



UNIVERSIDAD AUTÓNOMA CHAPINGO

DEPARTAMENTO DE FITOTECNIA

INSTITUTO DE HORTICULTURA

**IDENTIFICACIÓN DE REGIONES GENÓMICAS ASOCIADAS A
CALIDAD DE FRUTO Y RENDIMIENTO EN 95 LÍNEAS DE TOMATE
(*Solanum lycopersicum* L.)**

TESIS

**QUE COMO REQUISITO PARCIAL PARA OBTENER EL GRADO DE:
DOCTOR EN CIENCIAS EN HORTICULTURA**

PRESENTA:

VIOLETA CALVO SALAZAR

BAJO LA DIRECCIÓN DE:

DR. JUAN ENRIQUE RODRIGUEZ PEREZ



APROBADA



CHAPINGO, ESTADO DE MÉXICO, MAYO 2021.

IDENTIFICACIÓN DE REGIONES GENÓMICAS ASOCIADAS A CALIDAD DE FRUTO Y RENDIMIENTO EN 95 LÍNEAS DE TOMATE (*Solanum lycopersicum* L.)

Tesis realizada por Violeta Calvo Salazar, bajo la dirección del Comité Asesor indicado, aprobada por el mismo y aceptada como requisito parcial para obtener el grado de:

DOCTORADO EN CIENCIAS EN HORTICULTURA

PRESIDENTE:



Dr. Juan Enrique Rodríguez Pérez

ASESOR:



Dr. Jaime Sahagún Castellanos

ASESOR:



Dr. Mateo Vargas Hernández

LECTOR EXTERNO:



Dra. Dulce Jazmín Hernández Melchor

Mayo 2021

CONTENIDO

LISTA DE CUADROS	v
LISTA DE FIGURAS	vi
DEDICATORIAS	vii
AGRADECIMIENTOS	viii
DATOS BIOGRÁFICOS.....	ix
RESUMEN GENERAL	1
GENERAL ABSTRACT	2
CAPITULO 1: AVANCES E IMPACTOS DE LAS TÉCNICAS DE GENOTIPIFICACIÓN Y ESTUDIOS DEL GENOMA EN EL TOMATE ¹	3
ADVANCES AND IMPACTS OF GENOTYPING TECHNIQUES AND STUDIES OF THE GENOME IN TOMATO ²	4
1.1 Introducción	5
1.2 Revisión de Literatura	7
1.2.1 Antecedentes	7
1.3 Discusión	10
1.4 Conclusiones	13
1.3 Literatura citada	14
CAPITULO 2: IDENTIFICACIÓN DE REGIONES GENÓMICAS ASOCIADAS A CALIDAD DE FRUTO Y RENDIMIENTO EN 95 LÍNEAS DE TOMATE (<i>Solanum lycopersicum</i> L.) ¹	18
IDENTIFICATION OF GENOMIC REGIONS ASSOCIATED WITH FRUIT QUALITY AND YIELD IN 95 LINES OF TOMATO (<i>Solanum lycopersicum</i> L.) ²	19
2.1 Introducción	20
2.2 Objetivos	21
2.2.1 General	21
2.2.2 Particulares	21
2.3 Hipótesis	22

2.4 Revisión de literatura	23
2.4.1 Cultivo de tomate y su importancia	23
2.4.2 Mejoramiento genético.....	23
2.4.3 Marcadores genéticos.....	24
2.4.4 Marcadores SNPs.....	24
2.4.5 Genotipificación por Secuenciación (Genotyping by Sequencing, GBS)	25
2.4.6 Análisis de asociación del genoma (GWAS).....	27
2.5 Materiales y métodos	28
2.5.1 Genotipos evaluados	28
2.5.2 Evaluación fenotípica.....	28
2.5.3 Caracteres evaluados	28
2.5.4 Análisis de datos fenotípicos.....	29
2.5.5 Análisis molecular (Extracción de ADN)	29
2.5.6 Mapeo y genotipificación de SNP	30
2.6 Resultados y discusión	31
2.6.1 Resultados de la secuenciación del genoma y estructura de la población ...	33
2.6.2 Resultados de asociación	35
2.7 Conclusiones	40
2.8 Literatura citada	41

LISTA DE CUADROS

	Pág.
Cuadro 1. Las réplicas, la media, el coeficiente de variación (CV), la varianza genotípica y la heredabilidad de los datos fenotípicos se estimaron a partir del análisis de varianza utilizando un análisis de prueba en múltiples entornos con el software R (Alvarado et al. 2015).	22
Cuadro 2. El coeficiente de correlación de Pearson entre los componentes de calidad y rendimiento.	22

LISTA DE FIGURAS

	Pág.
Figura 1. Árbol de enlace vecino basado en el cálculo de distancias genéticas desde la identidad por estado (IBS) estimadas a partir de SNP de alta calidad. El árbol se construyó con 95 líneas de tomate élite (<i>Solanum lycopersicum</i> L.), con 14.328 marcadores SNP. El dendrograma muestra 3 grupos, el primer (I) color azul donde se encuentran líneas de crecimiento mayoritariamente indeterminadas, el segundo (II) color naranja que contiene líneas mayoritariamente heterogéneas, y finalmente el grupo mayoritariamente con líneas donde el color rojo del fruto predomina.	24
Figura 2. Estructura de la población: enfoque basado en modelos para el análisis de la estructura de la población K = 2.	25
Figura 3. Gráficos de Manhattan para cuatro rasgos fenotípicos contruidos a partir de valores p de todo el genoma obtenidos de TASSEL v5.2.60 basados en un modelo lineal general ponderado con 95 líneas élite de tomate (<i>Solanum lycopersicum</i> L), con 14.328 marcadores SNP. Los genomas A, B, C y D son el peso del fruto, el número de frutos, el peso del racimo y finalmente el diámetro ecuatorial. En los gráficos de Manhattan, la línea azul indica el umbral para una asociación SNP significativa.	26

DEDICATORIAS

A mi padre Víctor Manuela Calvo Torres, quien siempre me procuro y se ocupó por mi formación, por ser el más claro ejemplo de entrega y generosidad, por su inmenso amor, dedicación, apoyo y comprensión.

AGRADECIMIENTOS

A la Universidad Autónoma Chapingo y al Instituto de Horticultura, por mi formación como Doctora en Ciencias en Horticultura.

A mi Comité Asesor integrado por el Dr. Juan Enrique Rodríguez Pérez, Dr. Jaime Sahagún Castellanos, Dr. Jorge Alexander Duitama Castellanos y el Dr. Mateo Vargas Hernández, quienes me apoyaron en todas las etapas de la investigación, compartieron sus conocimientos, me orientaron para tomar las mejores decisiones y sobre todo me brindaron su amistad y confianza.

Al Dr. Jorge Alexander Duitama Castellanos, Profesor investigador de la Universidad de los Andes, Bogotá, Colombia por brindarme su asesoría y apoyo en los análisis bioinformáticos en mi estancia profesional.

A mi padre Víctor Manuel Calvo Torres y a mis compañeros de doctorado, muy especialmente a M.C Sonia Bernabé y M.C Monserrat Tuxtla. Y a los trabajadores académicos o administrativos, trabajadores de campo y laboratorio, en especial a C. Jorge Sánchez Galicia y Quim. Ricardo Gaspar Hernández que participaron en alguna etapa de la investigación.

DATOS BIOGRÁFICOS

La autora nació en ciudad de México, el 13 de abril de 1984. Es Ingeniera en Biotecnología, graduado de la Universidad Tecnológica de Tecámac en 2011, realizó estudios en el Colegio de Postgraduados, departamento de Recursos y Productividad Genética, graduada en 2015, donde obtuvo el grado de Maestra en Ciencias.

Cuenta con 9 años de experiencia laboral en el Centro Internacional de Mejoramiento de Maíz y Trigo (CIMMYT), en el área de mejoramiento genético para resistencia a Royas de trigo.

En enero de 2017 ingresó al Instituto de Horticultura de la Universidad Autónoma Chapingo para realizar estudios de Doctorado en Ciencias en Horticultura.

RESUMEN GENERAL

IDENTIFICACIÓN DE REGIONES GENÓMICAS ASOCIADAS A CALIDAD DE FRUTO Y RENDIMIENTO EN 95 LÍNEAS DE TOMATE (*Solanum lycopersicum* L.)

La tecnología de secuenciación de ADN de nueva generación ha llevado al desarrollo de aplicaciones de detección rápida de polimorfismo de nucleótido único (SNP) en todo el genoma en varias especies de plantas para la creación de mapas de asociación (GWAS). El uso de esta técnica ha permitido definir la participación de genes en los cromosomas de caracteres cuantitativos, así como, reducir los ciclos de selección para la creación de nuevas variedades. La comprensión y recopilación de información sobre los avances e impactos del estudio del genoma completo es importante para poder aplicar esta técnica en investigaciones, por tal motivo, primero se recopiló información sobre los principios básicos de genotipificación y asociación del genoma completo. Posteriormente se llevó a cabo la aplicación de las técnicas antes descritas en 95 líneas avanzadas (F10 a F12) del programa de mejoramiento genético de la Universidad Autónoma Chapingo (UACH), producto de cruces entre variedades e híbridos comerciales de tomate para caracteres fenotípicos de calidad y rendimiento del fruto de tomate, generando 88 marcadores se asociaron con 4 variables, número de frutos (NF), peso del fruto del racimo (CW), peso del fruto (FW) y diámetro ecuatorial del fruto (DE).

Palabras clave: SNPs, GWAS, rendimiento, calidad.

¹Tesis de Doctorado en Ciencias en Horticultura, Universidad Autónoma Chapingo.
Autor: Violeta Calvo Salazar. Director: Dr. Juan Enrique Rodríguez Pérez.

GENERAL ABSTRACT

IDENTIFICATION OF GENOMIC REGIONS ASSOCIATED WITH FRUIT QUALITY AND YIELD IN 95 LINES OF TOMATO (*Solanum lycopersicum* L.)

Next-generation DNA sequencing technology has led to the development of applications for rapid detection of single nucleotide polymorphism (SNP) throughout the genome in various plant species for the creation of association maps (GWAS). The use of this technique has made it possible to define the participation of genes in the chromosomes of quantitative characteristics, as well as to reduce the selection cycles for the creation of new varieties. Understanding the compilation and be informed on the advances and impacts of the study of the whole genome is important to be able to apply this technique in research, for this reason, information was first compiled on the basic principles of genotyping and association of the whole genome. Subsequently, the application of the techniques described above was carried out in 95 advanced lines (F10 to F12) of the genetic improvement program of the Universidad Autónoma Chapingo (UACH), product of crosses between varieties and commercial hybrids of tomato for phenotypic characteristics and yield. of tomato fruit, generating 88 markers were associated with 4 variables, number of fruits (NF), bunch fruit weight (CW), fruit weight (FW) and equatorial diameter of the fruit (DE).

Keywords: SNPs, GWAS, yield, quality.

¹Tesis de Doctorado en Ciencias en Horticultura, Universidad Autónoma Chapingo.
Autor: Violeta Calvo Salazar. Director: Dr. Juan Enrique Rodríguez Pérez.

CAPITULO 1: AVANCES E IMPACTOS DE LAS TÉCNICAS DE GENOTIPIFICACIÓN Y ESTUDIOS DEL GENOMA EN EL TOMATE¹

RESUMEN

El tomate (*Solanum lycopersicum* L.) es una de las especies en las que se ha realizado mayor inversión en el mejoramiento genético, ya que es considerada la segunda hortaliza de mayor importancia a nivel mundial. Actualmente, los consumidores se han interesado en variedades con un perfil nutraceútico y sensorial superior; situación que ha repercutido en los objetivos de los programas de mejoramiento genético, los cuales se han reorientado no solo a la obtención de variedades con altos rendimientos y resistencia a enfermedades, sino también a frutos de alta calidad nutricional. En la última década, las técnicas de secuenciación de genomas y de caracterización genotípica (genotipificación) han contribuido a detectar variantes genéticas asociadas a caracteres fenotípicos, entre ellos los de calidad de fruto de tomate, con el fin de detectar de forma más precisa marcadores SNPs (polimorfismos de un solo nucleótido), que, a su vez, permitan identificar combinaciones genéticas asociadas a los caracteres fenotípicos estudiados. La selección genómica realizada con estos métodos requiere de herramientas estadísticas, matemáticas y computacionales cada vez más potentes. El objetivo de la presente investigación fue proporcionar información actualizada sobre los avances e impactos de las técnicas de genotipificación y estudio del genoma completo en el tomate para los caracteres de rendimiento y calidad de fruto.

