



UNIVERSIDAD NACIONAL DE ROSARIO
Facultad de Ciencias Agrarias

**‘GENÓMICA Y TRANSCRIPTÓMICA DE LA
MADUREZ DEL FRUTO DE TOMATE:
APLICACIONES EN EL MEJORAMIENTO
GENÉTICO’**

Licenciado en Genética Paolo Cacchiarelli

Tesis para optar al título de
DOCTOR EN CIENCIAS AGRARIAS

DIRECTORA: Dra. Tapia, Elizabeth

CO- DIRECTOR: Dr. Pratta, Guillermo R.

AÑO 2022

**‘GENÓMICA Y TRANSCRIPTÓMICA DE LA MADUREZ DEL FRUTO DE TOMATE:
APLICACIONES EN EL MEJORAMIENTO GENÉTICO’**

PAOLO CACCHIARELLI

Licenciado en Genética - Universidad Nacional de Morón

Esta Tesis es presentada como parte de los requisitos para optar al grado académico de Doctor en Ciencias Agrarias de la Universidad Nacional de Rosario y no ha sido previamente presentada para la obtención de otro título en ésta u otra Universidad. La misma contiene los resultados obtenidos en investigaciones llevadas a cabo en la Facultad de Ciencias Agrarias de la Universidad Nacional de Rosario, durante el período comprendido entre Mayo del 2016 y Julio del 2022, bajo la dirección de la Dra. Elizabeth Tapia y la co-dirección de Dr. Pratta, Guillermo R.



Lic. Cacchiarelli Paolo

Nombre y firma del Doctorando



Dra. Tapia Elizabeth

Nombre y firma de la Directora



Dr. Pratta Guillermo R.

Nombre y firma del Co-director

Defendida el 29 de Julio del 2022

AGRADECIMIENTOS

Quiero agradecer en primer lugar a mis directores Dra. Elizabeth Tapia y Dr. Guillermo Pratta, por toda la dedicación y guía en las actividades de esta tesis, pero principalmente por transmitir su pasión y el apoyo incondicional más allá de lo académico.

A los integrantes de la Facultad de Cs. Agrarias de la UNR, personal no docente y personal en general por hacer grato todo el transcurso de mi tesis.

A los integrantes del CIFASIS, Javi, Joaco, Nacho, Naty, José, Nico, Pilar, Laura, Eli por darme su apoyo constante en todas las etapas de mi aprendizaje y en especial a Flavio que estuvo presente en muchas actividades y correcciones de esta tesis.

A los integrantes de la Cátedra de Genética y todos sus ayudantes, Roxi, Liliana, Gusti, Fer, Graciela, Gaby, Javi, Vladi, y ayudantes Cami y Felipe, por estar presente en cualquier momento y a disposición para llevar adelante cualquier tarea.

A los integrantes de la Cátedra de Fitopatología, Rosanna, Miriam, Ale, Facu, Orne, y también Facu, Franco y Mario que ya no están pero siempre están presentes brindándome todo su apoyo y sus tiempos para poder llevar adelante mi trabajo doctoral.

A mis compañeros de la Sala de Becarios, Meli, Fede, Pau, Lu, Franco, Agu, Cele, Vale, Caro, Baby, Virgi y Flor que fueron y son tanto colegas como amigos en el cual compartimos un grato ambiente de trabajo, pero en especial por su apoyo y acompañamiento durante todo el doctorado.

A los integrantes del Acuario del Río Paraná, Vani, Vicky, Mari, Flor, Marti, Edu, y Sofí que en mi breve paso por la institución me hicieron sentir uno más del grupo y especialmente a Nacho y Marti que me brindaron todo su tiempo y apoyo en las tareas.

A la Dra. Debora Arce, por ser siempre parte del equipo ayudando con las diferentes tareas al igual que la Dra. Flavia Krsticevic que siempre estuvo desde los comienzos y sigue a pesar de la distancia. Agradezco también a todo el personal del INCA-RJ y CIT, por compartir momentos de trabajo y ser parte de esta tesis.

Al personal no docente de la Sección Horticultura, cátedra de Horticultura y a los CPA, en principal David y Toni, que siempre estuvieron a disposición de todo para llevar a cabo los ensayos y experimentos.

A CONICET y ANPCYT por las becas otorgadas para poder cumplir este sueño y hacerlo realidad.

A todos los que me ayudaron con el día a día en este desafío y hacerlo mucho más agradable. A Facu Hernandez en especial, que hicimos todo de manera conjunta y paralela,

desde la tecnicatura hasta el doctorado, y siendo uno el sostén del otro antes las situaciones más difíciles. Vicky que estuvo siempre desde mis comienzos hasta hoy en día, por ayudarme y enseñarme. Franco y David, por estar siempre de manera proactiva para ayudar, desde el traslado hasta el desarrollo de los experimentos y Valen por estar en esta última etapa y siempre dispuesta para colaborar con todo. Gracias por siempre estar ahí y poder contar con Uds, y ya saben que son más que colegas y amigos. Me llevo sus amistades para siempre.

Agradezco principalmente a mi familia que siempre me apoyaron en mis decisiones y están siempre presentes ante cualquier adversidad. Siempre se encuentran disponibles para ser el sostén que necesito. Les dedico todo este trabajo hecho a Uds que siempre me motivan y que son el pilar constante en todo lo que me planteo.

A mis amigos personales que son parte de mi vida cotidiana y están siempre para hacer todo mucho más divertido, con muchas sonrisas y carcajadas en todo momento.

Para Nico, Tanque y Rocio,

PUBLICACIONES Y PRESENTACIONES A CONGRESOS

- Detección de *QTLs* de interés agronómico mediante marcadores *InDel* desarrollados en base a enfoques ómicos en tomate (*Solanum lycopersicum*). Cacchiarelli P., Tapia, E. Pratta, G.R. Asociación Latinoamericana de Genética. 2021. Valdivia, Chile. (Premiado como mejor trabajo en el área 'Mejoramiento vegetal').
- Desarrollo de marcadores moleculares en base a estudios ómicos sobre dos progenitores de diferentes poblaciones base de un programa de mejoramiento genético de tomate. Cacchiarelli P., Tapia, E. Pratta, G.R. VI Jornadas de Ciencia y Tecnología 2021. Facultad de Ciencias Agrarias. Universidad Nacional de Rosario. 2021. Zavalla, Santa Fe.
- Transcriptome analyses of *Solanum pimpinellifolium* LA0722 at three ripening stages. Cacchiarelli P., Tapia, E. Pratta, G.R. SOL International Online Meeting 2020. Wageningen, Países Bajos.
- Ensamblado *de novo* del transcriptoma en tres estados de madurez del fruto de la accesión silvestre de tomate *Solanum pimpinellifolium* LA0722. Cacchiarelli P., Tapia, E. Pratta, G.R. SAG - Sociedad Argentina de Genética. 2020. Buenos Aires, Argentina. (Premiado como mejor trabajo en el área 'Genómica y genética molecular vegetal').
- Interactómica entre transcriptos en dos estados de madurez del fruto de dos genotipos de tomate (*Solanum* spp.) discrepantes para la calidad. Cacchiarelli P. Arce DP., Tapia E. y Pratta GR. Sociedad Nacional de Biología. 2020. Mendoza, Argentina.
- Estimación de heredabilidad en familias F_4 de tomate mediante análisis de la varianza. Cacchiarelli, P. Del Médico, A.P. Cabodevila, V., Iglesias, F. Tapia, E. Pratta, G.R. GAB - XXIV Reunión Científica del Grupo Argentino de Biometría. 2019. Mendoza, Argentina.
- Heredabilidad en sentido estricto para caracteres organolépticos en una generación F_4 de tomate (*Solanum* spp.). Cacchiarelli, Paolo; Spetale, Flavio E.; Tapia, Elizabeth; Pratta, Guillermo R. SBR 2019. Rosario, Argentina.
- Identificación del número de lóculos en frutos de tomate (*Solanum lycopersicum* L.) usando visión por computadora. Spetale, F; Vazquez, D; Murillo, J; Cacchiarelli, P; Rodríguez, G; Tapia, E. Sociedad Biología de Rosario. 2019. Rosario, Argentina.
- Inducción diferencial de HSPs evaluada por ARN-Seq en dos especies de tomate (*Solanum* spp.) durante la madurez del fruto. Arce D. P., P. Cacchiarelli, M. D. Giménez, G.R. Rodríguez, G.R. Pratta. ALAG 2019. Mendoza, Argentina.
- *de novo* assembly of tomato genomes with emphasis in chromosome 6 sHSPs. Cacchiarelli P, Krsticevic FJ, Ezpeleta J, Tapia E. and Pratta GR. IX Argentinian Congress of

Bioinformatics and Computational Biology 9CAB2C. Mar del Plata, Bs. As., Argentina. 2018.

- *In silico* analysis of gene copy number of small heat shock proteins in tomato wild species. Cacchiarelli P, Krsticevic FJ, Spetale F, Pratta GR, and Tapia E. V International Society for Computational Biology Latin America (ISCB-LA SOIBIO EMBnet 2018). Viña del Mar, Chile.
- Detección de HSEs en promotores sHSPs en dos genotipos parentales de un programa de mejoramiento genético de tomate argentino. Arce DP, Cacchiarelli P, Tapia E y Pratta GR. III Jornadas regionales de Genética del Litoral. 2018. Est. Exp. INTA Rafaela. Santa Fe, Argentina.
- Functional characterization of tandem duplicated sHSPs genes in tomato, potato and Arabidopsis genomes. Cacchiarelli P, Arce DP, Tapia E and Pratta GR. VIII Argentinian Bioinformatics and Computational Biology Congress. 2017, Posadas (Misiones), Argentina.
 - Exploración de efectos genéticos en genotipos de primer y segundo ciclo de tomate. Cacchiarelli, P.; Cambiaso, V.; Pereira Da Costa, J.; Tapia, E.; Rodríguez, G.R.; Pratta, G.R. Congreso Interamericano de Estadística. Sociedad Argentina de Estadística. Rosario, 2017.
- Tipificación de materiales homocigotas y heterocigotas de primer y segundo ciclo en tomate (*Solanum spp.*). Cacchiarelli, P.; Cambiaso, V.; Pereira Da Costa, J.; Tapia, E.; Rodríguez, G.R.; Pratta, G.R. XIX Congreso y XXXVIII Reunión Anual de la SBR. ISSN 2314-1484. Vol 1, Pág. 37. Rosario, 2017.
- The importance of being duplicated in the tomato sHSP gene family. Arce DP, Spetale FE, Cacchiarelli P, Krsticevic FJ, Ponce SD, Pratta GR and Tapia E. IV Congreso Colombiano de Bioinformática y Biología Computacional & VIII Conferencia Iberoamericana de Bioinformática 2017. Santiago de Cali, 2017. Colombia.
- The microsynteny of genes functionally related with the fruit ripening process in Solanaceae species. Cacchiarelli P, Arce DP, Ezpeleta J, Pratta GR, Tapia E and Krsticevic J. ISCB-Latin America Conference 2016 y 7^{mo} CA2BC, 2016. Buenos Aires, Argentina.
- Advances in exploring the diallel approach to dichotomic biological data. Macat, P.; Kovalevsky, L.; Cacchiarelli, P.; Quaglino, M.; Pratta, G.R. 2016. ISCB-Latin America Conference 2016 y 7^{mo} CA2BC, 2016. Buenos Aires, Argentina.

ARTÍCULOS PUBLICADOS EN REVISTAS CIENTÍFICAS

- Structural and functional analysis of two sHSP subfamilies in tomato ripening (2021). Cacchiarelli P., Arce D., Tapia E., Pratta G.R. Plant gene, ELSEVIER BV. Netherlands, Universities and research institutions in the Netherlands. <https://doi.org/10.1016/j.plgene.2021.100297>
- *LocAnalyzer*: A computer vision method to count locules in tomato fruits (2020). Flavio E. Spetale, Javier Murillo, Dana V. Vazquez, Paolo Cacchiarelli, Gustavo R. Rodríguez, Elizabeth Tapia. Computers and Electronics in Agriculture. ELSEVIER SCI-LTD. Amsterdam, Países Bajos. <https://doi.org/10.1016/j.compag.2020.105382>
- Regulatory motifs found in the small heat shock protein (sHSP) gene family in tomato (2018). Arce D., Spetale F., Krsticevic F., Cacchiarelli P., De Las Rivas J., Ponce S., Pratta G.R., Tapia E. ISSN: 1471-2164 - BMC Genomics - BIOMED CENTRAL LTD. Londres. <https://doi.org/10.1186/s12864-018-5190-z>
- Identificación de QTL en las generaciones segregantes de un híbrido de segundo ciclo de tomate (2017). Cabodevila, V.; Cacchiarelli, P.; Pratta, G.R. Revista de la Facultad de Ciencias Agrarias UNCuyo. Universidad Nacional de Cuyo. 49 (2): 1-18. Mendoza. <https://revistas.uncu.edu.ar/ojs3/index.php/RFCA/issue/view/220>
- Segregación y recombinación genética en las generaciones segregantes de un híbrido de segundo ciclo de tomate (2015). Cacchiarelli P., Cabodevila V.G., Pantuso F.S. y Pratta G.R. Revista de la Facultad de Ciencias Exactas, Químicas y Naturales de la Universidad de Morón. Vol: 13, Pag: 25-38. Morón. <https://www.unimoron.edu.ar/area/exactas/stream/dd9a715d7-2-pdf>

ÍNDICE

	Pág.
Agradecimientos	2
Publicaciones y Presentaciones a Congresos	5
Índice	8
Abreviaturas y Símbolos	9
Resumen	10
Abstract	12
Introducción	13
Hipótesis y Objetivo General	20
<u>Capítulo I: Caracterización de los progenitores en los niveles de expresión génica</u>	<u>21</u>
<i>molecular y fenotípica</i>	
Introducción	22
Objetivos Específicos	23
Materiales y Métodos	24
Resultados	30
Discusión	48
Conclusiones	51
<u>Capítulo II: Análisis de la expresión génica de sHSPs en los progenitores y desarrollo...</u>	<u>52</u>
<i>de marcadores moleculares que pongan en evidencia el polimorfismo molecular</i>	
Introducción	53
Objetivos Específicos	54
Materiales y Métodos	55
Resultados	58
Discusión	67
Conclusiones	69
<u>Capítulo III: Caracterización fenotípica y molecular de la población segregante F_4 de...</u>	<u>70</u>
<i>un HSC y detección de QTLs asociados a caracteres de calidad de fruto</i>	
Introducción	71
Objetivos Específicos	72
Materiales y Métodos	73
Resultados	77
Discusión	90
Conclusiones	93
<u>Capítulo IV: Exploración de la región génica del cromosoma 06 de las sHSPs por</u>	<u>94</u>
<i>secuenciación de tercera generación</i>	
Introducción	95
Objetivos Específicos	96
Materiales y Métodos	97
Resultados	100
Discusión	108
Conclusiones	111
Conclusión General	112
Consideraciones Finales	113
Bibliografía	114

ABREVIATURAS Y SÍMBOLOS

PCR: *Polymerase Chain Reaction* o reacción en cadena de la polimerasa

QTL: *Quantitative Trait Loci*, o *loci* de caracteres cuantitativos

SNP: *Single Nucleotide Polymorphism* o polimorfismos de nucleótido simple

SS: Sólidos Solubles

VP: Vida Poscosecha

VM: Verde Maduro

RM: Rojo Maduro

Pi: Pintón

HPC: Híbrido de Primer Ciclo

HSC: Híbrido de Segundo Ciclo

GO: Gene Ontology

PB: Procesos Biológicos

a/b o valor *chroma*: Relación entre absorbancia a una longitud de onda de 540 nm (a) y 675 nm (b)

AT: Acidez Titulable

GDG: Grado de Determinación Genética

InDel: Inserciones/Deleciones

L: Porcentaje de reflectancia

kb: kilobases

Log₂FC: Logaritmo en base 2 del *Fold Change*

RESUMEN

El tomate cultivado (*Solanum lycopersicum* L.) es una de las Solanáceas de mayor importancia económica. La superficie cultivada dedicada a este cultivo se distribuye en gran parte del territorio a nivel global. Uno de los destinos de su producción es el de consumo en fresco y un carácter de fundamental importancia es la prolongación de la vida poscosecha, ya que las pérdidas poscosecha en los países en desarrollo representan casi el 50 % de lo producido (Meli *et al.*, 2010). Un grupo de proteínas denominadas *small Heat Shock Proteins* (sHSPs) desempeñan diversas funciones de regulación en la maduración y por lo tanto podrían extender la vida poscosecha. Es de destacar, que estas sHSPs interactúan con la fitohormona clave de los frutos climatéricos, el etileno. En esta tesis se abordó el estudio de la participación de las sHSPs en el proceso de maduración del fruto de tomate a través de enfoques transcriptómicos y genómicos, en el cual se desarrollaron marcadores moleculares a partir de dicha información con el fin de obtener nuevas estrategias de mejoramiento genético de la especie y asistir al programa de mejora. Se realizaron estudios transcripcionales mediante secuenciación mediante tecnología Illumina® en el cv. Caimanta (C), la accesión silvestre LA0722 (P) y su F₁ (CxP) en los estados de maduración de Verde Maduro (VM), Pinton (Pi) y Rojo Maduro (RM). En total se obtuvieron un promedio de 22.541.809 lecturas por cada biblioteca. Se detectaron genes con Expresión Diferencial (ED) significativa entre los estados Pi vs. VM y RM vs. VM para los 3 genotipos considerados uniformes de manera global. Se realizó una caracterización y análisis de la funcionalidad de genes en donde se detectaron 4 términos GO de la ontología génica (GO, del inglés *gene ontology*) relacionados a procesos biológicos en donde participan genes relacionados a la maduración. Los ID de los genes en cada GO fueron identificados y comparados, obteniendo un total de 2744 genes únicos entre ellos, evidenciando una gran riqueza en este enfoque. A fin de refinar el análisis global de genes, se enfocó en la subfamilia de sHSPs previamente estudiadas *in silico*. Se identificaron una decena de genes de sHSPs con expresión diferencial en los progenitores y su cruzamiento. Siguiendo los análisis de las sHSPs, se estudiaron aquellas detectadas previamente por Arce *et al.* (2015, 2016) y Krsticevic *et al.* (2016). El gen Solyc06g076540.1 se destacó por su mayor expresión diferencial y significancia entre los estados madurativos. Con el fin de detectar efectos genéticos a través de estudios de expresión entre los progenitores se realizaron comparaciones de los transcriptomas en el mismo estado de maduración (C vs P para VM, C vs P para Pi y C vs P para RM). Esto evidenció que la sHSPs Solyc06g076540.1 mantuvo una expresión diferencial constante en los 3 estados. A partir de esta información del

análisis transcriptómico y genómico se desarrollaron 3 *InDels* y 1 *SNP* a través del polimorfismo detectado en las secuencias génicas de los progenitores. El enfoque genómico se desarrolló con tecnología de secuenciación de segunda y tercera generación. Se realizó la caracterización fenotípica y molecular de la población F_4 obtenida del cruzamiento ToUNR18 x ToUNR1. En la caracterización fenotípica se analizaron caracteres morfológicos y organolépticos. Los resultados de los caracteres morfológicos mostraron h^2 altas (mayores a 0,72) excepto para vida poscosecha que tuvo un valor intermedio (0,46). Por el lado de los caracteres organolépticos fue de medio a bajo (0,33 fue el máximo para dureza). Se caracterizó molecularmente una generación F_4 con *InDels* y *SNP* desarrollados y se detectaron *QTLs* robustos para peso, altura y pH. Por último, con el fin de relevar la región de estas sHSPs del cromosoma 6 con mayor certidumbre se realizó una secuenciación de la región en la plataforma ONT (del inglés, *Oxford Nanopore Technologies*) de secuenciación de tercera generación. Estudios *in silico* previos del grupo habían sugerido una variación en el número de copias de estas sHSPs que podrían distorsionar los análisis de expresión diferencial e intervenir en el proceso de maduración. Esta tecnología es capaz de reducir fuertemente las dificultades del ensamblaje en regiones repetitivas y duplicadas en *tándem*. Se desarrollaron 27 pares de cebadores para amplificar dicha región y se secuenció los genotipos C, P y su F_1 . Se lograron secuencias con longitudes entre 400 pb y 6000 pb. Se ensambló la región y se obtuvieron 2 *contigs* para el cv. Caimanta sumando un total de 12011 pb, la accesión silvestre LA0722 obtuvo 3 *contigs* con una longitud de 10006 pb y la F_1 4 *contigs* con un total de 13945 pb. Los resultados obtenidos permiten disponer de secuencia genómica más confiable a nivel de regiones con genes duplicados en *tándem* o regiones repetitivas con el fin de continuar con el desarrollo de marcadores moleculares que asistan al programa de mejora. Integrar la información transcriptómica mediante un enfoque basado en el conocimiento de la secuencia genómica permite estudiar posibles genes candidatos implicados en la determinación de los rasgos agronómicos de interés y así obtener nuevas variedades con larga vida poscosecha a través del mejoramiento genético de la especie.

PALABRAS CLAVES

Solanum spp. - expresión génica - mejoramiento genético vegetal - maduración del fruto

ABSTRACT

Genomic and transcriptomic of the tomato fruit ripening: application to plant breeding

Tomato (*Solanum lycopersicum* L.) is one of the most economically important Solanaceae. An important trait is the prolongation of post-harvest life since postharvest losses in developing countries represent 50% of production. The small Heat Shock Proteins (sHSP) gene family shows a high copy number of members, involved in stress and ripening processes. We have studied the genomic structure and function of sHSP subfamily members in *Solanum lycopersicum* cv. Caimanta, *Solanum pimpinellifolium* accession LA0722 and their F_1 , which are the parental lines of different breeding populations. The goal was to study the participation of sHSPs in the fruit ripening process and develop new strategies for genetic breeding. Transcriptomes were obtained in cv. Caimanta (C), the wild accession LA0722 (P) and its F_1 (CxP) at the mature green (MG), breaker (B) and red ripe (RR), main ripening stages. For C, 6159 and 6922 genes with differential gene expression (DGE) were detected between B vs MG and RR vs MG respectively, in P there were 7313 and 6929 genes with differential gene expression (DGE) were detected between B vs MG and RR vs MG respectively, and for the F_1 it was 2715 and 6994 genes with differential gene expression (DGE) were detected between B vs. MG and RR vs. MG respectively. Gene functionality analysis was performed and 4 Gene Ontology (GO) terms were detected. We identified 10 sHSP genes with DGE in C, 5 genes only in the RR state and 5 shared between the B and RR states. In P there were a total of 6 sHSP genes with DGE, detecting 4 shared genes between both states, and a unique gene in B and RR. Three *InDels* and a *SNP* were developed through the polymorphism detected in genome sequences of parental genotypes and a segregating population (F_4) was molecularly characterized. We have identified three Quantitative Trait Loci (QTL) for weight, height, and pH, respectively. Previous studies evidenced variation in the copy numbers of sHSP in chr06. This region was sequenced using 3rd generation technology. The region was assembled and we obtained 2 contigs for cv. Caimanta, 3 contigs for wild accession and 4 contigs for F_1 . Results obtained allowed integrating the information from the transcriptomic study with an approach based on the knowledge of the genomic sequence and thus developing molecular markers that allow assisting the genetic improvement of the species to obtain new elite varieties.

INTRODUCCIÓN

El tomate cultivado (*Solanum lycopersicum* L.) es la segunda Solanácea de mayor importancia económica en el mundo, luego de la papa. La superficie cultivada dedicada a este cultivo representa el 0,6% a nivel mundial, y en la Argentina se lo cultiva tanto en cinturones hortícolas como en zonas de producción especializada, siendo la especie con mayor superficie en producción bajo invernadero (Ferratto *et al.*, 2006). El mercado de tomate se divide principalmente en mercado fresco y tomate para industria. La producción promedio anual de tomate argentino de los últimos años se ubica en torno a 1.100.000 toneladas y con un alcance de 17.000 hectáreas destinadas a la producción. De este total, aproximadamente un 60-70 % se destina a mercado para consumo en fresco y un 30-40 % para industria. Nuestro país posee regiones donde se desarrolla la producción asegurando el abastecimiento local, recurriendo a importaciones desde países limítrofes ante situaciones climáticas desfavorables que dificultan cubrir la demanda nacional. El consumo nacional ronda los 16 kg/persona/año y dentro del rubro hortícola sólo es superado por la papa¹. Argentina exporta pequeñas cantidades de su producción de tomate para consumo en fresco, siendo Paraguay el destino principal, seguido por Brasil y Uruguay. Diferentes provincias de la Argentina son las que concentran el mayor porcentaje de esta superficie (Buenos Aires, Corrientes, Salta, Santa Fe y Mendoza), sin embargo, los destinos de los tomates que en esas zonas se cosechan no son los mismos y ahí radica la principal diferencia entre ellas.



Principales zonas productoras de tomate en Argentina. (Fuente: Elaborado por la Dirección de Producción Agrícola en base a datos de MCBA, SENASA y TOMATE 2000).

¹ www.argentina.gob.ar/sites/default/files/produccion-tomate-argentina-diciembre-2020.pdf

De los dos destinos mencionados que puede tener la producción nacional y mundial de tomate, pero principalmente en la destinada a consumo en fresco, un carácter de fundamental importancia es la prolongación de la vida poscosecha de los frutos, ya que las pérdidas poscosecha de frutas y hortalizas en los países en desarrollo representan casi el 50 % de lo producido (Meli *et al.*, 2010). Esta problemática se relaciona a su vez con el transporte, logística y comercialización de los mismos. En consecuencia, tomates larga vida son altamente apreciados para su producción y comercialización como fruto fresco, y así la vida poscosecha se considera un rasgo altamente importante en los programas de mejoramiento genético de esta especie. Este atributo es una variable agronómica de gran importancia biológica, histórica y económica para este cultivo. La obtención de nuevas variedades de larga vida poscosecha es la meta de muchas investigaciones, y hasta el momento estuvo limitada a la introgresión de genes mutantes para la maduración *nor* (*non ripening*), *rin* (*ripening inhibitor*), *alc* (*alcohaca*) y *Nr* (*never ripe*) en fondos genéticos élites a través de técnicas convencionales dentro de los programas de mejoramiento. Esta estrategia puede llevarse a cabo por transgénesis o cruzamientos con estos mutantes espontáneos que retrasan la madurez, pero poseen implicancias socioeconómicas desfavorables, al modificar otros caracteres agronómicos de interés de manera negativa o indeseable. En consecuencia es necesario buscar otras estrategias hacia la obtención de nuevas variedades.

En los programas de mejoramiento genético de tomate es usual realizar cruzamientos con especies silvestres para incrementar la variabilidad genética del tomate cultivado. El uso de la biodiversidad relacionada con un cultivo es una alternativa para incrementar la variabilidad disponible, pero junto a la ampliación de la base genética suelen aparecer caracteres agronómicamente indeseables como depresión en el rendimiento, disminución del peso de los frutos, presencia de espinas, pubescencia y/o metabolitos secundarios que afectan la calidad del producto (van de Wouw *et al.*, 2010).

La dispersión de semillas y la coevolución entre los frutos del tomate silvestre y los frugívoros que lo consumen puede explicarse en parte por la persistencia de ciertas características de calidad como aroma, color, sabor y textura del fruto. En términos evolutivos, los individuos poseedores de una senescencia retardada y de características organolépticas que los hagan más atractivos tendrían una ventaja evolutiva respecto a los demás, ya que probablemente sus semillas serían mayormente dispersas. Así, es de esperar que los genes que retrasan la maduración de los frutos hayan sido seleccionados positivamente durante la evolución de las especies silvestres de tomate. A su vez, estos

genes representan herramientas potenciales muy útiles en el mejoramiento vegetal de la variedad cultivada, evitando así los efectos adversos de las estrategias de mejoramiento genético.

El tomate es un cultivo de constitución genética diploide y con 12 cromosomas. Dentro de su clado, se conocen 13 especies de tomate pertenecientes al género *Solanum* sección *Lycopersicon*, 12 especies silvestres y una especie cultivada (Peralta *et al.*, 2008). Las especies silvestres crecen en distintas regiones andinas con los más diversos ambientes que dicha región dispone de manera característica. Esta se encuentra definida por la región Andina de América del Sur comprendida entre Ecuador y el norte de Chile. El tomate fue domesticado en dos etapas con dos centros independientes, una inicial que fue en Perú y Ecuador y una segunda en la región de la península de Yucatán en México (Blanca *et al.*, 2012 y 2015). Su dispersión por el mundo comenzó con la llegada de los conquistadores españoles a América, luego fue llevado de México a Europa y posteriormente se dispersó al resto del mundo actual.

Estudios experimentales llevados a cabo por el grupo de Mejoramiento Genético de Tomate de la Cátedra de Genética FCA-UNR desde los años '90 aportaron el conocimiento básico sobre las modificaciones en la estructura genómica del tomate cultivado *Solanum lycopersicum* (cultivar Caimanta, tipo platense) causadas por la incorporación del germoplasma silvestre de *Solanum pimpinellifolium* (accesión LA0722) para prolongar la vida poscosecha de los frutos. Por consiguiente, los cambios fenotípicos asociados a la introgresión pueden ser evaluados para la generación de nuevos materiales genéticos que satisfagan la demanda de los productores y los consumidores, y que además estén adaptados al ambiente y las formas de cultivo locales. El grupo de investigación tiene amplia trayectoria en dicha área y ha obtenido 18 líneas recombinantes (RILs, siglas del inglés: *Recombinant Inbred Lines*) de tomate obtenidas a partir del cruzamiento interespecífico (Híbrido de Primer Ciclo, HPC) mencionado, mediante selección antagónica y divergente para peso y vida poscosecha de los frutos (Rodríguez *et al.*, 2006). Así, se obtuvieron 16 RILs y otras 2 particulares, una de color amarillo (a pesar de que ambos progenitores poseen color final rojizo) y una con forma perita (a pesar de sus progenitores poseen forma redonda o aplanada). La alta variabilidad presente entre las RILs (Pratta *et al.*, 2011) y la aparición de distintos tipos de efectos génicos en los Híbridos de Segundo Ciclo (HSC) (Pereira da Costa *et al.*, 2016, Cabodevila *et al.*, 2017, Cambiaso *et al.*, 2019^a) demuestra que la variación fenotípica observada responde en gran medida a componentes genéticos. Las generaciones segregantes de los HSC pueden presentar nuevas combinaciones

genéticas entre los alelos favorables que fueron acumulados en las generaciones de autofecundación y selección a través de las cuáles se originaron sus líneas progenitoras (De Toledo *et al.*, 1984; Pooni and Jinks, 1985; Ipsilandis *et al.*, 2006), tanto en especies autógamas (Ferraz de Toledo, 1987) como en alógamas, por lo que representan una población promisoría para realizar nueva selección y detección de nuevos *QTLs* (*Quantitative Trait Loci*, o *loci* de caracteres cuantitativos) o validación de *QTLs* identificados. Se han analizado ya las generaciones F_2 de tres HSC promisorios (Cabodevila *et al.*, 2013), apareciendo el HSC (ToUNR18 x ToUNR1) como el más apropiado para avanzar a F_3 . Mediante el uso de marcadores moleculares es posible estimar con precisión el número, tipo y magnitud de los efectos génicos involucrados en la determinación de un carácter cuantitativo. Varias características del fruto fueron ya asociadas a marcadores moleculares (Tanksley *et al.*, 1996; Grandillo y Tanksley 1996; Bernacchi *et al.*, 1998; Fulton *et al.*, 2000; Causse *et al.*, 2001; Gonzalo y van der Knaap 2008). Estos autores señalaron la importancia de los marcadores moleculares para identificar genes de interés agronómico en especies de interés.

Los caracteres de peso y vida poscosecha de los frutos están estrechamente asociados a la maduración. Esta última se define como un proceso de desarrollo específico en el que la regulación de la expresión génica causa cambios metabólicos y fisiológicos que varían entre individuos, generando una amplia diversidad en las poblaciones naturales y experimentales. El cambio de color es causado por la degradación de la clorofila, la síntesis de pigmentos carotenoides en los cloroplastos y cromoplastos, y los pigmentos fenólicos: antocianinas, flavonoides y pro-antocianidinas (Lancaster *et al.*, 1997). Este cambio se vincula con la evolución de la textura, aroma y sabor que esperan los consumidores de estos productos en el momento de consumo. Durante la maduración de los frutos ocurren cambios fisiológicos, bioquímicos y estructurales que influyen en la producción de compuestos químicos que generan el sabor y aroma correspondientes. Las características sensoriales, principalmente de olor y sabor, juegan un papel central en la calidad de los frutos y dependen del gusto y aroma percibidos por cada consumidor.

El tomate es un fruto climatérico ya que el proceso de maduración de los frutos es regulado por su principal fitohormona, el etileno. Esta dependencia con el etileno ha posicionado la maduración del fruto del tomate como modelo para estudiar la regulación de la síntesis y percepción de esta fitohormona. Este proceso ocurre tanto en los frutos que están aún en la planta, o sea, antes de la cosecha, como en los cosechados antes que maduren. La vida poscosecha está definida por el tiempo de maduración del fruto

transcurrido una vez que es cosechado de la planta hasta que ya no es comercialmente aceptado (Beckles *et al.*, 2012). Se ha definido como un fenómeno de estrés oxidativo que requiere el intercambio continuo de especies reactivas del oxígeno (EROs) (Jiménez *et al.*, 2002). El balance entre la producción de EROs y su eliminación por el sistema antioxidante indicarían que, un grupo de proteínas denominadas *small Heat Shock Proteins* (sHSPs), podrían desempeñar diversas funciones de regulación durante dicho proceso de maduración y por lo tanto extender su vida poscosecha. Este fenómeno es un proceso de estrés en el cual intervienen las sHSPs (Arce *et al.*, 2015; Ré *et al.*, 2017) y en el que estas sHSPs interaccionan con la fitohormona antes mencionada, el etileno (Upadhyay *et al.*, 2020; Shukla *et al.*, 2017). Estas proteínas, que se expresan en respuesta al estrés oxidativo, constituyen una familia multigénica que se caracteriza por el bajo peso molecular (entre 15 y 42 kDa). Las sHSPs presentan actividad de chaperonas moleculares en respuesta al estrés térmico (*heat shock* o HS) previniendo la agregación proteica irreversible dentro de la célula (Basha *et al.*, 2012; Poulain *et al.*, 2010). Sin embargo, el HS no es el único estímulo que induce la expresión de las sHSPs. La síntesis de estas proteínas se induce durante la maduración de los frutos (Lawrence *et al.*, 1997; Löw *et al.*, 2000; Neta-Sharir *et al.*, 2005) y en ciertos estados del desarrollo en *Arabidopsis* y diferentes Solanáceas (Faurobert *et al.*, 2007; Prasinós *et al.*, 2005). De forma característica, esta familia génica es antigua (450 millones de años) ya que se encuentra en la base del árbol de la vida. Así, su presencia está ampliamente distribuida en los tres dominios de organismos, Archaea, Bacteria y Eukarya (Waters, 2012). En la actualidad es conocido que su número es variable entre los organismos. Notablemente en plantas, comparado con otros organismos, su frecuencia es alta: 19 genes en *Arabidopsis thaliana* (Scharf *et al.*, 2001), 39 en *Oryza sativa* (Ouyang *et al.*, 2009) y 51 en *Glycine max* (Lopes-Caitar *et al.*, 2013). Durante la era pre-genómica, el status de la familia sHSP en tomate era difuso, siendo posible caracterizar experimentalmente a sólo 14 sHSPs (Alba *et al.*, 2005; Baniwal *et al.*, 2004; Frank *et al.*, 2009; Lee *et al.*, 2012; Sanmiya *et al.*, 2004). Además, dentro de la secuencia génica de estos genes de sHSPs, el extremo 5' está enriquecido en motivos de secuencia que responden a diferentes fitohormonas relacionadas con la maduración (Goyal *et al.*, 2012), como lo son ABA, etileno, GA y jasmonato de metilo. En la era post-genómica el número de estas sHSPs caracterizadas se duplicó, identificándose un total de 26 genes correspondientes a sHSPs involucrados en diferentes situaciones de estrés en diversos tejidos, incluyendo HS en hojas (Fragkostefanakis *et al.*, 2015) y microsporas (Frank *et al.*, 2009). Arce *et al.* (2015) y Krsticevic *et al.* (2016) propusieron una nueva caracterización de

estos genes de sHSPs con 33 integrantes totales, y además la variación en su expresión a lo largo de la madurez de estas proteínas cumpliría un rol homeostático en relación al mantenimiento de las características de los frutos durante mayores períodos de poscosecha.

