UNIVERSIDAD ANDRÉS BELLO

Facultad de Ciencias Biológicas

Escuela de Ingeniería en Biotecnología

IDENTIFICACIÓN DE GENES CANDIDATOS ASOCIADOS A HARINOSIDAD EN *Prunus persica* (L.) Batsch USANDO QTL DE EXPRESIÓN

Tesis presentada como parte de los requisitos para optar al Grado de **Magíster en Biotecnología**

Director de tesis: **Dr. Reinaldo Campos**

Centro de Biotecnología Vegetal

Facultad de Ciencias Biológicas

Universidad Andrés Bello

Director de tesis: **Dr. Claudio Meneses**

Centro de Biotecnología Vegetal

Facultad de Ciencias Biológicas

Universidad Andrés Bello

CRISTOBAL ALEJANDRO BALLADARES MOYA

Santiago-Chile

Julio 2017



UNIVERSIDAD ANDRÉS BELLO
Facultad de Ciencias Biológicas
Ingeniería en Biotecnología

"IDENTIFICACIÓN DE GENES CANDIDATOS ASOCIADOS A HARINOSIDAD EN *Prunus persica* (L.) Batsch USANDO QTL DE EXPRESIÓN"

Cristobal Alejandro Balladares Moya

Este trabajo fue elaborado bajo la supervisión de los directores de Tesis Dr. Reinaldo Campos y Dr. Claudio Meneses, en el Centro de Biotecnología Vegetal, Universidad Andrés Bello, aprobado por los miembros de la Comisión de evaluación.

_____
Dr. Reinaldo Campos
Director de Tesis

_____
Dr. Claudio Meneses
Director de Tesis

_____
Dr. Eduardo Castro
Comisión de Tesis

_____
Dr. Patricio Hinrichsen
Comisión de Tesis

SANTIAGO-CHILE
Año 2017

Financiamiento:

Esta tesis fue llevada a cabo gracias al financiamiento de:

- Proyecto CORFO Consorcio Biofrutales 13CTI-21520-SP04.
- Proyecto Genoma G13I0005.

Índice:

Resumen: 3

Introducción: 4

a) Marco teórico: 4

b) Planteamiento del problema: 4

c) Antecedentes bibliográficos: 6

d) Formulación de hipótesis: 15

e) Objetivo general y específico: 15

Materiales y métodos: 16

a) Diseño experimental: 16

b) Obtención del Material Vegetal: 18

c) Caracterización de la información fenotípica de una población segregante: 18

d) Selección de individuos contrastantes para el carácter de harinosidad: 20

e) Extracción de ARN: 21

f) Construcción, calidad y secuenciación de librerías: 21

g) Mapeo de las lecturas y análisis de expresión diferencial: 22

h) Análisis de ontología génica: 22

i) Análisis de QTLy eQTL: 23

j) Validación del transcriptoma: 23

Resultados: 24

a) Caracterización fenotípica de los frutos de una población segregante para harinosidad: 24

b) Construcción de un mapa genético de la población OxN: 25

c) Análisis de QTL convencionales para harinosidad: 27

d) Selección de segregantes para secuenciación: 30

e) Métrica de secuenciación de librerías de ARN: 32

f) Análisis de expresión diferencial: 34

g) Filtro de resultados de expresión diferencial previo al análisis de eQTL: 37

h) Análisis de eQTL para harinosidad: 40

i) Identificación de genes candidatos para harinosidad: 41

j) Validación del transcriptoma de la población OxN por medio de RT-PCR: 43

Discusión: 45

Conclusión: 54

Referencias: 55

Anexos: 64

Resumen:

El duraznero (*Prunus persica* (L.) Batsch) es una especie frutal caducifolia de clima templado. A nivel mundial es la cuarta especie en importancia en términos de producción. Por otra parte, Chile es el principal exportador desde el hemisferio sur hacia EE.UU., Europa y Asia. Sin embargo, para transportar duraznos a los mercados de destino, que se encuentran a distancias considerables, es necesario utilizar condiciones de almacenamiento refrigerado por largos períodos a bajas temperaturas, lo que genera desórdenes fisiológicos conocidos como daño por frío o "chilling injury*"* (CI). Los síntomas del CI son pardeamiento, "bleeding", pérdida de sabor y harinosidad. Este último es de gran importancia ya que ha sido identificado como uno de los principales problemas en la fruta por parte de los consumidores. La harinosidad se expresa como falta de jugo en la pulpa del fruto con la consecuente disminución de la calidad organoléptica. Es por este motivo que resulta muy relevante estudiar y comprender el CI con el objeto de generar tecnologías para evitar su aparición o reducir su efecto.

El objetivo general de este trabajo fue identificar genes candidatos involucrados con harinosidad utilizando la asociación de marcadores genéticos con los niveles de expresión de transcritos (fenotipo molecular) y los niveles de harinosidad (fenotipo), combinando la estrategia de análisis de QTLs con la secuenciación masiva de transcriptomas de individuos de una población F1 ('O'Henry' x 'NR-053'; OxN) contrastantes para harinosidad. Para esto, se realizó una caracterización fenotípica de la población para seleccionar los individuos contrastantes. Posteriormente, se secuenció el transcriptoma de tres frutos en cuatro estadíos. Adicionalmente, con información proveniente de un mapa genético previamente construido por el grupo de trabajo, se realizó un análisis de QTLs convencionales y de expresión para identificar los genes candidatos asociados al fenotipo.

Se encontraron genes asociados a producción de antocianinas, respuesta a estímulos, modificación de pared celular y fitohormonas. En base a ellos fue posible hipotetizar que la harinosidad es gatillada a partir de un estrés por frío durante la etapa de ablandamiento para madurez de consumo del fruto. Esta situación genera una respuesta celular deficiente, no pudiendo reclutar correctamente factores que gatillan el ablandamiento o concentraciones suficientes de hormonas o metabolitos, lo cual finalmente culmina en el desarrollo del fenotipo harinoso en *Prunus persica*.

Introducción:

a) Marco teórico:

Prunus persica (L.) Bastch se considera una especie modelo dentro de la familia de las rosáceas debido a su pequeño genoma diploide (n=*x*=8) de 290 Mpb y su corto periodo de juvenilidad (2-3 años) comparado con otras especies del género como *Prunus armeniaca* o *Prunus avium* (Arús *et al*. 2012). Su genoma se encuentra secuenciado y disponible (Verde *et al*. 2013), así como también una serie de información genética, genómica, de ontología y anotación en bases de datos como "Genome Database for Rosaceae" (GDR; http://www.rosaceae.org) y "Phytozome" (http://www.phytozome.org).

Luego del manzano y el peral, *P. persica* es el tercer frutal de clima templado en producción mundial con 19.9 millones de toneladas durante 2016 (USDA, 2016). Chile es el exportador más importante del hemisferio sur con 80 mil toneladas exportadas para el año 2016 (USDA, 2016). A pesar del posicionamiento de Chile en cuanto a exportación y dada la competencia en el mercado y la necesidad de ofrecer productos de alta calidad orientado a satisfacer al consumidor, es indispensable para Chile realizar un esfuerzo constante en investigación para optimizar la producción.

b) Planteamiento del problema:

El tamaño del fruto varía al comienzo del desarrollo por división celular y posteriormente por el tamaño de las células (Batjer y Westwood 1958). En durazno y nectarín se ha demostrado que el conteo celular del mesocarpo o pulpa presenta mayores diferencias entre frutos grandes y pequeños en distintos cultivares (Scorza *et al*. 1991), siendo más relevante para el tamaño del fruto el cambio en el número de células por sobre el tamaño de las mismas (Bradley y Muriel 1959). Este cambio en número y tamaño tiene influencias sobre la estructura de cada célula a nivel de su pared, cambiándola y adaptándola.

Posteriormente, durante la maduración y ablandamiento de los frutos ocurren diversos cambios bioquímicos a nivel de pared celular. La pared celular de los dicotiledones se compone de una matriz de glicanos, pectinas y glicoproteínas estructurales (McCann y Roberts 1992; Carpita y Gibeaut 1993). Básicamente, se depolimerizan azúcares estructurales de la pared, generando

pérdida de firmeza y causando el ablandamiento del mesocarpo (Brummell y Harpster 2001). La enzima pectin metil esterasa (PME), por ejemplo, participa en la remoción de los grupos metilo (CH_3) de las pectinas, lo que favorece la actividad de la enzima poligalacturonasa (PG), dentro de la cual existen dos tipos: endoPG y exoPG (Hadfield y Bennett 1998). La endoPG cataliza la hidrólisis de los enlaces glicosídicos internos de las cadenas de los homogalacturonanos, que forman parte de las pectinas en la pared celular. La exoPG, en cambio, hidroliza enlaces glicosídicos desde el extremo no reductor de los homogalacturonanos. La enzima PME se ve comúnmente aumentada durante el ablandamiento del fruto (Payasi *et al*. 2009) al igual que la enzima endoPG y exoPG (Brummell *et al*. 2004). El problema de harinosidad ha sido abordado principalmente desde el punto de vista de estas enzimas en ensayos que incluyen el concepto de daño por frío (del inglés "chilling injury").

Durante el periodo de transporte los duraznos y nectarinas se mantienen a bajas temperaturas para extender la vida del fruto en el mercado (Crisosto *et al.* 1999). Por esto, muchas variedades pueden presentar daño por frío, alterando características de calidad en los frutos y provocando disminución en sus propiedades sensoriales (Campos-Vargas *et al.* 2006).

El daño por frío se define como la disfunción fisiológica del fruto al exponerse a temperaturas bajas no congelantes (sobre 0 ºC) (Lyons 1973). En duraznos y nectarinas los síntomas del daño por frío se expresan luego de una a dos semanas de almacenaje en frío (2 a 5 ºC) dependiendo del cultivar (Crisosto *et al*. 1999) y posee distintos síntomas como pardeamiento (oscurecimiento de la pulpa), sangrado (coloración rojiza del mesocarpo), textura de tipo cuero (textura de la pulpa dura o firme y a la vez flexible) y harinosidad o sequedad (sensación de falta de agua) (Lurie y Crisosto 2005). Dentro de los síntomas del daño por frío en los durazneros y nectarinas, destaca la harinosidad, el cual es un desorden a nivel de textura definido como 'ablandamiento anormal del fruto y ausencia de jugosidad en la boca' (Barreiro *et al*. 2000) y 'tejido disgregado y sensación de falta de jugo' (Valero *et al*. 2005)**.** Este se genera por un ablandamiento incorrecto de la pared celular, donde se ha reportado que la enzima endoPG disminuye su nivel de actividad en contraste con su nivel de actividad normal (Brummell *et al*. 2004). No se ha logrado concluir el rol de la exoPG en daño por frío ya que la bibliografía es contradictoria al respecto (Lurie y Crisosto 2005).

c) Antecedentes bibliográficos:

Se han realizado numerosos esfuerzos desde el punto de vista científico para reducir, mitigar y comprender el problema de daño por frío en duraznos y nectarinas tales como fertilización (Daane *et al.* 2008), regímenes de riego (Crisosto *et al*. 1994; Goldhammer *et al*. 2002), atmósferas controladas (Streif *et al.* 1994; Lurie 2007), aplicaciones de la hormona vegetal etileno (Zhou *et al.* 2001), estudios a nivel genético (Vizoso *et al*. 2009; Cantín *et al*. 2010; Nuñez-Lillo *et al*. 2015; Sanhueza *et al*. 2015), etc.

El problema del daño por frío en *Prunus persica* se ha asociado a muchos factores que involucran cambios en niveles de actividad de enzimas remodeladoras de pared celular como PG, PME, expansina y XTH (Trainotti *et al*. 2003; Brummell *et al*. 2004; Trainotti *et al*. 2006; Sanhueza *et al*. 2015), cambios a nivel transcripcional de genes que se expresan diferencialmente entre duraznos jugosos y harinosos (Gonzalez-Aguero *et al.* 2008; Ogundiwin *et al.* 2008) y cambios en niveles de ARN mensajeros de expansina (Obenland *et al.* 2003). Muchas de las enzimas asociadas a ablandamiento y textura del fruto se codifican por familias multigénicas (Brummell y Harpster 2001) y estarían asociadas al problema de daño por frío (Dagar *et al.* 2013). Por consiguiente, el uso de técnicas que permitan el estudio de un gran número de transcritos facilitaría el entendimiento de la problemática de la inducción y desarrollo de este desorden.

La transcriptómica es el estudio del set completo de transcritos (ARN o ARNm) de una célula o tejido, en una etapa especifica de su desarrollo o condición fisiológica (Wang *et al.* 2009). Este se puede obtener, por ejemplo, por medio de métodos de alto rendimiento tales como análisis de microarreglos (Horgan y Kenny 2011) y últimamente por medio de herramientas de secuenciación masiva (Mutz *et al.* 2012).

Con el desarrollo de las nuevas técnicas de secuenciación de última generación (NGS; del inglés “Next Generation Sequencing”) es posible secuenciar el genoma de organismos completos con gran precisión y en muy poco tiempo. Gracias a esto es factible analizar transcritos de manera masiva, tanto ARN mensajero, ARN no codificante y ARN pequeños (Wang *et al.* 2009). Los procedimientos de secuenciación de ARN también pueden cuantificar la cantidad de transcritos existentes, permitiendo realizar análisis comparativos entre distintas etapas o estados de distintas líneas celulares. A través de estos estudios es posible identificar un sinnúmero de moléculas de

ARN que pueden estar involucradas en diversas situaciones biológicas (Anders *et al*. 2010; Tarazona y García-Alcalde 2011).

Ejemplos de estos estudios en *Prunus persica* se encuentran en publicaciones como las de Vizoso *et al*. (2009), en donde se secuenciaron 50.625 EST (del inglés "Expressed Sequence Tag") de mesocarpo de la variedad de duraznero fundente 'O'Henry' y se observaron genes como poligalacturonasa, pectin metilesterasa, endoquitinasas básicas y endopoligalacturonasas (que aumentan sus niveles de expresión en frío) e invertasas, inhibidores de pectin metilesterasa y pectato liasa (con niveles de expresión disminuidos en frío). Los transcritos asociados a harinosidad serían poligalacturonasa y proteínas que inhiben poligalacturonasa. Estos se encuentran con niveles de expresión disminuidos en comparación con frutos sin harinosidad.

Posteriormente, Sanhueza *et al*. (2015), estudió nectarinas de la variedad 'Red Pearl' en cosecha y en tratamiento con frío a 4 ºC (para promover harinosidad en base a estrés por frío) con posterior espera a madurez de consumo en condiciones de atmosfera controlada (alto CO_2 y bajo O_2) y condiciones normales (sin alterar atmosfera). Se encontraron diferencias a nivel de transcritos de enzimas relacionadas a remodelamiento de pared celular como aumento de ARNm de poligalacturonasa durante la maduración en ambas condiciones mencionadas anteriormente, sin embargo, la exo-poligalacturonasa presenta una mayor actividad en frutos con atmosfera controlada y una tasa poligalacturonasa/pectin esterasa menor en frutos en condiciones normales.

Uno de los grandes avances conseguidos junto con los métodos de secuenciación masiva es la identificación en forma simultánea de un gran número de marcadores moleculares. El más popular de estos son los polimorfismos nucleotídicos únicos o SNP por sus siglas en inglés ("Single Nucleotide Polymorphism"). Estos son diferencias a nivel de secuencias nucleotídicas que consisten en el cambio de una única base en la secuencia de ADN. Ocurren con alta frecuencia, siendo el polimorfismo más abundante en el genoma humano con una frecuencia de 1 SNP cada 20 a 300 pares de bases (Wang *et al.* 1998) implicando una cobertura casi total del genoma. Gracias a esto, los SNP son idóneos para generar y construir mapas genéticos completos de distintos organismos a diferencia de otros tipos de marcadores como RAPD (del inglés "Random Amplification of Polimorphic DNA"), los cuales consumen mucho tiempo, son costosos, requieren grandes cantidades de material genético y son poco reproducibles, impidiendo la obtención de un gran número de marcadores; o los microsatélites que a pesar de ser simples, robustos y

reproducibles requieren grandes cantidades de tiempo para su obtención (Collard *et al*. 2005). Se han construido mapas genéticos para *Prunus persica* utilizando RAPD y microsatélites en el año 1994 (Chaparro *et al*. 1994), utilizando 83 marcadores RAPD y otro mapa genético utilizando 82 microsatélites en 2006 (Dirlewanger *et al*. 2006), sin embargo, el mapa de referencia actual de *Prunus persica* posee unos 103.342 marcadores disponibles (https://www.rosaceae. org/organism/Prunus/persica).

Para construir un mapa genético es deseable tener una población de mapeo segregante proveniente de parentales contrastantes, la cual entregará información de marcadores genéticos una vez haya sido secuenciada. Estos marcadores (típicamente SNP) serán asociados estadísticamente para conformar grupos de ligamientos, los cuales representan los cromosomas (Paterson 1996). Es necesario tener una gran cobertura del genoma con estos marcadores, ya que de otra forma los grupos de ligamiento generados pueden diferir respecto al número de cromosomas de la especie.

Los loci de carácter cuantitativo o QTL (del inglés "Quantitative Trait Loci"), son una región del genoma que contiene uno o más factores genéticos que controlan un carácter cuantitativo explicando una parte de la variación fenotípica total (Collard *et al*. 2005). Los QTL pueden ser utilizados en mejoramiento genético y selección asistida por marcadores o MAS, (del inglés "Marker Assisted Selection"). Un QTL se obtiene asociando información genética proveniente de marcadores a nivel de ADN (SNP, microsatélites, entre otros) con información fenotípica (el carácter cuantitativo) (Collard *et al.* 2005).

Al unir la información de algún carácter fenotípico de interés y asociarla con los marcadores genéticos mapeados dentro de un mapa de ligamiento se puede, entonces, obtener QTL. Esta asociación debe ser significativa, lo cual se define por el valor de LOD (del inglés "Logarithm of Odds") el cual es el logaritmo de la probabilidad estadística. Un valor de LOD superior o igual a 3 indica con una relación de 1 en 1000 que la asociación entre los marcadores y el QTL existe y que esta asociación no es generada por azar.

Varios estudios, que han utilizado el análisis de QTL en diversas poblaciones de *Prunus persica* contrastantes para harinosidad, han encontrado genes candidatos ubicados en el grupo de ligamiento 4 y 5.

Peace *et al*. en 2005 investigó el control genético diferencial entre dos fenotipos: de carozo suelto y de pulpa fundente. Se pensaba que estos eran controlados por dos loci asociados entre sí o por un locus multi alélico. Ellos lograron asociar el ablandamiento de frutos fundentes a la enzima poligalacturonasa, ubicada en el grupo de ligamiento 4 a partir de dos poblaciones segregantes de duraznero. Esta enzima tiene un pico importante durante el ablandamiento en poscosecha.

Posteriormente en 2006 (Peace *et al*. 2006), Peace investigó los factores genéticos asociados al daño por frío en una población segregante entre 'Dr. Davis', que genera frutos no fundentes y 'Georgia Belle', que genera frutos fundentes particularmente susceptibles a daño por frío. La población segregante F1 fue llamada 'Pop-DG'. Esta fue almacenada en frío por 2 a 3 semanas y se encontró la enzima endopoligalacturonasa en el grupo de ligamiento 4 como principal responsable de la susceptibilidad a daño por frío, hipotetizando una desregulación de esta enzima a causa de las bajas temperaturas y una actividad normal de la pectin metilesterasa, resultando en componentes alterados de pectinas que generan un gel, causando harinosidad.

Estudios más detallados se llevaron a cabo en el año 2007 por Ogundiwin (Ogundiwin *et al*. 2007), donde se buscaron los responsables genético-moleculares de los síntomas del daño por frío (harinosidad, pardeamiento y sangrado) en la misma población mencionada anteriormente: 'Pop-DG'. Se construyó un mapa de ligamiento parcial para esta población, utilizando 142 marcadores a lo largo de 10 grupos de ligamiento, con una densidad promedio de 4,8 cM por marcador. Se ubicaron varios QTL asociados a los síntomas generados por daño por frío en los grupos de ligamiento 1, 2, 4, 5 y 6, encontrándose principalmente los responsables de sangrado y harinosidad en el grupo de ligamiento 4 y de pardeamiento en el grupo de ligamiento 5.

El mapa de ligamiento de la población 'Pop-DG' de Ogundiwin en 2007 (Ogundiwin *et al*. 2007) consistía de 10 grupos de ligamiento por falta de robustez en sus marcadores. Este problema fue solucionado por Martínez-García *et al*. en 2012 utilizando una F1 entre 'Pop-DG' y 'Pop-DF' (proveniente de una cruza entre 'Dr. Davis' y un híbrido de duraznero x almendro llamado 'F8,1-42') con 962 marcadores tipo SNP para 'Pop-DG' y 1400 SNP para 'Pop-DF'. Con esto se generó un mapa consenso utilizando 588 SNP repartidos en 8 grupos de ligamiento con una densidad de 0,81 cM por marcador. Gracias a esto se identificaron QTL asociados a harinosidad y sangrado en el grupo de ligamiento 4 y pardeamiento en el grupo de ligamiento 5.

Por otra parte, Cantín *et al*. en 2010, encontró QTL asociados a harinosidad, granulosidad (sensación fenotípica visual de la harinosidad) y textura tipo cuero en el grupo de ligamiento 4 a lo largo de dos temporadas de evaluación en una población F1 de nectarinas de pulpa fundente 'Venus' x 'BigTop'. Además, en la región denotada por el QTL en el grupo de ligamiento 4, se encontraron genes asociados a biosíntesis de etileno como ACO3 (ACC oxidasa) y C-0212 (Acetil-CoA acetiltransferasa).

Nuñez-Lillo *et al*. en 2015 en una población de 'Venus' x 'Venus', encontró QTL asociados a harinosidad y fecha de cosecha en tres temporadas en el grupo de ligamiento 4 en un mapa construido con 1.984 marcadores tipo SNP, logrando una cobertura de 0,21 cM por marcador. Se ha encontrado evidencia de que la fecha de cosecha está asociada a la calidad del fruto, siendo los de cosecha tardía (que pasan más tiempo en el árbol) más susceptibles a harinosidad que aquellos de cosecha temprana (Mitchell 1989; Crisosto y Valero 2008). Además, se encontraron factores de transcripción asociados a la familia NAC, ERF4 y CPK9. Los factores de transcripción NAC están involucrados en fecha de cosecha, maduración, estrés biótico y abiótico, respuestas asociadas a deshidratación dependiente de ABA, especies reactivas de oxígeno y defensa. Los factores de transcripción ERF4 se asocian a madurez de consumo en manzana, tomate y kiwi y en Arabidopsis como represor transcripcional modulando respuestas de ABA y etileno. Finalmente, los factores de transcripción CPK9 se asocian a fosforilación de residuos de serina de ACS en tomates (ácido 1-aminociclopropano-1-carboxilico sintasa, asociados a problemas en la producción de etileno en tomates harinosos). El grupo de ligamiento 4 de esta publicación se detalla en la Figura 1.