Palabras clave: Genotipificación, genoma, rendimiento, calidad.

¹Tesis de Doctorado en Ciencias en Horticultura, Universidad Autónoma Chapingo.
Autor: Violeta Calvo Salazar. Director: Dr. Juan Enrique Rodríguez Pérez.

**ADVANCES AND IMPACTS OF GENOTYPING TECHNIQUES AND STUDIES OF
THE GENOME IN TOMATO²
ABSTRACT**

Tomato (*Solanum lycopersicum* L.) is one of the species in which the greatest investment has been made in genetic improvement, since it is considered the second most important vegetable worldwide. In recent times, consumers have been interested in varieties with a superior nutraceutical and sensory profile. This situation has had an impact on the objectives of the improvement programs, which have been reoriented not only to obtaining varieties with high yields and resistance to diseases, but also to fruits of high nutraceutical quality. In the last decade, genome sequencing and genotypic characterization (genotyping) techniques have contributed to detecting genetic variants associated with phenotypic characters, including tomato fruit quality, in order to more accurately detect SNP markers (single nucleotide polymorphisms), which in turn, seek to identify genetic combinations associated with the phenotypic characters studied. Genomic selection carried out with these methods requires increasingly powerful statistical, mathematical and computational tools. The objective of the present investigation was to provide truthful and updated information on the advances and impacts of genotyping techniques and study of the current complete genome in tomato for the characteristics of yield and fruit quality.

Keywords: Genotyping, genome, yield, quality.

² Thesis, Ph. D. in Horticultural Sciences. Universidad Autónoma Chapingo.
Author: Violeta Calvo Salazar. Advisor: Dr. Juan Enrique Rodríguez Pérez.

1.1 Introducción

La disponibilidad de técnicas de genotipificación rápidas y rentables ha hecho que los polimorfismos de un solo nucleótido (SNPs) sean los marcadores con mayor potencial para utilizarse en los análisis que involucra el estudio de todo el genoma de una especie. Las estrategias para la identificación de SNPs involucran la creación de bibliotecas de ADN, la secuenciación de estas y la aplicación de análisis bioinformáticos (Deperi et al., 2018), con el fin de detectar variantes en las secuencias de bases en puntos específicos del genoma.

En los últimos años se han generado gran cantidad de información acerca de la selección genómica y su implementación en los programas de mejoramiento genético de especies vegetales. Dentro de los estudios que se han realizado en tomate se encuentra el proyecto “The 100 Tomato Genome Sequencing Consortium” donde se exploró la variación genética mediante la secuenciación de 84 accesiones de tomate y especies silvestres directamente relacionadas con *Solanum lycopersicum*; lo que generó la secuenciación del genoma completo de esta especie; información que permitió el estudio de la diversidad del genoma del tomate con mayor precisión (Aflitos et al., 2014). Así mismo, este trabajo ha servido como base para continuar con el estudio de la diversidad de esta especie (Blanca et al., 2015; Bauchet et al., 2017; Mata et al., 2020; Razifard et al., 2020 y Florez et al., 2021).

Para impulsar la investigación y facilitar el descubrimiento de genes que confieren rasgos favorables, el Consorcio del Genoma del Tomate publicó la secuencia del genoma de alta calidad de *Solanum lycopersicum* cv. “Heinz 1706” (The Tomato Genome Consortium, 2012).

La disponibilidad de la secuencia del genoma del tomate cultivado ha dado lugar a varios avances importantes, como la identificación de genes candidatos (CG) asociados a caracteres fenotípicos específicos con una alta probabilidad; por ejemplo, aquellos asociados a caracteres de calidad (Zhong et al., 2013; Liu Ruie, 2016) y al rendimiento de fruto de tomate (Tieman et al., 2017). Esto gracias a la identificación de polimorfismos de un solo nucleótido (SNP) (Viquez et al., 2014). La secuenciación del genoma del tomate ha facilitado también el diseño del sistema de edición de genes (Brooks et al., 2014).

Ante los avances y mejoras constantes en el desarrollo de técnicas de genotipificación, es necesario revisar y verificar la forma de aplicación de estas para obtener mayores beneficios y facilitar su aplicación e interpretación. Por esta razón, el objetivo de la presente revisión fue recopilar información actualizada sobre los avances e impactos de las técnicas de genotipificación y estudio del genoma completo en el tomate para el rendimiento y calidad de fruto.

La revisión se realizó en artículos científicos en los idiomas inglés y español de revistas científicas de prestigio como: Biology, BMC Genomics, Front Plant, Molecular Biology, Natural Biotechnology, Plant Genome, PLoS Computational Biology, PLoS One, Science, Theoretical Applied Genetics. Se emplearon en mayor cuantía artículos publicados en el período de 2012 a 2021, aunque fue necesario usar publicaciones de años anteriores como aporte a la historia del tema.

Se obtuvo un total de 37 artículos, a los cuales se les aplicaron criterios consistentes en vigencia y aporte a la temática a tratar, se efectuó un análisis bibliométrico para su clasificación por tema de interés, autores y fechas de publicación.

1.2 Revisión de Literatura

1.2.1 Antecedentes

Los marcadores moleculares constituyen una herramienta poderosa que facilitan la selección de genotipos o poblaciones y la caracterización de la diversidad genética; así mismo, son de gran utilidad en la construcción y saturación de mapas genéticos (diagrama descriptivo de la ubicación de los genes en cada cromosoma) y en la identificación y aislamiento de genes y/o grupos de éstos. Lo anterior ha brindado información relevante sobre la estructura del genoma y la localización de genes mayores específicos que controlan caracteres de interés en diversas especies (Hernández, 2002). Existen diferentes tipos de marcadores genéticos, los más utilizados actualmente son los SNPs (Single Nucleotide Polymorphism), ya que son abundantes (alta densidad) y pueden presentarse en forma continua a lo largo de todo el genoma, los cuales son heredables (Lindblad et al., 2000; Kennedy et al., 2003).

La secuencia de ácidos nucleicos permite establecer el orden de los nucleótidos presentes en las moléculas de ADN o ARN, razón por la cual, su uso ha aumentado en los últimos años exponencialmente en investigaciones; es así como las herramientas de secuenciación de ADN de nueva generación (NGS) se ha implementado gracias a la posibilidad que brindan para realizar secuenciaciones masivas y simultaneas de millones de fragmentos del ADN y/o ARN presente en la muestra; gracias al empleo de tecnología de punta de bajo costo y con alto eficiencia, por lo que, actualmente, es posible secuenciar un genoma completo en un solo día (Tagliotti et al., 2020).

Las tecnologías NGS han permitido el desarrollo de aplicaciones de detección rápida de SNPs en todo el genoma en varias especies de plantas; a este proceso de caracterización del genoma se le ha llamado “genotipificación por secuenciación” (GBS) (Deschamps et al., 2012).

La GBS se basa en la creación de librerías genómicas y la posterior secuenciación de estas en un número alto de individuos, lo cual permite identificar variación en el material genético de una población o especie.

A partir de los resultados del GBS es posible realizar el mapeo de asociación o mapeo de disequilibrio de ligamiento con alta precisión o resolución y puede realizarse sobre todo el genoma completo, a esta técnica se le conoce como “genome-wide association mapping” (GWAS) (Mateos, 2018). La cual intenta asociar un carácter fenotípico a alguno de los cientos de miles de marcadores moleculares SNPs distribuidos por todo el genoma (Zhu et al., 2008). La mayor ventaja que ofrecen la aplicación de GBS y GWAS, es que se explora la totalidad del genoma basándose en el estudio de numerosos eventos de recombinación y, por tanto, ofrece la posibilidad de interpretar con mayor precisión del control genético de un carácter (Korte et al., 2013; Burghardt et al., 2017).

El mapeo asociativo (GWAS) es una de las herramientas más utilizadas para estudiar rasgos gobernados por varios genes (Yang et al., 2010). Esta técnica, también conocida como mapeo de disequilibrio de ligamiento (LD), surge como una alternativa eficaz para el mejoramiento genético de caracteres complejos, ya que permite asociar rasgos fenotípicos cuantitativos con la variación genética al explorar los eventos de recombinación históricos y evolutivos a nivel de la población (Risch, y Merikangas, 1996;

Nordborg y Tavaré, 2002).

En función de la escala y el enfoque de un estudio en particular, el mapeo de asociación generalmente se divide en dos categorías amplias, 1) mapeo de asociación de un gen candidato para un rasgo específico que está relacionado a un polimorfismo; y 2) mapeo de asociación de todo el genoma (GWAS), o exploración del genoma, que analiza la variación genética en todo el genoma para encontrar asociaciones de caracteres cuantitativos (Tagliotti et al., 2020).

La mayoría de las características heredables de importancia económica, tales como componentes de rendimiento, calidad nutricional y tolerancia a estrés, resultan de la acción conjunta de varios genes y son llamados caracteres poligénicos, cuantitativas o de herencia compleja. Poco se sabe acerca del número, posición cromosómica, magnitud del efecto e interacciones de los *loci* que controlan la expresión de dichas características. Estos *loci* son denominados QTL (Quantitative trait loci), es decir, *loci* controladores de características cuantitativas (Tagliotti et al., 2020).

Los mapas de ligamiento son basados en el principio de que los genes (marcadores o *loci*) segregan vía recombinación de los cromosomas durante la meiosis, permitiendo así su análisis en la progenie; de esta manera, los mapas de ligamiento permiten realizar un análisis completo de los genotipos, descomponer características complejas en sus componentes mendelianos, localizar regiones genómicas que controlan caracteres de importancia agronómica, además de cuantificar el efecto de estas regiones (Barrios, 2011)

El mapeo por GWAS requiere de la información detallada sobre las relaciones genéticas

sobre los individuos dentro de una población con el fin de minimizar asociaciones falsas entre el carácter bajo estudio y los SNPs (Sharma et al., 2018). Adicionalmente, a fin de evitar la dificultad en diferenciar los efectos de los polimorfismos funcionales (efectos de SNPs) de número de alelos, el nivel de ploidía debe ser considerado, especialmente si la población de estudio contiene introducciones silvestres (Zhu et al., 2008). La diversidad genética, el LD, el parentesco de individuos dentro de la población, la densidad de marcadores, los métodos estadísticos a emplear, determinan la resolución del mapeo (Sharma et al., 2018).

El análisis del desequilibrio de ligamiento (LD) en el genoma de un organismo es un factor clave que afecta la precisión que proporciona el mapeo de asociación y, a su vez, se ve afectado por muchos factores genéticos y no genéticos, como la recombinación, la deriva génica, la selección natural y artificial, el patrón de apareamiento, y el número de alelos (Yang et al., 2010). En particular el LD busca determinar el grado de asociación no aleatoria entre marcadores moleculares con alelos de un *locus* específico o entre alelos de diferentes *loci* (Tagliotti et al., 2020).

El análisis mediante el mapeo asociativo (GWAS) explota la diversidad existente en el germoplasma estudiado sin la necesidad de construir poblaciones de mapeo. Resultados recientes parecen indicar el potencial de esta metodología en los programas de mejoramiento de tomate ayudando a identificar marcadores asociados con caracteres de rendimiento y calidad de fruto en tomate (Ruggieri et al., 2014; Sacco et al., 2015; Zhang et al., 2015; Zhang et al., 2016).