Por su parte, la secuenciación del genoma de tomate ha marcado un antes y un después en la historia de este cultivo. En la actualidad, varios genomas vegetales ya han sido secuenciados y entre ellos se encuentra el genoma de tomate, evento que ha convertido a la especie en un referente mundial. Dicha secuenciación fue llevada a cabo en el cultivar Heinz 1706 de la especie *Solanum lycopersicum* y permitió la construcción de un genoma de referencia (The Tomato Genome Consortium, 2012). Sus secuencias genómicas se encuentran disponibles públicamente en grandes bases de datos y pueden usarse como referencia en el ensamblaje de datos generados a partir de nuevas secuenciaciones como también para el desarrollo de marcadores polimórficos. La secuenciación de su pariente evolutivamente más cercano *Solanum pimpinellifolium* LA1589 provee una fuente de recursos genómicos importantes para avanzar en el estudio de ambas especies. Además, hasta la fecha se han llevado a cabo dos grandes proyectos de resecuenciado de genomas de tomate (The 100 Tomato Genome Sequencing Consortium, 2014) alineando respecto del genoma de referencia más de 400 accesiones de tomate tanto de especies silvestres como cultivadas. Por su parte, el grupo de investigación en tomate de la Cátedra de Genética cuenta con las secuencias genómicas de los progenitores utilizados en el programa de mejoramiento genético, el cv. Caimanta de *Solanum lycopersicum* y la accesión LA0722 de *Solanum pimpinellifolium*, obteniendo el conocimiento de la información genética de ambos materiales y de gran utilidad para el desarrollo de marcadores moleculares que permitan continuar con el proceso de mejora genética de la especie (Cambiaso *et al.*, 2015). Toda la información genómica del tomate y de otras especies pertenecientes a la familia de las Solanáceas se encuentra almacenada y disponible públicamente en el sitio de internet de la red genómica de Solanáceas (The Sol Genomics Network²; Mueller *et al.*, 2005; Fernández-Pozo *et al.*, 2015). Este sitio fue creado para almacenar información genómica y fenotípica de cultivos como el tomate, papa, pimiento, berenjena y tabaco, y también provee herramientas bioinformáticas para estudios genómicos y de mejoramiento vegetal, disponible para toda la comunidad científica.

Los avances en las nuevas tecnologías de secuenciación o NGS (del inglés, *Next Generation Sequencing*) y el auge de la técnica *RNA-seq* (Secuenciación del ARN) permiten desarrollar estudios transcripcionales de gran alcance, siendo el objetivo principal la

² <http://solgenomics.net>

detección genes expresados diferencialmente en diversas situaciones de desarrollo, como es el caso de la maduración de los frutos, o bien durante situaciones de estrés abióticos y/o bióticos. Pueden observarse este tipo de experimentos en la web *Tomato Functional Genomics Database*³, base de datos disponible que presenta gran cantidad de información generada a partir de la técnica de *RNA-seq* en diferentes situaciones de estrés y estados madurativos en fruto de tomate, y en el que pueden encontrarse secuencias de genes, secuencias de promotores, perfiles de expresión, y otros resultados de interés. En el caso del material vegetal previamente descrito (progenitores del cruzamiento interespecífico y RILs derivadas del programa de mejoramiento), si bien ya han sido caracterizadas por diferentes tipos de marcadores de ADN, ARN y polipeptídicos, permanece como trabajo vacante integrar los conocimientos generados en un proyecto que utilice herramientas bioinformáticas para optimizar el mejoramiento genético. Esto permite el uso con mayor eficiencia de los recursos disponibles y afrontar el programa de cruzamientos y selección con mayor certidumbre sobre los caracteres a modificar, en especial los antes mencionados, la vida poscosecha de los frutos. La utilización de la técnica de *RNA-seq* previamente nombrada permite caracterizar los niveles de expresión génica global en los estados madurativos del fruto y ponen de manifiesto posibles genes candidatos, que en conjunto con los datos originados a partir de la secuenciación de diferentes genomas de tomates, permiten desarrollar nuevas estrategias para poder continuar con el programa de mejoramiento genético de la especie.

La combinación de conocimiento experto y adopción de nuevas aplicaciones ómicas desarrolladas en plataformas de secuenciación de alto rendimiento provee una nueva vía de desarrollo con foco en caracteres de interés agronómicos predefinidos. Estudios del tipo transcriptómico y genómico son capaces de brindar información sobre la potencial localización en los materiales genéticos de aquellas variantes genéticas relacionadas a estos caracteres de interés. Esta estrategia de diseño tiene por objetivo final la reducción de los tiempos de desarrollo de marcadores moleculares que asistan al mejoramiento genético para la obtención de nuevas variedades, como así su sistematización a la vista de la información de contexto provista por las técnicas ómicas. De esta forma, es posible continuar con esta investigación traslacional dentro del programa de mejoramiento y avanzando hacia la obtención del ideotipo más adecuado a la demanda del mercado nacional e internacional.

³ <http://ted.bti.cornell.edu/>

HIPÓTESIS DE TRABAJO

Las hipótesis de trabajo planteadas son:

1. Durante la maduración del fruto de tomate varía la representación de un grupo de transcritos relacionados a las sHSPs.
2. Existe polimorfismo molecular en la secuencia genómica del cv. Caimanta de *S. lycopersicum* y *S. pimpinellifolium* LA0722 asociadas a la representación diferencial de dichos transcritos.
3. Los marcadores moleculares diseñados en base a este polimorfismo ponen de manifiesto la existencia de variabilidad genética en poblaciones segregantes derivadas de estos progenitores.
4. La variabilidad genética detectada con estos marcadores moleculares está asociada con variabilidad fenotípica para caracteres de interés agronómico.

OBJETIVO GENERAL

Estudiar la participación de las sHSPs en el proceso de maduración del fruto de tomate y planificar nuevas estrategias de mejoramiento genéticos tendientes a obtener genotipos de larga vida poscosecha y calidad comercial.

Para facilitar la presentación y el análisis de los resultados obtenidos, este trabajo de tesis se presenta en cuatro Capítulos en los que se plantean y desarrollan diferentes objetivos específicos:

Capítulo I: Caracterización de los progenitores en dos niveles de expresión: molecular y fenotípico.

Capítulo II: Análisis de la expresión génica de sHSPs en los progenitores y desarrollo de marcadores moleculares en base a su polimorfismo molecular.

Capítulo III: Caracterización fenotípica y molecular de la población segregante F_4 de un HSC y detección de *QTLs* asociados a caracteres de calidad de fruto.

Capítulo IV: Exploración de la región genómica de las sHSPs del cromosoma 06 mediante tecnología de secuenciación de tercera generación.

CAPÍTULO I

CARACTERIZACIÓN DE LOS PROGENITORES EN DOS NIVELES DE EXPRESIÓN: MOLECULAR Y FENOTÍPICO

INTRODUCCIÓN

La información biológica es heredable en todos los seres vivos y está codificada en secuencias de ADN. Estas secuencias son decodificadas y traducidas en unidades estructurales y funcionales, y se denomina expresión génica. La activación de la expresión génica es un proceso estrictamente regulado que depende de las señales que el organismo reciba, tanto de tipo bióticas como abióticas. Un transcriptoma es capaz de brindar información completa acerca de la expresión génica en diferentes condiciones o etapas del desarrollo (Wen *et al.*, 2019). Para dar inicio a esta tesis y comenzar a revelar qué genes están involucrados en el proceso de maduración del fruto que intervengan en la vida poscosecha de los mismos, se desarrolló un análisis de la expresión global de genes a través de la técnica de secuenciación masiva *RNA-seq* como también una caracterización fenotípica en los materiales de primer ciclo, identificando luego posibles asociaciones entre los transcriptos relacionados a la subfamilia proteica objeto de estudio de esta Tesis (las sHSPs) y la variabilidad fenotípica encontrada en los materiales genéticamente uniformes. El análisis de la expresión global mediante la técnica *RNA-seq* genera millones de secuencias de manera simultánea, obteniéndose así transcriptomas de cada genotipo en cada etapa principal de maduración (Shinozaki *et al.*, 2018). La información genética de cada parental fue reorganizada mediante selección artificial a partir de su cruzamiento biparental en cada paso de autofecundación llevado a cabo durante la obtención de las RILs (Rodríguez *et al.*, 2006). Dos de estas RILs fueron nuevamente cruzadas para dar inicio al segundo ciclo de mejoramiento y de donde proviene la población segregante F_4 , cuyo estudio será presentado en Capítulos posteriores a fin de validar las asociaciones eventualmente detectadas entre la variación para los transcriptos asociados a las sHSPs y la variabilidad fenotípica para caracteres de fruto en los materiales uniformes. Estudiar estos materiales uniformes a través de sus transcriptomas permite detectar genes con expresión diferencial y las vías metabólicas representativas del proceso biológico de la maduración. Los genes expresados de forma significativa pueden ser utilizados como marcadores moleculares (MM) para los rasgos agronómicos de interés en la F_4 . Estos MM permiten identificar nuevas variantes génicas asociadas a los caracteres agronómicos y así disminuir los tiempos en la obtención de nuevas variedades con los rasgos deseados. Este proceso, dentro de los programas de mejoramiento genético, es conocido como selección asistida por marcadores. En este Capítulo se planteó un primer abordaje en donde se caracteriza tanto en el nivel molecular como también en el nivel fenotípico a los progenitores del programa de mejoramiento genético y su cruzamiento.

OBJETIVOS ESPECÍFICOS

1. Caracterización de los transcriptomas de los frutos Verde Maduro (VM), Pintón (Pi) y Rojo Maduro (RM) de cv. Caimanta de *S. lycopersicum*, *S. pimpinellifolium* LA0722 y su F₁ interespecífica (CxP).
2. Comparación de los transcriptomas de frutos de *S. pimpinellifolium* LA0722, *S. lycopersicum* cv. Caimanta y su F₁ (CxP) en los tres estados de madurez.
3. Estudio de procesos biológicos relacionados a la madurez del fruto en los que participan transcriptos de expresión diferencial en estos genotipos.
4. Caracterización fenotípica de estos materiales vegetales para atributos de interés agronómicos relacionados a la madurez del fruto.
5. Detección de asociaciones entre la variación en transcriptos específicos de las sHSPs y la variabilidad fenotípica en los progenitores.

MATERIALES Y MÉTODOS

Material Vegetal

Los progenitores utilizados fueron el cultivar Caimanta de *Solanum lycopersicum* L. y la accesión LA0722 de la especie silvestre *S. pimpinellifolium*. El cv. Caimanta se caracteriza por ser un cultivar para consumo en fresco desarrollado en la Estación Experimental INTA Cerrillos de Salta, Argentina, según el esquema de cruzamientos mostrado en la Figura I-1. Este cultivar fue seleccionado ya que presenta un buen perfil fitosanitario frente a diferentes tipos de patógenos. Sus plantas son compactas, con hábito de crecimiento determinado y los frutos son aplanados (mayor diámetro que altura) con un peso promedio de $98,5 \pm 9,9$ g y una vida poscosecha de $9,7 \pm 0,9$ días (Rodríguez *et al.*, 2006). Por su parte, la especie silvestre *S. pimpinellifolium* LA0722 presenta frutos de menor tamaño y peso que los cultivares comerciales, pero se diferencia de estos por su alta calidad nutritiva. Esta accesión fue recolectada en el año 1959 en Trujillo, La Libertad, Perú y presenta plantas con crecimiento indeterminado, elevado número de flores por inflorescencia y frutos redondos y pequeños con un peso promedio de $0,9 \pm 0,1$ g y una vida poscosecha de $18,7 \pm 0,4$ días (Rodríguez *et al.*, 2006).

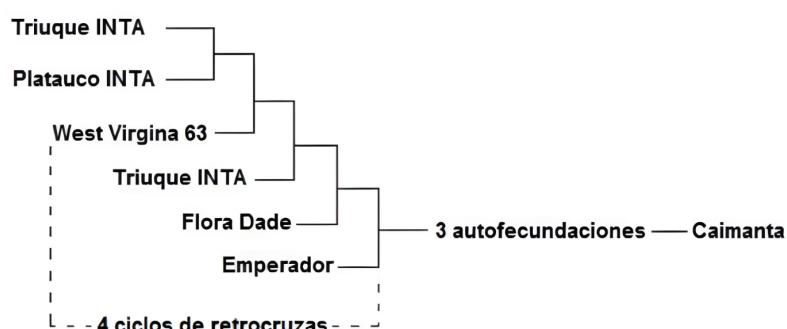


Figura I-1: Genealogía del cv. Caimanta de *Solanum lycopersicum* obtenido en la Estación Experimental INTA Cerrillos.

Para desarrollar tanto el nivel tanto molecular (del *RNA-seq*) como el fenotípico, se sembraron 10 semillas de los 8 genotipos considerados uniformes: cv. Caimanta de *Solanum lycopersicum* (C), LA0722 de *Solanum pimpinellifolium* (P) y su F_1 interespecífica (CxP). La siembra se realizó en bandejas de pre-germinación de 72 celdas de 55 cm^3 utilizando el sustrato N°2 de *Dynamics*. A los 45 días las plántulas se trasplantaron a macetas de 10 litros de capacidad, sembrando 8 plantas de cada genotipo en un invernadero climatizado siguiendo un diseño completamente aleatorizado. Las plantas fueron regadas diariamente por sistema automático de riego durante todo el ciclo del cultivo,

variando su cantidad en función a la demanda ambiental y el tamaño de las plantas con el fin de evitar el estrés hídrico.

Caracterización fenotípica

Los caracteres evaluados se obtuvieron de 10 frutos en 8 plantas por genotipo, y fueron: peso (Pe), altura (A), diámetro (D), vida poscosecha (VP, en días), forma (F, altura/diámetro), color (mediante los índices de absorbancia a y b, el cociente entre ellos o valor chroma (a/b) y el porcentaje de reflectancia L), dureza (Du), contenido en sólidos solubles (SS), pH y acidez titulable (AT). Para la obtención del ARN, el tejido utilizado fue el fruto de tomate, órgano del cual se extrajeron las muestras de pericarpio de su plano ecuatorial. En cada planta se cosecharon diez frutos en Pi y se evaluaron los caracteres: peso (en g), diámetro (en cm), altura (en cm), forma (cociente entre la altura y el diámetro) y vida poscosecha (VP), expresada como los días transcurridos desde la cosecha hasta el inicio del ablandamiento del fruto, de acuerdo Schuelter *et al.* (2002). En seis frutos de cada planta en RM se midió la firmeza con un durómetro manual tipo Shore A Durofel y el color con un cromámetro MINOLTA CR400, a través del porcentaje de reflectancia (L), variando desde 0 para negros a 100 para blancos y el cociente a/b, siendo “a” la absorbancia a longitudes de onda de 540 nm y “b” la absorbancia a longitudes de onda de 675 nm. Se obtuvo el jugo del pericarpio a partir de frutos cosechados en estado rojo maduro y se evaluaron los siguientes caracteres: contenido en sólidos solubles (SS en °Brix), medido con un refractómetro manual; acidez, medida a través del pH del jugo homogeneizado y de la acidez titulable (AT en gramos de ácido cítrico/100 gramos de jugo).

Extracción de ARN para secuenciación

La extracción del ARN se realizó en los tres estados discrepantes de la maduración: Verde Maduro (VM, estado de maduración en el que el fruto alcanza el máximo tamaño con



Figura I-2: Ejemplos de frutos en los 3 estados principales de maduración.

coloración verde), Pintón (Pi, cuando el 10% de la superficie del fruto vira hacia el color de madurez) y Rojo Maduro (RM, estado en el que el fruto presenta más del 90% de su superficie con el color final de la madurez) de cada genotipo (Figura I-2). Para caracterizar el nivel molecular, las muestras del pericarpio del fruto (en el plano ecuatorial) obtenidas fueron colocadas inmediatamente en tubos eppendorf de 1,5 ml en Nitrógeno líquido y reservadas en freezer a -80°C hasta el momento de su utilización. Se tomaron 3 réplicas biológicas de cada genotipo tal explica Liu *et al.* (2014). La extracción del ARN se llevó a cabo de acuerdo a las instrucciones del protocolo del kit Direct-zol™ RNA Miniprep Plus (Zymo Research Corporation - Cat #R2071). La calidad del ARN se verificó dos veces, tanto en espectrofotómetro (A260/A280 ratio ≥ 1.8) como en geles de agarosa al 3% p/v; la muestra que superó este doble chequeo fue enviada a secuenciar.

Secuenciación del ARN

La secuenciación mediante *RNA-seq* fue generada en un instrumento de secuenciación NovaSeq 6000 S1 de Illumina en la empresa MacroGen® de Corea, con el uso de la librería TruSeq Stranded mRNA LT Sample Prep Kit® (Guía de preparación - Parte 15031047 - Rev. E), enviando 2 ug de ARN total en una concentración de 30 ng/ul como lo establece el laboratorio. Las muestras enviadas fueron 27 en total (P30V, P30P, P30R, P64V, P64P, P64R, P27V, P27P, P27R, C34V, C34P, C34R, C61V, C61P, C61R, C62V, C62P, C62R, CxP23V, CxP23P, CxP23R, CxP37V, CxP37P, CxP37R, CxP38V, CxP38P, CxP38R) que fueron identificadas de la siguiente manera: la primera letra es el genotipo (C para el cv. Caimanta, P para *S. pimpinellifolium* y CxP para la F₁), el número siguiente pertenece a la réplica biológica muestreada en dicha posición del invernadero, y la última letra refiere al estado de maduración (V para el estado Verde maduro, Pi para el estado Pintón y R para el estado Rojo maduro). Las lecturas generadas fueron de 2 x 101 *paired-end reads* (lecturas apareadas), obtenidas con la construcción de bibliotecas de tipo *strand-specific* con el protocolo 'TruSeq Stranded mRNA Sample Preparation Guide, Part #15031047 Rev. E'. Una vez que se llevó a cabo el armado de las librerías, estas fueron secuenciadas en instrumento de secuenciación Illumina por laboratorio tercerizado.

Análisis y comparación de transcriptomas

Los archivos del tipo FASTQ fueron enviados por el laboratorio en donde se realizó la secuenciación. Los datos obtenidos a partir del *RNA-seq* fueron procesados en la plataforma online Galaxy Project® (Giardine *et al.*, 2005). Se sometieron a evaluaciones de calidad mediante la versión 0.72 del programa FastQC (Andrews, 2010). Al detectar mediante esta

herramienta incoherencia en los datos obtenidos, se realizó una limpieza bioinformática de los mismos mediante los software *Trimmomatic* y *SortMeRNA*. El software *Trimmomatic* versión 0.32.3 (Bolger *et al.*, 2014^a) permitió remover aquellas secuencias que presenten contaminación con adaptadores propios de la secuenciación, descartar las lecturas que no poseen buena calidad en su secuencia, y también descartar aquellas lecturas de longitud relativamente más cortas a las solicitadas (101 pb) que probablemente se originaron por artefactos de la misma corrida en el secuenciador. Por su parte, el software *SortMeRNA* Versión 2.1b.6 (Kopylova *et al.*, 2012) permitió filtrar y eliminar la contaminación de ARN ribosomales en datos transcriptómicos secuenciados en esta tesis. A las secuencias limpias se les realizó una segunda comprobación de calidad utilizando nuevamente el software *FastQC* antes mencionado y una vez realizado esto, se mapearon las lecturas de cada biblioteca frente al genoma de referencia de tomate mediante el uso del software HISAT2 Versión 2.1.0 (Kim *et al.*, 2015). El genoma de referencia de tomate pertenece al cv. Heinz 1706 de *Solanum lycopersicum* y fue descargado de la plataforma online Sol Genomics Network

(ftp://ftp.solgenomics.net/tomato_genome/assembly/build_3.00/S_lycopersicum_chromosomes.3.00.fa.tar.gz) como así también la anotación génica perteneciente al mismo (ftp://ftp.solgenomics.net/tomato_genome/annotation/ITAG3.2_release/ITAG3.2_gene_model.s.gff). Una vez alineadas las lecturas frente a la referencia se utilizó el software *featureCounts* Versión 1.6.3+ (Liao *et al.*, 2013) para medir la expresión génica global en cada una de las muestras estudiadas. A continuación, estos datos se analizaron con el paquete edgeR Versión 3.24.1 (Robinson *et al.*, 2009) en lenguaje R de BioConductor para establecer comparaciones entre los perfiles de expresión en cada genotipo y estado de maduración, y así obtener una lista de todos los genes con expresión diferencial (ED). Se llevó a cabo el análisis de los genes expresados diferencialmente teniendo como objetivo la cuantificación de los niveles de expresión en las diferentes muestras con el siguiente *pipeline* o flujo de trabajo. En la Figura I-3 se muestra el diagrama de trabajo seguido.

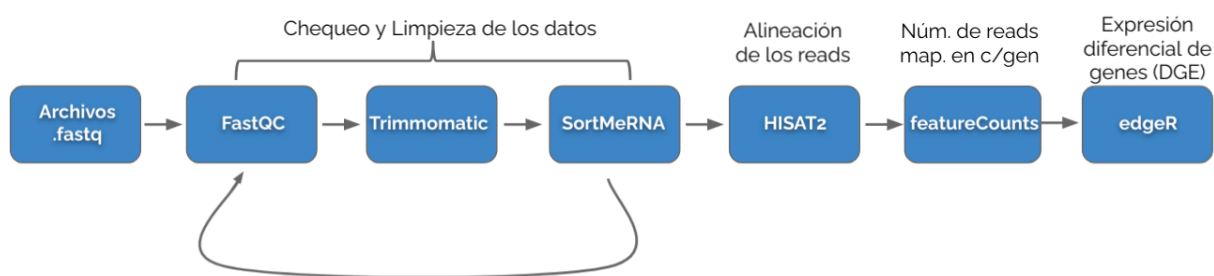


Figura I-3: Flujo de trabajo o *pipeline* llevado a cabo para la Caracterización de los transcriptomas de los frutos en los diferentes estados de maduración de los genotipos en estudio.

El total del flujo de trabajo se llevó a cabo en la plataforma online Galaxy <https://usegalaxy.org/> (Enis Afgan *et al.*, 2018). Se consideraron diferenciales aquellos genes con un p-valor ajustado (FDR) < 0,05. El estado de maduración VM fue tomado como una referencia constante en la expresión diferencial de los genes. Es por eso que se plantearon en el software edgeR los contrastes entre los estados de maduración RM frente a VM y Pintón frente a VM, para distinguir aquellos genes que se expresan diferencialmente entre los mismos. Aquellos genes que presentaron expresión diferencial por encima de esta referencia se los consideraron sobre-expresados, y aquellos que se encontraban por debajo de la referencia, como inhibidos.

Análisis de Ontología génica

Para enriquecer el análisis de la expresión global de los genes se utilizó la herramienta online AgriGO (<http://systemsbiology.cau.edu.cn/agriGOv2/>) (Zhou Du *et al.*, 2010; Tian *et al.*, 2017), un instrumento poderoso para describir la funcionalidad de los genes, anotación funcional y el análisis de enriquecimiento en la que se encuentran representados significativamente un grupo o familia de genes. Se listaron aquellos genes con expresión diferencial significativa obtenidos a través del análisis transcripcional previamente llevado a cabo en los 3 estados de maduración para cada genotipo. Esta lista de genes con expresión diferencial están involucrados en el proceso de maduración del fruto, y así se obtuvieron dos listas de genes con ED en cada genotipo por la comparación entre el estado Pi vs VM y RM vs VM. Estos genes fueron utilizados como archivo de entrada en esta herramienta, y donde se obtuvieron los gráficos y términos GO significativos seleccionando el dominio de procesos biológicos. Así, una vez realizado este análisis, se consideraron significativos los términos GO con p-valor ajustado (FDR) < 0,05, y se seleccionaron los GO que superan este umbral del total de la salida de esta herramienta. Posterior a esto, se compararon manualmente estos términos GO significativos y se seleccionaron aquellos que se comparten o están presentes en los tres genotipos bajo estudio. Luego, se extrajeron los IDs de todos los genes que están dentro de cada termino GO y se genero una lista global concatenando estos genes que provienen de cada GO seleccionado. Utilizando los comandos *sort* (ordenar) y *unique* (únicos) de linux, se obtuvo una nueva lista con los genes únicos que provienen de cada uno de los términos GO seleccionados previamente. Una vez obtenida esta lista de genes se seleccionaron de manera aleatoria 20, para poder desarrollar los análisis genéticos posteriores relacionados a la estimación de la dominancia.

Análisis genéticos y de asociación entre la variación en la expresión génica de las sHSPs y la variación fenotípica

Se realizaron, tanto para los niveles de expresión genética molecular como fenotípica, análisis genéticos y de asociación de la variabilidad en ambos niveles. Los análisis genéticos consisten en estimar las acciones génicas y a partir de ellas, calcular los grados de dominancia (d/a , Kearsey and Pooni, 1996). En el caso del nivel de expresión genética molecular, se seleccionaron aleatoriamente genes vinculados a GO significativos detectados en el punto anterior. Del total de genes se analizaron 20 genes al azar. Para cada progenitor se obtuvo en la comparación entre los estados VM vs Pi, extrayendo el valor del *logCPM* (del inglés, *logarithm of counts per million reads* o logaritmo de cantidad de lecturas por millón de lecturas). Se calculó el Punto de Origen (PO) o valor medio entre los genotipos homocigotas, siendo el valor 'a' la diferencia entre el valor del padre de máxima expresión para cada gen seleccionado al azar, el valor '-a', la diferencia entre el padre de menor expresión y el PO y el valor 'd', la diferencia entre la F_1 y el PO. Luego, a fin de determinar la acción génica existente, se estimó el grado de dominancia (d/a), correspondiendo dominancia incompleta o aditividad intralocus a valores de $d/a = 0$, dominancia parcial cuando $0 < d/a < 1$, dominancia completa cuando $d = 1$ y sobredominancia cuando $d/a > 1$ o $d/a < -1$. Estos análisis se aplicaron de igual manera a los caracteres fenotípicos evaluados. Luego, a fin de detectar asociación entre transcriptos específicos de sHSPs de los cromosomas 6 y 9, sobre las cuáles -como se explicará en próximos Capítulos- se centra el resto de la Tesis, se calcularon sus d/a y se aplicó un Análisis de Procrustes Generalizado (APG, Gower 1975) entre los valores de expresión de estas sHSPs y los valores medios de los caracteres cuantitativos evaluados en cv. Caimanta, LA0722 y su F_1 . El APG permite analizar tablas múltiples con individuos por variables y por condiciones, para detectar posibles asociaciones entre los mismos. Se analizaron 6 genes de sHSPs (Solyc06g076520, Solyc06g076540, Solyc06g076560, Solyc06g076570, Solyc09g015000 y Solyc09g015020) por separado del transcriptoma global de cada genotipo, para observar su efecto particular que quedaría encubierto de no ser aisladas del resto del transcriptoma. Para estos genes se realizó también un análisis para estimar el grado de dominancia y dilucidar las asociaciones encontradas con los datos fenotípicos. Estos 6 genes de sHSPs presentaron en análisis previos *in silico* expresión diferencial significativa en los estados de maduración del fruto (Arce *et al.*, 2015; Krsticevic *et al.*, 2016), y aquí se da comienzo a su exploración para entrar en un análisis directo en el siguiente Capítulo de esta tesis.

RESULTADOS

Extracción de ARN para secuenciación

En relación al análisis transcriptómico, se extrajeron 2 ng de ARN en una concentración mayor a 30 ug/ul, cantidad y concentración mínima requerida de ARN por el servicio de secuenciación, extraído a partir de la totalidad de los genotipos uniformes de primer ciclo (Figura I-4). Las muestras fueron enviadas a Macrogen Korea ® para llevar a cabo la secuenciación correspondiente.

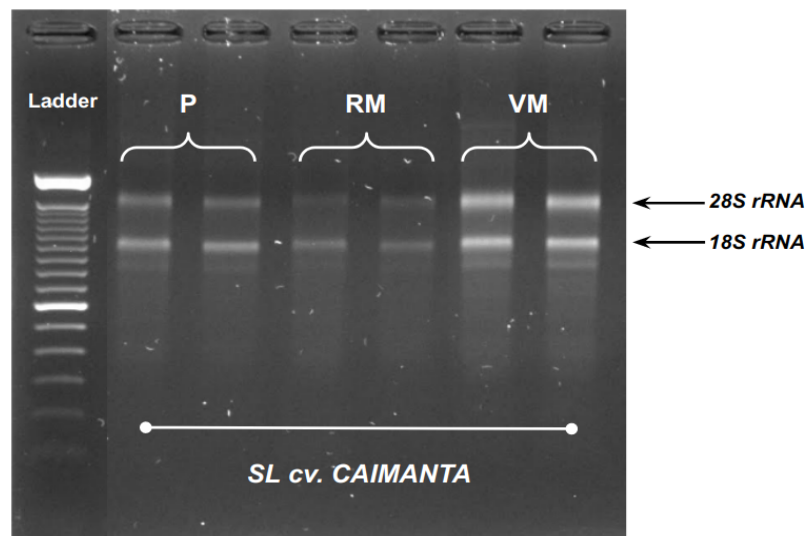


Figura I-4: Gel de agarosa al 1,5% para chequeo de la concentración de ARN en las muestras y validación de la cuantificación previa de las mismas en espectrofotómetro Perkin Elmer Lambda Bio+ ®.

Análisis de secuenciación del ARN en los tres genotipos uniformes bajo estudio

A partir del experimento de *RNA-seq* de los genotipos, se obtuvieron 51.761.745 lecturas en promedio por genotipo en cada estado de maduración. Del total de muestras analizadas, la muestra P64P fue la réplica biológica con mayor cantidad de lecturas, mientras que la C34R fue la réplica biológica con menor cantidad de lecturas (Tabla I-1).

ID muestra	Total de bases secuenciadas	Total de lecturas	GC(%)	AT(%)	Q20(%)	Q30(%)
P30V	5.602.502.320	55.470.320	44,06	55,94	97,74	93,63
P30P	4.418.828.376	43.750.776	43,62	56,38	97,84	93,74
P30R	5.227.746.466	51.759.866	43,83	56,17	98,07	94,09
P64V	5.770.446.534	57.133.134	43,90	56,10	98,07	94,13
P64P	6.947.271.062	68.784.862	44,01	55,99	97,97	93,91
P64R	4.884.644.618	48.362.818	44,99	55,01	98,06	94,10
P27V	4.231.440.248	41.895.448	44,00	56,00	97,71	93,52
P27P	5.183.529.676	51.322.076	43,15	56,85	97,87	93,83
P27R	4.687.772.792	46.413.592	47,30	52,70	98,02	94,11
C34V	5.771.060.210	57.139.210	43,62	56,38	97,98	93,92
C34P	5.896.398.988	58.380.188	45,75	54,25	97,47	93,01
C34R	4.162.963.864	41.217.464	45,97	54,03	98,11	94,23
C61V	5.101.147.410	50.506.410	43,40	56,60	98,10	94,18
C61P	4.681.268.594	46.349.194	45,04	54,96	98,02	94,00
C61R	5.098.301.028	50.478.228	46,20	53,80	98,03	94,07
C62V	6.106.998.532	60.465.332	43,86	56,14	97,86	93,70
C62P	4.976.227.984	49.269.584	44,25	55,75	97,94	93,84
C62R	5.354.304.112	53.012.912	44,57	55,43	97,98	93,97

Tabla I-1: Tabla parcial del reporte de Macrogen con cantidad de bases por lecturas, cantidad total de lecturas por muestra, porcentaje de GC y AT, puntaje de calidad (Q) 20 (1 base incorrecta cada 100 bases) y 30 (1 base incorrecta cada 1000 bases).

A partir de estos resultados preliminares se observó la calidad de las muestras y a partir de esto se procedió a la limpieza de las mismas. Luego de la limpieza de los datos generados con tecnología NGS se obtuvo un promedio de 22.541.809,11 lecturas por cada biblioteca generada. En el mejor de los casos esta limpieza solo redujo un 3.80% del total de lecturas cosechadas en la secuenciación, mientras que en el peor de los casos se perdió un 25.91% del total de lecturas crudas obtenidos (Tabla I-2).

Muestra	Nº lecturas originales	Nº lecturas post trimming	Nº lecturas post filtrado rRNA	% total reducción
C34V_1	28.569.605	27.657.529	27.483.999	3,80
C34V_2	28.569.605	27.657.529	27.483.999	3,80
C34P_1	29.190.094	28.206.758	21.801.214	25,31
C34P_2	29.190.094	28.206.758	21.801.214	25,31
C34R_1	20.608.732	19.953.782	15.526.543	24,66
C34R_2	20.608.732	19.953.782	15.526.543	24,66
C61V_1	25.253.205	24.565.500	24.215.438	4,11
C61V_2	25.253.205	24.565.500	24.215.438	4,11
C61P_1	23.174.597	22.230.491	19.610.529	15,38
C61P_2	23.174.597	22.230.491	19.610.529	15,38
C61R_1	25.239.114	24.485.547	18.699.198	25,91
C61R_2	25.239.114	24.485.547	18.699.198	25,91
C62V_1	30.232.666	28.896.184	28.668.483	5,17
C62V_2	30.232.666	28.896.184	28.668.483	5,17
C62P_1	24.634.792	23.809.287	21.141.619	14,18
C62P_2	24.634.792	23.809.287	21.141.619	14,18
C62R_1	26.506.456	25.614.029	22.976.282	13,32
C62R_2	26.506.456	25.614.029	22.976.282	13,32
P27V_1	20.947.724	20.072.532	19.405.510	7,36
P27V_2	20.947.724	20.072.532	19.405.510	7,36
P27P_1	25.661.038	24.669.971	24.329.346	5,19
P27P_2	25.661.038	24.669.971	24.329.346	5,19
P27R_1	23.206.796	22.206.870	14.551.506	37,30
P27R_2	23.206.796	22.206.870	14.551.506	37,30
P30V_1	27.735.160	26.843.745	25.916.482	6,56
P30V_2	27.735.160	26.843.745	25.916.482	6,56
P30P_1	21.875.388	21.191.393	19.856.934	9,23
P30P_2	21.875.388	21.191.393	19.856.934	9,23
P30R_1	25.879.933	24.933.798	23.616.437	8,75
P30R_2	25.879.933	24.933.798	23.616.437	8,75
P64V_1	28.566.567	27.593.326	26.942.249	5,69
P64V_2	28.566.567	27.593.326	26.942.249	5,69
P64P_1	34.392.431	32.842.356	30.451.392	11,46
P64P_2	34.392.431	32.842.356	30.451.392	11,46
P64R_1	24.181.409	23.422.242	20.559.403	14,98
P64R_2	24.181.409	23.422.242	20.559.403	14,98

Tabla I-2: Tabla parcial de la cantidad de lecturas originales por muestras y número final de lecturas por muestra luego del procesamiento de limpieza de la misma. En la columna final se observa el porcentaje de reducción de cada una de ellas.

Análisis y comparación de transcriptomas

Observando de manera grupal los transcriptomas de los 3 genotipos (Caimanta como C, la accesión LA0722 como P, y el híbrido entre estos como CxP) en los 3 estados de maduración, el análisis de componentes principales muestra que la primera componente explicó el 58% de la variabilidad total detectada, mientras que la segunda componente explicó 14% de la misma, dando en total un 72% (Figura I-5).

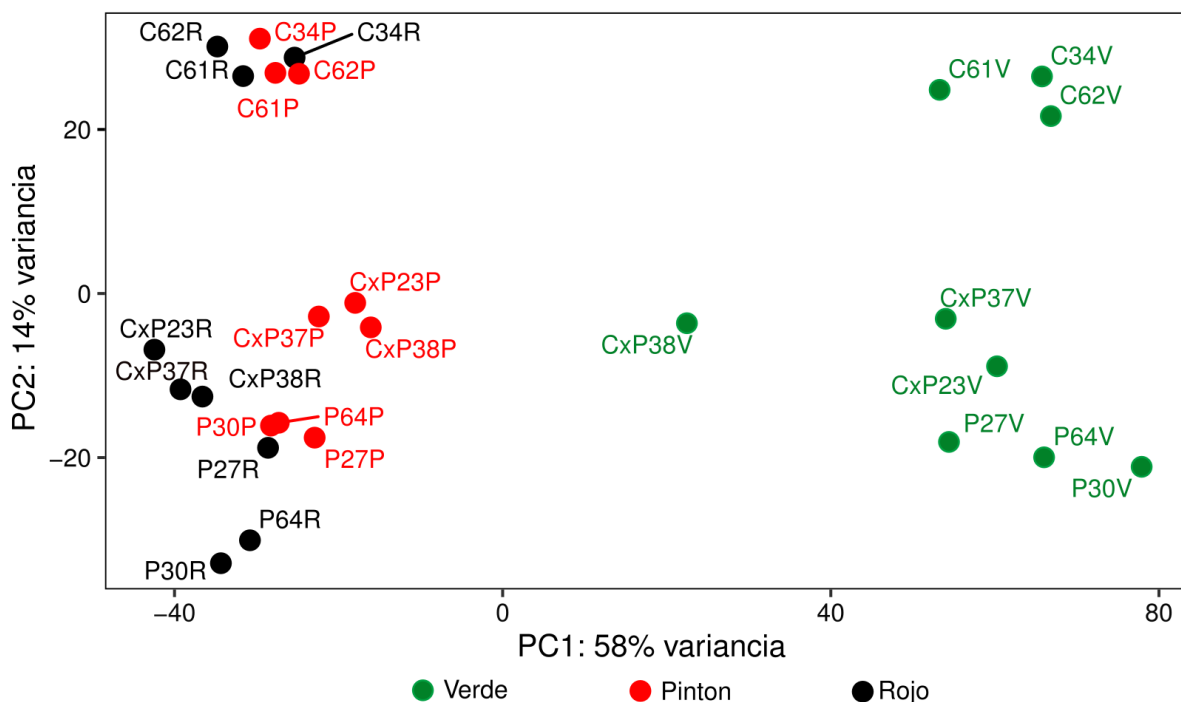


Figura I-5: Análisis de Componentes Principales del conjunto de datos transcriptómicos. En Verde se muestran los datos pertenecientes al estado de maduración VM; en Rojo se muestran los del estado Pi y en negro los datos pertenecientes al estado RM.

