



UNIVERSIDAD DE BUENOS AIRES
Facultad de Ciencias Exactas y Naturales

Departamento de Fisiología y Biología Molecular y Celular

Análisis funcional de los determinantes genéticos y epigenéticos del contenido de vitamina E en tomate

**Tesis presentada para optar al título de Doctor de la
en el área CIENCIAS BIOLÓGICAS**

Estanislao Burgos

Director de Tesis: Dr. Fernando Carrari
Director asistente: Dr. Leandro Quadrana

Consejero de estudios: Dr. Norberto Iusem

Lugar de trabajo: Instituto de Biotecnología, CICVyA, INTA

Buenos Aires, 2018

Resumen

Análisis funcional de los determinantes genéticos y epigenéticos del contenido de vitamina E en tomate

El objetivo de esta tesis es dilucidar los determinantes genéticos de los contenidos de vitamina E (VTE) en tomate y su relación con el ambiente. Por un lado, se analizaron los contenidos de VTE en frutos de 150 entradas provenientes de distintos bancos de germoplasma y de 126 líneas recombinantes derivadas de un cruzamiento multiparental cultivados en ambientes contrastantes. Mediante estudios de asociación para el genoma completo (GWAS por sus siglas en inglés) y mapeo de QTLs, se identificaron numerosos *loci* potencialmente involucrados en la determinación de este carácter. Por otro lado, resultados previos del laboratorio indican que la expresión de los genes duplicados *VTE3(1)* y *VTE3(2)*, cuyos productos son responsables de los últimos pasos de la biosíntesis de VTE, son capaces de modular la acumulación de VTE en hojas y frutos de tomate. En particular, la expresión de *VTE3(1)* es regulada epigenéticamente debido a la metilación variable del ADN de la secuencia promotora. Con el objetivo de caracterizar las bases moleculares de esta regulación, como así también la co-regulación transcripcional de ambos parálogos, se generó y caracterizó molecularmente una población de 80 líneas recombinantes derivada del cruzamiento de líneas de introgresión portando los alelos silvestres (*S. pennellii*) y domesticados (*S. lycopersicum*) de dichos genes. Adicionalmente, se caracterizó la variación natural de expresión y metilación del ADN de *VTE3(1)* y *VTE3(2)* en una población de variedades andinas de tomate. Estos experimentos revelaron que la modulación transcripcional de *VTE3(1)* y *VTE3(2)*, mediada en parte por la variación en los niveles de metilación de sus promotores, es en efecto, un blanco importante para el mejoramiento de este carácter nutricional. En conjunto, esta tesis pone en evidencia la importancia de explorar tanto la variación intraespecífica como interespecífica para caracterizar las bases genéticas de fenotipos metabólicos y la necesidad de incluir la variación epigenética como un factor determinante de caracteres de interés agronómico.

Palabras Claves: Metabolismo secundario, Vitamina E, Epigenética, Parálogos, Tomate.

Abstract

Functional analysis of genetic and epigenetic determinants of Vitamin E content in tomato

The objective of this thesis is to elucidate the genetic determinants of the vitamin E (VTE) contents in tomato and their relationship with the environmental conditions. On one hand, the VTE fruit's contents were analyzed for 150 entries, from different germplasm banks, and for 126 recombinant lines derived from a multiparental crossing cultivated in contrasting environments. Through Genome-wide association study (GWAS) and QTL mapping, several loci were identified potentially involved in the determination of this character. On the other hand, previous results of our group, indicate that the expression of the duplicated genes *VTE3(1)* and *VTE3(2)*, whose products are responsible for the last steps of VTE biosynthesis, are capable of modulating the accumulation of VTE in leaves and tomato fruits. In particular, the expression of *VTE3(1)* is naturally and epigenetically regulated through the variable DNA methylation of the promoter region. In order to characterize the molecular basis of this regulation, as well as the transcriptional co-regulation of both paralogs, a population of 80 recombinant lines, derived from the crossing of introgression lines carrying the wild alleles (*S. pennellii*) and domesticated (*S. lycopersicum*) of both paralogs genes, were generated and molecularly characterized. Moreover, the natural variation of *VTE3(1)* and *VTE3(2)* expression and DNA methylation was characterized in a population of Andean tomato landraces. These experiments revealed that the transcriptional modulation of *VTE3(1)* and *VTE3(2)*, partially regulated by the variation on the methylation levels of their promoters, is in fact an important target for the improvement of this nutritional character. Taken together these results, this thesis highlights the importance of exploiting both intraspecific and interspecific variation to characterize the (epi)genetic basis of metabolic phenotypes, and remarks the significance of include epigenetic variation as a determinant of agronomical important traits.

Key Words: Secondary Metabolism, Vitamin E, Epigenetics, Paralogs, tomato.

Índice

Listado de abreviaturas	1
CAPITULO I - Introducción	3
<i>I.1 El metabolismo vegetal; su importancia como fuente de moléculas nutracéuticas</i>	3
<i>I.1.1 El metabolismo de tococromanos. Su importancia en la dieta humana y en las plantas</i>	5
<i>I.2 El fruto de tomate (<i>Solanum lycopersicum</i>) como modelo de estudio del metabolismo</i>	10
<i>I.3 Recursos genéticos y mejoramiento en tomate</i>	13
<i>I.3.1 Líneas de introgresión; poblaciones multiparentales; Mapeo genético; GWAS y mapeo de QTLs</i>	17
<i>I.3.2 Ventajas, limitaciones y complementación entre el mapeo de QTLs y GWAS</i>	20
<i>I.3.3 Criterio de evaluación de análisis de GWAS</i>	22
<i>I.4 Bases genéticas del metabolismo</i>	23
<i>I.4.1 Interacción génica y alélica. Mecanismos epigenéticos asociados</i>	23
<i>I.4.2 Heterosis y epistasis</i>	27
<i>I.5 Definición del problema, hipótesis y objetivos</i>	29
<i>I.5.1 Objetivo general</i>	31
<i>I.5.2 Objetivos específicos</i>	32
CAPITULO II- Búsqueda de nuevos determinantes genéticos asociados a los contenidos de VTE, abordando dos estrategias complementarias GWAS y mapeo de QTLs	33
<i>II.1 Introducción a las poblaciones estudiadas de GWAS y mapeo de QTL</i>	33
<i>II.1.2 Análisis de estructura génica y LD de ambas poblaciones de estudio</i>	33

<i>II.2. Resultados.....</i>	<i>37</i>
<i>II.2.1 Analisis de los QTLs identificados a través de 6 analisis de GWAS distintos y contraste con el Mapeo de la población F7 multiparental.....</i>	<i>37</i>
<i>II.2.2 Analisis de enriquecimiento en categorías funcionales de genes para los análisis de Mapeo</i>	<i>38</i>
<i>II.2.3 Estudio del enriquecimiento en factores de transcripción para los genes identificados en los análisis de mapeo.....</i>	<i>41</i>
<i>II. 2.4 Genes identificados por GWAS análisis y mapeo de QTL</i>	<i>44</i>
<i>II.2.5 Arquitectura del carácter heredabilidad VTE (disección de la variabilidad asociada al carácter VTE).....</i>	<i>47</i>
<i>II.3. Discusion.....</i>	<i>48</i>
CAPITULO III- Interacción génica y alélica entre VTE3(1) y VTE3(2).....	51
<i>III.1 Introducción.....</i>	<i>51</i>
<i>III.2.Resultados.....</i>	<i>57</i>
<i>III.2.1 Obtencion de una población F2 recombinante para los alelos silvestres y cultivados de VTE3(1) y VTE3(2).....</i>	<i>57</i>
<i>III.2.2 Contenidos de tocoferoles y niveles de metilación y expresión de líneas de introgresión, parentales de la F2, desafiadas en dos ambientes contrastantes</i>	<i>57</i>
<i>III.2.3. Caracterización fenotípica y molecular de individuos de la población F2 con genotipos parentales.....</i>	<i>61</i>
<i>III.2.4 Análisis de la población completa F2 y evauacion de la contribución de facotres genéticos y epigenéticos, ambiente, y sus interacciones en los niveles de tocoferol y expresión de VTE3(1) y VTE3(2).....</i>	<i>62</i>
<i>III.2.5 Variación natural de los mecanismos epigenéticos de regulación de genes, generales.....</i>	<i>70</i>

CAPITULO IV- Integración del conocimiento	74
IV.1 Conclusiones.....	74
CAPÍTULO V – Materiales y Métodos	76
V.1 Descripción de los materiales y diseño experimental de una población de GWAS y 124 líneas recombinantes derivadas de un cruzamiento multiparental cultivados en ambientes contrastantes para la búsqueda de QTLs.....	76
V.1 .1 Materiales GWAS.....	76
V.1 .2 Materiales MAGIC POP.....	77
V.1 .3 Mapeo y datos genotípicos GWAS.....	77
V.1 .4 Mapeo y datos genotípicos MAGIC POP.....	77
V.1 .5 Diseño experimental y condiciones de riego materiales GWAS y MAGIC-POP.....	78
V.2 Descripción de la obtención de una población F2 proveniente de la cruce de dos líneas de introgresión.....	78
V.2.2 Cruzamientos.....	78
V.2.3 Desarrollo y aplicación de marcadores moleculares para el genotipado de la población recombinante F2 Cruza IL 3-1 x IL9-2-6-1 (Ensayo apilamiento de dos loci).....	79
V.3 Medición de datos fenotípicos (contenido de tococromanoles) por HPLC..	80
V.4.1 PCR en tiempo Real.....	81
Glosario	83
CAPÍTULO VI – Referencias bibliográficas	85
Anexo	96
Tabla S1. Vision global del conjunto de materiales utilizados en esta tesis a traves de los distintos capitulos de esta Tesis.....	96

<i>Tabla S2. Lista de primers utilizados.....</i>	<i>97</i>
<i>Tabla S3. Genes candidatos identificados a través de los distintos análisis de GWAS y el mapeo de QTLs reportados en la literatura</i>	<i>98</i>
<i>Tabla S4. Listado de Factores de transcripción identificados en base de datos de unión a sitios CIS del set genes VTE reportados en la literatura.....</i>	<i>101</i>
<i>Tabla S5. Niveles de expresión y anotación funcional de los 186 ORFs dentro de la IL3-1.....</i>	<i>104</i>

Listado de abreviaturas

DMPBQ: 2,3-dimetil 6-fitol 1,4-benzoquinona

GGDP: geranil geranil difosfato

HGA: ácido homogenístico

HGGT: homogentisato geranil geranil transferasa

LD: desequilibrio de ligamiento

MEP: ruta metabólica del metileritritol fosfato

MGGBQ: 2-metil 6-geranilgeranil benzoquinol

MP: metabolitos primarios

MPBQ: 2-metil 6-fitol 1,4-benzoquinona

MS: Metabolitos secundarios

PDP: Fitol di fosfato

QTLs: caracteres cuantitativos asociados a una región genómica de un determinado locus del Inglés, *Quantitative Trait Loci*.

SK: Vía metabólica del Shikimato.

TF: factores de transcripción.

VTC: Vitamina C

VTE: Vitamina E

VTE1: enzima tocoferol Ciclase

VTE2: enzima homogentizato fitil Transferasa

VTE3(1): gen que codifica para la 2-metil-6-itolquinol metiltransferasa

VTE3(2): gen que codifica para 2-metil-6-fitolquinol metiltransferasa (de su sigla en inglés).

VTE5: enzima fitol quinasa

VTE6: enzima fitil fosfato quinasa

MAGIC POP: población avanzada de múltiples intercruces de su sigla en inglés).

CAPITULO I- Introducción

1.1 El metabolismo vegetal; su importancia como fuente de moléculas nutraceuticas

Las plantas como organismos superiores se caracterizan por la complejidad de su red metabólica, aunque varían según la especie. Son organismos sésiles que han tenido que desarrollar diversas estrategias para su supervivencia. La síntesis de azúcares, aminoácidos, nucleótidos, lípidos, también conocidos como metabolitos primarios (MP), son los sustratos precursores la síntesis de los metabolitos secundarios (MS), los cuales juegan funciones ecológicas específicas claves en el mantenimiento de la planta frente a animales herbívoros, infecciones microbianas o virales y la radiación UV, que incluyen a las lectinas, inhibidores de amilasa o de proteinasas y glucósidos o glucosinolatos cianogénicos, y a otras sustancias aditivas que atraen a los polinizadores, etc.

La diversidad de MS producidos en el reino de las plantas es muy grande, existiendo cerca de 200.000 especies químicas estructurales distintas (Dixon y Strack 2003). Ciertos MS están restringidos a unas pocas especies de plantas, siendo un número muy pequeño redundante en un dado tejido y/o especie (Ferne 2007). Sin embargo, algunas especies han evolucionado a una diversidad particular presumiblemente por su adaptación a condiciones ambientales en las que se desarrollan (Li y col. 2015, Futuyama y Agrawal 2009). Esta es quizá la razón por la cual nuestro conocimiento de la estructura metabólica global es fragmentada para una gran cantidad de vías (Ferne y Takayuki 2017).

Los MS se pueden dividir en tres grandes grupos: compuestos fenólicos, terpenoides isoprenoides y alcaloides glucosinolatos, que se producen a partir de diferentes vías del metabolismo primario: la glucólisis, el ciclo de Krebs, aminoácidos alifáticos, la ruta de las pentosas fosfato, via del Shikimato y aminoácidos aromáticos (figura 1).

Existe una estrecha relación entre los MP y los MS debido a que las vías del metabolismo secundario han evolucionado de forma parcial o completa a partir de otras vías del propio metabolismo secundario, pero también del primario. Es así, que muchos de los pasos en las vías son compartidos, tanto por sustratos, como enzimas las cuales difieren en la especificidad y/o su localización celular. Un buen ejemplo es el paso de generación de la cadena ramificada en leucina/metionina y la síntesis de glucosinolatos (Strassman y Ceci 1963).

La hipótesis del balance carbono/nitrógeno en plantas afirma que, bajo condiciones de disponibilidad limitada de nitrógeno, el metabolismo secundario se dirige a producir metabolitos ricos en carbono y viceversa, aunque existen excepciones (Coley y col. 1985). Se ha demostrado por ejemplo, como un incremento la relación azúcar/aminoácidos, resulta en una elevación en fenilpropanoides y en un decaimiento en los contenidos de alcaloides nicotínicos (Matt y col. 2002). Otro ejemplo de regulación entre MP y MS, es el caso del azufre, un componente esencial para plantas que cumple una serie de funciones en diferentes reacciones enzimáticas, en la regulación redox, metilación de DNA (dadores de grupos metilos), síntesis de poliaminas, como también es incorporado en una serie de metabolitos secundarios. Cuando una planta crece en un ambiente deficiente en azufre, se produce el reciclado catabólico del azufre, siendo este un mecanismo crucial para responder a una serie de situaciones de estrés.

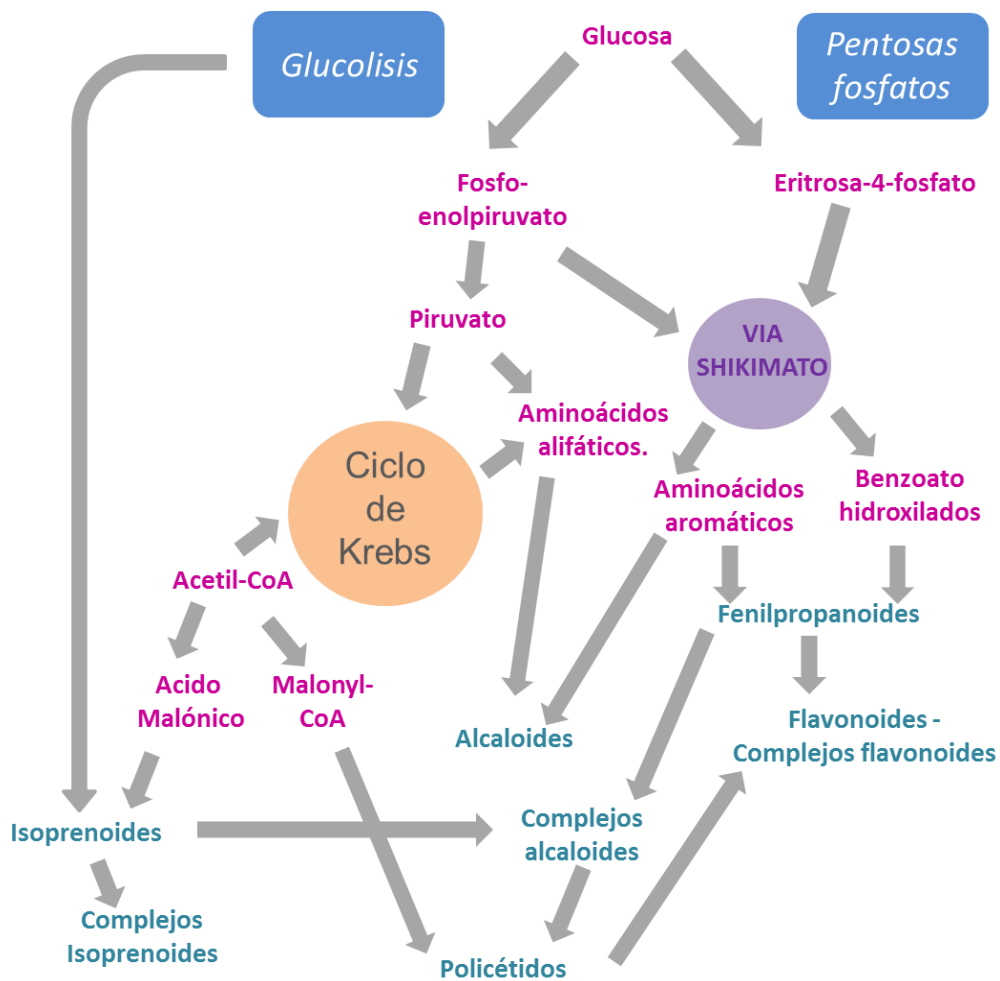


Figura 1. Vías del metabolismo primario y secundario. La producción de metabolitos secundarios asociada a vías del metabolismo primario y central como la glicólisis, la vía del Shikimato, la síntesis de aminoácidos aromáticos y alifáticos. Adaptada de Aharoni y Galili 2011.

En los últimos años se han realizado importantes avances en la comprensión de la regulación de flujos metabólicos en las plantas. La expresión de genes que codifican para varios factores de transcripción (TF) asociados al metabolismo, ha demostrado que regula tanto la expresión de genes que codifican enzimas que participan del metabolismo primario como a genes que codifican enzimas específicas involucradas en el metabolismo secundario relacionadas entre sí, indicando la existencia de redes altamente coordinadas entre MP y MS en las plantas (Maruyama y col. 2006, Malitsky y col. 2007, Sonderby y col. 2007).

Desde hace ya tiempo, el principal desafío para la biotecnología vegetal es aumentar, mediante ingeniería metabólica, los niveles de moléculas activas en plantas, ya que estos metabolitos secundarios son útiles para múltiples propósitos, desde el empleo como tintas, saborizantes, fragancias, a estimulantes alucinógenos, insecticidas, agentes terapéuticos y nutricionales. La mejor forma para abordar este objetivo será de manera integrada entre los MP, MS y su regulación a distintos niveles.

1.1.1 El metabolismo de tococromanoles. Su importancia en la dieta humana y en las plantas

Los seres humanos necesitan ingerir nutrientes que les permitan desarrollar todos sus procesos vitales, siendo su fuente principal el consumo de vegetales. Entre estas moléculas esenciales que el organismo humano no tiene la capacidad intrínseca de sintetizar, y debe incorporar en la dieta, se incluyen las vitaminas, cuya función general es mantener la homeostasis del metabolismo celular promoviendo el correcto funcionamiento fisiológico. Las vitaminas pueden clasificarse por su solubilidad, siendo hidrosolubles las vitaminas del complejo B y la vitamina C, mientras que la vitamina A (Retinol), D (Calciferol), K (Filoquinona) y E (Tococromanoles: tocoferoles y tocotrienoles) son liposolubles (Fitzpatrick y col. 2012). Hasta la fecha, 13 compuestos diferentes han sido clasificados como vitaminas. Cada una de estas moléculas tiene una función específica en las células vegetales, que puede ser diferente del mecanismo de acción que ejercen en células de mamíferos, por el cual se las considera esenciales. A diferencia de los nutrientes principales, las vitaminas son requeridas en pequeñas cantidades (lo que las define como micronutrientes). En este sentido, la razón por la que algunos vegetales sintetizan (y acumulan) grandes cantidades de estas moléculas cuando su mecanismo de acción, si es el mismo que se describe para células de mamíferos, requeriría de pequeñas cantidades. Para el caso de los antioxidantes, como las vitaminas C (VTC) y E (VTE) la explicación surge de la

explosión oxidativa que implican la fotosíntesis y la capacidad de lidiar con patógenos que poseen las células vegetales.

En particular, la VTE es el mayor antioxidante lipofílico incorporado en la dieta, y está constituida por un conjunto de moléculas denominadas tococromanoles (tocoferoles y tocotrienoles), sintetizadas exclusivamente por organismos fotosintéticos. Los tocoferoles son moléculas anfipáticas constituidas por una cabeza polar cromanol, derivada del ácido homogentisato, y una cola lipofílica que deriva del fitol. Por su parte, los tocotrienoles son moléculas que difieren de los tocoferoles en la longitud e insaturación de la cola lipofílica (derivada del geranilgeranil difosfato), pero que encuentran amplia relación en su biosíntesis. Ambos grupos de moléculas se clasifican naturalmente en cuatro especies (α , β , γ y δ) metiladas diferencialmente en el anillo cromanol (Figura 2).

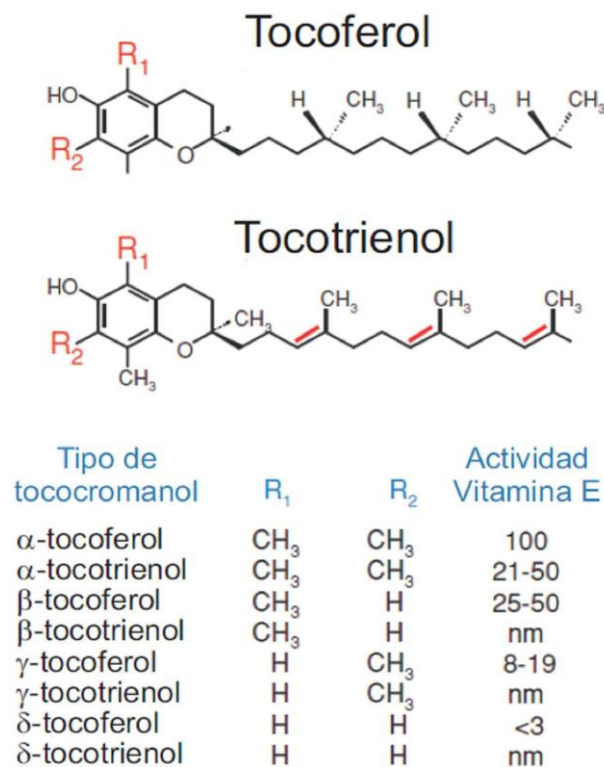


Figura 2. Estructura de tococromanoles. La figura muestra las moléculas de tocoferol y tocotrienol, indicando en rojo las diferencias entre las diferentes isoformas. En la parte inferior se especifica las sustituciones en el anillo cromanol para cada isoforma y su capacidad antioxidante relativa a α -tocoferol. "nm" refiere a que la capacidad no ha sido medida. Adaptado de DellaPenna y Pogson 2006.

