cambios de osmolaridad y funciona como una histidina kinasa (Urao *et al.*, 1999). Otro posible

osmosensor es NtC7 de *Nicotiana tabacum*, una proteína de membrana tipo receptor cuya expresión es inducida por sequía y estrés salino, y cuya sobreexpresión confiere mayor tolerancia a estos factores abióticos (Tamura *et al.*, 2003).

Complejos eventos conectan la percepción del estrés hídrico con los posteriores cambios regulatorios en la expresión génica, estos incluyen la acumulación de ácido abscísico (ABA), la redistribución del calcio intercelular, la señalización de fosfolípidos, y la fosforilación de proteínas. El calcio funciona como un segundo mensajero en plantas para acoplar el estímulo extracelular a respuestas intracelulares; la fosforilación y defosforilación de proteínas son mecanismos generales de integración de señales y regulación de muchas rutas en plantas; moléculas de señalización como la espermidina y el óxido nítrico cuya acumulación está relacionada con la sobre expresión de proteínas de respuesta a sequía (Bartels *et al.*, 2007).

El ácido abscísico (ABA) es reconocido como la hormona vegetal más importante en la integración de cambios ambientales de la disponibilidad de agua y las respuestas adaptativas en plantas (Raghavendra *et al.*, 2010). Estudios experimentales evidencian la existencia de sistemas regulatorios que gobiernan la expresión de genes inducibles por sequía y estrés salino que son dependientes e independientes de la presencia de ABA (Shinozaki y Yamaguchi-Shinozaki, 2007). La sequía desencadena la producción de ABA que es sintetizado por una enzima inducida en células radicales o en células del parénquima a partir de un carotenoide. El ABA sintetizado en las raíces entra al xilema en forma libre o en forma conjugada con glucosa y desde ahí es transportado a las hojas, causando el cierre de los estomas e induciendo la expresión de genes relacionados con la sequía (Yokota *et al.*, 2006).

Los productos de los genes inducibles por sequía, pueden clasificarse en dos. El primero comprende factores de transcripción, kinasas, fosfatasa y otras moléculas de señalización. El segundo grupo comprende proteínas funcionales de resistencia a sequía e incluyen las proteínas de embriogénesis tardía (LEA),

osmotina, proteínas anticongelantes, enzimas claves para la biosíntesis de osmolitos, proteínas de canales de agua, transportadores de prolina y varias proteasas (Shinozaki & Yamaguchi-Shinozaki, 2007).

Estudios con microarreglos en tomate (*Solanum pennellii*), un cultivo muy relacionado a la papa, permitieron identificar cerca de 400 genes candidatos de respuesta al estrés por sequía, entre estos, factores de transcripción, proteínas de señalización, genes involucrados con la estructura de la pared celular, biosíntesis de cera, y peso de la planta, además enzimas claves de la gluconeogénesis, de la biosíntesis de nucleótidos, degradación de triptófano, almidón y la remoción de radicales superóxido (Gong *et al.*, 2010).

Reportes anteriores con accesiones de *Solanum tuberosum andigena* han revelado que genotipos de papa mejor adaptadas a la sequía tienen una expresión constitutiva de genes codificantes de la ruta de flavonoides (Watkinson *et al.*, 2006), mayores niveles de expresión de proteínas de choque térmico (HSP), de la proteínas fosfatasa 2C y de los factores de transcripción DREB (Vasquez-Robinet *et al.*, 2008; Schafleitner *et al.*, 2007b). Evers *et al.* (2010), empleando microarreglos de cDNA, encontraron que variedades tolerantes a sequía, reprimen fuertemente la expresión de genes relacionados a la fotosíntesis, y al metabolismo de carbohidratos. También se ha identificado un factor de transcripción MYB tipo R1 (StMYB1R-1) cuyos niveles de expresión refuerza la respuesta en papa frente a la sequía; plantas transgénicas para este gen tienen menor tasa de pérdida de agua, y un cierre de estomas mucho más rápido frente a la sequía (Shin *et al.*, 2011)

Pocos son los estudios de expresión génica de papa en respuesta a sequía, los cuales han empleado principalmente plataformas de microarreglos de cDNA, que tienen la desventaja de estar limitados a la evaluación de secuencias ya conocidas, por lo que no permiten estudiar completamente los transcriptomas. La última versión de microarreglos de papa “TIGR Potato cDNA Microarray 10K.v4” contiene solo 15,264 cDNAs (TIGR Solanaceae Genomics Resource, 2014) mientras que el secuenciamiento del genoma de papa ha revelado la

existencia de 39,031 genes codificantes de proteínas (PGSC, 2011). Los avances en la tecnología del secuenciamiento han mejorado los métodos para el estudio de los transcriptomas mediante el secuenciamiento del mRNA. De acá la necesidad de emplear estas nuevas tecnologías que permitan secuenciar y conocer todo el transcriptoma de este importante cultivo, para futuras aplicaciones en su mejoramiento.

En este trabajo de investigación se buscó caracterizar el transcriptoma de papa en respuesta a sequía empleando la tecnología de secuenciamiento de RNA (RNA-Seq). Esta técnica es una herramienta revolucionaria de la transcriptómica que permite mapear transcritos y cuantificarlos de manera precisa, presentando muchas ventajas frente a otros enfoques de análisis transcriptómico. El RNA-Seq no está limitado a la detección de transcritos conocidos, pues es posible la reconstrucción completa de estos y de sus niveles de expresión en diversos tipos celulares bajo cualquier condición o tratamiento y de cualquier especie, no tiene un límite superior de cuantificación y tiene altos niveles de reproducibilidad (Mortazavi *et al.*, 2008; Morozova & Marra, 2008; Wang *et al.*, 2009; Filichkin *et al.*, 2010).

Mediante el secuenciamiento de los transcriptomas de las variedades tolerante y susceptible de *Solanum tuberosum* subsp. *andígena* y el uso de varias herramientas bioinformáticas, se buscó caracterizar la expresión diferencial entre ellas y entre tejidos (raíz y hoja) durante las etapas del estrés hídrico (respuesta temprana, tardía y recuperación). Además, el trabajo tuvo como objetivo asignar funcionalidad a los genes más inducidos, así como encontrar módulos de genes altamente correlacionados, que nos den una referencia de su participación en conjunto de determinados procesos celulares.

II. ANTECEDENTES

1. La sequía

La sequía es un componente del clima que ocurre en áreas con elevados y bajos niveles de precipitación y tiene origen cuando la cantidad de precipitación está debajo de los niveles normales (Wilhite, 2007).

No tiene definición universal, para lo cual debemos tener en cuenta los tipos de sequía: meteorológica, hidrológica, agrícola y socioeconómica (National Drought Mitigation Center, 2014). Los cuatro tipos de sequía están estrechamente relacionados e interconectados, y todos se originan a partir de niveles menores de precipitación a lo esperado.

- a. Sequía meteorológica: Es específica de una región, y se refiere al grado de desecación en comparación a una cantidad “normal” o promedio de precipitación y a la duración del periodo seco. Todos los tipos de sequía se originan a partir de una deficiencia en las precipitaciones.
- b. Sequía hidrológica: Es el déficit en el suministro de agua de superficies o subsuperficies como ríos, reservorios, lagos y lagunas y aguas subterráneas. La sequía hidrológica se origina después de la ocurrencia de una sequía meteorológica y agrícola.
- c. Sequía agrícola y sus impactos económicos: Es un período sin lluvia que por su duración daña el cultivo y reduce significativamente los rendimientos económicos, ésta comienza cuando se agota el agua disponible del suelo en la zona de la raíz de la planta (Kramer, 1987). La sequía agrícola es muy compleja, ya que sus impactos dependen de la magnitud, duración y frecuencia de la sequía, la cantidad de agua que necesita una planta depende además de las características biológicas de la especie, la etapa de su crecimiento y de las propiedades físicas y biológicas del suelo.

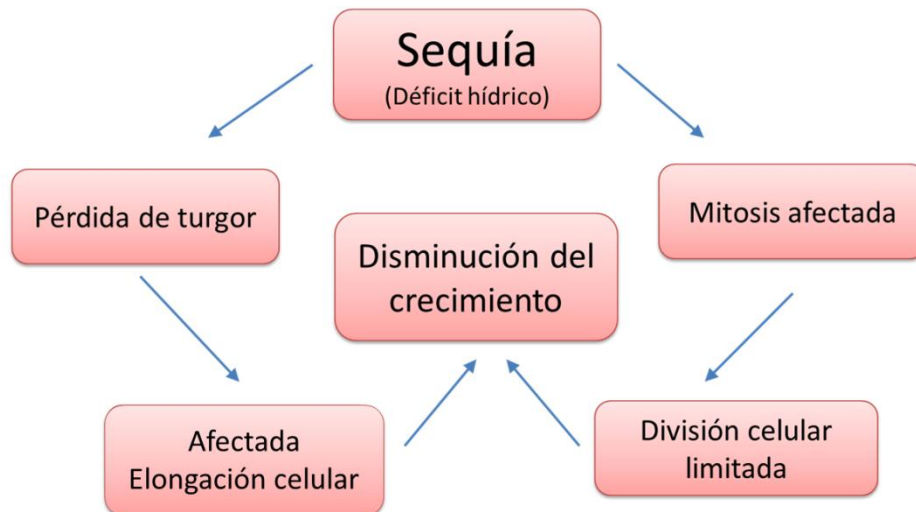
- d. Sequía socioeconómica: Ocurre cuando un periodo extendido de clima seco sin lluvias provoca una demanda que excede el suministro de servicios como agua potable o poder hidroeléctrico, los cuales son dependientes del clima.

2. Efectos de la sequía en las plantas

La sequía es uno de los factores medio ambientales más importantes en la agricultura, siendo la amenaza más crítica para la seguridad alimentaria del mundo. Los efectos de la sequía en las plantas van desde los morfológicos hasta efectos a nivel molecular y son evidentes en todas las etapas del crecimiento de la planta.

a. Crecimiento y rendimiento del cultivo

La división celular, crecimiento y diferenciación celular son procesos involucrados en el desarrollo de las plantas. El crecimiento de las células es uno de los procesos fisiológicos más sensibles a la sequía, debido a la reducción de la presión de turgencia. Además, el déficit hídrico severo puede inhibir la elongación celular por interrupción de flujo de agua desde el xilema hacia las células en elongación del entorno (Nonami, 1998). La expansión foliar es uno de los procesos más sensibles al déficit hídrico, muchos estudios indican que esta sensibilidad se debe al pequeño tamaño de las células y a la reducción en el número de células producidos por los meristemos foliares (Alves & Setter, 2004). Por tanto, bajo condiciones de sequía, la disminución de mitosis, de expansión y elongación celular resultan en una reducción del tamaño de las plantas, del área foliar y del rendimiento del cultivo (Farooq *et al.*, 2009).



(Modificado de Farroq *et al.* 2009)

Figura 1. Posibles mecanismos responsables de la reducción del crecimiento en plantas bajo condiciones de estrés

b. Crecimiento de las raíces

La sequía también afecta el crecimiento y desarrollo de las raíces, los que se ven favorecidos por la inhibición de la expansión foliar, pues una mayor proporción de las sustancias asimiladas por la planta pueden dirigirse hacia el sistema radicular. En este sentido se da un crecimiento preferencial de las raíces hacia las zonas del suelo que aún permanecen húmedas. El crecimiento de raíces profundas en zonas húmedas del suelo es considerado un tipo de defensa de las plantas frente a la sequía (Taiz & Zeiger, 2006).

c. Relaciones hídricas

El contenido relativo de agua, el potencial hídrico de las hojas, la resistencia estomática, la tasa de transpiración, la temperatura de la hoja son importantes características que influyen en el estado hídrico de la planta. La exposición de una planta a la sequía disminuye significativamente el potencial

hídrico de las hojas, el contenido relativo de agua en sus hojas, la tasa de transpiración y un subsecuente incremento de la temperatura de la planta.

La eficiencia del uso de agua (relación entre la materia seca producida y la cantidad de agua consumida) de las plantas es afectada por la sequía. Las plantas estresadas tienen una mayor eficiencia del uso de agua, debido en gran parte a que disminuyen su conductancia estomática y por tanto la transpiración; además, disminuyen su área foliar y de manera relativa su rendimiento (Farooq *et al.*, 2009).

d. Fotosíntesis

- Limitaciones estomáticas: El cierre de estomas es una respuesta temprana de las plantas frente a la sequía, para evitar una gran pérdida de agua que podría resultar en una deshidratación celular, cavitación del xilema y muerte (Chaves *et al.*, 2003). El cierre de los estomas es una de las principales causas de la disminución de la tasa fotosintética en plantas bajo estas condiciones, pues disminuye la concentración interna de CO₂ en el mesófilo (Athar & Ashraf, 2005).

- Limitaciones no estomáticas: El estrés por sequía produce cambios en los componentes y pigmentos fotosintéticos, dañando el aparato de fotosíntesis y disminuyendo la actividad de las enzimas del ciclo de Calvin, causas importantes del reducido rendimiento de los cultivos. En las plantas superiores, la tasa fotosintética depende de la actividad de la enzima Rubisco; diversos estudios han demostrado la pérdida de la actividad de la enzima bajo condiciones de sequía, la cantidad y actividad de esta enzima limitan la capacidad fotosintética al controlar la asimilación de carbono (Reddy *et al.*, 2004). Otro efecto de la sequía que inhibe el crecimiento y capacidad fotosintética de las plantas es el desbalance en la producción de especies reactivas de oxígeno y la defensa antioxidante, lo cual termina en la inducción de estrés oxidativo en proteínas, lípidos de membrana y otros componentes celulares (Farooq, 2009).

3. La resistencia a la sequía

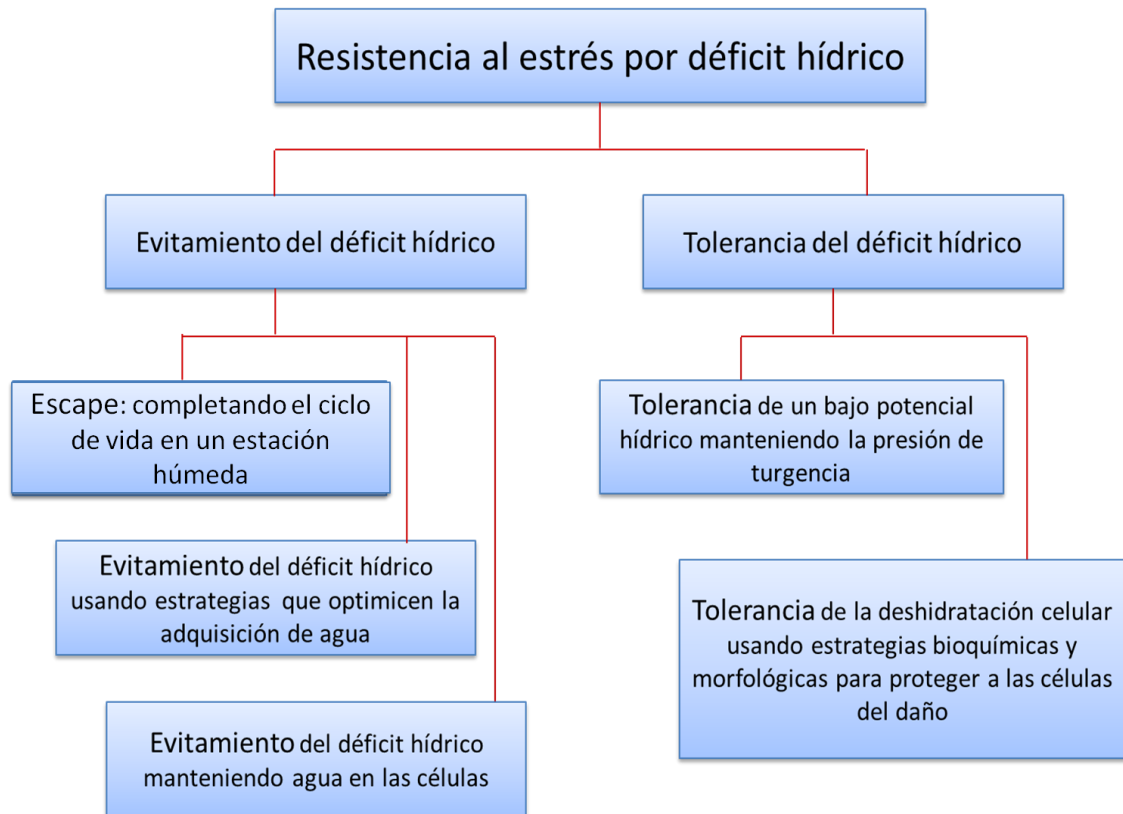
Las plantas tienen la capacidad de responder, adaptarse y sobrevivir bajo condiciones de sequía, gracias a la inducción de varias respuestas de tipo morfológicas, bioquímicas y fisiológicas. Se dice que un genotipo es resistente a la sequía cuando produce un cultivo rentable dentro de los límites de su potencial de producción a pesar de la disponibilidad limitada de agua (Ekanayake, 1993).

La resistencia a la sequía se origina a partir de la capacidad de las plantas para tolerar o evitar el déficit hídrico (Bray, 2007), de acá se originan los dos componentes de la resistencia a la sequía: la tolerancia y el evitamiento. Algunos autores consideran como otro mecanismo a la recuperación siendo esta la capacidad de las plantas para restablecer su aparato fotosintético y sus principales funciones vitales, bajo irrigación o lluvia luego de un periodo de sequía (Ekanayake, 1989).

a. Evitamiento

El evitamiento de la deshidratación es definido como la capacidad de la planta para mantener relativamente alto el nivel de hidratación bajo condiciones de déficit hídrico del suelo o de la atmósfera. En este sentido, si la planta retiene un alto nivel de hidratación en sus tejidos, los procesos fisiológicos bioquímicos y metabólicos involucrados en el crecimiento y rendimiento no estarán expuestos al déficit hídrico.

Un componente del evitamiento es el escape, el cual permite a la planta completar su ciclo de vida durante el periodo de suficiente suministro de agua antes del inicio de la sequía. Por otra parte, las plantas consiguen evitar el estrés hídrico con una eficiente absorción de agua por parte de las raíces o por la reducción de la evapotranspiración de las partes aéreas de la planta.



(Modificado de Bray, 2007)

Figura 2. Componentes de la resistencia a la sequía en plantas.

b. Tolerancia

La tolerancia es la capacidad de la planta, como un todo o de alguno de sus componentes de funcionar a un bajo potencial hídrico. Ésta es una propiedad inherente a los componentes celulares y debe (a) limitar el daño en la planta hasta un nivel reparable, (b) mantener la integridad celular durante la deshidratación, y (c) luego de la rehidratación, se deben movilizar rápidamente mecanismos que resuelvan el daño sufrido durante la deshidratación (Oliver *et al.*, 2007).

4. Transducción de señales y expresión génica en respuesta a la sequía

La respuesta de la planta a la sequía es dependiente de la tasa de pérdida de agua. Una menor tasa de pérdida de agua puede permitir la aclimatación de la planta al déficit hídrico y limitar la extensión del daño, mientras que una tasa rápida podría evitar la aclimatación (Bray, 1997).

a. El ácido abscísico (ABA)

La importancia del ABA en la respuesta de las planta a múltiples estreses ambientales es indiscutible. Las señales de déficit hídrico liberadas por las raíces son llevadas hacia las hojas a través de una o más rutas de señalización, siendo la molécula señal más importante el ácido abscísico (Raghavendra *et al.* 2010). Frente a la pérdida de agua de los suelos, los tejidos de las raíces sintetizan ABA, a partir de un precursor carotenoide, y es transferido hacia las hojas a través de los tejidos vasculares, provocando el cierre de estomas, lo cual reduce la pérdida de agua por transpiración, y la inducción de muchos genes de respuesta a sequía (Schachtman & Goodger, 2008). Es tal la importancia del ABA en la respuesta génica de las plantas, que los sistemas regulatorios que gobiernan la expresión génica inducible por sequía han sido clasificados como dependientes o independientes de ABA.

b. Percepción celular del déficit hídrico

Las células son capaces de percibir alteraciones y de regular su contenido hídrico, de tal forma que la célula puede reconocer la pérdida de agua y transducir una condición física en una respuesta bioquímica. Este proceso, conocido como “*osmosensing*” o mecanotransducción ha sido muy bien estudiado en sistemas modelo como la levadura y *E. Coli*, pero en plantas aún queda mucho por ser elucidado (Bray, 2007).

La ausencia de precipitaciones en los ambientes naturales causa la desecación de la atmósfera y del suelo debido a la evaporación del agua de la

superficie del suelo durante el día. Las plantas, como otros organismos, deben de percibir estas señales ambientales vía receptores específicos, que luego deben desencadenar una cascada de eventos que lleven a la modificación de la actividad celular y metabólica, incluyendo la regulación de la expresión de determinados genes.

Es así que las plantas necesitan de sistemas adecuados tanto en hojas como en raíces para percibir el déficit hídrico (Yokota *et al.*, 2006). La molécula con el mayor potencial para ser un osmosensor es la histidina kinasa ATHK1, identificada en *Arabidopsis*, cuyo rol primario es monitorear fluctuaciones en la osmolaridad externa e iniciar una cascada de señales para una osmo-adaptación (Reisser *et al.*, 2003), ATHK1 es un homólogo del osmosensor SLN1 en levaduras y es sobreexpresado en raíces por cambios de osmolaridad (Urao *et al.*, 1999). Otro posible osmosensor es NtC7 de *Nicotiana tabacum*, una proteína de membrana tipo receptor cuya expresión es inducida por sequía y estrés salino, y cuya sobreexpresión confiere mayor tolerancia a estos estreses (Tamura *et al.*, 2003).

c. Segundos mensajeros y moléculas señal

Luego de la percepción del estrés, tienen lugar una compleja cascada de eventos que permiten los cambios en la expresión génica. Estos eventos incluyen, la acumulación de ABA, la redistribución intracelular de calcio, la señalización por fosfolípidos y la fosforilación de proteínas. Se ha observado en *Arabidopsis* un rápido incremento de calcio libre en el citosol durante el estrés por sequía y salinidad. El calcio funciona como un segundo mensajero en las plantas para acoplar el estímulo extracelular y las respuestas intracelulares (Knight *et al.*, 1997). En las plantas se han caracterizado tres clases de sensores de calcio: proteínas calmodulina (CaM), proteínas relacionadas a calmodulina, proteínas tipo calcineurina B (CBL), y proteínas kinasas dependientes de calcio (CDPKs) (Knight & Knight. 2001).

La fosforilación y defosforilación de proteínas son mecanismos generales de la integración de señales y de rutas de regulación en las plantas. Las cascadas de proteínas kinasa activadas por mitógeno (MAPK) están involucradas en la respuesta al estrés osmótico (Sinha *et al.*, 2011). En *Arabidopsis* se ha encontrado una MAPK fosfatasa 1 que incrementa la tolerancia a la salinidad, lo cual sugiere la importancia de la actividad fosfatasa de estas proteínas en la integración de las respuestas a los desafíos ambientales. La fosforilación no solo tiene un rol en los procesos de señalización, sino que también es importante para la función de protección de las proteínas abundantes de embriogénesis tardía (LEA).

La activación de la señalización por fosfolípidos en respuesta a una variedad de estrés así como también en muchos procesos de desarrollo de las plantas está muy bien documentada. La actividad de las dos fosfolipasas C y D generan ácido fosfatídico, el cual funciona como un segundo mensajero para activar blancos corriente abajo (Bartels & Sukar, 2005).

Entre otras moléculas señal, además del importante ABA, existe la poliamina espermidina, cuya sobreexpresión en hojas de *Arabidopsis* refuerza la tolerancia a la sequía y al estrés salino. Además, otros estudios mencionan al óxido nítrico como una molécula de señalización durante la respuesta adaptativa de las plantas al estrés por sequía (García-Mata & Lamattina, 2001).

d. Control transcripcional

Uno de los efectos de la transducción de señales desencadenada por sequía es la activación de los factores de transcripción, cada uno de los cuales puede activar o no a un conjunto de genes blanco requeridos para la respuesta de la planta. Como se ha mencionado, los genes de respuesta a sequía pueden ser divididos en dos grupos: los genes dependientes de ABA, y los genes independientes de ABA para su inducción (Shinozaki & Yamaguchi- Shinoshaki, 2007). Los elementos más estudiados y más distribuidos son los elementos de respuesta a ABA (ABREs) que son dependientes de ABA, y los elementos de

respuesta a la deshidratación (DREs) que no dependen de ABA para su inducción (Yamaguchi-Shinozaki & Shinozaki, 2005).

En *Arabidopsis* los genes DREB1 y DREB2 codifican proteínas estructuralmente diferentes y son inducidos específicamente por baja temperatura y por sal o sequía, respectivamente. DREB2A y DREB2B son sintetizados en raíces solo en respuesta a salinidad, pero son producidas en tallo y raíces luego de un tratamiento de sequía (Knight & Knight, 2001).

Por otro lado, ABRE es el elemento en *cis* más importante del promotor de muchos genes de respuesta a ABA. Este elemento fue primer identificado en el gen *Em* en el trigo, y desde entonces en otras especies como el maíz, arroz, tabaco y *Arabidopsis*. La expresión de los genes de respuesta a ABA es activada por la unión de un factor de transcripción bZIP a la secuencia ABRE, otra familia de factores de transcripción dependientes de ABA, son los factores MYC y MYB encontrados (Nakashima *et al.*, 2014).

e. Proteínas Funcionales

El último paso durante la cascada de señalización es la activación de la diversidad de genes que codifican proteínas funcionales como enzimas de detoxificación, de biosíntesis de osmolitos como prolina, betaina y azúcares, chaperonas, proteínas LEA, proteasas para la renovación de proteínas, transportadores y canales de agua para el movimiento de agua a través de las membranas celulares y enzimas relacionadas al daño oxidativo (Shinozaki & Yamaguchi-Shinozaki, 1997).

- Enzimas necesarias para la biosíntesis de osmolitos

Las plantas expuestas a estrés osmótico como resultado de la sequía acumulan solutos conocidos como osmolitos para mantener el turgor en las células. Estos osmoprotectores son confinados en el citoplasma. En respuesta al bajo potencial hídrico de las células, se da un ajuste osmótico que involucra la acumulación de solventes orgánicos e inorgánicos. Entre estos osmolitos se tienen prolina,

betaina, dimetil-sulfoniopropionato, manitol, sorbitol, pinitol, trealosa y fructanos (Serraj & Sinclair, 2002).

- Proteínas de embriogénesis tardía

Las proteínas LEA son consideradas importantes para la tolerancia a la sequía de las semillas, y de tejidos vegetativos, conociéndose al menos cinco grupos. Se desconocen las bases de la función protectora de las proteínas LEA, sin embargo varios estudios sugieren que estas proteínas podrían proteger a las células de la sequía y de un estrés por frío actuando como amortiguadores, secuestrando iones, estabilizando proteínas, membranas y la estructura de la cromatina o renaturando proteínas desplegadas (Ingram & Bartels, 1997).

- Secuestro de especies reactivas de oxígeno (ROS)

Una de las consecuencias de muchos estreses, incluida la sequía, es el incremento en la concentración de ROS como el oxígeno libre, radicales hidroxilo, peróxido de hidrógeno y aniones superóxido, los cuales causan daños irreversibles en las membranas, proteínas, DNA y RNA. La generación de antioxidantes y el establecimiento de un sistema de secuestro de ROS que incluye a las enzimas superóxido dismutasa y la catalasa enfrentan la actividad y el daño de las ROS en las plantas (Xiong & Zhu, 2002). La tolerancia de las plantas a la sequía, muchas veces depende de la eficiencia de estos sistemas.

- Aquaporinas

Son proteínas localizadas en la membrana plasmática y en la membrana vacuolar de las células vegetales, pertenecen a la familia de proteínas intrínsecas mayores y facilitan el flujo de agua a través de las membranas celulares, siguiendo gradientes de presión osmótica o hidrostática. La importancia biológica de las aquaporinas para las plantas es su capacidad para modular el transporte transmembrana de agua en situaciones como la sequía o salinidad en las que el flujo de agua es fisiológicamente crítico (Alexandersson *et al.*, 2005).

5. La Papa

La papa es una planta herbácea que produce tubérculos como parte de su sistema de tallos, siendo éstos los principales órganos de almacenamiento. Pertenece a la gran familia de las Solanaceas que incluye más de 3000 especies. El género *Solanum* (dividido en siete subgéneros) es uno de los más importantes económicamente dentro de la familia, e incluye además de papa (*Solanum tuberosum* L.), tomate (*S. lycopersicum*), berenjena (*S. melongena*), ajíes, pimientos y otras frutas menores como el pepino (*S. muricatum*) y el aguaymanto. Dentro de este género, las papas silvestres y cultivadas están incluidas en el subgénero Potatoe - Sección Petota que comprende aproximadamente 188 especies silvestres y siete especies cultivadas (Spooner & Salas, 2006).

La mayor parte de las especies de papas silvestres, son raras, endémicas y están distribuidas en 16 países, desde el sur oeste de los Estados Unidos, hasta la costa central de Chile. Las papas cultivadas nativas crecen principalmente en los Andes, desde Venezuela hasta Argentina y en la parte central de Chile. Entre las especies silvestres y cultivadas, se encuentran marcadas diferencias morfológicas, así como dentro las especies cultivadas que pueden ser afectada por factores ambientales como la temperatura, el fotoperiodo, la humedad y la fertilidad del suelo (Huamán, 1986).

a. Las papas cultivadas

La papa cultivada tiene sus orígenes en los andes centrales de Perú y Bolivia, donde comenzó la domesticación de las papas silvestres. Se cree que tiene un origen único a partir de una especie silvestre del complejo *S. brevicaulle* (Spooner & Hetterscheid, 2005). Su expansión como un cultivo de importancia alimenticia mundial comenzó durante la conquista española, quienes viendo el gran potencial que tenía, la introdujeron a Europa.

Luego de su introducción a Europa, la papa inicialmente fue una curiosidad botánica estudiada en invernaderos con intereses y propósitos médicos. Su potencial como cultivo alimenticio fue por primera vez visto en Irlanda, a finales del siglo XVII y durante el siglo XVIII, país en el cual representó un cultivo de mucha importancia alimenticia. Su cultivo se expandió en todo el mundo durante el siglo XIX, en China y la India durante la segunda mitad del siglo XX (Bradshaw & Bonierbale, 2010).