Puede visualizarse que la primera componente separa las muestras por estado de maduración, quedando el estado verde a la derecha del gráfico y el resto hacia la izquierda, y por otro lado la segunda componente disgrega las muestras por genotipo, visualizándose al cv. Caimanta en los valores positivos de esta componente, y el resto de los genotipos hacia los valores negativos (parte inferior del gráfico).

El análisis de componentes principales para cada genotipo individual puede observarse en los siguientes análisis realizados. En el genotipo de Caimanta (Figura I-6) la primera Componente Principal (CP) explicó el 81% de la variabilidad total detectada, mientras que la segunda CP explicó el 6% de la misma. Para la accesión silvestre LA0722 de *S. pimpinellifolium*, el análisis de componentes principales para estos datos, la primera CP explicó el 79% de la variabilidad total detectada, mientras que la segunda CP explicó 9%

de la misma (Figura I-6). Por último, para la F_1 (CxP), la primera componente explicó el 67% y la segunda el 23% (Figura I-7).

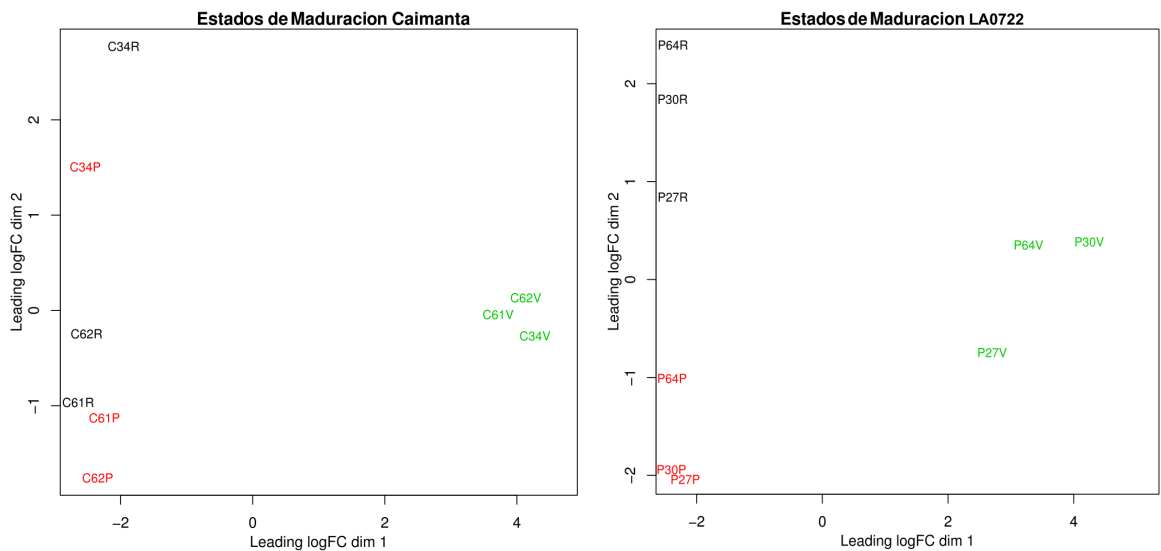


Figura I-6: Gráficos de CP en los progenitores del PMG. En Verde se muestran los datos pertenecientes al estado de maduración VM; en Rojo se muestran los del estado P y en negro los datos pertenecientes al estado RM en cada genotipo

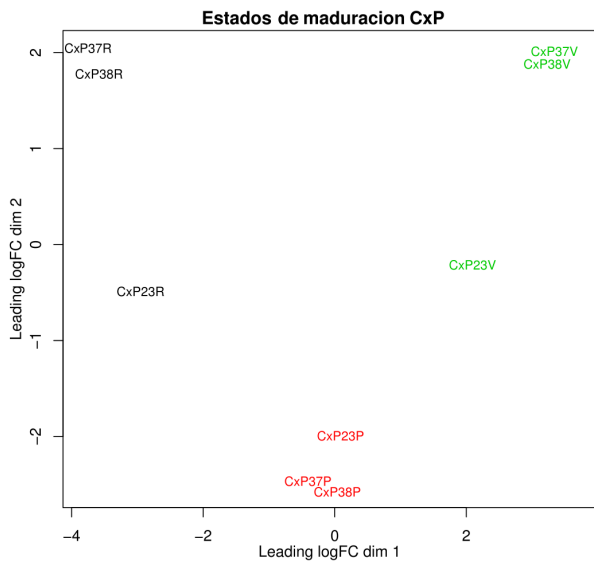


Figura I-7: Gráficos de CP en la F_1 . En Verde se muestran los datos pertenecientes al estado de maduración VM; en Rojo se muestran los del estado P y en negro los datos pertenecientes al estado RM.

La comparación entre los estados Pintón y Rojo Maduro frente al estado de maduración de referencia en el genotipo perteneciente a cv. Caimanta presento el siguiente plot (Figura I-8):

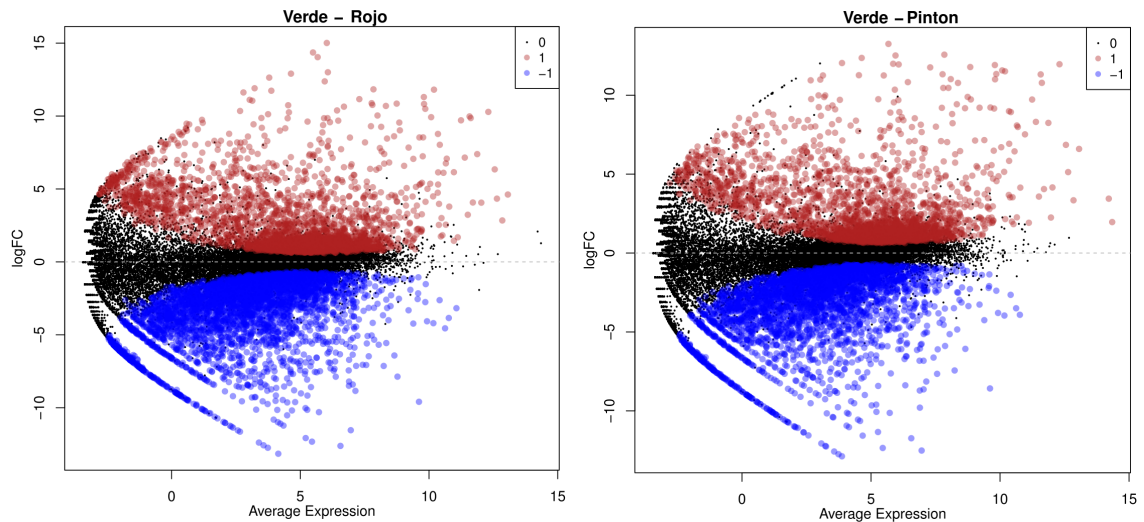


Figura I-8: Gráficos obtenidos a partir de los contrastes estudiados para la detección de genes expresados diferencialmente en el cv. Caimanta. Cada punto representa un gen. Los puntos rojos representan aquellos genes sobre-expresados; los puntos azules aquellos genes inhibidos; los puntos negros representan los genes sin expresión diferencial.

Por su parte, la comparación entre los mismos estados para el genotipo silvestre perteneciente a la accesión LA0722 presento el siguiente gráfico (Figura I-9):

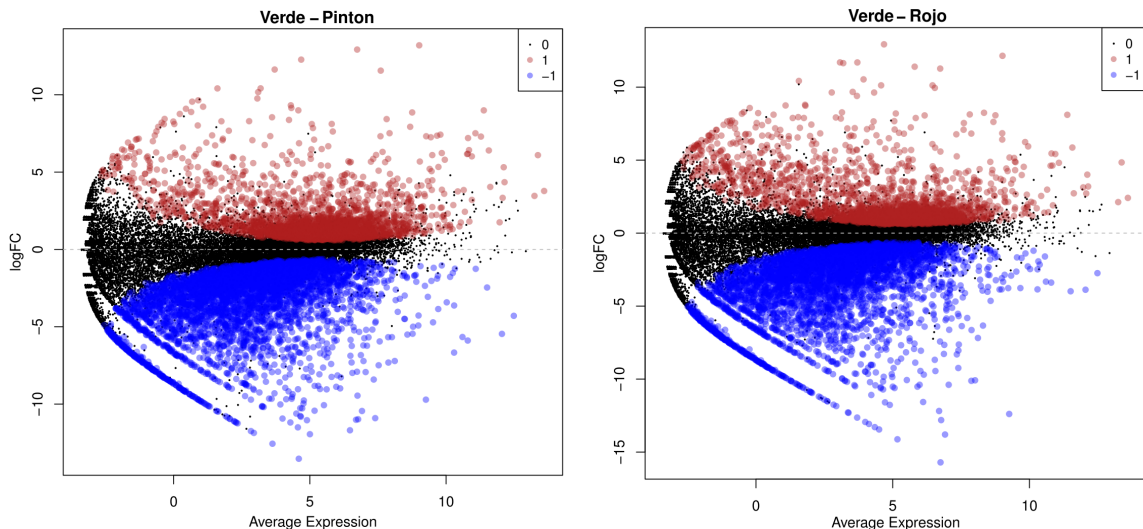


Figura I-9: Gráficos obtenidos a partir de los contrastes planteados para la detección de genes expresados diferencialmente en LA0722. Cada punto representa un gen. Los puntos rojos representan aquellos genes sobre-expresados; los puntos azules aquellos genes inhibidos; los puntos negros representan los genes sin expresión diferencial.

La comparación entre los estados Pintón y Rojo Maduro frente al estado de maduración de referencia en el genotipo perteneciente a la F_1 CxP presentó los siguientes gráficos (Figura I-10).

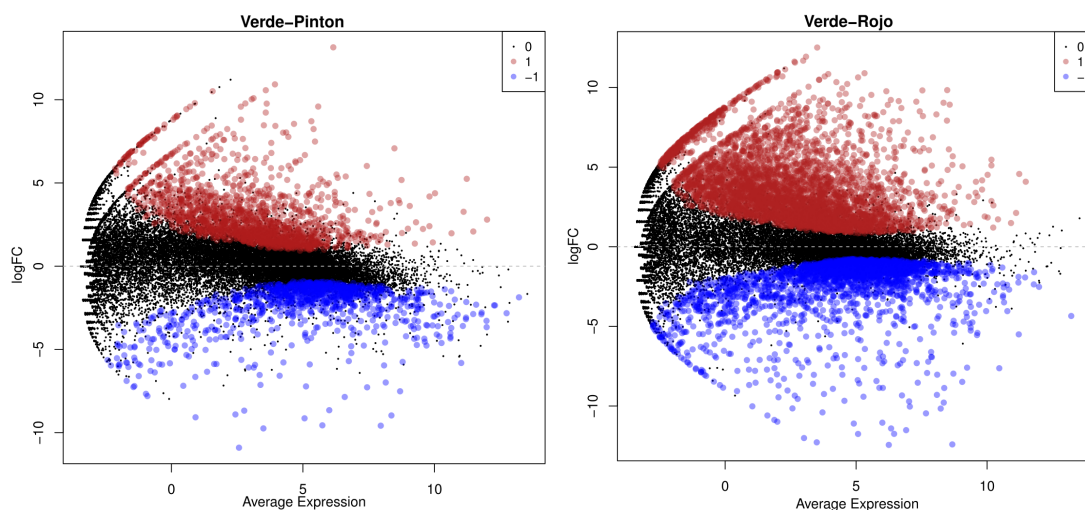


Figura I-10: Gráficos obtenidos a partir de los contrastes estudiados para la detección de genes expresados diferencialmente en la F_1 (CxP). Cada punto representa un gen. Los puntos rojos representan aquellos genes sobre-expresados; los puntos azules aquellos genes inhibidos; los puntos negros representan los genes sin expresión diferencial.

La comparación de los transcriptomas de *S. pimpinellifolium* LA0722, *S. lycopersicum* cv. Caimanta y su F_1 en los frutos en diversos estados de desarrollo y obtención de perfiles de expresión a partir del estudio de *RNA-seq* detectó lo siguiente:

Para *Solanum lycopersicum* cv. Caimanta:

- 6159 genes (2704 sobreexpresados y 3455 inhibidos) con expresión diferencial (ED) significativa de 27216 genes que mostraron ED entre los estados Pintón vs. Verde (de un total de 35768 genes)
- 6922 genes (3111 sobreexpresados y 3811 inhibidos) con ED significativa de 27328 genes que mostraron ED entre los estados Rojo vs. Verde (de un total de 35768 genes)

Por su parte, para *Solanum pimpinellifolium* accesión LA0722:

- 7313 genes (2762 sobreexpresados y 4551 inhibidos) con ED significativa de 26609 genes que mostraron ED entre los estados Pintón vs. Verde (de un total de 35768 genes)
- 6929 genes (2739 sobreexpresados y 4190 inhibidos) con ED significativa de 26644 genes que mostraron ED entre los estados Rojo vs. Verde (de un total de 35768 genes)

Para la F₁ CxP:

- 2715 genes (1054 sobreexpresados y 1661 inhibidos) con ED significativa de 26593 genes que mostraron ED entre los estados Pintón vs. Verde (de un total de 35768 genes)
- 6994 genes (2815 sobreexpresados y 4179 inhibidos) con ED significativa de 26681 genes que mostraron ED entre los estados Rojo vs. Verde (de un total de 35768 genes)

Los 5 genes con mayor expresión diferencial en cada genotipo se muestran en la Tabla I-3.

Genotipo	Estado	GeneID	logFC	logCPM	F	P Value	FDR
Caimanta	Pintón vs Verde	<i>Solyc01g086920.3</i>	-12,5133	6,9616	331,0405	1,40E-09	1,37E-05
		<i>Solyc07g063350.3</i>	-10,5544	4,4936	320,4475	1,45E-09	1,37E-05
		<i>Solyc02g079500.4</i>	-7,8627	4,7160	303,2988	1,95E-09	1,37E-05
		<i>Solyc11g065320.2</i>	-4,3343	4,5105	293,8660	2,31E-09	1,37E-05
		<i>Solyc01g058320.3</i>	-5,2136	5,3362	284,9232	2,73E-09	1,37E-05
	Rojo vs Verde	<i>Solyc01g058320.3</i>	-6,6070	5,3362	389,2202	5,06E-10	1,25E-05
		<i>Solyc10g007690.3</i>	-6,7166	6,8962	321,1926	1,43E-09	1,25E-05
		<i>Solyc01g100460.3</i>	7,1843	4,6341	319,2853	1,48E-09	1,25E-05
		<i>Solyc01g086920.3</i>	-11,5415	6,9616	316,8038	1,77E-09	1,25E-05
		<i>Solyc07g063350.3</i>	-9,6361	4,4936	306,2833	1,85E-09	1,25E-05
LA0722	Pintón vs Verde	<i>Solyc10g050110.1</i>	-11,9467	5,0019	396,0089	2,58E-10	8,00E-06
		<i>Solyc09g092330.2</i>	-5,8467	6,4940	359,1010	4,47E-10	8,00E-06
		<i>Solyc12g100120.2</i>	-6,4389	5,5182	321,6483	8,30E-10	8,94E-06
		<i>Solyc06g083580.3</i>	-7,9290	7,0472	311,1205	1,00E-09	8,94E-06
		<i>Solyc06g083590.3</i>	-7,9838	6,5612	290,0503	1,48E-09	1,04E-05
	Rojo vs Verde	<i>Solyc06g083580.3</i>	-10,3884	7,0472	407,3292	2,20E-10	6,28E-06
		<i>Solyc10g050110.1</i>	-10,5207	5,0019	373,2947	3,60E-10	6,28E-06
		<i>Solyc06g083590.3</i>	-9,4463	6,5612	343,1132	5,78E-10	6,28E-06
		<i>Solyc09g092330.2</i>	-5,5010	6,4940	328,7493	7,34E-10	6,28E-06
		<i>Solyc07g005960.3</i>	-6,3179	3,7870	310,4349	1,01E-09	6,28E-06
F ₁ (CxP)	Pintón vs Verde	<i>Solyc01g108880.3</i>	-8,9673	8,3646	218,8780	9,83E-10	3,52E-05
		<i>Solyc02g068877.1</i>	6,2925	4,7087	175,5511	3,99E-09	7,13E-05
		<i>Solyc02g068880.2</i>	6,2153	4,7283	164,5304	6,00E-09	7,15E-05
		<i>Solyc10g084600.2</i>	-7,2215	5,1178	139,2766	3,20E-08	0,00026
		<i>Solyc01g111270.3</i>	4,3837	5,2346	122,7520	3,69E-08	0,00026
	Rojo vs Verde	<i>Solyc07g054800.1</i>	-11,7544	5,0742	341,4320	5,57E-11	1,49E-06
		<i>Solyc01g111270.3</i>	9,2442	5,2346	305,9892	1,14E-10	1,49E-06
		<i>Solyc07g054745.1</i>	-11,7497	6,4514	284,8602	1,83E-10	1,49E-06
		<i>Solyc05g053310.3</i>	7,2423	5,2135	277,6401	2,13E-10	1,49E-06
		<i>Solyc04g016220.3</i>	8,8391	5,1514	273,1158	2,37E-10	1,49E-06

Tabla I-3: Top 5 de genes de sHSPs detectados en los transcriptomas de los frutos en los diferentes estados de maduración de los genotipos en estudio.

Análisis de ontología génica en los materiales uniformes

Para profundizar y enriquecer aún más el estudio global de genes obtenido con los perfiles transcripcionales de cada genotipo en los 3 estados madurativos, se prosiguió a un nuevo análisis con la herramienta online AgriGO a fin de caracterizar la funcionalidad de los genes con expresión diferencial. Además, este análisis de enriquecimiento permite explorar los procesos biológicos relacionados a la madurez del fruto en los que participan transcritos de expresión diferencial tal se plantea en el Objetivo Específico 3.

A través del análisis de expresión diferencial se obtuvieron 2 listas por genotipo con los ID de los genes expresados diferencialmente frente a las comparaciones Pi vs. VM y RM vs VM, en total 6 listas. A partir de estas listas y el uso de la herramienta AgriGO se seleccionó la comparación entre Pi y VM antes nombrada, ya que arrojó mayor cantidad de términos GO significativos en la primera exploración. Al poseer mayor cantidad de términos GO significativos relacionados a los procesos biológicos (en comparación con la de RM y VM), fue seleccionada por brindar mayor información de interés. Así, se encontraron 11 términos GO significativos para el cultivar Caimanta como se observa en la Figura I-11 (GO:0044699, GO:0015979, GO:0008152, GO:0044710, GO:0006091, GO:0044763, GO:0009765, GO:0019684, GO:0005975, GO:0055114, GO:0044281), mientras que para la accesión LA0722 se encontraron 25 GO significativos (GO:0008152, GO:0044699, GO:0071704, GO:0044238, GO:0044710, GO:0044763, GO:0005975, GO:0006629, GO:0044281, GO:0044711, GO:0006790, GO:0051186, GO:0006793, GO:0006082, GO:0043436, GO:0019752, GO:0006520, GO:0044283, GO:0036211, GO:0006464, GO:0006468, GO:0051188, GO:0016310, GO:0009765, GO:0006796) entre estos estados de maduración (Figura I-12), y finalmente la F₁ fue de 6 términos GO significativos (GO:0005975, GO:0044699, GO:0006468, GO:0044710, GO:0016310, GO:0008152) durante en este análisis (Figura I-13). Estos pueden observarse en los siguientes grafos en los 3 genotipos:

Ontología génica: términos GO significativos para SL cv. CAIMANTA

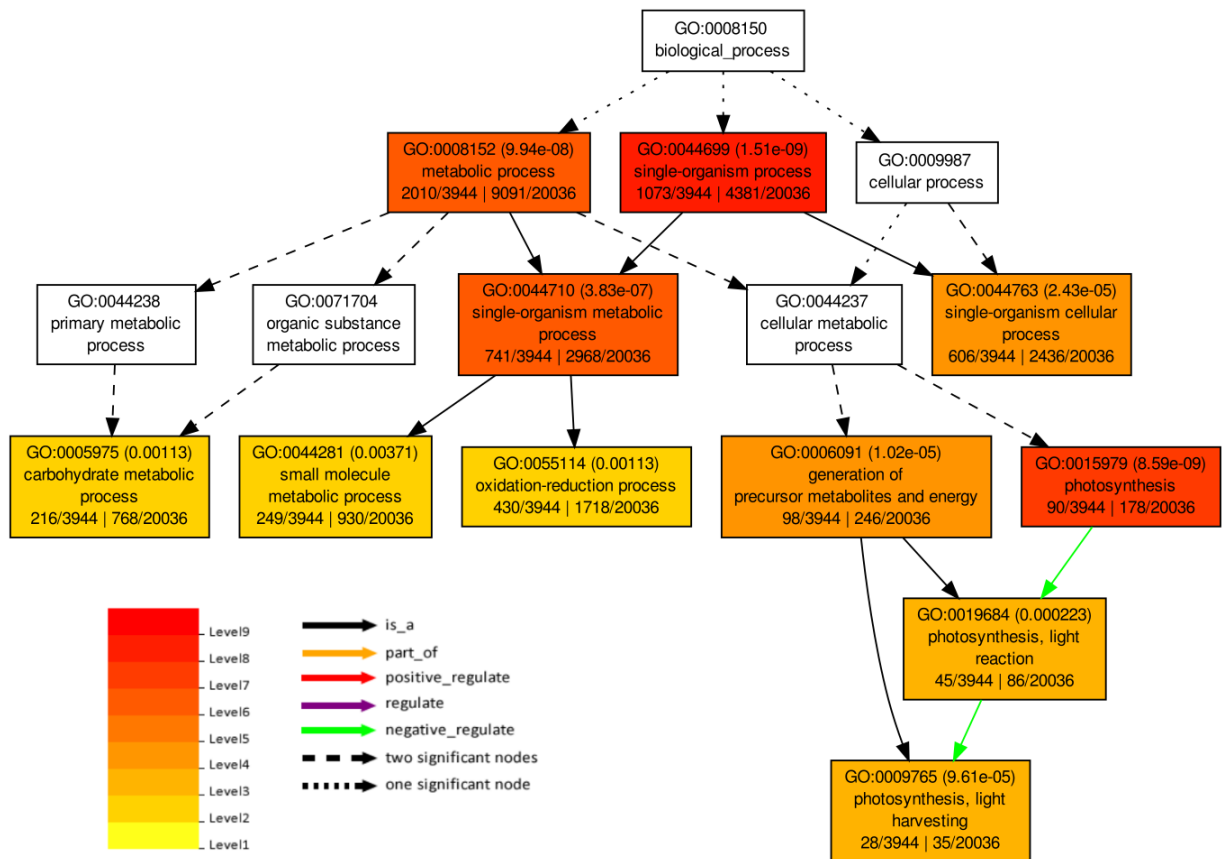


Figura I-11: Grafo GO resultante del dominio Procesos Biológicos del cv. Caimanta a partir de genes expresados diferencialmente entre los estados VM y Pi.

Ontología génica: términos GO significativos para **SP acc. LA0722**

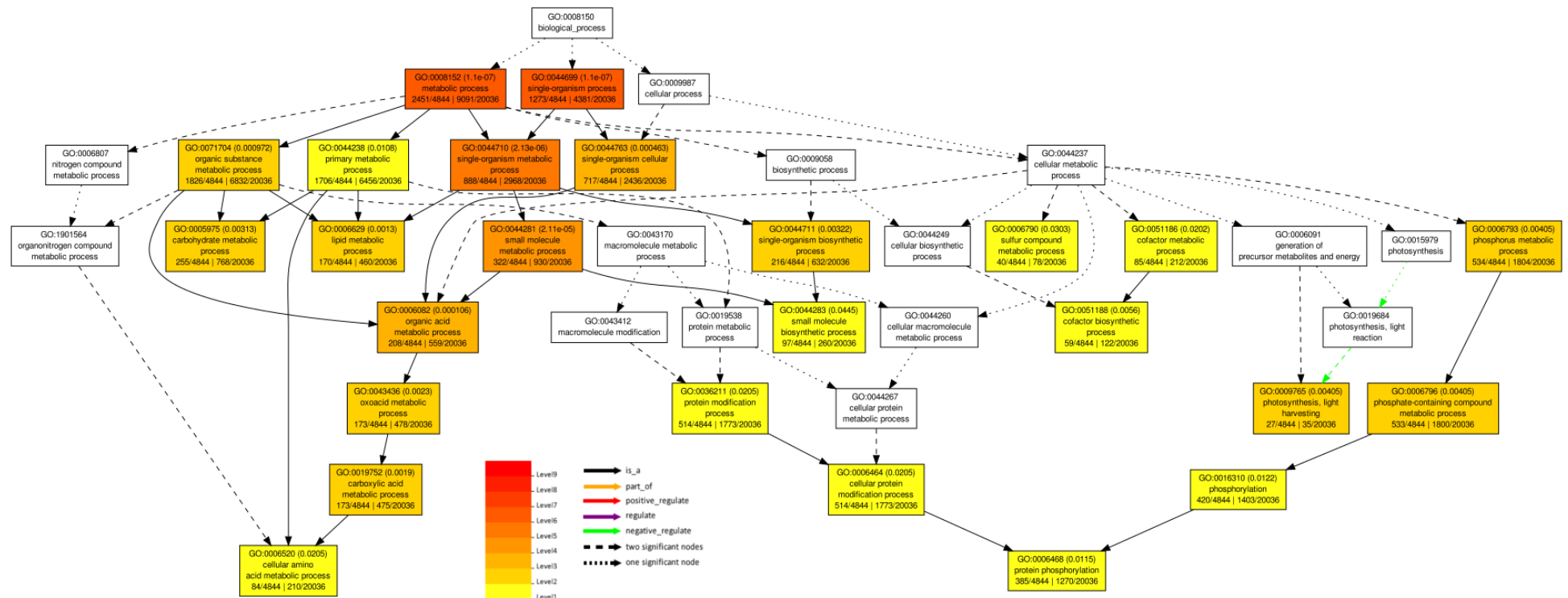


Figura I-12: Grafo GO resultante del dominio Procesos Biológicos de la acc. LA0722 a partir de genes expresados diferencialmente entre los estados VM y Pi.

Ontología génica: términos GO significativos para F_1 (CxP)

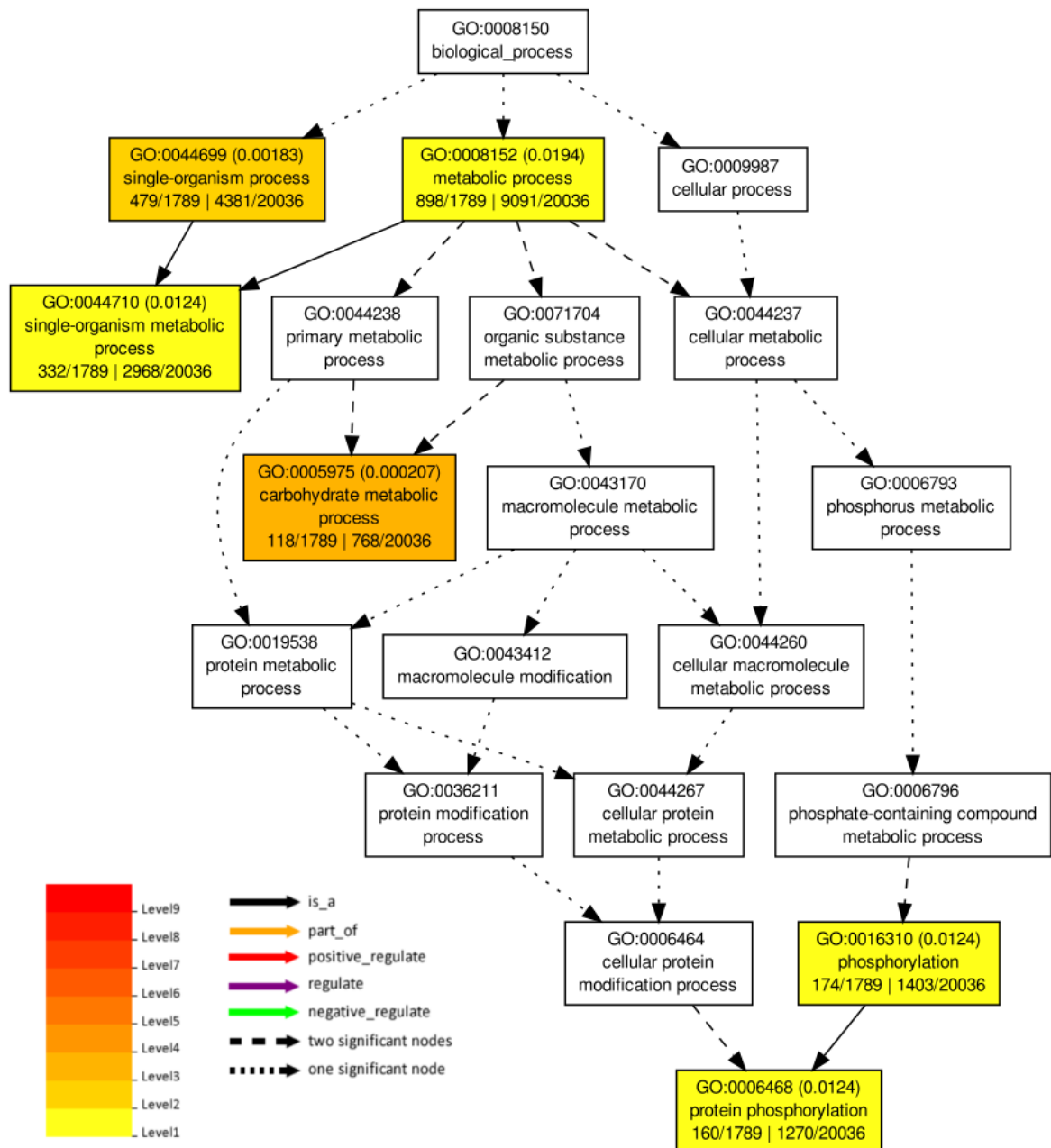


Figura I-13: Grafo GO resultante del dominio Procesos Biológicos de la F_1 (CxP) a partir de genes expresados diferencialmente entre los estados VM y Pi.

A partir de estos términos GO significativos (con p-valor ajustado < 0,05) obtenidos en cada genotipo uniforme, se profundizó el análisis comparando los términos GO de estos y dilucidar aquellos que participan en el proceso de maduración del fruto. La comparación entre el dominio de procesos biológicos entre los genotipos progenitores y su F₁ generó una lista de términos GO consenso entre los mismos (Tabla I-4), en el que se detectaron 4 en total.

#	GO	Ontología	Descripción
1	GO:0044699	PB	<i>Single-organism process</i>
2	GO:0008152	PB	<i>Metabolic process</i>
3	GO:0044710	PB	<i>Single-organism metabolic process</i>
4	GO:0005975	PB	<i>Carbohydrate metabolic process</i>

Tabla I-4: Tabla comparaciones entre el dominio de Procesos biológicos (PB) con términos GO compartidos entre los genotipos progenitores y su F₁. (CxP).

Dado que cada una de estos términos GO presenta metainformación, en donde genes que participan en un proceso biológico puntual se agrupan en un término GO, se extrajeron los ID de los genes en cada uno de estos y se los comparó con el fin de obtener una lista única de aquellos genes involucrados en este proceso biológico. Esta comparación arrojó un total de 2744 genes únicos entre ellos. A partir de esta lista de genes, se analizaron los 20 genes seleccionados para llevar a cabo la estimación de la dominancia en estos genes, respecto a cada uno de los progenitores y al valor que obtuvo el híbrido entre ellos. Esto permite estimar qué progenitor aporta en mayor o menor medida en la progenie.

Estimación del grado de dominancia de los 20 genes seleccionados

A continuación se realizó el análisis antes detallado y se llevó a cabo la estimación de la dominancia de 20 genes seleccionados dentro del dominio de procesos biológicos.

Gen	CAIMANTA	LA0722	F ₁ CxP	PO	a	-a	d	d/a	Grado Dominancia	Progenitor dominante	Descripción funcional del gen*
<i>Solyc01g059870.4</i>	4,364	4,536	3,369	4,45	0,086	-0,086	-1,080	-12,529	Sobredominancia negativa	CAI	Phytochrome B1
<i>Solyc01g087210.3</i>	8,340	8,148	7,633	8,24	0,096	-0,096	-0,611	-6,352	Sobredominancia negativa	LA0722	Cellulose synthase family protein
<i>Solyc01g095080.3</i>	8,744	7,592	7,441	8,17	0,576	-0,576	-0,727	-1,262	Sobredominancia negativa	LA0722	1-aminocyclopropane-1-carboxylic acid synthase-2
<i>Solyc02g080160.3</i>	3,220	4,727	2,134	3,97	0,753	-0,753	-1,839	-2,441	Sobredominancia negativa	CAI	Xyloglucan endotransglucosylase/hydrolase
<i>Solyc02g091920.3</i>	5,201	6,696	4,661	5,95	0,747	-0,747	-1,287	-1,723	Sobredominancia negativa	CAI	Xyloglucan endotransglucosylase-hydrolase 7
<i>Solyc02g093150.3</i>	4,177	3,829	3,257	4,00	0,174	-0,174	-0,747	-4,292	Sobredominancia negativa	LA0722	APETALA2c
<i>Solyc03g031800.3</i>	3,998	4,776	3,034	4,39	0,389	-0,389	-1,353	-3,478	Sobredominancia negativa	CAI	Xyloglucan endotransglucosylase/hydrolase
<i>Solyc03g031860.3</i>	12,553	12,146	11,979	12,35	0,203	-0,203	-0,370	-1,822	Sobredominancia negativa	LA0722	Phytoene synthase 1
<i>Solyc03g044300.3</i>	8,057	7,855	7,701	7,96	0,101	-0,101	-0,255	-2,519	Sobredominancia negativa	LA0722	APETALA2a
<i>Solyc04g071650.3</i>	6,925	7,525	6,277	7,23	0,300	-0,300	-0,949	-3,164	Sobredominancia negativa	CAI	Cellulose synthase
<i>Solyc04g081300.3</i>	6,251	6,965	6,411	6,61	0,357	-0,357	-0,196	-0,549	Dominancia parcial hacia menores valores	CAI	Endoglucanase
<i>Solyc05g005080.3</i>	8,240	7,882	7,429	8,06	0,179	-0,179	-0,632	-3,532	Sobredominancia negativa	LA0722	Endo-1,4-beta-glucanase
<i>Solyc05g012020.3</i>	9,807	9,842	9,228	9,83	0,018	-0,018	-0,597	-33,426	Sobredominancia negativa	CAI	Ripening inhibitor
<i>Solyc06g083580.3</i>	6,881	7,047	6,209	6,96	0,083	-0,083	-0,755	-9,079	Sobredominancia negativa	CAI	Pectate lyase
<i>Solyc07g052980.3</i>	1,320	2,669	0,658	1,99	0,675	-0,675	-1,336	-1,981	Sobredominancia negativa	CAI	Xyloglucan endotransglycosylase/hydrolase 16
<i>Solyc07g053920.3</i>	2,988	3,070	2,235	3,03	0,041	-0,041	-0,794	-19,234	Sobredominancia negativa	CAI	Mannan endo-1,4-beta-mannosidase 2
<i>Solyc08g061100.3</i>	8,527	8,565	7,971	8,55	0,019	-0,019	-0,575	-30,458	Sobredominancia negativa	CAI	Cellulose synthase
<i>Solyc10g078550.3</i>	2,783	1,108	1,596	1,95	0,838	-0,838	-0,350	-0,418	Dominancia parcial hacia menores valores	LA0722	Glutamate dehydrogenase
<i>Solyc11g072600.2</i>	2,567	1,120	0,395	1,84	0,724	-0,724	-1,449	-2,002	Sobredominancia negativa	LA0722	APETALA2d
<i>Solyc12g008840.2</i>	6,068	7,976	6,414	7,02	0,954	-0,954	-0,608	-0,637	Dominancia parcial hacia menores valores	CAI	Beta-galactosidase 4

Tabla I-5: Tabla con ID genes y estimación del grado de dominancia de los 20 genes seleccionados a partir del análisis de enriquecimiento. (*según anotación ITAG3.2)

En este análisis de la estimación de la dominancia (Tabla I-5) puede observarse que de los 20 genes, 17 genes mostraron sobredominancia negativa superando al progenitor de menor valor, y los otros 3 mostraron dominancia parcial hacia menores valores. Disgregando aún más, el 60% de los genes fue hacia el valor del progenitor dominante, en este caso el cv. Caimanta, y el 40% restante hacia el progenitor silvestre.

Caracterización fenotípica de los tres materiales genéticamente uniformes

Se realizó un análisis de componentes principales de estos 3 genotipos con las variables de interés agronómico. Este estudio puso de manifiesto la variabilidad generada para caracteres de interés agronómico y la contribución de tales caracteres para la diferenciación de los genotipos bajo estudio (Figura I-14).