1.3 Discusión

La importancia de mejorar la calidad del fruto ha sido tema de investigación y de interés

agronómico, aunque los estudios recientes que emplean las técnicas antes mencionadas son pocos. A pesar de ello, se han ido identificando regiones genéticas candidatas para la mejora de la calidad de fruto del tomate (Ruggieri et al., 2014; Sacco et al., 2015; Zhang et al., 2015; Zhang et al., 2016).

Uno de los trabajos más significativos de asociación para calidad de fruto fue el que realizó Ruggieri V. en el 2014 donde identificó 20 SNP que se asociaron significativamente con siete de los diez caracteres. El análisis reveló dos marcadores asociados con la concentración de compuestos fenólicos, tres con las del ácido ascórbico, β -caroteno y trans-licopeno, seis con acidez titulable y uno con el pH y el peso fresco de fruto. Además, se encontraron 19 genes putativos en desequilibrio de ligamiento y, además, se identificaron combinaciones alélicas favorables entre *loci* asociados que mejoraron la expresión del carácter estudiado.

Por otra parte, en el año 2015 Adriana Sacco identificó 36 marcadores asociados a la variación de 15 caracteres de calidad de fruto y rendimiento de tomate; 30 de ellos se asociaron con la morfología del fruto, incluidos el grosor del pericarpio y el peso del fruto. Esto corroboró la fuerte participación de dos regiones genómicas en los cromosomas 2 y 11 para determinar la forma del fruto.

A partir de un mapeo de asociación para caracteres nutricionales y de calidad de fruto Zhang (2015) utilizó 174 accesiones de tomate con 182 marcadores moleculares de Secuencias Simples Repetidas (SSRs), detectó 111 marcadores SSRs asociados a diez caracteres medidos; la mayoría de ellos con el diámetro ecuatorial del fruto, el ácido ascórbico y el peso del fruto, con presencia en los cromosomas 2 y 11. Posteriormente

Zhang (2016) con este mismo esquema detecto QTLs para el contenido de sólidos solubles en los cromosomas 2 y 9.

Para incrementar el rendimiento, el peso de la fruta es un criterio de selección importante en los programas de mejora de tomate. Este carácter es de herencia cuantitativa controlada por cerca de 28 *loci* (Grandillo et al., 1999). Los *loci* fw1.1, fw2.2, fw2.3, fw3.1, fw4.1 y fw9.1 explican en conjunto más de 20 % de varianza fenotípica del peso de fruto. Además, los genes fw1.1, fw2.1, fw2.2, fw3.1, fw3.2 y fw11.3 se han identificado en al menos cuatro estudios independientes (Grandillo et al. 1999). El QTL de peso de la fruta del cromosoma 2 (fw2.2), fue el primer gen asociado a QTL que se identificó y clonó (Alpert et al., 1995; Alpert y Tanksley, 1996; Frary et al., 2000). Estos trabajos mencionados fueron realizados con marcadores moleculares de baja precisión.

El estudio GWAS con la utilización de SNPs, marcadores de alta resolución, también identificaron asociaciones significativas con regiones localizadas en los cromosomas 2, 8 y 11 (Huang et al., 2011; Ruggieri et al., 2014; Sacco et al., 2015).

En el 2011 Huang, estudio el peso de la fruta al asociarlo con el cromosoma 11 por medio de mapas de ligamiento, y determinó la posición de los genes *fas* y *fw11.3*. Posterior a este descubrimiento en el 2014 y 2015 Ruggieri V y Sacco A corroboran la participación del cromosoma 11 en caracteres de peso de fruto que es un componente de rendimiento del tomate.

A partir de lo anterior, las perspectivas de futuras investigaciones deben dirigirse a la búsqueda de asociaciones de caracteres cuantitativos que también están involucrados en la calidad y rendimiento del fruto del tomate como lo es la concentración de vitamina A y licopeno, firmeza, diámetro polar y ecuatorial, numero de frutos, etc. Es importante

mencionar que se pueden asociar otros caracteres cuantitativos de interés agronómico como lo es la resistencia a enfermedades o a estrés hídrico. El uso y aplicación de estas técnicas permite facilitar los criterios de selección en programas de mejoramiento genético acortando los ciclos para la obtención de nuevas variedades.

1.4 Conclusiones

A pesar del esfuerzo que se ha hecho en la investigación aplicando las técnicas de genotipificación, hace falta impulsar su uso pues ha demostrado ser una técnica confiable y precisa. La identificación de SNPs asociados a caracteres cuantitativos, permite entender la estructura del genoma y las regiones involucradas en caracteres cuantitativos, pudiendo así, definir la participación de los genes para posteriormente ser clonados y usados para crear nuevas variedades con una carga genética favorable para diferentes caracteres cuantitativos de interés agronómico.

1.3 Literatura citada

- Aflitos, S., Schijlen, E., De Jong, H., De Ridder, D., Smit, S. (2014). The 100 Tomato Genome Sequencing Consortium, Exploring genetic variation in the tomato (*Solanum* section *lycopersicon*) clade by whole-genome sequencing. *Plant J.* 80, 136–148.
- Alpert, K.B., Grandillo, D., Tanksley, S.D. (1995). *fw2.2*: a major QTL controlling fruit weight is common to both red- and green-fruited tomato species. *Theor Appl Genet.* 91:994–1000.
- Alpert, K.B., Tanksley, S.D. (1996). High-resolution mapping and isolation of a yeast artificial chromosome contig containing *fw2.2*: a major fruit weight quantitative trait locus in tomato. *Proc Natl Acad Sci USA.* 93:15503–15507.
- Bai, Y. and Lindhout, P. (2007). Domestication and breeding of tomatoes: what have we gained and what can we gain in the future? *Annals of Botany* 100:1085-1094.
- Barrios M., R.P. (2011). Identificación de qtls asociados a características agronomicas de interes, en una retrocruza avanzada de frijol comun (*Phaseolus vulgaris*). Tesis de Licenciatura. Universidad del Tolima. Facultad de ciencias, Programa de Biologia. pp. 128.
- Bauchet, G., Grenier, S., Samson, N., Bonnet, J., Grivet, L., Causse, M. (2017). Use of modern tomato breeding germplasm for deciphering the genetic control of agronomical traits by Genome Wide Association study. *Theor Appl Genet.* May;130(5):875-889.
- Bruce, A., Johnson, A., Lewis, J., Raff, M., Keith, R., Walter, P. (1994). *The Molecular Biology of the Cell*. New York: Garland Science. ISBN-10: 0-8153-4072-9.
- Brooks, C., Nekrasov, V., Lippman, Z.B., Van, E.J. (2014). Efficient gene editing in tomato in the first generation using the clustered regularly interspaced short palindromic repeats/CRISPR-associated9 system. *Plant Physiol* 166:1292–1297.
- Blanca, J., Montero, P.J., Sauvage, C., Bauchet, G., Illa, E., Díez, M.J., Francis, D., Causse, M., van der Knaap, E., Cañizares, J. (2015). Genomic variation in tomato, from wild ancestors to contemporary breeding accessions. *BMC Genomics.* Apr 1;16(1):257.
- Burghardt, L.T., Young, N.D., Tiffin, P. (2017) A guide to genome-wide association

- mapping in plants in *Current Protocols in Plant Biology*, 2:22-38.
- Bush, W.S., Moore, J.H. (2012). Chapter 11: Genome-wide association studies. *PLoS Computational Biology*, 8(12), e1002822.
- Deperi, S.I., Tagliotti, M.E., Bedogni, M.C., Manrique C., N.C., Coombs, J., Zhang, R., Huarte, M. A. (2018). Discriminant analysis of principal components and pedigree assessment of genetic diversity and population structure in a tetraploid potato panel using SNPs. *PloS one* 13, (3), 1-14.
- Deschamps, S., Llacá, V., May, G. (2012). Genotyping-by-Sequencing in Plants. *Biology*. 1. 460-83.
- Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., Mitchell, S.E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS one* 6:e19379.
- Florez R., A.M., Scharmann, M., Roth, M., Städler, T. (2021). Population Genomics of the "Arcanum" Species Group in Wild Tomatoes: Evidence for Separate Origins of Two Self-Compatible Lineages. *Front Plant Sci.* Mar 19;12:624442
- Frary, A., Nesbitt, T.C., Frary, A., Grandillo, S., van der Knaap, E., Cong, B., Liu, J., Meller, J., Elber, R., Alpert, K.B. (2000) *fw2.2*: a quantitative trait locus key to the evolution of tomato fruit size. *Science* 289:85–88.
- Grandillo, S., Ku, H.M., Tanksley, S.D. (1999). Identifying the loci responsible for natural variation in fruit size and shape in tomato. *Theor Appl Genet* 99:978–987.
- Hernández C, M.T. (2002). *Marcadores Moleculares: nuevos horizontes en la genética y la selección de las plantas*. La Habana: Editorial Félix Varela, Pp 366.
- Huang, Z., van der Knaap, E., (2011). Tomato fruit weight 11.3 maps close to fasciated on the bottom of chromosome 11. *Theor Appl Genet.*;123: 465–474.
- Korte, A., Farlow, A., (2013) The advantages and limitations of trait analysis with GWAS: a review in *Plant Methods*, 9:29.
- Liu Ruie. (2016). Functional analysis of active DNA demethylation in tomato. *Vegetal Biology*. Université de Bordeaux, English. ffNNT : 2016BORD0273ff. fftel-01806333f.
- Mata, N.E., Montero, P.J., Gimeno, P.E., Garcia C., V., Ziarsolo, P., Menda, N., Mueller, L.A., Blanca, J., Cañizares, J., van der Knaap, E., Díez, M.J. (2020). Exploiting the

- diversity of tomato: the development of a phenotypically and genetically detailed germplasm collection. *Hortic Res.*;7(1):66.
- Mateos S, B. (2018). Identificación de regiones genómicas implicadas en el control del tamaño y peso de la semilla judía mediante GWAS. Tesis Maestría. Universidad Politécnica de Valencia, Programa Genética Vegetal. pp. 46.
- Nordborg, M., Tavaré, S., (2002). Linkage disequilibrium: what history has told us in *Trends in Genetics*, 18(2): 83-90.
- Razifard, H., Ramos, A., Della V., A.L., Bodary, C., Goetz, E., Manser, E.J. (2020). Genomic Evidence for Complex Domestication History of the Cultivated Tomato in Latin America. *Mol Biol Evol.*;37: 1118–1132.
- Risch, N., Merikangas, K., (1996). The future of genetic studies of complex human diseases. *Science*. 273(5281):1516–1517.
- Ruggieri, V., Francese, G., Sacco, A., D'Alessandro, A., Rigano, M.M., Parisi, M., (2014). An association mapping approach to identify favourable alleles for tomato fruit quality breeding. *BMC Plant Biol.*;14.
- Sacco, A., Ruggieri, V., Parisi, M., Festa, G., Rigano, M.M., Picarella, M.E. (2015). Exploring a tomato landraces collection for fruit-related traits by the aid of a high-throughput genomic platform. *PLoS One.*;10: 1–20.
- Sharma, S.K., MacKenzie, K., McLean, K., Dale, F., Daniels, S., Bryan, G.J. (2018). Linkage Disequilibrium and Evaluation of Genome-Wide Association Mapping Models in Tetraploid Potato. *G3: Genes, Genomes, Genetics*, 8(10), 3185-3202
- Sim, S.C., Durstewitz, G., Plieske, J., Wieseke, R., Ganai, M.W., Van Deynze, A., Hamilton, J.P., Buell, C.R., Causse, M., Wijeratne, S., Francis, D. (2012). Development of a large SNP genotyping array and generation of high-density genetic maps in tomato. *PLoS One* 7(7):e40563.
- Tagliotti, M.E, Bedogni, M.C., Cendoya, M.G., Ortego, J., Huarte, M. (2020). Multi-trait differential response in a diverse potato panel under contrasting water regimes. *Crop Science*. *Crop Science*. 60. 10.1002/csc2.20031.
- Tieman, D., Zhu, G., Resende, M.F., Lin J., Nguyen, T., Bies, C., Rambla D., Beltran, J.L., Taylor, K.S., Zhang, M., Ikeda, B., Liu H., Fisher Z., Zemach, J., Monforte I., Zamir, A., Granell, D., Kirst, A., Huang, M., and Klee, H. (2017). A chemical genetic

- roadmap to im-proved tomato flavor. *Science* 355: 391–394.
- Víquez Z., M., Vosman, B., van de Geest, H., Bovy, A., Visser R, G.F., Finkers, R., van Heusden, A.W. (2013). Tomato breeding in the genomics era: Insights from a SNP array. *BMC Genomics*, 14(1):354.
- Xu, Z., Taylor, J.A. (2009). SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. *Nucleic acids research* 37:W600-W605.
- Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., Goddard, M.E., Visscher, P.M. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nat Genet.* 2010 Jul;42(7):565-9.
- Zhang, J., Zhao, J., Xu, Y., Liang, J., Chang, P., Yan, F. (2015). Genome-wide association mapping for tomato volatiles positively contributing to tomato flavor. *Front Plant Sci.*;6: 1–13.
- Zhang, J., Zhao, J., Liang, Y., Zou, Z. (2016). Genome-wide association-mapping for fruit quality traits in tomato. *Euphytica.*;207: 439–451.
- Zhong, S., Fei, Z., Chen, Y.R. et al. (2013) Single-base resolution methylomes of tomato fruit development reveal epigenome modifications associated with ripening. *Nat Biotechnol* 31(2): 154–159.
- Zhu, C., Gore, M., Buckler, E. S., Yu, J., (2008) Status and Prospects of Association Mapping in Plants in *Plant Genome* 1:5-20.

