PhD THESIS
David Alonso Martín

Linking genetic resources, genomes and phenotypes of Solanaceous crops

2023

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

COMAV

G2P-SOL

Linking genetic resources, genomes and phenotypes of Solanaceous crops

PhD THESIS
David Alonso Martín

Advisors:
Dr. María José Díez Niclós
Dr. Jaime Prohens Tomás
Dr. Santiago Vilanova Navarro

VALENCIA, SEPTEMBER 2023

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Linking genetic resources, genomes and phenotypes of Solanaceous crops

Ph.D. dissertation by

David Alonso Martín

Advisors

Dr. María José Díez Niclós

Dr. Jaime Prohens Tomás

Dr. Santiago Vilanova Navarro

A DISSERTATION FOR THE DEGREE OF DOCTOR OF PHILOSOPHY WITHIN THE BIOTECHNOLOGY PROGRAM OF THE UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Valencia, September 2023

Alto soto de torres que al ponerse tras las encinas que el celaje esmaltan dora a los rayos de su lumbre el padre Sol de Castilla; bosque de piedras que arrancó la historia a las entrañas de la tierra madre, remanso de quietud, yo te bendigo, ¡mi Salamanca!

Fragmento 'Mi Salamanca', Miguel de Unamuno
Versión musical: Música para un códice salmantino, Joaquin Rodrigo

Table of Contents

Acknowledgements

Ha pasado mucho tiempo que comencé la aventura del doctorado, un camino largo y duro, en el que hubo un tiempo en el que me flojearon los ánimos y pensé que no llegaría a conseguirlo. Sin embargo, aquí estamos, yo escribiendo y tu leyendo. Por ello, pido perdón porque me voy a extender ya que tengo la suerte de poder agradecer el apoyo a grandes personas y científicos que no me dejaron caer y quienes me han acompañado en algún periodo de este camino.

Quisiera comenzar los agradecimientos con mis directores, **María José**, **Jaime** y **Santi**. Siento un gran respeto y admiración por vosotros, tanto en lo profesional como en lo personal. Estaré eternamente agradecido por darme la oportunidad que me brindasteis de trabajar en el proyecto G2P-SOL. Me habéis formado, guiado, corregido, animado y sobre todo apoyado en todo el proceso. No hay hornazos ni dulces suficientes para compensaros.

María José, es un privilegio aprender de ti y trabajar contigo. Sin duda, eres inspiradora. Gracias por darme acceso al maravilloso mundo de los recursos fitogenéticos. Estoy muy agradecido de poder haber compartido café y dulces acompañados de charlas de toda índole. Espero que tras tu jubilación podamos repetirlo. Gracias también a tus hijas, quienes tanto me ayudaron con el trasiego de semillas.

Jaime, gracias por darme la libertad de investigar y ayudarme a aprender a no perder el rumbo. Eres un pozo de sabiduría y un auténtico líder. Gracias por las oportunidades que me das. Espero que podamos seguir colaborando y crear anécdotas de viajes.

Santi, gracias por permitirme ser tu Oompa Loompa de la ciencia. Trabajar contigo en el laboratorio es apasionante y divertido. Gracias por recordarme día tras día lo que me gusta la biología. Tu método formativo creo que es el necesario en un laboratorio: diseñar un experimento, probar, fallar e investigar todas las partes del proceso de manera autónoma hasta llegar a discernir la solución. Gracias por formarme como científico.

Quiero también dar las gracias a mis compañeros de los laboratorios de Berenjen@s, Tomatitos, Melones, Pimenteros y del Banco de germoplasma (pasados y presentes): **Andrea A., Andrea S., Gloria, Marina M., Marina S., Joan, Jose, Ana, Edgar, Ximo, Giulio, Juan P., Juan B., Javi, Silia, Silvia, Martín, Oussama, Sandra, Esther, Loles, Mariola, Pietro, Cris C., Elena S., Resu, Leandro, Elena R., María F., Estela, Marisa, Neus, Ana A., Carla, Alicia S., Cristina E., María L., Clara, Cris S., Eva S., Alba R., Estefanía.**... sois muchos los que habéis contribuido de alguna u otra manera a esta tesis y que pasarme horas en el laboratorio no haya sido tan pesado.

Es justo destacar a **Mariola** y **Pietro**, gracias por cada uno de vuestros consejos que he intentado seguir fielmente; a **Esther** por ser una fantástica compañera de fenotipado y DJ; a **Cris C.** y **Elena S.** por sus risas y ayuda en todo momento; a **Carla** por ser la mejor compañera de despacho y abrazos; y a **Gloria** y **Andrea A.** por ser las mejores coach motivacionales.

También quiero agradecer a todas y cada una de las personas que conforman el COMAV, especialmente a **María José Pérez** y **Pascual** quienes siempre han estado ahí para mí.

Por otro lado, fuera del ámbito académico tengo que agradecer a las maravillosas personas que desde hace años están conmigo interesándose por mi tesis.

A los amigos que me dio la Universidad de Salamanca: **Jorge, Joana, María, Bea, Noemí, Mario, Pablo, Yohana, Lucía, Tess, Javi, Fredo, Alba (x2), Rodri (x2), Carmen, Becky** y **Víctor**. Aunque la vida nos ha llevado por caminos diferentes, os siento muy cerca.

A **Alba G.V.,** por estar siempre disponible para tomar una cerveza y hablar.

A mis amigas mejoradoras: **Pilar, Carolina** y **Cris S.** Por las cerve-skype, las cenas turrón-tequila, las innumerables clases de fitness y por las presentaciones oficiales de cada uno de nuestros tiestos en casa.

A mis amigas de Encinas: **Lucía, Rebe, Leti, Patri, Sara, Natalia** y **Rocío.** Por devolverme cada verano a la realidad del pueblo y sus pasodobles.

Al **Orfeó Universitari de València**, en especial a **Pilar, Paloma, Olga, Alba, Alexandra, Nuria, Andrea, Elena** y **Cloe** por permitirme una vía de escape de las largas jornadas de laboratorio y poder desahogarme mediante el canto.

A la bicicleta: **Cris, Carmen, Eli, Carlinhos, Patri, Àlex** y **Diego.** Por haber vivido esta tesis de primera mano, sábado tras sábado. Espero que os haya enseñado a valorar la ciencia.

A **Lorena y Sergio**, por tenerme siempre presente. Nunca unas huellas de dinosaurio unieron tanto.

Para el final he dejado a las personas más importantes y las culpables de la realización de esta tesis, mis padres. Cuando empecé la tesis siempre os imaginé sentados durante la defensa en primera fila. Sin embargo, la vida no siempre es como se quiere. **Mamá,** sé que te hubiera encantado tener la tesis en tus manos y leer lo agradecido que estoy de la educación que me has dado, por escucharme, apoyarme y enseñarme lo que significa la resiliencia. **Papá,** gracias por enseñarme lo que es el esfuerzo del trabajo y la responsabilidad. Gracias por trabajar tanto para que no me faltara de nada. Tus noches sin dormir para ir a regar el campo y las largas jornadas en el almacén han dado como resultado esta tesis y mi formación.

A mi familia (Salamanca, Madrid y Valencia), me siento muy afortunado de teneros. Prima **Yara**, gracias por ser mi mentora, la otra bióloga de la familia y mi más fiel confidente. Tía **Charo**, gracias por cuidarnos y permitir que estuviera más tranquilo durante la tesis. Tía **Cati**, gracias por tus ánimos y preocuparte de tu ahijado.

Finalmente, me gustaría dedicar esta tesis a **Miguel Ángel**, mi compañero de vida y aventuras. Gracias por tu apoyo incondicional, tus abrazos son la energía que he necesitado para acabar esta tesis. Te debo mucho tiempo. Ahora comienza nuestra aventura (lo dejo por escrito para que me puedas reclamar en un futuro).

Para terminar, aquí os dejo la playlist de la tesis. Cada canción me recuerda a cada uno de vosotros, disfrutad.

https://open.spotify.com/playlist/2fAnBph6E7PV4EWtvt5gPm?si=ddd13d16383b4eb7

ABSTRACT

The impact of climate change on horticultural crops is increasingly evident, leading to drastic loss and erosion of genetic diversity. This poses significant challenges for crop improvement, which requires the exploration of plant genetic resources conserved in germplasm banks and the development of technologies that allow the evaluation of the phenotypic and genotypic value of these materials. However, the current situation of germplasm collections is characterized by the existence of unidentified duplicates among collections, taxonomic mislabelling, insufficient and unavailable documentation for researchers and breeders, and a lack of funding for proper conservation and management. This greatly hampers the utilization of these resources. This thesis addresses this problem by starting with the unification of passport, phenotyping, and image data from the main collections of tomato, pepper, and eggplant into a single repository. Genotyping the collection and jointly analysing the generated genotyping data has enabled the construction of core collections that have been phenotyped in detail, thereby increasing knowledge of genotypic and phenotypic variability and, consequently, enhancing their utilization by researchers and breeders.
In the first chapter, an inventory, standardization, and review of available passport and phenotypic data of tomato, pepper, and eggplant accessions conserved in major European and non-European germplasm banks were conducted to improve the efficiency of plant genetic resource management.

The second chapter focuses on the development and optimization of a high-quality, fast, and cost-effective genomic DNA extraction method that combines the advantages of the CTAB-based extraction method with nucleic acid purification on a silica matrix. It is a universal method that can be used for different species and tissues. The efficiency of the resulting genomic DNA was evaluated on different sequencing platforms, such as Single Primer Enrichment Technology (SPET) and Oxford Nanopore, yielding promising results. This facilitates the prerequisite step of DNA extraction before genotyping the collections.

Chapter three addresses the genotyping of the collections. The high number of accessions for each crop, particularly tomato, poses an often insurmountable economic problem. Therefore, chapter three is focused on evaluating the potential of SPET sequencing technology, which is more cost-effective than other known methods, for high-throughput genotyping of tomato and eggplant germplasm collections. SPET is a targeted genotyping technology based on sequencing a region flanking a unique primer, thereby sequencing the SNP and the surrounding regions, allowing the discovery of thousands of closely related novel SNPs. The results reveal

that SPET genotyping is a robust and high-performance technology for genetic studies, including the identification of duplicates and taxonomic misclassifications in the accessions stored in the germplasm banks. Based on the information generated in the first three chapters, core collections were established for each crop, encompassing maximum genetic and phenotypic diversity in a set of 450 individuals.

Finally, in the fourth chapter, the genetic and phenotypic analysis of the tomato core collection is examined and described using an approach based on establishing genetic groups based on their genetic proximity. Genetic and phenotypic diversity analysis revealed distinct patterns of variation among different genetic groups, contradicting previous claims of a decrease in genetic diversity due to genetic improvement and uncovering unique correlations between morphological traits within different groups. The study highlights the importance of considering both genetic and phenotypic diversity in tomato breeding initiatives, with a particular emphasis on aspects such as fruit size, shape, color, and quality.

In conclusion, the work carried out in this thesis increases the knowledge and accessibility to the major Solanaceae collections conserved in germplasm banks, while also generating molecular tools for their genotypic evaluation. Furthermore, the findings emphasize the crucial role of germplasm banks as reservoirs of genetic diversity, as well as the challenges they face, including limited data availability, duplications, inaccurately classified accessions, and lack of consensus in characterizations, underscoring the need to maintain efforts in data collection, review, and standardization. In summary, these advancements provide a foundation for the future, offering valuable information for the conservation of the collections themselves and for their use in breeding programs.

RESUMEN

El impacto del cambio climático en los cultivos hortícolas es cada vez más evidente, lo que ha llevado a la pérdida y erosión de diversidad genética de manera drástica. Esto plantea importantes desafíos para la mejora de los cultivos, que requiere la exploración de los recursos fitogenéticos conservados en los bancos de germoplasma y el desarrollo de tecnologías que permitan evaluar el valor fenotípico y genotípico de estos materiales. Sin embargo, la situación actual de las colecciones de germoplasma es la existencia de duplicados no identificados entre colecciones, errores en la clasificación taxonómica, documentación insuficiente y no disponible para investigadores y mejoradores, añadido a la falta de financiación para la conservación y gestión adecuadas. Esto dificulta enormemente la utilización de estos recursos. En la presente Tesis se aborda este problema comenzando por la unificación de datos de pasaporte, fenotipado e imágenes de las principales colecciones de tomate, pimiento y berenjena en un mismo repositorio. El genotipado de la colección y el análisis conjunto de los datos de genotipado generados ha permitido la construcción de colecciones nucleares que han sido fenotipadas con detalle, aumentando su conocimiento acerca de la variabilidad genotípica y fenotípica y, consecuentemente, su aprovechamiento por parte de los investigadores y mejoradores.

En el primer capítulo, se ha realizado el inventariado, estandarización y revisión de los datos de pasaporte y fenotípicos disponibles de las accesiones de tomate, pimiento y berenjena conservados en los principales bancos de germoplasma europeos y no europeos, con el objeto de mejorar la eficiencia del manejo de los recursos fitogenéticos.

El segundo capítulo se centra en el desarrollo y optimización de un método de extracción de ADN genómico de alta calidad, rápido y económico que combina las ventajas del método de extracción basado en el CTAB, añadido a la purificación de los ácidos nucleicos en una matriz de sílice. Es un método universal que puede utilizarse para diferentes especies y tejidos. Se ha evaluado la eficiencia del ADN genómico resultante en diferentes plataformas de secuenciación como SPET (*Single Primer Enrichment Technology*) y Oxford Nanopore, generando resultados muy prometedores. Esto facilita el paso previo al genotipado de las colecciones que es la extracción de ADN.

En el tercer capítulo se aborda el genotipado de las colecciones. El elevado número de accesiones de cada cultivo, en particular el tomate, supone un problema de tipo económico, en ocasiones irresoluble. Por ello, el tercer capítulo está orientado a la evaluación del potencial de la tecnología de secuenciación SPET, más económica que otras conocidas, para el genotipado de alto rendimiento de colecciones de

germoplasma de tomate y berenjena. SPET es una tecnología de genotipado dirigida basada en la secuenciación de una región que flanquea a un cebador único, secuenciando al SNP y las regiones que rodean al SNP diana, lo que permite descubrir miles de SNP nuevos estrechamente relacionados. Los resultados revelan que el genotipado SPET es una tecnología robusta y de alto rendimiento para estudios genéticos, incluyendo la posibilidad de identificación de duplicados y errores de clasificación taxonómica en las entradas conservadas en los bancos. Con la información generada en los primeros tres capítulos se establecieron las colecciones nucleares para cada cultivo, abarcando la máxima diversidad genética y fenotípica en un conjunto de 450 individuos.

Finalmente, en el cuarto capítulo, se analiza y describe la colección nuclear de tomate a nivel genético y fenotípico, mediante un enfoque basado en el establecimiento de grupos genéticos basados en su proximidad genética. El análisis de la diversidad genética y fenotípica reveló patrones de variación distintos entre diferentes grupos genéticos, contradiciendo afirmaciones anteriores que proponían una disminución en la diversidad genética como consecuencia de la mejora genética y descubriendo correlaciones entre rasgos morfológicos únicas dentro de los diferentes grupos. El estudio destaca la importancia de considerar tanto la diversidad genética como la fenotípica en las iniciativas de mejora del tomate, con especial énfasis en aspectos como el tamaño, la forma, el color y la calidad del fruto.

En definitiva, los trabajos realizados en esta Tesis aumentan, por un lado, el conocimiento y la accesibilidad a las principales colecciones de solanáceas conservadas en los bancos de germoplasma, y por otro, generan herramientas moleculares que permiten su evaluación genotípica. Además, los hallazgos obtenidos destacan el papel crucial de los bancos de germoplasma como reservorios de diversidad genética, así como los desafíos a los que se enfrentan: disponibilidad limitada de datos, duplicaciones, accesiones erróneamente clasificadas y la falta de consenso en las caracterizaciones, que enfatizan la necesidad de mantener los esfuerzos en la recopilación, revisión y estandarización de los datos. En resumen, estos avances suponen una base para el futuro, proporcionando información valiosa para la propia conservación de las colecciones y para su uso en programas de mejora.

RESUM

L'impacte del canvi climàtic en els cultius hortícoles és cada vegada més evident, la qual cosa ha portat a la dràstica pèrdua i erosió de la diversitat genètica. La reduïda diversitat genètica planteja importants reptes per a la millora dels cultius. Sent necessari l'exploració dels recursos genètics vegetals conservats en els bancs de germoplasma i el desenvolupament de tecnologies que permeten avaluar el valor fenotípic i genotípic d'aquests materials. Pel que fa a les col·leccions de germoplasma presenten duplicats no identificats entre col·leccions, errors en la classificació taxonòmica, falta de finançament per a la conservació i gestió adequades a banda de documentació insuficient i no disponible (investigadors i milloradors vegetals). En la present tesi doctoral s'aborda aquest problema unificant les dades de passaport, fenotipat i imatges de les principals col·leccions de tomaca, pebre i albergínia en un mateix repositori. El genotipat de la col·lecció i l'anàlisi conjunt dels genotips ha permès la construcció de col·leccions nuclears (que han sigut fenotipats amb detall), augmentant el coneixement per part dels investigadors i milloradors vegetal sobre la variabilitat genotípica i fenotípica de la/les col·leccions nuclears generades.

En el primer capítol, s'ha realitzat l'inventariat, estandardització i revisió de les dades de passaport i fenotípiques disponibles de les accessions de tomaca, pebre i albergínia conservats en els principals bancs de germoplasma a nivell mundial, a fi de millorar l'eficiència del maneig dels recursos genètics vegetals.

El segon capítol es focalitza en el desenvolupament i optimització d'un mètode d'extracció de ADN genòmic d'alta qualitat, ràpid i econòmic que combina els avantatges del mètode d'extracció basat en el CTAB amb l'ús de matrius de sílice. El mètode desenvolupat pot utilitzar-se de manera universal per a diferents espècies i teixits vegetals. S'ha avaluat l'eficiència del ADN genòmic resultant en diferents plataformes de seqüenciació com SPET (*Single Primer Enrichment Technology*) i *Oxford Nanopore*, generant resultats molt prometedors. Això facilita el pas previ al genotipat de les col·leccions que és l'extracció d'ADN.

En el tercer capítol aborda l'optimització del procés de genotipat de les col·leccions generades. L'elevat nombre d'accessions de cada cultiu, en particular la tomaca, suposa un problema de tipus econòmic, a vegades irresoluble. Per això, aquest capítol està orientat a l'avaluació del potencial de la tecnologia de seqüenciació SPET per al genotipat d'alt rendiment de col·leccions de germoplasma de tomaca i albergínia a un preu econòmic . SPET és una tecnologia de genotipat dirigida basada en la seqüenciació d'una regió que flanqueja a un encebador únic, seqüenciant el SNP diana i les regions genòmiques que l'envolten, permetent descobrir milers de

SNP nous estretament relacionats. Els resultats revelen que el genotipat SPET és una tecnologia robusta i d'alt rendiment per a estudis genètics, incloent-hi la possibilitat d'identificació de duplicats i errors de classificació taxonòmica en les entrades conservades en els bancs de germoplasma. La informació generada va permetre establir col·leccions nuclears per a cada cultiu, abastant la màxima diversitat genètica i fenotípica en un conjunt de 450 individus.

Finalment, en el quart capítol, s'analitza i descrigué la col·lecció nuclear de tomaca a nivell genètic i fenotípic, focalitzant-se en l'establiment de grups genètics basats en la seua proximitat genètica. L'anàlisi de la diversitat genètica i fenotípica va revelar patrons de variació diferents entre diferents grups genètics, contradient afirmacions anteriors que proposaven una disminució en la diversitat genètica a conseqüència de la millora genètica. També és descobriren noves correlacions entre trets morfològics únics dins dels diferents grups. L'estudi destaca la importància d'abordar les iniciatives de millora de la tomaca tenint en compte tant la diversitat genètica com la fenotípica, amb especial èmfasi en aspectes com la grandària, la forma, el color i la qualitat del fruit.

En definitiva, els treballs realitzats en aquesta tesi doctoral augmenten, d'una banda, el coneixement i l'accessibilitat a les principals col·leccions de solanàcies conservades en els bancs de germoplasma. Per un altre, generen eines moleculars que permeten l' avaluació genotípica de les col·leccions analizades. A més, dels resultats obtinguts destaca el paper crucial dels bancs de germoplasma com a reservoris de diversitat genètica, així com els desafiaments als quals s'enfronten: disponibilitat limitada de dades, duplicacions, accessions erròniament classificades i la falta de consens en les caracteritzacions. Aquests reptes als que fan front el bancs de germoplasma, emfatitzen la importància dels esforços en la recopilació, revisió i estandardització de les dades. En resum, aquests avanços suposen una base per al futur, proporcionant informació valuosa per a la pròpia conservació de les col·leccions i el seu ús en programes de millora.

GENERAL INTRODUCTION

I was taught that the way of progress is neither swift nor easy.
Marie Curie

1. Genetics resources and Genebanks

Collections of plant genetic resources (PGRs) began in the mid-20th century, triggered by the accelerated loss of agronomic diversity when traditional landraces started to be replaced by the first improved cultivars (Peres, 2016). Indeed, seed banks were conceived as a response to preserve genetic material solving the genetic erosion issue due to the globalization of crops, resulting from the introduction of new high-yielding agricultural varieties during the Green Revolution promoted by Norman Borlaug in 1970 (Khush, 1999).
The main purpose of gathering and inventorying all available genetic diversity was their further use in the future (Fowler & Hodgkin, 2004), both as a source of diversity for breeding programmes and directly through the reintroduction of landraces to face possible environmental changes and society's needs (Mascher et al., 2019). PGRs stored in genebanks are considered historical documents that have the potential to elucidate the genetic history of agriculture (Fowler & Hodgkin, 2004; Tripodi et al., 2021). Seed banks have evolved throughout history to suit the demands of users.

However, the accessibility and use of genetic variation present in genebanks and its transfer into crop improvement is currently limited due mainly to the incompleteness of the available information and a lack of data allowing comparability of genebank accessions held in different countries (Weise et al., 2020). To better serve the user community, genebanks must strengthen their core activities by providing access to more comprehensive, uniform and pertinent information on a crop's gene pool. There is a need to provide novel services, such as the introduction of specific user-oriented collection types, including research populations and genetically purified lines, and the development of novel information services, allowing end-users to access the phenotypic and experimental information on accessions generated by genebanks during collection management but also by the community of genebank users. These services will accelerate Solanaceae Genetic Resources (SGR) mobilization stored in genebanks to foster crop improvement, diversification, and better adaptability to changing environments.

2. G2P-SOL initiative

G2P-SOL (*Linking genetic resources, genomes and phenotypes of Solanaceous crops*) is a global research alliance bringing together the major European and International repositories hosting germplasm of the four major Solanaceous food crops (potato, tomato, pepper and eggplant). The production value of these crops amounts to 66% of the total European horticultural crop production value (FAO, 2021). The G2P-SOL project will focus on harmonizing information on Solanaceous germplasm resources and performing sequence-based genotyping and phenotyping for traits of strategic importance for each crop. These activities will lead to a significantly enhanced understanding of the variability held in these highly valuable collections, thus greatly increasing their utility in Solanaceous crop improvement, which will have downstream impacts on food security in the face of rapid environmental changes. The G2P-SOL project aims to set up a pipeline to generate novel and curated phenotypic and genotypic data and promote linking the genetic code underlying Solanaceae biodiversity with the genotypes stored in germplasm bank and with the traits that improve productivity and adaptation to different stresses.

The advances in sequencing technologies and its cost reduction have made the genotyping of hundreds of thousands of accessions stored affordable. Nevertheless, the utilization of genetic diversity stored in genebanks and its integration into cultivated crops is constrained by the inadequate and insufficient information available across various countries. To address these limitations, genebanks need to enhance their efforts by offering broader, standardized, and concise access to comprehensive information regarding their crop collections.

To attain the objectives of this project, a range of specifics goals were established to accomplish:

(1) Defining and maintaining genetic pools for crop improvement.
(2) Phenomic and genomic data: collection, generation, analysis, harmonization, storage and linkage with genebanks.
(3) Pre-breeding and germplasm enhancement: introduction of important traits from wild non-adapted germplasm into elite germplasm.
(4) Dissemination, valorisation and training.

The project will redefine how to manage and organize SGRs to make them accessible to the community. The utilization of this knowledge has facilitated the development of cultivars with enhanced contributions to sustainability, human, nutrition, and health.

3. Relevance of solanaceous crops: understanding the importance of tomato and eggplant

The *Solanaceae* family includes more than 2.800 species, organized into 96 genera, defined by their extraordinary morphological and ecology diversity (Peralta et al., 2008). Moreover, from an economic point of view, it is one of the most important families with some of the largest cultivated worldwide crops such as tomato (*Solanum lycopersicum*), eggplant (*Solanum melongena*), pepper (*Capsicum annuum*) and potato (*Solanum tuberosum*). Specifically, tomato and eggplant crops ranked first and fifth, respectively, in terms of annual world total production (FAOSTAT, 2021). Production has drastically increased by $189 \cdot 10^6$ t in tomato and $58.6 \cdot 10^6$ t in eggplant over the last decades to cover feed requirements (**Figure 1**). The relevance of tomato and eggplant crops have been increasing in the past decades and, consequently, their harvested area and production volume have increased during that same period.

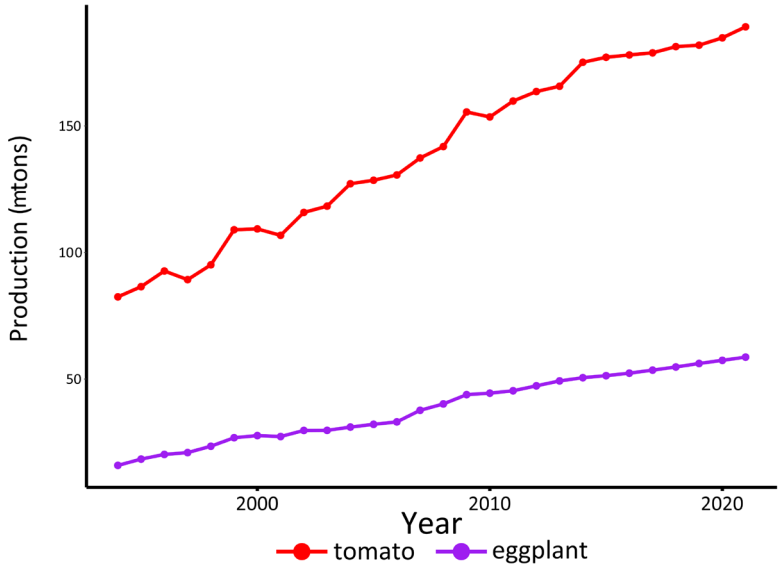


Figure 1. World eggplant and tomato production (mtons) per year between 1994 and 2021. Based on data from FAOSTAT, 2021.

According to the available data from FAO in 2021, more than half of tomato production was produced in Asia (63 %), more precisely in China (35.7 %), which is responsible for a third of the world's tomato production (**Figure 2ab**). Followed by Europe (12.9 %), where the most important growers are Turkey ($1.3 \cdot 10^7$ t), Italy ($6.6 \cdot 10^6$ t) and Spain ($4.7 \cdot 10^6$ t), and America (12.5 %), with USA as the largest producer ($1.1 \cdot 10^7$ t).

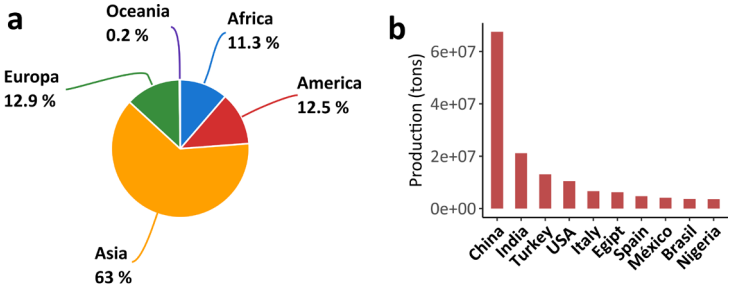


Figure 2. (a) Global tomato production by region. **(b)** Ranking of top 10 tomato-producing (tons) countries in the world in 2021. Adapted and modified from FAOSTAT, 2021.

Regarding the eggplant crop, the Asian continent accounts for most of global production, contributing approximately 94.2 % of the total output (**Figure 3a**). Among the Asian countries, China and India are the largest producers, contributing 63.8 % and 21.9 % of the total production, respectively (**Figure 3b**). Notably, Europe only accounts for 1.7 % of production. Spain ranks second in European eggplant production with a total output of $2.6 \cdot 10^5$ t and ranks tenth globally. Conversely, Italy has the highest eggplant production in Europe ($3.1 \cdot 10^5$ t) and ranks eighth globally (**Figure 3b**).

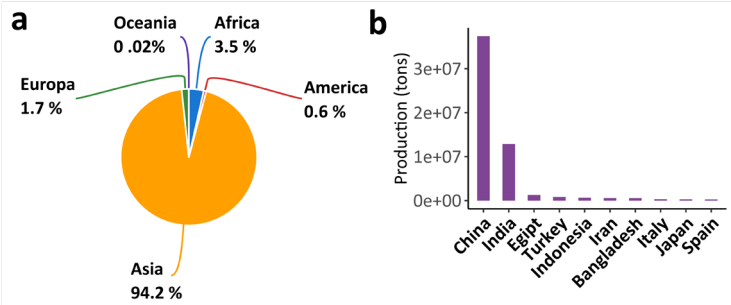


Figure 3. **(a)** Global eggplant production by region. **(b)** Ranking of top 10 eggplant-producing (tons) countries in the world in 2021. Adapted and modified from FAOSTAT, 2021.

Furthermore, the significance of tomato and eggplant is due to their extensive utilization by humans for a multitude of purposes over countless years. These purposes encompass their ornamental value, nutraceutical properties, and their role as model species in biological and genetic research. However, their primary and most prevalent use has been as a food source**.** As such, tomato and eggplant continue to play a vital role in global agriculture and culinary traditions, making them indispensable crops in many regions around the world.

4. The case of tomato

Introduction

The tomato crop is globally recognized owing to its economic and nutritional importance (FAO, 2021). The nutritional value is due to its richness of fiber, minerals, vitamins, and being an essential source of antioxidants, making it a highly appreciated product worldwide (Rosa-Martínez et al., 2021).

The taxonomic classification of tomato has been a topic of extensive discussion among scientists due to the large number of species comprising the *Solanum* genus. The currently accepted taxonomic classification is that presented in Peralta et al., (2008). In a very simplified vision, the *Lycopersicon* clade encompasses the cultivated tomato (*S. lycopersicum*) and its 12 closest wild relatives (Causse et al., 2017; Peralta & Spooner, 2005), all of which are native to South America and occupy a vast range of ecological niches, including dry deserts, coastal regions, and high-altitude Andean valleys (**Figure 4**)

The origin and domestication of tomato has also been the subject of debate for decades, however, the latest research established that cultivated tomato originated in Mesoamerica (Blanca et al., 2015), specifically in the region now known as Mexico. It is the result of the domestication of its semi-domesticated ancestor *S. lycopersicum* var. *cerasiforme* (Blanca et al., 2012; Razifard et al., 2020), which in turn was pre-domesticated from the wild species *S. pimpinellifolium*, native to Peru (Blanca et al., 2015) and Ecuador (Blanca et al., 2022a; Zuriaga et al., 2009). The tomato was introduced to Europe from Latin America at the beginning of the 16th century (Bauchet & Causse, 2012). However, it was not until three centuries later that the tomato became a popular food for human consumption, which led to an explosion in its diffusion throughout Europe via Spain, North America, and Asia. The Spanish introduced tomatoes to their colonies, including the Philippines, which served as a gateway to reach other parts of Asia (Peralta et al., 2008). The widespread acceptance of tomatoes in the United States did not occur until the 19th century, during which a period of "tomato fever" emerged. First breeding programs were launched in the 1930s to satisfy this growing demand, creating more productive tomato varieties (Ronga et al., 2019). In addition, the emergence of seed companies and trained breeders allowed the commercialization of the first tomato hybrids (Cheema & Dhaliwal, 2004). The advent of hybrids was a pivotal moment in tomato cultivation. It resulted in a rise in commercial yield due to farmers needing to purchase new seeds yearly, given trait segregation in the offspring. In response to the rising demand for tomatoes, production experienced a significant boost in the 20th century, mainly due to the adoption of chain packaging. Besides that, in the 1940s-1960s, there were the first introgressions with wild tomato species to

develop more productive varieties better adapted to the industry's demands and the incipient effect of climate change on crops.

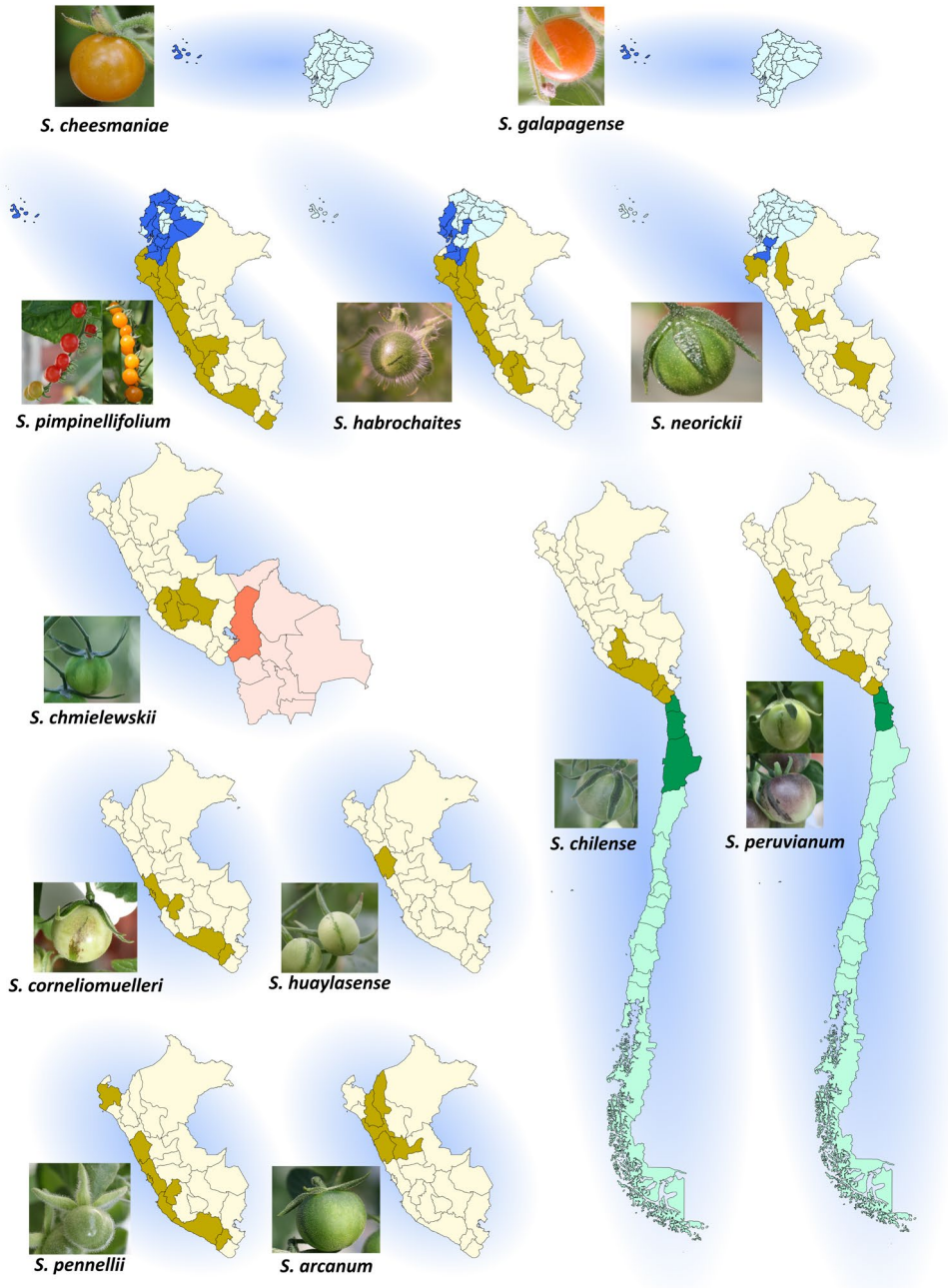


Figure 4. Geographic distribution of wild tomato species belonging to the *Lycopersicon* clade. Based on data from the TGRC database and Peralta et al. (2008). The highlighted countries correspond to the countries where that species is found.

Tomato germplasm collections

Tomato germplasm collections play a crucial role in preserving the genetic diversity for further uses. These collections serve as repositories of genetic resources, encompassing wild relatives, landraces and cultivated varieties, and breeding materials. They provide valuable material for research and breeding efforts. Additionally, these collections serve as a safeguard against the loss of genetic diversity, thereby ensuring the long-term sustainability of tomato cultivation. They provide a solid foundation for developing improved tomato varieties that can meet the evolving demands of the agricultural industry and consumer preferences. The availability and accessibility of well-curated germplasm collections, and collaborative research efforts, are essential for advancing tomato research, fostering innovation, and ensuring a sustainable future for tomato cultivation.

Currently, some of the most notable international germplasm collections for tomato and its wild relatives include the World Vegetable Center in Taiwan (14,398 accessions), the United States Department of Agriculture - Agricultural Research Service (USDA-ARS; 12,589 accessions), Plant Genetic Resources Unit (PGRU; 6,600 accessions), and the C.M. Rick Tomato Genetics Resources Center (TGRC; 4,153 accessions) in the United States (**Table 1**). In terms of germplasm collections in Europe, noteworthy institutions include the Centre for Genetic Resources at Wageningen University in the Netherlands, the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK: 1227 accessions) in Germany, and the two Spanish collection held in the Institute for the Conservation and Breeding of Agricultural Biodiversity at Universitat Politècnica de València (COMAV-UPV; 4189 accessions) and the Banco de Germoplasma de Especies Hortícolas del Centro de Investigación y Tecnología Agroalimentaria (BGHZ; 3131 accessions)— additionally, the N.I. Vavilov All-Russian Scientific Research Institute of Plant Industry (VIR) in Russia is renowned for its valuable materials, albeit with limited accessibility.

Considering the information included in Genesys and WIEWS databases, as well as the individual databases of the different germplasm banks or national inventories such as the Spanish Inventory of Plant Genetic Resources (INIA-CRF) (**Table 1**), the majority of the conserved material corresponds to the cultivated species *S. lycopersicum* var. *lycopersicum* (SLL), followed by *S. lycopersicum* var. *cerasiforme* (SLC). Among the wild species, *S. pimpinellifolium* (SP) is the most abundant, followed by *S. peruvianum*, *S. habrochaites* and *S. chilense*. Apart from these species, the remaining species belonging to the *Lycopersicon* clade, such as *S. arcanum*, *S. galapagense*, and *S. cornelliomulleri*, are very poorly represented (**Table 2**). The species with the least conserved representativeness is *S. huaylasense*.

Table 1. Inventory of accessions for tomato and its wild relatives belonging to *Lycopersicon* clade across diverse germplasm banks, databases, and inventories.

Institutions	Description	Country	Total*	SLL	SLC	SP	WR
WorldVeg[1]	Genebank	Taiwan	14,398	8,070	396	333	790
TGRC[2]	Genebank	USA	4,153	2,869	421	331	532
USDA-ARS[3]	Genebank	USA	12,589	10,515	591	576	907
INRAE[4]	Genebank	France	3,380	3,095	*nd*	*nd*	285
IPK[5]	Genebank	Germany	1,227	*nd*	*nd*	*nd*	*nd*
UPV-COMAV	Genebank	Spain	4,189	3,629	253	120	187
BGHZ	Genebank	Spain	3,131	3,025	102	2	2
INIA-CRF[6]	Databases and inventory	Spain	13,048	11,893	535	228	392
Genesys[7]	Databases	USA		27,806	1,654	1,028	1,411
WIEWS[8]	Databases	USA	65,207	42,480	1,691	1,951	2,451

The asterisk (*****) refers to the total number of accessions for tomato and wild relatives belonging to *Lycopersicon* clade. SLL = *Solanum lycopersicum* var. *lycopersicum*; SLC = *Solanum lycopersicum* var. *cerasiforme*; SP = *Solanum pimpinellifolium*; WR = Wild relatives. and *nd* = no data. The table was compiled using data from the following databases: [1]https://genebank.worldveg.org , [2]https://tgrc.ucdavis.edu/ , [3]https://www.ars-grin.gov/ , [4]https://urgi.versailles.inra.fr/siregal/siregal/grc.do , [5]https://www.ipk-gatersleben.de/forschung/genbank , [6]https://bancocrf.inia.es/es/ , [7]https://www.genesys-pgr.org/ and [8]https://www.fao.org/wiews/data/ex-situ-sdg-251/search/es/#results.

Table 2. Inventory of tomato wild relatives among germplasm banks, databases, and inventories.

Specie	Genesys	TGRC	AVRDC	ARS-GRIN	INIA-CRF	WIEWS	Total
S. galapagense	59	28	27	32	11	70	227
S. cheesmaniae	71	41	17	55	6	164	354
S. neorickii	75	47	12	61	38	124	357
S. arcanum	27	37	4	19	24	45	156
S. chmielewskii	42	16	11	27	28	85	209
S. peruvianum	492	66	116	295	54	934	1957
S. huaylasense		14		3	9	8	34
S. cornelliomulleri	42	56	10	35	18	75	236
S. chilense	163	114	46	116	105	382	926
S. habrochaites	367	114	137	202	89	584	1493
S. pennellii	135	47	59	62	23	175	501

Significant attention has been dedicated in recent years to conducting comprehensive phenotypic evaluations on tomato germplasm collections, which display an extensive range of natural diversity. These traits variation included not only standardized descriptors developed by IPGRI (1996), such as fruit shape, colour and plant growth habit but also included analysis of fruit quality traits as sugar content (°Brix), volatiles organic compound, secondary metabolites, among other; and stresses tolerances analysis against the diseases and pest.

Numerous studies have been conducted to assess the morphological diversity of tomato germplasm, aiming to understand and characterize the wide range of variations present within different tomato cultivars and related genetic resources (Farinon et al., 2022; García-Martínez et al., 2013; Grozeva et al., 2021; Mazzucato

et al., 2008; Pons et al., 2022; Schouten et al., 2019). All studies revealed the extensive diversity of collections stored in germplasm banks, ranging from morphological descriptors such as fruit shape, size, and colour to agronomic traits such as yield output, flowering, and maturate dates. Additionally, there was diversity observed in metabolic compounds of interest for fruit quality improvement, such as VOCs (volatile organic compounds), lipids, and semi-polar compounds. On the other hand, the evaluation of wild relatives of tomato has been conducted to a lesser extent. The characterization and genotyping of a total of 163 tomato accessions by Mata-Nicolás et al., (2020), which include the cultivated tomato, *S. lycopersicum* var. *cerasiforme*, and the wild species *S. pimpinellifolium*, representing the morphological and genetic variability of their center of origin and domestication, resulted in a wide diversity in most of the evaluated traits, primarily related to leaf characteristics, fruit shape, size, colour, and floral morphology. Another study conducted by Blanca et al., (2015) on approximately 1000 accessions, of which the majority were cultivated tomatoes (350 *S. lycopersicum* var. *lycopersicum* and 316 *S. lycopersicum* var. *cerasiforme*), also included wild species such as *S. pimpinellifolium*, *S. galapagense*, *S. neorickiii*, *S. chmielewskii*, and hybrids between tomato and *S. pennellii* and *S. pimpinellifolium*. The study revealed a wide range of variation and identified genes associated with desirable traits such as disease resistance and fruit quality. These findings provide insights into the history of tomato breeding and underscore the importance of preserving genetic diversity.

Furthermore, huge efforts have been made to store these data in easily accessible online databases that can be navigated to explore variations in traits. These databases serve as repositories for the passport data, allowing researchers to efficiently search and analyse trait variations in cultivated tomato and other related species. As a reflection of this, three major completed international initiatives, the National Natural Science Foundation of USA Varitome Project (NSF IOS 1564366; https://www.solgenomics.net/projects/varitome), and the European H2020 Projects TRADITOM (No. 634561; https://traditom.eu/es/), G2P-SOL (No. 677379; http://www.g2p-sol.eu/) and BRESOV (No. 774244; https://bresov.eu/) have demonstrated tremendous potential and provided valuable information on the genetic and morphological diversity of tomato, as well as its potential for future research and breeding programs.

Genetic and phenotypic diversity: an asset for enhancing genetic resources

The morphologic and genetic diversity of tomatoes has been enhanced over time through the processes of domestication and natural selection, resulting in its adaptation to various environmental conditions and significant morphological variability (Schouten et al., 2019). Nevertheless, domestication also caused a bottleneck effect that limited genetic diversity, resulting in a dramatic reduction in

early modern varieties lacking introgressions (Cebolla-Cornejo et al., 2007; Gao et al., 2019; Tamburino et al., 2020), thereby decreasing their ability for adaptation. To prevent these problems it is essential to broaden the genetic base. To achieve this, a good source of genetic diversity is the wild relatives. In fact, some gene sets associated with biotic and abiotic stresses lost during domestication are present in wild species (Kapazoglou et al., 2023). Wild species are much more genetically diverse compared to cultivated tomatoes. It has been observed that wild species and taxonomically closely related varieties to cultivated tomatoes exhibit different genetic variabilities, such as the case of the wild *S. pimpinellifolium* and the variety *S. lycopersicum* var. *cerasiforme* (Blanca et al., 2012). On the other hand, within the species *S. pimpinellifolium*, different levels of genetic diversity are also found depending on their geographical origin (Blanca et al., 2015). Populations in the northwest of Peru show the highest variability, progressively decreasing towards the south of Peru and Ecuador (Zuriaga et al., 2009).

Over the years, significant progress has been made in understanding the diversity of cultivated tomatoes (Bai & Lindhout, 2007). According to Schouten et al., (2019), the genetic diversity of cultivated tomato varieties was minimal in the 1950s and even lower in the 1960s. However, from the 1970s onward, the diversity expanded significantly, increasing eightfold compared to the previous decade. This expansion is thought to be a result of the introduction of introgressions, which offset the decline in diversity caused by selective breeding. Notably, chromosomes 4, 5, 6, 9, 11, and 12 exhibit a notable increase in diversity in modern varieties due to the incorporation of large introgressions from related wild species.

Further detailed studies on existing tomato materials have unveiled the causal introgressions and the donor species responsible associated with the response to major biotic and abiotic stresses. As a reflection, a large introgression from *S. peruvianum* on chromosome 9 carries the Tomato mosaic virus (ToMV) resistance genes *Tm-2* and *Tm-2*^{2} (Alexander, 1963; Laterrot & Pecaut, 1969; Schroeder et al., 1967), while genes *Cf-4* and *Cf-9* located in chromosome 1 and *Cf-2* and *Cf-5* in chromosome 6 for resistance to leaf mold disease (*Cladosporium fulvum*) were introgressed from *S. pimpinellifolium* (Jones et al., 1993). Additionally, the resistance gene *Mi-1*, providing resistance to the root-knot nematode (*Meloidogyne incognita*), is found on chromosome 6 and came from *S. peruvianum* (Bailey, 1941; Ho et al., 1992). These findings shed light on the specific genetic contributions from different donor species that confer important stress tolerance traits in cultivated tomatoes.

Furthermore, using wild species has proved an efficient strategy to improve fruit quality. For instance, the species *S. pimpinellifolium* serves as a valuable genetic

source for desirable traits associated with fruit quality. Notably, it exhibits a high lycopene and soluble solid content (Y. D. Sun et al., 2012), fruit colour variations (Barrantes et al., 2016; Grandillo & Tanksley, 1996), sugar content (Grandillo & Tanksley, 1996), and volatile organic compounds (VOCs) (Capel et al., 2015; Rambla et al., 2016).

Regarding the phenotypic diversity of tomatoes, it has not been reduced over the years; rather, it has been modified in response to changing market demands and consumer preferences. Morphological variations have emerged to facilitate crop and fruit management. In the case of tomatoes intended for industrial purposes, modifications have been made to select more manageable shapes for transportation and harvesting, both mechanically and manually. This involves selecting tomatoes without pistil scars, round, with high firmness, and with fewer ribs to facilitate transportation and minimize mechanical injuries that could lead to pathogen proliferation and shorten their post-harvest life (Sun et al., 2017; van der Knaap et al., 2014). While there is still diversity in morphologies, the predominant focus is on forms that are easier to handle.

Moreover, the genetic basis of fruit characteristics has been explored, and specific quantitative trait loci (QTLs) associated with fruit shape, such as *ovate*, *fs8.1*, or *sun*, have been identified in populations resulting from crosses between cultivated tomatoes and *S. pimpinellifolium* (Wang et al., 2019). In terms of fruit size, Víquez-Zamora et al., (2013) demonstrated that the development of cherry-like fruit sizes was achieved through the introgression of large segments of chromosomes 4, 5, and 12 from *S. pimpinellifolium.* These findings highlight the genetic mechanisms underlying the morphological diversity observed in tomato fruits and provide valuable insights for breeding programs aiming to enhance fruit characteristics.

Tomato Genetic Improvement: Current Trends and Strategies

The needs and requirements for tomato production, aiming to adapt to different cultural contexts as market demand and population growth, have been modifying the objectives of breeding programs. The initial goal of breeding programs was to increase yield by reducing damage caused by pests and diseases, which remains relevant today (Savary et al., 2012). Both pests and diseases to which cultivated tomatoes are susceptible, such as those caused by fungi, bacteria, viruses, nematodes, insects, and mites, can cause significant economic losses, as they affect characteristics such as fruit quality and plant production (Stout et al., 2018). From the 1970s onwards, globalization of the market led to the transportation of tomatoes to distant regions from their place of production. Consequently, breeding programs in the 1980s focused on extending the post-harvest life of fruits and

enhancing their firmness to withstand handling and transportation (Bai & Lindhout, 2007). However, this emphasis on post-harvest traits resulted in a decline in tomato quality (Díez & Nuez, 2008), as delayed ripening led to a reduction in the accumulation of carotenoids and volatile compounds (Baldwin et al., 2000). As a result, consumer preference for high-quality tomatoes became prominent in the 1990s. Subsequently, breeding programs redirected their efforts towards improving flavor and fruit quality (Bai & Lindhout, 2007).

Currently, breeding objectives vary depending on the market type, consumer preferences, and the intended use of the fruit. These factors are closely interconnected and share common goals. With regards to the fruit's utilization, it can be categorized for either fresh consumption or processing. In the case of fresh consumption, the focus is on fruit shape and colour, uniform appearance, extended post-harvest life, flavour, and texture. On the other hand, for processing tomatoes, such as sauces, purees, or canned products, the aim is to have compact growth and absence of abscission zone to facilitate mechanical harvesting, determinate growth habit, and traits related to fruit quality such as pH, acidity, soluble solids content, lipids, VOCs (volatile organic compounds) and inorganic compound. In addition to these quality aspects, disease resistance and adaptability of the crops to different environments are common objectives for both markets (Foolad, 2007).

Additionally, new objectives are being incorporated, which are equally important. These include the generation and search for new varieties that exhibit efficient nitrogen utilization and a broader spectrum of stress tolerance, enabling a 50% reduction in pesticide usage and a 20% decrease in fertilizer application. Furthermore, there is a current drive to promote traditional varieties due to their predominantly local characteristics and high risk of disappearance (Figàs et al., 2015). This trend aligns with the growing preference for proximity trade, highlighting the need to preserve these traditional varieties.

Tomato breeding programs have undergone significant transformations to meet the evolving needs and requirements of tomato production (Bohra et al., 2021). These programs have been driven by market demands, population growth, and consumer preferences. In order to achieve desired traits and address challenges, various types of breeding programs have been implemented. These programs encompass traditional breeding methods, such as mass selection, pedigree breeding, hybridization, and grafting, as well as modern approaches that incorporate gene technologies and bioinformatics. Each of these breeding programs plays a crucial role in developing improved tomato varieties with enhanced yield, quality, disease resistance, and adaptability to different environments.

Thanks to the introduction of new plant varieties, advancements in technologies, and improved crop management practices, tomato productivity has witnessed remarkable increases. Over the past 50 years, tomato yields have surpassed a 240% increase, while in the last 30 years alone, they have risen by 88% (Hernández et al., 2021). In 1970, the average tomato yield in Spain stood at 25 tons per hectare, whereas in 2018, it reached an impressive 85 tons per hectare. These significant productivity gains highlight the positive impact of advancements in tomato breeding, cultivation techniques, and agricultural practices on meeting the growing demand for this valuable crop.

Genomic and Biotechnological Innovations for Genetic Resources Enhancement

The establishment and maintenance of germplasm collections must be coupled with an active exploitation of them, which requires that both the phenotypic and the genetic diversity of these collections are well described (Mascher et al., 2019). The development and implementation in PGRs of novel genomic tools over the last decades has been considerably increased mainly due to their affordability and the high potential to generate meaningful information, shifting from using a few dozen of markers to using several thousands in a single experiment (Sim et al., 2012a), as well as the possibility of genetically evaluating interesting population and the subsequent selection of the most promising individuals.
Regarding that, morphological and biochemical markers such as allozymes have been widely used in tomato breeding. However, in the 1990s, the advent of different DNA molecular markers such as RAPDs (*Randomly Amplified Polymorphic DNAs*), SSR (*microsatellites*), AFLPs (*Amplified Fragment Length Polymorphisms*) and SNPs (*Single Nucleotide Polymorphisms*) enabled individual genotyping and the construction of genetic maps (Ashrafi et al., 2009; Grandillo & Tanksley, 1996; Sim et al., 2012a). In tomato crop, since the first map was released in 1968 based on morphological and physiological markers (Butler, 1968), several genetic linkage maps have been created based on interspecific crosses between cultivated tomato and its wild relatives, in particular *S. pennellii* (Foolad, 2007), as well as using different molecular approaches.

Concerning sequencing procedures, Frederick Sanger was the pioneer in DNA sequencing during the late 1970s with the development of the Sanger sequencing method (Sanger et al., 1977), whose technic relied on the utilization of chain-terminating fluorescently-labeled, which were incorporated into the single-stranded DNA chain. Subsequently, the DNA fragments were separated using an electrophoresis-based method. The automation of the process enabled the development of the first generation of sequencers, which exhibited varying reading and resolution capacities. This milestone has led to significant advancements in genomics research, particularly with the completion of the whole genome

sequencing of tomato in 2012 (Sato et al., 2012). This milestone has triggered a rapid evolution in next-generation sequencing (NGS) technologies, enabling the efficient identification and screening of an immense number of SNPs distributed throughout the genome. In 2005, 454 Life Sciences Roche introduced the first NGS machine (Daigle et al., 2011), revolutionizing the sequencing process relied on pyrosequencing (Parameswaran et al., 2007). Subsequently, in 2006-2007, two additional platforms became available to the scientific community: SOLiD by Thermo-Fisher Scientific and Solexa, which was later acquired by Illumina, which were based on sequencing by ligation and subsequent amplification (Huang et al., 2012; Kircher et al., 2011). Another notable second-generation platform that emerged was Ion Torrent (Rothberg et al., 2011), which utilized a sequencing-by-synthesis approach without the need for fluorescence or luminescence techniques. This innovation not only reduced costs but also accelerated the sequencing process, further advancing the field of NGS. The shift from second-generation to third-generation sequencing methods has revolutionized the scale of analysis, dramatically increasing throughput, reducing errors and generating a huge amounts of data (Schadt et al., 2010). Pacific Biosciences (PacBio) (Roberts et al., 2013; Xiao & Zhou, 2020) and Oxford Nanopore Technologies (MinION) (Loit et al., 2019) are two prominent technologies that have implemented long-read sequencing.

Consequently, this remarkable progress in sequencing has opened up unprecedent opportunities to develop novel applications and facilities that can accelerate processes in genetics research and breeding programs (Tripodi, 2022). In relation to these developments, microarrays emerged as the pioneering tool employed in plant research to study gene expression patterns and conduct genome-wide analyses (GWAS). Microarrays have played a significant role in investigating diverse biological processes in plants, including plant development, response to biotic or abiotic stress tolerance. Moreover, Microarray technology has also been utilized for genotyping and mapping purposes. In recent years, microarray technology has evolved, leading to the emergence of new platforms, such as DNA microarrays (Hoheisel, 2006), and transcriptome profiling arrays, further expanding the possibilities for plant research. On the other hand, the genotyping-by-sequencing (GBS) methodology has made substantial contributions to the scientific community and has become one of the most prevalent genotyping strategies. Its widespread adoption can be attributed to its versatility and broad range of applications (Elshire et al., 2011; Poland & Rife, 2012; Scheben et al., 2017; Sonah et al., 2013). The GBS method has enabled the discovery of single nucleotide polymorphisms (SNPs), high-density genetic mapping, genome-wide association studies (GWAS), diversity studies, and genomic selection, among other valuable applications. Its utilization has revolutionized genotyping practices and significantly advanced genetic research. Furthermore, SPET (Single Primer Enrichment Technology), a new

genotyping technology for targeted sequencing of single nucleotide polymorphisms (SNPs) in plants, has been successfully introduced recently. However, there is currently limited information on its performance for high-throughput plants genotyping (Barchi et al., 2019). This advancement holds great potential for driving advancements in agricultural research and enhancing breeding programs.

5. The case of eggplant

Review article

Conventional and new genetic resources for an eggplant breeding revolution

Pietro Gramazio[1,†], David Alonso[1,†], Andrea Arrones[1], Gloria Villanueva[1], Mariola Plazas[1], Laura Toppino[2], Lorenzo Barchi[3], Ezio Portis[3], Paola Ferrante[4], Sergio Lanteri[3], Giuseppe Leonardo Rotino[2], Giovanni Giuliano[4], Santiago Vilanova[1] and Jaime Prohens[1,*]

[†] These authors contributed equally to this work.

[1]Instituto de Conservación y Mejora de la Agrodiversidad Valenciana, Universitat Politècnica de València, Camino de Vera 14, 46022 Valencia, Spain
[2]CREA Research Centre for Genomics and Bioinformatics, Via Paullese 28, 26836 Montanaso Lombardo (LO), Italy
[3]Dipartimento di Scienze Agrarie, Forestali e Alimentari (DISAFA), Plant Genetics, University of Turin, Largo P. Braccini 2, 10095 Grugliasco (TO), Italy
[4]Agenzia Nazionale Per Le Nuove Tecnologie, L'energia e Lo Sviluppo Economico Sostenibile (ENEA), Casaccia Research Centre, Rome, Italy

[*] Correspondence: jprohens@btc.upv.es

Ph.D. candidate contribution

David Alonso had a main role in the following activities: literature revision, data visualization, drafting manuscript and manuscript review.

Citation: Gramazio, P., Alonso, D., Arrones A., Villanueva, G. et al. Conventional and new genetic resources for an eggplant breeding revolution. Journal of Experimental Botany. erad260 (2023). doi: https://doi.org/10.1093/jxb/erad260

Highlight

This review emphasizes the importance of genetic resources, including germplasm accessions and new experimental populations, for a breeding revolution in eggplant in the context of climate change.

Abstract

Eggplant (*Solanum melongena*) is a major vegetable crop with great potential for genetic improvement owing to its large and mostly untapped genetic diversity. Eggplant is closely related to over 500 species of *Solanum* subgenus *Leptostemonum,* belonging to its primary, secondary, and tertiary genepools, and exhibits a wide range of characteristics, including adaptive traits to climate change, that are useful for eggplant breeding. Germplasm banks worldwide hold more than 19,000 accessions of eggplant and related species, most of which have yet to be evaluated. Nonetheless, eggplant breeding using the cultivated *S. melongena* genepool has yielded significantly improved varieties. To overcome current breeding challenges and adaptation to climate change, a qualitative leap forward in eggplant breeding is necessary. The initial findings from introgression breeding in eggplant indicate that unleashing the diversity present in eggplant relatives can greatly contribute to a revolution in eggplant breeding. The recent creation of new genetic resources, such as mutant libraries, core collections, recombinant inbred lines (RILs), and sets of introgression lines (ILs) will be another crucial element for an eggplant breeding revolution, which will require the support of new genomics tools and biotechnological developments. The systematic utilization of eggplant genetic resources supported by international initiatives will be critical for a much-needed eggplant breeding revolution to address the challenges posed by climate change.

Keywords: eggplant, genepools, genetic resources, genomics, germplasm banks, introgression breeding, *Solanum melongena*, wild relatives

Introduction

Eggplant (*Solanum melongena* L.), also known as common eggplant, brinjal or aubergine, was domesticated 9,000-10,000 years ago from its wild ancestor *S. insanum* L. in the Indian subcontinent and southeast Asia (Page et al., 2019*a;* Barchi et al., unpublished). Apart from the intra-specific diversity of eggplant over 500 species of *Solanum* subgenus *Leptostemonum* (Vorontsova and Knapp, 2016; Knapp et al., 2019), belonging to its primary, secondary and tertiary genepools provide a formidable source of additional diversity (Syfert et al., 2016). As the sixth most important vegetable crop in production globally, following tomato, onion, watermelon, cucumber and cabbage (FAOSTAT, 2023), eggplant holds significant agricultural value. Its global production has increased by 24% over the past decade, from $47.3 \cdot 10^6$ t in 2012 to $58.6 \cdot 10^6$ t in 2021 (FAOSTAT, 2023). As a warm-climate crop, it is mostly cultivated in tropical, subtropical, and temperate regions of the world, and is an important part of the cuisine from East Asia, the Indian subcontinent, Southern Asia, the Middle East, and the Mediterranean basin. Eggplant contributes to alleviating "hidden hunger" by providing significant dietary amounts of K, P, Mn,Cu and folate, but also high concentrations of bioactive phenolics beneficial for human health (Plazas et al., 2013; Rosa-Martínez et al., 2021).

In recent years, the impact of climate change on agricultural production has become a significant concern, and eggplant cultivation is expected to be strongly affected by its effects (del Pozo et al., 2019; Habib-ur-Rahman et al., 2022). Increased spells of extreme events such as intense drought periods may have a dramatic impact on crops sensitive to water stress, such as eggplant (Plazas et al., 2022).

Although significant advances have occurred in eggplant breeding (Daunay and Hazra, 2012; Chapman, 2020; Toppino et al., 2021; Arafa et al., 2022) they have not kept pace with other major vegetable crops, such as tomato. Tomato breeding programs have made extensive use of genetic resources, including the introgression of multiple genes from crop wild relatives (CWRs) that have been incorporated into modern cultivars (Díez and Nuez, 2008; Schouten et al., 2019). While eggplant breeding programs have been successful in developing new improved cultivars, until now only a few of them have involved the introgression of useful traits. For example, resistance to the fungal wilts *Fusarium oxysporum* f. sp. *melongenae* (*Fom*) and *Verticillium dahliae* from *S. aethiopicum* (Toppino et al., 2007, 2008b, 2009) and *S. linnaeanum* (Liu et al., 2015) has been exploited with considerable success and new resistant materials of eggplant to these diseases have been obtained (Toppino et al., 2021).

In addition to brinjal eggplant, two other related minor crops, namely the scarlet eggplant (*S. aethiopicum* L.) and the gboma eggplant (*S. macrocarpon* L.), were domesticated in Africa (Page et al., 2019b) and are mostly grown in the sub-

Saharan region (Schippers, 2000). Although they have mostly local importance, they are relevant crops, particularly *S. aethiopicum*, in some parts of the world such as Brazil and the Caribbean as well as in Southern Italy, where a Protected Denomination of Origin exists for the *S. aethiopicum* landrace 'Melanzana Rossa di Rotonda' (Schippers, 2000; Sunseri et al., 2010). The fact that both species are cross-compatible with common eggplant (Bletsos et al., 2004; Oyelana and Ugborogho, 2008; Rotino *et al.*, 2014) is of great relevance for the breeding of the latter crop, as they share many domestication traits, facilitating the introduction of resistance to some pests and diseases from these species to common eggplant without dragging undesirable traits present in wild species (Toppino et al., 2021; Arafa et al., 2022).

Eggplant displays a wide morphological diversity, even within a single varietal group, particularly for fruit traits such as size, color and shape (**Figure 1**), as well as for agronomic traits and adaptation to different environments (Cericola et al., 2013; Taher et al., 2017; Chapman, 2020; Kouassi et al., 2020; Ro et al., 2022; Salinier et al., 2022; Toppino et al., 2022). The diversity present within the cultivated eggplant is of great interest for eggplant breeding and has allowed the development of significantly improved modern cultivars (Daunay and Hazra, 2012). However, as occurs with many other vegetables, this high morphological diversity is mostly the consequence of genetic variation in a few major genes, particularly those related to fruit traits (Daunay et al., 2004; Portis et al., 2015; Toppino et al., 2016; Mangino et al., 2021; Arrones et al., 2022; Guan et al., 2022), and the overall genetic variation of the crop is narrow (Acquadro et al., 2017; Barchi et al., 2019a; Liu et al., 2019). Therefore, the exploitation of the genetic diversity present in other cultivated eggplants (*S. aethiopicum* and *S. macrocarpon*) and wild relatives represents a promising avenue for developing improved eggplant materials by incorporating unique traits from these exotic materials (Oyelana and Ugborogho, 2008; Prohens et al., 2012; Toppino et al., 2021).

Figure 1. Diversity for fruit morphology in the cultivated (*S. melongena*) gene pool **(a),** within a particular cultivar type (striped eggplant) **(b)** and in eggplant wild relatives from the primary (GP1), secondary (GP2) and tertiary (GP3) genepools **(c)**.

Eggplant breeding challenges for the present and the future: the need for a breeding revolution

Eggplant yield increased dramatically from a global average of 10.2 t/ha in 1961-1970 to 28.0 t/ha in the 2012-2021 decade (FAOSTAT, 2023). There is not much information on the genetic vs. agronomic factors that have driven this dramatic yield increase, although Muñoz-Falcón et al. (2009) found that modern varieties of black eggplants yielded on average 29.8% more than landraces, suggesting that improvement of cultural techniques such as protected cultivation, irrigation, improved fertilization and pest and pathogens management, may have had a major role in yield increase in the last decades. This suggests that major genetic advances that occurred in the past in other major staple and vegetable crops (Hedden, 2003; Díez and Nuez, 2008) may still have to occur in the coming decades for eggplant breeding, resulting in dramatically improved varieties, adapted to the new conditions posed by climate change. So far, breeding advances and actual exploitation of genetic resources in eggplant, particularly those from related species, are not comparable to those obtained in other major vegetable crops such as tomato (Schouten et al., 2019). Although tomato has a narrow genetic diversity and exhibits limited crossability with only a few CWRs, considerable broadening of the genetic base and genetic advances have been achieved in this species through introgression breeding using wild relatives as donors. Among these achievements, the introgression of multiple genes for tolerance to diseases and fruit quality traits, the development of heterotic hybrids, the improvement of shelf-life, the diversification of varietal types, and the adaptation to multiple environments (Díez and Nuez, 2008), enabled the production of a large number of highly productive tomato varieties of many different typologies, resistant to the major diseases and suited to different environments. The success obtained in the extensive use of genetic resources in tomato breeding suggests that in eggplant, which exhibits an even greater diversity of cross-compatible relatives, the advances in breeding for adaptation to climate change and other traits, achievable with a systematic use of its genetic resources may be extraordinary.

Like tomato, eggplant is self-compatible and mostly autogamous (Daunay and Hazra, 2012). Indeed, in a study involving eggplant and tomato accessions genotyped by Single Primer Enrichment Technology (SPET) - a genotyping technique that employs targeted amplification of specific genomic regions using a single specific primer - (Scaglione et al., 2019), the heterozygosity of eggplant and tomato was reported to be 0.67% and 0.65%, respectively (Barchi et al., 2019a), confirming the mostly autogamous reproduction of both species, which in turn impacts on the applicable breeding methods. However, high levels of cross-pollination can occur when the circumstances are favourable, such as in open field conditions with the presence of pollinators (Quamruzzaman, 2021). Avoiding cross-pollination is highly

relevant for maintaining purity in the case of reproduction of landraces or germplasm accessions.

Breeding in eggplant traditionally relies on selection from both within and among landraces as well as in the development of F1 hybrids, which are predominant in high-value markets (EU Plant Variety Database, 2022). It is known since long ago that F1 hybrids in eggplant generally display heterosis (i.e., when a hybrid displays superior quantitative traits, such as yield, with respect to the standard parent) (Kakizaki, 1931; Sambandam, 1964) and heterobeltiosis (i.e., when a hybrid displays superior quantitative traits with respect to the best parent)) is also common (Rodríguez-Burruezo et al., 2008; Kumar et al., 2020). Selection of parents for heterotic hybrids is possible by evaluating the parents' combining ability, as well as by selecting parents with high genetic distance using molecular markers (Rodríguez-Burruezo et al., 2008). It is worth remembering that landraces and pure line selections of eggplant with excellent yields are also available and extensively cultivated (Muñoz-Falcón et al., 2009; Taher et al., 2017). However, further improvement of the yield potential is a significant challenge in eggplant breeding, which will undoubtedly benefit from the incorporation of new genetic diversity to allow additional genetic advances (Muñoz-Falcón et al., 2009; Daunay and Hazra, 2012).

One of the major current challenges in eggplant breeding is the development of breeding lines with an improved tolerance or resistance to major pests and diseases (Toppino et al., 2021), which may cause crop losses of up to 100% (Daunay and Hazra, 2012; Arafa et al., 2022). Eggplant is affected by numerous diseases, although the most relevant in terms of economic impact is the bacterial wilt caused by *Ralstonia solanacearum*, which is highly prevalent in tropical regions (Lebeau et al., 2013; Barik et al., 2020). In many cases, bacterial wilt prevents eggplant cultivation unless plants are grafted onto resistant rootstocks (Namisy et al., 2019). Verticillium and Fusarium wilts, as well as nematodes, are also important eggplant pathogens in many regions of the world (Arafa et al., 2022). So far, most eggplant modern commercial varieties do not carry genes for disease resistance (Srinivasan, 2009). Introgression breeding from the multiple sources of resistance found in eggplant-related species can result in the development of a new generation of materials with resistance to the main eggplant pathogens, mimicking the process occurred in tomato breeding,, where the incorporation of disease-resistant genes introgressed from wild relatives are crucial technical innovations for the success of modern commercial varieties (Díez and Nuez, 2008; Schouten et al., 2019).

The eggplant fruit and shoot borer (*Leucinodes orbonalis*), is the most damaging and difficult pest to control in the Indian subcontinent, Southern and East Asia, where multiple insecticide sprays are used to partially control it (Srinivasan, 2008). This pest is such a damaging and limiting factor in eggplant cultivation that two countries (Bangladesh and the Philippines) have authorized the use of genetically modified *Bt* eggplants expressing the cry1Ac gene from *Bacillus thuringiensis* to

control the eggplant fruit and shoot borer (Shelton et al., 2018; Gonzalvo et al., 2022). Additional pests attacking *S. melongena* are spider mites, whiteflies and aphids, which affect other solanaceous crops as well (Srinivasan, 2009). To this purpose, the development of eggplant hairless materials such as CleanLeaf® (Rijk Zwaan, De Lier, The Netherlands) has improved biological pest control in greenhouse cultivation, as the pests are more accessible to their predators and parasites. The development of new resistant or tolerant varieties can benefit from the use of eggplant genetic resources, as sources of variation to the main diseases are available in these materials (Arafa et al., 2022).

Abiotic stresses are expected to increase in the areas where eggplant is cultivated due to climate change (Toppino et al., 2022; Khalid et al., 2023). Eggplant is mildly tolerant to water and salinity stresses (Heuer et al., 1986; Díaz-Pérez and Eaton, 2015; Kouassi et al., 2020; Toppino et al., 2022); however, developing new varieties with better resilience is needed, particularly in drought-prone areas or where water and soil salinity is a problem for eggplant cultivation. Tolerance to extreme temperatures as well as to soil flooding are also important breeding objectives. Despite being a warm-climate plant, high temperatures affect pollen viability and fruit set (Toppino et al., 2022) and heat-tolerant varieties are needed for production in the warm seasons. Tolerance to cold is also important in off-season production in temperate areas, as growth and development are arrested, and fruit set is impaired by low temperatures (Toppino et al., 2022). To this purpose, parthenocarpic materials have been developed which can set fruit even under cold conditions affecting pollen viability (Kikuchi et al., 2008). Improving water and nutrient use efficiencies is also necessary for a more sustainable agriculture. In this context, breeding for better root systems could lead to more sustainable production (Chapman, 2020).

Diversification and improvement of fruit quality (Daunay and Hazra, 2012) represent other important challenges in breeding. Eggplant displays a large diversity of fruit sizes, shapes, and colors, facilitating breeding for outer fruit quality and appearance traits. QTLs (quantitative trait loci) have been identified for fruit morphological traits (Portis et al., 2015; Toppino et al., 2016, 2020; Barchi et al., 2019c; Mangino et al., 2021), although few causative genes have been identified. One exception is the *APRR2* gene (Arrones et al., 2022), which controls the synthesis of fruit peel chlorophyll, as well as several genes involved in anthocyanin synthesis (Florio et al., 2021; He et al., 2022; Li et al., 2022). The identification of causative genes underlying other important traits for fruit appearance, such as the presence of fruit stripes, fruit netting or prickliness, will provide additional tools for eggplant breeding. Eggplant is one of the vegetables with the highest antioxidant and bioactive properties, resulting from its high content of phenolic acids (Kaushik et al., 2015), which are also associated with increased browning of the fruit flesh (Mishra et al., 2013; Docimo et al., 2016; Kaushik et al., 2017), a non-desirable trait. Breeders, by directly selecting genotypes with low fruit browning, indirectly

selected for low content in phenolic acids (Prohens et al., 2007). Selection for low polyphenol oxidase (PPO) activity has been proposed to improve the phenolic acid content while limiting the effects of browning, (Plazas et al., 2013). Indeed, CRISPR/CAS9 editing of *PPO* genes expressed in the fruit has been shown to reduce fruit flesh browning without affecting phenolic acid content (Maioli et al., 2020; Kodackattumannil et al., 2023). Parthenocarpic fruit set is also of interest for reducing fruit browning, as browning is more intense in the tissues surrounding the seeds (Sarengaowa et al., 2022). Finally, saponins present in the fruit flesh tissues contribute to the bitterness of some accessions, which is also an undesirable trait (Aubert et al., 1989). In summary, research on the above traits for a more efficient development of improved eggplant cultivars.

Eggplant CWRs often exhibit concentrations of glycoalkaloids above those considered safe for human consumption (Aubert et al., 1989; Rosa-Martínez et al., 2022a). This represents a challenge in introgression breeding of eggplant, although several works showed that most introgression lines display glycoalkaloid concentrations similar to those of the cultivated recurrent parent (Mennella et al., 2010; Rosa-Martínez et al., 2022a). This indicates that, although levels of glycoalkaloids have to be monitored in the introgressed breeding lines, most of them will be safe for consumption.

Rootstock development is another important field in eggplant breeding. Rootstocks with robust root systems have been shown to improve yield and confer tolerance to soil diseases and abiotic stresses in eggplant (Gisbert et al., 2011; Barik et al., 2020). In this way, wild eggplant relatives, as well as interspecific hybrids have demonstrated high potential as rootstocks for improving eggplant production (Sabatino et al., 2018; Toppino et al., 2021). For example, the eggplant wild relative *S. torvum*, which is resistant to most soil diseases and nematodes, and hybrids between eggplant and scarlet eggplant, which provide vigor and good performance under cold conditions, are used as rootstocks at the commercial level (King et al., 2010; Schwarz et al., 2010; Calvo-Asensio et al., 2014; Ranil et al., 2015).

The systematic exploitation of genetic diversity and the use of modern technologies, such as molecular markers, for introgression breeding in eggplant will facilitate the development of highly productive and resilient varieties with traits such as disease and pest resistance, yield heterosis through genetic diversity, tolerance to abiotic stresses, including improved rootstocks, removal of undesirable traits such as prickliness, and the development of long shelf-life or seedless materials (Daunay and Hazra, 2012; Chapman, 2020; Arafa et al., 2022; Toppino et al., 2022). To achieve a successful breeding revolution, systematic efforts must be made to efficiently and rapidly utilize the high genetic diversity present in eggplant and its close wild relatives (CWRs). In particular, so far, the large genetic diversity present in CWRs has been barely exploited and used in eggplant breeding. Moreover, speed breeding techniques, such as cold priming at the expanded cotyledon stage, K fertilization supplementation, and embryo rescue, have proven

to be efficient tools for reducing generation cycles in tomato and pepper (Manzur et al., 2014; Ayenan et al., 2019; Gimeno-Páiz et al., 2023), need to be developed for a faster and more efficient eggplant breeding revolution.

The eggplant genepools and their potential for eggplant breeding enhancement

The large diversity present in cultivated eggplant for morphological and agronomic traits of interest (**Figure 1**) has facilitated the development of new varieties with improved performance and new combinations of traits (Taher et al., 2017). However, intra-specific variation is reduced for some traits, particularly tolerance to some biotic and abiotic stresses (Arafa et al., 2022) and improving such traits will require accessing inter-specific diversity. In addition, the vast number of eggplant relatives, with their diverse phenotypic (**Figure 1**) and physiological characteristics and environmental adaptation greatly expands the access to exotic and wild genetic diversity for eggplant breeding. Indeed, eggplant can be hybridized with many wild relatives from the subgenus *Leptostemonum*, which are adapted to a wide range of environments of all tropical and subtropical regions of the world (Vorontsova and Knapp, 2016; Knapp et al., 2019). Many of these wild relatives can be crossed with eggplant (Daunay and Hazra, 2012; Rotino et al., 2014; Plazas et al., 2016), facilitating conventional breeding methods to introgress the traits of interest in eggplant from allied species. Interspecific hybrids between eggplant and wild relatives as well as backcrosses with eggplant have been obtained through sexual crosses using several wild and allied species (Daunay and Hazra, 2012; Rotino et al., 2014; Premabati Devi et al., 2015; Plazas et al., 2016; Daunay et al., 2019). This include many species from the Old World (Rotino et al., 2014; Plazas et al., 2016; Toppino et al., 2021), as well as American species such as *S. elaeagnifolium, S. torvum*, *S. viarum* and *S. sisymbriifolium* (Daunay and Hazra, 2012; Rotino et al., 2014; Kouassi et al., 2016; Plazas et al., 2016), which diverged from eggplant approximately 6.7, 7.7, 8.3 and 8.9 million years ago, respectively (Särkinen et al., 2013). The accessibility for breeding of the available genetic diversity of eggplant-related species depends mainly on the genepool (primary, secondary, or tertiary) they belong to (Prohens et al., 2017), although there are significant differences within the secondary and tertiary genepools in the crossability and ease of hybridization and subsequent introgression breeding (Kouassi et al., 2016; Plazas et al., 2016).