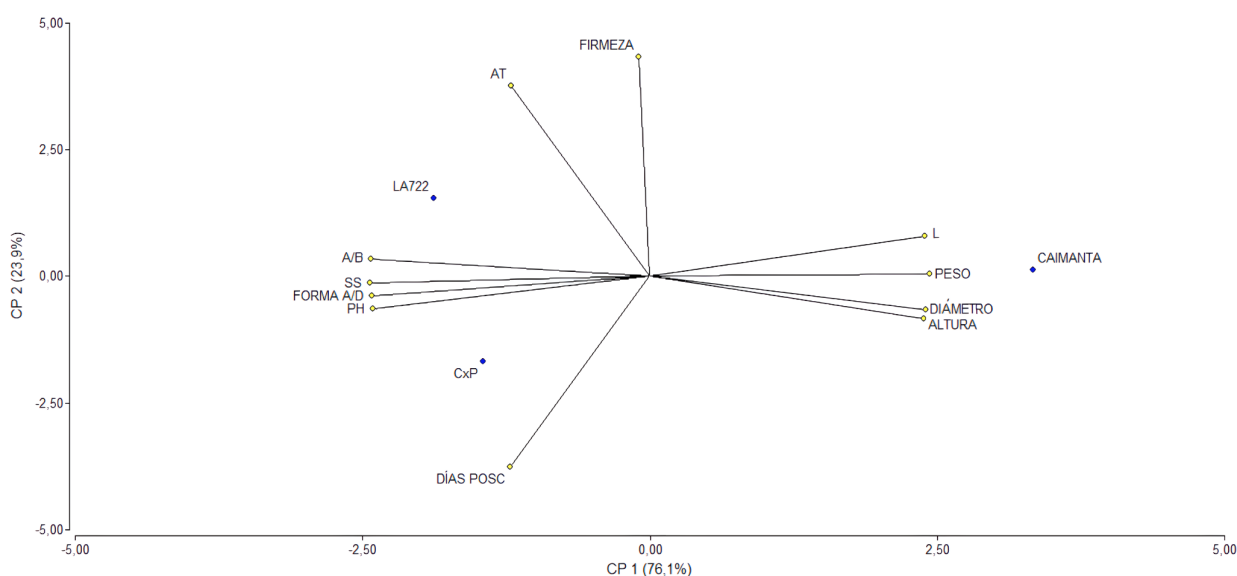


Figura I-14: Gráficos de componentes principales en los 3 genotipos estudiados con los atributos de interés agronómico.

El gráfico muestra que la primera componente principal explica el 76,1% de la variabilidad, mientras que la segunda explica el 23,9% de la misma (Figura I-14). Por su parte, cv. Caimanta se encuentra en los valores positivos de la CP1 y cercano a las variables que aporta de manera mayoritaria, como lo son peso, altura, diámetro y el coeficiente L (Figura I-14). Este material se caracteriza por su alto peso y luminosidad, en concordancia con lo observado en el gráfico. Por el otro lado, en los valores negativos se observan los genotipos LA0722 y F₁ (CxP), cada uno cercano a las variables que más explican.

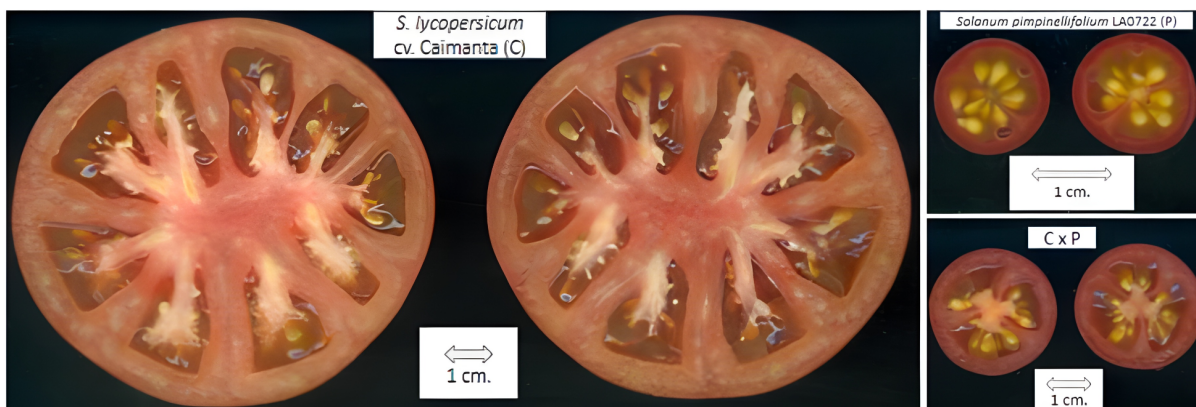


Figura I-15: Frutos de los 3 materiales de primer ciclo en el estado RM. Corte transversal y escaneo de los mismos.

En el caso de la F_1 la vida poscosecha y para el genotipo silvestre AT y firmeza, quedando entre estas dos variables que comparten como de pH, SS, forma y el coeficiente a/b. Los autovectores del análisis de componentes principales se detallan en la Tabla I-6.

Variables	CP1	CP2
SS	-0,35	-0,02
PH	-0,34	-0,09
AT	-0,17	0,54
A/B	-0,34	0,05
L	0,34	0,11
DUREZA	-0,01	0,62
DIÁMETRO	0,34	-0,09
ALTURA	0,34	-0,12
FORMA (a/d)	-0,34	-0,06
PESO	0,35	0,01
DÍAS POSC	-0,17	-0,53

Tabla I-6: Valores de los autovectores obtenidos en el análisis de CP. (Correlación cofenética= 1,00)

Análisis de asociación entre la variabilidad en la expresión génica de las sHSPs y la variabilidad fenotípica

El Análisis Procrustes Generalizado se realizó entre los genotipos en diferentes estados de maduración y con el estudio de diferentes variables agronómicas antes mencionadas, y además con los niveles de expresión génica de los genes de sHSPs del cromosoma 6 y 9 en cada uno de los estados de maduración (Figura I-16). Estos genes de sHSPs en particular fueron estudiados previamente *in silico*, y aquí se focalizó en estos ya que su aporte en general en los transcriptomas completos no es abundante de manera absoluta. Hacer luz en aquellos genes de interés frente al resto del transcriptoma es un paso previo importante para determinar y detectar asociaciones significativas entre los niveles fenotípicos y moleculares.

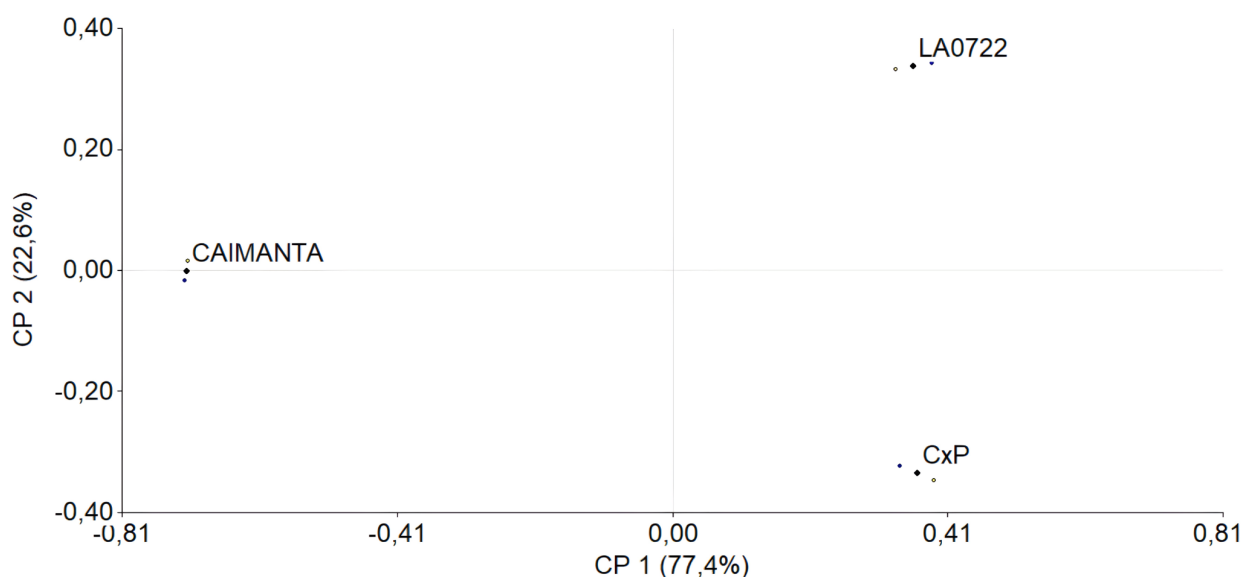


Figura I-16: Análisis Procrustes Generalizado en cv. Caimanta, LA0722 y su F_1 CxP. El punto del medio representa el genotipo, el punto gris las sHSPs del cromosoma 06 y el otro punto las del 09.

Puede observarse cómo la primera componente principal explica el 77,4% de la variabilidad, mientras que la segunda explica el 22,6% (Figura I-16). Este análisis simultáneo generó una estructura consenso capaz de sintetizar toda la información disponible. Los puntos consenso obtenidos para cada genotipo son acordes a lo esperado, agrupando cada valor de las variables en conjunto a su genotipo representativo y generando 3 grupos (Figura I-16). Esto evidencia la asociación entre los dos conjuntos de datos en cada genotipo uniforme, variables cuantitativas referidas a la caracterización fenotípica y la expresión de estas sHSPs de interés en los progenitores y su F_1 .

Estimación del grado de dominancia de genes de sHSPs

Se llevó a cabo una estimación de la dominancia en estas sHSPs pertenecientes al cromosoma 6 y 9, con la misma metodología usada en con los genes provenientes de la expresión global. Los resultados de la estimación de la dominancia para los 6 genes de sHSPs se muestran en la siguiente tabla (tabla I-6).

GeneID	CAIMANTA	LA0722	F ₁ CxP	PO	a	-a	d	d/a	Grado Dominancia	Progenitor dominante
<i>Solyc06g076520.1</i>	10,841	12,190	12,265	11,52	0,674	-0,674	0,750	1,112	Sobredominancia positiva	LA0722
<i>Solyc06g076540.1</i>	7,147	12,572	11,984	9,86	2,713	-2,713	2,124	0,783	Dominancia parcial de los mayores valores	LA0722
<i>Solyc06g076560.2</i>	10,088	12,633	11,795	11,36	1,272	-1,272	0,435	0,342	Dominancia parcial de los mayores valores	LA0722
<i>Solyc06g076570.3</i>	10,383	12,077	11,785	11,23	0,847	-0,847	0,554	0,655	Dominancia parcial de los mayores valores	LA0722
<i>Solyc09g015000.3</i>	9,039	10,949	11,201	9,99	0,955	-0,955	1,207	1,264	Sobredominancia positiva	LA0722
<i>Solyc09g015020.1</i>	8,640	10,292	10,130	9,47	0,826	-0,826	0,664	0,804	Dominancia parcial de los mayores valores	LA0722

Tabla I-7: Estimación de la dominancia en genes de sHSPs del cromosoma 06 y 09.

En la totalidad de los genes de sHSPs en estudio, el valor observado de la F₁ fue en dirección al progenitor silvestre (tabla I-7), siendo este el progenitor con mayor valor. Los genes *Solyc06g076520* del cromosoma 6 y *Solyc09g015000* del cromosoma 9 mostraron sobredominancia positiva superando al progenitor de mayor valor (accesión LA0722), y el resto de los genes mostraron dominancia parcial hacia mayores valores (tabla I-7).

DISCUSIÓN

En el presente Capítulo se planteó un análisis global de los transcriptomas a fin de caracterizar la expresión génica de los progenitores que iniciaron el programa de mejoramiento genético de la especie. El grupo de trabajo cuenta con diferentes estudios en dichos materiales, como marcadores moleculares de ADN, ARN y proteínas (Pereira da Costa *et al.*, 2016, Cabodevila *et al.*, 2017; Di Giacomo *et al.*, 2020), secuenciación de sus genomas (Cambiaso *et al.*, 2019^b), y a fin de avanzar con estudios en los 3 niveles del dogma central de la biología se realizó la secuenciación de los transcriptomas de los progenitores.

El análisis global de los genes evidenció diferencias en la cantidad de transcritos expresados en los 3 estados madurativos en cada progenitor. El análisis de componentes principales muestra concordancia con la caracterización fenotípica y su gráfico de componentes principales. El grupo ha evidenciado este tipo de comportamiento, en donde en caracterizaciones fenotípicas el cv. Caimanta localiza de manera distante al resto de los genotipos, ya sea el genotipo silvestre, la F_1 o familias derivadas de ellos, y en donde agrupa con la variable peso y baja vida poscosecha, característica principal de este cultivar (Zorzoli *et al.*, 2000). Por otro lado, el análisis molecular en conjunto de todos los genotipos en los 3 estados de maduración evidenció que la primera componente principal separa por estado madurativo, quedando el estado VM hacia los mayores valores y el estado Pi y RM hacia los menores. Esto puede deberse a que el estado VM es una etapa del desarrollo en el que se desencadenan la mayor parte de la expresión génica para el desarrollo del mismo (observable en la figura I-4). Esto indica que el estado de madurez es importante además del genotipo para determinar las diferencias principales tal lo detallado por Gallo *et al.* (2010). La segunda componente separó a cv. Caimanta por un lado y la accesión silvestre por otro en conjunto con la F_1 similar a lo evidenciado en la primera componente del análisis fenotípico de estos materiales. Aquí se identifica el aporte de cada genotipo a la variabilidad genética subyacente en cada uno de los estados de maduración.

La cantidad de genes con expresión diferencial significativa fue, para el cv. Caimanta, del 17,22% y del 19,35% para los estados Pi y RM respectivamente, para LA0722, del 20,45% y 19,37% para los estados Pi y RM respectivamente, y para su F_1 , del 7,59% y 19,55% para los estados Pi y RM respectivamente. Esto pone de manifiesto las diferencias en la expresión de los genes en cada estado de maduración y en cada genotipo. Este comportamiento puede deberse a variantes en la secuencia génica que regula el proceso de expresión molecular en cada genoma en estudio. Además, esta variación puede ser

generada no desde el punto de vista biológico, sino por el alineamiento de estos transcriptomas frente a la referencia mundial de tomate. El progenitor cultivado pertenece al mismo género y especie que dicha referencia, pero no es el caso de la accesión silvestre y la F₁. Esto puede generar variación en el recuento de transcriptos que no se deban a la biología de dicho proceso sino al proceso de alineamiento propiamente dicho, en el que se setean parámetros para que un transcripto sea considerado alienado o no frente a la referencia. Llevar a cabo el alineado y mapeo de los transcriptomas de todos los genotipos frente a un mismo genoma de referencia es un primer abordaje en el cual permite obtener una visión exploratoria hacia la identificación de genes con expresión diferencial frente a un estado particular del desarrollo. Su principal ventaja es permitir la posterior comparación entre estos todos los genotipos y entre todos los estados de maduración estudiados.

Concluido el análisis global de genes, se continuó con un primer enfoque específico en las sHSP, objetivo general de esta tesis. En el análisis global de los transcriptomas, las sHSPs pueden quedar encubiertas o enmascaradas por la expresión de otros genes de mayor longitud o genes que presenten niveles de expresión constante en todas las células y condiciones. Los genes de sHSPs del cromosoma 6 tienen una longitud de 451 pb (sin intrones) y de 154 aminoácidos. En la estimación de la dominancia, el total de los miembros de sHSPs analizados (6) fue hacia el progenitor silvestre (tabla I-7). Dado las características particulares de estas sHSPs, en especial su longitud y funcionalidad se estudiaron puntualmente en el siguiente Capítulo.

En la caracterización funcional de los genes con expresión diferencial, se evidenciaron genes relacionados a procesos biológicos que involucran el proceso de maduración. Todos los genotipos mostraron variabilidad en la cantidad de términos GO significativos. A diferencia de lo esperado la accesión silvestre LA0722 fue la que presentó la mayor cantidad de términos GO significativos (25 en total) relacionados a procesos biológicos. La mayoría de las ontologías desarrolladas en ese tipo de software se realizan a través de los datos obtenidos en los genomas de referencia mundial, por lo que era esperable que el cv. Caimanta presenta mayor cantidad de términos GO significativos, por ser evolutivamente más cercana a la referencia que el progenitor silvestre. De todas maneras, el progenitor silvestre evidenció mayor cantidad de vías metabólicas significativas relacionadas al proceso de interés.

La caracterización fenotípica de los progenitores y su cruzamiento manifestó valores medios de los caracteres agronómicos semejantes a los informados y se evidenciaron diferencias significativas entre los progenitores para todos los caracteres analizados, en

concordancia con resultados obtenidos en trabajos del grupo (Rodríguez *et al.*, 2006; Pratta *et al.*, 2011; Pereira da Costa *et al.*, 2013). Disponer de materiales que presentan variabilidad significativa para los rasgos de interés es crucial para el desarrollo de poblaciones en donde poder estudiar los comportamientos fenotípicos y las bases moleculares que rigen los caracteres de calidad del fruto en tomate.

Por último, las asociaciones a través del Análisis Procrustes Generalizado (Figura I-16) evidencian una correspondencia entre los datos de expresión de genes de sHSPs tanto del cromosoma 6 como del 9 con las variables cuantitativas referidas a la caracterización fenotípica. Esto indica la presencia de asociación entre los datos moleculares y fenotípicos y entre la subfamilia de sHSPs con los rasgos de interés agronómicos que son deseados mejorar por el grupo. La estimación de la dominancia de estos genes en cada genotipo evidencia para las sHSPs como para los caracteres fenotípicos, que el progenitor dominante suele ser la accesión LA0722. La asociación de la variabilidad en ambos niveles muestra concordancia y de manera alta con LA0722, en donde las sHSPs actúan como chaperonas y podrían controlar las homeostasis celular previniendo la degradación del fruto y extendiendo así su vida poscosecha. Esto también puede observarse en la dirección de la primera componente principal de la parte fenotípica y en donde la accesión silvestre (en conjunto con la F_1) se encuentran junto a caracteres de calidad del fruto (incluida la vida poscosecha) y explicando en mayor medida la variabilidad detectada.

CONCLUSIONES

- La secuenciación del ARN de los materiales uniformes de primer ciclo en los tres estados principales de maduración permitió la obtención y caracterización de perfiles de expresión génica en cada uno de ellos.
- La comparación entre los transcriptomas de los genotipos de primer ciclo pone de manifiesto una lista de genes con expresión génica diferencial en los 3 estados principales de maduración del fruto.
- El estudio de la funcionalidad de los genes con ED detecta una lista de genes involucrados en procesos biológicos relacionados a la maduración en los que participan transcritos de expresión diferencial.
- La caracterización fenotípica de estos materiales de primer ciclo muestra una amplia variabilidad generada en el programa de mejoramiento genético, cuya diferenciación es posible de acuerdo a sus valores para los caracteres de interés agronómico evaluados.
- El análisis de la variabilidad entre los genotipos y la asociación entre la variabilidad genética en los niveles molecular y fenotípico pusieron en evidencia atributos diferenciales potenciales para continuar y asistir con mayor certidumbre al programa de mejoramiento de la especie.

CAPÍTULO II

ANÁLISIS DE LA EXPRESIÓN GÉNICA DE sHSPs EN LOS PROGENITORES Y DESARROLLO DE MARCADORES MOLECULARES EN BASE A SU POLIMORFISMO MOLECULAR

INTRODUCCIÓN

El análisis de la expresión global de genes a través de la secuenciación del ARN genera millones de secuencias de manera simultánea, obteniéndose un perfil transcripcional de un organismo en un momento en particular. Esto representa una gran oportunidad para aplicar diferentes técnicas de extracción de conocimiento a partir de toda la información y cuyo objetivo final es determinar la funcionalidad de los genes y establecer la relación entre el genotipo y fenotipo. Esta herramienta posee un potencial destacable y robusto para estudiar todos los genes, aunque para detectar la expresión de una familia en particular, como las de sHSPs, suele enfocarse directamente en dicha familia *target* para evitar que su expresión quede oculta detrás de la expresión de otras familias génicas con unidades transcripcionales de mayor longitud o con niveles de expresión constante en todas las etapas. Es por eso que en este Capítulo se hace énfasis en el estudio de la subfamilia de genes previamente estudiadas *in silico*, las *small heat shock proteins*.

Estas proteínas constituyen una subfamilia multigénica que se caracteriza por el bajo peso molecular (aproximadamente de 20 kDa) de las proteínas codificadas por sus miembros. Esta familia es antigua (de más 450 m.a) y se encuentra en la base del árbol de la vida, distribuidas en los dominios Archaea, Bacteria y Eukarya (Waters, 2012). Presentan actividad de chaperonas moleculares en respuesta no solo al estrés térmico (Guo *et al.*, 2016; Jacob *et al.*, 2017), previniendo la agregación proteica irreversible dentro de la célula (Basha *et al.*, 2012; Poulain *et al.*, 2010), sino que su síntesis fluctúa considerablemente durante la maduración de los frutos (Lawrence *et al.*, 1997; Löw *et al.*, 2000; Neta-Sharir *et al.*, 2005). La expresión diferencial de las secuencias sHSPs dan indicios acerca de su comportamiento durante el proceso de maduración y posibilita establecer potencialmente genes candidatos que intervengan en la extensión de la vida poscosecha de los frutos, tal lo descrito por Arce *et al.* (2015) y Krsticevic *et al.* (2016).

Del total de 33 sHSPs en el genoma de referencia mundial de tomate (cv. Heinz 1706), se dirigió con principal atención en 4 sHSPs de sus miembros pertenecientes al cromosoma 6 de la especie y el cual representa aproximadamente el 57% del total de transcritos en el estado RM de este genoma de referencia. La identificación de estas sHSPs con expresión significativa entre las etapas de maduración pueden ser utilizadas como una región *target* en donde se desarrollen marcadores moleculares que permitan caracterizar poblaciones segregantes y luego identificar asociaciones con caracteres agronómicos de interés en la producción de este cultivo.

OBJETIVOS ESPECÍFICOS

1. Análisis y mapeo de los transcritos de las sHSPs diferencialmente expresadas en la secuencia genómica de cv. Caimanta de *S. lycopersicum* y la accesión LA0722 de *S. pimpinellifolium*.
2. Detección de polimorfismo molecular en la secuencia genómica correspondientes a dichos transcritos diferencialmente expresados en el cv. Caimanta y la accesión LA0722.
3. Desarrollo de marcadores moleculares que pongan en evidencia el polimorfismo detectado en la secuencia genómica.
4. Validación de los marcadores moleculares desarrollados en los materiales uniformes cv. Caimanta de *S. lycopersicum*, *S. pimpinellifolium* LA0722 y su F₁ CxP.

MATERIALES Y MÉTODOS

Análisis de la expresión génica de sHSPs

Tal lo mencionado en el Capítulo anterior, se consideraron diferenciales aquellos genes con un p-valor ajustado (FDR) < 0,05. El estado de maduración VM fue tomado como una referencia constante en la expresión diferencial de los genes. Es por eso que se plantearon en el software edgeR los contrastes entre los estados de maduración RM frente a VM y P frente a VM, para distinguir aquellos genes que se expresan diferencialmente entre los mismos. Los genes que presentan expresión diferencial por encima de esta referencia se los consideran sobre-expresados, y aquellos por debajo de la referencia inhibidos. Esta sección se centró en el análisis de las 33 sHSPs previamente exploradas por Arce *et al.* (2015) y Krsticevic *et al.* (2016), usando el comando *grep* frente a las tablas obtenidas con la expresión diferencial global de los genes y así poder seleccionar estos genes de interés. Luego se enfocó en las 4 sHSPs del cromosoma 6 (Soly06g076520, Soly06g076540, Soly06g076560 y Soly06g076570) ya que soportan el 57% de la expresión en el fruto rojo maduro tal se detalla también por Krsticevic *et al.* (2016).

Mapeo de los genes diferencialmente expresados, análisis de polimorfismos y desarrollo de cebadores

A fin de desarrollar marcadores moleculares que pongan en evidencia el polimorfismo presente en esta región de ~16,5 kb (SL4.0ch06:45227381-45243522, desde inicio del gen Soly06g076520 hasta el fin del gen Soly06g076570) donde mapean las sHSPs del cromosoma 6, se realizó la detección y localización de cada polimorfismo génico entre los genomas a partir de la información de expresión génicas obtenidas con los transcriptomas. La comparación de las secuencias nucleotídicas se realizó en un principio con el uso de la herramienta BLAST (Basic Local Alignment Search Tool) (Altschul *et al.*, 1997) para establecer las coordenadas de cada gen de sHSPs (\$ blastall -p blastn -d genoma.fasta -i sHSPs.fasta -m8 -o salida.txt). La extracción de la región se realizó con la herramienta *faidx* perteneciente al toolkit de SAMtools (Li *et al.*, 2009), siguiendo las coordenadas que se obtuvieron con anterioridad mediante la herramienta BLAST (\$ faidx genoma.fasta chr06:coordenada_inicio-coordenada_fin). La comparación de estas regiones en donde se encuentran las sHSPs se realizó utilizando los genomas de los progenitores alineados frente a la referencia de tomate del cv. Heinz 1706, desarrollado previamente por Cambiaso *et al.*, 2019^a. Luego, las regiones extraídas de ambos progenitores se alinearon con la herramienta *needle* que se encuentra disponible de manera online en el sitio web del EBI, del inglés

'European Bioinformatics Institute' (Rice et al., 2000) (https://www.ebi.ac.uk/Tools/psa/emboss_needle/). La comparación de estas regiones a través del alineamiento de sus secuencias permitió poner de manifiesto los polimorfismos que las diferencian. A partir de esto se detectaron Inserciones/Deleciones como también cambios de única base, y se desarrollaron marcadores moleculares (MM) del tipo *InDel* y *SNP* (polimorfismos de nucleótido único, del inglés *Single Nucleotide Polymorphism*) para posterior uso en la caracterización molecular de la población segregante. Los MM de tipo *InDel* son marcadores que detectan inserciones o deleciones provenientes de un genotipo o del otro. Por su lado, los *SNP* (pronunciado "esnips") son un tipo de polimorfismo que produce una variación en un único par de bases. Tanto los *InDels* como los *SNPs* pueden estar asociados a caracteres de interés y por eso pueden utilizarse como marcadores moleculares. Para desarrollar estos, la herramienta *online* Primer3 (Untergasser et al. 2012) (<https://primer3.ut.ee/>) permitió el diseño de los cebadores, considerando los parámetros 'any_th', '3'_th' y 'hairpin' cercanos o igual a 0, que indican alta calidad y especificidad.

Validación de los marcadores desarrollados

Para validar los marcadores desarrollados, se realizó la extracción de ADN siguiendo un protocolo estándar a partir de hojas jóvenes pertenecientes a cada uno de los genotipos uniformes: cv. Caimanta de *Solanum lycopersicum* (C), LA0722 de *Solanum pimpinellifolium* (P) y su F_1 interespecífica (CxP). El ADN extraído de estos genotipos de primer ciclo se chequeo en geles de agarosa al 1% p/v realizando corridas electroforéticas de 30 minutos a 140 V con buffer TAE 1X y tinción con SYBR® Safe (Thermo Fisher Scientific®, Waltham, MA, EEUU). La amplificación a partir de ADN genómico se realizó a través de ciclos de PCR. El protocolo de amplificación para el *InDel* 1 y 3 utilizado fue el que se detalla a continuación:

- Desnaturalización inicial de 2 minutos a 95 °C.
- 30 ciclos de amplificación:
 - Desnaturalización del ADN 1 minuto a 92 °C.
 - Hibridación del cebador de 1 minuto a 54 °C.
 - Extensión del amplicón 1 minutos a 72 °C.
- Elongación final de 5 minutos a 72 °C.

Para el *InDel* 2 el único cambio fue la temperatura de hibridación a 51 °C.

La separación de los fragmentos amplificados de cada marcador se llevó a cabo con geles de agarosa al 2,5 % p/v realizando corridas electroforéticas de 2 horas a 140 V constantes

con buffer TAE en concentración de 1X y se los visualizó mediante tinción con SYBR® Safe, a fin de obtener mejor discrepancia entre las bandas provenientes de los diferentes cultivares. Respecto a los marcadores de tipo SNP, se amplificó con PCR de 30 ciclos siguiendo el protocolo detallado a continuación:

- Desnaturalización inicial de 2 minutos a 95 °C.
- 30 ciclos de amplificación:
 - Desnaturalización del ADN 1 minuto a 92 °C.
 - Hibridación del cebador de 1 minuto a 52 °C.
 - Extensión del amplicón 1 minutos a 72 °C.
- Elongación final de 5 minutos a 72 °C.

Posterior a la amplificación se chequeó su buen funcionamiento en geles de agarosa al 1% p/v realizando corridas electroforéticas de 60 minutos a 140 V con buffer TAE 1X y tinción con SYBR® Safe, para su posterior secuenciación mediante tecnología *Sanger*®, que se explicará en el siguiente Capítulo.

RESULTADOS

Análisis de la expresión génica diferencial de sHSPs

El desarrollo de diagramas de *Venn* puso en evidencia genes de sHSPs que presentan expresión diferencial entre los estados de maduración para cada genotipo. Algunos de ellos se expresan de manera compartida y otros de manera única entre los estados Pintón y Rojo Maduro, en cada genotipo. Se detectaron 10 genes de sHSPs con expresión diferencial en el cv. Caimanta, 5 genes solo en el estado RM y 5 compartidos entre el estado P y RM (Figura II-1). La accesión silvestre hubo un total de 6 genes de sHSPs con expresión génica diferencial, detectandose 4 genes compartidos entre ambos estados, y 1 gen único tanto en P como en RM (Figura II-1).

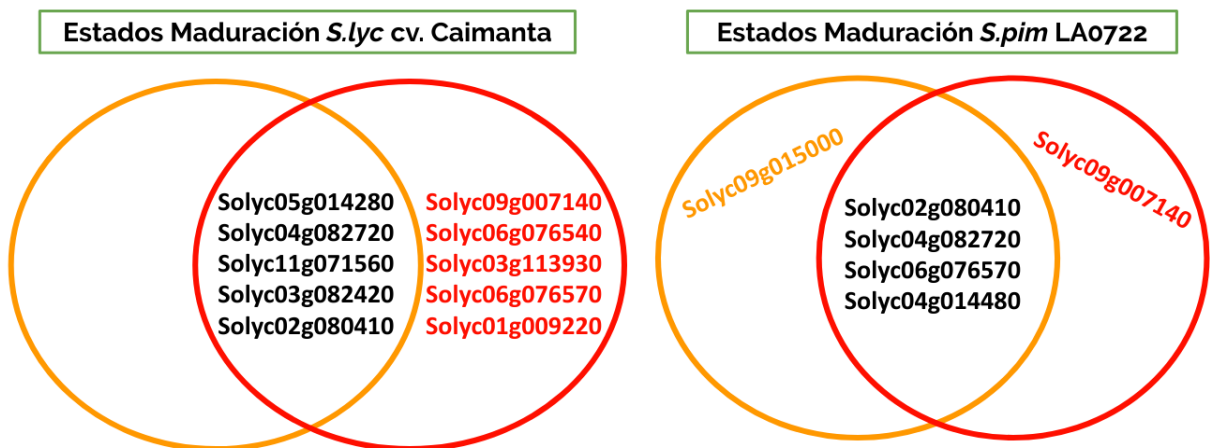


Figura II-1: Diagrama de Venn con genes de sHSPs expresados diferencialmente en los estados principales de maduración del fruto de cada genotipo progenitor. En color rojo genes de sHSPs con ED en RM y en naranja en el estado pintón.

Por otro lado, se colapsaron los estados de maduración para ver las sHSPs diferenciales entre los progenitores del programa de mejoramiento genético. Se detectaron 12 genes con expresión diferencial (Figura II-2). Dos de estos fueron únicos para el parental silvestres, 6 para el cultivar Caimanta y 4 genes de sHSPs fueron compartidos entre ambos. Además, se observaron los 5 genes con mayor significancia y ED en cada genotipo y estado (Figura II-2).

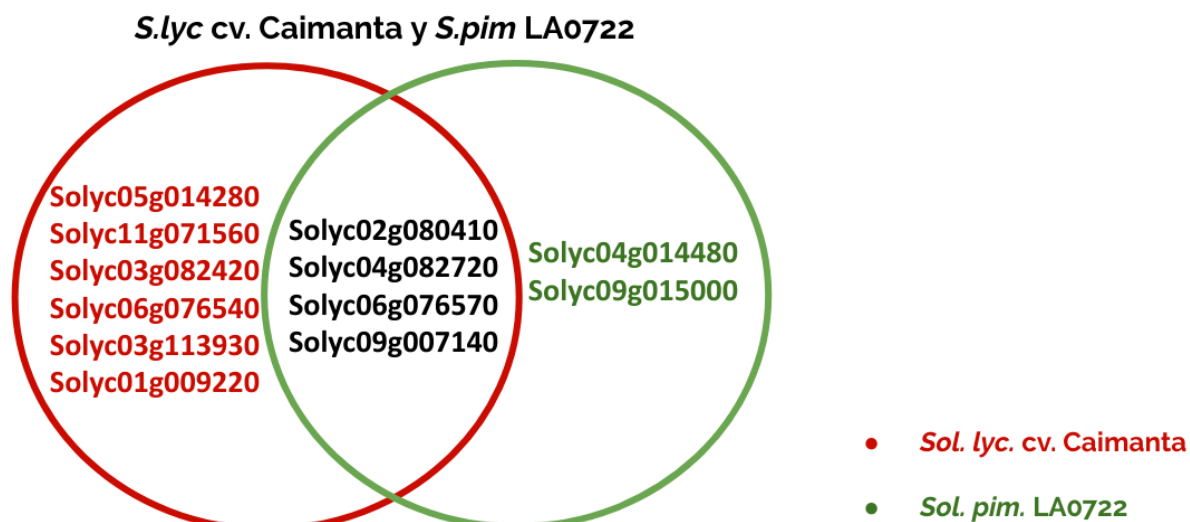


Figura II-2: Diagrama de Venn con genes de sHSPs expresados diferencialmente tanto únicos como compartidos entre los genotipos progenitores, unificando los estados de maduración del fruto. En color rojo genes de sHSPs en genotipo Caimanta y en verde en LA0722.

Top 5 de genes de sHSPs expresados diferencialmente en los genotipos uniformes

En el genotipo cultivado el gen Solyc05g014280.3 fue el gen con expresión diferencial más significativo de las 33 sHSPs para los dos estados madurativos en el material cultivado (Tabla II-1). Para el progenitor silvestre, el gen Solyc02g080410.3 fue el que presentó mayor significancia también en los dos estados. Para la F_1 no hubo concordancia entre los estados, ya que para el estado Pi vs VM el gen fue el Solyc08g062340.3 mientras que para el estado RM vs VM fue el gen Solyc03g113930.2 (Tabla II-1).

Los 5 genes con mayor significancia dentro de las 33 sHSPs en cada genotipo se muestran en la Tabla II-1.

Genotipo	Estado	GeneID	logFC	logCPM	F	P Value	FDR
Caimanta	Pintón vs Verde	Solyc05g014280.3	2,959480	9,721362	28,743723	0,000231	0,004222
		Solyc04g082720.3	1,068370	6,030240	16,578560	0,001796	0,016776
		Solyc11g071560.2	-2,839602	-0,703972	14,080967	0,003124	0,024437
		Solyc03g082420.3	3,327956	9,157060	15,593989	0,003313	0,025444
		Solyc02g080410.3	-3,084217	0,315559	10,277906	0,008300	0,048751
	Rojo Maduro vs Verde	Solyc05g014280.3	3,263746	9,721362	34,041871	0,000114	0,002457
		Solyc04g082720.3	1,286604	6,030240	23,825413	0,000467	0,006062
		Solyc02g080410.3	-5,069213	0,315559	20,294897	0,000879	0,009319
		Solyc09g007140.3	-4,767273	-0,647661	17,052485	0,001627	0,014045
		Solyc11g071560.2	-3,027144	-0,703972	15,519090	0,002256	0,017715
LA0722	Pintón vs Verde	Solyc02g080410.3	-3,991542	0,622837	35,594026	0,000076	0,001560
		Solyc04g082720.3	1,704297	6,892898	23,750730	0,000424	0,004971
		Solyc06g076570.3	3,336753	12,077104	16,364311	0,001897	0,014531
		Solyc04g014480.3	2,741457	9,494758	12,328102	0,005901	0,033202
		Solyc09g015000.3	3,204520	10,948753	10,918529	0,008427	0,043235
	Rojo Maduro vs Verde	Solyc02g080410.3	-2,166127	0,622837	13,618491	0,003273	0,023028
		Solyc04g082720.3	1,227509	6,892898	12,629597	0,004184	0,027426
		Solyc06g076570.3	2,715925	12,077104	11,438144	0,006047	0,035616
		Solyc09g007140.3	-2,785667	-0,279277	10,970824	0,006484	0,037483
		Solyc04g014480.3	2,663704	9,494758	11,713071	0,006827	0,038992
F ₁ (CxP)	Pintón vs Verde	Solyc08g062340.3	-4,215804	10,188178	61,100742	0,000002	0,001328
		Solyc03g113930.2	-3,570295	9,185512	60,412288	0,000002	0,001328
		Solyc11g020330.1	-4,615067	10,268703	54,734630	0,000004	0,001842
		Solyc01g102960.3	-3,592804	11,080206	46,146800	0,000010	0,002920
		Solyc05g014280.3	-4,029482	10,605024	44,531106	0,000013	0,003112
	Rojo Maduro vs Verde	Solyc03g113930.2	-2,621091	9,185512	35,351653	0,000041	0,001117
		Solyc06g076570.3	-2,363512	11,784570	25,968338	0,000182	0,003009
		Solyc11g020330.1	-2,838018	10,268703	24,362610	0,000243	0,003647
		Solyc02g080410.3	4,041807	0,426356	21,693912	0,000405	0,005129
		Solyc09g007140.3	4,085032	-0,315743	21,547529	0,000417	0,005242

Tabla II-1: Tablas con los 5 de genes de sHSPs con mayor significancia detectados en los transcriptomas de los frutos en los diferentes estados de maduración de los genotipos en estudio.