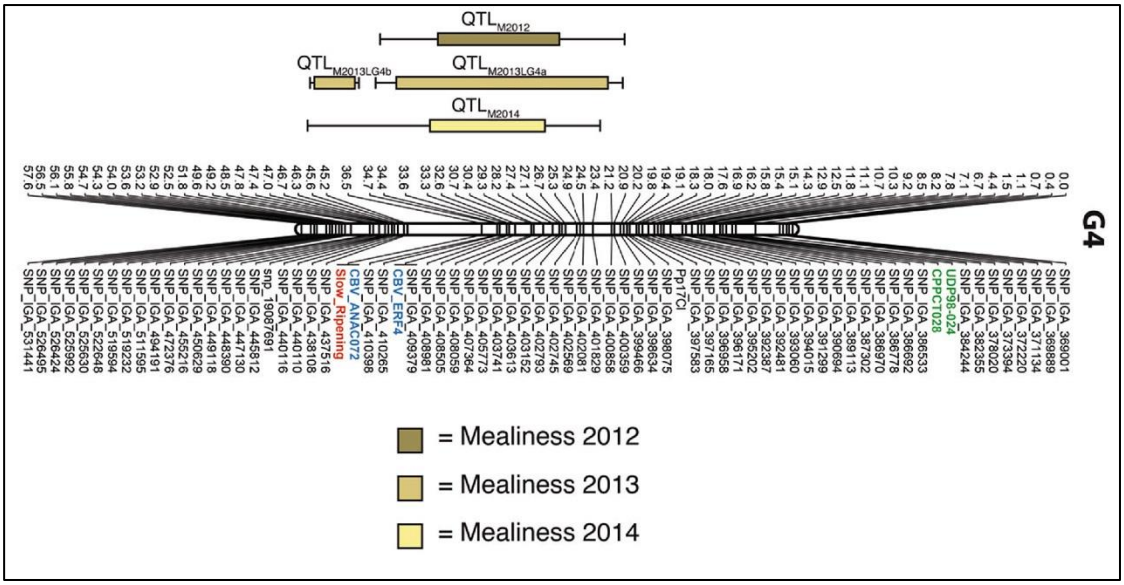


Figura 1: QTL detectados para el carácter de harinosidad (‘*mealiness*’, en inglés) en el grupo de ligamiento 4 durante tres temporadas de evaluación (2012, 2013 y 2014) (modificado de Nuñez-Lillo et al. 2015). En la figura se representa el grupo de ligamiento 4 del mapa genético ‘Venus x Venus’, mostrando los marcadores tipo SNP (en negro), SSR (en verde), marcadores de PCR (en azul) y marcadores morfológicos (en rojo) que lo conforman. Sobre el grupo de ligamiento se observan las barras que indican la detección de una región asociada a harinosidad con los datos fenotípicos de las temporadas 2012 (marrón oscuro), 2013 (marrón claro) y 2014 (mostaza).

Por otra parte, el RNA-Seq describe el conjunto de métodos computacionales y experimentales para identificar la abundancia de secuencias de ARN en muestras biológicas. Experimentalmente implica la extracción de las muestras desde el sujeto biológico, la preparación y secuenciación de librerías y el análisis bioinformático de éstas (Korpelainen *et al.* 2015).

El principal objetivo de un experimento RNA-Seq es identificar las secuencias, estructuras y abundancias de los ARN en una muestra específica. La secuencia implica el orden sucesivo de los nucleótidos, la estructura es ubicar las distintas zonas que conforman una secuencia de ARN como la posición del promotor, frontera intron-exon, regiones sin traducir (UTR; del inglés “Untranslated Regions”), etc. La abundancia es meramente la cantidad de transcrito. Con esta información es posible encontrar genes codificantes de proteínas o reguladores como ARN no codificantes, los cuales son moléculas de ARN que no son traducidos en una proteína y están involucrados en regulación de traducción de ARN mensajeros y procesos similares (Korpelainen *et al.* 2015).

Para el duraznero, la secuencia completa de ARN obtenida mediante el proceso de RNA-Seq puede ser mapeada directamente a su genoma ya que éste es de libre acceso (Verde *et al*. 2013). Con esto, es posible descubrir nuevos genes codificantes de proteínas gracias a predicciones computacionales, siempre que existan bases de datos de proteínas ortólogas (Mathews *et al.* 1999).

A través de la abundancia de transcritos detectada mediante un experimento de RNA-Seq, es posible realizar comparaciones entre distintos estados, etapas fisiológicas o celulares y analizar qué transcritos priman sobre una condición u otra, pudiendo comparar sus distintos niveles. Esto es lo que se conoce como un análisis de expresión diferencial y es el paso siguiente luego de la anotación de los genes codificantes de proteínas (Korpelainen *et al.* 2015).

Al asociar información de mapas genéticos con datos provenientes de RNA-Seq se pueden obtener QTL de expresión o eQTL. Conceptualmente, si un genotipo en un locus se correlaciona con la expresión de un gen, entonces esa región contiene, potencialmente, un regulador especifico de la expresión de aquel gen. Para detectar estas regiones se genotipifica (utilizando marcadores genéticos) una población de individuos contrastantes (fenotípicamente distintos) y se miden los niveles de expresión de los transcritos utilizando técnicas como el RNA-Seq. Luego se correlacionan los marcadores genéticos con los niveles de expresión de los transcritos. Una correlación en base a un valor de LOD superior o igual a 3 entre un marcador y la expresión de un transcrito especifico constituyen un QTL de expresión o eQTL (Michaelson *et al.* 2009).

Dicho de otra forma, un eQTL se encuentra cuando se asocia la información de los niveles de transcritos (o expresión) de cada gen anotado en el experimento de RNA-Seq a un mapa de ligamiento genético construido con marcadores. Luego se realiza una asociación estadística para detectar en qué regiones de este mapa se ubican los genes responsables de aquellos niveles de transcritos. Los eQTL, entonces, serán loci responsables por el control genético de variantes de ADN que alteran los niveles de expresión (Majewski y Pastinen 2011).

Los eQTL pueden ser de dos tipos: *cis* y *trans*. Los eQTL en *cis* son aquellos cuyo marcador responsable de un cierto nivel de expresión se encuentra físicamente en el mismo grupo de ligamiento que el eQTL, es decir, actúa localmente. Por otra parte, los eQTL en *trans* son aquellos cuyo marcador responsable se encuentra físicamente en otro grupo de ligamiento distinto al eQTL, es decir, actúa de forma remota (Figura 2).

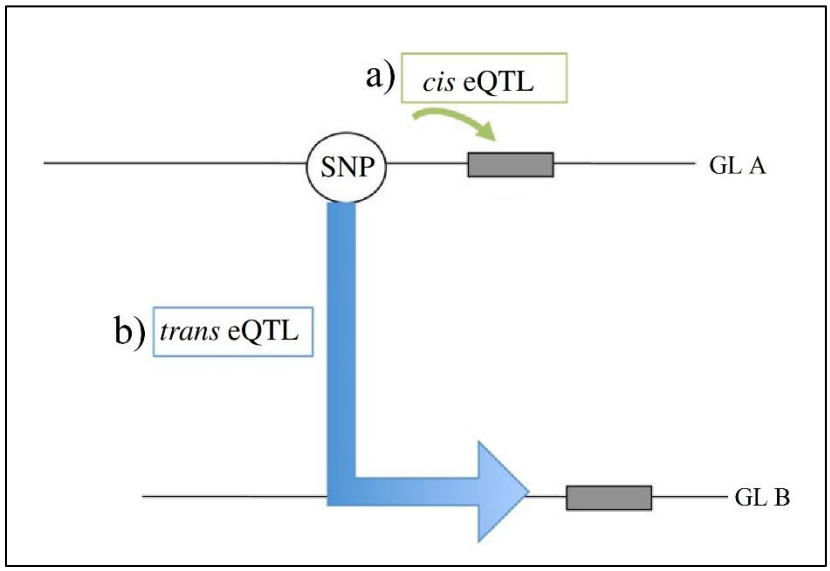


Figura 2: Representación gráfica de regulación génica por medio de QTL de expresión (eQTL) (modificado de Nica y Dermitzakis 2013). a) regulación de tipo *cis*, donde la posición del eQTL y el marcador que causa el fenotipo observado (abundancia de transcrito) colocalizan. Esta es la forma más probable de regulación. b) regulación de tipo *trans* en donde el eQTL no colocaliza, por lo que este eQTL representa el locus que controla la expresión de un marcador asociado a un gen.

Muchos estudios confirman la utilidad y efectividad de los eQTL. En un estudio llevado a cabo por Li en el año 2014 (Li *et al*. 2014) se buscaron eQTL en muestras obtenidas desde bases de datos de ARN mensajero de tejidos tumorales de mamas, colon, riñón, pulmón y próstata. Se asoció la expresión del ARNm a 149 loci que se conocía que presentaban riesgo para los 5 tipos de tumores. Esta información fue derivada de GWAS (del inglés "Genome-Wide Association Study"). La asociación entre loci y niveles de expresión reveló 42 eQTL significativamente asociados, los cuales fueron luego relacionados a transcritos que estarían entonces regulando la expresión y/o generación de tumores en los distintos tipos de tejidos. Se encontraron 8 transcritos asociados a

riesgo de Cáncer de mamas, 31 transcritos para riesgo de cáncer a la próstata, 1 para riesgo de cáncer de pulmón, 1 transcrito para riesgo de cáncer de riñón y 1 transcrito para cáncer de colon. Estos transcritos están listos para ser estudiados de manera funcional y determinar su rol en el aumento del riesgo de distintos tipos de canceres.

Con la intención de buscar los genes responsables del color morado intenso de la uva de vino, Huang en el año 2013 (Huang *et al*. 2013) realizó un estudio en la población compuesta por 191 individuos de F1 entre 'Syrah' x 'Grenache', ambos cultivares de uva vinífera. Habiendo un mapa de ligamiento disponible (Doligez *et al*. 2006) se extrajo ARN total de bayas congeladas y se asociaron los niveles de expresión de VvUFGT provenientes de un experimento de PCR en tiempo real con el mapa de ligamiento de la población consenso y con los parentales individuales. Se encontraron 2 eQTL, uno señalando al transcrito VvUFGT y otro a factores de transcripción VvMYBA. VvUFGT ha sido identificado previamente como el mayor locus involucrado en contenido de antocianinas de la baya (Fournier-Level *et al*. 2009), mientras que VvMYBA1 y VvMYBA2 son dos principales reguladores de síntesis de antocianinas en uva, parte de la familia MyB de reguladores transcripcionales (Kobayashi *et al*. 2002).

La identificación de genes candidatos y sus niveles de expresión asociados a harinosidad en durazno y nectarina es un proceso que requiere grandes esfuerzos de análisis genómicos. Por este motivo se realizó un análisis de RNA-Seq de una población de durazneros proveniente de individuos contrastantes para el carácter de harinosidad para luego asociarlos a datos de marcadores y conseguir eQTL que se correlacionen con el fenotipo de harinosidad.

d) Formulación de hipótesis:

Acorde a lo descrito anteriormente se describe la siguiente hipótesis:

Es posible identificar genes candidatos no mencionados previamente en la literatura para el carácter de harinosidad mediante la comparación de individuos jugosos y harinosos provenientes de una población segregante de duraznero (*Prunus persica* (L.) Batsch) utilizando análisis de RNA-Seq, QTL convencionales y QTL de expresión.

e) Objetivo general y específico:

Como objetivo general se plantea:

Identificar genes candidatos no mencionados previamente en la literatura involucrados en el carácter de harinosidad de la pulpa mediante el análisis de QTL, eQTL y la secuenciación masiva de transcriptomas de individuos contrastantes para el fenotipo.

Cuyos objetivos específicos son:

1. Realizar una caracterización fenotípica de los frutos de una población segregante para harinosidad de *Prunus persica* en cosecha y poscosecha.
2. Secuenciar y analizar el transcriptoma de frutos en individuos de una población segregante contrastantes para harinosidad.
3. Identificar regiones de los cromosomas asociadas al fenotipo harinoso usando el análisis de QTL.
4. Identificar genes candidatos para harinosidad utilizando la estrategia de QTL de expresión.
5. Validar los genes expresados diferencialmente por medio de RT-PCR cuantitativo.

Materiales y métodos:

a) Diseño experimental:

El diseño experimental consta de la generación de una población contrastante para el carácter de harinosidad, cuyos individuos fueron genotipados. Con aquellos datos se generó el mapa de ligamiento genético de la población en estudio. Este mapa fue cruzado con la informacion fenotipica del parametro de harinosidad de las todas las temporadas de evaluacion para obtener QTL convencionales asociados a harnisodidad en cada una de aquellas temporadas. En base a la informacíon fenotípica de las temporadas Choapinos 2014-2015 y Rinconada 2014-2015 se seleccionaron los segregantes a secuenciar. El material biológico provino de los frutos de Rinconada 2014-2015. Sobre este material se realiazon los análisis bioinformáticos para obtener los genes expresados diferencialmente, cuya abundancia de transcrito fue utilizada para buscar QTL de expresión. Posterior a eso se cruzaron los datos de los QTL convencionales y los QTL de expresión obteniendo así los genes candidatos (ver Figura 3).

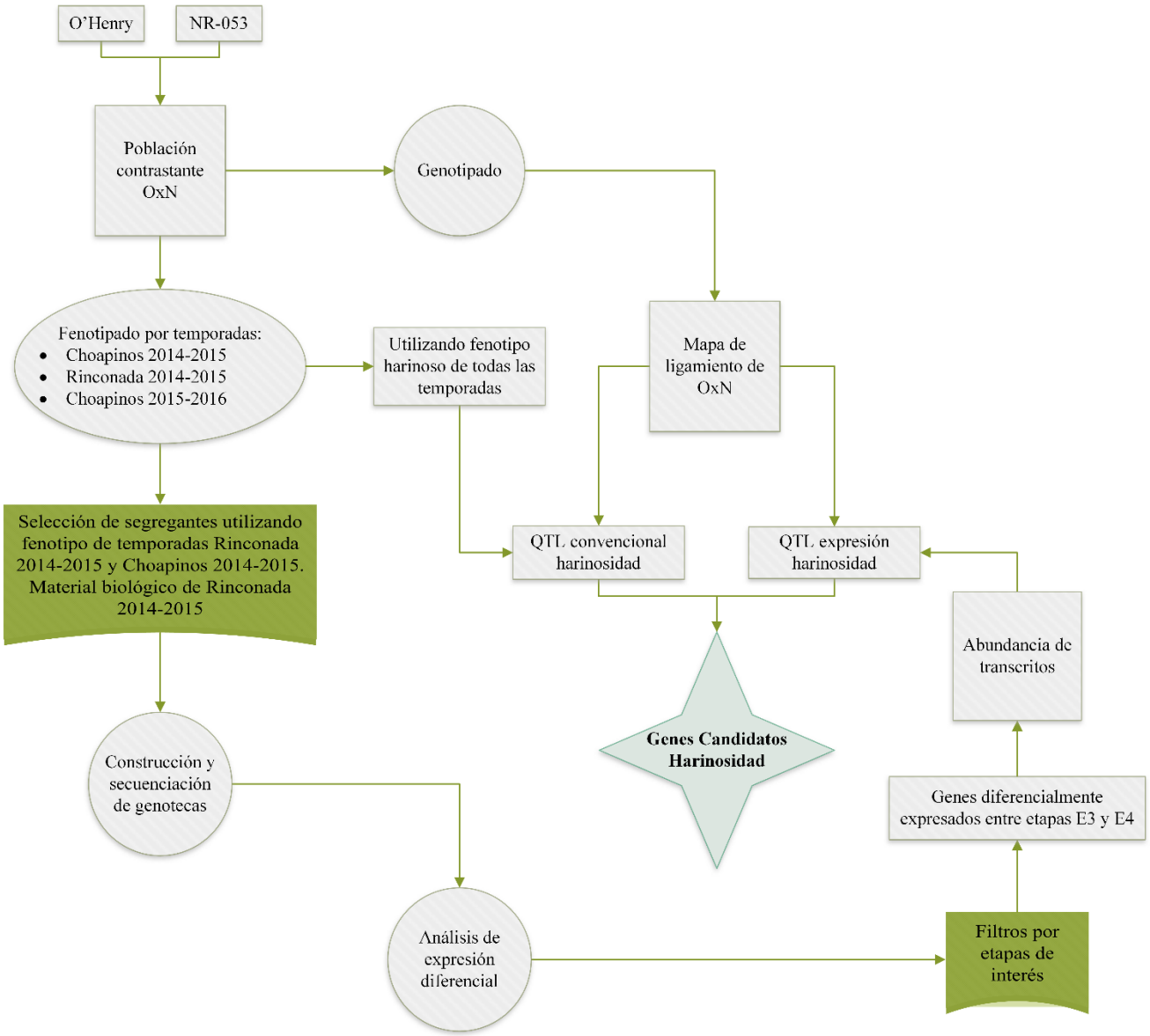


Figura 3: Esquema general del diseño experimental utilizado.

b) Obtención del Material Vegetal:

Se utilizó una población de 194 individuos segregantes provenientes del cruce entre la variedad de duraznero (*Prunus persica* (L.) Batsch) de pulpa amarilla fundente 'O'Henry' y la variedad de nectarina de pulpa blanca y fundente 'NR-053' proveniente del Programa de Mejoramiento Genético (PMG) de la Universidad de Chile. Trabajos previos de evaluación de estas variedades han sido realizados por el equipo de trabajo de la presente propuesta y han mostrado un desempeño contrastante de estas variedades en la vida de poscosecha ('O'Henry' tiene un pobre desempeño mientras que 'NR-053' uno excepcional). De esta manera, se obtuvo una población segregante para el carácter de harinosidad. Esta población fue nombrada como OxN.

Los 194 individuos están ubicados en dos localidades: Rinconada, comuna de Maipú (Región Metropolitana), en dependencias de la Universidad de Chile y Choapinos, comuna de Rengo (Región del Libertador General Bernardo O'Higgins) en INIA-Rayentué.

c) Caracterización de la información fenotípica de una población segregante:

A los frutos de esta población, en la temporada 2014-2015 se le midieron los parámetros de color de cubrimiento, color de fondo, contenido de jugo, firmeza de pulpa, peso del fruto, contenido de sólidos solubles en pulpa y fecha de cosecha.

El color de cubrimiento fue determinado por medio de una plantilla de color para duraznos y nectarinas perteneciente a la Asociación de Exportadores de Chile (ASOEX). El color de fondo se midió por medio de un DA-meter (T.R Turoni s.r.l ©) para medir la degradación de clorofila en los frutos (Costa *et al.* 2011), el cual al estar entre un valor de 0,8 y 1,2 se considera listo para cosechar. El parámetro de fecha de cosecha se determinó contando desde la fecha promedio de floración de los árboles (1 de septiembre) hasta el día en que estos se cosecharon.

El contenido de jugo fue medido por medio del método de los rodillos (Infante *et al.* 2009) en el cual se toma un corte de pulpa pesado previamente y se cuantifica la cantidad de jugo liberado al presionar el trozo de pulpa entre dos rodillos metálicos (Pasta Machine PPD001, Shule) por medio del cálculo de diferencia de masa antes y después de haber presionado el tejido. La firmeza se determinó utilizando un penetrómetro manual (Fruit Pressure Tester FT 011, Facchini srl®). Para

cuantificar el peso se utilizó una balanza granataria (ES1000HA, Belltronic®) y finalmente para medir el contenido de sólidos solubles de la pulpa se utilizó un refractómetro digital (PAL-1, ATAGO®).

Las distintas etapas de evaluación de los caracteres fenotípicos se definen como: etapa 1 (E1) o cosecha, etapa 2 (E2) o cosecha más "shelf life" (SL) donde "shelf life" es el periodo que corresponde al tiempo en que el fruto se mantuvo a 20 ºC en espera de alcanzar la firmeza de consumo, la cual va entre 2-4 lb. Etapa 3 (E3) o 30 días de almacenamiento en frío a 0 ºC y finalmente la etapa 4 (E4) que corresponde al ablandamiento después de frío (Figura 4).

Todos los parámetros fenotípicos fueron obtenidos desde cinco frutos en E1 por árbol (Figura 4, Tabla 1) a excepción de firmeza que se midió a lo largo de todas las etapas y la jugosidad que fue medida para nueve frutos por árbol en la etapa E4.

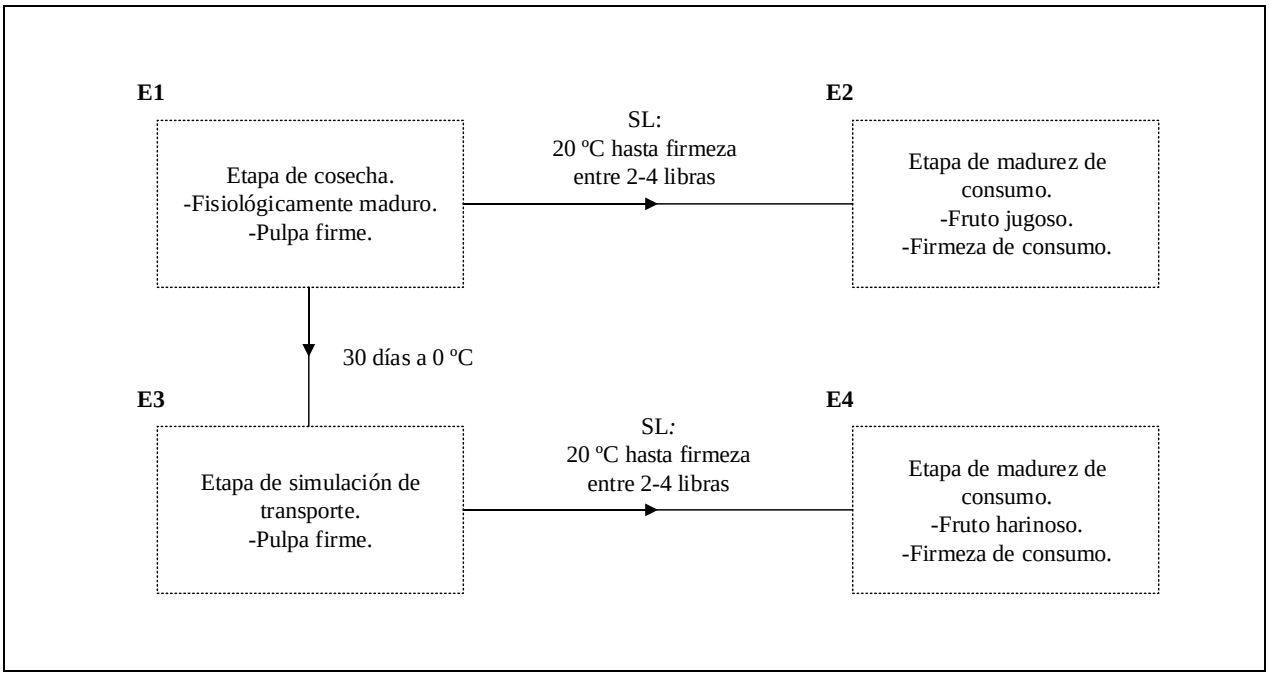


Figura 4: Momentos de evaluación fenotípica (E1, E2, E3 y E4) y transcriptómica de los frutos de la población segregante.

