UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Development and exploitation of MAGIC experimental populations in eggplant and tomato genetics and breeding

Ph.D. dissertation by

Andrea Arrones Olmo

Advisors

Dr. Santiago Vilanova Navarro

Dr. Pietro Gramazio

Valencia, 15 July 2024

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Development and exploitation of MAGIC experimental populations in eggplant and tomato genetics and breeding

Ph.D. dissertation by

Andrea Arrones Olmo

Advisors

Dr. Santiago Vilanova Navarro

Dr. Pietro Gramazio

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

Valencia, 15 July 2024

Andrea Arrones Olmo is grateful to Spanish Ministerio de Ciencia, Innovación y Universidades for a pre-doctoral (FPU18/01742) contract.

The work presented in this thesis was also supported by the Ministerio de Ciencia, Innovación y Universidades, Agencia Estatal de Investigación and Fondo Europeo de Desarrollo Regional (grant RTI-2018-094592-B-I00 from MCIU/AEI/FEDER, UE; grants PID2020-118627RB-I00 and PID2021-128148OB-I00 from MCIN/AEI/10.13039/501100011033 and "ERDF A way of making Europe"; grants TED2021-129296B-I00 and PDC2022-133513-I00 from MCIN/AEI/10.13039/501100011033 and European Union Next Generation EU/ PRTR), Conselleria d'Innovació, Universitats, Ciència i Societat Digital (Generalitat Valenciana, Spain) grant CIPROM/2021/020, European Union's Horizon, 2020 Research and Innovation Programme under grant agreement No. 677379 (G2P-SOL project: Linking genetic resources, genomes and phenotypes of Solanaceous crops) and No. 101000716 (HARNESSTOM: Harnessing the value of tomato genetic resources for now and the future), and Horizon Europe Programme under grant agreement No. 10194738 (PRO-GRACE: Promoting a plant genetic resource community for Europe).

Agradecimientos

Ahora sí hay una doctora en la sala. He superado esta odisea de berenjenas y tomates. – *"Vos ho conte:* cinc *anys morint per un projecte. Consciència neta i actes que sí que són exemples, perquè el soroll distrau, i ací hem buscat silenci. Respecte per la gent que observa el temple."* – Pero el mérito no es solo mío, sino también de los que me han acompañado en el camino.

Mi familia. Gracias por soportarme. – *"Toquen a la porta i alegría. Era* Andrea*, com venia. Com elefant en catxarreria."*

Juan. Gracias por estar siempre. – *"T'estime, t'estimo, t'estim."*

Mis directores de tesis. Gracias por confiar en mí y por haberme ayudado a crecer como investigadora. – *"Crea, lluita maniquea. El mon plorant, tu davant, quet. Crea, puja la marea, obri les vel·les i arrea."*

Mis compañeros y buenos amigos del COMAV. Gracias por compartir conmigo siensia, cotilleos, risas y café. Mucho café. ¡Ánimo! – *"Que no tot en la vida és faena, que morir és no viure lluitant."*

Mis mejoradores. Gracias por llegar pisando fuerte. – *"Mis amigos saben bien quien soy y de que pie cojeo. Mis amigos saben bien quien soy, con quien me codeo. Mis amigos saben bien quien soy, donde merodeo."*

Mis biotec, OHANA, aunque hayan pasado la friolera de siete años desde que acabamos la carrera. Gracias por seguir ahí. – *"En las malas y en las buenas, las de cal y las de arena, sé con quién puedo contar. No quiero echaros de menos, mueran la reja y la pena, venga el sereno y reparta bondad."*

Mi equipo. Gracias por aguantar los "*Andrea nunca broma*". ¡Amunt! – *"L'esperança als ulls, les banderes al vent i les ganes de viure de la meua gent."*

Gracias por haber formado parte de esto. Gracias por vuestro apoyo incondicional durante estos años. Lo mejor está aún por venir. – *"Se'n va el dolor amb la olor a mar, les primaveres demanen vida, l'estiu arriba."*

– Obrint pas, La gossa sorda, ZOO & SFDK

TABLE OF CONTENTS

LIST OF ABBREVIATIONS v

ABSTRACT ix

RESUMEN xi

INTRODUCTION 17

General introduction of breeding in Solanaceae crops 19

Breeding in eggplant 22

Breeding in tomato 27

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 33

Abstract 34

1. Introduction 35

2. Overview of experimental populations and germplasm collection for traits dissection 36

3. Advantages and limitations of MAGIC populations 41

4. MAGIC development strategies 45

5. Analysis software for genetic gap construction and QTL mapping 50

6. An appraisal of MAGIC populations developed and evaluated 56

8. Conclusions 64

OBJECTIVES 75

RESULTS 81

Chapter I: Eggplant MAGIC population for multiple fruit traits breeding 83

Newly developed MAGIC population allows identification of strong associations and candidate genes for anthocyanin pigmentation in eggplant 85

Abstract 86

1. Introduction 87

2. Materials and methods 89

3. Results 94

4. Discussion 103

5. Conclusion 106

Mutations in the *SmAPRR2* transcription factor suppressing chlorophyll pigmentation in the eggplant fruit peel are key drivers of a diversified colour palette 117

Abstract 118

1. Introduction 119

2. Materials and methods 120

3. Results 124

4. Discussion 133

The irregular fruit green netting: An eggplant domestication trait controlled by the *SmGLK2* gene with implications in fruit colour diversification 147

Abstract 148

1. Introduction 149

2. Materials and methods 150

3. Results 156

4. Discussion 163

Chapter II: Development of a novel inter-specific tomato MAGIC population 173

A novel tomato inter-specific (*Solanum lycopersicum* var. *cerasiforme* and *S. pimpinellifolium*) MAGIC population facilitates trait association and candidate gene discovery in untapped exotic germplasm 175

1. Introduction 177

2. Materials and methods 178

3. Results 183

4. Discussion 197

GENERAL DISCUSSION 213

GENERAL CONCLUSIONS 225

GENERAL REFERENCES 231

LIST OF ABBREVIATIONS

AB advanced backcross
ABA abscisic acid
AMOVA analysis of molecular variance

BC backcross
bHLH basic-helix-loop-helix
BIL backcross inbred line
BLINK Bayesian-information and linkage-disequilibrium iteratively nested keyway
BLUP best linear unbiased prediction
BSA-Seq bulked segregant analysis by sequencing
BWA-MEM Burrows-Wheeler aligner - maximal exact match

CDS coding sequence
CSLM confocal laser scanning microscopy
CWR crop wild relative

DAS *Solanum dasyphyllum*
DH doubled haploid
DHY double hybrid

F_x filial
FA fruit anthocyanin
FC fruit chlorophyll
FDR false discovery rate
FN fruit netting

G_x generation
GAPIT genome association and prediction integrated tool
GBS genotyping-by-sequencing
GLM general linear model
GMS genetic male sterility
GP genepool
GQ genotype quality
GWA genome-wide association
GWAS genome-wide association study

H heterozygotes
HRM high resolution melting

IC inter-cross generation
IGV integrative genomics viewer
IL introgression line
INC *Solanum incanum*
INS *Solanum insanum*

KNNi k-nearest neighbour genotype imputation method

LD linkage disequilibrium
LOD limit of detection

MAGIC multi-parent advanced generation intercross
MAGReS multi-parent advanced generation recurrent selection
MAS marker-assisted selection
MEL *Solanum melongena*
MLM mixed linear model
MpCD multi-parent cross designs
MVMP multivariate multi-parent
MYB myeloblastosis

NIL near isogenic line
NCBI national center for biotechnology information

PA plant anthocyanin
PC principal component
PCA principal component analysis
PCR polymerase chain reaction
PPB participatory plant breeding

PPP public–private partnerships
PUC pigmentation under the calyx

QTL quantitative trait locus

RIL recombinant inbred line
ROS reactive oxygen species

S_x generation of selfing
S3MEGGIC S3 MAGIC eggplant incanum
SILEX silica matrix extraction
SLC *Solanum lycopersicum* var. *cerasiforme*
SNP single-nucleotide polymorphism
SP *Solanum pimpinellifolium*
SPET single primer enriched technology
SSD single-seed descent

TASSEL trait analysis by association, evolution and linkage
TF transcription factor
ToMAGIC tomato MAGIC

UTR untranslated region

VIGS virus-induced gene silencing

WGAIM whole-genome average interval mapping
WGS whole genome sequencing

GENES:

ANT aintegumenta
AP1/FUL apetalla 1/fruitfull
AP2/ERF apetala 2/ethylene response factor
AP3/DEF apetalla 3/deficiens
APRR2 *Arabidopsis* pseudo response regulator 2

bHLH **TF** basic-helix-loop-helix transcription factor

COP1 constitutive photomorphogenic 1

FT flowering locus T
FW2.2 fruit weight 2.2

GLK2 golden 2-like MYB

IAA9 indole-3-acetic acid 9

J jointless

MYB113 myeloblastosis 113
MYB-ATV myeloblastosis-atroviolacea

SFT single-flower truss
SP1 self-pruning interacting protein 1
SPA suppressor of phya

U uniform ripening

ABSTRACT

The Solanaceae family includes crops of global economic importance such as eggplant (*Solanum melongena* L.) and tomato (*S. lycopersicum* L.). These crops play an important role in the diet of most of the world population and are characterised by their nutraceutical properties thanks to their rich composition of bioactive compounds. Although both crops show positive trends in production, yield, and consumption, they are currently under constant threat from multiple abiotic and biotic stresses, due to the genetic narrowing and uniformity of modern high-yielding commercial varieties. As a consequence of domestication processes and plant breeding programs, usually there has been a significant loss of genetic variability in cultivated species, leading to an increased risk of yield losses when facing threats and a decrease in sources of variability for crop improvement. In this respect, the study and introduction of crop wild relative species in breeding programs represent a source of novel genetic diversity.