The primary genepool (GP1) of eggplant consists of the cultivated eggplant *S. melongena* and its ancestor *S. insanum* L. (Syfert et al., 2016), which was previously considered a botanical variety of *S. melongena* (*S. melongena* var. *insanum*) (Knapp et al., 2013; Ranil et al., 2017). Although two genetic groups, named Occidental (predominantly grown in the Middle East, Europe and Africa) and Oriental (mostly grown in the Indian subcontinent, Southeast Asia and eastern Asia), have been recognized within *S. melongena* (Vilanova et al., 2012; Cericola et al., 2013) no

genetic barriers exist between them or with *S. insanum*, and hybridization within and between *S. melongena* groups or between *S. melongena* and *S. insanum* is equally successful (Plazas et al., 2016; Daunay et al., 2019). *S. insanum* grows as a wild or weedy species in a wide range of environments in its natural distribution (Indian subcontinent, Southeast and Eastern Asia, Madagascar and some Indian Ocean islands) (Ranil et al., 2017). In these areas, *S. melongena* and *S. insanum* form a genetic continuum with intermediate forms resulting from hybridization, and genetic flow between both species has been documented (Knapp et al., 2013; Davidar et al., 2015; Mutegi et al., 2015; Page et al., 2019a). *Solanum insanum* has a high potential for the development of improved cultivars (Ranil *et al.*, 2017). Nonetheless, due to the natural genetic flow between *S. insanum* and *S. melongena*, it is plausible that some unknown introgressions from the former have been already inadvertently incorporated and utilized in eggplant breeding. This species, therefore, represents a reservoir of potential superior untapped alleles for traits of interest, including those related to climate change, which could be easily transferred to the *S. melongena* genepool.

The secondary genepool (GP2) is very broad in terms of number of species (Eggplant clade, Anguivi grade, and Climbing clade), geographic distribution (Africa, Indian subcontinent, Southeast and Eastern Asia), and environmental adaptation (from desertic areas to wet forests; from sea level to 3,300 m) (Vorontsova and Knapp, 2016; Syfert et al., 2016; Knapp et al., 2017). The wild ancestor of eggplant (*S. insanum*) diverged from all GP2 species between 1.5 and 4.6 million years ago (Särkinen et al., 2013). Within the GP2, eggplant hybridization and introgression are easier with Eggplant clade species, showing a higher hybridization success, hybrid seed viability and pollen fertility than in the *Anguivi* grade and Climbing clade (Rotino et al., 2014; Plazas et al., 2016). Generally, embryo rescue is unnecessary to obtain hybrids and backcrosses with *S. melongena*, although hybridization with GP2 species is more challenging than with GP1 materials (Kouassi et al., 2016; Plazas et al., 2016; Daunay et al., 2019) and sometimes alternative breeding strategies such as somatic hybridization were necessary to obtain fertile hybrids (Rotino et al., 1998; Särkinen et al., 2013). Several species belonging to the GP2 such as *S. anguivi*, *S. dasyphyllum*, *S. incanum*, *S. linnaeanum* and *S. tomentosum* (**Table 1**) have been identified as of great interest for eggplant breeding due to their tolerance to biotic and abiotic stresses and high contents of bioactive compounds beneficial for human health (Syfert et al., 2016; Kaushik et al., 2017; Arafa et al., 2022; Toppino et al., 2022) and for some of them, introgressed and backcrossed populations have been obtained, while many other interesting GP2 species still unexploited in breeding hold great potential. Moreover, the two cultivated eggplants (*S. aethiopicum* and *S. macrocarpon*) are also valuable for eggplant breeding, since aside from presenting characteristics of interest for eggplant breeding, they display the typical traits associated to the domestication syndrome, which facilitates their use in breeding (Särkinen et al., 2013; Plazas et al., 2014).

Table 1. *Solanum* species from the primary (GP1), secondary (GP2) and tertiary (GP3) genepools (according to Syfert et al., 2016) for which introgression breeding with eggplant has been reported.

Species	Main traits of interest for eggplant breeding	Most advanced type of generations obtained with *S. melongena*	References
Primary genepool (GP1)			
S. insanum	Drought and salinity tolerance, phytochemical composition	Advanced backcrosses	Ranil et al. (2017); Brenes et al. (2020); Plazas et al. (2020); Nadeeshani et al. (2021); González-Orenga et al. (2023)
Secondary genepool (GP2)			
S. aethiopicum	Resistance or tolerance to *Fusarium* and bacterial wilts and nematodes, vigor of F1 hybrids as rootstocks, spider mite resistance	Lines with introgressed resistance to Fusarium and Verticillium wilt	Collonnier et al. (2001); Toppino et al. (2008); Prohens et al. (2012); Calvo-Asensio et al. (2014); Barbierato et al. (2016); Barchi et al. (2018); Taher et al. (2019); Zhuang & Wang (2009)
S. anguivi	Drought tolerance, high content of phenolics	Second backcross generation	Kaushik et al. (2019); Plazas et al. (2020); Kouassi et al. (2021)
S. dasyphyllum	Drought tolerance, two-spotted spider mite and silverleaf whitefly tolerance	Advanced backcrosses	Plazas et al. (2020); Kouassi et al. (2021); Taher et al. (2020); Villanueva et al. (2023)
S. incanum	Drought tolerance, bacterial wilt resistance, fruit and shoot borer resistance, silverleaf whitefly tolerance, high content of phenolics	Introgression lines	Bletsos and Olympios (2008); Prohens et al. (2013); Gramazio et al. (2017); Namisy et al. (2019); Mangino et al. (2020); Taher et al. (2020)
S. lichtensteinii	Drought tolerance, silverleaf whitefly tolerance	Second backcross generation	Vorontsova and Knapp (2016); Plazas et al. (2020); Taher et al. (2020)
S. lidii	Unexplored so far	Second backcross generation	Plazas et al. (2020)
S. linnaeanum	Salinity tolerance, Verticillium wilt resistance	Lines with introgressed resistance to Verticillium wilt	Mennella et al. (2010) Acciarri et al. 2007; Zhuang et al. (2014); Liu et al. (2015)
S. tomentosum	Fusarium and Verticillium wilts and nematodes resistance, silverleaf whitefly tolerance	Introgression lines	Toppino et al. (2018); Taher et al. (2020); Caliskan et al. (2023)
Tertiary genepool (GP3)			
S. elaeagnifolium	Drought tolerance, high content of phenolics	Advanced backcrosses	García-Fortea et al. (2019); Plazas et al. (2020); Villanueva et al. (2021)

Hybridization of eggplant with around 20 GP2 species has been achieved, including the *Anguivi* grade cultivated species *S. aethiopicum* and *S. macrocarpon*, as well as with *S. linnaeanum*, *S. incanum* and *S. tomentosum* (Daunay and Hazra, 2012; Särkinen et al., 2013; Rotino et al., 2014; Plazas et al., 2016; Daunay et al., 2019; Toppino et al., 2021). Different kinds of introgression materials were obtained with eggplant relatives from the GP2, mostly aimed at exploiting resistance traits to pathogens and adverse environmental conditions. The tertiary genepool (GP3) is genetically very diverse, including species found in Africa and Madagascar, as well as in Australia, Pacific Islands, Asia and in distant American species of subgenus *Leptostemonum* (**Figure 2**) (Knapp et al., 2013; Syfert et al., 2016). As expected, the success of hybridization of eggplant with GP3 species is very low, although attempts to obtain interspecific hybrids with eggplant have been successful in several cases, including the Madagascar species *S. pyracanthos* and the American *S. elaeagnifolium*, *S. sisymbriifolium*, *S. torvum*, and *S. viarum* (Rotino et al., 2014; Kouassi et al., 2016; Plazas et al., 2016; Daunay et al., 2019). In many cases, embryo rescue was necessary, especially in crosses with American species. Although interspecific hybrids between eggplant and American species are highly sterile, some backcrosses to eggplant were obtained when the interspecific hybrid with *S. elaeagnifolium* was used as maternal parent, suggesting the possibility to exploit previously untapped GP3 genetic material for introgression breeding (Plazas et al., 2016; García-Fortea et al., 2019).

Overall, the large genetic, phenotypic and physiological diversity present in the three genepools represents an enormous potential for eggplant breeding, which has been barely explored, particularly in the case of wild species (Daunay and Hazra, 2012; Rotino et al., 2014; Taher et al., 2017; Toppino et al., 2021, 2022; Arafa et al., 2022; Salinier et al., 2022). Unlocking this high diversity will be essential for developing new materials with adaptation to climate change and meeting the urgent need for an eggplant breeding revolution.

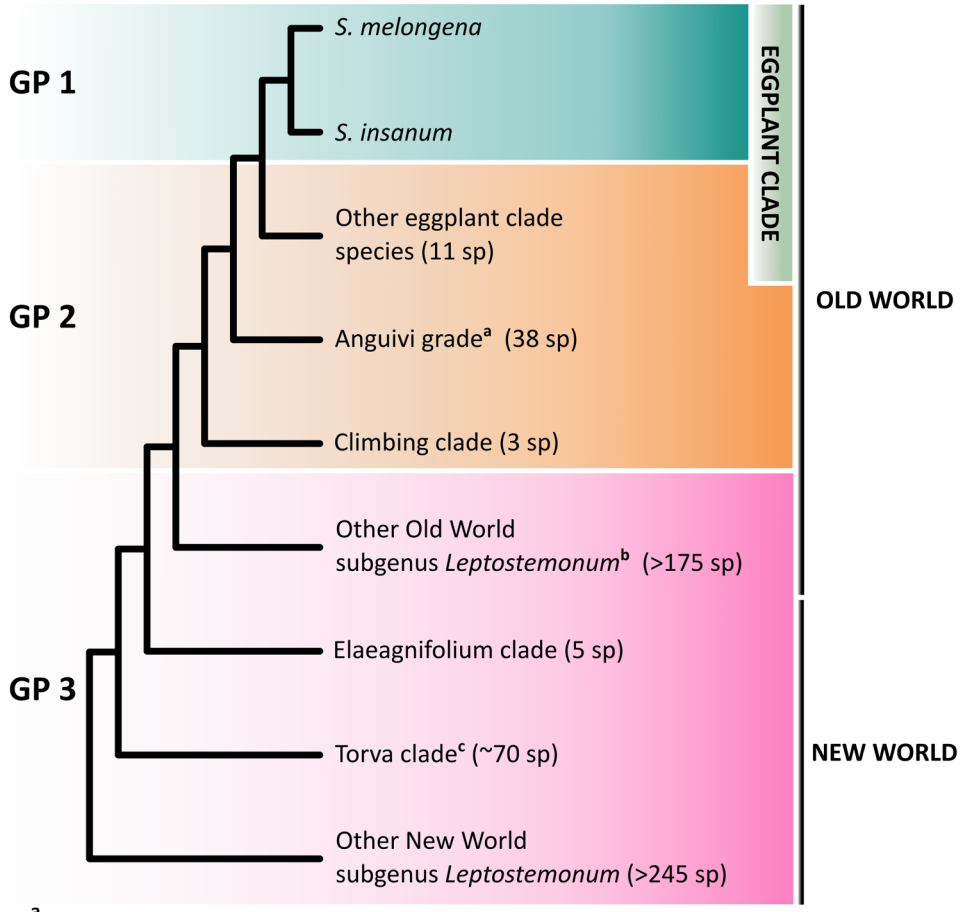


Figure 2. Dendrogram representing relationships of the most relevant groups of the primary (GP1), secondary (GP2) and tertiary (GP3) genepools of S. melongena. Based on. Whalen, 1984; Vorontsova et al., 2013; Aubriot et al., 2016; Knapp and Vorontsova, 2016 and Knapp et al., 2019.

Eggplant germplasm collections

Based on the recent Global Strategy for the Conservation and Use of Eggplants (Solberg et al., 2022), 19,020 accessions of cultivated eggplants and relatives are conserved in 110 germplasm banks and collections around the world (**Figure 3**) (FAO, 2010). The largest genebank collections of eggplant are conserved at the National Bureau of Plant Genetic Resources (India; 4,236 accessions), the World Vegetable Center (an international organization with eggplant germplasm collections headquartered in Taiwan; 3,036 accessions), the INRAE Genebank of

France (2,388 accessions), the National Genebank for Vegetable Germplasm Resources of China (1,601 accessions) and the NARO Genebank of Japan (1,501 accessions) (Taher et al., 2017; Salinier et al., 2022; Solberg et al., 2022).

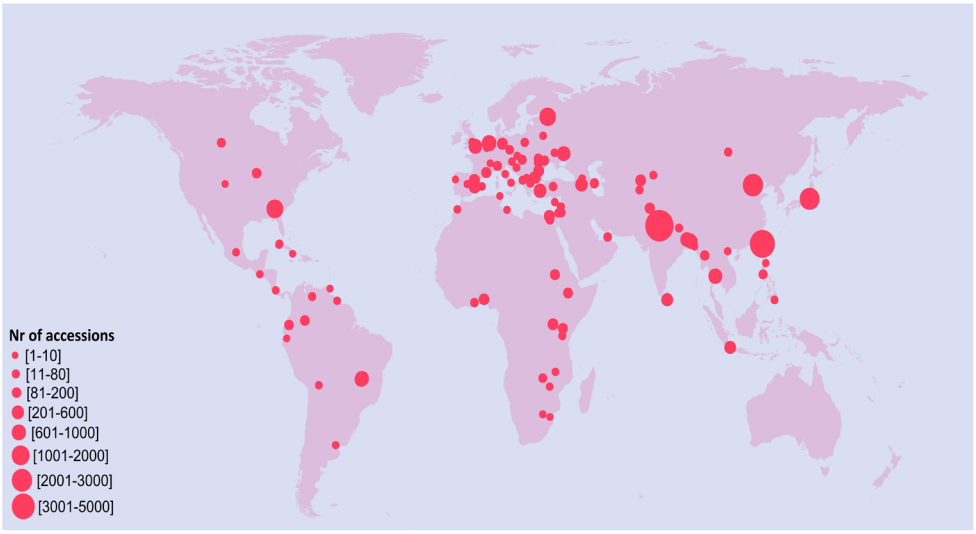


Figure 3. Map of global distribution of cultivated eggplant and its wild relatives in genebank holdings. Map elaborated according to data from FAO et al. (2010), Taher et al. (2017), Salinier et al. (2022) and Solberg et al. (2022).

When considering the Genesys (https://www.genesys-pgr.org/) and WIEWS (World Information and Early Warning System on Plant Genetic Resources for Food and Agriculture; https://www.fao.org/wiews/en/) global germplasm databases, most of the conserved materials of the eggplant genepools correspond to cultivated *S. melongena* (12,665 accessions), *S. aethiopicum* (1,004) and *S. macrocarpon* (208) while the wild species of the GP1, GP2 and GP3 genepools are much less represented (2,351 accessions in total) (Solberg *et al.*, 2022). Among the wild species, *S. incanum* is the most abundant (GP2; 423 accessions), followed by *S. torvum* (GP3; 358 accessions), *S. aculeatissimum* (GP3; 210 accessions), *S. virginianum* (GP2; 187 accessions) and *S. grandiflorum* (GP3; 184 accessions). However, apart from these five wild species, the number of remaining wild species accessions from GP2 and GP3 of eggplant is dramatically low, with just 14 species having more than 10 accessions conserved, while for many others no accessions are conserved at all (Solberg et al., 2022). This is particularly evident for the 14 eggplant CWRs classified as at risk of extinction (one critically endangered, nine threatened, three near threatened, and one extinct in the wild), for six of them (including *S. ruvu*, which is considered extinct in the wild) no accessions are conserved in germplasm banks and for the remaining, up to just four accessions are conserved *ex situ* (Syfert et al., 2016).

Relevant information for the *in situ* conservation*, i.e.* the on-site management of genetic resources, is available thanks to Syfert et al. (2016). The study identified hotspots of diversity of eggplant crop wild relatives in southern and eastern Africa and the Indian subcontinent. These hotspots, found in protected areas of Kenya, Tanzania, and Uganda, are potential areas of interest for establishing *in situ* conservation policies and collecting genetic resources to fill germplasm gaps in *ex situ* collections. However, few *in situ* programmes are ongoing. A total of five eggplant wild relatives (*S. lidii*, *S. linnaeanum*, *S. marginatum*, *S. sisymbriifolium*, and *S. torvum*) are included in the European priority CWR taxa (Rubio Teso et al., 2021), although none of them is native to continental Europe (Vorontsova *et al.*, 2013; Vorontsova and Knapp, 2016), and two (*S. sisymbriifolium* and *S. torvum*) are invasive (Alaniz et al., 2020; Musarella, 2020). Two of these species (*S. lidii* and *S. marginatum*) are found only in one European country, and specific conservation sites exist only for *S. lidii*, which is an endangered endemism of the Canary Islands (Gramazio et al., 2020; Rubio Teso et al., 2021).

The level of exploration of the cultivated eggplant germplasm is variable, depending on the traits. While passport data are available for most accessions conserved in germplasm banks, the availability of characterization data, generally performed using standardized descriptors such as those of Bioversity (IBPGR, 1990), UPOV (2002) or EGGNET (van der Weerden and Barendse, 2007), is much more limited. On the one hand, some phenotypic studies were performed using a large number of accessions (>150) and aiming at evaluating the morphological diversity of cultivated eggplant (Cericola et al., 2013; Kumar et al., 2013; Liu et al., 2018; Oladosu et al., 2021; Ro et al., 2022). These studies revealed a large diversity of morpho-agronomic characteristics in the cultivated eggplant genepool and provided relevant information for their utilization in breeding. Large screening for evaluation traits in germplasm collections of eggplant relatives is more limited. Seventy *S. aethiopicum* accessions, mostly belonging to the *gilo* group, were assessed for morpho-physiological traits, AFLP (amplified fragment length polymorphism) and SSR (simple sequence repeat) molecular markers and chlorogenic acid content, revealing a wide genetic diversity (Sunseri et al., 2010). A total of 125 accessions of *S. aethiopicum* and *S. macrocarpon* were evaluated by Taher et al. (2019) for resistance to the two-spotted spider mite (*Tetranychus urticae*), resulting in the identification of high levels of resistance in two accessions of *S. macrocarpon*. In another large evaluation study, Stommel and Whitaker (2003) studied the phenolic acid profiles of 115 accessions, mostly of cultivated *S. melongena*, but also including some accessions of *S. aethiopicum*, *S. anguivi*, *S. incanum* and *S. macrocarpon*. Another study on 73 accessions, most of which were of *S. melongena*, but also included *S. aethiopicum* and *S. macrocarpon*, also found large variations in total phenolics content (8.4-fold), and fruit flesh browning (7.3-fold), but less in ascorbic acid (2.3-fold) (Prohens et al., 2007).

Overall, given the large number of species in the GP2 and GP3 of eggplant, the Focused Identification of Germplasm Strategy (FIGS), which is based on the assumption that wild accessions growing in specific environments must have adaptive genes to these conditions (Street et al., 2016), might help in identifying putative species or accessions of interest for tolerance to a certain biotic or abiotic stress (Prohens et al., 2017). However, the exploration of eggplant and relatives germplasm collections for traits relevant to adaptation to climate change has been very scarce until now. To achieve a breakthrough in eggplant breeding, it is essential to systematically evaluate the available variation and identify sources of variation for adaptation to climate change.

Use of genetic resources in breeding: achievements and challenges

Selections of eggplants started very early in breeding, with accessions having improved characteristics already present in seed catalogues in the late 19th and early 20th centuries (Daunay and Janick, 2007). In addition, heterosis for yield was already reported in 1931 (Kakizaki, 1931), which opened the door for the development of hybrid varieties with improved features. Genetic improvements in eggplant have relied on the use of germplasm, and breeders have been using the eggplant germplasm (mostly of cultivated *S. melongena*) for breeding and developing new selections, lines and hybrids. According to a survey of germplasm banks (Solberg et al., 2022), the number of eggplant accessions distributed per year ranged between 0 and 503, revealing that some germplasm banks make a significant distribution to users, many of whom are breeders.

The genetic improvements of eggplant are evident in the characteristics of modern cultivars, which are considerably better in yield and overall quality than landraces. For example, by considering the western market modern F1 hybrid cultivars found in western markets have no prickles, greater earliness, intense black color and epidermal shininess, and lower fruit flesh browning (Prohens et al., 2007; Muñoz-Falcón et al., 2009) or increased yield (Sambandam, 1964; Rodríguez-Burruezo et al., 2008; Daunay and Hazra, 2012; Kaushik et al., 2018; Kumar et al., 2020). The development of modern eggplant cultivars has been mainly carried out employing the cultivated genepool. This resulted in a reduction of the genetic base of eggplant elite breeding lines and materials used for developing modern F1 hybrids. For instance, Muñoz-Falcón et al. (2009) evaluated the genetic diversity of black eggplants of different groups and found that modern F1 hybrids have a narrow genetic base and share a common genepool. This situation is in contrast to tomato, where the widespread use of CWRs, especially for introgressions of biotic resistance traits increased the genetic diversity of modern varieties (Díez and Nuez, 2008; Schouten et al., 2019). The exploitation of cultivated eggplant germplasm allowed the development of new cultivars and elite materials with improved resistance or tolerance to pests and diseases. Indeed, sources of resistance to the most significant

pests, including the eggplant fruit and shoot borer, leafhopper, aphids, spider mites, and whiteflies, as well as to the primary diseases such as bacterial wilt, Fusarium, and Verticillium wilts, have been identified (Taher et al., 2017; Arafa et al., 2022; Salinier et al., 2022). Many of these cultivated accessions have been transferred to researchers and breeders to incorporate them into their breeding pipelines (Taher et al., 2017). However, while some quantitative improvements have been achieved, resulting in cultivars with improved tolerance, the genetic diversity for resistance to these biotic stresses present in the primary genepool of eggplant seems to be limited (Taher et al., 2017).

Accessions of wild species of eggplant GP2 and GP3 species, as well as from the cultivated *S. aethiopicum*, have been employed for introgression breeding (Mennella et al., 2010; Liu et al., 2015; Gramazio et al., 2017; Plazas et al., 2020; Villanueva et al., 2021). Eggplant lines fully resistant to F. oxysporum f. sp. *melongenae* (Fom) have been obtained by introgressing the *Rfo-sa1* resistance locus from *S. aethiopicum* (Toppino et al., 2008b). Interestingly, the response mechanism to *Fom* inoculation triggered by this locus is also able to protect the plant from *Verticillium* wilt (Barbierato et al., 2016; Barchi et al., 2018) when the two fungi are used in a combined artificial inoculation. These elite *Fom*-resistant lines introgressed from *S. aethiopicum*, along with associated molecular markers, are of great interest for the development of commercial cultivars.

Solanum linnaeanum has also been used in introgression breeding for the development of early backcross eggplant materials with resistance to Verticillium wilt (Acciarri et al., 2004; Liu et al., 2015). However, no eggplant commercial cultivars with resistance derived from *S. linnaeanum* have been produced until now.

First backcross generations of eggplant with *S. aethiopicum* as a donor displayed a wide morphological variability (Prohens et al., 2012). Similarly, high morphological diversity and a wide range of values for phenolic acid contents were found in the first backcross generations using *S. incanum* as the donor parent (Prohens et al., 2013). However, introgression lines (ILs) derived from these early *S. incanum* backcrosses were largely similar to the recurrent parent, although two lines with higher plant vigour were identified (Mangino et al., 2020). Some advanced backcrosses with *S. elaeagnifolium* exhibited a higher yield than the recurrent *S. melongena* parent (Villanueva et al., 2021). However, these materials are still in an early stage of development and have not been used for the development of new cultivars. In addition, several species such as *S. aethiopicum*, *S. anguivi*, *S. grandiflorum*, *S. kurzii*, *S. violaceum* and *S. virginianum* have been used for the development of alloplasmic lines of eggplant that display cytoplasmic male sterility (Khan and Isshiki, 2016). These male-sterile lines have potential interest in the production of hybrids. However, to our knowledge, alloplasmic male sterility has not been used so far in the commercial production of eggplant hybrids.

Eggplant wild species and interspecific hybrids have also been explored for their use as rootstocks, mainly to obtain resistance to diseases and enhanced vigor (King

et al., 2010; Schwarz et al., 2010). In this way, apart from selections of *S. torvum* and interspecific hybrids between eggplant and *S. aethiopicum* used as commercial rootstocks, other wild species, such as *S. anguivi*, *S. incanum*, *S. insanum*, *S. palinacanthum* or *S. sisymbriifolium* (Gisbert et al., 2011; Rakha et al., 2020; Kumbar et al., 2021; Murata et al., 2022) were proposed as potential new rootstocks. However, no commercial rootstocks have been developed so far from these latter species. One potential reason is that some of these species with potential interest have prickly stems (Vorontsova et al., 2013; Vorontsova and Knapp, 2016), making the grafting procedure difficult and unusable for mass-scale grafting, typical of field cultivation (**Figure 4**).

It is worth considering that interspecific hybrids of eggplant with some wild species such as *S. tomentosum* or *S. elaeagnifolium* (**Figure 4**) are highly vigorous and have an extended root system (García-Fortea et al., 2019), making them exploitable as rootstocks. However, in some cases such as the hybrids between *S. melongena* and *S. elaeagnifolium*, the obtainment of hybrids is very challenging and requires embryo rescue (Kouassi et al., 2016), limiting their exploitation. Introgression breeding with wild species that display high contents of glycoalkaloids (solasonine and solamargine) might result in the inadvertent increase of these glycoalkaloids in the recurrent eggplant parents. However, studies performed by Mennella et al. (2010) with *S. aethiopicum* and *S. linnaeanum* and by Rosa-Martínez et al. (2022a) with *S. incanum* did not detect significantly higher levels of glycoalkaloids in ILs with these species than in the recurrent parents. However, given that these ILs did not represent the whole genome of the donor parents, the evaluation of glycoalkaloids should be performed in the elite materials obtained after the introgression process with eggplant relatives that exhibit high contents in potentially harmful glycoalkaloids. Similarly, given some toxic alkaloids synthesized in the roots can move in the epigeal part of the plant, in the case of using wild species or interspecific hybrids with high alkaloid contents as rootstocks, the content in the fruit should be checked to ensure the safety of the potential new commercial varieties.

Although considerable improvements have been made in eggplant breeding, the diversity used mostly relied on the cultivated eggplant *S. melongena* (Daunay and Hazra, 2012; Taher et al., 2017; Kumar et al., 2020). Furthermore, breeding efforts have been made in using eggplant CWRs for breeding, even if they did not have a major impact on the modern cultivars presently grown. As in other important crops, a qualitative leap forward in genetic advances in eggplant will require unleashing the huge potential of CWRs, which is still largely unexploited.

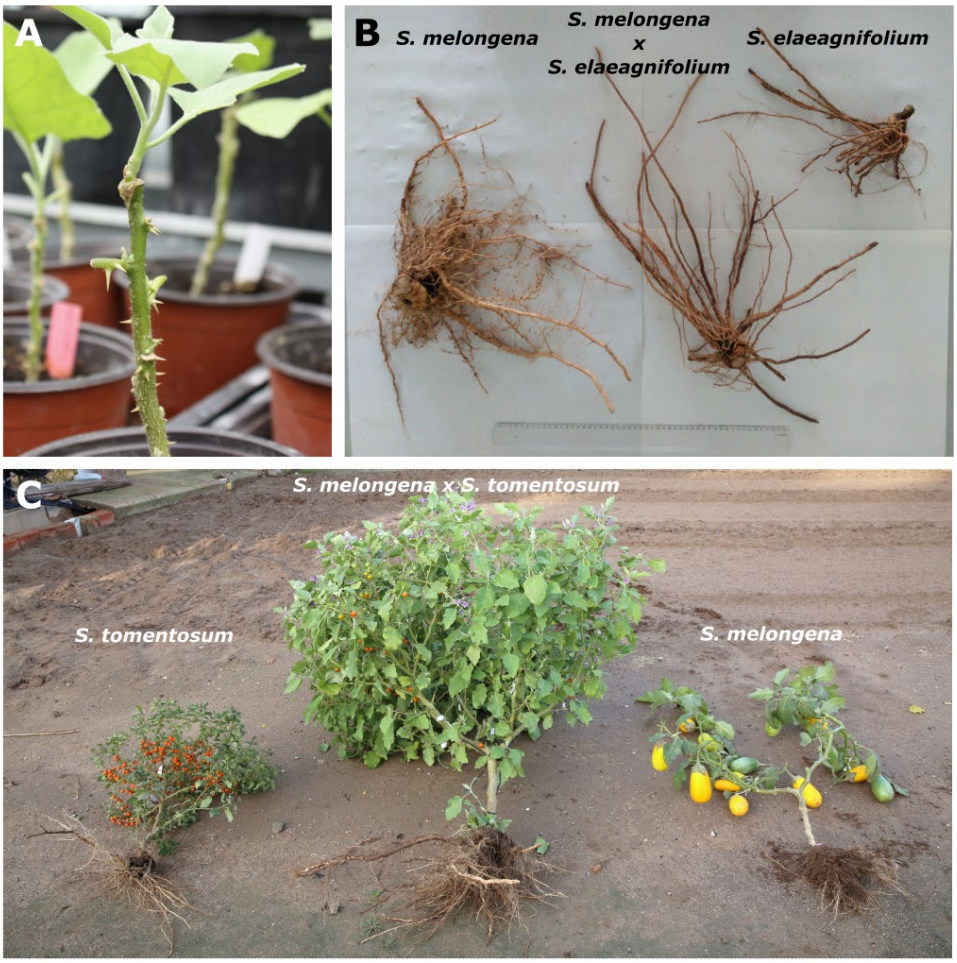


Figure 4. Unexploited eggplant wild relatives and interspecific hybrids as potential roortstocks for eggplant: highly prickly rootstocks are challenging for commercial rootstock utilization as prickles difficult the manual grafting process (above left; A); interspecific hybrids of eggplant (*S. melongena*) with some wild species such as *S. elaeagnifolium* (above, right; B) and *S. tomentosum* (below; C) are highly vigorous and/or have an extended root system which is great interest for improving resilience.

A new generation of genetic resources

Besides germplasm accessions of eggplant cultivated and CWRs, during the last years a new generation of eggplant genetic resources, consisting of core collections, recombinant inbred lines (RILs), and ILs (**Table 2**) have been generated (Toppino et al., 2008b, 2018, 2020; Gangopadhyay et al., 2010; Mennella et al., 2010; Lebeau et al., 2013; Gramazio et al., 2017; Barchi et al., 2018; Miyatake et al., 2019; Mishra et al., 2020; Arrones et al., 2022; Ro et al., 2022; Gaccione et al., 2022; Mangino et al.,

2022). These materials are considered immortal since they can be regenerated by selfing for seed propagation. This is in contrast to F2 and early backcross materials, for which several populations have been obtained in eggplant (Daunay and Hazra, 2012; Prohens et al., 2012, 2013; Clarke et al., 2014; Portis et al., 2014; Toppino et al., 2016; Boyaci et al., 2021; Qian et al., 2022), and where each individual has a variable degree of heterozygosis and can be thus maintained only by vegetative propagation.

We should also point out that few mutant collections exist so far for eggplant (Xi-ou et al., 2017; Du et al., 2022). Two ethyl methane sulfonate (EMS) mutant libraries of 400 and 790 M_2 lines (**Table 2**) were generated and used to identify mutants for phenotypic traits, including dwarf mutant plants (Xiao et al., 2016; Xi-ou et al., 2017; Lu et al., 2021; Du et al., 2022).

Core collections allow a representation of most of the diversity of large germplasm set in a reduced number of accessions (Odong et al., 2013) exploitable for genotype to phenotype studies. The first eggplant core collection of 181 eggplant accessions was developed by Gangopadhyay *et al.* (2010) from an original set of 1,798 accessions by using 14 morphological descriptors (**Table 2**). In a first attempt to apply a GWAS (genome-wide association study) approach in eggplant, Ge et al. (2013) were able to identify several phenotype/genotype associations related to eight fruit-related traits. Subsequently, a selected eggplant association panel of 191 selected accessions (Cericola et al., 2013), comprising a mixture of breeding lines, old varieties and landrace selections originating from Asia and the Mediterranean Basin, was genotyped and phenotyped. This allowed the identification and positioning of several marker/trait associations related to fruit, plant and leaf morphological traits relevant to eggplant breeding (Cericola et al., 2014; Portis et al., 2015) as well as to identify contrasting genotypes for nitrogen use efficiency (NUE) (Mauceri et al., 2020) and, most recently, to identify the gene networks responsible of such diversity (Mauceri et al., 2021).

Subsequently, Miyatake et al. (2019) genotyped 893 accessions, mostly from Asia, with 831 SNP and 50 SSR markers and established a core collection of 100 accessions (World Eggplant Core; WEC). More recently, a core collection of 288 accessions from an initial set of 587 accessions by using 52 SNP markers complemented with agro-morphological traits (Ro et al., 2022). The combination of both types of data resulted in the identification of significant associations of SNPs with six traits, which allowed the identification of several candidate genes. Another core collection of 322 *S. melongena* accessions was obtained from an original set of over 3,600 accessions (Gaccione et al., 2022), most of which were genotyped with the 5k probes eggplant SPET platform (Barchi et al., 2019a). This core collection has been re-sequenced and phenotyped at three locations for multiple agronomic and composition traits (Gaccione et al., 2022) and has already proved useful in identifying allelic variants for the *SmAPRR2* transcription factor responsible for chlorophyll pigmentation in the eggplant fruit peel (Arrones et al., 2022).

Table 2. New eggplant genetic resources, consisting of mutant libraries, core collections, biparental and multiparental recombinant inbred lines and introgression lines sets.

Plant material used	Number of lines or accessions	Conventional and biotechnological tools used for the development	Reference
Mutant libraries			
S. melongena accession E31-1	790	Ethyl methane sulfonate	Xi-ou et al. (2017)
S. melongena line 14-345	400	Ethyl methane sulfonate	Du et al. (2022)
Core collections			
1,798 accessions of *S. melongena*	181	14 morphological descriptors	Gangopadhyay et al. (2010)
392 accessions of *S. melongena*	191	314 SNPs, 33 morphological traits, NUE	Cericola et al. (2013), 2014, Portis et al. (2015), Mauceri et al. (2020);
893 accessions of *S. melongena*	100	831 SNPs and 50 SSRs	Miyatake et al. (2019)
587 accessions of *S. melongena*	288	52 SNPs and 17 agromorphological traits	Ro et al. (2022)
3,600 accessions of *S. melongena* and wild relatives	322	5k probes SPET platform	Gaccione et al. (2022)
Biparental recombinant inbred lines			
S. melongena lines MM378 and AG91-25	178 F6	AFLP, SSR and SRAP	Lebeau et al. (2013)
S. melongena lines 305E40 and 67/3	163 F7	GBS (10 k polymorphic markers)	Toppino et al. (2020)
S. melongena landrace Ramnagar Giant and *S. incanum* accession W-4	114 F8	282 polymorphic RAPD, ISSR, SCoT and SSR	Mishra et al. (2020)
Multiparental recombinant inbred lines			
Seven *S. melongena* accessions (MM1597, DH ECAVI, AN-S-26, H15, A0416, IVIA-371 and ASI-S-1) and one *S. incanum* accession (MM577)	420 (S3 MAGIC)	5k probes SPET platform	Mangino et al. (2022)
Introgression lines sets			
S. melongena lines 1F5(9), Dourga, Tal 1/1 and CCR3, two accessions of *S. aethiopicum* and one accession of *S. linnaeanum*	57	Selection for tolerance to Fusarium and Verticillium wilts	Acciarri *et al.* (2007) Mennella *et al.* (2010)
S. melongena AN-S-26 and *S. incanum* MM577	51	COSII, SSRs, SNPs (GBS and SPET)	Gramazio et al. (2017), Plazas et al. (2020)
S. melongena accession 67/3 and one *S. tomentosum* accession	90	HRM Molecular markers	Toppino et al. (2018)

RILs from bi-parental or multi-parental crosses are genetic resources of great relevance, as each of them is a different genetic mosaic of the parents (Arrones et al., 2020). Therefore, new genotypes of interest for breeding combining desirable characteristics present in the set of parents may be recovered in the set of RILs. In addition, in the absence of selection, bi-parental or multi-parental RIL sets do not present genetic structure, which makes them a powerful tool for the detection of major genes and QTLs involved in traits of interest (Cockram and Mackay, 2018). Several RIL populations of eggplant (**Table 2**), which have in common one eggplant relative (*S. aethiopicum* or *S. incanum*) in their pedigree, have been obtained from bi-parental crosses (Lebeau et al., 2013; Toppino et al., 2020). A first RIL population composed of 178 F6 lines was obtained by single seed descent from the F2 generation obtained after crossing an eggplant line (MM738) susceptible to bacterial wilt with a resistant breeding line (AG91-25) derived from the crossing of a resistant *S. melongena* and an *S. aethiopicum* accession (Lebeau et al., 2013). Genotyping of this RIL population with AFLP, SSR and SRAP (sequence-related amplified polymorphism) markers allowed the construction of a genetic map with 119 polymorphic markers in which a major dominant gene and several QTLs were detected. Interestingly, some RILs displayed better performance than the resistant parent (AG91-25) for some of the resistance traits evaluated (Lebeau et al., 2013). More recently, Toppino et al. (2020) developed a RIL population of 163 F7 lines derived from single seed descent of the F2 from the cross between eggplant lines '305E40' and '67/3'. The parent '305E40' derived from the repeated backcrossing of a doubled haploid of the somatic hybrid between *S. melongena* and *S. aethiopicum* to two eggplant lines and carries the *Rfo-sa1* gene from *S. aethiopicum*, which confers resistance to *Fom*, as well as tolerance to Verticillium wilt (Barbierato et al., 2016; Barchi et al., 2018; Toppino et al., 2018). This RIL population was sequenced at low coverage and employed to anchor the genome of the male parent '67/3' (Barchi et al., 2019b). More recently the same population was genotyped by GBS, resulting in over 10k polymorphic markers, which allowed the development of a high-density genetic map and the identification of a large number of QTLs, as well as candidate genes, for multiple morphological and metabolic traits (Toppino et al., 2020; Sulli et al., 2021), together with the characterization of two major QTLs for resistance to *Fom* (Tassone et al., 2022). Also, Mishra et al. (2020) developed a RIL population of 114 F8 RILs from the crossing between a cultivated landrace (Ramnagar Giant) and an accession of *S. incanum* (W-4), allowing the development of a genetic map after genotyping the population with 282 polymorphic RAPD (random amplified polymorphic DNA), ISSR (inter-simple sequence repeat), SCoT (start codon targeted) and SSR markers.

Following the intercrossing of eight parental lines (seven *S. melongena* of different origins and characteristics and one *S. incanum*), the only multiparental RIL population is a MAGIC (multi-parent advanced generation inter-cross) set of lines of eggplant (MEGGICS3), constituted of 420 S3 lines (**Table 2**) that were resequenced

at an average of an average depth of 20x (Gramazio et al., 2019). The MEGGICS3 population was developed following a funnel scheme and single seed descent from the S0 quadruple hybrid recombinant generation (Mangino et al., 2022) and has been genotyped with the eggplant 5k SPET panel (Barchi et al.), resulting in 7,724 high-confidence SNPs. The phenotyping of plant and fruit anthocyanic pigmentation as well as fruit peel chlorophyll presence has allowed the identification of several major QTLs and candidate genes for the traits evaluated (Arrones et al., 2022; Mangino et al., 2022). Interestingly, in combination with the core collection developed within the H2020 programme project G2P-SOL (Linking genetic resources, genomes and phenotypes of Solanaceous crops; https://www.g2p-sol.eu/), the MAGIC population allowed identifying the gene *SmAPRR2* as responsible for fruit chlorophyll pigmentation in the fruit peel (Arrones et al., 2022).

The occurrence of interspecific gene exchanges between eggplant and its relatives is a fundamental prerequisite for enlarging the eggplant genetic pool and, therefore, exploiting the variation resident in the allied germplasm for the improvement of the cultivated species. As an example, the occurrence of tetrasomic inheritance, including chromatid recombination, disclosed in the population of dihaploids obtained from anther culture of the tetraploid somatic hybrid between *S. melongena* and *S. aethiopicum* gr. Gilo (Toppino et al., 2008a) it opened up the possibility to introduce this material into advanced breeding programs. The first ILs of *S. melongena* with related species were obtained after recurrent backcrossing of two hybrids between *S. melongena* lines 1F5(9) and Dourga and *S. aethiopicum,* towards cultivated eggplant (Toppino et al., 2008b). Also, ILs were obtained after hybridization of several eggplant lines with *S. linnaeanum* (Mennella et al., 2010). In total, 57 ILs derived from these programmes after 6-7 cycles of backcrossing were studied for several health-related compounds and PPO activity (Mennella et al., 2010). The results revealed that both IL sets displayed similar glycoalkaloid levels to the recurrent parents, indicating their safety for human consumption, while a significant number of ILs displayed better values for antioxidant compounds. Subsequently, Gramazio et al. (2017) used marker-assisted selection in the repeated backcrossings (up to BC6) and subsequent selfings between *S. melongena* accession ANS26 and *S. incanum* accession MM577. This resulted in 25 ILs with single introgressions that covered 61.7% of the *S. incanum* genome, which was recently increased to over 70% of the *S. incanum* MM577 genome (Plazas et al., 2020). A subset of these ILs have been characterized for morphological and agronomic traits (Mangino et al., 2020; Rosa-Martínez et al., 2022b), fruit shape characteristics (Mangino et al., 2021), and composition (Rosa-Martínez et al., 2022a,b), highlighting several stable QTLs, as well as low levels of glycoalkaloids. Toppino et al. (2018) recently developed 90 ILs carrying introgressions from the wild relative *S. tomentosum*, which may be of great interest for breeding for resistance to several traits present in this wild relative, such as resistance to *Fusarium*, *Verticillium* or nematodes (Caliskan et al., 2023), as well as to whitefly (Taher et al., 2020). New

sets of ILs with *S. insanum*, *S. dasyphyllum* and *S. elaeagnifolium* are in advanced stages of development (Plazas et al., 2020) and will soon increase the diversity available to eggplant breeders from so far unexplored exotic genetic resources. Advanced backcrosses containing *S. elaeagnifolium* introgressions, screened under low N conditions have revealed a great potential for low-input agriculture (Villanueva et al., 2021).

These new generations of genetic resources make extant eggplant genetic diversity more accessible to breeders, allowing the development of new recombinant genotypes and representing powerful tools for identifying genes/alleles and QTLs associated with traits of interest, including complex traits such as those related to climate change (Prohens et al., 2017; Chapman, 2020). The extended use of these materials has already started to demonstrate their tremendous potential for eggplant breeding (Lebeau et al., 2013; Barchi et al., 2018; Mangino et al., 2020, 2022; Arrones et al., 2022).

Genomic and biotechnological tools to enhance the exploitation of genetic resources

New genomic tools such as high-throughput genotyping derived from NGS technologies, reference genomes, pangenomes and resequencing projects, can efficiently contribute to the enhancement of eggplant genetic resources and are essential for the breakthrough in eggplant breeding (Gramazio et al., 2018; Lanteri and Barchi, 2019; Simko et al., 2021). Although DNA molecular markers of different types, such as RAPDs, AFLPs and SSRs have been widely used for eggplant genotyping and genetic mapping since the early 1990s (Collonnier et al., 2001; Gramazio et al., 2014, 2018), the availability of NGS technologies allowed an easier genotyping of large sets of accessions and experimental populations with hundreds to thousands of markers, contributing to the evaluation of the eggplant and CWRs genetic diversity, the establishment of genetic relationships of germplasm sets and identification of QTLs (Barchi et al., 2019c; Liu et al., 2019; Miyatake et al., 2019; Toppino et al., 2020; Sulli et al., 2021; Mangino et al., 2022; Ro et al., 2022; Tassone et al., 2022; Gaccione et al., 2023), which is of interest in identifying materials for breeding and germplasm management (Lanteri and Barchi, 2019; Arafa et al., 2022; Toppino et al., 2022).

A first draft of the eggplant genome was published in 2014 (Hirakawa et al., 2014), but chromosome-anchored eggplant genome assemblies have not been available until recently (Barchi et al, 2019b; Wei et al., 2020; Barchi et al., 2021, 2022; Li et al., 2021) and this has delayed the application of the potential of resequencing and pangenome projects to enhance the management of eggplant genetic resources. The availability of resequencing data from eight accessions (Gramazio et al., 2019) allowed the development of an eggplant 5k.

SPET platform (Barchi et al., 2019a), which is the first specific eggplant genotyping platform. This platform has been used for the genotyping of germplasm of eggplant and wild relatives and the first MAGIC population (Barchi et al., 2019a, 2022; Gramazio et al., 2020; Arrones et al., 2022) as well as the marker-assisted selection for the development of ILs (Plazas et al., 2020; Villanueva et al., 2021). The first eggplant pangenome, which included the resequencing data of 23 accessions of *S. melongena* and two of CWRs (*S. incanum* and *S. insanum*) is very recent and allowed the identification of additional genes compared to the reference genome used, as well as of selective sweeps during domestication and the associated underlying candidate genes (Barchi et al., 2021).

Genebank genomics can help in the management and utilization of eggplant germplasm collections (Mascher et al., 2019), but so far no studies have been performed on eggplant. Similarly, the potential of landscape genomics (Li et al., 2017) to identify materials of eggplant with adaptive genes to specific environmental conditions has not been exploited yet. Both genomics approaches have a lot of potential for contributing to the eggplant breeding revolution. The genetic/genomic data and the phenotypic information available on the eggplant genetic resources (i.e. core collection and experimental populations) might lay the foundation to start applying genome-enable prediction methods to both accelerate eggplant breeding and increase the efficiency of the selection processes.

New Plant Breeding Techniques (NPBTs) such as those based on CRISPR (clustered regularly interspaced short palindromic repeats) and Cas (CRISPR-associated protein) genome editing represent valuable tools useful to create novel genetic variation as well as to determine the function of genes via targeted mutagenesis. However, only two studies have been published so far on CRISPR/Cas gene editing in eggplant (Maioli et al., 2020; Kodackattumannil et al., 2023), probably as a consequence of the recalcitrance of *S. melongena* to *in vitro* regeneration (García-Fortea et al., 2020) and genetic transformation. In the study of Maioli et al. (2020), polyphenol oxidase (PPO) genes *PPO4*, *PPO5*, and *PPO6* were knocked out, which resulted in reduced fruit flesh browning demonstrating how the creation of new allelic variation contributed to the improvement of an important trait. In a subsequent study, Kodackattumannil et al. (2023) found that CRISPR/Cas mutation of *PPO2* resulted in the inhibition of fruit flesh browning, but also had multiple pleiotropic effects in morphological and agronomic traits.

Conclusions and future perspectives

The exploration of the cultivated and wild eggplant germplasm both at the phenotypic and molecular levels is required for the identification of sources of variation for new traits barely explored so far, such as tolerance to new stresses caused by climate change as well as for improved sustainability, such as water and fertilizers use efficiencies. The establishment of core collections, such as the ones

already existing (Gangopadhyay et al., 2010; Miyatake et al., 2019; Gaccione et al., 2022; Ro et al., 2022), as well as the FIGS approach of identification of potentially useful germplasm (Street et al., 2016), genebank and landscape genomics (Li et al., 2017; Mascher et al., 2019) may help in facilitating the identification accessions of interest carrying novel and/or superior alleles.

The eggplant breeding revolution requires coordinated activities and programs for an improved exploitation of its genetic resources enhancement of its genetic resources. Two major international initiatives, the "Adapting Agriculture to Climate Change" (2011-2021) initiative of the Global Crop Diversity Trust (Dempewolf *et al.*, 2014), and the H2020 project G2P-SOL (2016-2021) have demonstrated the enormous potential of international collaboration in the improved conservation and utilization of eggplant genetic resources. In this way, the eggplant activities and projects performed under the "Adapting Agriculture to Climate Change" (2011-2021) initiative allowed the identification of gaps in the eggplant CWRs germplasm collections and proposed priorities for collection and *in situ* conservation (Syfert et al., 2016), as well as the collection of 474 new accessions of eggplant and CWRs (32 different species) for the completion of these gaps (Eastwood et al., 2022). Also, this initiative allowed the development of advanced backcrosses and ILs with four different eggplant CWRs (*S. dasyphyllum*, *S. elaeagnifolium*, *S. incanum* and *S. insanum*) (Gramazio et al., 2017; Plazas et al., 2020; Villanueva et al., 2021). Characterization of these sets of ILs for multiple traits and stress conditions is expected to result in new materials with improved adaptation to climate change (García-Fortea et al., 2019; Plazas et al., 2020). The "Adapting Agriculture to Climate Change" initiative has also contributed to the development of the Germinate platform (Raubach et al., 2021), which includes a database on eggplant (https://ics.hutton.ac.uk/cwr/eggplant) that contains 59 datasets with genotypic and phenotypic data from cultivated eggplant, wild species and pre-breeding materials.

The H2020 project G2P-SOL represents another landmark for the enhancement of genetic resources of *S. melongena* and the species of its genepools, in which the 5k SPET platform was developed (Barchi et al., 2019a) and used for the largest genotyping effort in eggplant germplasm (around 3,500 accessions), allowing the evaluation of diversity of the eggplant genepool, establishment of relationships, identification of duplicates, and in combination with historical characterization data the identification of hundreds of QTLs (Barchi et al., 2022; Gaccione et al., 2022). By using these data, a core collection of 322 eggplant accessions was created, which has been resequenced and phenotyped in multiple locations as well as evaluated for several biotic (*Fusarium* wilt, *Verticillium* wilt, *Meloidogyne* nematodes) and abiotic (drought tolerance and salinity tolerance) related to climate change (Gaccione et al., 2022; Salinier et al., 2022). In addition, fruit metabolomic analyses of the core collection have been performed (Sulli et al., in preparation).

Although these two initiatives represented a starting point for the eggplant breeding revolution, new international, preferably global, actions are needed for a coordinated and systematic exploitation of the advances obtained so far. In this way, the Global Strategy for the Conservation and Use of Eggplants (Solberg et al., 2022) calls for seven priority activities: (i) establishing a global eggplant working group, (ii) developing an Eggplant Knowledge Platform; (iii) improve passport data accuracy and completeness in the collection databases; (iv) facilitate and encourage collaborative plant health-related activities; (v) support collaborative activities associated with accessions regeneration and safety duplication, (vi) characterize the global eggplant collection morphologically and genetically, and (vii) encourage collaborative efforts to involve CWR in breeding programmes. However, the implementation of this strategy, which would represent an additional boost for an eggplant breeding revolution is still in the phase of funding acquisition for its effective fulfilment. In any case, the foundations are set for the international networks already established, together with new projects and developments in the fields of genomics and biotechnology, to bring forward the eggplant breeding revolution. As for other crops, we foresee this will result in dramatic genetic improvements in eggplant cultivars that will increase yield and quality and will allow the development of more resilient materials able to cope with climate change challenges.

Conflict of interest

The authors declare that there is no conflict of interest.

Funding

This work was supported by grants CIPROM/2021/020 from Conselleria d'Innovació, Universitats, Ciència i Societat Digital (Generalitat Valenciana, Spain), PID2021-128148OB-I00 funded by MCIN/AEI/10.13039/501100011033/ and by "ERDF A way of making Europe", PDC2022-133513-I00 funded by MCIN/AEI/10.13039/501100011033/, and by "European Union NextGenerationEU/PRTR", by Grant Agreement No. 677379 (G2P-SOL project: Linking genetic resources, genomes and phenotypes of Solanaceous crops) from European Union's Horizon 2020 Research and Innovation Programme, by the Grant Agreement No. 101094738 (PRO-GRACE project: Promoting a Plant Genetic Resource Community for Europe) from the European Union's Horizon Europe programme, as well as by the initiative "Adapting Agriculture to Climate Change: Collecting, Protecting and Preparing Crop Wild Relatives", which is supported by the Government of Norway. This latter project is managed by the Global Crop Diversity Trust with the Millennium Seed Bank of the Royal Botanic Gardens, Kew and implemented in partnership with national and international gene banks and plant

breeding institutes around the world. For further information, see the project website: http://www.cwrdiversity.org/. The overall work also partially fulfils some goals of the Agritech National Research Center and received funding from the European Union Next-Generation EU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR)–MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.4—D.D. 1032 17/06/2022, CN00000022). In particular, this study represents a review paper within: Spoke 4 (Task 4.1.1.) 'Next-generation genotyping and -omics technologies for the molecular prediction of multiple resilient traits in crop plants'; Spoke 1 (Task 1.2.1 Linking phenotype and genotype: discovery of loci/genes/alleles for traits of interest) and Spoke 2 (Task 2.2.1: 'Improved genetic materials to reduce the use of agrochemicals'). Pietro Gramazio is grateful to Spanish Ministerio de Ciencia e Innovación for a post-doctoral grant (RYC2021-031999-I) funded by MCIN/AEI /10.13039/501100011033 and the European Union through NextGenerationEU/PRTR. Andrea Arrones is grateful to Spanish Ministerio de Ciencia, Innovación y Universidades for a pre-doctoral (FPU18/01742) contract. Gloria Villanueva is grateful to Spanish Ministerio de Ciencia e Innovación for a pre-doctoral grant (PRE2019-103375) funded by MCIN/AEI /10.13039/501100011033.

References

Acciarri N, Rotino GL, Sabatini E, Valentino D, Sunseri F, Mennella G, Tamietti G. 2004. Improvement of eggplants for resistance to *Verticillium*. Proceedings of the XIIth Meeting on Genetics and Breeding of Capsicum and Eggplant. Noordwijkerhout, The Netherlands, 178.

Acquadro A, Barchi L, Gramazio P, Portis E, Vilanova S, Comino C, Plazas M, Prohens J, Lanteri S. 2017. Coding SNPs analysis highlights genetic relationships and evolution pattern in eggplant complexes. PLoS ONE 12, 1–20.

Alaniz AJ, Núñez-Hidalgo I, Carvajal MA, Alvarenga TM, Gómez-Cantillana P, Vergara PM. 2020. Current and future spatial assessment of biological control as a mechanism to reduce economic losses and carbon emissions: the case of *Solanum sisymbriifolium* in Africa. Pest Management Science 76, 2395–2405.

Arafa RA, Prohens J, Solberg SØ, Plazas M, Rakha M. 2022. Breeding and genome mapping for resistance to biotic stress in eggplant. In: Kole C, ed. Genomic designing for biotic stress resistant vegetable crops. Springer, 147–188.

Arrones A, Mangino G, Alonso D, Plazas M, Prohens J, Portis E, Barchi L, Giuliano G, Vilanova S, Gramazio P. 2022. Mutations in the *SmAPRR2* transcription factor suppressing chlorophyll pigmentation in the eggplant fruit peel are key drivers of a diversified colour palette. Frontiers in Plant Science 13, 1–14.

Arrones A, Vilanova S, Plazas M, Mangino G, Pascual L, Díez MJ, Prohens J, Gramazio P. 2020. The dawn of the age of multi-parent magic populations in plant breeding: Novel powerful next-generation resources for genetic analysis and selection of recombinant elite material. Biology 9, 1–25.

Aubert S, Daunay MC, Pochard E. 1989. Saponosides stéroïdiques de l'aubergine (*Solanum*

melongena L.) I. Intérêt alimentaire, méthodologie d'analyse, localisation dans le fruit. Agronomie **9**, 641–651.

Aubriot X, Singh P, Knapp S. 2016. Tropical Asian species show that the Old World clade of 'spiny solanums' (*Solanum* subgenus *Leptostemonum* pro parte: Solanaceae) is not monophyletic. Botanical Journal of the Linnean Society 181, 199–223.

Ayenan MAT, Danquah A, Hanson P, Ampomah-Dwamena C, Sodedji FAK, Asante IK, Danquah EY. 2019. Accelerating breeding for heat tolerance in tomato (*Solanum lycopersicum* L.): An integrated approach. Agronomy 9, 720.

Barbierato V, Toppino L, Rinaldi P, Sala T, Bassolino L, Valè G, Ferrarini A, Delledonne M, Bagnaresi P, Rotino GL. 2016. Phenotype and gene expression analyses of the Rfo-sa1 resistant aubergine interaction with Fusarium oxysporum f. sp. *melongenae* and Verticillium dahliae. Plant Pathology 65, 1297–1309.

Barchi L, Acquadro A, Alonso D, Aprea G, Bassolino L, Demurtas O, Ferrante P, Gramazio P, Mini P, Portis E, Scaglione D, Toppino L, Vilanova S, Díez MJ, Rotino GL, Lanteri S, Prohens J, Giuliano G. 2019*a*. Single Primer Enrichment Technology (SPET) for high-throughput genotyping in tomato and eggplant germplasm. Frontiers in Plant Science **10**, 1005.

Barchi L, Pietrella M, Venturini L, *et al.* 2019*b*. A chromosome-anchored eggplant genome sequence reveals key events in Solanaceae evolution. Scientific Reports **9**, 11769.

Barchi L, Portis E, Lanteri S, Alonso D, Diez MJ, Prohens J, Lefebvre V, Salinier J, Boyaci HF, Toppino L. 2022. Worldwide population structure of eggplant identified by SPET genotyping over 3,400 accessions. Virtual Plant and Animal Genome XXXIX Conference. San Diego, California, USA, .

Barchi L, Portis E, Toppino L, Rotino GL. 2019*c*. Molecular mapping, QTL identification, and GWA analysis. In: Chapman MA, ed. The Eggplant Genome. Springer, 41–54.

Barchi L, Rabanus-Wallace MT, Prohens J, Toppino L, Padmarasu S, Portis E, Rotino GL, Stein N, Lanteri S, Giuliano G. 2021. Improved genome assembly and pan-genome provide key insights into eggplant domestication and breeding. Plant Journal 107, 579–596.

Barchi L, Toppino L, Valentino D, Bassolino L, Portis E, Lanteri S, Rotino GL. 2018. QTL analysis reveals new eggplant loci involved in resistance to fungal wilts. Euphytica **214**, 1–15.

Barik S, Reddy AC, Ponnam N, Kumari M, C AG, Reddy D C L, Petikam S, Gs S. 2020. Breeding for bacterial wilt resistance in eggplant (*Solanum melongena* L.): Progress and prospects. Crop Protection **137**, 105270.

Bletsos FA, Olympios CM. 2008. Rootstocks and grafting of tomatoes, peppers and eggplants for soil-borne disease resistance, improved yield and quality. The European Journal of Plant Science and Biotechnology **2**, 62–73.

Bletsos F, Roupakias D, Tsaktsira M, Scaltsoyjannes A. 2004. Production and characterization of interspecific hybrids between three eggplant (*Solanum melongena* L.) cultivars and *Solanum macrocarpon* L. Scientia Horticulturae **101**, 11–21.

Boyaci HF, Kabas A, Aysan Y, Prohens J. 2021. Screening of eggplant genotypes for resistance to

bacterial wilt disease caused by *Clavibacter michiganensis* subsp. *michiganensis*. Plant Protection Science **57**, 112–121.

Brenes M, Solana A, Boscaiu M, Fita A, Vicente O, Calatayud Á, Prohens J, Plazas M. 2020. Physiological and biochemical responses to salt stress in cultivated eggplant (*Solanum melongena* L.) and in *S. insanum* L., a close wild relative. Agronomy **10**.

Caliskan S, Toppino L, Boyaci HF, Rotino GL, Cebeci E. 2023. Assessment of new genetic resources to uncover potential nematode resistance traits for eggplant (*Solanum melongena*) improvement. Phytoparasitica, 1–14.

Calvo-Asensio I, Prohens J, Gisbert C. 2014. Vigor for in vitro culture traits in *S. melongena* × *S. aethiopicum* hybrids with potential as rootstocks for eggplant. The Scientific World Journal **702071**, 1–8.

Cericola F, Portis E, Lanteri S, Toppino L, Barchi L, Acciarri N, Pulcini L, Sala T, Rotino GL. 2014. Linkage disequilibrium and genome-wide association analysis for anthocyanin pigmentation and fruit color in eggplant. BMC Genomics **15**, 896.

Cericola F, Portis E, Toppino L, Barchi L, Acciarri N, Ciriaci T, Sala T, Rotino GL, Lanteri S. 2013. The population structure and diversity of eggplant from Asia and the mediterranean basin. PLoS ONE **8**, e73702.

Chapman MA. 2020. Eggplant breeding and improvement for future climates. Genomic designing of climate-smart vegatable crops. Springer, 257–276.

Clarke CR, Hayes BW, Runde BJ, Wicker E, Vinatzer BA. 2014. Eggplant and related species are promising genetic resources to dissect the plant immune response to *Pseudomonas syringae* and *Xanthomonas euvesicatoria* and to identify new resistance determinants. Molecular Plant Pathology **15**, 814–822.

Cockram J, Mackay I. 2018. Genetic mapping populations for conducting high-resolution trait mapping in plants. In: Varshney RK, Pandey MK, Chitikineni A, eds. Plant genetics and molecular biology. Springer, 109–138.

Collonnier C, Fock I, Kashyap V, Rotino GL, Daunay MC, Lian Y, Mariska IK, Rajam M V., Servaes A, Ducreux G, Sihachakr D. 2001. Applications of biotechnology in eggplant. Plant Cell, Tissue and Organ Culture **65**, 91–107.

Daunay M-C, Aubert S, Frary A, Doganlar S, Lester RN, Barendse G, van der Weerden G, Hennart J-W, Haanstra J, Dauphin F, Evelyne J. 2004. Eggplant (*Solanum melongena*) fruit colour: pigments, measurements and genetics. Proceedings of the XIIth EUCARPIA meeting on genetics and breeding of capsicum and eggplant. Noordwijkerhout, The Netherlands: Plant Research International, 108–116.

Daunay M-C, Hazra P. 2012. Eggplant. In: Peter KV, Hazra P, eds. Handbook of vegetables. Studium Press, Houston, TX, USA, 257–322.

Daunay M-C, Janick J. 2007. History and iconography of eggplant. Chronica Horticulturae **47**, 16–22.

Daunay M-C, Salinier J, Aubriot X. 2019. Crossability and diversity of eggplants and their wild

relatives. The Eggplant Genome. Springer, 135–191.

Davidar P, Snow AA, Rajkumar M, Pasquet R, Daunay MC, Mutegi E. 2015. The potential for crop to wild hybridization in eggplant (*Solanum melongena*; Solanaceae) in Southern India. American Journal of Botany **102**, 129–139.

Dempewolf H, Eastwood RJ, Guarino L, Khoury CK, Müller J V., Toll J. 2014. Adapting agriculture to climate change: a global initiative to collect, conserve, and use crop wild relatives. Agroecology and Sustainable Food Systems **38**, 369–377.

Díaz-Pérez JC, Eaton TE. 2015. Eggplant (*Solanum melongena* L.) Plant growth and fruit yield as affected by drip irrigation rate. HortScience **50**, 1709–1714.

Díez MJ, Nuez F. 2008. Tomato. In: Prohens J, Nuez F, eds. Vegetable II. Handbook of Plant Breeding, vol 2. New York, NY: Springer, 249–326.

Docimo T, Francese G, De Palma M, Mennella D, Toppino L, Lo Scalzo R, Mennella G, Tucci M. 2016. Insights in the fruit flesh browning mechanisms in *Solanum melongena* genetic lines with opposite postcut behavior. Journal of Agricultural and Food Chemistry **64**, 4675–4685.

Du W, Lu Y, Luo S, Yu P, Shen J, Wang X, Xuan S, Wang Y, Zhao J, Li N, Chen X, Shen S. 2022. Genome-wide transcriptome analysis reveals that upregulated expression of *Aux/IAA* genes is associated with defective leaf growth of the *slf* mutant in eggplant. Agronomy **12**, 2647.

Eastwood RJ, Tambam BB, Aboagye LM, *et al.* 2022. Adapting agriculture to climate change: a synopsis of coordinated national crop wild relative seed collecting programs across five continents. Plants **11**, 1840.

EU Plant Variety Database. 2022. EU Plant Variety Database.

FAO. 2010. The second report on the state of the world's animal genetic resources for food and agriculture.

FAOSTAT. 2023. FAOSTAT database collections.

Florio FE, Gattolin S, Toppino L, Bassolino L, Fibiani M, Lo Scalzo R, Rotino GL. 2021. A SmelAAT acyltransferase variant causes a major difference in eggplant (*Solanum melongena* L.) peel anthocyanin composition. International Journal of Molecular Sciences **22**, 9174.

Gaccione L, Barchi L, Toppino L, Tumino G, Alonso D, Boyaci HF, Prohens J, Lanteri S, Portis E, Rotino GL, Giuliano G. 2022. Genome wide association analysis of agronomically relevant traits in an eggplant core collection representative of the worldwide genetic variation. Proceedings of the LXV SIGA Annual Congress. Piacenza, Italy, .

Gaccione L, Martina M, Barchi L, Portis E. 2023. A compendium for novel marker-based breeding strategies in eggplant. Plants **12**, 1016.

Gangopadhyay KK, Mahajan RK, Kumar G, Yadav SK, Meena BL, Pandey C, Bisht IS, Mishra SK, Sivaraj N, Gambhir R, Sharma SK, Dhillon BS. 2010. Development of a core set in brinjal (*Solanum melongena* L.). Crop Science **50**, 755–762.

García-Fortea E, Gramazio P, Vilanova S, Fita A, Mangino G, Villanueva G, Arrones A, Knapp S, Prohens J, Plazas M. 2019. First successful backcrossing towards eggplant (*Solanum melongena*) of a New World species, the silverleaf nightshade (*S. elaeagnifolium*), and characterization of interspecific hybrids and backcrosses. Scientia Horticulturae **246**, 563–573.

García-Fortea E, Lluch-Ruiz A, Pineda-Chaza BJ, García-Pérez A, Bracho-Gil JP, Plazas M, Gramazio P, Vilanova S, Moreno V, Prohens J. 2020. A highly efficient organogenesis protocol based on zeatin riboside for in vitro regeneration of eggplant. BMC Plant Biology **20**, 1–16.

Ge H, Liu Y, Jiang M, Zhang J, Han H, Chen H. 2013. Analysis of genetic diversity and structure of eggplant populations (*Solanum melongena* L.) in China using simple sequence repeat markers. Scientia Horticulturae **162**, 71–75.

Gimeno-Páiz E, Prohens J, Moreno-Cerveró M, Luis-margarit A De, María José Díez, Gramazio P. 2023. Agronomic treatments combined with embryo rescue for rapid generation advancement in tomato speed breeding. bioRxiv 2023.02.28.530438.

Gisbert C, Prohens J, Raigón MD, Stommel JR, Nuez F. 2011. Eggplant relatives as sources of variation for developing new rootstocks: Effects of grafting on eggplant yield and fruit apparent quality and composition. Scientia Horticulturae **128**, 14–22.

González-Orenga S, Plazas M, Ribera E, Pallotti C, Boscaiu M, Prohens J, Vicente O, Fita A. 2023. Transgressive biochemical response to water stress in interspecific eggplant hybrids. Plants **12**, 1–19.

Gonzalvo CM, Florendo Aala WJ, Maharjan KL. 2022. Is implementing a biotech ban correct or not? Analysis of farmer perceptions and attitudes on the Philippine supreme court's ban on biotech crops. Sustainability (Switzerland) **14**, 7919.

Gramazio P, Jaén-Molina R, Vilanova S, Prohens J, Marrero Á, Caujapé-Castells J, Anderson GJ. 2020. Fostering conservation via an integrated use of conventional approaches and high-throughput SPET genotyping: A case study using the endangered Canarian endemics *Solanum lidii* and *S. vespertilio* (Solanaceae). Frontiers in Plant Science **11**, 1–17.

Gramazio P, Prohens J, Plazas M, Andjar I, Herraiz FJ, Castillo E, Knapp S, Meyer RS, Vilanova S. 2014. Location of chlorogenic acid biosynthesis pathway and polyphenol oxidase genes in a new interspecific anchored linkage map of eggplant. BMC Plant Biology **14**, 350.

Gramazio P, Prohens J, Plazas M, Mangino G, Herraiz FJ, García-Fortea E, Vilanova S. 2018. Genomic tools for the enhancement of vegetable crops: A case in eggplant. Notulae Botanicae Horti Agrobotanici Cluj-Napoca **46**, 1–13.

Gramazio P, Prohens J, Plazas M, Mangino G, Herraiz FJ, Vilanova S. 2017. Development and genetic characterization of advanced backcross materials and an introgression line population of *Solanum incanum* in a *S. melongena* background. Frontiers in Plant Science **8**, 1477.

Gramazio P, Yan H, Hasing T, Vilanova S, Prohens J, Bombarely A. 2019. Whole-genome resequencing of seven eggplant (*Solanum melongena*) and one wild relative (*S. incanum*) accessions provides new insights and breeding tools for eggplant enhancement. Frontiers in Plant Science **10**, 1220.

Guan W, Ke C, Tang W, Jiang J, Xia J, Xie X, Yang M, Duan C, Wu W, Zheng Y. 2022. Construction

of a high-density recombination bin-based genetic map facilitates high-resolution mapping of a major QTL underlying anthocyanin pigmentation in eggplant. International Journal of Molecular Sciences **23**, 10258.

Habib-ur-Rahman M, Ahmad A, Raza A, Hasnain MU, Alharby HF, Alzahrani YM, Bamagoos AA, Hakeem KR, Ahmad S, Nasim W, Ali S, Mansour F, EL Sabagh A. 2022. Impact of climate change on agricultural production; Issues, challenges, and opportunities in Asia. Frontiers in Plant Science **13**, 925548.

He YJ, Li SH, Dong YX, Zhang XT, Li DL, Liu Y, Chen HY. 2022. Fine mapping and characterization of the dominant gene *SmFTSH10* conferring non-photosensitivity in eggplant (*Solanum melongena* L.). Theoretical and Applied Genetics **135**, 2187–2196.

Hedden P. 2003. The genes of the Green Revolution. Trends in Genetics **19**, 5–9.

Heuer B, Meiri A, Shalevet J. 1986. Salt tolerance of eggplant. Plant and Soil **95**, 9–13.

Hirakawa H, Shirasawa K, Miyatake K, Nunome T, Negoro S, Ohyama A, Yamaguchi H, Sato S, Isobe S, Tabata S, Fukuoka H. 2014. Draft genome sequence of eggplant (Solanum melongena L.) the representative Solanum species indigenous to the old world. DNA Research **21**, 649–660.

IBPGR. 1990. Descriptors for eggplant/Descripteurs pour l'aubergine. Rome, Italy: International Board for Plant Resources.

Kakizaki Y. 1931. Hybrid vigour in egg-plants and its practical utilization. Genetics **16**, 1–25.

Kaushik P. 2019. Genetic analysis for fruit phenolics content, flesh color, and browning related traits in eggplant (*Solanum melongena*). International Journal of Molecular Sciences **20**.

Kaushik P, Andújar I, Vilanova S, Plazas M, Gramazio P, Herraiz FJ, Brar NS, Prohens J. 2015. Breeding vegetables with increased content in bioactive phenolic acids. Molecules **20**, 18464–18481.

Kaushik P, Gramazio P, Vilanova S, Raigón MD, Prohens J, Plazas M. 2017. Phenolics content, fruit flesh colour and browning in cultivated eggplant, wild relatives and interspecific hybrids and implications for fruit quality breeding. Food Research International **102**, 392–401.

Kaushik P, Plazas M, Prohens J, Vilanova S, Gramazio P. 2018. Diallel genetic analysis for multiple traits in eggplant and assessment of genetic distances for predicting hybrids performance. PLoS ONE **13**, 1–20.

Khalid MF, Huda S, Yong M, Li L, Li L, Chen ZH, Ahmed T. 2023. Alleviation of drought and salt stress in vegetables: crop responses and mitigation strategies. Plant Growth Regulation **99**, 177–194.

Khan MMR, Isshiki S. 2016. Cytoplasmic male sterility in eggplant. Horticulture Journal **85**, 1–7.

Kikuchi K, Honda I, Matsuo S, Fukuda M, Saito T. 2008. Stability of fruit set of newly selected parthenocarpic eggplant lines. Scientia Horticulturae **115**, 111–116.

King SR, Davis AR, Zhang X, Crosby K. 2010. Genetics, breeding and selection of rootstocks for Solanaceae and Cucurbitaceae. Scientia Horticulturae **127**, 106–111.

Knapp S, Aubriot X, Prohens J. 2019. Eggplant (*Solanum melongena* L.): Taxonomy and relationships. In: Chapman MA, ed. The Eggplant Genome. Springer, 11–22.

Knapp S, Sagona E, Carbonell AKZ, Chiarini F. 2017. A revision of the *Solanum elaeagnifolium* clade (Elaeagnifolium clade; subgenus *Leptostemonum*, Solanaceae). PhytoKeys **84**, 1–104.

Knapp S, Vorontsova MS. 2016. A revision of the "African Non-Spiny" Clade of Solanum L. (*Solanum* sections *Afrosolanum* Bitter, *Benderianum* Bitter, *Lemurisolanum* Bitter, *Lyciosolanum* Bitter, *Macronesiotes* Bitter, and *Quadrangulare* Bitter: PhytoKeys **66**, 1–142.

Knapp S, Vorontsova MS, Prohens J. 2013. Wild relatives of the eggplant (*Solanum melongena* L.: Solanaceae): New understanding of species names in a complex group. PLoS ONE **8**, e57039.

Kodackattumannil P, Lekshmi G, Kottackal M, Sasi S, Krishnan S, Senaani S AL, Amiri KMA. 2023. Hidden pleiotropy of agronomic traits uncovered by *CRISPR - Cas9* mutagenesis of the tyrosinase CuA - binding domain of the *polyphenol oxidase 2* of eggplant. Plant Cell Reports, 2–5.

Kouassi AB, Kouassi KBA, Sylla Z, Plazas M, Fonseka RM, Kouassi A, Fonseka H, N'guetta ASP, Prohens J. 2020. Genetic parameters of drought tolerance for agromorphological traits in eggplant, wild relatives, and interspecific hybrids. Crop Science **61**, 55–68.

Kouassi AB, Kouassi KBA, Sylla Z, Plazas M, Fonseka RM, Kouassi A, Fonseka H, N'guetta ASP, Prohens J. 2021. Genetic parameters of drought tolerance for agromorphological traits in eggplant, wild relatives, and interspecific hybrids. Crop Science **61**, 55–68.

Kouassi B, Prohens J, Gramazio P, Kouassi AB, Vilanova S, Galán-Ávila A, Herraiz FJ, Kouassi A, Seguí-Simarro JM, Plazas M. 2016. Development of backcross generations and new interspecific hybrid combinations for introgression breeding in eggplant (*Solanum melongena*). Scientia Horticulturae **213**, 199–207.

Kumar A, Sharma V, Jain BT, Kaushik P. 2020. Heterosis breeding in eggplant (*Solanum melongena* L.): Gains and provocations. Plants **9**, 1–16.

Kumar G, Sivaraj N, Kamala V, Gangopadhyay KK, Pandey S, Tiwari SK, Panwar NS, Meena BL, Dutta M. 2013. Diversity analysis in eggplant germplasm in India using DIVA-GIS approach. Indian Journal of Horticulture **70**, 519–525.

Kumbar S, Narayanankutty C, Sainamole Kurian P, Sreelatha U, Barik S. 2021. Evaluation of eggplant rootstocks for grafting eggplant to improve fruit yield and control bacterial wilt disease. European Journal of Plant Pathology **161**, 73–90.

Lanteri S, Barchi L. 2019. Advances in eggplant genome sequencing. In: Chapman MA, ed. The Eggplant Genome. Springer, 65–70.

Lebeau A, Gouy M, Daunay MC, Wicker E, Chiroleu F, Prior P, Frary A, Dintinger J. 2013. Genetic mapping of a major dominant gene for resistance to *Ralstonia solanacearum* in eggplant. Theoretical and Applied Genetics **126**, 143–158.

Li L, Li S, Ge H, Shi S, Li D, Liu Y, Chen H. 2022. A light-responsive transcription factor *SmMYB35* enhances anthocyanin biosynthesis in eggplant (*Solanum melongena* L.). Planta **255**, 12.

Li D, Qian J, Li W, Yu N, Gan G, Jiang Y, Li W, Liang X, Chen R, Mo Y, Lian J, Niu Y, Wang Y. 2021. A high-quality genome assembly of the eggplant provides insights into the molecular basis of disease resistance and chlorogenic acid synthesis. Molecular Ecology Resources **21**, 1274–1286.

Li J, Ren L, Gao Z, Jiang M, Liu Y, Zhou L, He Y, Chen H. 2017. Combined transcriptomic and proteomic analysis constructs a new model for light-induced anthocyanin biosynthesis in eggplant (*Solanum melongena* L.). Plant Cell and Environment **40**, 3069–3087.

Liu W, Qian Z, Zhang J, Yang J, Wu M, Barchi L, Zhao H, Sun H, Cui Y, Wen C. 2019. Impact of fruit shape selection on genetic structure and diversity uncovered from genome-wide perfect SNPs genotyping in eggplant. Molecular Breeding **39**, 140.

Liu J, Yang Y, Zhou X, Bao S, Zhuang Y. 2018. Genetic diversity and population structure of worldwide eggplant (*Solanum melongena* L.) germplasm using SSR markers. Genetic Resources and Crop Evolution **65**, 1663–1670.

Liu J, Zheng Z, Zhou X, Feng C, Zhuang Y. 2015. Improving the resistance of eggplant (*Solanum melongena*) to Verticillium wilt using wild species *Solanum linnaeanum*. Euphytica **201**, 463–469.

Lu Y, Luo S, Li N, Li Q, Du W, Zhang W, Yu P, Xuan S, Wang Y, Zhao J, Chen X, Shen S. 2021. Candidate gene, *SmCPR1*, encoding CPR1 related to plant height of the eggplant dwarf mutant *dwf*. Horticulturae **7**, 196.

Maioli A, Gianoglio S, Moglia A, Acquadro A, Valentino D, Milani AM, Prohens J, Orzaez D, Granell A, Lanteri S, Comino C. 2020. Simultaneous CRISPR/Cas9 editing of three PPO genes reduces fruit flesh browning in *Solanum melongena* L. Frontiers in Plant Science **11**, 607161.

Mangino G, Arrones A, Plazas M, Pook T, Prohens J, Gramazio P, Vilanova S. 2022. Newly developed MAGIC population allows identification of strong associations and candidate genes for anthocyanin pigmentation in eggplant. Frontiers in Plant Science **13**, 1–15.