Luego de haber extraído los datos fenotípicos de las muestras recolectadas, estas fueron trozadas, rotuladas e inmediatamente congeladas en nitrógeno líquido para su almacenamiento en congelador a -80 ºC con el objetivo de mantener el estado de las muestras y que estas no generen cambios a nivel transcripcional a causa de su manipulación y que pudiese alterar los análisis posteriores.

Tabla 1: Parámetros de calidad del fruto y etapas de evaluación de la población segregante.

Caracteres	Etapas de evaluación			
	E1	E2	E3	E4
Contenido de sólidos solubles (ºBrix)	X			
Peso (g)	X			
Color de fondo y cubrimiento	X			
DA-Meter	X			
Firmeza (lb)	X	X	X	X
Jugosidad (%)				X

d) Selección de individuos contrastantes para el carácter de harinosidad:

Los datos fenotípicos para el carácter de contenido de sólidos solubles y fecha de cosecha fueron analizados estadísticamente utilizando la prueba t de Student con un nivel de confianza del 95% (p-value = 0,05) por medio del software R-Project 3.2.1 (R Development Core Team. 2015).

La selección de los frutos fue realizada por medio del software InfoGen (Balzarini y Rienzo 2014) en el cual se introdujeron los datos de 'jugosidad' (o porcentaje de jugo medido en porcentaje), 'firmeza' (libra-fuerza), 'fecha de cosecha' (días) y 'contenido de sólidos solubles' (ºBrix) para estimar diferencias por medio de modelos mixtos. Dentro del programa se determinó la variable dependiente como cada uno de los frutos y los criterios de clasificación fueron 'lugar' y 'porcentaje de jugo', siendo 'lugar' la variable aleatoria y 'porcentaje de jugo' la variable fija. La 'firmeza' fue incluida como covariable utilizando la función de 'varIdent()' otorgada por el software. Luego de calcular el modelo mixto, el programa automáticamente realizó una comparación entre los valores de jugosidad resultantes para cada fruto utilizando el estadístico LSD Fisher con un nivel de significancia de 0,05 (p-value = 0,05).

e) Extracción de ARN:

La molienda y preparación de las muestras se realizó utilizando un molino analítico (A11 Basic, IKA®) el cual fue previamente acondicionado con nitrógeno líquido. En este se pusieron las muestras trozadas de duraznos congelados provenientes del campo de Rinconada (temporada 2014-2015) y fueron molidas hasta obtener un polvillo. El polvillo fue rápidamente almacenado en tubos Falcon de 50 ml (previamente enfriados con nitrógeno líquido) e inmediatamente almacenados en congelador a -80 ºC para su posterior uso.

La extracción del ARN de los frutos seleccionados fue realizada por medio del kit comercial Spectrum Plant Total RNA Kit (Sigma-Aldrich®), siguiendo el protocolo provisto por el fabricante. La calidad del ARN extraído fue evaluado por medio del número de calidad del ARN (RQN, del inglés "RNA Quality Number") que otorga el equipo de electroforesis capilar (Fragment Analyzer). Luego, la concentración del ARN se determinó utilizando un fluorímetro (Qubit® 2.0 Fluorometer) siguiendo el protocolo recomendado por el proveedor.

f) Construcción, calidad y secuenciación de librerías:

Las librerías fueron confeccionadas utilizando el Kit TruSeq RNA Sample Preparation (Illumina®) a partir de una concentración de ARN inicial de 1000 ng/ul para todas las muestras siguiendo el protocolo del fabricante.

Posteriormente estas fueron analizadas para verificar la concentración y calidad utilizando un equipo de electroforesis capilar (Fragment Analyzer) por medio del kit High Sensitivity NGS Fragment Analysis Kit (Advanced Analytical®).

Finalmente, las librerías resultantes fueron empacadas y enviadas a temperatura ambiente a la empresa surcoreana de secuenciación Macrogen. La secuenciación fue llevada a cabo utilizando un equipo HiSeq 2500 de Illumina, con lecturas "paired-end" de 126 pares de bases de largo para un tamaño de librerías promedio de 300 pares de bases.

g) Mapeo de las lecturas y análisis de expresión diferencial:

Las secuencias provenientes de Macrogen fueron filtradas utilizando Flexbar aplicando un parámetro mínimo de valor de Phred (valor Q) de 20 y eliminando todas aquellas lecturas cuyo largo fuera menor a 35 pares de bases con motivos de mejorar su calidad y aumentar la tasa de alineamiento.

Posteriormente, las librerías fueron alineadas contra el genoma de referencia de *Prunus persica* disponible en la base de datos de Rosáceas, GDR (Genome Database for Rosaceae, http://www.rosaceae.org). El mapeo se llevó a cabo por medio del software Tophat versión 2.0.14 (Trapnell *et al*. 2009) siguiendo el manual que se encuentra en la página web del software (https://ccb.jhu.edu/software/tophat/index.shtml).

Se utilizaron los paquetes del software R: Limma (Law *et al*. 2014), edgeR (Robinson *et al*. 2009) y DESeq2 (Love *et al*. 2014) para realizar expresión diferencial. Para esto se calcularon las cuentas utilizando el paquete de Python HTSeq (Anders *et al*. 2015), específicamente HTSeq-Count, siguiendo el manual encontrado en la página web (http://www-huber.embl.de/users/anders/HTSeq/doc/count.html).

h) Análisis de ontología génica:

Los análisis de ontología génica fueron realizados con el paquete GOSeq (Young *et al*. 2010) para R, siguiendo el manual provisto junto con el paquete. Se utilizaron los datos de la base de datos de Rosáceas, en la cual se asocian los genes de *Prunus persica* a términos GO. El análisis se llevó a cabo con las cuentas normalizadas obtenidas del paquete edgeR.

i) Análisis de QTLy eQTL:

Por medio del software MapQTL (Van Ooijen 2009) se hicieron análisis de QTL y eQTL utilizando el fenotipo convencional de harinosidad en tres temporadas de evaluación (Choapinos temporadas 2014-2015 y 2015-2016; Rinconada temporada 2014-2015) para los QTL y las cuentas normalizadas de edgeR para los eQTL (material biológico de Rinconada 2014-2015). Los análisis fueron llevados a cabo entre las condiciones de harinoso contra jugoso punto por punto, es decir, para cada etapa se realizó un análisis de eQTL utilizando las cuentas normalizadas de edgeR de la comparación de frutos harinosos contra jugosos. La asociación entre loci y fenotipo se realizó con el estadístico Kruskal-Wallis.

j) Validación del transcriptoma:

El perfil de expresión obtenido mediante el RNA-Seq fue validado utilizando la técnica de RT-qPCR. Se eligieron cinco genes al azar provenientes de una publicación previa conducida por el grupo de trabajo (Sanhueza *et al*. 2015) para realizar esta validación en frutos jugosos y harinosos por separado (cinco genes para jugosos, cinco genes para harinosos); cada PCR se realizó con tres réplicas biológicas. Se usó el master mix SYBR Green PCR master mix (Applied Biosystems) y un sistema AriaMx Real-Time PCR (Agilent Technologies). La PCR involucra pasos de 95 °C por dos minutos seguido por 40 ciclos a 95 °C por 5 segundos y 60 °C por 30 segundos.

Para la curva de "melting" se realizó un PCR de 1 ciclo a 95 ºC por 30 segundos, 65 ºC por 30 segundos y finalmente 95 ºC por 1 segundo. Se verificó que la eficiencia de los partidores se encontrase entre 95 a 105. El gen normalizador fue el factor de elongación 2 (TEF2) y la condición normalizadora fue E1.

Resultados:

a) Caracterización fenotípica de los frutos de una población segregante para harinosidad:

La fenotipificación del material vegetal conllevó el análisis de jugosidad en frutos en E4, la medición de contenido de sólidos solubles en E1 (por árbol) y la anotación de la fecha de cosecha de cada árbol. Los resultados de este ejercicio se grafican en la Figura 5.

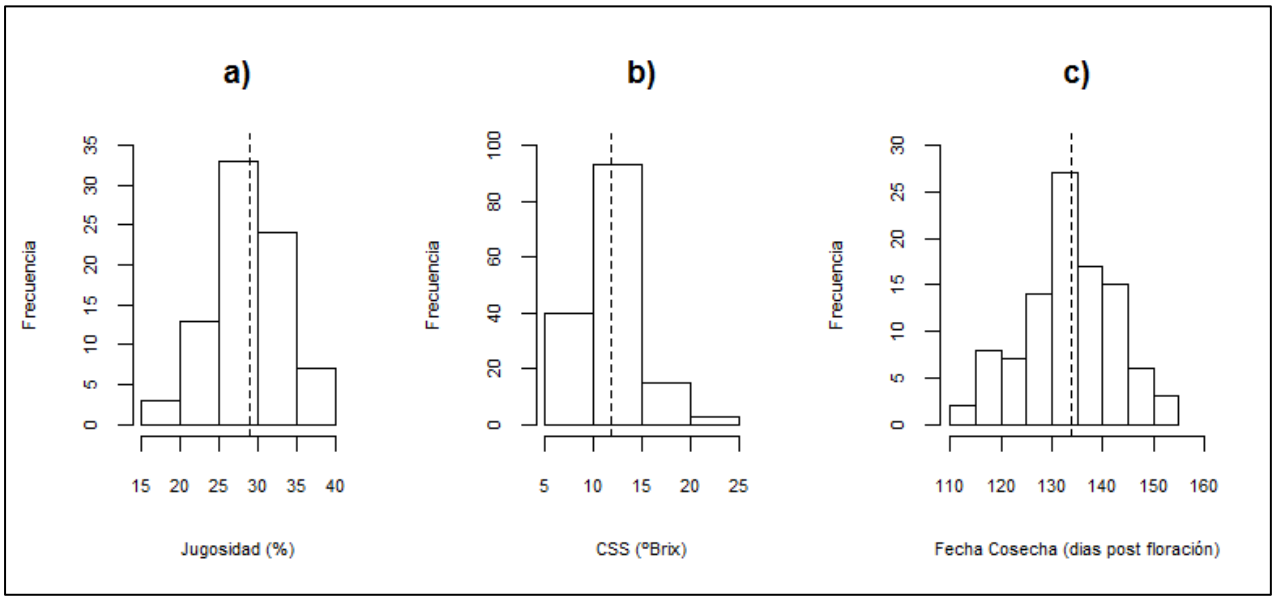


Figura 5: Histogramas de resultados de fenotipificación de los segregantes de duraznero. En a) se muestra la frecuencia de la jugosidad medida en porcentaje. La figura b) muestra el contenido de sólidos solubles medido en grados Brix y la figura c) grafica la frecuencia de la fecha de cosecha medida en días post floración. La línea punteada representa el promedio de la población para cada parámetro.

Respecto al porcentaje de jugosidad, el máximo de la población fue de 37,44% y un mínimo de 16,29%, siendo el promedio un valor de porcentaje de jugo de 28,84% (Figura 5a; línea punteada) y mostrando una distribución unimodal. El contenido de sólidos solubles presenta un máximo y mínimo de 20,61 y 7,45 grados brix, con un promedio de 11,77 grados brix (Figura 5b; línea punteada), también en una distribución unimodal. La fecha de cosecha fue contada a partir del 1 de septiembre (días post floración) y posee un promedio de 133 días (Figura 5c; línea punteada) con una cantidad de días mínimos de 114 días y máximo de 154 días en una distribución unimodal.

b) Construcción de un mapa genético de la población OxN:

Utilizando 485 marcadores tipo SNP, 11 microsatélites y dos marcadores morfológicos se construyó el mapa genético consenso de la población OxN. Posee un tamaño de 717,6 cM, una densidad de 1,44 cM por marcador y tiene nueve grupos de ligamiento. El grupo de ligamiento 5 no tenía suficientes marcadores heterocigotos por lo cual están separados en 5.1 y 5.2 (Figura 6).

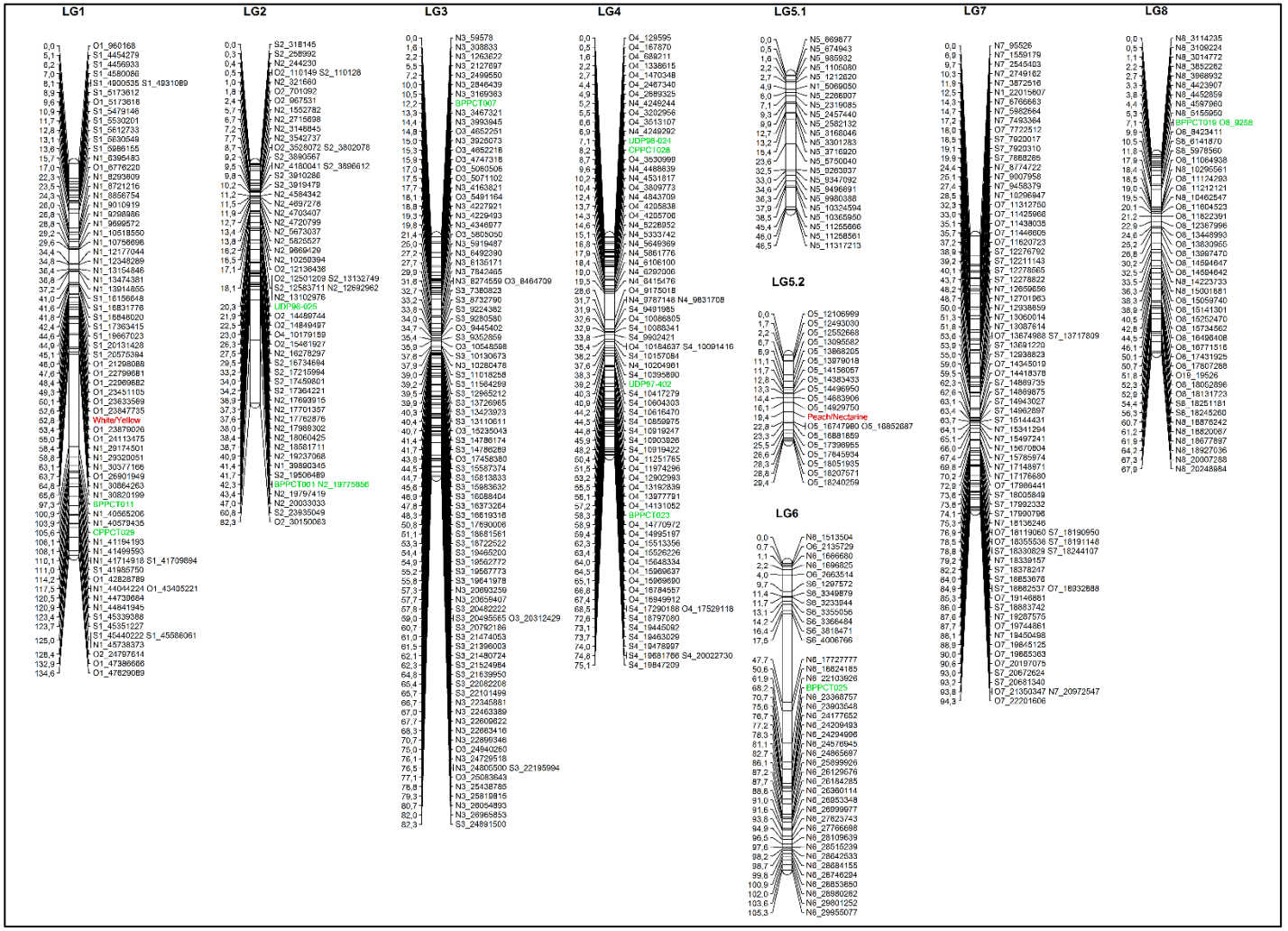


Figura 6: Mapa de ligamiento genético de la población OxN confeccionado con 498 marcadores en total. 485 de ellos son de tipo SNP (en color negro), 11 son microsatélites (en color verde) y dos son marcadores morfológicos (en color rojo).

c) Análisis de QTL convencionales para harinosidad:

Se asociaron los datos fenotípicos del porcentaje de jugosidad medido previamente a los marcadores del mapa de ligamientos de la población OxN (Figura 6) por medio de la prueba de Kruskal-Wallis en el software MapQTL. Esto reveló QTL convencionales para el fenotipo de harinosidad, cuyos resultados por grupo de ligamiento se muestran en la Figura 7.

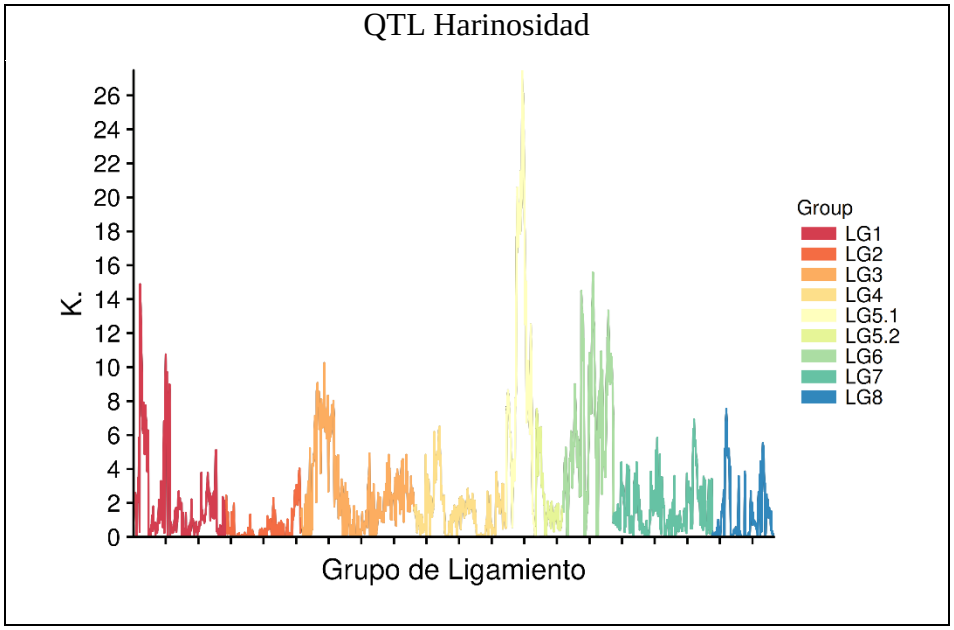


Figura 7: QTL convencionales para el carácter de harinosidad. Los grupos de ligamiento están marcados con color acorde a lo indicado en la escala.

Se observó una asociación estadística relevante entre marcador y fenotipo en los grupos de ligamiento 5.1 y 6 con una probabilidad del 99.95% (p-value= 0.05) cuando el valor de K es igual o mayor a 8,0. Los grupos de ligamiento 1, 3, 5.1 y 6 superan este valor en al menos una ocasión. En particular, el grupo de ligamiento 1 supera el valor de K=8 en dos ocasiones, sin embargo, es en distintas temporadas de evaluación (Figura 8). Únicamente los grupos de ligamiento 5.1 y 6 superan un K=8 en tres ocasiones simultáneamente (Figura 9). El detalle del resto de los grupos de ligamiento se encuentra en la Figura Suplementaria 1, en la sección de Anexos.

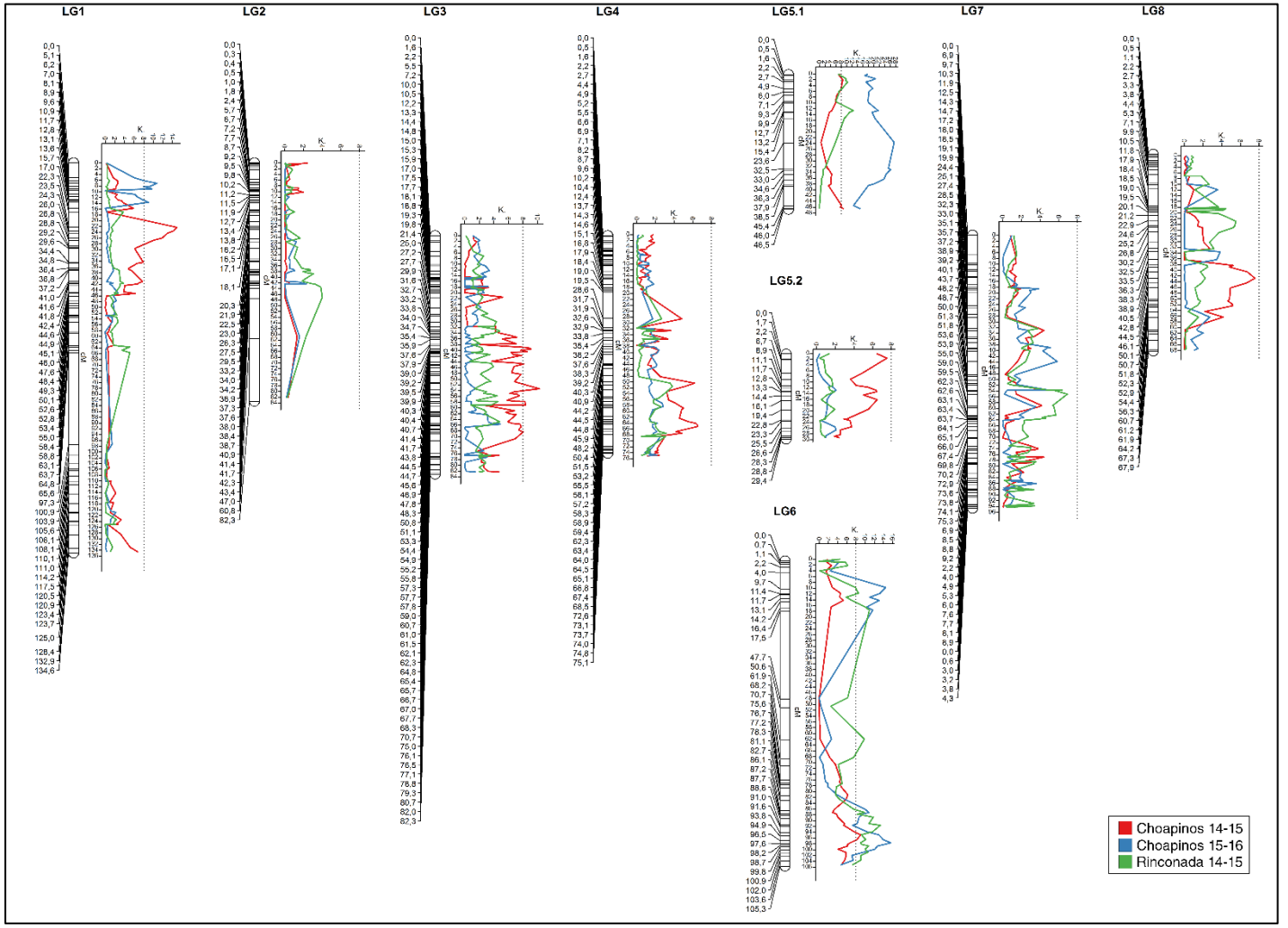


Figura 8: QTL convencionales sobrepuestos al mapa de ligamiento genético de la población OxN. Cada grupo de ligamiento muestra la posición en centiMorgan (cM) de sus marcadores. Los QTL presentan en su eje Y el valor de asociación estadística entregado por la prueba de Kluskal-Wallis (K) y en el eje X la distancia genética en cM. La línea roja representa la temporada 2014-2015 de evaluación fenotípica para el campo de Choapinos. La línea azul representa la temporada 2015-2016 para el mismo campo y la línea verde representa la temporada 2014-2015 de Rinconada.