Therefore, in the present doctoral thesis, it is proposed the leverage of the hidden potential of the wild relatives by constructing inter-specific Multi-parent Advanced Generation InterCross (MAGIC) experimental populations combining cultivated, semi-domesticated, and wild germplasm to contribute to rescuing some of the lost variability in eggplant and tomato and the identification of candidate genes and causative polymorphisms controlling interesting quantitative traits.

In the first chapter of this thesis, we aimed to present the first eggplant MAGIC population so far, developed from seven accessions of common eggplant from different origins and one wild species *S. incanum* accession. The eight founders used for the construction of this population showed a wide genetic diversity according to genome resequencing data, phenotypic, fruit, agronomic and stress resistance traits. This inter-specific population was phenotyped for different traits and genotyped through the single primer enriched technology (SPET) platform. Results were used for the elucidation of the genes responsible for different fruit pigmentation patterns including the biosynthesis of uniform anthocyanins and chlorophylls, and the irregular fruit green netting. Understanding the mechanisms by which fruit pigments are synthesized could be extremely useful to foster future breeding programs focused not only on specific consumers' demand for fruit diversification but also on the development of novel varieties exhibiting enhanced nutritional quality and resilience to tackle the forthcoming environmental challenges. This could be possible thanks to the relationship of pigments with antioxidant properties, photosynthesis, and the stress biology of plants, among others.

In the second chapter of this thesis, we addressed the development of a tomato MAGIC population using four accessions of the semi-domesticated closest wild relative *S. lycopersicum* var. *cerasiforme*, and four accessions of the *S. pimpinellifolium*, which is the ancestor of the cultivated tomato. These species have

been little used in breeding and their full potential has not been exploited despite their high genetic diversity, their wide climatic and ecological variability and their closeness to the cultivated tomato. Previous MAGIC populations have been developed combining cherry type and larger fruit accessions of common tomato. However, the results indicate that these eight founders do not represent the Andean genetic variability of *S. l.* var. *cerasiforme* lost during the domestication process and therefore, the proposed population is complementary to that already developed. The final inter-specific population was also genotyped through a SPET panel and a proof-of-concept for testing the potential of the population was conducted by phenotyping it for fruit size, plant pigmentation, leaf morphology, and earliness traits. As a result, previously identified and novel candidate genes were identified, demonstrating that the population is a valuable resource for genetic research in tomato.

Overall, both inter-specific eggplant and tomato MAGIC populations constitute a useful resource for genetic studies to the scientific community. Final individuals include a huge phenotypic and genetic diversity, which could be directly exploited not only for their intrinsic potential but also as donors carrying combinations of multiple traits of interest. These populations open up doors to future assays of screening for interesting alleles in a multi-environment or multi-trait background.

RESUMEN

La familia Solanaceae incluye cultivos de gran importancia económica mundial como la berenjena (*Solanum melongena* L.) y el tomate (*S. lycopersicum* L.). Estos cultivos desempeñan un papel importante en la dieta de la mayoría de la población mundial y se caracterizan por sus propiedades nutracéuticas gracias a su rica composición de compuestos bioactivos. Aunque ambos cultivos muestran tendencias positivas en producción, rendimiento y consumo, actualmente se encuentran bajo la amenaza constante de múltiples estreses abióticos y bióticos, debido al estrechamiento y uniformidad genética de las variedades comerciales modernas de alto rendimiento. Como consecuencia de los procesos de domesticación y de los programas de mejora, habitualmente se ha producido una importante pérdida de variabilidad genética en las especies cultivadas, lo que conlleva un aumento del riesgo de pérdidas de rendimiento frente a las amenazas y una disminución de las fuentes de variabilidad para la mejora del cultivo. En este sentido, el estudio e introducción de especies silvestres en los programas de mejora representa una fuente de nueva diversidad genética.

Por ello, en la presente tesis doctoral, se propone el aprovechamiento del potencial oculto de los parentales silvestres mediante la construcción de poblaciones experimentales interespecíficas Multi-parent Advanced Generation InterCross (MAGIC) que combinen germoplasma cultivado, semidomesticado y silvestre para contribuir en el rescate de parte de la variabilidad perdida en berenjena y tomate y la identificación de genes candidatos y polimorfismos causales que controlen caracteres cuantitativos de interés.

En el primer capítulo de esta tesis, nos propusimos presentar la primera población MAGIC de berenjena desarrollada hasta el momento, a partir de siete accesiones de berenjena común de diferentes orígenes y una accesión de la especie silvestre *S. incanum*. Los ocho parentales utilizados para la construcción de esta población mostraron una amplia diversidad genética según los datos de resecuenciación genómica, fenotípica, de fruto, agronómica y de caracteres de resistencia al estrés. Esta población interespecífica fue fenotipada para diferentes caracteres y genotipada mediante la plataforma de Single Primer Enrichment Technology (SPET). Los resultados se utilizaron para la elucidación de los genes responsables de los diferentes patrones de pigmentación del fruto, incluyendo la biosíntesis de antocianinas y clorofilas uniformes, y la malla verde irregular del fruto. Comprender los mecanismos por los que se sintetizan los pigmentos de los frutos podría ser de gran utilidad para impulsar futuros programas de mejora centrados no sólo en la demanda específica de diversificación de frutos por parte de los consumidores, sino también en el desarrollo de nuevas variedades que presenten

una mayor calidad nutricional y resistencia para hacer frente a los retos medioambientales que se avecinan. Esto es posible gracias a la relación de los pigmentos con las propiedades antioxidantes, la fotosíntesis y la biología del estrés de las plantas, entre otras.

En el segundo capítulo de esta tesis, se abordó el desarrollo de una población MAGIC de tomate utilizando cuatro accesiones de la especie silvestre más cercana semidomesticada *S. lycopersicum* var. *cerasiforme*, y cuatro accesiones de *S. pimpinellifolium*, el ancestro del tomate cultivado. Estas especies se han utilizado poco en la mejora genética del tomate y no se ha explotado todo su potencial a pesar de su gran diversidad genética, su amplia variabilidad climática y ecológica y su proximidad al tomate cultivado. Las poblaciones MAGIC anteriores se desarrollaron combinando accesiones de tomate común de tipo cherry y de fruto más grande. Sin embargo, los resultados indican que estos ocho parentales no representan la variabilidad genética andina de *S. l.* var. *cerasiforme* perdida durante el proceso de domesticación y, por lo tanto, la población propuesta es complementaria a la ya desarrollada. La población interespecífica final también se genotipó mediante un panel SPET y se realizó una prueba de concepto para comprobar el potencial de la población fenotipándola para tamaño de fruto, pigmentación de la planta, morfología de la hoja y precocidad. Como resultado, se identificaron genes candidatos previamente identificados y otros nuevos, lo que demuestra que la población es un recurso valioso para la mejora genética del tomate.

En conjunto, ambas poblaciones interespecíficas MAGIC de berenjena y tomate constituyen un recurso útil para estudios genéticos para la comunidad científica. Los individuos finales incluyen una enorme diversidad fenotípica y genética, que podría explotarse directamente no sólo por su potencial intrínseco, sino también como donantes portadores de combinaciones de múltiples caracteres de interés. Estas poblaciones abren las puertas a futuros ensayos de detección de alelos interesantes en un contexto multiambiental o multicarácter.

RESUM

La família Solanaceae inclou cultius de gran importància econòmica mundial com l'albergina (*Solanum melongena* L.) i la tomaca (*S. lycopersicum* L.). Aquests cultius exerceixen un paper important en la dieta de la majoria de la població mundial i es caracteritzen per les seues propietats nutracèutiques gràcies a la seua rica composició de compostos bioactius. Encara que tots dos cultius mostren tendències positives en producció, rendiment i consum, actualment es troben sota l'amenaça constant de múltiples estressos abiòtics i biòtics, a causa de l'estrenyiment i uniformitat genètica de les varietats comercials modernes d'alt rendiment. A conseqüència dels processos de domesticació i dels programes de millora, habitualment s'ha produït una important pèrdua de variabilitat genètica en les espècies cultivades, la qual cosa comporta un augment del risc de pèrdues de rendiment enfront de les amenaces i una disminució de les fonts de variabilitat per a la millora del cultiu. En aquest sentit, l'estudi i introducció d'espècies silvestres en els programes de millora representa una font de nova diversitat genètica.

Per això, en la present tesi doctoral, es proposa l'aprofitament del potencial ocult dels parentals silvestres mitjançant la construcció de poblacions experimentals interespecífiques Multi-parent Advanced Generation InterCross (MAGIC) que combinen germoplasma cultivat, semidomesticat i silvestre per a contribuir en el rescat de part de la variabilitat perduda en albergina i tomaca i la identificació de gens candidats i polimorfismes causals que controlen caràcters quantitatius d'interès.