La biosíntesis *de novo* de los tocoferoles se localiza en plástidos y algunas de las enzimas involucradas se encuentran en la membrana interna de cloroplastos y cromoplastos, mientras que otras se ubican en estructuras membranosas asociadas a los tilacoides, llamadas plastoglóbulos (Vidi y col. 2006, Ytterberg y col. 2006, Miret y Munné-Bosch 2015) (Figura 3). En plantas superiores, los tococromanoles se han encontrado en mayor concentración en plástidos, coincidentemente con la localización de las enzimas de la vía central (Sun y col. 2009).

La confluencia de las vías del Shikimato (SK) y Metileritritol fosfato (MEP) dan lugar al *core* de la VTE, o ruta central de su síntesis, la cual fue dilucidada por Soll y Schultz en 1979. Asimismo, el primer gen de esta vía fue clonado en *A. thaliana* en 1998, años siguientes integrando conocimientos obtenidos de organismos modelo *A. thaliana* y *Synechocystis sp*, permitió manipular diversos pasos de la biosíntesis (individuales o múltiples), a través del uso de mutantes y/o líneas transgénicas, modificando los contenidos y composición de VTE tanto en hojas como en semillas de *A. thaliana* (Della Penna y Pogson, 2006) sugiriendo que existe cierta plasticidad en la producción de VTE en plantas superiores (Falk y Munné-Bosch 2010, Mene-Saffrane y DellaPenna 2010).

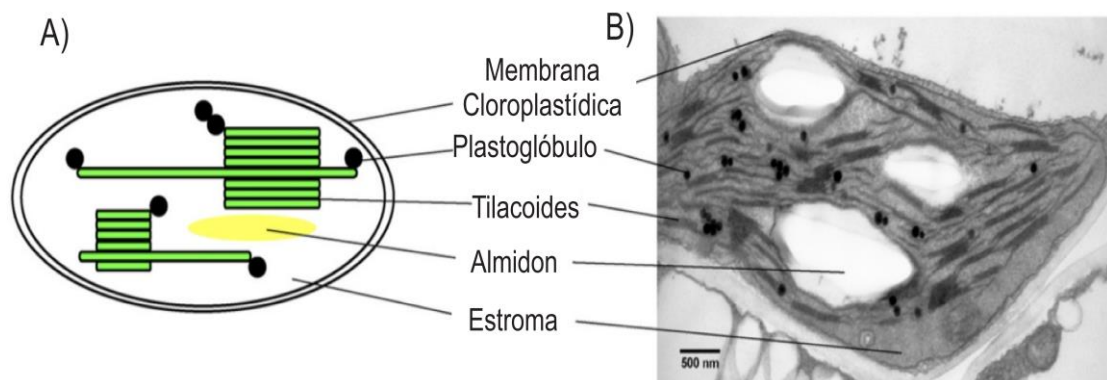


Figura 3. Esquema de compartimentalización de un cloroplasto. A) El esquema muestra el cloroplasto y su organización, delimitada por una doble membrana. El interior de los cloroplastos comprende el estroma acuoso, las membranas tilacoides (en verde), los gránulos de almidón (amarillos) y los plastoglóbulos (partículas negras), que son partículas lipídicas asociadas a la tilacoides. **B)** Esta imagen muestra en detalle un cloroplasto mediante microscopía electrónica de transferencia a partir de un corte ultrafino de tejido de hoja de *A. Thaliana*. Los plastoglóbulos son visibles luego de su tinción con tetroxido de osmio. En la parte inferior de la fotografía se muestra la escala que corresponde a 500 nm. Adaptada de Vidi 2007.

El primer paso del *core* ocurre a partir de PDP (fitil difosfato), uno de los sustratos de VTE2, la homogentizato fitil transferasa, la única enzima exclusiva de la síntesis de VTE. Ésta, convierte HGA (ácido homogentísico) en 2-metil 6-fitil 1,4-benzoquinona (MPBQ). Luego la tocoferol ciclasa (VTE1), cataliza la

síntesis del anillo cromanol permitiendo la formación de δ -tocoferol. Paralelamente, la metilación de MPBQ por parte de VTE3 (2-metil 6-fitol metiltransferasa) forma 2,3-dimetil 6-fitol 1,4-benzoquinona (DMPBQ), que luego de la ciclación por parte de VTE1, da como producto γ -tocoferol. El último paso de la vía que culmina con la adición de un grupo metilo en la posición 6 del anillo cromanol, es realizado por la enzima gama tocoferol metil transferasa (VTE4), la cual parte de δ -tocoferol y γ -tocopherol para producir β -tocopherol y α -tocopherol, respectivamente. El precursor PDP también puede provenir del reciclado de fitol como producto de la degradación de clorofila, por la acción de las enzimas fitol quinasa (VTE5) y la fitil fosfato quinasa (VTE6) (Valentin 2006, Dorp y col. 2015).

Por otro lado, los tocotrienoles son sintetizados por la condensación de HGA con geranil geranil difosfato (GGDP), por la acción de la homogentisato geranil-geranil transferasa (HGGT). Como resultado, se produce 2-metil 6-geranil-geranil benzoquinol (MGGBQ), el cual es modificado de igual manera que en los tocoferoles por VTE1, VTE3 y VTE4, para producir α -, β -, δ - y γ -tocotrienoles.

Aunque la mayoría de las enzimas y actividades envueltas en la biosíntesis de tococromanoles están localizadas en la membrana de las tilacoides (Spicher y Kessler 2015), el paso clave que es la ciclación por parte de VTE1 ocurre en los plastoglobulos (Vidi y col. 2006). Sin embargo, HGA se produce en el citosol (Wang y col. 2016), lo cual implica un tráfico de moléculas precursoras para que la biosíntesis de VTE ocurra. El tráfico debe ocurrir entre la membrana de los PG y la de los cloroplastos (Spicher y Kessler, 2015) y también entre el citoplasma y los plástidos (Pellaud y Mène Saffrané 2017). En este sentido, estudios previos de nuestro grupo, utilizando líneas de introgresión (ILs) de *S. Pennelli* permitieron identificar QTL para contenidos de tocoferol que mapeaban sobre el cromosoma 9. Este QTL colocaliza con un locus que codifica para SEC 14-like (Almeida y col. 2011). Además, ha sido reportado que algunos miembros de esta familia están relacionados al transporte de tocoferol en mamíferos y en levaduras (Saito y col. 2007, Bankaitis y col. 2010).

Gran cantidad de reportes han demostrado el rol protector de los tococromanoles a los ácidos grasos polinsaturados frente a las especies reactivas de oxígeno (ROS), en un proceso conocido como peroxidación lipídica (revisado por Wolf 2005). La actividad antioxidante de los tococromanoles resulta de su reacción química directa con compuestos reactivos tales como radicales lipídicos, especies reactivas de oxígeno y de nitrógeno (Kamal-Eldin y Appelqvist 1996, Munné-Bosch y Alegre 2002, Schneider 2005). En la actualidad, el mecanismo antioxidante de estos compuestos se cree que implica la transferencia de un átomo de hidrógeno

fenólico del anillo cromanol a un radical peroxil lipídico, resultando en la producción de un hidroperóxido lipídico y un radical tococromanol. El radical tococromanol es más estable debido a la deslocalización del electrón desapareado y por lo tanto reduciendo su reactividad. Sin embargo, puede convertirse nuevamente en su tococromanol correspondiente por la reacción con otro reductor (por ejemplo, ascorbato o ubiquinol) (Kobayashi y DellaPenna 2008). También se ha propuesto que los tocoferoles, junto con la plastoquinona 9 (PQ-9), capturan oxígenos en su forma singulete (1O_2), producidos durante la fotosíntesis por el fotosistema II (Miret y Munné-Bosch 2015, Kruk y col. 2005, Kruk y Trebst 2008).

La VTE más allá de su función antioxidante, también posee un rol ecofisiológico muy importante en las plantas, el cual se extiende desde la germinación, la partición de fotoasimilados, la división celular, el crecimiento, la senescencia foliar, hasta las respuestas al estrés biótico y abiótico (Falk y Munné-Bosch 2010). Además, varios estudios han demostrado una estrecha interacción entre el metabolismo de tocoferoles y otras vías. Frutos de tomate que sobre-expresan la fitoeno sintasa (PSY), una enzima clave en la biosíntesis de carotenoides, muestran mayores niveles de tocoferol (Fraser y col. 2007). Por otro lado, los contenidos de tococromanoles se ven afectados cuando la vía de post-corismato es manipulada. Plantas transgénicas de *A. thaliana* sobreexpresando el gen bacteriano que sintetiza la enzima bifuncional corismato mutasa / prefenato deshidratasa (CM/PDT), muestran niveles significativamente más altos de fenilalanina, así como de γ -tocoferol y γ -tocotrienol, además de otros metabolitos secundarios (Tzin y col. 2009). También se ha estudiado la variación de VTE mediada por señales endógenas y ambientales. Un claro ejemplo es el factor de respuesta a la luz DE-ETIOLATED1, el cual resulta en frutos de tomate con mayores niveles de antioxidantes, entre ellos carotenoides, flavonoides y tocoferol (Davuluri y col. 2005, Enfissi y col. 2010). Sin embargo, a pesar del gran número de trabajos sobre el metabolismo de VTE en organismos modelos, muy poco se sabe sobre la síntesis de tocoferol y la regulación de la vía biosintética en tomate, un modelo de estudio y producto hortícola cuyo aporte de VTE a la dieta humana es sustancial.

Desde el primer aislamiento de VTE del aceite de trigo (Evans y col. 1936), los tococromanoles han sido reconocidos como un poderoso componente de defensa contra el daño oxidativo a través de la eliminación de radicales libres poliinsaturados, evitando así la peroxidación lipídica.

Han sido descritas nuevas funciones de la VTE, así como también su influencia en la actividad enzimática, las cascadas de señalización y la

regulación génica, sin embargo los mecanismos moleculares in vivo aún no han sido dilucidados (Azzi 2007, Traber y Atkinson 2007, Briglius-Flohé 2009). En tanto, la forma alfa (α) de VTE es considerada la de mayor actividad dado a que se almacena y transporta de manera eficaz en células de mamíferos (Traber y Sies, 1996). A pesar de que es inusual la deficiencia en vitamina E en la población occidental (Dror y Allen, 2011), modelos experimentales animales evidenciaron síntomas específicos vinculados a la insuficiencia de esta vitamina, como por ejemplo resorción fetal, distrofia muscular y encefalomalacia. Esta situación específica se ha encontrado en personas que no pueden metabolizar grasas (patología digestiva), en individuos con desórdenes en el metabolismo de lípidos (patología genética) y en niños prematuros con un peso corporal bajo. Más aún, ensayos realizados bajo dietas pobres en vitamina E se han asociado a una mayor predisposición a enfermedades como cáncer, aterosclerosis y degeneración neuronal (Saldeen y Saldeen 2004, Terasawa y col. 2000, Yokota y col. 2001). Debido a que la dieta rica en grasas es la que proporciona la mayor absorción de vitamina E y en la actualidad hay una tendencia a reducir el consumo de grasas, esto conduciría a aumentar el consumo de tocoferol en la dieta de un adulto para así alcanzar los niveles recomendados de 15 mg/día (Bender y col. 2003). En este sentido, el consumo de vegetales que aporten micronutrientes lipofílicos como la VTE es altamente recomendable. Proveer las bases del conocimiento que permitan entender los procesos regulatorios del metabolismo de los mismos es un pre-requisito para la explotación y consumo de dichos vegetales.

1.2 El fruto de tomate (*Solanum lycopersicum*) como modelo de estudio del metabolismo

La importancia económica de esta especie, sumado a las investigaciones básicas y aplicadas desarrolladas sobre este cultivo, dan lugar a que sea un modelo de estudio ideal. El tomate cultivado pertenece a la familia de la Solanacea, que después de los cereales y las leguminosas, es la de mayor valor agregado en origen. Las mismas se diversificaron en America del sur mucho antes de la separación de Gondwana, pruebas de esto lo dio la identificación de un fósil de 52.2 millones años de un fruto de tomatillo petrificado, encontrado en la Patagonia (Wilf y col. 2017).

Esta hortaliza comprende más de 3.000 especies que se adaptan a los más diversos ambientes y dentro de esta familia, es la más cultivada mundialmente después de la papa. La batata, el aji, la berenjena y el tabaco son otros miembros de esta misma familia y constituyen otros modelos de estudio, como por ejemplo en el desarrollo de tubérculos (Fernie y col. 2001, Kolomiets y col. 2001) o en estudios de relaciones entre hospedantes y patógenos (Bogdanove

y Martin 2000, Chung y Felton 2011, Gebhardt y Valkonen 2001, Martin y col. 1993, Pedley & Martin 2003, Rathjen y col. 1999, Rossi y col. 1998).

Las principales ventajas de la familia de las Solanaceas del empleo como modelo para el estudio de la diversidad y la adaptación ambiental, radican en su amplia variabilidad fenotípica y su conservación genómica y cromosómica demostrados a partir del análisis de cariotipo de diversas especies (Wang y col. 2008, Wu y Tanksley, 2010).

La sección *Lycopersicon* del género *Solanun* está formada por 1.500 especies, originarias en su mayoría del oeste de América del Sur (Ecuador y Perú) hasta el norte de Bolivia y Chile, a excepción de dos que son endémicas de las Islas Galapagos (Peralta y col. 2008). Debido a su amplia aplicación y distribución, a pesar de ser nativa del oeste de Sudamérica, no es posible asegurar si derivó por selección artificial de alguna especie muy cercana (*S. pimpinellifolium*) o si existió en su forma silvestre en la naturaleza (Peralta y col. 2008). Las plantas ruderales de tomate cultivado se extendieron a otras partes de América, y en México fueron cultivadas y seleccionadas por los pueblos precolombinos (Jenkins, 1948). En la actualidad, el tomate está ampliamente distribuido a nivel mundial en cada uno de los continentes a excepción de la Antártida, ya que son cultivados o persisten en sus formas ruderales en ambientes alterados por la actividad del hombre.

El genoma de *S. lycopersicum* es diploide y relativamente pequeño (~900Mb). Esta es otra de las razones que lo posicionan como un modelo de investigación alternativo al de *Arabidopsis thaliana* para el estudio de procesos de formación, maduración y metabolismo de frutos carnosos y climatéricos (Carrari y Fernie, 2006, Giovannoni y col. 2004). Además, esta especie es naturalmente autógama, las poblaciones sintéticas son estables y es capaz de cruzarse con otras especies del género (con *S. pimpinellifolium*, *S. habrochaites* y *S. pennellii*) de forma eficaz, sencilla y asistida. Estas características, muy importantes para su adopción como modelo de estudio, permitieron contar con amplios recursos genéticos y genómicos. Sumado a que posee protocolos de transformación génica de alta reproducibilidad (The Tomato Genome Consortium, 2012). Recientemente, se han publicado varios trabajos que dan cuenta de la posibilidad de aplicar edición génica en genes con diferentes mecanismos de acción (Brooks y col. 2014, Yasuhiro Lto y col. 2015, Changtian y col. 2016, Soyk y col. 2018).

Dependiendo de las condiciones climáticas y de su crecimiento primario y secundario (por ejemplo la capacidad de desarrollar tallos secundarios), el tomate es una planta herbácea anual, bianual o perenne. La mayoría de las plantas poseen un crecimiento indeterminado, sin embargo algunas variedades

son semideterminadas o determinadas, formando arbustos más ramificados y compacto. El tallo principal es simpodial, formado ejes laterales con hojas alternadas dispuestas en espiral de 1/3, en la que cada unidad finaliza con una inflorescencia.

Por su parte, el fruto es una baya compuesta por tejido epidérmico, un pericarpio grueso y la placenta gelatinosa que envuelve a las semillas. El desarrollo del fruto se divide en 4 etapas, de dos a tres semanas de duración, bien determinadas que involucran, I) el inicio de la floración hasta la antesis, II) el desarrollo del fruto caracterizado por divisiones celulares, III) la expansión celular sin divisiones, y finalmente, IV) la maduración determinada por modificaciones tanto en la composición metabólica (azúcares, aminoácidos y pigmentos) como en la composición de las paredes celulares (firmeza del fruto). Esta última etapa ha sido ampliamente estudiada tanto a nivel molecular como bioquímico, considerando su importancia en la cadena productiva en especial luego de la cosecha (Alba y col. 2005; Alexander & Grierson, 2002; Bapat y col. 2010; Barry y col. 2005; Carrari y col. 2006; de Jong y col. 2009).

Está ampliamente descrito que este estadio de desarrollo se inicia una vez concluida la expansión celular, acompañado de un cambio abrupto en los perfiles metabólicos, un aumento de la hormona etileno y de la respiración (Carrari y Fernie, 2006). En tanto, se caracteriza por la des-diferenciación de cloroplastos y la concomitante degradación de clorofila, la re-diferenciación a cromoplastos con la acumulación de carotenos (licopeno y caroteno), dando lugar al color rojo característico de la maduración del fruto. Además, se producen modificaciones en la textura del fruto por variaciones en la turgencia celular y en la estructura de las paredes, acompañadas por cambios en la composición de los principales metabolitos (azúcares, ácidos orgánicos) y compuestos volátiles, que determinan su calidad nutricional y organoléptica (Carrari y Fernie, 2006, Tieman 2006).

El fruto del tomate es una de las principales fuentes de VTE para la dieta debido al elevado consumo en las poblaciones occidentales. Los datos publicados por el USDA en el año 2012 ubican al consumo de pasta de tomate en el 5to puesto entre 8.176 productos dietarios, por debajo de los cereales suplementarios (www.ars.usda.gov/ba/bhnrc/ndl). Además de VTE, el fruto del tomate aporta cantidades considerables de ácido ascórbico, licopenos, carotenos y compuestos fenólicos entre los principales antioxidantes. Existen numerosos estudios que describen las bases moleculares de la producción de estos compuestos en plantas superiores, aunque la mayoría hacen referencia a la VTC (Conklin y col. 2000, Jain y col. 1999) y licopenos (Auldrige y col. 2006, Simkim y col. 2004). En particular, han sido investigadas las diferentes vías de biosíntesis para la VTC mediante el clonado y mapeo de los genes

involucrados (Zou y col. 2015). Además, otros autores han identificado la co-localización de algunos de estos genes con *loci* específicos para caracteres cuantitativos o QTL (del inglés, Quantitative Trait Loci) para las variaciones en los contenidos de VTC en tres poblaciones diferentes de tomate silvestre (Stevens y col. 2006). Investigaciones realizadas en relación a los compuestos fenólicos aportados por el fruto del tomate señalan que los flavonoides y ácidos fenólicos, representados por quercetina y ácido clorogénico son los que se encuentran en mayor proporción (Slimestad y Verheul 2005). Estos antecedentes, sumado a la aplicación de estrategias de ingeniería metabólica, han permitido lograr frutos de tomate ricos en altos contenidos de estos compuestos (Giovinazzo y col. 2005, Muir y col. 2001, van der Rest y col. 2006). En contraste, el conocimiento desarrollado sobre la base de la regulación de las vías biosintéticas de tocoferoles en frutos de tomate es muy acotado (Almeida y col. 2011, Quadrana y col. 2014, Silvestre Lira y col. 2014). No obstante, los estudios realizados sobre el análisis de perfiles metabólicos en estos frutos indican la extensa variabilidad presentada entre las distintas especies y cultivares de tomate, proporcionando el conocimiento para el mejoramiento de estos caracteres (Schauer y col. 2005, Schauer y col. 2006; Quadrana y col, 2014).

En conjunto, estos antecedentes proponen al cultivar de tomate como un modelo ideal para el metabolismo de la VTE.

1.3 Recursos genéticos y mejoramiento en tomate

Hace más de 200 años en Europa, se comenzaron a reproducir mejoras en el cultivar de tomate para satisfacer varias necesidades (Foolad 2007). Sin embargo en términos de rendimientos e índices de cosecha, actualmente los cultivares y/o híbridos, seleccionados por sus rendimientos redujeron de manera considerable la base genética explotable del género *Solanum* (Bai y Lindhout 2007) y sólo un pequeño porcentaje de variables genéticas de la sección *Lycopersicon* son utilizadas para el mejoramiento. En consecuencia, no solo ha tenido un impacto negativo en la producción de compuestos de interés alimentario y farmacológico sino también en el estudio genético de las vías metabólicas de esta especie. En este sentido, un claro ejemplo es un trabajo que demuestra que el contenido de aminoácidos y pigmentos es entre 10 y 30 veces mayor en especies silvestres emparentadas respecto a los frutos de la especie domesticada (*S. lycopersicum*) (Schauer y col. 2005).

El mejoramiento genético en el rendimiento se obtuvo mediante cruzamientos de especies de parentales silvestres beneficiadas a partir del conocimiento de numerosos marcadores moleculares específicos (Swamy y Sarla 2008), potencialmente involucrados en la determinación de caracteres específicos que

impactan directamente en su producción. En este sentido, la identificación de distintos loci cuya variación alélica está asociada con un rasgo cuantitativo o también denominado QTL, implicados en el rendimiento y otros rasgos relacionados como peso del fruto, en los sólidos solubles totales o en el contenido de licopeno, entre otros, en *S. peruvianum* (Fulton y col. 1997), *S. pennelli* (Schauer y col. 2006; Kamenetzky y col. 2010) y *S. parviflorum* (Fulton y col. 2000), distribuidos al azar en el genoma del tomate. A partir de allí, estas regiones genómicas particulares se convirtieron en objetivos interesantes para la mejora del rendimiento del tomate (Swamy y Sarla 2008).