Entre las papas cultivadas, según Hawkes (1990), existen siete especies y siete subespecies siendo: *S. ajanhuiri*, *S. chaucha*, *S. curtilobum*, *S. juzepczukii*, *S. phureja* subsp. *phureja*, *S. phureja* subsp. *estradae*, *S. phureja* subsp. *hygrothermicum*, *S. stenotomum* subsp. *stenotomum*, *S. stenotomum* subsp. *goniocalyx*, *S. tuberosum* subsp. *andigena*, y *S. tuberosum* subsp. *tuberosum*; sin embargo esta taxonomía no es universalmente aceptada. En el 2002, Huamán & Spooner, estudiaron las diferencias morfológicas entre las papas cultivadas nativas, y sus resultados los llevaron a reconocer a todas las papas cultivadas como una sola especie *Solanum tuberosum*, con ocho grupos: Grupo Ajanhuiri, Andigenum, Chaucha, Chilotanum, Curtilobum, Juzepczukii, Phureja, y Stenotomum.

b. La especie *Solanum tuberosum*

- Posición taxonómica

Reino: Plantae

División: Magnoliophyta

Clase: Magnoliopsida

Subclase: Asteridae

Orden: Solanales

Familia: Solanaceae

Género: *Solanum*

Sección: Petota

Especie: *Solanum tuberosum* (Hawkes, 1990)

- Característica e importancia

Solanum tuberosum es una especie cultivada tetraploide, con número cromosómico de $2n=4x=48$. Es cultivada en todo el mundo y es la especie de mayor importancia económica dentro de las papas cultivadas. Se divide en dos subespecies, *S. tuberosum* subsp. *tuberosum* que hoy se cultiva en todo el mundo, está adaptada a días más prolongados y es altamente susceptible a la sequía, y *S. tuberosum* subsp. *andigena*, adaptada a las condiciones de días cortos, es cultivada principalmente en los andes, desde Venezuela hasta el norte de Argentina a altitudes tan altas como los 3500 msnm y está por tanto muy bien adaptada a condiciones adversas como la sequía.

6. La papa y la sequía

Comparada con otras especies, la papa es una especie sensible a la sequía. Para prevenir pérdidas en el rendimiento de la papa, su frecuencia de irrigación debe ser mayor a otros cultivos como el tomate, maíz o caña de azúcar (van Loon, 1981). La sequía influye en cada una de las etapas del cultivo y afecta el desarrollo y el crecimiento de las hojas, raíces y tubérculos (Deblonde & Ledent, 2000), la vulnerabilidad que tiene la papa al déficit hídrico durante todo su ciclo de crecimiento tiene como resultados bajos rendimientos y tubérculos de mala calidad.

a. Fisiología de la papa frente a la sequía

La sequía afecta directamente el rendimiento del cultivo de papa, debido a la reducción de la respiración y de la fotosíntesis, e indirectamente por la evaporación de agua del suelo y por la transpiración de las hojas. Los efectos de la sequía sobre su crecimiento y producción, se deben a la reducción en la producción de foliaje (menor emergencia y menor expansión foliar), de la tasa de fotosíntesis por unidad de área y al acortamiento del periodo vegetativo o una temprana senescencia (Ekanake, 1989).

- **Crecimiento y Rendimiento**

La sequía afecta inicialmente el área foliar, lo que como consecuencia afecta la altura de los tallos y la cobertura de la planta. Bansal & Nagarajan (1986) demostraron que el crecimiento de hojas de papa es menor en plantas estresadas que en plantas control, además la evaluación de varios genotipos indica que estos responden de manera diferente a la sequía y a la recuperación. Los investigadores encontraron plantas que presentan una mínima reducción del área foliar durante el estrés y que durante la recuperación el área foliar algunas veces supera a las plantas control, y plantas con grandes reducciones del área foliar durante el estrés y que no logran recuperarse al ser irrigadas.

El efecto negativo de la sequía sobre el rendimiento de tubérculos se debe parcialmente a la reducción del potencial de producción diaria de tubérculos. Esta reducción puede ser provocada, primero por la reducción del área foliar de modo que no hay cobertura total del suelo durante el periodo de tuberización, y segundo por la reducción de la tasa fotosintética por área foliar (van Loon, 1981). La sequía puede provocar pérdidas en el rendimiento de tubérculos que van del 25 al 69%, siendo la extensión de esta pérdida dependiente del genotipo (Schafleitner *et al.*, 2007b). La sequía también afecta a la calidad de los tubérculos, formándose alargados y deformados, debido a la alteración de su maduración y crecimiento.

- **Fotosíntesis**

Varios estudios indican los efectos negativos de la sequía sobre la tasa fotosintética en papa (van Loon, 1986, Schapendonk *et al.*, 1989). El déficit hídrico reduce la tasa fotosintética indirectamente por el cierre de estomas y directamente por la reducción de la capacidad fotosintética de las hojas. Además Schapendonk *et al.* (1989) mostraron que el déficit hídrico en papa inhibía el Ciclo de Calvin pero no la tasa de transporte de electrones. En esta misma investigación también se demostró la asociación de variación genética y la respuesta de variedades de papa a la sequía, encontrando diferencias significativas en la tasa de fotosíntesis, asociadas con diferencias en la conductancia estomática y en la transpiración.

- **Conductancia estomática**

Levy (1983) sugiere que, en papa, el ajuste osmótico es una importante característica relacionada al mantenimiento del turgor de las hojas, y que esta característica debe ser empleada para la evaluación de la tolerancia a sequía de cultivares de papa. Sin embargo, otro estudio muestra que la capacidad de ajuste osmótico es pequeña en comparación a otros cultivos como el algodón, arroz o maíz, y sugiere que otros mecanismos deber estar más involucrados en la resistencia a la sequía (Jefferies, 1992).

b. Expresión génica de plantas de papa en respuesta a la sequía

Trabajos anteriores han empleado bibliotecas específicas para estreses abióticos para identificar genes relacionados a la sequía en plantas de papa. Rensik *et al.* (2005) con la finalidad de entender la respuestas de la papa a estreses abióticos entre ellos la sequía, generaron 20756 ESTs (*“Expressed sequence tags”*). El análisis de estas secuencias reveló una similaridad con genes conocidos e involucrados en respuestas a estreses abióticos de otras especies, y también se identificaron genes con función aun desconocida, generando nuevos genes candidatos de papa asociados a estrés.

Dos años después, Schafleitner *et al.* (2007a) identificaron genes putativos de tolerancia a sequía en hojas de plantas de papa bajo este estrés. Entre ellos encontraron la inducción de los genes: LEA5 de embriogénesis tardía, el gen de una dehidrina, un factor de transcripción (AtHB-12), y un gen de biosíntesis de prolina. Reportaron además que la expresión de la proteína fosfatasa 2C, bajo condiciones de sequía, está asociada positivamente con el mantenimiento del rendimiento del cultivo, y que los cultivares tolerantes expresan el factor de transcripción DREB en mayores niveles que los cultivares susceptibles.

Un año más tarde, se estudiaron 2 genotipos de la subespecie *Solanum tuberosum* subsp. *andigena*, un genotipo Sullu más resistente a la sequía que Negra Ojosa. Mediante un análisis de microarreglos e información metabólica de hojas bajo condiciones de sequía, se encontró en Sullu, una mayor actividad metabólica mitocondrial, mayor inducción de antioxidantes localizados en cloroplastos y de genes codificantes de chaperonas; mientras que en Negra Ojosa fueron más inducidos factores de transcripción de respuesta a ácido abscísico como WRKY1 (Vasquez-Robinet *et al.*, 2008).

En el 2010, Evers *et al.* determinaron que los osmolitos prolina, análogos de prolina, poliaminas, inositol, galactinol y galactosa son compuestos de respuesta a sequía y están relacionados con la tolerancia a la sequía. Además se encontró una distorsión en la expresión de genes involucrados en la fotosíntesis y metabolismo de carbohidratos en plantas expuestas a sequía.

Massa *et al.* 2013 secuenciaron el transcriptoma de la respuesta *in vitro* a estrés biótico y abiótico de *Solanum tuberosum* Grupo Phureja DM1-3 516 R44, encontrando que 37% de genes mostraron una expresión diferencial en comparación a su control. Los genes con la mayor expresión diferencial que encontraron fueron aquellos inducibles por estrés o los involucrados en la regulación del estrés.

7. RNA-Seq: Secuenciamiento masivo de lecturas de RNA

Entender la complejidad de los procesos y funciones biológicas en cualquier organismo requiere una identificación de todos los genes, sus isoformas alternativas y la variación de sus perfiles de expresión entre tipos celulares, tejidos y órganos, frente a determinada condición externa. El estudio del transcriptoma completo de algún organismo bajo una condición o tratamiento era inimaginable antes de la era de los microarreglos y de las técnicas de secuenciamiento masivo de nueva generación; de hecho, el uso de estas nuevas tecnologías como el secuenciamiento de RNA (RNA-Seq) ha revolucionado la biología a través de la generación de millones y billones de lecturas a bajo costo.

El secuenciamiento de RNA (RNA-Seq) se refiere a un procedimiento experimental que genera lecturas de DNA a partir del RNA (transcriptos) de un organismo. Esta tecnología ha emergido como una herramienta poderosa para el estudio de transcriptomas, pues es posible la reconstrucción completa de los transcriptos y de sus niveles de expresión de diversos tipos celulares bajo cualquier condición o tratamiento y de cualquier especie (Wang *et al.*, 2009).

a. Metodología

A la fecha, se cuenta con varias tecnologías de secuenciamiento masivo de moléculas de DNA. Actualmente, los sistemas que dominan el campo son 454 GS-FLX de Roche Applied Science, AB Solid de Applied Biosystem, Pacific Biosciences de Helicos y Genome Analyzer II de Illumina (Bentley *et al.*, 2008). Estas diferentes tecnologías requieren de protocolos experimentales distintos, pero en esencia muy similares. De los sistemas antes mencionados, los equipos de Illumina son los más empleados, y la metodología incluye las siguientes etapas:

- a. Enriquecimiento de RNA informativo: Un experimento típico de RNA-Seq, comienza con la purificación de una fracción particular del RNA total, el RNA mensajero (mRNA). Este enriquecimiento es posible mediante una selección de las moléculas con una cola poli-A, mediante el empleo de perlas magnéticas de oligo poli-T, además de la eliminación del RNA ribosomal.
- b. Fragmentación del RNA.: Luego de la purificación, el RNA es cortado en pequeños fragmentos mediante una hidrólisis química o mediante el uso de una RNase III.
- c. Síntesis del cDNA de doble cadena: Los fragmentos de RNA son luego convertidos en DNA empleando una transcriptasa reversa que requiere la hibridización de iniciadores al azar en la cadena de RNA.
- d. Ligación de adaptadores: Los extremos salientes 3' de los cDNA son convertidos en extremos romos mediante el uso de enzimas. Se agrega una

base A a los extremos 3' modificados, preparando los fragmentos de cDNA para la ligación de los adaptadores que contienen una única T en el extremo saliente 3'.

- e. Selección de tamaño y amplificación por PCR: En el paso de la fragmentación, las moléculas de RNA son cortadas en diferentes tamaños. Para asegurar un tamaño uniforme se realiza una purificación por filtración en gel. Luego de esta selección se usan nuevos iniciadores para amplificar los fragmentos seleccionados.
- f. Secuenciamiento por síntesis: Inicia con la hibridización de iniciadores a cada molécula de DNA, la cual es luego complementada usando nucleótidos marcados por fluorescencia. El resultado de este proceso es una secuencia de imágenes, una para cada nucleótido que es incorporado. La combinación de esta secuencia de imágenes permite obtener la secuencia de nucleótidos para cada molécula. Finalmente, esta información es almacenada como un archivo de texto FASTQ, el cual contiene una identificación única para cada lectura, su secuencia de nucleótidos, y los valores de calidad por base nucleotídica.

Como se ha visto, con estas nuevas tecnologías de secuenciamiento se genera una enorme cantidad de información de lecturas, y para estudiarlas se requiere de eficientes herramientas bioinformáticas. Las etapas principales del análisis bioinformático son: a) Control de calidad de las secuencias, b) mapeo de las lecturas y c) cuantificación de la expresión, las cuales serán desarrolladas en esta tesis.

b. Estudios de estrés en plantas empleando RNA-Seq

A la fecha, pocos son los estudios que han empleado RNA-Seq para elucidar el transcriptoma de plantas en respuesta a la sequía u otro tipo de estrés. En el 2010, Mizuno *et al.* secuenciaron el transcriptoma de arroz en respuesta a estrés salino, mapearon lecturas de 36 pares de bases a su genoma de referencia y encontraron cerca a 50,000 transcritos expresados en condiciones de estrés.

Un año más tarde, Villar *et al.* (2011) secuenciaron el transcriptoma de ápices de eucalipto en respuesta a déficit hídrico, generaron 1.14 millones de lecturas y encontraron que el genotipo más productivo en condiciones de estrés mostró un mayor número de genes de respuesta al estrés hídrico. Por su parte, Kakumanu *et al.* (2012) estudiando el transcriptoma de maíz en respuesta a la sequía, encontraron una masiva disminución de la abundancia de transcritos de división celular y de ciclo celular, y la presencia de varios transcritos relacionados al metabolismo de carbohidratos y almidón.

En papa, solo se han publicado dos estudios transcriptómicos empleando RNA-Seq. En el 2011, Massa *et al.* secuenciaron 32 transcriptomas de varios tejidos y de plántulas de *Solanum phureja* bajo tratamientos abióticos y bióticos *in vitro* como sal, manitol, calor, infección con *Phytophthora infestans*, entre otros. El análisis de más de 550 millones de lecturas, les permitió identificar y cuantificar los niveles de expresión de más de 22,000 genes.

Posteriormente, Massa *et al.* 2013 secuenciaron el transcriptoma de la respuesta *in vitro* a estrés biótico y abiótico de *Solanum tuberosum* Grupo Phureja DM1-3 516 R44, encontrando que 37% de genes mostraron una expresión diferencial en comparación a su control, y donde los genes con la mayor expresión diferencial fueron aquellos inducibles por estrés o los involucrados en la regulación del estrés.

III. HIPÓTESIS

Existen diferencias en el número y nivel de expresión de genes relacionados al déficit hídrico, dependientes de la intensidad del estrés, del tejido y del nivel de tolerancia de *Solanum tuberosum* subsp. *andigena*.

IV. OBJETIVOS

1. Objetivo general

Caracterizar el transcriptoma de la respuesta a sequía de hojas y raíces de *Solanum tuberosum* subsp. *andigena*.

2. Objetivos específicos

- Reconstruir los transcriptomas de hojas y raíces de dos variedades de *Solanum tuberosum* subsp. *andigena* bajo condiciones de sequía.
- Identificar la expresión génica diferencial en hojas y raíces entre las variedades tolerante y susceptible, durante el tiempo de exposición a sequía y la recuperación de esta.
- Anotar funcionalmente los genes más sobreexpresados e identificar módulos de co-expresión génica.

V. MATERIALES Y MÉTODOS

1. Localización y condiciones climáticas

Se realizó el cultivo de plantas bajo un sistema aeropónico en la Estación Experimental Santa Ana, localizada en la provincia de Huancayo, en la sierra central del Perú (12°00' LS, 75°13' LO) a 3260 msnm, con temperaturas entre 6 y 18 °C.

2. Material vegetal

Se seleccionaron dos variedades de papa nativa, una tolerante (CIP 703671, luego referida como 703671) y otro susceptible a sequía (CIP 703248, luego referida como 703248) utilizada como control (Tabla 1), a partir de una base de datos generada por el CIP (Proyecto Papa Salud), la cual cuenta con un registro de todos los datos fenotípicos de clones nativos y de su respuesta a diferentes estreses bióticos y abióticos, entre ellos la sequía.

Tabla1. Descripción de las variedades de *S. tuberosum* subsp. *andigena* empleadas en este estudio.

Código	CIP 703671	CIP 703248
Respuesta a Sequía	Tolerante	Susceptible
Variedad	Negrita	Wila Huaka Lajra
Género	Solanum	Solanum
Especie	<i>S. tuberosum</i> subsp. <i>andigena</i>	<i>S. tuberosum</i> subsp. <i>andigena</i>
País de Origen	Perú	Bolivia
Altura (msnm)	3100	3800

a. Cultivo de *Solanum tuberosum* subsp. *andigena*

Mantenimiento *in vitro*: plántulas *in vitro* fueron transportadas al laboratorio para realizar su propagación en frascos y tubos, y aumentar el número de plantas por genotipos y posteriormente ser trasladados a un invernadero con condiciones climáticas controladas. Las plantas se propagaron en una cabina de flujo laminar, empleando el medio Murashige & Skoog (SIGMA), pH 5.6, (Tabla 2) y mantenidas en un cuarto de cultivo bajo condiciones controladas de temperatura y fotoperiodo.

Tabla 2. Composición del medio utilizado para la propagación de plántulas de *Solanum tuberosum* subsp. *andigena*.

Componentes	Cantidad
Medio basal Murashige & Skoog	4.3 g
Sucrosa	25.0 g
Phytigel o Agar	3.5 g
Volumen final (Agua destilada)	1 L
Ajustar el pH del medio a 5.6	

Cultivo aeropónico: Todas las plántulas obtenidas fueron transportadas a la Estación Experimental Santa Ana (INIA- Huancayo), localizada en la provincia de Huancayo. Inicialmente las plántulas fueron sembradas en tierra vegetal, sin embargo debido a los objetivos del estudio se cambió el sustrato por uno inerte (arena de cuarzo) que permitiera muestrear las raíces, obteniendo una alta mortalidad de las plantas en este sustrato. Debido a esto se eligió un sistema aeropónico (Figura 3), a fin de muestrear las raíces sin ningún contaminante del suelo y sin ser estresadas mecánicamente.

Para el riego se empleó una solución nutritiva de pH de 6.5 (Tabla 3), que fue aplicada a las raíces cada treinta minutos. Esta solución fue preparada con una concentración de fertilizantes y a un pH de 6.5, como se detalla en Tabla 3.

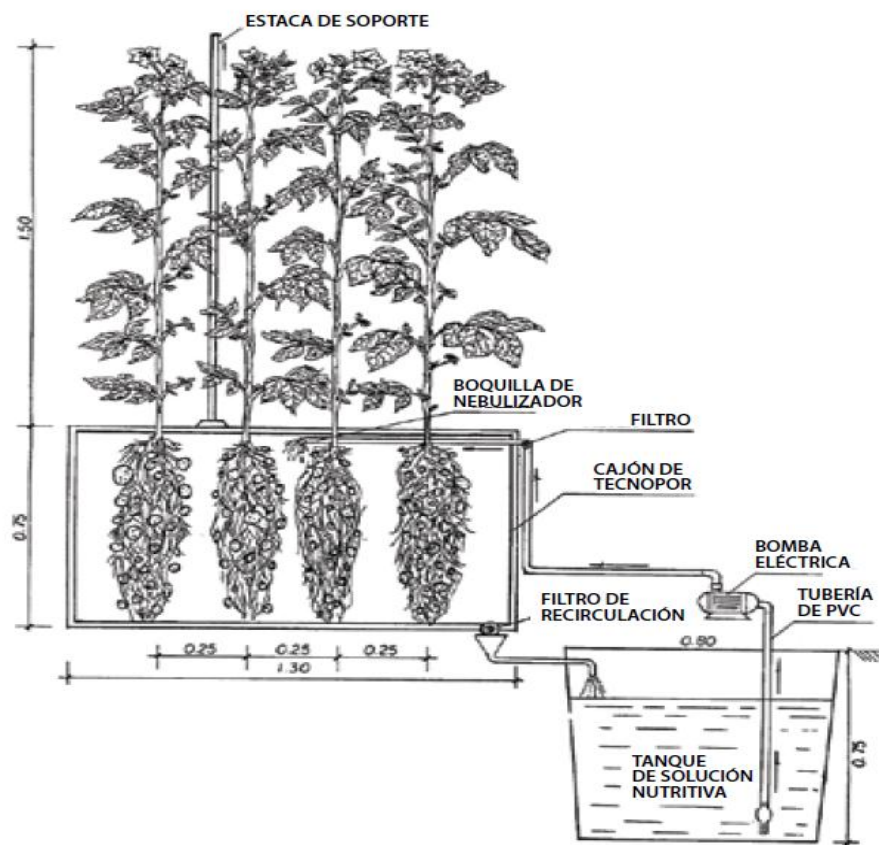


Figura 3. Representación gráfica del sistema aeropónico de cultivo de papa (Otazú, 2010)

Tabla 3. Composición de la solución nutritiva empleada para el cultivo aeropónico (Otazú, 2010).

Fertilizante	Fórmula	g/L
Nitrato de Potasio	KNO_3	0.101
Nitrato de amonio	NH_4NO_3	0.08
Sulfato de Potasio	K_2SO_4	0.087
Sulfato de Magnesio	$MgSO_4$	0.247
Fosfato de Potasio	KH_2PO_4	0.136
Superfosfato triple de Calcio	$Ca(H_2PO_4)_2$	0.117
Fetrilon Combi	Micronutrientes	0.0125

*Fetrilom combi es un nutriente foliar comercial en polvo que tiene la siguiente formulación: 9% MgO, 3% S, 4% Fe, 4% Mn, 1.5%, Cu, 1.5% Zn, 0.5% B, y 0.1% Mo.

3. Inducción del estrés por sequía

Las plantas recibieron un riego normal por aproximadamente 3 meses, tiempo en el que comienza la tuberización, para luego iniciar la inducción del estrés por sequía, lo cual consistió en la ausencia del riego por aspersión. Se consideraron cuatro tiempos de muestreo, de acuerdo a la disminución de la actividad fotosintética en la planta tolerante (Vásquez- Robinet *et al*, 2008).

<u>Tratamiento</u>	<u>Descripción</u>
T0 (Control):	Con Riego
T1 (Respuesta temprana):	Sin riego, la tasa de fotosíntesis de la variedad tolerante disminuye en un 25% respecto al control
T2 (Respuesta tardía):	Sin riego, la tasa de fotosíntesis de la variedad tolerante disminuye en un 50 – 60% respecto al control
T3 (Recuperación):	Reinicio del riego, la tasa de fotosíntesis de la variedad tolerante se recupera hasta un 80% del control.

4. Evaluaciones fisiológicas

Inicialmente, se evaluó la tasa de fotosíntesis en 703671 (variedad tolerante) en función del tiempo y exposición a sequía o riego, con la finalidad de determinar los tiempos de toma de muestra correspondientes a la respuesta temprana, tardía y la recuperación. La evaluación se realizó antes de la inducción de sequía, y durante todo el tratamiento de sequía empleando el equipo CI-340 Handheld Photosynthesis System (CI-340). Este equipo es un analizador infrarrojo de CO₂/H₂O que posee una cámara foliar en la cual se realizan mediciones de la tasa a la cual un área foliar conocida asimila una concentración de CO₂ en un tiempo dado.

5. Análisis de la expresión génica

a. Extracción de RNA

Se tomaron un total de 16 muestras, que incluyeron hojas y raíces de las variedades tolerantes y susceptibles en los 4 cuatro tiempos del estrés (Tabla 4). El RNA fue aislado a partir de aproximadamente 1 - 2 g de hojas o raíces usando el método basado en fenol - cloroformo (Buell Lab, Michigan State University-Anexo 1). Se evaluó la calidad en geles de agarosa al 1% y se realizó su purificación usando el Kit de Ambion DNA-free (Cat. Num. 1906).

Tabla 4. Códigos de las muestras colectadas de *S. tuberosum* subsp. *andigena* de las variedades tolerante y susceptible para la extracción de RNA.

Variedad	Tejido	Tratamiento	Muestra
Tolerante 703671	Hoja	Control	TLC
		Respuesta temprana	TL1
		Respuesta tardía	TL2
	Raíz	Recuperación	TL3
		Control	TRC
		Respuesta temprana	TR1
		Respuesta tardía	TR2
		Recuperación	TR3
		Susceptible 703248	Hoja
Respuesta temprana	SL1		
Respuesta tardía	SL2		
Raíz	Recuperación		SL3
	Control		SRC
	Respuesta temprana		SR1
	Respuesta tardía		SR2
	Recuperación		SR3

b. Secuenciación de RNA

Se empleó un secuenciador Hi-Seq™ 2000 de Illumina. La población de RNA fue convertida a una biblioteca de fragmentos de cDNA, cada molécula luego fue amplificada, cuantificada y secuenciada, con lecturas de 50 pares de bases (bp). Este procedimiento fue realizado en la Universidad de Michigan (Buell Lab). Se generaron 16 bibliotecas correspondientes a las 16 muestras tomadas.

c. Análisis bioinformático

Luego del secuenciamiento, toda la información de las lecturas fue recibida en archivos con formato FastQ (La "Q" se refiere a "quality" e implica que cada posición o base, además de tener una letra, tiene asociado un valor de calidad), obteniendo un archivo por cada biblioteca generada.

c.1. Control de calidad

FastQC

Se empleó el programa FastQC versión 0.10.0, una potente aplicación diseñada para proporcionar de manera simple controles de calidad a las lecturas procedentes de sistemas de secuenciamiento de nueva generación. Para ello se empleó el siguiente comando en la consola Linux:

```
>fastqc [input.fastq] [output_fastqc]
```

FASTX

Además, se empleó el kit de herramientas FastX, un conjunto de comandos para pre- procesar lecturas antes de los análisis de éstas. Para ello se emplearon las siguientes herramientas:

FASTQ Clipper (que elimina secuencias de adaptadores o linkers)

```
>fastx_clipper -v -Q33 -a [secuencia del adaptador] -l 20 -i [input] -o  
[output]
```

FASTQ Trimmed (que permite eliminar bases de mala calidad, si las hubiese, acortando el tamaño de las lecturas)

```
>fastx_trimmer -Q33 -v -f 3 -l 50 -i [input] -o [output]
```

c.2. Alineamiento de secuencias al genoma de referencia

Se emplearon los programas Bowtie versión 0.12.8 (Langmead *et al.*, 2009) y TopHat 2.0.4 (Trapnell *et al.*, 2009). En primer lugar se construyeron los archivos de acceso para el programa Bowtie, empleando el comando y el genoma de referencia de papa secuenciado (<http://www.potatogenome.net/>):

```
>bowtie-build reference.fa genome_potato
```

Este comando generó 6 archivos: genome_potato.1.ebwt,
genome_potato.2.ebwt,
genome_potato.3.ebwt,
genome_potato.4.ebwt,
genome_potato.rev.1.ebwt, y
genome_potato.rev.2.ebwt.

Posteriormente se corrió el programa Tophat, empleando el comando

```
> tophat -p 2 -i 5 -l 15000 -o output_thout potato_genome Input-clipped.fa
```

Se consideraron los tamaños mínimo y máximo de los intrones para papa de 5 y 15000 pares de bases respectivamente. El output de este programa fue un archivo accepted_hits.bam (o .sam), una lista de los alineamientos de cada lectura.

c.3. Ensamblaje de transcritos y cuantificación de su expresión

Se empleó el programa Cufflinks versión 2.0.2 (Trapnell *et al.*, 2012) para ensamblar los transcritos y determinar los niveles normalizados de la expresión génica, como fragmentos por kilobase de exón por millón de fragmentos mapeados (FPKM), una medición normalizada de la densidad de las lecturas

exónicas que permite comparar el nivel de los transcritos entre las 16 bibliotecas (Trapnell *et al.*, 2010).

El programa se corrió utilizando el comando:

```
>cufflinks [options]* <aligned_reads.(sam/bam)>
```

Este programa generó tres archivos de salida

- transcripts.gtf, que contiene las secuencias ensambladas.
- isoforms.fpk_tracking, contiene los valores de nivel de expresión de las secuencias.
- genes.fpk_tracking, contiene los valores estimados de nivel de expresión de los genes en fpkm.

Cuffdiff: Esta herramienta está incluida en el programa Cufflinks y con ella se encontraron los cambios significativos en el nivel de expresión de los genes entre los tiempos de muestreo. Los archivos de entrada para emplear este programa fueron: el archivo .gtf (los transcritos ensamblados), dos o más archivos .sam que contiene el alineamiento de 2 o más muestras o tratamientos. Con este programa se generaron varios archivos que contenían los resultados de cambios en la expresión a nivel de transcritos entre las muestras. A partir de estos archivos se pudo determinar cuáles fueron los genes inducidos o reprimidos en respuesta a la sequía temprana, tardía y en la recuperación, comparando los perfiles de expresión de 703248 y 703671.

c.4. Clasificación funcional de los transcritos

Para determinar las categorías de los genes se empleó la clasificación Gene Ontology (<http://www.geneontology.org>), empleando el programa Blast2GO version 2.3.6 (<http://www.blast2go.org>) se anotaron los primeros 200 genes inducidos por sequía de cada tratamiento (respuesta temprana, tardía y recuperación), en cada genotipo (703671 y 703248), de hojas y raíces.

Blast2go es una herramienta automatizada para la asignación de términos GO a cualquier secuencia de interés. Para obtener la anotación funcional, se siguieron los siguientes pasos:

- *Blast* para encontrar secuencias homólogas
- *Mapping* para asignar a las secuencias un término GO.
- *Anotación* para seleccionar las funciones más fiables, empleando la herramienta AgriGO (bioinfo.cau.edu.cn/agriGO).

C.5. Redes de co-expresión génica (Gene co-expression network analysis)

Se empleó el paquete WGCNA (Weighted Correlation Network Analysis) del programa “R”, con la finalidad de identificar los módulos de co-expresión de genes, utilizando los siguientes pasos:

- Se elaboraron las matrices de expresión de genes para hojas y raíces de 703248 y 703671 empleando el programa Cuffdiff, con la finalidad de generar perfiles de expresión de los genes a lo largo de los tratamientos.
- Posteriormente, los valores de expresión fueron transformados a su logaritmo (base 2) empleando un script de perl “log_transform_GSE_matrix”.
- Se filtraron los datos aplicando el coeficiente de variación usando un script de Perl “coefficient_of_variation_filtering_sample” para reducir el número de genes para el procesamiento final. La finalidad de este filtro es eliminar genes que sean constitutivamente expresados, que no tengan expresión o que esta varíe de manera no significativa durante los tiempos de estrés por sequía.
- Luego se realizó el clustering jerárquico para cada matriz y se generaron dendogramas donde se representan las relaciones de similitud entre los

genes analizados. A partir de estos dendogramas se determinó el número de los mismos.