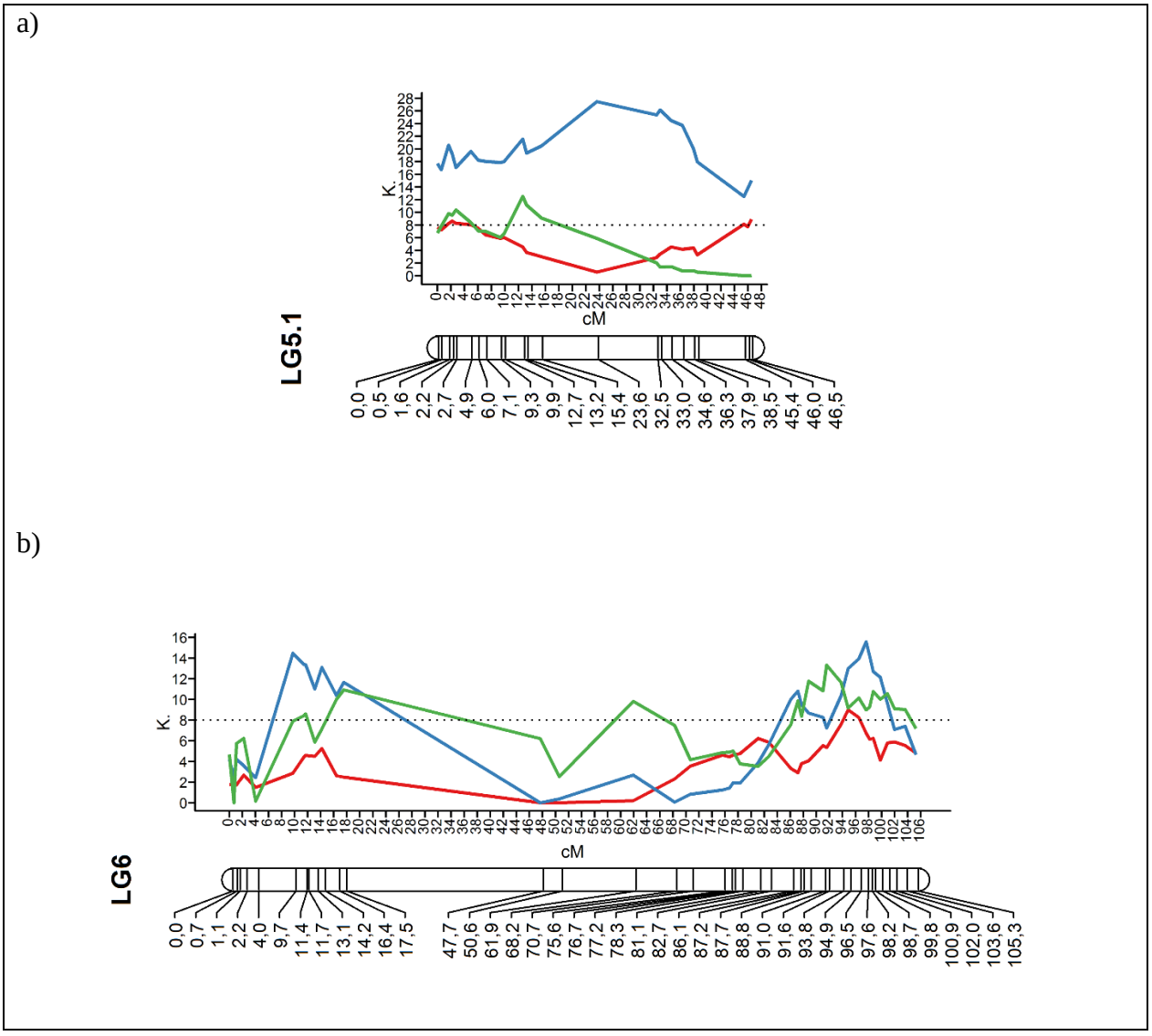


Figura 9: Detalle de grupos de ligamientos que poseen QTL convencionales en al menos tres temporadas de evaluación. La línea roja representa la temporada 2014-2015 de evaluación fenotípica para el campo de Choapinos. La línea azul representa la temporada 2015-2016 para el mismo campo y la línea verde representa la temporada 2014-2015 de Rinconada.

d) Selección de segregantes para secuenciación:

Utilizando los datos del fenotipado se realizó un análisis de modelos mixtos con la finalidad de seleccionar aquellos frutos que fueran estadísticamente distintos en base a su porcentaje de jugosidad. Los frutos seleccionados, el código del árbol del que provienen y el porcentaje de jugosidad se detallan en la Tabla 2. Aquellos frutos considerados jugosos son los que poseen un porcentaje de jugosidad mayor al 30%. En promedio, los frutos seleccionados poseían entre 35% a 39% de jugosidad mientras que los harinosos fluctuaron en promedio entre 15% a 22% de contenido de jugo.

Tabla 2: Resumen del fenotipo de jugosidad para cada fruto (replica biológica) perteneciente a cada segregante, así como su promedio y desviación estándar. DE: desviación estándar.

Segregante	Fruto	Jugosidad (%)	Promedio frutos	DE
19	1	35,0		
19	5	34,0	35,9	2,47
19	8	38,7		
37	7	13,9		
37	2	17,6	15,4	1,98
37	1	14,7		
60	1	38,5		
60	6	40,0	39,2	0,77
60	9	39,1		
124	2	37,0		
124	9	40,6	39,8	2,43
124	3	41,7		
161	4	36,4		
161	5	33,3	35,7	2,15
161	7	37,5		
172	4	20,9		
172	5	20,0	20,3	0,54
172	9	20,0		
173	5	38,5		
173	6	36,4	37,1	1,21
173	1	36,4		
180	3	22,2		
180	5	23,3	22,4	0,87
180	6	21,6		
197	2	15,0		
197	6	16,7	17,6	3,22
197	7	21,2		
200	8	19,0		
200	9	19,0	19,9	1,55
200	7	21,7		

e) Métrica de secuenciación de librerías de ARN:

Las librerías de ARN de los frutos seleccionados previamente en conjunto con las librerías de las otras etapas de evaluación ya mencionadas fueron secuenciadas utilizando el servicio que presta la empresa Macrogen. Estas secuencias fueron primeramente filtradas antes de alinear las lecturas contra el genoma de referencia.

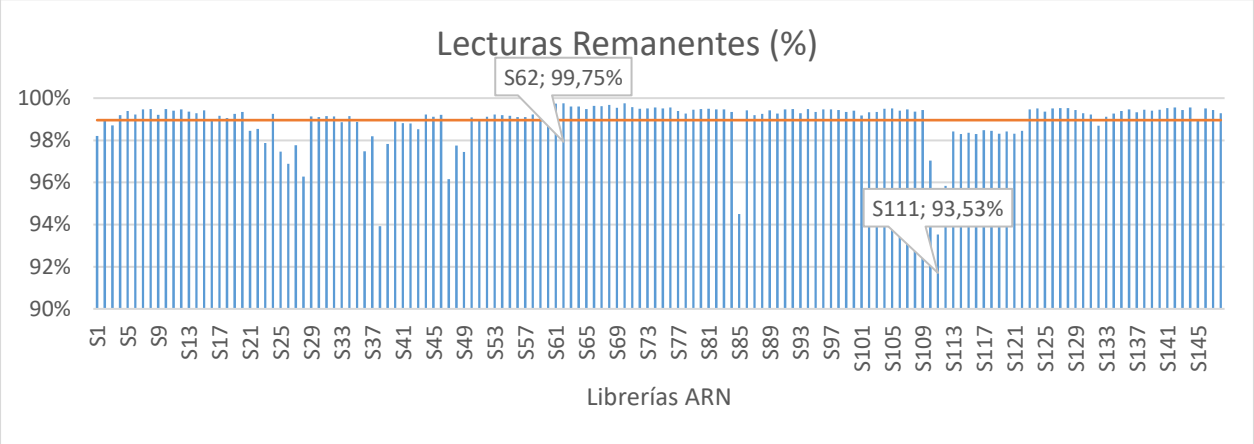


Figura 10: Cantidad de lecturas remanentes para cada librería secuenciada luego de aplicar filtros de calidad para un valor de Phred mínimo de 20 y remover todas las secuencias que tuviesen un largo menor a 35pb utilizando Flexbar. El valor promedio se indica con una línea color naranja.

Para mejorar la calidad de los alineamientos posteriores se aplicaron filtros utilizando Flexbar. Se eliminaron todas aquellas secuencias cuya calidad o valor de Phred fuese menor a 20. También se eliminaron las secuencias que tuviesen un largo menor a 35 pares de bases.

La librería que perdió más secuencias después del filtrado fue la S111, que mantuvo el 93,53% de sus lecturas, mientras que la S62 mantuvo un 99,75% de sus lecturas luego del filtro, siendo el promedio de lecturas remanentes de 98.96% para todas las librerías. Los valores exactos de lecturas remanentes se encuentran en la Tabla Suplementaria 1 en la sección de Anexos.

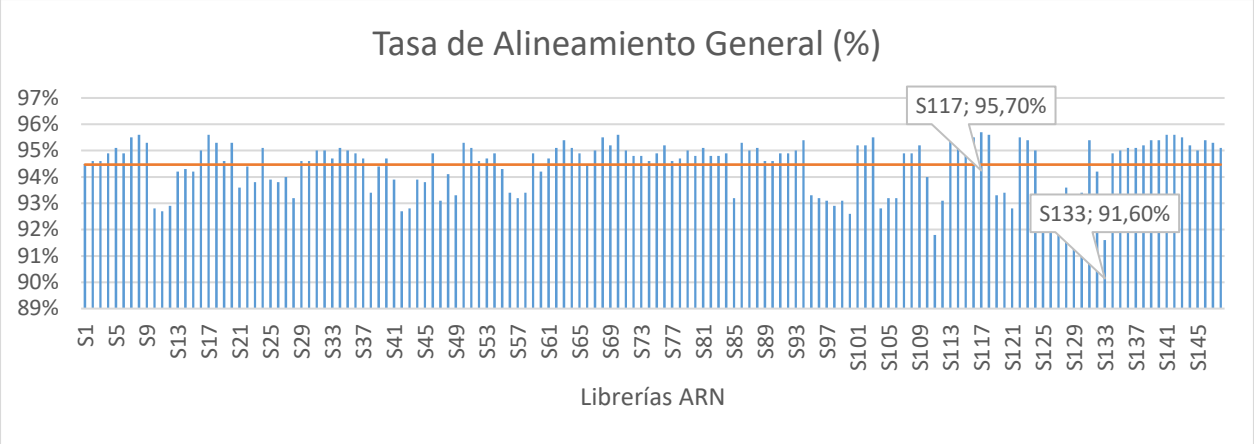


Figura 11: Tasa de alineamiento general para cada librería secuenciada utilizando Tophat2. El valor promedio se indica con una línea color naranja.

La tasa de alineamiento general tiene un promedio de 94,47%, su punto más alto fue la librería S117 con un 95,70% y el más bajo la librería S133 con 91,60% de alineamiento general. Este parámetro indica, de manera general, cuántas lecturas se alinearon contra el genoma de referencia. Los valores exactos del alineamiento general se encuentran en la Tabla Suplementaria 1 en la sección de Anexos.

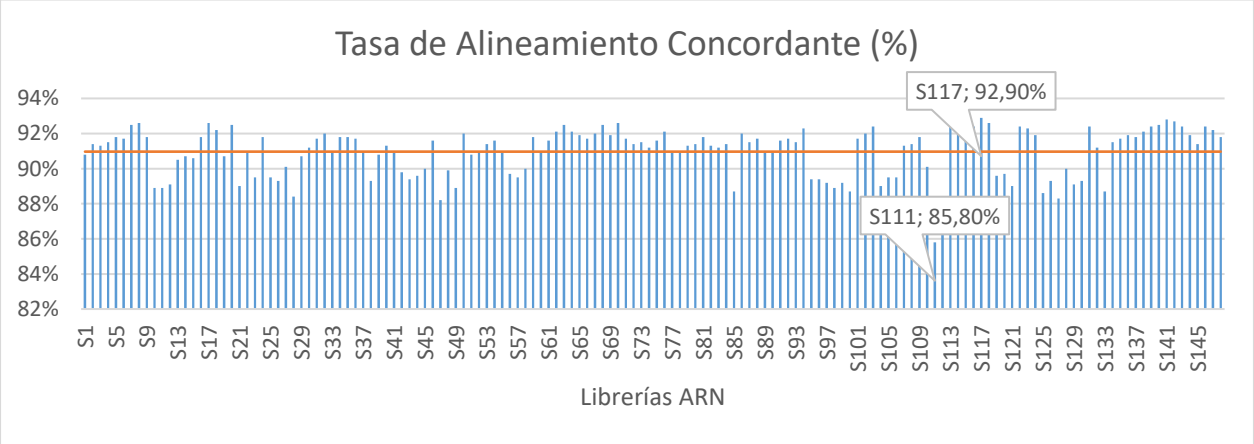


Figura 12: Tasa de alineamiento concordante para cada librería secuenciada utilizando Tophat2. El valor promedio se indica con una línea color naranja.

La tasa de alineamiento concordante explica cuántas lecturas se alinearon en ambos sentidos (izquierda y derecha) contra el genoma de referencia. En promedio, un 90,97% de las lecturas tienen un alineamiento concordante. El punto mayor y menor fueron las librerías S117 con 92,90% y S111 con 85,80% respectivamente. Los valores exactos del alineamiento concordante se encuentran en la Tabla Suplementaria 1 en la sección de Anexos.

f) Análisis de expresión diferencial:

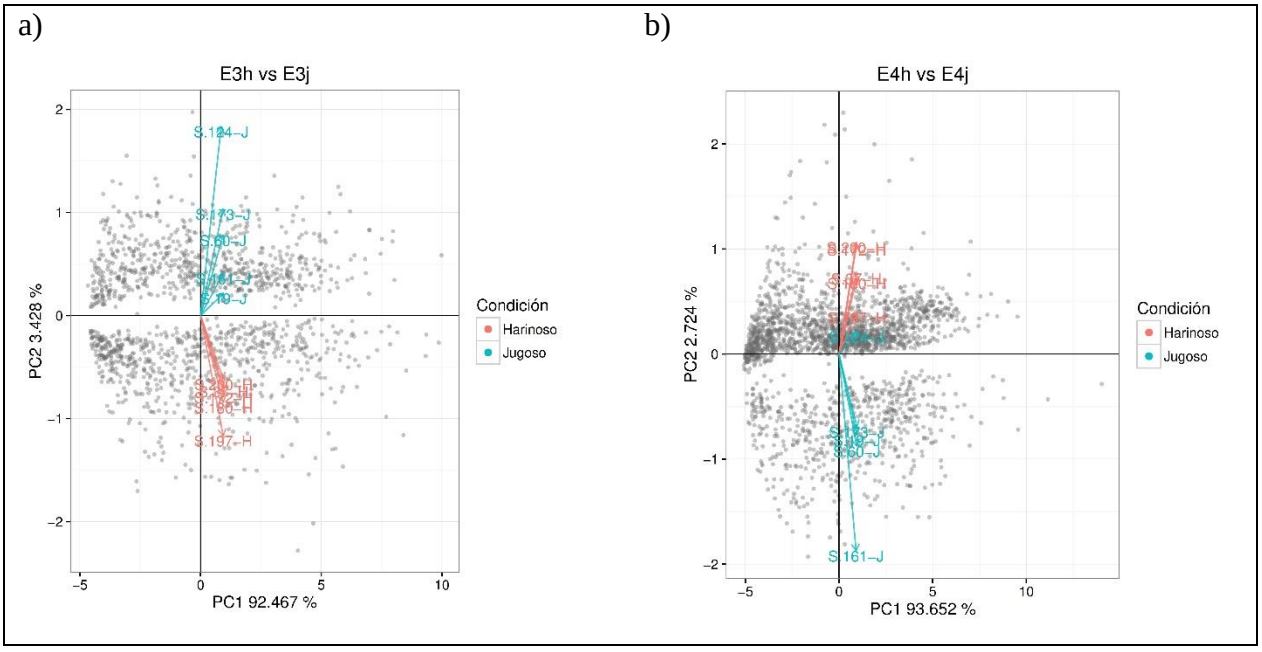


Figura 13: Análisis de componentes principales confeccionado a partir de los promedios de las cuentas normalizadas de los transcritos expresados diferencialmente obtenidos por medio del paquete de R 'edgeR' para los frutos de la etapa 3, harinosos (E3h) y jugosos (E3j) (a) y para los frutos de la etapa 4, harinosos (E4h) y jugosos (E4j) (b). Las flechas de color rojo denotan aquellos segregantes con frutos harinosos y las flechas color celeste aquellos con frutos jugosos.

Por medio de un análisis de componentes principales (PCA) utilizando las cuentas normalizadas que entrega el paquete de R 'edgeR' se pudo corroborar la diferencia entre aquellos frutos harinosos y jugosos que habían sido anteriormente seleccionados durante el fenotipado.

El PCA de E3h vs E3j (Figura 13a) posee las cuentas normalizadas de 1.409 transcritos expresados diferencialmente, y estos se diferenciaron claramente entre frutos jugosos y harinosos, mientras que para las cuentas normalizadas de 2.168 transcritos expresados diferencialmente del PCA de E4h vs E4j (Figura 13b) todos los frutos jugosos se agruparon en la parte inferior a lo largo del eje y, excepto el valor promedio de los frutos del segregante 124 (Tabla 2).

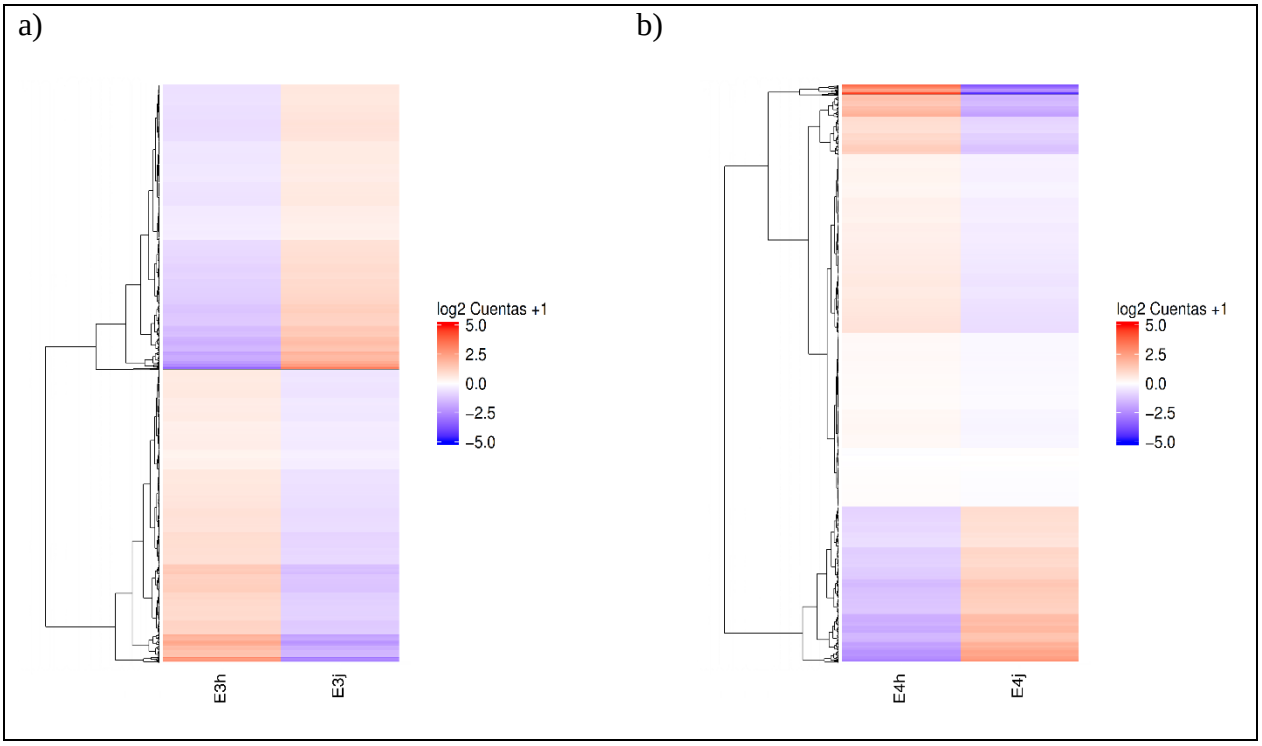


Figura 14: Mapas de calor para las cuentas de los transcritos expresados diferencialmente de la etapa 3 de los frutos harinosos (E3h) contra la etapa 3 de los frutos jugosos (E3j) (a) y de las cuentas de los transcritos expresados diferencialmente de los frutos harinosos en la etapa 4 (E4h) contra los jugosos en la misma etapa (E4j). Se utilizó el logaritmo en base dos de las cuentas normalizadas +1 de cada gen expresado diferencialmente en las comparaciones mencionadas anteriormente, siendo 1409 transcritos para E3h contra E3j y 2168 transcritos para E4h contra E4j. El dendrograma mostrado a la izquierda de cada mapa de calor denota la separación jerárquica entre los transcritos expresados diferencialmente.

Los mapas de calor muestran niveles de expresión distintos entre frutos jugosos y harinosos para los 1.409 transcritos expresados diferencialmente en la comparación E3h vs E3j. Del total, 714 transcritos están sobreexpresados y 695 están con expresión reprimida (Figura 14a). Por otro lado, en la comparación E4h vs E4j existen 2.168 transcritos expresados diferencialmente en total y 1.583 de ellos están sobre expresados, mientras que 695 de los transcritos se encuentran con expresión reprimida (Figura 14b).