Siguiendo con los análisis sobre las sHSPs, se estudiaron aquellas detectadas previamente por Arce *et al.* (2016) y Krsticevic *et al.* (2016) como determinantes en gran proporción entre los estados de maduración. Se enfocó en aquellas del cromosoma 6 y 9, características por estar duplicadas en *tándem* y no presentar intrones.

En dicho estudio se destacó por encima de las otras la sHSPs Solyc06g076540.1 como aquella con mayor expresión diferencial y significancia entre los estados madurativos en los genotipos. Este gen se distinguió principalmente en el cv. Caimanta, obteniendo un

valor de sobre-expresión 3,92 (Log_2FC) en el estado RM en comparación del VM y de de 2,56 (Log_2FC) para el estado pintón, tal se detalla en la figura II-3.

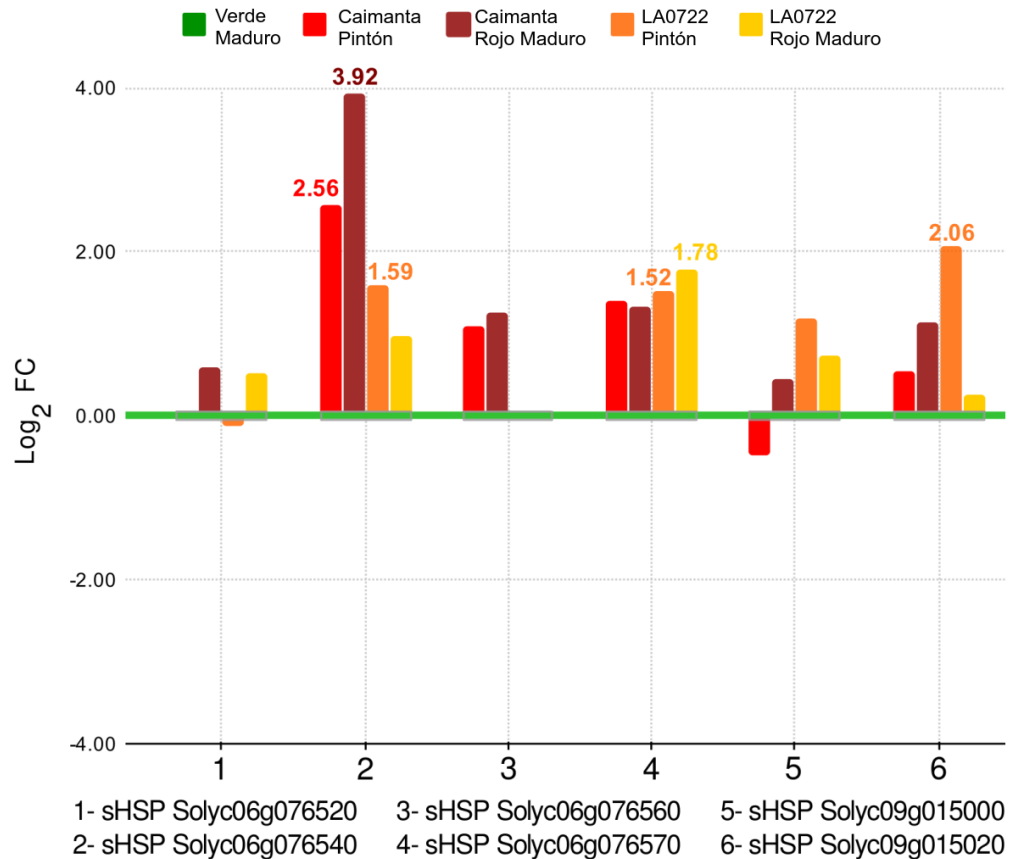


Figura II-3: Genes de small heat shock proteins (sHSPs) del cromosoma 6 y 9 con expresión génica diferencial (EGD) obtenidos a través de la secuenciación del ARN.

Siguiendo con el análisis hecho de sHSPs en esta sección (Tabla II-1 y Figura II-3), se plantearon nuevas comparaciones entre los transcriptomas a fin de dar énfasis a la expresión génica de estas sHSPs con el fin de seleccionar aquellas con sobre-expresión o inhibición de manera consistente. Como el fin de esta tesis es detectar efectos genéticos a través de estudios de expresión entre los progenitores y no desde un punto de vista fisiológico relacionado a la maduración del fruto, se realizaron comparaciones de los transcriptomas en el mismo estado de maduración pero entre los progenitores, el cultivado y el silvestre.

Estas comparaciones de la expresión génica dentro del mismo estado de maduración fueron: genotipo Caimanta vs LA0722 para VM, genotipo Caimanta vs LA0722 en P y genotipo Caimanta vs LA0722 en el estado RM. Esto evidenció que la sHSPs

Solyc06g076540 mantuvo una expresión diferencial (Tabla II-2) constante entre las 33 sHSPs en los 3 estados.

Posición dentro las 33 sHSPs	Estado	GeneID	logFC	logCPM	F	P Value	FDR
1	VM	Solyc06g076540.1	-7.46	9.77	28.77	0.0009	0.0357
5	P	Solyc06g076540.1	-6.66	12.46	14.26	0.0127	0.0893
1	RM	Solyc06g076540.1	-4.08	11.24	20.79	0.0016	0.0315

Tabla II-2: Genes de sHSPs expresados diferencialmente entre Caimanta y LA0722 dentro de cada estado. En color verde estado VM, en naranja estado P y en rojo estado RM.

Búsqueda de polimorfismo molecular en la secuencia genómica de los progenitores

Posteriormente, de acuerdo a los resultados obtenidos se comparó la región completa de estas sHSPs del cromosoma 6, ya que la expresión diferencial se observó en todos los genes de dicha región y los polimorfismos a detectar podrían estar posiblemente distribuidos en toda la región. Además, no solo la sHSPs Solyc06g076540.1 sino toda la región es fuertemente estudiada y representa una región clave con alta expresión génica (Arce *et al.*, 2015), como también presenta la particularidad que sus genes reaccionan frente a uno de los reguladores de crecimiento vegetal que participa dentro del proceso fisiológico de la maduración del fruto, como es el etileno, y en donde es bien reportado en su participación en conjunto con las sHSPs como se demuestra en el trabajo realizado por Shukla *et al.* (2017). A partir de esto se obtuvieron las regiones genómicas (Figura II-4), de esta región de aprox. 20 kb (más 1000 pb a fin de obtener las regiones promotoras de los genes de los extremos) aproximadamente para detectar polimorfismo y desarrollar MM.

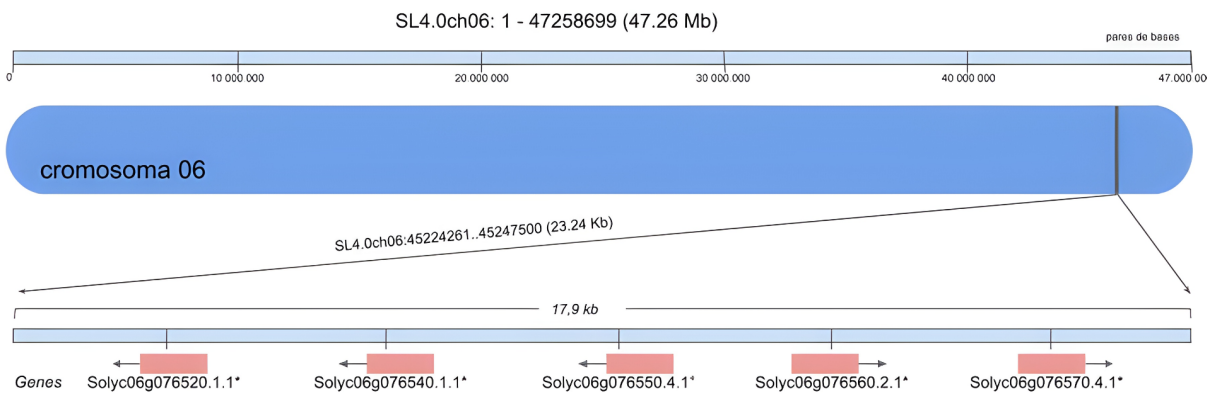


Figura II-4: Localización de genes de sHSPs en la región genómica de interés del genoma de referencia (SL4.0 - cv. Heinz 1706) para detección de polimorfismo molecular.

Identificación del polimorfismo detectado

La comparación se llevó a cabo con la herramienta *needle* y arrojó la siguiente información, como puede observarse en detalle en la figura II-5.

Aligned_sequences: 2 1: 47547349-19770 2: 47552636-19725 Matrix: EDNAFULL Gap_penalty: 10.0 Extend_penalty: 0.5		Length: 19787 Identity: 19651/19787 (99.3%) Similarity: 19656/19787 (99.3%) Gaps: 79/19787 (0.4%) Score: 97873.0	
47547349-1977	16841	TTGATGCTTTAAAATTACATTACGCATATACACCACATAGAGGTTACATT	16890
47552636-1972	16812	TTGATGCTTTAAAATTACATTACGCATA-----TACATT	16845
47547349-1977	16891	ATTTCAACTTCGCAGGCTTACAATACCAAGATTCATGTTATGGTACCATA	16940
47552636-1972	16846	ATTTCAACTTCGCAGGCTTACAATACCAAGATTCATGTTATGGTACCATA	16895
47547349-1977	16941	TATATTGAATTATTATATGTTATGTAGCAAATTATCAATGTTATTATTTT	16990
47552636-1972	16896	TATATTGAATTATTATATGTTATGTAGCAAATCAATGTTATTATTTT	16945

Figura II-5: Ejemplo de la detección de polimorfismos en dicha región. Con guiones (-) el programa detecta polimorfismos del tipo *InDels* mediante el alineamiento entre las secuencias, y con puntos (.) aquellos polimorfismos del tipo *SNP*. La secuencia 1 pertenece al cv. Caimanta y la n° 2 a la accesión LA0722.

Esta comparación demostró una similitud entre estas secuencias de *SL*. cv. Caimanta y LA0722 de *S. pimpinellifolium* del 99,3%. Dicha región tiene un tamaño de 19787 bases de largo, y se procedió a la búsqueda de polimorfismo del tipo *SNP* (del inglés, *Single Nucleotide Polymorphism*) y/o *InDel* (Inserción/Delección) a lo largo de toda su longitud.

Se detectaron 15 *InDels* a lo largo de la región. La mayoría de pequeño tamaño (1,2 o 3 bases). Solo 5 son de un tamaño medianamente mayor, de entre 8 a 16 bases. También se detectaron 53 *SNP* a lo largo de la longitud de la región de interés en estudio. Este polimorfismo es clave para el potencial desarrollo de marcadores moleculares por fuera de la frontera del polimorfismo, y así realizar la caracterización de la F_4 , ya caracterizada fenotípicamente. Debajo se muestran ejemplos de un *InDel* y un *SNP* para desarrollar con Software Primer3⁴ el marcador correspondiente.

⁴ <http://primer3.ut.ee/>

Ejemplos de polimorfismos en donde pueden desarrollarse marcadores moleculares podrían ser los siguientes:

-InDel:

```

                                cebador_sentido->
Seq_Caimanta      TTTTGTATAACACATCATAATAAACTCTTCTAATGATGTGTACAGGACAAGTGAAT
                   |||                               |||
Seq_LA0722        TTTTGTATAACACA-----TCTAATGATGTGTACAGGACAAGTGAAT
                                           <-cebador_antisentido

```

-SNP:

```

                                cebador_sentido->
Seq_Caimanta      CACCTGTAATAGTATTTATAAGGACTAGATAACAAATTAATGGGATAATGGACTAACA
                   |||                               |||
Seq_LA0722        CACCTGTAATAGTATTTATAAGGACTAGCTAACAAATTAATGGGATAATGGACTAACA
                                           <-cebador_antisentido

```

Diseño y desarrollo de cebadores a partir del tipo de polimorfismo detectado

A partir de toda la información identificada en la secuencia génica, se obtuvieron 3 *InDels* para caracterizar molecularmente luego la población F₄ segregante de tomate. La lista de *InDels* obtenida a partir de los análisis llevados a cabo fue la siguiente:

- **InDel #1** Izquierdo: 5'-GATGTTCTGATGTCTGATGTAGTCT-3' - Derecho: 5'-TGATTCTATTATGTATAAAGTAGCAGAGG-3' - Amplicón de 402 bases de tamaño. Alinea/empareja con la región codificante del gen Solyc06g076550.4.1 - Diferencia de 14 pb entre los genotipos
- **InDel #2** Izquierdo: 5'-TTAAGCCAGCAAATTAACAACACTGT-3' - Derecho: 5'-AAATAAGGTTTTTCATGTTTAGCCCT-3' - Amplicón de 336 bases de tamaño. Alinea/empareja con la región promotora del gen Solyc06g076570.4.1 - Diferencia de 8 pb entre los genotipos
- **InDel #3** Izquierdo: 5'-GGATGATGGTGTGTAAGAATGATGA-3' - Derecho: 5'-GTAAGCCTGCGAAGTTGAAATAATG-3' - Amplicón de 559 bases de tamaño. Alinea/empareja con la región promotora del gen Solyc06g076570.4.1 - Diferencia de 16 pb entre los genotipos.

Validación de los marcadores moleculares desarrollados

Una vez desarrollados dichos cebadores y sintetizados, se realizaron amplificaciones por PCR y corridas electroforéticas para probar su funcionamiento en los materiales de primer ciclo. Como puede observarse en la Figura II-6, para el *InDel* 1 se visualizaron las bandas esperadas tanto en Caimanta y LA0722 aunque la F_1 no mostró un patrón de herencia codominante sino un único amplicón de igual tamaño que el del cultivar. Por su parte, los *InDel* 2 y 3 sólo amplificaron en Caimanta y en la F_1 , por lo que en los tres casos la herencia resultó dominante, presentando la accesión silvestre las alternativas recesivas de los 3 *InDels*.

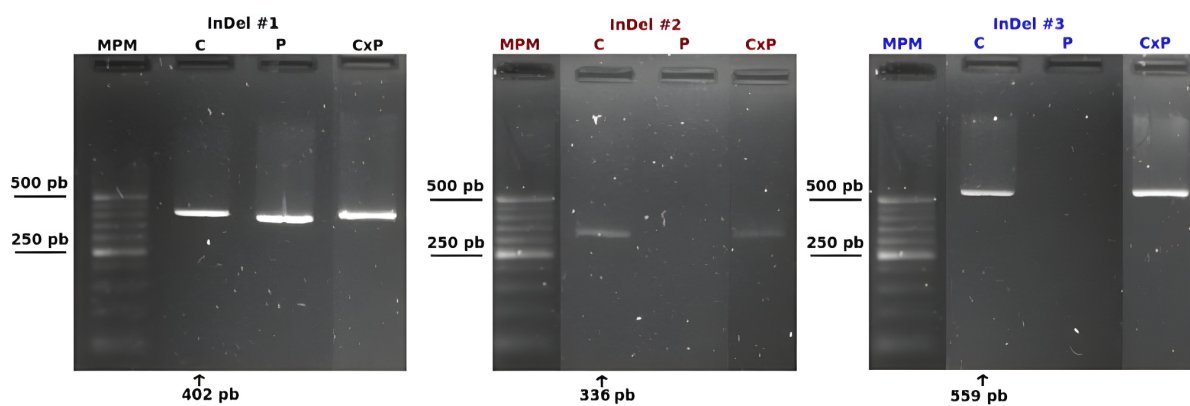


Figura II-6: Geles de agarosa con visualización de bandas de los 3 *InDels* desarrollados amplificados en los genotipos de interés. (MPM: marcador de peso molecular)

Por su parte, para desarrollar marcadores moleculares de tipo *SNP* fueron detectados cambios de nucleótidos en una posición puntual, y se desarrolló un par de cebadores para levantar un amplicón de 660 pares de bases que contenía en dicha secuencia 6 *SNP* y 5 *InDels* (menores a 10 bases). Este *SNP* se obtuvo para caracterizar molecularmente la población F_4 segregante. Este par de cebadores arrojó la siguiente información a partir de los análisis llevados a cabo:

- *SNP* #1 - Izquierdo: 5'-TTTGATGGTTGATGGAAGAGGA-3' - Derecho: 5'-GCATATTTGACCCTTTTCCCTTAA-3' - Amplicón de 660 bases de tamaño. Alinea/empareja con la región promotora del gen Solyc06g076540.1.1

Este MM de tipo *SNP* fue chequeado en gel de agarosa (Figura II-7) con las mismas condiciones que los *InDels* y luego secuenciado mediante tecnología *Sanger*® en laboratorio tercerizado, a fin de detectar polimorfismos asociados a caracteres de interés agronómicos en la población F_4 derivada del cruzamiento ToUNR18xToUNR1.

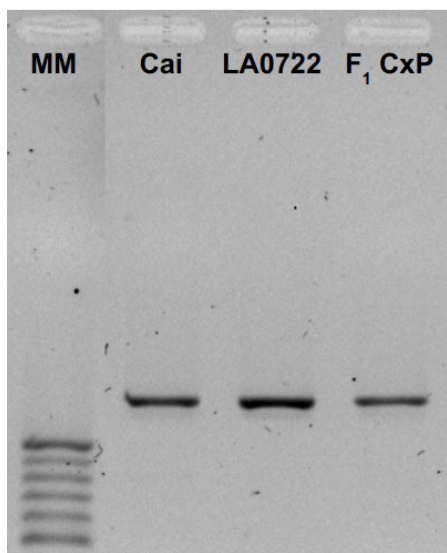


Figura II-7: Gel de agarosa con amplificación mediante el MM de *SNP* desarrollado.

DISCUSIÓN

La genómica tiene como una de sus potenciales metas identificar los *loci*, genes y/o variantes génicas responsables de la variabilidad presente en los rasgos fenotípicos. El cruzamiento interespecífico entre diferentes especies elegidas de tomate genera variación asociada con la introgresión de genes dentro del nuevo acervo genético de la población obtenida el mismo. Así es de esperar que la distribución de estas variantes y polimorfismo molecular sea en todo el genoma completo y en diferentes cantidades de acuerdo a cada región en particular. En este caso el polimorfismo detectado mediante la comparación de los genomas de ambos progenitores arrojó una similitud e identidad del 99,3% respecto a los 20 kb de la región de interés (Figura II-5). Esto representa una alta identidad para ambos genotipos, aunque a pesar de esto la variabilidad presente en dichas secuencias génicas permitió el desarrollo de marcadores que detecten este polimorfismo molecular.

Por su parte, la sHSP Solyc06g076540.1 es una de las sHSPs expresadas de manera más diferencial durante la maduración del fruto de tomate (Figura II-3). Estudios previos de Krsticevic *et al.* (2016) encontraron similares resultados respecto a este gen en sus análisis. Otro estudio detecta que este gen es producto de eventos de duplicación en *tándem*, el cual el primer evento tuvo lugar hace 13 millones de años aproximadamente (Goyal *et al.*, 2012). Este gen fue capaz de mantenerse frente a los diferentes pasos evolutivos y estar presente en la actualidad. Otros estudios sostienen que la capacidad de reacción de este gen frente a diferentes estrés de tipo biótico y abiótico refiere a que, al no presentar en su secuencia regiones intrónicas, su expresión se desencadena rápidamente al evitar los eventos de splicing alternativo en presencia de estos estímulos externos.

Estudios similares realizados obtienen resultados que ponen de manifiesto la importancia de este gen y la del resto de sus miembros. Estos genes de sHSPs del cromosoma 6, en especial el gen de Solyc06g076540.1 (Tabla II-2), presentan expresión diferencial significativa en las accesiones silvestres utilizadas por Upadhyay *et al.* (2020). Además, Shukla *et al.* (2017) realizaron experimentos con el uso de la fitohormona etileno, en donde se observó que la aplicación de etileno disminuye o inhibe la expresión de estas sHSPs, y por ende, indicaría que su expresión interviene en la maduración del fruto. La inhibición detectada de estas sHSPs en este Capítulo, en la cual estos genes son regulados negativamente durante el transcurso del tiempo desde el estado VM al RM pone de manifiesto que la síntesis natural de etileno en el fruto de tomate durante su maduración podría moderar el accionar de estas sHSPs y llevar a cabo el camino inverso, o sea, la

activación o síntesis de estas chaperonas podrían extender la vida poscosecha del fruto, tal lo descrito por Shukla *et al.* (2017) y Upadhyay *et al.* (2020).

Estudios previos en estas regiones cromosómicas de interés demostraron que existe variabilidad en el número y disposición de sus genes, en particular entre las accesiones silvestres presentes en el germoplasma de tomate. Esta variabilidad en el número y disposición de los genes de sHSPs en dicha región podría explicar la herencia dominante observada en la validación de los *InDels* frente a los progenitores del programa de mejoramiento y su cruzamiento. Podría ocurrir que frente a la condición heterocigota de los cruzamientos interespecíficos, los cebadores amplifican preferentemente sobre el genoma aportado por el progenitor cultivado. Cabe aclarar que la condición heterocigota de las F_1 ya había sido verificada previamente con otros MM por Cabodevila *et al.* (2021), y que a pesar de este comportamiento dominante no esperado, el uso de los 3 *InDel* para caracterizar poblaciones segregantes no impide su uso (Figura II-6), sino que a lo sumo sólo condiciona su análisis al de MM de naturaleza dominante.

Estos comportamientos detectados pueden sustentarse respecto a la idea previamente discutida en el Capítulo anterior, en el que los genomas de Caimanta y LA0722 fueron ambos alineados frente a la referencia mundial del cultivo, el cv. Heinz 1706, y esto pudo haber originado sesgos y/o forzar arreglos genéticos y/o cromosómicos que resultan en artefactos en el genoma de LA0722, tal como Cacchiarelli *et al.* (2021) demostraron para el análisis transcriptómico. El alineamiento de transcriptomas de diferentes especies frente a una especie en común y la normalización de los transcriptomas frente a los diferentes estados del desarrollo permite llevar a cabo las comparaciones que demande el estudio. Este tipo de análisis es un primer reflejo del comportamiento global de todas las familias o grupos de genes, aunque podría generar sesgos o efectos indeseables frente a una región en particular o genes *target* de interés.

CONCLUSIONES

- Existen genes de sHSPs diferencialmente expresados en los diferentes estados de maduración del fruto de tomate, identificados mediante el análisis de los transcriptos.
- Se detectó polimorfismo molecular en la secuencia génica de los progenitores en dicha región en donde las sHSPs se expresan diferencialmente entre los estados de maduración.
- El polimorfismo molecular detectado previamente permitió el desarrollo de marcadores moleculares de diferente naturaleza que ponen de manifiesto variantes génicas entre los progenitores.
- Se validaron los marcadores moleculares en los progenitores considerados uniformes con el fin de avanzar en la caracterización molecularmente de una población segregante F₄.

CAPÍTULO III

CARACTERIZACIÓN FENOTÍPICA Y MOLECULAR DE LA POBLACIÓN SEGREGANTE F₄ DE UN HÍBRIDO DE SEGUNDO CICLO Y DETECCIÓN DE *QTLs* ASOCIADOS A CARACTERES DE CALIDAD DE FRUTO

INTRODUCCIÓN

La obtención de nuevas variedades éliticas de tomate es el principal objetivo dentro de los programas de mejoramiento genético. Este proceso suele generarse mediante el uso de muchos recursos de diferente naturaleza, y uno de ellos es el tiempo. Estos programas son iniciados con la inspección y explotación del germoplasma de tomate. De las 13 especies pertenecientes al género *Solanum* sección *Lycopersicon*, la especie silvestre *S. pimpinellifolium* se caracteriza por presentar frutos de menor tamaño y peso pero de alta calidad nutritiva. Esta especie es capaz de mantener las propiedades organolépticas del fruto durante más largo tiempo, incluyendo la vida poscosecha. Uno de los desafíos dentro de los programas es detectar e identificar aquellas regiones introgresadas en la progenie que aporten nuevas variantes génicas asociadas a los caracteres de calidad. Este tipo de estrategia aumenta la variabilidad y evita el fenómeno de erosión genética estudiado en los materiales cultivados, por la uniformidad genética de sus semillas.

Los caracteres fenotípicos de interés agronómico son en su mayoría caracteres complejos, controlados por muchos genes, cada uno con un efecto aditivo relativamente pequeño sobre el carácter en estudio (Falconer y Mackay, 1996). Este tipo de caracteres son llamados cuantitativos mientras que la región genómica asociada se denomina QTL. En poblaciones segregantes derivadas del cruzamiento entre progenitores divergentes para los caracteres de interés es posible identificar nuevas regiones genómicas que controlan estos caracteres (Rodríguez *et al.*, 2013). Estas regiones de interés pueden identificarse mediante el uso de marcadores moleculares. En el Capítulo anterior se desarrollaron y diseñaron marcadores en base a la secuencia génica de los progenitores en donde se detectó polimorfismo molecular. En esta región mapean 4 sHSPs que intervienen en el proceso biológico de la maduración del fruto y que presentan expresión diferencial significativa frente a las etapas principales de este proceso.

Para detectar QTLs asociados a caracteres de interés es necesario realizar una caracterización fenotípica y una caracterización molecular en la población segregante. Este requisito fue llevado a cabo en el presente Capítulo en donde se caracterizó fenotípica y molecularmente una generación F_4 con la finalidad de obtener marcadores moleculares funcionales que asistan a los programas de mejora. Estos pueden utilizarse para realizar una selección asistida por marcadores y así disminuir significativamente los tiempos necesarios hasta la obtención de una nueva variedad con caracteres de calidad superiores.

OBJETIVOS ESPECÍFICOS

1. Caracterización fenotípica de la población F_4 generada a partir del híbrido de segundo ciclo (HSC) ToUNR18xToUNR1.
2. Caracterización molecular de la población F_4 a través de marcadores de tipo *InDels* y *SNP* desarrollados a través del polimorfismo molecular detectado.
3. Detección de *QTLs* asociados a caracteres de calidad de fruto en base a la variabilidad molecular y fenotípica relevada.
4. Comparación de los *QTLs* detectados en la caracterización molecular mediante el uso de los *InDels* y *SNP*.

MATERIALES Y MÉTODOS

Material Vegetal

La generación segregante F_4 obtenida a partir del cruzamiento ToUNR18xToUNR1 fue la caracterizada fenotípica y molecularmente en este Capítulo. Las plantas fueron seleccionadas a través de los datos fenotípicos de la F_3 , seleccionando de manera antagónica y divergente para las variables peso y vida poscosecha. Se sembraron 12 plantas de cada una de las 18 familias F_4 ($n=216$) más 10 plantas de los 8 genotipos uniformes como testigos ($n=80$) en bandejas de germinación de 72 celdas de 55 cm³ utilizando el sustrato N°1 de Dynamics (www.dynamics compost.com.ar), dando un total de 296 plántulas. A los 45 días se realizó el trasplante a tierra en invernadero semicontrolado, en un diseño completamente aleatorizado. Las plantas fueron regadas por goteo diariamente durante todo el ciclo del cultivo evitando el estrés hídrico de las mismas mediante la variación en la cantidad de agua suministrada. La temperatura establecida para el invernadero fue de 27 ± 4 °C, la intensidad de luz de 700 $\mu\text{mol.s}^{-1}\text{m}^{-2}$ y la humedad relativa de 60 %.

Caracterización fenotípica de la generación segregante F_4

Se cosecharon frutos de las 18 familias de dicha generación en la que se examinaron todos los caracteres de interés agronómicos. En diez frutos en estado pintón se evaluaron los siguientes caracteres: peso (en g), diámetro (en cm), altura (en cm), forma (cociente entre la altura y el diámetro) y vida poscosecha (VP), expresada como los días transcurridos desde la cosecha hasta el inicio del ablandamiento del fruto. Para evaluar este último carácter los frutos se almacenaron en estantería a 25 ± 3 °C de temperatura ambiente. Siguiendo la metodología propuesta por Schuelter *et al.* (2002) los frutos se examinaron tres veces por semana descartando aquellos comercialmente inaceptables por mostrar arrugamiento o excesivo ablandamiento. Aproximadamente a seis u ocho frutos de cada planta en estado rojo maduro (RM) se les midió la dureza con un durómetro manual tipo Shore A Durofel y el color con un cromámetro MINOLTA CR400, a través del porcentaje de reflectancia (L), variando desde 0 para negros a 100 para blancos y el cociente a/b, siendo “a” la absorbancia a longitudes de onda de 540 nm y “b” la absorbancia a longitudes de onda de 675 nm. Se obtuvo luego el jugo del pericarpio a partir de frutos cosechados en estado rojo maduro y se evaluaron los siguientes caracteres: contenido en sólidos solubles (SS en °Brix), medido con un refractómetro manual; acidez, medida a través del pH del jugo homogeneizado y de la acidez titulable (AT en gramos de ácido cítrico/100 gramos de jugo).

Posterior a esto se hizo tratamiento de semillas y se las almacenó en cámara de refrigeración para proseguir con el programa de mejoramiento. Con la información recolectada, se probó la distribución normal mediante la prueba propuesta por Shapiro y Wilk (1965). Los datos recolectados se compararon mediante un ANOVA a un criterio de clasificación, a partir de cuyos cuadrados medios se calculó la heredabilidad en sentido estricto (h^2). Debido al porcentaje de homocigosis esperado para esa generación, el coeficiente fue corregido multiplicándose por 7/6, producto que permitió finalmente estimar la h^2 para cada carácter de interés.

Caracterización molecular por InDels de la población F₄

La caracterización molecular se realizó a través del uso de los dos tipos de marcadores moleculares, en este caso los *InDels* que fueron previamente validados en los materiales uniformes de primer ciclo. La extracción de ADN se realizó a partir de discos de 40 mg de tejido foliar joven según protocolo estandarizado. Estos *InDels* se amplificaron mediante la técnica de reacción en cadena de la polimerasa (PCR, Polymerase Chain Reaction) siguiendo los pasos del protocolo correspondiente y modificando la temperatura de anillado de cada cebador *InDel*, tal se describió en el capítulo anterior. La separación de los fragmentos amplificados fue en geles de agarosa al 2,5 % p/v con corridas electroforéticas de 2 horas a 140 V constantes en buffer TAE 1X y se los visualizó mediante tinción con SYBR® Safe.

Caracterización molecular por SNP de la población F₄

El marcador molecular tipo *SNP* desarrollado en el Capítulo anterior fue utilizado en esta sección para la caracterización molecular. Este marcador se utilizó en la población de 90 plantas F₄ previamente caracterizadas con los *InDel*, siendo seleccionadas 30 plantas de alto peso, 30 de bajo peso y 30 de peso intermedio. Esta caracterización fue llevada a cabo a través del uso de la pirosecuenciación de tecnología *Sanger*®. Se utilizó el mismo ADN extraído en la caracterización con *InDels*. La amplificación por PCR se llevó a cabo según protocolo correspondiente y con modificación en la temperatura de anillado, descrito en el capítulo anterior. La calidad de las amplificaciones de estos *SNP* fue chequeada en geles de agarosa al 1,5 % p/v realizando corridas electroforéticas de 30 minutos a 140 V constantes, visualizando las bandas mediante tinción con SYBR® Safe. El total de muestras a laboratorio tercerizado (Centro Científico, Tecnológico y Educativo 'Acuario del Río Paraná' - Laboratorio Mixto de Biotecnología Acuática) en donde previo a la secuenciación se

corroboró la calidad de las muestras mediante espectrofotómetro NanoDrop® 2000 y se purificó por método enzimático.

Detección de QTL mediante el uso de marcadores tipo InDels

La detección de QTL se llevó a cabo utilizando el programa estadístico *Infostat* (<https://www.infostat.com.ar/>) (Di Rienzo *et al.*, 2010) empleando los marcadores moleculares *InDels*. La detección de estos QTLs se hizo con análisis de único punto, pero con dos modelos estadísticos diferentes, un ANOVA unifactorial por un lado, teniendo en cuenta sólo la acción del marcador molecular, y por otro lado, un ANOVA bifactorial teniendo en cuenta el efecto del marcador molecular, la familia y la interacción entre ellos. Todas las variables fueron consideradas significativas con un p-value < 0,05.

Detección de QTL mediante el uso de marcadores tipo SNP

En relación a *SNP*, primero se obtuvieron las secuencias, se eliminaron los caracteres especiales, y se corrigieron los IDs de cada planta para posterior identificación. La clusterización de los amplicones obtenidos en la secuenciación de los *SNP* fue llevada a cabo en la plataforma online del EBI utilizando la herramienta Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) que permite alinear múltiples secuencias a la vez, usando los parámetros de manera predeterminada. La detección de estos QTLs se hizo al igual que con *InDels*, con análisis de único punto pero con un ANOVA unifactorial y uno bifactorial (por un lado sólo con la acción del marcador molecular, y por otro con el efecto del marcador molecular, la familia y la interacción entre ellos). Además, como este es un marcador de tipo codominante se calcularon los grados de dominancia, empleando un test de Duncan para ver la comparación de medias y el efecto de cada *SNP*. Las variables fueron consideradas significativas con un p-value < 0,05.

Comparación entre las detecciones de QTLs con los MM desarrollados

Los valores de p-value de los QTLs detectados se compararon entre los *InDels* y los *SNP*, previamente utilizados para la caracterización molecular. El número de plantas analizadas fue distinto entre ambos MM, es por eso que se volvió a realizar el análisis estadístico de los *InDels* pero solo en aquellas plantas caracterizadas por *SNP*. Se estudió la concordancia y coincidencia entre los MM a fin de analizar su efectividad.

Re-análisis de los marcadores *InDels* en las plantas caracterizadas por *SNP*

En la última parte se re-analizaron los *InDels* desarrollados en las plantas seleccionadas a través del uso de los marcadores moleculares de tipo *SNP*, a fin de refinar la detección *QTLs*. Este análisis se realizó conservando el número de plantas caracterizadas por los *SNP* con el fin de que no se vea afectada por el número de individuos de la población (N).

Para finalizar, se estableció un flujo de trabajo en el que se engloban los diferentes experimentos realizados, entre ellos las caracterizaciones descritas en este Capítulo.



Figura III-1: Flujo de trabajo realizado para detección de *QTLs* agronómicos. (EGD: Expresión Génica Diferencial; WGS: secuenciación del genoma completo del inglés: *Whole Genome Sequencing*; MM: Marcador Molecular).

RESULTADOS

Caracterización fenotípica de la generación segregante F_4

En las plantas F_4 se probó la distribución normal mediante la prueba propuesta por Shapiro y Wilk (1965). En este análisis se observó la normalidad de todas las variables estudiadas. Todas las variables fueron normales, con valores de W por encima de 0,91, salvo Peso un valor de 0,86 por lo cual fue transformado por logaritmo natural.

Se cosecharon en promedio 2377 frutos para el estudio de los caracteres morfológicos de los mismos, y un promedio de 1471 frutos para el estudio de los caracteres organolépticos, dependiendo la cantidad de frutos por planta de acuerdo al tamaño de los mismos. Para medir pH, SS y AT se lograron 240 jugos en promedio a partir de los frutos cosechados en invernadero (Figura III-2).



Figura III-2: Plantas y frutos de tomate cosechados para obtención de tejido joven y datos fenotípicos en la población segregante perteneciente a la generación F_4 .

Caracteres morfológicos de los frutos

Se agruparon los caracteres peso (Pe), diámetro (D), altura (A), forma (F, altura/diámetro) y vida poscosecha (VP). Los resultados mostraron h^2 altas (mayores a 0,72) para los todos los caracteres excepto para VP que tuvo un valor intermedio (0,46). Los

valores de h^2 para esta generación F_4 fueron superiores a los estimados previamente en la generación F_3 por Cabodevila *et al.* (2021), lo que puede explicarse por el aumento en la proporción de homocigosis a medida que se avanza en las generaciones de autofecundación. Pueden observarse en la Tabla III-1 las familias que obtuvieron los valores medios máximos y mínimos para cada rasgo agronómico. Se destaca la familia 18P61 con los valores máximos para D, A, y Pe. La familia con mayor VP fue la 18P55 (Tabla III-1). A partir de estos datos puede notarse cómo se avanza hacia la obtención de nuevas líneas mejoradas que presentan caracteres agronómicos deseables, capaces de satisfacer las demandas de los mercados locales y las exigencias tanto de los productores como la de los consumidores.

Genotipos Uniformes					Población Segregante F_4			
CE	CAIMANTA	LA0722	$F_1(C \times P)$	h^2	Media Mín.	Familia	Media Máx.	Familia
D	6,39 $\pm$ 0,24	1,20 $\pm$ 0,01	2,59 $\pm$ 0,08	0,74	1,49	18P53	3,29	18P61
A	5,05 $\pm$ 0,18	1,09 $\pm$ 0,01	2,35 $\pm$ 0,08	0,77	1,30	18P53	3,15	18P61
F	0,80 $\pm$ 0,02	0,90 $\pm$ 0,00	0,90 $\pm$ 0,01	0,82	0,86	18P64	1,21	18P54
Pe	123,23 $\pm$ 9,90	0,95 $\pm$ 0,03	11,19 $\pm$ 0,86	0,72	1,71	18P53	20,72	18P61
VP	9,97 $\pm$ 1,60	13,79 $\pm$ 0,85	23,02 $\pm$ 1,64	0,46	17,00	18P49	36,48	18P55

Tabla III-1: Valores medios y su error estándar para los Caracteres Evaluados (CE) en los genotipos uniformes y valores medios máximos y mínimos de cada CE en las familias perteneciente a la generación F_4 .