Mangino G, Plazas M, Vilanova S, Prohens J, Gramazio P. 2020. Performance of a set of eggplant (*Solanum melongena*) lines with introgressions from its wild relative *S. incanum* under open field and screenhouse conditions and detection of QTLs. Agronomy **10**, 1–16.

Mangino G, Vilanova S, Plazas M, Prohens J, Gramazio P. 2021. Fruit shape morphometric analysis and QTL detection in a set of eggplant introgression lines. Scientia Horticulturae **282**, 110006.

Manzur JP, Oliva-Alarcón M, Rodríguez-Burruezo A. 2014. In vitro germination of immature embryos for accelerating generation advancement in peppers (Capsicum annuum L.). Scientia Horticulturae **170**, 203–210.

Mascher M, Schreiber M, Scholz U, Graner A, Reif JC, Stein N. 2019. Genebank genomics bridges the gap between the conservation of crop diversity and plant breeding. Nature Genetics **51**, 1076–1081.

Mauceri A, Abenavoli MR, Toppino L, Panda S, Mercati F, Aci MM, Aharoni A, Sunseri F, Rotino GL, Lupini A. 2021. Transcriptomics reveal new insights into molecular regulation of nitrogen use efficiency in *Solanum melongena*. Journal of Experimental Botany **72**, 4237–4253.

Mauceri A, Bassolino L, Lupini A, Badeck F, Rizza F, Schiavi M, Toppino L, Abenavoli MR, Rotino

GL, Sunseri F. 2020. Genetic variation in eggplant for Nitrogen Use Efficiency under contrasting NO3- supply. Journal of Integrative Plant Biology **62**, 487–508.

Mennella G, Rotino GL, Fibiani M, D'Alessandro A, Franceses G, Toppino L, Cavallanti F, Acciarri N, Lo Scalzo R. 2010. Characterization of health-related compounds in eggplant (Solanum melongena L.) lines derived from introgression of allied species. Journal of Agricultural and Food Chemistry **58**, 7597–7603.

Mishra BB, Gautam S, Sharma A. 2013. Free phenolics and polyphenol oxidase (PPO): The factors affecting post-cut browning in eggplant (*Solanum melongena*). Food Chemistry **139**, 105–114.

Mishra P, Tiwari SK, Kashyap SP, Tiwari KN, Singh M, Singh B. 2020. High-density genetic linkage map based on arbitrary and microsatellite markers using inter-specific recombinant inbred lines in eggplant (*Solanum melongena* L.). Journal of Plant Biochemistry and Biotechnology **29**, 427–438.

Miyatake K, Shinmura Y, Matsunaga H, Fukuoka H, Saito T. 2019. Construction of a core collection of eggplant (*Solanum melongena* L.) based on genome-wide snp and ssr genotypes. Breeding Science **69**, 498–502.

Muñoz-Falcón JE, Prohens J, Vilanova S, Nuez F. 2009. Diversity in commercial varieties and landraces of black eggplants and implications for broadening the breeders' gene pool. Annals of Applied Biology **154**, 453–465.

Murata G, Uehara T, Lee HJ, Mizutani M, Kadota Y, Shinmura Y, Saito T, Uesugi K. 2022. *Solanum palinacanthum* Dunal as a potential eggplant rootstock resistant to root-knot nematodes. Journal of Phytopathology **170**, 185–193.

Musarella CM. 2020. *Solanum torvum* Sw. (Solanaceae): a new alien species for Europe. Genetic Resources and Crop Evolution **67**, 515–522.

Mutegi E, Snow AA, Rajkumar M, Pasquet R, Ponniah H, Daunay MC, Davidar P. 2015. Genetic diversity and population structure of wild/ weedy eggplant (*Solanum insanum*, Solanaceae) in southern India: Implications for conservation. American Journal of Botany **102**, 140–148.

Nadeeshani H, Samarasinghe G, Wimalasiri S, Silva R, Hunter D, Madhujith T. 2021. Comparative analysis of the nutritional profiles of selected *Solanum* species grown in Sri Lanka. Journal of Food Composition and Analysis **99**, 103847.

Namisy A, Chen JR, Prohens J, Metwally E, Elmahrouk M, Rakha M. 2019. Screening cultivated eggplant and wild relatives for resistance to bacterial wilt (*Ralstonia solanacearum*). Agriculture (Switzerland) **9**, 1–12.

Odong TL, Jansen J, van Eeuwijk FA, van Hintum TJL. 2013. Quality of core collections for effective utilisation of genetic resources review, discussion and interpretation. Theoretical and Applied Genetics **126**, 289–305.

Oladosu Y, Rafii MY, Arolu F, Chukwu SC, Salisu MA, Olaniyan BA, Fagbohun IK, Muftaudeen TK. 2021. Genetic diversity and utilization of cultivated eggplant germplasm in varietal improvement. Plants **10**, 1–21.

Oyelana OA, Ugborogho RE. 2008. Phenotypic variation of F1 and F2 populations from three

species of *Solanum* L. (Solanaceae). African Journal of Biotechnology **7**, 2359–2367.

Page A, Daunay M-C, Aubriot X, Chapman MA. 2019*a*. Domestication of eggplants: A phenotypic and genomic Insight. In: Chapman MA, ed. The Eggplant Genome. Springer, 193–212.

Page A, Gibson J, Meyer RS, Chapman MA. 2019*b*. Eggplant domestication: Pervasive gene flow, feralization, and transcriptomic divergence. Molecular Biology and Evolution.1359–1372.

Plazas M, Andújar I, Vilanova S, Gramazio P, Javier Herraiz F, Prohens J. 2014. Conventional and phenomics characterization provides insight into the diversity and relationships of hypervariable scarlet (*Solanum aethiopicum* L.) and gboma (*S. macrocarpon* L.) eggplant complexes. Frontiers in Plant Science **5**, 1–13.

Plazas M, Andújar I, Vilanova S, Hurtado M, Gramazio P, Herraiz FJ, Prohens J. 2013. Breeding for chlorogenic acid content in eggplant: Interest and prospects. Notulae Botanicae Horti Agrobotanici Cluj-Napoca **41**, 26–35.

Plazas M, González-Orenga S, Nguyen HT, Morar IM, Fita A, Boscaiu M, Prohens J, Vicente O. 2022. Growth and antioxidant responses triggered by water stress in wild relatives of eggplant. Scientia Horticulturae **293**, 110685.

Plazas M, Gramazio P, Vilanova S, Kouassi AB, Fonseka RM, Rakha M, García-Fortea E, Mangino G, Kouassi KBA, Fonseka HH, Taher D, Kouassi A, Villanueva G, Arrones A, Alonso D, Prohens J. 2020. Introgression breeding from crop wild relatives in eggplant landraces for adaptation to climate change. Crop wild relative.1–12.

Plazas M, Vilanova S, Gramazio P, Rodríguez-Burruezo A, Fita A, Herraiz FJ, Ranil R, Fonseka R, Niran L, Fonseka H, Kouassi B, Kouassi A, Kouassi A, Prohens J. 2016. Interspecific hybridization between eggplant and wild relatives from different genepools. Journal of the American Society for Horticultural Science **141**, 34–44.

Portis E, Barchi L, Toppino L, Lanteri S, Acciarri N, Felicioni N, Fusari F, Barbierato V, Cericola F, Valè G, Rotino GL. 2014. QTL mapping in eggplant reveals clusters of yield-related loci and orthology with the tomato genome. PLoS ONE **9**, e89499.

Portis E, Cericola F, Barchi L, Toppino L, Acciarri N, Pulcini L, Sala T, Lanteri S, Rotino GL. 2015. Association mapping for fruit, plant and leaf morphology traits in eggplant. PLoS ONE **10**, e0135200.

del Pozo A, Brunel-Saldias N, Engler A, Ortega-Farias S, Acevedo-Opazo C, Lobos GA, Jara-Rojas R, Molina-Montenegro MA. 2019. Climate change impacts and adaptation strategies of agriculture in Mediterranean-climate regions (MCRs). Sustainability (Switzerland) **11**, 1–16.

Premabati Devi C, Munshi AD, Behera TK, Choudhary H, Vinod, Gurung B, Saha P. 2015. Cross compatibility in interspecific hybridization of eggplant, *Solanum melongena*, with its wild relatives. Scientia Horticulturae **193**, 353–358.

Prohens J, Gramazio P, Plazas M, Dempewolf H, Kilian B, Díez MJ, Fita A, Herraiz FJ, Rodríguez-Burruezo A, Soler S, Knapp S, Vilanova S. 2017. Introgressiomics: A new approach for using crop wild relatives in breeding for adaptation to climate change. Euphytica **213**, 158.

Prohens J, Plazas M, Raigón MD, Simarro JMS, Stommel JR, Vilanova S. 2012. Characterization of

interspecific hybrids and first backcross generations from crosses between two cultivated eggplants (*Solanum melongena* and *S . aethiopicum* Kumba group) and implications for eggplant breeding. Euphytica **186**, 517–538.

Prohens J, Rodríguez-Burruezo A, Raigón MD, Nuez F. 2007. Total phenolic concentration and browning susceptibility in a collection of different varietal types and hybrids of eggplant: Implications for breeding for higher nutritional quality and reduced browning. Journal of the American Society for Horticultural Science **132**, 638–646.

Prohens J, Whitaker BD, Plazas M, Vilanova S, Hurtado M, Blasco M, Gramazio P, Stommel JR. 2013. Genetic diversity in morphological characters and phenolic acids content resulting from an interspecific cross between eggplant, *Solanum melongena*, and its wild ancestor (*S. incanum*). Annals of Applied Biology **162**, 242–257.

Qian Z, Ji Y, Li R, Lanteri S, Chen H, Li L, Jia Z, Cui Y. 2022. Identifying quantitative trait loci for thousand grain weight in eggplant by genome re-sequencing analysis. Frontiers in Genetics **13**, 1–10.

Quamruzzaman A. 2021. The first GM crop in Bangladesh – Bt Eggplant. European Journal of Agriculture and Food Sciences **3**, 45–55.

Rakha M, Namisy A, Chen JR, El-Mahrouk ME, Metwally E, Taha N, Prohens J, Plazas M, Taher D. 2020. Development of interspecific hybrids between a cultivated eggplant resistant to bacterial wilt (*Ralstonia solanacearum*) and eggplant wild relatives for the development of rootstocks. Plants **9**, 1–13.

Ranil RHG, Niran HML, Plazas M, Fonseka RM, Fonseka HH, Vilanova S, Andújar I, Gramazio P, Fita A, Prohens J. 2015. Improving seed germination of the eggplant rootstock *Solanum torvum* by testing multiple factors using an orthogonal array design. Scientia Horticulturae **193**, 174–181.

Ranil RHG, Prohens J, Aubriot X, Niran HML, Plazas M, Fonseka RM, Vilanova S, Fonseka HH, Gramazio P, Knapp S. 2017. *Solanum insanum* L. (subgenus *Leptostemonum* Bitter, Solanaceae), the neglected wild progenitor of eggplant (*S. melongena* L.): A review of taxonomy, characteristics and uses aimed at its enhancement for improved eggplant breeding. Genetic Resources and Crop Evolution **64**, 1707–1722.

Raubach S, Kilian B, Dreher K, *et al.* 2021. From bits to bites: Advancement of the germinate platform to support prebreeding informatics for crop wild relatives. Crop Science **61**, 1538–1566.

Ro N, Haile M, Kim B, Cho GT, Lee J, Lee YJ, Hyun DY. 2022. Genome-wide association study for agro-morphological traits in eggplant core collection. Plants **11**, 1–20.

Rodríguez-Burruezo A, Prohens J, Nuez F. 2008. Performance of hybrids between local varieties of eggplant (*Solanum melongena*) and its relation to the mean of parents and to morphological and genetic distances among parents. European Journal of Horticultural Science **73**, 76–83.

Rosa-Martínez E, Adalid-Martínez AM, García-Martínez MD, Mangino G, Raigón MD, Plazas M, Gramazio P, Prohens J, Vilanova S. 2022*a*. Fruit composition of eggplant lines with introgressions from the wild relative *S. incanum*: Interest for breeding and safety for consumption. Agronomy **12**, 266.

Rosa-Martínez E, García-Martínez MD, Adalid-Martínez AM, Pereira-Dias L, Casanova C, Soler E, Figàs MR, Raigón MD, Plazas M, Soler S, Prohens J. 2021. Fruit composition profile of pepper, tomato

and eggplant varieties grown under uniform conditions. Food Research International **147**, 110531.

Rosa-Martínez E, Villanueva G, Şahin A, Gramazio P, García-Martínez MD, Raigón MD, Vilanova S, Prohens J, Plazas M. 2022*b*. Characterization and QTL identification in eggplant introgression lines under two N fertilization levels. Horticultural Plant Journal, 1–15.

Rotino GL, Perri E, D'Alessandro A, Mennella G. 1998. Characterization of fertile somatic hybrids between eggplant (*S. melongena* L.) and *S. integrifolium*. Proc. Eucarpia meeting on genetics and breeding of capsicum and eggplant.213–217.

Rotino GL, Sala T, Toppino L. 2014. Eggplant. Alien gene transfer in crop plants, Volume 2. Springer, 381–410.

Rubio Teso ML, Álvarez-Muñiz C, Gaisberger H, Kell SP, Lara-Romero C, Magos Brehm J, Maxted N, Phillips J, Iriondo JM. 2021. European crop wild relative diversity: towards the development of a complementary conservation strategy. Birmingham, UK.: Farmer's Pride, University of Birmingham.

Sabatino L, Iapichino G, D'Anna F, Palazzolo E, Mennella G, Rotino GL. 2018. Hybrids and allied species as potential rootstocks for eggplant: Effect of grafting on vigour, yield and overall fruit quality traits. Scientia Horticulturae **228**, 81–90.

Salinier J, Lefebvre V, Besombes D, Burck H, Causse M, Daunay MC, Dogimont C, Goussopoulos J, Gros C, Maisonneuve B, McLeod L, Tobal F, Stevens R. 2022. The INRAE Centre for vegetable germplasm: Geographically and phenotypically diverse collections and their use in genetics and plant breeding. Plants **11**, 347.

Sambandam CN. 1964. Heterosis in eggplant. Economic Botany **18**, 128–131.

Sarengaowa, Wang L, Liu Y, Yang C, Feng K, Hu W. 2022. Effect of ascorbic acid combined with modified atmosphere packaging for browning of fresh-cut eggplant. Coatings **12**, 1580.

Särkinen T, Bohs L, Olmstead RG, Knapp S. 2013. A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): A dated 1000-tip tree. BMC Evolutionary Biology **13**, 214.

Scaglione D, Pinosio S, Marroni F, Centa E Di, Fornasiero A, Magris G, Scalabrin S, Cattonaro F, Taylor G, Morgante M. 2019. Single primer enrichment technology as a tool for massive genotyping: A benchmark on black poplar and maize. Annals of Botany **124**, 543–551.

Schippers RR. 2000. African indigenous vegetables: An overview of the cultivated species. Chatham, UK: Natural Resources Institute/ACP-EU Technical Centre for Agricultural and Rural Cooperation.

Schouten HJ, Tikunov Y, Verkerke W, Finkers R, Bovy A, Bai Y, Visser RGF. 2019. Breeding has increased the diversity of cultivated tomato in The Netherlands. Frontiers in Plant Science **10**, 1–12.

Schwarz D, Rouphael Y, Colla G, Venema JH. 2010. Grafting as a tool to improve tolerance of vegetables to abiotic stresses: Thermal stress, water stress and organic pollutants. Scientia Horticulturae **127**, 162–171.

Shelton AM, Hossain MJ, Paranjape V, Azad AK, Rahman ML, Khan ASMMR, Prodhan MZH, Rashid MA, Majumder R, Hossain MA, Hussain SS, Huesing JE, McCandless L. 2018. Bt eggplant

project in Bangladesh: History, present status, and future direction. Frontiers in Bioengineering and Biotechnology **6**, 1–6.

Simko I, Jia M, Venkatesh J, Kang BC, Weng Y, Barcaccia G, Lanteri S, Bhattarai G, Foolad MR. 2021. Genomics and marker-assisted improvement of vegetable crops. Critical Reviews in Plant Sciences **40**, 303–365.

Solberg S, van Zonneveld M, Rakha M, Taher D, Prohens J, Jarret R, van Dooijeweert W, Giovannini P. 2022. Global strategy for the conservation and use of eggplants. Zenodo.

Srinivasan R. 2008. Integrated pest management for eggplant fruit and shoot borer (*Leucinodes orbonalis*) in south and southeast Asia: Past, present and future. Journal of Biopesticides **1**, 105–112.

Srinivasan R. 2009. Insect and mite pest on eggplant: A field guide for identification and management. Shanhua, Taiwan: AVTDC - The World Vegetable Center.

Stommel JR, Whitaker BD. 2003. Phenolic Acid Content and composition of eggplant fruit in a germplasm core subset. J. Amer. Soc. Hort. Sci. **128**, 704–710.

Street K, Bari A, Mackay M, Amri A. 2016. How the focused identification of Germplasm Strategy (FIGS) is used to mine plant genetic resources collections for adaptive traits. In: Maxted N, Dulloo ME, Ford-Lloyd BV, eds. Enhancing crop genepool use: capturing wild relative and landrace diversity for crop improvement. CABI Wallingford, 54–65.

Sulli M, Barchi L, Toppino L, Diretto G, Sala T, Lanteri S, Rotino GL, Giuliano G. 2021. An eggplant recombinant inbred population allows the discovery of metabolic QTLs controlling fruit nutritional quality. Frontiers in Plant Science **12**, 638195.

Sunseri F, Polignano GB, Alba V, Lotti C, Bisignano V, Mennella G, Alessandro AD, Bacchi M, Riccardi P, Fiore MC, Ricciardi L. 2010. Genetic diversity and characterization of African eggplant germplasm collection. African Journal Plant Science **4**, 231–241.

Syfert MM, Castañeda-Álvarez NP, Khoury CK, Särkinen T, Sosa CC, Achicanoy HA, Bernau V, Prohens J, Daunay M-C, Knapp S. 2016. Crop wild relatives of the brinjal eggplant (*Solanum melongena*): Poorly represented in genebanks and many species at risk of extinction. American Journal of Botany **103**, 635–651.

Taher D, Rakha M, Ramasamy S, Solberg S, Schafleitner R. 2019. Sources of resistance for two-spotted spider mite (*Tetranychus urticae*) in scarlet (*Solanum aethiopicum* L.) and Gboma (*S. macrocarpon* L.) eggplant germplasms. HortScience **54**, 240–245.

Taher D, Ramasamy S, Prohens J, Rakha M. 2020. Screening cultivated eggplant and wild relatives for resistance to sweetpotato whitefly (*Bemisia tabaci*) and to two-spotted spider mite (*Tetranychus urticae*). Euphytica **216**, 1–13.

Taher D, Solberg SØ, Prohens J, Chou Y, Rakha M, Wu T. 2017. World vegetable center eggplant collection: Origin, composition, seed dissemination and utilization in breeding. Frontiers in Plant Science **8**, 1484.

Tassone MR, Bagnaresi P, Desiderio F, Bassolino L, Barchi L, Florio FE, Sunseri F, Sirangelo TM, Rotino GL, Toppino L. 2022. A genomic BSAseq approach for the characterization of QTLs underlying

resistance to *Fusarium oxysporum* in eggplant. Cells **11**, 1–21.

Toppino L, Acciari N, Mennella G, Lo Scalzo R, Rotino GL. 2009. Introgression breeding in eggplant (*Solanum melongena* L.) by combining biotechnological and conventional approaches. Proc. of the 53rd Italian Society of Agricultural Genetics Annual Congress. Torino, Italy, Oral communication.

Toppino L, Barchi L, Mercati F, Acciarri N, Perrone D, Martina M, Gattolin S, Sala T, Fadda S, Mauceri A, Ciriaci T, Carimi F, Portis E, Sunseri F, Lanteri S, Rotino GL. 2020. A new intra-specific and high-resolution genetic map of eggplant based on a ril population, and location of QTLS related to plant anthocyanin pigmentation and seed vigour. Genes **11**, 1–29.

Toppino L, Barchi L, Rotino GL. 2022. Next generation breeding for abiotic stress resistance in eggplant. Genomic designing for abiotic stress resistant vegetable crops. Springer, 115–152. In: Kole, C. (eds) Genomic Designing for Abiotic Stress Resistant Vegetable Crops. Springer, Cham.

Toppino L, Barchi L, Lo Scalzo R, Palazzolo E, Francese G, Fibiani M, D'Alessandro A, Papa V, Laudicina VA, Sabatino L, Pulcini L, Sala T, Acciarri N, Portis E, Lanteri S, Mennella G, Rotino GL. 2016. Mapping quantitative trait loci affecting biochemical and morphological fruit properties in eggplant (*Solanum melongena* L.). Frontiers in Plant Science **7**, 256.

Toppino L, Mennella G, Rizza F, D'Alessandro A, Sihachakr D, Rotino GL. 2008*a*. ISSR and isozyme characterization of androgenetic dihaploids reveals tetrasomic inheritance in tetraploid somatic hybrids between Solanum melongena and Solanum aethiopicum group Gilo. Journal of Heredity **99**, 304–315.

Toppino L, Prohens J, Rotino GL, Plazas M, Parisi M, Carrizo, Garcia C, Tripodi P. 2021. Pepper and eggplant genetic resources. In: Carputo, D., Aversano, R., Ercolano, M.R. (eds) The Wild Solanums Genomes. Compendium of Plant Genomes. Springer, Cham.

Toppino L, Ribolzi S, Shaaf S, Bassolino L, Carletti G, Fadda S, Rossini L, Boyaci HF, Caliskan S, Unlu A, Rotino GL. 2018. Development of an introgression lines population and genetic mapping of novel traits linked to key breeding traits in eggplant. Proceedings of the 62th SIGA Congress Verona. Italy.

Toppino L, Valè G, Alberti P, Mennella G, Acciarri N, Rotino G. 2007. Introgression of resistance to *Fusarium oxysporum* f.sp. melongenae into cultivated eggplant and characterization of linked molecular markers for development of a markers assisted selection system. EUCARPIA, XIIIth Meeting on Genetics and Breeding of Capsicum and Eggplant. POL, 157–166.

Toppino L, Valè G, Rotino GL. 2008*b*. Inheritance of *Fusarium* wilt resistance introgressed from *Solanum aethiopicum* Gilo and Aculeatum groups into cultivated eggplant (*S. melongena*) and development of associated PCR-based markers. Molecular Breeding **22**, 237–250.

UPOV. 2011. Guidelines for the conduct of test for distinctness, uniformity and stability. Eggplant (*Solanum melongena* L.).

Vilanova S, Manzur JP, Prohens J. 2012. Development and characterization of genomic simple sequence repeat markers in eggplant and their application to the study of diversity and relationships in a collection of different cultivar types and origins. Molecular Breeding **30**, 647–660.

Villanueva G, Rosa-Martínez E, Şahin A, García-Fortea E, Plazas M, Prohens J, Vilanova S. 2021. Evaluation of advanced backcrosses of eggplant with *Solanum elaeagnifolium* introgressions under

low N conditions. Agronomy **11**, 1–18.

Villanueva G, Vilanova S, Plazas M, Prohens J, Gramazio P. 2023. Transcriptome profiles of eggplant (*Solanum melongena*) and its wild relative *S. dasyphyllum* under different levels of osmotic stress provide insights into response mechanisms to drought. Current Plant Biology **33**, 100276.

Vorontsova MS, Knapp S. 2016. A revision of the spiny solanums, Solanum subgenus *Leptostemonum* (Solanaceae) in Africa and Madagascar. Systematic Botany Monographs **99**, 1–436.

Vorontsova MS, Stern S, Bohs L, Knapp S. 2013. African spiny Solanum (subgenus *Leptostemonum*, Solanaceae): a thorny phylogenetic tangle. Botanical Journal of the Linnean Society **173**, 176–193.

van der Weerden GM, Barendse GWM. 2007. A web-based searchable database developed for the eggnet project and applied to the Radboud University Solanaceae database. Acta Hortic. **745**, 503–506.

Wei Q, Wang J, Wang W, Hu T, Hu H, Bao C. 2020. A high-quality chromosome-level genome assembly reveals genetics for important traits in eggplant. Horticulture Research **7**, 153.

Whalen M. 1984. Conspectus of species groups in *Solanum* subgenus *Leptostemonum*. Gentes Herbarium **12**, 282.

Xi-ou X, Wenqiu L, Wei L, Xiaoming G, Lingling L, Feiyue M, Yuge L. 2017. The analysis of physiological variations in M2 generation of *Solanum melongena* L. Mutagenized by ethyl methane sulfonate. Frontiers in Plant Science **8**, 1–14.

Xiao XO, Lin WQ, Li W, Liu GQ, Zhang XH, L. LL. 2016. Creating new eggplant germplasm by EMS Mutation. J. South Agric. **47**, 1247–1253.

Zhuang J, Zhang J, Hou XL, Wang F, Xiong AS. 2014. Transcriptomic, proteomic, metabolomic and functional genomic approaches for the study of abiotic stress in vegetable crops. Critical Reviews in Plant Sciences **33**, 225–237.

OBJECTIVES

Los bancos de germoplasma no están para guardar y acumular semilla. Debe tener también el propósito de estudiarlo y usarlo.

Maria José Díez Niclós

In order to ensure future productivity and quality of crops, it is necessary to preserve and use the valuable genetic resources stored in genebanks. However, for these collections to be useful and for the conserved materials to be actively and successfully utilized, a comprehensive morphological and genetic characterization is essential as well as the availability of these data in a complete and easily searchable databases. In the present doctoral thesis, we aimed at enhancing the value of the tomato, pepper, and eggplant germplasm collections from the G2P-SOL project, as well as to develop and implement molecular tools that allow for their in-depth study.

Therefore, we propose the following specific main objectives:

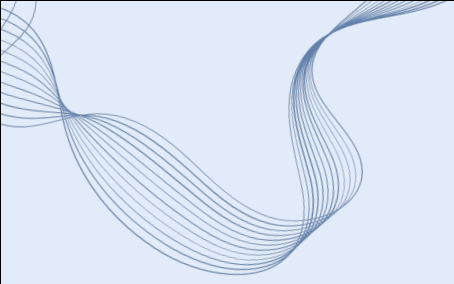
1. To compile and harmonize the passport and phenotypic data and images of the three major solanaceous crops, namely tomato, pepper, and eggplant, into a single and user-friendly database.

2. To develop and optimize a fast and affordable DNA extraction protocol that can be applied to solanaceous crops as well as to a wide range of species and tissues.

3. To develop SPET genotyping platforms for tomato and eggplant, and use them to investigate the genetic variation and population structure of an extensive collection of cultivated and wild tomato and eggplant accessions.

4. To evaluate the genotypic and phenotypic characteristics of a tomato core collection to make it more useful to breeders and researchers.

RESULTS



CHAPTER I. Origin and composition of the Solanaceae germplasm collection: tomato, pepper, eggplant and potato from the European project G2P-SOL

David Alonso[1], Jaime Prohens[1], María José Díez[1] and Giovanni Giuliano[2]

[1]Instituto de Conservación y Mejora de la Agrodiversidad Valenciana (COMAV). Universitat Politècnica de València. Camino de Vera 14, 46022 Valencia
[2]Agenzia Nazionale per le Nuove Tecnologíe, L'Energia e lo Sviluppo Economico Sostenible (ENEA). Via Anguillarese 301 – 00123 Roma, Italia (*In representation of the G2P-SOL consortium*)

*Corresponding author: daalmar8@upv.es

Ph.D. candidate contribution

David Alonso had a main role in the following activities: conceived and designed the research, data collection, data curation, data visualization, data analysis and drafting the manuscript.

AUTHOR'S VERSION.

BEING PREPARED FOR SUBMISSION TO A JOURNAL


1

CHAPTER I

Origin and composition of the Solanaceae germplasm collection: tomato, pepper, eggplant and potato from the european project G2P-SOL

David Alonso[1], Jaime Prohens[1], María José Díez[1] and Giovanni Giuliano[2]

[1]Instituto de Conservación y Mejora de la Agrodiversidad Valenciana (COMAV). Universitat Politècnica de València. Camino de Vera 14, 46022 Valencia
[2]Agenzia Nazionale per le Nuove Tecnologíe, L'Energia e lo Sviluppo Economico Sostenible (ENEA). Via Anguillarese 301 – 00123 Roma, Italia (*En representación del Consorcio G2P-SOL*)

ABSTRACT

The G2P-SOL project (*Linking genetic resources, genomes, and phenotypes of Solanaceous crops*), funded by the European program Horizon 2020, aims to rationalize the major collections of tomato, pepper, eggplant, and potato worldwide. This will be addressed by constructing a core collection for each crop using passport, genotyping, and phenotyping data, followed by the evaluation of various traits. The project includes 58,636 accessions from various European germplasm banks, as well as the International Potato Center (CIP, Peru), and the World Vegetable Center (WorldVeg, Taiwan). As a first step towards constructing the core collections, a meticulous study of the passport data has been conducted, enabling the detection of species misattributions, identifying duplicates within and between collections, and gaining detailed knowledge of the geographic origin of the collection and the relative participation of different types of biological materials.

Keywords: *Solanum lycopersicum*, *Solanum melongena*, *Capsicum annuum*, *Solanum tuberosum*, wild species, passport data.

INTRODUCTION

According to Indicator 2.5.1 of the Sustainable Development Goals 2 (SDGs) of the FAO on plant genetic resources, it is estimated that germplasm collections worldwide house around 5.8 million accessions, which are conserved in the medium or long term in 846 gene banks (Gil et al., 2019). These germplasm collections represent a source of diversity with great potential for genetic improvement. Accessibility to the genetic resources diversity contained in them is limited, mainly due to the incomplete status and lack of standardization of the databases, which significantly hinders their comparison. The European project G2P-SOL (*Linking genetic resources, genomes and phenotypes of Solanaceous crops*), funded by the European program Horizon 2020, aims to rationalize the major collections of Solanaceous crops worldwide. The germplasm collection included in the project consists of 58,937 accessions, and it involves 14 institutions from 8 European countries (Germany, Bulgaria, Spain, France, Netherlands, Italy, Poland, and the United Kingdom) and four non-European countries (Israel, Peru, Taiwan, and Turkey). Rationalization has been approached by creating a single database as an indispensable tool for efficient and unified information management. In the second step, most accessions from the four crops were genotyped, and core collections were established. These core collections have been characterized for agro-morphological traits, both nutraceutical and organoleptic quality, through metabolomic analysis and resistance to pathogens and abiotic stresses. This chapter describes the characteristics of the G2P-SOL collection based on passport and phenotypic data available in the germplasm banks, which are essential for the orderly pursuit of future project objectives. Although the project includes potato cultivation, this chapter will focus on tomato, pepper, and eggplant since our bank actively participated in these crops with data and plant material.

MATERIALS AND METHODS

Plant material

A germplasm collection composed of a total of 58,937 accessions has been analyzed. These accessions originate from the Leibniz Institut fuer Pflanzengenetik und Kulturpflanzenforschung (IPK), the Institut National de la Recherche Agronomique (INRAE), the Wageningen University & Research (WUR-DLO), the Instituto de Conservación y Mejora de la Agrodiversidad Valenciana de la Universitat Politècnica de València (COMAV-UPV), the Università degli Studi di Torino (UniTO), the Agenzia Nazionale Per le Nuove Tecnologie, L'Energia e lo Sviluppo Economico Sostenible (ENEA), the James Hutton Institute (JHI), the Instytut Hodowli i Aklimatyzacji Roslin – Panstwowy Instytut Badawczy (IHAR-PIB), the Consiglio per la Ricerca e la Sperimentazione in Agricoltura e l'Analisis dell'Economia Agraria (CREA) and the Maritsa Vegetable Crops Research Institute (MVCRI). In addition, four non-European institutions are participating in the project: the World Vegetable Center (WorldVeg), the Hebrew University of Jerusalem (HUJI), the Agricultural Research Organisation of Israel – The Volcanic Center (ARO) and the Bati Akdeniz Agricultural Research Institute (BATEM). Regarding crops, there are 24,767 tomato accessions, 14,117 pepper accessions and 6,574 eggplant accessions, including cultivated and wild species.

Gathering available passport and phenotypic data

To gather the available passport data stored in different germplasm banks, a subgroup of descriptors based on the Multicrop Passport Descriptor List (MCPD V2.1) developed by IPGRI (International Plant Genetic Resources Institute) and FAO (Alercia et al., 2015) was employed. The selection of descriptors was agreed upon and evaluated to fulfill the objectives of a) taxonomic identification of the accessions, b) understanding their origin, c) identifying the original donor of the accession and any possible duplicates existing in other banks, and d) determining the type of biological material for each entry.

Three sets of descriptors were developed to compile the available phenotypic data of the accessions included in the collections, one for each crop (tomato, pepper, eggplant). The descriptors were developed using the system designed by IPGRI-FAO, which categorizes them into four types: vegetative, inflorescence, fruit, and agronomic traits, considering the descriptors used in each germplasm bank (**Annex 1**).

To enhance the quality and utility of the database, images of a significant number of accessions were also included.

Data review

After receiving the available data from germplasm banks, they were standardized and unified based on the scales described in the phenotypic descriptors (**Annex 1)**. In cases where images were available, they were reviewed alongside the available phenotypic data, correcting any errors if present. Additionally, to enhance the quality of the database, information from different available databases such as Genesys, USDA (U.S. Department of Agriculture), TGRC (Tomato Genetics Resource Center, Davis, California), CGN (Center for Genetics Resources, Wageningen, Netherlands), IPK (Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany), and WorldVeg (Taiwan) was reviewed and added to the corresponding accessions.

All graphical representations were created using R software version 4.2.0 (R Core Team, 2020). The maps were generated using the *Tidyverse* 1.3.2 package, specifically *ggplot2* 3.4.0 (Wickham, 2016), using available longitudinal and latitudinal georeferences generated with the map_data function, assigning new ones if necessary to construct the map. The cladograms were designed in Newick format and visualized using the *ggtree* 3.2.1 package (Yu, 2020). The bubble plots comparing passport traits and mosaic plots for phenotypic data were created using the *ggplot2* 3.4.0 package.

RESULTS AND DISCUSSION

Passport data selection

A total of 22 passport descriptors were selected (**Table 1**), out of which nine are identification codes for the accessions from the original banks and their possible duplicates in other repositories. These data are essential for identifying duplicates among collections and for completing the information of the accessions if incomplete. Three descriptors refer to the taxonomic classification of the accessions, six to the geographical origin, two to the type of biological material and collection origin, and finally, two descriptors related to the collection and acquisition dates.

Table 1. Passport descriptors for identifying and classifying accessions in germplasm banks based on the Multicrop Passport Descriptor List (MCPD V2.1) developed by IPGRI and FAO (Alercia et al., 2015).

Acronym	Descriptor
SAMPLE.ID	Unique G2P-SOL code
INSTCODE	Institute code
ACCENUMB	Accession number
LOCALNAME	Local name
COLLNUMB	Collecting number
COLLCODE	Collection institute code
DONORCODE	Donor institute code
DONORNUMB	Donor accesión number
OTHERDONORCODE	Code of duplicates in other genebanks
DUPLINSTNAME	Code of the Genebank hoolder of the duplicates
GENUS	Genus
SPECIES	Species
SUBTAXA	Subtaxon
COLLSITE	Location of collecting site
PROVINCE	Province
COUNTRY	Country
LATITUDE	Latitude of collecting site
LONGITUDE	Longitude of collecting site
ALTITUDE	Elevation of collecting site
BIOSTATUS	Biological status of accession
COLLSRC	Collecting/acquisition sources
ACQDATE	Acquisition date
COLLDATE	Collecting date of sample

Characteristic of collections

The total number of accessions with complete information regarding taxonomic classification is 58,937, out of which 24,767 correspond to tomato, 14,117 to pepper and 6,574 to eggplant (**Table 2**).

Table 2. Number of accessions for different crops in the G2P-SOL collection provided by each participant.

Institutions	G2P-SOL total	Tomato (% wild relatives)	Pepper (%wild relatives)	Eggplant (% wild relatives)
Belonging to the European Union (EU)				
IPK (Germany)	11.679	3.835 (2,2%)	1.536 (0,7%)	112 (0%)
INRAE (France)	5.263	1.210 (18,9%)	1.300 (1,7%)	1.546 (24,8%)
COMAV-UPV (Spain)	4.076	2.390 (2,8%)	1.426 (0%)	260 (6,3%)
WUR-DLO (Netherlands)	4.043	1.332 (7,9%)	1.033 (1,7%)	510 (7,1%)
CREA (Italy)	660	-	300 (1,3%)	360 (7,9%)
MVCRI (Bulgaria)	211	52 (3,8%)	67(0%)	-
ENEA (Italy)	239	239 (0%)	-	-
UniTO (Italy)	136	-	136 (0%)	-
Non-EU countries				
WorldVeg (Taiwan)	20.254	8.545 (10,5%)	7.964 (0,5%)	3.745 (11.8%)
HUJI (Israel)	7.114	7.114 (5,2%)	-	-
ARO (Israel)	305	-	305 (1,3%)	-
BATEM (Turkey)	140	50 (0%)	50 (0%)	41 (0%)
G2P-SOL total	58.636	24.767 (6.9%)	14.117 (0.7%)	6.574 (13.9%)

Tomato

The tomato collection was contributed by nine institutions, with WorldVeg and HUJI being the major contributors, with over 7,000 accessions each (**Table 2**). These collections, along with those from INRAE, COMAV-UPV, IPK, and WUR-DLO, contain accessions from many countries and different types of materials, greatly enriching the genetic basis of the collection. On the other hand, other institutions contribute a smaller number of accessions, mostly from their own countries, such as MVCRI, ENEA, and BATEM. The cultivated tomato accessions come from 120 countries (**Figure 1a**), while those of related wild species are from Peru, Ecuador, and Chile. The number of accessions collected or originating from Spain and the United States is notable. COMAV-UPV mainly provides Spanish germplasm, while the American accessions are mostly from USDA and TGRC but preserved in other institutions.

The percentage of wild species concerning the total is 6.9%, with WorldVeg, HUJI, and INRAE being the gene banks that contribute the highest number of accessions. Within the collection, all wild species and the *cerasiforme* variety are represented (**Figure 1b**), ranging from those most divergent from cultivated tomato, such as species belonging to the Hirsutum group (estimated divergence time from *S. lycopersicum* var. *lycopersicum* (SLL): ~2.2 Ma), *S. habrochaites*, and *S. pennellii*, to those closest, such as *S. pimpinellifolium* and the variety *S. lycopersicum* var. *cerasiforme* (SLC) belonging to the Esculentum group. Wild tomato species have been widely used in genetic improvement due to the narrow genetic base of this crop, which makes it highly susceptible to pests and diseases (Roselló et al., 2001; Foolad et al., 2014), as well as abiotic stresses (Reimer et al., 2021). In smaller numbers, we also find a representation of the *Lycopersicoides* section (*S. lycopersicoides* and *S. sitiens*) and the *Juglandifolia* section (*S. juglandifolium* and *S. ochranthum*).

The tomato collection consists mainly of advanced or improved cultivars, traditional cultivars, wild species, and breeding lines (**Figure 1c**). The collection contributed by UPV-COMAV, which mostly includes traditional varieties prospecting throughout the country, deserves special mention. This collection is important since Spain was a center of tomato diversification and dissemination upon its arrival in Europe in the 16th century (Bergougnoux, 2014). The germplasm banks from Germany, the Netherlands, and HUJI hold the majority of advanced or improved cultivars in the collection.

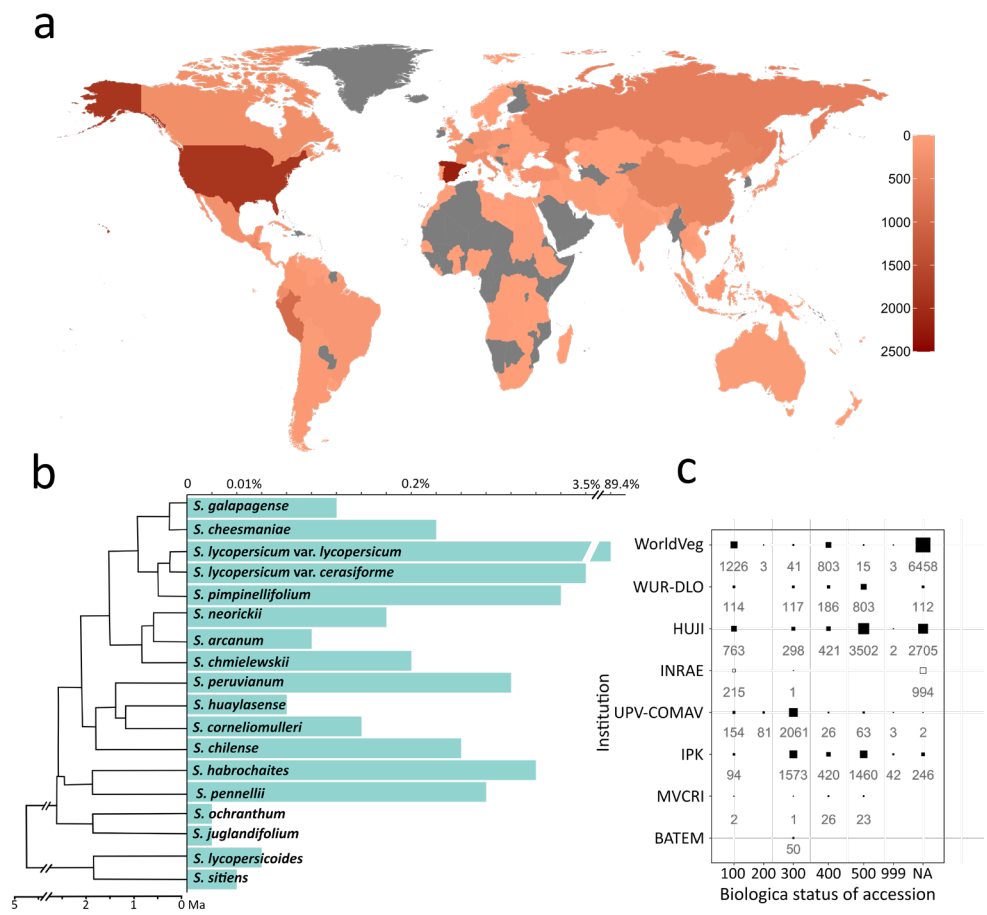


Figure 1. Description of passport data**. (a)** Representation of the country of origin for cultivated tomato (SLL/SLC) accessions included in the G2P-SOL collection. The gray color represents countries with no accessions. **(b)** Cladogram of the *Lycopersicon*, *Juglandifolia* and *Lycopersicoides* sections used in this study based on Beddows et al., 2017, Pease et al., 2016 and Särkinen et al., 2013. The bars indicate the abundance of each species in the collection. **(c)** Box plot showing the distribution of biological status of accessions across different institutions (100: Wild; 200: Weedy; 300: Traditional cultivar/landraces; 400: Breeding lines; 500: Advanced or improved cultivar; 999: Others; NA: Missing data).

Pepper

Regarding the pepper collection, WorldVeg is the most significant contributor, with over 7,000 accessions; with the other institutions, the total amount comes to 14,117 accessions. The contributions from banks such as IPK, INRAE, UPV-COMAV, and WUR-DLO are also noteworthy, with each contributing just over 1,000 accessions (**Table 2**). The genetic diversity of the *Capsicum* genus in the collection is represented by a wide range of accessions from cultivated species, including *Capsicum annuum*, *C. frutescens*, *C. chinense*, *C. baccatum* and *C. pubescens*, which account for 98,3% of the collection, originating from 180 countries (**Figure 2a,b**). Wild species such as *C. tovarii*, *C. praetermissum*, *C. chacoense*, *C. galapagoense*, *C. eximium* and *C. cardenasii*, from 13 countries, are present in a lower percentage.

Most of the pepper collection consists of traditional varieties, primarily contributed by UPV-COMAV and IPK (**Figure 2c**). These varieties are highly interested in breeding due to their genetic variability and the result of farmer selection over hundreds of years. The Spanish collection is particularly important, as Spain served as the main entry point for all plant materials, especially *C. annuum*, brought from Mexico, the Caribbean, and South America (Andrews, 1995; Nuez et al., 2003). The heterogeneous environments, soil conditions and diverse cultivation practices have given rise to locally adapted varieties suited to specific agroclimatic requirements (Nuez et al., 2003). Furthermore, improved cultivars and other breeding materials play a significant role in the collection.

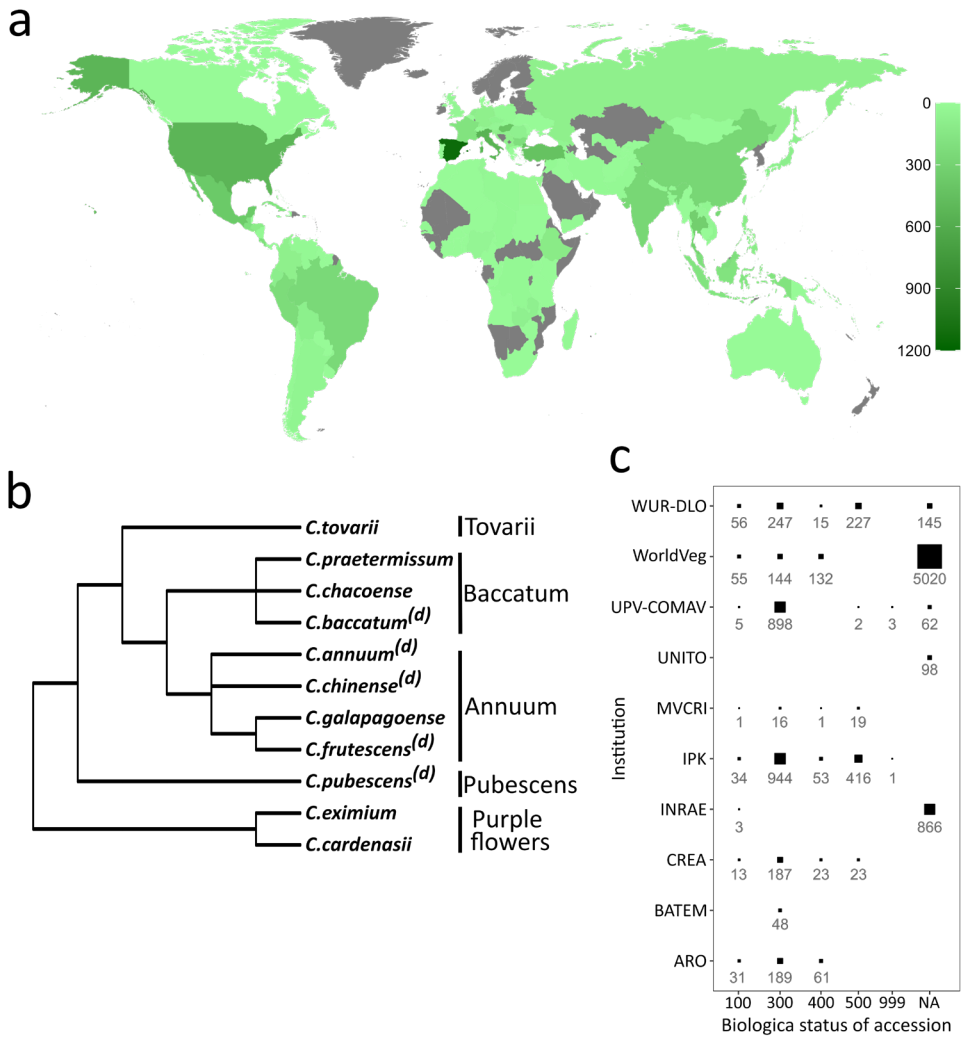


Figure 2. Description of passport data **(a).** Representation of the country of origin for domesticated pepper accessions (*Capsicum annuum, C. frutescens, C. chinense, C. baccatum* and *C. pubescens*). The gray color represents countries with no accessions. **(b)** Cladogram showing the species relationships in *Capsicum* based on Barboza et al., 2022 and Carrizo García et al., 2016. The letter (d) corresponds to cultivated species. **(c)** Box plot showing the representation of biological status for each institution (100: Wild; 300: Traditional cultivar/landraces; 400: Breeding lines; 500: Improved cultivar; 999: Others; NA: Missing data).

Eggplant

The eggplant collection comprises accessions sourced from seven institutions, with WorldVeg (Taher et al., 2017) once again being the largest contributor, supplying over half of the accessions (**Table 2)**. While WorldVeg and INRAE provide most of the wild species, traditional varieties are contributed by multiple banks. In eggplant, improved cultivars are not as prominent as in the previously mentioned collections.

Most of the eggplant collection (86,1%) consists of cultivated species (*Solanum melongena*, *S. macrocarpon* and *S. aethiopicum*) originating from 93 countries across Africa, Asia, and Europe (**Figure 3a**). Related species (854 accessions of 35 different species) are distributed among 65 countries (**Figure 3ab**). Scarlet eggplant (*S. aethiopicum*) and gboma eggplant (*S. macrocarpon*) are predominantly cultivated in Africa, with their respective wild ancestors, *S. anguivi* and *S. dasyphyllum*, represented in the collection. Notably, the collection encompasses a substantial number of accessions from the wild ancestor *S. insanum* and other closely related species, including *S. incanum*, *S. lichtensteinii* and *S. linnaeanum*, all of which have significant relevance to common eggplant (*S. melongena*) (Acquadro et al., 2017; Vorontsova et al., 2013). The collection also exhibits a representation of more distantly related American species, including *S. torvum* and *S. sisymbriifolium* (Syfert et al., 2016; Vorontsova et al., 2013), which hold potential interest for genetic improvement of eggplant due to their resistance to significant biotic stresses such as bacteria and nematodes (Daunay y Hazra, 2012; Rakha et al., 2020; Zhang et al., 2021). Species like *S. elaeagnifolium*, *S. incanum* and *S. dasyphyllum* also contribute to the genetic diversity of the collection, offering valuable traits related to abiotic stress tolerance (Christodoulakis et al., 2009; Gramazio et al., 2017; Villanueva et al., 2022).

The collection predominantly comprises traditional varieties, improved cultivars and breeding materials (**Figure 3c**). Notably, the UPV-COMAV collection of traditional varieties stands out for its composition of accessions adapted to diverse environments and cultivation systems (Martínez-Ispizua et al., 2021; San José et al., 2014).

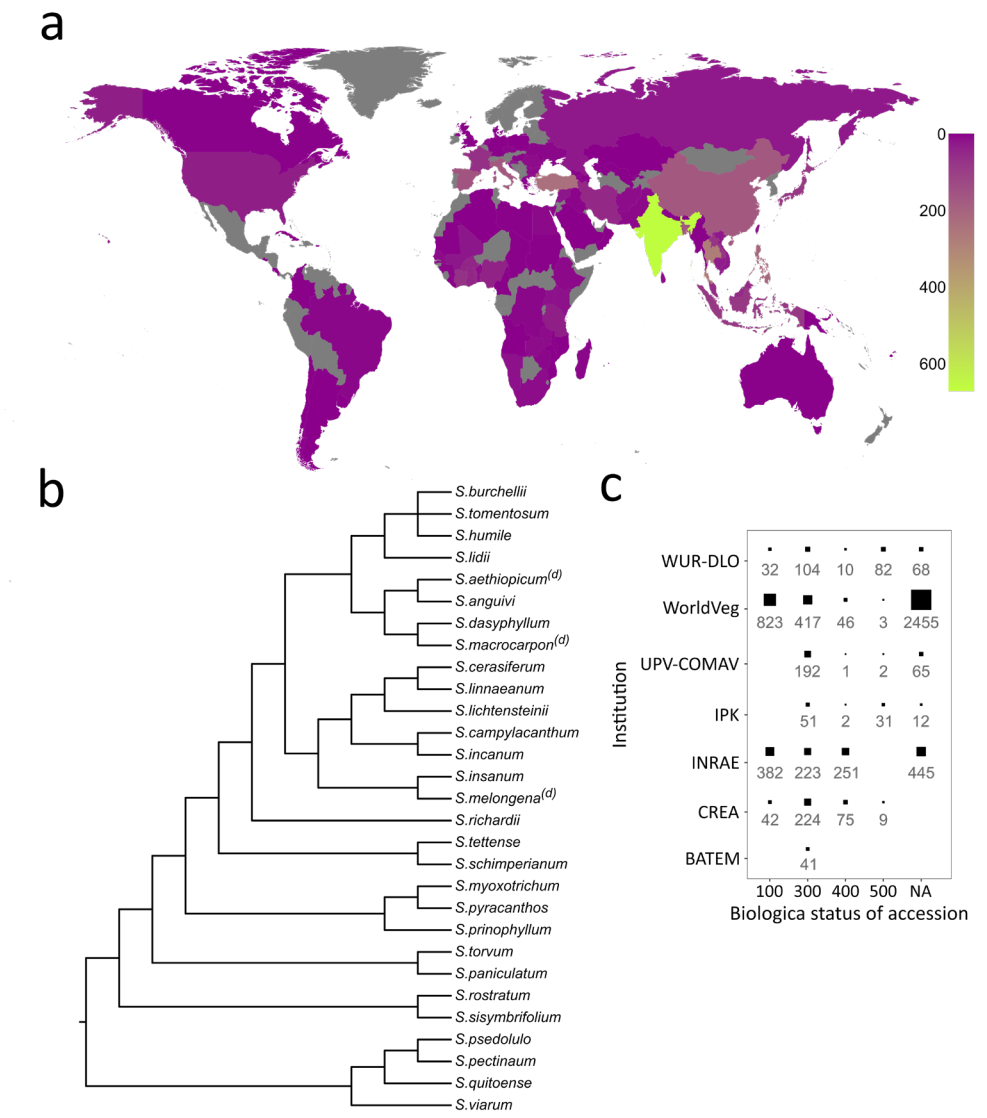


Figure 3. Description of passport data. **(a)** Representation of the country of origin of cultivated eggplant accessions (*S. melongena*, *S. aethiopicum*, and *S. macrocarpon*). The gray color represents countries with no accessions. **(b)** Cladogram of the main species in the *Solanum* taxa collection based on Särkinen et al., 2013 and Vorontsova et al., 2013. The letter (d) corresponds to cultivated species. **(c)** Box plot showing the representation of the biological status of the accession in each institution (100: Wild; 300: Traditional cultivar/landraces; 400: Breeding lines; 500: Advanced or improved cultivar; NA: Missing data).

Phenotypic variability of the collections

The final morphological descriptors for each crop were 25 for tomato, 26 for pepper, and 21 for eggplant (**Supplementary material 1**). Systematic characterization of germplasm through morphological and botanical descriptors (plant characteristics, leaf, inflorescence, and fruit) is crucial for adequately conserving and utilizing the material (Salinier et al., 2022). The descriptors are mainly focused on fruit traits of agronomic interest, such as color, shape and size. The main Solanaceae accessions contributed by different germplasm banks to establish the G2P-SOL project collection exhibit a high degree of morphological variability, encompassing all types of vegetative, inflorescence, fruit, and agronomic traits (**Figure 4**).

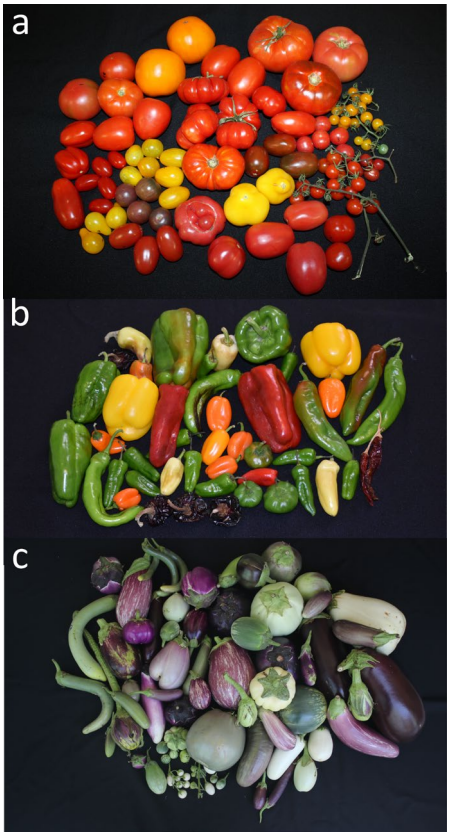


Figure 4. Illustration of the phenotypic diversity present in the three G2P-SOL collections. **(a)** Tomato**, (b)** Pepper and **(c)** Eggplant

The results presented below are based on the phenotyping provided by the germplasm banks. The analyses have focused on the most agronomic significant cultivated species.

Tomato

The cultivated tomato collection is characterized by a wide range of phenotypic variability, particularly in agronomically relevant traits. This extensive diversity is evident in the broad variation observed across the 25 qualitative descriptors compiled.

Among the significant vegetative descriptors, the growth habit is particularly noteworthy. The collection predominantly exhibits an indeterminate growth habit (55,2%), followed by a determinate growth habit (26,5%) (**Figure 5a**). This trait, regulated by the SP gene family (Carmel-Goren et al., 2003), is crucial in determining the preferred agronomic management strategy in greenhouse or open-field conditions and the intended application, whether for fresh consumption or industrial processing. The style position is another trait exhibiting variation within the collection (**Figure 5b**). Variation in style position, which determines the transition from the allogamy of wild species to the autogamy of cultivated species (Blanca et al., 2012), is observed in the collection. Most styles are inserted (62,4%) compared to slightly exerted and exerted forms. Plants with flowers possessing inserted styles have been selected by humans to fix genotypes adapted to new conditions, favoring self-crossing and minimizing cross-pollination (Chen and Tanksley, 2004). However, exerted styles are commonly found in local varieties.

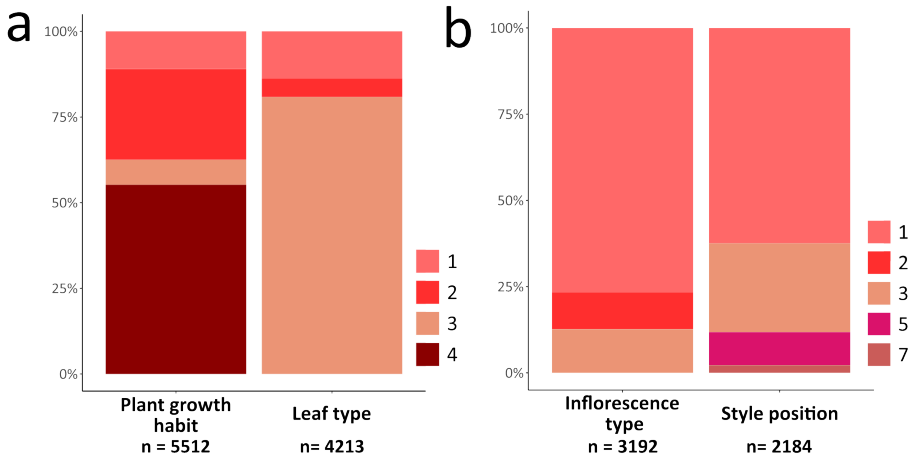


Figure 5. Stacked bar chart of vegetative traits **(a)** Plant growth habit (1: Dwarf; 2: Determinate; 3: Semi-determinate; 4: Indeterminate) and Leaf type (1: Dwarf; 2: Potato leaf type; 3: Standard; 4: Double feathered) and inflorescence traits **(b)** Inflorescence type (1: Uniparous; 2: Forked; 3: Multiparous or irregular) and Style position (1: Inserted; 3: Same level as stamen; 5: Slightly exerted; 7: Highly exerted).

Regarding descriptors related to fruit morphology, the collection exhibits a wide range of fruit shapes and sizes (**Figure 6a**). Flattened (49,1%) and round (32,9%)

shapes are the most abundant, aligning with historical diversification, crop domestication, and market preferences where ribbed forms decreased in favor of round and uniform shapes since the 17th century (Blanca et al., 2023, manuscript in preparation). Medium-sized fruits (50 to 200 g) dominate, but extreme sizes (< 50 g and > 200 g) are also represented. Moreover, there is significant diversity in fruit morphological traits of agronomic importance, such as fasciation, ribbing, and pistil scar (**Figure 6a**). Slightly fasciated fruits (41,8%) with mildly pronounced ribbing (49,9%) and a dot pistil scar shape (59.3%) are most prevalent, likely due to historical selection. Increased ribbing and irregular scars can lead to deformed fruits with numerous potential pathogen entry points, reducing post-harvest preservation and complicating transportation and handling (Bai and Lindhout, 2007).

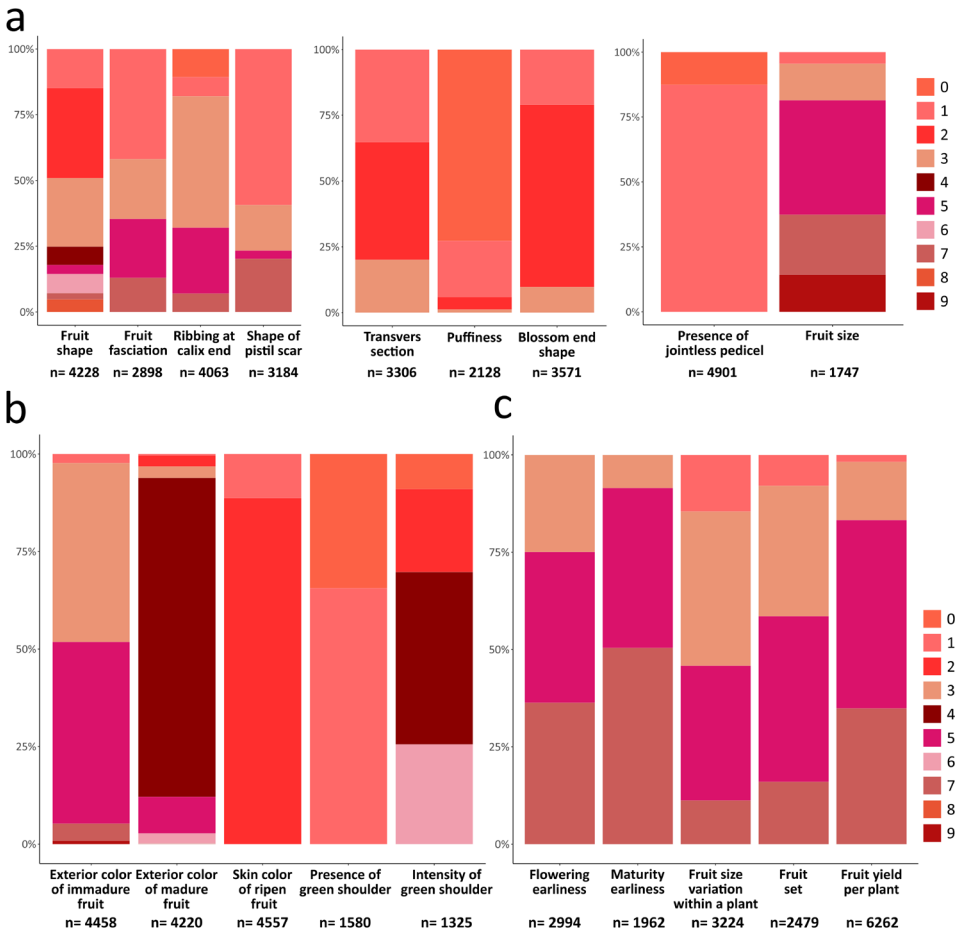


Figure 6. Stacked bar chart of fruit descriptors related to vegetative descriptors **(a)** Predominant fruit shape (1: Flattened; 2: Slightly flattened; 3: Rounded; 4: High rounded; 5: Heart-shaped; 6: Cylindrical; 7: Pear-shaped; 8: Plum-shaped; 9: Others), Fruit fasciation (1: Smooth; 3: Slight; 5: Medium; 7: Severe), Ribbing at calix end (0: Absent; 3: Slight; 5: Medium; 7: Strong), Shape of pistil scar (1: Dot; 3:

Stellate; 5: Linear; 7: Irregular), Transvers section (1: Round; 2: Angular; 3: Irregular), Puffiness (0: Absent; 3: Slight; 5: Medium; 7: Severe), Blossom end shape (1: Indented; 2: Flat; 3: Pointed), Presence of jointless pedicel (0: Absent; 1: Present), and Fruit size (1: Very small (less than 15 g); 3: Small (15 to 50 g); 5: Medium (50 to 400 g); 7: Big (400 to 800 g); 9: Very big (over 1000 g)), fruit color **(b)** Exterior color of immature fruit (1: Greenish-white; 3: Light green; 5: Green; 7: Dark green; 9: Very dark green), Exterior color of mature fruit (1: Green, 2: Yellow, 3: Orange, 4: Red, 5: Pink, 6: Orange-red, 7: Brown, 8: Violet, 9: Others), Skin color of ripen fruit (1: Colorless; 2: Yellow), Presence of green (shoulders) trips on the fruit (0: Absent (Uniform ripening); 1: Present), and Intensity of green shoulders (0: Absent; 3: Slight; 5: Intermediate; 7: Strong) and agronomic descriptors **(c)** Flowering earliness (3: Early; 5: Medium; 7: Late), Maturity earliness (3: Early; 5: Medium; 7: Late), Fruit size variation within a plant (1: Uniform; 3: Slight; 5: Medium; 7: High), Fruit set (1: Low; 3: Intermediate; 5: High; 7: Very high), and Fruit yield per plant (1: Very low; 3: Low; 5: Medium; 7: High; 9: Very high). The letter *n* corresponds to the number of accessions with data.

The descriptors related to fruit color reveal that most accessions in the collection produce red fruits (81,8%), although other colors such as yellow, orange, pink, and brown are also represented (**Figure 6a**). Another trait that exhibits significant variability is the presence and intensity of green shoulders (**Figure 6a**). The variability in fruit color at the maturity stage, skin color and the presence of green shoulders are depicted in **Figure 7ab**.

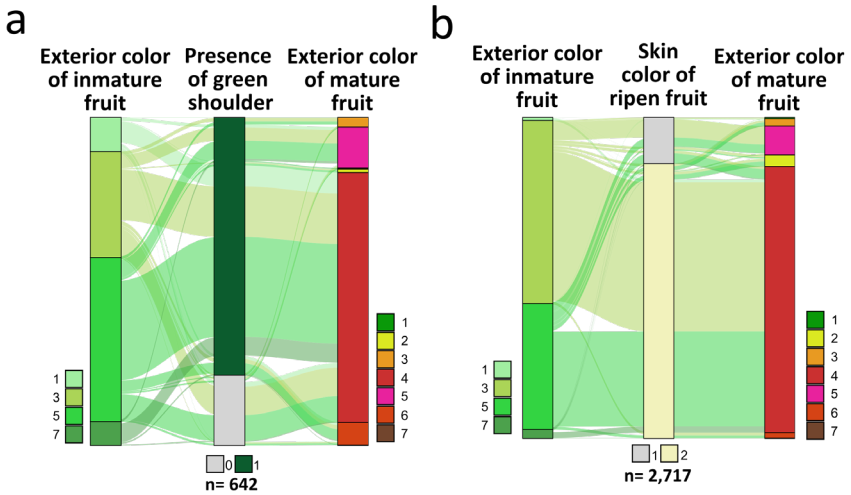


Figure 7. Flowchart of fruit color-related characters. **(a)** Variability of fruit color at an immature stage (1: Greenish-white; 3: Light green; 5: Green; 7: Dark green; 9: Very dark green), mature stage (1: Green, 2: Yellow, 3: Orange, 4: Red, 5: Pink, 6: Orange-red, 7: Brown, 8: Violet, 9: Others), and presence of green shoulders (0: Absent; 1: Present); **(b)** Variability of fruit color at immature stage, skin color (1: Colorless; 2: Yellow), and color at maturity stage. The letter n corresponds to the number of accessions with data.

The collection exhibits a remarkable diversity of agronomically relevant traits such as flowering and maturation earliness, fruit set, and yield (**Figure 6c**), highlighting its potential for selecting and breeding materials with specific characteristics and performance objectives.

For some descriptors, the variation is limited despite having all the ranges represented, such as inflorescence type (**Figure 5b**) and leaf type (**Figure 5a**), where the predominant forms are uniparous inflorescence (76,6%) and standard leaf type (80,8%).

Pepper

The pepper collection is predominantly composed of cultivated species from the three main complexes (**Figure 2b**): Annuum, which includes the widely consumed and highly diverse species *Capsicum annuum* (An), *C. chinense* (Ch), and *C. frutescens* (Fr); Pubescens, which includes *C. pubescens* (Pb), a less consumed species taxonomically more distant from the Annuum complex; and Baccatum, which contains *C. baccatum* (Ba), the species farthest from the other groups.

The extensive diversity within the pepper collection is evident in the wide variation observed across the 26 evaluated descriptors. Plant height and the presence of anthocyanins in the nodes are the vegetative descriptors that display the highest level of heterogeneity (**Figure 8**). Accessions belonging to **An** exhibit intermediate heights (50-100 cm), while those belonging to **Fr**, **Ch**, **Pb**, and **Ba** tend to be shorter and more compact. Regarding the accumulation of anthocyanins in the nodes, a significant percentage of purple nodes is observed in **An** (85,8%), **Ch** (84,1%), **Pb** (77,7%), and **Ba** (64,8%), whereas **Fr** exhibits a higher degree of absence (60,8%).

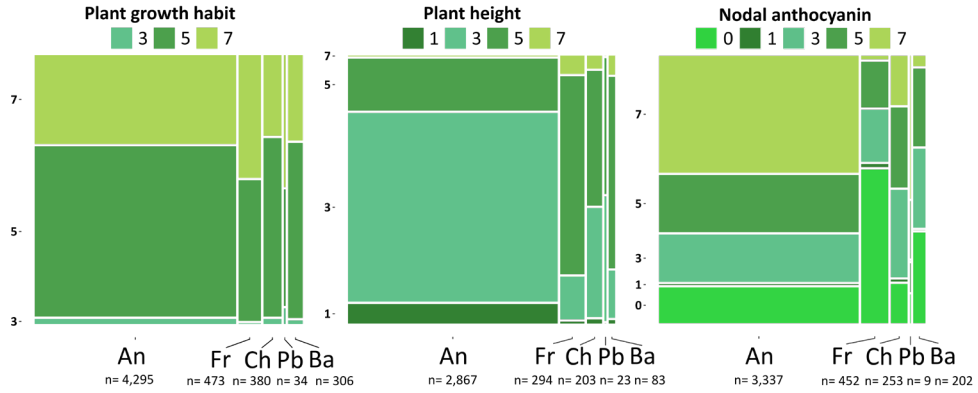


Figure 8. A mosaic plots of cultivated species (*C. annuum* (An), *C. frutescens* (Fr), *C. chinense* (Ch), *C. pubescens* (Pb) and *C. baccatum* (Ba) for vegetative descriptors. Plant growth habit (3: Prostrate; 5: Compact; 7: Erect), Plant height (1: Short (<50cm); 3: Intermediate (50-100cm); 5: Tall (>100cm); 7: Very tall (>200cm)), and Nodal anthocyanins (0: Green; 1: Very pale purple; 3: Pale purple; 5: Purple; 7: Dark purple). The letter *n* corresponds to the number of accessions with data.

In terms of descriptors related to fruit morphology, the collection shows all the shapes, sizes, and types of fruit cross-sections described (**Figure 9a**). The elongated shape (37,7%), conical (34,49%) and rectangular (19,4%) predominate in **An**, while

in **Fr** and **Ba** the most abundant shape is elongated (81,8% and 68,9%), bell-shaped in **Ch** (46,5%) and rectangular in **Pb** (53,3%). Regarding the size of the fruit, **An**, **Ch**, and **Ba** present all categories. On the contrary, the very small fruit size (<4 g) stands out in **Fr**.

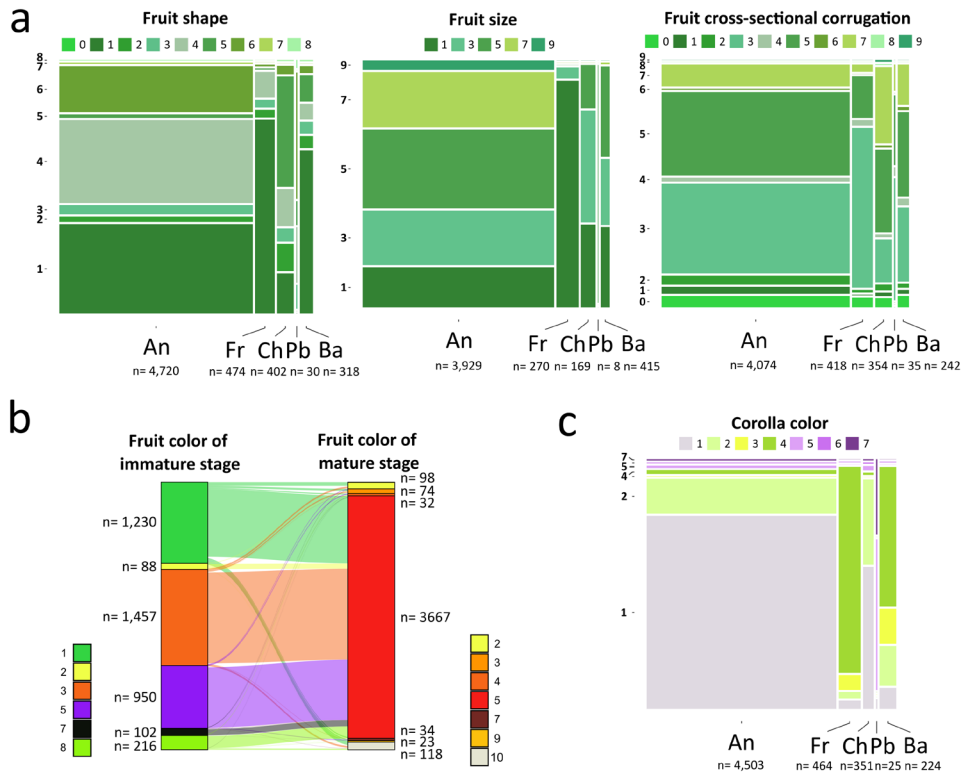


Figure 9. A mosaic plots of cultivated species (*C. annuum* (An), *C. frutescens* (Fr), *C. chinense* (Ch), *C. pubescens* (Pb), and *C. baccatum* (Ba)) for fruit descriptors related to their morphology. **(a)** Fruit shape (1: Elongated; 2: Oblate; 3: Round; 4: Conical; 5: Campanulate; 6: Bell or blocky; 7: Pumpkin shaped; 8: Other), Fruit size (1: Very small (<4 g); 3: Small (4 < x < 10 g); 5: Medium (10 < x < 40 g); 7: Big (40 < x < 150 g); 9: Very big (>150 g), and Fruit cross-sectional corrugation (0: Smooth; 3: Slightly corrugated; 5: Intermediate; 7: Corrugated); **(b)** Flow chart of fruit color in mature and immature states in *C. annuum*. Fruit color at immature stage (1: Green; 2: Yellow; 3: Orange; 4: Red; 5: Purple; 6: Brown; 7: Black; 8: Yellow-green; 9: Other) and Fruti color at mature stage (1: Green; 2: Yellow; 3: Orange; 4: Orange-red; 5: Red; 6: Purple; 7: Brown; 8: Black; 9: Yellow-orange; 10: White). **(c)** A mosaic plot of Corolla color (1: White; 2: Light yellow; 3: Yellow; 4: Yellow-green; 5: White with purple base; 6: White with purple margins; 7: Purple; 8: Other). The letter *n* corresponds to the number of accessions with data.

The flow chart in **Figure 9b** illustrates the extensive diversity observed in the collection, as it showcases the color transformation of fruits from their immature to mature states in **An**. The majority of fruits that are initially green, orange, or purple transition to a red hue when fully mature. Pepper, being naturally abundant

in a variety of beneficial compounds such as carotenoids, phenolic compounds, fats, oils, and minerals, holds great significance as one of the world's most important crops (Rodríguez-Burruezo et al., 2010; Rosa-Martínez et al., 2023). The dual-stage consumption of these fruits adds to their allure, as they exhibit distinct nutritional characteristics depending on their level of ripeness (Ribes-Moya et al., 2018). Consequently, diversity in these traits becomes crucial in identifying and selecting the most desirable attributes and the optimal stage for consumption (Rodriguez-Burruezo and Nuéz, 2006).

Analyzing the corolla color, which plays a pivotal role in taxonomic classification of different species (Ortiz et al., 2010), we observe that **An** and **Ch** predominantly display white and greenish corollas, while **Fr** and **Ba** exhibit green corollas, and **Pb** showcases shades of purple (**Figure 9c**). Additional significant traits for taxonomic differentiation include the presence of spots on the corolla, with **Ba** distinguished by a distinctive yellow-green spot. Likewise, the black seed color in **Pb** provides a clear distinction for this particular species (data not shown).

Among the agronomic and commercial traits, pungency emerges as noteworthy (**Figure 10**). **Fr**, **Ch**, **Pb**, and **Ba** species possess pungent fruits, whereas **An** substantial exhibits variability in pungency, ranging from absence to varying degrees (data not shown). Regarding flowering and maturation earliness (**Figure 10**), the collection demonstrates remarkable diversity. **An** predominantly displays a moderate level of flowering earliness (75,1%) and early to moderate maturation earliness (57,3% and 38,6%, respectively), while **Fr**, **Ch**, **Pb**, and **Ba** exhibit late flowering and moderate maturation earliness.

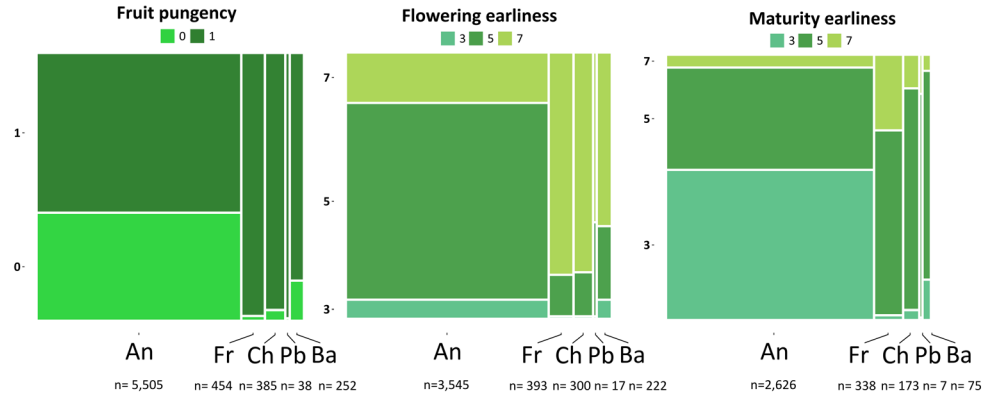


Figure 10. A mosaic plot of cultivated species (*C. annuum* (An), *C. frutescens* (Fr), *C. chinense* (Ch), *C. pubescens* (Pb), and *C. baccatum* (Ba)) for agronomic descriptors. Fruit pungency (0: Non-pungent (sweet); 1: Pungent), Flowering earliness (3: Early; 5: Medium; 7: Late), and Maturation earliness (3: Early; 5: Medium; 7: Late). The letter *n* corresponds to the number of accessions with data.