En el primer capítol d'aquesta tesi, ens vam proposar presentar la primera població MAGIC d'albergina desenvolupada fins ara, a partir de set accessions d'albergínia comuna de diferents orígens i una accessió de l'espècie silvestre *S. incanum*. Els huit parentals utilitzats per a la construcció d'aquesta població van mostrar una àmplia diversitat genètica segons les dades de reseqüenciació genòmica, fenotípica, de fruit, agronòmica i de caràcters de resistència a l'estrès. Aquesta població interespecífica va ser fenotipada per a diferents caràcters i genotipada mitjançant la plataforma de Single Primer Enrichment Technology (SPET). Els resultats es van utilitzar per a l'elucidació dels gens responsables dels diferents patrons de pigmentació del fruit, incloent-hi la biosíntesi d'antocianines i clorofil·les uniformes, i la malla verda irregular del fruit. Comprendre els mecanismes pels quals se sintetitzen els pigments dels fruits podria ser de gran utilitat per a impulsar futurs programes de millora centrats no sols en la demanda específica de diversificació de fruits per part dels consumidors, sinó també en el desenvolupament de noves varietats que presenten una major qualitat nutricional i resistència per a fer front als reptes mediambientals que s'aveïnen. Això és possible gràcies a la relació dels

pigments amb les propietats antioxidants, la fotosíntesi i la biologia de l'estrès de les plantes, entre altres.

En el segon capítol d'aquesta tesi, es va abordar el desenvolupament d'una població MAGIC de tomaca utilitzant quatre accessions de l'espècie silvestre més pròxima semidomesticada *S. lycopersicum* var. *cerasiforme*, i quatre accessions de *S. pimpinellifolium*, l'ancestre de la tomaca cultivada. Aquestes espècies s'han utilitzat poc en la millora genètica de la tomaca i no s'ha explotat tot el seu potencial malgrat la seua gran diversitat genètica, la seua àmplia variabilitat climàtica i ecològica i la seua proximitat a la tomaca cultivada. Les poblacions MAGIC anteriors es van desenvolupar combinant accessions de tomaca comuna de tipus cherry i de fruit més gran. No obstant això, els resultats indiquen que aquests huit parentals no representen la variabilitat genètica andina de *S. l.* var. *cerasiforme* perduda durant el procés de domesticació i, per tant, la població proposada és complementària a la ja desenvolupada. La població interespecífica final també es va genotipar mitjançant un panell SPET i es va realitzar una prova de concepte per a comprovar el potencial de la població fenotipant-la per a grandària de fruit, pigmentació de la planta, morfologia de la fulla i precocitat. Com a resultat, es van identificar gens candidats prèviament descrits i altres nous, la qual cosa demostra que la població és un recurs valuós per a la millora genètica de la tomaca.

En conjunt, totes dues poblacions interespecífiques MAGIC d'albergina i tomaca constitueixen un recurs útil per a estudis genètics per a la comunitat científica. Els individus finals inclouen una enorme diversitat fenotípica i genètica, que podria explotar-se directament no sols pel seu potencial intrínsec, sinó també com a donants portadors de combinacions de múltiples caràcters d'interès. Estes poblacions obrin les portes a futurs assajos de detecció d'al·lels interessants en un context multiambiental o multicaràcter.

INTRODUCTION

The wizarding world of MAGIC populations.

General introduction of breeding in Solanaceae crops

The Solanaceae family is the largest group of vegetable crops hosting approximately 90 genera and 4,000 species growing worldwide in habitats from rainforests to the driest deserts (Knapp, 2002). They include important food crops such as eggplant (*Solanum melongena* L.) and tomato (*S. lycopersicum* L.). Therefore, the Solanaceae family plays a key role in agricultural and food economics and scientific research (Li *et al.,* 2021a).

The Solanaceous fruits are very diverse in terms of morphology, exhibiting considerable variation in size, shape, and colour (Knapp, 2002). However, domestication bottlenecks and the replacement of traditional landraces with modern cultivars have generally resulted in the loss of variation in crops, which has been defined as genetic erosion (van de Wouw *et al.,* 2010). During domestication, human selection only benefited from a limited number of individuals and therefore, a subset of the huge genetic variation present in the wild, leading to loss of diversity (Dempewolf *et al.,* 2012; Meyer and Purugganan, 2013). In addition, crop breeding programs have usually been focused for years on producing highly inbred varieties with higher yields (Smýkal *et al.,* 2018; Kersey *et al.,* 2020). This has also led to a reduction of genetic variation due to a reduced number of modern varieties used under intensive agriculture (Fu and Dong, 2015).

The expected increasing demand for plant products within a climate change scenario together with the consumers' demand for fruit diversification, represents a challenge for plant breeders (Prohens *et al.,* 2017). In this respect, wild relatives and traditional landraces constitute essential germplasm materials for the rescue of this endangered diversity (Aubriot and Daunay, 2019). Breeders have long recognized the enormous potential of crop wild relatives (CWRs) for breeding owing to them represent a source of novel genetic diversity (Prohens *et al.,* 2017). Until now, CWRs have predominantly been exploited in breeding to introgress a small number of major genes, mainly related to disease resistance or tolerance (Smýkal *et al.,* 2018). However, they are also interesting for their tolerance to abiotic stresses, as most of them are able to grow with low inputs and in extreme environments (Gasparini *et al.,* 2021). It has been reported that the introduction of CWRs in breeding programs has improved crop performance for years (Hajjar and Hodgkin, 2007; Schouten *et al.,* 2019).

However, to successfully introduce the interesting genes present in CWRs into cultivated crops, it is crucial to detect these genes and strategically guide their introgression. In this context, mapping or segregating populations consisting of individual progenies originated from two or more parents (Boopathi, 2020), become essential tools. The development of experimental populations has been of great

relevance to the scientific community (Kumar *et al.,* 2017). Different types of bi-parental and multi-parental populations have been used in genetic studies to map and identify QTLs and major genes controlling key traits in many crops (Doerge, 2002). They integrate genomic, phenotypic and germplasm resources suitable for increasing crop diversity through the combination of different founder backgrounds (Scott *et al.,* 2020). An appropriate selection of founders is crucial for a successful mapping study. Normally, selected founders are highly homozygous and differ in a number of economic and agronomically traits of interest (Boopathi, 2020). The introduction of CWRs as founders of experimental populations could be a way of increasing the genetic diversity by including multiple genomic fragments or introgressions from CWRs species into a cultivated background genome. The final inter-specific populations carrying CWR introgressions could be easily incorporated into breeding programs with the aim of widening the genetic base and diversity of the crop (Figure 1).

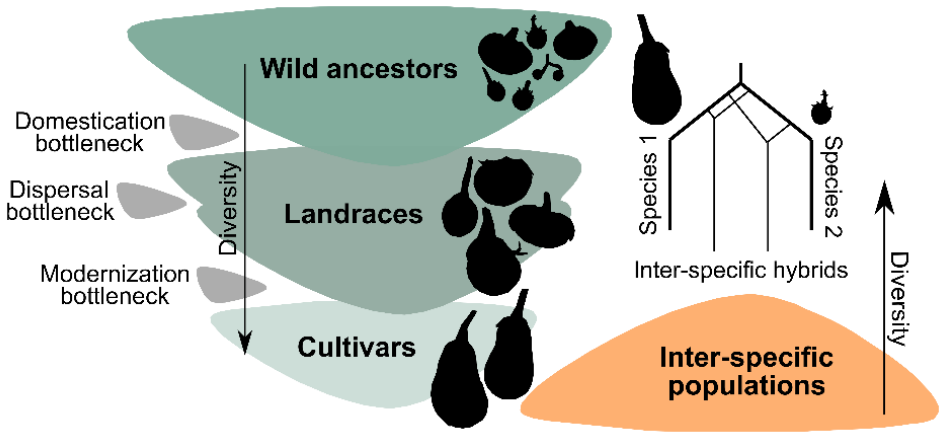


Figure 1. Crop diversity bottlenecks occurred from wild ancestors to modern cultivars and the expected diversity increased by developing inter-specific populations using crop wild relatives (CWRs) as donors of novel genetic diversity. Adapted from Van De Wouw *et al*. (2009).

However, the use of CWRs in breeding programs is often not feasible due to experimental drawbacks. Although their introduction poses difficulties in the development of experimental populations, different methodologies have been implemented to overcome them. On the one hand, inter-specific hybridization is often linked to crossing barriers, low hybrid fertility, or sterility. However, specific techniques could be used to potentially overcome the pre-zygotic (e.g., stigma excision or use of pollen mixtures) and post-zygotic (e.g., embryo rescue or in vitro

culture) reproductive barriers. Sometimes, the use of bridge species or polyploidization may be needed to restore fertility (Figure 2) (Prohens *et al.,* 2017). On the other hand, CWRs usually present some deleterious genes or undesirable agronomic traits such as low yield or unpleasant flavour. These wild traits could be introduced unintentionally along with the beneficial traits, the so-called linkage drag, which is of no interest to either breeders, farmers, or consumers (Meyer and Purugganan, 2013). Parallel high-throughput genotyping techniques could minimize linkage drag by providing molecular markers which precisely target the introgressed wild fragment in the very early stages of plant development (Voss-Fels and Snowdon, 2016).