Caracteres organolépticos de los frutos

Por otro lado, se analizaron los datos fenotípicos de las variables organolépticas de los frutos. Estas son sólidos solubles (SS), acidez titulable (AT), pH, dureza (Du), porcentaje de reflectancia (L), a (absorbancia a longitudes de onda de 540 nm), b (absorbancia a longitudes de onda de 675 nm) y el cociente a/b. En esta generación F_4 se detectó variación en todos los caracteres organolépticos analizados, presentando algunos un rango de variación que supera a los genotipos uniformes. Sin embargo, el porcentaje de variación fenotípica debido a variación aditiva es de medio a bajo, como lo demuestran los valores de h^2 estimados, lo que implica que la selección artificial presenta limitaciones para continuar modificando la media de estos caracteres. En consecuencia, será necesario recurrir a otras estrategias de mejoramiento, tales como nuevas hibridaciones entre las diferentes familias a fin de usufructuar el componente de variación no aditiva o bien con materiales de otro fondo

genético con la intención de incrementar la variación aditiva, en esta generación F_4 . Los resultados obtenidos se muestran en la Tabla III-2.

CE	<i>a</i>		<i>b</i>		<i>a/b</i>		<i>L</i>		<i>Du</i>		<i>SS</i>		<i>pH</i>		<i>AT</i>	
Ge	Me	D.E	Me	D.E	Me	D.E	Me	D.E	Me	D.E	Me	D.E	Me	D.E	Me	D.E
C	22,88	3,73	24,79	2,92	0,93	0,15	40,45	2,99	50,55	10,37	5,13	0,65	4,20	0,14	6,29	1,30
P	30,88	3,78	22,27	2,84	1,40	0,20	36,89	1,57	57,46	7,46	7,30	0,89	4,58	0,19	2,48	4,28
CxP	21,12	3,61	18,94	3,12	1,12	0,15	37,23	2,04	51,78	8,35	5,76	0,61	4,49	0,17	3,79	1,45
F_4	25,11	3,66	20,03	2,77	1,27	0,17	38,42	2,01	56,90	7,06	6,37	1,29	4,60	0,21	3,82	1,59
h^2	0,30		0,22		0,28		0,32		0,33		0,27		0,29		0,28	

Tabla III-2: Valores medios (Me) y desviación estándar (D.E.) para los caracteres evaluados (CE) en cada genotipo (Ge) y heredabilidad en sentido estricto (h^2) estimada en las familias F_4 .

A partir de estos datos recolectados, se calculó heredabilidad en sentido estricto (h^2) de cada carácter en estudio mediante un ANOVA a un criterio de clasificación, para verificar si la segregación fenotípica en esta F_4 se debe a componentes genéticos, y se los comparó con los obtenidos en la generación F_3 perteneciente a este mismo cruzamiento (Tabla III-3), caracterizada previamente por Cabodevila *et al.* (2017).

Variable	h^2 en la generación F_3	h^2 en la generación F_4
<i>Diámetro</i>	0,46	0,74
<i>Altura</i>	0,54	0,77
<i>Forma</i>	0,42	0,82
<i>Peso*</i>	0,60	0,72
<i>Vida Poscosecha</i>	0,28	0,46
<i>a</i>	N/C	0,30
<i>b</i>	N/C	0,22
<i>a/b</i>	0,52	0,28
<i>L</i>	0,36	0,32
<i>Dureza</i>	0,38	0,33
<i>SS</i>	0,77	0,27
<i>pH</i>	N/C	0,29
<i>AT</i>	0,58*	0,28

Tabla III-3: Tabla de heredabilidad en sentido estricto para las generaciones F_3 y F_4 en las variables en estudio. (*transformado por $\log_{10}$). N/C: No calculado.

Caracterización molecular de la población segregante F_4 (ToUNR18xToUNR1)

Caracterización por marcadores de tipo *InDels*

La caracterización molecular se inició con *InDels* en la población F_4 . Estos *InDels* permitieron diferenciar molecularmente, como se observa en la figura III-3, no solo a los genotipos uniformes previamente observados en el Capítulo anterior, sino a las 18 familias F_4 . Se verificó la distribución de las variables y fueron normales, salvo peso que se transformó con Logaritmo Natural. Este marcador resultó significativo en las variables peso, altura y pH para los 3 *InDel* desarrollados, como puede observarse en la Tabla III-4, donde se realizó un análisis unifactorial.

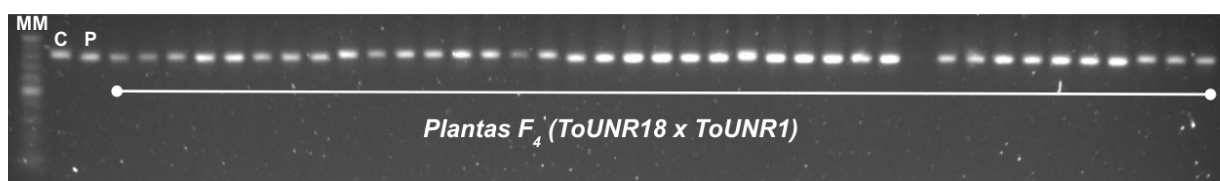


Figura III-3: Ejemplo de gel de agarosa 2,5 % p/v para corrida electroforética de los *InDels* en la generación F_4 .

Variable	F.V.	SC	gl	CM	F	p-valor	R ² Aj
ALTURA	InDel_1	1,25	1,00	1,25	5,80	0,02	0,04
	InDel_2	1,02	1,00	1,02	4,72	0,03	0,03
	InDel_3	1,01	1,00	1,01	4,66	0,03	0,03
pH	InDel_1	0,17	1,00	0,17	4,59	0,04	0,04
	InDel_2	0,17	1,00	0,17	4,63	0,03	0,04
	InDel_3	0,19	1,00	0,19	5,16	0,03	0,04
PESO*	InDel_1	2,83	1,00	2,83	8,07	0,01	0,05
	InDel_2	2,42	1,00	2,42	6,81	0,01	0,04
	InDel_3	2,50	1,00	2,50	7,06	0,01	0,05
*Transformado mediante logaritmo natural							

Tabla III-4: *InDels* asociado a variable de interés agronómico y detección de QTLs.

A partir del análisis bifactorial, teniendo a marcador y familia como fuentes de variación y su interacción entre ellos, se detectó que en todos los casos la interacción entre *InDel* y familia fue no significativa, como puede verse en la tabla III-5. Esto demuestra que el MM tiene efecto y asociación a los caracteres detectados de manera independiente a la familia analizada en esta F_4 .

Variable	F.V.	SC	gl	CM	F	p-valor	R ² Aj
ALTURA	InDel_1	1,25	1,00	1,25	16,37	0,00	0,66
	FAMILIA	18,58	17,00	1,09	14,31	<0,0001	
	InDel_1*FAMILIA	0,43	6,00	0,07	0,94	0,47	
	Error	7,72	101,00	0,08			
	InDel_2	1,02	1,00	1,02	14,18	0,00	0,68
	FAMILIA	19,02	17,00	1,12	15,50	<0,0001	
	InDel_2*FAMILIA	0,66	5,00	0,13	1,84	0,11	
	Error	7,43	103,00	0,07			
	InDel_3	1,01	1,00	1,01	14,09	0,00	0,68
	FAMILIA	19,09	17,00	1,12	15,65	<0,0001	
	InDel_3*FAMILIA	0,57	4,00	0,14	1,98	0,10	
	Error	7,46	104,00	0,07			
pH	InDel_1	0,17	1,00	0,17	5,59	0,02	0,21
	FAMILIA	1,15	17,00	0,07	2,25	0,01	
	InDel_1*FAMILIA	0,12	6,00	0,02	0,65	0,69	
	Error	1,95	65,00	0,03			
	InDel_2	0,17	1,00	0,17	5,46	0,02	0,19
	FAMILIA	1,14	17,00	0,07	2,17	0,01	
	InDel_2*FAMILIA	0,03	5,00	0,01	0,20	0,96	
	Error	2,05	66,00	0,03			
	InDel_3	0,19	1,00	0,19	6,15	0,02	0,20
	FAMILIA	1,13	17,00	0,07	2,17	0,01	
	InDel_3*FAMILIA	0,03	4,00	0,01	0,24	0,91	
	Error	2,05	67,00	0,03			
LN PESO	InDel_1	2,83	1,00	2,83	27,01	<0,0001	0,72
	FAMILIA	32,20	17,00	1,89	18,08	<0,0001	
	InDel_1*FAMILIA	0,74	6,00	0,12	1,18	0,33	
	Error	10,58	101,00	0,10			
	InDel_2	2,43	1,00	2,43	23,11	<0,0001	0,72
	FAMILIA	33,29	17,00	1,96	18,66	<0,0001	
	InDel_2*FAMILIA	0,36	5,00	0,07	0,69	0,63	
	Error	10,81	103,00	0,10			
	InDel_3	2,51	1,00	2,51	24,12	<0,0001	0,72
	FAMILIA	33,27	17,00	1,96	18,82	<0,0001	
	InDel_3*FAMILIA	0,30	4,00	0,08	0,72	0,58	
	Error	10,82	104,00	0,10			

Tabla III-5: InDels asociado a variable de interés. MM e interacción de MM por familia.

Caracterización molecular por marcadores de tipo SNP

La caracterización con *SNP* se realizó en 90 plantas de la generación F_4 más los genotipos uniformes como testigos. Estas 90 plantas están dentro de las 130 plantas que previamente fueron caracterizadas con los *InDels*. Se seleccionaron 30 individuos de alto peso, 30 de bajo peso y 30 de peso intermedio. Las variables agronómicas de estas plantas seleccionadas fueron normales, con valores de W mayores a 0,90, salvo peso (W de 0,81) que fue transformado mediante logaritmo natural.

De las 90 plantas totales, 5 fueron desestimadas por fallas en la secuenciación. La secuencia obtenida en cada planta F_4 fue utilizada para realizar un alineamiento múltiple, inclusive aquellas de los materiales uniformes usados como testigos en este análisis. A partir del alineamiento múltiple se obtuvo un dendrograma (Figura III-4) con las 85 plantas F_4 en conjunto con los materiales de primer ciclo, y se identificaron 3 grupos de acuerdo a cada *SNP*.

- *Grupo 1*

El grupo uno fue considerado para aquellas plantas que presentaban *SNP* como el progenitor cultivado (CAIMANTA) y fue la sección del dendrograma comprendida entre las plantas 18P53_V-2 y 18P50_II-31. Este grupo presentó un total de 30 individuos.

- *Grupo 2*

El grupo dos fue considerado para aquellas plantas que presentaban *SNP* como la F_1 (CxP) y fue la sección del dendrograma comprendida entre las plantas 18P60_VII-46 y 18P51_II-37. Este grupo presentó un total de 39 individuos.

- *Grupo 3*

El grupo tres fue considerado para aquellas plantas que presentaban *SNP* como el progenitor silvestre LA0722 y fue la sección del dendrograma comprendida entre las plantas 18P50_II-30 y la 18P55_III-60. Este grupo presentó un total de 16 individuos.

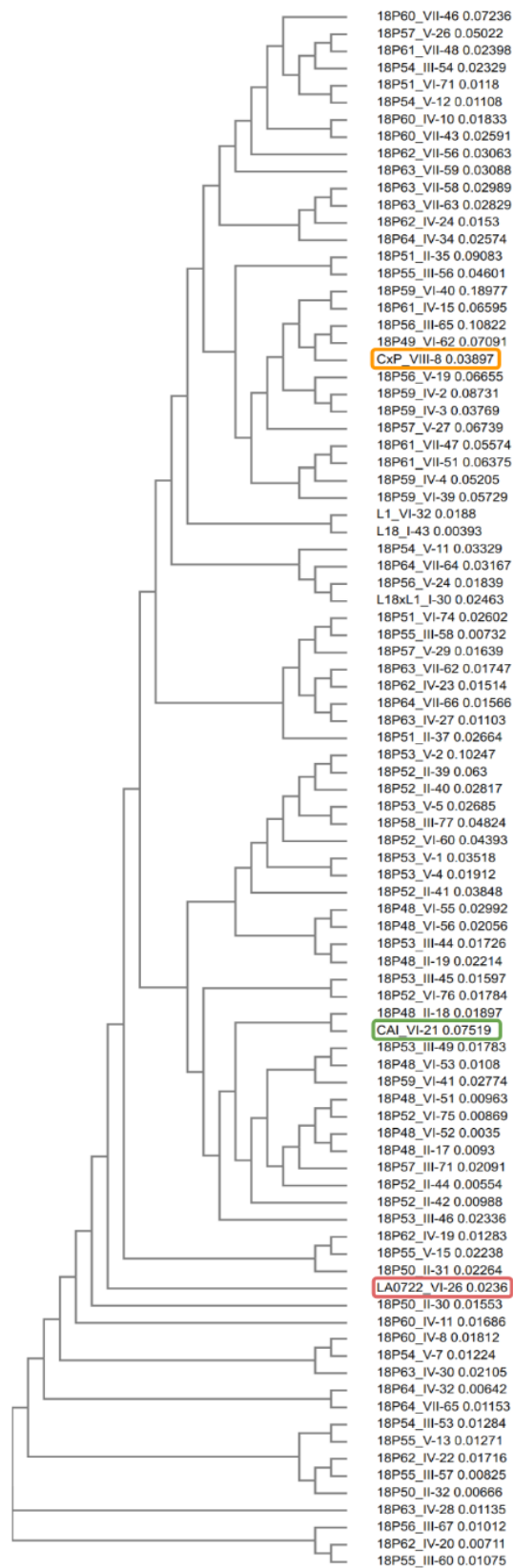


Figura III-4: Dendrograma obtenido a partir de la caracterización molecular con tecnología Sanger.

Esta información se utilizó para caracterizar cada planta con un identificador que define el *SNP* respecto a alguno de los progenitores o la F_1 . Se identificó con 1 si provenía del progenitor cv. Caimanta, 2 como la F_1 y 3 como el progenitor silvestre LA0722. Así se prosiguió con la detección de *QTL* utilizando la información aquí detallada como la de los caracteres fenotípicos.

A partir del ANOVA unifactorial se detectó el efecto del marcador *SNP*. El marcador fue significativo para las variables agronómicas de peso, diámetro, altura, *a*, dureza y pH del fruto. Al tener la información perteneciente al híbrido entre los progenitores se realizó la estimación de la dominancia para estos caracteres de interés agronómicos significativos.

Variable	F.V.	SC	gl	CM	F	p-valor	R ² Aj
PESO*	SNP	12,67	2	6,34	21,56	<0,0001	0,33
DIÁMETRO	SNP	6,51	2	3,26	17,35	<0,0001	0,28
ALTURA	SNP	5,76	2	2,88	15,56	<0,0001	0,26
<i>a</i>	SNP	57,92	2	28,96	3,74	0,030	0,08
DUREZA	SNP	193,33	2	96,66	4,13	0,021	0,09
pH	SNP	0,49	2	0,25	7,4	0,001	0,17
*Transformado mediante logaritmo natural							

Tabla III-6: Variables asociadas al marcador *SNP* en la población F_4 analizada.

Se llevó a cabo también ANOVA bifactorial, en donde se tuvo en cuenta a marcador y familia como fuentes de variación y también la interacción entre ellos. En todos los casos la interacción entre *SNP* y familia fue no significativa, como puede verse en la tabla III-7. Esto demuestra que el MM tiene efecto y asociación a los caracteres detectados de manera independiente a la familia analizada en esta F_4 de la misma manera que tuvo el comportamiento de los marcadores *InDels* (Tabla III-5).

Variable	F.V.	SC	gl	CM	F	p-valor	R² Aj
PESO*	Modelo.	31,67	28,00	1,13	12,42	<0,0001	0,79
	SNP	12,67	2,00	6,34	69,59	<0,0001	
	FAMILIA	18,48	16,00	1,16	12,69	<0,0001	
	SNP*FAMILIA	0,52	10,00	0,05	0,57	0,83	
	Error	5,10	56,00	0,09			
	Total	36,77	84,00				
DIÁMETRO	Modelo.	17,33	28,00	0,62	7,56	<0,0001	0,69
	SNP	6,51	2,00	3,26	39,79	<0,0001	
	FAMILIA	10,26	16,00	0,64	7,84	<0,0001	
	SNP*FAMILIA	0,55	10,00	0,05	0,67	0,75	
	Error	4,58	56,00	0,08			
	Total	21,91	84,00				
ALTURA	Modelo.	17,77	28,00	0,63	11,18	<0,0001	0,77
	SNP	5,76	2,00	2,88	50,77	<0,0001	
	FAMILIA	11,64	16,00	0,73	12,81	<0,0001	
	SNP*FAMILIA	0,37	10,00	0,04	0,65	0,77	
	Error	3,18	56,00	0,06			
	Total	20,94	84,00				
a	Modelo.	380,28	23,00	16,53	4,39	<0,0001	0,55
	SNP	57,92	2,00	28,96	7,70	0,00	
	FAMILIA	304,76	15,00	20,32	5,40	<0,0001	
	SNP*FAMILIA	17,60	6,00	2,93	0,78	0,59	
	Error	150,52	40,00	3,76			
	Total	530,80	63,00				
DUREZA	Modelo.	1100,86	22,00	50,04	4,12	0,00	0,53
	SNP	193,33	2,00	96,66	7,96	0,00	
	FAMILIA	867,28	15,00	57,82	4,76	<0,0001	
	SNP*FAMILIA	40,25	5,00	8,05	0,66	0,65	
	Error	473,69	39,00	12,15			
	Total	1574,55	61,00				
pH	Modelo.	1,49	24,00	0,06	2,37	0,01	0,35
	SNP	0,49	2,00	0,25	9,39	0,00	
	FAMILIA	0,87	15,00	0,06	2,22	0,02	
	SNP*FAMILIA	0,13	7,00	0,02	0,69	0,68	
	Error	1,00	38,00	0,03			
	Total	2,49	62,00				
*Transformado mediante logaritmo natural							

Tabla III-7: Variables asociadas al marcador *SNP* y análisis de la interacción MM por familia.

A partir de los datos observados con la caracterización por *SNP* se calcularon los grados de dominancia de las variables significativas (Tabla III-8), empleando un test de Duncan para ver la comparación de medias y el efecto de cada *SNP*.

Variable	CAIMANTA (1)	LA0722 (3)	F ₁ CxP (2)	Punto Origen	a	-a	d	d/a	Grado Dominancia	Progenitor dominante
PESO	5,26	10,11	11,28	7,69	2,43	-2,43	3,60	1,48	Sobredominancia positiva	LA0722
DIÁMETRO	2,04	2,58	2,64	2,31	0,27	-0,27	0,33	1,22	Sobredominancia positiva	LA0722
ALTURA	1,93	2,47	2,48	2,20	0,27	-0,27	0,28	1,04	Dominancia completa hacia los mayores valores	LA0722
a	26,32	23,49	25,47	24,91	1,42	-1,42	0,56	0,40	Dominancia completa hacia los mayores valores	CAIMANTA
DUREZA	55,12	55,78	58,84	55,45	0,33	-0,33	3,39	10,27	Sobredominancia positiva	LA0722
pH	4,54	4,80	4,66	4,67	0,13	-0,13	-0,01	-0,08	Dominancia parcial hacia los menores valores	CAIMANTA

Tabla III-8: Estimación del grado de la dominancia en variables significativas.

Estos resultados en la tabla reflejan que de las 6 variables agronómicas, 4 de ellas fueron hacia el progenitor silvestre como dominante, en el cual 3 (Peso, diámetro y dureza) mostraron sobredominancia positiva hacia este al superar su valor y en el caso de la variable altura mantuvo dominancia completa hacia el mismo. El resto de las variables, pH y a, mostraron tendencia hacia el progenitor cultivado como dominante. Finalmente, el análisis global determinó que el 77,7% de las variables fue hacia el valor del progenitor silvestre como dominante, mientras que el 33,3% restante hacia el progenitor cultivado.

Comparación entre las detecciones de QTLs con los MM desarrollados

La detección de QTLs con los MM fue satisfactoria. En el caso de los *InDels* se detectó 3 QTLs de interés agronómicos, los cuales son altura, peso y pH, mientras que en el caso de los *SNP* se detectaron QTLs de calidad del fruto, siendo éstos: altura, a, dureza, diámetro, peso y pH. La efectividad de los *SNP* en este caso fue mayor para la detección de QTLs.

Una vez realizados los análisis de ambos marcadores moleculares se estudió el comportamiento entre estos, a partir de las 85 plantas caracterizadas con los *SNP* que pertenecían al grupo de 130 plantas caracterizadas con *InDels*.

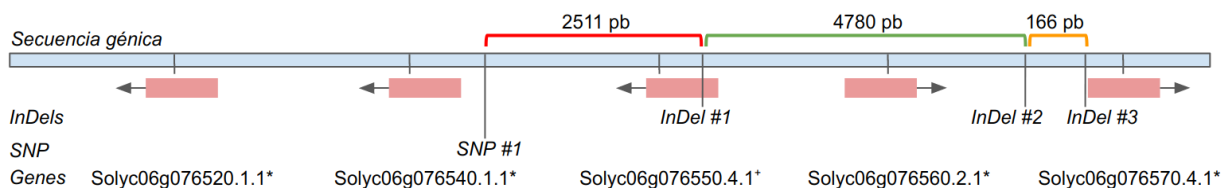


Figura III-5: Región génica del cromosoma 06 con distancia en pb de los marcadores *InDels* y *SNP* utilizados en para caracterizar molecularmente la población segregante.

A partir de la información en la localización de los marcadores moleculares desarrollados y utilizados en la generación F_4 para su caracterización molecular, se analizó el comportamiento de estos marcadores de tipo *InDel* y *SNP*. El comportamiento entre estos demostró un 67,06 % de coincidencia entre ellos (Tabla III-9), lo que indicaría que la región no presenta altos niveles de recombinación. Dentro de esta coincidencia, el material cultivado fue el que evidenció mayor concordancia entre estos marcadores moleculares con casi un 72% (Tabla III-9).

Comportamiento entre <i>InDels</i> y <i>SNP</i>	TOTAL	%
<i>Coincidencia entre InDel y SNP</i>	57 de 85	67,06
<i>Coincidencia de cv. Caimanta</i>	41 de 57	71,93
<i>Coincidencia de acc. LA0722</i>	16 de 57	28,07
<i>No Coincidencia entre InDel y SNP</i>	28 de 85	32,94

Tabla III-9: Cálculo del comportamiento de los MM en la caracterización molecular.

Re-análisis de los marcadores InDels en las plantas caracterizadas por SNP

Por último, se realizó un re-análisis de los MM *InDels* pero solo sobre aquellas 85 plantas caracterizadas por *SNP*, para visualizar posible efecto o sesgo a través del número de plantas utilizadas en dicha caracterización. En este nuevo análisis, los marcadores tipo *InDels* asociaron a un nuevo carácter agronómico, la altura del fruto. Así suman en total 4 las variables asociadas a este tipo de MM. Se obtuvieron nuevos valores de R^2 ajustado, en el que el *InDel* #1 generó los R^2 ajustados máximo (0,27 para peso, Tabla III-10) y mínimo (0,09 para pH, Tabla III-10).

Variable	F.V.	SC	gl	CM	F	p-valor	R² Aj
Peso*	Modelo.	10,41	1,00	10,41	32,80	<0,0001	0,27
	InDel_1	10,41	1,00	10,41	32,80	<0,0001	
	Error	26,34	83,00	0,32			
	Total	36,75	84,00				
Diámetro	Modelo.	4,57	1,00	4,57	21,89	<0,0001	0,20
	InDel_1	4,57	1,00	4,57	21,89	<0,0001	
	Error	17,34	83,00	0,21			
	Total	21,91	84,00				
Altura	Modelo.	4,63	1,00	4,63	23,53	<0,0001	0,21
	InDel_1	4,63	1,00	4,63	23,53	<0,0001	
	Error	16,32	83,00	0,20			
	Total	20,94	84,00				
pH	Modelo.	0,27	1,00	0,27	7,49	0,01	0,09
	InDel_1	0,27	1,00	0,27	7,49	0,01	
	Error	2,21	61,00	0,04			
	Total	2,49	62,00				
*Transformado mediante logaritmo natural							

Tabla III-10: ANOVA del *InDel* #1 en las 85 plantas previamente caracterizadas con SNP.

Variable	F.V.	SC	gl	CM	F	p-valor	R ² Aj
Peso*	Modelo.	9,11	1,00	9,11	27,37	<0,0001	0,24
	InDel_2	9,11	1,00	9,11	27,37	<0,0001	
	Error	27,64	83,00	0,33			
	Total	36,75	84,00				
Diámetro	Modelo.	3,77	1,00	3,77	17,26	0,00	0,16
	InDel_2	3,77	1,00	3,77	17,26	0,00	
	Error	18,14	83,00	0,22			
	Total	21,91	84,00				
Altura	Modelo.	4,04	1,00	4,04	19,84	<0,0001	0,18
	InDel_2	4,04	1,00	4,04	19,84	<0,0001	
	Error	16,90	83,00	0,20			
	Total	20,94	84,00				
pH	Modelo.	0,32	1,00	0,32	9,17	0,00	0,12
	InDel_2	0,32	1,00	0,32	9,17	0,00	
	Error	2,16	61,00	0,04			
	Total	2,49	62,00				
*Transformado mediante logaritmo natural							

Tabla III-11: ANOVA del *InDel* #2 en las 85 plantas previamente caracterizadas con SNP.

Variable	F.V.	SC	gl	CM	F	p-valor	R² Aj
Peso*	Modelo.	9,35	1,00	9,35	28,31	<0,0001	0,25
	InDel_3	9,35	1,00	9,35	28,31	<0,0001	
	Error	27,40	83,00	0,33			
	Total	36,75	84,00				
Diámetro	Modelo.	3,99	1,00	3,99	18,49	<0,0001	0,17
	InDel_3	3,99	1,00	3,99	18,49	<0,0001	
	Error	17,92	83,00	0,22			
	Total	21,91	84,00				
Altura	Modelo.	4,04	1,00	4,04	19,85	<0,0001	0,18
	InDel_3	4,04	1,00	4,04	19,85	<0,0001	
	Error	16,90	83,00	0,20			
	Total	20,94	84,00				
pH	Modelo.	0,36	1,00	0,36	10,41	0,00	0,13
	InDel_3	0,36	1,00	0,36	10,41	0,00	
	Error	2,12	61,00	0,03			
	Total	2,49	62,00				
*Transformado mediante logaritmo natural							

Tabla III-12: ANOVA del *InDel* #3 en las 85 plantas previamente caracterizadas con SNP.

Este nuevo análisis permite la detección de nuevos QTLs para caracteres de calidad de fruto, tanto morfológicos (Peso, Diámetro y Altura) y organolépticos (pH). La coincidencia entre los *InDels* dentro de este nuevo análisis fue del 95% (81 de 85). Estos marcadores pueden utilizarse en el programa de mejoramiento genético de la especie en otras poblaciones con diferentes estructuras genéticas para revalidar su accionar y su asociación a las variables nombradas. Aquellos marcadores que presenten la misma asociación a alguno o todos de estos caracteres podrían usarse en un futuro para realizar una selección asistida por marcadores y disminuir los tiempos que requiere el desarrollo y la obtención de una nueva variedad élite.

DISCUSIÓN

Los datos fenotípicos relevados en este Capítulo sobre las variables agronómicas de interés mostraron un aumento en el h^2 respecto a la generación anterior analizada por Del Medico *et al.* (2019). Esto puede deberse al aumento de la homocigosis dentro de cada familia con el paso de las generaciones, ya que para una generación F_4 ronda el 93,8% de homocigosis al ser una especie de naturaleza autógama, teniendo en cuenta que genera su progenie a través autofecundaciones.

En esta generación F_4 se detectó variación en todas las variables analizadas, presentando algunas un rango de variación que supera a los genotipos uniformes. Sin embargo, el porcentaje de variación fenotípica debido a variación aditiva fue alto para los caracteres morfológicos, no así para los caracteres organolépticos, con valores de medio a bajo. Esto lo demuestran los valores de h^2 estimados e implica que la selección artificial presenta limitaciones para continuar modificando la media de estos últimos caracteres antes mencionados.

La caracterización molecular fue efectiva con los marcadores desarrollados por lo cual permitió identificar y detectar *QTLs* para peso, altura, diámetro, dureza, coeficiente a , y pH. Esta variación molecular fue detectada por polimorfismos identificados en la secuencia génica de la región de interés por presentar expresión diferencial, que fueron validados en los progenitores y luego se verificó en esta población segregante. Esta observación de polimorfismos a partir de genes que presentan expresión génica relacionada a la maduración del fruto y detectados en los progenitores, fue un punto de partida para desarrollar MM capaces de optimizar los recursos económicos y así detectar *QTLs* mediante la asociación entre la variabilidad fenotípica y la molecular, permitiendo así poder asistir al programa de mejoramiento genético de la especie. Además, los *QTLs* detectados en la generación segregante de segundo ciclo entre caracteres cuantitativos con variancia genética significativa y marcadores desarrollados en base a polimorfismos en la expresión génica confirman las correspondencias identificadas por Análisis Procrustes Generalizado en el Capítulo I entre variabilidad en la expresión génica en los niveles transcriptómico y fenotípico en los materiales uniformes de primer ciclo.

Por su parte los MM de tipo *InDels* reflejaron un comportamiento dominante principalmente hacia el material cultivado, corroborando lo observado en el Capítulo II en la validación de los marcadores estructurales desarrollados en base a diferencias funcionales en los transcriptomas. Sin embargo, a pesar de comportarse como dominantes, los *InDels* permitieron detectar asociaciones significativas, pero debe tenerse en cuenta que en el caso

en que el estado genotípico es heterocigota (F_1 y plantas F_4 al estado heterocigota donde no fijaron uno de los alelos de los progenitores) sólo se observa un amplicón no siendo posible distinguir a estos individuos de los homocigotas para el alelo aportado por cv. Caimanta. Así, el efecto de cada *QTL* puede presentar un sesgo en su estimación, además, de no ser posible calcular el grado de dominancia para cada *QTL* detectado. Por otro lado, es destacable que estos marcadores mostraron asociación con variables agronómicas, pero con un R^2 ajustado en niveles bajos. Esto suele suceder ya que estos MM fueron desarrollados sobre la secuencia génica de las sHSPs, y estos genes poseen un bajo efecto sobre el fenotipo. Por su naturaleza las sHSPs actúan como chaperonas de manera conjunta, aportando cada una un efecto pequeño, sumando su acción como poligenes o genes menores y no como genes mayores. Es por eso que se encuentran duplicadas en *tándem* y evolutivamente conservadas, ya que le dan soporte a la planta frente a cambios por diferentes tipos de estrés, tanto biótico como abiótico, con efecto medio a bajo sobre todo el fenotipo. Además, nuestra hipótesis se refuerza con este accionar de las sHSPs en el que no presentan intrones, ya que antes cambios bruscos ambientales se expresan rápidamente al no sobrellevar empalme alternativo (del inglés *splicing alternativo*) y además, podría hipotetizarse que las especies silvestre presentan una mayor cantidad de copias de sHSPs que las especies cultivadas, por tener que enfrentarse a condiciones ambientales más extremas. Al realizar un segundo análisis con los *InDels* en las 85 plantas caracterizadas con *SNP* se observó un aumento en el R^2 ajustado y se detectó un nuevo *QTL*.

En el caso de los *SNP*, la caracterización molecular con estos MM fue de mayor dificultad, pero a la hora de detección de *QTLs* resultó ser más efectivo que los *InDels*. Este marcador al ser codominante presenta la posibilidad de calcular el grado de dominancia, y esto demostró que de las 6 variables agronómicas que resultaron significativas mediante el ANOVA, 4 de ellas fueron hacia el progenitor silvestre como dominante, en el cual 3 (Peso, diámetro y dureza) mostraron sobredominancia positiva hacia este al superar su valor y en el caso de la variable altura mantuvo dominancia completa hacia el mismo. En cuanto al pH y al coeficiente 'a', estos mostraron tendencia hacia el progenitor cultivado como progenitor dominante. El análisis global determinó que el 77,7% de las variables fue hacia el valor del progenitor silvestre como dominante, mientras que el 33,3% restante hacia el progenitor cultivado.

La comparación entre estos MM de *InDel* y *SNP* que detectaron *QTLs* de interés agronómico, fue concordante con 4 caracteres de interés agronómico (peso, diámetro, altura

y pH). Del total de plantas caracterizadas con estos dos MM, se observó que en más del 67,06 % de las mismas hubo coincidencia en comportamiento de los MM. Este alto nivel de coincidencia en los marcadores moleculares de esta región del cromosoma 6 (en donde las altas repeticiones) no es esperable encontrar elevada recombinación génica. Al juntar la información proveniente de ambos MM se esperaba asociación de la variable peso con la variante de cv. Caimanta y lo observado a través de la información de los *SNP* fue lo contrario. Las familias con *InDel* tipo LA0722 presentaron mayor peso.

Para finalizar, al observarse en la caracterización fenotípica un porcentaje de variación fenotípica asociado a variación del tipo aditiva con valores de medio a bajo para los caracteres organolépticos, este resultado implicaría que la selección artificial presenta limitaciones para continuar modificando la media de estos últimos caracteres. Es por eso, que será necesario recurrir a otras estrategias de mejoramiento, tales como nuevas hibridaciones entre las diferentes familias a fin de usufructuar el componente de variación no aditiva o bien con materiales de otro fondo genético a fin de incrementar la variación aditiva. A pesar de esto, se logró integrar la información obtenida en estudios previos de transcriptómica, desarrollo de cebadores para caracterización molecular de una población segregante como su caracterización fenotípica, y así asistir de manera eficiente el mejoramiento vegetal de la especie.

CONCLUSIONES

- Se caracterizó fenotípicamente una población segregante F_4 en donde la heredabilidad en sentido estricto reflejó un fuerte aumento entre esta generación y la anterior, para los caracteres de interés agronómico analizados.
- Los MM de tipo *InDels* y *SNP* desarrollados previamente permitió caracterizar molecularmente la generación F_4 y detectar *QTLs* asociados a caracteres de calidad de frutos.
- Se detectaron *QTLs* robustos para peso, altura, diámetro y pH del fruto de tomate mediante MM tipo *InDels*. En el caso de los *SNP*, se detectaron *QTLs* para altura, a, dureza, diámetro, peso y pH.
- La comparación entre los *QTLs* detectados con *InDels* y *SNP* mostraron una concordancia de más del 50% entre las plantas analizadas, y permitieron detectar 4 *QTLs* de manera coincidente.

CAPÍTULO IV

EXPLORACIÓN DE LA REGIÓN GENÓMICA DE LAS sHSPs DEL CROMOSOMA 06 MEDIANTE TECNOLOGÍA DE SECUENCIACIÓN DE TERCERA GENERACIÓN

INTRODUCCIÓN

Las diferencias entre individuos y especies se originan a partir de cambios en su genoma. Comparaciones de genomas de tomate y otras especies relacionadas respaldan la idea de que durante su especiación hubo eventos de expansión del genoma completo mediante casos de triplicación del mismo, como también la expansión de familias génicas a través de sucesos de duplicación en *tándem* (The Tomato Genome Consortium, 2012; Barker *et al.*, 2008). Estos eventos que generan aumento del número de la ploidía, como el de familias génicas dentro de cada cromosoma, son capaces de sobrevivir a la selección natural o artificial lo que representa un desafío a la hora de secuenciar y ensamblar.

La familia multigénica de sHSPs presenta copias múltiples de sus genes y un grupo de ellos, pertenecientes al cromosoma 6, no son la excepción. Las copias múltiples y también las secuencias repetitivas son un interrogante a la mayoría de las estrategias de ensamblaje. La alta identidad de secuencia entre genes duplicados en *tándem* dentro de una misma especie (genes parálogos) genera que las lecturas muchas veces colapsen en una única copia no siendo posible distinguirlas, principalmente aquellas de lectura corta (Krsticevic *et al.*, 2010). Análisis *in silico* de esta familia de sHSPs determinó que nueve eventos independientes de duplicación modificaron la arquitectura y funcionalidad de sus genes (Krsticevic *et al.*, 2016). Esto involucra el origen de nuevos genes o cambios en la expresión, o el camino inverso, a través de genes sin funcionalidad denominados pseudogenes. Otros estudios han observado que los genes duplicados en *tándem* presentan expresión génica mayor al doble de la expresión que su copia individual, y en algunos casos cinco veces más (Loehlin *et al.*, 2016).

Las nuevas tecnologías de tercera generación ONT (de moléculas largas) permiten abordar esta problemática de manera más certera y con mayor certidumbre. Algunos genomas de tomate, en especial los silvestres, fueron secuenciados con esta tecnología (Schmidt *et al.*, 2017). Un único genoma de referencia no es autosuficiente para identificar la amplia diversidad genética y, en consecuencia, en la actualidad se ha iniciado recientemente un proyecto para la obtención de un pangenoma. Iniciar una secuenciación con dicha tecnología en esta región permitirá dilucidar la organización de las sHSPs del cromosoma 06 que participan en el proceso de la maduración del fruto. Esta región génica fue la zona objetivo de esta tesis, en donde se realizaron estudios tanto transcripcionales como genéticos, y a partir de este capítulo se plantea un camino de ida y vuelta entre estos abordajes para identificar posibles genes candidatos y así optimizar los recursos que suelen utilizarse durante la obtención de una nueva variedad superior.