CAPITULO 2: IDENTIFICACIÓN DE REGIONES GENÓMICAS ASOCIADAS A CALIDAD DE FRUTO Y RENDIMIENTO EN 95 LÍNEAS DE TOMATE (*Solanum lycopersicum* L.)¹

RESUMEN

El objetivo de la presente investigación fue identificar las regiones genómicas asociadas con la calidad y rendimiento de fruto en tomate, los resultados de la genotipificación y las evaluaciones fenotípicas realizadas en 95 líneas avanzadas (F10 a F12) del programa de mejoramiento genético de la Universidad Autónoma Chapingo (UACH), producto de cruces entre variedades e híbridos comerciales de tomate. Estas líneas incluyen diversidad de formas, colores y tamaños de frutos, así como variedades de ensaladas, cerezas y bolas. Las evaluaciones fenotípicas de los caracteres relacionados con el rendimiento y la calidad del fruto se realizaron en condiciones de invernadero en un ensayo de bloques completos al azar con tres repeticiones y una unidad experimental de seis plantas. El análisis de genotipificación por secuenciación (GBS) generó 14,328 SNP funcionales después de un análisis de filtrado estricto. La agrupación de muestras mediante la unión de vecinos identificó tres grupos. Realizamos asociaciones de todo el genoma (GWAS) utilizando el modelo lineal mixto implementado en Tassel para relacionar los SNP descubiertos con GBS con la variabilidad fenotípica observada. Para los parámetros de rendimiento, las asociaciones fueron altamente significativas y consistentes en todas las repeticiones, 88 marcadores se asociaron con 4 variables, número de frutos (NF), peso del fruto del racimo (CW), peso del fruto (FW) y diámetro ecuatorial del fruto (DE). Con los resultados obtenidos se espera que permita la predicción genómica de futuros, esfuerzos de mejoramiento que combinen este panel de diversidad para producir variedades mejoradas.

Palabras clave: Calidad, Rendimiento, genotipificación por secuenciación, asociaciones de todo el genoma.

¹Tesis de Doctorado en Ciencias en Horticultura, Universidad Autónoma Chapingo.
Autor: Violeta Calvo Salazar. Director: Dr. Juan Enrique Rodríguez Pérez.

IDENTIFICATION OF GENOMIC REGIONS ASSOCIATED WITH FRUIT QUALITY AND YIELD IN 95 LINES OF TOMATO (*Solanum lycopersicum* L.)²
ABSTRACT

The aim of the present investigation was to identify genomic regions associated with fruit quality and yield traits in tomato, the results of genotyping and phenotypic evaluations performed on 95 advanced lines (F10 to F12) of the genetic improvement program at Universidad Autónoma Chapingo (UACH), generated from segregating selected populations, product of crosses among varieties and commercial tomato hybrids. These lines include diversity of shape, color and fruit size, as well as varieties of salad, cherry and ball types. Phenotypic evaluations of traits related to yield and fruit quality were performed under greenhouse conditions in a randomized complete blocks trial with three replicates and six plants experimental unit. GBS of this diversity panel generated 22,127 SNP markers but only 14,328 SNPs were functional after strict filtering analysis. Sample clustering using Neighbor Joining identified three groups. We performed Genome-Wide Associations (GWAS) using the Mixed Linear Model implemented in Tassel to relate SNPs discovered with GBS to the observed phenotypic variability. For the yield parameters the associations were highly significant and consistent across replicates. Yield traits associations were highly significant, 88 markers were associated with 4 variables, Number of fruits (NF), Cluster fruit weight (CW), Fruit weight (FW), and Fruit equatorial diameter (DE). We expect that these results serve as a basis to enable genomic prediction of phenotypes in future breeding efforts combining this diversity panel to produce improved varieties for different traits.

Keywords: Quality, yield, Genotyping by sequencing, Genome-wide association.

² Thesis, Ph. D. in Horticultural Sciences. Universidad Autónoma Chapingo.
Author: Violeta Calvo Salazar. Advisor: Dr. Juan Enrique Rodríguez Pérez.

2.1 Introducción

El tomate (*Solanum lycopersicum* L.) es después de la papa (*Solanum tuberosum*), la segunda hortaliza más importante a nivel mundial, el tomate demanda alta calidad de insumos y mano de obra para maximizar el rendimiento (Melomey et al., 2019).

Las líneas con alto rendimiento y calidad de fruto adecuada, con resistencia a enfermedades y otros factores adversos, se utilizan como progenitores en programas de hibridación para generar nuevas variedades con potencial sobresaliente para aprovechar efectos genéticos como heterosis y aditividad (Melomey et al., 2019).

En las últimas décadas, el mejoramiento de tomate con fines comerciales se ha orientado hacia la producción y liberación de variedades híbridas F1, ya que el mejoramiento de tomates se lleva a cabo principalmente por el sector privado y solo de manera limitada por programas públicos. La evaluación fenotípica para la selección de nuevos cultivares implica una alta inversión financiera. Para hacerlo más productivo es necesario implementar estrategias, como la predicción del rendimiento de frutos, que permitan realizar y evaluar solo aquellos cruces con mayor probabilidad de éxito (Hernández et al., 2017).

Las herramientas biotecnológicas desarrolladas recientemente, como los marcadores genéticos de alta densidad y las técnicas estadísticas especializadas, junto con las evaluaciones fenotípicas de líneas destacadas, permiten identificar regiones del genoma asociadas a caracteres fenotípicos y así obtener predicciones de su rendimiento per se y como progenitores de híbridos.

Estas metodologías pueden reducir significativamente el tiempo y los costos requeridos para obtener nuevas variedades con alto rendimiento y calidad de fruto y otros atributos de interés agronómico (Schaid et al., 2018), especialmente aquellas asociadas a efectos aditivos.

Bajo este escenario, una de las técnicas que se utilizan actualmente es el GWAS (Genome wide association study), basado en el uso de marcadores moleculares de polimorfismo de nucleótido único (SNP), cuyo objetivo es asociar la variación genética detectada por marcadores moleculares con las evaluaciones fenotípicas de un rasgo específico de interés para el mejoramiento de cultivos (Schaid et al., 2018).

Estas técnicas abren un abanico de posibilidades que permiten acelerar el mejoramiento genético del tomate, debido a la posibilidad de aplicarlo a diversos rasgos cuantitativos, que son importantes en variedades comerciales, como la tolerancia a factores adversos bióticos y abióticos, y en especial, al fruto; calidad evaluada a través de diferentes rasgos (firmeza, color, forma y concentración de sólidos solubles, vitamina C y ácidos orgánicos, por nombrar algunos). Algunos de estos son caracteres favorables asociados al rendimiento del fruto, como la forma, el peso y el tamaño del fruto, que ya han sido asociados con SNPs altamente significativos. (Schaid et al., 2018).

2.2 Objetivos

2.2.1 General

Localizar regiones genómicas asociadas a caracteres fenotípicos de herencia cuantitativa de la calidad y rendimiento de fruto en 95 líneas homocigóticas de tomate mediante el análisis de asociación de genoma completo GWAS (Genetic wide análisis).

2.2.2 Particulares

Identificar y evaluar caracteres de herencia cuantitativas asociados a calidad y rendimiento de fruto en 95 líneas de tomate.

Realizar análisis GBS (Genotyping by Sequencing), con cadenas cortas de 75pb para el alineamiento de secuencias con respecto al genoma de referencia e identificar marcadores SNPs (Single Nucleotide Polymorphism) altamente significativos a través del análisis de GWAS (Genome-wide association study) de 95 líneas de *Solanum lycopersicum* asociados a caracteres cuantitativos de calidad y rendimiento de fruto.

Investigar SNPs significativos, ver su relación en el genoma de referencia y obtener conocimiento de su estudio.

2.3 Hipótesis

Mediante el uso de análisis de asociación del genoma completo (GWAS), se permite la identificación de SNPs asociados a genes cuantitativos que regulan la calidad y rendimiento de fruto de tomate.

2.4 Revisión de literatura

2.4.1 Cultivo de tomate y su importancia

En México el tomate (*Solanum lycopersicum* L.) ha sido una hortaliza muy apreciada desde la época prehispánica (Escofet et al., 2013; Medina et al., 2017), representa un alimento con propiedades beneficiosas para la salud humana por su contenido de compuestos bio-activos como el licopeno, el ácido ascórbico, el tocoferol, el β -caroteno, los ácidos fenólicos, los flavonoides, folatos, fibra, heterósidos cumarínicos (esculósidos), fitoeno y fitoflueno, vitamina E, ácidos grasos, carbohidratos, potasio, calcio, magnesio y otros microelementos como zinc, hierro y manganeso, entre otros (Ahmed et al., 2011; Perveen et al., 2015).

México se ha mantenido como eficaz productor y principal proveedor de tomate en el mercado internacional (SAGARPA, 2017) donde el valor de las exportaciones en 2012-2017 aumentó de 15.6 % a 58.1 % equivalentes de 1,001 a 1,583 millones de dólares (SAGARPA/SIAP, 2018).

Por otro lado, la demanda de dicho cultivo coloca a Estados Unidos como el principal mercado de destino de las exportaciones mexicanas con 90.1 %, lo cual posiciona a nuestro país como el décimo productor mundial de tomate con una producción de 3'469,707 t·año⁻¹ (SAGARPA SIAP, 2018).

2.4.2 Mejoramiento genético

La manipulación de la función génica es una herramienta básica en la obtención de cultivares mejorados, ha sido utilizada por los programas de mejoramiento genético tradicional a través de la incorporación de mutaciones naturales o inducidas en genotipos élitos. A pesar de ser exitoso en la obtención de numerosas variedades, este enfoque es impreciso, trabajoso y azaroso (Concepción, 2018).

El desarrollo de las técnicas de secuenciación de ADN unido al aumento de la capacidad de procesamiento de los sistemas informáticos ha favorecido la disponibilidad de los genomas y transcriptomas de un número siempre creciente de organismos (Concepción, 2018).

2.4.3 Marcadores genéticos

Los marcadores moleculares constituyen una herramienta poderosa que facilitan la selección de genotipos o poblaciones y la caracterización de la diversidad genética; así mismo, son de gran utilidad en la construcción y saturación de mapas genéticos (diagrama descriptivo de la ubicación de los genes en cada cromosoma) y en la identificación y aislamiento de genes y/o grupos de éstos. Lo anterior ha brindado información valiosa sobre la estructura genómica y la localización de genes mayores específicos que controlan caracteres de interés en diversas especies (Cornide, 2002).