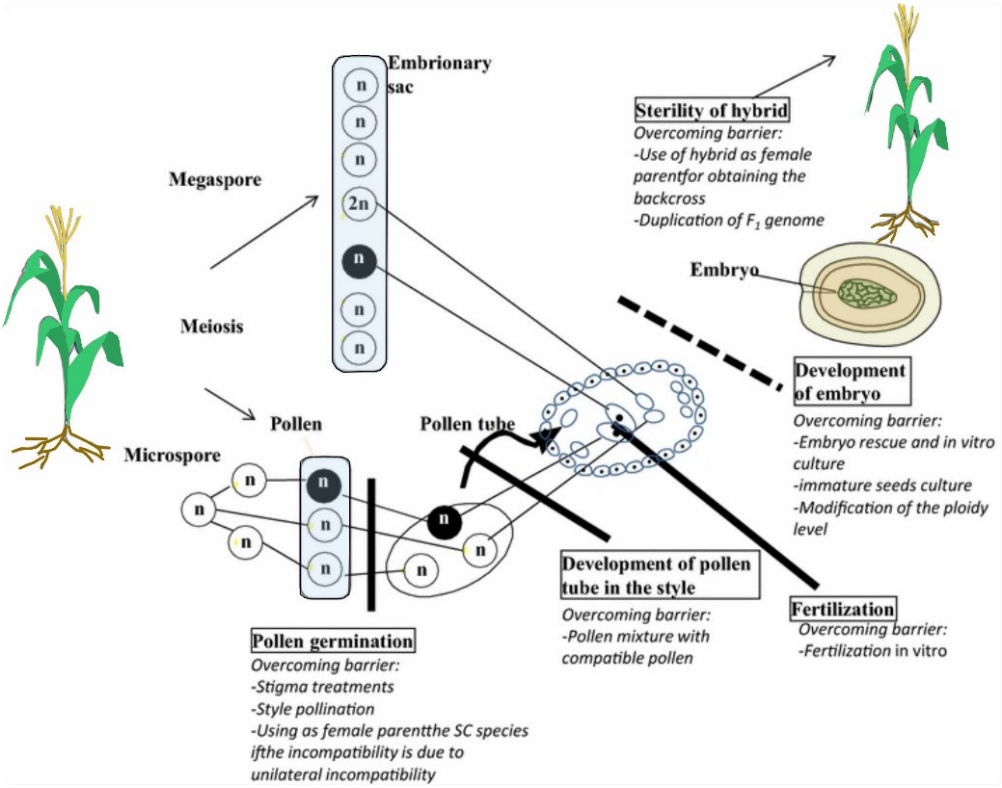


Figure 2. Pre- and post-zygotic barriers challenging inter-specific hybridization with CWRs and strategies to overcome them. Source: Prohens *et al.,* 2017.

In conclusion, although the incorporation of CWRs in breeding programs and population development schemes can be challenging, they represent an efficient use of the available genetic resources. The introgression of wild fragments into cultivated background together with high-throughput phenotyping and genotyping tools result in the development of new elite materials, recovering a lost part of

ancestral diversity and taking advantage of their multiple benefits (Prohens *et al.,* 2017).

Breeding in eggplant

Eggplant (*S. melongena* L.) is an important vegetable crop worldwide (Li *et al.,* 2021a). Among the Solanaceae family, eggplant is ranked in the third position in harvested area and production after tomato and potato (FAOSTAT, 2021). It is exclusively native to the Old World because, unlike other Solanaceae, it was domesticated in Africa, Asia, and Europe (Chapman, 2019). It is a member of the large clade *Leptostemonum* (subgenus *Leptosmonum* Bitter), commonly known as the 'spiny *Solanum*', defined by the possession of stellate trichomes, long tapering anthers and usually prickles (Vorontsova *et al.,* 2013).

Clarifying the routes of eggplant domestication has until today been limited due to partial knowledge on the identification of eggplant wild relatives, among other reasons (Page *et al*, 2019). It has been considered for years that eggplant was domesticated from the wild *S. incanum* due to its morphological similarity to and close affinity with cultivated eggplant and other quite similar wild relatives (Lester and Hasan, 1991; Knapp *et al.,* 2013). Studies also suggested that eggplant had two centres of domestication due to cultivated accessions were clearly split into Eastern (Chinese, Malaysian, Indonesian, Thai, and Vietnamese) and Western (Indian, Madagascan, and Sri Lankan) groups (Meyer *et al.,* 2012). However, the latest studies based on genotyping-by-sequencing (GBS) and RNAseq data of a broad sampling of wild, weedy, and cultivated eggplant revealed that a single domestication event occurred in South-East Asia from the wild progenitor *S. insanum* (Page *et al.,* 2019). Recent studies suggest a reduction of genetic diversity of about 50% during *S. melongena* domestication (Page and Chapman, 2021), which presents an opportunity to drive the field forward the genetic eggplant breeding.

Over the years, great efforts have been made to elucidate genetic relationships among the cultivated eggplant and related species to understand the evolutionary divergence of the crop. The unavailability of genomic sequences from wild relatives has made this task difficult (Oladosu *et al.,* 2021). Eggplant and its wild relatives have been classified according to their crossability with cultivated eggplant into primary, secondary, and tertiary genepools (Figure 3, Harlan and de Wet, 1971). The primary genepool (GP1) only consists of a single species, the wild ancestor *S. insanum*. *S. melongena* and *S. insanum* can be crossed without difficulty and hybrids are completely fertile (Plazas *et al.,* 2016). The secondary genepool (GP2) is constituted of more than 40 wild species within the subgenus *Leptostemonum*,

divided into the Eggplant clade, the Climbing clade, and the Anguivi grade (Syfert *et al*. 2016; Oladosu *et al.,* 2021), with a significant reduction in the crossability, fertility, and viability of the inter-specific hybrids (Kowassi *et al.,* 2016; Plazas *et al.,* 2016). The tertiary genepool (GP3) includes New World species distantly from the cultivated eggplant. Hybridization is rarely successful and specific cross-breeding techniques are required to succeed (Kouassi *et al*. 2016; Plazas *et al*. 2016). Although the use of the GP3 CWRs for crop improvement is challenging, GP1 and GP2 CWRs could be easily introduced in breeding programs, or they can even be used directly (Prohens *et al.,* 2017).

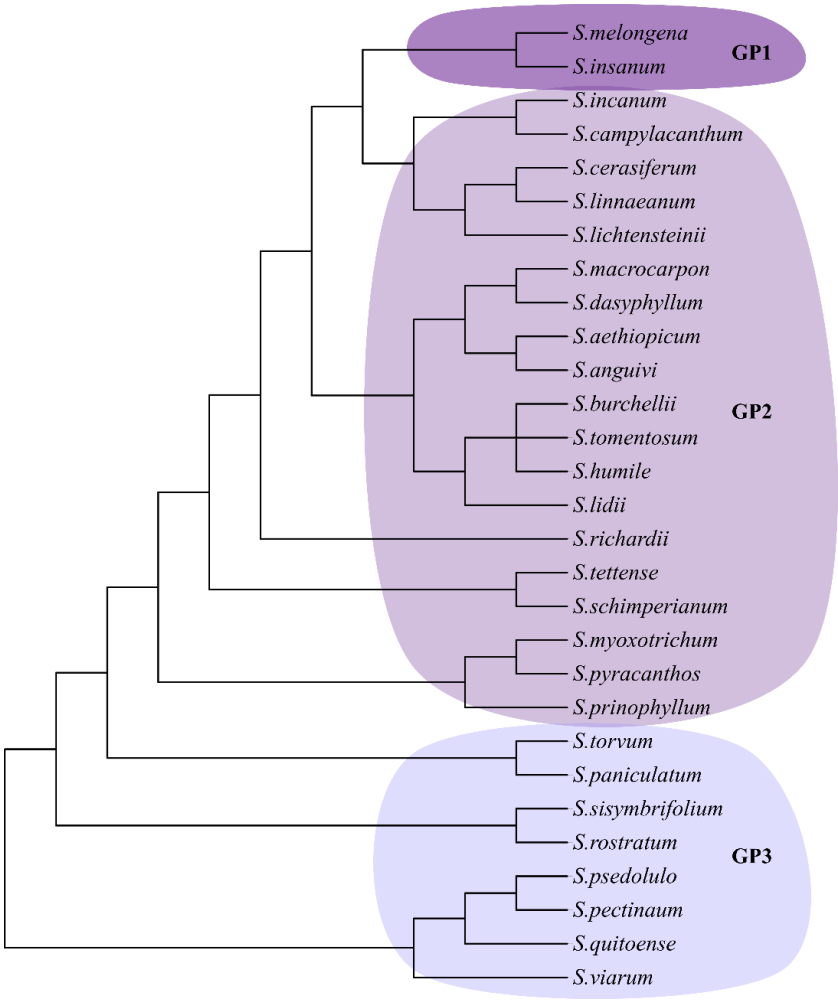


Figure 3. Phylogenetic tree representing relationships of the most relevant species from the primary (GP1), secondary (GP2), and tertiary (GP3) genepools of eggplant. Adapted from Vorontsova *et al*. (2013), Knapp *et al*. (2013, 2019), Syfert *et al*. (2016).

Traditional eggplant breeding programs have been mainly focused on agronomic traits such as fruit uniformity, increased yield, and resistance or tolerance to biotic and abiotic stress (Oladosu *et al.,* 2021). Recently, eggplants have received much attention owing to the increasing population demand for their beneficial nutritional properties (Yang *et al.,* 2022), representing an important source of dietary fiber, minerals, and antioxidants (Gürbüz *et al.,* 2018). These functional properties are mainly provided by the presence of anthocyanins in the peel and chlorogenic acid in the flesh (Plazas *et al.,* 2013; Rosa-Martínez *et al.,* 2022). More recent studies have been focused on fruit quality by enhancing chemical composition, including phenolic content, carotenoids, and antioxidants (Raigón *et al.,* 2008; Yang *et al.,* 2022). CWRs have not been widely used to improve nutritional quality due to the presence of undesirable compounds, such as glycoalkaloids, although some of them show high levels of chlorogenic acid (Rosa-Martínez *et al.,* 2022). They have been used mainly in breeding programs related to tolerance to biotic and abiotic stresses (Gramazio *et al.,* 2023).