OBJETIVOS ESPECÍFICOS

1. Caracterización mediante secuenciación de tercera generación de la región genómica en la que se ubican los genes de la subfamilia de sHSPs del cromosoma 6 de tomate.
2. Comparación de los polimorfismos encontrados por esta secuenciación entre los genotipos cv. Caimanta, LA0722 y su F₁ a fin de verificar el comportamiento genético de los marcadores moleculares desarrollados en esta Tesis.

MATERIALES Y MÉTODOS

Material vegetal

Para la secuenciación de tercera generación se extrajo ADN de 40 mg de tejido foliar joven homogeneizado, con el objetivo de obtener ADN de alto peso molecular a fin de asegurar estructuras de gran tamaño. La calidad del ADN de cada genotipo fue verificada con NanoDrop®. La cantidad de ADN utilizado en cada genotipo fue de 30 ng/ul. Al igual como fue mencionado en el Capítulo II, se desarrollaron cebadores para amplificar esta región en particular para secuenciar dicho segmento de ADN.

Extracción de la región de interés y desarrollo de cebadores

Con la finalidad de desarrollar cebadores para amplificar selectivamente dicha región génica, se realizó la extracción de la región genómica en cada progenitor con la herramienta *faidx* perteneciente al toolkit de SAMtools (Li *et al.*, 2009). Mediante la comparación a través del alineamiento con la herramienta *needle* (https://www.ebi.ac.uk/Tools/psa/emboss_needle/) se detectaron regiones conservadas a fin de desarrollar allí los cebadores para amplificar la región. Con el fin de profundizar el análisis, se estudiaron genes que se encuentran en la región, principalmente los genes Solyc06g07510 (*Phosphoserine phosphatase*), Solyc06g07550 (*Glycosyltransferase*) y Solyc06g07590 (*DUF3527 domain-containing protein*). Estos genes son miembros de la región y análisis previos evidenciaron que son genes de gran tamaño en comparación de las sHSPs y de copia única, lo que los hace regiones de mayor conservación en comparación a las sHSPs. Luego, se desarrollaron un conjunto de cebadores con la herramienta online Primal Scheme (<https://primalscheme.com/>). A fin de detectar otras posibles variaciones en la secuencia genómica principalmente del genotipo silvestre (LA0722 de *S. pimpinellifolium*), se comparó su secuencia con la accesión LA2093 del mismo género (<https://www.ncbi.nlm.nih.gov/nuccore/CM026540.1>) con la herramienta *AlignMiner* (Guerrero *et al.*, 2010) con la intención de detectar regiones conservadas entre estas accesiones pertenecientes al mismo género. Luego de esta primera comparación de las secuencias, se generaron cebadores extras para ampliar el espectro y tener alternativas a la hora de amplificar dicha región. Para ese proceso se utilizó además el software online *Primer3* (<http://primer3.ut.ee/>).

Amplificación de la región de interés

La amplificación del ADN genómico se realizó a través de PCR personalizadas de acuerdo a cada cebador (ya que se esperaba productos > 6 kb, fue útil aumentar el tiempo de extensión a 40–50 segundos/kb), por lo que el protocolo de PCR fue ajustado de la siguiente manera:

- Desnaturalización inicial de 2,5 minutos a 98 °C.
- 30 ciclos de amplificación:
 - Desnaturalización del ADN 20 segundos a 98 °C.
 - Hibridación del cebador de 1 minuto a 64 °C.
 - Extensión del amplicón 8 minutos a 72 °C.
- Elongación final de 10 minutos a 72 °C.

Se utilizó la *Taq polimerasa* Q5 (Q5 *High-Fidelity DNA Polymerase* - M0491S) que presenta alta fidelidad y gran potencial para amplificar regiones de gran tamaño. Se hicieron las amplificaciones de cada genotipo. A fin de evitar que los cebadores de los amplicones contiguos se solapen dentro de la reacción de PCR y se amplifiquen de manera masiva, los amplicones fueron enumerados de izquierda a derecha y se dividieron en amplicones pares por un lado y los impares por otro. Esta estrategia se llevó a cabo ya que la región se amplifica mediante un escalonamiento de varios fragmentos a fin de reconstruir la región completa a través de cada uno de estos fragmentos amplificados. Así, para cada genotipo las amplificaciones se llevaron a cabo con un grupo de amplicones pares por otro lado y un grupo de amplicones impares por el otro (Figura IV-1). Una vez realizada las amplificaciones, la separación de los fragmentos amplificados se llevó a cabo con geles de agarosa al 2% p/v realizando corridas electroforéticas de 2 horas a 140 V constantes con buffer TAE en concentración de 1X y se los visualizó mediante tinción con SYBR® Safe (Thermo Fisher Scientific®, Waltham, MA, EEUU). Toda esta sección es desarrollada como una primera ronda de PCR a fin de levantar los amplicones escalonados de esta región en cada genotipo, y en donde los cebadores contienen en su extremo 5' una secuencia consenso que es complementario a la región consenso de las etiquetas (*barcodes*) próximos a utilizarse en una segunda ronda de PCR, a fin de rotular a cada genotipo con una etiqueta.

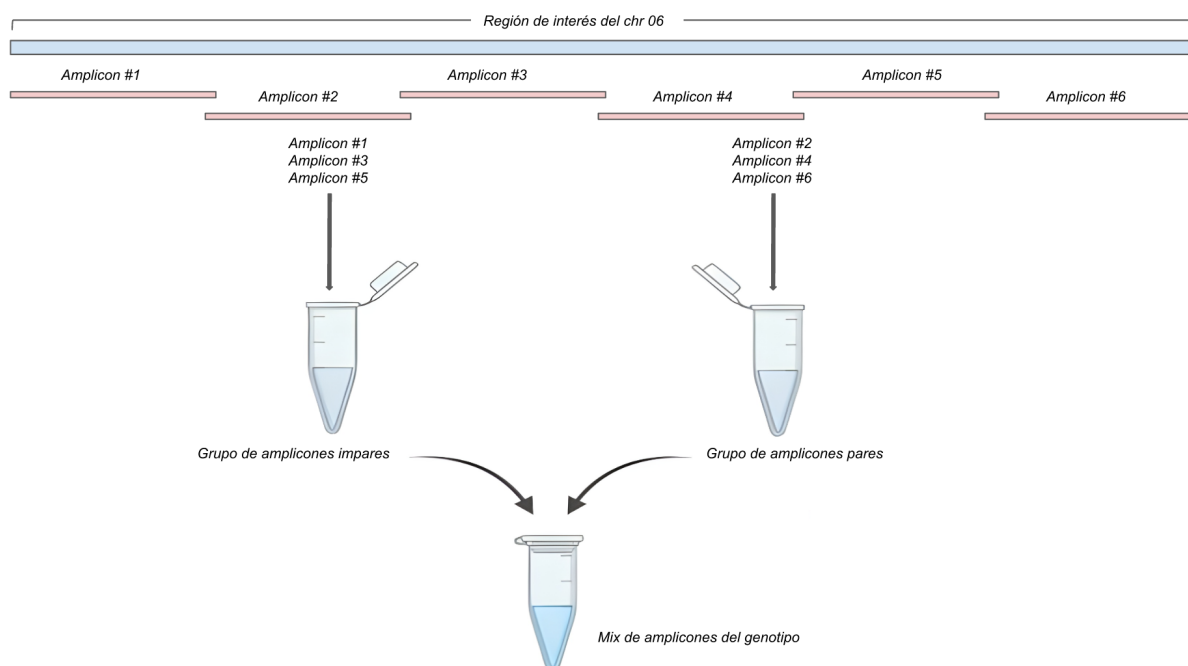


Figura IV-1: Esquema de amplificación mediante PCR de amplicones pares e impares de cada genotipo en estudio.

Etiquetado de los amplicones a través del uso de etiquetas (barcodes)

Se utilizaron 3 oligos como etiquetas (*barcodes*) provenientes de la empresa ArgenTAG® (<https://www.argentag.com/>) a fin de etiquetar los fragmentos genómicos de cada genotipo y luego realizar la secuenciación con la totalidad de las muestras de cada genotipo agrupadas en una muestra única. El genotipo cv. Caimanta se etiquetó con el oligo 'BC3', la accesión LA0722 con el oligo 'BC16', y la F₁ con 'BC26', tal lo descrito por Ezpeleta *et al.* (2022). El uso de etiquetas permite reducir los costos de secuenciación en hasta 20 veces menos, reducir los errores o artefactos provenientes del secuenciador y disminuir significativamente los tiempos requeridos (Ezpeleta *et al.*, 2022). Esta estrategia de marcado de las muestras con una etiqueta suele utilizarse para compartir el potencial de secuenciación que tiene intrínsecamente el instrumento de secuenciación, y es denominado multiplexación. Etiquetar cada muestra permite agrupar en una muestra única todas las muestras previo al ingreso del secuenciador, y una vez llevada a cabo la secuenciación recuperar las lecturas provenientes al total de muestras que se tenía previamente (Figura IV-2). A este proceso se lo conoce como demultiplexación, y es el paso en donde se recuperan las lecturas crudas de cada muestra individual. En la actualidad la mayoría de los secuenciadores poseen gran capacidad de secuenciación. Son instrumentos de alto rendimiento, pero en general este poder de secuenciación no puede utilizarse de manera

parcial, sino de manera total. Utilizar estos oligos para etiquetar varias muestras individuales para luego agruparlas en una muestra única y utilizar el poder de rendimiento del secuenciador de manera total es una de las principales ventajas de esta estrategia novedosa. A partir de esto, se hicieron limpiezas y filtrados mediante el uso de perlas magnéticas y se llevó a cabo una segunda ronda de PCR en donde se etiquetó a cada genotipo con una etiqueta distinta. Estas etiquetas poseen la secuencia consenso complementaria (5'-GCAAGCGGTACACTCAGATC-3') a la de los cebadores utilizados en la primera ronda de amplificación. Así, se obtuvieron los amplicones de esta región de interés en cada genotipo y se los etiquetó, para que ingresen al secuenciador en una única muestra. Previo al agrupamiento de todas las muestras en una única, se cuantifican estas muestras individuales y se las agrupó colocando la misma cantidad equimolar de cada genotipo. Se ligaron los adaptadores utilizando el kit NEBNext (New England BioLabs, EE. UU.) siguiendo el protocolo estándar. A continuación se realizó la purificación con 0,6 veces el volumen de perlas magnéticas NucleoMag y la preparación para la secuenciación de acuerdo con el protocolo estándar de Oxford Nanopore Technologies (SQK-LSK109). La cuantificación de 1 µL de la biblioteca en instrumento Qubit® mostró una concentración de ADN suficiente para ingresar al secuenciador MinION.

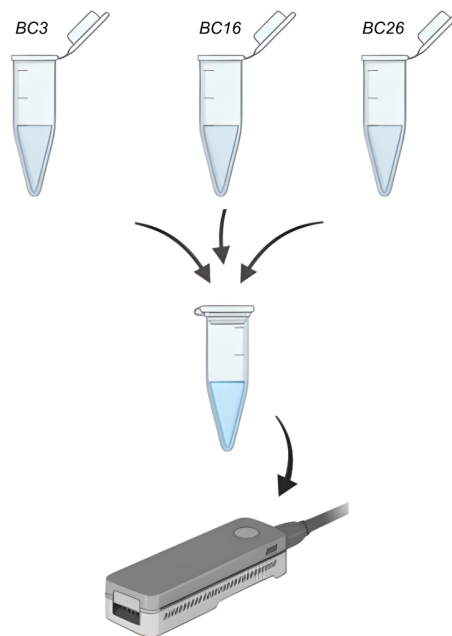


Figura IV-2: Esquema de etiquetado de genotipos en estudio para unificar en un tubo previo al ingreso del secuenciador.

Secuenciación de la región de interés por tecnología de 3^{ra} generación

Se utilizó tecnología de tercera generación para llevar a cabo la secuenciación. El instrumento con el que se generó dicho proceso fue el MinION de Oxford Nanopore®. Usualmente, a pesar de que el genoma de tomate es considerado medianamente pequeño, suelen utilizarse varias flow-cell para llevar a cabo una secuenciación total de esta especie. En este caso, como el objetivo de esta sección es secuenciar una región de interés (*target*) se utilizó una flow-cell completa con 1182 nanoporos detectados, ya que la capacidad del instrumento fue suficiente para visitar la misma con gran profundidad. Se hicieron dos corridas logrando un total de 12,5 horas de secuenciación, con el objetivo de obtener cantidad suficiente de lecturas para luego utilizar en el ensamblado. Se filtraron las lecturas por longitud (500 pb) y calidad (15) con el programa *fastp* (fast all-in-one preprocessing for FASTQ files, Chen *et al.*, 2018) en la plataforma online Galaxy (<https://usegalaxy.eu/>) y luego se ejecutó el programa *Canu Assembler* (Koren *et al.*, 2017) con los siguientes parámetros para el ensamblaje de dicha región en cada genotipo:

- Technology --nanopore
- Processing --corrected
- To restrict canu to only a specific stage --all
- Estimated genome size --20k
- Minimum read length --500
- Minimum overlap --10
- Target coverage for corrected reads --40
- Stop the assembly if read coverage is too low to be useful --10

Identificación de los MM de tipo InDels usados en la caracterización molecular de la F₄

Se analizó la presencia de los MM *InDels* en la región secuenciada con tecnología de tercera generación. Esta búsqueda fue llevada a cabo con la utilización de la herramienta BLAST (Basic Local Alignment Search Tool) (Altschul *et al.*, 1997) para establecer las coordenadas de cada marcador molecular y una búsqueda manual para corroborar los resultados obtenidos.

RESULTADOS

Mediante las comparaciones nombradas entre los diferentes genotipos silvestres, se pusieron en evidencia las regiones conservadas entre los mismos, con el fin de desarrollar y diseñar los cebadores sobre las mismas. En la comparación entre la accesión LA0722 y LA2093 de *Solanum pimpinellifolium* se observó que presenta variaciones entre sus secuencias en entre la región del cromosoma 06 donde mapean las 4 sHSPs de interés.

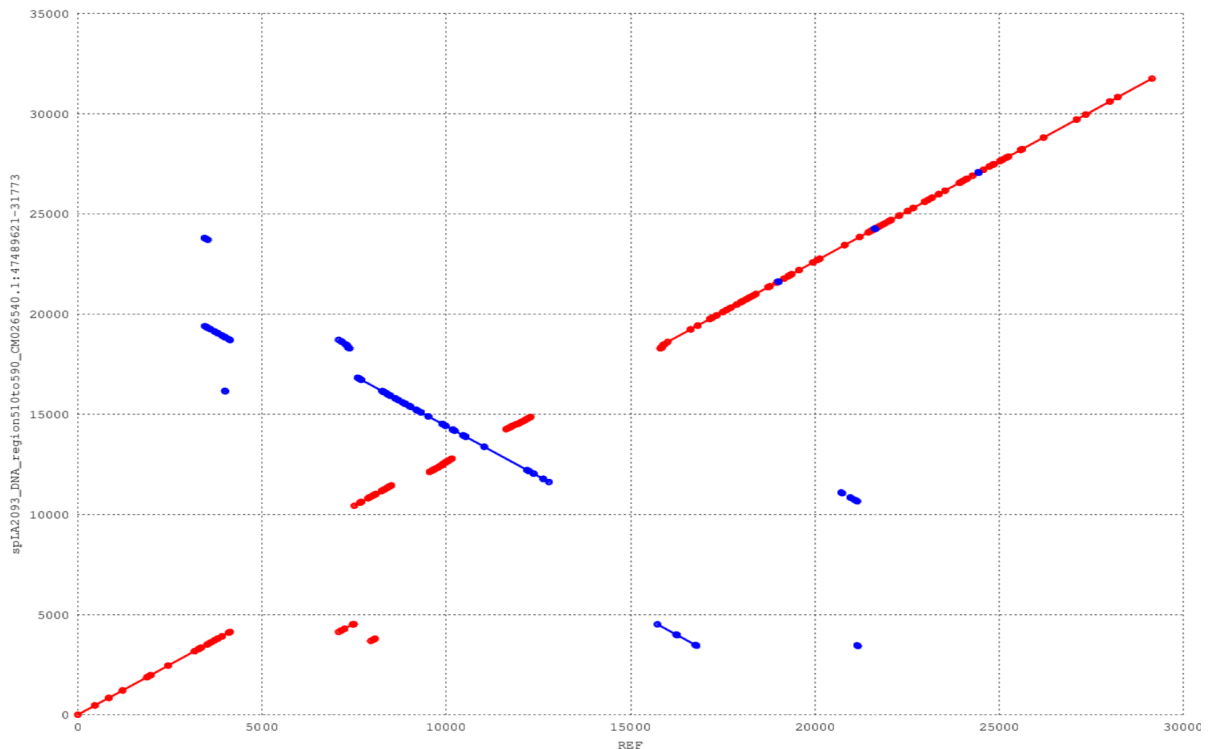


Figura IV-3: Salida de la herramienta AlignMiner en la comparación entre la región del cromosoma 06 de LA0722 (REF, 7:47550195-47579358) frente a accesión de *pimpinellifolium* LA2093 (CM026540.1:47489621-47521394).

El gráfico muestra grandes regiones conservadas entre los genotipos silvestres, como también regiones particulares y únicas de cada uno de ellos. Esto pone de manifiesto de manera global como es esta región en cada uno (Figura IV-3). Pueden observarse rearrreglos en dicha región como también inversiones entre ellas. Hay que tener en cuenta que el genotipo LA0722 fue secuenciado con tecnología de lectura corta y ensamblado frente a la referencia mundial mientras que LA2093 fue secuenciado con tecnología de lectura larga y ensamblado *de novo*. Esta comparación fue especialmente hecha para poder evidenciar la posible disposición de los genes que se encuentran en dicha región para la especie

pimpinellifolium. Así se desarrollaron cebadores para la accesión silvestre (Tabla IV-1 y Tabla IV-2).

Cebadores de cv. CAIMANTA - amplicón de 4 kb			
Nombre	Secuencia	Size	Tm
SL_Heinz_Cai_chr06_510to590_4kb_1_LEFT	5'-GCAAGCGGTACTACTCAGATCATCAAGACGGCAAAGCTCGG-3'	40	60
SL_Heinz_Cai_chr06_510to590_4kb_1_RIGHT	5'-GCAAGCGGTACTACTCAGATCGCAACTTGAATGTTGGAGGAAAATC-3'	45	57
SL_Heinz_Cai_chr06_510to590_4kb_2_LEFT	5'-GCAAGCGGTACTACTCAGATCCCCGTGTGGATGGGCTTAAT-3'	40	59
SL_Heinz_Cai_chr06_510to590_4kb_2_RIGHT	5'-GCAAGCGGTACTACTCAGATCGCCGGTGCTTCATCGTTACA-3'	40	59
SL_Heinz_Cai_chr06_510to590_4kb_3_LEFT	5'-GCAAGCGGTACTACTCAGATCTGGCGCTTTCTCCACTTC-3'	40	59
SL_Heinz_Cai_chr06_510to590_4kb_3_RIGHT	5'-GCAAGCGGTACTACTCAGATCTCGGTCCACTCATAAGGGTGT-3'	41	59
SL_Heinz_Cai_chr06_510to590_4kb_4_LEFT	5'-GCAAGCGGTACTACTCAGATCGCGAGGCTAATCAAACATTGGG-3'	43	59
SL_Heinz_Cai_chr06_510to590_4kb_4_RIGHT	5'-GCAAGCGGTACTACTCAGATCAAGCCCTGGAAGATCAGCCT-3'	40	59
SL_Heinz_Cai_chr06_510to590_4kb_5_LEFT	5'-GCAAGCGGTACTACTCAGATCGCAGATTGTGGGCCGTTGTA-3'	40	59
SL_Heinz_Cai_chr06_510to590_4kb_5_RIGHT	5'-GCAAGCGGTACTACTCAGATCCAGGCCACGTACAGATCAT-3'	40	59
SL_Heinz_Cai_chr06_510to590_4kb_6_LEFT	5'-GCAAGCGGTACTACTCAGATCGGACATGCAACCTCCCCAAT-3'	40	59
SL_Heinz_Cai_chr06_510to590_4kb_6_RIGHT	5'-GCAAGCGGTACTACTCAGATCAGCACTAGCATGTCCAGTTGAA-3'	42	58
SL_Heinz_Cai_chr06_510to590_4kb_7_LEFT	5'-GCAAGCGGTACTACTCAGATCGGTAGATAACCACCGGACCG-3'	40	60
SL_Heinz_Cai_chr06_510to590_4kb_7_RIGHT	5'-GCAAGCGGTACTACTCAGATCTCTTGCCAACGACGTTAGGC-3'	40	59
Cebadores de cv. CAIMANTA - amplicón de 6 kb			
Nombre	Secuencia	Size	Tm
SL_Heinz_Cai_chr06_510to590_6kb_1_LEFT	5'-GCAAGCGGTACTACTCAGATCAACCCGAGTGAGCTGAACCT-3'	40	59
SL_Heinz_Cai_chr06_510to590_6kb_1_RIGHT	5'-GCAAGCGGTACTACTCAGATCTCCCTTTTCATGTGTGGATGATAGA-3'	45	57
SL_Heinz_Cai_chr06_510to590_6kb_2_LEFT	5'-GCAAGCGGTACTACTCAGATCACGAAGATCGGATAGTTTTCGAACA-3'	45	57
SL_Heinz_Cai_chr06_510to590_6kb_2_RIGHT	5'-GCAAGCGGTACTACTCAGATCTGGGACATTTTGCATCGGA-3'	40	59
SL_Heinz_Cai_chr06_510to590_6kb_3_LEFT	5'-GCAAGCGGTACTACTCAGATCGTTTAGGACACCATGCGGACA-3'	41	59
SL_Heinz_Cai_chr06_510to590_6kb_3_RIGHT	5'-GCAAGCGGTACTACTCAGATCAAGCCCTGGAAGATCAGCCT-3'	40	59
SL_Heinz_Cai_chr06_510to590_6kb_4_LEFT	5'-GCAAGCGGTACTACTCAGATCGCAGATTGTGGGCCGTTGTA-3'	40	59
SL_Heinz_Cai_chr06_510to590_6kb_4_RIGHT	5'-GCAAGCGGTACTACTCAGATCTCTGCACCTGGTCAACGACT-3'	40	59
SL_Heinz_Cai_chr06_510to590_6kb_5_LEFT	5'-GCAAGCGGTACTACTCAGATCACACAAGCAGAAGCAGCGAC-3'	40	59
SL_Heinz_Cai_chr06_510to590_6kb_5_RIGHT	5'-GCAAGCGGTACTACTCAGATCCCTGCAGGATGTTTGACCCC-3'	40	59
SL_Heinz_Cai_chr06_510to590_6kb_6_LEFT	5'-GCAAGCGGTACTACTCAGATCAAAGCACCACGAAAGAGCGT-3'	40	58
SL_Heinz_Cai_chr06_510to590_6kb_6_RIGHT	5'-GCAAGCGGTACTACTCAGATCCTGGGCTCCCTCTGGATTCA-3'	40	60
Cebadores de LA0722 - amplicón de 6 kb			
Nombre	Secuencia	Size	Tm
SP_LA0722_2093_chr06_510to590_1_LEFT	5'-GCAAGCGGTACTACTCAGATCTCTGTTTCATGCTGTATCAAGACG-3'	43	58
SP_LA0722_2093_chr06_510to590_1_RIGHT	5'-GCAAGCGGTACTACTCAGATCTTGCAAGAAATCATGAGTGTGA-3'	43	57
SP_LA0722_2093_chr06_510to590_2_LEFT	5'-GCAAGCGGTACTACTCAGATCGCTCAATCAACCAGCTTTGTTC-3'	43	58
SP_LA0722_2093_chr06_510to590_2_RIGHT	5'-GCAAGCGGTACTACTCAGATCGTCTCTGATCCCAAGAATCTTCG-3'	43	59
SP_LA0722_2093_chr06_510to590_3_LEFT	5'-GCAAGCGGTACTACTCAGATCCTTCGATCTCCACTTTCACTTCC-3'	43	59
SP_LA0722_2093_chr06_510to590_3_RIGHT	5'-GCAAGCGGTACTACTCAGATCTTTTGCTTGCTTGCTTTCGATTT-3'	43	56
SP_LA0722_2093_chr06_510to590_4_LEFT	5'-GCAAGCGGTACTACTCAGATCATAACTACCGCCTCTCTTAAT-3'	43	58
SP_LA0722_2093_chr06_510to590_4_RIGHT	5'-GCAAGCGGTACTACTCAGATCATGATCGCCAATGACACTAATC-3'	43	58
SP_LA0722_2093_chr06_510to590_5_LEFT	5'-GCAAGCGGTACTACTCAGATCGTTGACCAGGTGCAGATATCTTC-3'	43	59
SP_LA0722_2093_chr06_510to590_5_RIGHT	5'-GCAAGCGGTACTACTCAGATCATTTAGCAGAAAACCTCCACCTGG-3'	43	58

Tabla IV-1: Cebadores desarrollados sobre la región del cromosoma 06 de cv. Caimanta y LA0722.

El total de cebadores para el genotipo silvestre fue de 5 pares, con un tamaño de amplicón variable para cada sector de la región, pero que supera los 4 kb (Tabla IV-2).

Por otra parte se desarrollaron cebadores para el cultivar Caimanta, pero para llevar a cabo esto se realizó la comparación, mediante un alineamiento múltiple, de la secuencia entre este genotipo y la referencia mundial (cv. Heinz 1706), ya que pertenecen a la misma especie *S. lycopersicum*. Así, se desarrollaron cebadores que amplifican la región de interés en el material cultivado tal como se muestra en la Tabla IV-1. El total de cebadores para este genotipo fue 7 pares, con un tamaño de amplicón de 4 kb, y de 6 pares de cebadores con un producto de amplificación de 6 kb.

Además, se diseñaron y sintetizaron cebadores alternativos (Tabla IV-2) para la accesión silvestre LA0722, para disponer en caso de que los cebadores de la Tabla IV-1 para esta ocasión presenten fallas o artefactos durante su uso. En este caso el total de cebadores fue de 9 pares con productos de amplificaciones que varían en su tamaño.

Cebadores de la acc. LA0722 - alternativos			
Nombre	Secuencia	Size	Tm
SP_InSerLA0722_chr06_510to590_1_LEFT	5'-GCAAGCGGTACACTCAGATCGATGGTTGATTGAAGAGGAGGTG-3'	43	59
SP_InSerLA0722_chr06_510to590_1_RIGTH	5'-GCAAGCGGTACACTCAGATCTGTCTTATTGTCGTTGGCAAGAA-3'	43	57
SP_InSerLA0722_Intermedio_chr06_510to590_1_LEFT	5'-GCAAGCGGTACACTCAGATCGATAGGCCCATCACACAATTGAA-3'	43	58
SP_InSerLA0722_Intermedio_chr06_510to590_1_RIGTH	5'-GCAAGCGGTACACTCAGATCAGGCAAGCATCCTTAATTTTCTCA-3'	44	57
SP_InSerLA0722_Intermedio_chr06_510to590_2_LEFT	5'-GCAAGCGGTACACTCAGATCTGAAATTGTTGCTAATGAGTCTCT-3'	45	57
SP_InSerLA0722_Intermedio_chr06_510to590_2_RIGTH	5'-GCAAGCGGTACACTCAGATCTTGGCAAGAAATCATGAGTGTGA-3'	43	57
SP_InSerLA2093_chr06_510to590_1_LEFT	5'-GCAAGCGGTACACTCAGATCACTCATGATTTCTTGCCAACGAT-3'	43	57
SP_InSerLA2093_chr06_510to590_1_RIGTH	5'-GCAAGCGGTACACTCAGATCATCTCCTCTTCCATCAACCATCA-3'	43	58
SP_InSerLA2093_Intermedio_chr06_510to590_1_LEFT	5'-GCAAGCGGTACACTCAGATCCGAAATTGGGCTTGTGTTTAGTCG-3'	43	58
SP_InSerLA2093_Intermedio_chr06_510to590_1_RIGTH	5'-GCAAGCGGTACACTCAGATCAGAACACACTCTCCCAATATCT-3'	43	58
SP_InSerLA2093_Intermedio_chr06_510to590_2_LEFT	5'-GCAAGCGGTACACTCAGATCAGTTCTTCTTTCCCTCTGTTTTGA-3'	45	57
SP_InSerLA2093_Intermedio_chr06_510to590_2_RIGTH	5'-GCAAGCGGTACACTCAGATCAAATAACTACGATCTCCTCTTCCA-3'	45	57
SP_InSerLA0722-2093_mix_chr06_510to590_1_LEFT	5'-GCAAGCGGTACACTCAGATCTGGAACAGACTGAAAAGGGTTAG-3'	43	58
SP_InSerLA0722-2093_mix_chr06_510to590_1_RIGTH	5'-GCAAGCGGTACACTCAGATCTTTTGCTTGCTTTCGTTTCGATTT-3'	43	56
SP_InSerLA0722-2093_mix_Intermedio_chr06_510to590_1_LEFT	5'-GCAAGCGGTACACTCAGATCTGAACATCAGAAACAGAGACAAC-3'	44	57
SP_InSerLA0722-2093_mix_Intermedio_chr06_510to590_1_RIGTH	5'-GCAAGCGGTACACTCAGATCGCTTGCTTGCTTTTCGATTTGA-3'	42	57
SP_InSerLA0722-2093_mix_Intermedio_chr06_510to590_2_LEFT	5'-GCAAGCGGTACACTCAGATCCGAGGCTAATCAAACATTGGGT-3'	43	58
SP_InSerLA0722-2093_mix_Intermedio_chr06_510to590_2_RIGTH	5'-GCAAGCGGTACACTCAGATCAGTCTATTCTGTGTGTTGGCAAAT-3'	43	57

Tabla IV-2: cebadores alternativos desarrollados para amplificación de la región en la acc. LA0722.

Se amplificó la región de interés del cromosoma 6 con los diferentes cebadores que amplifican esta región de manera escalonada. La tecnología de tercera generación a utilizar fue la tecnología MinION de Oxford Nanopore®. Esta tecnología es capaz de reducir fuertemente las dificultades del ensamblaje en regiones altamente repetitivas y con genes duplicados en tándem, situación que es dificultosa de resolver mediante tecnologías de

secuenciación de lectura corta. Se probaron todos los cebadores desarrollados en cada uno de los genotipos.

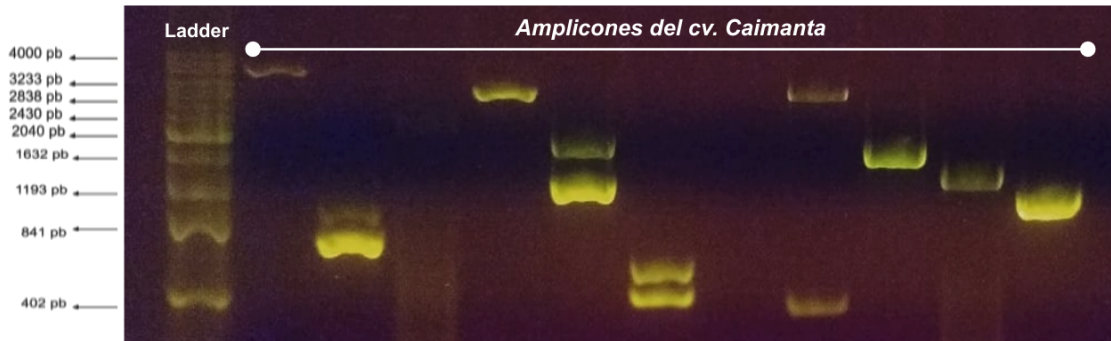


Figura IV-4: Gel con amplificación de región del cromosoma 06 en el cv. Caimanta (Ladder de 400 pb).

Los geles realizados revelaron la amplificación de fragmentos con distinto tamaño, que varían entre los 400 pb aproximadamente y hasta 4100 pb de bases en los casos más particulares (Figura IV-4). Una vez amplificados con los diferentes cebadores desarrollados, estos fueron cuantificados para tener una equimolaridad igual mediante diluciones. Se obtuvieron secuencias entre las 400 bases y los 5,5 kb, inclusive algunas que superaron los 6kb. La línea del secuenciador funcionó de manera correcta para esta región y arrojó gran cantidad de lecturas en total. En la actualidad, según detalla el fabricante, esta tecnología es capaz de generar de 10 a 20 Gb de lecturas largas, multiplexando pequeños genomas o secuenciando regiones de interés (*target*) en genomas de mayor tamaño. En este caso, la cantidad de lecturas totales obtenidas directamente del secuenciador fue de 3.09 millones, según mostraba la interfaz en tiempo real donde se conectó el instrumento. Posteriormente se realizó el uso del algoritmo de llamada a cada base, que transforma la señal obtenida en el secuenciador a la base nucleotídica propiamente dicha, y se demultiplexaron las muestras. Se obtuvieron así datos que comprenden 2.529.592 lecturas de calidad, según la salida del algoritmo de Ezpeleta *et al.* (2022).

Lecturas recuperadas 1^{era} corrida

- True positive barcodes: 3 in 1.670.686 reads. C_BC3: 563.666 reads; P_BC16: 331.238 reads; CxP_BC26: 775.782 reads
- False positive barcodes: 1692 in 13037 reads.
- Observed error rate: 0.77429 %
- Adjusted error rate: 0.00056 %

Lecturas recuperadas 2^{da} corrida

- True positive barcodes: 3 in 858.906 reads. C_BC3: 283.116 reads; P_BC16: 166.400 reads; CxP_BC26: 409.390 reads
- False positive barcodes: 1263 in 6784 reads.
- Observed error rate: 0.78365 %
- Adjusted error rate: 0.00057 %

Luego, se descartaron las lecturas de menos de 500 pb y calidad menor a 15. Así, el cv. Caimanta obtuvo un total de 208.764 lecturas, para la accesión silvestre LA0722 fue de 123.394 lecturas y para la F₁ CxP de 292.250 lecturas. Finalmente, posterior a estos pasos de filtrado, el ensamblado de cada uno de los genotipos en esta primera exploración arrojó los siguientes resultados:

- *Caimanta* - 2 contigs. Total: 12011 bases
- *LA0722* - 3 contigs. Total: 10006 bases
- *F₁ CxP* - 4 contigs. Total: 13945 bases

Además de estos *contigs* el software Canu Assembler dejó secuencias sin poder montar en el ensamblado, que son secuencias potenciales para rearmar la región de interés. En el caso del cv. Caimanta existen 261 secuencias con un promedio de 1596 pares de bases; para el genotipo silvestre fueron de 37 secuencias con un promedio de 1939 pares de bases, y por último, para la F₁ fue de 743 secuencias 1173 pares de bases en promedio.

Una vez realizado este avance, se chequearon los *InDels* desarrollados en el Capítulo III de este trabajo y utilizados para la caracterización molecular de la generación F₄. El fin de este análisis es intentar explicar el comportamiento dominante que fue observado en el Capítulo anterior durante la caracterización molecular de la población segregante. La Tabla IV-3 muestra los resultados obtenidos hasta el momento.

<i>InDel</i>	cv. CAIMANTA	LA0722	F ₁ CxP
- InDel #1 - Izquierdo: 5'-GATGTTCTGATGTCTGATGTAGTCT-3'	PRESENTE	AUSENTE	PRESENTE
- InDel #1 - Derecho: 5'-TGATTCTATTATGTATAAAGTAGCAGAGG-3'	PRESENTE	AUSENTE	PRESENTE
- InDel #2 - Izquierdo: 5'-TTAAGCCAGCAAATTAACAACACTGT-3'	AUSENTE	AUSENTE	AUSENTE
- InDel #2 - Derecho: 5'-AAATAAGGTTTTTCATGTTTAGCCCT-3'	AUSENTE	AUSENTE	AUSENTE
- InDel #3 - Izquierdo: 5'-GGATGATGGTGTGTAAGAATGATGA-3'	AUSENTE	AUSENTE	AUSENTE
- InDel #3 - Derecho: 5'-GTAAGCCTGCGAAGTTGAAATAATG-3'	AUSENTE	AUSENTE	AUSENTE

Tabla IV-3: Chequeo de *InDels* desarrollados previamente en base a las secuencias nucleotídicas.

La presencia o ausencia de los marcadores del tipo *InDel* utilizados en el Capítulo anterior, puede deberse a diferentes motivos (Tabla IV-3). Refinando los análisis, se detectó que el *contig* N° 1 del cv. Caimanta y el *contig* N° 2 tanto del genotipo silvestre como de la F₁

mapean en la misma región. Este fragmento pertenece al inicio de la región de interés abarcando desde el inicio del gen Solyc06g07510 hasta el final del gen Solyc06g076540. Al realizar un alineamiento múltiple entre estas secuencias, se identificó que la región presenta la misma arquitectura que el genoma de referencia, aunque se detectó la presencia de otros polimorfismos que no habían sido detectados en el desarrollo de los marcadores moleculares descritos en el Capítulo II. Uno de estos es una inserción o delección de 39 bases en la región promotora del gen de sHSPs Solyc06g076520. Esto podría utilizarse en un futuro cercano para desarrollar más marcadores a fin de saturar con más de estos en toda la extensión de esta región de interés. El resto de los contigs de cada genotipo mapean en la sección final de esta región, comprendida entre el gen de sHSPs Solyc06g076570 y el gen Solyc06g076590, quedan la región central sin examinar, salvo para el genotipo CxP en donde el *contig* N°1 abarca toda la región del gen Solyc06g076550 hasta el inicio del gen de sHSPs Solyc06g076560.