Más tarde, en los años 80, la vida útil era el problema más importante para los tomates frescos marcados y los principales avances se centraron en elucidar la etapa del desarrollo de la maduración, principalmente porque sus determinantes afectan directamente la vida útil. Diferentes estudios se concentraron en la identificación de los principales componentes implicados en la maduración y firmeza de la fruta, incluido el papel del etileno (Alba y col., 2005) y la enzima poligalacturonasa (Giovannoni 2001). Pocas mutantes de maduración se caracterizaron y utilizaron en líneas heterocigotas y en el desarrollo de cultivares que exhibieron una maduración retardada, una mayor vida útil y una mayor firmeza. Más recientemente, se caracterizaron y mapearon varios genes relacionados con la maduración y QTL, siendo uno de ellos fue el inhibidor de la maduración (*rin*) utilizado en la reproducción asistida por marcadores con resultados prometedores (Foolad 2007). Ya en los años 90, el sabor se convirtió en el objetivo principal. Éste es incluso un rasgo más complejo ya que no está claramente diseccionado y comprendido, principalmente porque está determinado por una gran cantidad de factores genéticos y no genéticos, no todos identificados o bien caracterizados (Causse y col. 2002). Se ha postulado que la relación entre azúcares y ácidos orgánicos es el principal determinante del sabor del tomate (Bucheli y col. 1999). Además, varios compuestos volátiles aromáticos tienen importantes influencias sobre el sabor (Goff y Klee 2006). A pesar de algunos ejemplos que explican el esfuerzo para mejorar esta propiedad de la fruta, esta tarea se basa principalmente en los incrementos de sólidos solubles y acidez obtenidos en algunos casos (Foolad 2007). Finalmente, en la última y actual etapa, el objetivo de la mejora se centra en la calidad nutricional del tomate (Bai y Lindhout 2007).

Hasta la fecha se han identificado gran cantidad de polimorfismos involucrados en el desarrollo y la composición del fruto en tomate para la mejora en la calidad. En la tabla se enumeran los principales, que afectan directa o indirectamente la calidad nutricional. Algunas de estas, pueden inducir variaciones en el color o aspecto de la fruta. Otras mutaciones, afectan a la arquitectura de la planta, como *sp*, que controla el crecimiento determinado/indeterminado, afectando la calidad (Pnueli y col. 1998). Otro

ejemplo es la mutación *rin* (gen involucrado en el proceso de maduración) en el nivel heterocigótico. La presencia de la mutación *rin* redujo la preferencia del consumidor. Las diferencias se detectaron por los perfiles sensoriales, los híbridos de *rin* tenían frutos un 17% menos dulces, con menos aroma de tomate y más "extraño" y frutos más harinosos, aunque la firmeza y el contenido de azúcar no fueron diferentes. Estos resultados aclaran la influencia negativa de la mutación *rin* en la preferencia del consumidor, pero también indicaron que cuando se transfiere a un híbrido con alto sabor, la influencia negativa de la mutación se reduce. Por lo tanto, la selección podría llevarse a cabo para obtener líneas mucho más dulces y perfumadas combinadas con la vida útil en híbridos *rin*.

Tabla 1. Genes clonados en tomate, su ubicación, fenotipos relacionados a calidad de fruto, hoja y arquitectura

Tabla 1. Genes clonados en tomate, su ubicación, fenotipos relacionados a calidad de fruto, hoja y arquitectura					
	Nombre en ingles	Chr.	Posicion	Fenotipo	Referencia
<i>y</i>	<i>Colorless epidermis</i>	1	71255600	Epidermis rosada	(Ballester y col. 2010)
<i>L-2</i>	<i>Lutescent-2</i>	10	61858478	Alteracion de cloroplastos	(Barry y col. 2012)
<i>Gr</i>	<i>green flesh</i>	8	60582066	frutos verdes	(Barry y col. 2008)
<i>C</i>	<i>potato leaf</i>	6	42804036	hojas simples	(Busch y col. 2011)
<i>r</i>	<i>phytoene synthase 1</i>	3	8606749	frutos amarillos	(Fray y Grienson 1993)
<i>cwp1</i>	<i>cuticular water permeability 1</i>	4	63765366	microfisuras /deshidratacion del fruto	(Hovav y col. 2007)
<i>t</i>	<i>carotenoid isomerase</i>	10	62006972	frutos naranja	(Isaacson y col. 2007)
<i>s</i>	<i>compound inflorescence</i>	2	36913957	inflorescencia	(Lippman y col. 2008)
<i>Cnr</i>	<i>colorless nonripening</i>	2	37323107	no madura	(Manning y col. 2006)
<i>j</i>	<i>jointless</i>	11	3640857	ausencia de pedicelo	(Mao y col. 2000)
<i>fa</i>	<i>falsifora</i>	3	61162449	inflorescencia frondosa	(Molinero Rosales y col. 2004)
<i>sft</i>	<i>single flower truss</i>	3	30564833	<i>single flower truss</i>	(Molinero Rosales y col. 2004)
<i>hp-2</i>	<i>de etiolated 1</i>	1	46495644	alta pigmentacion	(mustili y col. 1999)
<i>sp</i>	<i>self-pruning</i>	6	42361623	habito determinado	(Pnueli y col. 1998)
<i>u</i>	<i>uniform ripening</i>	10	2293088	incremento clorofila	(Powell y col. 2012)
<i>Del</i>	<i>Delta</i>	12	2285372	frutos naranja	(Ronen y col. 1999)
<i>B</i>	<i>Beta-carotene</i>	6	42288127	incremento Beta carotenos	(ronen y col. 2000)

<i>sucr</i>	<i>sucrose accumulator</i>	3	47401871	acumulacion de sacarosa en frutos	(Sato y col. 1993)
<i>ls</i>	<i>lateral suppresser</i>	7	64958148	supresion de corola parcialmente esteril el macho	(Schumacher y col. 1999)
<i>hp-3</i>	<i>Zeaxanthin epoxidase</i>	2	46947557	alta pigmentacion	(Thompson y col 2000a)
<i>rin</i>	<i>ripening inhibitor</i>	5	5217073	nunca madura	(Vrebalov y col. 2002)
<i>mc</i>	<i>macrocalyx</i>	5	5214073	Largos sepalos	(weller y col. 2001)
<i>phyB2</i>	<i>apophytochrome B2</i>	5	62648223	Recepción luz roja	(weller y col. 2001)
<i>phyA</i>	<i>apophytochrome A</i>	10	22854459	rojo lejano	Wikinson y col. 1995)
<i>Nr</i>	<i>Never ripe</i>	9	62631866	no madura	
<i>e</i>	<i>Entire leaf</i>	4	59354677	reducida complejidad hoja	(Zhang y col. 2007)

La mayoría de los caracteres implicados en la calidad del tomate se heredan cuantitativamente. Por esta razón, el tomate fue uno de los primeros cultivos a partir de los cuales se usaron marcadores moleculares para diseccionar la base genética de los rasgos cuantitativos o QTL (*Quantitative Trait Loci*) a partir del trabajo pionero de Tanksley (1993). Debido al muy bajo polimorfismo revelado a nivel de especie, la mayoría de estos estudios se realizaron en progenies inter-específicas derivadas de cruces entre especies de tomate silvestre y el tomate cultivado (excepto el trabajo de Causse y col. 2001; Saliba-Colombani y col. 2001).

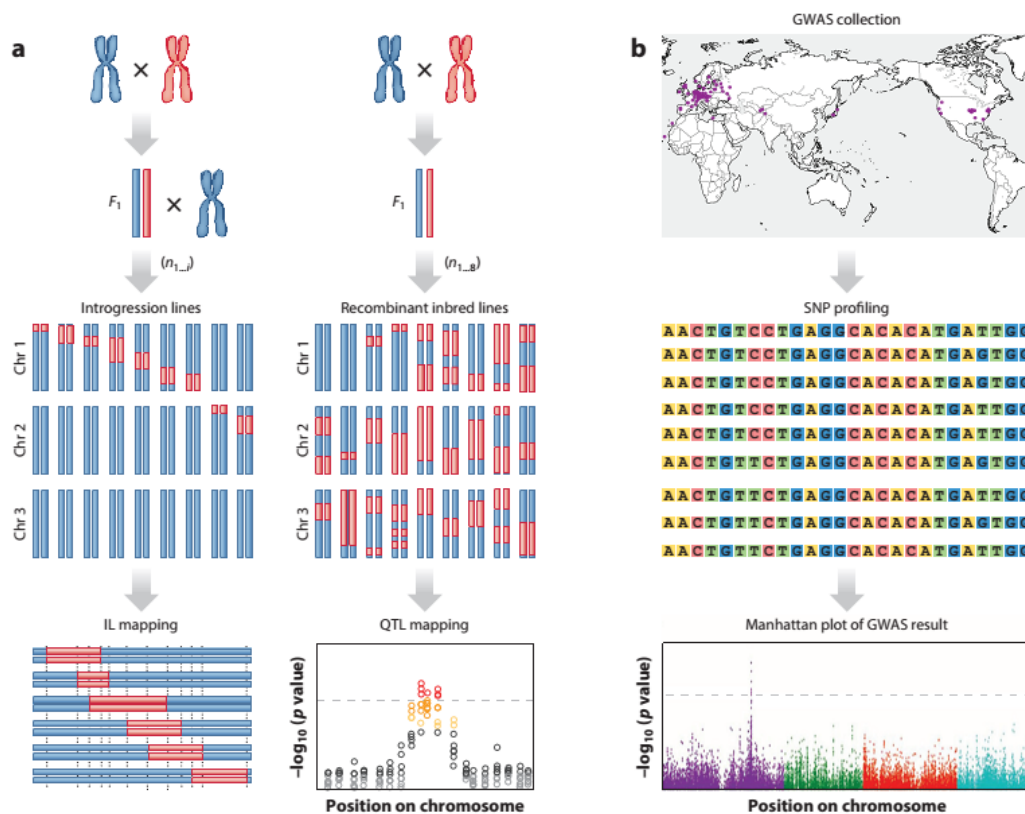
En la mayoría de los estudios, algunos QTL explicaron una gran fracción (20 a 50%) de la variación fenotípica, conjuntamente con QTL menores que no pudieron ser detectados. La mayoría de los QTL actuaron de manera aditiva, pero se detectaron QTL dominantes y superdominantes (Paterson y col. 1988, 1991, DeVicente y Tanksley 1993, Semel y col. 2006). La epistasia (interacción entre los QTL) rara vez se detectó a menos que se usara un diseño experimental específico (Tanksley y col. 1996, Causse y col. 2007).

La secuenciación reciente del genoma del tomate, seguida de la liberación de matrices de genotipado que permiten genotipar simultáneamente miles de SNP a escala de genoma completo, ha allanado el camino para utilizar diseños de mapeo QTL más complejos, en los niveles inter e intraespecífico, para acotar la brecha entre fenotipo y genotipo. Utilizando un conjunto de 192 marcadores SNP genotificados en 188 accesiones, Xu y col. (2013) identificaron 2, 16 y 17 loci asociados a la acidez, los sólidos solubles y el contenido de azúcar, respectivamente, lo que demuestra la viabilidad del mapeo de asociación de genoma amplio (GWA) para caracteres de calidad de fruta de tomate.

La combinación alélica favorable entre los loci asociados a la calidad de la fruta, como el pH, la acidez titulable, la SSC o la forma del fruto se identificaron mediante modelos lineales mixtos en colecciones de tomate cultivares y variedades nativas genotipadas para miles de SNP (Ruggieri y col. 2014, Sacco y col. 2015).

1.3.1 Líneas de introgresión; poblaciones multiparentales; Mapeo genético; GWAS y mapeo de QTLs

El mapeo de QTLs en poblaciones segregantes está basado en genotipar la progenie derivada de cruces entre distintos genotipos y asociar marcadores moleculares al fenotipo de interés en la descendencia (figura 4).



Las poblaciones homocigotas requeridas para estos propósitos, son obtenidas mediante autocruzamientos repetidos como es el caso de las RILs (del inglés, recombinant inbred lines) o de DHs (Doble haploids) que son genotipos producidos induciendo espontáneamente células doble haploides de gametofito y regenerado plantas de las resultantes células DH.

El uso de poblaciones estables de mediante el crecimiento de réplicas clonales permite la ventaja de múltiples análisis de individuos idénticos genéticamente en distintas condiciones ambientales. El establecimiento de las bases genéticas como los caracteres cuantitativos QTLs esta supeditados a una herencia multigenica con fuertes interacciones con el ambiente.

Las RILs son generadas por autocruces de una F1 por al menos 8 generaciones. Cada individuo de la población posee múltiples introgresiones y las líneas homocigotas se utilizan en el análisis de QTLs.

Otro tipo de población son las ILs, las cuales son obtenidas por repetidos retrocruzamientos y extensivo genotipado. La población de ILs consiste en un pequeño número de fragmentos genómicos de un padre donador sobre el fondo genético de otro. Las plantas homocigotas para cada uno de los locus se utilizan en los análisis de QTLs mediante el mapeo de marcadores moleculares que delimitan cada locus y datos fenotípicos para cada una de las ILs.

Para las RILs la posición del evento de recombinación es fijo, por lo tanto puede incrementarse mediante el agregado de más líneas. En contraste para las ILs la población puede ser mejorada minimizando el tamaño de la introgresion (Alseekh y col. 2013). Aunque superficialmente las ILs y RILs parecen similares, estas difieren totalmente en cuanto a su uso dado que la frecuencia de recombinación es considerablemente mayor para el mismo tamaño de población en las RILs en comparación a las ILs. Dada una determinada región genómica su proporción esta representada de igual forma por ambos parentales genómicos dentro de la población, con lo cual la replicación de líneas individuales no es necesaria.

Las ILs contienen una única introgresion por línea y son más indicadas para identificar QTLs de pequeño efecto. El hecho de que las RILs posean un mosaico de genomas linqueado estos materiales son los más apropiados para el estudio de interacciones epistaticas. La ventaja de las RILs frente a las DHs y es la alta tasa de recombinación dentro de la población (Keurentjes y Fernie 2011).

Este tipo de recursos genéticos son muy poderosos para exponer los determinantes genéticos subyacentes a los caracteres cuantitativos, ya que las variaciones fenotípicas que se observan en las ILs respecto del parental recurrente pueden ser atribuidas principalmente al segmento introgresado y/o a las interacciones de este con el resto del genoma.

Las ILs se presentan como una herramienta interesante para explorar los determinantes genéticos subyacentes a los caracteres cuantitativos, ya que son determinadas las variaciones fenotípicas de las IL respecto del parental cultivado (M82) principalmente por el segmento introgresado correspondiente. Más aún, el uso de este material facilita la incorporación de alelos salvajes a cultivares elite que permiten el mejoramiento del cultivo sin recurrir a cultivares transgénicos.

Con el fin de identificar *loci* para caracteres de interés, estas ILs han sido ampliamente caracterizadas mediante el mapeo de más de 2000 QTLs (Lippman y col. 2007). Entre estos compuestos se encuentran metabolitos del metabolismo primario (Schauer y col. 2008), secundario (Alseekh y col. 2015), volátiles (Tieman y col. 2006) y acil azúcares (Schilmiller 2010). El éxito de esta población llevó a que se resecuenciaran del parental *S. Pennnelli*. Más recientemente nuevos *backcrosses* de la línea de introgresión dieron lugar a una población de BILs (del inglés, Backcrosses introgresión lines) esta a sido utilizada para identificar genes relacionados a los acil azúcares y otras diferencias entre metabolitos de *S. Lycopersicum* y *S. pennelli*. Otras poblaciones de ILs que se crearon son: *S. Lycopersicum* x *S. chmielenskii* (Ballester y col. 2016, Do PT 2010), *S. Lycopersicum* x *S. Pimpinellifolium* (Barrantes y col 2016). Inclusive se han construido ILs entre distintos cultivares que tienen sus ventajas (Chai y col. 2006, Zanon y col 2009).

Uno de los principales intereses de la biología es la identificación de relaciones causales entre fenotipos y polimorfismos dentro de especies. La posibilidad de predecir un carácter agronómico como la tasa de crecimiento o rendimiento, requiere de entender los *loci* que subyacen a estos caracteres así como también su arquitectura. El abordaje básico en GWAS es evaluar asociaciones entre fenotipos de interés y cada marcador genotipado a lo largo de cada uno de los individuos estudiados.

Los comienzos de los análisis de GWAS (del inglés, Genome Wide associations) fueron inicialmente en el desarrollo o uso de la genética médica (Zeng y col. 2015). En plantas la implementación comenzó en *Arabidopsis* y dentro de los estudios de metabolismo asociados a defensa, en glucosinolato (Kliebenstein y col. 2001); enzimas del metabolismo primario y secundario (Mitchell-Olds y col. 1998), oligosacáridos solubles (Bentsink y col. 2000) y

eficiencia en el uso del nitrógeno (Loudet 2003). Más recientemente una serie de experimentos analizaron el metabolismo primario de un set de 94 ecotipos lo cual permitió llegar a importantes conclusiones acerca de las regulaciones genéticas relacionadas al crecimiento de estas plantas (Kleessen y col 2012, 2014). Estos estudios sugieren que el almidón y la proteína son los mayores determinantes de crecimiento. Más aun, modelos matemáticos de perfiles metabólicos y características de crecimiento revelan que el fenotipo metabólico muestra un patrón de robustez a partir de los fenotipos metabólicos lo cual muestra cuan robustos son.

1.3.2 Ventajas, limitaciones y complementación entre el mapeo de QTLs y GWAS

El mapeo de QTLs es una poderosa herramienta para identificar regiones del genoma que co-segregan para un carácter dado en poblaciones F2 de familias de líneas recombinantes. Un ejemplo de esto, son los genes claves que determinan el tiempo de floración los cuales han sido identificados de esta forma en *Arabidopsis* (Alonso-Blanco y col. 1998, Clarke y col. 1995, Kowalski y col. 1994). A pesar de su éxito, el mapeo de QTLs tiene dos limitaciones principales: i) Solo puede ser estudiada la diversidad que segrega proveniente del cruce de los parentales de la F2.ii) La cantidad de recombinaciones que ocurren durante la creación de las líneas recombinantes dan lugar a una limitante en la resolución de mapeo.

La primer limitante, diversidad alélica, puede ser incrementada hasta cierto punto intercruzando múltiples accesiones (diversidad alélica) antes de establecer la población de RIL. Este es el caso de MAGIC (del inglés, Multi-parent Advanced Generation Inter-Cross) y en *Arabidopsis* AMPRIL (en inglés, multi-parent RIL) (Kover y col. 2009, Huang y Col. 2011).

Estos nuevos tipos de poblaciones que involucran varias líneas parentales como MAGIC también se han demostrado útiles en el tomate para mapear QTLs para rasgos de calidad en pequeños intervalos de confianza. Combinado con la secuenciación de las líneas parentales, podría proponerse un acceso directo a polimorfismos putativos bajo los QTL (Pascual y col. 2015).

Recientemente, Ofner y col. (2016) desarrollaron una población de 446 líneas endocriadas de retrocruzamiento (BIL) derivadas después de unas pocas generaciones de retrocruzas de la especie silvestre *S. pennellii* con el tomate cultivado, seguidas por más de siete generaciones de autopolinización. Este material genético fue mapeado con diez mil marcadores de SNP utilizando un conjunto de SNP de genoma amplio y debería constituir una herramienta

fantástica para confirmar y trazar los QTL de los rasgos de calidad del tomate en el futuro cercano.

El inconveniente de la capacidad de resolución de mapeo puede ser mejorado mediante la generación de inter cruces una vez establecida la población RIL, como por ejemplo en generaciones avanzadas de RILs (Balasubramanian y col. 2009). Vale hacer la salvedad de que las frecuencias de recombinación de las poblaciones RIL difieren de las de las poblaciones naturales. Para muchas aplicaciones esto no es una limitante. Pero si lo son, cuando hay análisis de epistasis en los cuales se pueden confundir ya que es limitada la diversidad observada frente a la diversidad natural.

Los análisis de GWAS y mapeo de QTLs son siempre complementarios cuando son conducidos juntos, mitigando de esta forma las limitaciones de uno en otro (Brachi y col 2010, Zhao y col. 2007). Por ejemplo los análisis de GWAS sirven como experimentos iniciales para explorar la arquitectura genética del carácter y a *posteriori* permiten elegir los padres para mapeo de QTLs para luego validar los genes candidatos por mutagénesis, transgénesis, edición génica, etc.

El poder del GWAS para identificar asociaciones verdaderas entre SNPs y caracteres depende de la varianza fenotípica de la población explicada por los SNPs. Esta, es determinada por cuan fuerte son las diferencias fenotípicas de las variantes alélicas y sus frecuencias en la muestra. Es por esto que variantes alélicas raras de baja frecuencia son un problema para los análisis de GWAS (Asimit y col 2010, Gibsom y col. 2011). Estas variantes raras pueden estar asociadas a otras dentro del genoma y que sean no causales “sintéticas” las cuales pueden ser “arrastradas” por el desequilibrio de ligamiento (Gibson y

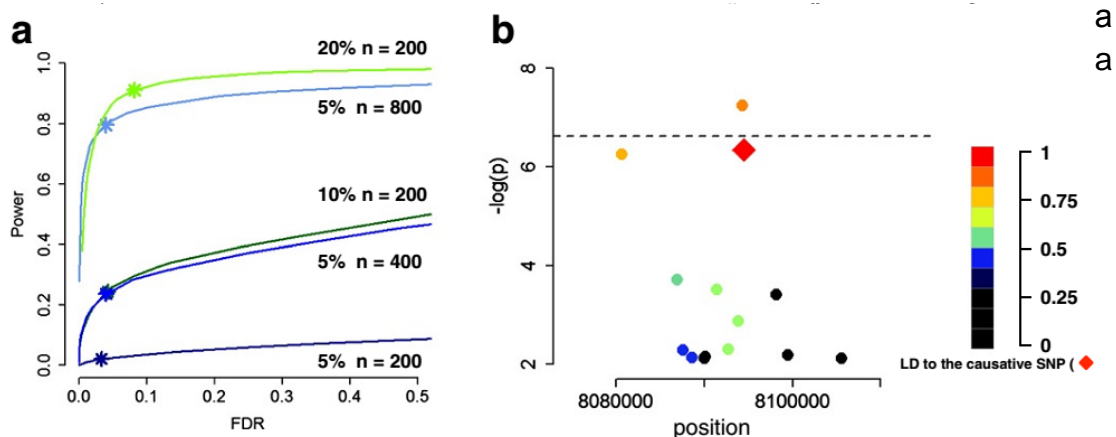


Figura 5. Efecto del fondo genético y la identificación de falsos positivos en Analisis de GWAS a) Ejemplo en la identificación de falsos positivos para un fenotipo ideal. Simulación en la cual un marcador SNP explica 5%, 10%, 20% de la varianza fenotípica (con una heredabilidad cercana a 0,75%). Los asteriscos indican el poder par identificar verdaderos positivos y el FDR o del ingles *False positive rate*. **b)** Ejemplo de una situación particular en la cual un SNP causal, rombo rojo, no es significativo en el Manthattan plot. La escala de desequilibrio de ligamiento esta a la izquierda con una graduación de colores de negro (0) a rojo (1), la línea de corte responde al test de bonferroni para 220000 marcadores

Hasta ahora la mayoría de los análisis desarrollados en *Arabidopsis* tienen solo unos cientos de individuos pero para cientos de caracteres, resultados significativos pueden ser obtenidos con menos de 100 accesiones (Atwell y col. 2010). Esto se debe a que los caracteres considerados en estos trabajos son explicados por unos pocos *loci* que expliquen una gran porción de la varianza fenotípica. La situación es diferente en humanos donde típicamente un gran número de *loci* con pequeño efecto fueron encontrados para la mayoría de los análisis con lo cual, miles de individuos son requeridos en estos casos (Manolio 2009 , Bopmer 2008).