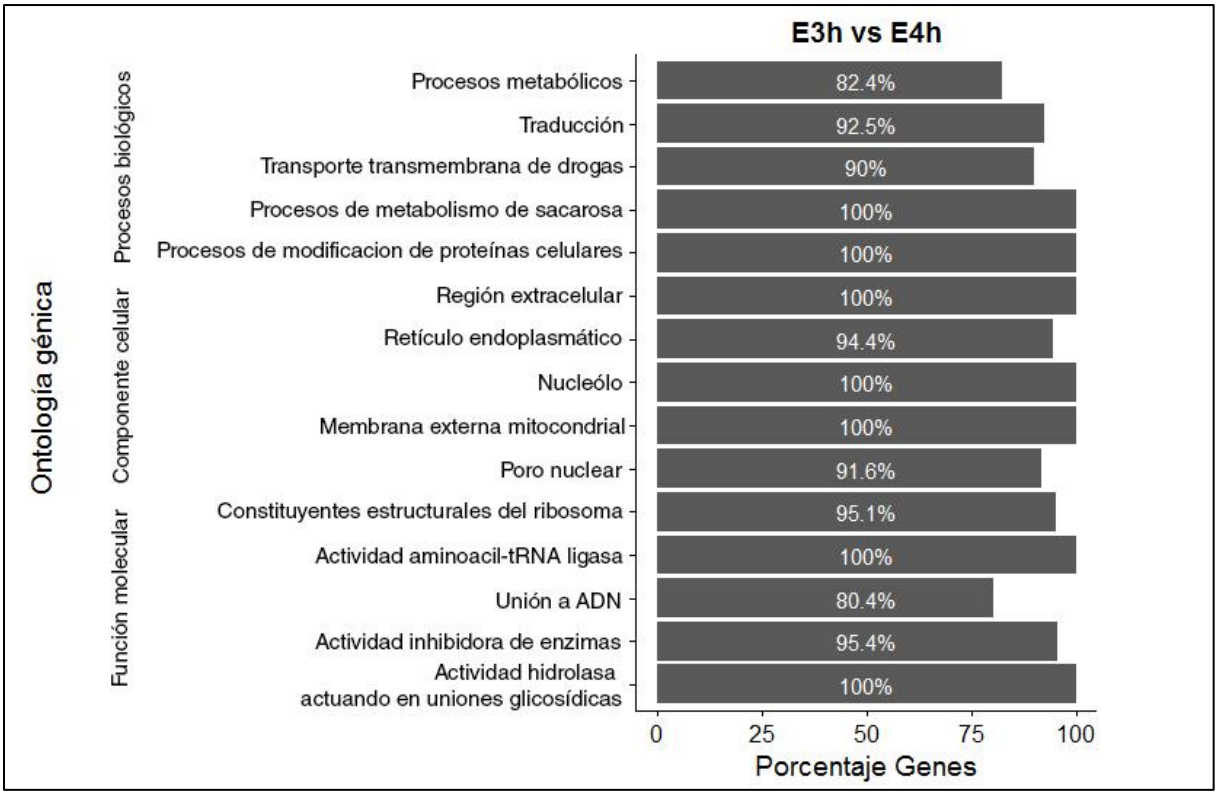


Figura 15: Análisis de ontología génica obtenido mediante el paquete de R 'Goseq' de todos los transcritos obtenidos en el análisis de expresión diferencial entre los frutos de la etapa 3 harinosos contra los frutos de la etapa 4 harinosos (E3h vs E4h). Se graficaron los 5 términos por categoría ontológica más enriquecidos en base su p-value. En las barras se indica el porcentaje de transcritos expresados diferencialmente respecto del total de transcritos asignados a esa categoría en particular.

Se realizó un análisis de ontología génica entre los frutos harinosos de la etapa 3 y los harinosos de la etapa 4 (E3h vs E4h, Figura 15). Del análisis se tomaron los 5 primeros términos más enriquecidos para cada categoría ontológica. Para los procesos biológicos se encuentran principalmente enriquecidos los términos que corresponden a procesos metabólicos, traducción, transporte transmembrana de drogas, metabolismo de sacarosa y procesos de modificación de proteínas celulares. Por otro lado, en la categoría de componente celular se encuentran enriquecidos los términos que corresponden a región extracelular, retículo endoplasmático, núcleo, membrana mitocondrial externa y poro nuclear. Los términos más enriquecidos para función molecular son los constituyentes estructurales del ribosoma, actividad ligada de una aminoacil-ARNt, unión a ADN, inhibidor de actividad enzimática y actividad hidrolasa sobre uniones glicosidicas.

g) Filtro de resultados de expresión diferencial previo al análisis de eQTL:

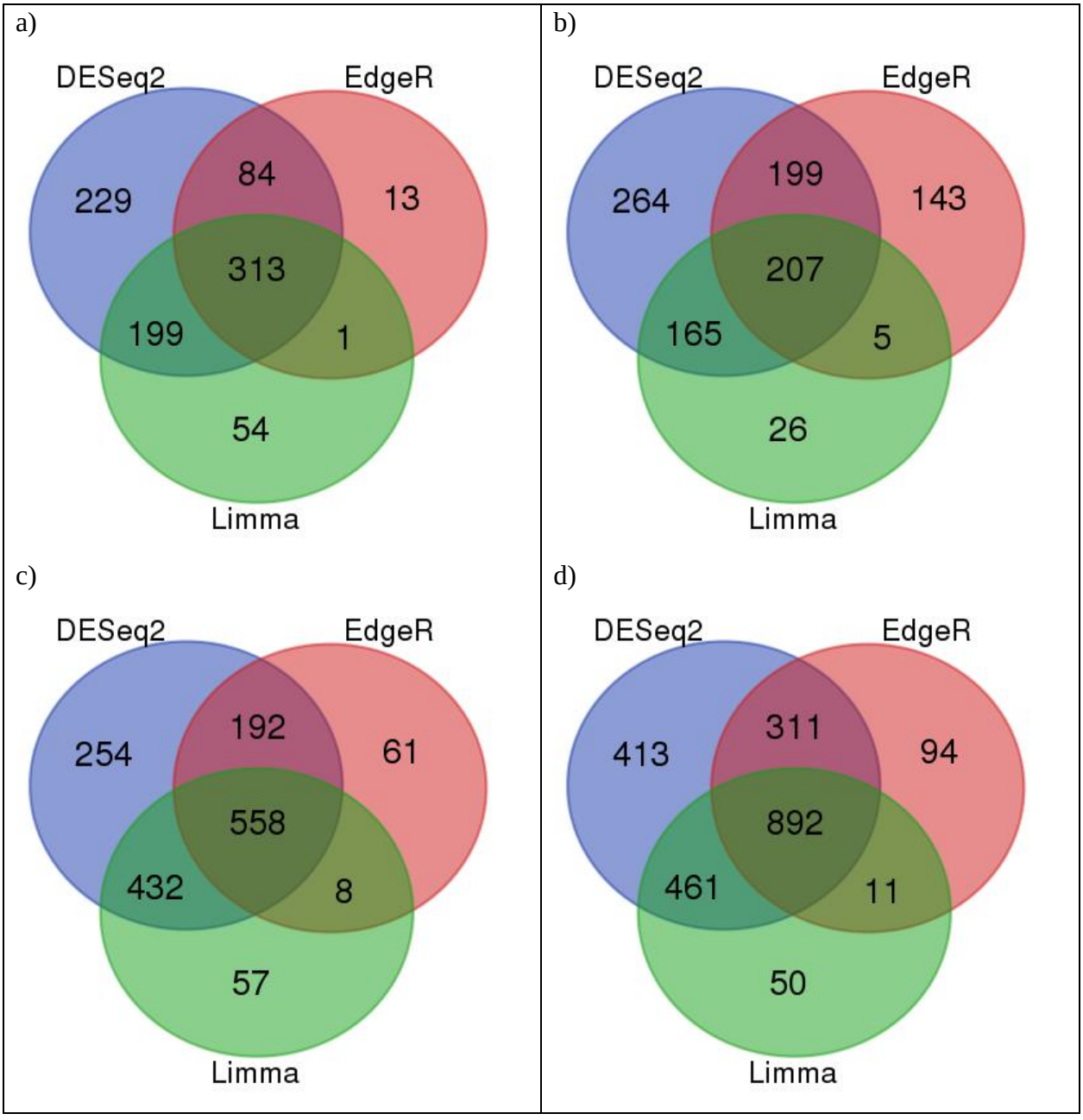


Figura 16: Diagramas de Venn de las comparaciones entre los tres paquetes de R utilizados para la expresión diferencial en cuatro comparaciones distintas siendo a) E1h vs E1j, b) E2h vs E2j, c) E3h vs E3j y d) E4h vs E4j.

Con el propósito de reducir la probabilidad de falsos positivos se utilizaron tres paquetes de R para realizar los análisis de expresión diferencial: DESeq2, edgeR y Limma. Estos difieren en cuanto a los resultados de transcritos expresados diferencialmente respecto a la manera de realizar la normalización de las cuentas crudas y de definir aquellos transcritos que se encuentran diferencialmente expresados de aquellos que no.

Particularmente, DESeq2 utiliza una distribución binomial negativa utilizando la media de la tasa de las cuentas observadas como estimador del tamaño de las cuentas (Anders *et al*. 2010), mientras que edgeR y Limma usan la media truncada de los valores de una tasa logarítmica de expresión en la normalización de las cuentas (normalización TMM) (Robinson y Oshlack 2010).

Estas diferencias indicarían que los transcritos expresados diferencialmente en la intersección de los tres paquetes poseen sin lugar a dudas una diferencia notable en sus niveles de expresión entre las etapas comparadas. Para comprobar esto, se confeccionaron cuatro diagramas de Venn con las comparaciones por estadio entre harinosos y jugosos de E1h vs E1j (Figura 16a), E2h vs E2j (Figura 16b), E3h vs E3j (Figura 16c) y finalmente E4h vs E4j (Figura 16d). De estos diagramas se tomaron los 313 transcritos en común de la comparación E1h vs E1j (Figura 16a), los 207 transcritos en común de la comparación E2h vs E2j (Figura 16b), los 558 transcritos en común de la comparación E3h vs E3j (Figura 16c) y los 892 transcritos en común de la comparación E4h vs E4j (Figura 16c).

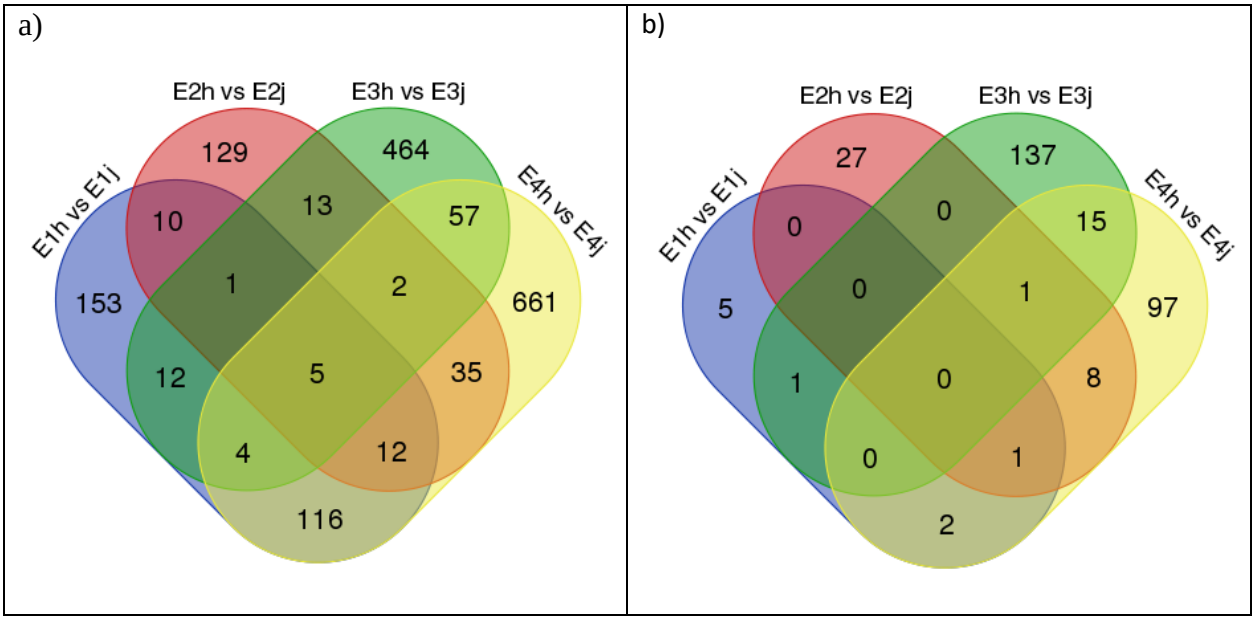


Figura 17: Diagrama de Venn confeccionado con a) los transcritos en común entre todas las comparaciones (E1h vs E1j, E2h vs E2j, E3h vs E3j, E4h vs E4j) y b) que posean veces de cambio menor a -1 y mayor a 1 expresados diferencialmente para cada comparación señalada en la figura. Ambos diagramas de Venn fueron hechos con los datos de los paquetes de R 'DESeq2', 'Limma' y 'edgeR'.

Los transcritos en común obtenidos mediante los diagramas de Venn (Figura 16) fueron utilizados para realizar otra selección en base a la comparación entre todos los periodos de evaluación descritos anteriormente (E1, E2, E3 y E4) (Figura 17a). Mediante este ejercicio se lograron remover aquellos transcritos que son propios del desarrollo del fruto normal y que se encuentran entre E1 y E2, quedándonos entonces con los transcritos únicos y comunes de las etapas E3 y E4 que son las etapas en donde es más probable encontrar los genes responsables del fenotipo harinoso debido a que estos han estado bajo estrés por frío tanto en E3 (período de 30 días a 0 ºC) como en E4 (etapa de madurez de consumo luego de tratemiento con frío) (Figura 17b).

Posterior a esa selección se aplicaron filtros de veces de cambio que eliminan todos aquellos transcritos con valores de veces de cambio entre -1 y 1 (Figura 17b) con la intención de remover aquellos transcritos con variaciones mínimas en sus niveles de expresión.

Finalmente se continuó trabajando con los 249 transcritos únicos y comunes en las comparaciones de E3h vs E3j y E4h vs E4j (Figura 17b).

h) Análisis de eQTL para harinosidad:

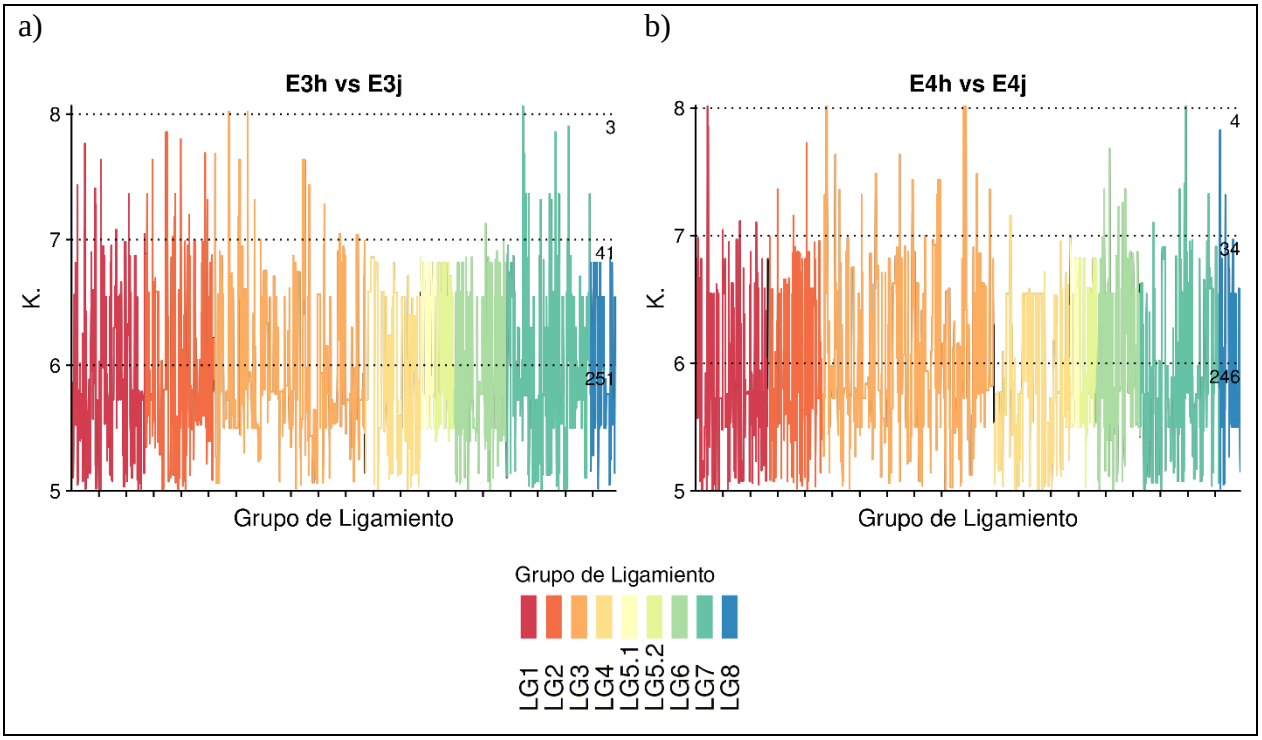


Figura 18: QTL de expresión para las comparaciones de frutos harinosos y jugosos de la etapa 3 (a, E3h vs E3j) y para los frutos harinosos y jugosos de la etapa 4 (b, E4h vs E4j). Los grupos de ligamiento están marcados con color acorde a lo indicado en la escala. Se establecieron tres puntos de corte, a K=6,0; K=7,0; y K=8,0 denotado por las líneas horizontales punteadas en cada gráfico.

Para el análisis de eQTL se utilizaron las cuentas ya normalizadas de todos los transcritos encontrados entre E3 y E4 a partir del diagrama de Venn filtrado (Figura 17b) y se utilizaron dentro del software MapQTL como si fuesen un fenotipo convencional (Figura 18) utilizando la prueba de Kruskal-Wallis. Se establecieron puntos de corte en K=6,0; K=7,0; y K=8,0. Sobre los puntos de corte establecidos se encuentran 251 transcritos para un valor de K=6,0; 41 transcritos para un valor de K=7,0 y tres transcritos para K=8,0 en el caso de la comparación E3h vs E3j (Figura 18a). Para la comparación E4h vs E4j se encontraron 246 transcritos para K=6,0; 34 transcritos para K=7,0 y cuatro transcritos para K=8,0 (Figura 18b). Los resultados se detallan en la Tabla 3.

Tabla 3: Métrica de los eQTL para harinosidad encontrados en las comparaciones E3h vs E3j y E4h vs E4j con un K>=6, indicando la cantidad de transcritos en Cis y Trans encontrados por cada comparación, así como también el detalle por cada uno de los grupos de ligamiento.

	Comparación E3h vs E3j		Comparación E4h vs E4j	
	Cantidad	Porcentaje (%)	Cantidad	Porcentaje (%)
Cis	30	11,9	34	13,9
Trans	221	88,1	212	86,1
Total	251	100,0	246	100,0

	Por Grupo de Ligamiento (GL)			
	Comparación E3h vs E3j		Comparación E4h vs E4j	
	Cis	Trans	Cis	Trans
GL 1	7	32	9	31
GL 2	3	37	2	35
GL 3	8	37	11	40
GL 4	0	19	1	22
GL 5.1	0	11	1	9
GL 5.2	2	12	1	11
GL 6	4	23	5	26
GL 7	3	38	4	27
GL 8	3	12	0	11

i) Identificación de genes candidatos para harinosidad:

Los QTL convencionales (Figura 7), señalan una mayor asociación estadística en los grupos de ligamiento 5.1 y 6 (Figura 8), por lo tanto, se analizaron los eQTL en las regiones identificadas por los QTL convencionales. Se lograron identificar genes candidatos para el fenotipo de harinosidad, los cuales se detallan en la Tabla 4. En total, consta de 22 transcritos y 20 transcritos únicos (es decir, que no se repiten entre E3h vs E3j y E4h vs E4j). Los transcritos comunes entre las dos etapas analizadas son C1 y CHI3 (Tabla 4).

Tabla 4: Genes candidatos. Descripción en Swissprot y tipo de eQTL encontrados sobre un valor de K=6.0 en los QTL de expresión. El tipo de eQTL indica si este está en 'Trans' (el gen se encuentra en un grupo de ligamiento y actúa en uno distinto) o en 'Cis' (el gen se encuentra en un grupo de ligamiento y actúa en el mismo grupo).