- Se extrajeron los módulos usando el método Dynamic Tree Cut del paquete WGCNA (Langfelder and Horvath, 2007). Los parámetros β y tree-cut empleados fueron 20 - 0.9 para hojas de 703671 y raíces de 703248 y 703671 y 15 - 0.95 para raíces de 703248. Los demás parámetros fueron usados por defecto.
- Se transformaron los valores de expresión de los genes de todos los módulos con un "z-score" con otro script de Perl "generate_z_scores_wrapper.p" para estandarizar los niveles de expresión, y se representaron gráficamente los valores de expresión normalizados de los módulos de genes identificados.
- Para cada módulo se calcularon los eigengenes usando la función moduleEigengenes del paquete WGCNA (Langfelder and Horvath, 2007), y se generaron los Heatmap de eigengenes.

VI. RESULTADOS

1. Evaluación de la tasa fotosintética

Se consideró como respuesta temprana a la sequía una disminución de la tasa de fotosíntesis en un 25%, como respuesta tardía una disminución del 60%, y como recuperación un incremento de hasta el 80% del valor control, tomando como referencia la variedad tolerante (Figura 4). La disminución de un 25% de la tasa fotosintética fue registrada 40 minutos luego de iniciado el estrés, del 60% a los 120 minutos, y la recuperación de la tasa fotosintética se registró luego de 20 minutos de reiniciado el riego (Tabla 5).

Para la variedad susceptible 703248 se observó que la disminución de su fotosíntesis es mayor que en la variedad tolerante, solo a los 100 minutos la tasa fotosintética disminuye en un 60% (Figura 5). Además, luego del riego, la tasa fotosintética de 703671 se recuperó hasta el 80% del valor inicial (antes del estrés), mientras que en 703248 solo hasta un poco más del 50% del valor inicial (antes del estrés).

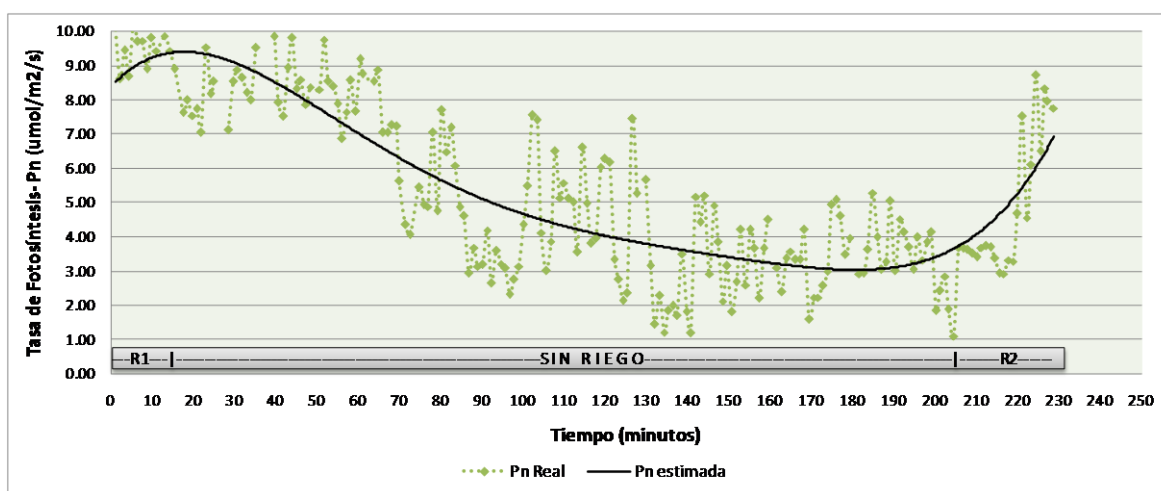


Figura 4. Variación de la tasa de fotosíntesis en la variedad 703671 (tolerante) durante las condiciones control (riego inicial: R1), sequía (sin riego) y recuperación (reinicio del riego: R2).

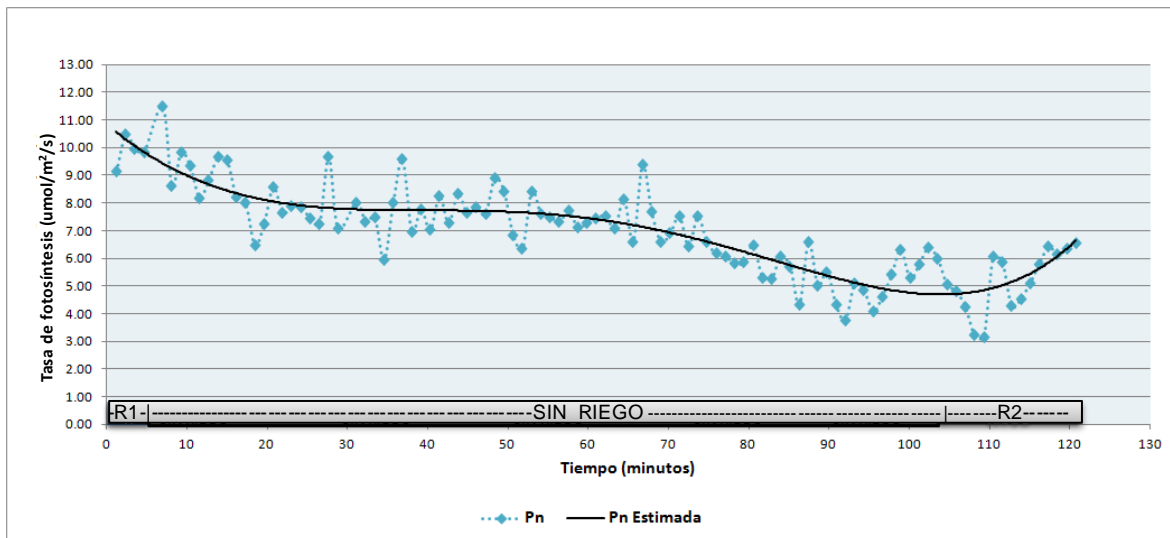


Figura 5. Variación de la tasa de fotosíntesis en la variedad 703248 (susceptible) durante las condiciones control (riego inicial: R1), sequía (sin riego) y recuperación (reinicio del riego: R2).

Tabla 5. Tiempos de muestreo de hojas y raíces calculados a partir de la variación de la tasa de fotosíntesis (Pn) durante la respuesta temprana (T1), tardía (T2) y recuperación del estrés (T3).

Tratamiento	Tiempo	Pn Promedio ($\mu\text{mol}/\text{m}^2/\text{s}$)
Control	Minuto cero, antes de la sequía	9.0
T1	40 minutos después de iniciar la sequía	6.8 - 7.0
T2	120 minutos luego de iniciar la sequía	3.6 - 3.8
T3	20 minutos luego de iniciar el riego	7 - 7.5

a. Evaluación visual durante la exposición a sequía

A partir de la inspección visual del aspecto externo de las plantas durante la exposición a sequía se observó que la variedad 703671 estuvo menos afectada (Figura 6), manteniendo su rigidez, a diferencia de 703248 que perdió rigidez rápidamente (Figura 7). Estas observaciones resultan congruentes con las mediciones de las tasas fotosintéticas.

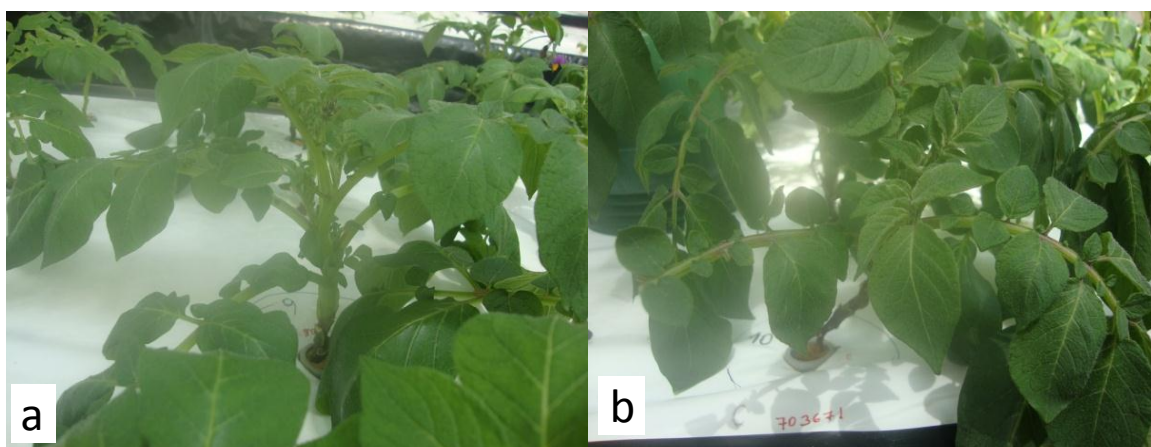


Figura 6. Plantas de la variedad 703671 luego de 60 minutos de exposición a sequía. (a) vista de planta entera, (b) vista de una hoja compuesta



Figura 7. Plantas de la variedad 703248 luego de 60 minutos de exposición a sequía. (a) pérdida de rigidez de la planta, (b) marchitez de una hoja compuesta.

2. Extracción de RNA

La cantidad y calidad del RNA extraído fue la adecuada pues se observan las bandas correspondientes a los rRNA 28S, 18S y 5S en geles de agarosa (Figura 8). Los valores de los índices A260/A280 y A260/A230, que indican la pureza de las muestras, resultaron dentro de los aceptados (A260/A280~2.0 y < A260/A230), como se muestra en la Tabla 6.

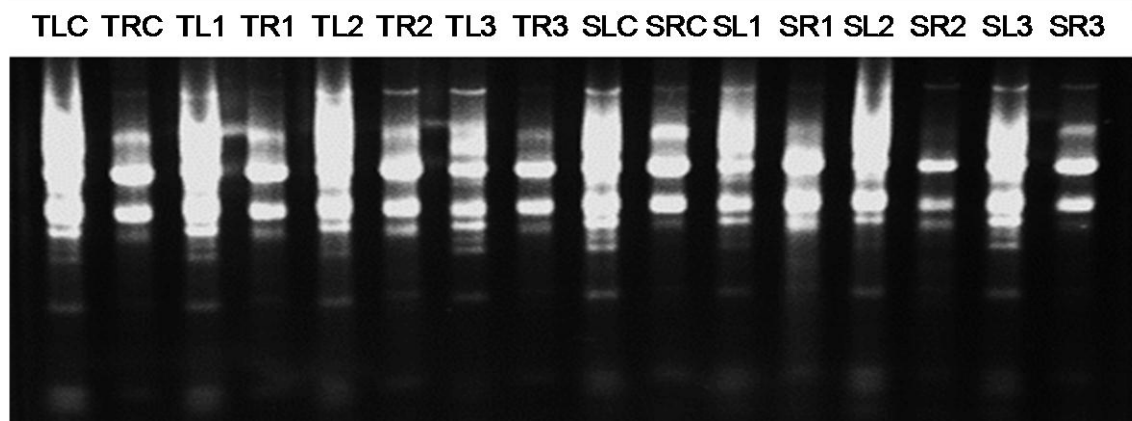


Figura 8. Evaluación de calidad de RNA extraído de hojas y raíces de *S. tuberosum* subsp. *andigena* empleando geles de agarosa al 1%.

Tabla 6. Cuantificación (ng/ μ L) y calidad (índices de absorbancia 260nm/280nm, 260nm/230nm) de las extracciones de RNA utilizando un equipo NanoDrop.

Variedad	Muestra	Concentración ng/μL	Índice 260/280	Índice 260/230
Susceptible	SLC	4559.63	1.98	2.01
CIP 703248	SRC	2853.82	2.12	2.32
	SL1	3042.89	2.05	2.05
	SR1	3527.75	2.08	2.27
	SL2	3875.48	2.02	2.1
	SR2	3042.53	2.13	2.32
	SL3	3824.2	2.04	2.08
	SR3	2571.61	2.16	2.35
	Tolerante CIP 703671	TLC	2832.37	2.06
TRC		2442.19	2.13	2.32
TL1		4728.81	1.9	2.03
TR1		3885.68	2.05	2.23
TL2		4698.38	1.96	1.98
TR2		4289.25	2	2.15
TL3		4145.04	1.98	2.02
TR3		4442.63	1.93	2.09

3. Análisis bioinformático de los datos generada por RNA-seq

a. Control de calidad de secuencias

Todos los datos del secuenciamiento fueron recibidos en 16 archivos, uno por cada biblioteca, en formato .FastQ. Se determinó el número total de secuencias obtenidas por cada biblioteca empleando el programa FastQC, el cual generó un archivo “FastQC Report” para cada una de las bibliotecas, y con la

inspección de estos archivos se determinó el número de lecturas de cada biblioteca. En total, se secuenciaron más de 392 millones de lecturas, con un número promedio de 23 millones de lecturas por biblioteca.

Tabla 7. Lista de bibliotecas correspondiente a cada variedad y tratamiento y número total de lecturas generadas.

Tejido	Muestra	Nombre de la biblioteca	Total de lecturas
Hoja	TLC	PGO_AA_CAGATC_L008_R1_001.fastq	26304956
	TL1	PGO_AI_AGTCAA_L006_R1_001.fastq	17079619
	TL2	PGO_AM_GTGAAA_L005_R1_001.fastq	22260826
	TL 3	PGO_AE_ATGTCA_L008_R1_001.fastq	25424875
	SLC	PGO_AC_AGTCAA_L008_R1_001.fastq	27976839
	SL1	PGO_AK_CCGTCC_L005_R1_001.fastq	28223430
	SL2	PGO_AO_TGACCA_L005_R1_001.fastq	23408804
	SL 3	PGO_AG_CAGATC_L006_R1_001.fastq	20067948
Raíz	TRC	PGO_AB_CTTGTA_L008_R1_001.fastq	30840996
	TR1	PGO_AJ_AGTTCC_L006_R1_001.fastq	19780525
	TR2	PGO_AN_CGATGT_L005_R1_001.fastq	23292392
	TR3	PGO_AF_GCCAAT_L006_R1_001.fastq	14717082
	SRC	PGO_AD_AGTTCC_L008_R1_001.fastq	27993309
	SR1	PGO_AL_GTCCGC_L005_R1_001.fastq	27415791
	SR2	PGO_AP_ACAGTG_L005_R1_001.fastq	24620110
	SR 3	PGO_AH_CTTGTA_L006_R1_001.fastq	16975442

El programa FastQC calculó los “Phred Scores” para cada base de las lecturas secuenciadas. Los valores Phred variaron entre 4 y 60, y se asignaron un valor de calidad a cada base de una secuencia, siendo más alta la calidad al más alto valor. Se consideró como umbral un valor Phred > 28 (considerando que un valor de Phred = 30 indica que la probabilidad de tener una base incorrecta es de 1 en 1000, o una precisión de base correcta de 99.9%). Con este procedimiento se optimizó la calidad de las lecturas de las 16 bibliotecas, haciendo posible el posterior análisis bioinformático de las mismas.

Respecto a la calidad de cada base de las lecturas de las bibliotecas, en 5 bibliotecas las 50 bases de las lecturas presentaron un valor Phred por encima del umbral (Fig. 9a), 6 bibliotecas presentaron las 3 primeras bases con un valor de Phred debajo del umbral (Fig. 9b) y 5 bibliotecas presentaron las 4 primeras bases con un valor Phred debajo del umbral (Figura 9c). En estos dos últimos casos se empleó la herramienta FastX-Trimmer para cortar las primeras 3 ó 4 bases de todas las lecturas de estas bibliotecas

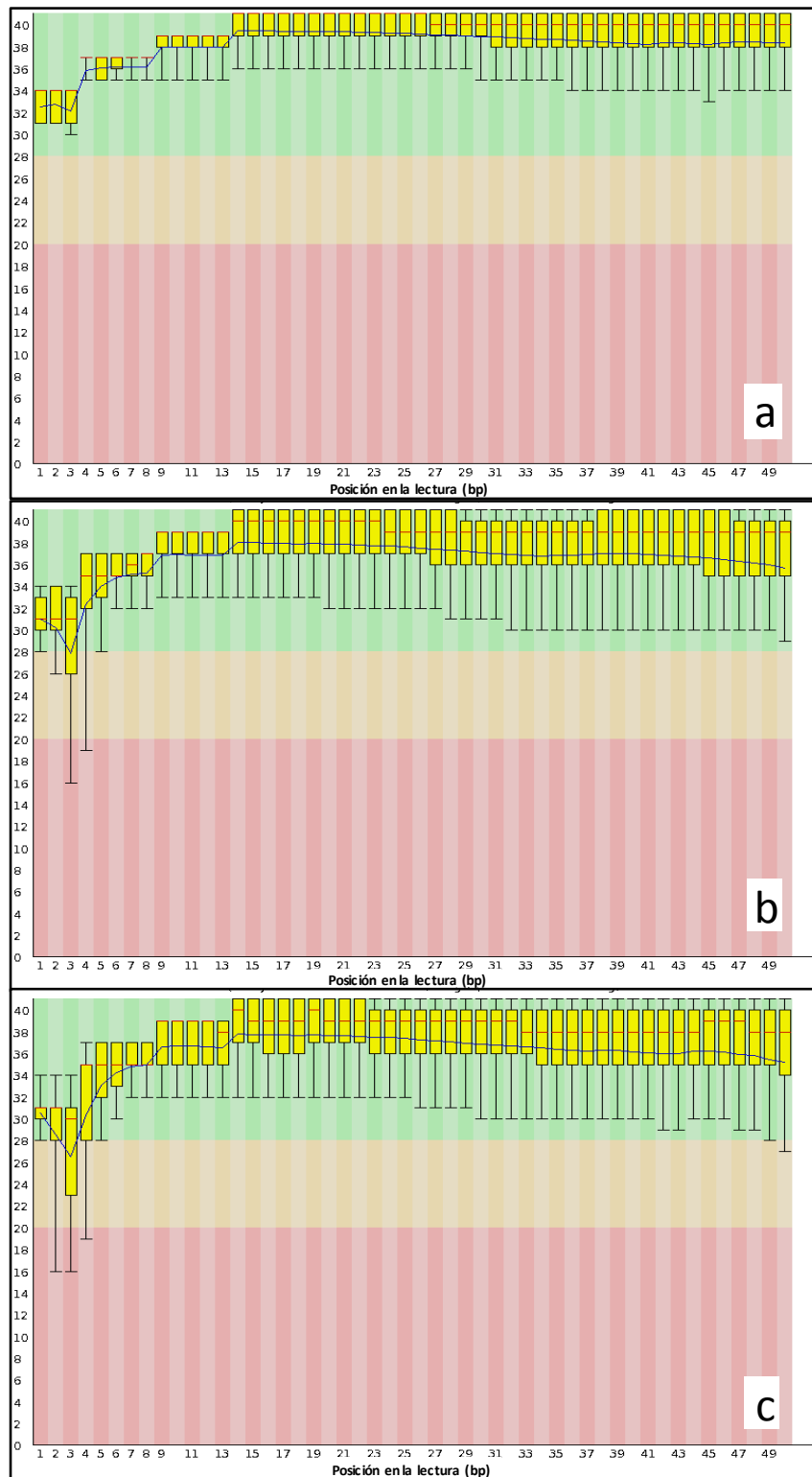


Figura 9. Valores Phred de calidad para cada base de las lecturas. (a) Lecturas con todas las bases con un valor Phred por encima del umbral, (b). Tres primeras bases con un valor Phred menor al umbral, (c) Cuatro primeras bases con un valor Phred menor al umbral.

b. Mapeo de las lecturas

Una de las tareas esenciales del RNA-Seq es mapear las lecturas o secuencias a un genoma de referencia. Esta parte del análisis se realizó empleando el programa TopHat para mapear todas las secuencias obtenidas al genoma de referencia de papa, que corresponde a *S. tuberosum* Grupo Phureja DM1-3 516R44 (PGSC, 2011).

En base al alineamiento, las secuencias fueron clasificadas en tres clases: lecturas únicas que son aquellas que mapean con una única posición en el genoma, lecturas múltiples que mapean a más de una posición en el genoma de referencia y las lecturas que no mapean con región genómica alguna. Empleando el script Perl (“mapping estadísticas”) se calculó el número y porcentaje de estas lecturas para cada biblioteca (Tabla 8).

Entre las 16 bibliotecas analizadas se encontró que el porcentaje de las lecturas únicas varió entre 75 - 83%, el de las lecturas múltiples entre 6 - 14%, y de las lecturas que no mapearon con el genoma de referencia entre 8 - 11%. Como un control interno de esta parte del análisis, se secuenció también una biblioteca de hojas de DM1-3 516R44 (hojas de DM), y se observó que los porcentajes de lecturas únicos, múltiples y que no mapean con el genoma resultaron similares.

Tabla 8. Porcentajes del tipo de lecturas según su mapeo al genoma de referencia de DM1-3 516R44.

Muestra	Total de lecturas	Lecturas únicas	%	Lecturas múltiples	%	No mapeado	%
TLC	26304956	21291611	80.94	2758808	10.49	2254537	8.57
TL1	17079619	14158415	82.90	1535509	8.99	1385695	8.11
TL2	22260826	18156792	81.56	2257435	10.14	1846599	8.30
TL3	25424875	20546901	80.81	2214432	8.71	2663542	10.48
TRC	30840996	25101293	81.39	2153567	6.98	3586136	11.63
TR1	19780525	16431906	83.07	1478617	7.48	1870002	9.45
TR2	23292392	19398147	83.28	1475718	6.34	2418527	10.38
TR3	14717082	12249428	83.23	1047903	7.12	1419751	9.65
SLC	27976839	21798560	77.92	3250580	11.62	2927699	10.46
SL1	28223430	21187144	75.07	4139513	14.67	2896773	10.26
SL2	23408804	18275867	78.07	2455980	10.49	2676957	11.44
SL3	20067948	15983176	79.65	2002679	9.98	2082093	10.38
SRC	27993309	22590737	80.70	2129629	7.61	3272943	11.69
SR1	27415791	22386601	81.66	1945125	7.09	3084065	11.25
SR2	24620110	20164169	81.90	1711811	6.95	2744130	11.15
SR3	16975442	13975326	82.33	1162503	6.85	1837613	10.83
Hojas de DM	15983851	12481572	78.09	1513613	9.47	1988666	12.44

c. Ensamblaje de transcritos

Con el programa Cufflinks se ensamblaron los transcritos y se determinó su abundancia en FPKM. Un transcripto fue considerado como expresado si: el límite inferior de un intervalo de confianza al 95% de la abundancia fue mayor de cero y si el nivel de expresión fue mayor a 0.001 ($f_{pkm} > 0.001$). En base a estos criterios, se encontró un promedio de 29000 transcritos por librería, desde 27188 transcritos como mínimo en raíces de 703671, hasta 32746 transcritos como máximo en raíces de 703248, ambos durante la recuperación.

Tabla 9. Número de transcritos identificados en las 16 librerías.

Muestra	Número de transcritos identificados
TLC	28819
TL1	27258
TL2	28899
TL3	30184
TRC	31151
TR1	29513
TR2	31487
TR3	27188
SLC	28622
SL1	28780
SL2	29159
SL3	28148
SRC	30584
SR1	32746
SR2	30339
SR3	28953

d. Expresión diferencial

Con la finalidad de identificar genes diferencialmente expresados entre dos condiciones, se comparó el perfil de expresión de los genes. Con el programa Cuff-diff y el scrip de Perl Cuffdiff_clean se determinaron las diferencias en la expresión de genes entre 703248 y 703671 bajo las diferentes condiciones de sequía.

d.1. Cambios de la expresión génica durante el estrés

El número de genes diferencialmente expresados entre tratamientos, varió entre 887 como mínimo y 1995 como máximo, se encontró que el número de genes que cambian su nivel de expresión aumenta conforme avanza la intensidad de la sequía, esto se encontró en hojas y raíces de ambas variedades (Figura 10).

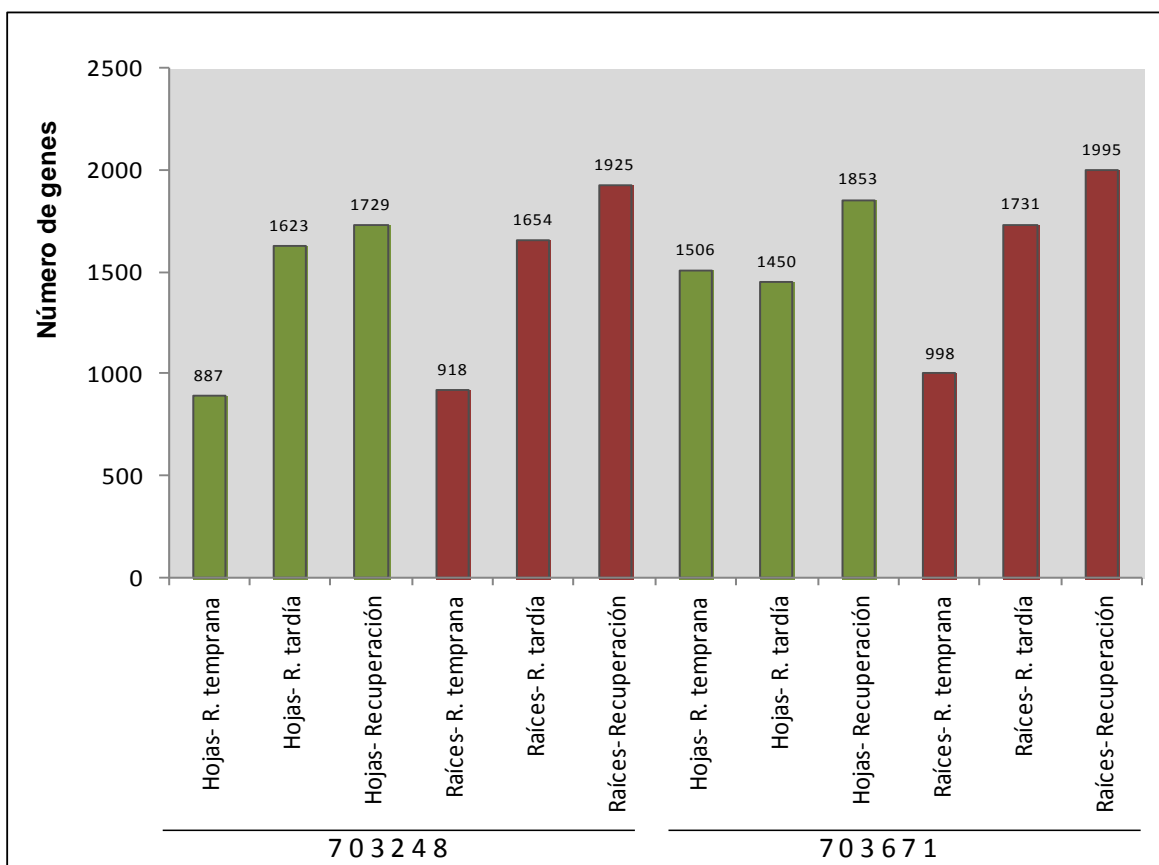


Figura 10. Número de genes diferencialmente expresados en hojas y raíces de las variedades 703248 (susceptible) y 703671 (tolerante).

En la recuperación de raíces se encontró el mayor número de genes con expresión diferencial: en 703248 (1925 genes) y en 703671 (1995 genes). Sin embargo, se observó que el mayor cambio ocurre en el paso de la respuesta temprana a tardía donde un mayor número de genes cambiaron sus niveles de expresión para hojas de 703248 (1179), hojas de 703671 (914), raíces de 703248 (1298) y raíces de 703671 (1296), mientras que en el paso de la respuesta tardía a recuperación un menor número de genes cambian su expresión (612 para hojas susceptibles, 839 para hojas tolerantes, 921 para raíces susceptibles y 1171 para raíces tolerantes). En la tabla 10, se muestran algunos de estos genes que son inducidos recién en la respuesta temprana, y que mantienen o reprimen su expresión luego en la recuperación.

Tabla 10. Niveles de expresión (en fpkm) de genes inducidos específicamente en la respuesta tardía, manteniendo o disminuyendo luego en la recuperación. T1: respuesta temprana, T2: respuesta tardía, T3: recuperación.

PGSC_id	Control	T1	T2	T3	UNIREF
Hojas tolerantes					
PGSC0003DMG400012770	0	0	2149.82	1.00	Calmodulina
PGSC0003DMG400025191	0	0	942.27	1.00	Proteína RD22
PGSC0003DMG400018807	0	0	2134.97	3.23	Rad7
PGSC0003DMG400004367	0	0	2.10	1.00	Proteína de respuesta a sal 1
PGSC0003DMG400008188	0	0	11.24	23.26	Factor de transcripción WRKY
Raíces de susceptibles					
PGSC0003DMG400015495	0	0	41.36	92.41	Dehidrina DH2a
PGSC0003DMG400003569	0	0	51.98	73.52	Proteína taumatina
PGSC0003DMG400017792	0	0	48.84	54.19	Ce-LEA
PGSC0003DMG400002731	0	0	11.24	19.56	Proteína LEA 3
PGSC0003DMG400000628	0	3.03	269513.87	558036.51	Proteína LEA
PGSC0003DMG400002731	0	0	11.24	19.56	Proteína LEA 3
PGSC0003DMG400015129	0	0	1845.76	1.00	Defensina

d.2. Genes inducidos y reprimidos

En la figura 11 se muestra la comparación numérica del número de transcritos inducidos y reprimidos bajo los tratamientos de sequía, en hojas y raíces de 703248 y 703671.