2.4.4 Marcadores SNPs

Existen diferentes tipos de marcadores genéticos: los morfológicos, proteicos y los basados en polimorfismo del ADN. Actualmente los marcadores moleculares más utilizados son los SNPs (Single Nucleotide Polymorphism), ya que son abundantes (alta densidad) y pueden localizarse en forma uniforme a lo largo de todo el genoma de un individuo. Se caracterizan por ser estables genéticamente y poseer baja tasa de mutación ($\sim 10^{-8}$) y requerir una mínima cantidad de muestra para su identificación; sin embargo, se necesita de una gran cantidad de marcadores SNPs para obtener un nivel de información adecuada y confiable. La gran ventaja de los SNPs por sobre el resto de los marcadores, es la posibilidad de identificarlos en forma automatizada, de manera que pueden encontrarse millones de SNPs simultáneamente (Lindblad et al., 2000).

Cabe destacar que han sido fundamentales en el análisis de genes y descubrimiento de la base genética molecular de importantes características agronómico-industriales. En el 2013 se realizó una genotipificación completa de 2,815 accesiones consanguíneas de maíz mostrando la distribución de 681,257 marcadores de SNP en todo el genoma, en el que algunos SNP están vinculados a

los genes candidatos conocidos para el color del núcleo, la dulzura y el tiempo de floración (Romay et al., 2013). También se ha identificado un conjunto de 205,614 SNP en 31 genotipos de soya, lo que constituye un valioso recurso genómico para los programas de mejoramiento de soya (Lam et al., 2010).

Estos marcadores permiten realizar investigaciones reduciendo los tiempos de análisis y por ende, la disminución de los costos para el descubrimiento de genes de interés (Kumar et al. 2012).

Los SNPs son identificados mediante el alineamiento de las secuencias con respecto a un genoma de referencia. Esta se puede llevar a cabo fácilmente en genomas de plantas pequeñas con genomas de referencia disponibles, como en el caso de arroz (Yamamoto et al., 2010) o Arabidopsis (Ossowski et al., 2008), aunque la identificación de SNPs en genomas complejos sin un genoma de referencia como el trigo, cebada y avena se puede lograr mediante técnicas de secuenciación de nueva generación (Next Generation Sequencing, NGS) o incluso, usando datos transcriptómicos o representaciones de complejidad reducidas, las que son facilitadas por las técnicas de secuenciación masiva (Kumar et al., 2012).

2.4.5 Genotipificación por Secuenciación (Genotyping by Sequencing, GBS)

Actualmente, dentro de las estrategias para identificación de SNPs, se encuentran la creación de bibliotecas de genoma reducido de ADN, para luego secuenciarlas y realizar los análisis bioinformáticos respectivos. Una biblioteca o librería genómica, es una compilación de secuencias de ADN, en la que se espera que esté incluido el mayor número de fragmentos posibles de un genoma (Bruce et al., 1994).

Para realizar la digestión, se utilizan enzimas de restricción que generan cortes en regiones específicas del genoma, cuyo objetivo es eliminar regiones altamente repetidas y poco informativas, debido a que existen enzimas que son sensibles a las regiones metiladas y no son capaces de cortar en estos sitios (Davey et al., 2011).

Los genomas complejos, presentan muchas secuencias repetidas, por ejemplo, aproximadamente la mitad del genoma humano corresponde a repeticiones, estas surgen a partir de los mecanismos biológicos de replicación, donde surgen copias extras de un fragmento del genoma; sin embargo, algunas de estas repeticiones podrían ser funcionales, actuando incluso como marcadores moleculares (Davey et al., 2011).

Genotipificación por Secuenciación, es un método de creación de librerías genómicas, que permite secuenciar un gran número de individuos de forma simultánea, se diferencia de otras metodologías (como RAD-seq y RRL) debido a que en el GBS no se realiza una selección de fragmentos por tamaño (los más grandes) antes de la secuenciación. NGS, requiere una menor cantidad de ADN, además la secuenciación es bastante rápida, simple y reproducible. La creación de librerías genómicas mediante este método presenta un bajo costo de secuenciación (Davey et al., 2011).

Esta técnica requiere de dos tipos de adaptadores, los primeros corresponden a oligonucleótidos que son complementarios a la secuencia que se encuentra unido a la celda de flujo del secuenciador y los segundos son denominados código de barras (barcodes), que corresponden a adaptadores que permiten la identificación de las muestras. La amplificación de los fragmentos con los adaptadores ocurre mediante PCR y se realiza una purificación para eliminar restos de adaptadores y reactivos, los que son evaluados mediante una electroforesis en gel de agarosa y un espectrofotómetro PicroGreen antes de la secuenciación (Elshire et al., 2011).

En el 2016 fue lanzado el paquete estadístico NGSEP (Next Generation Sequencing Experience Platform), el cual fue publicado por el Dr. Jorge Duitama, creador del Software, en la revista MCGenomics, bajo el artículo "Bioinformatic analysis of genotype by sequencing (GBS) data with NGSEP" en el cual se menciona la eficiencia de la aplicación al hacer un análisis preciso y eficiente (Duitama, 2015).

NGSEP hace pruebas estadísticas a partir de datos de secuencia sin procesar generando e identificando marcadores SNPs, posteriormente se crea un archivo

con información relativa a: identificación del cromosoma, posición del marcador y calidad de la muestra. Esta funcionalidad reduce enormemente la cantidad de trabajo requerido para obtener conjuntos de datos de variación sobre las poblaciones ya que crea un filtro de datos de alta calidad con la información obtenida de los marcadores SNPs para la lectura de alineación y detección de variantes a través de diferentes muestras (Duitama, 2015).

2.4.6 Análisis de asociación del genoma (GWAS)

Los estudios de asociación genética tienen como objetivo identificar variación en el material genético de una población o especie que están relacionadas con una característica de interés.

El reciente desarrollo y disponibilidad de diferentes genotipos mediante protocolos de secuenciación (GBS), proporcionó un enfoque rentable para realizar análisis genómicos de alta resolución de poblaciones enteras en diferentes especies. El componente central de todos estos protocolos es la digestión del ADN con enzimas de restricción para generar fragmentos de secuenciación. (Duitama, 2015).

Los GWAS (Genome-wide association study) utilizan tecnologías de genotipificación a gran escala (Illumina o Affymetrix) para evaluar la asociación de cientos de miles de SNPs con diferentes variables. Actualmente existen diferentes software para analizar GWAS como PLINK, SNPTEST, snpMatrix, SNPpassco, pero el más actual y usado es Tassel y R. (Causse, 2016).

TASSEL es un software de uso común para el mapeo de desequilibrio de ligamiento (LD, Linkage disequilibrium) en plantas, actualizado con frecuencia con métodos recientemente desarrollados, incluyen teoría sobre el modelo lineal general y modelos de regresión múltiple. TASSEL también puede calcular y visualizar la gráfica del estadístico de desequilibrio de ligamiento, análisis de la estructura de la población mediante PCA (análisis de componentes principales) y diagramas de distancia genética de árboles. TASSEL puede manejar los marcadores SSR y SNP, pero la versión más reciente solo acepta SNP (Braulio et al., 2012).

2.5 Materiales y métodos

2.5.1 Genotipos evaluados

Los genotipos evaluados fueron 95 líneas avanzadas de tomate (F10 a F12) del Proyecto de Mejoramiento Genético del Tomate de la Universidad Autónoma Chapingo (UACH), México, generadas a partir de la segregación de poblaciones seleccionadas originadas de cruces entre híbridos comerciales.

2.5.2 Evaluación fenotípica

La evaluación fenotípica se realizó en 2018 en condiciones de invernadero en la Universidad Autónoma Chapingo (19 ° 29'35 " latitud norte, 98 ° 52 '19 " longitud oeste y 2267 msnm).

Los genotipos se sembraron en bandejas de poliestireno con turba como sustrato. El trasplante se realizó 35 días después de la siembra en invernadero bajo sistema hidropónico. Se utilizaron bolsas negras de polietileno de 18 L con sustrato de espuma volcánica. La solución nutritiva utilizada fue la propuesta por Cadahia (2000) para tomate; la cantidad aplicada varió según la etapa fenológica y las condiciones climáticas mediante un sistema de riego por goteo con goteros cada 50 centímetros. La densidad de plantas fue de 3,3 por m². La unidad experimental consistió en 2 macetas, conteniendo dos plantas cada una (4 plantas por parcela experimental). Se utilizó un diseño de bloques completos al azar con tres réplicas para evaluar los caracteres de calidad y 6 réplicas para los caracteres de rendimiento.

2.5.3 Caracteres evaluados

De tres frutos de cada réplica se extrajeron 30 g de pulpa, la cual se congeló a -20 ° C y posteriormente se determinaron los siguientes parámetros:

Acidez titulable (TA,%): Determinada según el método A.O.A.C. (1990). Para lo cual se neutralizaron 5 mL de filtrado de muestra de tomate con NaOH 0.1 N, se utilizó timolftaleína como indicador.

Vitamina C (VC, mg): Determinada realizando una curva tipo con concentraciones conocidas de ácido ascórbico y titulada con solución de Tillman según la metodología establecida por la A.O.A.C. (1990).

Firmeza (F, N): Se registró en la línea ecuatorial del fruto y en tres puntos diferentes. Se utilizó un penetrómetro digital de la marca Mecmesin CE.

Potencial de hidrógeno (pH): La medición se realizó mediante potenciometría utilizando un medidor de pH automático.

Sólidos solubles totales (SST, ° Brix): Determinado según el método descrito por la AOAC (1990), con un refractómetro digital marca ATAGO modelo PAL-1 con una escala de 0 a 53%.

Para caracteres fenotípicos, los caracteres evaluados fueron:

Distancia internodal (DI, cm), Diámetro polar del fruto (PD, mm), Diámetro ecuatorial del fruto (DE, mm), Peso del fruto (FW, g), Peso del fruto del racimo CW, g) y Número de frutos (NF).

2.5.4 Análisis de datos fenotípicos

Los componentes de la varianza, las heredabilidades, los mejores estimadores lineales insesgados (BLUE) y los mejores predictores lineales insesgados (BLUP) se obtuvieron con los valores de los datos fenotípicos utilizando Meta R (Multi Environment Trial Analysis with R para Windows) v.6.0.1 utilizando un diseño de bloques completos al azar. Las pruebas normales, los valores mínimos, máximo, medio y correlaciones se determinaron utilizando el paquete estadístico Meta R Software.

2.5.5 Análisis molecular

(Extracción de ADN)

Para realizar la extracción de ADN, los genotipos se sembraron en bandejas de poliestireno de dos cavidades con turba como sustrato. El tejido vegetal de hojas verdes y tiernas, sin daño aparente se obtuvo a los 35 días después de la siembra.

El tejido vegetal se pulverizó con nitrógeno líquido y se trituró. La extracción de ADN se realizó con el método CTAB (bromuro de alquiltrimetilamonio), para verificar que el ADN estaba intacto, limpio y sin contaminación, se corrió en un gel de agarosa a una concentración de 1.5% y se cuantificó en una nanogota Thermo Scientific, 2000/ Espectrofotómetro 2000c en el laboratorio de Marcadores Moleculares del Departamento de Ciencias de los Cultivos de la Universidad Autónoma Chapingo.

La digestión de las muestras se realizó utilizando las enzimas BglII y DdeI, obteniendo fragmentos de 75 pb, la preparación de las bibliotecas etiquetadas, la cuantificación y secuenciación se realizó con el equipo Illumina NextSeq 2x75 del Laboratorio Nacional de Genómica para la Biodiversidad (Langebio).