Comparación	Código del Gen	Anotación Swissprot	Abreviación Swissprot	Tipo eQTL	Ubicación física	Ubicación acción	GO Asociado	Código Uniprot
E3h vs E3j	Prupe.2G324700	Antocianidina 3-O-glucosiltransferasa 2	FGT	Trans GL 6	GL 2	GL 6	Procesos metabólicos	Q5UL10
	Prupe.1G405400	Proteina C1 reguladora de antocianinas	C1	Trans GL 6	GL 1	GL 6	Unión a cromatina	P10290
	Prupe.2G263900	Probable chalcona-flavona isomerasa 3	CHI3	Trans GL 6	GL 2	GL 6	Actividad liasa intramolecular	Q8VZW3
	Prupe.1G262000	Probable galacturonosiltransferasa 9	GATL9	Trans GL 6	GL 1	GL 6	Actividad transferasa de grupos glicosídicos	O04536
	Prupe.7G070000	Probable poligalacturonasa	GSVIVT00026920001	Trans GL 6	GL 7	GL 6	Actividad poligalacturonasa	A7PZL3
	Prupe.1G249400	Proteina kinasa 2 receptor rico en cisteina	CRK2	Trans GL 6	GL 1	GL 6	Actividad proteina kinasa	Q9LJL9
	Prupe.8G056100	Receptor serina/treonina kinasa rico en repeticiones de leucina	At1g17230	Trans GL 6	GL 8	GL 6	Unión a proteína	Q9SHI2
	Prupe.5G223900	Probable serina/treonina kinasa	At1g18390	Trans GL 6	GL 5	GL 6	Unión a ATP	P0C5E2
	Prupe.1G340500	Factor de transcripcción Myb putativo	At1g14600	Trans GL 5.1	GL 1	GL 5.1	Unión a ADN	Q700D9
	Prupe.3G126500	Proteina de union a GTP tipo Rac	RAC2	Trans GL 6	GL 3	GL 6	Unión a GTP	Q40220
	Prupe.6G229000	Factor de transcripcción	MYB44	Trans GL 5.1	GL 6	GL 5.1	Unión a ADN	Q9FDW1
E4h vs E4j	Prupe.1G405400	Proteina C1 reguladora de antocianinas	C1	Trans GL 6	GL 1	GL 6	Unión a cromatina	P10290
	Prupe.5G203600	Flavonoide 3'-monooxigenasa	CYP75B2	Trans GL 6	GL 5	GL 6	Unión a iones de hierro	Q9SBQ9
	Prupe.2G263900	Probable chalcona-flavona isomerasa 3	CHI3	Trans GL 6	GL 2	GL 6	Actividad liasa intramolecular	Q8VZW3
	Prupe.6G058900	UDP glicosiltransferasa 71D1	UGT71D1	Cis GL 6	GL 6	GL 6	Procesos metabólicos	O82383
	Prupe.1G375700	Jasmonato O-metiltransferasa	JMT	Trans GL 6	GL 1	GL 6	Actividad metiltransferasa	Q9SBK6
	Prupe.1G111900	Giberelina 2-beta-dioxigenasa 2	GA2OX2	Trans GL 5.1	GL 1	GL 5.1	Actividad oxidorreductasa	Q9XHM5
	Prupe.8G201000	Xiloglucano endotransglicosidasa/ proteína hidrolasa 2	XTH2	Trans GL 6	GL 8	GL 6	Hidrolización componentes O-glicosidicos	Q9SV60
	Prupe.1G275900	Factor de transcripción APL de la familia Myb	APL	Trans GL 6	GL 1	GL 6	Unión a ADN	Q9SAK5
	Prupe.4G158000	Probable receptor LRR serina/treonina kinasa	At1g53430	Trans GL 5.1	GL 4	GL 5.1	Actividad proteína kinasa	C0LGG8
	Prupe.6G274800	Probable receptor LRR serina/treonina kinasa	At3g47570	Trans GL 5.1	GL 6	GL 5.1	Unión a proteína	C0LGP4
	Prupe.1G527200	Proteína serina/treonina kinasa	At1g28390	Trans GL 6	GL 1	GL 6	Unión a ATP	Q9SGN7

j) Validación del transcriptoma de la población OxN por medio de RT-PCR:

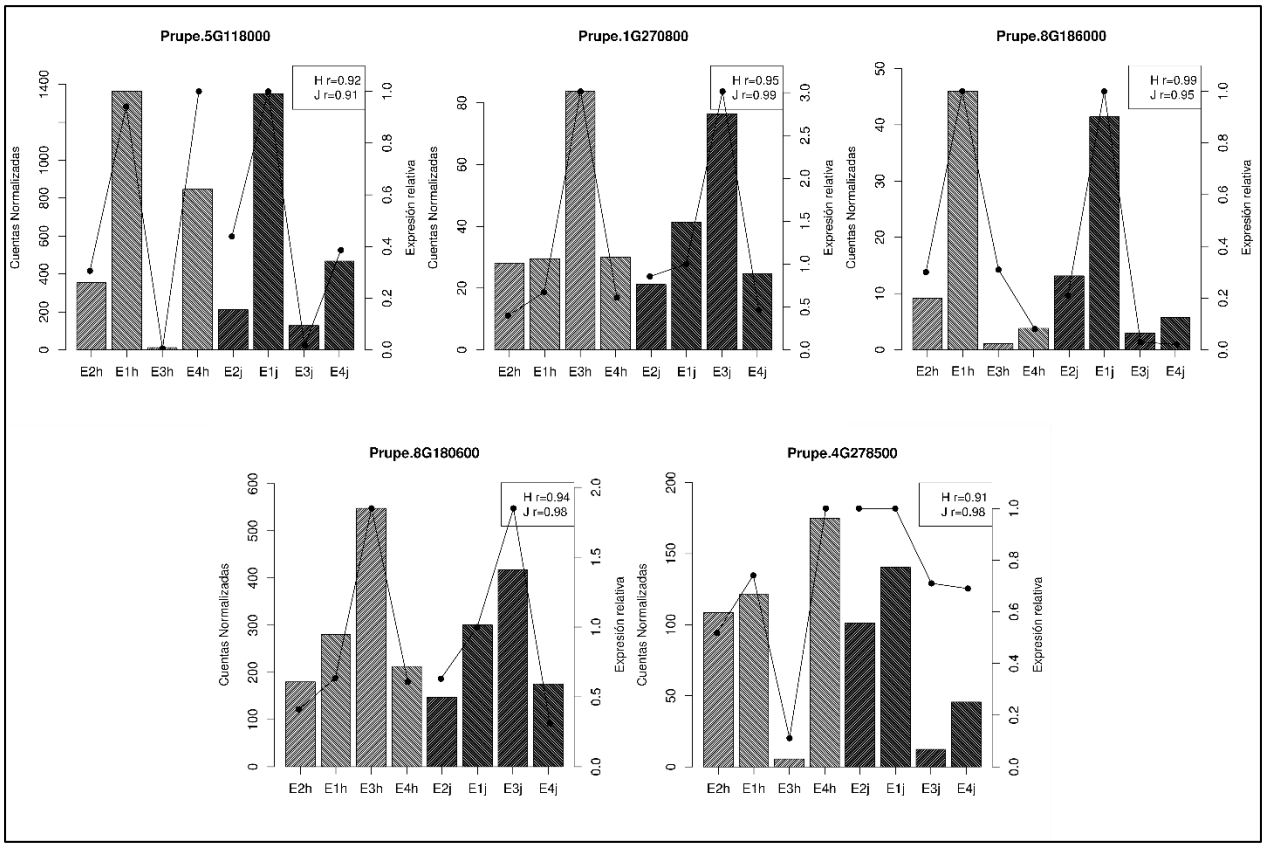


Figura 19: Validación de resultados de secuenciación de ARN por medio de PCR en tiempo real. Se tomaron cinco genes al azar dentro de los genes observados por Sanhueza (Sanhueza *et al*. 2015) y se analizaron para cada una de las cuatro etapas en cada condición (harinoso (h) o jugoso (j)). En los gráficos se comparan con barras ralladas color plomo las cuentas normalizadas para harinosidad y en barras negras las cuentas normalizadas para jugosidad, mientras que las barras plomas claras representan la expresión relativa del qPCR en harinosidad y las barras plomas oscuras representan la expresión relativa en jugosidad. Se calculó una correlación entre los valores de cuentas normalizadas y expresión relativa, el p-value resultante de esta correlación se muestra en la caja superior derecha en cada gráfico.

Como validación de la secuenciación de ARN llevado a cabo en este trabajo, se realizó un PCR en tiempo real con motivo de verificar que las cuentas calculadas se comportasen de manera similar a la expresión relativa calculada por medio del PCR en tiempo real utilizando cinco genes tomados al azar desde una población previa llevada a cabo por el grupo de trabajo (Sanhueza *et al*. 2015): Prupe.5G118000 (GH9C2, glicosil hidrolasa 9C2), Prupe.1G270800 (UTR-3, transportador de UDP-Galactosa), Prupe.8G186000 (UDP-GT, uridin difosfato glicosiltransferasa),

Prupe.8G180600 (TMT2, transportador de monosacáridos del tonoplasto) y Prupe.4G278500 (B-Gal, beta-galactosidasa 10).

En este análisis las tendencias se mantuvieron en cada gen con un valor de correlación superior a 0,90 entre las cuentas normalizadas de la secuenciación de ARN y la expresión relativa del qPCR, lo que implica una asociación directa entre las tendencias de ambos valores con una similitud mayor al 90%.

Discusión:

Población OxN:

La población de OxN utilizada en este trabajo se encuentra en dos localidades: Rinconada (RM) y Choapinos (VI). Además, las poblaciones se encuentran en condiciones de establecimiento y manejo diferentes: En Rinconada se lleva adelante un programa de mejoramiento genético, por esto los segregantes están plantados muy cerca uno de otro (doble seto; 0,4 m x 0,7 m), mientras que Choapinos es un huerto establecido a distancia de plantación comercial (4 m x 2 m). Esto genera que los frutos de esta localidad sean más homogéneos entre sí respecto a los de Rinconada, principalmente porque el efecto posición del fruto (iluminación) es menos relevante en árboles con una mayor canopia distribuida de forma horizontal.

Esto finalmente genera una variación fenotípica muy alta, que sin embargo permite encontrar información a través de análisis genómicos y transcriptómicos, identificando un factor genético significativo.

La harinosidad se define específicamente como un ablandamiento incorrecto de la pared celular durante la maduración del fruto y un tejido disgregado que genera una sensación de falta o ausencia de jugosidad (Barreiro *et al*. 2000; Valero *et al*. 2005). Este ablandamiento incorrecto se atribuye principalmente al efecto del almacenamiento en frío, debido al problema de daño por frío (Lurie y Crisosto 2005).

Mapa genético de la población OxN y análisis de QTL convencionales:

El mapa genético de la población de OxN utilizada en este trabajo consiste en nueve grupos de ligamiento generados a partir de marcadores consenso para cada población por separado. Estos grupos de ligamiento son correspondientes a los ocho cromosomas de *Prunus persica,* sin embargo, en el grupo de ligamiento 5 no existen suficientes marcadores heterocigotos, por lo que no se pudo generar un grupo de ligamiento único. En lugar de ello se denotan como LG5.1 y LG5.2 (Figura 6). La falta de marcadores también genera una baja densidad del mapa, siendo esta de 1,44 cM por marcador, a diferencia de mapas genéticos de *Prunus persica* anteriores que poseen por ejemplo 0,81 cM/marcador (Martínez-García *et al*. 2012) con 588 marcadores SNP y 0,21 cM/marcador (Nuñez-Lillo *et al*. 2015) con 1.948 marcadores SNP.

Por medio de la asociación entre marcadores y porcentaje de jugosidad fue posible encontrar QTL convencionales en los grupos de ligamiento 5.1 y 6. Distintas temporadas de evaluación fenotípica resaltan los dos grupos de ligamiento ya mencionados (Figura 8; Figura 9).

En distintas publicaciones se han encontrado genes asociados a harinosidad principalmente en los grupos de ligamiento 4 y 5 (Peace *et al*. 2006; Cantín *et al*. 2010; Martínez-García *et al*. 2012; Nuñez-Lillo *et al*. 2015) y algunas en los grupos 1, 2, 4, 5 y 6 (Ogundiwin *et al*. 2007). Sin embargo, no se observaron QTL convencionales en el grupo de ligamiento 4, esto puede deberse a la baja densidad del mapa o a la falta de marcadores, ya que finalmente es debido a estos que es posible asociar información fenotípica a un mapa genético.

Fenotipificación de la población:

Es posible confirmar el comportamiento contrastante de la población de OxN al analizar la fenotipificación ya que esta muestra que los datos medidos (Figura 5) poseen un comportamiento unimodal, tal como publicaciones anteriores (Campos-Vargas *et al*. 2006; Cantín *et al*. 2010; Martínez-García *et al*. 2012; Nuñez-Lillo *et al*. 2015), con un promedio poblacional que coincide para todos los datos en la cima de la distribución normal.

En los frutos fenotipificados de la población OxN se observa que cerca de la mitad de ellos son considerados ‘jugosos’ y la otra mitad ‘harinosos’, teniendo en cuenta que el parámetro de corte para esta característica es un porcentaje de jugosidad del 30% (Infante *et al*. 2009; Figura 5a).

Análisis de componentes principales y mapas de calor:

Específicamente, los frutos seleccionados para la secuenciación de su material genético presentan valores de jugosidad categoricamente distintos, promediando un 19% de jugosidad para todos los frutos harinosos y un 37% de jugosidad para todos los frutos jugosos (Tabla 2). Esto confirma el carácter contrastante de la población OxN respecto a harinosidad considerando los frutos seleccionados para su secuenciación y también la diferencia al observar el análisis de componentes principales (Figura 13), en donde los frutos forman dos grupos claramente separados para la comparación de E3h vs E3j, explicado por el PC2 con 3,4% (Figura 13a). En caso de la comparación de E4h vs E4j existe un único dato de jugosidad que se acerca hacia el grupo de condición harinosa, sin embargo, a pesar de su cercanía, este dato sigue perteneciendo al grupo de los jugosos ubicado en la zona inferior de la figura y que se explica por el PC2 de 2.7% (Figura 13b). Conjuntamente, los mapas de calor mostraron niveles de expresión con distintos patrones entre frutos jugosos y harinosos en E3 (Figura 14a) y también E4 (Figura 14b).

Análisis de ontología génica:

Con los transcritos expresados diferencialmente entre la comparación E3h vs E4h se realizó un análisis de ontología génica por medio del paquete de R GOSeq, el cual reveló que algunas de las categorías predominantes son el metabolismo de sacarosa (como proceso biológico), región extracelular (como componente celular) y actividad hidrolasa sobre enlaces glicosídicos (como función molecular) (Figura 15).

Respecto a esto, al modificar o alterar la pared celular por efectos bióticos, abióticos o como parte del desarrollo, es posible que se genere una señalización en forma de oligogalacturónidos derivados de pectina en caso de modificación de pared por agentes bióticos (Galletti *et al*. 2009) o fragmentos de carbohidratos de la pared celular u oligosacarinas en respuesta a modificaciones generales de la pared (Albersheim *et al*. 1983; Fry *et al*. 1990). Estas señales se podrían transducir de manera de, eventualmente, generar una respuesta compensatoria a nivel transcripcional y postraduccional para corregir estas alteraciones (Seifert y Blaukopf 2010). Esto ocurre específicamente en la matriz extracelular (o región extracelular) de la célula vegetal por kinasas tipo receptor o RLKs (del inglés Receptor-like Kinases) (Shiu y Bleecker 2001; Shiu y Bleecker 2003). Existe también señalización por azúcares como galactosa y sacarosa, el cual es independiente de su rol en el metabolismo (Hanson y Smeekens 2009). Esto podría estar explicando

por qué el metabolismo de sacarosa es uno de los principales resultados de GOSeq. Además, la sacarosa es el principal azúcar de los duraznos y nectarinas (Desnoues *et al*. 2014). Esta azúcar tiene funciones estructurales por medio de su hidrolisis en glucosa y fructosa (Vidal *et al*. 2001), energéticas (Nonis *et al*. 2007), en desarrollo (Lo-Bianco *et al*. 1999), etc.

Análisis de QTL de expresión:

Para reducir los falsos positivos dentro de los transcritos expresados diferencialmente se realizaron comparaciones de los transcritos expresados diferencialmente por etapa entre la condición harinoso contra jugoso obtenido mediante DESeq2, edgeR y Limma. Esto se detalla en la Figura 16 en donde sólo se consideraron aquellos transcritos expresados diferencialmente que fuesen comunes entre los 3 paquetes de R mencionados.

DESeq2 utiliza una estimación reducida de la dispersión de las veces de cambio y la dispersión de las cuentas dentro del método de Bayes (inferencia bayesiana) para estimar los transcritos expresados diferencialmente (Love *et al*. 2014). Por su parte, edgeR modera la dispersión de la expresión de cada gen respecto a un estimado común a través de todos los transcritos o utilizando una condición de peso o relevancia en un estimado local de transcritos con expresión similar (Robinson y Smyth 2007; McCarthy *et al*. 2012). Limma utiliza la función *voom()* dentro de R para transformar las cuentas normalizadas logarítmicas y estima la relación entre la media y la varianza para generar un modelamiento lineal (Seyednasrollah *et al*. 2013).

Con estos transcritos comunes entre los tres paquetes de R: DESeq2, edgeR y Limma se realizó una comparación entre todas las etapas y se eligieron sólo los transcritos de las etapas E3h vs E3j y E4h vs E4j (Figura 17). Esto nos permitió corroborar los transcritos expresados diferencialmente entre los tres paquetes en conjunto y además eliminar aquellos transcritos responsables del desarrollo común del fruto que se encuentran entre las etapas E1, correspondiente a cosecha y E2, correspondiente a madurez de consumo sin que el fruto haya pasado por almacenamiento en frío. De esto se obtuvieron 249 transcritos en total, 137 transcritos para E3h vs E3j, 97 para E4h vs E4j y 15 transcritos en común entre ambas comparaciones (Figura 17b).

A partir de estos 249 transcritos se realizó el análisis de eQTL, para lo cual se obtuvieron las cuentas normalizadas de cada uno de ellos en las comparaciones E3h vs E3j y E4h vs E4j. Las cuentas fueron asociadas a los marcadores del mapa de ligamiento dentro del software MapQTL

utilizando la prueba de Kluskal-Wallis (Figura 18). Para la comparación E3h vs E3j se encontró que el 11,95% de los transcritos eran tipo Cis y el resto (88,05%) eran tipo Trans. Una tendencia similar se observó en la comparación E4h vs E4j donde 13,82% eran Cis y 86,18% eran Trans. Estos datos son similares a resultados obtenidos anteriormente en el grupo de trabajo en *Prunus persica*, donde un 14,54% fueron Cis y 85,45% fueron Trans (Carrasco 2016).

Se buscaron genes candidatos específicamente en los grupos de ligamiento 5.1 y 6, los cuales habían sido previamente identificados por medio de QTL convencionales como los grupos de ligamiento que poseen una mayor probabilidad de albergar los loci que son responsables del fenotipo harinoso (Figura 8).

La información de los genes candidatos encontrados se detalla en la Tabla 4.

Genes candidatos involucrados en desarrollo de fenotipo de harinosidad:

En los QTL de expresión en estas regiones específicas es posible observar principalmente transcritos asociados a estrés abiótico, generación de pigmentos antioxidantes (antocianinas), respuesta a estímulos (entiéndase cascadas de señalización o receptores y transductores de señales) y en menor medida transcritos asociados a modificación estructural de la pared celular y fitohormonas (Tabla 4).

- **Genes candidatos asociados a biosíntesis de antocianinas:**

Las antocianinas son pigmentos predominantes en frutos negros, rojos y morados, principalmente en bayas. Su acumulación es normal durante el desarrollo, sin embargo se ha encontrado evidencia de que estos pigmentos están asociados a actividades defensivas antioxidantes en plantas (Miller *et al*. 2007; İşeri *et al*. 2015). Además, uno de los fenotipos característicos del daño por frío en *Prunus persica* es el sangrado, que se define como coloración rojiza del mesocarpo (Lurie y Crisosto 2005). Esta coloración rojiza es generada por pigmentos de antocianinas, especialmente por genes como la chalcona sintasa (Jiao *et al*. 2014). En el trabajo presente no hubo fenotipificación de este carácter.

Los transcritos asociados a biosíntesis de antocianinas aumentan su presencia en frutos harinosos a partir de su almacenamiento a bajas temperaturas, teniendo su punto mayor una vez el fruto ya ha madurado para su consumo luego del almacenamiento. Estos transcritos se asocian a

proteínas reguladoras de antocianinas que controlan la expresión de genes involucrados en biosíntesis de antocianinas, los cuales poseen un dominio de unión a Myb (código Uniprot: P10290); antocianidina 3-O-glucosiltransferasa, que en presencia de otros factores de color necesarios genera una glicosilación que permite la acumulación de pigmentos de antocianinas (código Uniprot: Q5UL10), esta es la enzima inicial de la vía de biosíntesis de antocianinas a partir de pelargonidina, cianidina y delfinidina. La chalcona-flavonola isomerasa (código Uniprot: Q8VZW3) forma parte de la biosíntesis de flavonoides en al menos cuatro reacciones enzimáticas de esa vía. Está presente también una flavonoide-3'-monooxigenasa (código Uniprot: Q9SBQ9), la cual está involucrada en biosíntesis de flavonoides en al menos seis procesos enzimáticos distintos dentro de esta vía y al menos dos procesos enzimáticos en la vía de biosíntesis de flavonas y flavonolas. Finalmente, la UDP-glicosiltransferasa (código Uniprot: O82383) regula el metabolismo de dos productos de síntesis de flavonoides, el kaempfenol y la queracitina, dentro de la vía de biosíntesis de flavonas y flavonolas.

Todas estas proteínas y/o reguladores se encuentran con veces de cambio que van en aumento desde E3h hasta E4h en comparación con sus contrapartes jugosas y que además está diferencialmente expresada en frutos harinosos. Cabe mencionar que a partir de la síntesis de flavonoides se generan los sustratos para la síntesis de antocianinas y de flavonas y flavonolas.

Además, se encontraron cuatro transcritos asociados a la familia de factores de transcripción Myb, asociados a la síntesis y regulación de flavonoides (Ravaglia *et al*. 2013). Estas son las proteínas reguladoras C1 (código Uniprot: P10290), factor de transcripción Myb (códigos Uniprot: Q700D9, Q9SAK5) y factor de transcripción Myb44 (código Uniprot: Q9FDW1).

- **Genes candidatos asociados a respuesta a estímulos:**

Respecto a la respuesta a estímulos, dentro de los transcritos encontrados por medio de eQTL se encontraron muchos receptores o transductores de membrana sin mayor descripción alguna, tales como proteínas serina/treonina kinasa (código Uniprot: Q9SGN7) y receptores ricos en leucina (códigos Uniprot: C0LGG8, C0LGP4, Q9SHI2). También receptores involucrados en transducción de señales en rutas que usan calcio como segundo mensajero (código Uniprot: Q9LJL9), inactivación de GTPasas (código Uniprot: Q40220), regulador de expresión del factor de transcripción NAC045 (código Uniprot: Q9SAK5) y factores de respuesta a ABA (códigos Uniprot: Q9FDW1, P0C5E2). Los receptores ricos en leucina, proteinas serina/treonina kinasa han

sido identificados dentro de la base de datos “ChillPeach” como genes que controlan o están involucrados en el desarrollo del daño por frío en *Prunus persica* (Ogundiwin *et al*. 2008).

Cabe destacar que los factores de transcripción NAC están asociados, entre otras cosas, a fecha de cosecha, maduración, estrés biótico y abiótico y vías asociadas a la fitohormona etileno. Esto indica que es altamente probable que estén asociados a susceptibilidad a harinosidad debido al rol que tienen todos estos factores respecto a ablandamiento, especialmente el rol de etileno (Zhou *et al*. 2001; Hayama *et al*. 2006).

- **Genes candidatos asociados a remodelación de pared celular:**

Por otro lado, los genes asociados a remodelación o modificación de pared celular se encuentran con una expresión mayor en frutos harinosos para el caso de la galacturonosiltransferasa, la cual está involucrada en biosíntesis de pectinas y/o xilanos de la pared celular (código Uniprot: O04536). La poligalacturonasa (código Uniprot: A7PZL3) sin embargo se encuentra reprimida en la etapa E3 de frutos harinosos, pero tiene tendencia a un aumento en su expresión en la etapa E4 de frutos harinosos. Se ha observado una disminución en la actividad enzimática de la poligalacturonasa en frutos harinosos, indicando que el frío genera una disminución en enzimas reguladas por etileno, las cuales son requeridas para el ablandamiento normal del fruto (Brummell *et al*. 2004).

La xiloglucano endotransglicosidasa (código Uniprot: Q9SV60), en cambio se ve con una expresión reprimida y disminuyendo entre estas etapas. La acción de esta enzima se ha asociado a ablandamiento durante la madurez de consumo del fruto por medio de la despolimerización de los xiloglucanos (Brummell y Harpster 2001).

- **Genes candidatos asociados a fitohormonas:**

Respecto a las fitohormonas Jin en el 2009 (Jin *et al*. 2009) utilizó una combinación de aire caliente y metiljasmonato en frutos en fase de madurez de consumo. Estos mostraron menor índice de harinosidad, explicado en parte por el aire caliente, que causa la síntesis de proteínas de choque térmico, mientras que el metiljasmonato suscita mecanismos de defensa contra daño por frío a través de metabolitos secundarios, niveles de transcritos de proteínas de choque térmico y genes relacionados a patogénesis. Además, promueve la actividad poligalacturonasa y pectin metilesterasa, causando que la célula genere un ablandamiento correcto de su pared.