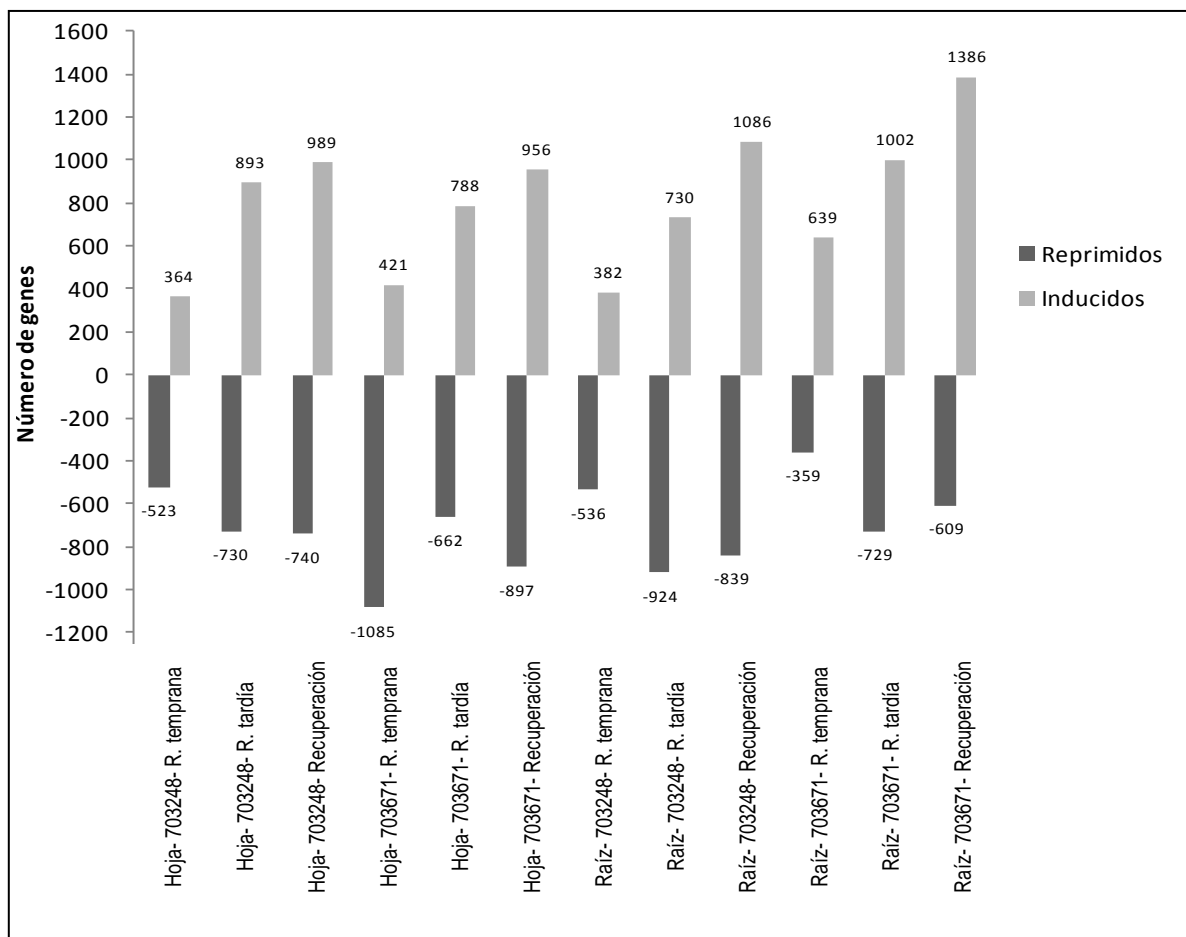


Figura 11. Distribución del número de genes suprimidos e inducidos por sequía.

Se observó una tendencia creciente en la inducción de genes conforme el estrés avanza. En hojas de 703248 se encontró que 364 genes fueron inducidos significativamente durante la respuesta temprana, 893 en la sequía tardía y 989 en la recuperación. En hojas de 703671 se encontró un patrón similar, 421, 788 y 956 genes inducidos en la respuesta temprana, tardía y de recuperación. En raíces de 73671 se encontró el mayor número de genes inducidos (639, 1002 y 1386) a lo largo del estrés. Entre los genes más sobreexpresados se encontraron

aquellos de: proteína LEA, factores de transcripción DREB, AP2/CBF, AP2/ERF fosfatasa, fosfolipasa, peroxidasa, proteína de choque térmico, inhibidor de proteasa, entre otros (Tabla 11).

Tabla 11. Algunos de los genes inducidos en raíces de la variedad 703671 durante la recuperación. PGSC-id: códigos de los transcritos de papa, UNIREF: Uniprot reference cluster, TRC: nivel de expresión en fpkm en la raíces control, TR3: nivel de expresión en fpkm en las raíces en recuperación.

PGSC_id	UNIREF	TRC (fpkm)	TR3 (fpkm)	log ₂ cambio
PGSC0003DMG40000628	Proteína abundante de embriogénesis tardía	0.483	413.79	9.743
PGSC0003DMG400012525	DREB1	0.412	66.690	7.337
PGSC0003DMG400006369	AP2/ERF	0.568	80.764	7.153
PGSC0003DMG400009112	Proteína fosfatasa 2C	0.827	101393	6.937
PGSC0003DMG400014055	Peroxidasa	2.860	267832	6.549
PGSC0003DMG400032259	Inhibidor de proteasa	1.844	89.726	5.605
PGSC0003DMG400026257	Fosfolipasa A1	0.281	13.172	5.549
PGSC0003DMG400022894	Lipoxigenasa	3.414	110304	5.014
PGSC0003DMG400027614	Peroxidasa de pared celular	0.358	9.893	4.790
PGSC0003DMG400022759	Proteína de unión a DREB1	2.047	50375	4.621
PGSC0003DMG400002291	Triacilglicerol lipasa	0.513	11593	4.499

Respecto al número de genes reprimidos, no se encontró una tendencia de cambio en las etapas del estrés, ya que solo en hojas de 703248 la represión de genes aumenta con el nivel de estrés, mientras que en hojas de 703671 se encontró el mayor número de genes reprimidos (1085) durante la respuesta temprana, y una disminución (662) en la respuesta tardía. En raíces, la represión de genes aumenta del paso de respuesta temprana a tardía y disminuye en la recuperación, encontrándose en la respuesta temprana de 703671 el menor número de genes reprimidos (359).

Entre los genes reprimidos en la etapa temprana de hojas tolerantes se encontraron aquellos asociados a división celular, a crecimiento celular, genes

SAUR de respuesta a auxina, varios de la familia citocromo P450, entre otros (Tabla 12).

Tabla 12. Algunos de los genes reprimidos en hojas de la variedad 703671 durante la recuperación. PGSC-id: códigos de los transcritos de papa, UNIREF: Uniprot reference cluster, TLC: nivel de expresión en fpkm en hojas control, TL1: nivel de expresión en fpkm en las hojas en la respuesta temprana.

PGSC_id	UNIREF	TLC (fpkm)	TL1 (fpkm)	log ₂ cambio
PGSC0003DMG400008001	Ciclina tipo D, familia 3, subgrupo3	23.7214	0.446651	-5.7309
PGSC0003DMG400022397	Ciclina tipo D6	3.99505	0.16112	-4.63194
PGSC0003DMG400009983	Ciclina B	8.28002	0.52757	-3.97218
PGSC0003DMG401022702	Ciclina B1	8.02544	0.60299	-3.73437
PGSC0003DMG400030415	Expansina18	29.9529	2.1316	-3.81269
PGSC0003DMG400019507	Expansina	13.0423	0.7938	-4.03828
PGSC0003DMG400034808	Extensina clase 1	30.3295	1.43628	-4.40031
PGSC0003DMG400004796	Extensina	44.0351	1.38566	-4.99000
PGSC0003DMG400001668	Proteína de la familia SAUR	27.5755	6.19471	-2.15428
PGSC0003DMG400001667	Proteína de la familia SAUR	55.2983	9.76017	-2.50226
PGSC0003DMG400003773	Proteína de la familia SAUR	5.00274	0.71540	-2.80588
PGSC0003DMG400002078	Citocromo P450	34.6827	0.64583	-5.74691
PGSC0003DMG400046813	Citocromo P450	10.6695	0.61646	-4.11333

d.3. Diferencias en la expresión génica entre las variedades

Comparando el perfil de expresión de ambas variedades, se observó en 703671 que un mayor número de genes mostraron cambios significativos en la respuesta a la sequía (Figura 12). En la respuesta temprana de hojas se encontró que el número de genes específicos y diferencialmente expresados en 703671 (1079) fue más del doble que en 703248 (460). Entre estos se encontraron genes que codifican para proteínas de unión a DRE, proteínas RD22 de respuesta a deshidratación, proteína LEA5, factor de transcripción MYB139 y a una proteína taumatina (Tabla 13).

Tabla 13. Algunos genes específicos de hojas de la variedad 703671 durante la etapa temprana. PGSC-id: códigos de los transcriptos de papa, UNIREF: Uniprot reference cluster, TLC: nivel de expresión en fpkm en la hojas control, TL1m: nivel de expresión en fpkm en las hojas en la respuesta temprana.

PGSC_id	UNIREF	TLC (fpkm)	TL1 (fpkm)	log ₂ cambio
PGSC0003DMG401023951	Proteína de unión de DRE	0.78240	3.37408	2.10851
PGSC0003DMG400025192	Proteína RD22 de respuesta a deshidratación	2.40771	11.8799	2.30279
PGSC0003DMG400015054	Proteína RD22 de respuesta a deshidratación	0.22797	2.00062	3.13348
PGSC0003DMG400017936	Proteína LEA 5	3.34748	15.9614	2.25344
PGSC0003DMG402010883	Factor de transcripción MYB139	1.51574	8.10369	2.41855
PGSC0003DMG400003569	Proteína taumatina	0.27162	23.1136	6.41099

Por otro lado, en la respuesta tardía y recuperación no existe marcada diferencia. En las raíces, la mayor diferencia se encontró en la recuperación, donde el número de genes diferencialmente expresados específicos de 703671 (852) es mucho mayor al de 703248 (460).

Al considerar solo los genes inducidos, la mayor diferencia entre las variedades se encontró en las raíces, donde el número de genes específicos de 703671 duplicó al de 703248 en todas las etapas del estrés (469 vs. 212 en la respuesta temprana, 655 vs. 383 en la respuesta tardía, y 594 vs. 294 en la recuperación). Por otro lado, inspeccionando la inducción de genes en hojas tolerante durante la recuperación se observó la marcada sobreexpresión de un grupo de genes codificantes de pequeñas proteínas de choque térmico en cloroplastos en comparación a hojas susceptibles (Tabla 14).

Tabla 14. Niveles de expresión de transcritos de pequeñas proteínas de choque térmico en cloroplastos en hojas de la variedad tolerante 703671 y la variedad susceptible 703248 durante la recuperación. PGSC-id: códigos de los transcritos de papa, UNIREF: Uniprot reference cluster, TLC: nivel de expresión en fpkm en la hojas control, TR3: nivel de expresión en fpkm en las hojas en la recuperación.

PGSC_id	Variedad	TLC (fpkm)	TR3 (fpkm)	log ₂ cambio
PGSC0003DMG400011630	Hojas	3.81369	37.19	3.28559
PGSC0003DMG400011628	susceptibles	3.98048	42.34	3.41109
PGSC0003DMG400011632		15.2504	170655	3.48416
PGSC0003DMG400011631		6.63899	90.89	3.77513
PGSC0003DMG400011627		5.17384	84.55	4.03054
PGSC0003DMG400011630	Hojas	0.823321	552532	9.39039
PGSC0003DMG400011628	tolerantes	1.07701	285019	8.04788
PGSC0003DMG400011632		1.87656	435525	7.85853
PGSC0003DMG400011631		3.64153	822774	7.81981
PGSC0003DMG400011627		3.5994	803299	7.80204

En cuanto a los genes reprimidos, la mayor diferencia entre las variedades se encuentra en la respuesta temprana de hojas, donde se encontró 809 genes específicos de 703671, cantidad mucho mayor a los 247 genes de 703248. En cuanto a raíces, el número de genes reprimidos específicos de 703671 es menor que 703248 en todas las etapas de la sequía.

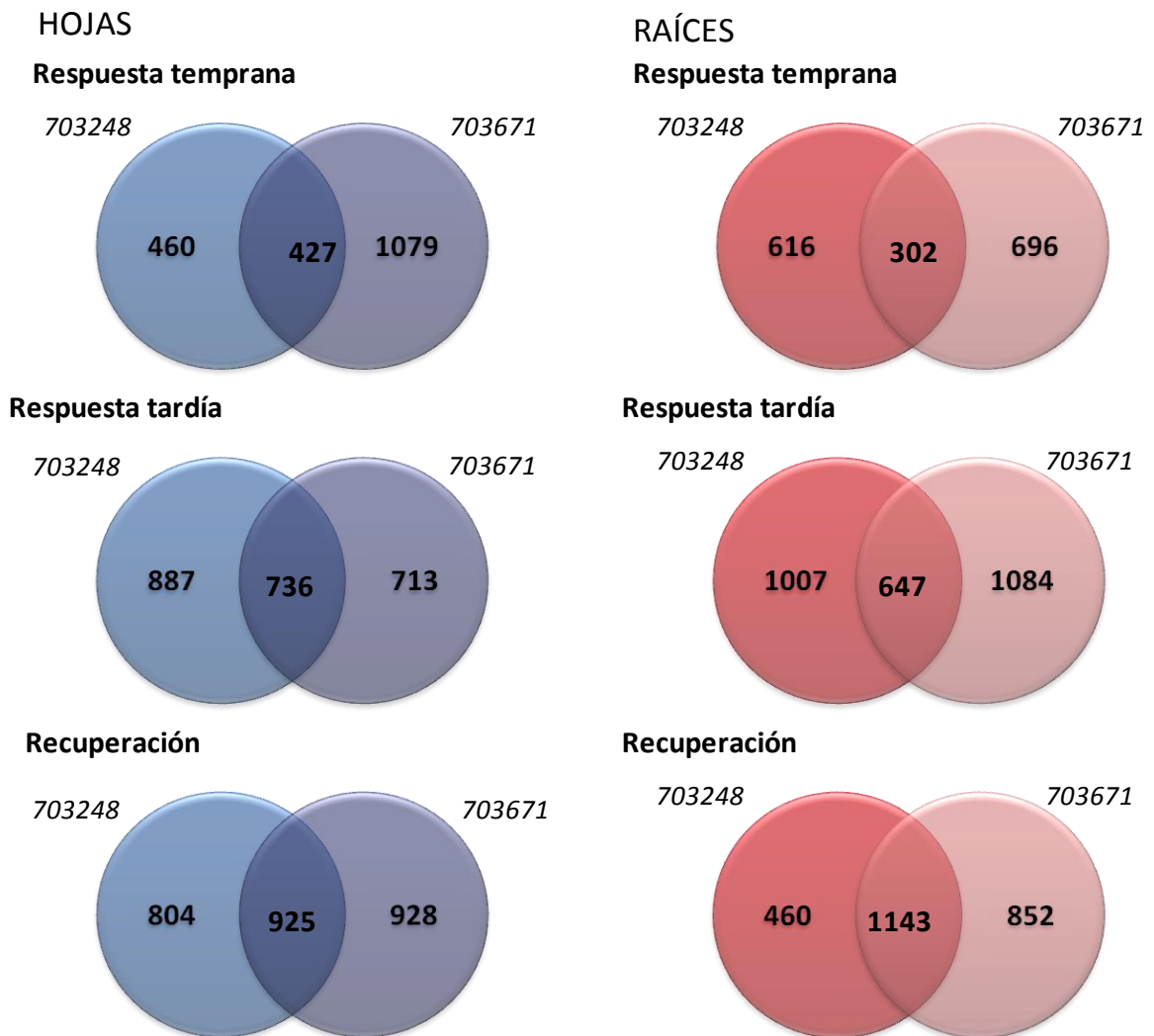
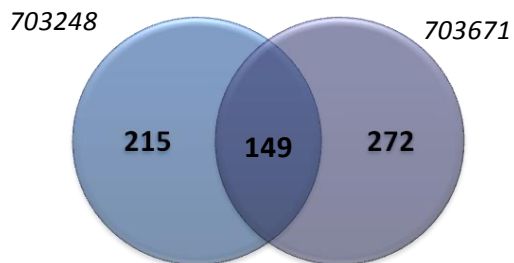


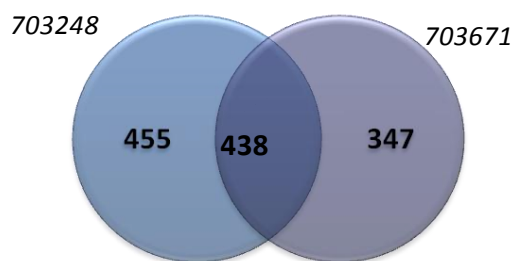
Figura 12. Número total de genes diferencialmente expresados por sequía (respuesta temprana y tardía) y en la recuperación. Se muestra el número de genes compartidos por la variedad susceptible (703248) y tolerante (703671) y aquellos de expresión única.

HOJAS

Respuesta temprana



Respuesta tardía

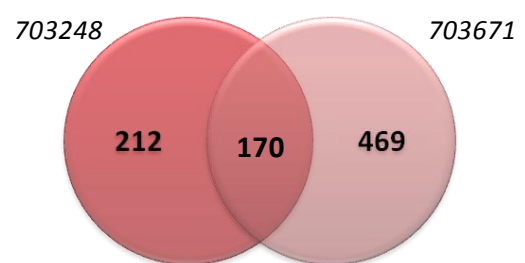


Recuperación



RAÍCES

Respuesta temprana



Respuesta tardía



Recuperación

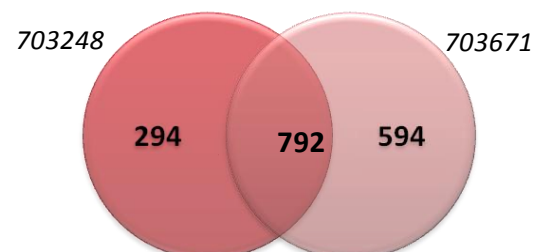
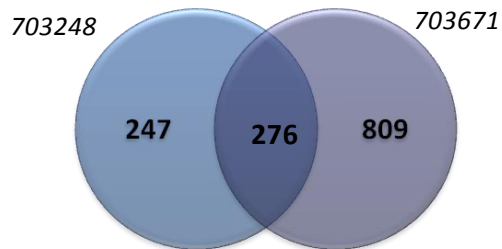


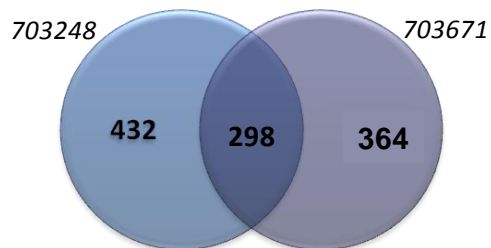
Figura 13. Número de genes inducidos por sequía (respuesta temprana y tardía) y en la recuperación. Se muestra el número de genes compartidos la variedad susceptible (703248) y tolerante (703671) y aquellos de expresión única.

HOJAS

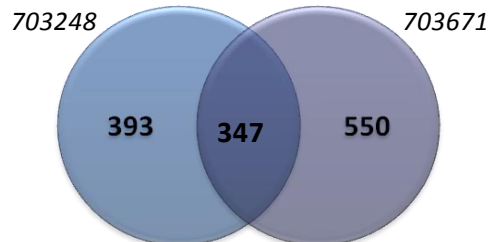
Respuesta temprana



Respuesta tardía



Recuperación



RAÍCES

Respuesta temprana



Respuesta tardía



Recuperación

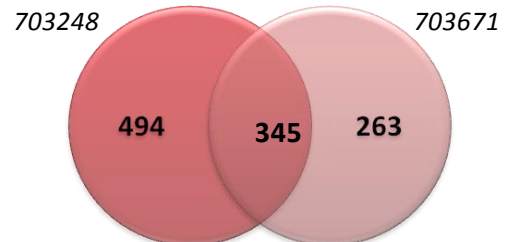


Figura 14. Número de genes reprimidos por sequía (respuesta temprana y tardía) y en la recuperación. Se muestra el número de genes compartidos la variedad susceptible (703248) y tolerante (703671) y aquellos de expresión única.

d.4. Diferencias en la expresión de genes entre tejidos

Se realizó una comparación entre hojas y raíces, para identificar aquellos genes comunes y específicos de tejido que son expresados durante la sequía. Se observó que durante la respuesta temprana los genes con expresión diferencial fueron en su mayoría específicos de tejido, pues se encontraron solo 79 genes comunes a los tejidos de variedad susceptible y 141 en la tolerante.

En cuanto a la inducción de genes, se encontró también que en la respuesta temprana la mayoría de los sobre-expresados fueron tejido-específicos; conforme el estrés avanza en hojas y raíces se comienzan a expresar genes comunes, es decir aquellos que ejercen su función en ambos tejidos, siendo el número de genes comunes en promedio 25 - 33% del total. En cuanto a la represión, se encontró que estos genes de respuesta a estrés son en su mayoría específicos de tejido, siendo el 5 - 14% el promedio de genes comunes del total de reprimidos (Tabla 15).

Entre los genes específicos de hojas se encontraron proteínas de unión a clorofila (15 transcritos reprimidos en la respuesta tardía de hojas de 703671), pequeñas proteínas de choque térmico de cloroplastos (5 transcritos inducidos), y pequeñas proteínas de choque térmico 17.6 kD clase I (3-5 transcritos inducidos) presentes en todas las etapa del estrés, mientras que en raíces solo se encontraron en la respuesta tardía y la recuperación. Por otro lado, entre los genes específicos de raíces, se encontró un número elevado de transcritos codificantes de extensinas (23 y 28 en la respuesta temprana de raíces susceptibles y tolerantes, respectivamente), y uno de exostonsina que son reprimidos en respuesta al estrés.

Tabla 15. Número de genes en los tejidos diferencialmente expresados en respuesta a la sequía. Se muestra el número de genes específicos y comunes a los tejidos.

Variedad	Etapa de estrés	Genes específicos		Genes comunes
		Hojas	Raíces	
Genes totales				
Susceptible	Respuesta temprana	808	839	79
	Respuesta tardía	1245	1276	378
	Recuperación	1332	1528	397
Tolerante	Respuesta temprana	1365	857	141
	Respuesta tardía	1086	1368	363
	Recuperación	1435	1578	418
Genes inducidos				
Susceptible	Respuesta temprana	318	336	46
	Respuesta tardía	641	478	252
	Recuperación	718	815	271
Tolerante	Respuesta temprana	349	567	72
	Respuesta tardía	527	741	261
	Recuperación	632	1062	324
Genes reprimidos				
Susceptible	Respuesta temprana	492	505	31
	Respuesta tardía	626	820	104
	Recuperación	650	749	90
Tolerante	Respuesta temprana	1060	334	25
	Respuesta tardía	573	640	89
	Recuperación	825	537	72

e. Clasificación funcional de los transcritos

Se realizó la clasificación de los 200 genes más sobre-expresados en cada tratamiento y de cada tejido. En general, cerca del 60% de los genes fueron mapeados a una o más ontologías, mientras que al resto de genes no fue posible asignarles algún término GO. La mayoría de los términos asignados fueron aquellos referidos a la categoría de procesos biológicos, con un total de 6076 anotaciones, seguidos de la categoría componente celular (3224) y función molecular (3054). Es importante indicar que para los genes sobreexpresados durante la respuesta temprana en raíces de 703671, se encontraron 955 términos GO de componentes celular, valor superior al encontrado en los demás tratamientos (Figura 15).

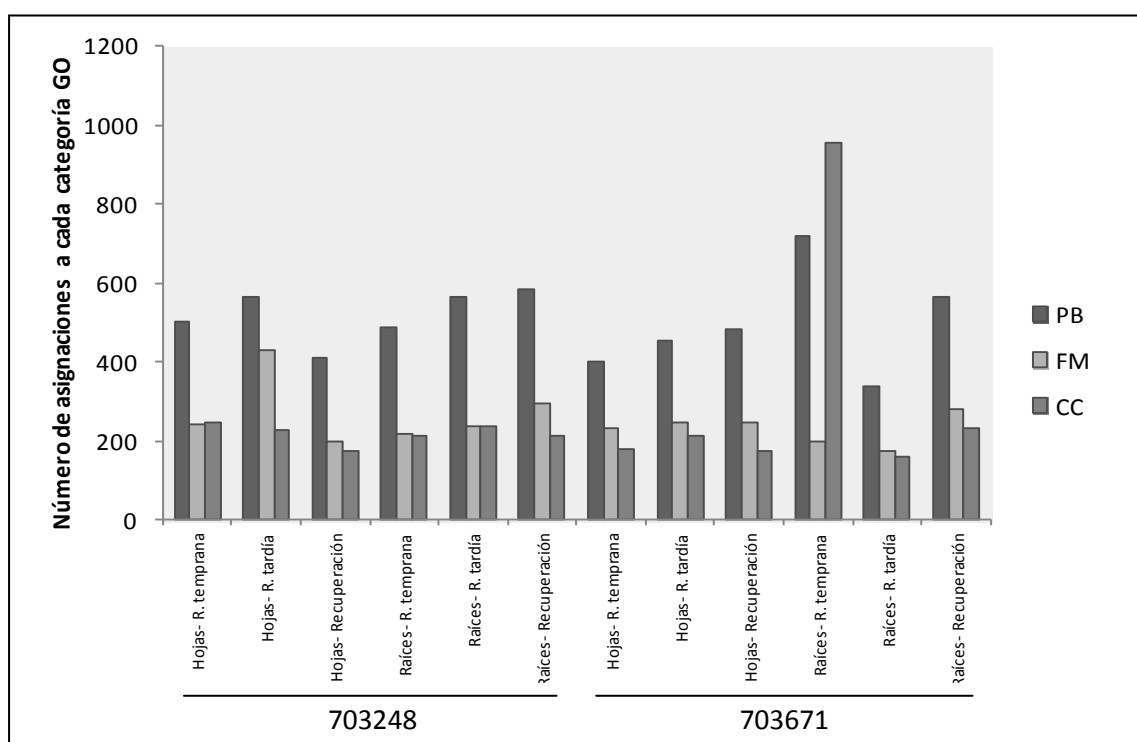


Figura 15. Número de asignaciones a las tres categorías de Gene Ontology: PB (proceso biológico), FM (función molecular) y CC (componente celular), para las dos variedades de *S. tuberosum* subsp. *andigena*.

- **Procesos biológicos**

Durante la respuesta temprana a la sequía en 703248 los términos más enriquecidos fueron de regulación biológica (GO:0065007, p-value = 5.51E-18), procesos metabólicos (GO:0008152, p-value = 2.90E-18), regulación de procesos metabólicos de RNA (GO:0051252, p-value = 6.30E-11), transcripción génica (GO:0006351, p-value = 5.70E-12), y regulación de la transcripción (GO:0006355, p-value = 6.30E-11). Estos términos también se encontraron enriquecidos en 703671, pero adicionalmente a ellos se encuentran otros como el de reducción de oxidación (GO:0055114, p-value = 5.510E-22) y de respuesta a estrés (GO:0006950, p-value = 5.50E-08) (Figura 16).

Durante la respuesta tardía y recuperación a la sequía, el término GO más enriquecido en hojas de ambas variedades fue el de respuesta a estrés (GO:0006950, p-value = 1.60E-36, 4.60E-24), seguido de respuesta a estímulos (GO:0050896, p-value = 1.30E-24), y de procesos metabólicos (GO:0008152, p-value = 4.00E-14) de proteínas y carbohidratos (Figura 17, 18).

- **Función molecular**

Durante la respuesta temprana, las funciones moleculares más enriquecidas en 703671 fueron de actividad catalítica (GO:003824, p-value = 7.30E-29), de unión de moléculas (GO:0005488, p-value = 3.0E-29), de actividad oxidoreductasa (GO:0016491, p-value = 1.50E-16), y de actividad hidrolasa (GO:0016787, p-value = 3.5E-14), lo mismo que en 703248. Estas funciones están relacionadas y son parte importante durante los procesos de señalización celular para desencadenar la respuesta a la sequía.

Durante la respuesta tardía las funciones más enriquecidas fueron aquellas de unión a ribonucleótidos (GO:0032553, p-value = 2.8E-19), de unión a nucleótidos como el ATP (GO:0032553, p-value = 1.30E-16), y proteínas de choque térmico (GO:0031072, p-value = 8.8E-05).

- **Componente celular**

En cuanto a términos de componente celular, se encontró que durante la respuesta temprana de hojas de 703671 solo un término fue enriquecido correspondiente a región extracelular (GO:0005576, p-value = 7.30E-05), mientras que en 703248 el de núcleo (GO:0005634, p-value = 7.70E-05). En cuanto a raíces se encontró una marcada diferencia entre las variedades, pues en 703248 solo se encontraron términos de membrana (GO:0044425, p-value = 0.0018) y núcleo (GO:0005634, p-value = 0.001), mientras que en 703671 se encontraron por lo menos 33 términos de componente celular enriquecidos y entre ellos: plastidios, membrana fotosintética, complejo proteico, región extracelular, región integral de membrana, entre otras (Figura 19).

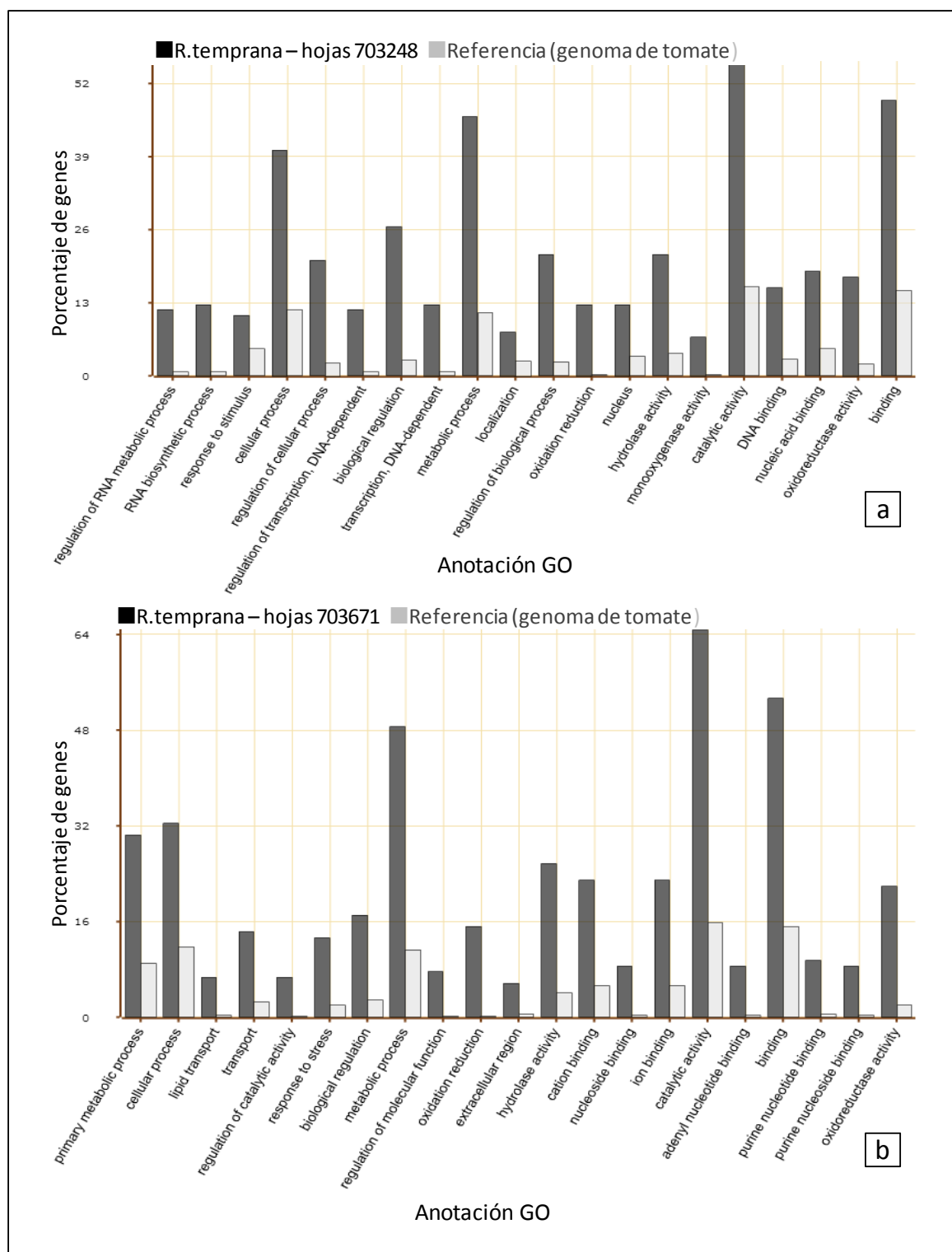


Figura 16. Categorías GO de los 200 transcritos más inducidos por sequía durante la respuesta temprana en hojas de 703248 (a) y 703671 (b). Lista referencia de agriGO: genoma de tomate.