2.5.6 Mapeo y genotipificación de SNP

El genoma de referencia (*Slycopersicum* 390) se descargó de la página Phytozome, <https://phytozome.jgi.doe.gov/pz/portal.html>). Se utilizó el software NGSEP 4.0.1 y Bowtie 2.3.5.1 para alinear las lecturas con la referencia y descubrir SNP e indeles. Se utilizó el software NGSEP para la anotación funcional de las variantes, el filtrado y la conversión de VCF a otros formatos para su posterior análisis. Se utilizó el software Flapjack 1.19.09.04 para visualizar los genotipos de SNP a través de las muestras.

La estructura de la población se determinó mediante el software STRUCTURE 2.3.4 y el software iTOL v5 (<http://itol.embl.de>) se utilizó para la construcción y visualización de dendrogramas. Las distancias genéticas entre pares de genotipos se obtuvieron por medio de la matriz de parentesco NGSEP calculada a partir de identidad por estado (IBS). La estrategia de construcción de dendrogramas fue el algoritmo de unión de vecinos.

Los datos genotípicos se evaluaron utilizando TASSEL v5.2.60, el mapeo de asociación se calculó siguiendo el PCA (análisis de componentes principales) y GLM (modelo lineal general).

2.6 Resultados y discusión

Los programas de mejoramiento genético del tomate buscan genes o grupos de genes que puedan contribuir a la expresión de rasgos relacionados con la calidad nutricional del fruto y el rendimiento (Bhandari et al., 2016). La fruta se caracteriza por su calidad en el alto contenido de compuestos bioactivos como sólidos solubles (fructosa y glucosa), carotenoides, betacarotenos, licopeno, vitamina C y polifenoles (Lahoz et al., 2016), compuestos que se han demostrado tener posibles efectos beneficiosos sobre la salud, desde el punto de vista antioxidante, anticanceroso, antihipertensivo, anti mutagénico e inhibidor de la angiogénesis (Vinha et al., 2014).

El mejoramiento genético, especialmente en tomate (a pesar de su gran valor metodológico), tiene una corta historia de predicción de características de rendimiento. Esto posiblemente se deba a que el mejoramiento vegetal lo lleva a cabo el sector privado y, en cierta medida, los programas públicos. El coste de producción del tomate es muy elevado, especialmente si se elabora en invernadero. La evaluación de campo de nuevos cultivares implica una alta inversión financiera (Hernández I, L. et al., 2017).

En este estudio, analizamos 95 genotipos, que fueron seleccionados para representar una amplia gama de diversidad fenotípica en tomate. Se utilizaron datos fenotípicos y genómicos (genotipificación basado en SNP) para determinar las correlaciones entre variables, estructura poblacional y GWAS.

Se evaluaron 5 fenotipos relacionados con la calidad y 6 con el rendimiento en 95 líneas de tomate élite, con diferentes características como forma (redonda, alargada, en forma de riñón, cuadrada, en forma de pera), tamaño (pequeño, mediano, grande) tipo de crecimiento (determinado, semideterminado e indeterminado) y color (amarillo, naranja, rojo).

Los rasgos de calidad del fruto, pH y ° Bx mostraron los valores más bajos de varianza genotípica (0.68, 0.66) y Heredabilidad (0.93, 0.92). Para las variables de rendimiento ID 2-3 y NF fueron las que presentaron los valores más bajos de 0.55 para varianza genotípica en ambos casos y 0.88 para heredabilidad, mismo valor para ambas variables (Cuadro 1).

Cuadro 1. Las réplicas, la media, el coeficiente de variación (CV), la varianza genotípica y la heredabilidad de los datos fenotípicos se estimaron a partir del análisis de varianza utilizando un análisis de prueba en múltiples entornos con el software R (Alvarado et al. 2015).

Variable	Replicas	Media	CV %	Varianza Genotípica σ_g^2	Heredabilidad h^2
TA	3	6.51	4.4	0.91	0.98
VC	3	50.36	17.97	0.81	0.96
F	3	1.56	15.77	0.89	0.98
pH	3	4.88	2.49	0.68	0.93
°Bx	3	0.68	10.7	0.66	0.92
ID 2-3	6	24.4	18.34	0.55	0.88
NF	6	6.08	30.26	0.55	0.88
CW	6	797.39	24.29	0.61	0.9
FW	6	158.25	23.98	0.77	0.95
PD	6	63.88	8.28	0.8	0.97
DE	6	62.87	9.15	0.8	0.96

Acidez titulable (TA), Vitamina C (VC), Firmeza (F), Potencial de hidrógeno (pH), Sólidos solubles totales (° Bx), Distancia entre nodos del segundo al tercer racimo (ID 2-3), Número de frutos (NF), Peso del fruto del racimo (CW), peso del fruto (FW), diámetro polar del fruto (PD), diámetro ecuatorial del fruto (DE).

Los coeficientes de correlación de Pearson entre todos los rasgos revelaron que el pH está asociado negativamente con TA (Cuadro 2).

Cuadro 2. El coeficiente de correlación de Pearson entre los componentes de calidad y rendimiento.

	TA	VC	F	pH	°Bx	ID 2-3	NF	CW	FW	PD
TA	1									
VC	0.14	1								
F	-0.072	-0.175	1							
pH	-0.549**	-0.06	-0.051	1						
°Bx	0.431*	0.318*	-0.135	-0.373**	1					
ID 2-3	0.003	0.256*	-0.155	-0.023	0.317*	1				
NF	0.232*	0.157	-0.242**	-0.209**	0.205*	0.2	1			
CW	-0.021	0.023	0.190	0.073	0.011	0.249*	0.187	1		
FW	-0.314**	-0.23**	0.192	0.296*	-0.273**	-0.179	-0.696**	0.213	1	
PD	-0.249**	-0.148	0.300*	0.212*	-0.163	0.024	-0.303**	0.149	0.208*	1
DE	-0.283**	-0.135	0.170	0.271*	-0.264**	-0.171	-0.690**	0.205*	0.908*	-0.038

Los valores con ** y * indican correlaciones con $p < 0.01$ y $p < 0.05$, respectivamente. Acidez titulable (TA), Vitamina C (VC), Firmeza (F), Potencial de hidrógeno (pH), Sólidos solubles totales (° Bx), Diferencia entre los nodos del segundo al tercer racimo (ID 2-3), Número de frutos (NF), Peso del fruto del racimo (CW), Peso del fruto (FW), Diámetro polar del fruto (PD), Diámetro ecuatorial del fruto (DE).

La acidez titulable se expresa como porcentaje de ácido cítrico (ácido orgánico) y su valoración se realizó mediante técnica de titulación. El pH es una medida de la concentración de iones de hidrógeno que expresa el grado de acidez o alcalinidad. El pH es importante para el control de microorganismos en los alimentos y la acidez en los alimentos depende de las concentraciones de ácidos orgánicos en la fruta, de los cuales el predominante es el ácido cítrico, seguido del ácido málico y luego el ácido tartárico (Kader, 2008).

El número de frutos y el peso del fruto por racimo son un carácter importante que contribuye al rendimiento, la DE ayuda a determinar la forma del fruto, que puede ser alargada, aplanada, grande, mediana o pequeña. En este estudio, la NF también se asocia negativamente con FW y DE con un valor de -0,696 en ambos.

2.6.1 Resultados de la secuenciación del genoma y estructura de la población

Obtuvimos información genotípica de las líneas evaluadas siguiendo el protocolo Genotype-By-Sequencing, se generaron 22,127 marcadores SNP y las líneas fueron genotipadas en más del 90%, pero solo 14,328 SNPs fueron funcionales después de un estricto análisis de filtrado.

Se construyó un dendrograma con 14.328 marcadores SNP generando tres grupos. El grupo I está conformado por 44 líneas, este grupo está conformado mayoritariamente por líneas de crecimiento indeterminadas (93%), el segundo grupo está conformado por 24 genotipos en su mayoría heterogéneos, agrupados con líneas de crecimiento determinado e indeterminado, saladette, bola, frutos en forma de pera y de color amarillo y rojo.

Finalmente, el grupo III más alejado con 28 genotipos donde predomina el color rojo del fruto con un 96% (Figura 1).

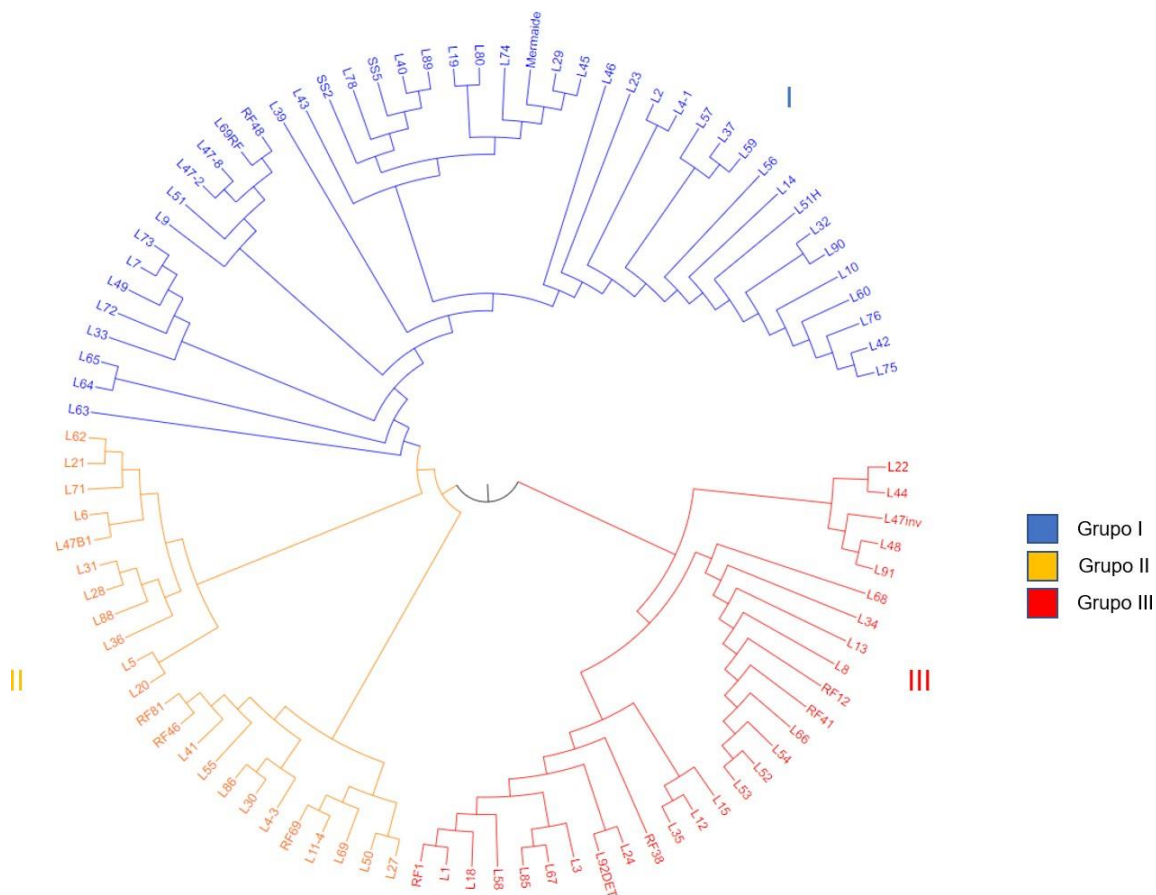


Figura 1. Árbol de enlace vecino basado en el cálculo de distancias genéticas desde la identidad por estado (IBS) estimadas a partir de SNP de alta calidad. El árbol se construyó con 95 líneas de tomate elite (*Solanum lycopersicum* L.), con 14.328 marcadores SNP. El dendrograma muestra 3 grupos, el primer (I) color azul donde se encuentran líneas de crecimiento mayoritariamente indeterminadas, el segundo (II) color naranja que contiene líneas mayoritariamente heterogéneas, y finalmente el grupo (III) mayoritariamente con líneas donde el color rojo del fruto predomina.