Eggplant

The eggplant collection, which includes cultivated varieties *S. melongena*, *S. macrocarpon* and *S. aethiopicum*, exhibits a wide variation in most studied traits. This analysis is primarily based on *S. melongena*, the most consumed cultivated species worldwide. The extensive diversity of the eggplant collection is evident in the broad range of variation observed in the 21 analyzed qualitative descriptors.

The collection shows significant variability in vegetative descriptors, encompassing all possible types (**Figure 11a**). Intermediate growth habit (60%) and erect growth habit (30,6%) are the most abundant. The vast majority exhibit varying degrees of anthocyanin pigmentation in stems and leaves, contrasting with their absence (88,7%). Regarding the presence or absence of prickles on leaves, non-prickles accessions predominate (51,5%), which results from domestication and selection processes (Hilgenhof et al., 2023). Additionally, as depicted in **Figure 11b**, the absence of leaf prickles does not influence the presence or absence of prickles in the calyx, giving rise to a wide variety of combinations.

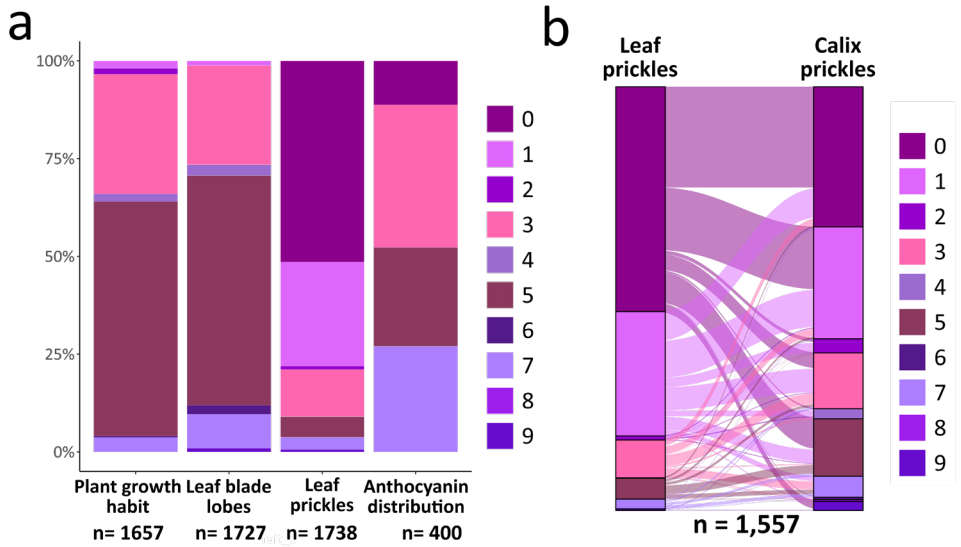


Figure 11. (a) Stacked bar chart of vegetative traits: Plant growth habit (1: Very upright; 2; 3: Upright; 4; 5: Intermediate; 6: Slightly prostrate; 7: Prostrate) and Leaf blade lobes (1: Very weak; 2; 3: Weak; 4; 5: Intermediate; 6; 7: Strong; 9: Very strong), Leaf prickles (0: None; 1: Very few (1-2); 2; 3: Few (3-5); 4; 5: Intermediate (6-10); 7: Many (11-20); 9: Very many (>20)), and General anthocyanin distribution in the plant (apex and stem) (0: Absent; 3: Low; 4; 5: Medium; 7: High). **(b)** Bar chart showing the flow of prickles presence degree in leaf and calyx. *Leaf prickles* (0: None; 1: Very few (1-2); 2; 3: Few (3-5); 4; 5: Intermediate (6-10); 7: Many (11-20); 9: Very many (>20)) and Calyx prickles (0: None; 1: Very few (<3); 3: Few (~5); 5: Intermediate (~10); 7: Many (~20); 9: Very many (>30)). The letter *n* corresponds to the number of accessions with data.

The fruit morphology-related traits (**Figure 12**), particularly the fruit's shape, cross-section, and curvature, demonstrate the significant heterogeneity and diversity present in the collection, primarily caused by preferences in different global markets (Frary et al., 2006). Regarding fruit color in the mature stage, green (35,5%), purple (25,5%), dark purple (11,2%), and fire red (8,3%) stand out, as well as fruits with a uniform color distribution (45,5%) and striped pattern (36,1%)

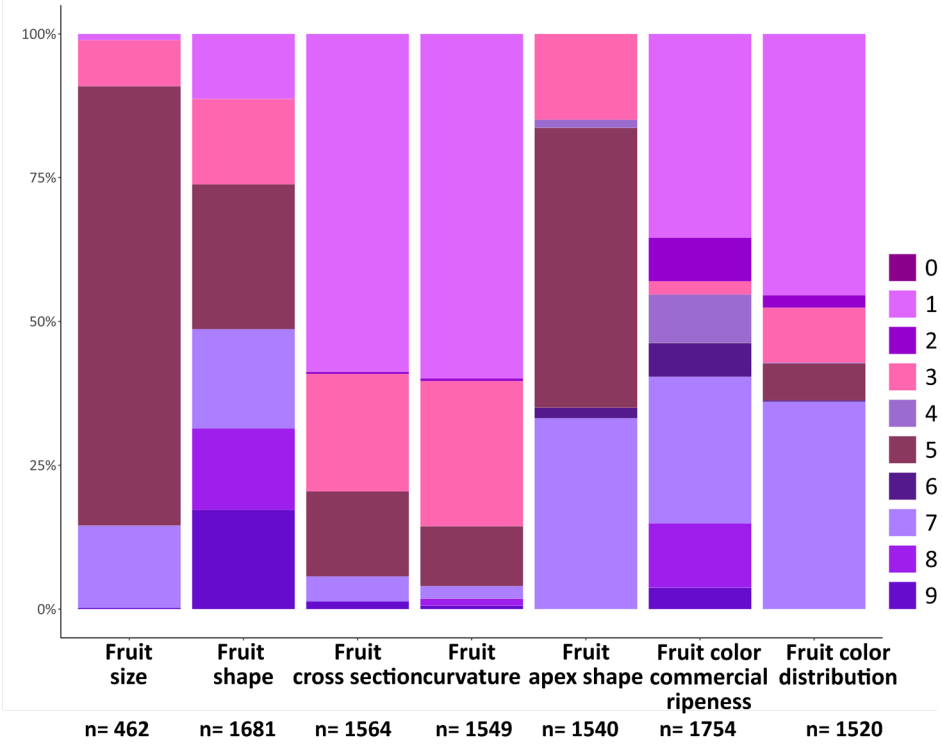


Figure 12. Stacked bar chart of fruit descriptors related to morphology. Fruit size (1: Very small (less than 15 g), 3: Small (15 to 50 g), 5: Medium (50 to 400 g), 7: Big (400 to 800 g), 9: Very big (over 1000 g)), Fruit size (1: Broader than long; 3: As long as broad; 5: Slightly longer than broad; 7: Twice as long as broad; 8: Three times as long as broad; 9: Several times as long as broad), Fruit cross section (1: Circular, no grooves; 3: Elliptic, no grooves; 5: Few grooves (~4); 7: Many grooves (~8); 9: Very irregular), Fruit curvature (1: None; 2: Slight curvature; 3: Slight curvature; 5: Curved; 7: Snake-shaped; 8: Sickle-shaped; 9: U-shaped), Fruit apex shape (3: Protruded; 5: Rounded; 7: Depressed), Fruit color at commercial ripeness (1: Green; 2: Milk white; 3: Deep yellow; 4: Fire red; 5: Scarlet red; 6: Lilac gray; 7: Purple; 8: Dark purple; 9: Black), and Fruit color distribution at commercial ripeness (1: Uniform; 3: Mottle; 5: Netted; 7: Striped). The letter n corresponds to the number of accessions with data.

In the agronomic descriptors, such as flowering earliness and yield, we find all the options represented (**Figure 13a**). These traits are related and, as revealed by the data, are also associated with fruit size (**Figure 12**) and the number of flowers per inflorescence (**Figure 13b**).

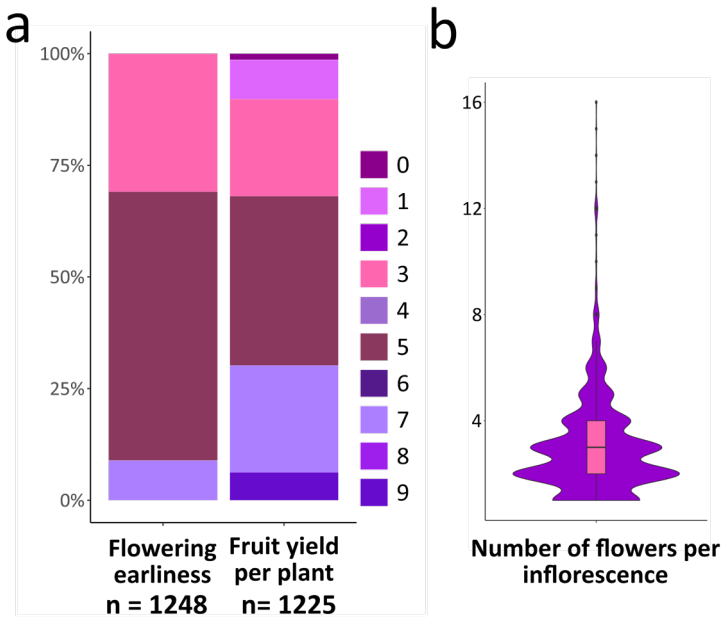


Figure 13. (a) Stacked bar chart of agronomic descriptors. Flowering earliness (3: Early; 5: Medium; 7: Late) and Fruit yield per plant (1: Very low; 3: Low; 5: Medium; 7: High; 9: Very high). **(b)** Violin plot displaying quantitative data on the number of flowers per inflorescence. The letter *n* corresponds to the number of accessions with data

Strengths and weaknesses of the documentation associated with germplasm collections

Understanding and disseminating the genetic diversity stored in germplasm banks is an essential task for the efficient management and subsequent utilization of plant material, both in breeding programs and basic research. Under the frame of the G2P-SOL project, the availability of passport and phenotypic data and images of the complete set of accessions was essential to know the potential of the germplasm to be studied, the extent of phenotypic variability, and the geographical coverage. All this information constitutes the most complete inventory of the most economically relevant Solanaceous crops. The detailed analysis of this vast amount of data is essential for determining duplication among collections, germplasm exchange among genebanks, and the variability of phenotypic characteristics, including different types of materials, cultivated varieties, and wild relatives. Consolidating this huge amount of information into a single database is also relevant for constructing core collections in conjunction with genotyping data.

The joint analysis of passport data from accessions distributed among major European and Asian germplasm banks has provided us a more accurate comprehension of the distribution and composition of genetic resources in these crops. Taking tomato as an example, which has the highest number of accessions (24,767), a significant number of duplicates among collections is evident. To explain this phenomenon, it is necessary to consider the history of germplasm banks worldwide, the interest and leadership in tomato genetic improvement since its inception, and the policies for protecting genetic resources in different countries. In this regard, the predominant role of the United States must be highlighted. The most important tomato collections are housed at the Tomato Genetic Resources Center (Davis, California) and the USA National Plant Germplasm System. Nearly 6,000 accessions from US genebanks and over 1,500 accessions from the TGRC are duplicated in many genebanks worldwide. An exceptional case is WorldVeg from Taiwan, which holds 4,289 accessions from the USA and 615 from the TGRC. Similarly, other important European genebanks also contain numerous accessions from these collections. This indicates that the same germplasm is utilized globally for tomato breeding. Fortunately, the incorporation of wild relatives in breeding programs helps prevent excessive narrowing of the genetic background in modern cultivars. From the analysis of passport data provided by the banks, the weakness of the databases becomes evident, as they lack key information such as the biological status of the accession, cultivar names, collecting date, or codes that facilitate the detection of duplicates. Depending on the research objectives, all this information is of crucial interest. Despite numerous international initiatives (Genesys, EURISCO, WIEWS databases), a considerable amount of work still needs

to be done, as complete and detailed information on conserved germplasm exponentially enhances its utility.

The availability of these data in the Phenome database (https://unity.phenome-networks.com/) developed in this project has facilitated their joint study. In the case of pepper, combining this information with genotyping data has led to the detection of 1,618 duplicates among and within collections and the determination of the global expansion of this crop (Tripodi et al., 2021). Similar studies are about to be published for tomato and eggplant.

Regarding the phenotypic data provided by the germplasm banks, the most notable finding of this study is likely the low proportion of phenotyped accessions, especially in genebanks that have contributed a large number of accessions (**Table 3**). Furthermore, despite the efforts made at the European level by the ECPGR (European Cooperative Programme on Plant Genetic Resources) to standardize the management of plant genetic resources, and by other institutions such as UPOV (The International Union for the Protection of New Varieties of Plants), there is a lack of uniformity in the collected data, requiring significant efforts to rationalize the received data.

Table 3. Number of accessions for different crops in the G2P-SOL collection and the number of accessions with at least one available phenotypic data.

	Tomato		**Pepper**		**Eggplant**	
Institutions	Total number of accessions	With phenotypic data	Total number of accessions	With phenotypic data	Total number of accessions	With phenotypic data
IPK	3.835	1205	1.536	308	112	70
INRAE	1.210	1205	1.300	2.611	1.546	1.057
COMAV-UPV	2.390	1385	1.426	703	260	147
WUR-DLO	1.332	1300	1.033	1.003	510	498
CREA	-	-	300	300	360	360
MVCRI	52	52	67	67	-	-
ENEA	239	0	-	-	-	-
UniTO	-	-	136	98	-	-
WorldVeg	8.545	4711	7.964	6380	3.745	2.269
HUJI	7.114	4251	-	-	-	-
CIP	-	-	-	-	-	-
ARO	-	-	305	0	-	-
BATEM	50	50	50	50	41	41
Total	24.767	14.159	14.117	11.520	6.574	4.442

The conducted study has also highlighted the significant morphological variability contained within the collections. For the majority of the studied descriptors, all

possible variants have been found among the set of accessions. The availability of these data in a single database, which also includes genotyping data and information regarding the constructed core collections for each crop, will enable the qualitative and quantitative utilization of this germplasm in a previously unexplored manner. In addition to passport and phenotypic data, the thousands of images provided by the germplasm banks further enhance the value of the collection.

REFERENCES

Acquadro, A., Barchi, L., Gramazio, P., Portis, E., Vilanova, S., Comino, C., et al. (2017). Coding SNPs analysis highlights genetic relationships and evolution pattern in eggplant complexes. *PLoS One* 12, e0180774.

Alercia, A.; Diulgheroff, S.; Metz, T. (2001). FAO/IPGRI Multi-Crop. Passport Descriptors (MCPD)

Andrews, J. (1995). *Peppers: The domesticated capsicums.* University of Texas Press.

Bai, Y., and Lindhout, P. (2007). Domestication and breeding of tomatoes: What have we gained and what can we gain in the future? *Ann. Bot.* 100, 1085–1094.

Barboza, G. E., García, C. C., Bianchetti, L. de B., Romero, M. V., and Scaldaferro, M. (2022). Monograph of wild and cultivated chili peppers (*Capsicum* L., Solanaceae). *PhytoKeys* 200, 1–423.

Beddows, I., Reddy, A., Kloesges, T., and Rose, L. E. (2017). Population genomics in wild tomatoes—The interplay of divergence and admixture. *Genome Biol. Evol.* 9, 3023–3038.

Bergougnoux, V. (2014). The history of tomato: From domestication to biopharming. *Biotechnol. Adv.* 32, 170–189.

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

Carmel-Goren, L., Liu, Y. S., Lifschitz, E., and Zamir, D. (2003). The self-pruning gene family in tomato. *Plant Mol. Biol.* 52, 1215–1222.

Carrizo García, C., Barfuss, M. H. J., Sehr, E. M., Barboza, G. E., Samuel, R., Moscone, E. A., et al. (2016). Phylogenetic relationships, diversification and expansion of chili peppers (*Capsicum* , Solanaceae). *Ann. Bot.* 118, 35–51.

Chen, K.-Y., and Tanksley, S. D. (2004). High-Resolution mapping and functional analysis of *se2.1*. *Genetics* 168, 1563–1573.

Christodoulakis, N. S., Lampri, P.-N., and Fasseas, C. (2009). Structural and cytochemical investigation of the leaf of silverleaf nightshade (*Solanum elaeagnifolium*), a drought-resistant alien weed of the Greek flora. *Aust. J. Bot.* 57, 432.

Foolad, M. R., Sullenberger, M. T., Ohlson, E. W., and Gugino, B. K. (2014). Response of accessions within tomato wild species, *Solanum pimpinellifolium* to late blight. *Plant Breed.* 133, 401–411.

Frary, A., Doganlar, S., and Daunay, M. C. (2006). Eggplant. Kole C.(ed.) Genome Mapping & Molecular Breeding in Plants, Volume V: Vegetables.

Gil, J. D. B., Reidsma, P., Giller, K., Todman, L., Whitmore, A., and van Ittersum, M. (2019). Sustainable development goal 2: Improved targets and indicators for agriculture and food security. *Ambio* 48, 685–698.

Gramazio, P., Prohens, J., Plazas, M., Mangino, G., Herraiz, F. J., and Vilanova, S. (2017). Development and genetic characterization of advanced backcross materials and an introgression line population of *Solanum incanum* in a *S. melongena* background. *Front. Plant Sci.* 8.

Hilgenhof, R., Gagnon, E., Knapp, S., Aubriot, X., Tepe, E. J., Bohs, L., et al. (2023). Morphological trait evolution in *Solanum* (*Solanaceae*): evolutionary lability of key taxonomic characters. *bioRxiv*

Martínez-Ispizua, E., Calatayud, Á., Marsal, J. I., Mateos-Fernández, R., Díez, M. J., Soler, S., et al. (2021). Phenotyping local eggplant varieties: commitment to biodiversity and nutritional quality preservation. *Front. Plant Sci.* 12.

Nuez, F., Ortega, R. O., and García, J. C. (2003). El Cultivo de Pimientos, Chiles y Ajies. *Mundi-Prensa*.

Ortiz, R., Delgado de la Flor, F., Alvarado, G., and Crossa, J. (2010). Classifying vegetable genetic resources-A case study with domesticated Capsicum spp. *Sci. Hortic.* 126, 186–191.

Pease, J. B., Haak, D. C., Hahn, M. W., and Moyle, L. C. (2016). Phylogenomics Reveals Three Sources of Adaptive Variation during a Rapid Radiation. *PLoS Biol.* 14, 1–24.

Rakha, M., Namisy, A., Chen, J.-R., El-Mahrouk, M. E., Metwally, E., Taha, N., et al. (2020). Development of Interspecific Hybrids between a Cultivated Eggplant Resistant to Bacterial Wilt (Ralstonia solanacearum) and Eggplant Wild Relatives for the Development of Rootstocks. *Plants* 9, 1405.

Reimer, J. J., Thiele, B., Biermann, R. T., Junker-Frohn, L. V., Wiese-Klinkenberg, A., Usadel, B., et al. (2021). Tomato leaves under stress: a comparison of stress response to mild abiotic stress between a cultivated and a wild tomato species. *Plant Mol. Biol.* 107, 177–206.

Ribes-Moya, A. M., Raigón, M. D., Moreno-Peris, E., Fita, A., and Rodríguez-Burruezo, A. (2018). Response to organic cultivation of heirloom Capsicum peppers: Variation in the level of bioactive compounds and effect of ripening. *PLoS One* 13, e0207888.

Rodríguez-Burruezo, A., Kollmannsberger, H., González-Mas, M. C., Nitz, S., and Fernando, N. (2010). HS-SPME comparative analysis of genotypic diversity in the volatile fraction and aroma-contributing compounds of capsicum fruits from the annuum ? chinense ? Frutescens complex. *J. Agric. Food Chem.* 58, 4388–4400.

R**odriguez-Burruezo, A., and Nuéz, F.** (2006). *Mejora de la calidad del pimiento.* In*: Mejora genética de la calidad en plantas*. Universitat Politècnica de València.

Rosa-Martínez, E., Arnaud, B., Plazas, M., Tikunov, Y., Prohens, J., and Pereira-Dias, L. (2023). Genetics and breeding of phenolic content in tomato, eggplant and pepper fruits. *Front. Plant Sci.* 14, 1135237.

Roselló, S., Ricarte, B., José Díez, M., and Nuez, F. (2001). Resistance to tomato spotted wilt virus introgressed from *Lycopersicon peruvianum* in line UPV 1 may be allelic to *Sw-5* and can be used to enhance the resistance of hybrids cultivars. *Euphytica* 119, 357–367

Salinier, J., Lefebvre, V., Besombes, D., Burck, H., Causse, M., Daunay, M. C., et al. (2022). The INRAE centre for vegetable germplasm: geographically and phenotypically diverse collections and their use in genetics and plant breeding. *Plants* 11, 347.

San José, R., Sánchez-Mata, M.-C., Cámara, M., and Prohens, J. (2014). Eggplant fruit composition as affected by the cultivation environment and genetic constitution. *J. Sci. Food Agric.* 94, 2774–2784.

Särkinen, T., Bohs, L., Olmstead, R. G., and Knapp, S. (2013). A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): A dated 1000-tip tree. *BMC Evol. Biol.* 13.

Syfert, M. M., Castañeda-Álvarez, N. P., Khoury, C. K., Särkinen, T., Sosa, C. C., Achicanoy, H. A., et al. (2016). Crop wild relatives of the brinjal eggplant (*Solanum melongena*): Poorly represented in genebanks and many species at risk of extinction. *Am. J. Bot.* 103, 635–651.

Taher, D., Solberg, S. Ø., Prohens, J., Chou, Y., Rakha, M., and Wu, T. (2017). World Vegetable Center Eggplant Collection: Origin, Composition, Seed Dissemination and Utilization in Breeding. *Front. Plant Sci.* 8, 1–12.

Tripodi, P., Rabanus-Wallace, M. T., Barchi, L., Kale, S., Esposito, S., Acquadro, A., et al. (2021). Global range expansion history of pepper (*Capsicum* spp.) revealed by over 10,000 genebank accessions. *Proc. Natl. Acad. Sci. U. S. A.* 118, 1–9.

Villanueva, G., Vilanova, S., Plazas, M., Prohens, J., and Gramazio, P. (2022). Transcriptome profiles of eggplant (*Solanum melongena*) and its wild relative S. dasyphyllum under different levels of osmotic stress provide insights into response mechanisms to drought. *bioRxiv*, 2022.11.02.514830.

Vorontsova, M. S., Stern, S., Bohs, L., and Knapp, S. (2013). African spiny *Solanum* (subgenus *Leptostemonum* , Solanaceae): a thorny phylogenetic tangle. *Bot. J. Linn. Soc.* 173, 176–193.

Wickham, H. (2016). *Elegant Graphics for Data Analysis: ggplot2*.

Yu, G. (2020). Using ggtree to Visualize Data on Tree-Like Structures. *Curr. Protoc. Bioinforma.* 69, e96.

Zhang, M., Zhang, H., Tan, J., Huang, S., Chen, X., Jiang, D., et al. (2021). Transcriptome analysis of eggplant root in response to root-knot nematode infection. *Pathogens* 10, 470.


1

CHAPTER II. SILEX: A fast and inexpensive high-quality DNA extraction method suitable for multiple sequencing platforms and recalcitrant plant species

Santiago Vilanova[1,*,+], David Alonso[1,+], Pietro Gramazio[2], Mariola Plazas[1], Edgar García-Fortea[1], Paola Ferrante[3], Maximilian Schmidt[4], María J. Diez-Niclós[1], Björn Usadel[4,5], Giovanni Giuliano[3] and Jaime Prohens[1]

[+] Equal contribution

[1]Instituto de Conservación y Mejora de la Agrodiversidad Valenciana, Universitat Politècnica de València, Camino de Vera 14, 46022 Valencia, Spain
[2]Faculty of Life and Environmental Sciences, University of Tsukuba, 1-1-1 Tennodai, 305-8572, Tsukuba, Japan
[3]ENEA, Italian National Agency for New Technologies, Energy and Sustainable Economic Development, Rome, Italy
[4]BG-4 Bioinformatics, Forschungszentrum Jülich, 52428 Jülich, Germany
[5]Heinrich Heine University Düsseldorf, CEPLAS, Institute for Biological Data Science, 40225 Düsselforf, Germany

*Corresponding author: sanvina@upv.es

Ph.D. candidate contribution

David Alonso had a main role in the following activities: optimization of the protocol, DNA extraction, data analysis, and manuscript draft writing.

Citation: Vilanova, S., Alonso, D., Gramazio, P. et al. SILEX: a fast and inexpensive high-quality DNA extraction method suitable for multiple sequencing platforms and recalcitrant plant species. Plant Methods 16, 110 (2020). doi: https://doi.org/10.1186/s13007-020-00652-y


2

CHAPTER II

SILEX: A fast and inexpensive high-quality DNA extraction method suitable for multiple sequencing platforms and recalcitrant plant species

Santiago Vilanova[1,*,+], David Alonso[1,+], Pietro Gramazio[2], Mariola Plazas[1], Edgar García-Fortea[1], Paola Ferrante[3], Maximilian Schmidt[4], María J. Diez-Niclós[1], Björn Usadel[4,5], Giovanni Giuliano[3] and Jaime Prohens[1]

[+] Equal contribution

ABSTRACT

Background: The use of sequencing and genotyping platforms has undergone dramatic improvements, enabling the generation of a wealth of genomic information. Despite this progress, the availability of high-quality genomic DNA (gDNA) in sufficient concentrations is often a main limitation, especially for third-generation sequencing platforms. A variety of DNA extraction methods and commercial kits are available. However, many of these are costly and frequently give either low yield or low-quality DNA, inappropriate for next generation sequencing (NGS) platforms. Here, we describe a fast and inexpensive DNA extraction method (SILEX) applicable to a wide range of plant species and tissues.

Results: SILEX is a high-throughput DNA extraction protocol, based on the standard CTAB method with a DNA silica matrix recovery, which allows obtaining NGS-quality high molecular weight genomic plant DNA free of inhibitory compounds. SILEX was compared with a standard CTAB extraction protocol and a common commercial extraction kit in a variety of species, including recalcitrant ones, from different families. In comparison with the other methods, SILEX yielded DNA in higher concentrations and of higher quality. Manual extraction of 48 samples can be done in 96 min by one person at a cost of 0.12 €/sample of reagents and consumables. Hundreds of tomato gDNA samples obtained with either SILEX or the commercial kit were successfully genotyped with Single Primer Enrichment Technology (SPET) with the Illumina HiSeq 2500 platform. Furthermore, DNA extracted from *Solanum elaeagnifolium* using this protocol was assessed by Pulsed-field gel electrophoresis (PFGE), obtaining a suitable size ranges for most sequencing platforms that required high-molecular-weight DNA such as Nanopore or PacBio.

Conclusions: A high-throughput, fast and inexpensive DNA extraction protocol was developed and validated for a wide variety of plants and tissues. SILEX offers an easy, scalable, efficient and inexpensive way to extract DNA for various next-generation sequencing applications including SPET and Nanopore among others.

Keywords: DNA extraction, CTAB protocol, Silica matrix, Contaminant-free DNA, High-molecular-weight DNA, Next-generation sequencing, High-throughput genotyping, recalcitrant species, SPET, Nanopore

BACKGROUND

In the last decade, sequencing and genotyping technologies have become routine, allowing to generate a wealth of genomic information even in non-model plant species and neglected crops (Scheben et al., 2017; Jung et al., 2019). Nowadays, genome sequencing, as well as the most common high-throughput genotyping strategies, like Genotyping-by-Sequencing (GBS; Elshire et al., 2011), Restriction Associated DNA Sequencing (RADseq; Baird et al., 2008) and Single Primer Enrichment Technology (SPET; Scaglione et al., 2019; Barchi et al., 2019), are conducted using next-generation sequencing (NGS) platforms. However, despite the advances, DNA quality is still a main bottleneck, mostly for the third-generation sequencing platforms where high-molecular-weight DNA free of contaminants is required (Vaillancourt and Buell, 2019). Unlike bacteria and mammalian cells, fungi and plant cells are protected by rigid polysaccharide cell walls that hamper the extraction of unfragmented DNA (Anderson et al., 2018). Furthermore, plants produce a wide array of compounds and secondary metabolites (e.g., pigments, phenols, carbohydrates, waxes, among others) that tend to co-precipitate with the DNA and interfere with the subsequent enzymatic reactions (Rana et al., 2019).

So far, the CTAB DNA extraction protocol developed by Doyle and Doyle (Doyle and Doyle, 1990) is one of the most widely used by plant researchers. Several modifications of this protocol have been implemented in order to minimize contamination by other compounds of specific tissues of species (Vaillancourt and Buell, 2019; Healey et al., 2014; Martínez-González et al., 2017). These modifications, apart from being species or tissue-specific and frequently not removing completely interfering compounds, are time-consuming due to many handling steps, and thus are not suitable for high-throughput applications (Barbier et al., 2019; Souza et al., 2019).

Conversely, commercial kits based on silica matrices avoid many of these issues by optimizing the conditions in which only DNA can bind to the silica surface. Therefore, contaminants such as polysaccharides, polyphenols and proteins can be easily removed (Kovačević et al., 2016). They also tend to be faster than the standard CTAB protocol, being the preferred option for sequencing studies in which many samples must be evaluated (Rana et al., 2019; Martín et al., 2019). Usually, commercial kits rely on the reversible interaction between DNA and a silica or silicate support, either in the form of a filter membrane or of silica-coated magnetic particles (Zhou et al., 2018; Park et al., 2019). The adsorption of DNA to the silica

surface is facilitated by buffers with low pH, high concentrations of chaotropic salts (such as guanidinium hydrochloride, guanidinium thiocyanate, or sodium iodide) and ethanol (Boom et al., 1990; Carter and Milton, 1993; Carvalho et al., 2018; Branton et al., 2019). Under these conditions, the surface of the silica can interact with the negative surface of DNA via ionic interactions (Cheng et al., 2016; Shi et al., 2015). After several washes with high concentrations of ethanol to eliminate contaminants, DNA is generally eluted with water or TE at pH 8.0. At this higher pH value, the negatively charged silica surface and DNA repeal each other, releasing the DNA (Katevatis et al., 2017; Green et al., 2018; Toole et al., 2019).

However, commercial kits are usually expensive, with reagent costs commonly ranging between 2 and 9 US$ per sample (Anderson et al., 2018; Pistaka et al., 2019), and many times provide low yields, insufficient for some NGS applications (Anderson et al., 2018; Xia et al., 2019). Furthermore, for some commercial kits, the DNA quality and quantity obtained in recalcitrant species is low (Akkurt, 2012; Marsal et al., 2011; Abdel-Latif et al., 2017). DNA extraction methods relying on silica matrices and chaotropic salts have been reported (Huang et al., 2000; Rogstad et al., 2012; Li et al., 2010; Li and Sheen, 2012); however, chaotropic salts can inhibit subsequent enzymatic reactions which are essential for NGS applications (Vandeventer et al., 2012; Boesenberg-Smith et al., 2012; Emaus et al., 2018).

In this study, we present a novel, fast and inexpensive DNA extraction protocol that combines the advantages of CTAB-based extraction coupled with a purification on a silica matrix. The new method was assessed on different species, including recalcitrant ones and different tissues. To test its suitability for different NGS applications, the method was compared with commercial kits for Single Primer Enrichment Technology (SPET) genotyping (Barchi et al., 2019). The method was also used to extract high-molecular-weight DNA from a recalcitrant wild species (*Solanum elaeagnifolium*). The DNA obtained was successfully used to construct long insert size Nanopore libraries for a de novo genome assembly, which can be difficult for recalcitrant species (Dumschott et al., 2020), thus proving its suitability for third-generation sequencing platforms.

We demonstrate that this new method combines the advantages of commercial kits (high-quality DNA, fast and broad range of species spectrum) with those of a CTAB-based method (high yield and inexpensive) being suitable for routinely DNA screening and NGS platforms.

MATERIALS

Plant material

To test our proposed protocol (hereafter named SILEX, for SILica matrix EXtraction), leaf and fruit tissue from non-recalcitrant species and leaf tissue from recalcitrant species was sampled for four different trials. In a first trial, leaf tissue from a total of 1,860 accessions of tomato (*S. lycopersicum*) and its wild relatives was extracted to compare the quality, quantity and integrity of DNA extracted using SILEX and the commercial kit sbeadex maxi plant kit (hereafter SMP kit; LGC Genomics, Teddington, UK) for SPET genotyping. Extractions were performed on different days over several months

In a second trial, in order to evaluate the appropriateness of SILEX in different plant tissues, 50 mg of fresh and 30 mg of lyophilized fruit tissue of tomato, eggplant (*S. melongena*) and pepper (*Capsicum annuum*) were extracted. The fruit tissue was collected, immediately frozen in liquid N_2 and lyophilized. In a third trial, the suitability of SILEX for DNA extraction in recalcitrant species was assessed using leaves tissue of six species, cassava (*Manihot esculenta)*, grapevine (*Vitis vinifera*), loquat (*Eriobotrya japonica*), banana (*Musa × paradisiaca*), naranjillo (*Solanum bonariense*), and strawberry (*Fragaria × ananassa*), selected to represent a wide range of recalcitrant species presenting different contaminants and secondary metabolites that interfere with DNA extraction. Extractions from recalcitrant plants made by SILEX were compared with those carried out using the standard CTAB protocol (Doyle and Doyle, 1990) and the commercial SMP kit following the manufacturer's instructions. Finally, the suitability of SILEX to extract clean and high-molecular-weight DNA for third-generation sequencing was assessed in the silverleaf nightshade (*S. elaeagnifolium*), a wild relative of eggplant (Knapp et al., 2017), that we selected for the difficulty to obtain contaminant-free DNA due to its high content in phenolics (García-Fortea et al., 2019).

Solutions, reagents and consumables

- 2 ml Sarstedt Microtube (www.sarstedt.com, Cat No. 72.691)
- 5 mm Glass beads (www.vwr.com, Cat No. MARI4901005)
- N-Cetyl-*N,N,N*-trimethylammonium bromide, CTAB (www.itwreagents.com, Cat No. A6284.0500)
- Polyvinylpyrrolidone, PVP-40 (www.sigmaaldrich.com, Cat No. PVP40-500G)
- Ethylenediaminetetraacetic acid, EDTA (www.itwreagents.com, Cat No. 131026.1210)
- Trizma® hydrochloride, Tris HCl (www.sigmaaldrich.com, Cat No. 93363-500G)

- Sodium Chloride (www.itwreagents.com, Cat No. 131659.1211)
- β-Mercaptoethanol (www.sigmaaldrich.com, Cat No. M6250-100ML)
- RNase A (www.vwr.com, Cat No. NA-03)
- Chloroform Essent® (www.scharlab.com, Cat No. CL1981000)
- Isoamyl alcohol Essent® (www.scharlab.com, Cat No. AL02851000)
- Polyethylene Glycol 8000 (www.itwreagents.com, Cat No. 146224.1211)
- Silicon dioxide (www.fishersci.se, Cat No. S5631-500G)
- Hydrochloric acid (www.sigmaaldrich.com, Cat No. H1758-500ML)
- Absolute Ethanol ExpertQ® (www.scharlab.com, Cat No. ET0002025P)
- Tris Base (www.itwreagents.com, Cat No. 14194.1211)

Equipment

- Qiagen TissueLyser II (www.qiagen.com, Cat No. 85300)
- Thermoblock (www.vwr.com, Cat No. 12621-096)
- Eppendorf Centrifuge 5424 (www.eppendorf.com, Cat. No. 5404000010)

Reagent setup

Extraction buffer: 2% (w/v) CTAB, 2% (w/v) PVP-40, 20 mM EDTA, 100 mM Tris HCl (pH 8.0) and 1.40 M NaCl. The buffer may be stored for several months at room temperature.

Protein precipitation buffer: 24 parts of chloroform and 1 part of isoamyl alcohol. It may be stored for several months at room temperature.

Binding buffer: 2.5 M NaCl and 20% PEG 8000. It may be stored for several months at room temperature.

Silica matrix buffer: Mix 5 g of silicon dioxide with 50 ml of MilliQ water and let stand for 24 hours. Discard the supernatant and resuspend the pellet in 50 ml of MilliQ water and let stand for another 5 hours. Discard the supernatant and resuspend the pellet in 1:1 (v/v) MilliQ water. Finally, add 10 µl of HCl 36% per ml of silica matrix solution obtained. It may be stored for several months at room temperature.

Washing buffer: Fresh prepared ethanol 70%. It may be stored for several months at room temperature.

Elution buffer: 10 mM Tris HCl (pH 8.0) and 1 mM EDTA (pH 8.0). It may be stored for several months at room temperature.

PROTOCOL

SILEX DNA extraction protocol

1. Add a 5 mm glass bead to a 2 ml tube, place around 50 mg of fresh or lyophilized tissue into the tube and immediately frozen in liquid nitrogen.
2. Place the tubes in the Qiagen TissueLyser adapters, grind the samples for 60 secs at 13,000 rpm and immediately return the samples in liquid nitrogen.

NOTE: This is a critical step and can greatly influences the final DNA recovery. Avoid sample thaw and pre-chill Qiagen TissueLyser plate adapters. Check that the plant material has become a fine powder.

3. Take the tube from the liquid nitrogen and gently tap it vertically to settle the plant material at the bottom of the tube. Add 1 ml of extraction buffer, add 14 µl of β-mercaptoethanol and gently mix the tube until complete homogenization. Then, add 2 µl of RNase (10 mg/ml) and incubate in a thermoblock for 30 min at 65 °C.

NOTE: Avoid sample thaw before adding the extraction buffer. Preheat the thermoblock at 65 °C.

4. Put the samples on ice for 5 min. Add 700 µl of protein precipitation buffer and gently vortex or (for very high molecular weight DNA) gently invert by hand thoroughly for approximately 20 seconds.
5. Centrifuge at 11,000 rpm for 5 min at room temperature, carefully recover around 800 µl of the supernatant phase and transfer it to a new 2 ml tube.

NOTE: Pipette carefully to avoid dragging the interphase. Interphase contaminants can largely affect the final quality of the DNA.

6. Add 480 µl of binding buffer and gently invert the tube by hand until complete mixing. Subsequently, add 720 µl of absolute ethanol and gently invert the tube again for a few seconds until complete mixing.

NOTE: The amount of binding buffer plus ethanol must be 1.5 volumes of the supernatant recovered in step 5. In the binding buffer plus ethanol mix, 40% should be the binding buffer and 60% should be absolute ethanol.

7. Add 20 µl of silica matrix buffer and mix gently during 5 min (by hand or using an orbital shaker).
8. Spin down the silica for 5 to 6 seconds and discard the supernatant by decantation.

NOTE: Longer centrifugation times will make it difficult to resuspend the silica in the subsequent steps.

9. Add 700 µl of washing buffer and shake gently by hand until a uniform dispersion of the silica is obtained.
10. Spin down the silica for 5 to 6 seconds, gently discard the supernatant by decantation and let dry at room temperature for 5 min.

NOTE: Make sure that all ethanol is completely evaporated.

11. Add 100 µl of elution buffer, shake gently by hand until the pellet is resuspended and incubate 5 min at 65 ºC.
12. Centrifuge at 14,000 rpm for 10 min at room temperature and transfer 90 µl of the supernatant to a new tube.

DNA concentration and quality

DNA integrity was checked by electrophoresis on a 0.8% agarose gel (Condalab, Madrid, Spain) in 1X TAE buffer (GenoChem World, Valencia, Spain) stained with GelRed® (Biotium, Fremont, CA, USA) at a constant voltage of 100 V for 50 min. Gel Doc XR+ System transilluminator (Bio-Rad, Hercules, CA, USA) was used to visualized agarose gels. For high-molecular-weight DNA, the size and integrity were tested by pulse-field gel electrophoresis was run at 3.3 V/cm in 15-second cycles with an angle of 120º for 24 h at 4 ºC with 0.8% agarose in TB buffer.

DNA yield and quality were measured spectrophotometrically using NanoDrop™ ND-1000 (Thermo Scientific, Waltham, MA, USA). A_{260}/A_{280} and A_{260}/A_{230} ratios were measured to determine, respectively, protein and polysaccharide contamination. DNA quantity was also quantified with a Qubit™ 2.0. Fluorometer (Thermo Scientific, Waltham, MA, USA). An aliquot of 2 µl of each sample was examined using the Qubit™ dsDNA BR Assay Kit (Thermo Scientific, Waltham, MA, USA) according to the instructions of the manufacturer.

In addition, the concentration of DNA obtained from the 1,860 tomato samples was measured fluorometrically using Quant-iT™ PicoGreen™ dsDNA Assay Kit (hereafter PicoGreen, Thermo Scientific, Waltham, MA, USA) and a 96-wells plate reader VICTOR3 1420 (PerkinElmer, Waltham, MA, USA) equipped with an excitation filter F485 and emission filter F535.
To check the suitability of the DNA extraction method for sequencing applications where DNA is fragmented, approximately 1 µg of DNA was digested for 1 h at 37 ºC followed by 20 min at 65 ºC with restriction enzymes *Eco*RI (New England Biolabs, Ipswich, MA, USA). The digestion was evaluated through 1% agarose electrophoresis as above.

High-throughput genotyping quality check

For the first trial, sequencing of tomato samples for genotyping by SPET was performed with an Illumina NextSeq 500 platform (Illumina Inc., San Diego, CA, USA), following the manufacturer protocol. *Phred* values were obtained using FastQC Version 0.11.8. and plotted in R (Ihaka and Gentleman, 1996) using the package ggplot2 (Wickham, 2016).

Suitability of extracted DNA for third-generation sequencing platforms

For the third trial, 5 µg of *S. elaeagnifolium* DNA from a single extraction were size-selected using the Circulomics SRE-XL-Kit (Circulomics Inc., Baltimore, MD, USA). For

library preparation, 1 µg of the size-selected DNA was used to prepare each of the three Nanopore LSK-109 libraries. Two of these libraries were sequenced on a MinION R9.4.1 (Oxford Nanopore, Oxford, UK) and the third was loaded on a PromethION PRO-002 (Oxford Nanopore, Oxford, UK). All three sequencing runs were basecalled using Oxford Nanopores Guppy basecaller version 3.2.2 (Oxford Nanopore, Oxford, UK) using the high accuracy basecalling models.

RESULTS

Comparison of DNA extraction methods

Tomato leaf samples

Total DNA yield extracted through the SMP kit and estimated by NanoDrop ranged from 14.5 ng/mg to 366.9 ng/mg with a mean of 38.3 ng/mg and a standard deviation (SD) of 29.2 ng/mg. DNA extracted by SILEX showed higher output, ranging from 86.1 ng/mg to 1,698.1 ng/mg with an average of 382.9 ng/mg and a SD of 205.3 ng/mg (**Table 1**). Despite higher SD, the coefficient of variation (CV) of SILEX (53.6%) was lower than that of the SMP kit (76.1%).

Table 1. Mean value, standard deviation (SD), range and coefficient of variation (CV) of the DNA yield (ng/mg) using SILEX and SMP kit and quantified by NanoDrop (ND) and PicoGreen (PG).

	SILEX		**SMP kit**	
	ND	**PG**	**ND**	**PG**
Samples (n)	1,380		480	
Mean	382.9	141.3	38.3	41.7
SD	205.3	36.8	29.2	26.4
Range	86.1–1,698.1	37.9–231.2	14.5–366.9	1.2–134.8
CV (%)	53.6	26.1	76.1	63.4
Ratio ND/PG	2.7		0.9	

The A_{260}/A_{280} ratio, which indicates protein contamination, was very variable in the SMP kit protocol, ranging from 1.15 to 2.32, with an average of 1.76 and a SD of 0.33 (**Figure 1a**). In contrast, SILEX showed a more consistent ratio with less variation (from 1.91 to 2.12) and with an average value of 2.03 and a SD of 0.05. Similarly, for the A_{260}/A_{230} ratio, which indicates salt and carbohydrates contamination, SMP kit showed a greater dispersion, with a ratio between 0.27 and 2.43 with an average of 1.09 and a SD of 2.55, compared to SILEX, which ranged from 1.16 to 2.16 with an average of 1.66 and a SD of 0.25 (**Figure 1b**) (**Table 2**).

Table 2. Mean value, standard deviation (SD), range and coefficient of variation (CV) of NanoDrop absorbance ratios (A_{260}/A_{280} and A_{260}/A_{230}) using SILEX and SMP kit.

	SILEX		**SMP kit**	
	A_{260}/A_{280}	**A_{260}/A_{230}**	**A_{260}/A_{280}**	**A_{260}/A_{230}**
Samples (n)	1,380		480	
Mean	2.03	1.66	1.76	1.17
SD	0.05	0.22	0.22	0.45
Range	1.91–2.12	1.16–2.16	1.15–2.32	0.27–2.43
CV (%)	2.5	13.3	12.5	38.5

n: Samples indicate the number of independent extractions performed

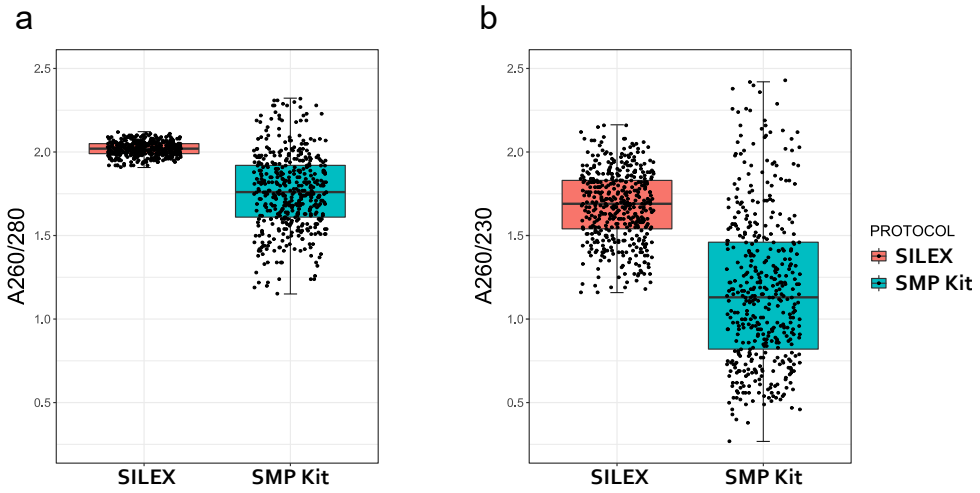


Figure 1. Quality control of DNA extracted with SILEX (orange) and SMP kit (blue) in 1,380 and 480 tomato samples respectively. Box and whisker plots are based on **(a)** A_{260}/A_{280} and **(b)** A_{260}/A_{230}. Each dot represents a sample, the median is indicated by a horizontal line, the box represents the upper and lower quartiles and the whiskers show the variability outside the quartiles.

Since spectrophotometric measurements with NanoDrop tend to overestimate DNA yield due to likely interferences of proteins (Ponti et al., 2018), those measures were compared with the fluorometric ones performed with PicoGreen. Yields estimated by the latter ranged from 1.2 ng/mg to 134.8 ng/mg with a mean of 41.7 ng/mg and a SD of 26.4 ng/mg in the case of DNA extracted by SMP kit. On the other hand, SILEX had higher yields, ranging from 37.9 ng/mg to 231.2 ng/mg with a mean of 141.3 ng/mg and a SD of 36.8 ng/mg (**Table 1**). In addition, yields estimated by PicoGreen had greater variation between samples in DNA extracted by SMP kit (63.4%) in comparison with that extracted by SILEX (26.1%).

To assess the overestimation of DNA yield extracted using the different protocols, we compared the ratios obtained by NanoDrop and PicoGreen measurements. Estimation of yield by NanoDrop of DNA extracted with the SMP kit showed an estimation of 0.9-fold compared to PicoGreen, suggesting that for this commercial kit NanoDrop measurements were comparable with the PicoGreen ones. In contrast, NanoDrop measurements from SILEX tended to overestimate DNA yield 2.7-fold compared to PicoGreen, suggesting contamination with a molecule absorbing at 260 nm. One possible explanation for this overestimation is that remnants of degraded RNA were present in our samples, since nanodrop is unable to discriminate among free nucleotides, RNA, single-stranded DNA, and double-stranded DNA. However, even with this overestimation, the average yield obtained with SILEX (141.3 ng/mg with PicoGreen) was 3.4 times higher than with SMP kit (41.7 ng/mg).

DNA extraction from dry and fresh fruit tissues

The amount of DNA obtained from dry and fresh fruit tissues was similar to that achieved using leaf tissue and ranged from 116.4 to 920.3 ng/mg. In general, higher yield of DNA was obtained with lyophilized tissue (**Table 3**). Regardless of the tissue used, A_{260}/A_{280} ratios were above 2.0 which indicates no protein contamination. On the other hand, lower ratios were observed in A_{260}/A_{230} ratio, suggesting the presence of some organic contaminants. Despite these ratios, DNA obtained was successfully digested by *Hind*III restriction enzyme.

Yield and quality control of DNA extracted by lyophilized and fresh fruit tissue of tomato, eggplant and pepper measured using NanoDrop.

Table 3. Yield and quality control of DNA extracted by lyophilized and fresh fruit tissue of tomato, eggplant and pepper measured using NanoDrop.

Species	Fruit sample	Yield (ng/mg)	Absorbance ratios	
			A_{260}/A_{280}	A_{260}/A_{230}
Tomato	Lyophilized	116.4 ± 65.3	2.08 ± 0.08	1.30 ± 0.11
	Fresh	237.6 ± 40.2	2.22 ± 0.04	0.98 ± 0.08
Eggplant	Lyophilized	863.0 ± 289.0	2.10 ± 0.05	1.36 ± 0.13
	Fresh	253.0 ± 63.8	2.11 ± 0.04	1.38 ± 0.06
Pepper	Lyophilized	920.3 ± 248.7	2.10 ± 0.04	1.46 ± 0.14
	Fresh	271.6 ± 54.0	2.08 ± 0.03	1.18 ± 0.15

Mean value and standard deviation of 12 independent extractions are shown.

DNA extraction from recalcitrant species

Overall, SILEX resulted in higher DNA yields, ranging (with fluorimetric determination) from 46.4 ng/mg in strawberry to 318.0 ng/mg in grapevine, than those obtained with the standard CTAB method or the SMP kit (**Table 4**). In addition, A_{260}/A_{280} ratios obtained with SILEX were above 2.0, which is considered a protein-free DNA, except in strawberry where values were on average 1.86, even though they were higher than in standard CTAB and SMP kit. In the same way, SILEX A_{260}/A_{230} ratios were higher than in the two other protocols, from 1.71 in loquat to 2.16 in banana, except in strawberry, where the results were similar to standard CTAB and SMP kit (**Table 4**). The differences were very noticeable in banana and grapevine, where the A_{260}/A_{230} ratios were 2.7 and 4.3-fold lower for standard CTAB and 1.7 and 3.3-fold lower for SMP kit, respectively (**Table 4**). NanoDrop/Qubit ratio estimated with SILEX for recalcitrant species seemed to be species-dependent, and ranged from 1.4-fold in grapevine to 6.6-fold in naranjillo with a mean of 3.9-fold in comparison to Qubit. However, even though the SMP kit provided lower NanoDrop/Qubit ratios, SILEX performed better than the standard CTAB, which on average had a NanoDrop/Qubit ratio of 18-fold.

Table 4. Yield and quality control of DNA extracted from six recalcitrant species using three different methods: SILEX, standard CTAB, and SMP kit quantified by NanoDrop (ND) and Qubit (Q).

Species/ Methods	Yield (ng/mg)		Absorbance ratios		Ratio ND/Q
	ND	Q	A_{260}/A_{280}	A_{260}/A_{230}	
Cassava					
SILEX	556.0 ± 76.6	106.9 ± 17.4	2.04 ± 0.02	2.13 ± 0.05	5.3
CTAB	512.7 ± 61.1	37.7 ± 27.7	2.12 ± 0.02	1.96 ± 009	13.6
SMP kit	75.4 ± 9.4	85.5 ± 13.9	1.80 ± 0.09	1.44 ± 0.10	0.9
Grapevine					
SILEX	442.8 ± 24.1	318.0 ± 115.2	2.07 ± 0.02	1.93 ± 0.01	1.4
CTAB	394.3 ± 18.7	21.5 ± 6.1	1.60 ± 0.05	0.44 ± 0.03	18.4
SMP kit	50.7 ± 30.2	0.0 ± 0.0	1.73 ± 0.20	1.09 ± 0.09	∞
Loquat					
SILEX	284.4 ± 24.9	92.7 ± 35.5	2.02 ± 0.05	1.71 ± 0.06	3.1
CTAB	112.7 ± 7.5	4.3 ± 2.1	1.58 ± 0.17	0.57 ± 0.08	25.9
SMP kit	66.7 ± 36.0	7.7 ± 6.1	1.78 ± 0.25	1.15 ± 0.41	8.6
Banana					
SILEX	267.8 ± 5.1	59.2 ± 8.2	2.13 ± 0.01	2.16 ± 0.05	4.5
CTAB	202.0 ± 37.5	19.1 ± 5.6	2.20 ± 0.14	0.78 ± 0.13	10.6
SMP kit	31.4 ± 37.5	1.3 ± 2.3	1.47 ± 0.31	0.74 ± 0.32	23.3
Naranjillo					
SILEX	1,184.1 ± 484.3	180.4 ± 60.6	2.06 ± 0.02	1.85 ± 0.03	6.6
CTAB	525.4 ± 126.2	26.0 ± 7.7	1.91 ± 0.11	0.86 ± 0.20	20.2
SMP kit	105.4 ± 35.8	42.6 ± 19.8	1.82 ± 0.04	1.72 ± 0.07	2.5
Strawberry					
SILEX	193.8 ± 5.1	46.4 ± 6.1	1.86 ± 0.05	1.02 ± 0.08	4.2
CTAB	405.3 ± 108.5	24.6 ± 4.0	1.77 ± 0.08	1.14 ± 0.11	16.5
SMP kit	25.9 ± 2.8	20.0 ± 14.3	1.69 ± 0.16	0.97 ± 0.02	1.3

Ratios of the latter are reported. Mean value and standard deviation are based on a minimum of three independent extractions.

In order to test if the presence of contaminants could inhibit the enzyme activity, DNA was digested with the restriction enzyme *Eco*RI. Agarose gels, such as the one shown in **Figure 2** indicated efficient endonuclease activity in all the DNA extracted from the six recalcitrant species even though in some cases A_{260}/A_{230} ratios were below 1.8 (strawberry and loquat). Also, strawberry samples showed yellow and brown coloration and high viscosity even after two washing steps.

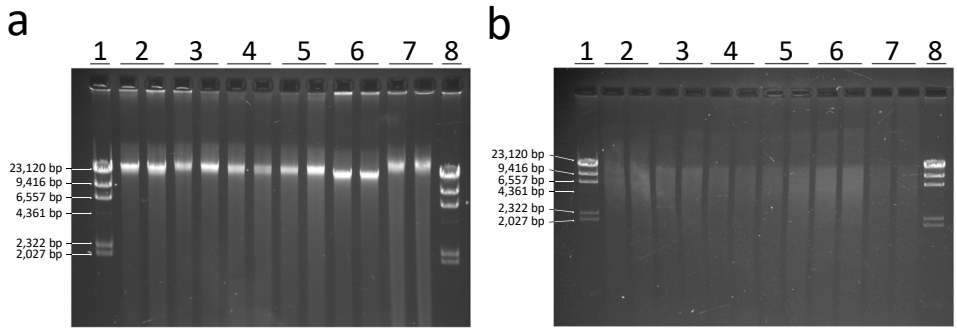


Figure 2. Agarose gel electrophoresis of uncut genomic DNA extracted from six recalcitrant species with SILEX **(a)** and the same DNA cut with *Eco*RI enzyme. **(b)** Two biological replicates for each species are shown. *Lambda* DNA restricted with *Hind*III (lane 1 and 8); cassava (lane 2); grapevine (lane 3); loquat (lane 4); banana (lane 5); naranjillo (lane 6) and strawberry (lane 7).

SILEX Timing and cost

The time needed to extract 48 samples, without taking into account the sampling of the plant material) is approximately 96 minutes (around 2 min/sample; **Figure 3**). The estimated cost of all consumables required to extract high-molecular-weight gDNA using SILEX is approximately 0.12 € per sample (Supplementary data S1).

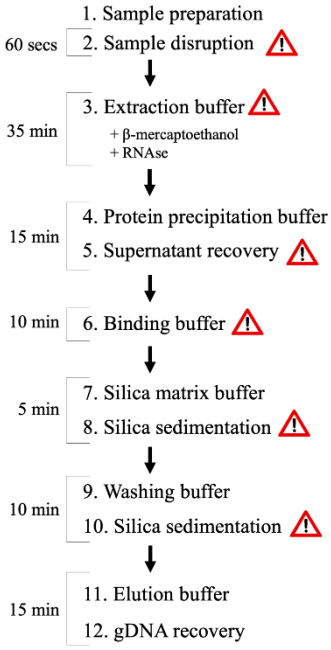


Figure 3. Flowchart of the twelve steps of SILEX DNA extraction method and timing estimation to perform 48 samples in tubes. The warning signal indicates key points (see in the text the notes in the protocol section).

High-throughput genotyping platforms

In order to evaluate the suitability of the gDNA obtained with SILEX for high-throughput genotyping, 1,380 tomato samples were genotyped using SPET (Barchi et al., 2019). The reads obtained showed excellent *Phred*-quality scores along the 150 bp, with a mean value of 33.1 (**Figure 4**). Similar results were obtained with 480 samples extracted using the SMP kit with a mean value of 33.5. The mean *Phred* score along the 150 bp sequenced was always over 30, indicating good sequencing quality in both methods, with the SILEX method providing more DNA per equal amount of tissue.

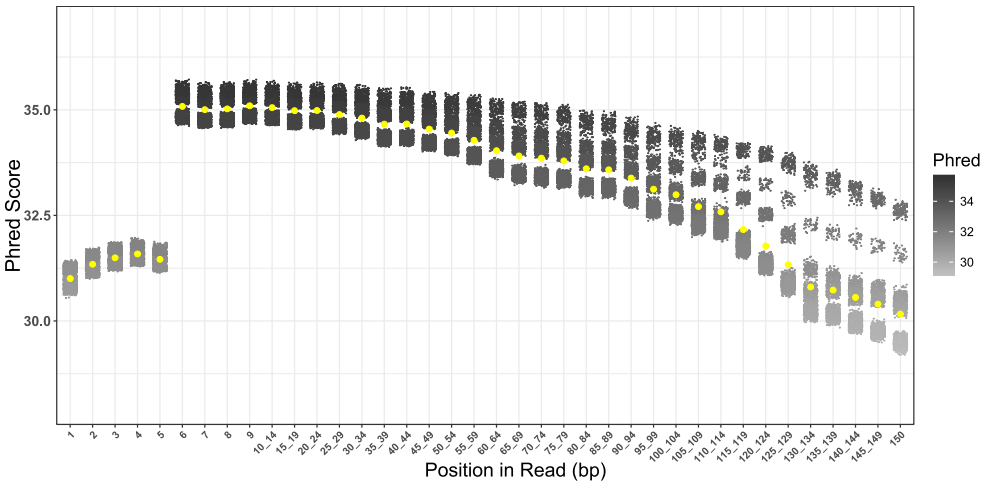


Figure 4. Summary of *Phred* values for 1,380 tomato samples extracted with the SILEX protocol and genotyped with SPET along the 150 bp sequence. Each dot corresponds to one sample. Yellow spots indicate the mean value.

High molecular weight DNA extraction

To test the suitability of using SILEX for NGS platforms requiring high-molecular-weight DNA, *S. elaeagnifolium* DNA was size-selected using the Circulomics short read eliminator kit, recovering 3.5 µg, and analysed using Pulsed-field gel electrophoresis (PFGE) (**Figure 5**). The size-selected DNA ranged from 20 to 100 Kb and contained relatively little small fragments. The sizes obtained were suitable for most sequencing platforms that require high molecular weight DNA free of impurities such as Nanopore or PacBio. Two libraries were sequenced with a MinION sequencer and yielded 6.8 and 7.5 Gbp, with N50 values of 22.8 and 28.2 kbp, respectively, while the third library, sequenced with a PromethION sequencer, resulted in 55.7 Gbp of raw data with N50 of 24.2 kbp. After the base-calling, the total sequencing yield was 70.1 Gbp.

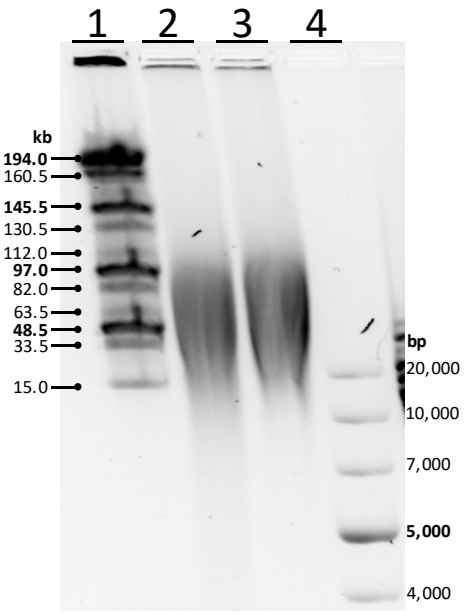


Figure 5. *Solanum elaeagnifolium* gDNA size estimation using PFGE after extraction with SILEX. Line 1 is MidRange PFG Marker (New England Biolabs, Ipswich, USA) and line 4 is GeneRuler 1 kb Plus DNA (Thermo Fisher Scientific, Waltham, USA). Line 2 shows the gDNA before size selection and line 3 the gDNA before size selection after using Short Read Eliminator XL Kit.

DISCUSSION

One of the main advantages of the SILEX protocol for DNA extraction is the use of common and inexpensive reagents and its simplicity. No toxic salts such as guanidinium thiocyanate or sodium iodide at high concentrations are used. Several authors reported that the use of NaCl at concentrations higher than 2 M facilitated the DNA binding to the silica surface (Lakshmi et al., 1999; Taylor et al., 2000; Prodělalová et al., 2004). Also, it is known that the addition of polyethylene glycol (PEG) to the binding solution increases the adsorption due to the compact globular structure of DNA adopted under these conditions (Shan et al., 2015). For these reasons, we use a binding buffer composed by the non-toxic, inexpensive NaCl and PEG compound to facilitate the DNA binding to the silica surface. The total cost of reagents and consumables is only 0.12 € per sample and for multiple simultaneous manual extractions, each sample requires less than 2 min per person. In this respect, in the SILEX method, the silica matrix used for each extraction cost less than 0.001 € and the washing buffer is only water and ethanol, a common non-toxic reagent in most molecular biology laboratories.

The protocol presented here has been tested on many samples of different species with similar satisfactory results, confirming its wide applicability. The quality and quantity parameters obtained also indicate that SILEX is at least as effective as commercial kits even when recalcitrant species were used. In recalcitrant species, the presence of polysaccharides and phenols was significantly lower in SILEX compared to the standard CTAB protocol where several samples showed yellow and brown coloration and high viscosity, indicating the presence of oxidized polyphenols and high concentration of polysaccharides. One of the reasons for this difference could be the absence of a precipitation step in SILEX, as polysaccharides and polyphenols tend to co-precipitate with DNA when isopropanol or ethanol is added (Greco et al., 2014). This is important since the presence of polysaccharides such as carrageenan, pectin and xylan are strong inhibitors of PCR (Schrader et al., 2012; Demeke et al., 1992). We also observed that in species with very high polyphenol and polysaccharide compounds, such as strawberry (Asami et al., 2003), a second washing step increases the A_{260}/A_{230} ratio. Our DNA extraction protocol has been tested by other research groups and it has been found to provide high-quality DNA in high concentrations in other plant species as different as silver fir (*Abies alba*), watermelon (*Citrullus lanatus*), melon (*Cucumis melo*), summer squash (*Cucumis pepo*), common fig (*Ficus carica*), lettuce (*Lactuca sativa*), European larch (*Larix decidua*), Spanish stonecrop (*Sedum hispanicum*), avocado (*Persea americana*) and sweet cherry (*Prunus avium*). Applications have included genotyping by SSRs (Single Sequence Repeats), HRM (High-Resolution Melting), and GBS among others.

Although DNA is usually extracted from fresh leaf tissue, it is sometimes necessary to use other types of material such as fresh or freeze-dried fruit. Our protocol was flexible enough to successfully extract high DNA quantities from lyophilized and fresh fruit tissues obtaining A_{260}/A_{280} ratios above 2.0.

Thousands of samples of tomato and wild relatives were successfully genotyped using SPET high-throughput genotyping, that relays on DNA fragmentation, target probe annealing, PCR amplification and NGS sequencing (Barchi et al., 2019). The quality of the reads produced had a mean *Phred* value over 30, which represents a base call accuracy of 99.9%. Also, hundreds of samples of grapevine and watermelon were genotyped using GBS, obtaining similar results (C. Esteras, personal communication). This indicates the suitability of SILEX to yield DNA of enough quality to be used in different genotyping platforms. Furthermore, our DNA extraction method could be used in applications requiring high molecular weight genomic DNA, such as long-read single molecule Nanopore sequencing (Schimidt et al., 2017) without any additional steps.

CONCLUSIONS

The SILEX protocol presented here is very robust and can be used in a wide variety of plants (including recalcitrant ones) and several tissues. It is based on common reagents without the need of expensive salts or equipment. This makes it inexpensive (0.12 € of reagents and consumables per sample) and accessible to most laboratories. It is also a fast method, where a trained person could process up to 48 samples in 96 minutes using Eppendorf tubes or 192 min if the extraction is performed in 96-well plates. The protocol is also amenable to automatization specially in labs that already have automatic DNA extraction robots. This could save hands-on time and increase the number of samples processed per day. We demonstrate that this new method gathers the advantages of commercial kits (high-quality DNA, fast and broad species spectrum) with those of the CTAB-based method (high yield and inexpensive), being suitable for routine DNA extraction for multiple applications, including NGS platforms.

SUPPLEMENTARY MATERIAL

Supplementary information accompanies this paper at: https://doi.org/10.1186/s13007-020-00652-y . Additionally, supplemental tables can be found at the end of this chapter.

AUTHORS' CONTRIBUTIONS

SV conceptualized the study and developed the protocol. SV, DA, PG, EG and MP optimized the protocol. DA extracted tomato samples and PF and GG performed SPET genotyping. MS and BU performed Nanopore sequencing. DA, EG and MP performed the rest of extractions. MJD, BU, GG and JP supervised the study and provides resources. SV, DA, PG, MP, PF, MS, MJD and JP analyzed the data. SV, DA, PG and JP wrote the manuscript draft. All author read and approved the final manuscript.

FUNDING

This research has been funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No 677379 (Linking genetic resources, genomes and phenotypes of Solanaceous crops; G2P-SOL). David Alonso is grateful to Universitat Politècnica de València for a predoc- toral (PAID-01-16) contract under the Programa de Ayudas de Investigación y Desarrollo initiative. Mariola Plazas is grateful to Generalitat Valenciana and Fondo Social Europeo for a postdoctoral grant (APOSTD/2018/014). Pietro Gramazio is grateful to Japan Society for the Promotion of Science for a Postdoctoral Grant (P19105, FY2019 JSPS Postdoctoral Fellowship for Research in Japan (Standard)). The Spanish Ministerio de Educación, Cultura y Deporte funded a predoctoral fellowship granted to Edgar García-Fortea (FPU17/02389).

REFERENCE

Scheben, A., Batley, J., Edwards, D. (2017). Genotyping-by-sequencing approaches to characterize crop genomes: choosing the right tool for the right application. *Plant Biotechnol J.,* 15, 149-161.

Jung, H., Winefield, C., Bombarely, A., Prentis, P., Waterhouse, P. (2019). Tools and strategies for long-read sequencing and de novo assembly of plant genomes. *Trends Plant Sci.* 24, 700-724.

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.

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A., Johnson, E.A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One*. 3, e3376.

Scaglione, D., Pinosio, S., Marroni, F., Centa, E., Fornasiero, A., Magris, G., Scalabrin, S., Cattonaro, F., Taylor, G., Morgante, M. (2019). Single primer enrichment technology as a tool for massive genotyping: A benchmark on black poplar and maize. *Ann Bot.* 124, 543-551.

Barchi, L., Acquadro, A., Alonso, D., Aprea, G., Bassolino, L., Demurtas, O., Ferrante, P., Gramazio, P., Mini, P., Portis, E., Scaglione, D., Toppino, L., Vilanova, S., Díez, M.J., Rotino, G., Lanteri, S., Prohens, J., Giuliano, G. (2019). Single primer enrichment technology (SPET) for high-throughput genotyping in tomato and eggplant germplasm. *Front Plant Sci*. 10, 1005.

Vaillancourt, B., Buell, C.R. (2019). High molecular weight DNA isolation method from diverse plant species for use with Oxford Nanopore sequencing. *BioRxiv.* 2019.783159.

Anderson, C.B., Franzmayr, B.K., Hong, S.W., Larking, A.C., van Stijn, T.C., Tan, R., Moraga, R., Faville, M., Griffiths, A. (2018). Protocol: a versatile, inexpensive, high-throughput plant genomic DNA extraction method suitable for genotyping-by-sequencing. *Plant Methods*. 14, 75.

Rana, M.M., Aycan, M., Takamatsu, T., Kaneko, K., Mitsui, T., Itoh, K. (2019). Optimized nuclear pellet method for extracting next-generation sequencing quality genomic DNA from fresh leaf tissue. *Methods Protoc.* 2, 54.

Doyle, J.J., Doyle, J.L. (1990). Isolation of plant DNA from fresh tissue. *Focus.* 12, 13-15.

Healey, A., Furtado, A., Cooper, T., Henry, R.J. (2014). Protocol: a simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species. *Plant Methods*. 10, 21.

Martínez-González, C.R., Ramírez-Mendoza, R., Jiménez-Ramírez, J., Gallegos-Vázquez, C., Luna-Vega, I. (2017). Improved method for genomic DNA extraction for *Opuntia* Mill. (Cactaceae). *Plant Methods*. 13, 82.

Barbier, F.F., Chabikwa, T.G., Ahsan, M.U., Cook S.E., Powell R., Tanurdzic, M., Beveridge, C. (2019). A phenol/chloroform-free method to extract nucleic acids from recalcitrant, woody tropical species for gene expression and sequencing. *Plant Methods*. 15, 62.

Souza, D.C., Teixeira, T.A. (2019). A simple and effective method to obtain high DNA quality and quantity from Cerrado plant species. *Mol Biol Rep*. 46, 4611-4615.

Kovačević, N. (2016). Magnetic beads based nucleic acid purification for molecular biology applications. Sample Preparation Techniques for Soil, Plant, and Animal Samples. In*: Micic M. Springer Protoc Handb*. 53-67

Martin, S.L., Parent, J.S., Laforest, M., Page, E., Kreiner, J.M., James, T. (2019). Population genomic approaches for weed science. *Plants.* 8, 354.

Zhou, Y., Zhang, Y., He, W., Wang, J., Peng, F., Huang, L., Zhao, S., Deng, W. (2018). Rapid regeneration and reuse of silica columns from PCR purification and gel extraction kits. *Sci Rep*. 8, 12870.

Park, H.J., Cho, H., Jung, H.S., Cho, B.H., Lee, M.Y. (2019). Development of a DNA isolation device using poly(3,4-dihydroxy-L-phenylalanine)-coated swab for on-site molecular diagnostics. *Sci Rep.* 9, 8144.

Boom, R., Sol, C.J., Salimans, M.M., Jansen, C.L., Wertheim-van Dillen, P.M., van der Noordaa, J. (1990). Rapid and simple method for purification of nucleic acids. *J Clin Microbiol.* 28, 495–503.

Carter, M.J., Milton, I.D. (1993). An inexpensive and simple method for DNA purifications on silica particles. *Nucleic Acids Res.* 21, 1044.

Carvalho, J., Puertas, G., Gaspar, J., Azinheiro, S., Diéguez, L., Garrido-Maestu, A., Vázquez, M., Barros-Velázquez, J., Cardoso, S., Padro, M. (2018). Highly efficient DNA extraction and purification from olive oil on a washable and reusable miniaturized device. *Anal Chim Acta.* 1020, 30–40.

Branton, D., Deamer, D., Quick, J., Loman, N.J. (2019). DNA extraction strategies for nanopore sequencing. *Nanopore Seq. World Sci.* 1, 91–105.

Cheng, H., Zhang, K., Libera, J., De La Cruz, M., Bedzyk, M. (2016). Polynucleotide adsorption to negatively charged surfaces in divalent salt solutions. *Biophys J.* 90, 1164–1174.

Shi, B., Shin, Y., Hassanali, A., Singer, S. (2015). DNA Binding to the Silica Surface. *J Phys Chem B.* 119, 11030–11040.

Katevatis, C., Fan, A., Klapperich, C.M. (2017). Low concentration DNA extraction and recovery using a silica solid phase. *PLoS One.* 12, e0176848.

Green, M.R., Sambrook, J. (2018). Isolation and quantification of DNA. *Cold Spring Harb Protoc. 2018*. 2018:403–14.

Toole, K., Roffey, P., Young, E., Cho, K., Shaw, T., Smith, M., Blagojevic, N. (2019). Evaluation of commercial forensic DNA extraction kits for decontamination and extraction of DNA from biological samples contaminated with radionuclides. *Forensic Sci Int.* 302, 109867.

Piskata, Z., Servusova, E., Babak, V., Nesvadbova, M., Borilova, G. (2019). The quality of DNA isolated from processed food and feed via different extraction procedures. *Molecules.* 24, 1188.

Xia, Y., Chen, F., Du, Y., Liu, C., Bu, G., Xin, Y., Boye, L. (2019). A modified SDS-based DNA extraction method from raw soybean. *Biosci Rep.* 39.

Akkurt M. (2012). Comparison between modified DNA extraction protocols and commercial isolation kits in grapevine (*Vitis vinifera* L.). *Genet Mol Res.* 11, 2343–2351.

Marsal, G., Baiges, I., Canals, J.M., Zamora, F., Fort, F. (2011). A Fast, efficient method for extracting DNA from leaves, stems, and seeds of *Vitis vinifera* L. *Am J Enol Vitic.* 62, 376–81.

Abdel-Latif, A., Osman, G. (2017). Comparison of three genomic DNA extraction methods to obtain high DNA quality from maize. *Plant Methods.* 13, 1.

Huang, J., Ge, X., Sun, M. (2000). Modified CTAB protocol using a silica matrix for isolation of plant genomic DNA. *Biotechniques.* 28, 432–434.

Rogstad, S.H. (2012). Plant DNA extraction using silica. *Plant Mol Biol Report.* 21, 463–463.

Li, J-F., Li, L., Sheen, J. (2010). Protocol: a rapid and economical procedure for purification of plasmid or plant DNA with diverse applications in plant biology. *Plant Methods.* 6, 1.

Li, J-F., Sheen, J. (2012). DNA purification from multiple sources in plant research with homemade silica resins. *Humana Press.* 862, 53–59.

Vandeventer, P.E., Lin, J.S., Zwang, T.J., Nadim, A., Johal, M.S., Niemz, A. (2012). Multiphasic DNA adsorption to silica surfaces under varying buffer, pH, and ionic strength conditions. *J Phys Chem B.* 116, 5661–5670.

Boesenberg-Smith, K.A., Pessarakli, M.M., Wolk, D.M. (2012). Assessment of DNA yield and purity: an overlooked detail of PCR troubleshooting. *Clin Microbiol Newsl.* 34, 1–6.

Emaus, M.N., Clark, K.D., Hinners, P., Anderson, J.L. (2018). Preconcentration of DNA using magnetic ionic liquids that are compatible with real-time PCR for rapid nucleic acid quantification. *Anal Bioanal Chem*. 410, 4135–4144.

Dumschott, K., Schmidt, M.H.W., Chawla, H.S., Snowdon, R., Usadel, B. (2020). Oxford Nanopore Sequencing: New opportunities for plant genomics? *J Exp Bot.* 18, 5313-5322

Knapp, S., Sagona, E., Carbonell, A.K.Z., Chiarini, F. (2017). A revision of the *Solanum elaeagnifolium* clade (Elaeagnifolium clade; subgenus *Leptostemonum*, Solanaceae). *PhytoKeys.* 84, 1–104.

García-Fortea, E., Gramazio, P., Vilanova, S., Fita, A., Mangino, G., Villanueva, G., Arrones, A., Knapp, S., Prohens, J., Plazas, M. (2019). First successful backcrossing towards eggplant (*Solanum melongena*) of a New World species, the silverleaf nightshade (*S. elaeagnifolium*), and characterization of interspecific hybrids and backcrosses. *Sci Hort.* 246, 563–573.

I**haka, R., Gentleman, R.** (1996). R: A Language for data analysis and graphics. *J Comput Graph Stat.* 5, 3299–314.

Wickham H. (2016). ggplot2: Elegant graphics for data analysis. In: *Springer-Verlag New York*

Ponti, G., Maccaferri, M., Manfredini, M., Kaleci, S., Mandrioli, M., Pellacani, G., Ozben, T., Depenni, R., Bianchi, G., Pirola, G., Tomasi, A. (2018). The value of fluorimetry (Qubit) and spectrophotometry (NanoDrop) in the quantification of cell-free DNA (cfDNA) in malignant melanoma and prostate cancer patients. *Clinica Chimica Acta.* 479, 14–19.

Lakshmi, R., Baskar, V., Ranga, U. (1999). Extraction of superior-quality plasmid DNA by a combination of modified alkaline lysis and silica matrix. *Anal Biochem.* 272, 109–112.

Taylor, J.I., Hurst, C.D., Davies, M.J., Sachsinger, N., Bruce, I.J. (2000). Application of magnetite and silica–magnetite composites to the isolation of genomic DNA. *J Chromatogr A*. 890, 159–166.