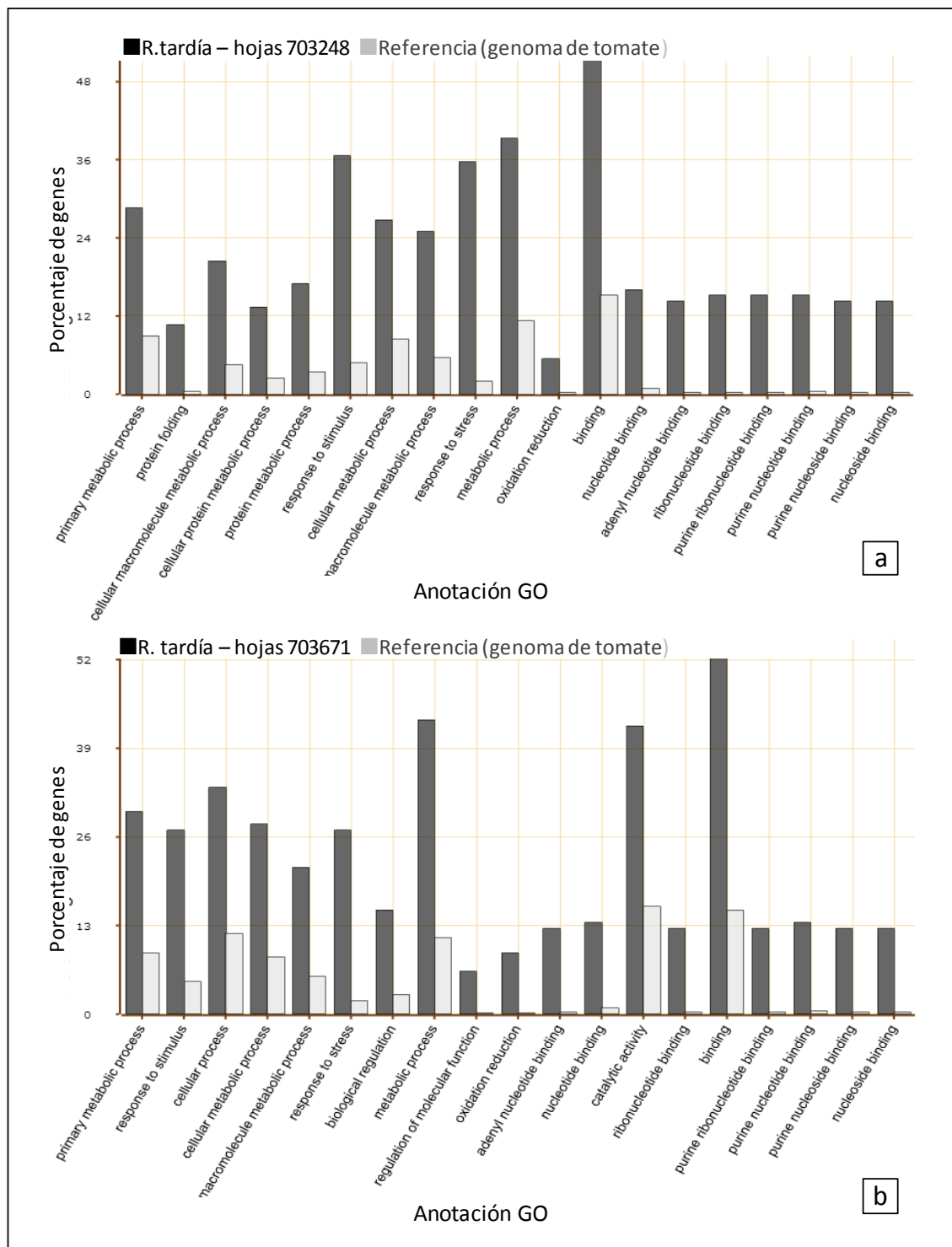


Figura 17. Categorías GO de los 200 transcritos más inducidos por sequía durante la respuesta tardía en hojas de 703248 (a) y 703671 (b). Lista referencia de agriGO: genoma de tomate.

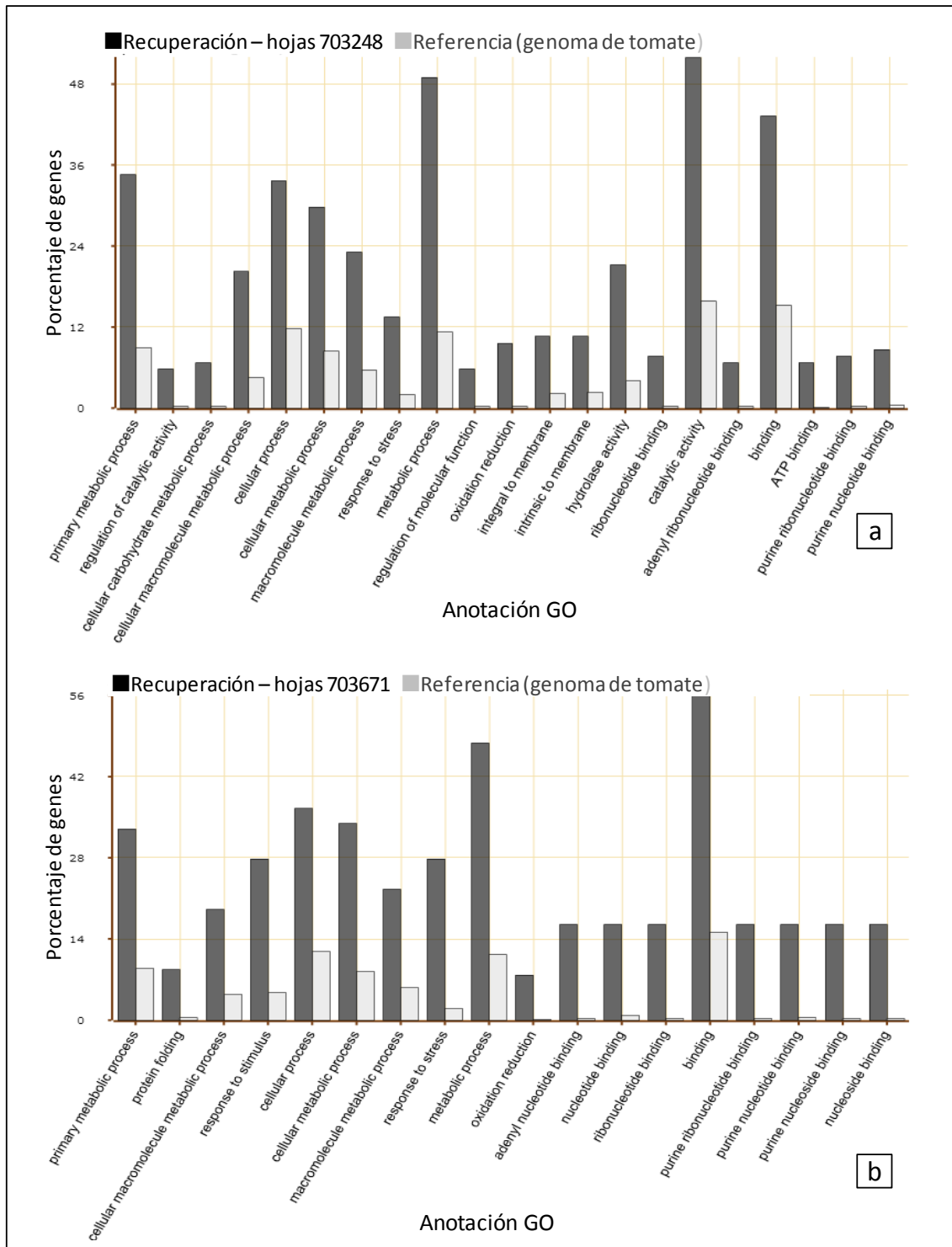


Figura 18. Categorías GO de los 200 transcritos más inducidos por sequía durante la recuperación en hojas de 703248 (a) y 703671 (b). Lista referencia de agriGO: genoma de tomate.

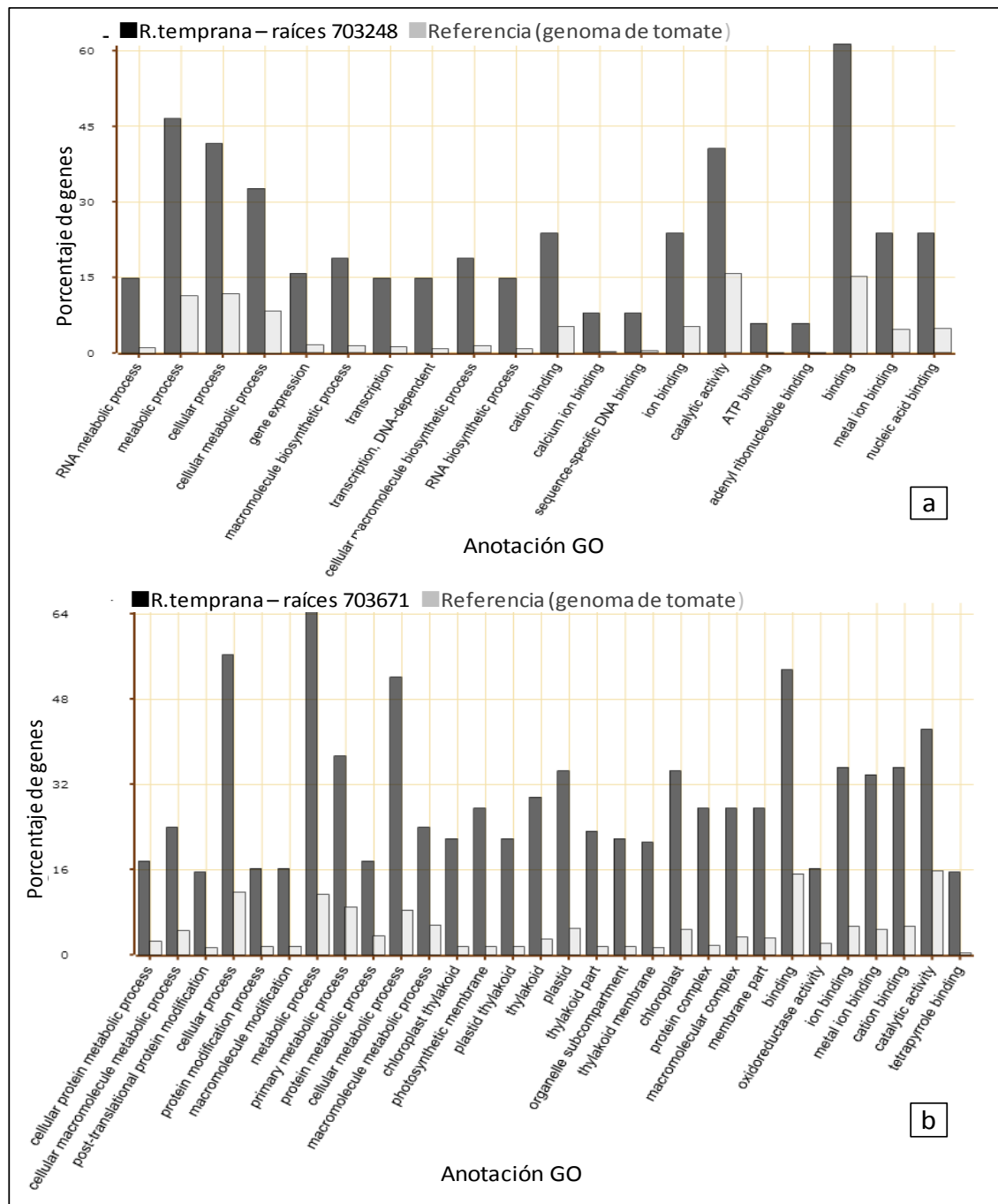


Figura 19. Categorías GO de los 200 transcritos más inducidos por sequía durante la respuesta temprana en raíces de 703248 (a) y 703671 (b). Lista referencia de agriGO: genoma de tomate.

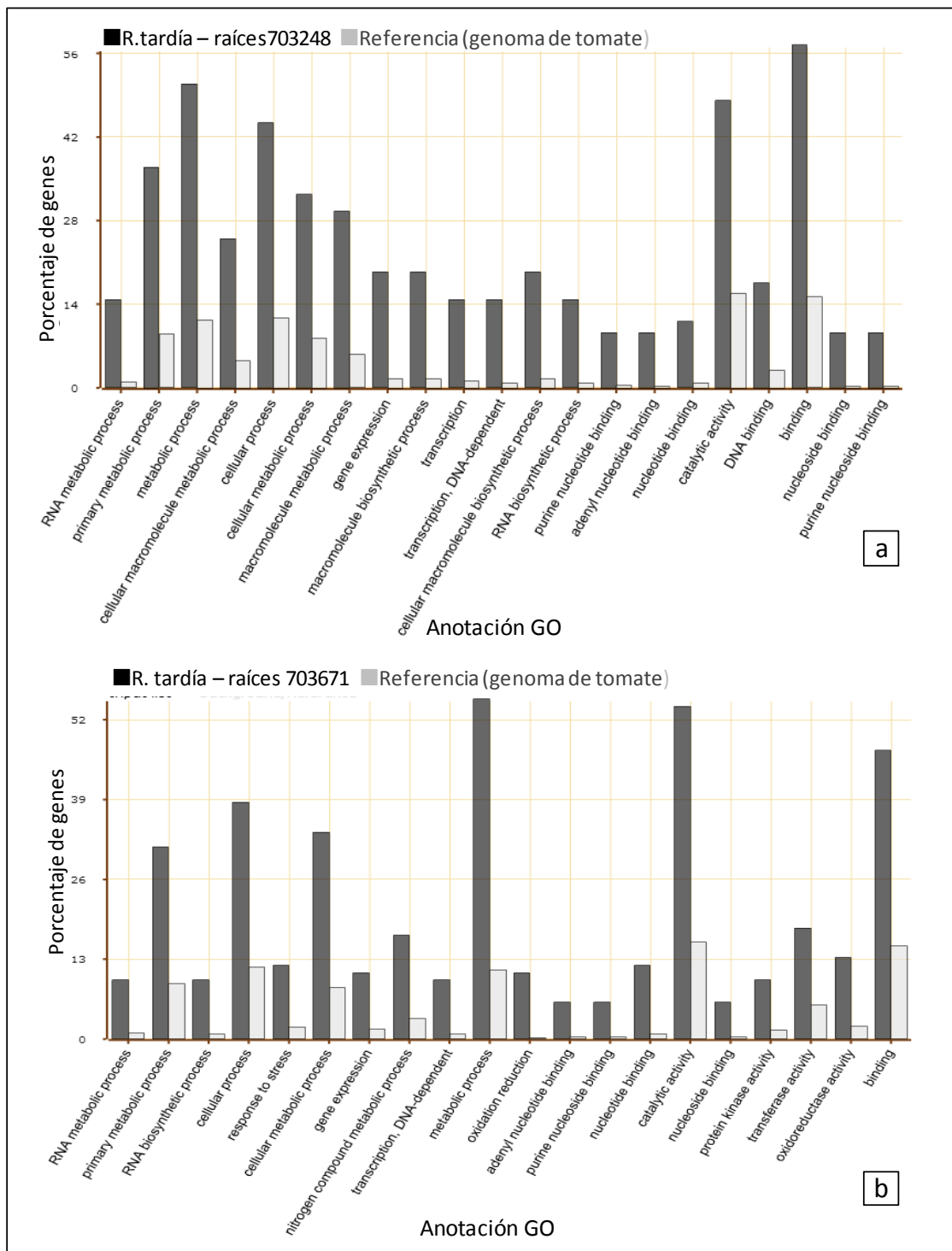


Figura 20. Categorías GO de los 200 transcritos más inducidos por sequía durante la respuesta tardía en raíces de 703248 (a) y 703671 (b). Lista referencia de agriGO: genoma de tomate.

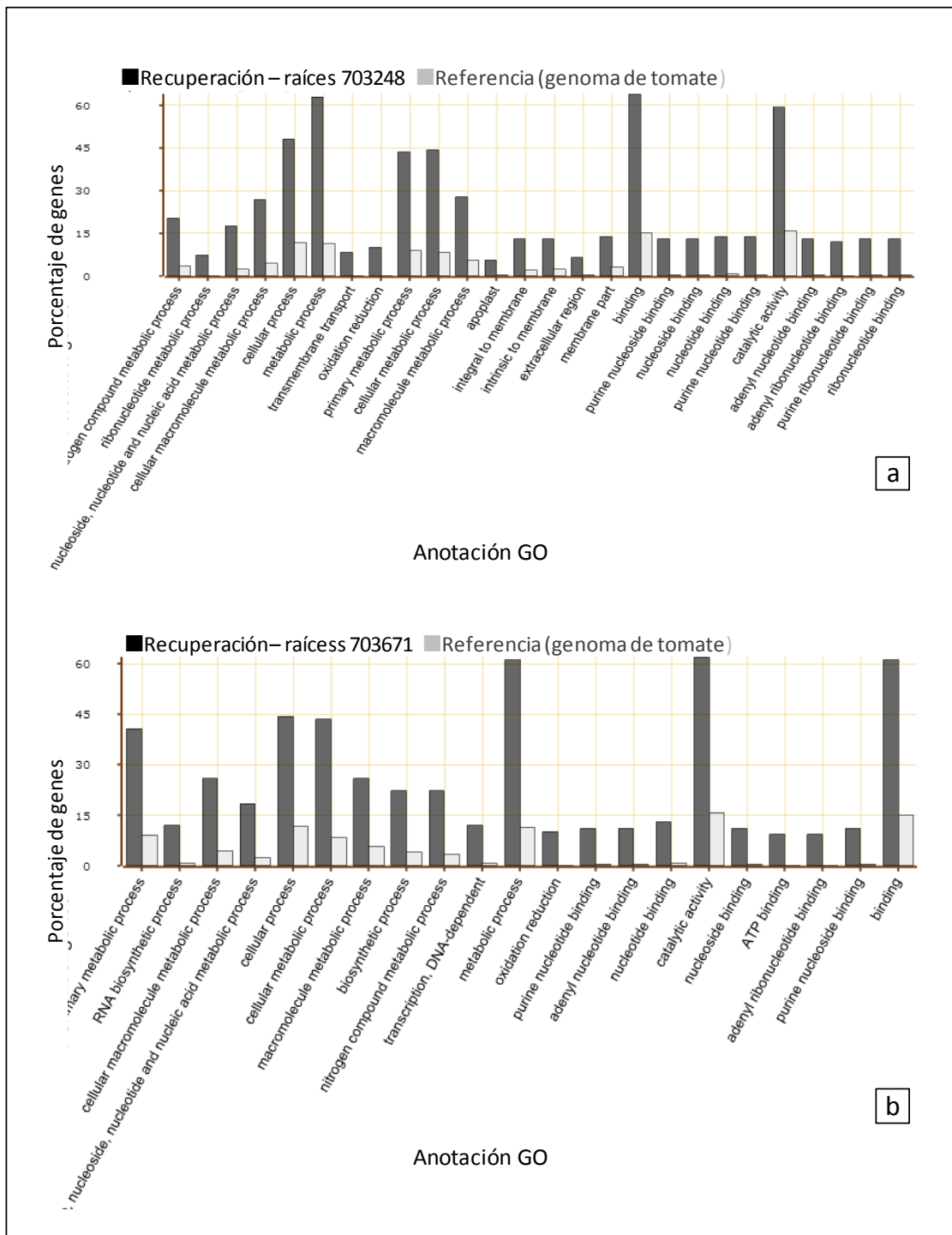


Figura 21. Categorías GO de los 200 transcritos más inducidos por sequía durante la recuperación en raíces de 703248 (a) y 703671 (b). Lista referencia de agriGO: genoma de tomate.

f. Redes de co-expresión génica

Con la finalidad de evaluar el perfil de expresión de genes o grupo genes a lo largo de la sequía y en la recuperación de esta, se trabajó con el perfil de expresión génica de cuatro grupos: hojas de 703248, raíces de 703248, hojas de 703671 y raíces de 703671

Luego de filtrar los datos (usando el coeficiente de variación= 0.7) se obtuvieron datos de expresión de 3234 genes de hojas de 703248, 3236 genes de raíces de 703248, 2790 genes de hojas de 703671 y 2991 genes de raíces de 703671, a lo largo de las etapas del estrés.

A partir de este filtro de datos, se generó para cada grupo un "HeatMap" en el que los genes más correlacionados fueron agrupados mediante un clusterig jerárquico, donde se puede observar y determinar preliminarmente el número de módulos formados (Figuras 22, 23). El color más rojo indica mayor correlación entre los genes. Se observaron 6 módulos en hojas de 703248, 8 módulos en hojas de 703671, 9 módulos en raíces de 703248 y 8 módulos en raíces de 703671.

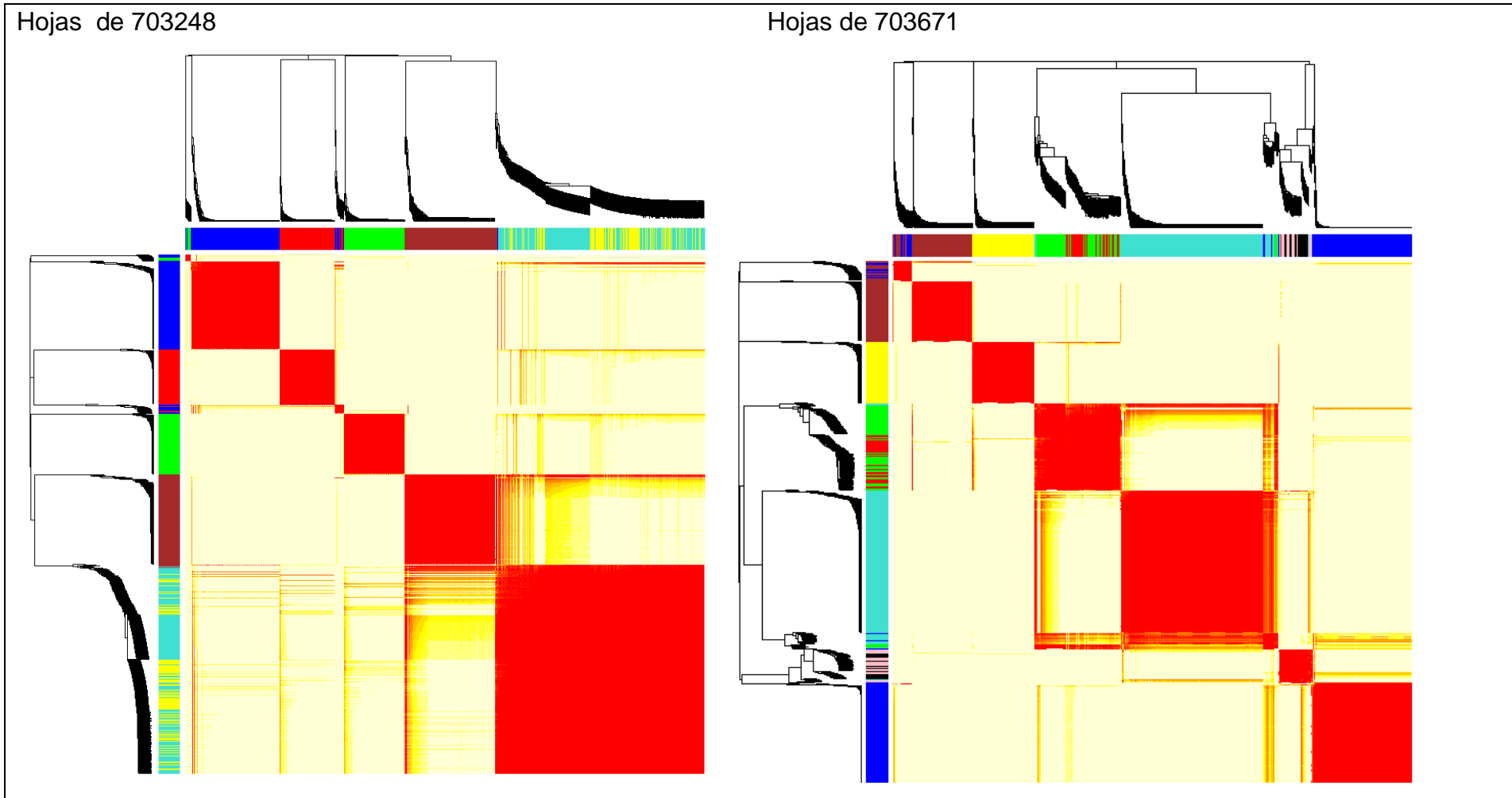


Figura 22. Heat Map de hojas de 703248 y 703671. El color indica el grado de correlación (rojo: mayor correlación).

Raíces de 703248

Raíces de 703671

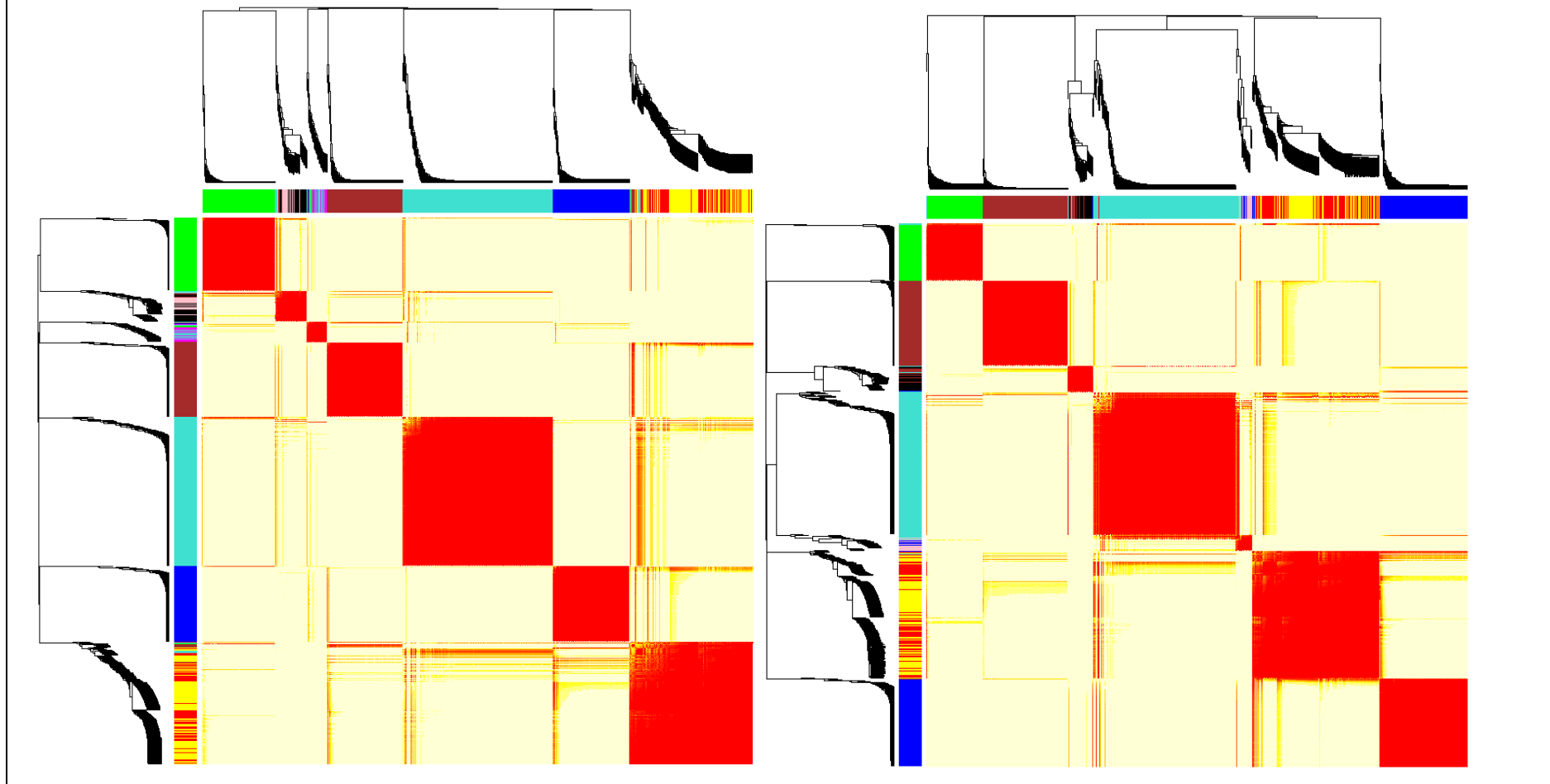


Figura 23. Heat Map de raíces de 703248 y 703671. El color indica el grado de correlación (rojo: mayor correlación).

El número de módulos identificados dentro de los tratamientos varió entre 6 y 9, conteniendo entre 41 y 931 genes, con un tamaño promedio de genes por módulo de 395. Cada módulo representa genes con un perfil de expresión altamente correlacionado, que probablemente pueden ser atribuidos a procesos específicos, esto quiere decir que los genes correlacionados dentro de un módulo podrían participar en procesos biológicos similares o de una misma ruta metabólica, lo cual representa información de mucho valor para estudios posteriores.