La arquitectura genética en casos en los que son muchos los determinantes y de bajo efecto fueron observados en otros animales (Flint y Mackay 2009) y en maíz (Tian y col. 2011). Esto muestra la existencia de diferentes arquitecturas genéticas de acuerdo al tipo de reproducción. *Loci* con efectos intermedios como el color de ojos y piel han sido estudiados (Sulem y col. 2007, Beleza y col. 2013).

A pesar del éxito de GWAS en *Arabidopsis* muchos caracteres son poligenicos con un pequeño efecto, lo cual incrementa el tamaño muestral para poder incrementar el poder en la identificación de asociaciones significativas. Dado esto, que hay que tener en cuenta para seleccionar los genotipos de la población de mapeo?

1.3.3 Criterio de evaluación de analisis de GWAS

El más básico e informativo de los abordajes es corregir por multiples test (usualmente el test de Bonferroni es muy usado. La inspección de QQplots y Manhattan plots permiten evidenciar efectos de inflexión. Ambos abordajes QQplots y Manhattan plots dan exploración global de los datos. Esto permite ver si hay muchas o pocas asociaciones significativas en relación a las esperadas. La principal limitación de esta corrección es asumir que cada SNP es independiente uno de otro. La estructura en la población de *Arabidopsis* claramente viola este supuesto y hace que muchas falsas asociaciones persistan gracias al LD.

El criterio más informativo es ver la proporción de falsos positivos y negativos simulada la cual tiene en cuenta el real desequilibrio de ligamiento de la población. Este es expresado como FDR “false discovery rate”. Teniendo en cuenta el propósito de estudio uno considera altas o bajas tasas para algunos proyectos. Por ejemplo si se está estudiando la arquitectura de genética de la

población o bajas tasas de falsos positivos para la identificación de genes candidatos a validar con otros estudios.

Otra característica a evaluar es la robustez de los GWAS es calculando el enriquecimiento *a priori* de en función del p-valor de corte de significancia del análisis. Evaluando genes anotados dentro de categorías funcionales (Kawakatsu y col. 2016). Un gráfico representativo de esta estrategia es graficar el enriquecimiento de estos genes en función del p-valor del análisis. Esta función debería ser creciente hasta determinado p-valor en el cual llega a un *plato* y luego decrece. Presentando un máximo en el enriquecimiento de genes identificados para un p-Valor de corte en el cual no se determinan más asociaciones a genes relacionados al metabolismo de la vitamina E

1.4 Bases genéticas del metabolismo

1.4.1 Interacción génica y alélica. Mecanismos epigenéticos asociados

El problema o limitación fundamental del GWAS es identificar variantes de poco efecto o visto de otra forma, la limitante es que la muestra sea pequeña y heterogénea. La inclusión de todos los datos de secuencia y incrementar el tamaño de la muestra podría en teoría sobre llevar esta limitación sin embargo también está sujeto a que el estado epigenético también contribuye a la heredabilidad (Johannes y col. 2009, Bergelson y Roux 2011).

Dada la epigenética como una variable que modula la conexión entre SNPs y el carácter sería importante considerarla a esta como información biológica. Esto podría ser implementado integrando la genética, epigenética, expresión génica y fenotipo dentro de un solo modelo. Estimar los parámetros de este modelo requeriría un sustancial tamaño de muestra.

El epigenoma orchestra la accesibilidad al genoma, su funcionalidad. Impacta en la transcripción y en el fenotipo lo cual podría contribuir a la adaptación. Aunque las bases de la variación de la metilación son realmente complejas. El origen geográfico es uno de los mejores predictores de la metilación a nivel global del genoma (Taiji Kawakatsu 2016).

La metilación de citosinas y la modificación de histonas, son marcas epigenéticas que tiene efecto sobre la actividad de los elementos transponibles, la transcripción de genes y formación de la heterocromatina.

En plantas la metilación de DNA ocurre en contextos simétricos CG and CHG (H = C, A, o T) y en contextos asimétricos CHH (Law y Jacobsen, 2010). Los sitios de metilación CG se propagan durante la replicación de DNA mientras que la metilación de sitios CHG y CHH se mantienen by self-reinforcing loops (Kawashima y Berger, 2014) (figura 6).

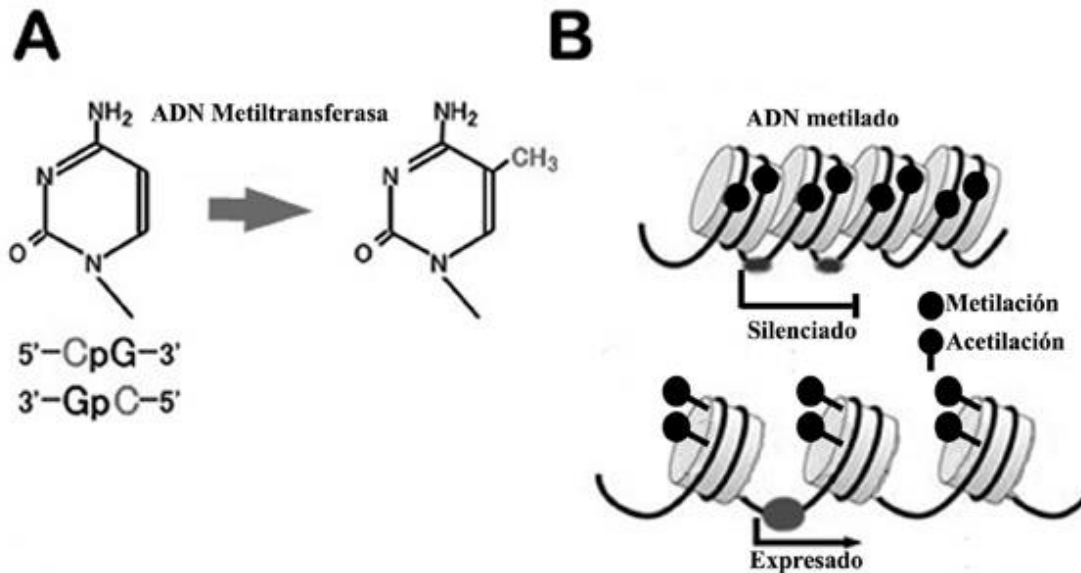


Figura 6. Contextos genómicos de la metilación del DNA **A)** Agregado del grupo metilo al C 5' de la citosina. **B)** Pasaje del estado de cromatina condensada y silenciada, por desatilación de histonas e hipermetilación del ADN, al estado descondensado y activado en el que la cromatina está desmetilada y acetilada. Tomado de Masuelli RW y Marfil CF 2011.

Las bases genéticas asociadas a las variaciones de metilación del DNA incluyen: I) variaciones estructurales como la inserción y deleción (Indels) de elementos transponibles. II) rearrreglos cromosómicos. III) mutaciones en factores de la metilación (Pecinka y col. 2013). IV) Factores ambientales se encuentran la temperatura y otros estreses (Downen y col. 2012, Dubin y col. 2015, Secco y col. 2015).

Los estudios de plantas han proporcionado numerosas contribuciones fundamentales al campo de la epigenética. En este sentido, la primera evidencia de comportamientos epigenéticos no mendelianos se describe sobre la interacción entre dos alelos de un solo locus en determinados heterocigotos, que resulta en un cambio heredable de la expresión de un gen de uno de los alelos implicados, un fenómeno conocido como para-mutación observado en tomate y maíz (Arteaga-Vázquez y Chandler 2010). La impronta parental de genes individuales, es decir, la expresión de un único alelo de origen materno o paterno, que es un proceso cuya mala regulación es la base de múltiples trastornos genéticos en humanos (Zoghbi y Beaudet 2014), se observó por

primera vez en maíz (Alleman y Doctor 2000). La repetida ocurrencia de individuos con simetría de flor alterada (de "peloria"), descrita por primera vez por Carl von Linne en el siglo XVIII, es causada por la formación de un epialelo silenciado cuya secuencia de ADN es idéntica a los alelos expresados del gen (Cubas y col. 1999). Los epialelos pueden afectar cambios de desarrollo; el gen FWA en *Arabidopsis* es un claro ejemplo en el cual el silenciamiento en algunos ecotipos naturales retrasa la floración (Soppe y col. 2000). Además, un trabajo pionero sobre elementos transponibles en el maíz realizado por Barbara McClintock y otros en la década de 1940 reveló numerosos vínculos entre el comportamiento genético y la regulación epigénica (Lisch 2009) determinada por su capacidad de movilizarse favoreciendo la aparición de determinados fenotipos. De hecho, los transposones proporcionan una base para establecer nuevos elementos reguladores y modificaciones epigenéticas en los genomas eucariotas.

A medida que la tecnología transgénica en plantas como el tabaco, la petunia y *Arabidopsis* se convirtió en rutina a finales de la década de 1980, se hizo evidente que la expresión del transgén era impredecible (Wassenegger 2002, Matzke y Matzke 2004). Esto condujo a constatar que el silenciamiento génico dependiente de homología de secuencia que ocurre cuando se introducen transgenes que llevan secuencias idénticas a genes ya presentes en el genoma, con un "enhanced" de ARN mensajero (ARNm) (silenciamiento génico postranscripcional, PTGS) o la represión de la transcripción (silenciamiento génico transcripcional, TGS) son la base molecular para silenciar. Un ejemplo sorprendente de PTGS ocurrió cuando se intentó la sobreexpresión del gen que codifica la chalcona sintasa (CHS), un regulador de la pigmentación de las flores en la petunia, e inesperadamente resultó en flores variegadas o incluso completamente blancas en lugar de flores púrpuras más oscuras, fenómeno denominado "co-supresión". La falta de pigmentación en los sectores blancos se debió al silenciamiento genético coordinado tanto del transgén CHS como del gen CHS endógeno (Jorgensen y col. 2006). PTGS es una forma de interferencia de ARN (ARNi), descubierta posteriormente en *Caenorhabditis elegans* y otros organismos.

A mediados de la década de 1990, se habían establecido vínculos entre el PTGS y la resistencia a los virus. Se demostró que los mecanismos epigenéticos que implican moléculas de ARN protegen naturalmente a las plantas de la replicación incontrolada de virus, que pueden ser inductores y como blancos de PTGS. Pronto se descubrió que estos mecanismos podrían explotarse para disminuir la expresión de genes de las plantas mediante la construcción de vectores virales que contienen secuencias de genes vegetales, aprovechando así la vía de silenciamiento génico inducido por el virus (VIGS, Baulcombe y Dean 2014; Senthil-Kumar y Mysore 2011). Otro mecanismo

descubierto en plantas infectadas con viroides es la especificación guiada por ARN de la metilación del ADN (metilación del ADN dirigida por ARN, RdDM) y otras marcas heterocromáticas en regiones de ADN homólogas (Wassenegger y col. 1994). Se demostró que la metilación de la citosina se dirigía al promotor, dando como resultado el silenciamiento transcripcional dependiente de la homología del gen correspondiente (Mette y col. 2000) (figura 7).

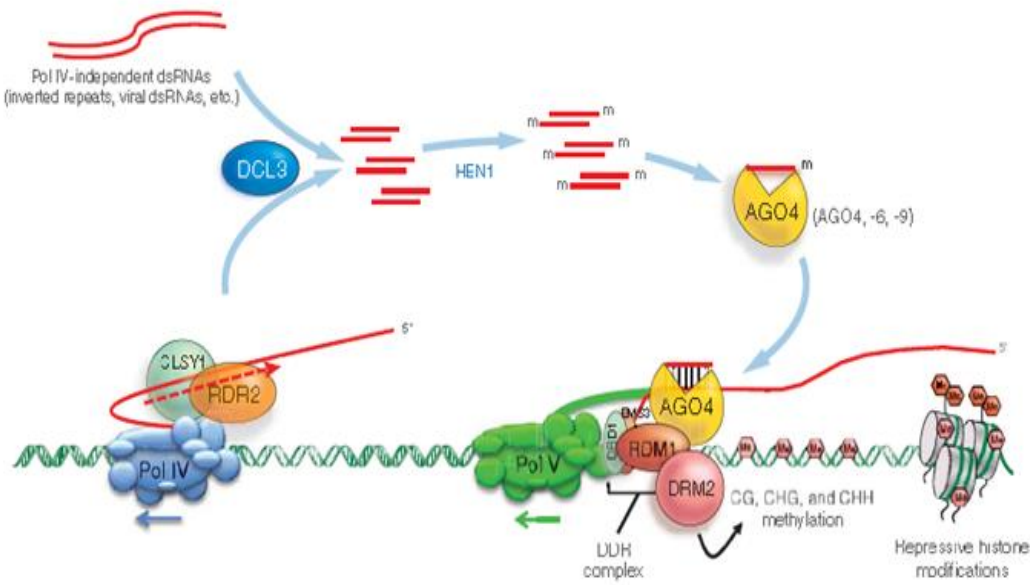


Figura7. RdDM. Los retrotransposones, virus, transgenes o genes repetitivos están sujetos a silenciamiento por RdDM. Se cree que RNA Pol IV genera RNA monocatenarios (extremo inferior izquierdo) que sirven como plantillas para la RNA polimerasa dependiente de RNA, RDR2. Los dsRNAs resultantes se cortan en mediante DCL3 para generar dúplex de siRNA de 24 nt que luego se metilan por HEN1 y se cargan en AGO4, o sus miembros familiares estrechamente relacionados, AGO6 o AGO9. dsRNA derivado de transgenes repetidos invertidos o virus (arriba a la izquierda) puede eludir la necesidad de RNA Pol IV y RDR2. Los complejos AGO-siRNA se reclutan en sus sitios de acción uniéndose a transcripciones generadas por RNA Pol V, así como mediante interacciones físicas con el dominio carboxi-terminal de la subunidad más grande de RNA Pol V. En algunos loci, se cree que RNA Pol II sustituye RNA Pol V para la producción de transcripciones de andamios a las que se unen los complejos AGO-siRNA. El complejo DDR (DRD1, DMS3 y RDM1) permite la transcripción de RNA Pol V. La subunidad RDM1 del complejo DDR también interactúa con AGO4 y la citosina metiltransferasa de novo, DRM2, lo que podría servir como un puente que recluta DRM2 a sitios de transcripción ARN Pol V.

Otra contribución destacada de la investigación epigenética de plantas fue la aplicación de la secuenciación profunda para realizar análisis genómicos (Zhang y col. 2006), incluido el primer análisis de metilación genómica en resolución de una sola base (Cokus y col. 2008; Lister y col. 2008). Estos estudios requirieron el desarrollo de herramientas bioinformáticas que desde entonces se han aplicado ampliamente para la investigación biomédica

1.4.2 Heterosis y epistasia

La heterosis, también conocida como vigor híbrido, se define como la capacidad de los híbridos para superar las características presentadas por sus parentales. Evidencias obtenidas para diversas especies permiten hipotetizar que alteraciones en el estado epigenético, mediadas por cambios en el patrón de metilación del ADN tengan un papel en la determinación del vigor híbrido.

Es sabido que cuando dos líneas con alto grado de homocigosis y distantes genéticamente, son utilizadas como parentales para la obtención de híbridos, como producto del “*estrés genómico*” resultante de la hibridación, ocurren interacciones de ciertos factores genéticos del tipo *cis* y *trans* con la maquinaria de metilación del ADN y los componentes necesarios para mantener y regular el estado epigenómico (por ejemplo, el nivel de metilación del ADN). En los últimos años se han reportado evidencias que indican que mecanismos epigenéticos relacionados a la metilación de la cromatina son fundamentales en esta re-programación (Ni y col. 2009, Ha y col. 2009; He y col. 2010; Groszmann y col. 2011, Shen y col. 2012, Chen 2013) determinando finalmente la regulación génica diferencial que se observa en los híbridos (Cortijo y col. 2014).

Goff (2011) propone un modelo de heterosis centrado en la eficiencia del uso de energía por parte de las células híbridas, integrando algunos de los mecanismos epigenéticos mencionados anteriormente. Según este modelo, en los híbridos diversos procesos celulares actuarían coordinadamente sintetizando ARNm y proteínas de alta calidad y conservando energía para destinarla a mayores tasas de crecimiento (Hauben y col. 2009, Keene 2007, Komili y Silver 2008).

La heterosis, es un fenómeno que viene siendo estudiado y explotado desde hace más de un siglo. Ya en 1876, Charles Darwin en su libro “*The effects of cross and self fertilisation in the vegetable kingdom*”, después de diversas observaciones de cruzamientos en más de 60 especies, concluía que la endogamia resulta generalmente deletérea y, por el contrario, la exogamia genera efectos benéficos para el *fitness* de la descendencia. No mucho tiempo después, en 1908, Shull y East describieron por primera vez el fenómeno de la depresión por endogamia y el vigor híbrido en maíz. Después de múltiples generaciones de autofecundación obtenían líneas homocigotas que mostraban una fuerte depresión endogámica y al ser cruzadas producían híbridos de alto rendimiento. Como resultado de estos estudios se acuñó el término “heterosis” como simplificación de la expresión original en inglés “*stimulus of heterozygosity*” (Goff 2011). De esta manera, el vigor híbrido o “grado de

heterosis” sería proporcional a las diferencias genéticas entre las dos líneas parentales aumentando el vigor de la progenie a medida que aumenta la distancia genética entre los progenitores (Chen 2010). Estudios realizados por Ashby y colaboradores (1930, 1932) permitieron proponer que la heterosis es un fenómeno básico y evolutivamente conservado, donde los individuos que la presentan exhiben aceleradas tasas de crecimiento (división celular) sin ningún impacto evidente en el patrón de desarrollo (Goff 2011, Jiang y col. 2013).

Específicamente en plantas, la heterosis es definida como un proceso complejo multigénico que puede ser extrapolado como la suma total de numerosos caracteres fisiológicos y fenotípicos tales como la tasa de crecimiento vegetativo, tiempo a floración y los distintos componentes del rendimiento (por ejemplo: número de inflorescencias, flores por inflorescencia, número y peso de frutos o granos). En el caso de la resistencia a estreses ambientales, todos estos factores contribuyen en grados diferenciales a los fenotipos transgresivos de los híbridos (Lippman y Zamir 2007, Baranwal y col. 2012, Chen 2013).

Hasta la fecha, se han propuesto diversas teorías para explicar la heterosis: 1) la de la sobredominancia propuesta por East (1908) y Shull (1908) sostiene que la interacción alélica que acontece a partir del cruzamiento de líneas endocriadas, resulta en una “función superior” (por ejemplo: actividad enzimática o regulatoria más eficiente) al estado homocigota en un mismo *locus*. 2) La teoría de la dominancia propuesta por Davenport (1908) postula que las líneas endocriadas presentan alelos recesivos deletéreos para diversos *loci*, los cuales, en los híbridos, serían complementados por los alelos superiores dominantes contribuidos por el otro parental.

Esta teoría fue ampliada por Jones (1925) explicando la heterosis como el resultado de un grupo de elementos dominantes promotores de crecimiento en vez del número de caracteres heterocigotos; esta teoría es comúnmente identificada como la complementación de alelos recesivos deletéreos (Goff, 2011). 3) Estudios recientes realizados en tomate sugirieron un nuevo modelo que fue denominado pseudo-dominancia. En este caso la heterosis está asociada a la complementación de dos o más caracteres recesivos y dominantes ligados en repulsión, donde los alelos dominante y recesivo se encuentran en diferentes cromosomas homólogos (Semel y col. 2006). 4) Finalmente, el efecto de la epistasis, que explica la interacción entre alelos de diferentes *loci*, fue propuesta por Powers (1944) como otro posible fenómeno para explicar el vigor híbrido.

Reportes recientes que usan abordajes a escala genómica permiten asegurar que los mecanismos anteriormente mencionados, efectivamente operan en

híbridos. Sin embargo, sugieren que no son suficientes para explicar todas las observaciones experimentales encontradas. Por ejemplo, estos modelos genéticos no explican completamente la heterosis observada en plantas poliploides ya que las dosis génicas y alélicas pueden jugar un papel mucho más importante que la complementación alélica o sus interacciones. Así, los cambios en la expresión génica dependiente del “dosaje génico” pueden ser mucho más profundos que la alteración en las interacciones alélicas (Chen, 2010, 2013). Consecuentemente, los mecanismos moleculares que determinan este fenómeno no han sido identificados y este desconocimiento impacta negativamente en la eficacia y dirección de los programas de mejoramiento genético.

1.5 Definición del problema, hipótesis y objetivos

La falta de inversiones económicas en programas de mejoramiento genético sobre caracteres nutricionales complejos como la vitamina E en cultivo de tomate; se debe no solo a que estos han sido delegados frente a otros caracteres relacionados al rendimiento sino también a la falta de conocimiento de su regulación a nivel molecular y a su uso potencial para ser aplicado como herramienta. Para que esto suceda, el primer paso es dilucidar acerca del entendimiento de la regulación de la vitamina E y el rol de los factores genéticos, ambientales y epigenéticos que la determinan.

La mayoría de los estudios relacionados a antioxidantes en plantas se ha realizado principalmente en VTC (Conklin y col. 2000) y licopenos (Auldrige y col. 2006, Simkin y col. 2004; Zou y col. 2005), carotenoides (Tan y col. 2003), flavonoides y ácidos fenólicos (Giovinazzo y col. 2005; Van der Rest y col. 2006). Contrariamente a las numerosas publicaciones mencionadas acerca de estos compuestos, el conocimiento acerca de la regulación de la biosíntesis de tococromanos en frutos de tomate es escaso menos aún la información “codificada” en el epigenoma y los mecanismos moleculares que operan en su regulación.

Se ha estimado que el carácter presenta baja heredabilidad (h^2) en tomate (Schauer y col. 2008) y más recientemente que la regulación de la biosíntesis está relacionada a proteínas de la maquinaria fotosintética que forman canales metabólicos en las membranas tilacoidales de *A. thaliana* (Tanaka y col. 2010).

Hasta el momento los únicos reportes centrados en el metabolismo de tocoferoles en tomate, permitieron identificar y mapear 16 QTL como candidatos en la variación del contenido de VTE: genes involucrados en las vías del MEP, SK y la vía principal del tocoferol (Almeida y col. 2011).

Otro estudio mostró que la biosíntesis de tocoferol está regulada temporal y espacialmente manteniendo un balance constante del tocoferol total. Esto fue posible a través del análisis de ARNm de 47 genes involucrados en las vías biosintéticas mencionadas previamente en hojas y frutos en distintos estadios de desarrollo (Quadrana 2013). Además, el análisis de las regiones promotoras de estos genes sugirió que los mismos son co-regulados por elementos comunes que actúan en cis (Quadrana y col. 2013).

Por otra parte el silenciamiento específico de *VTE3(1)* y *VTE4* en frutos (Quadrana y col. 2011, Quadrana y col. 2014), sugiere que la regulación de la vía puede explicarse, al menos en parte, a nivel de la actividad transcripcional de estos genes (Quadrana y col. 201).