Eggplant breeding has lagged behind other solanaceous crops, such as tomato or potato, in terms of genomic tool development (Page *et al.,* 2019). The study of molecular and genetic mechanisms underlying important eggplant traits has been delayed due to the reduced number of molecular markers and the relatively low density of genetic maps available in the last decade (Wei *et al.,* 2020a). Different eggplant reference genomes have been recently released contrasting for several traits (Hirakawa *et al.,* 2014; Barchi *et al.,* 2019; Wei *et al.,* 2020b; Li *et al.,* 2021a), leading to better dissection of different traits of interest (Wei *et al.,* 2020a; Qian *et al.,* 2021; Arrones *et al.,* 2022). Therefore, the availability of reference genomes is of great relevance as it allows the study of genetic crop diversity, promoting the use of the available plant genetic resources (Wambugu *et al.,* 2018). In fact, the construction of an eggplant high-quality reference sequence has also led to the development of novel and powerful genomic tools (Barchi *et al.,* 2019).

During the last years, a new generation of eggplant experimental populations has been generated from bi-parental crosses, including recombinant inbred lines (RILs) and introgression lines (ILs) (Lebeau *et al.,* 2013; Gramazio *et al.,* 2017; Toppino *et al.,* 2018, 2020; Mishra *et al.,* 2020). Unlike traditional F_2 and early backcross materials, these populations are considered immortal since final individuals present a high degree of homozygosis and can be regenerated by selfing for seed propagation (Gramazio *et al.,* 2023). The genome of the final individuals is a mixture of the parents as a result of genetic recombination events occurred during the early stages of population development (Arrones *et al.,* 2020). These resources have demonstrated their potential for mapping QTL related to traits of interest (Cockram and Mackay, 2018).

The first eggplant RIL population was generated by the cross of a *S. melongena* line susceptible to *Ralstonia solanacearum* and a resistant breeding line derived from a *S. melongena* × *S. aethiopicum* inter-specific cross (Lebeau *et al.,* 2013). The final 178 F_6 individuals were evaluated for bacterial wilt resistance leading to the identification of a major dominant gene and several related QTLs. Toppino *et al.* (2020) developed a second RIL population from a cross between a *S. melongena* line susceptible to *Fusarium oxysporum* and Verticillium wilt and a resistant breeding line derived from a doubled haploid of the somatic hybrid between *S. melongena* × *S. aethiopicum*. The final 163 F_7 individuals were phenotyped for anthocyanin content in different organs. Ultimately, from 7 to 17 QTLs were associated with the anthocyanin biosynthetic process and regulation. The same individuals were analysed through metabolic profiling and tested against *Fusarium*. Several QTLs influencing the regulation of different compounds involved in fruit nutritional quality were identified (Sulli *et al.,* 2021). Two major QTLs associated with complete and partial resistance to *Fusarium* were also identified (Tassone *et al.,* 2022). The last available RIL population was developed from the inter-specific cross between a *S. melongena* landrace and a *S. incanum* accession (Mishra *et al.,* 2020). The 114 F_8 final individuals were used to develop an improved high-density genetic linkage map.

The first ILs populations were obtained from the inter-specific cross of different *S. melongena* lines with *S. aethiopicum* and *S. linneanum* accessions (Toppino *et al.,* 2008; Mennella *et al.,* 2010). The final 57 ILs (BC_{6-7}) were characterized for health-related compounds to identify transgressive ILs with increased antioxidant values (Manella *et al.,* 2010). Subsequently, another IL population was developed from the inter-specific cross between *S. melongena* × *S. incanum* accessions (Gramazio *et al.,* 2017). The population was firstly evaluated for drought tolerance and 68 candidate genes were identified. The final 25 ILs (up to BC_6) were also characterized for morphological and agronomic traits, fruit shape characteristics, and composition (Mangino *et al.,* 2020, 2021; Rosa-Martínez *et al.,* 2022). Several putative QTLs were identified for the studied traits and potential candidate genes were spotted for most of them. A set of nine ILs were also evaluated under water stress conditions and low nitrogen fertilization levels (Flores-Saavedra *et al.,* 2023; Rosa-Martínez *et al.,* 2023). More recently, Toppino *et al.* (2018) released an inter-specific IL population from a *S. melongena* × *S. tomentosum* cross. The final 73 ILs (BC_{5-6}) have been used for the identification of QTLs related to fruit pigmentation near the calyx and, reduced plant height due to short internodes, and resistance to soil-borne diseases (nematodes, fusarium and verticillium wilt).

Currently, the *S. melongena* × *S. incanum* (GP2) population developed by Gramazio *et al.* (2017) is more advanced and it is almost finished. In addition, three different ILs populations are being developed from the inter-specific cross between

S. melongena × *S. insanum* (GP1), *S. melongena* × *S. dasyphyllum* (GP2), and *S. melongena* × *S. eleagnifolium* (GP3) accessions (Figure 4, unpublished). Advanced backcrosses (ABs) from the intermediate stages of these three ILs population development have been evaluated under low nitrogen conditions (Villanueva *et al.,* 2021, 2023). A set of ABs from the *S. melongena* × *S. eleagnifolium* ILs population is also being analysed for chlorogenic acid and glycoalkaloids content (unpublished).

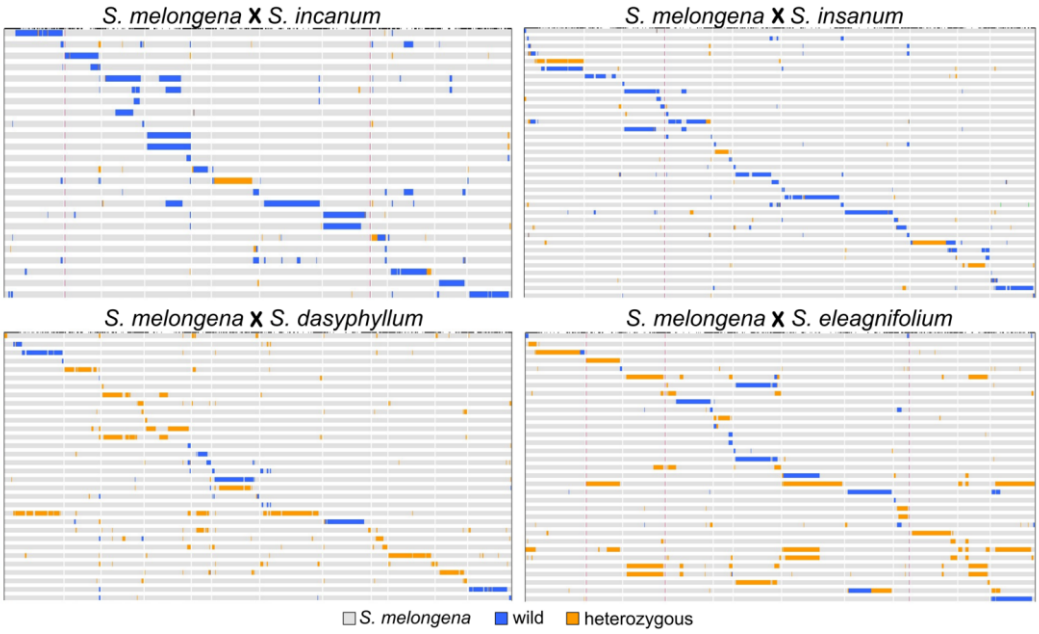


Figure 4. The four populations of introgression lines (ILs) currently under development in the Polytechnic University of Valencia (unpublished data) combining one *Solanum melongena* accession and one wild species. From more to less advanced: *S. melongena* × *S. incanum* (GP2), *S. melongena* × *S. insanum* (GP1), *S. melongena* × *S. dasyphyllum* (GP2), and *S. melongena* × *S. eleagnifolium* (GP3). In grey the *S. melongena* background, in blue the homozygous wild introgressions, and in orange the heterozygous introgressions.

However, so far, no multi-parent populations have been developed in the eggplant genepool. In order to fill this gap, here we present the first multi-parent advanced generation inter-cross (MAGIC) population developed by the inter-cross of seven *S. melongena* accessions from different origins and phenotypes and one from its wild relative *S. incanum* (Mangino *et al.,* 2022). The derived 420 S3 individuals were assessed for anthocyanin-related traits in vegetative plant tissues and fruit epidermis (Mangino *et al.,* 2022), and for the uniform and irregular

distribution of chlorophylls in the fruit peel, the latter being referred to as reticulation, variegation, or netting (Arrones *et al.,* 2022, 2023).

Eventually, the eggplant MAGIC population has demonstrated its potential usefulness as a permanent immortal mapping population for precise QTL identification. The combination of multiple founders and a large number of accumulated recombinant events has improved QTL mapping accuracy and resolution to subcentimorgan scale (Zheng, *et al.,* 2014; Pascual *et al.,* 2015). Although some drawbacks related to the introduction of a wild species appeared during the development of the population, such as a reduction of the hybrid fertility, the CWR founder represents an increased source of variability (Lester and Hasan, 1991). For instance, the genetic control of the netting trait was elucidated thanks to the contribution of the phenotype by the wild founder (Arrones *et al.,* 2023). The introduction of this wild trait rarely observed in commercial eggplants not only brings diversity in terms of fruit appearance (dark purple colour) but could directly improve fruit nutritional quality (Frary *et al.,* 2014; Nguyen *et al.,* 2014).