Al detectar esto, puede revelarse que la presencia de estos marcadores moleculares depende de varios motivos. En primer lugar los cebadores utilizados para la caracterización molecular de la F₄ fueron desarrollados y diseñados frente a la comparación de los genotipos progenitores cuyo genoma fue alineado frente a la referencia mundial. Podría ocurrir que las secuencias en donde se desarrollaron los mismos no coincidan con lo que realmente está presente en dicha región y podría explicar el comportamiento no codominante de los marcadores. Lo mismo ocurre con los cebadores desarrollados en este Capítulo en donde se hicieron comparaciones, en especial en el genotipo silvestre, en donde la secuencia utilizada para desarrollar los cebadores podrá presentar modificaciones con lo que realmente está en esa región. Desarrollar nuevos cebadores a partir de los contigs obtenidos en esta sección es un potencial evento que podría obtener nuevas secuencias genómicas que una y cierre aquellas secciones de esta región que no pudieron ser amplificadas o inspeccionadas por el secuenciador.

En esta primera exploración de esta región de interés que presenta duplicaciones en tándem y secuencias repetidas, como también variación en el número de copias de los genes que la componen, se logró poner a punto una novedosa metodología para visitar y amplificar regiones de interés o *target* que están involucradas a procesos biológicos de interés o presenten asociación con rasgos de interés agronómico. Esto puede ayudar a dilucidar los muchos cambios fenotípicos cuyo motivo aún se desconoce.

DISCUSIÓN

La disponibilidad de solo un genoma de referencia para la mayoría de los cultivos limita el estudio de la expresión génica en especies silvestres comúnmente empleadas como recursos fitogenéticos. El programa de mejoramiento genético de tomate presenta dos parentales: uno perteneciente a *Solanum lycopersicum* (cv. Caimanta) de la misma especie que el genoma de referencia mundial (cv. Heinz 1706) y el otro a *Solanum pimpinellifolium* (accesión LA0722). Este último, no posee un genoma de referencia ya que es un genotipo silvestre. En consecuencia, en la actualidad se están realizando diferentes proyectos de secuenciación de genomas de otras especies de tomate (Bolger *et al.*, 2014^b; Razali *et al.*, 2018), con el fin de cubrir un rango más amplio de variación biológica, y en muchos casos con tecnología de tercera generación (Alonge *et al.*, 2020; Takei *et al.*, 2021; Wang *et al.*, 2020). Uno de estos es el caso pertenece a *Solanum pennellii*, en donde se secuenció dicha especie con tecnología de segunda y tercera generación, a fin de disponer de un genoma anotado completamente (Bolger *et al.*, 2014^b; Schmidt *et al.*, 2017). En algunos casos solo una parte de esta valiosa información está disponible públicamente hacia la comunidad científica.

En el caso del tomate y en especial en donde se sitúan las sHSPs del cromosoma 06, esta región repetitiva y con genes duplicados en *tándem* plantea un desafío computacional tanto biológico como bioinformático. En el Capítulo anterior un comportamiento de los marcadores moleculares no ideal o esperado, y por lo que uno de los problemas que puede ocurrir es que esos MM están desarrollados en lugares en donde los métodos de alineación convencionales basados en un genoma de referencia son insuficientes para identificar con precisión el número real de copias de un gen o colapsar regiones repetitivas que realmente existe pero que el software informático de lectura corta no puede resolver. Es por eso que en este último Capítulo se planteó la secuenciación de dicha región con tecnología de tercera generación, a fin de realizar una primera exploración de la región de interés y así obtener una mejor aproximación para detectar variantes genéticas que se perdieron.

El híbrido entre estos dos progenitores fue el que obtuvo el mayor número de *contigs* con el largo mayor de longitud de la secuencia génica. El progenitor cultivado y el silvestre presentaron menor cantidad de *contigs*. Los marcadores tipo *InDels* no estuvieron en su totalidad representados en la secuencia génica obtenida en estos genotipos. Solo el *InDel* #1 estuvo presente el material cultivado como en la F₁. En el caso de la accesión silvestre, ninguno de los 3 *InDels* fue identificado en esta nueva secuenciación en la región genómica de interés (Tabla IV-3).

La amplificación de esta región es un desafío al intentar amplificarla de principio a fin en cada genotipo, principalmente por la posible estructura genética que tenga cada uno de ellos en dicha región de interés. Todos los cebadores desarrollados fueron diseñados frente a los genomas que se encuentran disponibles en la actualidad, llevando a cabo diferentes análisis y alineamiento a fin de detectar las regiones más conservadas entre ellos y diseñar cebadores que funcionen en todos los genotipos. Como esto no fue posible al detectar pocas regiones conservadas, se realizaron cebadores para el genotipo cultivado por un lado y para el genotipo silvestre por otro. Incluso para este último se desarrollaron cebadores alternativos complementarios, ya que la comparación con otra accesión de su misma especie (LA2093) reveló que la sección central de esta región presenta rearrreglos en la estructura o arquitectura de sus secuencias génicas (Figura IV-3). A pesar de esto el diseño de los cebadores fue posible y se realizó frente a los genomas disponibles. Esto puede explicar tentativamente los resultados obtenidos, frente a la cantidad de *contigs* en cada genotipo y frente a cada sección de la región que fue posible armar. La región central fue la más dificultosa de ensamblar y recuperar a través de las lecturas obtenidas. Esta situación podría deberse a que los cebadores presentan algún tipo de variación que impide la hibridación completa de los mismos a la secuencia objetivo (*target*). La competencia entre los cebadores y una mayor afinidad de estos frente a uno de los genomas podría explicar los resultados obtenidos. Desarrollar nuevos cebadores a partir de los *contigs* obtenidos es una potencial herramienta para obtener nuevas secuencias genómicas que permita obtener los fragmentos faltantes de esta región que no llegaron a ser secuenciados en este experimento.

Recursos pangenómicos se está desarrollando en la actualidad y promoverá aún más los estudios de este tipo como también otros análisis del tipo evolutivos y/o funcionales en el clado del tomate (Gao *et al.*, 2019). En este clado se obtuvo una primera exploración hacia la obtención de un pangenoma para *Solanum* spp. Este recurso es crucial ya que en muchos programas de mejoramiento genético como el presentado en este trabajo de tesis suele utilizarse especies silvestres. *Solanum pimpinellifolium* es el progenitor silvestre del tomate cultivado. En los programas es usual ya que presenta tolerancia al estrés y un sabor intenso, y por eso es un recurso valioso como estrategia de mejoramiento aportando características importantes y notables de la calidad del fruto.

En la actualidad la mayoría de los secuenciadores poseen gran capacidad de secuenciación. Son instrumentos de alto rendimiento, pero en general este poder de secuenciación no puede utilizarse de manera parcial, sino de manera total. La estrategia

aquí planteada es capaz de dilucidar regiones genómicas de interés en donde existe una asociación entre lo genético y lo fenotípico. La utilización de nuevas tecnologías como son los *barcodes* para etiquetar varias muestras individuales y luego ingresarlas al secuenciador como una única muestra reduce sustancialmente los costos de la secuenciación propiamente dicha y permite utilizar el poder de rendimiento del secuenciador de manera total. Esta es así una de las principales ventajas que tiene esta estrategia y posibilita desarrollar otro tipo de asistencia que complementa a los programas de mejoramiento genético para la obtención de ideotipos superiores.

CONCLUSIONES

- Se caracterizaron los materiales de primer ciclo cv. Caimanta, accesión silvestre LA0722 y su F₁ CxP mediante secuenciación de tercera generación en la región de las sHSPs del cromosoma 6.
- Esta secuenciación permitió comparar los polimorfismos previamente identificados en la secuencia génica de los progenitores y así verificar el comportamiento genético de los marcadores moleculares desarrollados previamente de tipo *InDels* y SNP.

CONCLUSION GENERAL

En este trabajo de tesis se validó la participación de las sHSPs en el proceso de maduración del fruto de tomate. A través de enfoques ómicos se detectó la presencia de polimorfismo molecular en las secuencias genómicas de los progenitores del programa de mejoramiento en la región en donde dichas sHSPs presentan expresión diferencial. Ese polimorfismo fue clave y potencial para el desarrollo de marcadores moleculares en base a la secuencia génica de estos materiales de primer ciclo y así utilizarlos para detectar variabilidad genética existente en la población segregante F_4 obtenida a partir del cruzamiento biparental entre las RILs ToUNR18xToUNR1. Esa variabilidad genética detectada con estos marcadores moleculares desarrollados estuvo asociada con caracteres de interés agronómico como son peso del fruto, altura y diámetro del fruto, coeficiente α y pH del mismo. Estos marcadores moleculares desarrollados son herramientas potenciales que pueden asistir al programa de mejoramiento genético para la obtención de nuevas variedades élites optimizando de manera significativa los recursos disponibles. La exploración de la región de interés en donde localizan estas sHSPs con expresión diferencial en las etapas principales de maduración del fruto evidenció nuevas variantes que permitirán en un futuro desarrollar nuevos marcadores moleculares y continuar con el mejoramiento genético con una mayor profundidad y eficacia.

CONSIDERACIONES FINALES

En los programas de mejoramiento genético vegetal es usual utilizar el cruzamiento con un progenitor silvestre con el fin de potenciar la variabilidad e introgresar nuevos cambios fenotípicos en la descendencia. Los cruzamientos interespecíficos son ideales para estudiar las bases moleculares que rigen los caracteres de calidad del fruto en tomate. Esto se debe al elevado polimorfismo existente, capaz de localizar un alto número de marcadores moleculares en las regiones de interés que controlan los caracteres de interés; pero fundamentalmente porque las especies silvestres (como uno de los progenitores del cruzamiento) no han sido suficientemente exploradas en su aporte para mejorar las características de calidad del fruto. Los cambios fenotípicos asociados a la introgresión pueden ser evaluados a través de diferentes aplicaciones ómicas. A partir de los objetivos e hipótesis planteados en la presente tesis se logró integrar la información de estudios genómicos y transcripcionales con un enfoque traslacional. La identificación de genes con expresión diferencial durante el proceso de maduración del fruto permitió desarrollar marcadores moleculares en base a la secuencia de estos genes que asisten de manera eficiente el programa de mejoramiento genético de la especie. En esta tesis se desarrollaron 4 marcadores moleculares que asociaron con variables agronómicas de interés. Numerosos genes contribuyen igualmente hacia un determinado efecto fenotípico como las encontradas en la variación continua de muchos rasgos cuantitativos, cada uno con pequeños e iguales efectos y sujeto a la influencia del medio ambiente. Para descubrir genes que controlan estas variables de interés agronómico es necesario realizar caracterizaciones tanto de naturaleza molecular como fenotípica y aplicar diferentes análisis genéticos en las poblaciones segregantes con el fin de establecer dichas asociaciones. La exploración de la región de interés en donde localizan estas sHSPs con expresión diferencial en la maduración del fruto y que interactúan con el etileno, evidenció nuevas variantes que permitirán en un futuro desarrollar nuevos marcadores moleculares y continuar con el programa. Estas nuevas herramientas del mejoramiento genético permiten planificar nuevas estrategias y utilizar de manera efectiva los recursos con el fin de desarrollar en el mediano plazo nuevas variedades de calidad comercial y larga vida poscosecha tendientes a satisfacer las diversas demandas de los consumidores como los productores.

BIBLIOGRAFÍA

Alba, R., Payton, P., Fei, Z., McQuinn, R., Debbie, P., Martin, G. B., ... James J. Giovannoni. (2005). Transcriptome and Selected Metabolite Analyses Reveal Multiple Points of Ethylene Control during Tomato Fruit Development. *Plant Cell*, 17(November), 2954-2965. <http://doi.org/10.1105/tpc.105.036053.1>

Alonge, M., Wang, X., Benoit, M., Soyk, S., Pereira, L., Zhang, L., ... & Lippman, Z. B. (2020). Major impacts of widespread structural variation on gene expression and crop improvement in tomato. *Cell*, 182(1), 145-161.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research*, 25(17), 3389-3402.

Andrews S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>

Arce D.P., Krsticevic F.J., Bertolaccini M.R., Ezpeleta J., Ponce S.D., Tapia E. 2015. Analysis of Small Heat Shock Protein Gene Family Expression (RNA-Seq) during the Tomato Fruit Maturation. In: Braidot A, Hadad A, eds. VI Latin American Congress on Biomedical Engineering CLAIB 2014, Paraná, Argentina 29, 30 & 31 October 2014, Cham: Springer International Publishing, pp. 679-682.

Arce, D.P.; Krsticevic, F.J., Ezpeleta, J.; Ponce, S.D.; Pratta, G.R.; Tapia E. 2016. Heterogeneous expression pattern of tandem duplicated sHsps genes during fruit ripening in two tomato species. *Journal of Physics: Conference Series* 705 (2016) 012004 doi:10.1088/1742-6596/705/1/012004

Baniwal, S. K., Bharti, K., Chan, K. Y., Fauth, M., Ganguli, A., Kotak, S., ... von Koskull- Döring, P. (2004). Heat stress response in plants: a complex game with chaperones and more than twenty heat stress transcription factors. *Journal of Biosciences*, 29, 471–487. <http://doi.org/10.1007/BF02712120>

Barker, M. S., Kane, N. C., Matvienko, M., Kozik, A., Michelmore, R. W., Knapp, S. J., & Rieseberg, L. H. (2008). Multiple paleopolyploidizations during the evolution of the Compositae reveal parallel patterns of duplicate gene retention after millions of years. *Molecular biology and evolution*, 25(11), 2445-2455.

Basha, E., O'Neill, H., & Vierling, E. (2012). Small heat shock proteins and α -crystallins: dynamic proteins with flexible functions. Trends in Biochemical Sciences.

<http://doi.org/10.1016/j.tibs.2011.11.005>

Beckles D. M., "Factors affecting the postharvest soluble solids and sugar content of tomato (*Solanum lycopersicum* L.) fruit." Postharvest Biology and Technology, vol. 63, no. 1, pp. 129–140, 2012.

Bernacchi, D., Beck-Bunn, T., Eshed, Y., Lopez, J., Petiard, V., Uhlig, J., ... & Tanksley, S. (1998). Advanced backcross QTL analysis in tomato. I. Identification of QTLs for traits of agronomic importance from *Lycopersicon hirsutum*. Theoretical and Applied Genetics, 97(3), 381-397.

Blanca, J., Cañizares, J., Cordero, L., Pascual, L., Diez, M. J., & Nuez, F. (2012). Variation revealed by SNP genotyping and morphology provides insight into the origin of the tomato. PloS one, 7(10), e48198.

Blanca, J., Montero-Pau, J., Sauvage, C. et al. Genomic variation in tomato, from wild ancestors to contemporary breeding accessions. BMC Genomics 16, 257 (2015). <https://doi.org/10.1186/s12864-015-1444-1>

Bolger^a, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, 30(15), 2114–2120. Retrieved from <http://dx.doi.org/10.1093/bioinformatics/btu170>

Bolger^b, A., Scossa, F., Bolger, M. E., Lanz, C., Maumus, F., Tohge, T., ... & Fernie, A. R. (2014). The genome of the stress-tolerant wild tomato species *Solanum pennellii*. Nature genetics, 46(9), 1034-1038.

Cabodevila, V.G.; Cacchiarelli, P.; Pratta, G.R. (2013). Phenotypic and molecular characterization of a tomato (*Solanum lycopersicum*) Second Cycle Hybrid and QTL detection in the F₂ segregating population. 4^{to} Congreso Argentino de Bioinformática y Biología Computacional, Rosario.

Cabodevila, V.G.; Picardi, L.A.; Pratta, G.R. (2017). A multivariate approach to explore the genetic variability in the F₂ segregating population of a tomato second cycle hybrid. Basic and Applied Genetics 28(1): 7-18.

Cabodevila, V. G., Cambiaso, V., Rodríguez, G. R., Picardi, L. A., Pratta, G. R., Capel, C., Capel, J. (2021). A segregating population from a tomato second cycle hybrid allows the identification of novel QTL for fruit quality traits. *Euphytica*, 217(1), 1-14.

Cacchiarelli P., Arce D., Tapia E., Pratta G.R. (2021). Structural and functional analysis of two sHSP subfamilies in tomato ripening. *Plant Gene*, 27, 100297. Plant gene, ELSEVIER BV. Netherlands. <https://doi.org/10.1016/j.plgene.2021.100297>

Cambiaso, V.; Pereira da Costa, J.H.; Rodríguez, G.R.; Pratta, G.R.; Picardi, L.A.; Francis, D.M.; Zorzoli, R. (2015). Polimorfismo en la secuencia genómica completa entre un cultivar argentino y una especie silvestre de tomate (*Solanum* spp.). XLIV Congreso Argentino de Genética, Mar del Plata, Buenos Aires.

Cambiaso^a, V., Gimenez, M. D., da Costa, J. H. P., Vazquez, D. V., Picardi, L. A., Pratta, G. R., & Rodríguez, G. R. (2019). Selected genome regions for fruit weight and shelf life in tomato RILs discernible by markers based on genomic sequence information. *Breeding science*, 69(3), 447-454.

Cambiaso^b, V., Pratta, G. R., da Costa, J. H. P., Zorzoli, R., Francis, D. M., & Rodríguez, G. R. (2019). Whole genome re-sequencing analysis of two tomato genotypes for polymorphism insight in cloned genes and a genetic map construction. *Scientia Horticulturae*, 247, 58-66.

Causse, M., Saliba-Colombani, V., Lesschaeve, I., & Buret, M. (2001). Genetic analysis of organoleptic quality in fresh market tomato. 2. Mapping QTLs for sensory attributes. *Theoretical and Applied Genetics*, 102(2), 273-283.

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. <https://doi.org/10.1101/274100>

Cruz-Mendívil, A., López-Valenzuela, J. A., Calderón-Vázquez, C. L., Vega-García, M. O., Reyes-Moreno, C., & Valdez-Ortiz, A. (2015). Transcriptional changes associated with chilling tolerance and susceptibility in 'Micro-Tom' tomato fruit using RNA-Seq. *Postharvest biology and technology*, 99, 141-151.

De Toledo J, Pooni H and Jinks J. 1984. Predicting the properties of second cycle hybrids produced by intercrossing random samples of recombinant inbred lines. *Heredity* 53:283-292.

Del Medico, A. P., Cabodevila, V. G., Vitelleschi, M. S., & Pratta, G. R. (2019). Multivariate estimate of heritability for quality traits in tomatoes by the multiple factor analysis. *Pesquisa Agropecuária Brasileira*, 54.

Di Giacomo, M., Luciani, M. D., Cambiaso, V., Zorzoli, R., Rodríguez, G. R., & Pereira da Costa, J. H. (2020). Tomato near isogenic lines to unravel the genetic diversity of *S. pimpinellifolium* LA0722 for fruit quality and shelf life breeding. *Euphytica*, 216(8), 1-15.

Di Rienzo, J., Balzarini, M., Gonzalez, L., Casanoves, F., Tablada, M., & Walter Robledo, C. (2010). Infostat: software para análisis estadístico.

Enis Afgan, Dannon Baker, Bérénice Batut, Marius van den Beek, Dave Bouvier, *et al.* 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses (2018, update). *Nucleic Acids Research*, Volume 46, Issue W1, 2 July 2018, Pages W537–W544, doi:10.1093/nar/gky379

Ezpeleta, J., Garcia Labari, I., Villanova, G.V., Bulacio Pilar, Lavista-Llanos Sofía, Posner Victoria, Krsticevic Flavia, Arranz Silvia & Tapia Elizabeth. Robust and scalable barcoding for massively parallel long-read sequencing. *Sci Rep* 12, 7619 (2022). <https://doi.org/10.1038/s41598-022-11656-0>

Faurobert, M., Mihr, C., Bertin, N., Pawlowski, T., Negroni, L., Sommerer, N., & Causse, M. (2007). Major proteome variations associated with cherry tomato pericarp development and ripening. *Plant Physiology*, 143(3), 1327–46. <http://doi.org/10.1104/pp.106.092817>

Fernandez-Pozo, N., Menda, N., Edwards, J. D., Saha, S., Tecle, I. Y., Strickler, S. R., & Mueller, L. A. (2015). The Sol Genomics Network (SGN)—from genotype to phenotype to breeding. *Nucleic acids research*, 43(D1), D1036-D1041.

Ferratto J, Mondino MC, Longo A, Grasso R (2006). Diagnóstico y necesidades de estrategias de intervención del Proyecto Hortícola de Rosario. 40.

Ferraz de Toledo J. 1987. Predicting the inbreeding and the outcrossing potential of soybeans (*Glycine max* (L.) Merrill) varieties. *Braz J Genet* 10:543-558.

Fragkostefanakis, S., Simm, S., Paul, P., Bublak, D., Scharf, K. D., & Schleiff, E. (2015). Chaperone network composition in *Solanum lycopersicum* explored by transcriptome profiling and microarray meta-analysis. *Plant, Cell and Environment*, 38(4), 693–709. <http://doi.org/10.1111/pce.12426>

Frank, G., Pressman, E., Ophir, R., Althan, L., Shaked, R., Freedman, M., ... Firon, N. (2009). Transcriptional profiling of maturing tomato (*Solanum lycopersicum* L.) microspores reveals the involvement of heat shock proteins, ROS scavengers, hormones, and sugars in the heat stress response. *Journal of Experimental Botany*, 60(13), 3891–3908. <http://doi.org/10.1093/jxb/erp234>

Fulton, T. M., Grandillo, S., Beck-Bunn, T., Fridman, E., Frampton, A., Lopez, J., ... & Tanksley, S. D. (2000). Advanced backcross QTL analysis of a *Lycopersicon esculentum* × *Lycopersicon parviflorum* cross. *Theoretical and Applied Genetics*, 100(7), 1025-1042.

Gallo, M., Rodríguez, G. R., Zorzoli, R., Picardi, L. A., & Pratta, G. R. (2010). Proteómica de la madurez del tomate: Identificación de dos estados de madurez del fruto por perfiles proteicos totales del pericarpio en RILs de tomate. *Revista de la Facultad de Ciencias Agrarias*, 42(2), 119-133.

Gao, L., Gonda, I., Sun, H., Ma, Q., Bao, K., Tieman, D. M., ... & Fei, Z. (2019). The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. *Nature genetics*, 51(6), 1044-1051.

Giardine, B., Riemer, C., Hardison, R. C., Burhans, R., Elnitski, L., Shah, P., ... & Nekrutenko, A. (2005). Galaxy: a platform for interactive large-scale genome analysis. *Genome research*, 15(10), 1451-1455.

Gonzalo, M. J., & Van Der Knaap, E. (2008). A comparative analysis into the genetic bases of morphology in tomato varieties exhibiting elongated fruit shape. *Theoretical and Applied Genetics*, 116(5), 647-656.

Goyal RK, *et al.* (2012). Features of a unique intron less cluster of class I small heat shock protein genes in tandem with box C/D snoRNA genes on chromosome 6 in tomato (*Solanum lycopersicum*) *Planta*. 235:453–471. doi:10.1007/s00425-011-1518-5

Gower, J. C. (1975). Generalized procrustes analysis. *Psychometrika*, 40(1), 33-51.

Grandillo, S., & Tanksley, S. D. (1996). QTL analysis of horticultural traits differentiating the cultivated tomato from the closely related species *Lycopersicon pimpinellifolium*. *Theoretical and Applied Genetics*, 92(8), 935-951.

Guerrero, D., Bautista, R., Villalobos, D. P., Cantón, F. R., & Claros, M. G. (2010). AlignMiner: a Web-based tool for detection of divergent regions in multiple sequence alignments of conserved sequences. *Algorithms for Molecular Biology*, 5(1), 1-16.

Guo, M., Liu, J. H., Ma, X., Luo, D. X., Gong, Z. H., & Lu, M. H. (2016). The plant heat stress transcription factors (HSFs): structure, regulation, and function in response to abiotic stresses. *Frontiers in plant science*, 7, 114.

Ipsilandis CG, Totaklidis IS, Vafias V and Stefanis, D. 2006. Criteria for developing second-cycle hybrids in maize. *Asian J Plant Sci* 5:680-685.

Jacob, P., Hirt, H., & Bendahmane, A. (2017). The heat-shock protein/chaperone network and multiple stress resistance. *Plant biotechnology journal*, 15(4), 405-414.

Jimenez A, Creissen G, Kular B, Firmin J, Robinson S, Verhoeyen M, Mullineaux P. 2002. Changes in oxidative processes and components of the antioxidant system during tomato fruit ripening. *Planta* 214(5): 751-758

Kearsey, M. J., Pooni, H. S., & Bulmer, M. (1996). The Genetical Analysis of Quantitative Traits. *Genetical Research*, 68(2), 183.

Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*, 12(4), 357-360. Retrieved from <http://dx.doi.org/10.1038/nmeth.3317>

Kopylova, E., Noé, L., & Touzet, H. (2012). SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics*, 28(24), 3211–3217. <https://doi.org/10.1093/bioinformatics/bts611>

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., & Phillippy, A. M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome research*, 27(5), 722-736.

Krsticevic, F. J., Santos, H. L., Januário, S., Schrago, C. G., & Carvalho, A. B. (2010). Functional copies of the Mst77F gene on the Y chromosome of *Drosophila melanogaster*. *Genetics*, 184(1), 295-307.

Krsticevic F. J., Arce D.P., Ezpeleta J., Tapia E. 2016. Tandem Duplication Events in the Expansion of the Small Heat Shock Protein Gene Family in *Solanum lycopersicum* (cv. Heinz 1706). *G3: GENES, GENOMES, GENETICS* October 1, 2016 vol. 6 no. 10 3027-3034; <https://doi.org/10.1534/g3.116.032045>.

Lancaster, J. E., Lister, C. E., Reay, P. F., & Triggs, C. M. (1997). Influence of pigment composition on skin color in a wide range of fruit and vegetables. *Journal of the American Society for Horticultural Science*, 122(4), 594-598.

Lawrence, S. D., Cline, K., & Moore, G. A. (1997). Chromoplast development in ripening tomato fruit: identification of cDNAs for chromoplast-targeted proteins and characterization of a cDNA encoding a plastid-localized low-molecular-weight heat shock protein. *Plant Molecular Biology*, 33, 483–492.

Lee, J. M., Joung, J.-G., McQuinn, R., Chung, M.-Y., Fei, Z., Tieman, D., ... Giovannoni, J. (2012). Combined transcriptome, genetic diversity and metabolite profiling in tomato fruit reveals that the ethylene response factor SIERF6 plays an important role in ripening and carotenoid accumulation. *The Plant Journal : For Cell and Molecular Biology*, 70(2), 191–204. <http://doi.org/10.1111/j.1365-3113X.2011.04863.x>

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16), 2078-2079.

Liao, Y., Smyth, G. K., & Shi, W. (2013). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30(7), 923-930. Retrieved from <http://dx.doi.org/10.1093/bioinformatics/btt656>

Liu Y., Zhou J., White K.P. 2014. RNA-seq differential expression studies: more sequence or more replication?. *Bioinformatics*. 2014 Feb 1;30(3):301-4. doi: 10.1093/bioinformatics/btt688. Epub 2013 Dec 6.

Loehlin, D. W., & Carroll, S. B. (2016). Expression of tandem gene duplicates is often greater than twofold. *Proceedings of the National Academy of Sciences*, 113(21), 5988-5992.

Lopes-Caitar, V. S., de Carvalho, M. C. C. G., Darben, L. M., Kuwahara, M. K., Nepomuceno, A. L., Dias, W. P., ... Marcelino-Guimarães, F. C. (2013). Genome-wide analysis of the Hsp20 gene family in soybean: comprehensive sequence, genomic organization and expression profile analysis under abiotic and biotic stresses. *BMC Genomics*, 14, 577. <http://doi.org/10.1186/1471-2164-14-577>

Löw, D., Brändle, K., Nover, L., & Forreiter, C. (2000). Cytosolic heat-stress proteins Hsp17.7 class I and Hsp17.3 class II of tomato act as molecular chaperones in vivo. *Planta*, 211, 575–582. <http://doi.org/10.1007/s004250000315>

Meli VS, Ghosh S, Parva TN, Chakraborty S and Datta A. 2010. Enhancement of fruit shelf life by suppressing N-glycan processing enzymes. *Proc Natl Acad Sci* 107:2413-2418.

Mueller, L. A., Solow, T. H., Taylor, N., Skwarecki, B., Buels, R., Binns, J., ... & Tanksley, S. D. (2005). The SOL Genomics Network. A comparative resource for Solanaceae biology and beyond. *Plant physiology*, 138(3), 1310-1317.

Neta-Sharir I, Isaacson T, Lurie S, Weiss D. 2005. Dual Role for Tomato Heat Shock Protein 21: Protecting Photosystem II from Oxidative Stress and Promoting Color Changes during Fruit Maturation. *Plant Cell* 200017:1829-1838.

Ouyang, Y., Chen, J., Xie, W., Wang, L., & Zhang, Q. (2009). Comprehensive sequence and expression profile analysis of Hsp20 gene family in rice. *Plant Molecular Biology*, 70(3), 341–357. <http://doi.org/10.1007/s11103-009-9477-y>

Peralta, I. E., Spooner, D. M., & Knapp, S. (2008). Taxonomy of wild tomatoes and their relatives (Solanum sect. Lycopersicoides, sect. Juglandifolia, sect. Lycopersicon; Solanaceae). *Systematic botany monographs*, 84.

Pereira da Costa, J. H., Rodríguez, G. R., Pratta, G. R., Picardi, L. A., & Zorzoli, R. (2013). QTL detection for fruit shelf life and quality traits across segregating populations of tomato. *Scientia Horticulturae*, 156, 47-53.

Pereira da Costa, J.H.; Mahuad, S.L.; Liberatti, D.R.; Marchionni Basté, E.M.; Rodríguez, G.R., Zorzoli, R.; Picardi, L.A. and Pratta, G.R. (2016). Tomato Second Cycle Hybrids as a new source of genetic variability for fruit quality traits. *Crop Breeding and Applied Biotechnology* 16: 289-297.

Pooni HS and Jinks JL. 1985. Predicting the properties of first cycle inbreds and second cycle hybrids extractable from two, three and four parent crosses. *Heredity* 54:397-411.

Poulain, P., Gelly, J.-C., & Flatters, D. (2010). Detection and Architecture of Small Heat Shock Protein Monomers. *PLoS ONE*, 5(4), e9990. <http://doi.org/10.1371/journal.pone.0009990>

Prasinos, C., Krampis, K., Samakovli, D., & Hatzopoulos, P. (2005). Tight regulation of expression of two Arabidopsis cytosolic Hsp90 genes during embryo development. *Journal of Experimental Botany*, 56, 633–644. <http://doi.org/10.1093/jxb/eri035>

Pratta, G.R.; Rodriguez, G.R.; Zorzoli, R.; Valle, E.M. and Picardi, L.A. 2011. Phenotypic and molecular characterization of selected tomato recombinant inbred lines derived from a cross *Solanum lycopersicum* x *S. pimpinellifolium*. *Journal of Genetics* 90: 229-237

Razali, R., Bougouffa, S., Morton, M. J., Lightfoot, D. J., Alam, I., Essack, M., ... & Negrão, S. (2018). The genome sequence of the wild tomato *Solanum pimpinellifolium* provides insights into salinity tolerance. *Frontiers in Plant Science*, 9, 1402.

Ré M.D., González C., Escobar M.R., Sossi M.L., Valle E.M., Boggio S.B. 2017. Small heat shock proteins and the postharvest chilling tolerance of tomato fruit. *Physiol Plant* 159, 148-160

Rice, P., Longden, I., & Bleasby, A. (2000). EMBOSS: the European molecular biology open software suite. *Trends in genetics*, 16(6), 276-277.

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2009). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), 139-140. Retrieved from <http://dx.doi.org/10.1093/bioinformatics/btp616>

Rodriguez, G.R.; Pratta, G.R.; Zorzoli, R. and Picardi, L.A. (2006). Recombinant lines obtained from an interspecific cross among *Lycopersicon* species selected by fruit weight and fruit shelf life. *Journal of the American Society for Horticultural Science* 131(5): 651-656.

Rodríguez, G. R., Kim, H. J., & Van Der Knaap, E. (2013). Mapping of two suppressors of OVATE (sov) loci in tomato. *Heredity*, 111(3), 256-264.

Sanmiya, K., Suzuki, K., Egawa, Y., & Shono, M. (2004). Mitochondrial small heat-shock protein enhances thermotolerance in tobacco plants. *FEBS Letters*, 557(1), 265–268

Scharf, K., Siddique, M., & Vierling, E. (2001). The expanding family of *Arabidopsis thaliana* small heat stress proteins and a new family of proteins containing α -crystallin domains (Acd proteins). *Cell Stress & Chaperones*, 6, 225–237.

Schmidt, M. H. W., Vogel, A., Denton, A. K., Istace, B., Wormit, A., van de Geest, H., Usadel, B. (2017). De novo assembly of a new *Solanum pennellii* accession using nanopore sequencing. *The Plant Cell*, 29(10), 2336-2348.

Schuelter, A.R.; Finger, F.L.; Casali, V.W.D.; Brommonschenkel, S.H.; Otoni, W.C. Inheritance and genetic linkage analysis of a firm-ripening tomato mutant. *Plant Breeding*, v.121, p.338-342, 2002.

Shinozaki, Y., Nicolas, P., Fernandez-Pozo, N., Ma, Q., Evanich, D. J., Shi, Y., ... & Rose, J. K. (2018). High-resolution spatiotemporal transcriptome mapping of tomato fruit development and ripening. *Nature communications*, 9(1), 1-13.

Shukla, V., Upadhyay, R. K., Tucker, M. L., Giovannoni, J. J., Rudrabhatla, S. V., & Mattoo, A. K. (2017). Transient regulation of three clustered tomato class-I small heat-shock chaperone genes by ethylene is mediated by SIMADS-RIN transcription factor. *Scientific reports*, 7(1), 1-12.

Takei, H., Shirasawa, K., Kuwabara, K., Toyoda, A., Matsuzawa, Y., Iioka, S., & Ariizumi, T. (2021). De novo genome assembly of two tomato ancestors, *Solanum*

pimpinellifolium and *Solanum lycopersicum* var. *cerasiforme*, by long-read sequencing. *DNA Research*, 28(1), dsaa029.

Tanksley, S. D., Grandillo, S., Fulton, T. M., Zamir, D., Eshed, Y., Petiard, V., ... & Beck-Bunn, T. (1996). Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative *L. pimpinellifolium*. *Theoretical and applied genetics*, 92(2), 213-224.

The Tomato Genome Consortium. 2012. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485, 635–641 (31 May 2012) doi:10.1038/nature11119.

The 100 Tomato Genome Sequencing Consortium. 2014. Exploring genetic variation in the tomato (*Solanum* section *Lycopersicon*) clade by whole-genome sequencing. *Plant J.* 2014 Oct;80(1):136-48. doi: 10.1111/tpj.12616. Epub 2014 Sep 3.

Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., ... & Su, Z. (2017). agriGO v2. 0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic acids research*, 45(W1), W122-W129.

Untergasser A., Nijveen H., Rao X., Bisseling T., Geurts R., Leunissen J.A. Primer3Plus, an enhanced web interface to Primer3 *Nucleic Acids Research* 2007 35: W71-W74; doi:10.1093/nar/gkm306.
<http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi>

Upadhyay, R. K., Tucker, M. L., & Mattoo, A. K. (2020). Ethylene and ripening inhibitors modulate expression of SIHsp17. 7A, B class I small heat shock protein genes during tomato fruit ripening. *Frontiers in plant science*, 975.

Van de Wouw, M., van Hintum, T., Kik, C., van Treuren, R., & Visser, B. (2010). Genetic diversity trends in twentieth century crop cultivars: a meta analysis. *Theoretical and applied genetics*, 120(6), 1241-1252. <https://doi.org/10.1007/s00122-009-1252-6>

Wang, X., Gao, L., Jiao, C., Stravoravdis, S., Hosmani, P. S., Saha, S., ... & Fei, Z. (2020). Genome of *Solanum pimpinellifolium* provides insights into structural variants during tomato breeding. *Nature communications*, 11(1), 1-11.

Waters, E. R. (2012). The evolution, function, structure, and expression of the plant sHSPs. *Journal of Experimental Botany*, 64(2), 391–403. <http://doi.org/10.1093/jxb/ers355>

Wen, J., Jiang, F., Weng, Y., Sun, M., Shi, X., Zhou, Y., ... & Wu, Z. (2019). Identification of heat-tolerance QTLs and high-temperature stress-responsive genes through conventional QTL mapping, QTL-seq and RNA-seq in tomato. *BMC plant biology*, 19(1), 1-17.

Zhou Du, Xin Zhou, Yi Ling, Zhenhai Zhang, Zhen Su. 2010. agriGO: a GO analysis toolkit for the agricultural community, *Nucleic Acids Research*, Volume 38, Issue suppl_2, 1, Pages W64–W70, <https://doi.org/10.1093/nar/gkq310>

Zorzoli, R., Pratta, G. R., Picardi, L. A. (2000). Variabilidad genética para la vida postcosecha y el peso de los frutos en tomate para familias F₃ de un híbrido interespecífico. *Pesquisa Agropecuária Brasileira*, 35(12), 2423-2427.