

## RESUMEN

La mejora de las especies cultivadas es un área dinámica, sujeta a las necesidades y requerimientos de los diferentes cultivos en cada momento y de las exigencias de productores y consumidores. Una limitación en este proceso de mejora es la diversidad genética existente en los cultivos. Muchas de las especies cultivadas se han visto sometidas a diversos cuellos de botella poblacionales durante su domesticación, ocasionando una notable pérdida de diversidad genética. En estos casos, las especies silvestres emparentadas revisten un especial interés, pudiéndose utilizar como fuente de genes en la mejora genética. Esta circunstancia se da en el caso del tomate cultivado (*Solanum lycopersicum* var. *lycopersicum*, SLL), una de las hortalizas con mayor producción a nivel mundial. Esta especie surgió a partir de *Solanum lycopersicum* var. *cerasiforme* (SLC) que a su vez fue pre-domesticada a partir de *Solanum pimpinellifolium* (SP). Pese al potencial para la mejora de estas especies debido a su mayor diversidad genética y proximidad al tomate, su uso está limitado en gran medida por la falta de información de las colecciones mantenidas en los bancos de germoplasma, principalmente en SLC. Es por ello necesario abordar estudios exhaustivos que hagan posible disponer de una información amplia y detallada de la diversidad de estas especies presente en el centro de origen y difusión del tomate cultivado, con el fin de ser utilizadas de una forma eficiente en mejora y éste es, por tanto, el objetivo general de esta tesis.

Para lograr el objetivo anteriormente expuesto se ha empleado una colección de germoplasma compuesta por 15 accesiones de SLL, procedentes de México; 27 de SP, procedentes de Perú y Ecuador y 121 de SLC, procedentes de Perú, Ecuador, México y Mesoamérica. Esta colección ha sido sometida a una extensa caracterización fenotípica y genética, lo que ha permitido identificar las regiones de genoma asociadas a caracteres fenotípicos mediante estudios de asociación del genoma completo (GWAS). Además, se han creado poblaciones segregantes F2 para cada una de las entradas de la colección a partir de los híbridos obtenidos con tres parentales diferentes, y estos cruces se han puesto a disposición de la comunidad científica. Este recurso permitirá el estudio del control genético de mutantes de interés en mejora. Esta utilidad se explora en el último capítulo de la tesis, donde se estudia el control genético de la densidad de tricomas en una accesión de la especie SP, característica de interés por conferir resistencia a artrópodos.

El estudio realizado con los datos morfológicos ha demostrado la presencia de una gradación morfológica continua entre los tres grupos taxonómicos, y ha permitido la identificación de una serie de caracteres que permiten la diferenciación entre grupos y también entre las distintas procedencias geográficas dentro de cada grupo taxonómico. Los datos analizados indican que, de entre todos los caracteres analizados, el aumento en el tamaño del fruto y la variación en su forma, así como la posición del estigma de las flores, serían los principales cambios asociados a la domesticación de la especie y, por lo tanto, diferencian a la especie silvestre de SLC y SLL. El análisis ha revelado también la existencia de variabilidad dentro de los grupos geográficos de la especie silvestre, conforme se fue expandiendo desde Perú a Ecuador, como consecuencia de su adaptación a diferentes ambientes. Por otro lado, SLC se caracteriza por una enorme variabilidad morfológica en caracteres de planta, inflorescencia y fruto, lo que la convierte en una variedad con un gran potencial para la mejora.

Tras esta caracterización morfológica, se exploró la diversidad genética que presentaba la colección. Para ello, se emplearon las secuencias genómicas públicas de

estos materiales obtenidas dentro del marco del proyecto VARITOME, dentro del que se enmarca esta tesis. Los diferentes polimorfismos de nucleótido único (SNPs) fueron anotados con base a la predicción de sus efectos en la secuencia codificantes mediante el programa SnpEff y se llevaron a cabo estudios de GWAS. El análisis genético reveló a Perú y Ecuador como las regiones con mayores niveles de diversidad, tanto en la especie SP como en SLC. Se observó una reducción en el número de variantes SNP en SLC México, que concuerda con la hipótesis de que Mesoamérica sea centro de domesticación y difusión del tomate cultivado. Por último, se constató el proceso de domesticación como la causa de la menor diversidad presente en la especie cultivada. Los estudios GWAS permitieron la identificación de correlaciones genotipo-fenotipo, revelando la asociación entre 107 SNPs y ocho caracteres cuantitativos y un total de 30 SNPs asociados a 7 caracteres cualitativos. Una parte de los SNPs detectados fueron localizados cerca de regiones genómicas ya asociadas con genes y QTLs relacionados con el carácter, como las asociaciones entre el peso del fruto con los QTLs ya descritos como *fw2.2* ó *fw9.2*. Otra parte se localizaron en regiones con posibles genes candidatos anotados, como fue el caso del color amarillo del fruto, lo que refuerza la validez de la colección para la detección de caracteres de interés agronómico. Por último, el resto de SNPs detectados se corresponden a regiones no descritas previamente, lo que abre el camino a estudios para la detección de nuevos genes candidatos.

La disponibilidad de datos de caracterización morfológica y genética revalorizan el conjunto de materiales estudiados. Con el objeto de aumentar su utilidad, se llevó a cabo el desarrollo de familias segregantes que facilitan el estudio de caracteres de interés por parte de la comunidad científica. Para ello, se cruzó toda la colección con una entrada representativa de cada uno de los tres grupos taxonómicos, SP, SLC y SLL; autofecundando las F1 conseguidas para obtener las poblaciones F2. Estas familias segregantes constituyen una poderosa herramienta como fuente de variantes alélicas naturales, al tiempo que permiten la caracterización de mutantes y la identificación de los genes implicados de una manera rápida y eficiente al disponer tanto de la caracterización genética como de las poblaciones F2.

La utilidad de dichas familias se demuestra en el último capítulo de esta tesis mediante el estudio de la base genética de la densidad de tricomas tipo IV. Durante la caracterización de la colección, se detectó una entrada de la especie SP (BGV016047) con una alta densidad de tricomas glandulares de tipo IV, en los que se ha descrito la presencia de distintos tipos de sustancias relacionadas con resistencia a algunas plagas. En esta tesis se muestra la caracterización fenotípica de la densidad de tricomas en función de la edad de la planta y de la hoja. Por otro lado, se estudia el control genético de la presencia y alta densidad de tricomas tipo IV en dos fondos genéticos diferentes, SLL x SP y SP x SLC. En ambos fondos se han detectado dos QTLs principales en los cromosomas 9 y 11. Además, se han detectado señales en los cromosomas 2, 5, 6, 7 y 8, en función de la familia segregante estudiada. Las regiones detectadas en los cromosomas 9 y 11 habían sido previamente descritas en otras especies por regular la densidad de este tipo de tricomas o los niveles de acilazúcares. En ambas regiones hay genes anotados que intervienen en el desarrollo de los tricomas o en la formación de acilazúcares, como aciltransferasas, glicosiltransferasas o los factores de transcripción MYB.