Breeding in tomato

Tomato (*Solanum lycopersicum* L.) is one of the most important crops, being the most produced among vegetables and the most important economically in the Solanaceae family (FAOSTAT, 2021; van Rengs *et al.,* 2022). It is a New World crop with evidence of origin in South America (Chapman, 2019). Tomato is easily distinguished from any other *Solanum* species by its bright yellow flowers and bipinnate compound non-spiny leaves (Knapp and Peralta, 2016).

The process of tomato domestication has led to much controversy, and it has been not long ago elucidated (Figure 5, Blanca *et al.,* 2022). *Solanum pimpinellifolium* is thought to be the wild ancestor of the cultivated tomato *S. lycopersicum* var. *lycopersicum* (Zuriaga *et al.,* 2009). It seems that tomato domestication occurred in two consecutive phases, one in South America and a later one in Mesoamerica, drastically reducing the genetic diversity along the way (Blanca *et al.,* 2015, 2022). Although an increase in morphological variability was experimented during the Mesoamerican tomato travel to Spain and Italy in the 16th century due to consumer preferences, tomato faced several genetic bottlenecks and its diversity remained very low. Nevertheless, recent breeding efforts have attempted to broaden the genetic base of cultivated tomato (Figàs *et al.,* 2015; Massaretto *et al.,* 2018). However, there is a special need to increase the knowledge of available germplasm resources and existing variability.

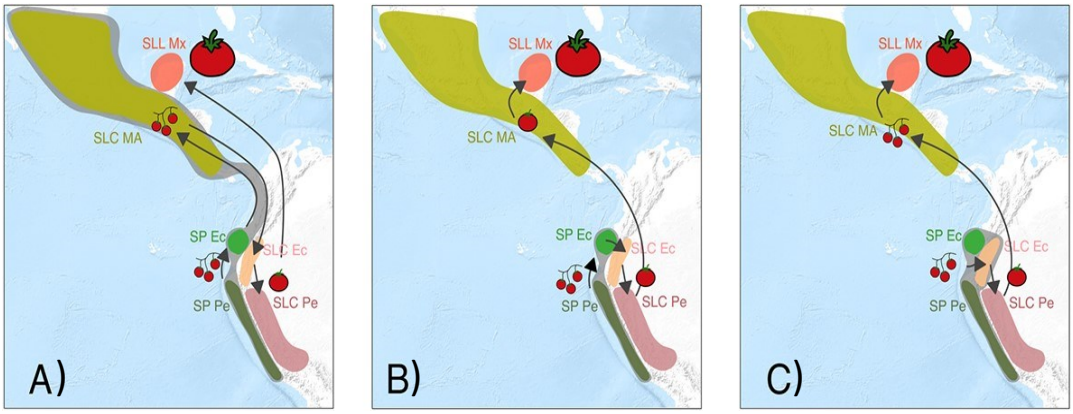


Figure 5. The different tomato evolution hypothesis proposed by (A) Blanca *et al.* (2022), (B) Blanca *et al.,* (2012), and (C) Razifard *et al.,* 2020. Source: Blanca *et al.* (2022).

In the case of tomato CWRs, there are several resource collections established around the world which contribute to a better understanding of the distribution of its diversity (Bauchet and Causse, 2012). However, greater efforts need to be made because natural habitats of wild species are shrinking as a result of globalisation and populations of wild tomatoes are being severely reduced (Grandillo *et al.,* 2011). The tomato clade consists of thirteen *Solanum* relative species divided into four subgroups (Figure 6, Peralta and Spooner, 2001; The 100 Tomato Genome Sequencing Consortium, 2014). The Lycopersicon section includes the wild ancestor, *S. pimpinellifolium*, and the closest species of cultivated tomato, *S. cheesmaniae* and *S. galapagense*, all self-compatible. These species occur on the western slopes of the Andes in dry desert or pre-desert environments (Knapp and Peralta, 2016). All the species included in the other subgroups (Arcanum, Ericopersicon, and Neolycopersicon) are green-fruited species characteristic of different altitudes of the Andean regions of South America, which explain their adaptation to different environments (Gonzali and Perata, 2021). Overall, inter-crosses between cultivated tomato and almost all the wild relative species are possible, although with varying degrees of difficulty (Knapp and Peralta, 2016). Comparisons of molecular data indicated relatively low genetic diversity between close species within the phylogenetic groups of the tomato clade (Miller and Tanksley, 1990). In fact, genome divergence between the wild *S. pimpinellifolium* and cultivated tomato was estimated to be only 0.6%, with most SNPs distributed in the gene-poor regions (Sato *et al.,* 2012). In the case of SNP frequencies for *S. pennellii*, which is the most distantly related species to tomato, is approximately 10% (The 100 Tomato Genome Sequencing Consortium, 2014). Thus, even if specific

cross-breeding techniques are required to succeed, the use of distant species will provide a positive transgressive variation within inter-specific progenies (Bauchet and Causse, 2012).

Traditional tomato breeding has been focused on higher productivity and adaption to different cultivation systems, followed by very firm fruits and long shelf-life varieties, giving rise to its economic success (The 100 Tomato Genome Sequencing Consortium, 2014). This led to a consumers request for more sensory and nutritional quality fruits, traits that were partially lost during tomato breeding (Causse *et al.,* 2010). Nowadays, specific breeding objectives are adaptation to growth constraints, disease and pest resistances, fruit productivity and quality, among others. Following these premises, CWRs have been primarily used as a source of resistance to face abiotic and biotic stresses (Grandillo *et al.,* 2011). Modern cultivars can carry more than 12 different wild introgressions since dominant resistance genes were found in the wild relatives (Bauche and Causse, 2012). Primary metabolites, nutritional, antioxidant, and volatile compounds associated with desired quality traits have recently received more attention and have also been identified in the CWRs (Schauer *et al.,* 2005; Grandillo *et al.,* 2013; Gascuel *et al.,* 2017).

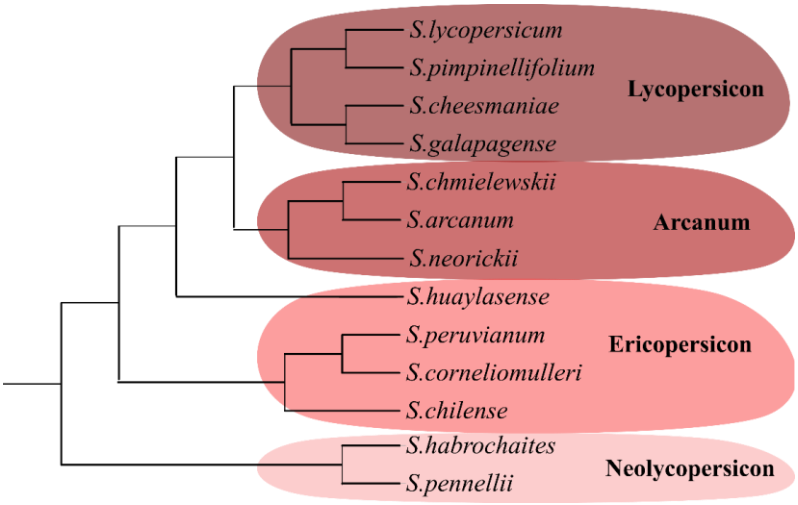


Figure 6. Phylogenetic tree of *Solanum* species belonging to the tomato clade, divided into the four subgroups identified by Peralta and Spooner (2001), and The 100 Tomato Genome Sequencing Consortium (2014).

The tomato clade has been exploited in many fields of plant research including biodiversity, evolution, adaptation, human domestication and nutrition perspectives (Peralta and Spooner, 2007). In terms of genomic tools, tomato is considered an important model species for scientific research since its genome sequence was published in The Tomate Genome Consortium (2014), and there are available extensive resequencing data in the Sol Genomics Network database (https://solgenomics.net) which is continuously developing (Bombarely *et al.,* 2010). In addition, several resequencing studies have been published including wild relative accessions from several geographical origins (Causse *et al*. 2013; Lin *et al.,* 2014; The 100 Tomato Genome Sequencing Consortium, 2014; Kevei *et al*. 2015; Gramazio *et al.,* 2019). The large amount of genetic information available has provided the opportunity to study gene function, diversity, and evolution in tomato and other Solanaceae through synteny (Suresh *et al.,* 2014).

Tomato has been used extensively for genetic studies because of its short generation time and the availability of numerous homozygous inbred lines, among other reasons (Suresh *et al.,* 2014). To harness the diversity found in both cultivated and wild species, tomato breeders have developed multiple permanent mapping populations aiming to identify and transfer desirable traits from wild to cultivated germplasm (Ranjan *et al.,* 2012). Over decades, several tomato experimental populations have been released, including ILs, RILs, ABs, backcross inbred lines (BILs), and even MAGIC populations (e.g., Eshed and Zamir, 1995; Paran *et al.,* 1995; Tanksley and Nelson, 1996; Doganlar *et al*. 2002; Lippman *et al.,* 2007; Salinas *et al.,* 2013; Fulop *et al.,* 2016; Ofner *et al.,* 2016). Many of them are inter-specific populations using different related species as parents such as *S. pimpinellifolium, S. pennellii,* or *S. cheesmaniae.* In these tomato populations, the introduction of CWRs has mainly been exploited for identifying QTLs linked to yield and fruit morphology parameters.