Prodělalová, J., Rittich, B., Španová, A., Petrová, K., Beneš, M.J. (2004). Isolation of genomic DNA using magnetic cobalt ferrite and silica particles. *J Chromatogr A*. 1056, 43–48.

Shan, Z., Jiang, Y., Guo, M., Bennett, J.C., Li, X., Tian, H., Oakes, K., Zhang, Zhou. Y., Huang, Q., Chen, H. (2015). Promoting DNA loading on magnetic nanoparticles using a DNA condensation strategy. *Colloids Surfaces B Biointerfaces*. 125, 247–254.

Greco, M., Sáez, C., Brown, M., Bitonti, M. (2014). A simple and effective method for high quality co-extraction of genomic DNA and total RNA from low biomass *Ectocarpus siliculosus*, the model brown alga. *PLoS One*. 9, e96470.

Schrader, C., Schielke, A., Ellerbroek, L., Johne, R. (2012). PCR inhibitor – occurrence, properties and removal. J Appl Microbiol. 113, 1014–1026

Demeke, T., Adams, R.P. (1992) The effects of plant polysaccharides and buffer additives on PCR. *BioTechniques*. 12, 332–334.

Asami, D.K., Hong, Y.J., Barrett, D.M., Mitchell, A.E. (2003). Comparison of the total phenolic and ascorbic acid content of freeze-dried and air-dried marionberry, strawberry, and corn grown using conventional, organic, and sustainable agricultural practices. *J Agric Food Chem.* 51, 1237–1241.

Schmidt, M., Vogel, A., Denton, A., Istace, B., Wormit, A., van de Geest, H., Bolger, M., Alseekh, S., Maß, J., Pfaff, C., Schurr, U., Chetelat, R., Maumus, F., Aury, J., Koren, S., Fernie, A., Zamir, D., Bolger, A., Usadel, B. (2017). De novo assembly of a new *Solanum pennellii* accession using nanopore sequencing. *Plant cell.* 29, 2336–


2

CHAPTER III. Single primer enrichment technology (SPET) for high-throughput genotyping in tomato and eggplant germplasm

Lorenzo Barchi[1], Alberto Acquadro[1], David Alonso[2], Giuseppe Aprea[3], Laura Bassolino[4], Olivia Demurtas[3], Paola Ferrante[3], Pietro Gramazio[2], Paola Mini[3], Ezio Portis[1], Davide Scaglione[5], Laura Toppino[4], Santiago Vilanova[2], María José Díez[2], Giuseppe Leonardo Rotino[4], Sergio Lanteri[1*], Jaime Prohens[2*] and Giovanni Giuliano[3]

[*] Correspondence: Sergio Lanteri (sergio.langeri@unito.it) and Jaime Prohens (jprohens@btc.upv.es)
[1]DISAFA, University of Turin, Turin, Italy
[2]Instituto de Conservación y Mejora de la Agrodiversidad Valenciana, Universitat Politècnica de València, Camino de Vera 14, 46022 Valencia, Spain
[3]ENEA, Italian National Agency for New Technologies, Energy and Sustainable Economic Development, Rome, Italy
[4]CREA-GB, Research Centre for Genomics and Bioinformatics, Montanaso Lombardo, Italy
[5]IGA Technology Services Srl, Udine, Italy

Ph.D. candidate contribution

David Alonso had a main role in the following activities: providing the plant material and extraction of DNA samples, data analysis and writing of the manuscript draft.

Citation: Barchi L., Acquadro A., Alonso D., Aprea G., Bassolino L., Demurtas O., Ferrante P., Gramazio P., Mini P., Portis E., Scaglione D., Toppino L., Vilanova S., Díez M.J., Rotino G.L., Lanteri S., Prohens J. and Giuliano G. (2019) Single Primer Enrichment Technology (SPET) for High-Throughput Genotyping in Tomato and Eggplant Germplasm. *Front. Plant Sci.* 10:1005. doi: https://doi.org/10.3389/fpls.2019.01005


3

CHAPTER III

Single primer enrichment technology (SPET) for high-throughput genotyping in tomato and eggplant germplasm

Lorenzo Barchi[1], Alberto Acquadro[1], David Alonso[2], Giuseppe Aprea[3], Laura Bassolino[4], Olivia Demurtas[3], Paola Ferrante[3], Pietro Gramazio[2], Paola Mini[3], Ezio Portis[1], Davide Scaglione[5], Laura Toppino[4], Santiago Vilanova[2], María José Díez[2], Giuseppe Leonardo Rotino[4], Sergio Lanteri[1*], Jaime Prohens[2*] and Giovanni Giuliano[3]

[*] Correspondence: Sergio Lanteri (sergio.langeri@unito.it) and Jaime Prohens (jprohens@btc.upv.es)

ABSTRACT

Single primer enrichment technology (SPET) is a new, robust and customizable solution for targeted genotyping. Unlike genotyping by sequencing (GBS), and like DNA chips, SPET is a targeted genotyping technology, relying on the sequencing of a region flanking a primer. Its reliance on single primers, rather than on primer pairs, greatly simplifies panel design, and allows higher levels of multiplexing than PCR-based genotyping. Thanks to the sequencing of the regions surrounding the target SNP, SPET allows the discovery of thousands of closely linked, novel SNPs. In order to assess the potential of SPET for high-throughput genotyping in plants, a panel comprising 5k target SNPs, designed both on coding regions and introns/UTRs, was developed for tomato and eggplant. Genotyping of two panels composed of 400 tomato and 422 eggplant accessions, comprising both domesticated material and wild relatives, generated a total of 12,002 and 30,731 high confidence SNPs, respectively, which comprised both target and novel SNPs in an approximate ratio of 1:1.6, and 1:5.5 in tomato and eggplant, respectively. The vast majority of the markers was transferrable to related species that diverged up to 3.4 million years ago (*Solanum pennellii* for tomato and *S. macrocarpon* for eggplant). Maximum Likelihood phylogenetic trees and PCA outputs obtained from the whole dataset highlighted genetic relationships among accessions and species which were congruent with what was previously reported in literature. Better discrimination among domesticated accessions was achieved by using the target SNPs, while better discrimination among wild species was achieved using the whole SNP dataset. Our results reveal that SPET genotyping is a robust, high-throughput technology for genetic fingerprinting, with a high degree of cross-transferability between crops and their cultivated and wild relatives, and allows identification of duplicates and mislabeled accessions in genebanks.

Keywords: SPET, genotyping, tomato, eggplant, germplasm

INTRODUCTION

Single nucleotide polymorphisms (SNPs) are the most abundant type of sequence variation in eukaryotic genomes and have emerged as the most widely used genotyping markers (Mammadov et al., 2012). Genotyping methods rely on different technologies, including next-generation sequencing (Davey et al., 2011), DNA microarrays (Hoheisel, 2006), and polymerase chain reaction (PCR) (Semagn et al., 2014). A widely used method for high-throughput SNP discovery and genotyping is genotyping by sequencing (GBS) based on different reduced-representation sequencing (RRS) approaches, the majority of which are based on the use of restriction enzymes (Elshire et al., 2011; Scheben et al., 2017). One major limitation of GBS is the random distribution of restriction enzyme sites on the genome, and thus the inability to target markers localized within genes, or having a functional significance.

Recently Nugen® developed the single primer enrichment technology (SPET, United States Patent 9,650,628), which is a customizable solution for targeted sequencing at an affordable price. SPET requires a priori genomic or transcriptomic information and identification of SNPs for probe design. SPET probes are around 40-bases long and are designed adjacent to a region containing a sequence variant, thus enabling detection of both the SNPs and the discovery of additional ones, surrounding the target one (**Figure 1**). Up to now, SPET has been applied for medical purposes (Scolnick et al., 2015; Nairismägi et al., 2016). Its application to plant materials is still largely unexplored, with one recent exception (Scaglione et al., 2019) assessing SPET application to *Zea mays* L. and to *Populus nigra* L. To date, SPET has not been applied for genotyping of germplasm sets, or of genepools including several related species. The relatively high sequence conservation of exons should facilitate the hybridization of SPET probes designed on these regions across different related species and thus increase the chances to identify novel SNPs, especially if the region downstream of the probe falls in less conserved regions such as introns and Untranslated regions UTR (Castle, 2011). Application of SPET to plant materials for which the genetic diversity and relationships are already known would allow its validation as a reliable and robust high-throughput genotyping method in germplasm sets.

Tomato (*Solanum lycopersicum* L.) and eggplant (*S. melongena* L.) are amongst the economically most important vegetables, and the diversity and genetic relationships of their genepools has been extensively studied. Several studies have applied sequence-, PCR- or microarray-based genotyping in tomato and eggplant, to analyze the genetic diversity and population structure of a limited number of cultivars, breeding lines, landraces or cultivated, and wild relatives (e.g., Vilanova et

al., 2012; Cericola et al., 2013; Acquadro et al., 2017; Pailles et al., 2017; Tranchida-Lombardo et al., 2018).

Cultivated tomato materials maintained in germplasm collections include traditional varieties and heirlooms which, compared to its wild relatives, display a narrow genetic diversity resulting from several bottlenecks during domestication and spread (Blanca et al., 2015). Regarding wild species, many studies have been performed evaluating the relationships between them and cultivated tomato (Rodriguez et al., 2009; Aflitos et al., 2014; Lin et al., 2014; Dodsworth et al., 2016; Beddows et al., 2017). The general consensus, using different molecular approaches, is that within the core tomato clade (*Solanum* section *Lycopersicon*), the wild species genetically closest to cultivated tomato are those of the "Lycopersicon" group (Peralta et al., 2008), including *S. pimpinellifolium* L. and the Galápagos Islands endemisms *S. cheesmaniae* (L. Riley) Fosberg and *S. galapagense* S.C. Darwin & Peralta. While the genetic diversity of the latter two species is limited (Pailles et al., 2017), *S. pimpinellifolium* is much more diverse than cultivated tomato heirlooms (Caicedo and Schaal, 2004; Razali et al., 2018). The next closest wild species to the "Lycopersicon" group are those of the "Arcanum" group, which includes *S. arcanum* Peralta, *S. chmielewskii* (C.M. Rick, Kesicki, Fobes & M. Holle) D.M. Spooner, G.J. Anderson & R.K. Jansen, and *S. neorickii* D.M. Spooner, G.J. Anderson & R.K. Jansen, followed by the five species of the "Eriopersicon" group (*S. huaylasense* Peralta, *S. chilense* (Dunal) Reiche, *S. corneliomulleri* J.F. Macbr, *S. peruvianum* L. and *S. habrochaites* S. Knapp & D.M. Spooner), and *S. pennellii* Correll, the only species included in the monotypic "Neolycopersicon" group (Rodriguez et al., 2009; Aflitos et al., 2014; Lin et al., 2014; Dodsworth et al., 2016; Beddows et al., 2017).

Unlike tomato, eggplant belongs to the *Solanum* subgenus *Leptostemonum*, collectively known as the "spiny solanums" group (Vorontsova et al., 2013). Several species from the eggplant clade, such as the direct wild ancestor *S. insanum* L. and several close relatives such as *S. incanum* L., *S. lichtensteinii* Willd., and *S. linnaeanum* Hepper & P-M.L. Jaeger are closely related to eggplant (Vorontsova et al., 2013; Acquadro et al., 2017). Other more distant species include some from the Anguivi grade, which comprises the two other cultivated species: scarlet eggplant (*S. aethiopicum* L.) and gboma eggplant (*S. macrocarpon* L.) and their respective wild ancestors *S. anguivi* Lam. and *S. dasyphyllum* Schumach. & Thonn. as well as other species of potential for breeding such as *S. tomentosum* L. (Kouassi et al., 2016; Plazas et al., 2016). A much more distant group includes American species (Vorontsova et al., 2013; Syfert et al., 2016; Acquadro et al., 2017), among which *S. torvum* Sw. and *S. sisymbriifolium* Lam. represent a potential source of tolerance to diseases for eggplant breeding (Daunay and Hazra, 2012).

Here, we report the application of SPET genotyping, assessing its reliability and using both target and non-target SNPs for studying the genetic variation and population structure of a large set of accessions from cultivated and wild genepools of tomato and eggplant, and to validate them against previous results on their diversity and genetic relationships. The results obtained highlight the potential of the SPET technology for genotyping and management of germplasm collections.

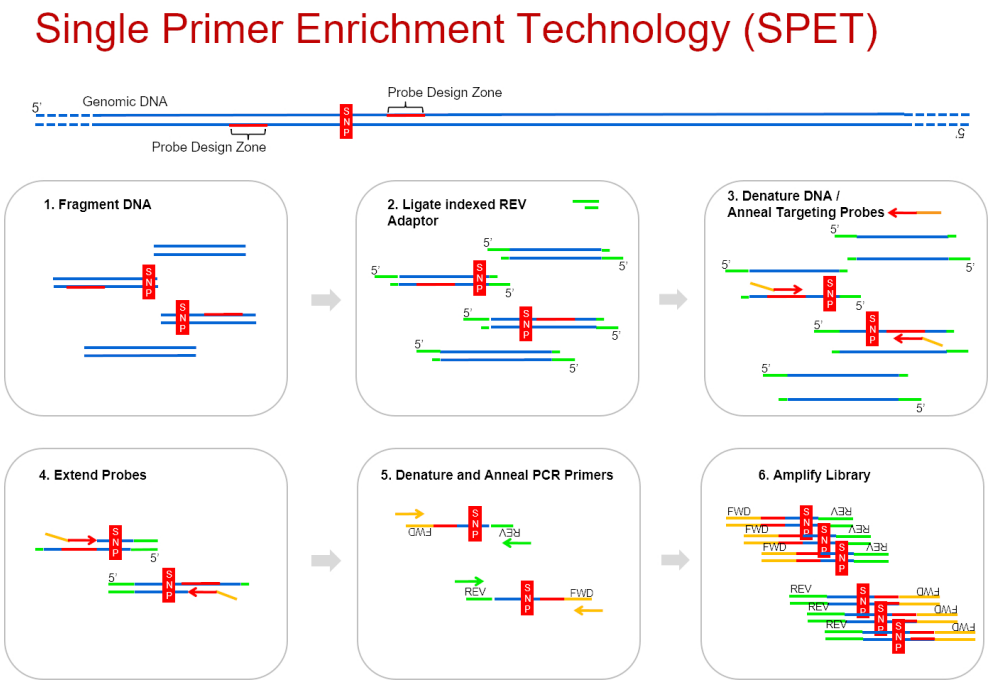


Figure 1. The six main steps of the SPET workflow. Probes can be designed up or downstream the identified SNP.

MATERIALS AND METHODS

Single Primer Enrichment Technology (SPET) set up

Tomato SNP data were retrieved from the SOL Genomics portal (http://solgenomics.net) and specifically from the "150 Tomato Genome Resequencing Project" (Aflitos et al., 2014) and the "AGIS Tomato 360 Resequencing Project" (Lin et al., 2014). Respectively, 52 samples from "150 Tomato Resequencing Project" and 184 samples from "AGIS Tomato 360 Resequencing Project" were used to mine alleles of *S. lycopersicum*. To identify intervarietal alleles across *S. lycopersicum*, *S. pimpinellifolium*, and *S. lycopersicum* var. *cerasiforme*, 159 additional samples were used from the "AGIS Tomato 360 Resequencing Project". All resequenced accessions used in this work are available through the SOL Genomics FTP site (ftp://ftp.solgenomics.net/genomes/) and listed in **Supplementary Table S1**.

Accessions from different VCF files were merged, retaining only simple biallelic SNPs. Since SNP calling was based on ITAG SL2.50 genome build, this was maintained as reference through the analysis with respective gene models. SNP selection was then made based on the following criteria: (i) only positions with alternative cohort-wise allele count greater than 8 (summing two from homozygous loci or one from heterozygous loci), (ii) SNPs within introns and UTRs had to be at least 15 kbp apart from each other or with SNPs in CDS, (iii) SNPs within CDS had to be at least 5 kbp apart from other selected SNPs, and (iv) SNPs have to reside on anchored chromosomes.

For the eggplant panel design sequencing data provided by the Universitat Politècnica de València and by the Italian Eggplant Consortium, which included the whole-genome resequencing of eight *S. melongena* and one *S. incanum* accessions (Gramazio et al., 2019), were aligned against the "67/3" eggplant reference genome (Barchi et al., 2019) with BWA-MEM aligner with default parameters (Li, 2013), discarding multiple-mapping reads. Samtools was used for variant calling (Li et al., 2009). Homozygous/heterozygous SNP calls were considered only with phred-scaled genotype likelihood equal to zero. Similarly to tomato, only biallelic SNPs with an alternative allele count of 4 and a minor allele frequency greater than 0.25 were retained.

From the eligible polymorphic sites previously identified in both the species, a randomly selected panel of SNPs were forwarded for probe design to NuGen (San Carlos, CA, United States). These filtered tomato and eggplant panels were then tested for their sequencing performances and reproducibility, and to identify a final set of about 5k probes for genotyping via SPET, commercialized under the name of

Allegro®. To this purpose, 24 accessions of both *S. lycopersicum* and *S. melongena* were used in a pilot experiment. In eggplant, the re-call performance was assessed with filtered VCF files using as parameters: (i) homozygous states called with a minimum of 10 reads; (ii) MAF (minor allele frequency) >0.04; (iii) heterozygosity comprised between MAF 0.25–0.5; (iv) a minimum genotyping ratio of 80% (i.e., 80% of samples must satisfy the above constraints). This allowed to retain 7,662 out of 11,625 SNPs (72%).

Besides cross-species genomic sequence conservation and the frequency of expected SNP detection, in both crops, probes were first filtered based on their ability to hit the target SNPs in the first 25 bp after 3'-end of the probe. After the first pilot run with about twice the number of probes, a coverage analysis was used to select the final set of 5,000 and 5,082 probes in tomato and eggplant respectively. Sites showing an average coverage ranging from 46× to 90× and with less than four samples having five or less mapped reads were retained for tomato, while sites with a coverage range of 79–130× and excluding all those probes with three or more individual showing a coverage below 5× were selected for eggplant. Furthermore, 82 extra probes were added to the final eggplant set for specific functional purposes. Sequencing yields, coverage analysis, and filtering for probe selection are provided in **Supplementary File S1**.

Plant Material

For tomato, a set of 400 G2P-SOL project (http://www.g2p-sol.es/) accessions maintained at Universitat Politècnica de València (Valencia, Spain) were included in the study. They comprise 361 accessions of *S. lycopersicum*, 20 of *S. pimpinellifolium*, the closest wild ancestor of the cultivated tomato (Blanca et al., 2015), which has repeatedly served as a source of valuable traits for its improvement (Caicedo and Schaal, 2004), and 19 accessions of six other wild relatives. The latter include three species belonging to the "Arcanum" group (i.e., *S. arcanum*, *S. chmielewskii*, and *S. neorickii*), two to the "Eriopersicon" group (i.e., *S. huaylasense* and *S. habrochaites*) and *S. pennellii* belonging to the monotypic "Neolycopersicon" group (**Supplementary Table S2**). Two DNA samples of the tomato inbred line "Heinz 1706" (CTR_H1_122 and CTR_H1_123) were used as controls.

For eggplant, a set of 422 accessions from the G2P-SOL project were included in the study. According to passport data, the accessions, maintained at Universitat Politècnica de València (Spain) and at the CREA-GB (Montanaso Lombardo, Italy), comprise 362 accessions of *S. melongena* of the Occidental and Oriental groups (Vilanova et al., 2012), 36 and 9 accessions, respectively, of the cultivated *S. aethiopicum* and *S. macrocarpon*, as well as 15 accessions belonging to seven wild

relatives (**Supplementary Table S3**). The latter include four species from the Old World, of which two (*S. incanum* and *S. linnaeanum*), together with *S. melongena*, are part of the "Eggplant" clade, and two other (*S. anguivi* L. and *S. tomentosum* L.), which together with cultivated *S. aethiopicum* and *S. macrocarpon*, are part of the "Anguivi" grade, while three are from American origin (*S. paniculatum* L., *S. sisymbriifolium*, and *S. torvum*) (Miz et al., 2008; Syfert et al., 2016). As control, three DNA samples of the eggplant inbred line "67/3" (i.e., GPE001970, GPE001970b, and control 31) were also genotyped.

DNA extraction, library construction preparation, and sequencing

DNA was extracted using the Qiagen plant mini-prep, the LGC Sbeadex kit or by a modified CTAB method. Libraries were prepared according to the Ovation Rapid Library Systems (Nugen) specifications. The streamlined workflow consists of six main steps, as shown in **Figure 1**.

For the pilot test, sequencing was performed with the Illumina HiSeq 2500 platform (Illumina Inc., San Diego, CA, United States), following the manufacturer protocol and using 75SE chemistry. For the genotyping of the whole set of accessions with the custom 5K probe sets, sequencing was performed with Illumina HiSeq 2500 platform (Illumina Inc., San Diego, CA, United States), following the manufacturer protocol and using 150SE chemistry. The sequencing raw data are available at NCBI SRA (BioProject ID PRJNA542237 for tomato data and BioProject ID: PRJNA542231 for eggplant data).

SPET sequencing and SNP calling

Base calling and demultiplexing were carried out using the standard Illumina pipeline. The read quality check and adapter trimming was carried out using ERNE (Del Fabbro et al., 2013) and Cutadapt (Martin, 2011) software. After alignment to the reference eggplant and tomato genomes, using BWAMEM (Li, 2013) with default parameters, the uniquely aligned reads were selected (i.e., reads with a mapping quality >10). SNP calling was obtained with GATK 4.0 (DePristo et al., 2011), following the software best practices in June 2018 for germline short variant discovery(https://software.broadinstitute.org/gatk/bestpractices/workflow?id=11145). Main steps of the analysis were: (i) per-sample variants calling on target regions using HaplotypeCaller (Poplin et al., 2017), resulting in GVCFs file for each sample; (ii) GVCFs consolidation across multiple samples, in order to improve scalability and speed up the following step using ImportGenomicsDB; (iii) joint genotyping based on GenotypeGVCFs to produce a set of joint-called variants; and (iv) selection of SNPs (using SelectVariants) and quality filtering of SNPs using

VariantFiltration (filter expression used: QD < 2.0 || MQ < 40.0 || MQRankSum < -12.5).

To extract high confidence SNPs, Vcftools (Danecek et al., 2011) was applied to both the eggplant and tomato generated VCFs, using the following parameters: min-meanDP 30, max-missing 0.95 (0.80 for tomato) and non-ref-ac-any 1.

Genetic relationships analysis

The polymorphic information content (PIC) of each SNP was evaluated by applying the following equation, as suggested by Anderson et al. (1993): $PIC = 1 - \sum Pij^2$, where Pij represented the frequency of the jth allele at the ith SNP and the summation was extended over n alleles.

Genetic relationships were described by constructing a phylogenetic tree by maximum likelihood (ML) method using the IQ-TREE software (Nguyen et al., 2015); data are available in **Supplementary File S2**. Branch supports were obtained with the ultrafast bootstrap (Hoang et al., 2018). Comparison between ML trees was assessed using the Robinson-Foulds distance (Robinson and Foulds, 1981) calculated with ETE 3 (Huerta-Cepas et al., 2016). A principal component analysis (PCA) was obtained with SNPrelate (Zheng et al., 2012) program. Analyses were performed using target SNPs only or using all (target plus non-target) SNPs.

RESULTS

SPET assay design and robustness

In tomato, 344,373 eligible SNPs were identified within *S. lycopersicum* (including *S. lycopersicum* var. *cerasiforme*) and *S. pimpinellifolium* using the data from the "150 Tomato Genome Resequencing Project" and the "AGIS Tomato 360 Resequencing Project", of which 14,566 sites were selected for probe design. Of these, 40% were localized in coding regions and the rest in introns or UTRs. In eggplant, 72,739 eligible SNPs were identified, using a resequencing panel including eight *S. melongena* and one *S. incanum* accessions (Gramazio et al., 2019), of which 11,928 were selected for probe design, being 60% localized in coding regions and the rest in introns or UTRs.

A pilot study was run on 24 genotypes for each species using 75-bases long sequencing, of which 40 bp corresponded to the probe sequence. After sequencing, 13,615 (93%) and 11,265 (94%) probes were found to target SNPs within the first 25 bp after the probe 3' end in tomato and eggplant, respectively.

Based on sequencing coverage and missing data (see section "Materials and Methods"), a final set of 5,000 probes for tomato (of which 2,254 in CDSs and 2,746 in Introns/UTRs), and 5,082 for eggplant (of which 3,619 in CDSs and 1,463 in Introns/UTRs) were chosen (**Supplementary File S1**). The probes localized to the gene-rich chromosome arms in both species (**Figure 2**), and their sequence is reported in **Supplementary File S3**.

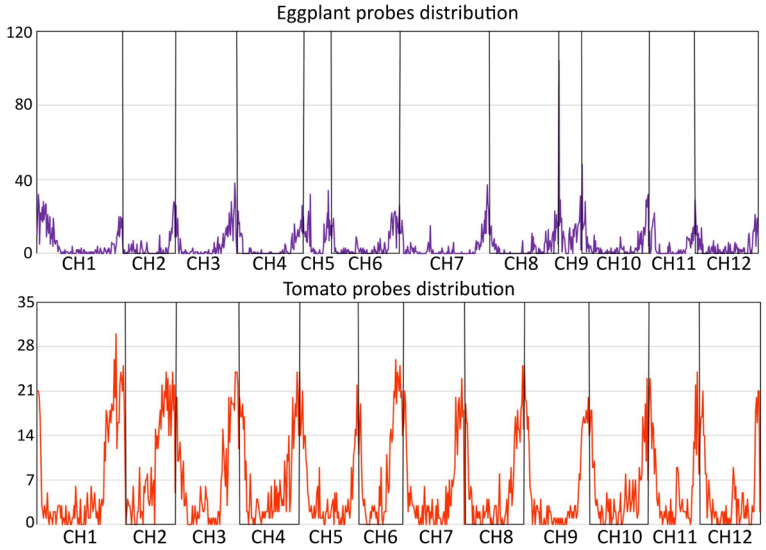


Figure 2. Eggplant and tomato distribution of the 5k SNPs panels according to their genomic position.

To assess the robustness and transferability of the SPET method, the numbers of reads obtained with the final probe panels and their mapping efficiency were assessed using different DNA preparations (**Supplementary Table S4**). As can be seen, both the number of reads and the mapping percentage were relatively stable when DNAs prepared with different DNA extraction methods and by different laboratories were used.

SPET diversity assessment of the tomato germplasm set

Using the final panel, around 215 million 150-bp single reads were produced in tomato, yielding 110 bp of useful sequence after probe trimming. After quality filtering, 198 high-quality million reads were retained (8% discarded) for the alignment to the "Heinz 1706" tomato genome sequence version 2.5 (The Tomato Genome Consortium, 2012) with an average mapping rate of 92.2%. Three out of 400 accessions gave an average read depth <10. By applying stringent criteria, 12,002 SNPs were identified among the 397 accessions included in the study. Of these, 4,577 SNPs were those originally targeted by the 5k probes set, while the remaining 7,425 were accessory non-target SNPs. By using the whole set of identified SNPs, the PIC ranged from 0.002 to 0.539 with an average of 0.094 (**Supplementary Table S5**), while by considering the target SNP panel, and the PIC average raised to 0.147 (**Supplementary Table S6**).

The 358 *S. lycopersicum* accessions with a read depth >10 showed very low levels of missing data (1.2%), high identity with the Heinz 1706 reference sequence (96.5%) and a low level of heterozygosity (0.65%), compatible with the autogamous reproduction of cultivated tomato (**Supplementary Table S2**). The missing data were slightly higher for the wild species, due to the sequence polymorphisms underlying the probes, ranging from 1.7% in *S. pimpinellifolium* to 4.2% in *S. neorickii*. Only *S. habrochaites* displayed high missing data (average of 10.4%). Conversely, the identity with the Heinz 1706 reference sequence was lower in wild species, ranging on average from 66.0% in *S. pimpinellifolium* to 43.2% in *S. huaylasense*. The heterozygosity level of wild species was higher than *S. lycopersicum*, ranging from 4.6% in *S. pimpinellifolium* to 27.5% in *S. pennellii* on average, consistent with the partial or total allogamy reported for these species (Chen and Tanksley, 2004). As expected, in the two replicates of the inbred line "Heinz 1706," more than 99.8% of SNPs showed the same allele of the reference sequence and just 18 and 19 sites, respectively, had alternative/heterozygous SNPs.

SPET diversity assessment of the eggplant germplasm set

In eggplant, more than 252 million single reads were produced. Sequences were trimmed and quality filtered to 242 million useful reads (4% discarded), corresponding to about 600K reads per sample on average. The latter were then aligned to the recently produced reference "67/3" eggplant genome (Barchi et al., 2019) with an average mapping rate of 95.8%. Three out of 422 accessions gave an average read depth <10. By applying stringent criteria, a total of 30,731 polymorphic sites were identified among the 422 accessions included in the study. Among them, 4,628 were SNPs targeted by the 5k probes set, while the remaining 26,103 were accessory non-target SNPs. By using the whole set of identified SNPs the PIC ranged from 0.002 to 0.607 with an average of 0.105 (**Supplementary Table S7**), while by considering the target SNP panel, the PIC ranged from 0.002 to 0.607 with an average of 0.381 (**Supplementary Table S8**).

The 360 *S. melongena* accessions with a read depth >10 showed, on average, extremely low levels of missing data (0.02%), high identity with the "67/3" reference sequence (93.6%) and a low level of heterozygosity (0.67%), compatible with the autogamous reproduction of cultivated eggplant (**Supplementary Table S3**). As for tomato, the missing data were slightly higher for the wild species, ranging from 0.5% in *S. incanum* to 4.5% in *S. macrocarpon*. Only the distantly related species *S. torvum* and *S. sisymbriifolium* displayed high missing data (21.2 and 22.7%, respectively). The identity with the "67/3" reference genome was lower in the eggplant relatives, ranging from 85.9% in *S. incanum* to 50.6% in *S. sisymbriifolium*. The heterozygosity level of wild species and cultivated relatives *S. aethiopicum* and *S. macrocarpon* was higher than that of *S. melongena*, ranging from 1.7% in *S. tomentosum* and *S. aethiopicum* to 9.4% in *S. sisymbriifolium*, consistent with the partial allogamy reported or suggested for these species (Daunay et al., 2001; Vorontsova and Knapp, 2016; Acquadro et al., 2017). As expected, in the three replicates of the eggplant inbred line "67/3" used as a control, over 99.9% of SNPs showed the same allele of the reference genome, and just from 26 to 35 sites showed alternative/heterozygous SNPs.

Genetic relationships in the tomato and eggplant germplasm set

The maximum likelihood (ML) dendrograms (**Figures 3**, **4**) show the genetic relationships between the tomato accessions in the study. As expected, the replicated samples of the same reference genotypes (CTR_H1_122 and CTR_H1_123) clustered together. In the dendrogram obtained using the whole set of SNPs (**Figure 3**), the accessions of the "Eulycopersicon" group (i.e., *S. lycopersicum* and *S. pimpinellifolium*) cluster together in a main branch, in which *S. pimpinellifolium* accessions are basal to the monotypic *S. lycopersicum* cluster. The *S. pimpinellifolium* accessions do not intermingle with the ones of *S. lycopersicum*, and they are divided in two branches, one which contains only Peruvian accessions,

and another one basal to all *S. lycopersicum* that contains all accessions from Ecuador and three accessions from Peru. In the dendrogram, a second branch includes the other wild species and two main sub-clusters, of which one contains the three "Arcanum" group accessions (*S. arcanum*, *S. chmielewskii*, and *S. neorickii*), with *S. arcanum* basal to the two other species, and the other accessions of *S. huaylasense* and *S. habrochaites* from the "Eriopersicon" group, and *S. pennelli* from the "Neolycopersicon" group, with *S. huaylasense* basal to the two other species.

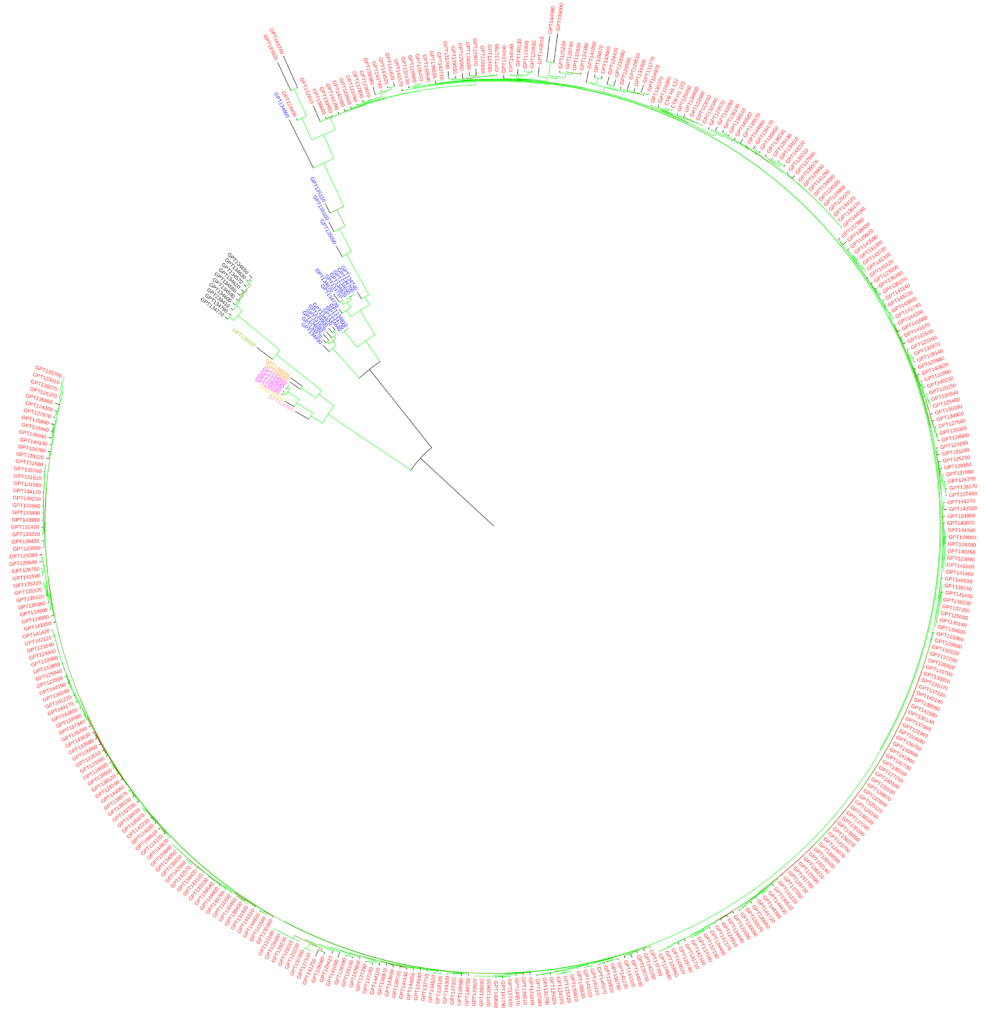


Figure 3. Maximum likelihood phylogenetic tree obtained with IQ-TREE, based on the whole set of SNPs, illustrating the genetic architecture of tomato and wild related species accessions in study. *Solanum lycopersicum* entries are in red, *S. pimpinellifolium* in blue, *S. chmielewskii* in yellow, *S. neorickii* in purple, *S. arcanum* in pink, *S. pennellii* in green, *S. habrochaites* in dark and *S. huaylasense* in orange. Branches are colored according to the boostrap values (red = 15 and green = 100)

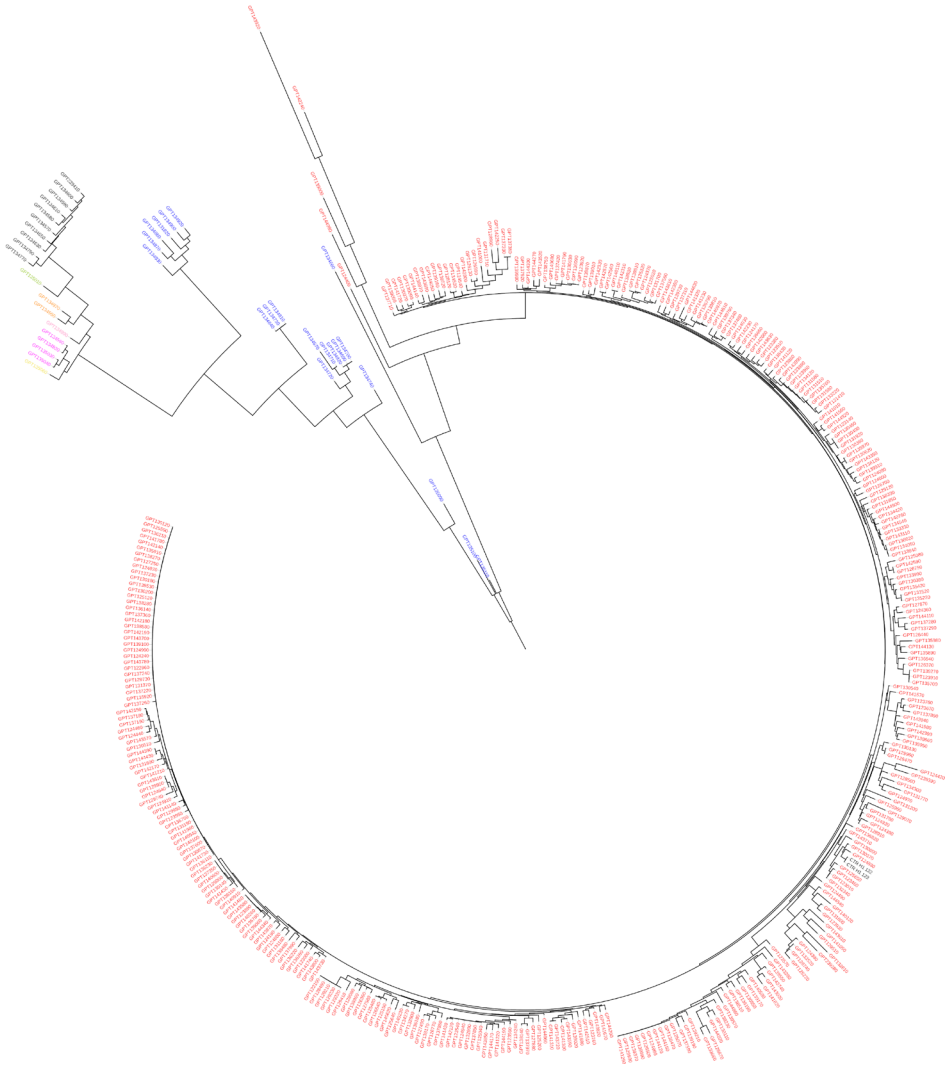


Figure 4. Maximum likelihood phylogenetic tree obtained with IQ-TREE, based on the target SNPs, illustrating the genetic architecture of tomato and wild related species accessions in study. *Solanum lycopersicum* entries are in red, *S. pimpinellifolium* in blue, *S. chmielewskii* in yellow, *S. neorickii* in purple, *S. arcanum* in pink, *S. pennellii* in green, *S. habrochaites* in dark and *S. huaylasense* in orange. Branches are colored according to the boostrap values (red = 15 and green = 100)

Some of the main findings obtained using the whole set of SNPs were confirmed using target SNPs (**Figure 4**), as in both cases the different species are not intermingled. Notwithstanding, the two dendrograms present some differences in terms of topology and branch length, confirmed by the Normalized Robinson-Foulds

distance being 0.63. In the dendrogram based on target SNPs, *S. pimpinellifolium* is basal to the rest of accessions which are split in the two major clusters, one containing the *S. lycopersicum* accessions and one accession of *S. pimpinellifolium*, and the other containing the rest of wild species. In this latter, the major branch *S. pimpinellifolium* is spread in different branches which are basal to the rest of wild species, which largely display the same topology than in the dendrogram obtained with all the SNPs (**Figure 3**).

PCA analysis for the tomato set, based on the whole set of SNPs (**Figure 5a**) confirmed the grouping of the ML dendrogram. The first and second components accounted for 21.7 and 13.6% of the genetic variation. *S. lycopersicum* and *S. pimpinellifolium* were clearly separated from the other species and all in all well differentiated with a minor degree of overlap. The other wild species are not intermingled and cluster in a different area of the PCA plot (**Figure 5a**). The three species of the "Arcanum" group (*S. arcanum*, *S. chmielewskii*, and *S. neorickii*) are the closest to the "Eulycopersicon" (*S. lycopersicum* and *S. pimpinellifolium*), the species *S. habrochaites* ("Eriopersicon"), and *S. pennellii* ("Neolycopersicon") are the most genetically distant, while *S. huaylasense* accessions occupy an intermediate position (**Figure 5a**). The first and second axes of the PCA using only target SNPs account for 38.3 and 7.1%, of the genetic variation, respectively (**Figure 5b**), and largely confirm the results of the PCA based on all SNPs. However, *S. lycopersicum* and *S. pimpinellifolium* display a greater dispersion as highlighted in **Figure 5b**.

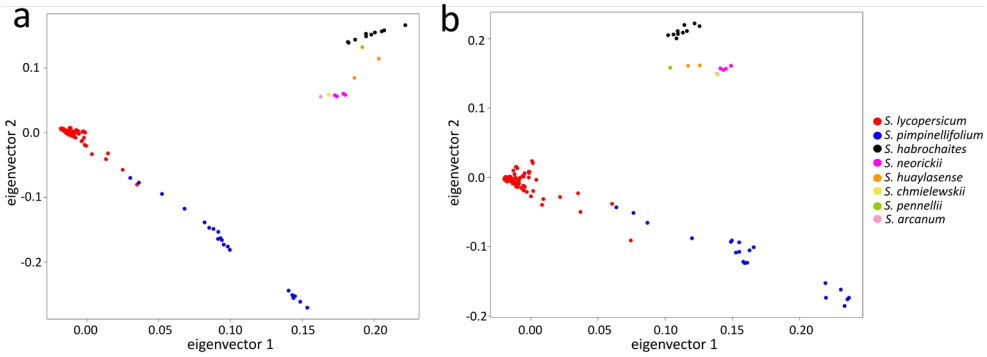


Figure 5. PCA visualization of the genetic relationships among the accessions of tomato and wild related species in study, based on the whole (a) or target (b) SNP Datasets.

In eggplant, the ML-based dendrograms (**Figures 6**, **7**) display the genetic relationships between the accessions in study. As expected, the three replicated samples of the reference genotype cluster together. The dendrogram obtained using the whole SNPs panel identifies two main branches, of which one includes the two accessions from American species (*S. sisymbriifolium* and *S. torvum*), and the

other includes the rest of species native to the Old World (**Figure 6**) together with the two accessions labeled as *S. paniculatum*. In this latter cluster, two major branches are distinguishable: one containing the four species of the "Anguivi" grade (*S. aethiopicum*, *S. anguivi*, *S. macrocarpon*, and *S. tomentosum*), and one including the three species of the "Eggplant" clade (*S. melongena*, *S. incanum* and *S. linnaeanum*), plus the two *S. paniculatum* accessions. In the cluster of the "Anguivi" grade, the four species are separated in different sub-branches, except *S. aethiopicum* and *S. anguivi*, which are intermingled. In the other cluster, a branch contains the "Eggplant" clade cluster, in which all the accessions of *S. linnaeanum* are separated in a sub-branch and *S. melongena* and *S. incanum* accessions in another sub-branch, with the latter basal to the monotypic eggplant cluster. The two accessions labeled as *S. paniculatum* are basal to the "Eggplant" clade branch. Three *S. melongena* accessions initially mis-labeled as *S. aethiopicum* in the germplasm bank list clustered correctly with the rest of *S. melongena* accessions.

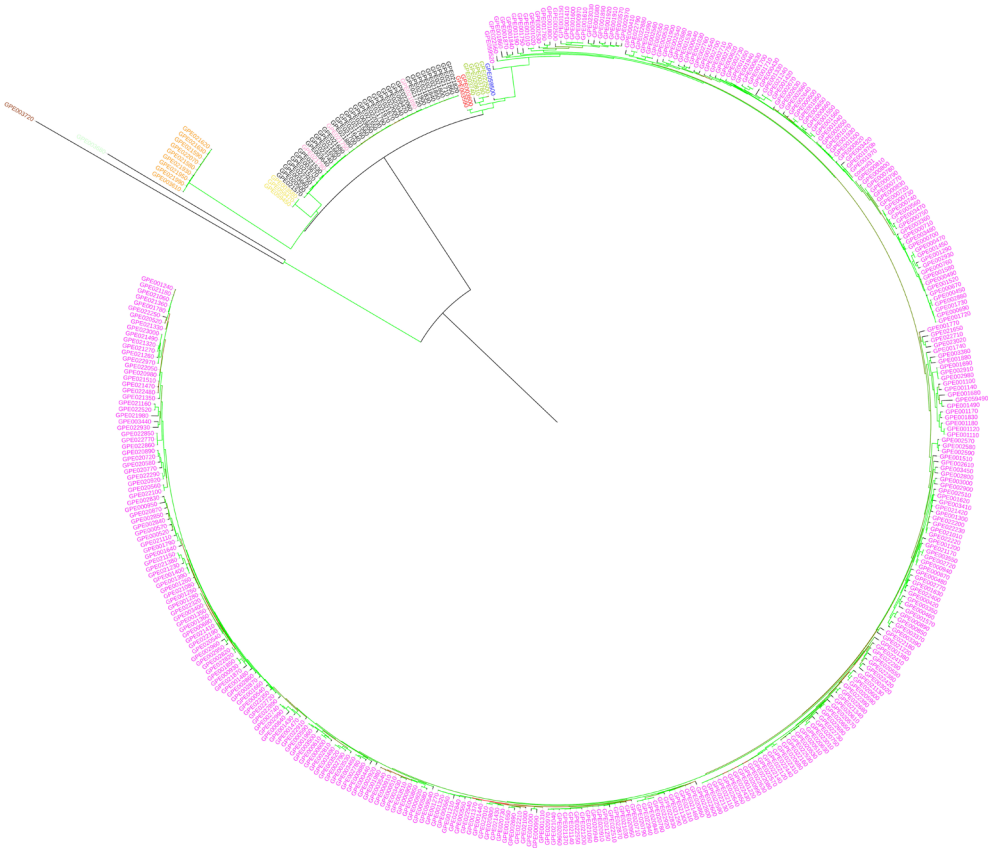


Figure 6. Maximum likelihood phylogenetic tree obtained with IQ-TREE, based on the whole set of SNPs, illustrating the genetic architecture of eggplant and wild related species in study. *Solanum melongena* entries are in purple, *S. aethiopicum* in black, *S. macrocarpon* in orange,

S. anguivi in pink, *S. incanum* in blue, *S. linnaeanum* in dark green, *S. paniculatum* (mislabeled in the germplasm collection; actually *Solanum* sp.) in red, *S. sisymbriifolium* in brown, *S. tomentosum* in yellow, and *S. torvum* in light green. Branches are colored according to the bootstrap values (red = 25 and green = 100)

Figure 7. Maximum likelihood phylogenetic tree obtained with IQ-TREE, based on the target SNPs, illustrating the genetic architecture of eggplant and wild related species in study. *Solanum melongena* entries are in purple, *S. aethiopicum* in black, *S. macrocarpon* in orange, *S. anguivi* in pink, *S. incanum* in blue, *S. linnaeanum* in dark green, *S. paniculatum* (mislabeled in the germplasm collection; actually *Solanum* sp.) in red, *S. sisymbriifolium* in brown, *S. tomentosum* in yellow, and *S. torvum* in light green. Branches are colored according to the bootstrap values (red = 25 and green = 100)

The eggplant dendrogram based on target SNPs (**Figure 7**) confirms a good separation among species and the intermingling in the same cluster of *S. aethiopicum* and *S. anguivi*. Discrepancies, confirmed by the Normalized Robinson-Foulds distance equal to 0.53, were observed in the topology and branch lengths compared to the dendrogram obtained with the whole SNPs dataset. The dendrogram based on target SNPs identifies two main branches, one of which includes only *S. melongena* accessions, while the other contains the rest of *S. melongena* accessions, and in a sub-cluster the other cultivated and wild species. In this latter, the two American species *S. sisymbriifolium* and *S. torvum* cluster together and are basal to all the others (**Figure 7**), while, unexpectedly, the *S. linnaeanum* accession GPE003740 (for which a low number of reads was obtained) shows an odd position, and it appears genetically differentiated (basal) from the others.

The PCA analysis based on the whole eggplant SNPs dataset (**Figure 8a**), largely confirms the grouping of genotypes obtained in the ML-based dendrogram. The first and second principal axes account, respectively, for 13.5 and 11.5% of the genetic variation. In the PCA graph, the two American species *S. sisymbriifolium* and *S. torvum* are clearly separated from the rest of entries. The species of the "Eggplant" clade and "Anguivi" grade are clearly separated by the first component of the PCA, with two exceptions: the *S. linnaeanum* accession GPE003740 as well as the entries of *S. aethiopicum* and *S. anguivi* clustering together. The closest species to *S. melongena* are the other "Eggplant" clade species *S. incanum* and *S. linnaeanum*, as well as the two accessions labeled as *S. paniculatum*, followed by the "Anguivi" grade *S. aethiopicum* and *S. anguivi*, the entries of *S. macrocarpon* and the ones of *S. tomentosum*.

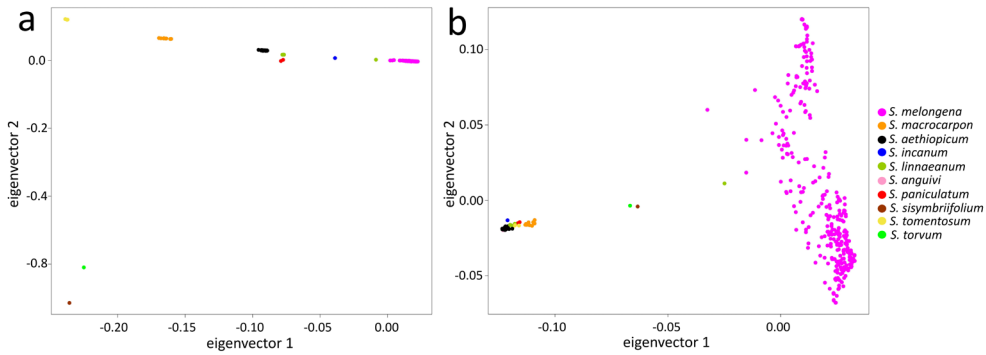


Figure 8. PCA visualization of the genetic relationships among the accessions of eggplant as well as cultivated and wild related species in study, based on the whole **(a)** or target **(b)** SNP Datasets.

The first and second components of the eggplant PCA based on target SNPs account, respectively, for 29.1 and 11.3% of the genetic variation. As observed with the ML-

based dendrogram, the PCA displays considerable differences with the one obtained with the whole SNPs dataset (**Figure 8b**). Indeed, the first component separates *S. melongena* and, surprisingly, *S. linnaeanum* accession GPE003740 from the other species. *Solanum melongena* is spread over a large area of the PCA graph with a wide range of values for both the components. The rest of species of the "Eggplant" clade and the "Anguivi" grade cluster together and in some cases are overlapped. Among this group of species, the closest to *S. melongena* are *S. macrocarpon*, followed *S. tomentosum*, *S. linnaeanum*, *S. incanum*, and finally *S. aethiopicum* and *S. anguivi*. Surprisingly, the American species *S. sisymbriifolium* and *S. torvum* plot in an intermediate area between *S. melongena* and the rest of Old World species (**Figure 8b**).

DISCUSSION

The single primer enrichment technology, recently developed by Nugen, has been used up to now for biomedical applications (Scolnick et al., 2015; Nairismägi et al., 2016), and in plants, for genotyping in monocot (*Z. mays*) lines and in a natural black poplar (*P. nigra*) population (Scaglione et al., 2019). Despite its potential interest, no information is available on its performance for characterizing large germplasm sets from crop plants. Our main goal was to assess the reliability and efficiency of the SPET technique for high-throughput genotyping of a large (822) set of accessions of tomato, eggplant and their cultivated and wild relatives. For this purpose, we evaluated the robustness of the technique and the diversity, heterozygosity and genetic relationships within the germplasm included in this study, comparing with the data reported in the literature.

Single primer enrichment technology is a robust method, performing well with DNA samples prepared by different laboratories using different DNA mini-preparation protocols, which is a prerequisite for large multicenter, and collaborative studies on plant genetic resources. Indeed, based on tomato samples extracted with different protocols, both the number of reads and the mapping percentage were relatively stable. Additionally, a very low level of missing data (1.18% for tomato, and 0.02% for eggplant) was observed when genotyping an intra-specific diversity panel, and a still acceptable (<5%) level was observed when genotyping species such as *S. pennellii* or *S. macrocarpon* which show 2.7 and 3.4 million years of divergence from tomato and eggplant, respectively (Kamenetzky et al., 2010; Särkinen et al., 2013).

Single primer enrichment technology combines in a single approach both targeted analysis of SNPs, thus being comparable with genotyping arrays, and complexity reduction typical of GBS approaches (Scheben et al., 2017). Furthermore, SPET provides the ability of multiplexing thousands of samples in a single sequencing run, which can be genotyped with tens of thousands of probes, and with a good

coverage at target sites. Finally, thanks to the sequencing of the genomic regions around the target SNPs, SPET allows the discovery of thousands of novel SNPs not originally included in the panel.

Compared to other crop species, both tomato and eggplant are known to have experienced a dramatic reduction of the genetic variability due to anthropogenic selection (Williams and St Clair, 1993; Cericola et al., 2013; Flint-Garcia, 2013), and hence present a lower frequency of SNPs than their wild species. The tomato and eggplant panels designed for this work, which originally targeted 5k SNPs for each species, allowed the discovery of 7,427 and 26,103 additional non-target SNPs, respectively, in the tomato and eggplant sets. Of these, 2,224 and 3,292 were detected only in *S. lycopersicum* and *S. melongena*, respectively, while 7,130 and 24,892 were found to be shared in the remaining species, belonging to the tomato and eggplant genepools, respectively. This indicates that the technique enables the discovery of high numbers of novel polymorphisms, even in gene pools that came across severe bottlenecks during domestication, migration and selection.

Most of the accessions of the largely autogamous *S. lycopersicum* and *S. melongena* (Chen et al., 2007; Daunay and Hazra, 2012; Acquadro et al., 2017) showed a low heterozygosity (on average 0.65 and 0.67%, respectively), in agreement with previous reports (Sim et al., 2012; Vilanova et al., 2012; Aflitos et al., 2014; Acquadro et al., 2017). Only a few accessions of both crops displayed higher values, up to 8.7% for tomato and 7.1% for eggplant, which probably reflect recent events of hybridization either before collection or during germplasm multiplication. Thus, the technique allows the identification of segregating accessions and may guide sampling during seed multiplication. For the tomato germplasm set, the accessions of *S. pimpinellifolium*, the closest wild relative of *S. lycopersicum*, exhibited on average an intermediate (4.7%) level of heterozygosity between *S. lycopersicum* and the other wild species, some of which are self-incompatible (Chen and Tanksley, 2004). In the two additional cultivated species of eggplant, i.e., *S. aethiopicum* and *S. macrocarpon*, a higher heterozygosity was detected, being on average, of 1.9 and 2.5%, respectively. This is attributable to the more limited anthropogenic selection on these species and to their higher rate of allogamy (Daunay et al., 2001). These values are slightly lower than the ones previously reported (Acquadro et al., 2017), based on the RAD-sequencing technique which provides a randomized representation of the genome.

Genetic relationships among the accessions in the study were explored by constructing ML phylogenetic trees and PCA analyses, based on both the whole set or just the target SNPs. The stringent criteria adopted to select the set of SNPs consistently reduced the frequency of missing data. A total of 358 of 361 accessions of *S. lycopersicum* displayed on average 1,18% of missing data and their frequency

was up to 3.3% in the closely related species *S. pimpinellifolium*. Higher values were detected in three tomato accessions (GPT141120, GPT130370 and GPT124880) due to the low number of reads obtained following sequencing. However, this did not affect their clustering with the other accessions of *S. lycopersicum* when ML tree and PCA were based on both the whole SNPs dataset or only the target SNPs. Indeed, several studies have explored how missing data may impact phylogenetic analyses using both empirical and simulated data, and suggest that it is possible to include taxa that have large amounts of missing data without ill effects (Wiens and Moens, 2008; Lynch and Wagner, 2010; Thomson and Shaffer, 2010; Wiens and Morrill, 2011). In the remaining accessions, the missing data ranged from 3.0 to 17.7% and reached the highest values in *S. habrochaites*, the most evolutionary divergent wild species among those evaluated (Aflitos et al., 2014; Beddows et al., 2017).

In almost all accessions of *S. melongena*, and the ones of the cultivated *S. aethiopicum*, very low levels (0.02 and 0.7%, respectively) of missing data were detected; in the other cultivated eggplant (*S. macrocarpon*) their frequency was on average 4.5%. The missing data varied in entries of the wild species and reached the highest values in the New World native species *S. torvum* (21.2%) and *S. sisymbriifolium* (22.7%), characterized by greater evolutionary divergence (Vorontsova et al., 2013; Acquadro et al., 2017). In two accessions of *S. melongena* and one of *S. linnaeanum*, an unexpected high frequency of missing data was observed, which was attributable to the low number of reads obtained. As observed for tomato, the two *S. melongena* accessions always clustered with the ones of the other accessions of this species, while the *S. linnaeanum* accession clustered with the others of the same species when the whole SNP dataset was used, but separately when only target SNPs were used. This suggests that, when a limited number of reads is obtained, the analysis based on the whole set of SNPs is less prone to misclassification of highly diverse genotypes.

Single primer enrichment technology genotyping with the whole set of SNPs or just the target SNPs proved to be a powerful tool for the identification of duplicates and mislabeled accessions. The two replicates of the *S. lycopercicum* accession "Heinz 1706," as well as the three replicates of the *S. melongena* breeding line "67/3" clustered together, with SNP polymorphism ranging from 0.1 to less than 0.2%. Furthermore, the two accessions initially labeled as *S. paniculatum*, a close relative of *S. torvum* (Miz et al., 2008), did not cluster with the latter nor with the other New World species *S. sisymbriifolium*. Instead, the SPET genotyping data suggested that these two accessions labeled as *S. paniculatum* correspond to an undetermined species (*Solanum* sp.) closely related to the eggplant clade or to the closely related "Anguivi" grade (Vorontsova et al., 2013). Closer passport and phenotypic inspection confirmed the taxonomic mislabeling of these two accessions.

Additionally, three *S. melongena* accessions initially mislabeled as *S. aethiopicum* in the germplasm bank listing were correctly clustered with the rest of *S. melongena* accessions, demonstrating that SPET is also a powerful technique (Mason et al., 2015) for detecting misclassified accessions for which no or few characterization data are available.

Tomato and eggplant diversity

Our results of the tomato and eggplant genotyping with the SPET platform are largely congruent with previous results on the knowledge of diversity in these groups (Lester and Thitai, 1989; Peralta et al., 2008; Rodriguez et al., 2009; Aflitos et al., 2014; Dodsworth et al., 2016; Acquadro et al., 2017; Gramazio et al., 2017). On the basis of both ML dendrogram and PCA analysis with all SNPs, the *S. lycopersicum* accessions clustered separately from all the other species and grouped in a single branch, revealing a low diversity. *S. pimpinellifolium*, native to Peru and Ecuador (Grandillo et al., 2011), is the only red-fruited wild species and it is the nearest wild relative to the cultivated tomato. In agreement with previous findings (Rodriguez et al., 2009; Aflitos et al., 2014; Dodsworth et al., 2016), its accessions were found to cluster close to *S. lycopersicum* in the ML dendrogram. It has been reported that the *S. pimpinellifolium* accessions are divided in three main genetic groups, corresponding to the environmental differences found in the coastal regions of Northern Ecuador, in the mountain region of Southern Ecuador and Northern Peru, and the coastal region of Peru (Blanca et al., 2015). Due to the smaller number of accessions of this species under study, we were not able to identify these three groups, however, the ML dendrogram detected two *S. pimpinellifolium* sub-clusters which partially correlate with the Ecuadorian and Peruvian origin of the entries. On the basis of PCA analysis, some *S. pimpinellifolium* and *S. lycopersicum* accessions were intermingled. This is in agreement with previous findings that the genome of the two species shows only 0.6% nucleotide divergence and signs of recent admixture (The Tomato Genome Consortium, 2012). Furthermore, due to the absence of crossing barriers between the two species, introgression events from *S. pimpinellifolium* into *S. lycopersicum* have probably occurred throughout the history of tomato domestication (The Tomato Genome Consortium, 2012).

The position of the rest of wild species in the dendrogram and PCA analyses with all SNPs is in agreement with previous studies (Rodriguez et al., 2009; Aflitos et al., 2014; Dodsworth et al., 2016; Beddows et al., 2017), i.e., “Arcanum” group species were found to be closer to the “Lycopersicon” group than the “Eriopersicon” and “Neolycopersicon”, and the two sister species *S. chmielewskii* and *S. neorickii* clustered together (Dodsworth et al., 2016).

Solanum melongena accessions of both Oriental and European origin, producing fruits of different shape and color, clustered separately from all the other cultivated and wild species in both the ML dendrogram and PCA obtained by using the whole SNP dataset. The remaining species fell in the dendrogram and PCA plots according to their expected positions (Acquadro et al., 2017; Gramazio et al., 2017). Among the eggplant relatives, our data showed that *S. incanum* was the closest species to eggplant, followed by the other "Eggplant" clade species (*S. linnaeanum*), as reported by other studies (Lester and Hasan, 1991; Weese and Bohs, 2010).

Contrasting results (Sakata et al., 1991; Sakata and Lester, 1997; Isshiki et al., 2008; Meyer et al., 2012; Vorontsova et al., 2013) have been reported on the relationships between *S. melongena* and the other two cultivated eggplants: *S. aethiopicum* (scarlet eggplant) and *S. macrocarpon* (gboma eggplant). In recent studies, it has been reported that the latter is genetically closer to eggplant (Acquadro et al., 2017). However, by applying SSR or SNP genotyping, different results were obtained (Gramazio et al., 2017). Our SPET-based clustering highlighted that the three cultivated species belong to clearly separate groups that probably diverged at similar times. As expected (Lester and Niakan, 1986; Lester and Hasan, 1991; Syfert et al., 2016), the accessions of *S. aethiopicum* resulted intermingled with those of its wild ancestor *S. anguivi*, indicating a probable genetic flux between the two species (Plazas et al., 2014). Also, *S. tomentosum* clustered in the "Anguivi" grade, confirming previous results (Vorontsova et al., 2013; Acquadro et al., 2017). As previously reported (Acquadro et al., 2017), the wild *S. sisymbriifolium* and *S. torvum*, native of South and Central America and representing sources of resistance to several diseases (Daunay and Hazra, 2012) formed a separate group following both the ML tree and PCA analyses.

In both the tomato and eggplant genepools, the ML dendrograms and PCA analyses performed on total SNP datasets provided a clustering similar to the one obtained on the basis of the target SNPs. However, some differences were evident: target SNPs provided a higher intra-specific discrimination within tomato and eggplant, accompanied by a less clear clustering of the other cultivated and wild species. The higher discrimination of different eggplant accession was particularly evident in the PCA analysis (**Figure 8**), and their genetic differentiation was maximized in the ML dendrogram as well (**Figure 7**). On the other hand, the use of target SNPs decreased the discrimination between the cultivated *S. aethiopicum* and *S. macrocarpon*, and the wild species belonging to *S. melongena* genepool clustered into a single branch of the ML tree, showing a strong clustering in the PCA graph.

The differences observed when using the target vs. whole SNP datasets are attributable to the number and the genetic distance of the resequenced genotypes used for the initial identification of the polymorphic sites for SPET set up. In tomato,

the SPET panel was based on a majority of resequencing data from *S. lycopersicum* (including var. *cerasiforme*) and some *S. pimpinellifolium*, while the eggplant panel was based on resequencing of eight *S. melongena* and one *S. incanum* accessions. Thus, in *S. lycopersicum* a clear assessment of the genetic relationships among species of the tomato genepool was obtained when SPET was based on both the whole dataset and target SNPs, although the genetic differentiation within tomato accessions was to a certain extent increased by the use of target SNPs. Vice versa, in eggplant, the whole SNP dataset was more informative for phylogenetic studies, while the use of target SNPs allowed a huge increase in the detection of genetic variation within accessions of S. melongena, as required for GWA studies.

CONCLUSIONS

Recently, a targeted genotyping approach (SPET) was released, but very limited information of its performance for highthroughput genotyping in plants is presently available. We assessed the efficiency and robustness of this technique in analyzing the genetic diversity in hundreds of S. lycopersicum and S. melongena germplasm accessions and of their cultivated and wild relatives maintained in genebanks. A high number of both target and non-target polymorphisms were analyzed by ML dendrograms and PCA analyses and the two approaches were complementary in the interpretation of data.

Our results demonstrate that SPET represents a valid alternative to random complexity reduction methods and arrays, and it allows users to customize the panel of target markers and provides reliable fingerprinting of accessions maintained in genebanks. Our results also demonstrate the transferability of the panels to closely related species, and that the number and genetic distances of the resequenced genotypes used for identifying the target SNPs play a significant role. We designed panels enriched in intra-specific SNPs reasoning that additional inter-specific ones would be more likely to be discovered in the surrounding sequenced regions, which was indeed the case. The use of the whole SNP dataset is more appropriate for broad phylogenetic studies, while the use of target SNPs, boosting the intra-specific discrimination, for domestication and genome wide association studies. The SPET technology made it also possible to clearly separate the different species, confirmed established phylogenetic relationships among different species and taxonomic groups, resolved the mislabeling of entries, and demonstrated its high reproducibility when applied on replicates of the same accession, proving its usefulness for the high-throughput genotyping, management, and enhancement of genebank collections.

DATA AVAILABILITY

The datasets generated for this study can be found in the BioProject ID: PRJNA542237 for tomato data and BioProject ID: PRJNA542231 for eggplant data.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2019.01005/full#supplementary-material

Table S1 | All the resequenced accessions of tomato used for SNPs mining.
Table S2 | Tomato genepool accessions and SNPs metrics.
Table S3 | Eggplant genepool accessions and SNPs metrics.
Table S4 | Robustness of the final tomato SPET assay using DNAs prepared with different methods and by different laboratories.
Table S5 | PIC values of the whole SNPs identified in the tomato genepool.
Table S6 | PIC values of the target SNPs identified in the tomato genepool.
Table S7 | PIC values of the whole SNPs identified in the eggplant genepool.
Table S8 | PIC values of the target SNPs identified in the eggplant genepool.
File S1 | Sequencing yields, coverage analysis and filtering for probe selection in tomato and eggplant.
File S2 | Phylip formatted input files used for generating phylogenetic tree by maximum likelihood (ML).
File S3 | Sequences of tomato and eggplant probes used for SPET assay.

AUTHORS' CONTRIBUTIONS

GG, SL, JP, and GLR conceived and designed the research. LT, LaB, GLR, DA, PG, SV, MD, OD, PM, and PF provided the plant material and extracted the DNA samples. DS performed the SPET panel design. LoB, GA, EP, and AA analyzed the data. LoB, GG, JP, and SL wrote the manuscript. PG, SV, AA, EP, MD, GLR, and LT reviewed and edited the manuscript. All authors read and approved the manuscript

FUNDING

This work has been funded by the European Union's Horizon 2020 Research and Innovation Programme under the grant agreement number 677379 (G2P-SOL project: Linking genetic resources, genomes, and phenotypes of solanaceous crops).

ACKNOWLEDGMENTS

We thank Joe Don Heath and Stephanie Bruns (Nugen-Tecan) for the help in SPET panel design.

REFERENCES

Acquadro, A., Barchi, L., Gramazio, P., Portis, E., Vilanova, S., Comino, C., et al. (2017). Coding SNPs analysis highlights genetic relationships and evolution pattern in eggplant complexes. *PloS One* 12, e0180774.

Aflitos, S., Schijlen, E., de Jong, H., de Ridder, D., Smit, S., Finkers, R., et al. (2014). Exploring genetic variation in the tomato (*Solanum* section *Lycopersicon*) clade by whole-genome sequencing. *Plant J.* 80, 136–148.

Anderson, J. A., Churcill, G. A., Autrique, J. E., Tanksley, S. D., and Sorrels, M. E. (1993). Optimizing parental selection for genetic linkage maps. *Genome* 36, 181–186.

Barchi, L., Pietrella, M., Venturini, L., Minio, A., Toppino, L., Acquadro, A., et al. (2019). A chromosome-anchored eggplant genome sequence reveals key events in *Solanaceae* evolution. *Sci Rep.* 18:42098.

Beddows, I., Reddy, A., Kloesges, T., and Rose, L. E. (2017). Population genomics in wild tomatoes—The interplay of divergence and admixture. *Genome Biol. Evol.* 9, 3023–3038.

Blanca, J., Montero-Pau, J., Sauvage, C., Bauchet, G., Illa, E., Díez, M. J., et al. (2015). Genomic variation in tomato, from wild ancestors to contemporary breeding accessions. *BMC Genomics* 16, 257.

Boyd, J. W., Murray, D. S., and Tyrl, R. J. (1984). Silverleaf Nightshade, *Solanum elaeagnifolium,* Origin, Distribution, and Relation to Man. *Econ. Bot.* 38, 210–217.

Caicedo, A. L., and Schaal, B. A. (2004). Population structure and phylogeography of *Solanum pimpinellifolium* inferred from a nuclear gene. *Mol. Ecol.* 13, 1871–1882.

Castle, J. C. (2011). SNPs occur in regions with less genomic sequence conservation. *PLoS One* 6, e20660.

Cericola, F., Portis, E., Toppino, L., Barchi, L., Acciarri, N., Ciriaci, T., et al. (2013). The population structure and diversity of eggplant from Asia and the Mediterranean basin. *PLoS One* 8, e73702.

Chen, K.-Y., Cong, B., Wing, R., Vrebalov, J., and Tanksley, S. D. (2007). Changes in regulation of a transcription factor lead to autogamy in cultivated tomatoes. *Science* 318, 643–5.

Chen, K.-Y., and Tanksley, S. D. (2004). High-resolution mapping and functional analysis of *se2.1*: a major stigma exsertion quantitative trait locus associated with the evolution from allogamy to autogamy in the genus Lycopersicon. *Genetics* 168, 1563–73.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158.

Daunay, M. . C., and Hazra, P. (2012). "Eggplant," in *Handbook of Vegetables.*, ed. H. P. In: Peter KV (Houston: Springer, New York.), 257–322.

Daunay, M. C., Lester, R. N., and Ano, G. (2001). Cultivated eggplants. *Trop. Plant Breed.*, 200–225.

Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., and Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat. Rev. Genet.* 12, 499–510.

Del Fabbro, C., Scalabrin, S., Morgante, M., and Giorgi, F. M. (2013). An Extensive Evaluation of Read Trimming Effects on Illumina NGS Data Analysis. *PLoS One* 8, e85024.

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V, Maguire, J. R., Hartl, C., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491–498.

Dodsworth, S., Chase, M. W., Särkinen, T., Knapp, S., and Leitch, A. R. (2016). Using genomic

repeats for phylogenomics: a case study in wild tomatoes (*Solanum* section *Lycopersicon* : Solanaceae). *Biol. J. Linn. Soc.* 117, 96–105.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6, 1–10.

Flint-Garcia, S. A. (2013). Genetics and consequences of crop domestication. *J. Agric. Food Chem.*

Gramazio, P., Prohens, J., Borràs, D., Plazas, M., Herraiz, F. J., and Vilanova, S. (2017). Comparison of transcriptome-derived simple sequence repeat (SSR) and single nucleotide polymorphism (SNP) markers for genetic fingerprinting, diversity evaluation, and establishment of relationships in eggplants. *Euphytica* 213, 264.

Gramazio, P., Yan, H., Hasing, T., Vilanova, S., Prohens, J. and Bombarely, A. (2019). Whole-genome resequencing of seven eggplant (*Solanum melongena*) and one wild relative (*S. incanum*) accessions provides new insights and breeding tools for eggplant enhancement. *Front. Plant Sci*. 10:1220.

Grandillo, S., Chetelat, R., Knapp, S., Spooner, D., Peralta, I., Cammareri, M., et al. (2011). "From Solanum sect. Lycopersicon," in *Wild Crop Relatives: Genomic and Breeding Resources*, ed. C. Kole (Springer), 129–215.

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* 35, 518–522.

Hoheisel, J. D. (2006). Microarray technology: beyond transcript profiling and genotype analysis. *Nat. Rev. Genet.* 7, 200–210.

Huerta-Cepas, J., Serra, F., and Bork, P. (2016). ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Mol. Biol. Evol.* 33, 1635–1638.

Isshiki, S., Iwata, N., and Khan, M. M. R. (2008). ISSR variations in eggplant (Solanum melongena L.) and related Solanum species. *Sci. Hortic. (Amsterdam).* 117, 186–190.

Kamenetzky, L., Asís, R., Bassi, S., de Godoy, F., Bermúdez, L., Fernie, A. R., et al. (2010). Genomic analysis of wild tomato introgressions determining metabolism- and yield-associated traits. *Plant Physiol.* 152, 1772–86.

Kouassi, B., Prohens, J., Gramazio, P., Kouassi, A. B. A., Vilanova, S., Galán-Ávila, A., et al. (2016). Development of backcross generations and new interspecific hybrid combinations for introgression breeding in eggplant (Solanum melongena). *Sci. Hortic. (Amsterdam).* 213, 199–207.

Lester (1991). "Origin and domestication of the brinjal eggplant, Solanum melongena, from S. incanum, in Africa and Asia.," in *Solanaceae III: taxonomy, chemistry, evolution.*, ed. T. L. S. of London (London), 369–387.

Lester, R. N., and Niakan, L. (1986). "Origin and domestication of the scarlet eggplant, Solanum aethiopicum, from S. anguivi in Africa," in *Solanaceae: Biology and systematics.*, ed. W. G. (ed. . D'Arcy (Columbia University Press, New York, NY, USA.), 433–456.

Lester, R. N., and Thitai, G. N. W. (1989). Inheritance in Solanum aethiopicum, the scarlet eggplant. *Euphytica* 40, 67–74.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Available at: http://arxiv.org/abs/1303.3997 [Accessed April 17, 2019].

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–9.

Lin, T., Zhu, G., Zhang, J., Xu, X., Yu, Q., Zheng, Z., et al. (2014). Genomic analyses provide insights into the history of tomato breeding. *Nat. Genet.* 46, 1220–6.

Lynch, V. J., and Wagner, G. P. (2010). Did egg-laying boas break Dollo Law? Phylogenetic evidence for reversal to oviparity in sand boas (ERYX: BOIDAE). *Evolution (N. Y).* 64, 207–216.

Mammadov, J., Aggarwal, R., Buyyarapu, R., and Kumpatla, S. (2012). SNP markers and their impact on plant breeding. *Int. J. Plant Genomics* 2012.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing

reads. *EMBnet.journal; Vol 17, No 1 Next Gener. Seq. Data Anal.*

Mason, A. S., Zhang, J., Tollenaere, R., Vasquez Teuber, P., Dalton-Morgan, J., Hu, L., et al. (2015). High-throughput genotyping for species identification and diversity assessment in germplasm collections. *Mol. Ecol. Resour.* 15, 1091–1101.

Meyer, R. S., Karol, K. G., Little, D. P., Nee, M. H., and Litt, A. (2012). Phylogeographic relationships among Asian eggplants and new perspectives on eggplant domestication. *Mol. Phylogenet. Evol.* 63, 685–701.

Miz, R. B., Mentz, L. A., and Souza-Chies, T. T. (2007). Overview of the phylogenetic relationships of some southern Brazilian species from section Torva and related sections of "spiny Solanum" (Solanum subgenus Leptostemonum, Solanaceae). *Genetica* 132, 143–158.

Nairismägi, M.-L., Tan, J., Lim, J. Q., Nagarajan, S., Ng, C. C. Y., Rajasegaran, V., et al. (2016). JAK-STAT and G-protein-coupled receptor signaling pathways are frequently altered in epitheliotropic intestinal T-cell lymphoma. *Leukemia* 30, 1311–9.

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274.

Pailles, Y., Ho, S., Pires, I. S., Tester, M., Negrão, S., and Schmöckel, S. M. (2017). Genetic diversity and population structure of two tomato species from the Galapagos Islands. *Front. Plant Sci.* 8, 138.

Peralta, I. E., Spooner, D. M., and Knapp, S. (2008). Taxonomy of wild tomatoes and their relatives (*Solanum* sect. *Lycopersicoides*, sect. *Juglandifolia*, sect. *Lycopersicon*; *Solanaceae*). *Syst. Bot. Monogr.* 84, 186 pp.

Plazas, M., Andujar, I., Vilanova, S., Gramazio, P., Herraiz, F. J., and Prohens, J. (2014). Conventional and phenomics characterization provides insight into the diversity and relationships of hypervariable scarlet (Solanum aethiopicum L.) and gboma (S. macrocarpon L.) eggplant complexes. *Front. Plant Sci.* 5, 318.

Plazas, M., Vilanova, S., Gramazio, P., Rodríguez-Burruezo, A., Fita, A., Herraiz, F. J., et al. (2016). *Interspecific hybridization between eggplant and wild relatives from different genepools*. [American Society for Horticultural Science] Available at: http://journal.ashspublications.org/content/141/1/34.short [Accessed January 3, 2019].

Poplin, R., Ruano-Rubio, V., DePristo, M. A., Fennell, T. J., Carneiro, M. O., Van der Auwera, G. A., et al. (2017). Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv*, 201178.

Razali, R., Bougouffa, S., Morton, M. J. L., Lightfoot, D. J., Alam, I., Essack, M., et al. (2018). The genome sequence of the wild tomato *Solanum pimpinellifolium* provides insights into salinity tolerance. *Front. Plant Sci.* 9, 1402.

Rodriguez, F., Wu, F., Ané, C., Tanksley, S., and Spooner, D. M. (2009). Do potatoes and tomatoes have a single evolutionary history, and what proportion of the genome supports this history? *BMC Evol. Biol.* 9, 191.

Sakata, Y., and Lester, R. N. (1997). Chloroplast DNA diversity in brinjal eggplant (Solanum melongena L.) and related species. *Euphytica* 97, 295–301.

Sakata, Y., Nishio, T., and Matthews, P. J. (1991). Chloroplast DNA analysis of eggplant (Solanum melongena) and related species for their taxonomic affinity. *Euphytica* 55, 21–26.