En hojas de 703248 se encontraron 6 módulos (Figura 24), cada uno con 591, 651, 630, 411, 538 y 413 genes. En hojas de 703671 se formaron 8 módulos (Figura 26), con 845, 289, 324, 47, 384, 86, 209 y 606 genes. En raíces de 703248 se encontraron 9 módulos (Figura 28), cada uno con 931, 288, 65, 49, 470, 493, 98, 427 y 415 genes. En raíces de 703671 se encontraron 8 módulos (Figura 30), cada uno con 896, 330, 526, 41, 71, 330, 350 y 447 genes.

Para estos módulos formados, se consolidó el perfil de expresión de cada módulo por un gen representativo: el denominado eigengene del módulo (Langfelder & Horvath, 2007). En las figuras 25, 27, 29 y 31 se observa el Heatmap de los eigengenes de hojas de 703248, hojas de 703671, raíces de 703248 y raíces de 703671 respectivamente.

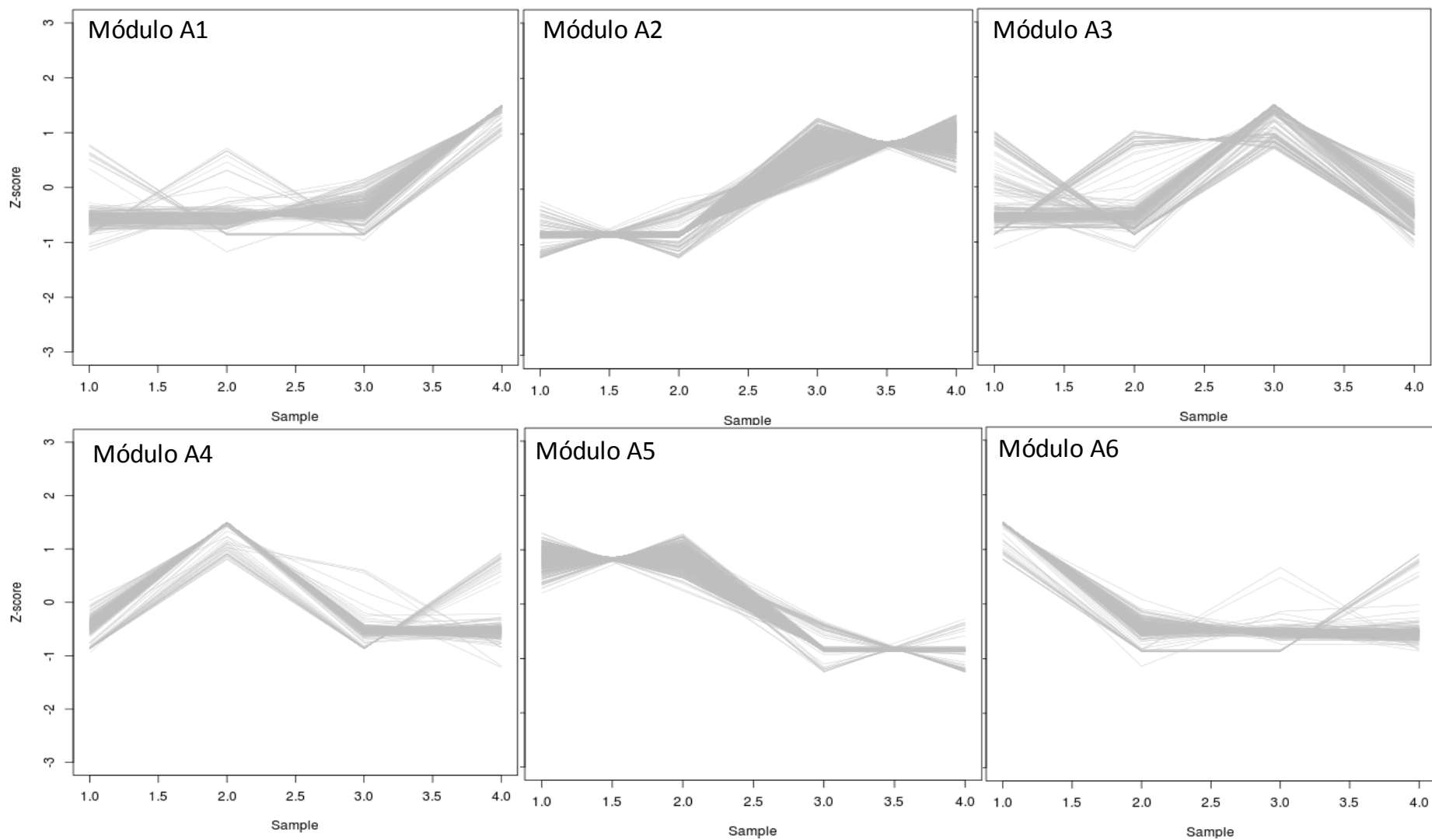


Figura 24. Gráficos de la tendencia de expresión de los módulos de genes formados en Hojas de 703248.

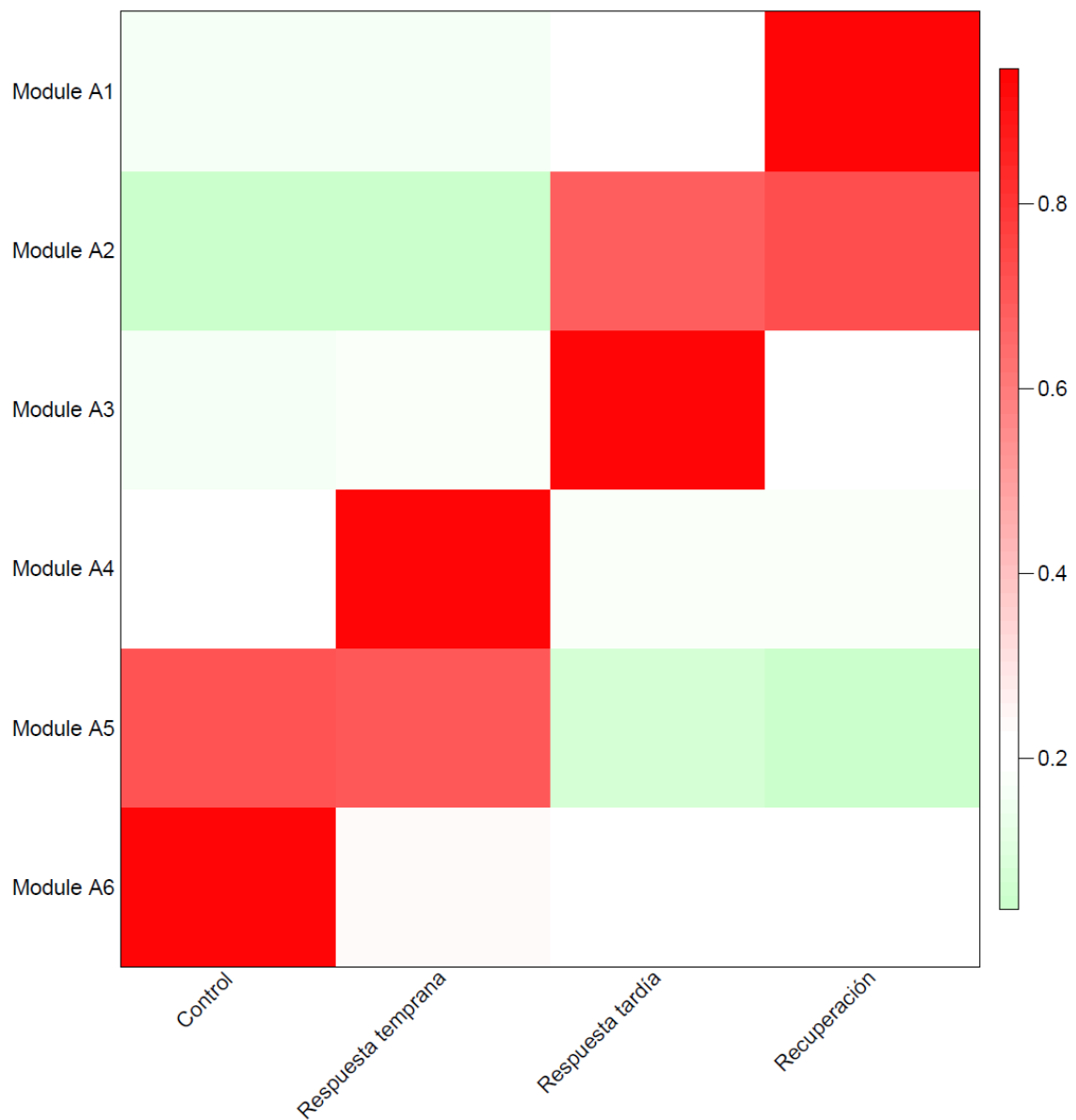


Figura 25. Heatmap de los eigengenes de los 6 módulos de co-expresión génica de hojas de 703248. Las columnas representan a los eigengenes, y las filas a los tiempos del experimento (Control, respuesta temprana, tardía y recuperación). La escala de colores indican los niveles relativos de expresión de todos los genes en el módulo.

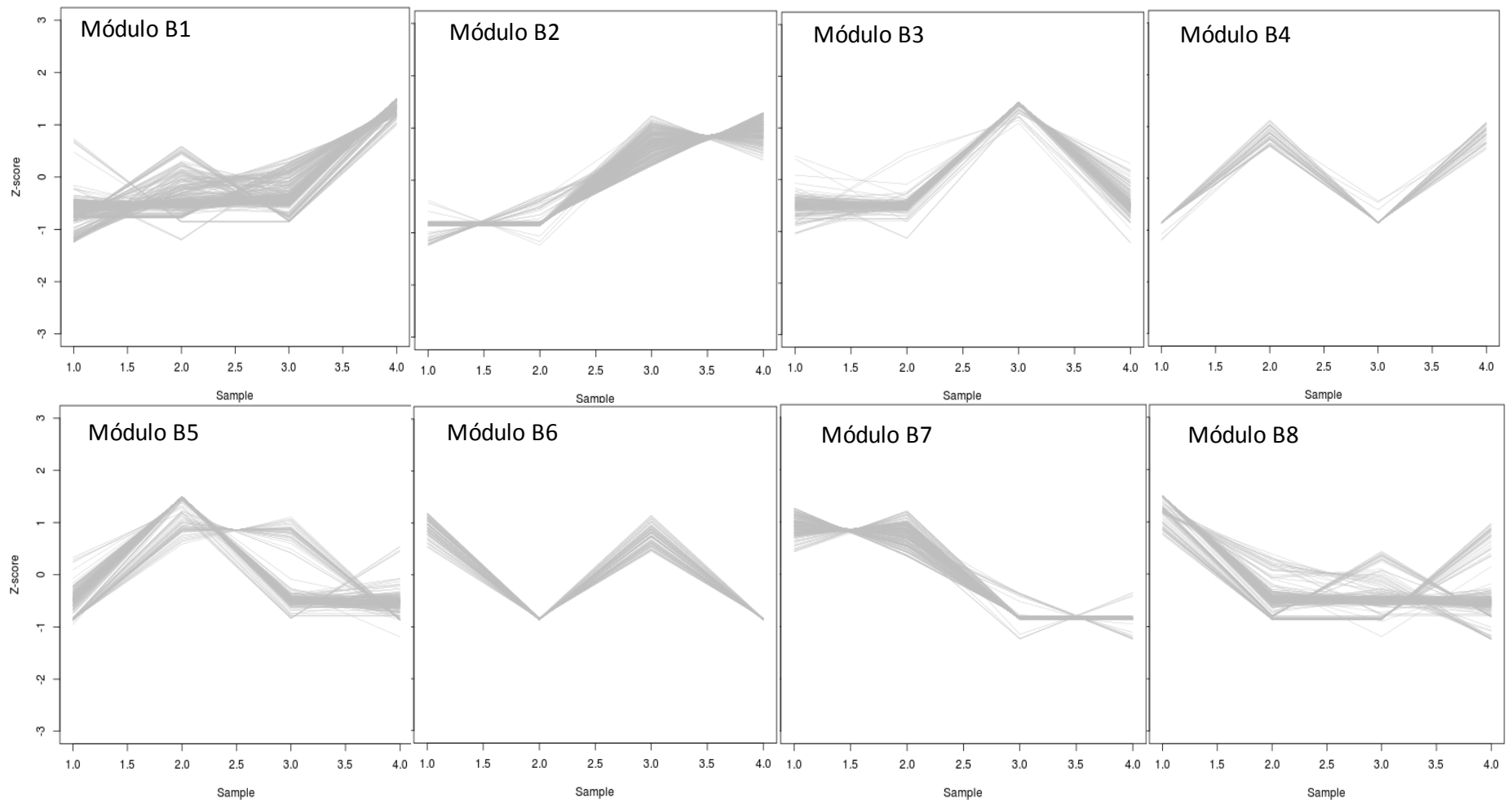


Figura 26. Gráficos de la tendencia de expresión de los módulos de genes formados en Hojas de 703671.

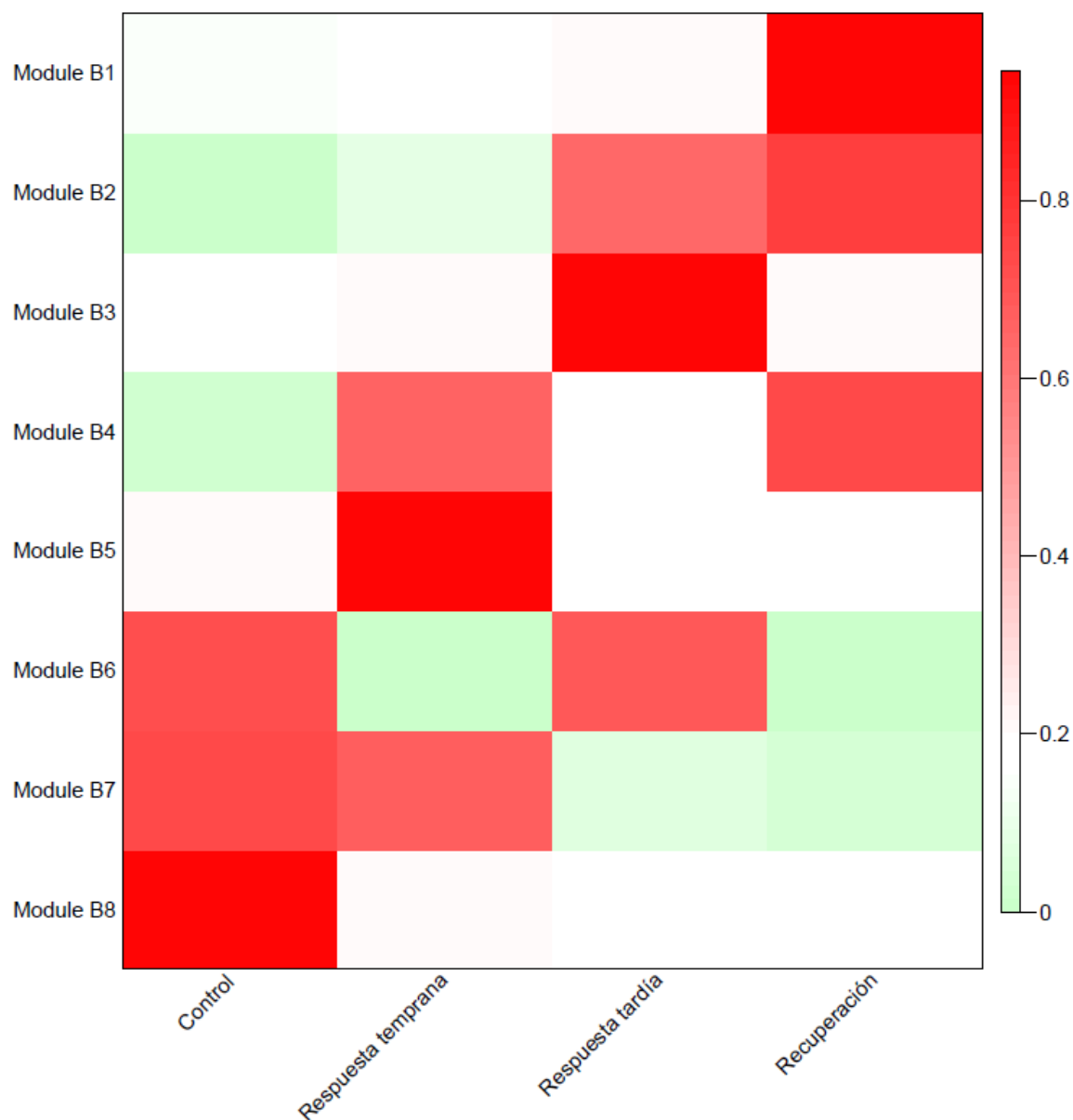


Figura 27. HeatMap de los eigengenes de los 6 módulos de coexpresión génica de hojas de 703671. Las columnas representan a los eigengenes, y las filas a los tiempos del experimento (Control, respuesta temprana, tardía y recuperación). La escala de colores indican los niveles relativos de expresión de todos los genes en el módulo.

En hojas de 703671, el módulo B1 presentó el mayor número de genes (845) y corresponde a aquellos que son solo inducidos significativamente en la etapa de recuperación. Entre los genes de este módulo se encontraron 29 de proteínas de choque térmico, 13 genes ACRE (elicitados por Avr9/Cf-9), 7 de citocromo P450, 6 proteínas LEA, algunas proteínas de unión a calcio y calmodulina, quinasas, fosfatasa y más de 100 genes de función desconocida (Tabla 16)

Tabla 16. Algunos genes del módulo de co-expresión B1 de hojas de 703671. Los valores indican el nivel de expresión expresado como \log_2 fpkm, TLC: Control, TL1: respuesta temprana, TL2: respuesta tardía, TL3: recuperación

Gene id	TLC	TL1	TL2	TL3	Anotación
PGSC0003DMG400030339	0	5.65	6.12	18.66	17.6 kD proteína pequeña de choque térmico clase I
PGSC0003DMG400030340	0	2.82	3.4	7.83	17.6 kD proteína pequeña de choque térmico clase I
PGSC0003DMG400030426	2.96	5.29	5.43	18.5	17.6 kD proteína pequeña de choque térmico clase I
PGSC0003DMG400019137	0	0	0	1.42	18.1 kDa proteína de choque térmico clase I
PGSC0003DMG400011073	0	0	1.36	4.12	Protein 137 elicitada por Avr9/Cf-9
PGSC0003DMG400010173	2.47	2.7	3.29	13.57	Protein 140 elicitada por Avr9/Cf-9
PGSC0003DMG400016498	0	1.05	0	1.68	Protein 146 elicitada por Avr9/Cf-9
PGSC0003DMG400016523	0	0	0	11.92	Protein 146 elicitada por Avr9/Cf-9
PGSC0003DMG400021563	1.14	1.7	3.4	15.45	Proteina de unión a calcio
PGSC0003DMG400030608	3.45	3.61	3.76	14.47	Protein CML24 de unión a calcio
PGSC0003DMG400008202	1.55	2.96	3.67	14.95	Proteina de unión a calmodulina
PGSC0003DMG400000067	3.79	4.22	4.84	15.91	Proteína LEA
PGSC0003DMG400000628	0	3.36	5.45	18.53	Proteína LEA

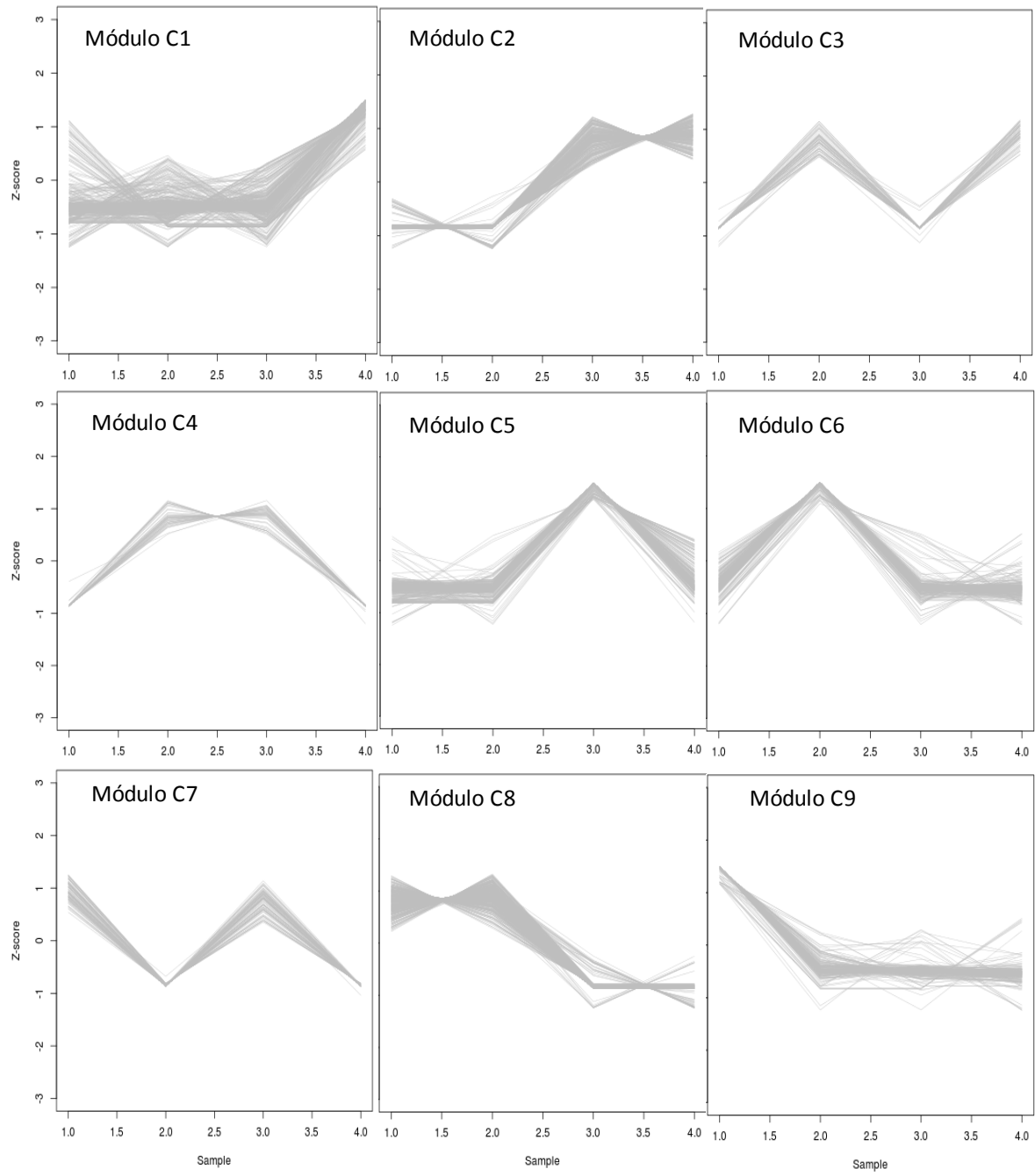


Figura 28. Gráficos de la tendencia de expresión de los módulos de genes formados en raíces de 703248.

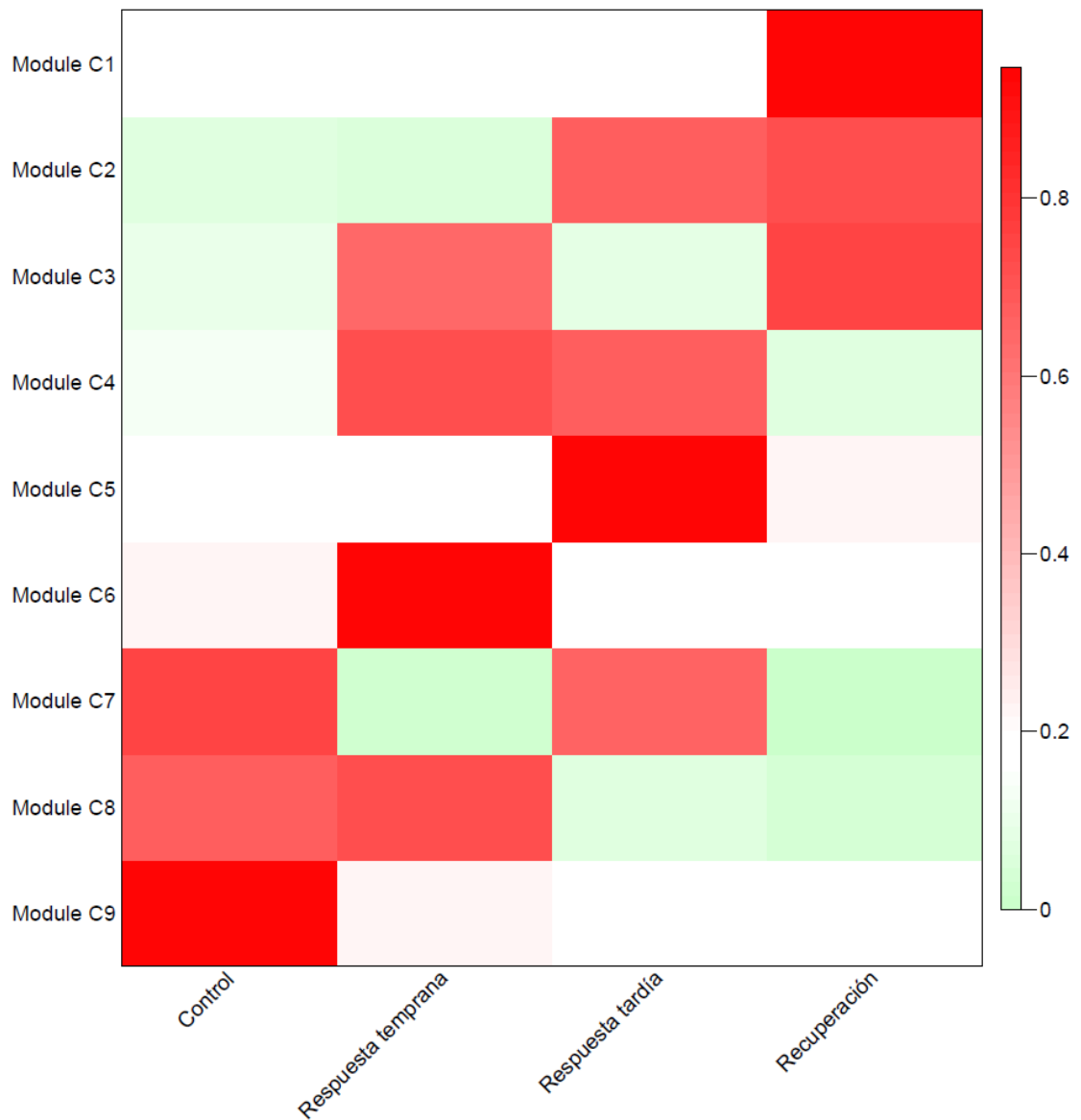


Figura 29. HeatMap de los eigengenes de los 6 módulos de coexpresión génica de raíces de 703248. Las columnas representan a los eigengenes, y las filas a los tiempos del experimento (Control, respuesta temprana, tardía y recuperación). La escala de colores indican los niveles relativos de expresión de todos los genes en el módulo.

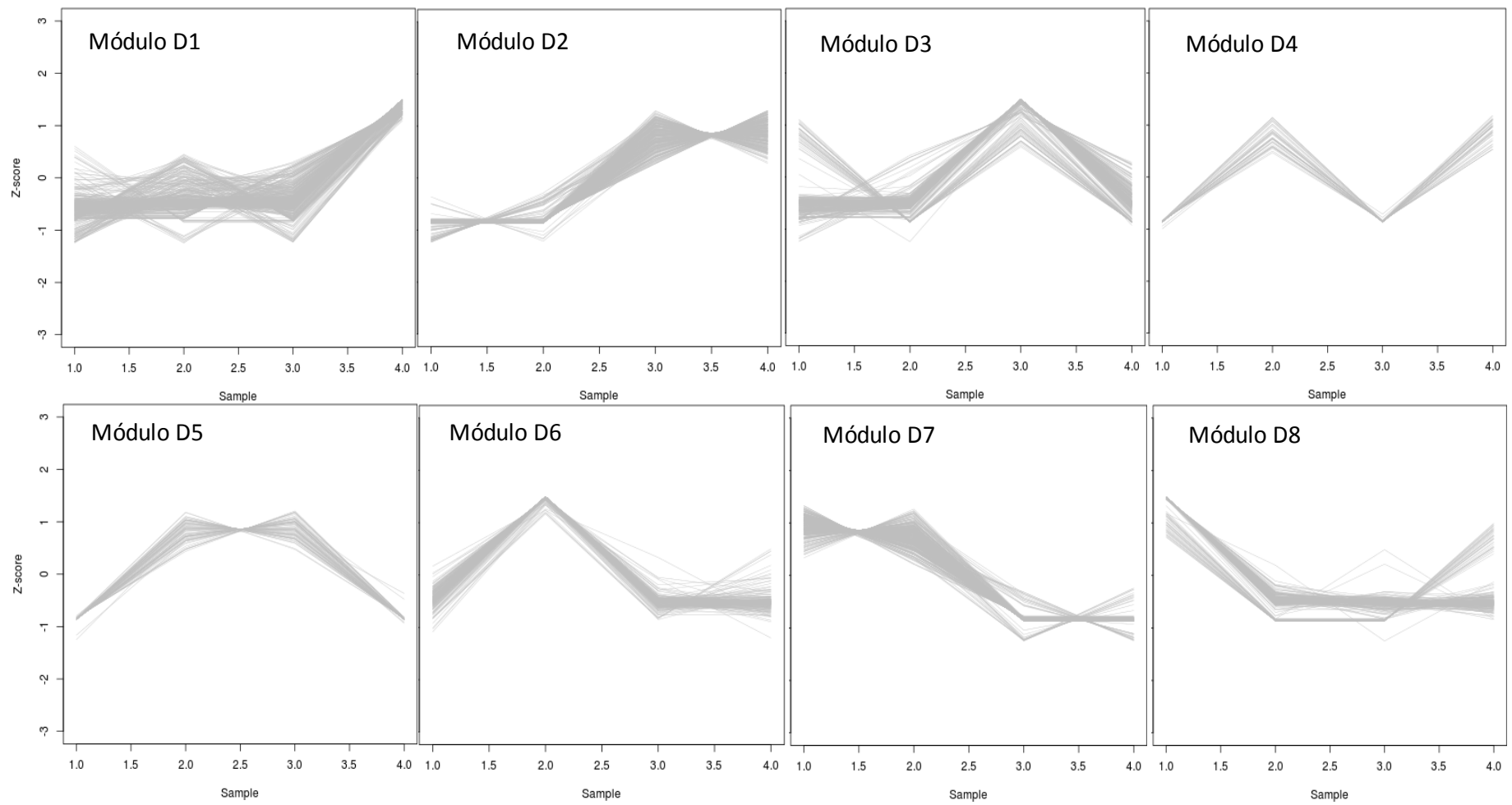


Figura 30. Gráficos de la tendencia de expresión de los módulos de genes formados en raíces de 703671.