Dentro de los transcritos analizados se encuentra una jamonato O-metiltransferasa (código Uniprot: Q9SBK6), la cual cataliza la metilación de jasmonato a metiljasmonato, que es la forma biológicamente activa de esta hormona. Se observó un aumento en su expresión entre las etapas E3 y E4 de frutos harinosos, sin embargo, no se observaron transcritos de proteínas de choque térmico expresadas diferencialmente respecto a frutos normales ni transcritos asociados a la enzima pectin metilesterasa. Por otra parte, los niveles de transcritos de poligalacturonasa en frutos harinosos comparados con jugosos aumentan levemente desde E3 a E4, pero aun manteniéndose reprimidos.

Se encontró también una giberelina 2-beta-dioxigenasa, la cual genera la hidroxilación de giberelinas biológicamente activas, generando una regulación homeostática de sus niveles endógenos (código Uniprot: Q9XHM5). Su expresión se encuentra reprimida y disminuyendo entre E3h y E4h respecto a su contraparte jugosa. En un estudio del año 2010 se encontró que en frutos con aplicación temprana de giberelinas (GA3) existía una prevención de harinosidad luego de su almacenamiento en frío. Aquellos frutos sin aplicación de esta hormona tuvieron menores cantidades de ARN mensajeros de genes asociados con el ablandamiento de pared y síntesis de etileno (Pegoraro *et al*. 2010).

En base a los datos recabados y la bibliografía existente al respecto es posible discernir que la harinosidad en *Prunus persica* es generada a partir de estrés por frío (daño por frío). Esto genera una desregulación en el metabolismo, impidiendo que el fruto se defienda eficientemente del estrés durante la etapa de ablandamiento para madurez de consumo y generando un ablandamiento incorrecto.

Esto tiene soporte en base a los transcritos encontrados tanto en este trabajo como en bibliografía, en donde al otorgarle al fruto hormonas como jasmonato o giberelina durante el estrés por frío o en etapas tempranas de desarrollo del fruto este tiene mejor desempeño en cuanto a generación de harinosidad. Por otra parte, en este y otros estudios se ha observado niveles disminuidos en transcritos y niveles enzimáticos de la enzima poligalacturonasa, la cual cataliza la hidrólisis de los enlaces glicosídicos de las cadenas de los homogalacturonanos en la pared celular, generando ablandamiento. Al no poderse efectuar un ablandamiento normal, reclutar las enzimas necesarias o producirse concentraciones suficientes de hormonas o metabolitos secundarios se estaría generando la harinosidad.

En base a la bibliografía existente y teniendo en cuenta que la harinosidad es generada a partir de estrés por frío durante la etapa de ablandamiento para madurez de consumo del fruto, es posible hipotetizar en base a los resultados que este estrés causa una respuesta celular deficiente respecto al reclutamiento de enzimas o acumulación de hormonas o metabolitos. Esto genera una respuesta pobre frente al estrés que culmina en el desarrollo del fenotipo harinoso en *Prunus persica*.

Conclusión:

Se identificaron QTL convencionales para harinosidad en los grupos de ligamiento 5.1 y 6 utilizando tres temporadas de evaluación provenientes de dos localidades distintas: Rinconada (RM) y Choapinos (región de O'Higgins).

Dentro de los QTL de expresión, en los grupos de ligamiento mencionados anteriormente, se identificaron 20 transcritos candidatos únicos (entre las comparaciones de E3h vs E3j y E4h vs E4j) que se correlacionan con el fenotipo de la harinosidad en *Prunus persica*. Entre estos, hay transcritos asociados a producción de antocianinas, respuesta a estímulos, modificación de pared celular y fitohormonas.

Referencias:

Albersheim, Darvill, McNeil, Valent, Sharp, Nothnagel, Davis, *et al.* (1983). Oligosaccharins: Naturally Occurring Carbohydrates with Biological Regulatory Functions. En *Structure and Function of Plant Genomes*, 293–312. Boston, MA: Springer New York.

Anders, Huber, Nagalakshmi, Wang, Waern, Shou, Raha, *et al*. (2010). Differential expression analysis for sequence count data. *Genome Biology*, 1-12.

Anders, Pyl, Huber (2015). HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics*, 166–169.

Arús, Verde, Sosinski, Zhebentyayeva, Abbott (2012). The peach genome. *Tree Genetics and Genomes*, 531–547.

Balzarini, Rienzo (2014). InfoGen. Córdoba, Argentina. https://www.info-gen.com.ar.

Barreiro, Ortiz, Ruiz-Altisent, Ruiz-Cabello, Fernández-Valle, Recasens, Asensio (2000). Mealiness assessment in apples and peaches using MRI techniques. *Magnetic Resonance Imaging*, 1175–1181.

Batjer, Westwood (1958). Size of Elberta and J.H. Hale peaches during the thinning period as related to size at harvest. *Proceedings of the American Society for Horticultural Science*, 101–105.

Bradley, Muriel (1959). Mean cell size in the mesocarp of mature peaches of different sizes. *Proceedings of the American Society for Horticultural Science*, 120–124.

Brummell, Cin, Lurie, Crisosto, Labavitch (2004). Cell wall metabolism during the development of chilling injury in cold-stored peach fruit: Association of mealiness with arrested disassembly of cell wall pectins. *Journal of Experimental Botany*, 2041–2052.

Brummell, Harpster (2001). Cell wall metabolism in fruit softening and quality and its manipulation in transgenic plants. *Plant Molecular Biology*, 311–340.

Campos-Vargas, Becerra, Baeza-Yates, Cambiazo, Gonzáez, Meisel, Orellana, Retamales, Silva, Defilippi (2006). Seasonal variation in the development of chilling injury in 'O Henry' peaches. *Scientia Horticulturae*, 79–83.

Cantín, Crisosto, Ogundiwin, Gradziel, Torrents, Moreno, Gogorcena (2010). Chilling injury susceptibility in an intra-specific peach [Prunus persica (L.) Batsch] progeny. *Postharvest Biology and Technology*, 79–87.

Carpita, Gibeaut (1993). Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. *The Plant journal : for cell and molecular biology*, 1–30.

Carrasco (2016). Identificación de genes candidatos asociados a la pérdida de firmeza del fruto en Prunus persica (L.) Batsch utilizando QTL y QTL de expresión (eQTL). Tesis de Magister, Facultad de Ciencias Biológicas, Universidad Andrés Bello, Chile.

Chaparro, Werner, O'Malley, Sederoff (1994). Targeted mapping and linkage analysis of morphological isozyme, and RAPD markers in peach. *Theoretical and Applied Genetics*, 805-815.

Collard, Jahufer, Brouwer, Pang (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica*, 169–196.

Costa, Bonora, Fiori, Noferini (2011). Innovative non-destructive device for fruit quality assessment. En *VII International Symposium on Kiwifruit*, 575–581. International Society for Horticultural Science - ISHS.

Crisosto, Johnson, Luza, Crisosto (1994). Irrigation Regimes Affect Fruit Soluble Solids Concentration and Rate of Water Loss of ' O ' Henry ' Peaches. *Horticultural Science*, 1169–1171.

Crisosto, Mitchell, Ju (1999). Susceptibility to chilling injury of peach, nectarine, and plum cultivars grown in California. *Horticultural Science*, 1116–1118.

Crisosto, Valero (2008). Harvesting and postharvest handling of peaches for the fresh market. En *The peach: botany, production and uses*, 575–596. Wallingford: CABI.

Daane, Johnson, Michailides, Crisosto, Dlott, Ramirez, Yokota, Morgan (2008). Excess nitrogen raises nectarine susceptibility to disease and insects. *California Agriculture*, 13–18.

Dagar, Puig, Ibanez, Ziliotto, Bonghi, Crisosto, Friedman, Lurie, Granell (2013). Comparative transcript profiling of a peach and its nectarine mutant at harvest reveals differences in gene expression related to storability. *Tree Genetics and Genomes*, 223–235.

Desnoues, Gibon, Baldazzi, Signoret, Génard, Quilot-Turion (2014). Profiling sugar metabolism during fruit development in a peach progeny with different fructose-to-glucose ratios. *BioMed Central Plant Biology*, 1-13.

Dirlewanger, Cosson, Boudehri, Renaud, Capdeville, Tauzin, Laigret, Moing (2006). Development of a second-generation genetic linkage map for peach [Prunus persica (L.) Batsch] and characterization of morphological traits affecting flower and fruit. *Tree Genetics & Genomes*, 1–13.

Doligez, Adam-Blondon, Cipriani, Gaspero, Laucou, Merdinoglu, Meredith, Riaz, Roux, This (2006). An integrated SSR map of grapevine based on five mapping populations. *Theoretical and Applied Genetics*, 369–382.

Fournier-Level, Cunff, Gomez, Doligez, Ageorges, Roux, Bertrand, Souquet, Cheynier, This (2009). Quantitative genetic bases of anthocyanin variation in grape (Vitis vinifera L. ssp. sativa) berry: A quantitative trait locus to quantitative trait nucleotide integrated study. *Genetics*, 1127–1139.

Fry, McDougall, Lorences, Biggs, Smith (1990). Oligosaccharins from xyloglucan and cellulose: modulators of the action of auxin and H+ on plant growth. *Symposia of the Society for Experimental Biology*, 285–298.

Galletti, De-Lorenzo, Ferrari (2009). Host-derived signals activate plant innate immunity. *Plant Signaling and Behavior*, 33–34.

Goldhammer, Salinas, Crisosto, Day, Soler, Moriana (2002). Effects of regulated deficit irrigation and partial root zone drying on late harvest peach tree perfomance. En *V International Peach Symposium*, 343–350. International Society for Horticultural Science - ISHS.

Gonzalez-Aguero, Pavez, Ibanez, Pacheco, Campos-Vargas, Meisel, Orellana, *et al*. (2008). Identification of woolliness response genes in peach fruit after post-harvest treatments.

Journal of Experimental Botany, 1973–1986.

Hadfield, Bennett (1998). Polygalacturonases: Many Genes in Search of a Function. *Plant Physiology*, 337–343.

Hanson, Smeekens (2009). Sugar perception and signaling - an update. *Current Opinion in Plant Biology*, 562–567.

Hayama, Shimada, Fujii, Ito, Kashimura (2006). Ethylene-regulation of fruit softening and softening-related genes in peach. *Journal of Experimental Botany*, 4071–4077.

Horgan, Kenny (2011). 'Omic' technologies: genomics, transcriptomics, proteomics and metabolomics. *The Obstetrician & Gynaecologist*, 189–195.

Huang, Bertrand, Guiraud, Vialet, Launay, Cheynier, Terrier, This (2013). Expression QTL mapping in grapevine-Revisiting the genetic determinism of grape skin colour. *Plant Science*, 18–24.

Infante, Meneses, Rubio, Seibert (2009). Quantitative determination of flesh mealiness in peach [Prunus persica L. (Batch.)] through paper absorption of free juice. *Postharvest Biology and Technology*, 118–121.

İşeri, Körpe, Sahin, Haberal (2015). High salt induced oxidative damage and antioxidant response in tomato grafted on tobacco. *Chilean Journal of Agricultural Research,* 192-201.

Jiao, Shen, Yan, Yu (2014). Gene regulation of anthocyanin biosynthesis in two blood-flesh peach (Prunus persica (L .) Batsch) cultivars during. *Journal of Zhejiang University. Science*, 809–819.

Jin, Zheng, Tang, Rui, Wang (2009). A combination of hot air and methyl jasmonate vapor treatment alleviates chilling injury of peach fruit. *Postharvest Biology and Technology*, 24–29.

Kobayashi, Ishimaru, Hiraoka, Honda (2002). Myb-related genes of the Kyoho grape (Vitis labruscana) regulate anthocyanin biosynthesis. *Planta*, 924–933.

Korpelainen, Tuimala, Somervuo, Huss, Wong (2015). RNA-seq Data Analysis A Practical Approach. Oxfordshire, UK: CRC Press.

Law, Chen, Shi, Smyth (2014). voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology*, 1-17.

Li, Stram, Chen, Kar, Gayther, Pharoah, Haiman, Stranger, Kraft, Freedman (2014). Expression QTL-based analyses reveal candidate causal genes and loci across five tumor types. *Human molecular genetics*, 5294–5302.

Lo-Bianco, Rieger, Sung (1999). Carbohydrate metabolism of vegetative and reproductive sinks in the late-maturing peach cultivar 'Encore'. *Tree physiology*, 103–109.

Love, Huber, Anders (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, 1-21.

Lurie (2007). Controlled atmosphere storage to decrease physiological disorders in nectarines. *International Journal of Food Science & Technology* 507–514.

Lurie, Crisosto (2005). Chilling injury in peach and nectarine. *Postharvest Biology and Technology*, 195–208.

Lyons (1973). Chilling Injury in Plants. *Annual Review of Plant Physiology*, 445–466.

Majewski, Pastinen (2011). The study of eQTL variations by RNA-seq: From SNPs to phenotypes. *Trends in Genetics*, 72–79.

Martínez-García, Parfitt, Ogundiwin, Fass, Chan, Ahmad, Lurie, Dandekar, Gradziel, Crisosto (2012). High density SNP mapping and QTL analysis for fruit quality characteristics in peach (Prunus persica L.). *Tree Genetics & Genomes*, 19–36.

Mathews, Sabina, Zuker, Turner (1999). Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *Journal of molecular biology*, 911–940.

McCann, Roberts (1992). Architecture of the primary cell wall. En *The Cytoskeletal Basis of Plant Growth and Form*, 109–129.

McCarthy, Chen, Smyth (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research*, 4288–4297.

Michaelson, Loguercio, Beyer (2009). Detection and interpretation of expression quantitative

trait loci (eQTL). *Methods*, 265–276.

Miller, Suzuki, Rizhsky, Hegie, Koussevitzky, Mittler (2007). Double mutants deficient in cytosolic and thylakoid ascorbate peroxidase reveal a complex mode of interaction between reactive oxygen species, plant development, and response to abiotic stresses. *Plant physiology*, 1777–1785.

LaRue, Johnson (1989). Peaches, plums and nectarines: Growing and handling for fresh market. En Factors affecting deteriorating rate, 165-178. California, USA: University of California.

Mutz, Heilkenbrinker, Lönne, Walter, Stahl (2012). Transcriptome analysis using next-generation sequencing. *Current Opinion in Biotechnology*, 22–30.

Nica, Dermitzakis (2013). Expression quantitative trait loci: present and future. *Philosophical Transactions of the Royal Society*, 1-6.

Nonis, Ruperti, Falchi, Casatta, Enferadi Thamasebi, Vizzotto (2007). Differential expression and regulation of a neutral invertase encoding gene from peach (Prunus persica): Evidence for a role in fruit development. *Physiologia Plantarum*, 436–446.

Nuñez-Lillo, Cifuentes-Esquivel, Troggio, Micheletti, Infante, Campos-Vargas, Orellana, Blanco-Herrera, Meneses (2015). Identification of candidate genes associated with mealiness and maturity date in peach [Prunus persica (L.) Batsch] using QTL analysis and deep sequencing. *Tree Genetics & Genomes*, 1-13.

Obenland, Crisosto, Rose (2003). Expansin protein levels decline with the development of mealiness in peaches. *Postharvest Biology and Technology*, 11–18.

Ogundiwin, Martí, Forment, Pons, Granell, Gradziel, Peace, Crisosto (2008). Development of ChillPeach genomic tools and identification of cold-responsive genes in peach fruit. *Plant Molecular Biology*, 379–397.

Ogundiwin, Peace, Gradziel, Dandekar, Bliss, Crisosto (2007). Molecular Genetic Dissection of Chilling Injury in Peach Fruit. En *International Symposium on Biotechnology of Temperate Fruit Crops and Tropical Species*, 633–638. International Society for Horticultural Science - ISHS.

Paterson (1996). Plant Genome Mapping: Strategies and Applications. *Biotechnology*, 23–39.

Payasi, Mishra, Chaves, Singh (2009). Biochemistry of fruit softening: An overview. *Physiology and Molecular Biology of Plants*, 103–113.

Peace, Crisosto, Garner, Dandekar, Gradziel, Bliss (2006). Genetic controls of Internal Breakdown in Peach. En *VI International Peach Symposium*, 489–496. International Society for Horticultural Science - ISHS.

Peace, Crisosto, Gradziel (2005). Endopolygalacturonase: A candidate gene for Freestone and Melting Flesh in peach. *Molecular Breeding*, 21–31.

Pegoraro, Zanuzo, Chaves, Brackmann, Girardi, Lucchetta, Nora, Silva, Rombaldi (2010). Physiological and molecular changes associated with prevention of woolliness in peach following pre-harvest application of gibberellic acid. *Postharvest Biology and Technology*, 19–26.

R Development Core Team (2015). R: A Programming Environment for Data Analysis and Graphics. R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org/.

Ravaglia, Espley, Henry-Kirk, Andreotti, Ziosi, Hellens, Costa, Allan (2013). Transcriptional regulation of flavonoid biosynthesis in nectarine (Prunus persica) by a set of R2R3 MYB transcription factors. *BioMed Central Plant Biology*, 1-14.

Robinson, McCarthy, Smyth (2009). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 139–140.

Robinson, Oshlack (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology*, 1-9.

Robinson, Smyth (2007). Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics*, 2881–2887.

Sanhueza, Vizoso, Balic, Campos-Vargas, Meneses (2015). Transcriptomic analysis of fruit stored under cold conditions using controlled atmosphere in Prunus persica cv. “Red Pearl”. *Frontiers in Plant Science*, 1–12.

Scorza, May, Purnell, Upchurch (1991). Differences in number and area of mesocarp cells between small- and large-fruited peach cultivars. *Journal of the American Society for Horticultural Science*, 861–864.

Seifert, Blaukopf (2010). Irritable walls: the plant extracellular matrix and signaling. *Plant Physiology*, 467–478.

Seyednasrollah, Laiho, Elo (2013). Comparison of software packages for detecting differential expression in RNA-seq studies. *Briefings in Bioinformatics*, 59–70.

Shiu, Bleecker (2001). Plant Receptor-Like Kinase Gene Family: Diversity, Function, and Signaling. *Science Signaling*, 1-13.

Shiu, Bleecker (2003). Expansion of the receptor-like kinase/Pelle gene family and receptor-like proteins in Arabidopsis. *Plant Physiology*, 530–543.

Streif, Retamales, Cooper (1994). Preventing cold storage disorders in nectarines. En *International Symposium on Postharvest Treatment of Horticultural Crops*, 160–166. International Society for Horticultural Science - ISHS.

Tarazona, García-Alcalde (2011). Diferential expression in RNA-seq: a matter of depth. *Genome Research*, 2213–2223.

Trainotti, Pavanello, Zanin (2006). PpEG4 is a peach endo-{beta}-1,4-glucanase gene whose expression in climacteric peaches does not follow a climacteric pattern. *Journal of Experimental Botany*, 589–598.

Trainotti, Zanin, Casadoro (2003). A cell wall-oriented genomic approach reveals a new and unexpected complexity of the softening in peaches. *Journal of Experimental Botany*, 1821–1832.

Trapnell, Pachter, Salzberg (2009). TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics*, 1105–1111.

Valero, Barreiro, Ruiz-Altisent, Cubeddu, Pifferi, Taroni, Torricelli, Valentini, Johnson, Dover (2005). Mealiness detection in apples using time resolved reflectance spectroscopy. *Journal of Texture Studies*, 439–458.

Van Ooijen (2009). MapQTL 6, Software for the mapping of quantitative trait loci in experimental populations of dipoid species. https://www.kyazma.nl/index.php/MapQTL/.

Verde, Abbott, Scalabrin, Jung, Shu, Marroni, Zhebentyayeva, *et al*. (2013). The high-quality draft genome of peach (Prunus persica) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nature genetics*, 487–494.

Vidal, Williams, O'Neill, Pellerin (2001). Polysaccharides from grape berry cell walls. Part I: Tissue distribution and structural characterization of the pectic polysaccharides. *Carbohydrate Polymers*, 315–323.

Vizoso, Meisel, Tittarelli, Latorre, Saba, Caroca, Maldonado, *et al*. (2009). Comparative EST transcript profiling of peach fruits under different post-harvest conditions reveals candidate genes associated with peach fruit quality. *BioMed Central Genomics*, 5-18.

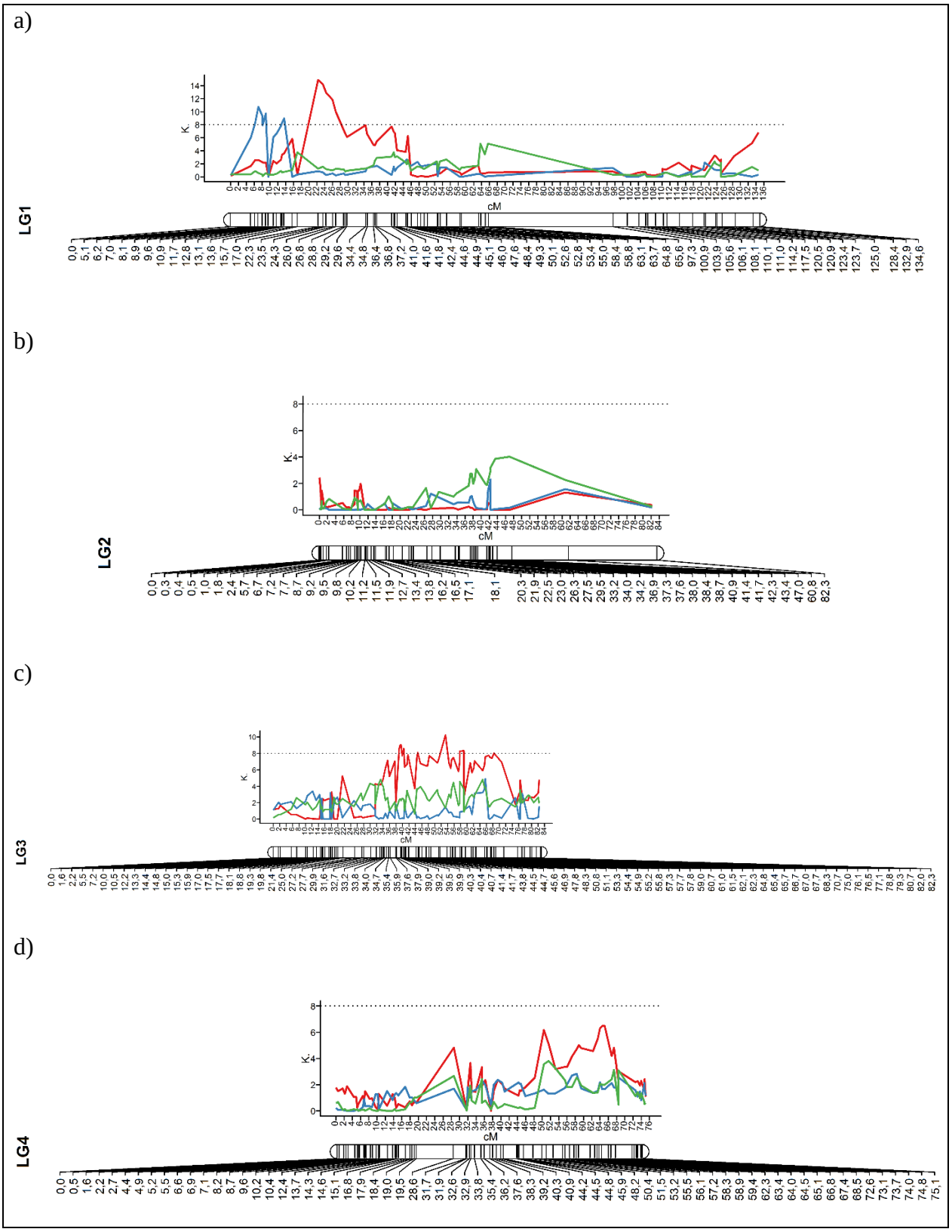
Wang, Fan, Siao, Berno, Young, Sapolsky, Ghandour, *et al*. (1998). Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science*, 1077–1082.