Scaglione, D., Pinosio, S., Marroni, F., Centa, E. Di, Fornasiero, A., Magris, G., et al. (2019). Single primer enrichment technology as a tool for massive genotyping: a benchmark on black poplar and maize. *Ann. Bot.*

Scheben, A., Batley, J., and Edwards, D. (2017). Genotyping-by-sequencing approaches to characterize crop genomes: choosing the right tool for the right application. *Plant Biotechnol. J.* 15, 149–161.

Scolnick, J. A., Dimon, M., Wang, I.-C., Huelga, S. C., and Amorese, D. A. (2015). An efficient method for identifying gene fusions by targeted RNA sequencing from fresh frozen and FFPE samples. *PLoS One* 10, e0128916.

Semagn, K., Babu, R., Hearne, S., and Olsen, M. (2014). Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement. *Mol. Breed.* 33, 1–14.

Sim, S.-C., Van Deynze, A., Stoffel, K., Douches, D. S., Zarka, D., Ganal, M. W., et al. (2012). High-density SNP senotyping of tomato (*Solanum lycopersicum* L.) reveals patterns of genetic variation due to breeding. *PLoS One* 7, e45520.

Syfert, M. M., Castañeda-Álvarez, N. P., Khoury, C. K., Särkinen, T., Sosa, C. C., Achicanoy, H. A., et al. (2016). Crop wild relatives of the brinjal eggplant (*Solanum melongena*): Poorly represented in genebanks and many species at risk of extinction. *Am. J. Bot.* 103, 635–651.

The Tomato genome Consortium (2012). The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485, 635–41.

Thomson, R. C., and Shaffer, H. B. (2010). Sparse supermatrices for phylogenetic inference taxonomy, alignment, rogue taxa, and the phylogeny of living turtles. *Syst. Biol.* 59, 42–58.

Tranchida-Lombardo, V., Mercati, F., Avino, M., Punzo, P., Fiore, M. C., Poma, I., et al. (2018). Genetic diversity in a collection of Italian long storage tomato landraces as revealed by SNP markers array. *Plant Biosyst. - An Int. J. Deal. with all Asp. Plant Biol.*, 1–10.

Vilanova, S., Manzur, J. P., and Prohens, J. (2012). Development and characterization of genomic simple sequence repeat markers in eggplant and their application to the study of diversity and relationships in a collection of different cultivar types and origins. *Mol. Breed.* 30, 647–660.

Vorontsova, M. S., and Knapp, S. (2016). *A Revision of the "spiny Solanums," Solanum Subgenus Leptostemonum (Solanaceae), in Africa and Madagascar*. American Society of Plant Taxonomists Available at: https://books.google.it/books?id=4rrdnAAACAAJ.

Vorontsova, M. S., Stern, S., Bohs, L., and Knapp, S. (2013). African spiny *Solanum* (subgenus *Leptostemonum* , Solanaceae): a thorny phylogenetic tangle. *Bot. J. Linn. Soc.* 173, 176–193.

Weese, T. L., and Bohs, L. (2010). Eggplant Origins: Out of Africa, into the Orient. *Taxon* 59, 49–56.

Wiens, J. J., and Moens, D. S. (2008). Missing data and the accuracy of Bayesian phylogenetics. *J. Syst. Evol.* 46, 307–314.

Wiens, J. J., and Morrill, M. C. (2011). Missing Data in Phylogenetic Analysis: Reconciling Results from Simulations and Empirical Data. *Syst. Biol.* 60, 719–731.

Williams, C. E., and Clair, D. A. St. (1993). Phenetic relationships and levels of variability detected by restriction fragment length polymorphism and random amplified polymorphic DNA analysis of cultivated and wild accessions of *Lycopersicon esculentum*. *Genome* 36, 619–630.

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., and Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 28, 3326–3328.

CHAPTER IV. Analysis of the genetic and morphological variability of the tomato core collection of the G2P-SOL project

David Alonso[1], Jaime Prohens[1], María José Díez[1] and Giovanni Giuliano[2]

[1]Instituto de Conservación y Mejora de la Agrodiversidad Valenciana (COMAV). Universitat Politècnica de València. Camino de Vera 14, 46022 Valencia
[2]Agenzia Nazionale per le Nuove Tecnologíe, L'Energia e lo Sviluppo Economico Sostenible (ENEA). Via Anguillarese 301 – 00123 Roma, Italia (*In representation of the G2P-SOL consortium*)

*Corresponding author: daalmar8@upv.es

Ph.D. candidate contribution

David Alonso had a main role in the following activities: conceived and designed the research, data collection, data curation, data visualization, data analysis and drafting the manuscript.

AUTHOR'S VERSION.

BEING PREPARED FOR SUBMISSION TO A JOURNAL


4

CHAPTER IV

Analysis of the genetic and morphological variability of the tomato core collection of the G2P-SOL project

INTRODUCTION

Tomato (*Solanum lycopersicum* L.) is a crop of great economic and nutritional importance worldwide (FAO, 2021). The tomato was domesticated from its ancestor *S. lycopersicum* var. *cerasiforme* (Blanca et al., 2012; Razifard et al., 2020), which in turn was pre-domesticated from the wild species *S. pimpinellifolium*, native to Peru and Ecuador (Blanca et al., 2022a; Zuriaga et al., 2009). The most notable morphological changes during the domestication process affected the shape and size of the fruits (Frary et al., 2000; Liu et al., 2002; Cong et al., 2008; Xiao et al., 2008; Muños et al., 2011), resulting in increased plant vigor to support the development of significantly larger fruits. However, associated with domestication, there was also a significant loss of genetic diversity (Rick, 1978; Miller & Tanksley, 1990; Young et al., 2004), making cultivated tomatoes susceptible to pests and diseases. The adaptation to thrive in unfavourable environments, characteristic of wild species, was also lost in this process, further increasing the vulnerability of this crop. However, nowadays, tomatoes are cultivated in extensive areas throughout the world. This has been facilitated by the genetic improvement that has utilized the conserved genetic resources in germplasm banks, including related wild species. The use of biotechnological techniques has overcome the crossability barriers between tomatoes and some of these species included in the *Lycopersicon* section of the *Solanum* genus, as well as more distant species included in the *Lycopersicoides* section, such as *Solanum lycopersicoides* and *Solanum sitiens* (Smith, 1944; Canady et al., 2005; Chetelat et al., 2019). Despite these successes, the plant genetic resources conserved in germplasm banks remain underexploited, primarily due to the lack of accessible information.

The fast development of genotyping techniques has enabled the joint analysis of large collections, thus opening up possibilities for their utilization for various purposes, such as studies on domestication and diffusion of this species (Blanca et al., 2012, 2022a), population structure and genetic differentiation associated with breeding history and selection (Sim et al., 2011, 2012;Blanca et al., 2015), association mapping for traits related to fruit quality (Xu et al., 2013), genome-wide association mapping for agronomic, fruit quality, and root architectural traits (Tripodi et al., 2021a), history of tomato breeding (Lin et al., 2014; Zhu et al., 2018)), discovery of new genes (Gao et al., 2019), genomic diversity (Esposito et al., 2020), and the diversity and relationships between European traditional varieties (Blanca et al., 2022b; Pons et al., 2022).

The establishment of the core collections (subsets of the complete collection selected to include the majority of the diversity (van Hintum et al., 2000) and their genotyping and/or sequencing is an efficient way to optimize the utilization of genetic resources conserved in germplasm banks. This is the case of the core collection consisting of 227 entries constructed from 1499 entries of European traditional varieties (Pons et al., 2022). This collection has been used to explore the basis of their phenotypic variation, Genotype x Environment (GxE) interactions, and stability for more than thirty agro-morphological traits. Roohanitaziani et al., (2020) selected a core collection of 122 resequenced accessions to explore its phenotypic and genotypic variation in vegetative and fruit quality traits. Other collections have been strategically curated for diverse aims, subjected to resequencing, and extensively characterized to ensure breeders access to these resources and all the accompanying information (Mata-Nicolás et al., 2020). The availability of these collections facilitates the fine mapping of the genes underlying the most important agronomic traits.

With the aim of continuing to provide high-quality resources to the scientific community and effectively harnessing the natural variability preserved in germplasm banks, the G2P-SOL project (Linking genetic resources, genomes, and phenotypes of Solanaceous crops) was initiated in 2016. This project involved approximately 24,000 accessions of tomato and other Solanaceae crops. A core collection consisting of 450 accessions of tomato and wild relative's species selected from a subset of 15,504 genotyped accessions was established. In this study, we present the genetic and phenotypic analysis of the core collection, focusing on the changes in phenotypic and genotypic variation that have emerged through different processes of evolution and genetic improvement in this crop.

MATERIAL AND METHODS

Construction of the core collection and its composition

The complete tomato collection of the G2P-SOL project (http://www.g2p-sol.eu/) consisted of approximately 24,000 tomato accessions, with 6.9% corresponding to wild species. **CHAPTER I** of this thesis includes all relevant information about the collection. A subset of 15,054 accessions from the complete collection was sequenced at low density using Single Primer Enrichment Technology (SPET) (Barchi et al., 2019) (see **CHAPTER III**). Tomato SNPs panel were obtained from the SOL Genomics portal, specifically from the "150 Tomato Genome Resequencing Project" (52 tomato samples) (Aflitos et al., 2014) and the "AGIS Tomato 360 Resequencing Project" (184 tomato samples) (Lin et al., 2014). To identify interspecific alleles between *S. lycopersicum*, *S. pimpinellifolium*, and *S. lycopersicum* var. *cerasiforme*, an additional 159 samples were taken from the "AGIS Tomato 360 Resequencing Project". From the identified SNPs, 5,000 were selected (85% polymorphic in *S. lycopersicum* and 15% polymorphic in *S. pimpinellifolium*) for genotyping the G2P-SOL collection. The resulting 150,472 SNPs were filtered based on the percentage of missing data (<5%) and a minimum allelic frequency of MAF>0.01, resulting in a set of 4,249 polymorphic SNPs.

The core collection was established using the distance matrix obtained from the VCF file, employing Euclidean distance. The process began by selecting an entry randomly and grouping it with adjacent accessions within a small radius around it. This process was repeated with the remaining accessions, iteratively adjusting the chosen radius until all accessions were grouped into approximately 400 clusters. Simultaneously, a total of 79 accessions were preselected based on their agronomic performance, resistance to pathogens, quality traits or historical purpose. For each of the established 400 clusters, if it contained one or more preselected accessions, those accessions were chosen to be included in the core collection. For the remaining clusters that did not contain any preselected accessions, a representative accession was chosen. Additionally, 50 accessions from wild species were added to this collection.

Molecular characterization

For the analysis of genetic diversity and population structure of the collection, the genotypic information of the 450 tomato accessions belonging to the core collection was extracted from the VCF file obtained through SPET (Barchi et al., 2019) from the 15,054 accessions (Blanca et al., 2023, manuscript in preparation), using SAM-tools/BCFtools version 1.3.1. This subset of data was used for further analysis.

Phenotypic characterization

The phenotyping trial of the tomato core collection was carried out in a greenhouse in Alboraya, Valencia [39°30'33.3"N 0°20'00.9"W] from August 28, 2019, to January 10, 2020. Due to the difference in agronomic management between determinate and indeterminate tomato plants, the trial was divided into two plots based on the available prior information. A completely randomized three-block design was used for each set of plants (determinate or indeterminate), with three plants per block, resulting in nine plants representing each genotype. Seed germination was carried out in a commercial nursery, and the plantlet were transplanted into a field with certified organic cultivation at the 3-4 true leaf stage. The planting spacing was 1.25 m between rows and 0.45 m between plants for determinate plants, and 0.30 cm between plants for indeterminate plants.

A total of 25 descriptors were evaluated, based on the IPGRI descriptors (1996) and the authors' previous experience in the study. These descriptors included vegetative, inflorescence, fruit, and agronomic interest. Out of the 25 descriptors (**Table 1**), 11 were qualitative, and 15 were quantitative. Fruit firmness was measured using a model 53125 Fruit Hardness Tester (TR Turoni srl, Forli, Italy). The CIE 1976 L*a*b* parameters related to fruit color, were evaluated using a CR-400/410 colorimeter from Konica Minolta. Soluble solid content (Brix degrees) was measure using 0.5 mL of tomato liquid extract with a HI 96801 digital refractometer (HANNA instrument, Padua, Italy).

Table 1. List of evaluated quantitative and qualitative descriptors, their corresponding abbreviation, scale or unit, and a brief description.

Descriptor	Abbrevation	Units	Scale/Description
Vegetative descriptor			
Qualitatives			
Plant growth habit	grohab		1: Determinate; 2: Semideterminate; 3: Indeterminate
Inflorescence descriptors			
Qualitatives			
Inflorescence type	flotype		1: Uniparous; 2: Forked; 3: Irregular; 4: Compound
Inflorescence with leaf and shoot	floleaftype		1: Absent; 2: With leaf; 3: With shoots; 4: With leaf and shoots
Fruit descriptors			
Qualitatives			
External immature fruit color	frucolim		1: Whitish; 2: Light green; 3: Medium green; 4: Dark green
External mature fruit color	frucolma		1: Yellow; 2: Orange; 3: Pink; 4: Red; 5: Dark red; 6: Purple; 7: Brown; 8: Other
Green shoulder	greensho		0: Uniform; 1: Light green; 2: Medium green; 3: Dark green
Blossom end scar	bloshp		1: Dot; 2. Stellate; 3: Lineal; 4: Irregular
Ribbing at calyx end	ribcalend		1: Smooth; 2: Weak; 3: Intermediate; 4: Strong
Fruit predominat shape	frushp		1: Very flat; 2: Slightly flat; 3: Round; 4: Oxheart; 5: Heart, 6: Rectangular; 7: Bell pepper; 8: Ellipsoid; 9: Obovoid; 10: Long pepper
Fruit set sequence	fruset		1: Very low; 3: Low; 5: Intermediate; 7: High; 9: Very high
Quantitatives			
Fruit firmness	frufirm	kg/cm^2	Measured at the longitudinal midpoint of a ripe fruit from one plant.
Color L,a,b parameter			Measured at the longitudinal midpoint of a ripe fruit from one plant.
Luminosity	lcol		0: Black; 100: White
Red-green balance	acol		Chromatic coordinates from red to green.
Blue-yellow balance	bcol		Chromatic coordinates from yellow to blue.
Brix degree	brix		Measured in one fruit per plant
Pericarp thickness	fruperi	mm	Measured at the equatorial section at the point of maximum fruit width in one fruit per plant.
Fruit length	frulen	mm	Measured from the base to the beginning of the pedicel in one fruit per plant.
Fruit width	fruwid	mm	Measured at the longest transverse diameter in one fruit per plant.
Ratio length/width	ratiolenwid	mm	Obtained by dividing fruit length by fruit width.
Number of locules	nrloc		Measured in one fruit per plant

Descriptor	Abbrevation	Units	Scale/Description
Agronomic descriptors			
	Qualitatives		
Puffiness	puff		0: Absent; 3: Light; 5: Intermediate; 7: Severe
	Quantitatives		
Number of fruits	nr_fruits		The fruits of three plants are counted, and the block average is calculated.
Fruit weight	fruweight_mean	g	Two fruits per block are weighed, and the block average is calculated.
Yield	yld	g	Calculated by multiplying the average number of fruits by the average fruit weight.
Flowering earliness	nr_sowtoflo		Number of days from sowing to the opening of the first flower. One measurement per block.
Ripening earliness	nr_sowtoripe		Number of days from sowing to the first ripe fruit. One measurement per block.

Data analysis

Genotypic data

The analysis of the genomic diversity was inferred using several approaches. The VCF file containing the genetic data obtained through SPET sequencing was used for this analysis In order to study the genetic structure and relationship among the accessions in the core collection, a principal component analysis (PCA) was conducted with the 'snpgdsLDprunning' and 'snpgdsPCA' functions from the *SNPRelate* 1.28.0 library (Zheng et al., 2012) in R. The 'ggplot' function from the *ggplot*2 3.4.0 library (Wickham, 2016) was used for visualizing the results in two dimensions. To generate more informative successive PCAs in two and three dimensions, the 'plot_ly' function from the *plotly* 4.10.1 library (Sievert, 2020) was employed.

The study of genetic diversity was conducted by calculating Nei's genetic diversity index (Nei, 1987) using the 'basic.stats' function from the *hierfstat* 0.5.11 package (Jerome and Thibaut, 2022) and the poppr 2.9.3 package (Kamvar et al., 2015; Kamvar and Grünwald, 2023) in R. For calculating the percentage of missing SNPs and the proportion of polymorphic loci in which the most frequent allele is lower than 0.95 frequency, TASSEL v5 (Bradbury et al., 2007) was used.

Phenotyping data

Firstly, the uniformity among the plants of each accessions was checked by removing accessions that showed heterogeneity for two or more traits. The generated data during the characterization underwent pre-processing to detect outliers. Outliers were identified using the inter-quartile range (IQR) and those falling outside the range were removed. These range values were defined as those below Q1-1.5*IQR or above Q3+1.5*IQR, where IQR is equal to Q3-Q1. With the curated data, the average, standard deviation, variance, maximum, and minimum were calculated for each quantitative trait using the *summarytools* 1.0.1 library (Comtois, 2019).

With the aim of exploring the existing variability among the studied accessions and their relationship with the established genetic groups, a principal component analysis (PCA) was conducted using both qualitative and quantitative data. For the qualitative data, a scale was used, where higher values corresponded to greater expression of the trait. The PCAs were performed using the 'prcomp' function from the *stats* 4.1.3 library in R, and the obtained results were visualized using the 'ggplot' function and the ggbiplot extension. Qualitative data was represented using mosaic plots, while boxplots were used for quantitative data. Additionally, the correlation between traits was calculated using the 'corrplot' function (Wei et al., 2021), aiming to identify patterns and relationships among the different evaluated traits, thereby facilitating a better understanding of the morphological variability of the studied accessions.

RESULTS AND DISCUSSION

Composition of the G2P-SOL core collection

The core collection included 450 accessions sourced from the germplasm banks of WorldVeg, INRAE, UPV-COMAV, MVCRI, WUR-DLO, IPK, and BATEM. According to the available passport data (**Supplementary Table 1**), out of the 450 accessions, 354 corresponded to the cultivated species (*S. lycopersicum* var. *lycopersicum* (SLL)), 46 to *S. lycopersicum* var. *cerasiforme* (SLC), and 50 to wild species (**Table 2**). The geographical range of origin of the core collection encompasses the center of origin and domestication and centers of diversity (Europe, Asia), including wild species originating from Ecuador, Chile, and Peru (**Figure 1**). Furthermore, this collection includes representatives of all wild species related to tomatoes included in the *Lycopersicon* section and relevant for breeding purposes. Regarding the geographical origin of SLC accessions, the majority came from Mexico, Ecuador, Peru, and Bolivia. As for SLL, they are distributed worldwide. On the other hand, the core collection showed a balanced distribution of accessions with different biological status (**Table 3**). Approximately 34.0% of the accessions were traditional cultivars, originating mainly from Spain, Italy, and Central and South America. These cultivars represented the rich heritage and traditional farming practices associated with the crop. On the other hand, 28.2% of the accessions were classified as modern cultivars, reflecting the influence of advancements in breeding from regions such as Northern Europe, Asia, and North America. Another 15.5% of the accessions consisted of breeding materials, indicating the continuous efforts to improve the crop's traits and adaptability. This diverse set of genetic resources was further enhanced by the inclusion of the 12 closest wild species to tomato.

Table 2. Species composition of the tomato core collection provided by different germplasm banks

Species	AVRDC	INRAE	COMAV-UPV	MVCRI	WUR-DLO	IPK	BATEM	Total
S. lycopersicum	143	32	55	15	4	97	8	354
S.l. var. cerasiforme	23	5		2		16		46
S. pimpinellifolium	8		1		1	6		16
S. galapagense	1							1
S. cheesmaniae		1						1
S. chmielewskii		2						2
S. arcanum			1					1
S. peruvianum	5			1	1			7
S. corneliomulleri		1			1			2
S. chilense		5						5
S. pennellii			1					1
S. habrochaites	1	11	1			1		14
Total	181	57	59	18	7	120	8	450

Table 3. Biological status of accessions composition of the core collection provided by different germplasm banks

Biological status of accession	AVRDC	INRAE	COMAV-UPV	MVCRI	WUR-DLO	IPK	BATEM	Total
100) Wild	15	20	4	1	3	8		51
200) Weedy	7	3	1					11
300) Traditional cultivar	53	4	54			34	8	153
400) Breeding material	38	9		10		13		70
500) Modern cultivar	46	11		7	4	59		127
900) Other						1		1
No data	22	10				5		37
Total	181	57	59	18	7	120	8	450

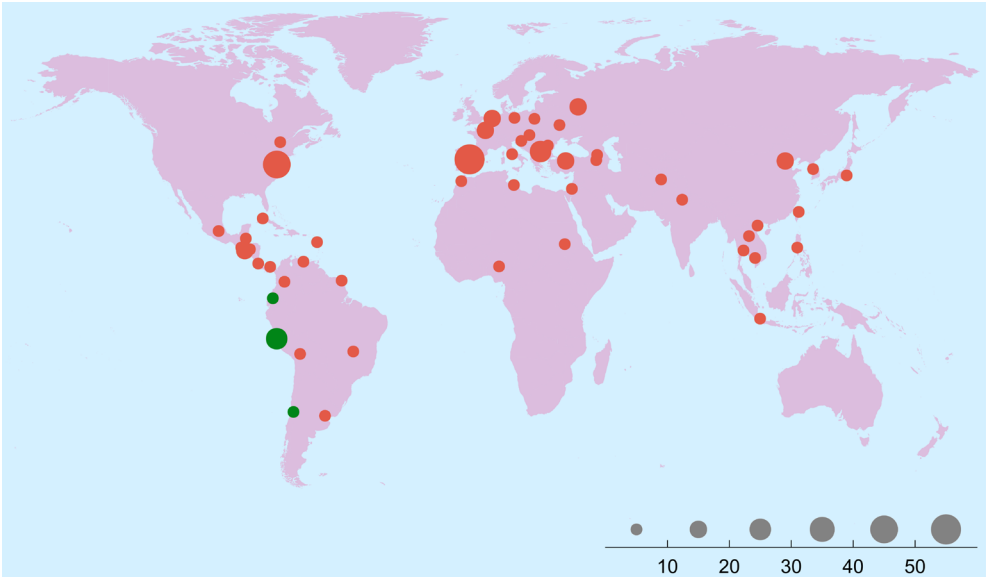


Figure 1. Geographical distribution of the 450 accessions included in the G2P-SOL tomato core collection. The red color corresponds to cultivated tomatoes and *S.l.* var. *cerasiforme*, while the green color represents all wild species.

Genetic structure of the core collection

Outside the scope of this doctoral thesis, a preliminary genetic analysis was conducted using all genotyped accessions from the *Lycopersicon* section, using SPET. As a result, the accessions were structured into a series of genetic groups through the implementation of consecutive Principal Component Analysis (PCAs) (Blanca et al., 2023, manuscript in preparation). These groups were formed based on the genetic proximity of the included accessions. In addition, they were profiled using passport data (geographical origin, variety name, and biological status of accessions, primarily (**Supplementary Table 1**)), morphological characteristics provided by germplasm banks and supplying institutions, and the evolutionary history of tomatoes known from previous studies conducted by the group (Blanca, et al., 2022a). The established groups were named as follows and included the following types of materials:

- **green**: wild species with green fruit (*Solanum chmielewskii*, *Solanum arcanum*, *Solanum peruvianum*, *Solanum corneliomulleri*, *Solanum chilense*, *Solanum pennellii* and *Solanum habrochaites*).
- **gal_chees**: accessions from *Solanum galapagense* and *Solanum cheesmaniae* species.
- **sp**: accessions from *Solanum pimpinellifolium* species.
- **slc_ecu_per_bol**: accessions of *Solanum lycopersicum* var. *cerasiforme* considered the earliest local cultivated varieties from Ecuador, Peru, and Bolivia.
- **slc_ma**: accessions of *Solanum lycopersicum* var. *cerasiforme* considered the earliest local varieties from Mesoamerica.
- **early_vint**: initial local tomato varieties.
- **vint_esp**: traditional or local varieties from Spain.
- **vint_ita**: traditional or local varieties from Italy.
- **eim**: first improved varieties without wild species introgressions.
- **mim**: modern improved materials, including varieties and breeding materials with wild species introgressions.
- **im_cherry**: modern improved varieties of "Cherry" type. Although they could have been included in the previous group, they have been separated as a distinct group due to their specific characteristics and use.

For the analysis of the core collection, we considered this genetic classification to enhance the collection's interpretation, discussion, and utilization.

The PCA performed on the core collection was carried out using 4,249 SNPs obtained through SPET genotyping (**Figure 2**). PC1 accounted for 41.65% of the observed variation and showed a clear separation between all wild species and *S. lycopersicum*. Similarly, PC2 (5.63% of the variation) separated the wild species into three groups: green-fruited species (**green**), red-fruited *Solanum pimpinellifolium* (**sp**), and orange-fruited species *S. galapaguense* and *S. cheesmaniae* (**gal_chees**). Finally, in the representation of the accessions in the plane defined by PC1 and PC3, a distinct separation among the green-fruited species was also observed. Positioned at the upper part of the green cluster is the "Arcanum" group, comprising the species *S. arcanum* and *S. chmielewskii*. In the middle section, we find the species *S. peruvianum*, *S. corneliomulleri*, and *S. chilense*, forming the "Peruvianum" group. In this area, *S. pennellii* is also located, slightly more distanced from the main cluster. Finally, at the bottom, we have *S. habrochaites*, belonging to the "Hirsutum" group. The remaining accessions are scattered, the representation on the axes formed by PC2 and PC3 providing a better interpretation. It can be observed that the accessions comprising the main core are those of **vint_esp**, **vint_ita**, and the majority of **eim**. Additionally, some of the **early_vint** accessions are included in this group, although many of them scatter and join the Mesoamerican *S.l.* var. *cerasiforme* (**slc_ma**). The group of accessions classified as **early_vint** predominantly originates from countries in Central and South America. These are local varieties that have not undergone any breeding process and display considerable genetic diversity. In an analysis conducted with a larger representation of *S.l.* var. *cerasiforme* materials (data not shown), these accessions exhibit close proximity to this variety, suggesting their ancestral nature and sharing of a significant portion of the genome with *S.l.* var. *cerasiforme*. Furthermore, they were cultivated and traded locally. These materials stand in contrast to the varieties included in **vint_esp**, **vint_ita**, and the early improved materials (**eim**), which form a compact nucleus, indicating the limited genetic diversity present within this group. The loss of genetic diversity in cultivated tomatoes is a well-established and demonstrated phenomenon by several authors (van der Beek et al., 1992; Tanksley & McCouch, 1997; Villand et al., 1998; Young et al., 2004; Tam et al., 2005; García-Martínez et al., 2006; Blanca et al., 2022a), and is reflected in this extensive collection of genetically unimproved or only intraspecifically improved varieties, such as the early improved materials (**eim**).

In order to further investigate the relationships between groups of traditional local varieties and the oldest improved cultivars, a new PCA has been conducted, incorporating the **early_vint** group (the oldest tomato core), **vint_esp**, **vint_ita**, and the early improved cultivars (**eim**) (**Figure 3**). The depiction of the accessions with the first two components reveals a distinct division between the **early_vint** and the traditional varieties from Spain (**vint_esp**) and Italy (**vint_ita**). Furthermore, a notable differentiation exists between the traditional varieties from Spain

(**vint_esp**) and Italy (**vint_ita**) and the close connection observed among them and the early improved cultivars (**eim**). It is worth noting that the initial progress in tomato breeding relied on harnessing intraspecific variation, with a primary objective of combine advantageous traits found in diverse varieties through controlled crossbreeding without the introduction of interspecific germplasm (Bai and Lindhout, 2007).

Furthermore, modern improve cultivars (**mim**) also exhibit dispersion, although in a different manner than the **early_vint** (**Figure 2**). The wide distribution of these materials indicates their higher genetic diversity resulting from the incorporation of interspecific variation. In fact, the key criterion for classifying these accessions as improved modern cultivars is the presence of wild species introgressions (data not shown), encompassing both commercially marketed cultivars and breeding materials.

Finally, **slc_ecu_per_bol** is also dispersed beyond the main core, as it comprises entries of the *cerasiforme* variety (**Figure 2**). A special case is represented by the modern "Cherry" cultivars (**im_cherry**). It is widely recognized that the wild species *S. pimpinellifolium* has been involved in the development of these cultivars, contributing to their compact fruit size, abundant fruit yield per inflorescence, and high Brix degree levels (Casals et al., 2018), which results in a notable resemblance to the *S. pimpinellifolium* (**sp**) accessions.

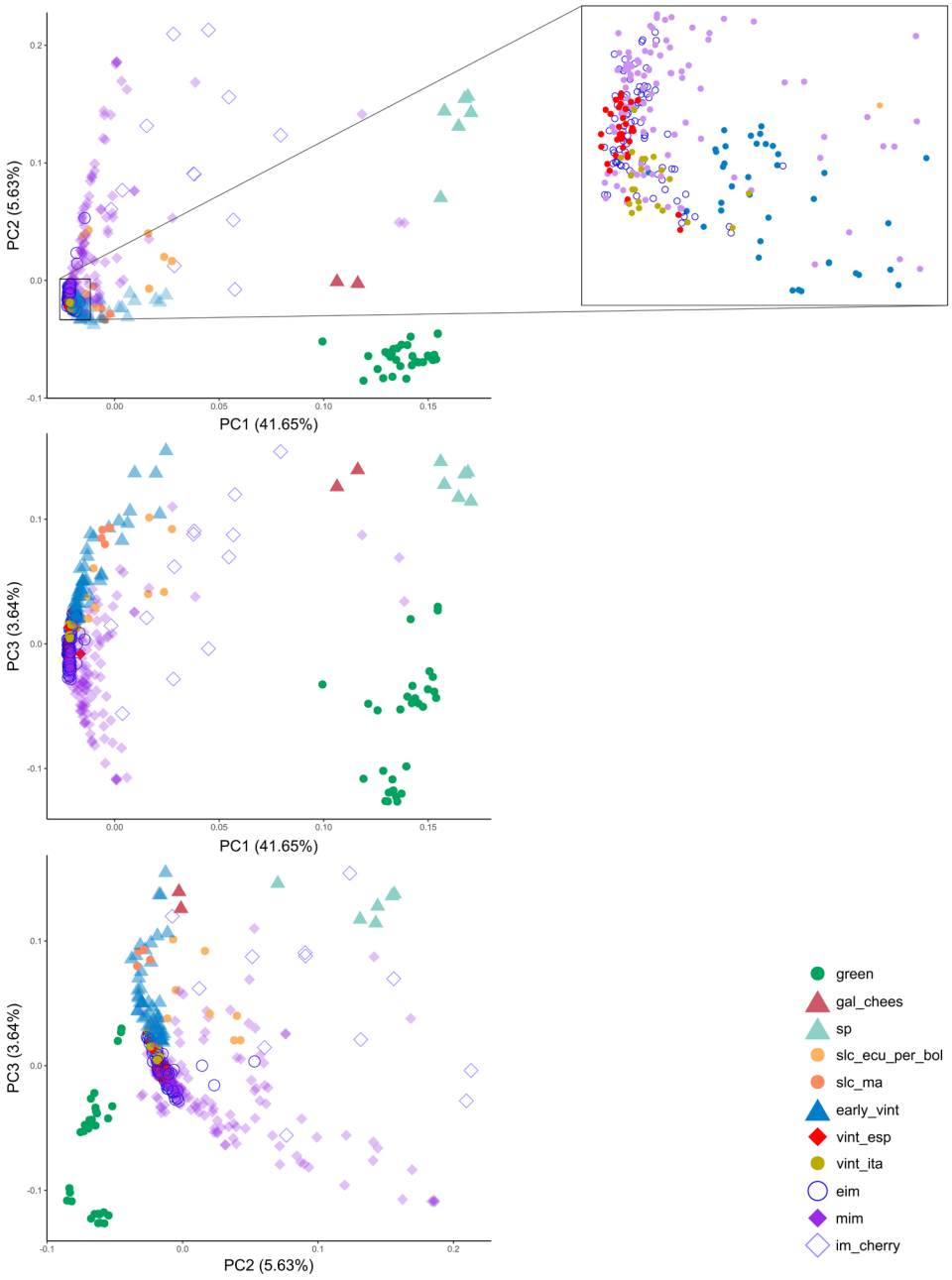


Figure 2. Principal component analysis of the 450 tomato accessions based on 4,249 SNPs.

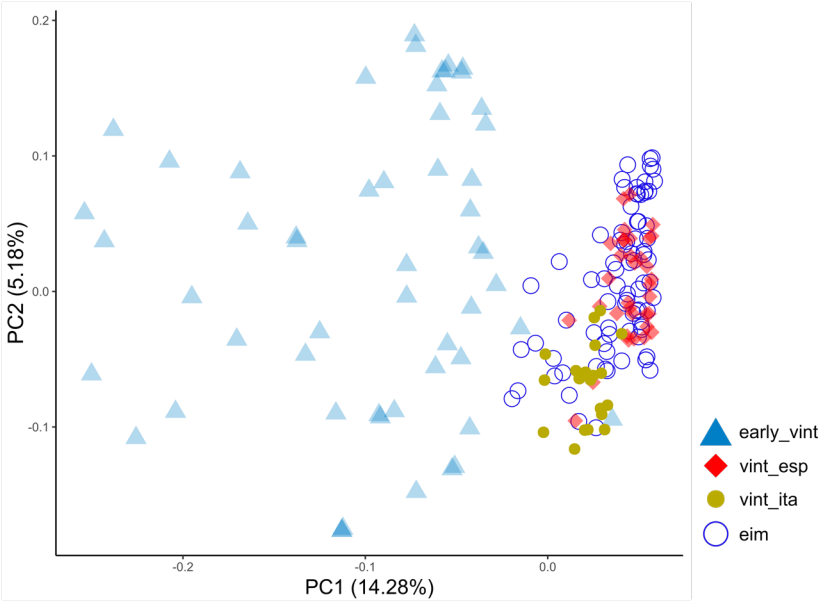


Figure 3. Principal component analysis of the 172 accessions from the **early_vint**, **vint_esp**, **vint_ita**, and **eim** groups of the core collection based on 4,249 SNPs.

Genetic diversity of the core collection

The results from the diversity indices further support the findings of the PCA analysis (**Figures 2,3**). In terms of the Nei's diversity index (**Figure 4a**), the accessions from **sp** and **slc_ecu_per_bol** showed considerable diversity. The discovery of lower diversity in wild species with green fruit (**green**) is a noteworthy finding, which contradicts the results reported by Schouten et al., 2019. This discrepancy may be attributed to the method employed for SNP selection panel, which focused on their polymorphism in *S. lycopersicum* (85%) and *S. pimpinellifolium* (15%) (Barchi et al., 2019). In fact, the selection of SNPs based on this criterion led to a significant number of missing SNPs in the more distantly related species, ranging from 3.22% in *S. chmielewskii* to 12.77% in *S. chilense* (**Figure 4b**). Similar results were obtained by Barchi et al., (2019), with the percentage of missing SNPs varying in this case from 1.7% in *S. pimpinellifolium* to 4.2% in *S. neorickii*, and reaching a maximum value of 10.4% in *S. habrochaites*. There is a significant decline in diversity when transitioning from the accessions of *S.l.* var. *cerasiforme* from South American countries (**slc_ecu_per_bol**) to those of the same species from Mesoamerica (**slc_ma**). This fact is crucial as it indicates where the highest genetic diversity is found within this variety and where the bottleneck phenomenon occurred as it moved away from the center of origin. The genetic diversity of these cultivars has already been highlighted in previous studies, emphasizing their potential significant

role in tomato breeding (Rick & Holle, 1990; Blanca et al., 2012, 2015). The genetic diversity decreases, reaching its lowest values in the traditional varieties from Spain (**vint_esp**) and Italy (**vint_ita**). Studies conducted by Blanca et al., (2022b) with a comprehensive collection of traditional varieties from Spain, France, Italy, and Greece support this finding. The early improved materials (**eim**), which were predominantly developed from traditional cultivars, maintain this low level of genetic diversity. The modern cultivars (**mim**), particularly the Cherry types (**im_cherry**), increase genetic diversity by incorporating wild species into breeding programs. This increase in genetic diversity resulting from genetic improvement has been highlighted in recent studies (Schouten et al., 2019b). This publication aims to provide evidence of this increase in diversity to refute a previous publication by Zsögön et al., (2018), which highlights the opposite, stating that '*the focus on yield improvement has been accompanied by a loss of genetic diversity and nutritional quality and flavor*'. Schouten et al., (2019) provide evidence to counter this claim by conducting a comprehensive study on the changing dynamics of varieties in Dutch greenhouses over the last seven decades, demonstrating the opposite effect. Indeed, the loss of diversity was evident in the 1950s and 1960s, but it has now increased eightfold. According to these authors, this increase can be attributed to two new breeding objectives pursued: disease resistance breeding, which led to the widespread utilization of wild species in breeding programs, and quality improvement. The utilization of wild species had both positive and adverse effects. On the positive side, it facilitated the identification of numerous useful disease resistance genes. However, it also had a negative impact on fruit quality, resulting in significant consumer rejection. As a result, seed companies made fruit quality breeding a top priority. In this case as well, wild species were employed, focusing on volatile compound profiles (VOCs) and organoleptic and nutritional quality components. The changes observed in these compounds in the improved cultivars confirm this phenomenon. Moreover, the significant increase in diversity, specifically among Cherry-type cultivars (**im_cherry**), was already highlighted by Sim et al., (2012), who found the highest diversity exhibited by modern cultivars for fresh and processing market compared to traditional varieties, with Cherry-type cultivars showing the highest diversity values. The proportion of polymorphic loci corroborates the results demonstrated by the genetic diversity index (**Table 3**).

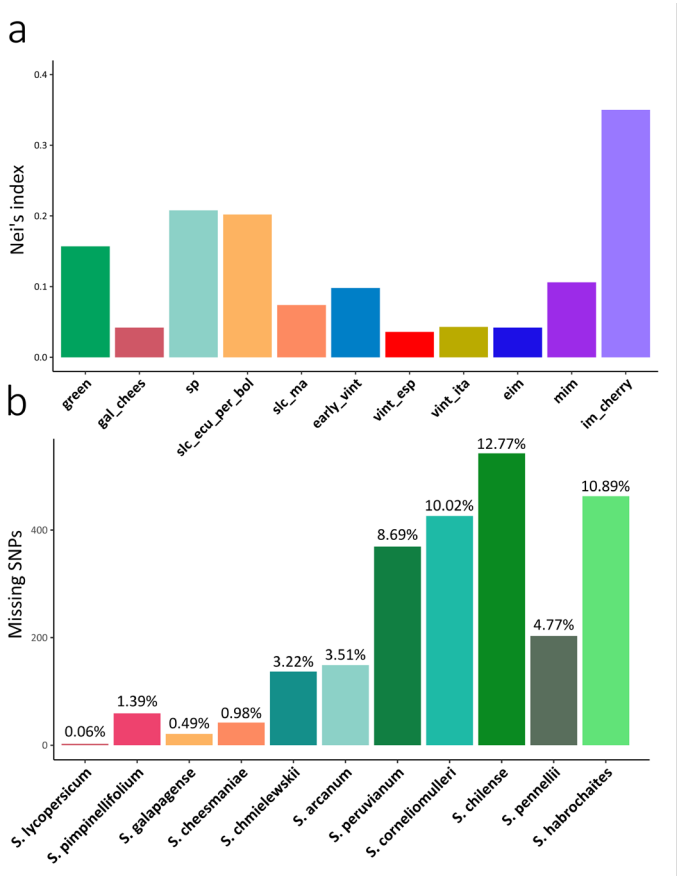


Figure 4. **(a)** Nei's diversity index (1987) per genetic groups and, **(b)** Number and percentage of missing SNPs per species.

Table 3. Proportion of Polymorphic SNPs.

Specie	Genetic group	Nr. Sample	nr. Alleles	P(0.95)
All collection	total	448	3193	75.2%
S. lycopersicum	total	392	1302	30.6%
	early_vint	53	1321	31.1%
	vint_esp	35	402	9.4%
	vint_ita	22	427	10.1%
	EIM	83	460	10.8%
	MIM	190	1334	31.4%
	im_cherry	9	2252	53.0%
S. l. cerasiforme	total	18	3760	88.5%
	slc_ecu_per_bol	9	2842	66.9%
S. pimpinellifolium	sp	7	2437	57.4%
Wild relatives	green	31	1812	42.5%

nr.Alleles. Numbers of polymorphic SNPs out of 4,248 SNPs in each genetic group
P(0.95). Percentage of markers with a frequency of the most common allele below 0.95.

Phenotypic characterization of the core collection

The complete tomato collection consisted of 450 accessions. However, 18 accessions could not be included due to phytosanitary requirements. Of the 432 accessions evaluated in the field, 384 were cultivated tomatoes and 49 were wild species. The wild species were not phenotyped; they were only used for seed regeneration and taxonomic confirmation of the species, except for *S. pimpinellifolium*, where Brix degrees were assessed. Out of the 384 tomato accessions phenotyped, 31 accessions were considered heterogeneous due to a high range of segregation based on the phenotypic assay and therefore were excluded from subsequent analysis. Finally, 353 accessions were characterized for 26 traits, including 11 qualitative and 15 quantitative traits.

Analysis of morphological variability among established groups

To explore the existing variability in the studied set of accessions and its relationship with the established genetic groups, a principal component analysis (PCA) was conducted using 11 qualitative and 15 quantitative traits, encompassing vegetative, inflorescence, fruit, and agronomic traits (**Table 1**). The wild species (**green** and **gal_chees**) were not included in these analyses. Only *S. pimpinellifolium* (**sp**) was included in the analysis of °Brix, as it is a quality trait of interest in breeding where high variability has been found in this species (Mata et al., 2000; Casals et al., 2018). **Figure 5** depicts the contribution of the traits to the first two principal component, with further details on the traits and their contribution to the first three components provided in **Supplementary Table 2**. PC1 and PC2 accounted for 32.49% and 16.89% of the observed variance, respectively. PC1 showed a positive correlation with fruit traits such as **fruwid**, **fruweight_mean**, **ribcalend**, **bloshp**, and **nrloc**, while negatively correlating with **nr_fruits** and **brix**. As a result, accessions with larger, multilocular, and ribbed fruits were positioned towards the right side of the figure, while those with smaller fruits and higher °Brix values were located towards the left side. On the other hand, PC2 exhibited positive correlations with traits like **ratiolenwid**, **frushp**, **frufirm**, **fruperi**, and **frulen**, which characterize processing-type cultivars with elongated fruits, firmer flesh, and thicker pericarp. **grohab** and **greensho** traits displayed a negative correlation with PC2, indicating the determinate growth habit and absence of green shoulders typically found in these cultivars. The PC3 were related to the fruit color traits.

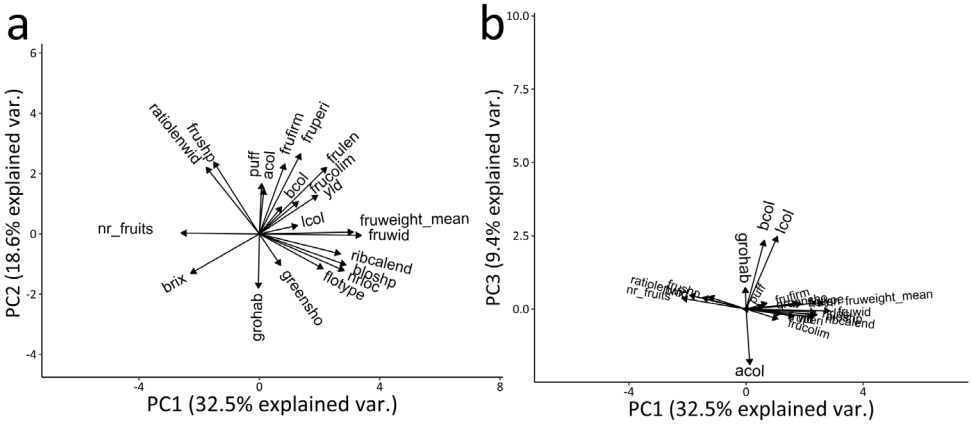


Figure 5. Biplot graph illustrating the contribution of the traits in (**a**) PC1 and PC2; (**b**) PC1 and PC3.

To provide a visual representation of the accession distribution, **Figure 6a** showcases the projection of the accessions onto the plane defined by the first two principal components. The accessions belonging to each genetic group, determined

by the genotyping data, are highlighted with different colors. Although there is a continuous gradient, a clear separation can be observed between the SLC accessions (**slc_ma** and **slc_ecu_per_bol**), as well as the modern Cherry-type cultivars (**im_cherry**), and some of the **early_vint** accessions. This separation is mainly attributed to the small size of their fruits and their high °Brix values (**Figure 5,6a**). The Spanish traditional varieties (**vint_esp**) are located in a wide part of the figure, predominantly in the center-right, indicating a high morphological variability and a predominance of ribbed and large-sized fruits. The traditional Italian varieties (**vint_ita**) are positioned higher in the graph, primarily due to the elongated shape of their fruits, as the majority of them are of the conserved type. The collection of early improved cultivars (**eim**) exhibits a notable range of morphological variability, as evidenced by their distribution across the entire graph. Within the collection, the modern cultivars (**mim**) can be categorized into two distinct groups. The first group, located at the upper portion of the graph, comprises determinate growth plants and elongated-shaped fruits, which are characteristic of varieties for tomato processing. The second group consists of larger-sized fruits with non-elongated shapes, more suitable for fresh consumption. Furthermore, the third principal component is mainly associated with color-related traits (**acol**, **bcol**, and **lcol**). The projection of accessions onto the plane defined by components 1 and 3 clearly separates accessions with yellow and orange fruits, as well as those with darker colorations (**Figure 6b**).

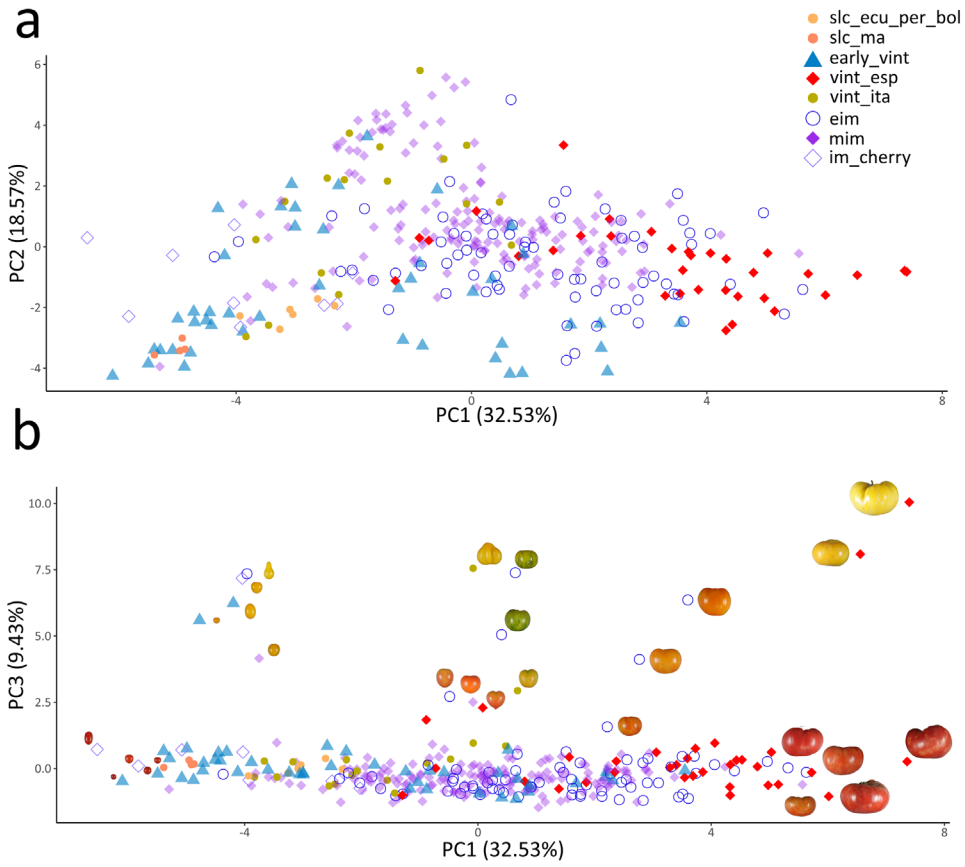


Figure 6. **Principal Component Analysis of 353 accessions from the tomato core collection using 25 phenotypic descriptors**. **(a)** First and second principal components. **(b)** First and third principal components. The tomatoes in the graph represent the contribution of color (lcol, acol, bcol) and fruit size to components 1 and 3.

Correlations between traits

The calculated correlations (**Figure 7**, **Supplementary Figure 1**) with the studied set of accessions allow establishing certain relationships between traits.

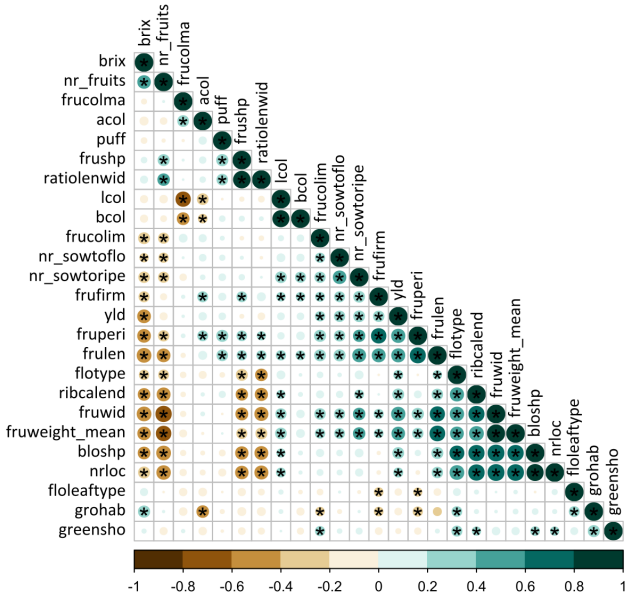


Figure 7. Correlation matrix of 25 vegetative, inflorescence, fruit and agronomic traits. Bonferroni's correction was applied to calculate the significance (p-value). Significance for $p \leq 0.05$ is marked with an asterisk.

Traits related to yield

Yield (**yld**) is one of the traits that shows a higher number of correlations with other traits, except for those related to fruit color and shape, and fruit number. Thus, yield depends much more on fruit weight (**fruweight_mean**, r=0.50) and, consequently, on other closely correlated traits. On the other hand, both yield and fruit weight are negatively correlated with Brix degree (r=-0.52 and r=-0.49, respectively). This may be partly due to a dilution of fruit soluble solids with increasing fruit weight or increasing total yield (Roohanitaziani et al., 2020). **Figure 8** illustrates the relationship between Brix degree and yield (**a**) and Brix degree and fruit weight (**b**). However, the described negative correlation is not constant when considering each of the genetic groups considered (**Supplementary Figure 1**). It should be noted that the variation for both traits is very different within each genetic group, and this has a significant effect on the correlation. Thus, in the **slc_ecu_per_bol** and **im_cherry** groups, the correlation between Brix degree and fruit weight is zero and very low for **slc_ma**, while for the rest of the groups, it is high. Similar results with some differences were obtained for the correlations between Brix degree and yield.

Roohanitaziani et al., (2020), studying a core collection composed of 122 accessions, which also included high variability for the studied traits, also observed an inverse correlation between fruit weight and soluble solids. In their case, it was observed in accessions with a fruit weight below 30 g, including Cherry types and wild species, while in fruits weighing over 30 g, the variation was very small, with values not exceeding 5.6 °Brix in the range of 30 to 300 g. In our case, the range of variation for both traits is larger, with varieties having fruits up to 497 g and 8.3 °Brix. The different variability for the considered traits within each group is responsible for this fact, indicating that within each group, the best combination of fruit weight/°Brix must be selected. One of the accession that accumulates a relatively high fruit weight and acceptable °Brix is GPT003310, a selection of the cultivar Montfavet, with fruits of 119.8 g and 6.7 °Brix. Roohanitaziani et al., (2020) also found cultivars with high production and °Brix, suggesting, according to these authors, that soluble solid content can be influenced by genetic factors that do not affect total yield. This was also shown for the *Brix9-2-5* QTL, caused by variation in the *Lin5* gene, which was introgressed from *S. pennellii* in a cultivated tomato background and led to a significant increase in Brix degree without an adverse effect on fruit size and yield (Fridman et al., 2002, 2004).

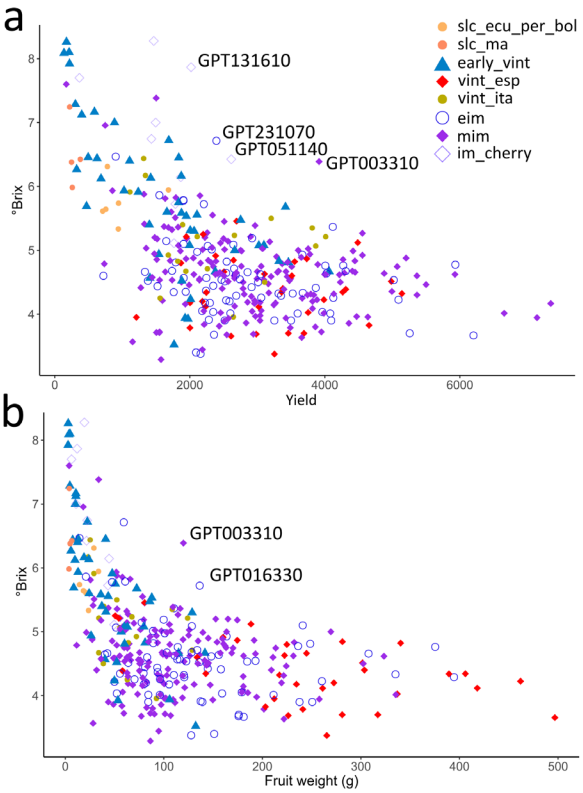


Figure 8. Scatter plot of yield (**a**) and fruit weight (**b**) versus brix degree (°Brix)

Fruit traits

Regarding fruit traits, they can be grouped into two clusters. On one hand, **frushp** and **ratiolenwid**, which show higher values in the case of cylindrical, pear-shaped, or ellipsoidal fruits, and are strongly correlated with each other. On the other hand, the remaining traits, **ribcalend**, **fruwid**, **fruweight_mean**, **bloshp**, and **nrloc**, show higher values in the case of rounded or flattened fruits, which are smaller in size and often have larger scars and ribs. Both groups of traits exhibit negative correlations among them.

Flowering and ripening earliness

Agronomic traits related to flowering earliness (**nr_sowtoflo**) and ripening earliness (**nr_sowtoripe**) show a clear negative correlation with the number of fruits per plant (**nr_fruit**, r=-0.27 and r=-0.32, respectively) and Brix degree (r=-0.22 and r=-0.34, respectively), indicating that smaller-sized fruits are the earliest to mature. Conversely, the correlations between both measures of earliness and traits associated with larger-sized fruits are positive, suggesting the need for a longer period of time to reach maturity in larger-sized fruits. Therefore, fruit size is an important trait to consider when applying the principles of 'speed breeding' in breeding programs. Nowadays, competition among companies to launch new hybrids to the market as quickly as possible has led to research related to what is known as speed breeding. It has been demonstrated that cold treatment of cotyledons, specific fertilization doses, or illumination are factors that significantly influence the number of days elapsed between sowing and obtaining the first mature fruit (Gimeno-Páez et al., 2023). According to our findings, fruit size should be also considered.

Morphological diversity by genetic groups

To gain a more comprehensive understanding of the morphological traits exhibited by the accessions in the core collection and explore the practical implications of including these entries within each established genetic group, a morphological description of the materials is conducted, taking into account their genetic clustering. This approach provides valuable insights into the evolution of morphological characteristics from the oldest materials to the most modern varieties and offers practical implications for further use. The results presented below are based on **Figures 9**, **10** and **11**, representing the quantitative and qualitative traits, respectively, as well as **Table 4**, which provides the most important statistics of each genetic group.

1. Accessions of *S. lycopersicum* var. *cerasiforme* considered as the first local varieties from Ecuador, Peru and Bolivia (slc_ecu_per_bol)

The accessions included in this group (n = 6) show indeterminate growth habit, with uniparous inflorescences. The fruits exhibit varying shades of green in their immature state, turning red when ripe, with variable presence of green shoulders. Generally, they have a slightly flattened or rounded shape and a smooth surface. The fruits are very small, with an average weight of 23.3 grams, thin pericarp, without scars or puffiness, and are bi- or trilocular. The high value of Brix degrees is noteworthy, surpassed only by Cherry-type cultivars (**im_cherry**) and *S.l.* var. *cerasiforme* from Mesoamerica (**slc_ma**), although the highest Brix degrees values are found in *S. pimpinellifolium* accessions (**sp**). The majority of accessions in this group were collected from markets in Bolivia and Peru in the 1930s. Hence, these are highly ancient materials that, in genetic analyses, exhibit the closest resemblance to entries of the *S.l.* var. *cerasiforme* variety, thus justifying the designation of this genetic group. Their substantial genetic variability, surpassed only by accessions of *S. pimpinellifolium*, makes them highly valuable for breeding purposes.

2. Accessions of *S. lycopersicum* var. *cerasiforme* considered as the first local varieties from Mesoamerica (slc_ma)

This group includes 4 accessions from El Salvador and Honduras. The plants have an indeterminate growth habit, with predominantly uniparous inflorescences. The immature fruits color is light green that turning red when ripe. The fruits generally lack green shoulders or exhibit only slight intensity, exhibiting a smooth and rounded shape, with a very small size. The Brix degree content for this set of accessions are comparable to those of accessions belonging to the group of modern Cherry-type cultivars (**im_cherry**). Most of the accessions included in this group were collected from roadside locations. As seen in the description of genetic diversity, there is a significant loss of genetic diversity compared to the previous group, although the limited number of entries included in this group should be taken into consideration.

3. Initial local tomato varieties (early_vint)

The plants included in this group (n = 45) are mainly indeterminate plants, with predominantly uniparous inflorescences and, to a lesser proportion, compound inflorescences. The fruits exhibit variation in the intensity of green shoulders. Regarding fruit color, in immature stage presents light or medium green color, and in mature stage displays a slightly wide range of colors, but they are primarily red

or pink. Notably, there is extraordinary variability in the shape and size of the fruits. They range from round and small fruits to larger, heavily ribbed, segmented, and deformed ones, as well as elongated fruits with prominent locules. The fruit weight, along with its associated traits, shows significant variation. This group ranks fourth in terms of Brix degrees, following groups more similar to *S.l*. var. *cerasiforme* (**slc_ecu_per_bol** and **slc_ma**), modern cultivars of the Cherry type (**im_cherry**), and *S. pimpinellifolium* (**sp**). A large portion of this collection was gathered from local markets, mainly in Guatemala and El Salvador, in 1960. It is a highly interesting material due to its extensive morphological variability. The genetic diversity is also high, comparable to improved modern cultivars.

4. Traditional or local varieties from Spain (vint_esp)

The majority of the accessions included in this group (n = 33) show indeterminate growth habit, with a smaller portion being semi-determinate, and they exhibit predominantly irregular inflorescences. The fruits show variation in the presence and intensity of green shoulders, as well as in the color of the immature fruit. This group stands out for its extraordinary variability in terms of fruit shape, color, ribbing, and pistil scar, with an increase in the number of fruits with irregular scars. The occurrence of puffiness fruits is also observed. It is the group with the heaviest fruits with an average weight of 248.5 grams and, consequently, the traits related to fruit weight, such as the number of locules. The yield is also the highest among the groups with an approximately 3,132.1 grams of fruits per plant. However, it is one of the groups with the lowest Brix degrees values. Most accessions included in this group were collected in the 1980s and serve as a good representation of the existing variability in traditional cultivars in this country. Despite the significant morphological variability found, it is important to note the enormous loss of genetic diversity in this group, making it the group with the lowest value.

5. Traditional or local varieties from Italy (vint_ita)

This group includes 17 accessions. In contrast to the groups described earlier, this group has a considerable proportion of accessions with determinate growth plants, as well as semi-determinate, although the majority still consists of indeterminate growth. Nearly 100% of the inflorescences are uniparous. The green shoulder exhibit varying degrees of green intensity, as well as the immature fruits. The fruits are smooth or only slightly ribbed, without scars, red when ripe, and somewhat display puffiness. A notable feature of this group is the presence of elongated fruits, whether elliptical, pear-shaped, or obovoid, typical of varieties intended for industrial processing. This group represents a diverse range of collection years, from

the 1930s to the 2000s. The genetic diversity is very low and comparable to that of traditional Spanish cultivars (**vint_esp**).

6. Early improve cultivar (eim)

This group includes 73 accessions from various countries worldwide, without a common origin. Within this group, there is an increased variability in plant growth habit type, with approximately 50% consisting of determinate or semi-determinate entries. The complexity of inflorescence architecture also increases, with all of them being irregular or compound. The fruits exhibit variability in color at the immature stage and the presence/intensity of green shoulders. All fruit shapes considered in this study are present in this group, as well as the fruit color at maturity and ribbing at calix end. Additionally, a high percentage of fruits display linear or stellate-shaped pistil scars. Regarding fruit size, where the Spanish traditional cultivars (**vint_esp**) have the largest size, the accessions of this group follow in second position. Both groups (**vint_esp** and **eim**) have the lowest soluble solids content (°Brix) with an average of 4.4 and 4.6 Brix degrees, respectively. The genetic diversity remains low, although slightly higher than that of traditional Spanish (**vint_esp**) and Italian cultivars (**vint_ita**).

7. Modern improve cultivars (mim)

The group of modern varieties is the largest, comprising 166 accessions. This group encompasses accessions acquired by germplasm banks from 1950 to 2013, originating from numerous countries, developed by companies or public research centers. There is a notable increase in accessions with determinate and semi-determinate growth habit. The majority of inflorescences are uniparous, although irregular ones are also present. Unlike previous groups, fruits without green shoulders or with slight intensity prevail. In their immature state, fruits are mostly light green. The pistil scar is usually dot or linear, without any irregularities. Fruits are predominantly smooth or with slight ribbing, turning red when ripe, compact, and exhibiting various shapes, with slightly flattened, round, and elliptical forms predominating. This indicates a lower morphological variability of the fruits. The fruit weight is slightly lower compared to the early improved varieties (**eim**), with an average value of approximately 108.3 grams. It is one of the groups with higher yield and lower Brix degree content (alongside Spanish varieties (**vint_esp**), and early improve cultivar (**eim**)). The genetic diversity is higher as a result of incorporating wild species into the breeding process.

8. Modern improved varieties of the "Cherry" type (im_cherry)

This group comprises only 9 accessions, consisting of improved cultivars that share the characteristic of being "Cherry" type, although they exhibit variation in other traits. While the majority exhibit indeterminate growth habit, one of them display determinate type. The inflorescences are uniparous. Immature fruits display a light green color with varying intensity of green shoulders. They have a smooth texture and develop into compact, round, flattened, or elliptical shapes, turning red or yellow when ripe. These fruits are very small in size, characterized by a thin pericarp, medium firmness, low yield, and a higher soluble solids content (with an average of 6.8 Brix degree) compared to the **slc_ma** cultivars, albeit significantly lower than *S. pimpinellifolium* (**sp**) (with an average of 10.7 Brix degree). Additionally, this group stands out for its remarkable genetic diversity, surpassing that of modern improved cultivars.

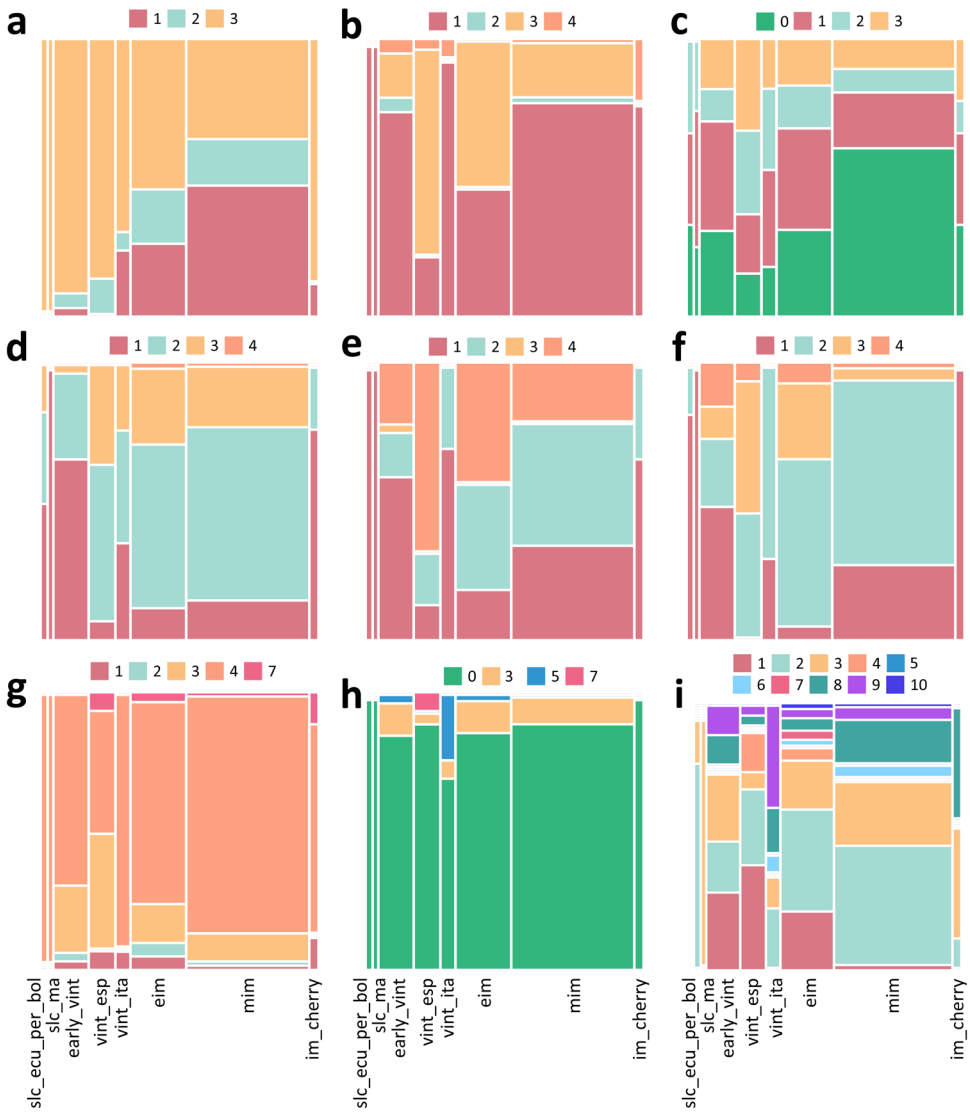


Figure 9. Distribution of qualitative morphological traits for each genetic group. **(a)** Plant growth habit [grohab] (1: Determinate; 2: Semideterminate; 3: Indeterminate), **(b)** Inflorescence type [flotype] (1: Uniparous; 2: Forked; 3: Irregular; 4: Compound), **(c)** Green shoulder [greensho] (0: Uniform; 1: Light green; 2: Medium green; 3: Dark green), **(d)** External immature fruit color [frucolim] (1: Whitish; 2: Light green; 3: Medium green; 4: Dark green), **(e)** Blossom end scar [bloshp] (1: Dot; 2. Stellate; 3: Lineal; 4: Irregular), **(f)** Ribbing at calyx end [ribcalend] (0: Smooth; 3: Weak; 5: Intermediate; 7: Strong), **(g)** External mature fruit color [frucolma] (1: Yellow; 2: Orange; 3: Pink; 4: Red; 5: Dark red; 6: Purple; 7: Brown; 8: Other), **(h)** Puffiness [puff] (0: Absent; 3: Light; 5: Intermediate; 7: Severe) and **(i)** Fruit predominat shape [frushp] (1: Very flat; 2: Slightly flat; 3: Round; 4: Oxheart; 5: Heart, 6: Rectangular; 7: Bell pepper; 8: Ellipsoid; 9: Obovoid; 10: Long pepper).

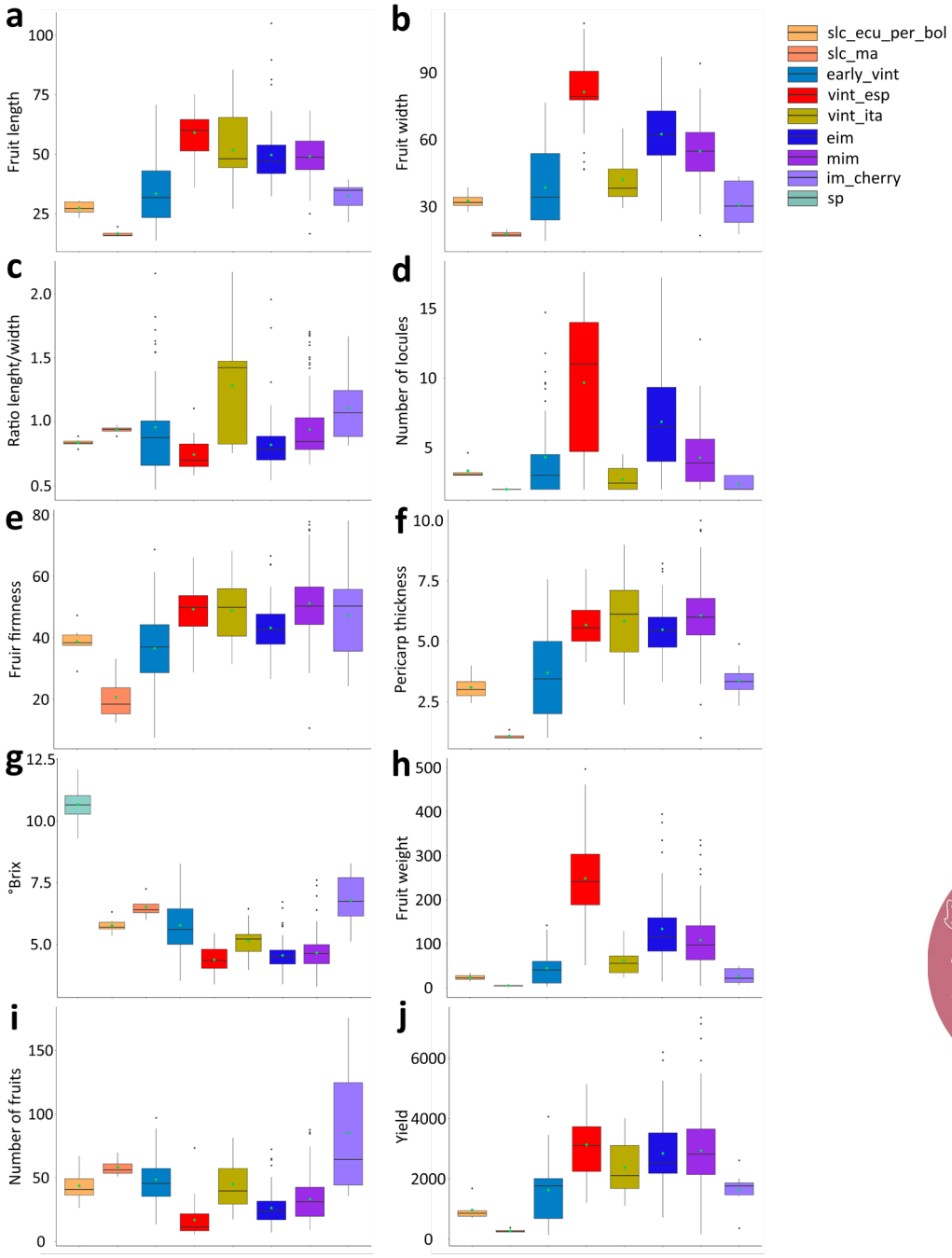


Figure 10. Boxplot of quantitative morphological traits for each genetic group. (**a**) Fruit length [frulen], (**b**) Fruit width [fruwid], (**c**) Ratio length/width [ratiolenwid], (**d**) Number of locules [nrloc], (**e**) Fruit firmness [frufirm], (**f**) Pericarp thickness [fruperi], (**g**) °Brix [brix], (**h**) Fruit weight [fruweight_mean], (**i**) Number of fruits [nr_fruits] and (**j**) Yield [yld].

Table 4. Statistical summary of the phenotypic traits evaluated in the different genetic groups. Range (minimum – maximum), Mean ± Standard deviation, and Coefficient of variation.