Más recientemente mediante el mapeo fino en la IL9-2-6 del mayor QTL para los contenidos de α - y tocoferol total permitieron identificar al gen *VTE3(1)* como el responsable de este QTL. *VTE3(1)* codifica para una dimetil-fitilquinol metil transferasa, la cual se encuentra en el cromosoma 9. Además este trabajo mostró que la regulación de su expresión diferencial entre el alelo silvestre y domesticado se debe a la presencia de un elemento transponible el cual se encuentra en el alelo domesticado y aquí se acumulan mayor cantidad de SiRNA direccionando la metilación de su región promotora mediante posible mecanismo de RdDM (Wassenegger y col. 1994) e inhibiendo su expresión, como consecuencia una disminución en la acumulación de VTE en frutos de tomate (Quadrana y col. 2014).

Además, estos niveles revierten en respuesta a condiciones ambientales. Esto fue evidenciado al comparar los niveles de metilación que presenta la región proximal de este promotor tanto en frutos maduros cultivados dentro de un invernáculo y a campo realizado en la Provincia de Mendoza (Argentina 32° 50'S, 68° 52'W and 900 masl). En este mismo trabajo también se exploró la variabilidad epialelica de *VTE3(1)* en una población de landraces bajo un mismo ambiente, encontrando una plétora de epialelos los cuales presentan distintos niveles de expresión/metilación para la región del elemento SINE.

Por otro lado, un parálogo de aquel *VTE3(2)* mapeado en el cromosoma 3, se expresa preferencialmente en hojas (Quadrana 2013). A su vez, resultados disponibles en bases de datos (<http://ted.bti.cornell.edu/>), sugieren que *VTE3(2)* co-localiza con otro QTL para variaciones en los contenidos y composición de VTE.

Aún cuando las evidencias resumidas más arriba sugieren que las variantes alélicas *S. pennellii* de los genes *VTE(s)* podrían explicar las variaciones en los contenidos de VTE de los frutos, el efecto de otros loci involucrados en mecanismos aún no descritos no puede descartarse.

Los problemas planteados guiaron a abordar el objetivo general de esta tesis:

1.5.1 Objetivo General:

Generar nuevo conocimiento y herramientas acerca de los mecanismos epi/genéticos que regulan los contenidos de vitamina E en frutos de tomates para su aplicación en programas de mejoramiento enfocados a aumentar su calidad nutricional.

A su vez, sobre la base de los antecedentes previamente descritos se planteó la siguiente hipótesis de trabajo:

Hipótesis 1: La biosíntesis de VTE es regulada diferencialmente a nivel de las enzimas involucradas. Los genes *VTE3(1)* y *VTE3(2)* localizados sobre QTLs para los contenidos de tocoferol tienen un efecto en los niveles de VTE tanto en fruto como en hoja de tomate. Teniendo en cuenta que ambos parálogos de *VTE3* poseen patrones de expresión disímiles en los diferentes tejidos y durante el desarrollo del fruto de tomate y que ambos se encuentran asociados a la biosíntesis de tocoferol, es posible suponer la existencia de interacciones entre estos genes. La comprensión de las mismas permitirá implementar mejores estrategias de mejoramiento de este carácter.

Hipótesis 2: Los contenidos de tocoferol tanto en hojas como en frutos están asociados al grado de metilación de los promotores de los genes *VTE3(1)* y *VTE3(2)* los cuales determinan la expresión de estos genes.

Hipótesis 3: Existe una plétora de epialélos naturales para ambos parálogos *VTE3(1y2)*, cada uno de estos poseen un comportamiento particular determinado por su *background* genético así como también de su interacción con el ambiente. Esta interacción tiene un efecto sobre el nivel de metilación de sus promotores, en consecuencia su nivel de expresión y el contenido de tocoferol tanto en hojas como en fruto

Hipótesis 4: El efecto de otros loci involucrados en mecanismos aún no descritos no puede descartarse. En este sentido, con el fin de explorar una mayor variabilidad genética e identificar nuevos loci candidatos asociados a los contenidos de VTE hemos abordado dos análisis complementarios de GWAS y mapeo de QTLs, para la identificación de nuevos genes candidatos

Para poner a prueba estas hipótesis se plantearon 5 objetivos específicos que se desarrollan en los capítulos II, III, IV y V de esta tesis:

1.5.2 Objetivos específicos

- 1)** Analizar el efecto fenotípico sobre individuos F1 y F2 derivados de un cruzamiento entre la IL conteniendo el alelo silvestre de *VTE3(1)*(IL9-2-6-1) X la IL conteniendo el alelo silvestre de *VTE3(2)* (IL3-1) en ambientes contrastantes.
- 2)** Analizar los patrones de metilación y los niveles de expresión de los genes *VTE3(1y2)* en las líneas recombinantes productos del cruzamiento entre IL9-2-6-1 x IL3-1 en ambientes contrastantes.
- 3)** Evaluar la variabilidad epialélica de *VTE3(1y2)* en tejidos de frutos y hojas mediante la evaluación de los patrones de metilación, expresión y contenidos de vitamina E en una población de Landraces en ambientes contrastantes.
- 4)** Explorar otros determinantes genéticos de la variabilidad natural de VTE mediante análisis un GWAS de una población de 150 genotipos y mapeo por asociación de una población de 124 líneas F7 multiparental (MAGIC-POP).

Los resultados obtenidos en los objetivos **1)** y **2)** proveerán conocimiento, modelo sobre un aspectos poco explorados, como es la regulación del metabolismo a partir de la metilación de las regiones promotoras y la expresión de genes codificantes de enzimas de la vía de síntesis de vitamina E. Además de la existencia o no de interacciones epistáticas entre genes parálogos y su efecto sobre la síntesis de VTE así como también el rol del ambiente sobre la regulación de estos genes.

La comprensión de estos mecanismos permitirá diseñar mejores estrategias a ser utilizadas en programas de mejoramiento tendientes a seleccionar cultivares de tomate con mayores contenidos de vitamina E.

Asimismo las herramientas desarrolladas a partir del tercer objetivo serán de fundamental relevancia para entender la variabilidad natural de los mecanismos de regulación de la biosíntesis de vitamina E.

Adicionalmente el desarrollo de dos estrategias complementarias como GWAS y mapeo por asociación en distintas poblaciones permitirá el estudio de la arquitectura genética del carácter e identificar nuevos *loci* o genes candidatos a ser validados mediante transgénesis o edición génica para su inmediata aplicación en programas de mejoramiento destinados a aumentar los contenidos de Vitamina E.

CAPITULO II- Búsqueda de nuevos determinantes genéticos asociados a los contenidos de VTE, abordando dos estrategias complementarias GWAS y mapeo de QTLs

II.1.1 Introducción a las poblaciones estudiadas de GWAS y mapeo de QTL

Las metodologías de GWAS y mapeo por asociación, hoy por hoy dan cuenta de una considerable cantidad de casos exitosos identificando genes candidatos relacionados a caracteres poligénicos varios de ellos validados a través de otras técnicas experimentales (Curtin y col. 2017) e inclusive asociados contenidos de VTE (Sauvage y col. 2014, Hong Wang 2018, Diepenbrock y col. 2017).

Este capítulo se propone presentar la complementación ambos análisis y mostrar como su abordaje permite detectar nuevos *loci* asociados a la VTE, estudiando a su vez, la disección de su arquitectura genica. Con este proposito, se analizaron los contenidos de α - β - γ - δ -tocoferol, α -tocotrienol y tocoferol total en frutos maduros para 150 accesiones de una la colección de GWAS sometidas a dos ambientes contrastantes, de estrés hidrico y control. Complementariamente se realizo el mapeo por asociación de estos mismos caracteres, sobre 124 líneas F7 proveniente de 8 parentales lo cual permitio una mayor resolución y análisis de los haplotipos. Más detalles de la descripción de los materiales se puede encontrar en los incisos, V.1.1 y V1.2 correspondientes a la construcción de la colección de GWAS y construcción de la F7 Multiparental, así mismo las condiciones de los ensayos y datos genéticos de ambas poblaciones.

II.1.2 Analisis de estructura génica y LD de ambas poblaciones de estudio

La disección de caracteres cuantitativos es un desafio especialmente para especies como el tomate, con su poca diversidad genetica y fuerte estructura de población (Blanca y col. 2015, Ranc y col. 2008). La falta de variación intraespecífica es una limitante en tomate y ha conducido al estudio de progenies interespecificas, entre especies relacionadas. La varicion intraespecífica ha sido pobremente estudiada.

La población de MAGICPOP es la primer población multiparental de tomate, con un mapa de alta densidad para el estudio de la variabilidad intraespecífica. Esta, permite el estudio de nuevas combinaciones alélicas, además de la identificación de el/los polimorfismos causales del mapeo de QTLs.

El estudio de la estructura en la población para MAGIC-POP a partir de los 1345 marcadores SNPs incluidos en su mapa genético de acuerdo al test de Evanno y col. 2005, indico ausencia de subgrupos. En contraste la colección GWAS si presenta estructura poblacional (figura 8) como fue reportado en Albert 2017 .

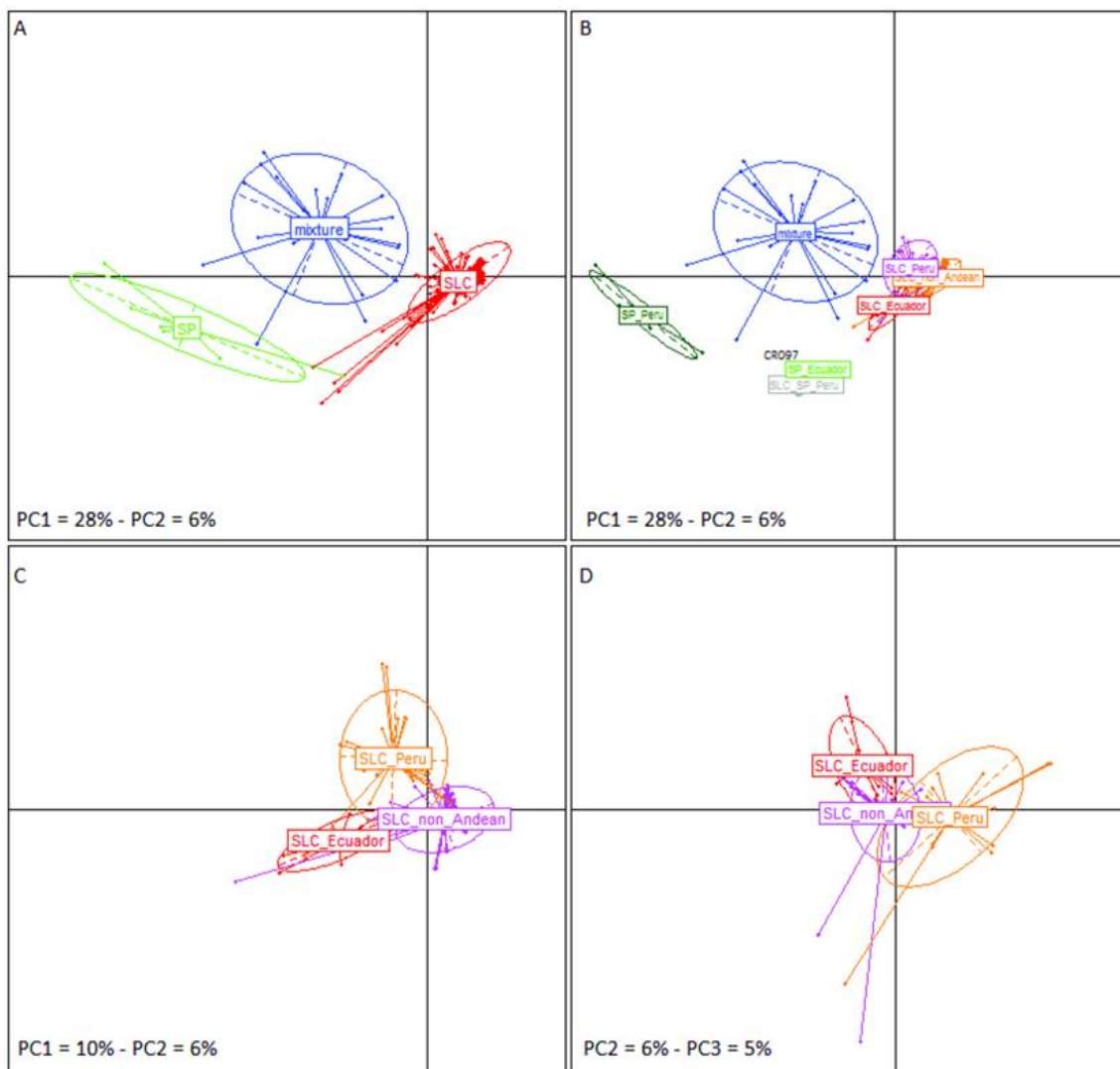


Figura 8. Estructura genética de la colección de GWAS. A) Analisis de coordenadas principales a partir de los 6000 SNPs del análisis de GWAS. En verde se distingue el grupo correspondiente a los individuos *Solanum pimpinellifolium* (SP); en rojo aquellos comprendidos dentro de *S. Lycopersicum* var. *Cerasiforme* y un tercer grupo genético fue determinado como “mixto” en color azul, el cual comprende mas de una especie ver materiales y metodos. **B)** El analisis de la población permite distinguir subgrupos de acuerdo a su lugar de origen (non. Andean SLC (violeta), SLC Peru (naranja), SLC Ecuador (rojo), SLC-SP Peru (gris claro), SP Peru (verde oscuro), SP Ecuador (verde claro) y la accesión CR097, la cual quedo indefinda. **C y D)** El análisis reduce la accesión SLC de acuerdo a su sub-grupo. Las elipses de dispersión comprenden el 67% de los individuos para cada grupo. Adaptado de Albert v col. 2017

En efecto las especies pueden ser divididas en dos grandes grupos: *S. lycopersicum* y *S. lycopersicum var. cerasiforme*. La MAGIC-pop fue construida tomando como líneas parentales a cuatro accesiones de cada uno de estos grupos. De grupo de *S. lycopersicum* con frutos alargados Levovil, Stupicke, LA0147 y Ferum, la contraparte de *S. lycopersicum var. cerasiforme* con frutos pequeños Cervil, Criollo, Plovdiv, LA1420.

Es por esto que la población muestra un amplio rango de fenotipos, sin dejar subgrupos. El diseño utilizado para desarrollar esta población mezclando los ocho genomas parentales permitió además un balance en las frecuencias alélicas a lo largo de todo el genoma (figura 9). En contraste a otras MAGIC-POPs (Huang y col. 2012b), donde la segregación fue distorsionada inclusive algunas líneas fundadoras llevan introgresiones de especies salvajes distantes (Causse et al., 2013). La excepción en cuanto a la homogeneidad de frecuencias se presenta en los cromosomas 5 y 11 donde las 3 líneas parentales Stupicke PR, LA0147 y Levovil fueron difíciles de discriminar.

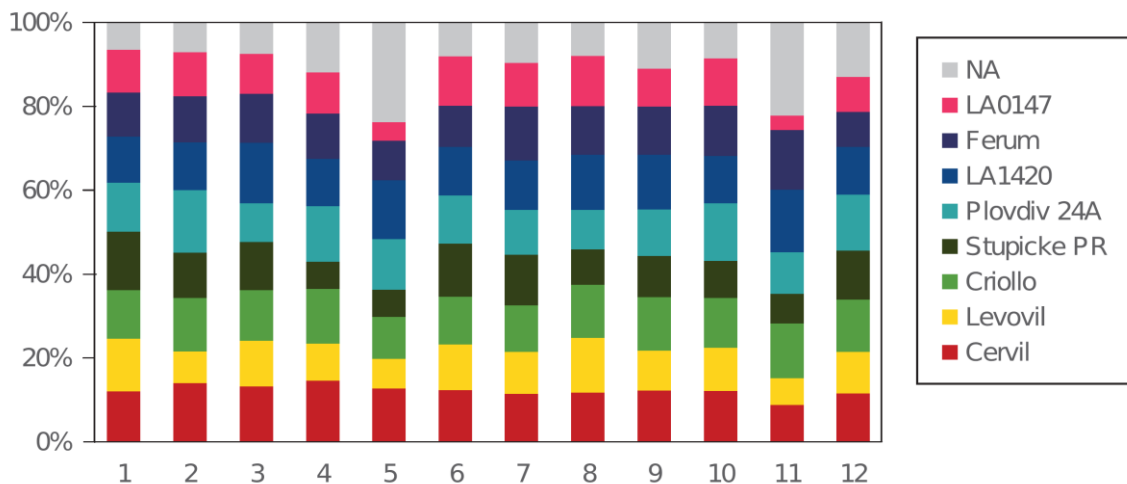


Figura 9. Efecto fundador de cada uno de las 8 líneas parentales sobre la población F7.

En esta figura se puede observar un balance de frecuencias alélicas homogéneas para cada uno de los 12 cromosomas. En gris claro, está indicada la proporción del genoma a la cual no ha sido identificada la procedencia alélica. En fucsia corresponde a LA0147, Azul oscuro para Ferum, Azul marino LA1420, celeste Plovdiv 24, verde oscuro Stupicke PR, verde claro Criollo, amarillo Levovil y finalmente Cervil

El desequilibrio de ligamiento así como el tamaño específico y estructura genómica de los centrómeros impacta en el número de marcadores necesarios para detectar QTLs. Estos deben ser tenidos en cuenta a la hora de estrategias de genotipado. El genoma del tomate se caracteriza por una baja tasa de recombinación en las regiones centroméricas y esta comprende cerca del 70% de los cromosomas (Sim y col. 2012) lo cual realmente afecta el

LD entre marcadores. El LD es menor en la MAGICPOP en comparación con la población de GWAS donde la línea de base se hace asintótica antes de los 50 cM (Sauvage y col. 2014; Sim y col. 2012). Por otro lado el mapa genético de la MAGIC-POP es un 87% y 105% mas grande en comparación con el EXPEN 2012map y el EXPIM 2012map. Esto se debe al aumento en la tasa de recombinación de poblaciones multiparentales frente a biparentales como son estos casos de de cruces del cultivar domesticado *S. lycopersicum* x *S. pennelli* y *S. lycopersicum* x *S. pimpinellifolium*, respectivamente (Figura 10). En la figura 10, se puede ver como se acentua mas este incremento en la frecuencia de recombinación en las porciones distales de los cromosomas de la MAGIC-POP (línea verde), frente a las poblaciones biparentales (roja y azul). Sin embargo en el cromosoma 3 este incremento se observa casi por completo.

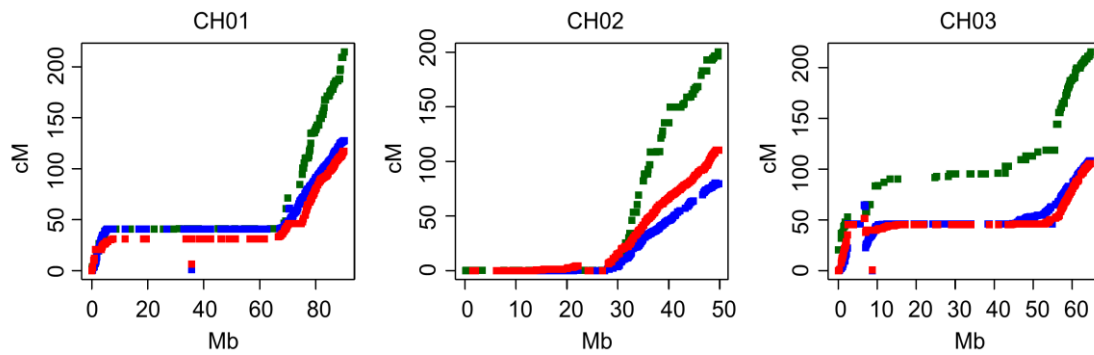


Figura 10. La figura ilustra la relación entre la distancia física y genética de las posiciones de tres cromosomas. La recombinación se incrementa en las regiones distales de los cromosomas y esta casi suprimida en las regiones centroméricas.

de cada cromosoma y la correlación de frecuencias alélicas r^2 en función de la distancia genética y física entre locis. El LD dentro de los cromosomas cae desde 0.7 a 5 cM hasta 0.3 a 25 cM, interesantemente esta es asintótica recién a 90 cM. Con respecto a la distancia física, donde tiene un valor promedio de $r^2=0,47$ a 1kb hasta menos de 0.2 a 2 Mb llegando a un mínimo de 0.08 a 20 Mb. Sin embargo marcadores más distantes (40Mb), el LD se incrementa nuevamente más de 0.13 para caer nuevamente cerca de 50Mb. Otra forma de ver esto es también en la matriz de Kinship o parentesco tiende a ser menor en el tercer cuartil para valores menores a 0.042, inclusive estos valores alcanzan 0.8 para algunas líneas (0.01%).

II.2. Resultados

II.2.1 Analisis de los QTLs identificados a través de 6 analisis de GWAS distintos y contraste con el Mapeo de la población F7 multiparental

Seis análisis de GWAS fueron realizados sobre en base a los datos fenotípicos para contenidos de α - β - γ - δ -tocoferol, α -tocotrienol y tocoferol total en frutos de los 150 materiales, seleccionando el mejor modelo MLMM para cada uno de estos de acuerdo a los test de BIC y extBIC. El seteo de estos análisis vario principalmente en dos variables: i) numero de pasos o cofactores del modelo

que explican la varianza residual asociada al carácter fenotípico (contenido de isómero de VTE) ii) El p-valor de significancia de la asociación que vario entre 10^{-4} y 10^{-5} (ver materiales y métodos para más detalle).

Se identificaron 83 QTLs asociados para 6 caracteres estudiados (Figura 11). Muchos de estos solapados, lo cual tiene lógica ya que las 5 isoformas de tocoferol e incluso el α -tocotrienol comprenden la mayoría de los pasos enzimáticos comunes. Sobre todos los cromosomas se encontró al menos un QTL. El cromosoma 1 fue el que mayor número de QTLs mapeo, con once, mientras que el cromosoma 8 solo uno. Los intervalos de estos QTL fueron calculados en base al LD entre marcadores poniendo un límite de corte promedio de r^2 0,28 bajo el supuesto de que valores por encima de este valor, los marcadores se encuentran ligados. Los intervalos fueron de lo mas variable, desde 68 Mpb comprendiendo por 1940 genes hasta cientos de pare cientos pares de bases con asociaciones con un único gen. De esta forma se determino el numero genes dentro de cada intervalo de confianza para cada análisis se pueden apreciar en la Tabla 2.