Focusing on multi-parent populations, two inter-specific tomato MAGIC populations have already been developed. The first one was an intra-specific population combining *S. lycopersicum* lines with large fruits and *S. lycopersicum* var. *cerasiforme* cherry-type accessions (Pascual *et al.,* 2015). This MAGIC population was used to construct a linkage map which showed an 87-105% increase in recombination frequencies in distal parts of the chromosomes compared to two inter-specific F_2 populations (Sim *et al.,* 2012), and a 69% increase compared to an intra-specific RIL population (Saliba-Colombani *et al.,* 2001). This increase is translated into a reduction of QTL support intervals in physical maps, thus increasing the accuracy of QTL fine mapping and candidate genes selection (Pascual *et al.,* 2015). The 397 S3 individuals were used to analyse fruit weight distribution in two environmental conditions. They identified already cloned QTLs for fruit weight (Frary *et al.,* 2000; Chakrabarti *et al.,* 2013), demonstrating the precision of QTL

mapping, and new QTLs thanks to the increased variability given by the multiple founders.

The second MAGIC population was constructed by inter-crossing seven *S. lycopersicum* accessions and one wild accession of *S. cheesmaniae* (Campanelli *et al.,* 2019). This population was developed to take advantage of the wild capacity to adapt to different environments and low-input agricultural practices. The 400 S4 individuals were grown in fields under organic management techniques and molecular analysis for resistance genes to fungi, bacteria, and viruses, and fruit shape was performed on the S5 seedlings.

Here, we present a novel inter-specific MAGIC population combining four *S. lycopersicum* var. *cerasiforme* and four *S. pimpinellifolium* accessions (Arrones *et al.,* 2024). The novelty of our population is based on the use of these species as founders which are the closest relative and the ancestor of tomato, respectively. They were selected due to their tolerance to fungi and bacteria and their ability to tolerate abiotic stresses since they indifferently colonise environments with high humidity and desert areas, from sea level to over 1,500 m altitude (Blanca *et al.,* 2015). The final set of 354 S5 lines were assessed for different fruit, plant, and flower traits, including lobule number, fruit weight, plant anthocyanin, and flowering time (Arrones *et al.,* 2024).

This latter inter-specific tomato MAGIC population represents a valuable pre-breeding resource which has demonstrated its potential to fine map previously identified causative genes and novel candidates responsible for controlling traits of interest. Previous founders' information together with high-throughput genotyping has allowed to significantly enhance recombination detection, haplotype prediction, and causal variants identification within the MAGIC population (Gramazio *et al.,* 2020). The introduction of CWR founders will be useful for shedding light on the genetics of agronomic and adaptation traits related with production, response to biotic and abiotic stresses, and earliness, among others. In addition, the identification of MAGIC lines showing a pyramiding of genes of interest holds direct utility for incorporation into breeding pipelines.

Review article

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

Andrea Arrones[1], Santiago Vilanova[1*], Mariola Plazas[1], Giulio Mangino[1], Laura Pascual[2], María José Díez[1], Jaime Prohens[1] and Pietro Gramazio[3*]

[1]Instituto de Conservación y Mejora de la Agrodiversidad Valenciana, Universitat Politècnica de València, Camino de Vera 14, 46022 Valencia, Spain

[2]Department of Biotechnology-Plant Biology, School of Agricultural, Food and Biosystems Engineering, Universidad Politécnica de Madrid, 28040 Madrid, Spain

[3]Faculty of Life and Environmental Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba 305-8572, Japan

[*]Corresponding authors

Ph.D. candidate contribution

A.A. had a main role in the following activities: literature revision, data compilation, drafting manuscript, manuscript review and editing.

Citation: Arrones, A., Vilanova, S., Plazas, M., Mangino, G., Pascual, L., Díez, M. J., *et al*. (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,* 229. doi: 10.3390/biology9080229

Abstract

The compelling need to increase global agricultural production requires new breeding approaches that facilitate exploiting the diversity available in the plant genetic resources. Multi-parent advanced generation inter-cross (MAGIC) populations are large sets of recombinant inbred lines (RILs) that are a genetic mosaic of multiple founder parents. MAGIC populations display emerging features over experimental bi-parental and germplasm populations in combining significant levels of genetic recombination, a lack of genetic structure, and high genetic and phenotypic diversity. The development of MAGIC populations can be performed using "funnel" or "diallel" cross-designs, which are of great relevance choosing appropriate parents and defining optimal population sizes. Significant advances in specific software development are facilitating the genetic analysis of the complex genetic constitutions of MAGIC populations. Despite the complexity and the resources required in their development, due to their potential and interest for breeding, the number of MAGIC populations available and under development is continuously growing, with 45 MAGIC populations in different crops being reported here. Though cereals are by far the crop group where more MAGIC populations have been developed, MAGIC populations have also started to become available in other crop groups. The results obtained so far demonstrate that MAGIC populations are a very powerful tool for the dissection of complex traits, as well as a resource for the selection of recombinant elite breeding material and cultivars. In addition, some new MAGIC approaches that can make significant contributions to breeding, such as the development of inter-specific MAGIC populations, the development of MAGIC-like populations in crops where pure lines are not available, and the establishment of strategies for the straightforward incorporation of MAGIC materials in breeding pipelines, have barely been explored. The evidence that is already available indicates that MAGIC populations will play a major role in the coming years in allowing for impressive gains in plant breeding for developing new generations of dramatically improved cultivars.

Keywords: MAGIC population, mapping populations, RILs, QTLs, association analysis, breeding resources, public–private partnerships, climate change

1. Introduction

The Green Revolution of the 1960s, together with substantial investment in plant breeding since then, enabled a significant increase in crop yields, especially in staple grain crops like wheat, rice, and maize (Pingali, 2012; Bailey-Serres *et al.,* 2019). Despite this, by 2050, agricultural production may need to be increased by 60–110% compared to current levels (Ray *et al.,* 2013; Hunter *et al.,* 2017). Though crop yields continue to increase globally, climate change represents a tremendous challenge for achieving this objective. In fact, recent research suggests that some major crop yields have already stagnated or even been reduced by the impact of climate change (Ray *et al.,* 2019).

The success of conventional breeding has been enhanced by genetic transformation technologies that were applied since the 1980s, allowing for the achievement of important advances in plant breeding (Ray *et al.,* 2013). Nevertheless, many of those technologies have been mostly focused on monogenic traits, while many major agronomic traits of interest are quantitative, controlled by multiple loci, and generally have a large environmental influence (Falconer and Mackay, 1996; Doerge, 2002; Holland, 2007; Wang *et al.,* 2016), thus posing a challenge to breeders due to the limited efficiency of breeding based on phenotypic selection.

Geldermann (1975) introduced the term quantitative trait locus (QTL) to describe "a region of the genome associated with an effect on a continuous trait." The identification of QTLs that significantly contribute to improve relevant yield and quality traits is a key factor in promoting a new Green Revolution (Pradhan *et al.,* 2019). By using marker-assisted selection (MAS), introgression breeding and the pyramiding of QTLs can be efficiently achieved to obtain dramatically improved cultivars (Mei *et al.,* 2020; Muthu *et al.,* 2020). However, the identification of QTLs underlying quantitative traits remains a challenge for plant breeders (Sehgal and Dreisigacker, 2019). Four basic elements are required to detect QTLs in crops: (i) an appropriate segregating population with a high genetic variability and contrasting parents for the target phenotype; (ii) marker systems that allow for the genotyping of the population; (iii) reproducible quantitative phenotyping methodologies; and (iv) an appropriate experimental design to evaluate the environmental effects and statistical methods to detect and locate QTLs (Doerge, 2002; Collard *et al.,* 2005; Cavanagh *et al.,* 2008; Huang *et al.,* 2015; Jaganathan *et al.,* 2020).

The use of experimental and germplasm populations has been of great relevance for mapping QTLs related to agronomic traits of interest (Kumar *et al.,* 2017). A new singular type of multi-parent population, the so-called multi-parent advanced generation inter-cross (MAGIC), was firstly proposed by Mackay and

Powell (Mackay and Powell, 2007). MAGIC populations are a set of immortal fixed lines with a genome that is an admixture of the genomes of multiple founder parents (Mackay and Powell, 2007; Cavanagh *et al.,* 2008; Rakshit *et al.,* 2012). Already available theoretical and experimental studies on MAGIC populations have revealed that they can make a relevant contribution for a dramatic improvement in detection of QTLs and the development of elite breeding material, as well as to a dramatic improvement in genetic advances in plant breeding (Zaw *et al.,* 2019; Han *et al.,* 2020).

Though some recent reviews on multi-parent populations have been published due to the growing interest of MAGIC populations (Scott *et al.,* 2020; Shekhawat *et al.,* 2020), our focal point is on the potential of MAGIC populations for plant breeding, including comprehensive information on the different types of mapping populations for QTL detection and the advantages and limitations of MAGIC populations, as well as strategies for development and available software for their genetic analysis. We also provide an overview of MAGIC populations already developed in plants and their utility for the detection of QTLs and breeding. In addition, we propose new approximations and prospects opening new avenues for MAGIC population expansion to inter-specific or MAGIC-like populations, as well as their incorporation into breeding pipelines.