Trait	slc_ecu_per_bol *n*= 6	slc_ma *n*= 4	early_vint *n*= 45	vint_esp *n*= 33	vint_ita *n*= 17	eim *n*= 73	mim *n*= 166	im_cherry *n*= 9	sp* *n*= 6
frufirm	29.1 - 47.3 38.7 ± 4.9 12.7	12.4 - 33.2 20.6 ± 4.8 23.1	7.5- 68.8 36.6 ± 4.6 12.6	28.8 - 66.1 49.3 ± 5.0 10.2	31.4 - 68.3 48.9 ± 3.0 6.2	26.6 - 66.7 43.3 ± 4.6 10.7	10.6 - 77.8 51.2 ± 4.8 9.3	24.4 - 78.2 47.3 ± 2.5 5.3	
lcol	32.4 - 34.6 33.2 ± 1.1 3.2	30.9 - 33.1 31.9 ± 0.5 1.7	30.6 - 42.8 33.5 ± 1.0 3.1	31.5 - 56.8 36.5 ± 1.0 2.7	28.7 - 48.8 33.9 ± 0.6 1.8	29.9 - 48.5 34.8 ± 1.3 3.8	30.6 - 40.5 34.0 ± 0.8 2.3	28.9 - 46.3 33.0 ± 1.4 4.4	
acol	15.6 - 20.2 18.3 ± 0.9 4.8	17.3 - 22.9 20.0 ± 0.8 4.0	(-0.5) - 25.6 19.2 ± 1.2 6.3	(-3.4) - 24.4 18.9 ± 1.2 6.3	(-0.9) - 26.6 18.6 ± 2.5 13.7	(-6.5) - 25.4 19.3 ± 1.2 6.3	1.5 - 31.9 22.3 ± 1.9 8.4	5.2 - 26.5 17.7 ± 1.3 7.3	
bcol	11.1 - 13.3 11.9 ± 0.4 3.6	12.4 - 14.3 13.1 ± 0.2 1.9	8.7 - 22.6 12.6 ± 0.7 5.6	9.6 - 30.3 14.4 ± 0.8 5.5	10.1 - 24.3 13.6 ± 0.7 5.2	8.7 - 27.1 14.0 ± 1.1 7.7	8.8 - 19.9 13.7 ± 0.6 4.1	5.5 - 23.7 13.0 ± 0.7 5.1	
brix	5.3 - 6.3 5.8 ± 0.3 5.6	6.0 - 7.2 6.5 ± 0.2 2.9	3.5 - 8.3 5.8 ± 0.5 8.4	3.4 - 5.5 4.4 ± 0.2 5.7	4.0 - 6.4 5.1 ± 0.3 6.2	3.4 - 6.7 4.6 ± 0.3 6.5	3.3 - 7.6 4.7 ± 0.3 6.0	5.1 - 8.3 6.8 ± 0.6 9.1	9.3 - 12.1 10.7 ± 0.9 8.0
fruperi	2.4 - 4.0 3.1 ± 0.4 11.6	1.0 - 1.3 1.1 ± 0.3 23.1	1.0 - 7.6 3.7 ± 0.4 10.3	4.1 - 8.0 5.7 ± 0.6 10.5	2.4 - 9.0 5.8 ± 0.5 7.9	3.3 - 8.2 5.5 ± 0.5 8.9	1.0 - 10.0 6.1 ± 0.5 8.5	2.33 - 4.9 3.4 ± 0.4 11.8	
frulen	23.1 - 30.5 27.4 ± 0.5 1.8	15.6 - 19.6 16.7 ± 0.8 4.7	13.7 -70.8 33.4 ± 2.2 6.5	36.0 - 75.2 59.1 ± 2.9 4.9	27.1 - 85.6 51.7 ± 3.1 6.0	32.3 - 105.0 49.7 ± 3.0 6.0	16.6 - 68.3 49.1 ± 2.0 4.1	21.6 -39.4 32.5 ± 1.0 3.0	
fruwid	27.7 -38.7 32.5 ± 1.5 4.7	16.6 - 19.9 17.7 ± 0.4 2.4	14.7 - 76.6 38.5 ± 2.8 7.3	46.7 - 112.0 81.3 ± 3.6 4.4	29.3 - 65 42.1 ± 6.3 14.9	23.4 - 97.2 62.5 ± 3.9 6.2	17.0 - 94.1 54.8 ± 3.3 6.0	17.7 - 43.5 30.7 ± 1.3 4.2	
ratiolenwid	0.8 - 0.9 0.8 ± 0.1 7.6	0.9 - 1.0 0.9 ± 0.0 3.1	0.5 -2.2 1.0 ± 0.0 5.2	0.6 - 1.1 0.7 ± 0.0 3.9	0.8 - 2.2 1.3 ± 0.1 9.0	0.5 - 2.0 0.8 ± 0.1 6.8	0.7 - 1.7 0.9 ± 0.0 5.2	0.8 - 1.7 1.1 ± 0.0 3.8	
nrloc	3.0 - 5.0 3.3 ± 0.5 14.9	2.0 -2.0 2.0 ± 0.0 0.0	2.0 - 15.0 4.3 ± 1.0 23.1	2.0 - 18.0 9.7 ± 1.4 14.3	2.0 - 5.0 2.7 ± 0.9 32.0	2.0 -17.0 6.9 ± 1.1 15.8	2.0 - 13.0 4.3 ± 0.8 19.4	2.0 -3.0 2.4 ± 0.2 7.0	
	41.3 - 63.0	35.7 - 47.7	32.7 -62.0	44.3 - 62.0	33.0 - 65.3	33.0 - 65.0	33.0 - 81.0	33.0 -46.3	

Trait	slc_ecu_per_bol *n*= 6	slc_ma *n*= 4	early_vint *n*= 45	vint_esp *n*= 33	vint_ita *n*= 17	eim *n*= 73	mim *n*= 166	im_cherry *n*= 9	sp* *n*= 6
nr_sowtoflo	46.3 ± 12.4 26.8	40.3 ± 1.7 4.2	43.1 ± 5.9 13.8	50.3 ± 4.0 7.9	49.0 ± 7.7 15.8	47.0 ± 4.4 9.4	48.4 ± 7.1 14.7	42.2 ± 1.6 3.8	
nr_sowtoripe	84.7 - 101.0 91.1 ± 2.2 2.5	86.0 - 95.3 88.5 ± 6.0 6.7	78.3 - 105.0 90.2 ± 2.3 2.5	89.3 - 171.0 110.9 ± 10.9 9.8	81.7 - 143.0 100.2 ± 6.2 6.2	84.0 -144.0 99.2 ± 6.2 6.2	78.7 - 164.0 98.5 ± 6.2 6.3	79.0 - 89.3 86.0 ± 3.1 3.6	
nr_fruits	26.0 - 67.0 43.6 ± 7.2 16.6	51.0 -70.0 58.3 ± 8.8 15.1	13.0 - 97.0 48.6 ± 10.5 21.6	5.0 - 73.0 16.8 ± 8.4 49.8	17.0 - 81.0 45.1 ± 12.6 28.0	7.0 -72.0 26.3 ± 7.8 29.5	9.0 -88.0 33.2 ± 9.8 29.5	36.0 - 176.0 85.1 ± 17.0 20.0	
fruweight_mean	14.4 - 33.7 23.3 ± 1.8 7.7	3.7 - 6.6 4.8 ± 0.4 9.0	2.9 - 142.0 44.4 ± 9.0 20.1	50.7 - 497.0 248.5 ± 60.5 24.3	23.1 - 129.0 61.8 ± 14.8 23.9	14.4 - 394.0 133.8 ± 32.5 24.3	3.9 - 335.0 108.3 ± 21.0 19.4	6.3 - 49.5 25.5 ± 3.2 12.6	
yld	710.0 - 1,684.0 968.8 ± 216.4 22.3	220.0 - 378.0 276.3 ± 40.7 14.7	136.0 -4,064.0 1628.5 ± 506.8 31.1	1208.0 - 4,012.0 3132.1 ± 797.9 25.5	1,112.0 -4,012.0 2373.4 ± 633.6 26.7	720.0 - 6,202.0 2846.1 ± 992.8 34.9	170.0 - 7,346.0 2936.3 ± 1078.7 36.7	364.0 - 2,614.0 1645.0 ± 240.4 14.6	

***sp was evaluated only for the sugar content in Brix degree.

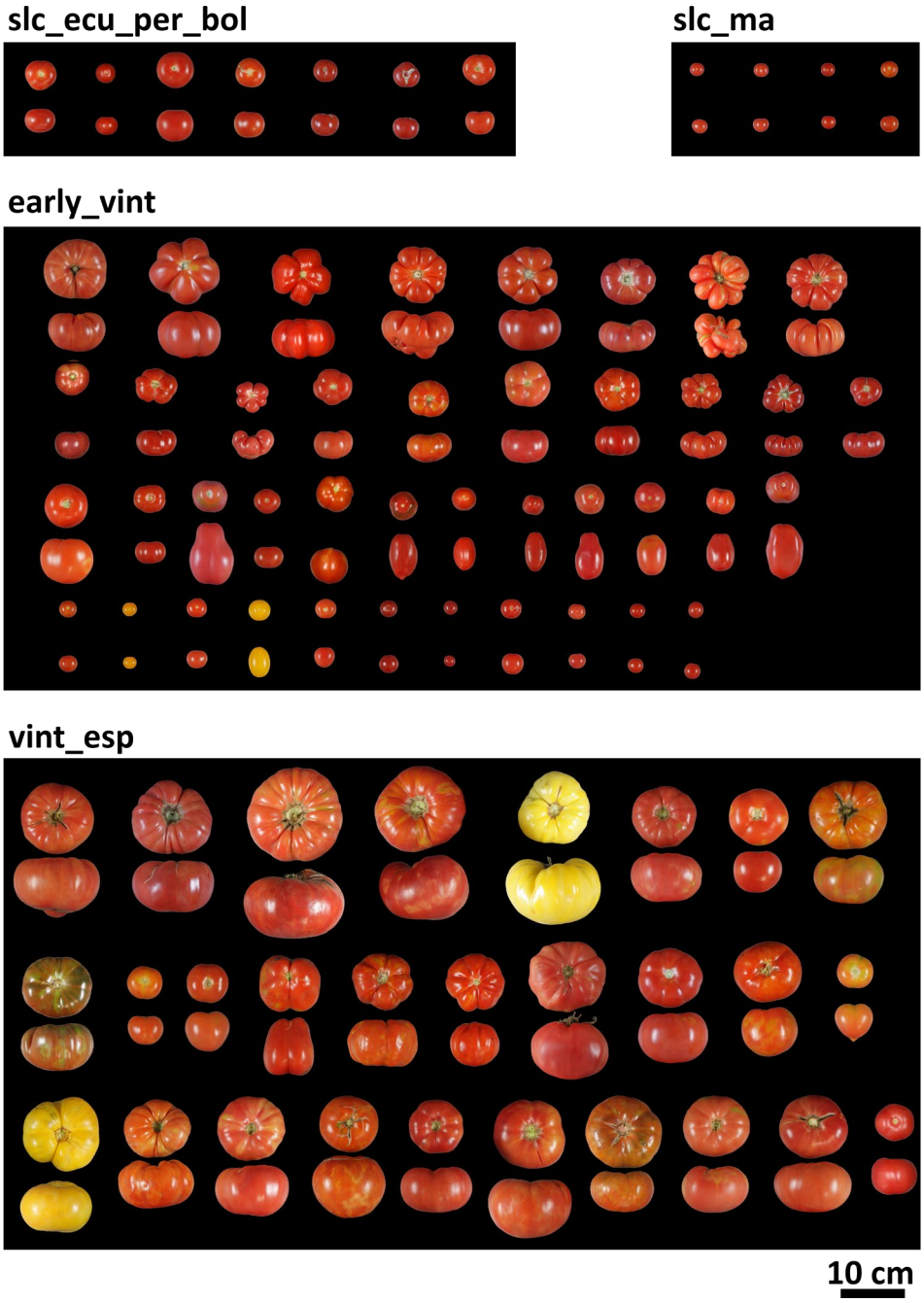


Figure 11. Morphological variation of the fruits of the accessions in the core collection according to their genetic group. The picture shows variation in fruit color, shape, and size, illustrating the variation between and within genetic groups. Scale bar is 10 cm. (Continue)

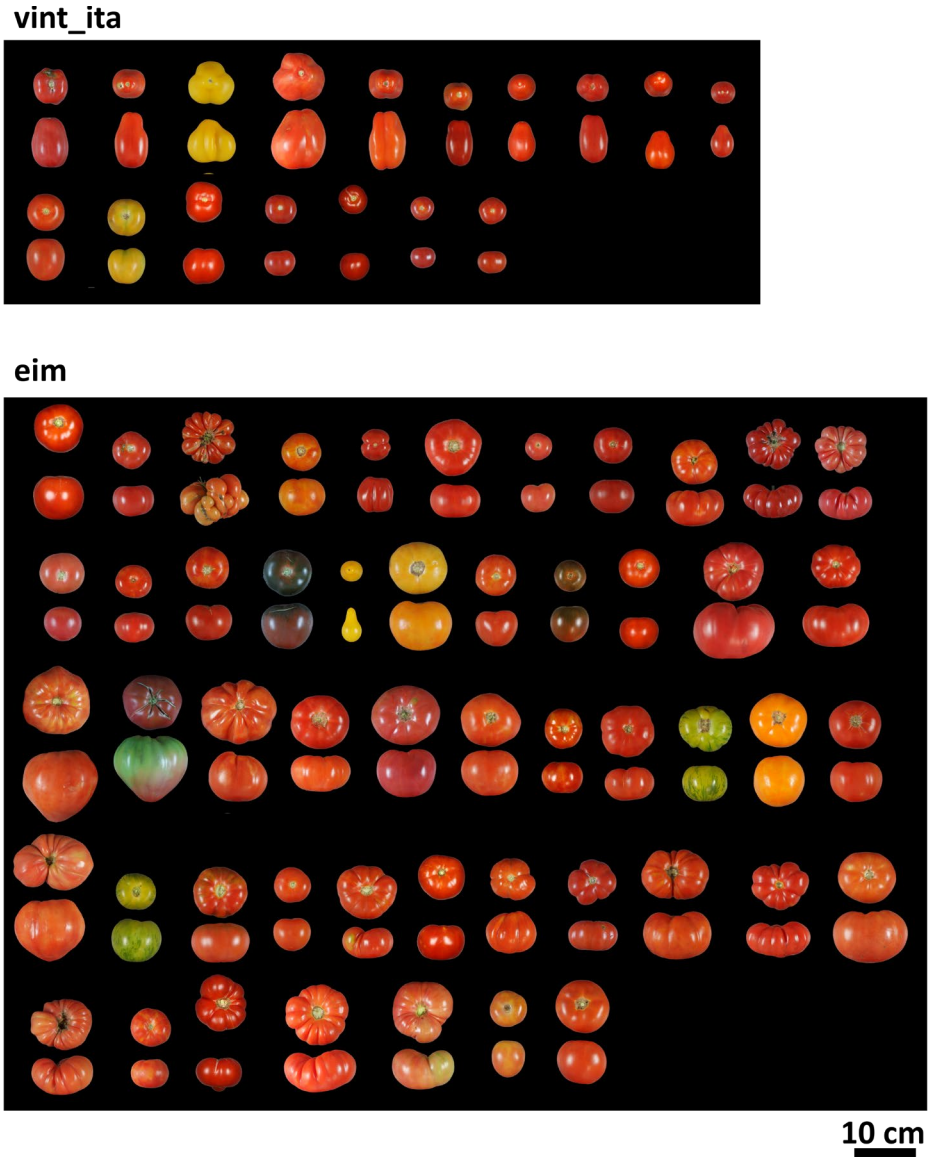


Figure 11. Morphological variation of the fruits of the accessions in the core collection according to their genetic group. The picture shows variation in fruit color, shape, and size, illustrating the variation between and within genetic groups. Scale bar is 10 cm. (Continue)

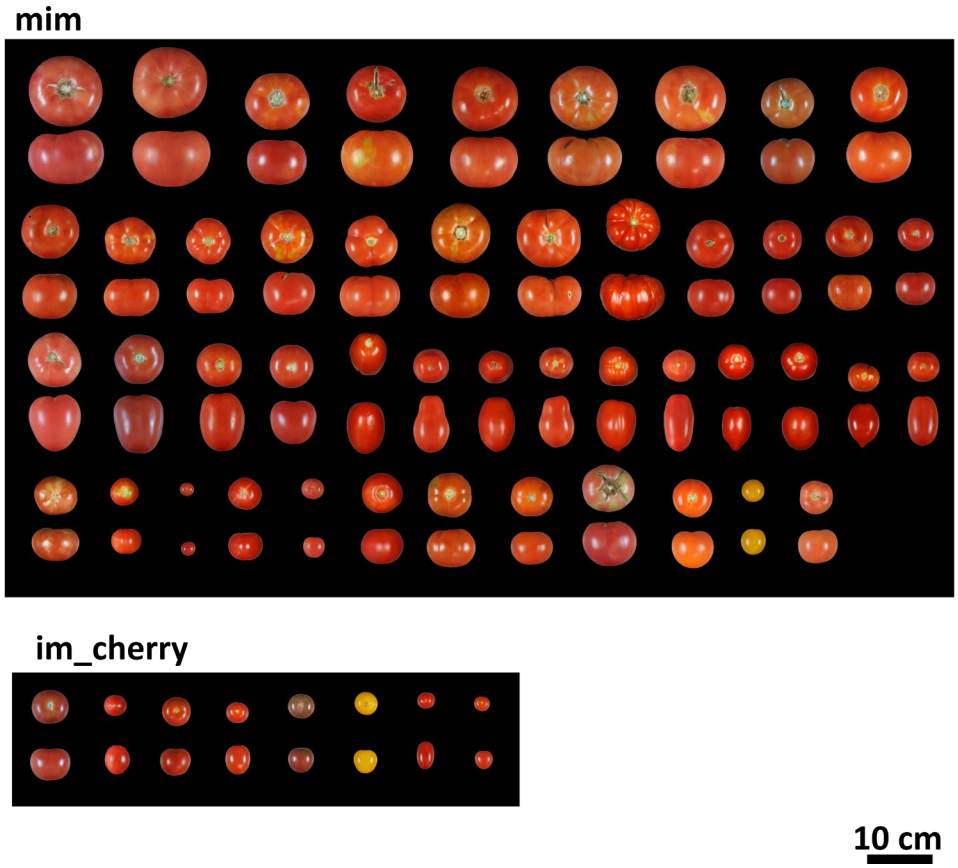


Figure 11. Morphological variation of the fruits of the accessions in the core collection according to their genetic group. The picture shows variation in fruit color, shape, and size, illustrating the variation between and within genetic groups. Scale bar is 10 cm.

Evolution of genetic diversity and morphological variation over time

The approach of this study, clustering the accessions of the core collection based on their genetic similarity complemented with passport data available in the databases, has allowed us to highlight the value of this collection, identifying the most genetically diverse accessions groups and the hotspots of phenotypic variability (**Figure 11**).

Despite the limited number of accessions in some of the established genetic groups, valuable insights can still be obtained regarding the changes in genetic diversity during the evolution from *S.l.* var. *cerasiforme* and the early tomato cultivars, their migration outside their center of origin, and the breeding efforts undertaken by breeders.

The accessions of **slc_ecu_per_bol** exhibit a high genetic diversity, comparable to *S. pimpinellifolium* (**sp**), among the accessions included in this study. This exceptional genetic richness was already demonstrated by Blanca et al., (2022a). The greatest loss of diversity occurs during the migration to Mesoamerica from the regions of Peru and Ecuador. Another significant bottleneck is the transition from America to Europe, where diversity dropped drastically. Based on these results, the importance of *S.l.* var. *cerasiforme* from Ecuador, Peru, and Bolivia, as well as the accessions considered as **early_vint**, should be highlighted. These **early_vint** materials represent the initial introduction into local markets of these tomatoes and still maintain a genetic background very similar to *S.l.* var. *cerasiforme*. As observed in the performed PCA (**Figure 2, 3**), these accessions are distinct from the central cluster of Spanish and Italian vintages (**vint_esp** and **vint_ita**, respectively), as well as from **eim** and **mim**, highlighting the different genetic background of these materials. Regarding the traditional varieties from Spain (**vint_esp**) and Italy (**vint_ita**), which are considered secondary centers of diversity, their genetic diversity is low; however, their significant morphological variability reflects a high allelic variation in genes responsible for fruit phenotypic variation and other plant traits such as growth habit, a key trait for the adaptation of tomato cultivation to mechanical harvesting in tomatoes destined for industrial processing. The modern improve varieties group (**mim**) represents genetically variable materials due to the involvement of wild species in their improvement and serves as a distinct source of variation compared to **slc_ecu_per_bol**.

The variation among the established groups for the phenotypic traits is also of great interest (**Figure 9, 10**). The **slc_ecu_per_bol** and **slc_ma** groups exhibit considerable uniformity in terms of their fruit characteristics, which are predominantly round, red, and small in size, contrasting with their high genetic diversity. The explosion of morphological variability is clearly evident in the **early_vint** group. This variability is reminiscent of the descriptions provided by botanists about the earliest tomatoes that arrived from America. Van Andel et al., (2022) conducted a meticulous research work on existing woodcuts illustration, names, and morphology of the oldest tomatoes. From this study, it becomes clear that different landraces of tomatoes were introduced early on in Europe, representing a wide variety in flower and fruit shape, size, and color, as suggested by (Daunay et al., 2006; Peralta et al., 2008). Some tomato illustrations show a deeply furrowed (segmented) fruits. These descriptions resemble the accessions GPT019700, GPT049250, GPT192580 and GPT198820 included in this group (**Figure 12**). In a book published in 1937, Boswell (1937) provided a detailed account of the early tomato improvement efforts in the United States, describing the morphology of unimproved materials as follows: '*Sixty years ago, the only large sorts were rough, ugly, heavily ribbed, variable 'varieties' of indifferent quality, although some good small ones of the greenhouse type had*

been brought in from England.'. This supports the diversity of existing forms before the beginning of the first genetic breeding efforts.

Figure 12. Morphology of deeply furrowed fruits described by Daunay et al., 2006, Peralta et al., 2008 and Van Ander et al., 2022 in the core collection. The included accessions are (left to right) GPT019700, GPT049250, GPT192580, and GPT198820.

The **vint_esp** group is another group with high morphological diversity, distinct from that found in the **early_vint** group. It may seem paradoxical that such a rich diversity of shapes, colors, sizes, uses, and other agronomic traits can be maintained with such a limited genetic diversity. In fact, Blanca et al., (2022b) found in a study conducted with 1254 varieties genotyped using GBS that only 298 SNPs were found to be polymorphic out of the 64,943 tested. According to these authors, this could be attributed to the selection carried out by traditional farmers in favor of this agronomic diversity, resulting in a lack of variation with only a few scattered polymorphic loci that confer fruit shape, size, and color adapted to the preferences of local communities. There is, therefore, a high allelic diversity for the genes controlling these traits. Moreover, farmers have incorporated further diversity into the traditional pool through the process of "traditionalization," by integrating varieties developed by breeding companies between the 18th and 21st centuries (Gentilcore, 2010). It is now widely accepted that Spanish local varieties (and also Italian traditional varieties) are the result of occasional crosses between truly traditional varieties and improved varieties, although the majority of the genome remains "traditional" due to successive backcrosses that naturally occur towards these varieties (Casañas et al., 2017). This was also observed in the study performed by Blanca et al., (2022b), where introgressions coming from wild species were found in some of the accessions considered as vintage. Similarly to the **vint_esp**, the early improve material (**eim**) are highly phenotypically variable as they have been improved through intraspecific crosses using existing heirlooms. In this previous breeding, the objectives were higher production and, above all, the round shape of the fruits, which allowed for better fruit handling and reduced losses. However, there still existed considerable morphological variability (Boswell, 1937).

In the group of improved modern cultivars (**mim**), a notable distinction can be observed between those intended for industrial processing and those cultivated for fresh consumption. This results in significant variation in traits that determine suitability for each type of use, such as growth habit, fruit shape and size, absence of green shoulders in the case of tomatoes for industrial processing, and higher pericarp thickness for fresh consumption. However, within each group, there is a certain degree of uniformity, with the most predominant type for fresh consumption being slightly flattened, medium-sized tomatoes without shoulders, aligning with the preferences of many markets. It should be noted that tomatoes of various colors have not been included in the core collection, as it is a current breeding objective to diversify the offerings and attract new markets.

Interest of the core collection in breeding

The core collection exhibited a high level of diversity across all investigated traits, as demonstrated for fruit size, color, and shape in **Figure 9, 11** and the coefficients of variation in **Table 4**. The complete core collection is currently undergoing re-sequencing, which will enable the detection of mutations or allelic variants in genes that determine fruit traits, such as the fasciated (*fas*) gene and locule number (*lc*), which influence both the final size and shape of the fruit (Rodríguez et al., 2011). Furthermore, it will allow the evaluation of the major loci involved in fruit size control, namely *fw1.1*, *fw2.1*, *fw2.2*, *fw3.1*, *fw3.2*, and *fw11.3*, which collectively account for approximately 67% of the total phenotypic variation, resulting in size changes with minimal impact on shape (Lippman and Tanksley, 2001). Similarly, three major loci (*ovate*, *sun*, and *fs8.1*) modulate fruit shape with minimal effect on size (Rodríguez et al., 2011). *Ovate* and *sun* lead to the development of elongated or pear-shaped fruits, while *fs8.1* contributes to increased fruit length by augmenting the number of cells in the proximal-distal direction. The core collection also exhibits significant diversity in fruit color (**Figure 11**), which is attributed to various mutations encountered during the domestication and improvement of this crop. These mutations have been characterized and are located in genes involved in carotenoid or flavonoid biosynthesis or chlorophyll degradation. For instance, the yellow flesh color is determined by the *yellow flesh* (*r*) gene (Kachanovsky et al., 2012), the orange color by the *tangerine* (*t*) gene (Isaacson et al., 2002), the green color by the *green flesh* (*gf*) gene (Barry et al., 2008), and the deep red color by the *old gold crimson* (og^c) gene (Ronen et al., 2000).

In addition to the existing variation in fruit traits, other traits related to plant architecture, inflorescences, and yield are also highly important. The variation observed in plant architecture and earliness allows plants to adapt to different environments and cultivation systems. For example, it enables the cultivation of very early varieties with clustered flowering and fruiting in cold environments where the tomato growing season is limited. The type of inflorescences also plays a crucial role, as it, along with fruit size, determines the number of fruits that can reach commercial maturity, a trait closely correlated with yield. There is a significant range of variability in these traits among the studied varieties, which can be utilized in breeding programs.

As indicated in the materials and methods section, the core collection included accessions that were specifically proposed by researchers due to their diverse characteristics of interest. These entries were selected based on their tolerance to abiotic stresses such as salinity, heat, cold, resistance to various viral, bacterial, or fungal diseases, as well as traits related to fruit quality such as firmness, high °Brix content, β-carotene, lycopene, ascorbic acid, and high yield potential. Additionally,

accessions were included for their relevance in studies of the demographic history of tomato. This increases the value of this collection.

The tomato breeding continues to rely on the utilization of wild species due to the absence of resistance genes and the lack of adaptation to abiotic stresses in cultivated varieties. This approach will remain necessary due to the continuous emergence of diseases (Salem et al., 2023) and the progressive desertification of cultivation areas. However, there is also a clear trend towards expanding the genetic base of modern tomatoes through introgressions of favourable traits from old cultivars and traditional varieties. The loss of fruit quality in commercial hybrids has compelled companies to turn to old varieties in search of flavour and all the associated traits. The availability of phenotypic and genotypic data from the studied collection, which largely consists of these types of cultivars, represents a source of genes of extraordinary interest for further tomato improvement in many aspects.

REFERENCES

Aflitos, S., Schijlen, E., De Jong, H., De Ridder, D., Smit, S., Finkers, R., et al. (2014). Exploring genetic variation in the tomato (*Solanum* section *Lycopersicon*) clade by whole-genome sequencing. *Plant J.* 80, 136–148.

Bai, Y., and Lindhout, P. (2007). Domestication and breeding of tomatoes: what have we gained and what can we gain in the future? *Ann. Bot.* 100, 1085.

Barchi, L., Acquadro, A., Alonso, D., Aprea, G., Bassolino, L., Demurtas, O., et al. (2019). Single Primer Enrichment Technology (SPET) for high-throughput genotyping in tomato and eggplant germplasm. *Front. Plant Sci.* 10, 1005.

Barry, C. S., McQuinn, R. P., Chung, M.-Y., Besuden, A., and Giovannoni, J. J. (2008). Amino acid substitutions in homologs of the STAY-GREEN protein are responsible for the green-flesh and chlorophyll retainer mutations of tomato and pepper. *Plant Physiol.* 147, 179–187.

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

Blanca, J., Montero-Pau, J., Sauvage, C., Bauchet, G., Illa, E., Díez, M. J., et al. (2015). Genomic variation in tomato, from wild ancestors to contemporary breeding accessions. *BMC Genomics* 16, 257.

Blanca, J., Sanchez-Matarredona, D., Ziarsolo, P., Montero-Pau, J., van der Knaap, E., Díez, M. J., et al. (2022a). Haplotype analyses reveal novel insights into tomato history and domestication driven by long-distance migrations and latitudinal adaptations. *Hortic. Res.* 9, 1–14.

Blanca, J., Pons, C., Montero-Pau, J., Sanchez-Matarredona, D., Ziarsolo, P., Fontanet, L., et al. (2022b). European traditional tomatoes galore: a result of farmers' selection of a few diversity-rich loci. *J. Exp. Bot.* 73, 3431–3445.

Boswell, V. R. (1937). Improvement and genetics of tomatoes, peppers and eggplant. *Improv. Genet. tomatoes, peppers eggplant.*

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633–2635.

Canady, M. A., Meglic, V., and Chetelat, R. T. (2005). A library of *Solanum lycopersicoides* introgression lines in cultivated tomato. *Genome.* 48, 685–697.

Casals, J., Rivera, A., Sabaté, J., del Castillo, R. R., and Simó, J. (2018). Cherry and fresh market tomatoes: differences in chemical, morphological, and sensory traits and their implications for consumer acceptance. *Agron.* 9(1), 9.

Casañas, F., Simó, J., Casals, J., and Prohens, J. (2017). Toward an evolved concept of landrace. *Front. Plant Sci.* 8, 145.

Chetelat, R. T., Qin, X., Tan, M., Burkart-Waco, D., Moritama, Y., Huo, X., et al. (2019). Introgression lines of *Solanum sitiens* , a wild nightshade of the Atacama Desert, in the genome of cultivated tomato. *Plant J.* 100, 836–850.

Comtois, D. (2019). Summary tools: tools to quickly and neatly summarize data R package version 0.9.4.

Cong, B., Barrero, L. S., and Tanksley, S. D. (2008). Regulatory change in *YABBY-like* transcription factor led to evolution of extreme fruit size during tomato domestication. *Nat. Genet.* 40, 800–804.

Daunay, M.-C., Laterrot, H., and Janick, J. (2006). Iconography of the Solanaceae from antiquity to the XVIIth century: A rich source of information on genetic diversity and uses. In: *VI International*

Solanaceae Conference: Genomics Meets Biodiversity 745, 59–88.

Esposito, S., Cardi, T., Campanelli, G., Sestili, S., Díez, M. J., Soler, S., et al. (2020). ddRAD sequencing-based genotyping for population structure analysis in cultivated tomato provides new insights into the genomic diversity of Mediterranean 'da serbo' type long shelf-life germplasm. *Hortic. Res.* 7, 134.

FAOSTAT. (2021). FAOSTAT database collections.

Frary, A., Nesbitt, T. C., Frary, A., Grandillo, S., Van Der Knaap, E., Cong, B., et al. (2000). *fw2.2*: A quantitative trait locus key to the evolution of tomato fruit size. *Science.* 289, 85–88.

Fridman, E., Carrari, F., Liu, Y. S., Fernie, A. R., and Zamir, D. (2004). Zooming in on a quantitative trait for tomato yield using interspecific introgressions. *Science.* 305, 1786–1789.

Fridman, E., Liu, Y. S., Carmel-Goren, L., Gur, A., Shoresh, M., Pleban, T., et al. (2002). Two tightly linked QTLs modify tomato sugar content via different physiological pathways. *Mol Genet Genomics.* 266, 821–826.

Gao, L., Gonda, I., Sun, H., Ma, Q., Bao, K., Tieman, D. M., et al. (2019). The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. *Nat. Genet.* 51, 1044–1051.

García-Martínez, S., Andreani, L., Garcia-Gusano, M., Geuna, F., and Ruiz, J. J. (2006). Evaluation of amplified fragment length polymorphism and simple sequence repeats for tomato germplasm fingerprinting: Utility for grouping closely related traditional cultivars. *Genome* 49, 648–656.

Gentilcore, D. (2010). Pomodoro!: a history of the tomato in Italy. In: Pomodoro!. *Columbia University Press*.

Gimeno-Páez, E., Prohens, J., Moreno-Cerveró, M., Luis-Margarit, A. de, Díez, M. J., and Gramazio, P. (2023). Agronomic treatments combined with embryo rescue for rapid generation advancement in tomato speed breeding. B*ioRxiv.*

IPGRI (1996). Descriptors for Tomato (*Lycopersicum spp.*). *Bioversity Internationa*l. 21.

Isaacson, T., Ronen, G., Zamir, D., and Hirschberg, J. (2002). Cloning of tangerine from tomato reveals a carotenoid isomerase essential for the production of β-Carotene and xanthophylls in plants. *Plant Cell* 14, 333–342.

Jerome, G., and Thibaut, J. (2022). Hierfstat: Estimation and tests of hierarchical F-statistics.

Kachanovsky, D. E., Filler, S., Isaacson, T., and Hirschberg, J. (2012). Epistasis in tomato color mutations involves regulation of phytoene synthase 1 expression by cis-carotenoids. *Proc. Natl. Acad. Sci. U. S. A.* 109, 19021–19026. doi: 10.1073/PNAS.1214808109/SUPPL_FILE/PNAS.201214808SI.PDF.

Kamvar, Z. N., Brooks, J. C., and Grunwald, N. J. (2015). Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Front. Genet.* 6, 208.

Kamvar, Z. N., and Grünwald, N. J. (2023). Analysis of Multilocus Genotypes and Lineages in poppr 2.9.3.

Lin, T., Zhu, G., Zhang, J., Xu, X., Yu, Q., Zheng, Z., et al. (2014). Genomic analyses provide insights into the history of tomato breeding. *Nat. Genet.* 46, 1220–1226.

Lippman, Z., and Tanksley, S. D. (2001). Dissecting the genetic pathway to extreme fruit size in tomato using a cross between the small-fruited wild species *Lycopersicon pimpinellifolium* and *L. esculentum* var. giant heirloom. *Genetics* 158, 413–422.

Liu, J., Van Eck, J., Cong, B., and Tanksley, S. D. (2002). A new class of regulatory genes underlying the cause of pear-shaped tomato fruit. *Proc. Natl. Acad. Sci.* 99, 13302–13306.

Mata-Nicolás, E., Montero-Pau, J., Gimeno-Paez, E., Garcia-Carpintero, V., Ziarsolo, P., Menda, N., et al. (2020). Exploiting the diversity of tomato: the development of a phenotypically and genetically detailed germplasm collection. *Hortic. Res.* 7.

Mata, M. C. S., Hurtado, M. C., Ripollés, S. R., Galiana-Balaguer, L., Isasa, M. E. T., and Viñals, F.

N. (2000). Breeding for flavour of fresh market tomato: sources for increasing acid content. *Acta Physiol. Plant.* 22, 250–253.

Miller, J. C., and Tanksley, S. D. (1990). RFLP analysis of phylogenetic relationships and genetic variation in the genus Lycopersicon. *Theor. Appl. Genet.* 80, 437–448.

Muños, S., Ranc, N., Botton, E., Bérard, A., Rolland, S., Duffé, P., et al. (2011). Increase in tomato locule number is controlled by two single-nucleotide polymorphisms located near *WUSCHEL*. *Plant Physiol.* 156, 2244–2254.

Nei, M. (1987). Molecular evolutionary genetics. *Columbia university press.*

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

Pons, C., Casals, J., Palombieri, S., Fontanet, L., Riccini, A., Rambla, J. L., et al. (2022). Atlas of phenotypic , genotypic and geographical. *Hortic. Res.* 9.

Razifard, H., Ramos, A., Della Valle, A. L., Bodary, C., Goetz, E., Manser, E. J., et al. (2020). Genomic evidence for complex domestication history of the cultivated tomato in Latin America. *Mol. Biol. Evol.* 37, 1118–1132.

Rick, C. M. (1978). The tomato. *Sci. Am.* 239, 76–89.

Rick, C. M., and Holle, M. (1990). Andean *Lycopersicon esculentum* var. *cerasiforme*: genetic variation and its evolutionary significance. *Econ. Bot.* 44, 69–78.

Rodríguez, G. R., Muños, S., Anderson, C., Sim, S. C., Michel, A., Causse, M., et al. (2011). distribution of *SUN, OVATE, LC*, and *FAS* in the tomato germplasm and the relationship to fruit shape diversity. *Plant Physiol.* 156, 275–285.

Ronen, G., Carmel-Goren, L., Zamir, D., and Hirschberg, J. (2000). An alternative pathway to β-carotene formation in plant chromoplasts discovered by map-based cloning of *Beta* and *old-gold color* mutations in tomato. *Proc. Natl. Acad. Sci.* 97, 11102–11107.

Roohanitaziani, R., de Maagd, R. A., Lammers, M., Molthoff, J., Meijer-Dekens, F., van Kaauwen, M. P. W., et al. (2020). Exploration of a resequenced tomato core collection for phenotypic and genotypic variation in plant growth and fruit quality traits. *Genes.* 11, 1278.

Salem, N. M., Jewehan, A., Aranda, M. A., and Fox, A. (2023). Tomato Brown Rugose Fruit Virus Pandemic. *Annu. Rev. Phytopathol.* 61.

Schouten, H. J., Tikunov, Y., Verkerke, W., Finkers, R., Bovy, A., Bai, Y., et al. (2019). Breeding has increased the diversity of cultivated tomato in The Netherlands. *Front. Plant Sci.* 10.

Sievert, C. (2020). Interactive web-based data visualization with R, plotly, and shiny. *CRC Press.*

Sim, S. C., Robbins, M. D., Van Deynze, A., Michel, A. P., and Francis, D. M. (2011). Population structure and genetic differentiation associated with breeding history and selection in tomato (*Solanum lycopersicum* L.). *Heredity.* 106, 927–935.

Sim, S. C., van Deynze, A., Stoffel, K., Douches, D. S., Zarka, D., Ganal, M. W., et al. (2012). High-Density SNP genotyping of tomato (*Solanum lycopersicum* L.) reveals patterns of genetic variation due to breeding. *PLoS One.* 7, e45520.

Smith, P. G. (1944). Embryo culture of a tomato species hybrid. *Proc. Amer. Sco. Hort. Sci.* 44, 413–416.

Tam, S. M., Mhiri, C., Vogelaar, A., Kerkveld, M., Pearce, S. R., and Grandbastien, M. A. (2005). Comparative analyses of genetic diversities within tomato and pepper collections detected by retrotransposon-based SSAP, AFLP and SSR. *Theor. Appl. Genet.* 110, 819–831.

Tanksley, S. D., and McCouch, S. R. (1997). Seed banks and molecular maps: Unlocking genetic potential from the wild. *Science.* 277, 1063–1066.

Tripodi, P., Soler, S., Campanelli, G., Díez, M. J., Esposito, S., Sestili, S., et al. (2021). Genome wide association mapping for agronomic, fruit quality, and root architectural traits in tomato under organic farming conditions. *BMC Plant Biol.* 21, 481.

van Andel, T., Vos, R. A., Michels, E., and Stefanaki, A. (2022). Sixteenth-century tomatoes in Europe: who saw them, what they looked like, and where they came from. *PeerJ* 10, e12790.

van der Beek, J. G., Verkerk, R., Zabel, P., and Lindhout, P. (1992). Mapping strategy for resistance genes in tomato based on RFLPs between cultivars: *Cf9* (resistance to *Cladosporium fulvum*) on chromosome 1. *Theor. Appl. Genet.* 84, 106–112.

van Hintum, T. J. L., Brown, A. H. D., and Spillane, C. (2000). *Core collections of plant genetic resources*. Bioversity International.

Villand, J., Skroch, P. W., Lai, T., Hanson, P., Kuo, C. G., and Nienhuis, J. (1998). Genetic variation among tomato accessions from primary and secondary centers of diversity. *Crop Sci.* 38, 1339–1347.

Wei, T., Simko, V., Levy, M., Xie, Y., Jin, Y., Zemla, J., et al. (2021). R package "corrplot": Visualization of a Correlation Matrix (Version Version 0.92).

Wickham, H. (2016). Elegant Graphics for Data Analysis: ggplot2.

Xiao, H., Jiang, N., Schaffner, E., Stockinger, E. J., and Van Der Knaap, E. (2008). A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. *Science (80-.).* 319, 1527–1530.

Xu, J., Ranc, N., Muños, S., Rolland, S., Bouchet, J. P., Desplat, N., et al. (2013). Phenotypic diversity and association mapping for fruit quality traits in cultivated tomato and related species. *Theor. Appl. Genet.* 126, 567–581.

Young, H. P., West, M. A. L., and St. Clair, D. A. (2004). Evaluation of AFLPs for germplasm fingerprinting and assessment of genetic diversity in cultivars of tomato (*Lycopersicon esculentum* L.). *Genome* 47, 510–518.

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., and Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 28, 3326–3328.

Zhu, G., Wang, S., Huang, Z., Zhang, S., Liao, Q., Zhang, C., et al. (2018). Rewiring of the fruit metabolome in tomato breeding. *Cell* 172, 249-261.

Zsögön, A., Čermák, T., Naves, E. R., Notini, M. M., Edel, K. H., Weinl, S., et al. (2018). De novo domestication of wild tomato using genome editing. *Nat. Biotechnol.* 36, 1211–1216.

Zuriaga, E., Blanca, J. M., Cordero, L., Sifres, A., Blas-Cerdán, W. G., Morales, R., et al. (2009). Genetic and bioclimatic variation in *Solanum pimpinellifolium*. *Genet. Resour. Crop Evol.* 56, 39–51.

GENERAL DISCUSSION

Those who cannot remember the past are condemned to repeat it.
Jorge Ruíz de Santayana

Over the past few decades, the impact of climate change on horticultural crops has become increasingly evident. The loss of genetic diversity and the erosion of genetic resources pose significant challenges for crop improvement, demanding the exploration of novel plant genetic resources and the development of associated technologies to expedite the breeding process (Díez et al., 2018). Recently, there have been remarkable advancements in the application of molecular tools to mitigate the effects of climate change and enhance our understanding of horticultural species. These tools have allowed researchers to assess both the phenotypic and genetic value of plant materials, providing valuable insights for crop improvement efforts. However, several challenges still persist. Issues such as duplicate storage of germplasm, insufficient documentation of the stored material and seed exchange have been raised as concerns that must be addressed as they hamper the efforts to conduct comprehensive genetic studies. Moreover, limited funding poses a barrier to conducting more extensive genetic investigations.

The objectives of this study were twofold. Firstly, it aimed to inventory and analyze the main solanaceous collections within the frame of the G2P-SOL project. The focus was gathering and examining the available phenotypic and passport data from the germplasm banks (**Chapter I**), aiming to establish a core collection for each crop that encompasses maximum genetic and phenotypic diversity (**Chapter IV**). To achieve this, a universal, rapid, and cost-effective DNA extraction method was developed and optimized for high-throughput genotyping platforms (**Chapter II**). Additionally, to enhance the value of the germplasm collections, the study explored and optimized the use of Single Primer Enrichment Technology (SPET) genotyping (**Chapter III**). This technique provided valuable insights into the diversity of the collections and facilitated the discrimination of accessions within germplasm banks. SPET technology emerged as a powerful tool for identifying duplicates, detecting mislabelled samples, and identifying potential errors during regeneration and storage processes.

The findings presented in this study shed light on critical aspects related to germplasm collections and provide valuable knowledge for future endeavours (Vos and Bellù, 2019). By comprehensively assessing the available materials stored in germplasm banks, the study emphasizes their crucial role as reservoirs of genetic diversity, serving as the foundation for genetic breeding initiatives. In conclusion, this research contributes to the ongoing efforts aimed at addressing the challenges posed by climate change and the need for sustainable horticultural crop improvement. By elucidating the issues surrounding duplicate storage, seed exchange, and genetic characterization, this study provides valuable insights to support future work in the field. The materials preserved in germplasm banks play a pivotal role in genetic conservation and hold immense potential for advancing the genetic improvement of horticultural crop.

G2P-SOL. ***Inventory and gathering of available passport and phenotypic data from plant material stored in the most relevant germplasm banks of the major solanaceous crops: tomato, pepper, and eggplant.***

G2P-SOL is a collaborative research initiative that brings together the leading European and international repositories hosting germplasm collections of the main solanaceous crops, including potato, tomato, pepper, and eggplant. This project represents the first comprehensive effort to examine such a large number of accessions, providing a worldwide sampling and a comprehensive view of the diversity and history of the main Solanaceous crops (Alonso et al., 2018). The high number of accessions stored in these germplasm banks highlights the importance of preserving and studying the genetic resources (Mascher et al., 2019). However, one of the challenges encountered during this inventory was the lack of data from a significant portion of the accessions. Many accessions needed proper characterization in terms of both phenotypic and genotypic data. This limited accessibility of reliable characterization data hampers the comprehensive understanding of the genetic diversity and traits present within the collections. These findings are similar to other studies conducted across numerous accessions, such as the one conducted by Milner et al., (2019), where a barley collection was genotyped and combined with insufficient and analogic passport data, available phenotypic data, and newly collected phenotypic data to dissect changes and selection of genes during crop evolution.

Furthermore, the presence of duplications between germplasm banks was observed to a considerable extent based on passport data. Duplicate accessions not only hinder the accurate assessment of the true diversity within the collections but also raise concerns regarding the authenticity and reliability of the available data. It is crucial to address these issues of duplications in order to ensure the accuracy and integrity of the germplasm collections. These findings have been supported by genomic data, as demonstrated by Tripodi et al., (2021) in pepper crop, where they identified 1,618 duplicate accessions within and between genebanks. The study also revealed that taxonomic ambiguity and misclassification often involve interspecific hybrids that pose challenges for morphological classification. Furthermore, additional unpublished studies in eggplant (Barchi et al., 2023, manuscript under review) and tomato (Blanca et al., 2023, manuscript in preparation) have also provided evidence of the existence of a high rate of duplicates and misclassification. Molecular passport data can reduce the workload through the exclusion of duplicates (Mascher et al., 2019).

Another challenge encountered during the compilation of data was the lack of consensus and agreement among the available characterizations (Gotor et al., 2008;

Figàs et al., 2018). Even when using universal descriptors, discrepancies were observed in the characterization data of various accessions. This lack of agreement can be attributed to variations in the methodologies employed, differences in the criteria used for characterizing traits, and the subjective interpretation of descriptors (Brewer et al., 2006; Figàs et al., 2018). Standardization and harmonization of characterization protocols would greatly enhance the comparability and reliability of the available data. The available data, although limited, offer insights into the genetic diversity and trait variations within the solanaceous crops. These data can be utilized to identify accessions with specific traits of interest, assess their potential for genetic enhancement, and guide breeding programs aiming to develop improved varieties.

In conclusion, the inventory and compilation of passport and phenotypic data of solanaceous germplasm have revealed the immense genetic diversity present in these collections. However, the challenges of limited data availability, duplications, mislabelling, and lack of consensus in characterizations highlight the need for continued efforts in data collection, curation, and standardization. By addressing these challenges, we can maximize the utilization of solanaceous genetic resources and enhance our understanding of the diversity and potential for crop improvement.

SILEX. ***Development and implementation of high molecular weight and high-quality genomic DNA extraction method for Next-Generation Sequencing platforms***

The availability of high-quality genomic DNA in sufficient concentration can sometimes be a limitation, especially for third-generation sequencing platforms (Pucker et al., 2022). There is a wide range of DNA extraction methods available, from commercial kits to homemade protocols (Vaillancourt and Buell, 2019). However, the cost of these kits and often their low yield and weak DNA quality for NGS platforms, as well as the use of toxic substances in other protocols, have led to the development of a fast and high-quality DNA extraction method, as shown in **Chapter II** of this thesis.

The SILEX protocol for genomic DNA extraction offers several advantages over other methods commonly used in molecular biology research. One of its main advantages is the use of common and inexpensive reagents, which makes it a cost-effective option. The total cost of reagents and consumables employ in the SILEX protocol is only 0.12 € per sample.

In contrast to certain protocols that utilize toxic chaotropic salts at high concentrations (Huang et al., 2000; Li et al., 2010; Li & Sheen, 2012; Rogstad, 2012),

which can inhibit enzymatic reactions relevant to NGS applications (Rana et al., 2019), the SILEX protocol utilizes non-toxic reagents. Specifically, the SILEX protocol employs a binding buffer composed of NaCl and PEG, which aids in the binding of DNA to the silica surface.

The effectiveness of the SILEX protocol has been demonstrated through extensive testing on samples from different species and the widespread adoption is evident through its successful application across various genomic applications. Research groups have validated the protocol in numerous studies involving different plant species, including snap bean (García-Fernández et al., 2022), cannabis (Galán-Ávila et al., 2021), pear (Simionca Mărcășan et al., 2023), almond (D'Amico-Willman et al., 2022) and the algae *Alaria esculenta* (Inaba et al., 2022). The quality and quantity parameters obtained using SILEX are comparable to those achieved with commercial kits, even when recalcitrant species are involved. Recalcitrant species often present challenges during DNA extraction due to the presence of polysaccharides and phenols, which can hinder downstream molecular analyses. In addition to its effectiveness in reducing contaminants, the SILEX protocol also adapts well to different sample types. While DNA extraction from fresh leaf tissue is the standard practice, there are instances where other types of material, such as fresh or freeze-dried fruit, need to be extracted. The SILEX protocol has demonstrated flexibility in successfully extracting high quantities of DNA from lyophilized and fresh fruit tissues. This adaptability expands the protocol's applicability to a broader range of species, type of samples and research areas (Inaba et al., 2022).

The suitability of SILEX-extracted DNA for genotyping applications, such as SSRs (Single Sequence Repeats), HRM (High-Resolution Melting), Nanopore sequencing and GBS among others, has been extensively evaluated (Barchi et al., 2019; Gramazio et al., 2020; Arrones et al., 2022; Mangino et al., 2022). These findings demonstrate the robustness of the SILEX protocol and its ability to yield high-quality DNA suitable for different genotyping platforms.

The simplicity of the SILEX protocol is another key advantage. The extraction process is straightforward and can be performed manually within a short time frame. For multiple simultaneous manual extractions, each sample requires less than 2 minutes per person. This efficiency is beneficial for large-scale studies where a considerable number of samples need to be processed. However, the automation and simultaneous multiplexing of 96-well plates, as seen in other extraction methods, should be considered for implementation in the near future.

In conclusion, the SILEX protocol offers a cost-effective, efficient, and versatile method for DNA extraction from various plant species and sample types. Its use of

common and inexpensive reagents, simplicity, and compatibility with different genotyping platforms make it a valuable tool for molecular biology research. With its wide applicability, robustness, and compatibility with advanced sequencing technologies, the SILEX protocol has the potential to significantly impact genetic research and contribute to a better understanding of plant genomes. As an example of the aforementioned, thousands of DNA samples from tomato and eggplant, as well as its wild relatives, were used to evaluated the efficiency and robustness of the genotyping technique (SPET) in **Chapter III** and **IV** of this Thesis, thus validating the methodology and the suitability of the extracted DNA.

SPET. ***Implementation and proof of concept of the SPET sequencing technique in horticultural crops, specifically tomato and eggplant.***

The SPET technique has been used in biomedical applications (Scolnick et al., 2015; Nairismägi et al., 2016; Wei et al., 2020) and genotyping in plants (Scaglione et al., 2019; Gonzalo et al., 2022), but its performance for characterizing large germplasm sets from crop plants was unknown. Furthermore, the implementation of this technique has led to the development of other applications based on this technology, such as its use as a method to study other organisms like the microbiota (Benjamino et al., 2021). The study presented in **Chapter III** of this thesis aimed to evaluate the robustness and reliability of the SPET technique and explore the diversity, heterozygosity, and genetic relationships within the germplasm included in the study.

The results showed that SPET is a robust method that performs well with DNA samples prepared by different laboratories using different protocols. The number of reads and mapping percentage remained relatively stable even with different sample extraction protocols. This result validates the reliability of the SILEX method (Vilanova et al., 2020) shown in **Chapter II**. The technique also exhibited a very low level of missing data, indicating its suitability for genotyping diverse plant species (Baccichet et al., 2022). It provided a comparable analysis of single nucleotide polymorphisms (SNPs) similar to genotyping arrays (Acquadro et al., 2017), while also allowing for complexity reduction typical of genotyping-by-sequencing (GBS) approaches. SPET enabled the discovery of thousands of novel SNPs not originally included in the panel. These results are consistent with other studies conducted on different species, such as apricot (Baccichet et al., 2022), where a panel of 25K target SNPs yielded a total of 32,492 polymorphic sites.

Tomato and eggplant, both subjected to human selection, have experienced a reduction in genetic variability (Blanca et al., 2012a; Cericola et al., 2013). The panels designed for this study allowed the discovery of additional non-target SNPs (7,427 and 26,103 non-target SNPs in tomato and eggplant, respectively), indicating

the technique's potential to identify novel polymorphisms even in gene pools that underwent severe bottlenecks during domestication, migration, and selection. Similar results have been obtained in Baccichet et al., (2022).

The study also analyzed the heterozygosity of the accessions. Most accessions of tomato and eggplant displayed low heterozygosity (0.67% and 0.65%, repectively), consistent with previous reports (Acquadro et al., 2017; Aflitos et al., 2014; Sim et al., 2012b). However, a few accessions showed higher heterozygosity, likely reflecting recent hybridization events. In the case of *S. pimpinellifolium*, the closest wild relative of *S. lycopersicum*, an intermediate level of heterozygosity (4.7 %) was observed, indicating its genetic position between *S. lycopersicum* and other self-incompatible wild species (Chen and Tanksley, 2004). These results are in line with similar results obtained in Blanca et al., (2022) using different genotyping platform approach. A similar outcome was observed in eggplant, where two additional cultivated species, *S. aethiopicum* and *S. macrocarpon*, exhibited high heterozygosity (1.9% and 2.5%, respectively). This can be attributed to the lesser extent of human-driven selection on these species and their higher rate of allogamy (Daunay et al., 2001, 2019). These findings show a slight variation compared to the previous report by Acquadro et al., (2017) using the RAD-sequencing technique.

Genetic relationships among the accessions were explored through phylogenetic trees and PCA analyses. The analysis based on both the whole set of SNPs and the target SNPs revealed the relationships and clustering patterns among the accessions. Missing data, which ranged from minimal to higher percentages in some wild species, did not significantly impact the clustering of accessions. Multiple studies have delved into the impact of missing data on phylogenetic analysis, employing both empirical and simulated data. They propose that including taxa with extensive missing data does not have detrimental effects (Wiens and Moen, 2008; Lynch and Wagner, 2010; Thomson and Shaffer, 2010; Wiens and Morrill, 2011; Xi et al., 2016).
In the case of tomato, the missing data varied from 0 to 17.7%. The cultivated tomato (*S. lycopersicum*) exhibited an average of 1.18% missing data, *S. pimpinellifolium* showed 3.3%, and the highest values were observed in *S. habrochaites* (17.7%), which is the most evolutionarily divergent wild species among the ones examined (Aflitos et al., 2014; Beddows et al., 2017). Those results are supported by the findings obtained in **Chapter IV** with a different set of accessions, where the species with the highest number of missing data is *S. chilense* (specie not included in this study), followed by *S. habrochaites*.
On the other hand, the analysis of missing data in different eggplant species revealed varying levels of absence. *S. melongena* and *S. aethiopicum* exhibited very low levels (0.02 and 0.7%, respectively) of missing data, while *S. macrocarpon* had a higher frequency (4.5%). Wild species such as *S. torvum* and *S. sisymbriifolium*

showed the highest proportions of missing data (21.2% and 22.7, respectively), indicating greater evolutionary divergence (Vorontsova et al., 2013; Acquadro et al., 2017; Knapp et al., 2019).

The technique also proved useful for identifying duplicates and mislabeled accessions, showcasing its potential for quality control in germplasm banks. These results are supported by the findings obtained in the analysis of the complete eggplant (Barchi et al., 2023, manuscript under review) and tomato (Blanca et al., 2023, manuscript in preparation) collection discussed in **Chapter I**. The results of the genotyping analysis were consistent with previous knowledge of diversity in tomato and eggplant (Blanca et al., 2022; Gao et al., 2019; Hilgenhof et al., 2023; Knapp et al., 2019).
On the basis of both ML dendogram and PCA analysis with all SNPs, *S. lycopersicum* accessions clustered separately, while *S. pimpinellifolium* showed a close relationship with cultivated tomato. This results are consistent with previous reports (Gao et al., 2019; Mata-Nicolás et al., 2020). The other wild species fell into expected positions (Aflitos, et al., 2014; Beddows et al., 2017) in the dendrogram and PCA plots.
On the other hand, *Solanum melongena* accessions from Oriental and European origins formed separate clusters from other cultivated and wild species in both the ML dendrogram and PCA plots (Acquadro et al., 2017; Gramazio et al., 2017). *S. insanum* was identified as the closest relative to eggplant, followed by *S. linnaeanum* within the "Eggplant" clade, consistent with previous (Lester and Hasan, 1991; Weese1 and Bohs1, 2010) and recent studies (Acquadro et al., 2017; Aubriot et al., 2018; Knapp et al., 2019).

The relationships between *S. melongena* and the other cultivated eggplants, *S. aethiopicum* and *S. macrocarpon*, have shown conflicting results (Sakata et al., 1991; Sakata & Lester, 1997; Isshiki et al., 2008; Meyer et al., 2012; Vorontsova et al., 2013). While recent studies suggest that *S. macrocarpon* is genetically closer to eggplant, different outcomes have been obtained using SSR or SNP genotyping (Gramazio et al., 2017). Our SPET-based clustering revealed distinct groups for each of the three cultivated species, indicating probable divergence at similar times. Accessions of *S. aethiopicum* were intermixed with its wild ancestor *S. anguivi*, suggesting genetic exchange between the two species (Plazas et al., 2014). *S. tomentosum* was placed in the "Anguivi" group, supporting previous findings (Vorontsova et al., 2013; Acquadro et al., 2017). Additionally, the wild species *S. sisymbriifolium* and *S. torvum*, known for disease resistance traits, formed a distinct group in both the ML tree and PCA analyses (Acquadro et al., 2017).

In conclusion, the study demonstrated the reliability and efficiency of the SPET technique for high-throughput genotyping of tomato, eggplant, and their relatives.

It allowed for the identification of novel SNPs, detection of segregating accessions, and quality control in germplasm banks. The findings provided valuable insights into the genetic diversity and relationships within the germplasm sets, allowing for the development of **Chapter IV** of this thesis.

TOMATO CORE COLLECTION. **Linking genetic resources, genotypes to phenotypes of the G2P-SOL tomato core collection.**

The core collection analyzed in this study comprised 450 accessions obtained from renowned germplasm bank, including WorldVeg, INRAE, UPV-COMAV, MVCRI, WUR-DLO, IPK and BATEM. Through the examination of available passport data, it was determined that out of the 450 accessions, 354 belonged to the cultivated species *S. lycopersicum* var. *lycopersicum* (SLL), while 46 were classified as *S. lycopersicum* var. *cerasiforme* (SLC), and 50 were wild species. The origins of this diverse collection encompassed Mesoamerica as the center of origin, along with Europe and Asia as centers of diversity. Notably, the inclusion of wild species from Ecuador, Chile, and Peru added to the genetic richness of the collection. Moreover, the core collection showcased a well-balanced distribution of accessions with different biological statuses. Traditional cultivars accounted for approximately 34.0% of the collection, originating from Spain, Italy, and Central and South America, representing the cultural heritage and traditional farming practices associated with the crop. Modern cultivars, comprising 28.2% of the accessions, reflected advancements in breeding and agricultural practices, predominantly from Northern Europe, Asia, and North America. Furthermore, 15.5% of the accessions consisted of breeding materials, highlighting ongoing efforts to enhance the crop's traits and adaptability.

The genetic structure of the core collection was investigated through the information generated in a preliminary genetic analysis (Blanca et al., 2023, manuscript in preparation) utilizing all genotyped accessions from the whole tomato collection described on **Chapter I**. The accessions were classified into different genetic groups based on their genetic proximity, as determined by Principal Component Analysis (PCA). These groups were further characterized using passport data, morphological traits, and previous studies on tomato evolution (Blanca et al., 2022a). The PCA analysis revealed clear separations between wild species and cultivated tomatoes, with additional differentiation among the different wild species groups. These results are consistent with previous research, such as the carried out by Blanca et al., (2012, 2015), indicating the reliability of the sequencing technology detailed in **Chapter III**. Moreover, the extrapolation of data from a large collection (n = 15,504) to a core collection (n = 450) indicates that the core collection displayed a balanced distribution of accessions with different biological statuses, including old varieties, traditional cultivars and modern cultivars. The analysis

highlighted the genetic diversity present within the collection and demonstrated the influence of breeding practices and the incorporation of wild species on modern improved cultivars. These results align with the conclusions drawn in earlier research developed by Schouten et al., 2019. Furthermore, the analysis emphasized the importance of early-vintage local varieties, including traditional varieties from Spain and Italy. Our results are in line with the literature (Blanca, et al., 2022b; García-Martínez et al., 2013; Pons et al., 2022), which has consistently demonstrated the suitability of these accessions due to their adaptation to specific soil and climatic conditions in the cultivation areas. The presence of wild species introgressions in modern cultivars, particularly in the "Cherry" type, was also observed. Our findings are in agreement with prior investigations that have reported by Sim et al., 2012b. This comprehensive understanding of the genetic structure of the core collection provides valuable insights for future research and breeding efforts aimed at enhancing tomato diversity and improving desirable traits.

The analysis of genetic diversity in the core collection provides further support for the findings obtained from the PCA analysis. The diversity indices reveal interesting patterns within different groups. Notably, the im_cherry group demonstrates the highest level of diversity, which is consistent with similar findings reported by Sim et al., (2012b). In addition, the accessions from sp and slc_ecu_per_bol also exhibit considerable diversity. These results are in concordance with Blanca et al., (2012, 2015) and Razifard et al., (2020). Surprisingly, the wild species with green fruit (green) display lower diversity, contrary to the results reported by Schouten et al., (2019). This discrepancy could be attributed to the SNP selection panel used, which focused on polymorphism in *S. lycopersicum* and *S. pimpinellifolium*. Consequently, a significant number of SNPs were missing in more distantly related species. The decline in diversity from slc_ecu_per_bol to slc_ma indicates a bottleneck phenomenon and highlights the regions of highest genetic diversity within the variety. Furthermore, the early improved materials (eim) derived from traditional cultivars exhibit low levels of genetic diversity. However, modern cultivars (mim), particularly the Cherry types (im_cherry), show increased diversity due to the incorporation of wild species in breeding programs. These findings refute previous claims regarding a loss of genetic diversity, as demonstrated by studies conducted by Schouten et al., (2019) in Dutch greenhouses. The utilization of wild species has had both positive and negative impacts on disease resistance and fruit quality. Notably, the increased diversity among modern cultivars, especially in the Cherry types, has been highlighted by Sim et al., (2012b), indicating higher diversity compared to traditional varieties. The proportion of polymorphic loci further supports the observed patterns of genetic diversity.

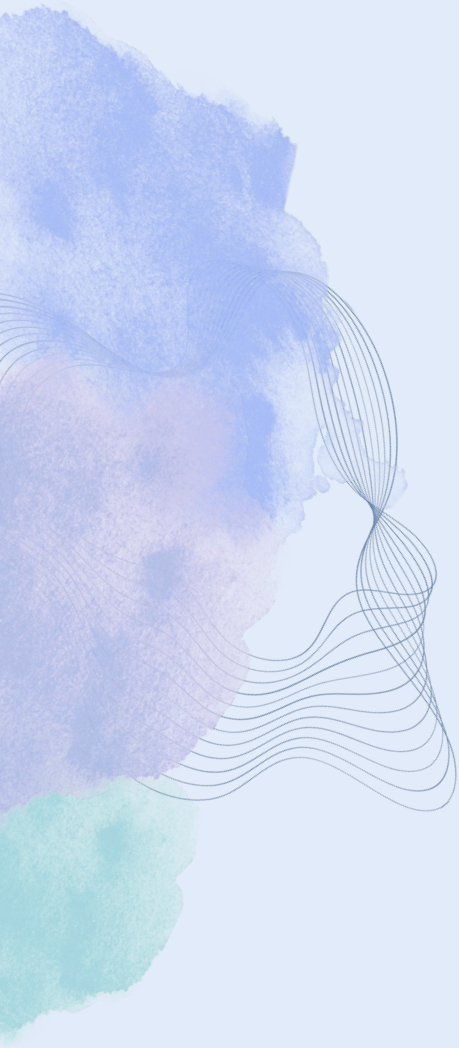
Phenotypic characterization of the core collection revealed valuable insights into the morphological variability among the genetic groups. PCA was performed on 11 qualitative and 15 quantitative traits, showing distinct patterns among different genetic groups. The PCA results highlighted the influence of fruit traits, such as size, shape, and ribbing, on the distribution of accessions. Additionally, correlations between traits provided further understanding of their relationships. Production was strongly correlated with fruit weight, while negatively correlated with Brix degree. However, the correlation varied across genetic groups, indicating the importance of selecting the appropriate combination of fruit weight and Brix degree within each group. Fruit traits were divided into two clusters, with certain traits associated with cylindrical or rounded fruits. Agronomic traits related to flowering and ripening earliness showed negative correlations with fruit size and Brix degree, suggesting that smaller-sized fruits tend to mature earlier. Despite the limited number of accessions in some groups, valuable insights were gained regarding the changes in genetic diversity over time. The slc_ecu_per_bol group exhibited high genetic diversity comparable to S. pimpinellifolium, with a significant loss of diversity observed during migration to Mesoamerica and Europe as was already demonstrated by Blanca et al., 2022a. The early_vint group, representing initial introductions into local markets, displayed distinct genetic backgrounds and a wide range of morphological variability. The substantial genetic and morphological variability within certain genetic groups, such as sp and slc_ecu_per_bol, suggests their potential value for breeding programs aimed at enhancing specific traits as Brix content (Casals et al., 2018). Traditional varieties from Spain (vint_esp) and Italy (vint_ita) exhibited low genetic diversity but demonstrated significant morphological variation, particularly in traits related to fruit phenotypes and growth habits. The presence of elongated fruits in the vint_ita group, typically used for industrial processing, underscores the importance of considering specific traits for different tomato market segments. The modern improved varieties group (mim) showed genetic variability due to the incorporation of wild species, with notable differences between varieties intended for industrial processing and fresh consumption. Overall, the findings highlight the importance of *S.l.* var. *cerasiforme* and early_vint accessions, emphasize the maintenance of morphological diversity despite limited genetic diversity in traditional varieties, and underscore the potential for genetic improvement in modern cultivars. These findings contribute to the knowledge of tomato morphological diversity and provide valuable insights for breeding programs aiming for specific trait combinations and accelerated breeding strategies. Further studies have also supported the importance of considering fruit size and quality in breeding programs to meet market demands (Fridman et al., 2002, 2004; Roohanitaziani et al., 2020; Gimeno-Páez et al., 2023).

In conclusion, the core collection exhibited high diversity in various fruit traits, including size, color, and shape. These findings reinforce the importance of the core

collection as a valuable resource for research and breeding programs aimed at enhancing tomato diversity and resilience. Ongoing re-sequencing of the collection will enable the identification of mutations or allelic variants in genes related to fruit traits, such as size and shape control. Major loci involved in fruit size and shape have been identified, contributing to the understanding of the genetic basis of these traits. Fruit color diversity in the core collection is attributed to mutations in genes involved in biosynthesis or degradation of carotenoids, flavonoids, and chlorophyll. In addition to fruit traits, variation in plant architecture, inflorescences, and yield-related traits is also significant. The inclusion of accessions with diverse characteristics in the core collection, including tolerance to stresses and traits related to fruit quality, enhances its value for breeding programs. The utilization of wild species and the integration of traits from old cultivars and traditional varieties contribute to expanding the genetic base of modern tomatoes. The availability of phenotypic and genotypic data from the core collection offers valuable genetic resources for further tomato improvement in various aspects, including flavor and stress tolerance.

GENERAL CONCLUSIONS

You have to believe in yourself, that's the secret.
Charlie Chaplin



1) A review has been performed on the tomato and eggplant germplasm collection and the status of the available information of these collections, as well as their relevance for breeding. This information serves as a crucial foundation for further research and breeding programs.

2) A database has been developed for the major solanaceous crops, namely tomato, pepper, and eggplant, containing standardized and unified passport, phenotyping, and image data, and made them readily accessible to end-users.

3) The analysis of the passport data has revealed the existence of duplicated accessions among and within genebanks, misclassifications, and insufficient documentation to establish traceability of these duplications in many cases. This contributes to an inefficient use of the genetic resources conserved.

4) The analysis of phenotypic data of the tomato, pepper, and eggplant collections provided by genebanks, has revealed the extensive diversity for plant, inflorescence, fruit, and agronomic traits. The availability of these data in a single place can contribute to enhance their use for researchers and breeders.

5) A robust genomic DNA extraction protocol has been developed, suitable for use in a wide range of species, including recalcitrant species, and various tissue samples. Furthermore, the SILEX protocol can be extensively utilized in laboratories due to the availability of common consumables, avoiding the need for expensive reagents and specialized equipment.

6) The SILEX protocol offers numerous advantages compared to commercial kits and the CTAB-based method. It is well-suited for routine DNA extraction across various applications, including Next-Generation Sequencing (NGS) platforms.

7) We have developed platforms for genotyping tomato and eggplant using the SPET technology, which combines targeted analysis of SNPs (like genotyping arrays) and complexity reduction (like GBS approach). The two platforms have bee used for assessing genetic diversity in tomato and eggplant accessions, including their wild relatives.

8) SPET allows the discovery of thousands of novel SNPs not originally included in the panel, thanks to the sequencing of the genomic regions around the target SNPs.

9) Through ML dendrograms and PCA analyses of SPET data, a substantial number of both target and non-target polymorphisms have been evaluated. The two approaches complemented each other in interpreting the data, providing valuable insights into the genetic diversity of the studied species. The utilization of the whole SNP dataset is preferable for comprehensive phylogenetic studies, whereas the use of target SNPs enhances intra-specific discrimination, particularly for studies on domestication and genome-wide association.

10) The SPET technology was successful in distinguishing between different species, confirmed established phylogenetic relationships, resolved mislabelling errors, and demonstrated its high reproducibility through replicate testing of the same accessions. These outcomes underscore the value of SPET for high-throughput genotyping, management, and improvement of genebank collections.

11) The genetic analysis of the core collection made up of 400 tomato accessions and 50 accessions of wild species reveals a reduction in genetic diversity in the vintage materials of tomatoes compared to wild species and *S.l.* var. *cerasiforme*. However, modern improved varieties exhibit higher genetic diversity primarily due to their enhancement through introgressions from wild species.

12) The core collection has been split in different genetic groups based on their genetic similarity. The phenotypic characterization of the core collection has demonstrated different patterns of diversity in each genetic group for plant, inflorescence, fruit, and agronomic traits, as a result of their evolutionary history and the level of genetic improvement.

13) The development of this doctoral thesis has significant implications on the conservation and use of genetic resources, representing a substantial advancement in terms of generation of information and development of biotechnological tools. The work carried out here, along with the generated knowledge, paves the way for new research avenues in plant breeding management, allowing for the identification of novel genes related to morphological traits of interest, as well as resistance or tolerance to biotic and abiotic stresses.

REFERENCES (GENERAL INTRODUCTION & DISCUSSION)

Acquadro, A., Barchi, L., Gramazio, P., Portis, E., Vilanova, S., Comino, C., et al. (2017). Coding SNPs analysis highlights genetic relationships and evolution pattern in eggplant complexes. *PLoS One.* 12, e0180774.

Aflitos, S., Schijlen, E., de Jong, H., de Ridder, D., Smit, S., Finkers, R., et al. (2014). Exploring genetic variation in the tomato (*Solanum* section *Lycopersicon*) clade by whole-genome sequencing. *Plant J.* 80, 136–148.

Alexander, L. J. (1963). Transfer of a dominant type of resistance to 4 known Ohio pathogenic strains of tobacco mosaic virus (TMV), from *Lycopersicon peruvianum* to *L. esculentum*. *Phytopathology.* 53(8), 869.

Alonso, D., Prohens, J., Díez, M. J., and Giuliano, G. (2018). Origen y composición de la colección de germoplasma de solanáceas: tomate, pimiento, berenjena y patata del proyecto europeo G2P-SOL. *Actas de Horticultura Sociedad Española de Ciencias Hortícolas*, Murcia. 26-29.

Arrones, A., Mangino, G., Alonso, D., Plazas, M., Prohens, J., Portis, E., et al. (2022). Mutations in the *SmAPRR2* transcription factor suppressing chlorophyll pigmentation in the eggplant fruit peel are key drivers of a diversified colour palette. *Front. Plant Sci.* 13, 4196.

Ashrafi, H., Kinkade, M., and Foolad, M. R. (2009). A new genetic linkage map of tomato based on a *Solanum lycopersicum* × *S. pimpinellifolium* RIL population displaying locations of candidate pathogen response genes. *Genome.* 52, 935–956.

Aubriot, X., Knapp, S., Syfert, M. M., Poczai, P., and Buerki, S. (2018). Shedding new light on the origin and spread of the brinjal eggplant (*Solanum melongena* L.) and its wild relatives. *Am. J. Bot*. 105, 1175–1187.

Baccichet, I., Chiozzotto, R., Scaglione, D., Bassi, D., Rossini, L., and Cirilli, M. (2022). Genetic dissection of fruit maturity date in apricot (*P. armeniaca* L.) through a Single Primer Enrichment Technology (SPET) approach. *BMC Genomics.* 23, 712.

Bai, Y., and Lindhout, P. (2007). Domestication and Breeding of Tomatoes: What have We Gained and What Can We Gain in the Future? *Ann. Bot.* 100, 1085.

Bailey, D. M. (1941). The seedling test method for root-knot-nematode resistance. *Proceedings of the American Society for Horticultural Science*. 38, 573–575.

Baldwin, E. A., Scott, J. W., Shewmaker, C. K., and Schuch, W. (2000). Flavor trivia and tomato aroma: biochemistry and possible mechanisms for control of important aroma components. *HortScience*. 35, 1013–1021.

Barchi, L., Acquadro, A., Alonso, D., Aprea, G., Bassolino, L., Demurtas, O., et al. (2019). Single Primer Enrichment Technology (SPET) for high-throughput genotyping in tomato and eggplant germplasm. *Front. Plant Sci.* 10, 1005.

Barrantes, W., López-Casado, G., García-Martínez, S., Alonso, A., Rubio, F., Ruiz, J. J., et al. (2016). Exploring new alleles involved in tomato fruit quality in an introgression line library of *Solanum pimpinellifolium*. *Front. Plant Sci.* 7, 1172.

Bauchet, G., and Causse, M. (2012). Genetic diversity in tomato (*Solanum lycopersicum*) and its wild relatives. *Genet. Divers. plants.* 8, 134–162.

Beddows, I., Reddy, A., Kloesges, T., and Rose, L. E. (2017). Population genomics in wild tomatoes—The interplay of divergence and admixture. *Genome Biol. Evol.* 9, 3023–3038.

Benjamino, J., Leopold, B., Phillips, D., and Adams, M. D. (2021). Genome-based targeted sequencing as a reproducible microbial community profiling assay. *Msphere*. 6(2), e01325-20.

Blanca, J., Cañizares, J., Cordero, L., Pascual, L., Diez, M. J., and Nuez, F. (2012). Variation revealed by SNP genotyping and morphology provides insight into the origin of the tomato. *PLoS One*. 7.

Blanca, J., Montero-Pau, J., Sauvage, C., Bauchet, G., Illa, E., Díez, M. J., et al. (2015). Genomic variation in tomato, from wild ancestors to contemporary breeding accessions. *BMC Genomics*. 16, 257.

Blanca, J., Sanchez-Matarredona, D., Ziarsolo, P., Montero-Pau, J., van der Knaap, E., Díez, M. J., et al. (2022a). Haplotype analyses reveal novel insights into tomato history and domestication driven by long-distance migrations and latitudinal adaptations. *Hortic. Res.* 9, 1–14.

Blanca, J., Pons, C., Montero-Pau, J., Sanchez-Matarredona, D., Ziarsolo, P., Fontanet, L., et al. (2022b). European traditional tomatoes galore: a result of farmers' selection of a few diversity-rich loci. *J. Exp. Bot.* 73, 3431–3445.

Bohra, A., Kilian, B., Sivasankar, S., Caccamo, M., Mba, C., Mccouch, S. R., et al. (2021). Reap the crop wild relatives for breeding future crops. *Trends Biotechnol.* 40(4), 412-431.

Brewer, M. T., Lang, L., Fujimura, K., Dujmovic, N., Gray, S., and Van Der Knaap, E. (2006). Development of a controlled vocabulary and software application to analyze fruit shape variation in tomato and other plant species. *Plant Physiol.* 141, 15–25.

Butler, L. (1968). Linkage summary. *Rep. Tomato Genet. Coop*. 18, 4–6.

Capel, C., Fernández del Carmen, A., Alba, J. M., Lima-Silva, V., Hernández-Gras, F., Salinas, M., et al. (2015). Wide-genome QTL mapping of fruit quality traits in a tomato RIL population derived from the wild-relative species *Solanum pimpinellifolium* L. *Theor. Appl. Genet.* 128, 2019–2035.

Casals, J., Rivera, A., Sabaté, J., del Castillo, R. R., and Simó, J. (2018). Cherry and fresh market tomatoes: differences in chemical, morphological, and sensory traits and their implications for consumer acceptance. *Agron.* 9(1), 9.

Causse, M., Giovannoni, J., Bouzayen, M., and Zouine, M. (2017). Compendium of plant genomes The Tomato Genome. *Springer*. Berlin.

Cebolla-Cornejo, J., Soler, S., and Nuez, F. (2007). Genetic erosion of traditional varieties of vegetable crops in Europe: tomato cultivation in Valencia (Spain) as a case Study. *Int. J. Plant Prod.* 1.

Cericola, F., Portis, E., Toppino, L., Barchi, L., Acciarri, N., Ciriaci, T., et al. (2013). The population structure and diversity of eggplant from Asia and the Mediterranean basin. *PLoS One.* 8, e73702.

Cheema, D. S., and Dhaliwal, M. S. (2004). Hybrid tomato breeding. *J. New Seeds.* 6, 1–14.

Chen, K.-Y., and Tanksley, S. D. (2004). High-resolution mapping and functional analysis of *se2.1*. *Genetics.* 168, 1563–1573.

D'Amico-Willman, K. M., Sideli, G. M., Allen, B. J., Anderson, E. S., Gradziel, T. M., and Fresnedo-Ramírez, J. (2022). Identification of putative markers of non-infectious bud failure in almond [*Prunus dulcis* (Mill.) D.A. Webb] through genome wide DNA methylation profiling and gene expression analysis in an almond × peach hybrid population. *Front. Plant Sci*. 13, 5.

Daigle, D., Simen, B. B., and Pochart, P. (2011). High-throughput sequencing of PCR products tagged with universal primers using 454 Life Sciences Systems. *Curr. Protoc. Mol. Biol.* 96, 7.5.1-7.5.14.

Daunay, M.C., Lester, R. N., and Ano, G. (2001). Eggplants. In *Tropical plant breeding*. *CRC Press.* 199-222.

Daunay, M.C., Salinier, J., and Aubriot, X. (2019). Crossability and diversity of eggplants and their wild relatives. *Springer*, Cham. 135–191.

Díez, M. J., and Nuez, F. (2008). Tomato. *Veg. II*, 249–323.

Díez, M. J., De la Rosa, L., Martín, I., Guasch, L., Cartea, M. E., Mallor, C., et al. (2018). Plant genebanks: Present situation and proposals for their improvement. The case of the Spanish network. *Front. Plant Sci.* 871, 1–13.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One.* 6, 1–10.

FAOSTAT. (2021). FAOSTAT database collections.

Farinon, B., Picarella, M. E., Siligato, F., Rea, R., Taviani, P., and Mazzucato, A. (2022). Phenotypic and genotypic diversity of the tomato germplasm from the Lazio region in central Italy, with a focus on landrace distinctiveness. *Front. Plant Sci.* 13, 2400.

Figàs, M. R., Prohens, J., Casanova, C., Fernández-de-Córdova, P., and Soler, S. (2018). Variation of morphological descriptors for the evaluation of tomato germplasm and their stability across different growing conditions. *Sci. Hortic*. 238, 107–115.

Figàs, M. R., Prohens, J., Raigón, M. D., Fernández-de-Córdova, P., Fita, A., and Soler, S. (2015). Characterization of a collection of local varieties of tomato (*Solanum lycopersicum* L.) using conventional descriptors and the high-throughput phenomics tool Tomato Analyzer. Genet. *Resour. Crop Evol.* 62, 189–204.

Foolad, M. R. (2007). Genome mapping and molecular breeding of tomato. *Int. J. Plant Genomics 2007*.

Fowler, C., and Hodgkin, T. (2004). Plant genetic resources for food and agriculture: Assessing global availability. *Annu. Rev. Environ. Resour.* 29, 143–179.

Fridman, E., Carrari, F., Liu, Y. S., Fernie, A. R., and Zamir, D. (2004). Zooming in on a quantitative trait for tomato yield using interspecific introgressions. *Science*. 305(5691), 1786–1789.

Fridman, E., Liu, Y. S., Carmel-Goren, L., Gur, A., Shoresh, M., Pleban, T., et al. (2002). Two tightly linked QTLs modify tomato sugar content via different physiological pathways. *Mol Genet Genomics*. 266, 821–826.

Galán-Ávila, A., Gramazio, P., Ron, M., Prohens, J., and Herraiz, F. J. (2021). A novel and rapid method for *Agrobacterium*-mediated production of stably transformed *Cannabis sativa* L. plants. *Ind. Crops Prod.* 170, 113691.

Gao, L., Gonda, I., Sun, H., Ma, Q., Bao, K., Tieman, D. M., et al. (2019). The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. *Nat. Genet.* 51, 1044–1051.

García-Fernández, C., Jurado, M., Campa, A., Brezeanu, C., Geffroy, V., Bitocchi, E., et al. (2022). A core set of snap bean genotypes established by phenotyping a large panel collected in Europe. *Plants.* 11, 577.

García-Martínez, S., Corrado, G., Ruiz, J. J., and Rao, R. (2013). Diversity and structure of a sample of traditional Italian and Spanish tomato accessions. *Genet. Resour. Crop Evol.* 60, 789–798.

Gimeno-Páez, E., Prohens, J., Moreno-Cerveró, M., Luis-Margarit, A. de, Díez, M. J., and Gramazio, P. (2023). Agronomic treatments combined with embryo rescue for rapid generation advancement in tomato speed breeding. B*ioRxiv.*

Gonzalo, M. J., da Maia, L. C., Nájera, I., Baixauli, C., Giuliano, G., Ferrante, P., et al. (2022). Genetic control of reproductive traits under different temperature regimes in inbred line populations derived from crosses between *S. pimpinellifolium* and *S. lycopersicum* accessions. *Plants.* 11, 1069.

Gotor, E., Alercia, A., Rao, V. R., Watts, J., and Caracciolo, F. (2008). The scientific information activity of Bioversity International: The descriptor lists. *Genet. Resour. Crop Evol.* 55, 757–772.

Gramazio, P., Pereira-Dias, L., Vilanova, S., Prohens, J., Soler, S., Esteras, J., et al. (2020). Morphoagronomic characterization and whole-genome resequencing of eight highly diverse wild and weedy *S. pimpinellifolium* and *S. lycopersicum* var. *cerasiforme* accessions used for the first interspecific tomato MAGIC population. *Hortic. Res.* 7, 174.