Tabla 2. Cantidad de genes identificados por GWAS y mapeo de QTLs		
Análisis de mapeo según su seteo	N genes común Mapeo de QTLs	Total genes IC GWAS
GWAS 1	44	9367
GWAS 2	0	3489
GWAS 3	619	12733
GWAS 4	0	3616
GWAS 5	1436	14077
GWAS 6	468	14045

Con el fin de evaluar la robustez entre los distintos análisis de GWAS (más y menos astringentes en cuanto a su seteo) se decidió comparar las listas de genes candidatos de estos análisis entre ellos y con el análisis de mapeo de

QTLs de los materiales MAGIC-pop (Tabla 2). En la misma se detallan la cantidad de genes en común entre estos, los análisis 1,3,5 y 6 tuvieron genes en común con el mapeo de QTLs.

II.2.2 Analisis de enriquecimiento en categorías funcionales de genes para los análisis de Mapeo

Para evaluar cuan efectivo habían sido realizados tanto los 6 análisis de GWAS como el Mapeo de QTLs de los materiales de la MAGICPOP. Se decidió analizar el enriquecimiento de cada análisis en base a la lista de genes en la categoría funcional Mapman como metabolismo secundario (MS) y genes VTEs (Tablas 3y4) considerados para este análisis de enriquecimiento conforman una lista de 94 genes a partir de 6 trabajos previos de este y otros grupos, tres de ellos en Maiz (Diepenbrock y col. 2017, Hong Wang y col. 2018, Alexander 2013) con lo cual previamente se realizó la búsqueda de ortólogos mediante OrthoDB v8, quedándonos solo con aquellos que exista alta identidad de secuencia (kriventseva y col.2015) algunos de ellos validados mediante mutantes u otros experimentos (ver en negrita Tabla S3) los otros dos reportes sumados son en tomate (Almeida y Col. 2011, Silvestre Lira y col. 2014).

Los resultados de los análisis de enriquecimiento en categorías funcionales no fueron contundentes ya que las diferencias entre los GWAS fue acotada. Sin embargo el Mapeo de QTLs de la MAGIC presento los mejores valores, con 1,3 veces lo esperado para la categoría MS y 1,4 veces más enriquecida en genes VTE. Sin embargo estos análisis se vieron afectados de gran manera por el tamaño de los QTLs que comprenden los genes candidatos en las regiones de alto ligamiento génico.

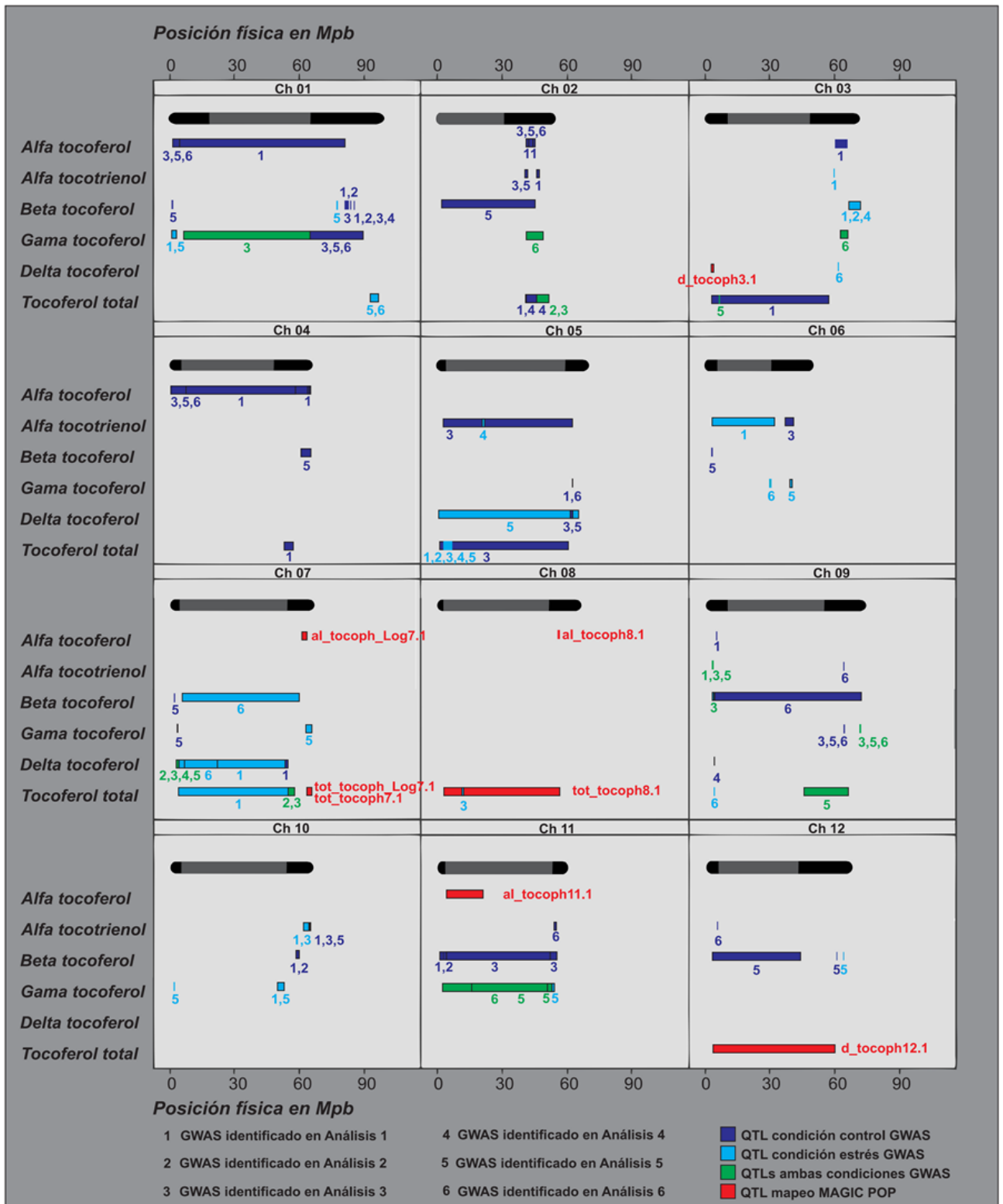


Figura 11. Mapeo físico de los QTLs detectados en GWAS y MAGICPOP. Las distancias están expresadas en millones de pb, sobre el genoma ensamblado SL2.5. Para cada cromosoma y carácter, los QTLs detectados están indicados con los números 1 a 6 dependiendo el análisis de GWAS. En color azul están los QTLs detectados en condición control, en celeste condición de estrés y en verde aquellos QTLs detectados en ambas condiciones

Tablas 3 y 4 - Análisis de enriquecimiento por categorías, vitamina E y Metabolismo secundario para cada una de las estrategias de mapeo.

Tabla 5. Análisis de enriquecimiento por categoría Mapman del BIN16. Metabolismo Secundario							
Analisis	GWAS1	GWAS2	GWAS3	GWAS4	GWAS5	GWAS6	QTL M
Total genes Analisis	9367	3489	12733	3616	14077	14045	2239
Genes BIN16.	170	76	217	72	226	215	24
Total genes todas Cat.	36756	36756	36756	36756	36756	36756	36756
Total genes BIN16.	524	524	524	524	524	524	524
Enriquesimiento MS	1.27	1.53	1.2	1.4	1.13	0.93	1.33
p-Valor Chi_cuadrado	6.30E-06	6.30E-06	6.30E-06	6.30E-06	6.30E-06	6.30E-06	6.30E-06

El total de genes anotados y BIN16 MS correspondiente a la categoría metabolismo secundario fue tomado de acuerdo a las librerías de Categorías funcionales MapMan (Thimm y col. 2004).

Tabla 4. Análisis de enriquecimiento por categoría VTE							
Analisis	GWAS1	GWAS2	GWAS3	GWAS4	GWAS5	GWAS6	QTL M
Genes VTE identif. ^(a)	24	12	50	12	50	58	9
Total genes Analisis	9367	3489	12733	3616	14077	14045	2239
Total genes VTE	94	94	94	94	94	94	94
Total genes anotados	36756	36756	36756	36756	36756	36756	36756
Enriquesimiento VTE ^(b)	0.8	1.1	1.2	1.1	1.1	1.3	1.4
p-Valor Chi cuadrado	1.4E-09	1.4E-09	1.4E-09	1.4E-09	1.4E-09	1.4E-09	1.4E-09

Los genes VTEs considerados para este análisis de enriquecimiento conforman una lista de 94 genes a partir de 6 trabajos previos de este y otros grupos, tres de ellos en Maiz (Diepenbrock y col. 2017, Hong Wang y col. 2018, Alexander 2013) con lo cual previamente se realizó la búsqueda de ortólogos mediante OrthoDB v8, quedándonos solo con aquellos que exista alta identidad de secuencia (kriventseva y col. 2015) algunos de ellos validados mediante mutantes u otros experimentos, en negrita, los otros dos reportes sumados son en tomate (Almeida y Col. 2011, Silvestre Lira y col. 2014). El total de genes anotados fue tomado de acuerdo a Categorías funcionales MapMan (Thimm y col. 2004). ^(a) Genes VTE identificados para cada uno de los respectivos análisis. ^(b) Enriquecimiento en genes de la ruta central de VTE.

En base a estos resultados decidimos explorar en mayor profundidad las asociaciones del análisis de GWAS 3 el cual comprende un número intermedio de genes en comparación con los otros análisis y buenos resultados en cuanto a su enriquecimiento en categorías funcionales metabolismo secundario y para genes relacionados a la vía de VTE (Tabla 5).

Para cada una de las asociaciones del análisis de GWAS 3, se calculó, la heredabilidad estimado en el paso 0 del modelo, la misma esta asociada a ese carácter para una condición dada, control o de estrés hídrico o ambas. Además también se estimó la heredabilidad "perdida" para ese carácter. Esta responde a la cantidad de varianza que podría ser explicada por otros SNPs pero no fue identificada. Otro parámetro que calculamos fue

II.2.3 Estudio del enriquecimiento en factores de transcripción para los genes identificados en los análisis de mapeo

Con el objetivo de identificar los locus correspondientes a los factores de transcripción identificados en los análisis de GWAS y por mapeo de QTLs de los materiales de la población MAGIC. Se realizó una búsqueda de enriquecimiento de factores de transcripción utilizando la herramienta provista

Tabla 5 .Resumen asociaciones analisis GWAS 3					
Carácter	Condicion	Heredabilidad	Heredabilidad "perdida"	PVE	Numero de Asociaciones
α - Tocoferol	Control	0.455	0.101	0.287	3
α - Tocotrienol	Control	0.572	0.028	0.276	4
α - Tocotrienol	Ambos	0.687	0.606	0.210	2
β - Tocoferol	Estrés	0.376	0.000	0.322	2
β - Tocoferol	Control	0.613	0.466	0.405	4
β - Tocoferol	Ambos	0.504	0.000	0.353	2
γ - Tocoferol	Control	0.485	0.190	0.325	4
γ - Tocoferol	Ambos	0.469	0.339	0.221	1
δ - Tocoferol	Control	0.150	0.000	0.167	1
δ - Tocoferol	Ambos	0.174	0.000	0.202	1
Toco total	Estrés	0.239	0.000	0.225	2
Toco total	Control	0.450	0.250	0.370	3
Toco total	Ambos	0.446	0.000	0.294	3

Resumen de asociaciones para contenidos de VTE, la heredabilidad de cada carácter fue calculada en el paso 0 del modelo MLMM, la "perdida de heredabilidad en el paso óptimo del modelo al igual que el porcentaje de la varianza explicado por el SNP

por base de datos PlantTFD, Esta última versión actualizó la notación e identificó 320370 TFs de 165 nuevas especies (Jin JP y col. 2017, http://plantregmap.cbi.pku.edu.cn/tf_enrichment_result.php).

La búsqueda se inició sobre un listado de 91 genes estructurales de la vitamina E (vías del core de VTE, MEP y SK. Este listado consiste de: 47 genes reportados en Almeida y col. (2012), 30 genes identificados por mapeo de asociación GWAS, 14 genes candidatos obtenidos por el mapeo de QTLs en los materiales de la población de MAGIC.

Esta herramienta de identificación de enriquecimiento, permite encontrar TFs sobrerrepresentados de unión a motivos CIS de las región promotora para conjuntos de genes targets. La búsqueda de estos TF se realiza a partir de regulaciones transcripciones que sido reportadas en la literatura, ensayos de ChiP-seq e infiriendo en base de datos, motivos de unión a TF y datos de elementos regulatorios. Para verificar si hay mayor proporción de genes a los cuales se unen TF se utilizo la prueba exacta de Fisher, contrastando contra todos los motivos de la base de datos de regulaciones previamente reportadas (ver materiales y métodos).

A partir de este análisis se identificaron 820 posibles regulaciones entre 216 TFs y 90 de los genes incluidos en la lista de 91 genes estructurales de la vitamina E y convergentes. Además 23 TFs se encuentran sobrerrepresentados con sitios de unión los 91 genes de la lista input (Tabla 6).

Esta herramienta de identificación de enriquecimiento, permite encontrar TFs sobrerrepresentados de unión a motivos CIS de las región promotora para conjuntos de genes targets. La búsqueda de estos TF se realiza a partir de regulaciones transcripciones que sido reportadas en la literatura, ensayos de ChiP-seq e infiriendo en base de datos, motivos de unión a TF y datos de elementos regulatorios. Para verificar si hay mayor proporción de genes a los cuales se unen TF se utilizo la prueba exacta de Fisher, contrastando contra todos los motivos de la base de datos de regulaciones previamente reportadas (ver materiales y métodos).

A partir de este análisis se identificaron 820 posibles regulaciones entre 216 TFs y 90 de los genes incluidos en la lista de 91 genes estructurales de la vitamina E y convergentes. Además 23 TFs se encuentran sobrerrepresentados con sitios de unión los 91 genes de la lista input (Tabla 6).

Tabla 6. Factores de transcripción sobrerrepresentados en el set de genes VTE						
<i>Factores de transcripcion^(a)</i>	<i>Genes totales^(b)</i>	<i>Genes con sitios de unión cis^(c)</i>	<i>Genes VTE^(d)</i>	<i>genesVTE c/unión cis^(e)</i>	<i>p_valor^(f)</i>	<i>FDR^(g)</i>
Solyc09g008390	34727	350	91	5	3,26E-04	4,14E-02
Solyc10g005760	34727	670	91	7	3,84E-04	4,14E-02
Solyc10g055410	34727	995	91	8	1,21E-03	8,68E-02
Solyc10g079050	34727	747	91	6	3,52E-03	1,90E-01
Solyc11g011050	34727	898	91	6	9,36E-03	3,18E-01
Solyc10g081350	34727	1160	91	7	1,12E-02	3,18E-01
Solyc05g014290	34727	949	91	6	1,24E-02	3,18E-01
Solyc04g078840	34727	1194	91	7	1,31E-02	3,18E-01
Solyc06g072890	34727	347	91	3	1,33E-02	3,18E-01
Solyc06g053350	34727	366	91	3	1,58E-02	3,34E-01
Solyc02g085160	34727	374	91	3	1,70E-02	3,34E-01
Solyc06g075140	34727	407	91	3	2,24E-02	3,55E-01
Solyc03g112890	34727	231	91	2	2,30E-02	3,55E-01
Solyc06g071520	34727	231	91	2	2,30E-02	3,55E-01
Solyc01g109880	34727	868	91	5	2,67E-02	3,84E-01
Solyc09g009490	34727	1149	91	6	3,13E-02	3,90E-01
Solyc02g062950	34727	1167	91	6	3,37E-02	3,90E-01
Solyc01g095460	34727	1197	91	6	3,78E-02	3,90E-01
Solyc06g062520	34727	4792	91	18	4,07E-02	3,90E-01
Solyc06g074320	34727	296	91	2	4,30E-02	3,90E-01
Solyc10g044680	34727	1809	91	8	4,75E-02	3,90E-01
Solyc08g005290	34727	1260	91	6	4,76E-02	3,90E-01
Solyc04g074170	34727	1814	91	8	4,82E-02	3,90E-01

(a) Factores de transcripción sobrerrepresentados en el listado de 91 genes estructurales de la VTE

(b) Total de TFs en la base de datos PlantTFDB para la inferencia del análisis de enriquecimiento

(c) Total de TFs en la base de datos PlantTFDB que poseen sitios de unión al listado de genes VTE input

(d) Listado de genes estructurales VTE (input)

(e) Número de genes con motivos de unión a TFs^(a), dentro del listado 91 genes

(f) p-valor Test exacto de Fisher

^(g) q-valor o p-valor ajustado a los falsos positivos FDR del Test exacto de Fisher

Entre los 216 TFs identificados, varios de ellos se encuentran compartidos como candidatos de los análisis de GWAS y mapeo de QTLs (Tabla 7). Este análisis identifico 216 TFs que podrían estar regulando los genes estructurales de la vitamina E reportados hasta el momento y nuevos, de los cuales además se encuentran compartidos dentro de los análisis de GWAS y mapeo de QTLs.

Tabla 7. Cantidad de TFs identificados por GWAS y mapeo de QTLs		
Análisis	Factores de transcripción	Total genes IC GWAS
GWAS 1	177	9367
GWAS 2	90	3489
GWAS 3	358	12733
GWAS 4	60	3616
GWAS 5	322	14077
GWAS 6	468	14045
Mapeo de QTLs	13	2239

El hecho que la mayoría de los motivos identificados muestren alta similitud con elementos ya descritos en plantas, que posean alta preferencia posicional y que estén sobre-representados en los promotores de genes de la vía de síntesis de vitamina E, sugiere que estos elementos en cis, son realmente reguladores funcionales.

Sin embargo es sabido que ni la presencia de motivos compartidos ni de motivos exclusivos redundan en un perfil de co-expresión. Un ejemplo es el caso de IPI(2)-CLH y VTE3(2)-TYRA(1). Ambos pares de genes comparten motivos comunes y exclusivos (VTE14MEME6 y VTE14MEME7, respectivamente) y mientras que el primer par muestra un perfil coordinado de expresión, el segundo no (Quadrona y col. 2013). Por otro lado, el número de motivos compartidos tampoco es un factor determinante para la co-regulación.

II. 2.4 Genes identificados por GWAS análisis y mapeo de QTL

El gen **CM(2) Solyc11g017240** que codifica para uno de los parálogos de la enzima Corismato Mutasa, que cataliza el paso de síntesis de corismato a prefenato. La Corismato mutasa está involucrada en la síntesis de aa aromáticos, fenilalanina, tirosina, triptófano y homogentisato precursor de la síntesis de VTE. Fue identificada en los análisis 3 y 6 de GWAs y a través del mapeo de QTLs sobre el cromosoma 11 comprendida en un intervalo de 11,8

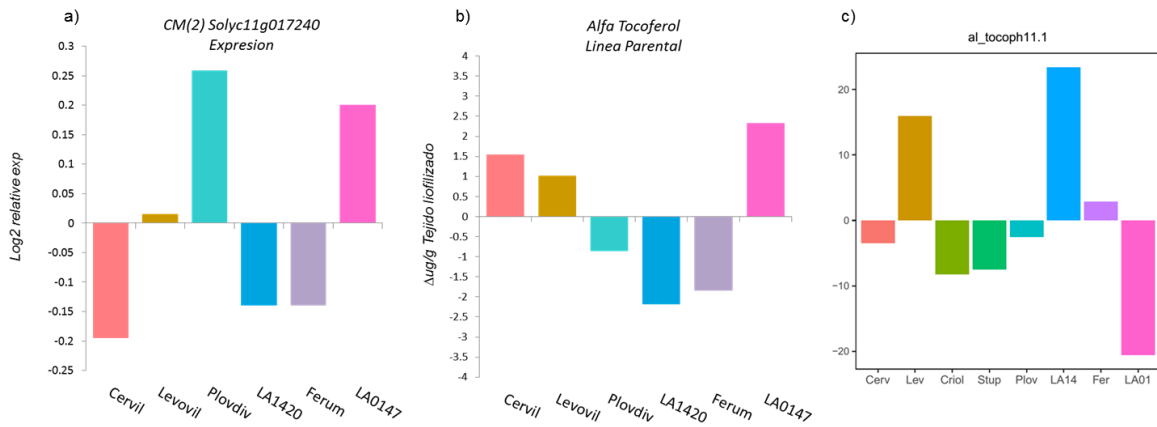


Figura 12. A) Expresion relativa de cada uno de los parentales para el gen que codifica para la Corismato Mutasa 2, en estadio *cell expansion* **B)** Contenidos de α -tocoferol en estadio de *cell expansion* para la línea parental **C)** Aporte parental de cada alelo parental sobre la Corismato Mutasa a partir de los datos de Mapeo de QTLs de la F7

cm en el qtl denominado al_tocoph11.1. Este QTL comprende un total de 619 genes, de los cuales 187 son mejores candidatos, ya que poseen polimorfismos congruentes con los alelos parentales que mas aportan, dentro de estos el único gen anotado dentro de la categoría metabolismo secundario Mapman es la Corismato Mutasa. Este gen presenta un total de 88 polimorfismos entre los alelos de la línea parental y *S. Lycopersicum* (cv.Heinz), 28 de estos se encuentran corriente abajo del gen, 33 corriente arriba, 25 dentro de intrones, 6 son Indels y 3 SNP sinónimos sobre exones.

Uno o más de estos polimorfismos son posibles candidatos a estar asociados a la modulación de la expresión. Si observamos los perfiles de la expresión de RNAseq de *CM(2)* en estadio *cell expansion* para la línea parental y los comparamos con los niveles de tocoferol para este mismo estadio. Podemos ver que para los genotipos Levovil, LA1420, Ferum y LA0147 (Figura 12 a), existe congruencia entre, contenidos de alfa tocoferol y niveles de expresión (Figura 12b).

Por otro lado para los genotipos Cervil y Ploudiv, esto no ocurre, lo cual tiene total sentido si miramos la figura 12 c), la cual muestra el aporte alélico de cada parental dentro de la población F7 analizada. Para el caso de Cervil y Ploudiv el aporte de ambos es mínimo en comparación a los demás parentales, ya la contribución alélica de este QTL viene dada principalmente a través de los genotipos LA1420, Levovil y LA0147.

Sin embargo siguiendo esta línea de análisis, esperaríamos ver los mayores niveles de alfa tocoferol respecto a la media para LA1420, niveles medios para Levovil y cantidades muy bajas en LA0147. Esta incongruencia entre los

contenidos de alfa tocoferol y el aporte de cada alelo, figuras b y c, respectivamente podría deberse a que en la figura c) el aporte de cada alelo parental, responde a en estadio maduro para la F7 mientras que en la figura b) los contenidos son en estadio de *cell expansión*.

Al explorar los niveles de SiRNAs en la región promotora de la CM(2) para la línea parental, esta presenta diferencias significativas ($p < 0,05$) en cuanto al número de lecturas que “pegan” sobre la misma entre los distintos parentales. Sin embargo, no hay correlación entre el nivel de expresión de las líneas parentales y la cantidad de SiRNAs.