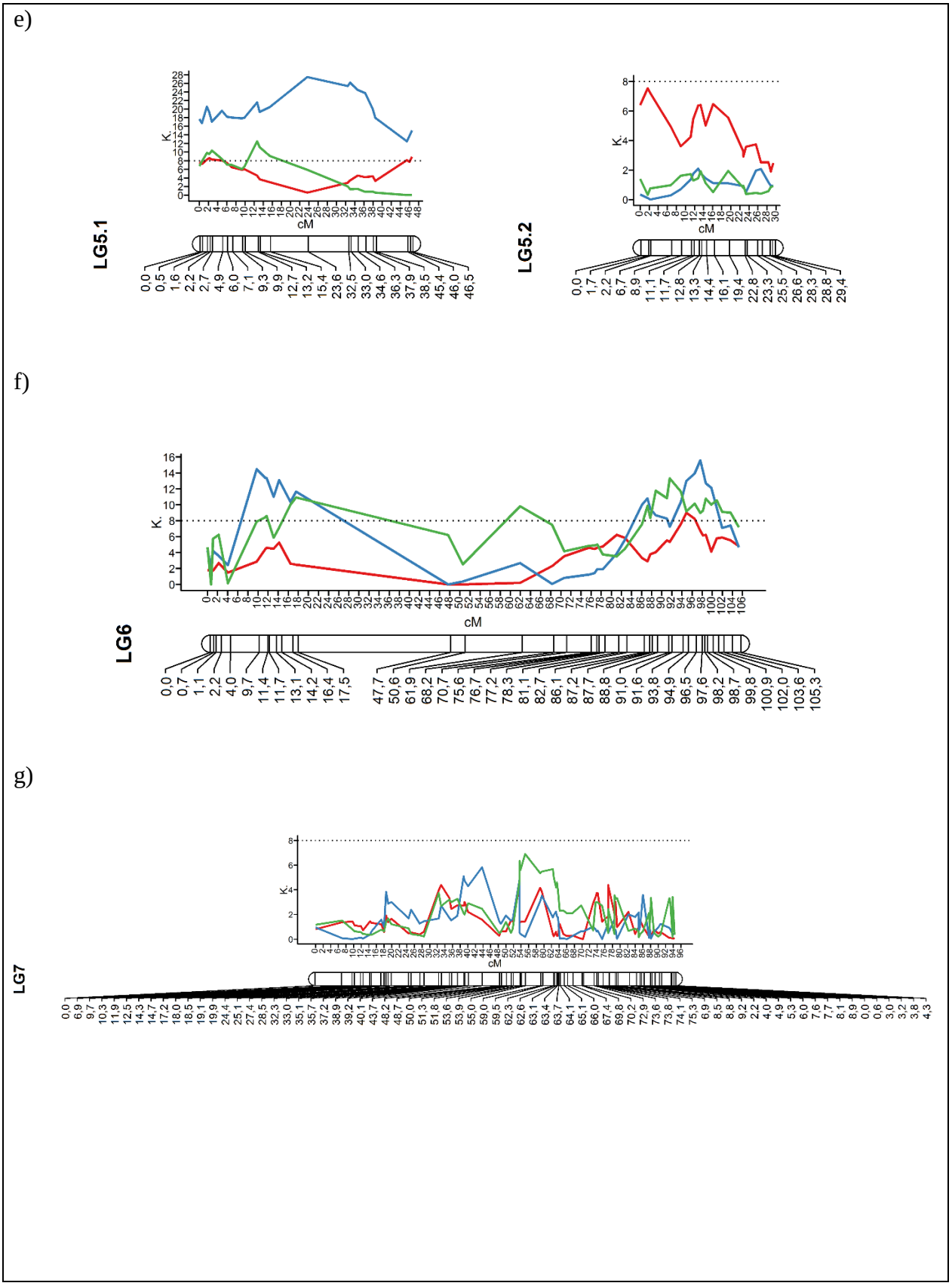
Wang, Gerstein, Snyder (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews. Genetics*, 57–63.

Young, Wakefield, Smyth, Oshlack (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome biology*, 1-12.

Zhou, Dong, Ben-Arie, Lurie (2001). The role of ethylene in the prevention of chilling injury in nectarines. *Journal of Plant Physiology*, 55–61.

Anexos:




e)
LG5.1
K
cM
0,0 0,5 1,6 2,2 2,7 4,9 6,0 7,1 9,3 9,9 12,7 13,2 15,4 23,6 32,5 33,0 34,6 36,3 37,9 38,5 45,4 46,0 46,5
LG5.2
K
cM
0,0 1,7 2,2 6,7 8,9 11,1 11,7 12,8 13,3 14,4 16,1 19,4 22,8 23,3 25,5 26,6 28,3 28,8 29,4
f)
LG6
K
cM
0,0 0,7 1,1 2,2 4,0 9,7 11,4 11,7 13,1 14,2 16,4 17,5 47,7 50,6 61,9 68,2 70,7 75,6 76,7 77,2 78,3 81,1 82,7 86,1 87,2 87,7 88,8 91,0 91,6 93,8 94,9 96,5 97,6 98,2 98,7 99,8 100,9 102,0 103,6 105,3
g)
LG7
K
cM

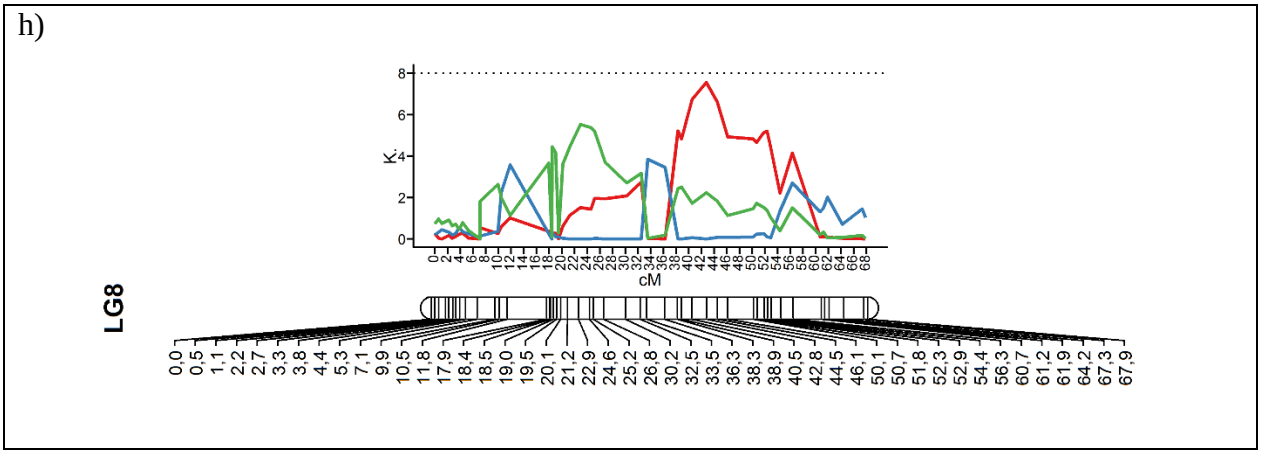


Figura Suplementaria 1: Detalle de los QTL convencionales para el carácter de harinosidad en tres temporadas de evaluación: La línea roja representa la temporada 2014-2015 de evaluación fenotípica para el campo de Choapinos. La línea azul representa la temporada 2015-2016 para el mismo campo y la línea verde representa la temporada 2014-2015 de Rinconada.

Tabla Suplementaria 1: Métrica alineamiento utilizando Tophat2 para cada una de las librerías.

Librería	Numero de Lecturas	Lecturas Remanentes luego de Flexbar (%)	Alineamiento de Lecturas	Tasa de Alineamiento General (%)	Alineamiento Concordante	Tasa de Alineamiento Concordante (%)
S1	42.612.180	98,21%	40.260.218	94,50%	19.392.091	90,80%
S2	38.958.418	98,98%	36.866.760	94,60%	17.813.627	91,40%
S3	43.036.094	98,71%	40.710.268	94,60%	19.660.309	91,30%
S4	43.237.372	99,19%	41.020.513	94,90%	19.810.553	91,50%
S5	41.322.098	99,39%	39.291.847	95,10%	18.993.740	91,80%
S6	42.629.332	99,22%	40.468.317	94,90%	19.559.883	91,70%
S7	40.219.910	99,47%	38.423.755	95,50%	18.610.480	92,50%
S8	36.997.192	99,49%	35.369.784	95,60%	17.147.543	92,60%
S9	37.679.052	99,21%	35.904.062	95,30%	17.355.505	91,80%
S10	35.659.614	99,48%	33.078.280	92,80%	15.866.110	88,90%
S11	35.825.036	99,40%	33.223.178	92,70%	15.941.431	88,90%
S12	37.462.362	99,46%	34.790.219	92,90%	16.707.798	89,10%
S13	37.832.012	99,35%	35.622.630	94,20%	17.154.409	90,50%
S14	38.415.700	99,28%	36.219.827	94,30%	17.444.371	90,70%
S15	37.709.396	99,43%	35.538.280	94,20%	17.114.938	90,60%
S16	38.591.032	98,95%	36.653.917	95,00%	17.720.464	91,80%
S17	40.113.378	99,16%	38.332.592	95,60%	18.578.754	92,60%
S18	42.790.794	99,05%	40.782.683	95,30%	19.728.989	92,20%
S19	45.161.614	99,25%	42.709.039	94,60%	20.610.950	90,70%
S20	38.859.822	99,35%	37.043.720	95,30%	17.998.664	92,50%
S21	38.560.072	98,45%	36.109.593	93,60%	17.291.218	89,00%
S22	40.436.170	98,54%	38.166.423	94,40%	18.403.044	90,90%
S23	39.743.590	97,86%	37.279.393	93,80%	17.869.185	89,50%
S24	45.210.698	99,26%	42.977.851	95,10%	20.785.817	91,80%
S25	46.087.918	97,45%	43.262.277	93,90%	20.714.350	89,50%
S26	42.482.432	96,89%	39.828.128	93,80%	19.038.765	89,30%
S27	44.928.866	97,77%	42.243.488	94,00%	20.295.486	90,10%
S28	42.118.528	96,27%	39.267.337	93,20%	18.708.439	88,40%
S29	52.727.474	99,13%	49.877.130	94,60%	24.062.493	90,70%
S30	48.856.862	99,10%	46.233.779	94,60%	22.315.711	91,20%
S31	48.597.838	99,14%	46.181.295	95,00%	22.317.503	91,70%
S32	52.347.798	99,13%	49.738.836	95,00%	24.103.127	92,00%
S33	44.180.696	98,86%	41.843.617	94,70%	20.184.139	91,00%
S34	45.542.618	99,15%	43.315.261	95,10%	20.949.570	91,80%
S35	43.489.208	98,87%	41.330.214	95,00%	19.972.343	91,80%
S36	39.939.946	97,47%	37.896.657	94,90%	18.322.404	91,70%
S37	43.366.948	98,19%	41.055.305	94,70%	19.771.330	91,00%
S38	37.889.728	93,92%	35.407.577	93,40%	16.919.289	89,30%
S39	39.511.176	97,83%	37.289.309	94,40%	17.948.796	90,80%
S40	40.561.566	98,91%	38.405.007	94,70%	18.518.693	91,30%

S41	39.344.596	98,81%	36.942.751	93,90%	17.903.225	90,90%
S42	40.750.832	98,80%	37.785.277	92,70%	18.300.740	89,80%
S43	35.068.468	98,53%	32.549.412	92,80%	15.691.064	89,40%
S44	39.658.326	99,22%	37.232.076	93,90%	17.792.225	89,60%
S45	41.026.504	99,12%	38.472.319	93,80%	18.480.949	90,00%
S46	39.254.000	99,21%	37.252.566	94,90%	18.008.192	91,60%
S47	34.570.032	96,15%	32.175.483	93,10%	15.351.023	88,20%
S48	40.461.862	97,76%	38.086.173	94,10%	18.302.263	89,90%
S49	37.608.948	97,44%	35.101.729	93,30%	16.793.602	88,90%
S50	35.956.766	99,09%	34.258.059	95,30%	16.565.881	92,00%
S51	41.270.090	98,97%	39.232.907	95,10%	18.944.542	90,80%
S52	29.410.848	99,11%	27.837.060	94,60%	13.407.082	91,00%
S53	30.743.386	99,22%	29.108.792	94,70%	14.059.760	91,40%
S54	30.848.270	99,20%	29.287.305	94,90%	14.143.092	91,60%
S55	35.536.696	99,16%	33.509.996	94,30%	16.163.379	90,90%
S56	39.621.036	99,10%	37.005.684	93,40%	17.789.894	89,70%
S57	37.828.716	99,10%	35.241.547	93,20%	16.959.038	89,50%
S58	38.295.040	99,22%	35.775.522	93,40%	17.245.784	90,00%
S59	32.833.426	99,71%	31.153.025	94,90%	15.077.147	91,80%
S60	35.623.592	99,64%	33.557.307	94,20%	16.223.116	91,00%
S61	35.604.390	99,74%	33.707.629	94,70%	16.303.501	91,60%
S62	35.322.952	99,75%	33.596.678	95,10%	16.272.497	92,10%
S63	37.670.070	99,61%	35.946.520	95,40%	17.434.222	92,50%
S64	36.678.642	99,60%	34.891.488	95,10%	16.901.224	92,10%
S65	37.061.834	99,47%	35.166.251	94,90%	17.038.441	91,90%
S66	38.618.326	99,63%	36.511.527	94,50%	17.709.384	91,70%
S67	36.723.176	99,62%	34.884.324	95,00%	16.910.273	92,00%
S68	32.839.366	99,69%	31.362.542	95,50%	15.211.378	92,50%
S69	35.831.116	99,54%	34.114.507	95,20%	16.510.196	91,90%
S70	42.613.352	99,75%	40.721.190	95,60%	19.742.929	92,60%
S71	38.052.080	99,57%	36.137.055	95,00%	17.460.034	91,70%
S72	46.124.972	99,50%	43.708.657	94,80%	21.085.937	91,40%
S73	34.718.482	99,52%	32.928.775	94,80%	15.893.423	91,50%
S74	43.261.850	99,56%	40.946.251	94,60%	19.736.835	91,20%
S75	36.701.004	99,50%	34.832.837	94,90%	16.817.821	91,60%
S76	35.138.112	99,56%	33.462.060	95,20%	16.191.492	92,10%
S77	43.202.910	99,39%	40.852.829	94,60%	19.663.580	90,90%
S78	46.302.828	99,27%	43.842.797	94,70%	21.081.619	90,90%
S79	51.208.374	99,45%	48.635.563	95,00%	23.434.184	91,30%
S80	35.795.572	99,49%	33.945.895	94,80%	16.383.196	91,40%
S81	47.035.512	99,49%	44.712.740	95,10%	21.601.294	91,80%
S82	46.980.082	99,47%	44.515.344	94,80%	21.470.029	91,30%
S83	42.812.632	99,46%	40.602.531	94,80%	19.589.667	91,20%

S84	41.151.956	99,34%	39.046.459	94,90%	18.838.727	91,40%
S85	42.823.060	94,51%	39.901.878	93,20%	19.004.135	88,70%
S86	43.547.022	99,42%	41.480.827	95,30%	20.070.492	92,00%
S87	44.490.838	99,20%	42.276.696	95,00%	20.403.176	91,50%
S88	43.144.452	99,26%	41.037.917	95,10%	19.830.800	91,70%
S89	38.465.326	99,43%	36.377.224	94,60%	17.504.421	90,90%
S90	34.592.010	99,27%	32.716.751	94,60%	15.747.803	90,90%
S91	35.778.072	99,46%	33.970.414	94,90%	16.401.191	91,60%
S92	37.002.646	99,49%	35.107.404	94,90%	16.969.693	91,70%
S93	46.943.558	99,28%	44.604.254	95,00%	21.491.230	91,50%
S94	35.697.088	99,49%	34.064.789	95,40%	16.483.146	92,30%
S95	39.681.454	99,34%	37.041.638	93,30%	17.762.966	89,40%
S96	38.499.914	99,47%	35.897.814	93,20%	17.221.973	89,40%
S97	41.657.276	99,47%	38.799.777	93,10%	18.591.813	89,20%
S98	40.104.122	99,44%	37.255.132	92,90%	17.858.804	88,90%
S99	44.585.632	99,35%	41.501.871	93,10%	19.903.770	89,20%
S100	39.702.924	99,40%	36.782.054	92,60%	17.621.618	88,70%
S101	41.462.774	99,18%	39.464.987	95,20%	19.085.315	91,70%
S102	38.741.402	99,33%	36.897.825	95,20%	17.855.943	92,00%
S103	45.298.650	99,35%	43.257.375	95,50%	20.958.077	92,40%
S104	39.901.318	99,49%	37.042.025	92,80%	17.781.580	89,00%
S105	40.191.454	99,50%	37.438.434	93,20%	17.994.854	89,50%
S106	40.993.120	99,40%	38.188.916	93,20%	18.353.123	89,50%
S107	34.416.078	99,46%	32.647.507	94,90%	15.721.442	91,30%
S108	29.854.362	99,36%	28.340.523	94,90%	13.652.058	91,40%
S109	26.476.746	99,44%	25.203.851	95,20%	12.168.651	91,80%
S110	38.191.094	97,03%	35.905.330	94,00%	17.260.625	90,10%
S111	33.761.664	93,53%	30.988.621	91,80%	14.610.512	85,80%
S112	34.675.756	95,83%	32.296.612	93,10%	15.401.692	87,80%
S113	36.542.940	98,42%	34.900.088	95,50%	16.916.922	92,50%
S114	37.804.964	98,30%	36.059.407	95,40%	17.456.104	92,30%
S115	40.156.830	98,35%	38.289.092	95,30%	18.536.588	92,20%
S116	35.310.754	98,29%	33.713.396	95,50%	16.329.071	92,40%
S117	47.899.892	98,47%	45.859.847	95,70%	22.257.844	92,90%
S118	35.146.828	98,45%	33.602.674	95,60%	16.286.550	92,60%
S119	45.046.502	98,31%	42.010.225	93,30%	20.196.369	89,60%
S120	40.143.112	98,41%	37.501.543	93,40%	18.029.549	89,70%
S121	44.480.172	98,32%	41.296.506	92,80%	19.801.176	89,00%
S122	39.936.410	98,45%	38.141.554	95,50%	18.469.552	92,40%
S123	41.983.984	99,47%	40.059.028	95,40%	19.381.279	92,30%
S124	35.179.132	99,50%	33.437.173	95,00%	16.177.088	91,90%
S125	40.804.752	99,36%	37.892.675	92,90%	18.138.612	88,60%
S126	41.444.386	99,51%	38.621.038	93,20%	18.536.318	89,30%

S127	52.991.596	99,53%	48.933.693	92,30%	23.426.764	88,30%
S128	36.956.262	99,52%	34.604.935	93,60%	16.652.763	90,00%
S129	40.666.354	99,43%	37.799.999	93,00%	18.122.215	89,10%
S130	46.972.288	99,29%	43.878.706	93,40%	21.070.445	89,30%
S131	48.578.546	99,23%	46.357.562	95,40%	22.455.372	92,40%
S132	37.985.746	98,69%	35.801.388	94,20%	17.351.745	91,20%
S133	38.175.788	99,11%	34.981.377	91,60%	16.950.397	88,70%
S134	37.601.558	99,26%	35.669.783	94,90%	17.218.997	91,50%
S135	35.371.430	99,39%	33.597.161	95,00%	16.228.720	91,70%
S136	36.578.588	99,46%	34.776.790	95,10%	16.823.347	91,90%
S137	38.423.648	99,34%	36.526.111	95,10%	17.644.370	91,80%
S138	35.628.436	99,45%	33.929.534	95,20%	16.423.372	92,10%
S139	41.830.378	99,40%	39.907.398	95,40%	19.353.987	92,40%
S140	39.802.296	99,44%	37.981.598	95,40%	18.432.692	92,50%
S141	44.754.320	99,53%	42.774.644	95,60%	20.783.743	92,80%
S142	44.239.408	99,56%	42.295.651	95,60%	20.535.519	92,70%
S143	39.094.060	99,43%	37.320.021	95,50%	18.067.160	92,40%
S144	40.258.918	99,55%	38.324.236	95,20%	18.509.596	91,90%
S145	47.039.916	99,02%	44.687.695	95,00%	21.570.716	91,40%
S146	41.645.518	99,51%	39.729.370	95,40%	19.257.124	92,40%
S147	37.019.204	99,44%	35.270.290	95,30%	17.075.665	92,20%
S148	39.564.942	99,29%	37.617.078	95,10%	18.176.437	91,80%

Tabla Suplementaria 2: Partidores utilizados en la comprobación del transcriptoma por medio de qPCR.

Código del Gen	Gen	Forward (5-3)	Reverse (5-3)
Prupe.5G118000	GH9C2	CCAACCAAAAGTAGCTCCACTACC	ACTCCCTTGGAAACCCAAGA
Prupe.8G186000	UDP-GT	AACAAGAGGCGGTGGATTGGAT	CTCCCAGTTGTCTTCATGACGCTT
Prupe.4G278500	BGAL10	CCTAACTCCACTTCTGTGGT	ACACTGGGACACAAATCCTC
Prupe.8G180600	TMT2	TCTCTGGCAGAAGGAGTTTGCT	TGCATTCACAACGCTGCCCATA
Prupe.1G270800	UTR3	ACCCCTTGGGTATGGACTGTGTTT	TTTCCCAGGCGCTTGTCTTT