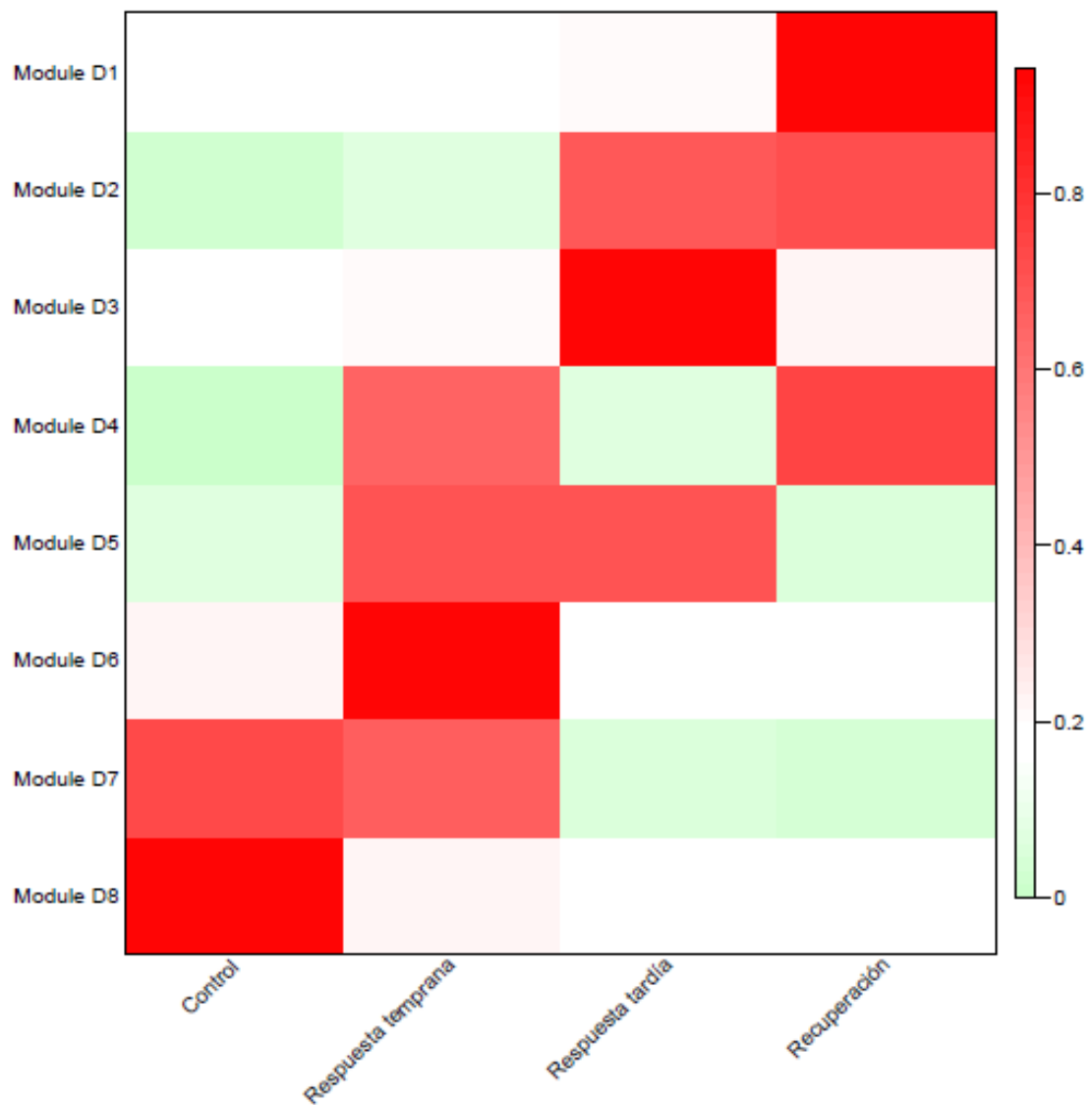


Figura 31. HeatMap de los eigengenes de los 6 módulos de coexpresión génica de raíces de 703671. Las columnas representan a los eigengenes, y las filas a los tiempos del experimento (Control, respuesta temprana, tardía y recuperación). La escala de colores indican los niveles relativos de expresión de todos los genes en el módulo.

VII. DISCUSIONES

El secuenciamiento del RNA es una herramienta valiosa que en pocos años, desde su aparición en el 2008, ha permitido caracterizar y cuantificar los transcriptomas de diversas especies (Nagalakshmi *et al.* 2008). En este trabajo se empleó un secuenciamiento completo del mRNA para caracterizar por primera vez los transcriptomas de hojas y raíces de dos variedades de *Solanum tuberosum* subsp. *andigena* bajo condiciones de sequía *in vivo*. La papa es considerada un cultivo susceptible a la sequía (van Loon, 1981; Yang *et al.*, 2003), por lo que el secuenciamiento de su transcriptoma brinda enormes recursos para futuras investigaciones en esta especie así como en otros miembros de la familia Solanaceae, que tengan por fin entender las bases moleculares de tolerancia a la sequía y mejorar los cultivos genéticamente.

La selección de las variedades bajo estudio se basó en resultados de la caracterización fisiológica del proyecto PapaSalud (CIP), que consideran a los clones 703671 y 703248 como variedades tolerante y susceptible a sequía, respectivamente. La tasa fotosintética de estos genotipos monitoreada antes durante y después del estrés, estuvo dentro de lo esperado (Vasquez-Robinet *et al.*, 2008), es decir, una disminución de la tasa fue más acelerada en la variedad susceptible (703248) que en la tolerante (703671). La fotosíntesis es uno de los procesos clave a ser afectados por la sequía, parámetro que fue considerado para la determinación de los tiempos de muestreo. Se han reportado diferencias en la respuesta fotosintética al estrés hídrico entre las variedades de papa, donde las tolerantes mantienen una mayor fotosíntesis debido a un ajuste estomático adecuado, permitiendo la eficiencia del uso del agua (Martínez & Moreno, 1992).

Se generaron en total más de 392 millones de lecturas con un número promedio de 23.5 millones de lecturas por biblioteca. En el 2011, Massa *et al.* secuenciaron el transcriptoma de plántulas de *Solanum tuberosum* Grupo Phureja clon DM1-3516R44 expuestas a manitol (260 μ M) por 24 horas, un tratamiento *in vitro* que simula una condición de sequía. Ellos generaron 15.5 millones de lecturas,

número menor al reportado en este trabajo, lo que nos permitió tener una mayor cobertura para la identificación de los genes relacionados a sequía.

La variación en el número de lecturas por biblioteca generadas en este trabajo (14 - 30 millones), fue menor en comparación a lo reportado por Massa *et al.* en el 2011, quienes secuenciando 32 transcriptomas de DM1-3516R44 obtuvieron entre 5 y 30 millones de lecturas por biblioteca. La uniformidad en la construcción de las bibliotecas generadas en un estudio resulta importante al momento de realizar comparaciones entre estas y se ve reflejada en la menor variación del número de lecturas (Head *et al.*, 2013).

El potencial de sesgo en las comparaciones que se realizaron entre las bibliotecas fue determinado en base al número total de lecturas mapeadas al genoma de referencia del clon de papa DM1-3516R44 (PGSC, 2011). La poca variación encontrada en el número mapeado (87% como mínimo y 91% como máximo), indica un desempeño similar en la construcción y secuenciamiento de las bibliotecas de este estudio. Del mismo modo, se obtuvo un 87.5% de lecturas mapeadas al secuenciar el transcriptoma de hojas de este clon y mapearlo a su propio genoma,

La cobertura del secuenciamiento o el porcentaje de genes identificados es otro factor importante a considerar en la técnica de RNA-Seq, ya que a mayor número de lecturas generado mayor será el número de genes que podrán ser identificados y cuantificados. Se sabe que el número de genes identificados por esta técnica permite alcanzar un 80% de cobertura con bibliotecas de 4 millones de lecturas (Mortazavi *et al.*, 2008). La cobertura del secuenciamiento en este trabajo fue la adecuada, pues no se observaron diferencias significativas en el número de transcritos identificados, encontrándose 27000 para la biblioteca con el menor número de lecturas (14 millones) y 31000 para la biblioteca con el mayor número (30 millones). Massa *et al.* (2011) tampoco encontraron una correlación entre el número de genes identificados/ biblioteca y el número de lecturas/biblioteca, aun inclusive teniendo una mayor variación entre sus bibliotecas.

Las comparaciones realizadas de cada tratamiento con su respectivo control permitieron detectar los cambios en la expresión génica asociados con el nivel de estrés. El número de genes con expresión diferencial a lo largo de la exposición a la sequía encontrado (887 como mínimo en la respuesta temprana de hojas de 703248 y 1995 como máximo en la recuperación de raíces de 703671) fue similar a lo reportado por Massa *et al.* (2013) para *Solanum tuberosum* Grupo Phureja, quienes empleando RNA-Seq reportaron 2262 genes sobreexpresados en respuesta a una combinación de parámetros de estrés abiótico como sal, calor y manitol.

Además, existe una relación directa entre el número de genes con expresión diferencial y la intensidad de la sequía en hojas y raíces, tolerantes y susceptibles. De estos, son los genes inducidos los que mantienen esta tendencia, no así en los genes reprimidos donde no se encontró una tendencia de cambio. Esto debido quizás a que las mayores exigencias ambientales requieren una serie de cambios moleculares en las células que permitan a las plantas responder o tolerar determinado estrés.

La etapa de recuperación presentó el mayor número de genes diferencialmente expresados tanto en hojas como raíces. La presencia del mayor número de genes inducidos en raíces de 703671 podría estar asociada a su rápido restablecimiento luego del estrés. Oono *et al.* (2003) reportan que los genes inducidos en la recuperación del estrés de *Arabidopsis* estuvieron conformados por proteínas regulatorias, proteínas funcionales responsables de la recuperación del estrés, y proteínas funcionales involucradas en el crecimiento y fotosíntesis. De igual forma, algunos de los genes más inducidos en la recuperación de 703671 se encuentran distribuidos entre proteínas regulatorias como factores de transcripción DREB, fosfatasa 2C y principalmente en las encargadas de la recuperación del estrés, como fosfolipasas, peroxidasas, proteínas de choque térmico, inhibidor de proteasas y proteínas LEA. En este trabajo el transcrito más inducido (PGSC0003DMG400000628) codifica una proteína LEA indicando su posible rol para la recuperación de *S. tuberosum* subsp. *andigena*. Se sabe que las proteínas LEA se acumulan en las plantas durante el déficit hídrico, siendo su

función la de proteger a enzimas, complejos proteicos y membranas (Vicré *et al.*, 2004).

Si bien la recuperación es la etapa con mayores cambios de expresión de genes, al inspeccionar el número de genes únicos a cada etapa del estrés se encontró en la respuesta tardía un mayor número de genes específicos inducidos o reprimidos respecto a la etapa anterior. La mayor intensidad de sequía desencadena mayores cambios en el transcriptoma de hojas y raíces de ambas variedades, donde posteriormente muchos de estos genes mantienen, aumentan o reprimen su expresión en la recuperación y un menor número de genes inducidos o reprimidos son específicos de esta etapa. En hojas tolerantes se encontró la sobreexpresión de una proteína RD22, también reportada como inducida por estrés salino e hídrico y por la aplicación de ABA en plantas de *Arabidopsis* (Yamaguchi-Shinozaki & Shinozaki, 1993). También se encontraron varias proteínas LEA recién inducidas en la respuesta tardía y cuya expresión se mantuvo durante la recuperación, debido a su rol de estabilización de proteínas, de membranas, de estructura de la cromatina, o renaturando proteínas desplegadas (Ingram & Bartels, 1997).

Al aparecer la regulación negativa de genes durante la etapa temprana tiene un efecto positivo en la tolerancia de la variedad 703671 a la sequía. El número de genes expresados diferencialmente en hojas tolerantes durante la respuesta temprana fue casi el doble (1506) respecto a las hojas susceptibles (887) y 50% mayor al encontrado en raíces tolerantes (998), siendo la mayor parte de ellos genes reprimidos (1085), de los cuales 809 fueron específicos de esta variedad. Algunos de estos genes están relacionados con la inhibición del crecimiento celular y de la fotosíntesis, como una respuesta de aclimatación a la sequía. Este comportamiento permitiría restringir la pérdida de agua en las etapas iniciales para mantener el estatus hídrico de la planta y la asimilación de carbono.

Entre los genes que podrían otorgar tolerancia a la sequía se encontraron varios asociados a la división celular (cdc45, ciclinas, reguladores de citoquinesis, extensina, expansina) en concordancia con una regulación del crecimiento

durante la sequía. También se encontraron genes SAUR de respuesta a auxina cuya función aún se desconoce, aunque algunos autores lo reportan presentes en tejidos blancos de elongación celular inducida por auxina y asociados por tanto a este proceso (Kant *et al.*, 2009; Gil *et al.*, 1994). Otro grupo de genes encontrados fueron los de la familia citocromo P450; en *Arabidopsis*, Kushiro *et al.* (2004) reportaron que un gen de esta familia (CYP707A) codifica a una hidroxilasa, una enzima clave del catabolismo y regulación de los niveles de ABA en condiciones normales y de estrés en plantas.

En la variedad tolerante se encontró un patrón interesante en cuanto a los genes reprimidos, presentado en hojas 1085 genes, y en raíces solo 359. Esto podría deberse a que plantas expuestas al déficit hídrico presentan un desarrollo del sistema radicular menos inhibido que el crecimiento de hojas, e incluso puede ser promovido. El mantenimiento del crecimiento radicular durante el estrés representa un beneficio para mantener un adecuado suministro de agua, proceso que se encuentra regulado molecularmente (Sharp *et al.*, 2004).

La comparación de los genes expresados diferencialmente entre las variedades nos permite identificar a aquellos que solo cambian su expresión en 703671, es decir específicos de esta variedad, y que podrían ser los responsables de su tolerancia. Las mayores diferencias entre las variedades se encontraron tanto en la respuesta temprana de hojas (donde el número de genes reprimidos en 703671 cuadruplica al número en 703248), como en todas las etapas del estrés en raíces donde el número de genes inducidos en 703671 duplica a 703248. El estudio de estos genes que son inducidos/reprimidos solo en la variedad tolerante podría dar luces sobre los mecanismos moleculares que hacen posible la resistencia de *S. tuberosun* subsp. *andigena* a la sequía.

Durante la respuesta temprana de hojas se encontraron varios genes específicos de la variedad tolerante que se sabe juegan un rol importante durante la sequía. Se encontraron proteínas de unión a DRE, proteínas RD22 de respuesta a deshidratación, una proteína LEA5, factor de transcripción MYB139 que regulan los movimientos del estoma permitiendo tolerar el estrés (Cominelli *et al.*, 2005), y

una proteína taumatina asociada a regulación osmótica. Por otro lado, se encontró durante la recuperación de hojas tolerantes una marcada sobreexpresión de genes que codifican proteínas de choque térmico de cloroplastos, en comparación a las susceptibles. Esto podría contribuir en el rápido restablecimiento de la variedad tolerante después de la sequía, pues se ha visto que estas proteínas están involucradas en la protección del transporte de electrones del fotosistema II durante varios estreses abióticos (Downs *et al.*, 1999).

La respuesta de hojas y raíces a la sequía mostró mayormente una expresión de genes específicos. En hojas se encontraron proteínas de unión a clorofila, cuya expresión es inducida específicamente; pequeñas proteínas de choque térmico de cloroplastos inducidas; así como pequeñas proteínas de choque térmico 17.6 kD clase I inducidas desde la respuesta temprana en hojas y en la respuesta de recuperación en raíces, mostrando nuevamente el rol clave de este tipo de proteínas para el restablecimiento de las funciones celulares en hojas y raíces.

En cuanto a genes específicos de raíz, se encontró una represión de varios genes de extensinas. En tomate, se ha reportado una expresión de estos genes predominantemente en raíces sobre otros tejidos, encontrándose que su expresión es afectada por factores medioambientales (Bucher *et al.*, 1997). Por tanto es posible que la supresión de estos, en papa, sea la de regular y mantener el crecimiento radicular durante el estrés. Finalmente, se encontraron genes de exostosinas reprimidos en raíces, cuya función en plantas aun no ha sido estudiada, sin encontrarse reportes de alguna relación de estos con la tolerancia a algún tipo de estrés.

Para un mejor entendimiento de los genes involucrados en la respuesta a sequía en papa, se determinaron las clases funcionales de los genes diferencialmente expresados empleando Blast2GO y la herramienta AgriGO. "Gene ontology" ofrece un vocabulario estructurado y organizado para la descripción de genes de acuerdo a tres ontologías: función molecular, proceso biológico y componente celular (The Gene Ontology Consortium, 2000).

Dentro de los términos de procesos biológicos más enriquecidos en la respuesta temprana se encontraron aquellos de procesos metabólicos, transcripción génica, y regulación génica. Esto indica la importancia y el rol que juegan los procesos de regulación génica durante el inicio del estrés en plantas para iniciar la señalización celular y desencadenar la activación o no de proteínas funcionales (Shinozaki & Yamaguchi-Shinozaki, 2007). Posteriormente, durante la respuesta tardía, los términos más enriquecidos fueron los de respuesta a estrés, entre ellos términos de respuesta a estrés osmótico que va de la mano con la sequía, de respuesta a la privación de agua y de respuesta a desecación.

Durante la respuesta tardía, algunos de los términos de función molecular más enriquecidos fueron: de unión a nucleótidos, actividad que no ha sido bien estudiada, pero se cree que ayuda a mantener la homeostasis celular durante un estrés hídrico que conlleva a un estrés salino (Zhu *et al.*, 2000); de unión a RNA que codifica dehidrasas (implicadas en la protección celular bajo condiciones de déficit hídrico) y que son moduladas por quinasas (Li *et al.*, 2002). Otro término enriquecido fue el de proteínas de choque térmico, cuyo rol funcional durante el estrés está bien documentado (Wang *et al.*, 2004).

El análisis de redes de co-expresión génica (WGCNA) correlaciona los niveles de expresión de genes para construir las redes y definir módulos de genes altamente interconectados (Langfelder & Horvath, 2008). Massa *et al.* (2013) encontraron en *S. tuberosum* Grupo Phureja bajo condiciones de estrés abiótico *in vitro* (sal, calor y manitol) 13 módulos de co-expresión conteniendo 4260 genes. En este análisis se encontraron 31 módulos de co-expresión conteniendo un total de 12251 genes, con un promedio de 395 genes por módulo. Los genes de cada módulo pueden ser atribuidos a procesos específicos, esto quiere decir que los genes correlacionados dentro de un módulo podrían participar en procesos biológicos similares o de una misma ruta metabólica, lo cual representa información de mucho valor para estudios posteriores.

El módulo B1 de hojas de 703671 comprende a aquellos que son solo inducidos significativamente durante la recuperación, encontrándose 13 genes ACRE

involucrados en una respuesta de defensa principalmente a factores abióticos; estos codifican proteínas regulatorias que incluyen quinasas y factores de transcripción (Rowland *et al.* 2005) y que probablemente también desempeñan un rol importante en las hojas de papa durante la etapa de recuperación. También forman parte de este módulo 29 genes que codifican para proteínas de choque térmico (incluyendo las pequeñas y de cloroplastos), un número bastante alto en comparación a otros módulos. Así por ejemplo, en el módulo B3 y B2 (genes inducidos en la respuesta tardía) se encontraron solo 3 y 2 genes, respectivamente, mientras que en el módulo A1 (genes inducidos en la recuperación de hojas de 703248) se encontraron 5 genes.

En plantas, las proteínas pequeñas de choque térmico se acumulan rápidamente en condiciones de sequía y se mantienen durante la recuperación del estrés para restablecer la conformación normal de las proteínas y enzimas, y por tanto la homeostasis celular (Wang *et al.* 2004). Estas proteínas podrían tener un rol importante también en papa, permitiendo la mejor recuperación de las plantas luego de la sequía.

Aunque la co-expresión no implica necesariamente que los genes estén funcionalmente relacionados, es muy probable que los genes dentro de los módulos generados en este trabajo tengan una mayor similaridad funcional entre sí comparado con aquellos genes que no comparten su perfil de expresión. La identificación de co-expresión de los más de 12000 genes analizados permite tener una visión general del transcriptoma de hojas y raíces de *S. tuberosum* subsp. *andigena*, información que sirve de base para futuras investigaciones en la expresión génica de papa y de otras especies cercanas.

VIII. CONCLUSIONES

- Se generaron más de 392 millones de lecturas de *Solanum tuberosum* subsp. *andigena* empleando el RNA-Seq con 23.5 millones por biblioteca e identificando desde 27188 transcritos de 703671 hasta 32746 transcritos de 703248 durante la recuperación de raíces.
- El menor número de genes con expresión diferencial a lo largo de la exposición a la sequía se encontró en la respuesta temprana de hojas de la variedad susceptible (887) y el mayor durante la recuperación de raíces de la tolerante (1995).
- La etapa de recuperación presenta el mayor número de genes diferencialmente expresados para ambos tejidos de las dos variedades, mientras que en la respuesta tardía existe un mayor número de genes específicos inducidos y reprimidos con respecto a la respuesta tardía.
- Las mayores diferencias de expresión de genes entre las variedades ocurren en la respuesta temprana de hojas (donde los genes reprimidos en 703671 cuaduplican en número a 703248); y durante todas las etapas del estrés en raíces (donde los genes inducidos en 703671 duplican en número a 703248).
- Los términos de procesos biológicos más enriquecidos en la respuesta temprana están asociados a procesos metabólicos, transcripción génica, y regulación génica; mientras que en la respuesta tardía están asociados a respuestas a estrés, a estrés osmótico, a la privación de agua y a desecación.
- Se generaron 31 módulos de genes de respuesta a sequía altamente correlacionados, conteniendo un total de 12251 genes, con un tamaño promedio de 395 genes/módulo.

IX. RECOMENDACIONES

- Para obtener conclusiones robustas del secuenciamiento de transcriptomas es necesario el uso de réplicas biológicas, que en este trabajo no fue posible debido al elevado costo de la técnica del RNA-Seq.
- Es muy importante el estudio y análisis posterior a profundidad de toda la data generada en esta tesis, que permita seleccionar genes que puedan ser utilizados en programas futuros de mejoramiento genético de *Solanum tuberosum* L.

X. REFERENCIAS BIBLIOGRÁFICAS

- Alves A. & Setter T. (2004). Response of Cassava Leaf Area Expansion to Water Deficit: Cell Proliferation, Cell Expansion and Delayed Development. *Annals of Botany* 94, 605–613.
- Alexandersson E., Fraysse L., Sjövall-Larsen S., Gustavsson S., Fellert M., Karlsson M. & Kjellbom P. (2005). Whole gene family expression and drought stress regulation of aquaporins. *Plant molecular biology*, 59(3), 469-484.
- Athar, H. & Ashraf M. (2005). Photosynthesis under drought stress. In: *Hand Book Photosynthesis*, 2nd edition, M. Pessarakli (ed.). CRC Press, New York, USA, 795-810.
- Bansal K. & Nagarajan S. (1986). Leaf water content, stomatal conductance and proline accumulation in leaves of Potato (*Solanum tuberosum* L.) in response to water stress. *Indian J. Plant Physiol.* 29, 397-404.
- Bartels D., Phillips J. & Chandler J. (2007). Desiccation Tolerance: Gene Expression, Pathways, and Regulation of Gene Expression. En Jenks M. & Wood A. (edit) *Plant dessication tolerance*, Iowa, USA (pp. 115-137).
- Bartels D., & Sunkar R. (2005). Drought and salt tolerance in plants. *Crit. Rev. Plant Sci.* 24, 23–58.
- Bentley D., Balasubramanian S., Swerdlow H., Smith G., Milton J., ... & Anastasi, C. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, 456(7218), 53-59.
- Bradshaw J. & Bonierbale M. (2010). Potatoes. En: Bradshaw J. (edit). *Root and Tuber Crops, Handbook of Plant Breeding Vol 7*, New York: Springer, pp. 1-52.

- Bray E. (2007). Molecular and physiological responses to water-deficit stress. En Jenks M., Hasegawa P. & Jain M. (eds.), *Advances in Molecular Breeding toward Drought and Salt Tolerant Crops*, The Netherlands, pp. 121–140.
 - Bucher M., Schroeder B., Willmitzer L. & Riesmeier, J. (1997). Two genes encoding extensin-like proteins are predominantly expressed in tomato root hair cells. *Plant molecular biology*, 35(4), 497-508.
- Chandler P. & Robertson M. (1994). Gene expression regulated by abscisic acid and its relation to stress tolerance. *Annual review of plant biology*, 45(1), 113-141.
- Chaves M., Maroco J. & Pereira J. (2003). Understanding plant responses to drought- from genes to the whole plant. *Functional Plant Biology*. 30: 239- 264.
- Cominelli E., Galbiati M., Vavasseur A., Conti L., Sala T., Vuylsteke M. & Tonelli C. (2005). A guard-cell-specific MYB transcription factor regulates stomatal movements and plant drought tolerance. *Current Biology*, 15(13), 1196-1200.
- Deblonde P. & Ledent J. (2000). Effects of moderate drought conditions on crop growth parameters and earliness of six potato cultivars under field conditions. *Agronomie* 20, 595–608.
- Downs C., Ryan S. & Heckathorn S. (1999). The chloroplast small heat-shock protein: evidence for a general role in protecting photosystem II against oxidative stress and photoinhibition. *Journal of plant physiology*, 155(4), 488-496.
- Evers D., Lefèvre I., Legay S., Lamoureux D., Hausman J., Rosales R., Marca L., Hoffmann L., Bonierbale M. & Schafleitner R. (2010). Identification of drought-responsive compounds in potato through a combined transcriptomic and targeted metabolite approach. *Journal of Experimental Botany*. 61(9), 2327–2343.

- Ekanayake I. (1989). Effect of drought on root system characters of potatoes. In: Resumen y programa. XVI Reunión de la Asociación Latinoamericana de la Papa. Mar del Plata, República Argentina, pp. 3
- Ekanayake I. (1993). Evaluación de resistencia a la sequía en genotipos de papa y camote. Guía de investigación. CIP 19.
- FAOSTAT (2008). Potato world: Production and consumption. International Year of the Potato. <<http://www.potato2008.org/en/world/>>. Acceso: 08/10/2008.
- Farooq M., Wahid A., Kobayashi N., Fujita D. & Basra S. (2009). Plant drought stress: effects, mechanisms and management. *Agronomy for Sustainable Development* 29: 185-212
- Filichkin S., Priest H., Givan S., Shen R., Bryant D., Fox S., Wong W., & Mockler T. (2010). Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*. *Genome Research*. 2010 20: 45-58.
- García-Mata, C., & Lamattina, L. (2001). Nitric oxide induces stomatal closure and enhances the adaptive plant responses against drought stress. *Plant Physiology*, 126(3), 1196-1204.
- Gene Ontology Consortium. (2004). The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Research*, Vol. 32, Database issue.
- Gil P., Liu Y., Orbovic V., Verkamp E., Poff K., & Green P. (1994). Characterization of the auxin-inducible SAUR-AC1 gene for use as a molecular genetic tool in *Arabidopsis*. *Plant Physiology*, 104(2), 777-784.
- Gong P., Zhang J., Li H., Yang Ch., Zhang Ch., Zhang X., Khurram Z., Zhang Y., Wang T., Fei Z. & Ye Z. 2010. Transcriptional profiles of drought-responsive genes in modulating transcription signal transduction, and biochemical pathways in tomato. *Journal of Experimental Botany*, 61(13), pp. 3563–3575.

- Hawkes J. (1990). The potato: evolution, biodiversity and genetic resources. Belhaven Press, Oxford. England/Smithsonian Institution Press, Washington, D. C. pp. 259.
- Hayano C., Calderón C., Ibarra E., Herrera L. & Simpson J. (2009). Analysis of gene expression and physiological responses in three mexican maize landraces under drought stress and recovery irrigation. PLoS ONE 4: e7531.doi:10.1371/journal.pone.0007531.
- Head, S. R., Komori, H. K., LaMere, S. A., Whisenant, T., Van Nieuwerburgh, F., Salomon, D. R., & Ordoukhanian, P. (2013). Library construction for next-generation sequencing: overviews and challenges. BioTechniques, 56(2), 61-4.
- Huamán Z. (1986). Systematic botany and morphology of the potato. Technical Information bulletin 6, International Potato Center, Lima, Peru.
- Huamán Z., & Spooner D. (2002). Reclassification of landrace populations of cultivated potatoes (*Solanum* sect. Petota). American Journal of Botany, 89(6), 947-965.
- Ingram J. & Bartels D. (1996). The molecular basis of dehydration tolerance in Plants. Annual Review of Plant Physiology and Plant Molecular Biology. 1996. 47:377–403.
- Jefferies R. (1992). Effects of drought on chlorophyll fluorescence in potato (*Solanum tuberosum* L.). I. Plant water status and the kinetics of chlorophyll fluorescence. Potato Research 35, 25- 34.
- Kakumanu A., Ambavaram M., Klumas C., Krishnan A., Batlang U., Myers E. & Pereira A. (2012). Effects of drought on gene expression in maize reproductive and leaf meristem tissue revealed by RNA-Seq. Plant physiology, 160(2), 846-867.

- Kant S., Bi Y., Zhu T., & Rothstein S. (2009). SAUR39, a small auxin-up RNA gene, acts as a negative regulator of auxin synthesis and transport in rice. *Plant physiology*, 151(2), 691-701.
- Knight H., Trewavas A. & Knight M. (1997). Calcium signalling in *Arabidopsis thaliana* responding to drought and salinity. *The Plant Journal* 12(5), 1067-1078.
- Knight H. & Knight M. (2001). Abiotic stress signalling pathways: specificity and cross-talk. *Trends in Plant Science* 6(6), 262-267.
- Kramer P. (1987). The role of water stress in tree growth. *Journal of Arboric*, 13:33-3.
- Kushiro T., Okamoto M., Nakabayashi K., Yamagishi K., Kitamura S., Asami T., Hirai N., Koshiha T., Kamiya Y. & Nambara, E. (2004). The *Arabidopsis* cytochrome P450 CYP707A encodes ABA 8'-hydroxylases: key enzymes in ABA catabolism. *The EMBO journal*, 23(7), 1647-1656.
- Lahlou O., Ouattar S. & Ledent J. (2003). The effect of drought and cultivar on growth parameters, yield and yield components of potato. *Agronomie* 23:257–268.
- Langfelder P. & Horvath S. (2007). Eigengene networks for studying the relationships between co-expression modules. *BMC Systems Biology*, 1:54.
- Langmead B., Trapnell C., Pop M. & Salzberg S. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*.10:25.
- Levy D. (1983). Varietal differences in the response of potatoes to repeated short periods of water stress in hot climates, 1. Turgor maintenance and stomatal behaviour. *Potato Research* 26: 303-313.