Otro de los genes identificados en el análisis de GWAS 5 y mediante mapeo, es el gen **VTE6 Solyc07g062180** que codifica para la Fitol Fosfato Quinasa que cataliza el paso de fitil monofosfato a difosfato, envuelto en la reutilización de la degradación de clorofila y la síntesis de vitamina E (Mach J y col. 2015). Este gen mapeo dentro de tres qtls anidados, *al_tocoph_Log7.1* para contenido de alfa tocoferol, *tot_tocoph_Log7.1* y *tot_tocoph71* para contenidos totales de tocoferol. El QTL para alfa tocoferol comprende, 292 genes dentro de los cuales 73 se diferencian en su aporte alélico parental, incluido VTE6 que es el único de la categoría metabolismo secundario.

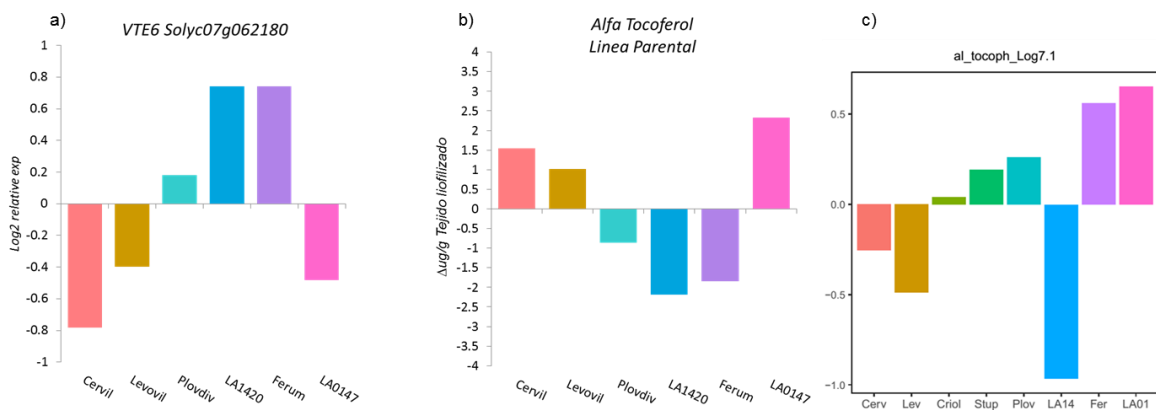


Figura 13. A) Expresión relativa de cada uno de los parentales para el gen que codifica para la VTE6, en estadio *cell expansión* **B)** Contenidos de α -tocoferol en estadio de *cell expansión* para la línea parental **C)** Aporte parental de cada alelo parental sobre VTE6 a partir de los datos de Mapeo de QTLs de la F7

del mapeo de QTLs de la F7, LA1420 y LA0147 presentan total congruencia mientras que Ferum y Levovil muestran un aporte en sentido opuesto al que se observa en la línea parental. Esto también podría deberse a que no se trata del mismo estadio de maduración del fruto.

II.2.5 Arquitectura del carácter heredabilidad VTE (disección de la variabilidad asociada al carácter VTE)

El análisis de la varianza para los contenidos de tocoferol total y sus isoformas fue homogéneo en cuanto a las proporciones de su distribución, gran parte no pudo ser explicada a través de este modelo mixto. Sin embargo, este análisis arroja algunos resultados interesantes como ser que entre un 5 y 15% de la varianza fenotípica se explica por la interacción genotipo-ambiente y entre un 20 y 35% por el genotipo. También se permitió ver como α -tocotrienol se diferencia en cuanto a los perfiles del resto de los caracteres, por ejemplo no presenta componente genotípica prácticamente. Es importante también resaltar que el lugar de medición tuvo un efecto sobre los contenidos de VTE (Figura 14).

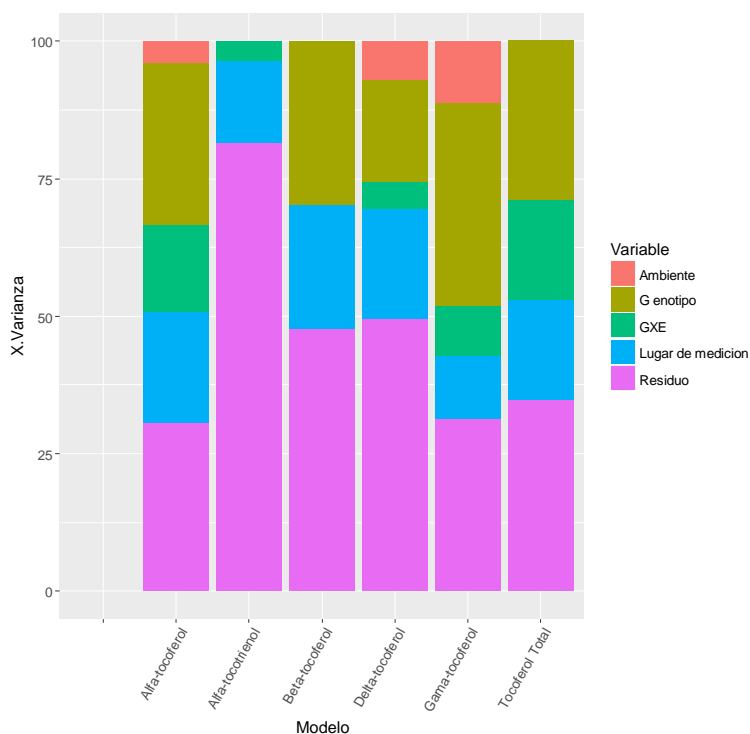


Figura 14. Disección de la varianza fenotípica para los 6 caracteres medidos en el análisis de GWAS. En naranja fue graficada la componente ambiente, en verde el genotipo, en rojo la interacción genotipo-ambiente, en azul el lugar de medición HPLC y en púrpura el residuo de la varianza.

II.3. Discussion

A partir de los resultados de este capítulo fue posible echar luz sobre la biosíntesis de tococromanos (figura 15). La reconstrucción del metabolismo de vitamina E en tomate dio como resultado la identificación de 83 nuevos QTLs y varios genes ya reportados en la vía y otros no.

Dentro de la vía del MEP, se identificó a la enzima GPPS, la cual desvía el metabolismo hacia la producción de monoterpenos, mientras que GGPS es responsable de la producción de geranylgeranyl difosfato mediante la unión secuencial de tres isopentenil difosfato (IPP) a una molécula de dimetilalil difosfato (DMAPP). El resultante geranylgeranyl difosfato sirve como precursor para la síntesis de carotenoides, clorofila y tococromanos (Aharoni et al., 2005; Joyard et al., 2009), indicando que GPPS no estaría involucrada en la producción de tococromanos. En este sentido, plantas de tomate silenciadas para GPPS muestran un fenotipo enano y reducción de los niveles de giberelinas, aunque no muestran alteración en los niveles de carotenoides, tocoferoles ni clorofila (Van Schie et al., 2007). Estos resultados sugieren que la enzima GPPS no tendría un rol importante en la biosíntesis de vitamina E.

En segundo lugar, la enzima HGGT responsable de la condensación de homogentisato con geranylgeranyl difosfato necesario para la síntesis de tocotrienol, fue solo identificada en gramíneas donde los tococromanos más abundantes son los tocotrienoles (Cahoon et al., 2003). Ni en *A. thaliana* en tomate se identificó ningún gen ortólogo a algunos de los identificados en gramíneas codificando para HGGT. Sin embargo, trazas de tocotrienol se han detectado en tomate y tabaco (Chun et al., 2006). Esto podría ser explicado por la promiscuidad en la aceptación de sustratos por parte de la enzima homogentisato solanesil transferasa (HST) y/o homogentisato fitil Transferasa (VTE2). En este sentido se ha visto que cambios en el flujo a través de SK, MEP y/o la vía central de tocoferol podría cambiar la preferencia de sustrato de estas enzimas (Falk & Munné-Bosch, 2010; Herbers, 2003). Esta hipótesis está apoyada por el hecho de que plantas de tabaco, que no posee HGGT endógena, co-expresando la HPPD de *A. thaliana* y la prefenato deshidrogenasa (PDH) de levadura, acumula altos niveles de tocotrienoles (Rippert et al., 2004). Este mismo experimento realizado sobre plantas de tomates resultó en similares patrones de acumulación de tocotrienoles en fruto (Baldet et al, datos no publicados). Por lo tanto, la hipótesis propuesta por Falk & Munné-Bosch (2010) acerca de que el aumento del suministro de homogentisato podría afectar a la especificidad de sustrato de HST favoreciendo la síntesis de

tocotrienoles pareciera plausible también en el caso de tomate. Además, HST de *A. thaliana*, esencial para la biosíntesis de plastoquinona-9, es capaz de catalizar la condensación desolanesil difosfato y homogentisato, como así también aceptar farnesil difosfato o geranilgeranil bifosfato como donantes de grupos prenilo (Sadre et al., 2006; Tian et al., 2007), demostrando que la producción de tocotrienoles puede suceder independientemente de la presencia de HGGT. Por último, dos pasos permanecieron durante mucho tiempo irresolutos en la red metabólica de plantas.

El primero es la conversión del intermedio de la vía del SK, prefenato en arogenato por la enzima PAT, la cual ha sido extensamente caracterizada en bacterias. Aunque su actividad había sido detectada en plantas, no se había logrado adjudicar ningún gen responsable (Tzin et al., 2009). El segundo punto oscuro de este metabolismo es la ausencia de una quinasa de fitil fosfato, que podría proporcionar fitil difosfato para síntesis de tocoferoles, como una alternativa a la vía MEP (Ischebeck et al., 2006; Valentin et al., 2006).

Casi la totalidad de los genes que codifican enzimas de la vía del MEP identificados en este estudio están en línea con trabajos previos en tomate (Ament et al., 2006; Botella-Pavía et al., 2004; Lois et al., 2000; Rodríguez-Concepción et al., 2001; Rodríguez-Concepción et al., 2003; Rohdich et al., 2000).

Más aún, en esta tesis se identificaron los genes que codifican para la 2- C- metil-D-eritritol 4-fosfato citidililtransferasa (CMS) y la geranilgeranil reductasa (GGDR), los cuales no habían sido reportados anteriormente en tomate. Por otro lado, también se identificaron dos loci adicionales para GGPS no reportados hasta el momento. Para la vía de SK, ya han sido reportados en tomate los genes que codifican las enzimas implicadas en reacciones río arriba del intermediario prefenato (Bischoff et al., 1996; Bischoff et al., 2001; Eberhard et al., 1996; Gasser et al., 1988; Görlach et al., 1993, 1995; Schmid et al., 1992). Sin embargo, identificamos nuevos loci para SDH / DHQ y CM, además de caracterizar por primera vez los loci que codifican para TYRA y TAT en tomate.

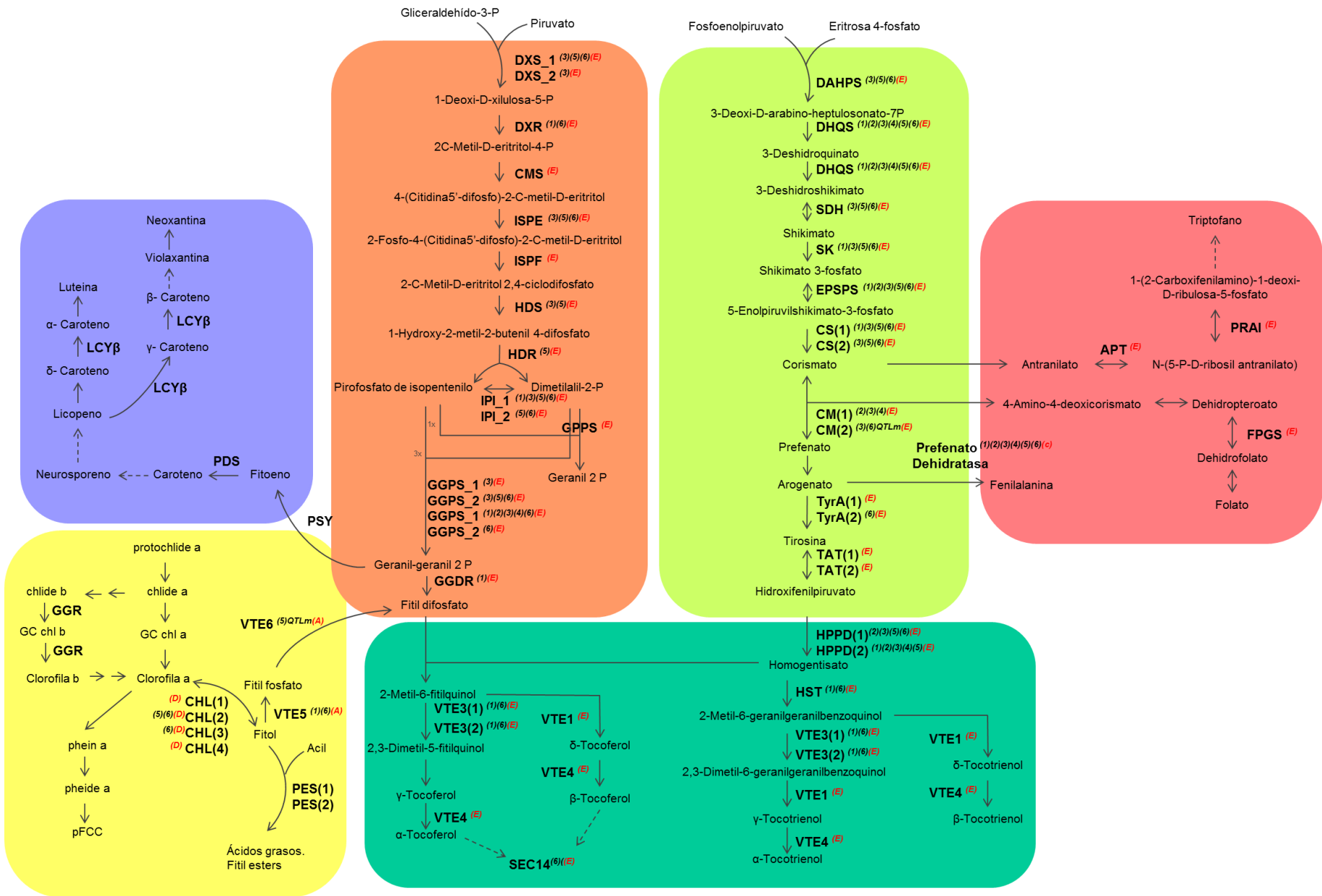


Figura 15. Genes candidatos identificados a partir de los análisis de GWAS1-6 y Mapeo de QTLs, en común con otros análisis previos reportados en la literatura de candidatos en Tomate y Maiz (ver tabla S3). Violeta via de biosíntesis de carotenoides, verde claro la via MEP, naranja via SK, rojo catabolismo y síntesis de aminoácidos fenilalanina y tirosina, amarillo degradación de clorofila, verde oscuro el core de tocoferol. Adaptado de Quadrana y col 2013, Hong Wang y col. 2018.

Páginas 51 a 84 eliminadas a pedido del autor.

CAPÍTULO VI – Referencias bibliográficas

Almeida J, Quadrana L, Asís R, Setta N, De Godoy F, Bermúdez L, Otaiza SN: Genetic dissection of vitamin E biosynthesis in tomato. *Journal of experimental botany* 2011, 1-18. doi:10.1093/jxb/err055

Almeida J, Azevedo S, Spicher L, Glauser G, Von Dorp K, Guyer L, Del Valle Carranza A, Asis R, de Souza AP, Buckeridge M, Demarco D, Bres C, Rothan C, Pers LE, Hortensteiner S, Kessler F, Dormann P, Carrari F, Rossi M: Down-regulation of tomato strongly impairs tocopherol biosynthesis and affects prenyl lipid metabolism in an organ-specific manner. *Journal of Experimental Botany* 2015, 67(3):919-34. Doi 10.1093/jxb/erv504

Alonso-Blanco C, El-Assal SE, Coupland G, Koornneef M: Analysis of natural allelic variation at flowering time loci in the Landsberg erecta and Cape Verde Islands ecotypes of *Arabidopsis thaliana*. *Genetics* 1998, 149(2):749-764.

Alseikh S, Ofner I, Pleban T, Tripodi P, Di Dato F y col: Resolution by recombination: breaking up *Solanum pennellii* introgressions. *Trends Plant Sci* 2013. 18:536-38

Alseikh S, Tohge T, Wendenberg R, Scossa F, Omranian N, et al. 2015. Identification and mode of inheritance of quantitative trait loci for secondary metabolite abundance in tomato. *Plant Cell* 27:485-512

Anders S, Pyl PT, Huber W: HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* 2015, 31(2):166-9.

Ashby E: Studies in the inheritance of physiological characters. I. A physiological investigation of the nature of hybrid vigor in maize. *Annals of Botany* 1930, 44: 457-476.

Asimit J, Zeggini E: Rare variant association analysis methods for complex traits. *Annu Rev Genet* 2010, 44:293-308.

Asprelli P: Recuperación, evaluación morfoagronómica y caracterización molecular de una colección de germoplasma de tomate criollo en Argentina. Tesis Doctoral, 2010, Universidad Nacional de Cuyo, PROBIOL.

Atwell S, Huang YS, Vilhjalmsón BJ, Willems G, Horton M, Li Y, Meng D, Platt A, Tarone AM, Hu TT y col: Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* 2010, 465(7298):627-631

Aufsatz W, Mette MF, van der Winden J, Matzke AJ, Matzke M: RNA-directed DNA methylation in *Arabidopsis*. *PNAS* 2002, 10: 16499-506.

Ausin I, Greenberg MV, Simanshu DK, Hale CJ, Vashisht AA, Simon SA, Lee TF, Feng S, Española SD, Meyers BC, Wohlschlegel JA, Patel DJ, Jacobsen SE: INVOLVED IN DE NOVO 2-containing complex involved in RNA-directed DNA methylation in *Arabidopsis*. *PNAS* 2012, 22: 8374-81.

Balasubramanian S, Schwartz C, Singh A, Warthmann N, Kim MC, Maloof JN, Loudet O, Trainer GT, Dabi T, Borevitz JO: QTL mapping in new *Arabidopsis thaliana* advanced intercross -recombinant inbred lines. *PLoS One* 2009, 4(2):e4318.

Ballester AR, Tikunov Y, Molthoff J, Grandillo S, Viquez-Zamora M, et al. Identification of loci affecting accumulation of secondary metabolites in tomato fruit of a *Solanum lycopersicum* × *Solanum chmielewskii* introgression line population. *Front. Plant Sci* 2016, 7:1428

Bao J, Lee S, Chen C, Zhang X, Zhang Y, Liu S, Clark T, Wang J, Cao M, Yang H, Wang SM, Yu J: Serial analysis of gene expression study of a hybrid rice strain (LYP9) and its parental cultivars. *Plant Physiology* 2005, 138: 1216-1231

Baranwal VK, Mikkilineni V, Zehr UB, Tyagi AK, Kapoor S: Heterosis: emerging ideas about hybrid vigour. *Journal Exp Bot* 2012, 63(18):6309-14.

Barrantes W, Lopez-Casado G, Garcia-Martinez S, Alonso A, Rubio F: Exploring new alleles involved in tomato fruit quality in an introgression line library of *Solanum pimpinellifolium*. *Front. Plant Sci* 2016, 7:1172

Beleza S, Johnson NA, Candille SI, Absher DM, Coram MA, Lopes J, Campos J, Araujo II, Anderson TM, Vilhjalmsson BJ y col: Genetic architecture of skin and eye color in an African-European admixed population. *PLoS Genet* 2013, 9(3):e1003372

Bentsink L, Alonso-Blanco C, Vreugdenhil D, Tesnier K, Groot SPC, Koornneef M: Genetic analysis of seed-soluble oligosaccharides in relation to seed storability of *Arabidopsis*. *Plant Physiol* 2000, 124:1595-

Bergelson J, Roux F: Towards identifying genes underlying ecologically relevant traits in *Arabidopsis thaliana*. *Nat Rev Genet* 2010, 11(12):867-879

Bermúdez L, de Godoy F, Baldet P, Demarco D, Osorio S, Quadrana L, Almeida J, Asis R, Gibon Y, Fernie AR, Rossi M, Carrari F: Silencing of the tomato sugar partitioning affecting protein (SPA) modifies sink strength through a shift in leaf sugar metabolism. *Plant Journal* 2014, 77(5):676-87.

Bermúdez L, Urias U, Milstein D, Kamenetzky L, Asis R, Fernie AR, Van Sluys MA, Carrari F, Rossi M: A candidate gene survey of quantitative trait loci affecting chemical composition in tomato fruit. *Journal of Experimental Botany* 2008, 59(10):2875-90.

Bodmer W, Bonilla C: Common and rare variants in multifactorial susceptibility to common diseases. *Nat Genet* 2008, 40(6):695-701

Bolger T y col: The genome of the highly stress tolerant wild tomato species *Solanum pennellii*. *Nature Genetics* 2014, 46, 1034-1038.

Brachi B, Faure N, Horton M, Flahauw E, Vazquez A, Nordborg M, Bergelson J, Cuguen J, Roux F: Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature. *PLoS Genet* 2010, 6(5):e1000940.

Brachi B, Faure N, Horton M, Flahauw E, Vazquez A, Nordborg M, Bergelson J, Cuguen J, Roux F: Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature. *PLoS Genet* 2010, 6(5):e1000940.

Causse M, Whole genome resequencing in tomato reveals variation associated with introgression and breeding events. *BMC Genomics* 2013, 14;14:791.

Chaib J, Lecomte L, Buret M, Causse M: Stability over genetic backgrounds, generations and years of quantitative trait locus (QTLs) for organoleptic quality in tomato. *Theor. Appl. Genet* 2006, 112:934-44

Chen Z: Genomic and epigenetic insights into the molecular bases of heterosis. *Nat Rev Genet* 2013, 14: 471-482.

Chen, Z: Molecular mechanisms of polyploidy and hybrid vigor. *Trends in plant science* 2010, 57-71.

Chodavarapu RK, Feng S, Ding B, Simon SA, Lopez D, Jia Y, Wang GL, Meyers BC, Jacobsen SE, Pellegrini M: Transcriptome and methylome interactions in rice hybrids. *PNAS* 2012, 109 (30),12040-5. doi:10.1073/pnas.1209297109

Clarke JH, Mithen R, Brown JK, Dean C: QTL analysis of flowering time in *Arabidopsis thaliana*. *Mol Gen Genet* 1995, 248 (3):278-286.

Coley P, Bryant J, Chapin Fr: Resource availability and plant antiherbivore defense. *Science* 1985, 230:895.12.

Cortijo S, Wardenaar R, Colomé-Tatché M, Gilly A, Etcheverry M, Labadie K, Caillieux E, Hospital F, Aury JM, Wincker P, Roudier F, Jansen RC, Colot V, Johannes F Mapping the epigenetic basis of complex traits. *Science* 2014. 7;343(6175):11458.

Davenport CB: Recessive characters. *Science* 1908,20; 28(725):729.