Los gráficos de barras de estructura poblacional generados por el software STRUCTURE en $K = 2$, 95 genotipos de Tomate agrupados en tres grupos genéticos o subpoblaciones utilizando el marcador de 14.328 SNP. Cada barra sólida representa un solo genotipo y diferentes colores (azul, rojo y una combinación de ambos) representan genotipos que pertenecen a tres subpoblaciones diferentes. Las proporciones de las barras de color representan las mezclas en las variedades (Figura 2).

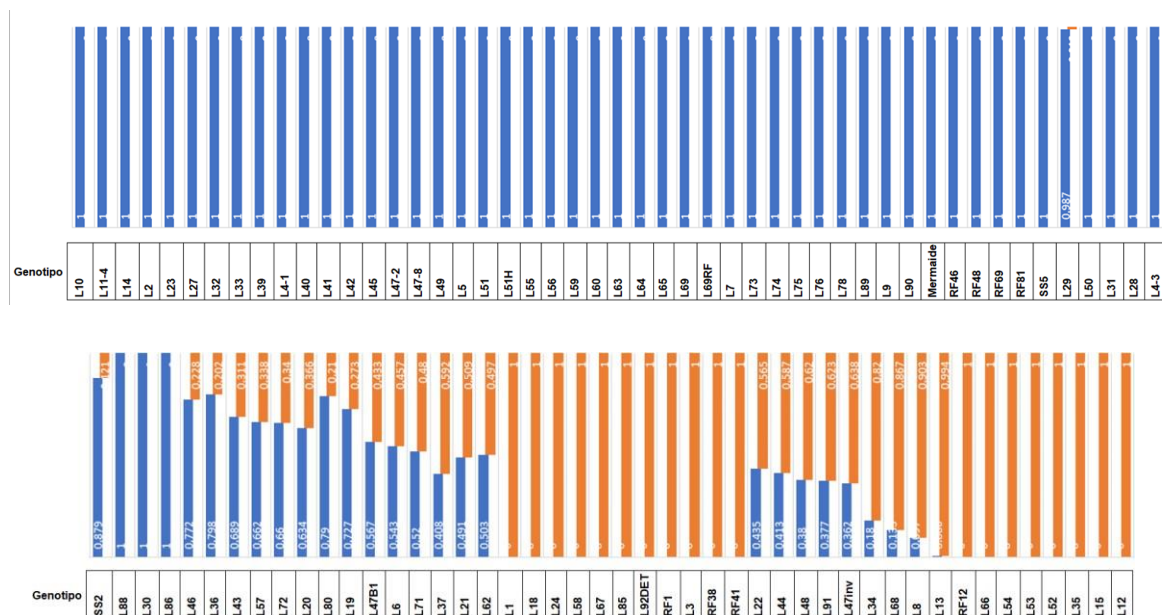


Figura 2. Estructura de la población: enfoque basado en modelos para el análisis de la estructura de la población K = 2.

De los 95 genotipos, 52 se agrupan en uno de los tres grupos representados en azul, 19 genotipos son del grupo en rojo y finalmente se mezclan 24 genotipos (obteniendo una tasa de heterocigotos del 24%) de los cuales el 56% están más influenciados por el grupo azul y el 46% tienen más influencia del grupo rojo.

El análisis de estructura mostró 3 subpoblaciones en el panel determinadas por el mejor K (K = 2). Los 95 genotipos de tomate evaluados son líneas élite. Se han mejorado mediante retro cruzamiento. Sin embargo, 20 de ellos todavía muestran algún grado de heterocigosidad.

2.6.2 Resultados de asociación

Los análisis de asociación se realizaron utilizando PCA y GLM con TASSEL para identificar señales de asociación. Se observó que, para los rasgos de rendimiento, las asociaciones más significativas fueron peso de fruto (FW), número de frutos por racimo (NF), peso de racimo (CW) y diámetro ecuatorial de fruto (DE).

Teniendo en cuenta estos 4 caracteres fenotípicos, tenemos un total de 88 asociaciones de marcadores SNP, de las cuales 61 están relacionadas con genes reportados en el genoma de referencia, 27 asociaciones no reportadas y 13 con

proteína desconocida. Los cromosomas 1 y 3 demostraron ser los más influyentes porque tenían marcadores significativos de 26 y 23 SNP, respectivamente.

El carácter NF tiene el mayor número de marcadores asociados, 44 marcadores en los cromosomas 2-4 y 7-10.

CW tenía 4 marcadores de SNP asociados y relacionados con NF, 2 marcadores de SNP en el cromosoma 3 y uno en los cromosomas 8 y 10.

El carácter FW tenía 31 marcadores SNP's con mayor influencia en el cromosoma 1 con 25 SNP's, 3 SNP's en el cromosoma 8 y finalmente un marcador en los cromosomas 2,5 y 12.

Finalmente, la DE tuvo 19 asociaciones, de las cuales 7 están relacionadas con FW en los cromosomas 1, 2, 5, 8 y 12. El cromosoma que tuvo mayor influencia en este carácter es el cromosoma 3 con 9 asociaciones (Figura 3).

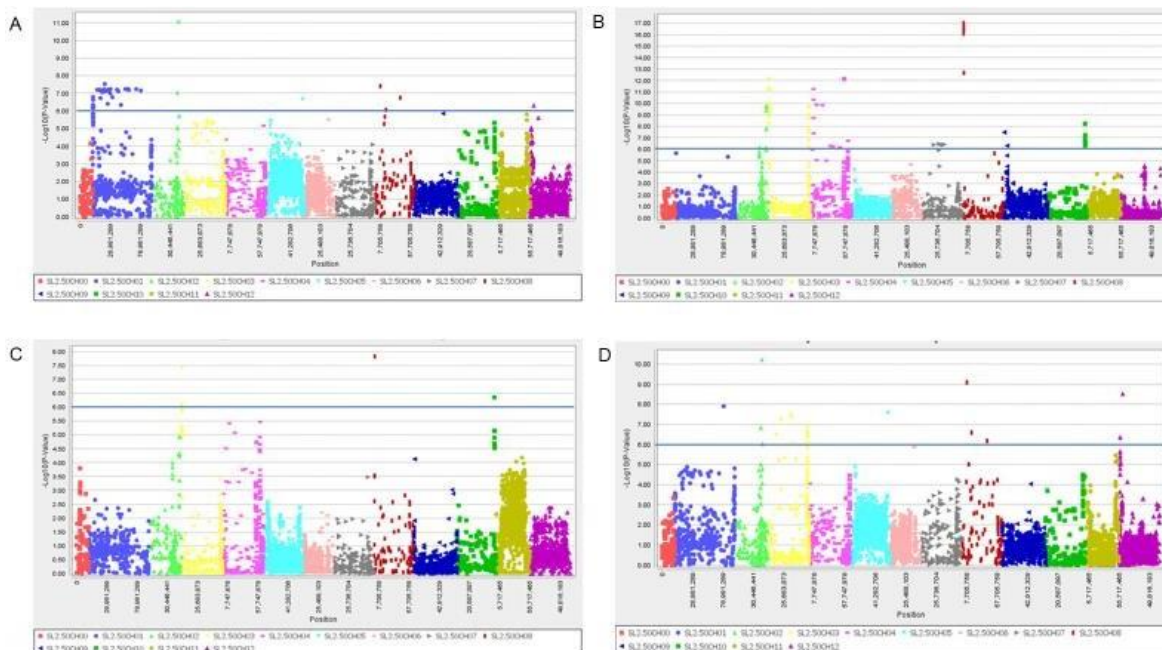


Figura 3. Gráficos de Manhattan para cuatro rasgos fenotípicos contruidos a partir de valores p de todo el genoma obtenidos de TASSEL v5.2.60 basados en un modelo lineal general ponderado con 95 líneas élite de tomate (*Solanum lycopersicum* L), con 14.328 marcadores SNP. Los genomas A, B, C y D son el peso del fruto, el número de frutos, el peso del racimo y finalmente el diámetro ecuatorial. En los gráficos de Manhattan, la línea azul indica el umbral para una asociación SNP significativa.

El peso de la fruta es un criterio de selección importante en los programas de mejora de muchos cultivos de frutas y hortalizas. Esta es una herencia cuantitativa que está controlada por hasta 28 *loci* de rasgos cuantitativos (QTL) (Grandillo et al., 1999). Los *loci* fw1.1, fw2.2, fw2.3, fw3.1, fw4.1 y fw9.1 explican cada uno más del 20% de varianza fenotípica. Además, fw1.1, fw2.1, fw2.2, fw3.1, fw3.2 y fw11.3 se han identificado en al menos cuatro estudios independientes (Grandillo et al., 1999). El QTL de peso de la fruta del cromosoma 2, número 2 (fw2.2), fue el primer gen asociado a QTL que se identificó y clonó (Alpert et al., 1995; Alpert y Tanksley, 1996; Frary et al., 2000).

Hay estudios donde el peso del fruto del tomate se asocia con el cromosoma 11, estudios posteriores refuerzan la asociación de este carácter fenotípico en el cromosoma 11. Sin embargo, también reflejan asociaciones significativas que mejoran los cromosomas 2 y 8 (Huang et al., 2011; Ruggieri et al. 2014; Sacco et al. 2015).

Para el cromosoma 2, Ruggieri et al. (2014) asocian el gen Solyc02g078790.2.1 2 en la posición 38.009.446 y Sacco, A. et al. (2015) asocian el gen Solyc02g083900 en la posición 47.148.187, en nuestro trabajo de investigación el gen asociado es Solyc02g080400.1 en una posición 44, 637,394, donde relativamente los genes están cerca.

En el trabajo de Valentino Ruggieri et al. (2014) el gen significativo para este carácter en el cromosoma 8 es Solyc08g006170.1.1 en una posición de 886,583-887435. En esta investigación tenemos un marcador asociado solo para FW en la posición 17,759,657 para el cual no hay información adicional en el genoma de referencia, por otro lado, tenemos dos asociaciones más en los mismos cromosomas que están relacionados con FW y DE en las posiciones 41,468,186 (sin información adicional) y 8.098.810 para el gen Solyc08g016630.

Barrantes W. et al. (2016) también estudiaron FW en tomate y encontraron 2 asociaciones QTL en el cromosoma 1 (fw1.1, fw1.2) en las posiciones 74,000 y 78,000. En el presente estudio, la mayoría de las asociaciones se obtuvieron en el cromosoma 1 con 26 de las 32 asociaciones, de las cuales 13 no se informan en

el genoma de referencia y solo una asociación en la posición 72,611,111 para el gen Solyc01g065920. 1.1 podría estar relacionado con el trabajo de Walter, por lo que podríamos considerar un descubrimiento importante con marcadores SNP relacionados con FW en este trabajo.

Adriana Sacco y col. (2015) también estudiaron el diámetro ecuatorial y tuvieron asociaciones en los cromosomas 2 y 11, en el cromosoma 2 con un marcador asociado al gen Solyc02g082030, en la posición 45,761,358, hemos encontrado que el gen Solyc02g081380.2 estaba asociado en la posición 45,365,766 , pues el mismo personaje que este es muy cercano al marcador que Adriana encontró en su obra para que pudieran estar relacionados.

Las proteínas de la familia OVATE (OFP) son una clase de proteínas con un dominio OVATE conservado. Esta proteína se identificó por primera vez en el tomate como un regulador clave de la forma de la fruta. (Wang et al., 2016).

El gen OVATE determina la conversión de frutos redondos a frutos con forma de pera en el tomate y es uno de los genes de forma de frutos del tomate más utilizados. Este se encuentra en el cromosoma 2 en la posición 48,300 (<https://solgenomics.net/locus/1094/view>), en esta investigación no se confirma la presencia del gen, pero sería importante en un estudio posterior confirmar su presencia o ausencia.

Sin embargo, en esta investigación el cromosoma 3 fue el más importante para este personaje con 9 de 12 asociaciones, ninguna de las cuales ha sido reportada y 3 de ellas tampoco están reportadas en el genoma de referencia. Por lo que también podría considerarse una contribución importante al carácter DE.