Gramazio, P., Prohens, J., Borràs, D., Plazas, M., Herraiz, F. J., and Vilanova, S. (2017). Comparison of transcriptome-derived simple sequence repeat (SSR) and single nucleotide polymorphism (SNP) markers for genetic fingerprinting, diversity evaluation, and establishment of relationships in eggplants. *Euphytica.* 213, 1–18.

Grandillo, S., and Tanksley, S. D. (1996). QTL analysis of horticultural traits differentiating the cultivated tomato from the closely related species *Lycopersicon pimpinellifolium*. *Theor. Appl. Genet*. 92, 935–951.

Grozeva, S., Nankar, A. N., Ganeva, D., Tringovska, I., Pasev, G., and Kostova, D. (2021). Characterization of tomato accessions for morphological, agronomic, fruit quality, and virus resistance traits. *Can. J. Plant Sci.* 101, 476–489.

Hernández, M., Inglada Ll., Millan, C., Bernabé, N., Ayala, J. (2021). Aportación social, económica y ambiental del sector obtentor al cultivo, distribución y consumo del tomate. In: Impactos en la cadena de valor agroalimentaria. *Institut Cerdá.*

Hilgenhof, R., Gagnon, E., Knapp, S., Aubriot, X., Tepe, E. J., Bohs, L., et al. (2023). Morphological trait evolution in *Solanum* (*Solanaceae*): evolutionary lability of key taxonomic characters. *BioRxiv*.

Ho, J., Weide, R., Ma, H. M., van Wordragen, M. F., Lambert, K. N., Koornneef, M., et al. (1992). The root-knot nematode resistance gene (*Mi*) in tomato: construction of a molecular linkage map and identification of dominant cDNA markers in resistant genotypes. *Plant J.* 2, 971–982.

Hoheisel, J. D. (2006). Microarray technology: beyond transcript profiling and genotype analysis. *Nat. Rev. Genet.* 7(3), 200–210.

Huang, J., Ge, X., and Sun, M. (2000). Modified CTAB protocol using a silica matrix for isolation of plant genomic DNA. *Biotechniques.* 28(3), 432-434.

Huang, Y. F., Chen, S. C., Chiang, Y. S., Chen, T. H., and Chiu, K. P. (2012). Palindromic sequence impedes sequencing-by-ligation mechanism. *BMC Syst. Biol.* 6, 1–7.

Inaba, M., Fort, A., Bringloe, T., Mols-Mortensen, A., Ni Ghriofa, C., and Sulpice, R. (2022). Branding and tracing seaweed: Development of a high-resolution genetic kit to identify the geographic provenance of *Alaria esculenta*. *Algal Res*. 67, 102826.

Isshiki, S., Iwata, N., and Khan, M. M. R. (2008). ISSR variations in eggplant (*Solanum melongena* L.) and related *Solanum* species. *Sci. Hortic.* 117, 186–190.

Jones, D. A., Dickinson, M. J., Balint-Kurti, P. J., Dixon, M. S., and Jones, J. D. G. (1993). Two complex resistance loci revealed in tomato by classical and RFLP mapping of the *Cf-2, Cf-4, Cf-5*, and *Cf-9* genes for resistance to *Cladosporium fulvum*. *Mol. Plant Microbe Interact.* 6, 348.

Kapazoglou, A., Gerakari, M., Lazaridi, E., Kleftogianni, K., Sarri, E., Tani, E., et al. (2023). Crop wild relatives: a valuable source of tolerance to various abiotic stresses. *Plants*. 12(2), 328.

Khush, G. S. (1999). Green revolution: Preparing for the 21st century. *Genome* 42(4), 646–655.

Kircher, M., Heyn, P., and Kelso, J. (2011). Addressing challenges in the production and analysis of illumina sequencing data. *BMC Genomics.* 12, 1–14.

Knapp, S., Aubriot, X., and Prohens, J. (2019). Eggplant (*Solanum melongena* L.): Taxonomy and relationships. In: Chapman MA, ed. *The Eggplant Genome*. Springer. 11-22.

Laterrot, H., and Pecaut, P. (1969). Gene *Tm-2*: new source. *Rep Tomato Genet Coop.* 19, 13–14.

Lester, R. N., and Hasan, S. M. Z. (1991). Origin and domestication of the brinjal eggplant, *Solanum melongena*, from *S. incanum*, in Africa and Asia. In *Solanaceae III: taxonomy, chemistry, evolution London: The Linnean Society of London*. 369–387.

Li, J. F., Li, L., and Sheen, J. (2010). Protocol: A rapid and economical procedure for purification of plasmid or plant DNA with diverse applications in plant biology. *Plant Methods*. 6, 6-1.

Li, J.-F., and Sheen, J. (2012). DNA Purification from multiple sources in plant research with homemade silica resins. In: *Plant DNA fingerprinting and barcoding: Methods and protocols. Humana Press.* 862, 53–59.

Loit, K., Adamson, K., Bahram, M., Puusepp, R., Anslan, S., Kiiker, R., et al. (2019). Relative performance of MinION (Oxford Nanopore Technologies) versus Sequel (Pacific Biosciences) third generation sequencing instruments in identification of agricultural and forest fungal pathogens. *Appl. Environ. Microbiol*. 85(21), e01368-19.

Lynch, V. J., and Wagner, G. P. (2010). Did egg-laying boas break Dollo's lay? Phylogenetic evidence for reversal to oviparity in sand boas (*Eryx: Boidae*). *Evolution.* 64, 207–216.

Mangino, G., Arrones, A., Plazas, M., Pook, T., Prohens, J., Gramazio, P., et al. (2022). Newly developed MAGIC population allows identification of strong associations and candidate genes for anthocyanin pigmentation in eggplant. *Front. Plant Sci.* 13, 1–15.

Mascher, M., Schreiber, M., Scholz, U., Graner, A., Reif, J. C., and Stein, N. (2019). Genebank genomics bridges the gap between the conservation of crop diversity and plant breeding. *Nat. Genet.* 51, 1076–1081.

Mata-Nicolás, E., Montero-Pau, J., Gimeno-Paez, E., Garcia-Carpintero, V., Ziarsolo, P., Menda, N., et al. (2020). Exploiting the diversity of tomato: the development of a phenotypically and genetically detailed germplasm collection. *Hortic. Res.* 7.

Mazzucato, A., Papa, R., Bitocchi, E., Mosconi, P., Nanni, L., Negri, V., et al. (2008). Genetic diversity, structure and marker-trait associations in a collection of Italian tomato (*Solanum lycopersicum* L.) landraces. *Theor. Appl. Genet.* 116, 657–669.

Meyer, R. S., Karol, K. G., Little, D. P., Nee, M. H., and Litt, A. (2012). Phylogeographic relationships among Asian eggplants and new perspectives on eggplant domestication. *Mol. Phylogenet. Evol.* 63, 685–701.

Milner, S. G., Jost, M., Taketa, S., Mazón, E. R., Himmelbach, A., Oppermann, M., et al. (2019). Genebank genomics highlights the diversity of a global barley collection. *Nat. Genet.* 51, 319–326.

Nairismägi, M. L., Tan, J., Lim, J. Q., Nagarajan, S., Ng, C. C. Y., Rajasegaran, V., et al. (2016). JAK-STAT and G-protein-coupled receptor signalling pathways are frequently altered in epitheliotropic intestinal T-cell lymphoma. *Leukemia*. 30, 1311–1319.

Parameswaran, P., Jalili, R., Tao, L., Shokralla, S., Gharizadeh, B., Ronaghi, M., et al. (2007). A pyrosequencing-tailored nucleotide barcode design unveils opportunities for large-scale sample multiplexing. *Nucleic Acids Res.* 35(19), e130.

Peralta, I. E., and Spooner, D. M. (2005). Morphological characterization and relationships of wild tomatoes (*Solanum* L. Sect. *Lycopersicon*). *Monographs In Systematic Botany*, 104, 227.

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

Peres, S. (2016). Saving the gene pool for the future: Seed banks as archives. *Stud. Hist. Philos. Sci. Part C Stud. Hist. Philos. Biol. Biomed. Sci.* 55, 96–104.

Plazas, M., Andújar, I., Vilanova, S., Gramazio, P., Herraiz, F. J., and Prohens, J. (2014). Conventional and phenomics characterization provides insight into the diversity and relationships of hypervariable scarlet (*Solanum aethiopicum* L.) and gboma (*S. macrocarpon* L.) eggplant complexes. *Front. Plant Sci*. 5., 318.

Poland, J. A., and Rife, T. W. (2012). Genotyping-by-Sequencing for plant breeding and genetics. *Plant Genome.* 5(3).

Pons, C., Casals, J., Palombieri, S., Fontanet, L., Riccini, A., Rambla, J. L., et al. (2022). Atlas of phenotypic, genotypic and geographical. *Hortic. Res.* 9.

Pucker, B., Irisarri, I., de Vries, J., and Xu, B. (2022). Plant genome sequence assembly in the era of long reads: Progress, challenges and future directions. *Quant. Plant Biol.* 3, e5.

Rambla, J. L., Medina, A., Fernández-del-Carmen, A., Barrantes, W., Grandillo, S., Cammareri, M., et al. (2016). Identification, introgression, and validation of fruit volatile QTLs from a red-fruited wild tomato species. *J. Exp. Bot.* 68(3), 429-442.

Rana, M. M., Aycan, M., Takamatsu, T., Kaneko, K., Mitsui, T., and Itoh, K. (2019). Optimized nuclear pellet method for extracting Next-Generation Sequencing quality genomic DNA from fresh leaf tissue. *Methods Protoc.* 2, 54.

Razifard, H., Ramos, A., Della Valle, A. L., Bodary, C., Goetz, E., Manser, E. J., et al. (2020). Genomic evidence for complex domestication history of the cultivated tomato in Latin America. *Mol. Biol. Evol.* 37, 1118–1132.

Roberts, R. J., Carneiro, M. O., and Schatz, M. C. (2013). The advantages of SMRT sequencing. *Genome Biol.* 14, 1–4.

Rogstad, S. H. (2012). Plant DNA extraction using silica. *Plant Mol. Biol. Report.* 21, 463–463.

Ronga, D., Francia, E., Rizza, F., Badeck, F. W., Caradonia, F., Montevecchi, G., et al. (2019). Changes in yield components, morphological, physiological and fruit quality traits in processing tomato cultivated in Italy since the 1930's. *Sci. Hortic.* 257, 108726.

Roohanitaziani, R., de Maagd, R. A., Lammers, M., Molthoff, J., Meijer-Dekens, F., van Kaauwen, M. P. W., et al. (2020). Exploration of a resequenced tomato core collection for phenotypic and genotypic variation in plant growth and fruit quality traits. *Genes.* 11, 1278.

Rosa-Martínez, E., García-Martínez, M. D., Adalid-Martínez, A. M., Pereira-Dias, L., Casanova, C., Soler, E., et al. (2021). Fruit composition profile of pepper, tomato and eggplant varieties grown under uniform conditions. *Food Res. Int.* 147, 110531.

Rothberg, J. M., Hinz, W., Rearick, T. M., Schultz, J., Mileski, W., Davey, M., et al. (2011). An integrated semiconductor device enabling non-optical genome sequencing. *Nature*. 475(7356), 348-352.

Sakata, Y., and Lester, R. N. (1997). Chloroplast DNA diversity in brinjal eggplant (*Solanum melongena* L.) and related species. *Euphytica*. 97, 295–301.

Sakata, Y., Nishio, T., and Matthews, P. J. (1991). Chloroplast DNA analysis of eggplant (*Solanum melongena*) and related species for their taxonomic affinity. *Euphytica* 55, 21–26.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci.* 74, 5463–5467.

Savary, S., Ficke, A., Aubertot, J. N., and Hollier, C. (2012). Crop losses due to diseases and their implications for global food production losses and food security. *Food Secur.* 4, 519–537.

Scaglione, D., Pinosio, S., Marroni, F., Di Centa, E., Fornasiero, A., Magris, G., et al. (2019). Single primer enrichment technology as a tool for massive genotyping: a benchmark on black poplar and maize. *Ann. Bot.* 124, 543–551.

Schadt, E. E., Turner, S., and Kasarskis, A. (2010). A window into third-generation sequencing. *Hum. Mol. Genet.* 19, 227–240.

Scheben, A., Batley, J., and Edwards, D. (2017). Genotyping-by-sequencing approaches to characterize crop genomes: choosing the right tool for the right application. *Plant Biotechnol. J.* 15, 149–161.

Schouten, H. J., Tikunov, Y., Verkerke, W., Finkers, R., Bovy, A., Bai, Y., et al. (2019). Breeding has increased the diversity of cultivated tomato in The Netherlands. *Front. Plant Sci.* 10, 1606.

Schroeder, W. T., Provvidenti, R., and Robinson, R. W. (1967). Incubation temperature and virus strains important in evaluating tomato genotypes for tobacco mosaic virus reactions. *Tomato Genet Coop Rep.* 17, 47–48.

Scolnick, J. A., Dimon, M., Wang, I.-C., Huelga, S. C., and Amorese, D. A. (2015). An efficient method for identifying gene fusions by targeted RNA sequencing from fresh frozen and FFPE samples. *PLoS One.* 10, e0128916.

Sim, S.C., Durstewitz, G., Plieske, J., Wieseke, R., Ganal, M. W., Van Deynze, A., et al. (2012a). Development of a large SNP genotyping array and generation of high-density genetic maps in tomato. *PLoS One*. 7, e40563.

Sim, S.C., van Deynze, A., Stoffel, K., Douches, D. S., Zarka, D., Ganal, M. W., et al. (2012b). High-density SNP genotyping of tomato (*Solanum lycopersicum* L.) reveals patterns of genetic variation due to breeding. *PLoS One.* 7, e45520.

Simionca Mărcășan, L. I., Oltean, I., Popa, S., Plazas, M., Vilanova, S., Gramazio, P., et al. (2023). Comparative analysis of phenotypic and molecular data on response to main pear diseases and pest attack in a germplasm collection. *Int. J. Mol. Sci.* 24, 6239.

Sonah, H., Bastien, M., Iquira, E., Tardivel, A., Légaré, G., Boyle, B., et al. (2013). An improved genotyping by sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. *PLoS One.* 8, 1–9.

Stout, M. J., Kurabchew, H., and Leite, G. L. D. (2018). Host-plant resistance in tomato. *Sustain. Manag. Arthropod Pests Tomato*. 217–236.

Sun, L., Chen, J., Xiao, K., and Yang, W. (2017). Origin of the domesticated horticultural species and molecular bases of fruit shape and size changes during the domestication, taking tomato as an example. *Hortic. Plant J.* 3, 125–132.

Sun, Y. D., Liang, Y., Wu, J. M., Li, Y. Z., Cui, X., and Qin, L. (2012). Dynamic QTL analysis for fruit lycopene content and total soluble solid content in a *Solanum lycopersicum* x *S. pimpinellifolium* cross. *Genet. Mol. Res.* 11, 3696–3710.

Tamburino, R., Sannino, L., Cafasso, D., Cantarella, C., Orrù, L., Cardi, T., et al. (2020). Cultivated tomato (*Solanum lycopersicum* L.) suffered a severe cytoplasmic bottleneck during domestication: Implications from chloroplast genomes. *Plants.* 9(11), 1443.

Thomson, R. C., and Shaffer, H. B. (2010). Sparse supermatrices for phylogenetic inference: taxonomy, alignment, rogue taxa, and the phylogeny of living turtles. *Syst. Biol.* 59, 42–58.

Tripodi, P. (2022). Next generation sequencing technologies to explore the diversity of germplasm resources: achievements and trends in tomato. *Comput. Struct. Biotechnol. J.* 20, 6250-6258.

Tripodi, P., Rabanus-Wallace, M. T., Barchi, L., Kale, S., Esposito, S., Acquadro, A., et al. (2021). Global range expansion history of pepper (*Capsicum* spp.) revealed by over 10,000 genebank accessions. *Proc. Natl. Acad. Sci.* 118, 1–9.

Vaillancourt, B., and Buell, C. R. (2019). High molecular weight DNA isolation method from diverse plant species for use with Oxford Nanopore sequencing. B*ioRxiv*, 783159.

van der Knaap, E., Chakrabarti, M., Chu, Y. H., Clevenger, J. P., Illa-Berenguer, E., Huang, Z., et al. (2014). What lies beyond the eye: The molecular mechanisms regulating tomato fruit weight and shape. *Front. Plant Sci.* 5, 227.

Vilanova, S., Alonso, D., Gramazio, P., Plazas, M., García-Fortea, E., Ferrante, P., et al. (2020). SILEX: A fast and inexpensive high-quality DNA extraction method suitable for multiple sequencing platforms and recalcitrant plant species. *Plant Methods.* 16, 1–11.

Víquez-Zamora, M., Vosman, B., van de Geest, H., Bovy, A., Visser, R. G. F., Finkers, R., et al. (2013). Tomato breeding in the genomics era: Insights from a SNP array. *BMC Genomics.* 14, 1–13.

Vorontsova, M. S., Stern, S., Bohs, L., and Knapp, S. (2013). African spiny *Solanum* (subgenus *Leptostemonum*, *Solanaceae*): a thorny phylogenetic tangle. *Bot. J. Linn. Soc*. 173, 176–193.

Vos, R., and Bellù, L. G. (2019). Global Trends and Challenges to Food and Agriculture into the 21st Century. In: *Sustainable Food and Agriculture*, eds. C. Campanhola and S. Pandey (London: Academic Press), 11–30.

Wang, Y., Clevenger, J. P., Illa-Berenguer, E., Meulia, T., Van Der Knaap, E., and Sun, L. (2019). A Comparison of *sun*, *ovate, fs8.1* and auxin application on tomato fruit shape and gene expression. *Plant Cell Physiol.* 60(5), 1067–1081.

Weese, T. L., and Bohs, L. (2010). Eggplant origins: Out of Africa, into the Orient. *Taxon*. 59, 49–56.

Wei, J., Rybczynska, A. A., Meng, P., Terpstra, M., Saber, A., Sietzema, J., et al. (2020). An all-in-one transcriptome-based assay to identify therapy-guiding genomic aberrations in non-small cell lung cancer patients. *Cancers.* 12, 2843.

Weise, S., Lohwasser, U., and Oppermann, M. (2020). Document or lose it—on the importance of information management for genetic resources conservation in genebanks. *Plants.* 9, 1–13.

Wiens, J. J., and Moen, D. S. (2008). Missing data and the accuracy of Bayesian phylogenetics. *J. Syst. Evol.* 46, 307.

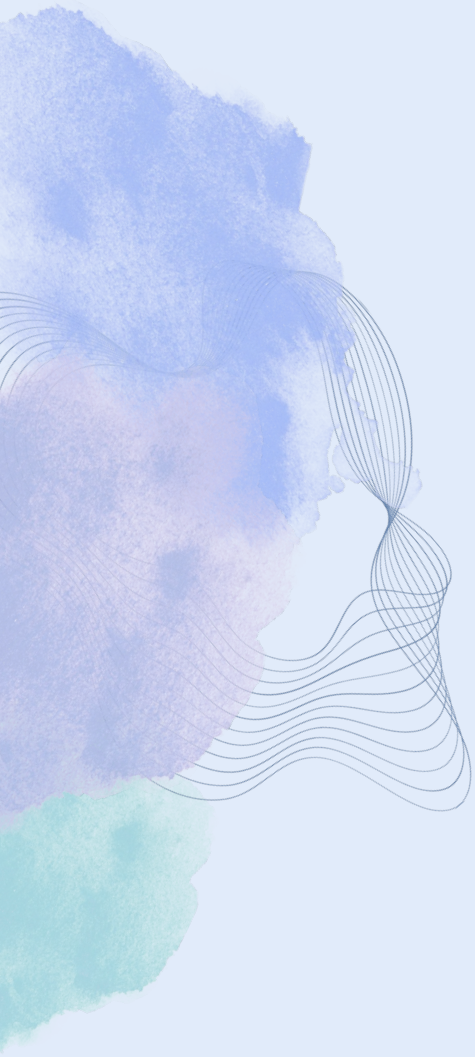
Wiens, J. J., and Morrill, M. C. (2011). Missing data in phylogenetic analysis: reconciling results from simulations and empirical data. *Syst. Biol.* 60, 719–731.

Xi, Z., Liu, L., and Davis, C. C. (2016). The impact of missing data on species tree estimation. *Mol. Biol. Evol.* 33, 838–860.

Xiao, T., and Zhou, W. (2020). The third generation sequencing: the advanced approach to genetic diseases. *Transl. Pediatr.* 9(2), 163.

Zuriaga, E., Blanca, J. M., Cordero, L., Sifres, A., Blas-Cerdán, W. G., Morales, R., et al. (2009). Genetic and bioclimatic variation in *Solanum pimpinellifolium. Genet. Resour. Crop Evol.* 56, 39–51.

SUPPLEMENTARY MATERIALS



CHAPTER I. Supplementary material

Supplementary material 1. Descriptors for tomato, pepper, and eggplant modified based on those developed by IPGRI-FAO for the collection of phenotypic data available in germplasm banks.

Tomato Descriptors

Characterization site descriptors

Country of characterization [Code of country in which the sample was originally characterized. Use the 3-letter ISO]

Cultivation period [Beginning and end month of cultivation]

Cultivation type

1. Greenhouse
2. Open air
3. Tunnel
4. Other (specify)

Plant descriptors

Vegetative data

1. **Plant growth habit**
   1. Dwarf
   2. Determinate
   3. Semi-determinate
   4. Indeterminate
2. **Leaf type**
   1. Dwarf
   2. Potato leaf type
   3. Standard
   4. Double feathered

Inflorescence data

3. **Inflorescence type**
   1. Uniparous
   2. Forked
   3. Multiparous or irregular
4. **Style position**
   1. Inserted
   3. Same level as stamens
   5. Slightly exerted
   7. Highly exerted

Fruit data

5. **Predominant fruit shape** (after the fruit turns color)
   1. Flattened
   2. Slightly flattened
   3. Rounded
   4. High rounded
   5. Heart-shaped
   6. Cylindrical

7. Pear-shaped
8. Plum-shaped
9. Other (specify)

6. Exterior color of immature fruit

1. Greenish-White
3. Light green
5. Green
7. Dark green
9. Very dark green

7. Presence of jointless pedicel

0. Absent
1. Present

8. Exterior color of mature fruit

1. Green
2. Yellow
3. Orange
4. Red
5. Pink
6. Orange-red
7. Brown
8. Violet
9. Other (specify)

9. Presence of green (shoulder) trips on the fruit

0. Absent (Uniform ripening)
1. Present

10. Intensity of greenback (green shoulder)

0. Absent
3. Slight
5. Intermediate
7. Strong

11. Fruit fasciation

1. Smooth
3. Slight
5. Medium
7. Severe

12. Fruit size

1. Very small
3. Small
5. Medium
7. Big
9. Very big

13. Shape of pistil car

1. Dot
2. Stellate
3. Linear
4. Irregular

14. Blossom end shape

1. Indented
2. Flat
3. Pointed

15. Ribbing at calix end

0. Absent
3. Slight
5. Medium
7. Strong

16. **Skin color of ripen fruit**
    1. Colourless
    2. Yellow
17. **Transvers section**
    1. Round
    2. Angular
    3. Irregular
18. **Number of locules**
19. **Puffiness**
    0. Absent
    3. Slight
    5. Medium
    7. Severe

Agronomic data

20. **Varietal type**
21. **Fruit size variation within a plant**
    1. Uniform
    3. Slight
    5. Medium
    7. High
22. **Fruit set (it refers to the proportion of flowers that set fruit)**
    1. Low
    3. Intermediate
    5. High
    7. Very high

23. **Flowering earliness** (measured in a qualitative basis, comparing each accessions with the other characterized in the same assay)
    3. Early
    5. Medium
    7. Late
24. **Maturity earliness** (measured in a qualitative basis, comparing each accessions with the other characterized in the same assay)
    3. Early
    5. Medium
    7. Late
25. **Fruit yield per plant** (measured in a qualitative basis, comparing each accessions with the other characterized in the same assay)
    1. Very low
    3. Low
    5. Medium
    7. High
    9. Very high

Images

Provide a photograph of the plant, flower and fruit (JPEG format). A picture of the fruit is the bare minimum, and should include a ruler and standardized color card.

Pepper Descriptors

Characterization site descriptors

Country of characterization [Code of country in which the sample was originally characterized. Use the 3-letter ISO]

Cultivation period [Beginning and end month of cultivation]

Cultivation type

1. Greenhouse
2. Open air
3. Tunnel
4. Other (specify)

Plant descriptors

Vegetative data

1. Plant growth habit

3. Prostrate
5. Compact
7. Erect

2. Plant height

1. Short (<50cm)
3. Intermediate (50-100 cm)
5. Tall (>100cm)
7. Very tall (>200cm)

3. Nodal anthocyanin (whole plant)

0. Green
1. Very pale purple
3. Light purple
5. Purple
7. Dark purple

Inflorescence data

4. Number of pedicel per axil

5. Pedicel position at anthesis

3. Pendant
5. Intermediate
7. Erect

6. Corolla colour

1. Whyte
2. Light yellow
3. Yellow
4. Yellow-green
5. White with purple base
6. White with purple margin
7. Purple
8. Other (specify)

7. Corolla spot

0. Absent
1. White
2. Yellow
3. Green-yellow
4. Green
5. Other

Fruit data

8. Calyx margin shape

3. Smooth
5. Intermediate
7. Dentate

9. Annular constriction at junction of calyx and peduncle (at mature green stage)

0. Absent
3. Not clear
5. Clear

10. Fruit position

3. Pendant
5. Intermediate
7. Erect

11. Fruit colour at immature stage

1. Green
2. Yellow
3. Orange
4. Red
5. Purple
6. Brown
7. Black
8. Yellow-green
9. Other (specify)

12. Fruit color at mature stage

1. Green
2. Yellow
3. Orange
4. Orange-red
5. Red
6. Purple
7. Brown
8. Black
9. Yellow-orange
10. Other (specify)

13. Fruit shape

1. Elongate
2. Oblate
3. Round
4. Conical
5. Campanulate
6. Bell or blocky
7. Pumpkin shaped
8. Other (specify)

14. Fruit shape at peduncle attachment

4. Acute
3. Obtuse
5. Truncate
7. Cordate
9. Lobate

15. Fruit shape at blossom end

3. Pointed

5. Blunt
7. Sunken
9. Sunked and pointed

16. Fruit pungency

0. Not pungent (sweet)
3. Low pungency
5. Intermediate pungency
7. High pungency

17. Fruit size

1. Very small (< 4g)
3. Small (4 < x < 10 g)
5. Medium (10 < x < 40 g)
7. Big (40< x < 150 g)
9. Very big (> 150 g)

18. Fruit cross-sectional corrugation (at 1/3 from pedicel end)

0. Smooth
3. Slightly corrugated
5. Intermediate
7. Corrugated

19. Fruit wall thickness

1. Very thin (< 1 mm)
3. Thin (1 < x < 2 mm)
5. Medium (2 < x <4 mm)
7. Thick (4 < x < 6 mm)
9. Very thick (> 6 mm)

Seed data

20. Seed colour

1. Straw
2. Black/brown
3. Other (specify)

Agronomic data

21. Varietal type

22. Fruit set (it refers to the proportion of flowers that set fruit)

1. Low
3. Intermediate
5. High
7. Very high

23. Flowering earliness (measured in a qualitative basis, comparing each accessions with the other characterized in the same assay)

3. Early
5. Medium
7. Late

24. Maturity earliness (measured in a qualitative basis, comparing each accessions with the other characterized in the same assay)

3. Early
5. Medium
7. Late

25. Fruit yield per plant (measured in a qualitative basis, comparing each accessions with the other characterized in the same assay)

2. Very low
3. Low

5. Medium
7. High
9. Very high

26. Use

1. Fresh consumption
2. Canned
3. Flavouring
4. Powder
5. Other (specify)

Images

Provide a photograph of the plant, flower and fruit (JPEG format). A picture of the fruit is the bare minimum, and should include a ruler and standardized color card.

Eggplant Descriptors

Characterization site descriptors

Country of characterization [Code of country in which the sample was originally characterized. Use the 3-letter ISO]

Cultivation period [Beginning and end month of cultivation]

Cultivation type

1. Greenhouse
2. Open air
3. Tunnel
4. Other (specify)

Plant descriptors

Vegetative data

1. **Plant growth habit**
   1. Very upright
   3. Upright
   5. Intermediate
   7. Prostrate
2. **Leaf blade lobes**
   1. Very weak
   3. Weak
   5. Intermediate
   7. Strong
   9. Very strong
3. **Leaf prickles**
   0. None
   1. Very few (1-2)
   3. Few (3-5)
   5. Intermediate (6-10)
   7. Many (11-20
   9. Very many (>20)

Inflorescence data

4. **Number of flower per inflorescence**
5. **Corolla colour**
   0. Yellow
   1. Greenish white
   3. White
   5. Pale violet
   7. Light violet
   9. Bluish violet

Fruit data

6. **Fruit size**
   1. Very small (less than 15g)
   3. Small (from 15 to 50 g)
   5. Medium (from 50 to 400 g)
   7. Big (400 to 800g)
   9. Very big (more than 1000 g)
7. **Fruit length/breadth ratio**
   1. Broader than long
   3. As long as broad
   5. Slightly longer than broad

7. Twice as long as broad
8. Three times as long as broad
9. Several times as long as broad

8. Fruit curvature

1. None
3 Slightly curved
5. Curved
7. Snake shaped
8. Sickle shaped
9. U shaped

9. Fruit shape (position of widest part of the fruit)

3. About ¼ way from base to tip
5. About ½ way from base to tip
7. About ¾ way from base to tip

10. Fruit apex shape

3. Protruded
5. Rounded
7. Depressed

11. Fruit color at commercial ripeness

1. Green
2. Milk white
3. Deep yellow
4. Fire red
5. Scarlet red
6. Lilac gray
7. Purple
8. Purple black
9. Black

12. Fruit color distribution at commercial ripeness

1. Uniform
3. Mottle
5. Netted
7. Striped

13. Fruit flesh colour

1. White
2. Intermediate
3. Green

14. Fruit calyx prickles

0. None
1. Very few (<3)
3. Few (~5)
5. Intermediate (~10)
7. Many (~20)
9. Very many (>30)

15. Fruit cross section

1. Circular, no grooves
3. Elliptic, no grooves
5. Few grooves (~4)
7. Many grooves (~8)

9. Very irregular

16. **General anthocyanin distribution in apex, stem, calix, leaf vein, leaf blade**
    0. Absent
    3. Low
    5. Medium
    7. High

Agronomic data

17. **Varietal type** [is related with the usage of the fruit. For instance, some eggplants are used for pickling. Other varietal types in eggplant are - Round black/purple - Oval black/purple - Semi Long B/P - Long B/P - White (round, U) - Stripped - Green and Mini eggplants]
18. **Fruit set** (number of fruits set compared with the number of flowers)
    1. Very low
    3. Low
    5. Medium
    7. High
    9. Very high
19. **Flowering earliness** (measured in a qualitative basis, comparing each accessions with the other characterized in the same assay)
    3. Early
    5. Medium
    7. Late
20. **Maturity earliness** (measured in a qualitative basis, comparing each accessions with the other characterized in the same assay)
    3. Early
    5. Medium
    7. Late
21. **Fruit yield per plant** (measured in a qualitative basis, comparing each accessions with the other characterized in the same assay)
    1. Very low
    3. Low
    5. Medium
    7. High
    9. Very high

Images

Provide a photograph of the plant, flower and fruit (JPEG format). A picture of the fruit is the bare minimum, and should include a ruler and standardized color card.

CHAPTER IV. Supplementary tables and figures

Supplementary Table 1. Passport data of the accessions included in the tomato core collection. Countries are labeled based on their ISO-3 code. ND = No data.

G2PSOL_code	Institute code	Genus	Specie	*Subtaxa*	Country	Biological status of accession	Genetic group
GPT000140	BATEM	*Solanum*	*lycopersicum*		TUR	300	eim
GPT000170	BATEM	*Solanum*	*lycopersicum*		TUR	300	vint_esp
GPT000210	BATEM	*Solanum*	*lycopersicum*		TUR	300	vint_esp
GPT000270	BATEM	*Solanum*	*lycopersicum*		TUR	300	vint_ita
GPT000280	BATEM	*Solanum*	*lycopersicum*		TUR	300	vint_esp
GPT000320	BATEM	*Solanum*	*lycopersicum*		TUR	300	vint_esp
GPT000430	BATEM	*Solanum*	*lycopersicum*		TUR	300	mim
GPT000500	BATEM	*Solanum*	*lycopersicum*		TUR	300	vint_esp
GPT002130	FRA011	*Solanum*	*lycopersicum*		USA	400	mim
GPT002240	FRA011	*Solanum*	*lycopersicum*		USA	400	mim
GPT002580	FRA011	*Solanum*	*lycopersicum*		BGR	ND	mim
GPT002680	FRA011	*Solanum*	*lycopersicum*		NLD	500	mim
GPT003010	FRA011	*Solanum*	*lycopersicum*		FRA	ND	mim
GPT003250	FRA011	*Solanum*	*lycopersicum*		FRA	ND	eim
GPT003260	FRA011	*Solanum*	*lycopersicum*		FRA	500	mim
GPT003270	FRA011	*Solanum*	*lycopersicum*		MAR	500	mim
GPT003290	FRA011	*Solanum*	*lycopersicum*		FRA	400	mim
GPT003310	FRA011	*Solanum*	*lycopersicum*		FRA	400	mim
GPT003350	FRA011	*Solanum*	*lycopersicum*		BGR	ND	im_cherry
GPT003580	FRA011	*Solanum*	*lycopersicum*		USA	400	mim
GPT003680	FRA011	*Solanum*	*lycopersicum*		CAN	400	mim
GPT004060	FRA011	*Solanum*	*lycopersicum*		FRA	ND	mim
GPT004340	FRA011	*Solanum*	*lycopersicum*		FRA	ND	mim
GPT004490	FRA011	*Solanum*	*lycopersicum*		BGR	400	mim
GPT004560	FRA011	*Solanum*	*lycopersicum*		ITA	ND	vint_ita
GPT004570	FRA011	*Solanum*	*lycopersicum*		FRA	500	mim
GPT004940	FRA011	*Solanum*	*lycopersicum*		USA	500	mim
GPT005160	FRA011	*Solanum*	*lycopersicum*		NLD	500	mim
GPT005810	FRA011	*Solanum*	*lycopersicum*		NLD	500	mim
GPT006670	FRA011	*Solanum*	*lycopersicum*		ITA	300	eim
GPT006680	FRA011	*Solanum*	*lycopersicum*		USA	500	mim
GPT006760	FRA011	*Solanum*	*lycopersicum*		USA	500	mim
GPT007460	FRA011	*Solanum*	*lycopersicum*		FRA	500	mim
GPT007510	FRA011	*Solanum*	*lycopersicum*		FRA	500	mim
GPT008160	FRA011	*Solanum*	*lycopersicum*		unknown	ND	eim
GPT008350	FRA011	*Solanum*	*lycopersicum*	*cerasiforme*	COL	200	early_vint
GPT008470	FRA011	*Solanum*	*lycopersicum*	*cerasiforme*	BRA	200	slc_ecu_per_bol
GPT008680	FRA011	*Solanum*	*lycopersicum*		PER	300	mim
GPT008870	FRA011	*Solanum*	*lycopersicum*		unknown	ND	mim
GPT009590	FRA011	*Solanum*	*lycopersicum*		unknown	ND	mim
GPT009990	FRA011	*Solanum*	*lycopersicum*		unknown	400	mim
GPT010210	FRA011	*Solanum*	*lycopersicum*	*cerasiforme*	MEX	300	early_vint
GPT010220	FRA011	*Solanum*	*lycopersicum*	*cerasiforme*	MEX	300	early_vint
GPT010280	FRA011	*Solanum*	*habrochaites*		PER	100	green
GPT010460	FRA011	*Solanum*	*chmielewskii*		PER	100	green
GPT010510	FRA011	*Solanum*	*habrochaites*		PER	100	green
GPT010590	FRA011	*Solanum*	*habrochaites*		ECU	100	green
GPT010600	FRA011	*Solanum*	*habrochaites*		ECU	100	green

G2PSOL_code	Institute code	Genus	Specie	*Subtaxa*	Country	Biological status of accession	Genetic group
GPT010610	FRA011	*Solanum*	*habrochaites*		ECU	100	green
GPT010630	FRA011	*Solanum*	*habrochaites*		ECU	100	green
GPT010720	FRA011	*Solanum*	*lycopersicum*	*cerasiforme*	BLZ	200	early_vint
GPT010730	FRA011	*Solanum*	*lycopersicum*		unknown	400	mim
GPT010780	FRA011	*Solanum*	*chilense*		PER	100	green
GPT010940	FRA011	*Solanum*	*chmielewskii*		PER	100	green
GPT011020	FRA011	*Solanum*	*habrochaites*		PER	100	green
GPT011220	FRA011	*Solanum*	*cheesmaniae*		ECU	100	gal_chees
GPT011440	FRA011	*Solanum*	*habrochaites*		ECU	100	green
GPT011450	FRA011	*Solanum*	*habrochaites*		PER	100	green
GPT011500	FRA011	*Solanum*	*corneliomulleri*		unknown	100	green
GPT012080	FRA011	*Solanum*	*chilense*		PER	100	green
GPT012090	FRA011	*Solanum*	*chilense*		PER	100	green
GPT012170	FRA011	*Solanum*	*chilense*		CHL	100	green
GPT012190	FRA011	*Solanum*	*chilense*		CHL	100	green
GPT012330	FRA011	*Solanum*	*habrochaites*		PER	100	green
GPT012470	FRA011	*Solanum*	*habrochaites*		PER	100	green
GPT012920	DEU146	*Solanum*	*lycopersicum*		GTM	300	early_vint
GPT014690	DEU146	*Solanum*	*lycopersicum*		USA	500	eim
GPT014700	DEU146	*Solanum*	*lycopersicum*		unknown	500	eim
GPT014710	DEU146	*Solanum*	*lycopersicum*		USA	400	mim
GPT014790	DEU146	*Solanum*	*lycopersicum*		unknown	500	mim
GPT015530	DEU146	*Solanum*	*lycopersicum*		unknown	500	mim
GPT015580	DEU146	*Solanum*	*lycopersicum*		DEU	500	mim
GPT015590	DEU146	*Solanum*	*lycopersicum*		unknown	400	eim
GPT015630	DEU146	*Solanum*	*lycopersicum*		NLD	500	mim
GPT016110	DEU146	*Solanum*	*lycopersicum*		RUS	500	eim
GPT016190	DEU146	*Solanum*	*lycopersicum*		unknown	500	eim
GPT016330	DEU146	*Solanum*	*lycopersicum*		ROU	500	eim
GPT016360	DEU146	*Solanum*	*lycopersicum*		NLD	500	eim
GPT016380	DEU146	*Solanum*	*lycopersicum*		NLD	500	mim
GPT016410	DEU146	*Solanum*	*lycopersicum*		RUS	500	eim
GPT016770	DEU146	*Solanum*	*lycopersicum*		RUS	500	mim
GPT016830	DEU146	*Solanum*	*lycopersicum*		NLD	500	mim
GPT017150	DEU146	*Solanum*	*lycopersicum*		BGR	500	mim
GPT017170	DEU146	*Solanum*	*lycopersicum*		BGR	500	mim
GPT017590	DEU146	*Solanum*	*lycopersicum*		RUS	300	mim
GPT017730	DEU146	*Solanum*	*lycopersicum*		RUS	300	mim
GPT017790	DEU146	*Solanum*	*lycopersicum*		ITA	300	vint_ita
GPT018970	DEU146	*Solanum*	*lycopersicum*		unknown	500	eim
GPT019250	DEU146	*Solanum*	*lycopersicum*		ITA	300	vint_ita
GPT019270	DEU146	*Solanum*	*lycopersicum*		ITA	300	eim
GPT019700	DEU146	*Solanum*	*lycopersicum*		unknown	300	early_vint
GPT019720	DEU146	*Solanum*	*lycopersicum*		unknown	ND	eim
GPT019950	DEU146	*Solanum*	*lycopersicum*		DEU	500	mim
GPT020030	DEU146	*Solanum*	*lycopersicum*		unknown	500	mim
GPT020190	DEU146	*Solanum*	*lycopersicum*		RUS	500	mim
GPT020450	DEU146	*Solanum*	*lycopersicum*		RUS	500	mim
GPT020770	DEU146	*Solanum*	*lycopersicum*	*cerasiforme*	USA	500	early_vint
GPT020910	DEU146	*Solanum*	*lycopersicum*		unknown	500	eim
GPT020920	DEU146	*Solanum*	*lycopersicum*		POL	500	eim
GPT020940	DEU146	*Solanum*	*lycopersicum*		DEU	500	mim
GPT021280	DEU146	*Solanum*	*lycopersicum*		ESP	300	eim
GPT021360	DEU146	*Solanum*	*lycopersicum*		unknown	ND	mim
GPT021460	DEU146	*Solanum*	*pimpinellifolium*		unknown	100	green

G2PSOL_code	Institute code	Genus	Specie	*Subtaxa*	Country	Biological status of accession	Genetic group
GPT021770	DEU146	*Solanum*	*lycopersicum*		ESP	300	mim
GPT021830	DEU146	*Solanum*	*lycopersicum*		unknown	ND	vint_ita
GPT022170	DEU146	*Solanum*	*lycopersicum*		unknown	500	mim
GPT023030	DEU146	*Solanum*	*lycopersicum*		USA	500	eim
GPT023250	DEU146	*Solanum*	*lycopersicum*		CHN	300	mim
GPT023530	DEU146	*Solanum*	*lycopersicum*		NLD	500	mim
GPT023680	DEU146	*Solanum*	*lycopersicum*	*cerasiforme*	CUB	300	early_vint
GPT023780	DEU146	*Solanum*	*lycopersicum*		unknown	500	mim
GPT023800	DEU146	*Solanum*	*lycopersicum*		unknown	300	early_vint
GPT023870	DEU146	*Solanum*	*lycopersicum*		unknown	400	eim
GPT023890	DEU146	*Solanum*	*lycopersicum*	*cerasiforme*	unknown	300	early_vint
GPT024020	DEU146	*Solanum*	*lycopersicum*		USA	500	mim
GPT024920	DEU146	*Solanum*	*lycopersicum*		unknown	500	eim
GPT025200	DEU146	*Solanum*	*lycopersicum*		unknown	999	mim
GPT025280	DEU146	*Solanum*	*lycopersicum*		unknown	100	mim
GPT025370	DEU146	*Solanum*	*lycopersicum*		MEX	300	vint_ita
GPT025480	DEU146	*Solanum*	*lycopersicum*		ITA	400	mim
GPT025600	DEU146	*Solanum*	*lycopersicum*		RUS	500	eim
GPT025710	DEU146	*Solanum*	*pimpinellifolium*		unknown	100	sp
GPT025800	DEU146	*Solanum*	*pimpinellifolium*		unknown	100	im_cherry
GPT025900	DEU146	*Solanum*	*lycopersicum*	*cerasiforme*	HRV	300	mim
GPT026050	DEU146	*Solanum*	*lycopersicum*		SDN	300	eim
GPT026150	DEU146	*Solanum*	*lycopersicum*		SDN	300	eim
GPT026210	DEU146	*Solanum*	*lycopersicum*		SDN	300	eim
GPT026380	DEU146	*Solanum*	*lycopersicum*	*cerasiforme*	ARG	300	im_cherry
GPT026390	DEU146	*Solanum*	*lycopersicum*	*cerasiforme*	ARG	300	mim
GPT026480	DEU146	*Solanum*	*lycopersicum*	*cerasiforme*	ARG	300	early_vint
GPT026750	DEU146	*Solanum*	*lycopersicum*		unknown	500	mim
GPT027240	DEU146	*Solanum*	*lycopersicum*	*cerasiforme*	CAN	400	early_vint
GPT027550	DEU146	*Solanum*	*lycopersicum*		unknown	400	mim
GPT027760	DEU146	*Solanum*	*lycopersicum*		unknown	400	vint_esp
GPT027970	DEU146	*Solanum*	*lycopersicum*	*cerasiforme*	CAN	300	mim
GPT027980	DEU146	*Solanum*	*lycopersicum*		unknown	500	mim
GPT028050	DEU146	*Solanum*	*lycopersicum*		unknown	500	mim
GPT028320	DEU146	*Solanum*	*lycopersicum*	*cerasiforme*	CAN	400	early_vint
GPT028390	DEU146	*Solanum*	*pimpinellifolium*		unknown	100	slc_ecu_per_bol
GPT028510	DEU146	*Solanum*	*lycopersicum*		unknown	500	eim
GPT028520	DEU146	*Solanum*	*lycopersicum*		USA	500	mim
GPT028550	DEU146	*Solanum*	*lycopersicum*		BGR	400	eim
GPT028790	DEU146	*Solanum*	*lycopersicum*		CAN	500	mim
GPT028940	DEU146	*Solanum*	*lycopersicum*		USA	500	eim
GPT029310	DEU146	*Solanum*	*lycopersicum*		DEU	500	mim
GPT029360	DEU146	*Solanum*	*lycopersicum*		HUN	400	mim
GPT029380	DEU146	*Solanum*	*lycopersicum*		HUN	500	mim
GPT029600	DEU146	*Solanum*	*lycopersicum*		NLD	412	mim
GPT029760	DEU146	*Solanum*	*lycopersicum*		BGR	500	mim
GPT030150	DEU146	*Solanum*	*lycopersicum*		unknown	500	mim
GPT030290	DEU146	*Solanum*	*lycopersicum*		DEU	500	mim
GPT030350	DEU146	*Solanum*	*lycopersicum*		BGR	412	eim
GPT030610	DEU146	*Solanum*	*lycopersicum*		unknown	400	mim
GPT030810	DEU146	*Solanum*	*lycopersicum*		unknown	500	mim
GPT031030	DEU146	*Solanum*	*lycopersicum*		unknown	500	mim
GPT031040	DEU146	*Solanum*	*lycopersicum*	*cerasiforme*	SDN	300	vint_ita
GPT031050	DEU146	*Solanum*	*lycopersicum*	*cerasiforme*	SDN	300	vint_ita
GPT031240	DEU146	*Solanum*	*lycopersicum*		DEU	500	im_cherry

G2PSOL_code	Institute code	Genus	Specie	*Subtaxa*	Country	Biological status of accession	Genetic group
GPT031520	DEU146	*Solanum*	*lycopersicum*		LBY	300	eim
GPT031590	DEU146	*Solanum*	*lycopersicum*		GEO	300	eim
GPT031600	DEU146	*Solanum*	*lycopersicum*		GEO	300	eim
GPT031630	DEU146	*Solanum*	*lycopersicum*		CAN	500	eim
GPT032120	DEU146	*Solanum*	*lycopersicum*	*cerasiforme*	GEO	300	early_vint
GPT032190	DEU146	*Solanum*	*lycopersicum*		CUB	500	mim
GPT032610	DEU146	*Solanum*	*lycopersicum*		CUB	300	mim
GPT032770	DEU146	*Solanum*	*lycopersicum*	*cerasiforme*	CUB	300	early_vint
GPT033220	DEU146	*Solanum*	*lycopersicum*		unknown	500	mim
GPT034780	DEU146	*Solanum*	*lycopersicum*		unknown	500	mim
GPT034930	DEU146	*Solanum*	*pimpinellifolium*		PER	100	sp
GPT035860	DEU146	*Solanum*	*lycopersicum*		unknown	500	eim
GPT035870	DEU146	*Solanum*	*lycopersicum*		unknown	500	eim
GPT036180	DEU146	*Solanum*	*lycopersicum*		ITA	300	mim
GPT036640	DEU146	*Solanum*	*lycopersicum*		RUS	500	eim
GPT037560	DEU146	*Solanum*	*lycopersicum*		RUS	500	mim
GPT041810	DEU146	*Solanum*	*lycopersicum*		UKR	500	eim
GPT041890	DEU146	*Solanum*	*lycopersicum*	*cerasiforme*	unknown	500	early_vint
GPT041900	DEU146	*Solanum*	*lycopersicum*		unknown	500	mim
GPT042000	DEU146	*Solanum*	*pimpinellifolium*		unknown	100	mim
GPT042490	DEU146	*Solanum*	*lycopersicum*		ITA	500	mim
GPT042800	DEU146	*Solanum*	*lycopersicum*		unknown	ND	mim
GPT043070	DEU146	*Solanum*	*lycopersicum*		unknown	ND	mim
GPT043290	DEU146	*Solanum*	*habrochaites*		PER	100	green
GPT044980	DEU146	*Solanum*	*lycopersicum*		SLV	300	early_vint
GPT049020	DEU146	*Solanum*	*lycopersicum*	*cerasiforme*	SLV	300	slc_ma
GPT049250	DEU146	*Solanum*	*lycopersicum*		GTM	300	early_vint
GPT050980	BGR030	*Solanum*	*lycopersicum*		BGR	410	eim
GPT051000	BGR030	*Solanum*	*lycopersicum*		BGR	410	vint_esp
GPT051040	BGR030	*Solanum*	*lycopersicum*		BGR	410	mim
GPT051060	BGR030	*Solanum*	*lycopersicum*		BGR	500	mim
GPT051070	BGR030	*Solanum*	*lycopersicum*		BGR	410	eim
GPT051080	BGR030	*Solanum*	*lycopersicum*		BGR	410	mim
GPT051100	BGR030	*Solanum*	*lycopersicum*		BGR	500	mim
GPT051110	BGR030	*Solanum*	*lycopersicum*		BGR	500	eim
GPT051120	BGR030	*Solanum*	*lycopersicum*		BGR	500	vint_esp
GPT051130	BGR030	*Solanum*	*lycopersicum*		BGR	410	im_cherry
GPT051140	BGR030	*Solanum*	*lycopersicum*	*cerasiforme*	BGR	410	im_cherry
GPT051150	BGR030	*Solanum*	*lycopersicum*	*cerasiforme*	BGR	500	im_cherry
GPT051180	BGR030	*Solanum*	*lycopersicum*		BGR	410	im_cherry
GPT051230	BGR030	*Solanum*	*lycopersicum*		BGR	410	eim
GPT051240	BGR030	*Solanum*	*peruvianum*		unknown	100	green
GPT051280	BGR030	*Solanum*	*lycopersicum*		BGR	500	mim
GPT051390	BGR030	*Solanum*	*lycopersicum*		BGR	500	mim
GPT051420	BGR030	*Solanum*	*lycopersicum*		BGR	410	mim
GPT123460	ESP026	*Solanum*	*lycopersicum*		ESP	300	mim
GPT123640	ESP026	*Solanum*	*lycopersicum*		ESP	300	vint_esp
GPT124040	ESP026	*Solanum*	*lycopersicum*		ESP	300	vint_esp
GPT124090	ESP026	*Solanum*	*lycopersicum*		ESP	300	vint_esp
GPT124380	ESP026	*Solanum*	*lycopersicum*		ESP	300	mim
GPT124610	ESP026	*Solanum*	*lycopersicum*		ESP	300	mim
GPT124940	ESP026	*Solanum*	*lycopersicum*		ESP	300	eim
GPT125570	ESP026	*Solanum*	*lycopersicum*		ESP	300	mim
GPT125690	ESP026	*Solanum*	*lycopersicum*		ESP	300	mim
GPT126040	ESP026	*Solanum*	*lycopersicum*		ESP	300	mim

G2PSOL_code	Institute code	Genus	Specie	*Subtaxa*	Country	Biological status of accession	Genetic group
GPT126140	ESP026	*Solanum*	*lycopersicum*		ESP	300	vint_esp
GPT126280	ESP026	*Solanum*	*lycopersicum*		ESP	300	mim
GPT126300	ESP026	*Solanum*	*lycopersicum*		ESP	300	vint_esp
GPT127470	ESP026	*Solanum*	*lycopersicum*		ESP	300	vint_esp
GPT127850	ESP026	*Solanum*	*lycopersicum*		ESP	300	vint_esp
GPT128010	ESP026	*Solanum*	*lycopersicum*		ESP	300	vint_esp
GPT128660	ESP026	*Solanum*	*lycopersicum*		ESP	300	mim
GPT128730	ESP026	*Solanum*	*lycopersicum*		ESP	300	mim
GPT129180	ESP026	*Solanum*	*lycopersicum*		ESP	300	mim
GPT129230	ESP026	*Solanum*	*lycopersicum*		ESP	200	eim
GPT129520	ESP026	*Solanum*	*lycopersicum*		ESP	300	eim
GPT130260	ESP026	*Solanum*	*lycopersicum*		ESP	300	eim
GPT130740	ESP026	*Solanum*	*lycopersicum*		ESP	300	mim
GPT130880	ESP026	*Solanum*	*lycopersicum*		ESP	300	mim
GPT131210	ESP026	*Solanum*	*lycopersicum*		ESP	300	mim
GPT131220	ESP026	*Solanum*	*lycopersicum*		ESP	300	eim
GPT131520	ESP026	*Solanum*	*lycopersicum*		ESP	300	eim
GPT131610	ESP026	*Solanum*	*lycopersicum*		ESP	300	im_cherry
GPT131700	ESP026	*Solanum*	*lycopersicum*		ESP	300	eim
GPT132040	ESP026	*Solanum*	*lycopersicum*		ESP	300	vint_esp
GPT132810	ESP026	*Solanum*	*lycopersicum*		ESP	300	mim
GPT133470	ESP026	*Solanum*	*lycopersicum*		ESP	300	vint_esp
GPT133710	ESP026	*Solanum*	*lycopersicum*		ESP	300	vint_esp
GPT133910	ESP026	*Solanum*	*lycopersicum*		ESP	300	vint_esp
GPT134160	ESP026	*Solanum*	*lycopersicum*		ESP	300	vint_esp
GPT134170	ESP026	*Solanum*	*lycopersicum*		ESP	300	vint_esp
GPT134420	ESP026	*Solanum*	*lycopersicum*		ESP	300	vint_esp
GPT134520	ESP026	*Solanum*	*lycopersicum*		ESP	300	vint_esp
GPT134860	ESP026	*Solanum*	*pimpinellifolium*		PER	100	mim
GPT134990	ESP026	*Solanum*	*arcanum*		PER	100	green
GPT135010	ESP026	*Solanum*	*pennellii*		PER	100	green
GPT135350	ESP026	*Solanum*	*lycopersicum*		ESP	300	im_cherry
GPT135920	ESP026	*Solanum*	*lycopersicum*		ESP	300	mim
GPT136160	ESP026	*Solanum*	*lycopersicum*		ESP	300	eim
GPT137710	ESP026	*Solanum*	*lycopersicum*		ESP	300	vint_esp
GPT138180	ESP026	*Solanum*	*lycopersicum*		ESP	300	vint_ita
GPT138690	ESP026	*Solanum*	*lycopersicum*		ESP	300	eim
GPT139240	ESP026	*Solanum*	*lycopersicum*		ESP	300	vint_esp
GPT140090	ESP026	*Solanum*	*lycopersicum*		ESP	300	vint_esp
GPT140450	ESP026	*Solanum*	*lycopersicum*		ESP	300	vint_esp
GPT140790	ESP026	*Solanum*	*lycopersicum*		ESP	300	mim
GPT140910	ESP026	*Solanum*	*lycopersicum*		ESP	300	vint_esp
GPT140950	ESP026	*Solanum*	*lycopersicum*		ESP	300	vint_esp
GPT141890	ESP026	*Solanum*	*lycopersicum*		ESP	300	vint_esp
GPT141940	ESP026	*Solanum*	*lycopersicum*		ESP	300	vint_esp
GPT142650	ESP026	*Solanum*	*lycopersicum*		ESP	300	eim
GPT142920	ESP026	*Solanum*	*lycopersicum*		ESP	300	vint_ita
GPT144270	ESP026	*Solanum*	*lycopersicum*		ESP	300	vint_esp
GPT147310	NLD037	*Solanum*	*peruvianum*		PER	100	green
GPT149330	NLD037	*Solanum*	*lycopersicum*		USA	500	mim
GPT152170	NLD037	*Solanum*	*lycopersicum*		USA	500	mim
GPT152370	NLD037	*Solanum*	*lycopersicum*		USA	500	mim
GPT152770	NLD037	*Solanum*	*corneliomulleri*		PER	100	green
GPT153840	NLD037	*Solanum*	*pimpinellifolium*		unknown	100	sp
GPT154450	NLD037	*Solanum*	*lycopersicum*		NLD	500	im_cherry

G2PSOL_code	Institute code	Genus	Specie	*Subtaxa*	Country	Biological status of accession	Genetic group
GPT160840	TWN001	*Solanum*	*lycopersicum*		PHL	400	mim
GPT161040	TWN001	*Solanum*	*lycopersicum*		PHL	400	mim
GPT161270	TWN001	*Solanum*	*pimpinellifolium*		PHL	100	eim
GPT161400	TWN001	*Solanum*	*lycopersicum*		TWN	ND	eim
GPT162120	TWN001	*Solanum*	*pimpinellifolium*		PER	100	vint_ita
GPT162150	TWN001	*Solanum*	*lycopersicum*		PER	300	mim
GPT162220	TWN001	*Solanum*	*lycopersicum*		USA	412	mim
GPT162700	TWN001	*Solanum*	*lycopersicum*		HKG	500	mim
GPT162770	TWN001	*Solanum*	*lycopersicum*		HKG	500	mim
GPT163140	TWN001	*Solanum*	*lycopersicum*		USA	500	mim
GPT163220	TWN001	*Solanum*	*lycopersicum*		USA	300	vint_ita
GPT163250	TWN001	*Solanum*	*lycopersicum*		USA	500	eim
GPT163400	TWN001	*Solanum*	*lycopersicum*		USA	500	eim
GPT163440	TWN001	*Solanum*	*lycopersicum*		USA	500	mim
GPT164460	TWN001	*Solanum*	*lycopersicum*		CHN	300	eim
GPT164560	TWN001	*Solanum*	*lycopersicum*		RUS	300	eim
GPT164660	TWN001	*Solanum*	*lycopersicum*		CHN	300	eim
GPT164900	TWN001	*Solanum*	*lycopersicum*		MAR	300	eim
GPT164920	TWN001	*Solanum*	*pimpinellifolium*		unknown	100	im_cherry
GPT165000	TWN001	*Solanum*	*pimpinellifolium*		ECU	100	slc_ecu_per_bol
GPT165240	TWN001	*Solanum*	*lycopersicum*		BRA	300	mim
GPT165480	TWN001	*Solanum*	*lycopersicum*	*cerasiforme*	VEN	300	early_vint
GPT165490	TWN001	*Solanum*	*lycopersicum*		VEN	300	early_vint
GPT165510	TWN001	*Solanum*	*lycopersicum*		VEN	300	mim
GPT165590	TWN001	*Solanum*	*lycopersicum*		TUR	300	eim
GPT165770	TWN001	*Solanum*	*lycopersicum*		TUR	300	eim
GPT165860	TWN001	*Solanum*	*lycopersicum*		IND	300	vint_ita
GPT166200	TWN001	*Solanum*	*lycopersicum*		PAN	300	vint_ita
GPT166350	TWN001	*Solanum*	*lycopersicum*		PER	300	vint_ita
GPT166850	TWN001	*Solanum*	*peruvianum*		PER	100	mim
GPT167050	TWN001	*Solanum*	*lycopersicum*		AFG	300	eim
GPT167470	TWN001	*Solanum*	*lycopersicum*	*cerasiforme*	BOL	300	slc_ecu_per_bol
GPT167530	TWN001	*Solanum*	*lycopersicum*	*cerasiforme*	BOL	300	slc_ecu_per_bol
GPT167760	TWN001	*Solanum*	*lycopersicum*	*cerasiforme*	BOL	300	slc_ecu_per_bol
GPT167770	TWN001	*Solanum*	*lycopersicum*	*cerasiforme*	BOL	300	slc_ecu_per_bol
GPT167950	TWN001	*Solanum*	*lycopersicum*	*cerasiforme*	BOL	300	early_vint
GPT167960	TWN001	*Solanum*	*lycopersicum*		BOL	300	vint_ita
GPT168180	TWN001	*Solanum*	*lycopersicum*		ARG	300	mim
GPT168920	TWN001	*Solanum*	*peruvianum*		CHL	100	green
GPT168980	TWN001	*Solanum*	*peruvianum*		CHL	100	green
GPT169710	TWN001	*Solanum*	*lycopersicum*	*cerasiforme*	COL	300	early_vint
GPT169830	TWN001	*Solanum*	*lycopersicum*	*cerasiforme*	USA	300	early_vint
GPT170280	TWN001	*Solanum*	*lycopersicum*		ARG	300	eim
GPT171560	TWN001	*Solanum*	*lycopersicum*		BRA	300	early_vint
GPT171710	TWN001	*Solanum*	*lycopersicum*	*cerasiforme*	PER	300	slc_ecu_per_bol
GPT172160	TWN001	*Solanum*	*lycopersicum*		TUR	300	eim
GPT172170	TWN001	*Solanum*	*lycopersicum*		TUR	300	eim
GPT173480	TWN001	*Solanum*	*lycopersicum*		IND	300	mim
GPT175600	TWN001	*Solanum*	*lycopersicum*		CRI	300	eim
GPT176410	TWN001	*Solanum*	*lycopersicum*		USA	400	vint_ita
GPT177380	TWN001	*Solanum*	*peruvianum*		PER	100	green
GPT178030	TWN001	*Solanum*	*lycopersicum*		ITA	400	mim
GPT178160	TWN001	*Solanum*	*lycopersicum*		USA	ND	vint_ita
GPT178600	TWN001	*Solanum*	*lycopersicum*		FRA	400	mim
GPT178830	TWN001	*Solanum*	*lycopersicum*		ESP	300	vint_esp

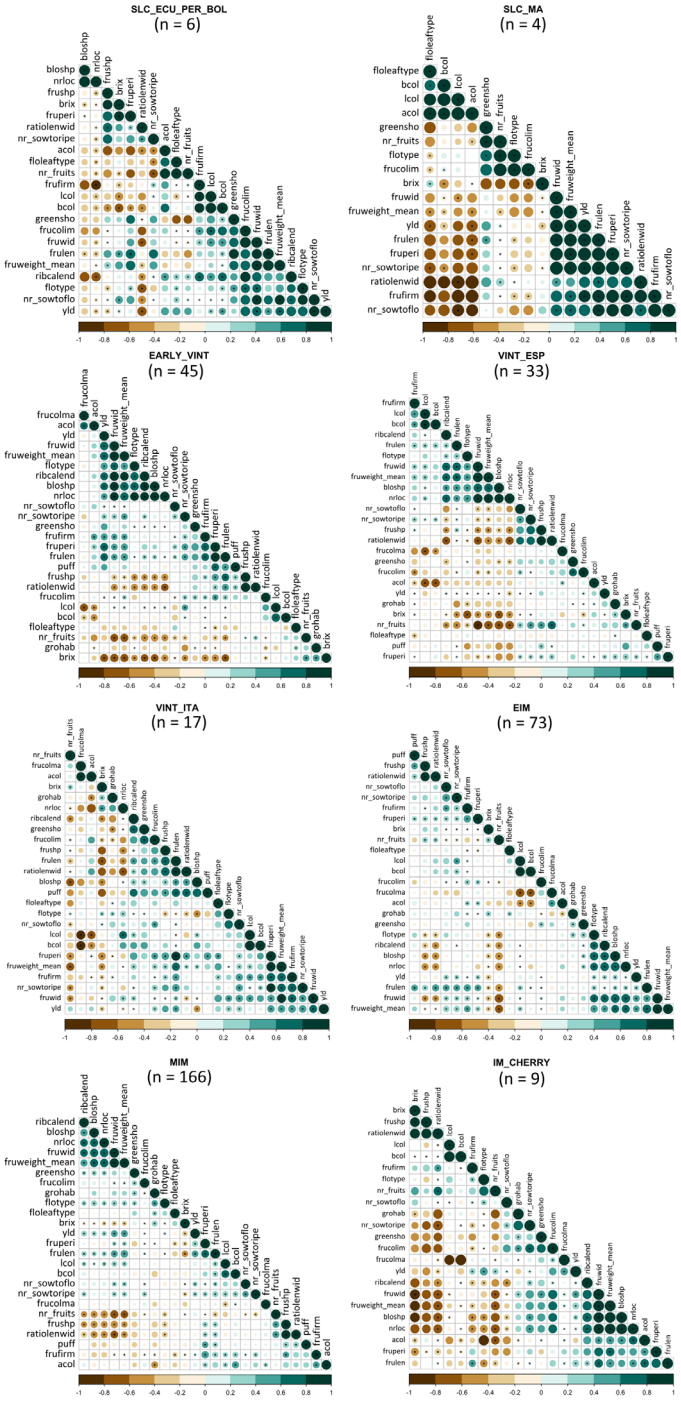
G2PSOL_code	Institute code	Genus	Specie	*Subtaxa*	Country	Biological status of accession	Genetic group
GPT181170	TWN001	*Solanum*	*lycopersicum*		IND	ND	eim
GPT181850	TWN001	*Solanum*	*lycopersicum*		SLV	300	early_vint
GPT184270	TWN001	*Solanum*	*lycopersicum*		SLV	300	early_vint
GPT184360	TWN001	*Solanum*	*lycopersicum*		SLV	300	early_vint
GPT184680	TWN001	*Solanum*	*lycopersicum*		SLV	300	early_vint
GPT184980	TWN001	*Solanum*	*lycopersicum*		SLV	300	early_vint
GPT185270	TWN001	*Solanum*	*lycopersicum*		SLV	300	early_vint
GPT185480	TWN001	*Solanum*	*lycopersicum*		SLV	300	mim
GPT185590	TWN001	*Solanum*	*lycopersicum*		SLV	300	early_vint
GPT185670	TWN001	*Solanum*	*lycopersicum*		SLV	300	early_vint
GPT186070	TWN001	*Solanum*	*lycopersicum*		SLV	300	early_vint
GPT186310	TWN001	*Solanum*	*lycopersicum*		GTM	300	early_vint
GPT186760	TWN001	*Solanum*	*lycopersicum*		BGR	400	mim
GPT187230	TWN001	*Solanum*	*lycopersicum*		unknown	ND	eim
GPT189070	TWN001	*Solanum*	*lycopersicum*		HUN	500	eim
GPT189470	TWN001	*Solanum*	*lycopersicum*	*cerasiforme*	HUN	500	early_vint
GPT189980	TWN001	*Solanum*	*lycopersicum*		CHN	ND	eim
GPT190020	TWN001	*Solanum*	*lycopersicum*		CHN	500	mim
GPT190140	TWN001	*Solanum*	*lycopersicum*		ISR	ND	mim
GPT190580	TWN001	*Solanum*	*lycopersicum*		USA	500	eim
GPT190630	TWN001	*Solanum*	*lycopersicum*		USA	500	eim
GPT190740	TWN001	*Solanum*	*lycopersicum*		USA	500	eim
GPT190810	TWN001	*Solanum*	*lycopersicum*		USA	500	mim
GPT190890	TWN001	*Solanum*	*lycopersicum*		USA	500	mim
GPT190910	TWN001	*Solanum*	*lycopersicum*		USA	500	mim
GPT191800	TWN001	*Solanum*	*lycopersicum*		USA	ND	eim
GPT192040	TWN001	*Solanum*	*pimpinellifolium*		unknown	100	sp
GPT192050	TWN001	*Solanum*	*peruvianum*		unknown	100	green
GPT192580	TWN001	*Solanum*	*lycopersicum*		GTM	300	early_vint
GPT195010	TWN001	*Solanum*	*lycopersicum*		GTM	500	early_vint
GPT195070	TWN001	*Solanum*	*lycopersicum*		RUS	500	mim
GPT195380	TWN001	*Solanum*	*lycopersicum*		USA	500	eim
GPT195430	TWN001	*Solanum*	*lycopersicum*		USA	500	mim
GPT195530	TWN001	*Solanum*	*lycopersicum*		USA	500	mim
GPT195650	TWN001	*Solanum*	*lycopersicum*		USA	500	mim
GPT195870	TWN001	*Solanum*	*lycopersicum*		HRV	500	mim
GPT195980	TWN001	*Solanum*	*lycopersicum*		HRV	500	vint_ita
GPT196040	TWN001	*Solanum*	*lycopersicum*		HRV	500	mim
GPT196520	TWN001	*Solanum*	*habrochaites*		ECU	100	mim
GPT197060	TWN001	*Solanum*	*lycopersicum*		BRA	500	early_vint
GPT197080	TWN001	*Solanum*	*lycopersicum*		BRA	500	vint_esp
GPT197100	TWN001	*Solanum*	*lycopersicum*		BRA	ND	eim
GPT197420	TWN001	*Solanum*	*lycopersicum*		BRA	ND	mim
GPT197630	TWN001	*Solanum*	*lycopersicum*		CAN	400	mim
GPT197890	TWN001	*Solanum*	*lycopersicum*		HRV	500	mim
GPT198250	TWN001	*Solanum*	*lycopersicum*		USA	500	mim
GPT198270	TWN001	*Solanum*	*lycopersicum*		USA	500	eim
GPT198340	TWN001	*Solanum*	*lycopersicum*		USA	500	mim
GPT198410	TWN001	*Solanum*	*lycopersicum*		USA	400	mim
GPT198490	TWN001	*Solanum*	*lycopersicum*		KOR	500	mim
GPT198820	TWN001	*Solanum*	*lycopersicum*	*cerasiforme*	NGA	400	early_vint
GPT199090	TWN001	*Solanum*	*lycopersicum*		USA	500	mim
GPT199780	TWN001	*Solanum*	*lycopersicum*		USA	500	mim
GPT199860	TWN001	*Solanum*	*lycopersicum*		TWN	400	mim
GPT200110	TWN001	*Solanum*	*lycopersicum*		IND	400	mim

G2PSOL_code	Institute code	Genus	Specie	*Subtaxa*	Country	Biological status of accession	Genetic group
GPT200240	TWN001	*Solanum*	*lycopersicum*		PAN	400	mim
GPT200700	TWN001	*Solanum*	*lycopersicum*		USA	500	mim
GPT200870	TWN001	*Solanum*	*lycopersicum*		USA	500	mim
GPT201180	TWN001	*Solanum*	*lycopersicum*		HRV	500	vint_ita
GPT201210	TWN001	*Solanum*	*lycopersicum*		HRV	500	eim
GPT201410	TWN001	*Solanum*	*lycopersicum*	*cerasiforme*	GUF	ND	early_vint
GPT201430	TWN001	*Solanum*	*lycopersicum*	*cerasiforme*	GUF	ND	early_vint
GPT201500	TWN001	*Solanum*	*lycopersicum*	*cerasiforme*	GLP	400	early_vint
GPT201980	TWN001	*Solanum*	*lycopersicum*		ECU	300	mim
GPT201990	TWN001	*Solanum*	*lycopersicum*	*cerasiforme*	ECU	300	slc_ecu_per_bol
GPT202340	TWN001	*Solanum*	*lycopersicum*		HRV	ND	eim
GPT202490	TWN001	*Solanum*	*lycopersicum*		HRV	ND	mim
GPT202680	TWN001	*Solanum*	*lycopersicum*		COL	300	mim
GPT203120	TWN001	*Solanum*	*lycopersicum*		PER	300	mim
GPT203350	TWN001	*Solanum*	*pimpinellifolium*		PER	100	sp
GPT203940	TWN001	*Solanum*	*lycopersicum*		CAN	ND	mim
GPT204460	TWN001	*Solanum*	*lycopersicum*		SLV	300	vint_ita
GPT204530	TWN001	*Solanum*	*lycopersicum*		SLV	300	early_vint
GPT205030	TWN001	*Solanum*	*lycopersicum*	cerasiforme	HND	200	slc_ma
GPT205290	TWN001	*Solanum*	*lycopersicum*	cerasiforme	HND	200	slc_ma
GPT205730	TWN001	*Solanum*	*lycopersicum*	cerasiforme	PAN	400	early_vint
GPT205740	TWN001	*Solanum*	*lycopersicum*		PAN	400	mim
GPT205750	TWN001	*Solanum*	*lycopersicum*		PAN	400	mim
GPT206160	TWN001	*Solanum*	*lycopersicum*		JPN	400	mim
GPT208160	TWN001	*Solanum*	*lycopersicum*		BRA	500	early_vint
GPT208450	TWN001	*Solanum*	*lycopersicum*		CHN	400	mim
GPT210380	TWN001	*Solanum*	*lycopersicum*		CHN	400	mim
GPT210570	TWN001	*Solanum*	*lycopersicum*		CHN	400	mim
GPT210590	TWN001	*Solanum*	*lycopersicum*		CHN	400	mim
GPT211460	TWN001	*Solanum*	*lycopersicum*		CHN	ND	early_vint
GPT211870	TWN001	*Solanum*	*lycopersicum*		HUN	400	mim
GPT212560	TWN001	*Solanum*	*lycopersicum*		ITA	ND	mim
GPT212730	TWN001	*Solanum*	*lycopersicum*		JPN	400	mim
GPT212960	TWN001	*Solanum*	*lycopersicum*		NLD	500	mim
GPT213070	TWN001	*Solanum*	*lycopersicum*		THA	400	eim
GPT213840	TWN001	*Solanum*	*lycopersicum*		NLD	400	mim
GPT213900	TWN001	*Solanum*	*lycopersicum*		NLD	412	mim
GPT214320	TWN001	*Solanum*	*lycopersicum*		CHN	500	mim
GPT214620	TWN001	*Solanum*	*lycopersicum*		COL	300	eim
GPT217020	TWN001	*Solanum*	*lycopersicum*		IDN	ND	mim
GPT217190	TWN001	*Solanum*	*lycopersicum*		IDN	500	early_vint
GPT218580	TWN001	*Solanum*	*pimpinellifolium*		PER	100	sp
GPT220940	TWN001	*Solanum*	*lycopersicum*		unknown	400	mim
GPT221790	TWN001	*Solanum*	*lycopersicum*		unknown	500	eim
GPT221990	TWN001	*Solanum*	*lycopersicum*		USA	400	mim
GPT222020	TWN001	*Solanum*	*lycopersicum*		USA	400	mim
GPT222750	TWN001	*Solanum*	*lycopersicum*		IND	400	eim
GPT222770	TWN001	*Solanum*	*lycopersicum*		USA	400	mim
GPT222880	TWN001	*Solanum*	*lycopersicum*		TWN	400	mim
GPT223690	TWN001	*Solanum*	*lycopersicum*		THA	400	mim
GPT224700	TWN001	*Solanum*	*lycopersicum*		USA	400	mim
GPT224720	TWN001	*Solanum*	*lycopersicum*		PHL	400	mim
GPT225120	TWN001	*Solanum*	*lycopersicum*		LAO	500	mim
GPT225620	TWN001	*Solanum*	*lycopersicum*		VNM	ND	early_vint
GPT226160	TWN001	*Solanum*	*lycopersicum*		KHM	300	early_vint

G2PSOL_code	Institute code	Genus	Specie	*Subtaxa*	Country	Biological status of accession	Genetic group
GPT228360	TWN001	*Solanum*	*lycopersicum*		ARM	500	mim
GPT228460	TWN001	*Solanum*	*lycopersicum*		PHL	ND	early_vint
GPT228650	TWN001	*Solanum*	*lycopersicum*		PHL	ND	early_vint
GPT229430	TWN001	*Solanum*	*lycopersicum*		unknown	500	mim
GPT231400	TWN001	*Solanum*	*lycopersicum*		TWN	400	mim
GPT232470	TWN001	*Solanum*	*lycopersicum*		ISR	400	mim
GPT235010	TWN001	*Solanum*	*galapagense*		ECU	100	gal_chees
GPT235480	TWN001	*Solanum*	*lycopersicum*		PER	ND	mim
GPT235560	TWN001	*Solanum*	*lycopersicum*		PER	ND	mim
GPT235990	TWN001	*Solanum*	*pimpinellifolium*		unknown	100	mim
GPT236510	TWN001	*Solanum*	*lycopersicum*	cerasiforme	GTM	200	early_vint
GPT236730	TWN001	*Solanum*	*lycopersicum*	cerasiforme	PER	200	early_vint
GPT236740	TWN001	*Solanum*	*lycopersicum*	cerasiforme	PER	200	early_vint
GPT237080	TWN001	*Solanum*	*lycopersicum*	cerasiforme	ECU	200	sp
GPT237980	TWN001	*Solanum*	*lycopersicum*	cerasiforme	HND	200	slc_ma
GPT238970	TWN001	*Solanum*	*lycopersicum*		USA	500	vint_ita
GPT239410	ESP026	*Solanum*	*habrochaites*		ECU	100	green

Supplementary Table 2. Contribution of the main components to the analyzed phenotypic traits.

Descriptor	Abbreviation	PC1	PC2	PC3
Plant growth habit	grohab	-0.01153	-0.29188	0.160337
Inflorescence	flotype	0.226722	-0.19836	0.043499
Green shoulder	greensho	0.076838	-0.18588	0.00781
External immature fruit color	frucolim	0.144721	0.164993	-0.04763
Fruit predominant shape	frushp	-0.15209	0.18615	0.164717
Blossom end scar	bloshp	0.313149	-0.18632	-0.05109
Ribbing at calyx end	ribcalend	0.296766	-0.13325	-0.08036
Puffiness appearance	puff	0.034254	0.232612	0.003251
Fruit firmness	frufirm	0.107026	0.349027	0.05261
L	lcol	0.139683	0.031718	0.602171
a	acol	0.025397	0.237772	-0.45145
b	bcol	0.084549	0.136242	0.573643
°Brix	brix	-0.25843	-0.19027	0.105191
Pericarp thickness	fruperi	0.166795	0.398083	-0.0364
Fruit length	frulen	0.260361	0.305764	0.048168
Fruit width	fruwid	0.371617	-0.02982	-0.00916
Ratio length/width	ratiolenwid	-0.17621	0.32002	0.083606
Number of Locules	nrloc	0.305595	-0.21814	-0.03397
Nº commercial fruits	nr_fruits	-0.28445	0.017238	0.079454
Fruit weight mean	fruweight_mean	0.343232	-0.01881	0.058524
Yield	yld	0.220836	0.186637	-0.05418



Supplementary Figure 1. Correlation matrix of 25 vegetative, inflorescence, fruit and agronomic traits considering the genetic groups. Bonferroni's correction was applied to calculate the significance (p-value). Significance for $p \leq 0.05$ is marked with an asterisk. *n* = number of samples.