De Godoy F, Bermúdez L, Lira BS, de Souza AP, Elbl P, Demarco D, Alseekh S, Insani M, Buckeridge M, Almeida J, Grigioni G, Fernie AR, Carrari F, Rossi M: Galacturonosyltransferase 4 silencing alters pectin composition and carbon partitioning in tomato. *Journal of Experimental Botany* 2013, 64 (8): 2449-2466.

Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB: Rare variants create synthetic genome-wide associations. *PLoS Biol* 2010, 8(1):e1000294

Dixon RA and Strack D: Phytochemistry meets genome analysis, and beyond. *Phytochemistry* 2003, 62:815-16

Do PT, Prudent M, Sulpice R, Causse M, Fernie AR: The influence of fruit load on the tomato pericarp metabolome in a *Solanum chmielewskii* introgression line population. *Plant Physiol* 2010, 154:1128-42

Dominguez PG, Frankel N, Mazuch J, Balbo I, Iusem N, Fernie AR, Carrari F. ASR1 mediates glucose-hormone cross talk by affecting sugar trafficking in tobacco plants. *Plant Physiol* 2013, Mar;161(3):1486-500

Feng T, Zhu X: Detecting rare variants. *Methods Mol Biol* 2012.

Fernie AR: The future of metabolic phytochemistry: larger numbers of metabolites, higher resolution, greater understanding. *Phytochemistry* 2007, 68:2861-80

Fitzpatrick T y col: Vitamin deficiencies in humans: Can Plant Science help? *Plant Cell* 2012, 24 (2):395-414.

Flint J, Mackay TF: Genetic architecture of quantitative traits in mice, flies, and humans. *Genome Res* 2009, 19(5):723-733

Frary A, Grandillo S, van der Knaap E, Cong B, Liu J, Meller J, Elber R, Alpert KB, Tanksley S: Quantitative Trait Locus key to the evolution of tomato fruit size. *Science* 2000, 289:85-88.

Frisch M, Thiemann A, Fu J, Schrag TA, Scholten S, Melchinger AE: Transcriptome-based distance measures for grouping of germplasm and prediction of hybrid performance in maize. *Theor. Appl. Genet* 2010, 120, 441-450.

Fritz C, Palacios-Rojas N, Feil R, Stitt M: Regulation of secondary metabolism by the carbon-nitrogen status in tobacco: nitrate inhibits large sectors of phenylpropanoid metabolism. *Plant J* 2006, 46:533-548.

Futuyma DJ, Agrawal AA.. Macroevolution and the biological diversity of plants and herbivores. *PNAS* 2009, 106:18054-61

Gibon Y, Blaesing OE, Hannemann J, Carillo P, Hohne M, Hendriks JHM, Palacios N, Cross J, Selbig J, Stitt M. A robot-based platform to measure multiple enzyme activities in *Arabidopsis* using a set of cycling assays: comparison of changes of enzyme activities and transcript levels during diurnal cycles and in prolonged darkness. *The Plant Cell* 2004, 16, 3304-3325.

Gibson G: Rare and common variants: twenty arguments. *Nat Rev Genet* 2011, 13(2):135-145.

Goff SA: A unifying theory for general multigenic heterosis: energy efficiency, protein metabolism, and implications for molecular breeding. *New phytol* 2011, 189(4), 923-37.

Greaves IK, Groszmann M, Ying H, Taylor JM, Peacock WJ, Dennis ES: Trans chromosomal methylation in *Arabidopsis* hybrids. *PNAS* 2012, 109, 3570-3575.

Groszmann M, Greaves IK, Albertyn ZI, Scofield GN, Peacock WJ, Dennis ES (2011). Changes in 24-nt siRNA levels in *Arabidopsis* hybrids suggest an epigenetic contribution to hybrid vigor. *PNAS, USA* 108, 2617-2622.

Guimerà R, Nunes Amaral LA: Functional cartography of complex metabolic networks. *Nature*. 2005, 24;433(7028):895-900.

Guo M, Rupe MA, Danilevskaya ON, Yang X, Hu Z: Genome-wide mRNA profiling reveals heterochronic allelic variation and a new imprinted gene in hybrid maize endosperm. *Plant J* 2003, 36: 30-44.

Guo M, Rupe MA, Yang X, Crasta O, Zinselmeier C, Smith OS, Bowen B: Genome-wide transcript analysis of maize hybrids: allelic additive gene expression and yield heterosis. *Theor. Appl. Genet* 2006, 113: 831-845.

Guo M, Rupe MA, Zinselmeier C, Habben J, Bowen BA, Smith OS (2004). Allelic variation of gene expression in maize hybrids. *Plant Cell* 2004, 16: 1707-1716.

Ha M, Lu J, Tian L, Ramachandran V, Kasschau KD, Chapman EJ, Carrington JC, Chen X, Wang X-J, Chen ZJ: Small RNAs serve as a genetic buffer against genomic shock in *Arabidopsis* interspecific hybrids and allopolyploids 2009, *PNAS*, USA 106, 17835-17840.

Hauben M, Haesendonckx B, Standaert E, Van Der Kelen K, Azmi A, Akpo H, Van Breusegem F, Guisez Y, Bots M, Lambert B, Laga B, De Block M: Energy use efficiency is characterized by an epigenetic component that can be directed through artificial selection to increase yield. *PNAS* 2009, USA 106, 20109-20114.

He G, Zhu X, Elling AA, Chen L, Wang X, Guo L, Liang M, He H, Zhang H, Chen F, Qi Y, Chen R, Deng XW: Global epigenetic and transcriptional trends among two rice subspecies and their reciprocal hybrids. *Plant Cell* 2010, 22, 17-33.

Hoecker N, Keller B, Muthreich N, Chollet D, Descombes P, Piepho HP, Hochholdinger F (2008a). Comparison of maize (*Zea mays* L.) F1-hybrid and parental inbred line primary root transcriptomes suggests organ-specific patterns of nonadditive gene expression and conserved expression trends. *Genetics* 179: 1275-1283.

Hoecker N, Lamkemeyer T, Sarholz B, Paschold A, Fladerer C, Madlung J, Wurster K, Stahl M, Piepho HP, Nordheim A, Hochholdinger F: Analysis of nonadditive protein accumulation in young primary roots of a maize (*Zea mays* L.) F(1)-hybrid compared to its parental inbred lines. *Proteomics* 2008, 8: 3882-3894.

Huang X, Zhao Y, Wei X, Li C, Wang A, Zhao Q, Li W, Guo Y, Deng L, Zhu C: Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nat Genet* 2012, 44(1):32-39.

Huang Y, Li L, Chen Y, Li X, Xu C, Wang S, Zhang Q: Comparative analysis of gene expression at early seedling stage between a rice hybrid and its parents using a cDNA microarray of 9198 unisequences. *Sci. China* 2006a, Ser. C. 49: 519-529.

Huang Y, Zhang L, Zhang J, Yuan D, Xu C, Li X, Zhou D, Wang S, Zhang Q: Heterosis and polymorphisms of gene expression in an elite rice hybrid as revealed by a microarray analysis of 9198 unique ESTs. *Plant Mol. Biol* 2006b, 62: 579-591.

Jahnke S, Sarholz B, Thiemann A, Kuhr V, Gutierrez-Marcos JF, Geiger HH, Piepho HP, Scholten S (2010). Heterosis in early seed development: a comparative study of F1 embryo and endosperm tissues 6 days after fertilization. *Theor. Appl. Genet.* 120: 389-400.

Jiang K, Liberatore KL, Park SJ, Alvarez JP, Lippman ZB: Tomato yield heterosis is triggered by a dosage sensitivity of the florigen pathway that fine-tunes shoot architecture. *PLoS Genet.* 2013, 9(12):e1004043.

Johannes F, Porcher E, Teixeira FK, Saliba-Colombani V, Simon M, Agier N, Bulski A, Albuissou J, Heredia F, Audigier P, et al: Assessing the impact of transgenerational epigenetic variation on complex traits. *PLoS Genet* 2009, 5(6):e1000530.

Kamenetzky L y col: Genomic analysis of wild tomato (*Solanum pennellii*) introgressions determining metabolic- and yield-associated traits. *Plant Physiol* 2010, 152(4):1772-86.

Kamenetzky L, Asís R, Bassi S, de Godoy F, Bermúdez L, Fernie AR, Van Sluys MA, Vrebalov J, Giovannoni JJ, Rossi M, Carrari F: Genomic Analysis of Wild Tomato Introgressions Determining Metabolism- and Yield-Associated Traits. *Plant Phys* 2010: 152: 1772-1786.

Keene JD (2007). RNA regulons: coordination of post-transcriptional events. *Nat. Rev. Genet* 2007, 8: 533-543.

Keurentjes JJB, Fernie AR. 2011. Genetics, genomics and metabolomics. In *Annual Plant Reviews*, ed. RD Hall, pp. 239-52. London: Wiley

Kleessen S, Antonio C, Sulpice R, Laitinen R, Fernie AR y col: Structured patterns in geographic variability of metabolic phenotypes in *Arabidopsis thaliana*. *Nat. Commun* 2012, 3:1319

Kleessen S, Laitinen R, Fusari CM, Antonio C, Sulpice R y col: Metabolic efficiency underpins performance trade-offs in growth of *Arabidopsis thaliana*. *Nat. Commun* 2014 5:3537

Kliebenstein DJ, Kroymann J, Brown P, Figuth A, Pedersen D y col: Genetic control of natural variation in *Arabidopsis* glucosinolate accumulation. *Plant Physiol* 2001. 126:811-25

Kollipara KP, Saab IN, Wych RD, Lauer MJ, Singletary GW (2002). Expression profiling of reciprocal maize hybrids divergent for cold germination and desiccation tolerance. *Plant Phys.* 129: 974-992.

Komili S, Silver PA (2008). Coupling and coordination in gene expression processes: a systems biology view. *Nat. Rev. Genet* 2008, 9: 38-48.

Kover PX, Valdar W, Trakalo J, Scarcelli N, Ehrenreich IM, Purugganan MD, Durrant C, Mott R: A Multiparent Advanced Generation Inter-Cross to fine-map quantitative traits in *Arabidopsis thaliana*. *PLoS Genet* 2009, 5(7):e1000551

Kowalski SP, Lan TH, Feldmann KA, Paterson AH: QTL mapping of naturally-occurring variation in flowering time of *Arabidopsis thaliana*. *Mol Gen Genet* 1994, 245(5):548-555

Krieger U, Lippman ZB, Zamir D: The flowering gene *SINGLE FLOWER TRUSS* drives heterosis for yield in tomato. *Nat. Genet* 2010, 42(5), 459-63.

Langmead B y col: Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 2010, 10:R25.

Lê Cao, K-A., González I. and Déjean S: intergrOmics: an R package to unravel relationships between two omics data sets. *Bioinformatics* 2009, 25(21):2855-2856

Li DP, Baldwin IT, Gaquerel E. Navigating natural variation in herbivory-induced secondary metabolism in coyote tobacco populations using MS/MS structural analysis. *PNAS* 2015, 112:E4147-55

Li H. y Durbin R: Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics* 2009, 25:1754-60.

Lippman ZB, Zamir D. Heterosis: revisiting the magic. *Trends Genet.* 2007, 23(2):60-6.

López M, Zanon I, Pratta GR, Stegmayer G, Boggio SB, Conte M, Bermúdez L, Coluccio Leskow C, Rodríguez GR, Picardi LA, Zorzoli R, Fernie A, Milone DH, Asís R, Valle EM, Carrari F: Metabolic analyses of interspecific tomato recombinant inbred lines for fruit quality improvement. *Metabolomics* 2015, DOI: 10.1007/s11306-015-0798-3.

López MG y col: Metabolic analyses of interspecific tomato recombinant inbred lines for fruit quality improvement. *Metabolomics.* 2015, DOI: 10.1007/s11306-015-0798-3.

Loudet O, Chaillou S, Merigout P, Talbotec J, Daniel-Vedele F: Quantitative trait loci analysis of nitrogen use efficiency in *Arabidopsis*. *Plant Physiol* 2003, 131:345-58

Louthan AM, Kay KM: Comparing the adaptive landscape across trait types: larger QTL effect size in traits under biotic selection. *BMC Evol Biol* 2011, 11:60.

Mach J: Phytol from degradation of Chlorophyll feeds biosynthesis of Tocopherols. *Plant Cell* 2015, 27(10):2676.

Malitsky S, Blum E, Less H, Venger I, Elbaz M, Morin S, Eshed Y, Aharoni A: The transcript and metabolite networks affected by the two clades of *Arabidopsis* glucosinolate biosynthesis regulators. *Plant Physiol* 2008, 148:2021-2049.

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A y col: Finding the missing heritability of complex diseases. *Nature* 2009, 461(7265):747-753.

Maruyama-Nakashita A, Nakamura Y, Tohge T, Saito K, Takahashi H: Arabidopsis SLIM1 is a central transcriptional regulator of plant sulfur response and metabolism. *Plant Cell* 2006, 18:3235-3251.

Matt P, Krapp A, Haake V, Mock H, Stitt M: Decreased Rubiscoactivity leads to dramatic changes of nitrate metabolism, amino acid metabolism and the levels of phenylpropanoids and nicotine in tobacco antisense RBCS transformants. *Plant J* 2002, 30:663-677.

Matzke M, Kanno T, Huettel B, Daxinger L, Matzke AJ: RNA-directed DNA methylation and Pol IVb in Arabidopsis. *Cold Spring Harb Symp Quant Biol* 2006, 71:449-59.

Meyer S, Pospisil H, Scholten S (2007). Heterosis associated gene expression in maize embryos 6 days after fertilization exhibits additive, dominant and overdominant pattern. *Plant Mol. Biol.* 63: 381-391.

Milone D y col: *omeSOM: a software for clustering and visualization of transcriptional and metabolite data mined from interspecific crosses of crop plants. *BMC Bioinformatics* 2010, 11:438.

Milone DH y col: Clustering biological data with SOMs: On topology preservation in non-linear dimensional reduction. *Expert Systems with Applications* 2013: 40(9)3841-3845.

Mitchell-Olds T, Pedersen D: The molecular basis of quantitative genetic variation in central and secondary metabolism in Arabidopsis. *Genetics* 2003, 149:739-47

Ni Z, Kim E-D, Ha M, Lackey E, Liu J, Zhang Y, Sun Q, Chen ZJ Altered circadian rhythms regulate growth vigour in hybrids and allopolyploids. *Nature* 2009, 457, 327-331.

Pascual L y col: Deciphering genetic diversity and inheritance of tomato fruit weight and composition through a systems biology approach. *J. Exp Bot* 2013, 64(18):5737-52.

Pascual, L. y col: Potential of a tomato MAGIC population to decipher the genetic control of quantitative traits and detect causal variants in the resequencing era. *Plant Biotechnol J* 2015, 13: 565-577. doi:10.1111/pbi.12282

Powers L.: An expansion of Jones's theory for the explanation of heterosis. *Am. Nat* 1944, 78: 275-280.

Quadrana L, Almeida J, Asís R, Duffy T, Dominguez PG, Bermúdez L, Conti G, Corrêa da Silva JV, Peralta IE, Colot V, Asurmendi S, Fernie AR, Rossi M, Carrari F: Natural

occurring epialleles determine vitamin E accumulation in tomato fruits. *Nature Commun.* 2014, 26; 5:3027. doi: 10.1038/ncomms5027.

Quadrana L, Almeida J, Otaiza S, Duffy T, Corrêa da Silva JV, de Godoy F, Asís R, Bermúdez L, Fernie AR, Carrari F, Rossi M: Transcriptional regulation of tocopherol biosynthesis in tomato. *Plant Mol. Biol* 2012.

Quadrana L, Rodriguez MC, López M, Bermúdez L, Nunes-Nesi A, Fernie AR, Descalzo A, Asís R, Rossi M, Asurmendi S, Carrari F: Coupling virus-induced gene silencing to exogenous green fluorescence protein expression provides a highly efficient system for functional genomics in *Arabidopsis* and across all stages of tomato fruit development. *Plant Phys* 2011, 156(3), 1278-91.

Robinson MD: "edgeR: a Bioconductor package for differential expression analysis of digital gene expression data". *Bioinformatics* 2010, 26, pp. -1.

Rossi M, Bermúdez L, Carrari F: Crop yield: challenges from a metabolic perspective. *Curr Opin Plant Biol* 2015, 20;25:79-89.

Schauer N, Semel Y, Balbo I, Steinfath M, Repsilber D, et al. 2008. Mode of inheritance of primary metabolic traits in tomato. *Plant Cell* 20:509-23

Schillmiller AL, Shi F, Kim J, Charbonneau AL, Holmes D y col: Mass spectrometry screening reveals widespread diversity in trichome specialized metabolites of tomato chromosomal substitution lines. *Plant J* 2010, 62:391-403

Segura V, Vilhjalmsón BJ, Platt A, Korte A, Seren U, Long Q, Nordborg M: An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nat Genet* 2012, 44 (7):825-830. 40.

Semel Y, Nissenbaum J, Menda N, Zinder M, Krieger U, Issman N, Pleban T, Lippman Z, Gur A, Zamir D Overdominant quantitative trait loci for yield and fitness in tomato. *PNAS* 2006, 103(35), 12981-6.

Shen H, He H, Li J, Chen W, Wang X, Guo L, Peng Z, He G, Zhong S, Qi Y, Terzaghi W, Deng XW Genome-wide analysis of DNA methylation and gene expression changes in two *Arabidopsis* ecotypes and their reciprocal hybrids. *Plant Cell* 2012, 24, 875-892.

Shull GH: The composition of a field of maize. *Am. Breeders Assoc* 1908 Rep. 4: 296-301.

Sonderby I, Hansen B, Bjarnholt N, Ticconi C, Halkier B, Kliebenstein D: A systems biology approach identifies a R2R3 MYB gene subfamily with distinct and overlapping functions in regulation of aliphatic glucosinolates. *PLoS ONE* 2007, 2:e1322.

Springer NM, Stupar RM: Allele-specific expression patterns reveal biases and embryo-specific parent-of-origin effects in hybrid maize. *Plant Cell* 2007. 19: 2391-2402.

Stegmayer G: A Biologically-inspired Validity Measure for Comparison of Clustering Methods over Metabolic Datasets. *IEEE/ACM Trans. Comput. Biol. Bioinf* 2012, 9 (3): 706-716.

Steinhauser MC: Enzyme Activity Profiles during Fruit Development in Tomato Cultivars and *Solanum pennellii*. *Plant Physiol* 2010, 153(1):80-98.

Stocks y col: The UEA sRNA workbench: a suite of tools for analysing and visualizing next generation sequencing microRNA and small RNA datasets. *Bioinformatics* 2012, Vol. 28,pp. 2059-2061

Strassman M, Ceci LN: Enzymatic formation of alpha-isopropylmalic acid, an intermediate in leucine biosynthesis. *J Biol Chem* 1963, 238:2445-2452.

Stupar RM, Gardiner JM, Oldre AG, HaunWJ, Chandler VL, Springer NM: Gene expression analyses in maize inbreds and hybrids with varying levels of Heterosis. *BMC Plant Biol* 2008, 8: 33.

Sulem P, Gudbjartsson DF, Stacey SN, Helgason A, Rafnar T, Magnusson KP, Manolescu A, Karason A, Palsson A, Thorleifsson G y col: Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nat Genet* 2007, 39(12):1443-1452.

Tarailo-Graovac M, Chen N: Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics* 2009, Chapter 4: Unit 4.10. doi: 10.1002/0471250953.bi0410s25.

The Tomato Genome Consortium: 312 authors. Tomato genome sequencing and comparative analysis reveal two consecutive triplications that spawned genes influencing fruit characteristics. *Nature* 2012, 485:635-641.

Tian F, Bradbury PJ, Brown PJ, Hung H, Sun Q, Flint-Garcia S, Rocheford TR, McMullen MD, Holland JB, Buckler ES: Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nat Genet* 2011, 43(2):159-16

Tieman DM, Zeigler M, Schmelz EA, Taylor MG, Bliss P y col: Identification of loci affecting flavour volatile emissions in tomato fruits. *J. Exp. Bot* 2006, 57:887-96

Uzarowska A, Keller B, Piepho HP, Schwarz G, Ingvarsdson C, Wenzel G, Lubberstedt T: Comparative expression profiling in meristems of inbred-hybrid triplets of maize based on morphological investigations of heterosis for plant height. *Plant Mol Biol* 2007, 63: 21-34.

Vidi PA, Kessler F y Bréhélin C: Plastoglobules: a new address for targeting recombinant proteins in the chloroplast. *BioMed Central* 2007, <https://doi.org/10.1186/1472-6750-7-4>

Wei G, Tao Y, Liu G, Chen C, Luo R, Xia H, Gan Q, Zeng H, Lu Z, Han Y, Li X, Song G, Zhai H, Peng Y, Li D, Xu H, Wei X, Cao M, Deng H, Xin Y, Fu X, Yuan L, Yu J, Zhu

Z, Zhu L. (2009): A transcriptomic analysis of superhybrid rice LYP9 and its parents. PNAS 2009, USA 106, 7695-7701.

Xu J, et al. (2013) An extensive proteome map of tomato (*Solanum lycopersicum*) fruit pericarp. Proteomics 2013, 13(20):3059-63.

Yang X, Kundariya H, Xu YZ, Sandhu A, Yu J, Hutton SF, Zhang M, Mackenzie SA: MutS HOMOLOG1-derived epigenetic breeding potential in tomato. Plant Physiol 2015, 168(1):222-32

Zanor MI, Rambla JL, Chaib J, Steppa A, Medina A, y col: Metabolic characterization of loci affecting sensory attributes in tomato allows an assessment of the influence of the levels of primary metabolites and volatile organic contents. J. Exp Bot. 2009, 60:2139-54

Zeng P, Zhao Y, Qian C, Zhang L, Zhang R y col: Statistical analysis for genome-wide association study. J. Biomed. Res 2015, 29:285-97

Zhang HY, He H, Chen LB, Li L, Liang MZ, Wang XF, Liu XG, He GM, Chen RS, Ma LG, Wang D : A genome-wide transcription analysis reveals a close correlation of promoter INDEL polymorphism and heterotic gene expression in rice hybrids. Mol Plant 2008, 1: 720-731.

Zhao K, Aranzana MJ, Kim S, Lister C, Shindo C, Tang C, Toomajian C, Zheng H, Dean C, Marjoram P y col: An Arabidopsis example of association mapping in structured samples. PLoS Genet 2007, 3(1):e4

Zhao K, Aranzana MJ, Kim S, Lister C, Shindo C, Tang C, Toomajian C, Zheng H, Dean C, Marjoram P: An Arabidopsis example of association mapping in structured samples. PLoS Genet 2007, 3(1):e4.

Zhong S, Fei Z, Chen YR, Zheng Y, Huang M, Vrebalov J, McQuinn R, Gapper N, Liu B, Xiang J, Shao Y, Giovannoni JJ: Single-base resolution methylomes of tomato fruit development reveal epigenome modifications associated with ripening. Nat Biotechnology 2013, 31(2):154-9.

Páginas 96 a 115 eliminadas a pedido del autor.