La FQ siempre estuvo asociada con NF en los cromosomas 3, 8 y 10, dos asociaciones en el cromosoma 3, en la posición 352 y 397 corresponden a los genes Solyc03g005460.2 y Solyc03g005520.1.1. En el cromosoma 8 existe una Asociación que corresponde al gen Solyc08g006650.2 y finalmente en el cromosoma 10 asociado al gen Solyc10g081070.1.1. Las asociaciones son muy

significativas; sin embargo, los genes aún no están relacionados con estos caracteres en ninguna literatura citada (Tabla 3).

NF mostró más asociaciones (44), 8 no tienen información en el genoma de referencia y 6 son genes con proteína desconocida.

El cromosoma 3 fue el más significativo con 12 asociados. El siguiente fue el cromosoma 4 con 11 genes, dos de ellos no están asociados a ningún gen. El cromosoma 7 con 5 asociaciones, pero ninguno de ellos relacionado con un gen, los cromosomas 8 y 10 tenían cada uno 3 genes asociados. Finalmente, el cromosoma 2 tenía cuatro genes asociados y el cromosoma 9 con dos genes asociados (uno que no está asociado a ningún gen).

El NF ha sido un carácter estudiado para incrementar el rendimiento, generalmente se ha incrementado el rendimiento de frutos, modificando uno o varios de los componentes primarios o secundarios: número de frutos / planta, peso medio de fruto, número de inflorescencias / planta, peso medio de fruto, número de inflorescencias / planta, flores / inflorescencia, número de lóculos / fruto o peso medio del lóculo (Vallejo, 2021). Sin embargo, no se han determinado los genes relacionados con este carácter.

Aquí hemos estudiado esta variación genética mediante la secuenciación del genoma de una selección de accesiones representativas de tomate pertenecientes al programa de mejoramiento de la Universidad Autónoma Chapingo, proporcionando un recurso ampliado para estudios genómicos comparativos. También presentamos resultados altamente confiables y la identificación de asociaciones para variables de rendimiento especialmente. Este estudio proporciona un conjunto de datos invaluable para estudios avanzados sobre las relaciones de los rasgos de secuencia y los mecanismos moleculares de la evolución del genoma del tomate, así como el desarrollo de enfoques genéticos para la mejora genética a través de la secuenciación.

2.7 Conclusiones

Se generaron un total de 14.328 *loci* SNP a partir de 95 líneas de tomate, y el enfoque GWAS reveló 88 marcadores asociados con 4 caracteres de rendimiento fenotípico. De los 88 SNP significativos, 44 se asociaron con el número de frutos. Nuestros resultados confirmaron la fuerte participación de las regiones genómicas del cromosoma 8 estando presentes en 4 caracteres (FW, NW, DE y CW), así como los cromosomas 2, 3 y 10 con asociaciones en al menos 2 de los 4 caracteres.

La importante participación del cromosoma 1 para FW y del cromosoma 7 para el carácter NF con asociaciones que no están vinculadas a ningún gen reportado puede usarse como referencia para futuras investigaciones y definir la participación de genes en los cromosomas de estos caracteres.

Finalmente, los cromosomas 5, 9 y 12 que se asociaron a un carácter, contribuyendo a un mejor conocimiento de los genes menores que definen la determinación del peso del fruto, número de frutos, peso del racimo y diámetro ecuatorial en el germoplasma de tomate de la Universidad Autónoma Chapingo. Esta información es importante y podría proporcionar una estrategia esencial para los productores de tomates.

2.8 Literatura citada

- Ahmed, L., Martin, D., Rico A.B., Barry, R.D. (2011). The antioxidant properties of whey permeate treated fresh-cut tomatoes. *Food Chemistry*, 124: 1451-1457.
- Alpert, K.B., Grandillo, D., Tanksley, S.D. (1995). *fw2.2*: a major QTL controlling fruit weight is common to both red- and green-fruited tomato species. *Theor Appl Genet.* 91:994–1000.
- Alpert, KB, Tanksley SD. (1996). High-resolution mapping and isolation of a yeast artificial chromosome contig containing *fw2.2*: a major fruit weight quantitative trait locus in tomato. *Proc Natl Acad Sci USA.* 93:15503–15507.
- A.O.A.C. (1990) Official Methods of Analysis. 15th Edition, Association of Official Analytical Chemist, Washington DC.
- Barrantes, W., López C.G., García, M.S., Alonso, A., Rubio, F., Ruiz, J.J., Fernández, M.R., Granell, A., Monforte A.J. (2016). Exploring New Alleles Involved in Tomato Fruit Quality in an Introgression Line Library of *Solanum pimpinellifolium*. *Front Plant Sci.* 17;7:1172.
- Bhandari, S.R., Cho, M.C., Lee, J.G. (2016). Genotypic variation in carotenoid, ascorbic acid, total phenolic, and flavonoid contents, and antioxidant activity in selected tomato breeding lines. *Hortic. Environ. Biote.*; 57(5):440-452.
- Braulio, J., Soto, C., Sylvie, C., (2012). Association Mapping in Plant Genomes, Genetic Diversity in Plants, Prof. Mahmut Caliskan (Ed.), ISBN: 978-953-51-0185-7.
- Bruce, A., Alexander, J., Julian, L., Martin, R., Keith R., Peter W. (1994). *The Molecular Biology of the Cell*. New York: Garland Science. ISBN-10: 0-8153-4072-9.
- Causse, M., Duffe, P., Gomez, M.C., Buret, M., Damidaux, R., Zamir, D. (2004). A genetic map of candidate genes and QTLs involved in tomato fruit size and composition. *J Exp Bot.*;55: 1671–1685. doi:10.1093/jxb/erh207.
- Concepción H.M., (2018). CRISPR/Cas: aplicaciones y perspectivas para el mejoramiento genético de plantas. Instituto de Biotecnología de las Plantas. UCLV. MES. eISSN 2074-8647, RNPS: 2154.
- Cornide, H.M.T. (2002). Marcadores Moleculares: nuevos horizontes en la genética y la selección de las plantas. La Habana: Editorial Félix Varela, Pp 366.
- Davey, J. W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M., Blaxter, M.L. (2011). Genomewide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, 12(7), 499–510.
- Duitama, J., Silva, A., Sanabria, Y., Cruz, D.F., Quintero, C., Ballen, C., (2015). Whole genome sequencing of elite rice cultivars as a comprehensive information resource for marker assisted selection. *PLoS One.*; 10.1371/ 0124617.
- Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*6:e19379 10.1371/0019379.
- Escofet, T.R. (2013). Importancia de la gastronomía prehispánica en el México actual. *Culinaria*, 6: 23-36.

- Frery, A., Nesbitt, T.C., Frery, A., Grandillo, S., van der Knaap E., Cong, B., Liu, J., Meller, J., Elber, R., Alpert, K.B. (2000) *fw2.2*: a quantitative trait locus key to the evolution of tomato fruit size. *Science* 289:85–88.
- Garrido, G.T. (2013). Métodos de análisis genéticos, espaciales y de conectividad en genética del paisaje Genetic, spatial and connectivity methods of analysis in landscape genetics.1031–1054. doi:10.7550/rmb.32500.
- Hernández I, L., Sahagún C, J., Rodríguez P, J E., Peña O, M.G. (2017). Prediction of fruit yield and firmness of tomato hybrids with BLUP and RR-BLUP using ISSR molecular markers.;23: 21–33.
- Huang, Z., van der Knaap, E., (2011). Tomato fruit weight 11.3 maps close to fasciated on the bottom of chromosome 11. *Theor Appl Genet.*;123: 465–474.
- Kader, A.A., (2008). Perspective Flavor quality of fruits and vegetables. *Journal of the Science of Food and Agriculture* 88:1863-1868.
- Kumar V., Akinleye A., Makkar P. S., Angulo M. A., Becker K., 2012. Growth performance and metabolic efficiency in Nile tilapia (*Oreochromis niloticus* L.) fed a diet containing *Jatropha platyphylla* kernel meal as a protein source. *Jornal of Animal Physiology and Animal.Nutrition.*, 96 (1): 37-46.
- Lam, H.M., Xu, X., Liu, X., Chen, W., Yang, G., Wong, F.L. (2010). Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nature Genetics*. 42,1053–1059.
- Lahoz, I., Pérez C, A., Valcárcel, M., Macua, J.I., Beltrán, J., Roselló S., Cebolla C, J., (2016). Effect of water deficit on the agronomical performance and quality of processing tomato. *Sci. Hort.*;200(1):55-65.
- Lindblad, T.K., Tanenbaum, D.M., Daly, M.J., Winchester, E. (2000). Loss of heterozygosity analysis of small cell lung carcinomas using single nucleotide polymorphism arrays. *Nature Biotechnology*. 18: 1001-1005.
- Medina, S.T., Arroyo, F.G., Dzul C, J.G. (2017). Origin and evolution of tomato production *Lycopersicon esculentum* in México. *Ciência Rural*, 47: e20160526.
- Melomey, L., Danquah, A., Offei, S., Ofori, K., Danquah E, O.M. (2019). Review on Tomato (*Solanum lycopersicum*, L.) Improvement Programmes in Ghana. 10.5772/intechopen.75843.
- Ossowski, S., Schwab, R., Weigel, D., (2008). Gene silencing in plants using artificial microRNAs and other small RNAs. *Journal Plant* 53: 674–690.
- Perveen, R., Suleria, H. A. R., Anjum, F. M., Butt, M. S., Pasha, I. and Ahmad, S. (2015). Tomato (*Solanum lycopersicum*) carotenoids and lycopenes chemistry; metabolism, absorption, nutrition, and allied health claims-A comprehensive review. *Critical Reviews in Food Science and Nutrition*, 55: 919-929.
- Romay, M.C., Millard, M.J., Glaubitz, J.C., Peiffer, J.A., Swarts, K.L., Casstevens, T.M. (2013). Comprehensive genotyping of the USA national maize in bred seed bank. *Genome Biology*. 14,R55.
- Ruggieri, V., Francese, G., Sacco, A., D'Alessandro, A., Rigano, M.M., Parisi, M. (2014). An association mapping approach to identify favourable alleles for tomato fruit quality breeding. *BMC Plant Biol.*;14.

- Sacco, A., Ruggieri, V., Parisi, M., Festa, G., Rigano, M.M., Picarella, M.E. (2015). Exploring a tomato landraces collection for fruit-related traits by the aid of a high-throughput genomic platform. *PLoS One*.;10: 1–20.
- SAGARPA (Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación) (2017). Planeación agrícola nacional 2017-2030; Jitomate mexicano. SAGARPA, México. 15 p. <https://www.gob.mx/cms/uploads/attachment/file/257077/Potencial- Jitomate.pdf>.
- SAGARPA, (2018). Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación / SIAP (Servicio de Información Agroalimentaria y Pesquera).
- Schaid, D.J., Chen, W., Larson, N.B., (2018). From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat Rev Genet.*;19: 491–504.
- Vallejo C, F.A. (2021). Avances en la producción de líneas de tomate *Solanum lycopersicum* Mill. con resistencia genética al perforador del fruto *Neoleucinodes elegantalis* (Guenée 1854) (Lepidoptera: Crambidae). *Rev. Acad. Colomb. Cienc. Ex. Fis. Nat.*, 45(174), 136-146.
- Vinha, A.F., Alves R.C., Barreira P, S.V., Castro A., Costa G, A.S., Oliveira M, B.P. (2014). Effect of peel and seed removal on the nutritional value and antioxidant activity of tomato (*Lycopersicon esculentum* L.) fruits. *LWT - Food Sci. Technol.*;55(1):197- 202.
- Wang, S., Chang, Y., Ellis, B. (2016). Overview of OVATE FAMILY PROTEINS, A Novel Class of Plant-Specific Growth Regulators. *Frontiers in plant science*, 7, 417.
- Yamamoto, T., Mochida, J., Kadota, J., Takeda, M., Bi, E., Tanaka, K. (2010). Initial polarized bud growth by endocytic recycling in the absence of actin cable dependent vesicle transport in yeast. *Molecular Biology Cell* 21(7):1237-52.