- Li J., Kinoshita T., Pandey S., Ng C., Gygi S., Shimazaki K. & Assmann S. (2002). Modulation of an RNA-binding protein by abscisic-acid-activated protein kinase. *Nature*, 418(6899), 793-797.
- Martinez C. & Moreno U. (1992). Expresiones fisiológicas de resistencia a la sequia en dos variedades de papa sometidas a estrés hídrico en condiciones de campo. *Revista Brasileira de Fisiología Vegetal*, 4, 33-38.
- Massa A., Childs K., Lin H., Bryan G., Giuliano G. & Buell R. (2011). The transcriptome of the reference potato genome *Solanum tuberosum* Group Phureja clone DM1-3 516R44. *PLoS One*, 6(10), e26801.
- Massa A., Childs K. & Buell R. (2013). Abiotic and Biotic Stress Responses in *Solanum tuberosum* Group Phureja DM1-3 516 R44 as Measured through Whole Transcriptome Sequencing. *The plant genome*, 6(3), 1-10.
- Manavalan L., Guttikonda S., Tran P. & Nguyen H. (2009). Physiological and Molecular Approaches to Improve Drought Resistance in Soybean. *Plant Cell Physiol.* 50(7): 1260–1276.
- Mizuno H., Kawahara Y., Sakai H., Kanamori H., Wakimoto H., Yamagata, H. & Matsumoto T. (2010). Massive parallel sequencing of mRNA in identification of unannotated salinity stress-inducible transcripts in rice (*Oryza sativa* L.). *BMC genomics*, 11(1), 683.
- Mortazavi A., Williams B, McCue K., Schaeffer L. & Wold B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*. 5(7), 621-628.
- Morozova O. & Marra M. (2008) Applications of next-generation sequencing technologies in functional genomics. *Genomics* 92, 255–264.

- Nagalakshmi U., Wang Z., Waern K., Shou C., Raha D., Gerstein M. & Snyder M. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science*, 320(5881), 1344-1349.
- Nakashima K., Yamaguchi-Shinozaki K., & Shinozaki K. (2014). The transcriptional regulatory network in the drought response and its crosstalk in abiotic stress responses including drought, cold, and heat. *Frontiers in Plant Science*, 5:170.
- National Drought Mitigation Center, 2014. <http://droughtmonitor.unl.edu/>. Fecha de acceso: Enero 2014.
- Nguyen H., Babu R. & Blum A. (1997) Breeding for drought resistance in rice: physiology and molecular genetics considerations. *Crop Science*. 37, 1426 – 1434.
- Nonami H. (1998). Plant water relations and control of cell elongation at low water potentials. *Journal of Plant Research*, 111(3), 373-382.
- Oliver M., Cushman J. & Koster K. (2007). Dehydration Tolerance in Plants. *Plant Stress Tolerance, Methods in Molecular Biology* 639.
- Oono Y., Seki M., Nanjo T., Narusaka M., Fujita M., Satoh R., Satou M., Sakurai T., Ishida J., Akiyama K., Iida K., Maruyama K & Shinozaki, K. (2003). Monitoring expression profiles of Arabidopsis gene expression during rehydration process after dehydration using ca. 7000 full-length cDNA microarray. *The Plant Journal*, 34(6), 868-887.
- Raghavendra A., Gonugunta V., Christmann A. & Grill E. (2010). ABA perception and signalling. *Trends in Plant Science* 15(7), 395-401.

- Reddy A., Chaitanya K. & Vivekanandan M. (2004). Drought-induced responses of photosynthesis and antioxidant metabolism in higher plants. *Journal of Plant Physiology*, 161, 1189-1202.
- Reisser V., Raitt D., & Sait H. (2003). Yeast osmosensor Sln1 and plant cytokinin receptor Cre1 respond to changes in turgor pressure. *The Journal of Cell Biology*, 161(6).
- Rensink W., Iobst S., Hart A., Stegalkina S., Liu J., & Buell C. (2005). Gene expression profiling of potato responses to cold, heat, and salt stress. *Functional & integrative genomics*, 5(4), 201-207.
- Rowland O., Ludwig A., Merrick C., Baillieul F., Tracy F., Durrant W., Fritz-Laylin L., Nekrasov V., Sjolander K., Yoshioka H. & Jones J. (2005). Functional analysis of Avr9/Cf-9 rapidly elicited genes identifies a protein kinase, ACIK1, that is essential for full Cf-9-dependent disease resistance in tomato. *The Plant Cell Online*, 17(1), 295-310.
- Schachtman D. & Goodger J. (2008). Chemical root to shoot signaling under drought. *Trends in Plant Science*, 13(6), 281-287.
- Schafleitner R., Gaudin A., Gutierrez R. Alvarado C. & Bonierbale M. (2007a). Proline accumulation and real time PCR expression analysis of genes encoding enzymes of proline metabolism in relation to drought tolerance in Andean potato. *Acta Physiologiae Plantarum*, 29, 19–26.
- Schafleitner R., Gutierrez R., Gaudin A., Alvarado C., Nomberto T., Tincopa L., Avila L., Delgado F., Simon R. & Bonierbale M. (2007b). Capturing candidate drought tolerance traits in two native Andean potato clones by transcription profiling of field grown plants under water stress. *Plant Physiology and Biochemistry* 45, 673–690.

- Schapendonk A., Spitters C. & Groot P. (1989). Effects of water stress on photosynthesis and chlorophyll fluorescence of five potato cultivars. *Potato Research* 32, 17-32.
- Schmieder R., & Edwards R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, 27(6), 863-864.
- Seki M., Ishida J., Narusaka M., Fujita M., Nanjo T., Umezawa T., Kamiya A., Nakajima M., Enju A. & Sakurai T. (2003). Monitoring the expression pattern of around 7,000 *Arabidopsis* genes under ABA treatments using a full-length cDNA microarray. *Functional e Integrative Genomics*, 2, (6), 282-291.
- Serraj R. & Sinclair T. (2002). Osmolyte accumulation: can it really help increase crop yield under drought conditions. *Plant, cell & environment*, 25(2), 333-341.
- Sharp R, Poroyko V, Hejlek L, Spollen W, Springer G, Bohnert H, Nguyen HT, Shin D., Moon S., Han S., Kim B., Park S., Yoon H., Lee H., Kwon H., Baek, Young Y. & Byu M. (2004). Root growth maintenance during water deficits: physiology to functional genomics. *Journal of Experimental Botany*, 55(407), 2343-2351.
- Shin D., Moon S., Han S., Kim , Park S., Lee S., Yoon H., Lee H., Kwon H., Baek D., Yi B. & Byun M. (2011). Expression of StMYB1R-1, a Novel Potato Single MYB-Like Domain Transcription Factor, Increases Drought Tolerance. *Plant Physiology* 155:421-432.
- Shinozaki K., & Yamaguchi-Shinozaki K. (1997). Gene expression and signal transduction in water stress response. *Plant Physiol.* 25:327–334.
- Shinozaki K., & Yamaguchi-Shinozaki K. (2007). Gene networks involved in drought stress response and tolerance. *Journal of Experimental Botany*, 58, (2), 221–227.

- Spooner D. & Hetterscheid W. (2005). Origins, evolution, and group classification of cultivated potatoes. In: Darwin's Harvest: New Approaches to the Origins, Evolution, and Conservation of Crops. T.J. Motley, N. Zerega, and H. Cross (eds). pp. 285-307. Colombia University Press, New York.
- Spooner D. & Salas, A. Structure, biosystematics, and genetic resources. En: J. Gopal, J. and Khurana, P (edit). Handbook of potato production, improvement, and post-harvest management. New York: Haworth's Press, 2006, p. 1-39.
- Steudle E. (2000). Water uptake by roots: effects of water deficit. *Journal of Experimental Botany*, 51, (350), 1531-1542.
- Tamura T., Hara K, Yamaguchi Y., Koizumi N., & Sano H. (2003). Osmotic Stress Tolerance of Transgenic Tobacco Expressing a Gene Encoding a Membrane-Located Receptor-Like Protein from Tobacco Plants. *Plant Physiology*, ,131, 454–462
- Taiz L. & Zeiger E. (2006) *Plant Physiology*, 4th edn., Sinauer, Massachusetts.
- TIGR Solanaceae Genomics Resource. (2014). TIGR Potato cDNA Microarray Description. Recuperado de http://jcvi.org/potato/sol_ma_microarrays.shtml.
- Trapnell C., Pachter L. & Salzberg S. (2009). TopHat: discovering splice junctions with RNA-seq. *Bioinformatics*, 25:1105–1111.
- Trapnell C., Roberts A., Goff L., Pertea G., Kim D., Kelley D., Pimentel H., Salzberg S., Rinn J., & Pachter L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols*, 7, (3), 562–578.
- Urao A., Yakubov B., Satoh R., Yamaguchi-Shinozaki K., Seki M., Hirayama T. & Shinozakib K. (1999). A Transmembrane Hybrid-Type Histidine Kinase in Arabidopsis Functions as an Osmosensor. *The Plant Cell*, 11, 1743–1754.

- Vasquez-Robinet C., Mane Sh., Ulanov A., Watkinson J., Stromberg V., Koeyer D., Schafleitner R., Willmot D., Bonierbale M, Bohnert H. & Grene R. (2008). Physiological and molecular adaptations to drought in Andean potato genotypes. *Journal of Experimental Botany*, 59, 8, 2109–2123.
- van Loon C. (1981). The effect of water stress on potato growth, development, and yield. *Stress Physiology Symposium*, sponsored by the Physiology Section of the Potato Association of America, presented July 23, 1979, at the 63rd Annual Meeting, Vancouver, B.C.
- van Loon C. (1986). Drought, a major constraint in potato production and possibilities for screening for drought resistance. *Potato research*, 5 - 16.
- Vitré M., Farrant J. & Driouich A. (2004). Insights into the cellular mechanisms of desiccation tolerance among angiosperm resurrection plant species. *Plant, Cell & Environment*, 27(11), 1329-1340.
- Villar E., Klopp C., Noirot C., Novaes E., Kirst M., Plomion C. & Gion J. (2011). RNA-Seq reveals genotype-specific molecular responses to water deficit in eucalyptus. *BMC genomics*, 12(1), 538.
- Wang W., Vinocur B., Shoseyov O. & Altman, A. (2004). Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. *Trends in plant science*, 9(5), 244-252.
- Wang Z., Gerstein M. & Snyder M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Review Genetics*, 10(1), 57-63.
- Watkinson J., Hendricks L., Sioson A., Vasquez-Robinet C., Stromberg V., Heath L., Schuler M., Bohnert H., Bonierbale M. & Grene R. (2006). Accessions of *Solanum tuberosum* ssp. *andigena* show differences in photosynthetic recovery

after drought stress as reflected in gene expression profiles. *Plant Science* 171, 745–758.

- Wilhite D., Svoboda M. & Hayes M. (2007). Understanding the complex impacts of drought: A key to enhancing drought mitigation and preparedness. *Water Resour Manage* 21:763–774.
- Xiong L. & Zhu J. (2002). Molecular and genetic aspects of plant responses to osmotic stress. *Plant, Cell & Environment*, 25(2), 131-139.
- Yamaguchi-Shinozaki K. & Shinozaki K. (1993). The plant hormone abscisic acid mediates the drought-induced expression but not the seed-specific expression of rd22, a gene responsive to dehydration stress in *Arabidopsis thaliana*. *Molecular and General Genetics MGG*, 238(1-2), 17-25.
- Yamaguchi-Shinozaki, K. & Shinozaki, K. (2005) Organization of cis-acting regulatory elements in osmotic- and cold-stress-responsive promoters. *Trends in Plant Science*. 10, 88-94.
- Yokota A., Takahara K. & Akasdi K. (2006). Water Stress. *Physiology and Molecular Biology of Stress Tolerance in Plants*, 15–39.
- Yuan B., Nishiyama S. & Kang Y. (2003). Effects of different irrigation regimes on the growth and yield of drip-irrigated potato. *Agricultural Water Management*, 63(3), 153-167.
- Zhu W., Ouyang Sh., Iovene M., O'Brien K., Vuong H., Jiang J. & Buell R. (2008). Analysis of 90 Mb of the potato genome reveals conservation of gene structures and order with tomato but divergence in repetitive sequence composition. 2008, *BMC Genomics*, 9:286
- Zhu J. & Bohnert H. (2000). Plant cellular and molecular responses to high salinity. *Annual Review of Plant Physiology and Plant molecular biology* 51:463–99.

XI. ANEXOS

Anexo I: Protocolo para la extracción de RNA de tejidos de papa

Componentes	Solución Stock	20 ml vol	100 ml vol
100mM LiCl	8M LiCl	0.5ml	2.5ml
100mM Tris pH8.5	1M Tris pH8.5	2.0ml	10.0ml
10mM EDTA	0.5M EDTA	0.4ml	2.0ml
1% SDS	10% SDS	2.0ml	10.0m
15mM DTT	1M DTT *	0.3ml	3.0ml
DEPC H2O		14.8 ml	74.0 ml

* El DithioDithreitol (DTT) se debe añadir cuando se va a usar el buffer de extracción.

1. Preparar el buffer de extracción (Tabla)
2. Añadir 4 ml de phenol pH 4.3 (Sigma) + 4 ml del buffer de extracción a un tubo de 15 ml y ponerlo en baño maría a 60°C.
3. Moler con nitrógeno líquido 1-2 g de tejido, transferirlo al tubo de 15ml, homogenizar por 60 segundos usando un vortex, y volver el tubo a baño maría.
4. Se agrega 4ml cloroformo: alcoholisoamílico (24:1) y se homogeniza durante 30 segundos.
5. Centrifugar 10 min a 1100rpm a +4°C, y transferir la fase acuosa a otro tubo con 4 ml de fenol:cloroformo/IAA (25:24:1).
6. Homogenizar por 30 segundos inmediatamente después de añadir la fase acuosa
7. Centrifugar 10 min a 1100rpm a +4°C, y transferir la fase acuosa a un tubo con 4 ml de fenol:cloroformo: IAA.
8. Homogenizar durante 30 segundos inmediatamente después de añadir la fase acuosa.
9. Centrifugar 10 min a 1100rpm, a +4°C, y transferir la fase acuosa a un tubo con 4 ml of cloroformo: IAA (24:1).
10. Repetir los pasos 8-9 dos veces.

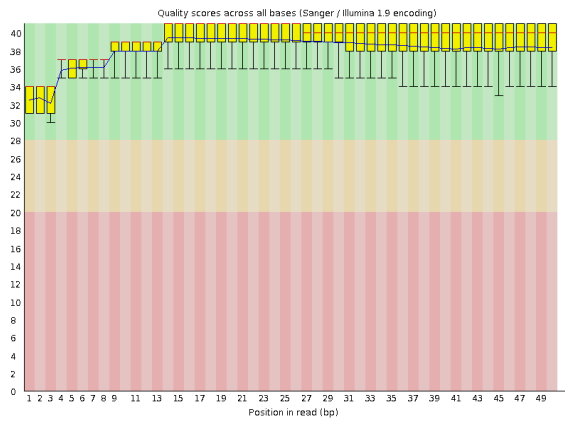
11. Transferir la fase acuosa a un tubo limpio y añadir 1 vol. de 4M LiCL (LiCl precipita el RNA pero no DNA).
12. Mezclar por inversion e incubar a 20°C for 2 hr o toda la noche.
13. Centrifugar 20min a 1100rpm a +4°C.
14. Remover el sobrenadante y lavar con 5ml de etanol 70% ethanol y centrifugar 200min a 1100rpm, a +4°C.
15. Remover el etanol y centrifugar 2min a1100rpm, a +4°C y remover cuidadosamente algún líquido presente
16. Dejar secar el tubo al ambiente por 30 minutos, y resuspender el pellet en 150µl de agua DEPC.
17. Medir la concentración de RNA concentración usando el método del Nanodrop.

Anexo II. Procedimiento para eliminar DNA contaminante usando el Kit DNA Free (Ambio)

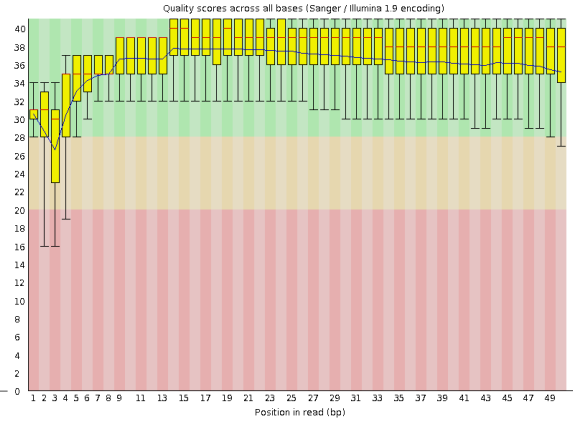
1. Agregar 0.1 volumen del Buffer DNase 10X y 1uL de rDNaseI al RNA que se va a tratar, y mezclar cuidadosamente.
2. Incubar la mecla a 37°C durante 20-30 minutos.
3. Agregar Reactivo de Inactivación de DNase (0.1 volumen) y mezclar bien.
4. Incubar 2 minutos a temperatura ambiente, y mezclar ocasionalmente.
5. Centrifugar a 10000 xg durante 1.5 minutos y transferir el RNA a un tubo fresco.

Anexo III: Resultados de calidad de todas las bibliotecas

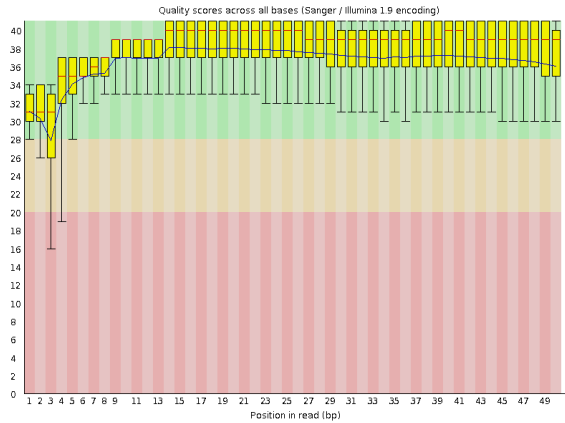
PGO_AA_CAGATC_L008_R1_001.fastq



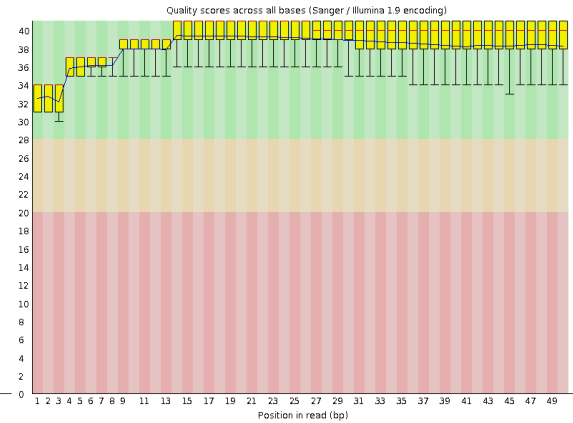
PGO_AI_AGTCAA_L006_R1_001.fastq



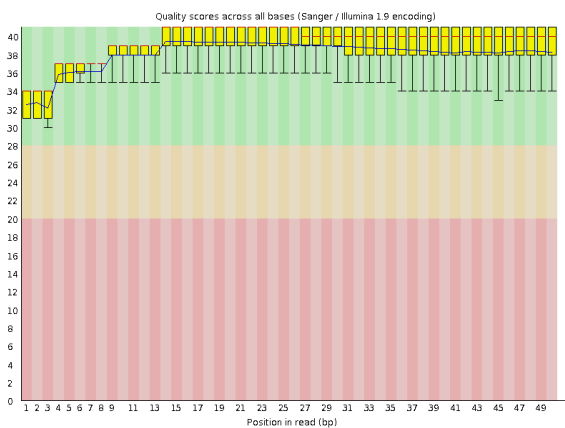
PGO_AM_GTGAAA_L005_R1_001.fastq



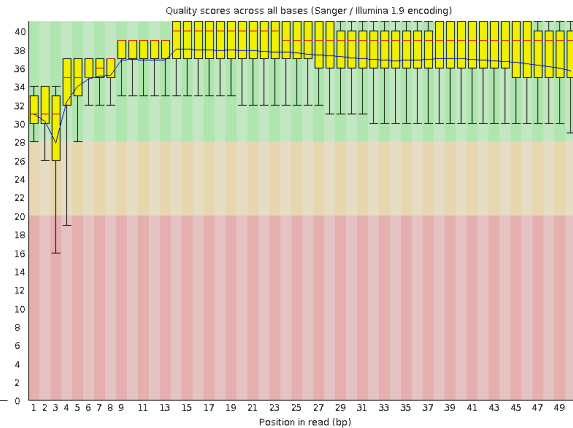
PGO_AE_ATGTCA_L008_R1_001.fastq



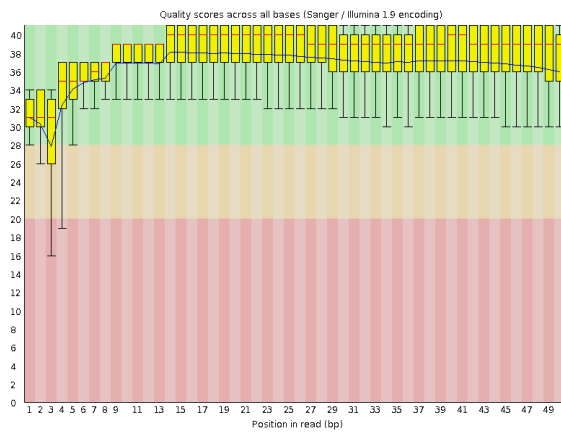
PGO_AC_AGTCAA_L008_R1_001.fastq



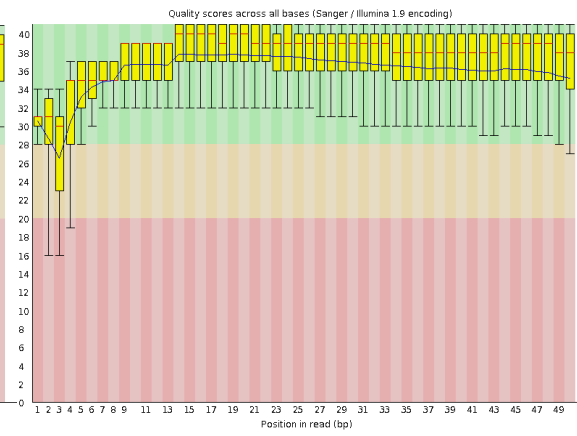
PGO_AK_CCGTCC_L005_R1_001.fastq



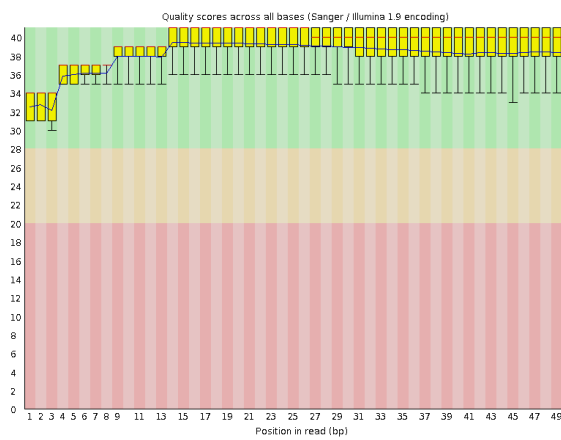
PGO_AO_TGACCA_L005_R1_001.fastq



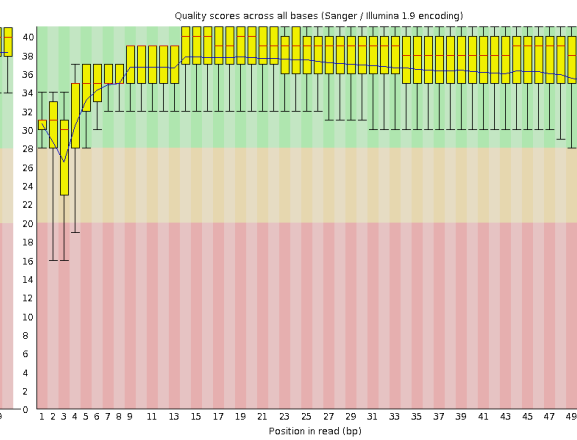
PGO_AG_CAGATC_L006_R1_001.fastq



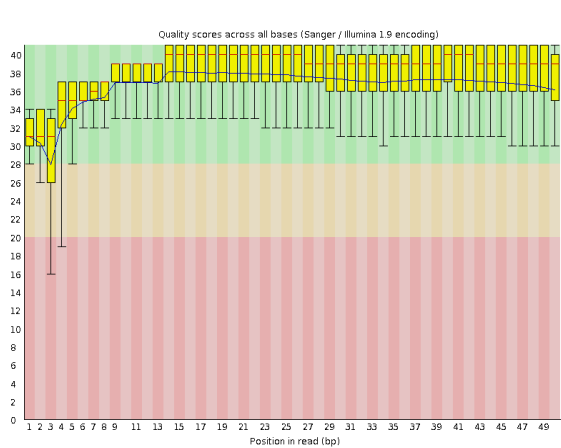
PGO_AB_CTTGTA_L008_R1_001.fastq



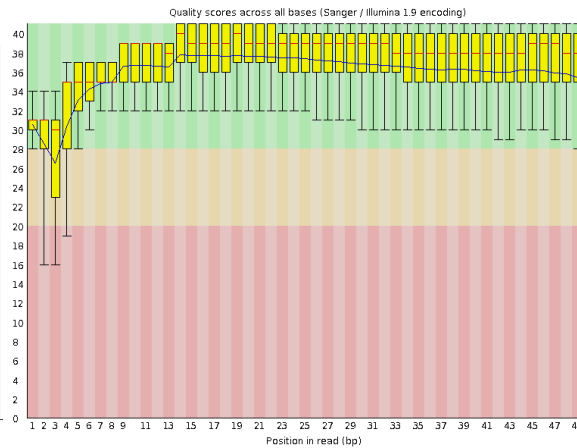
PGO_AJ_AGTTCC_L006_R1_001.fastq



PGO_AN_CGATGT_L005_R1_001.fastq

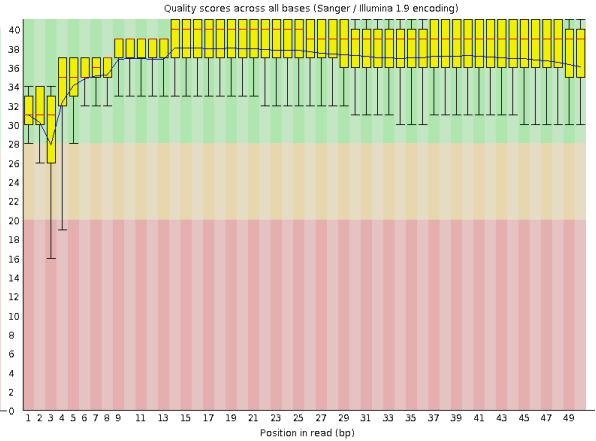
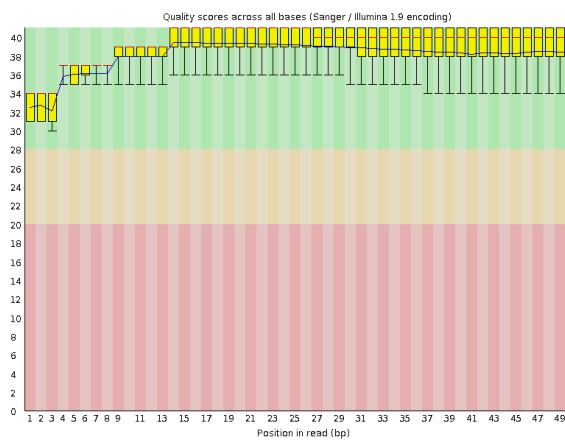


PGO_AF_GCCAAT_L006_R1_001.fastq



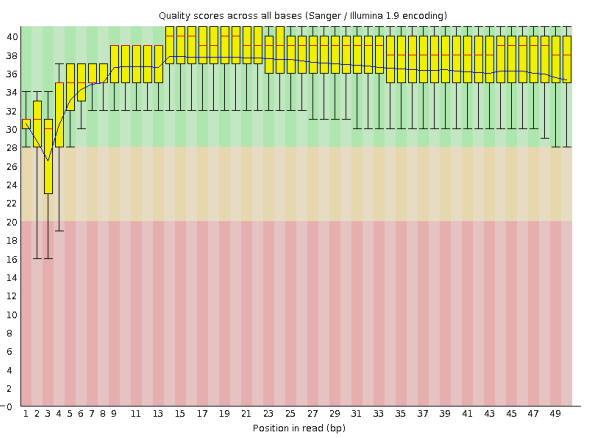
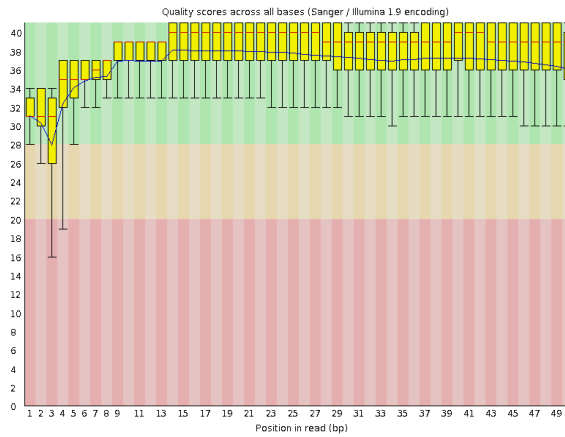
PGO_AD_AGTTC_L008_R1_001.fastq

PGO_AL_GTCCGC_L005_R1_001.fastq



PGO_AP_ACAGTG_L005_R1_001.fastq

PGO_AH_CTTGTA_L006_R1_001.fastq



Anexo IV. Número de Genes con asignación de términos GO