2. Overview of experimental populations and germplasm collection for traits dissection

2.1. Bi-Parental populations

Traditionally, genetic linkage mapping studies for QTL detection in plants have used data collected from experimental populations, especially from bi-parental populations, like F_2 and first backcross (BC_1) generations (Xu *et al.,* 2017). They can be directly analysed (Clarke *et al.,* 1995; Grandillo and Tanksley, 1996) or studied after fixing by selfing until homozygous immortal populations (Keurentejes *et al.,* 2007), such as recombinant inbred lines (RILs), backcross inbred lines (BILs), near isogenic lines (NILs), or doubled haploid populations (DHs), are obtained (Rakshit *et al.,* 2012; Collard *et al.,* 2015) (Figure 1). These methodologies have been used in many genetic studies to identify QTLs and to clone major genes underlying QTLs linked to key traits in many crops (Doerge, 2002; Price, 2006; Keurentejes *et al.,* 2007; Pandey *et al.,* 2016).

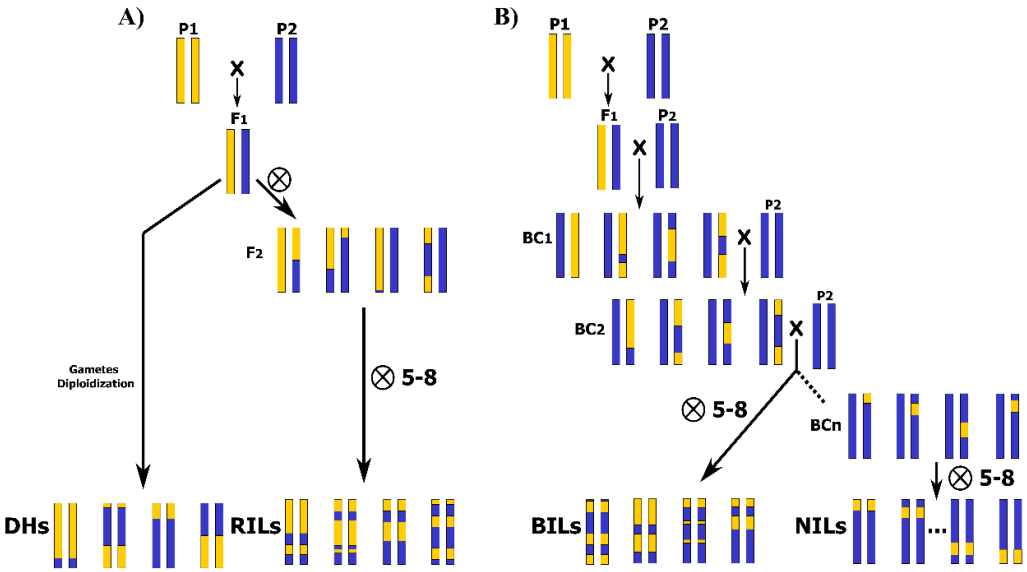


Figure 1. Bi-parental populations, from lowest to highest number of recombination events and required efforts in population development. Parental founders P1 and P2 are contrasting for the trait/s of interest, and F_1 is the simple hybrid derived from founder cross: (A) Doubled haploids (DHs), F_2, and recombinant inbred lines (RILs) developed by successive generations of selfing. (B) Backcross (BC), backcross inbred lines (BILs), and near isogenic lines (NILs) obtained by successive generations of selfing.

In bi-parental populations, two inbred lines are usually crossed to produce one or more segregating progenies (Xu *et al.,* 2017). Parents are selected based on their genetic and phenotypic diversity for a trait of interest, allowing for the detection of genomic regions associated with the target trait by the reconstruction of progeny genomes from the founder haplotypes (Dell'Acqua *et al.,* 2015). However, populations derived from bi-parental crosses only capture a small picture of the genetic factors that affect target traits in the species and suffer from a shortage of diversity due to the narrow genetic base that is limited to both parents (Huang *et al.,* 2012; Dell'Acqua *et al.,* 2015; Huang *et al.,* 2015). In addition, the limited opportunities for genetic recombination events, especially in F_2, BC_1, and DH populations, limits the resolution for QTL detection (Valdar *et al.,* 2006; Bandillo *et al.,* 2013). In these approaches, only two alleles are generally analysed, with a maximum resolution of 10–30 cm, as just a single opportunity of recombination has been possible during F_1 meiosis (Hall *et al.,* 2010). As a result, loci are mapped with low resolution and with large genetic support intervals (Huang *et al.,* 2012; Bandillo

et al., 2013). The same happens even in RILs because the number of efficient recombination decreases in advanced generations, as the proportion of the genome in heterozygosis is reduced to a half in each generation of selfing (Pascual *et al.,* 2015). Furthermore, the QTLs detected in a bi-parental population might not be expressed in other genetic backgrounds (Huang *et al.,* 2015).

2.2. Germplasm populations and germplasm collections

The continuous reduction of high-throughput genotyping costs together with the development of new genomic technologies has led to an emergence of QTL mapping resources, alternative to bi-parental populations, that demand an increase in marker density (Mackay *et al.,* 2014; Gegas *et al.,* 2016; Ongom and Ejeta, 2018). Genome-wide association studies (GWAS) have become one of the main genomic tools to dissect QTLs and genes underlying complex traits (Mitchell-Olds, 2010; Ongom and Ejeta, 2018). Association mapping takes advantage of high-genetic-diversity panels, including collections of selected individuals with unknown kinship and the historical accumulation of recombination events during thousands of generations (Korte and Farlow, 2013) (Figure 2). The application of GWAS for the more precise mapping of novel genes related to complex agronomic traits in crop plants has been demonstrated in many studies (Beck *et al.,* 2020; Tian *et al.,* 2020). In spite of that, the main limitations of GWAS are linkage disequilibrium (LD) (which may vary greatly among genome regions with different block length along chromosomes), population sub-structure (which can lead to inaccurate results), and unbalanced allele frequencies (Mitchell-Olds, 2010; Visscher *et al.,* 2012). Also, GWAS requires very large samples and many markers to have enough power to detect genomic regions of interest and its efficiency is limited by unknown pedigrees and missing parental information (Dell'Acqua *et al.,* 2015; Huang *et al.,* 2015). Furthermore, they may be suboptimal to identify rare alleles conferring interesting phenotypes (Kover and Mott, 2012).

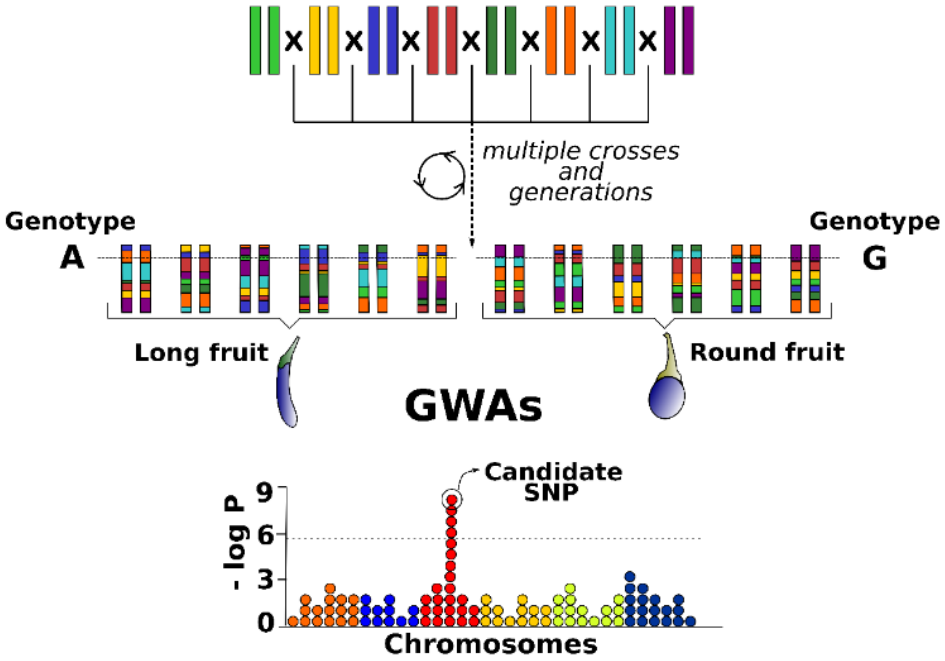


Figure 2. Historical recombination events and natural genetic diversity of germplasm collections used for genome-wide association studies (GWAS).

2.3. Multi-parent populations

Alternatively, multi-parent populations may offer solutions to the main drawbacks of bi-parental and germplasm populations (Mackay and Powell, 2007; Pascual *et al.,* 2015; Wang *et al.,* 2016; Scott *et al*, 2020). Multi-parent populations or multi-parent cross designs (MpCD) have been developed in many crop species throughout the history of scientific plant breeding (Harland and Martini, 1929; Suneson, 1956; Dell'Acqua *et al.,* 2015). The discovery of genetic male sterility (GMS) systems in the 1960s facilitated their adaptation to crops where artificial hybridization is challenging (Eberhart, 1972; Reddy and Stenhouse, 1994).

Multi-parent populations are produced by crossing more than two inbred founder lines, and they represent a bridge between linkage mapping (traditional bi-parental crosses) and association mapping (GWAS) approaches. This new approach dramatically increased mapping resolution by incorporating multiple founders with increased genetic and phenotypic diversity. Thanks to the increasing development of more sophisticated tools in high-throughput genomic technologies and statistical

analysis power, multi-parent populations can be employed in many genetic mapping studies (Mackay and Powell, 2007).

There are many types of MpCD, but here, we focus on MAGIC populations (Cavanagh *et al.,* 2008). MAGIC populations are panels of RILs that are fine-scale mosaics with theoretically equal proportions of the founder genomes. Therefore, MAGIC populations occupy an intermediate standing and represent a compromise between the much greater complexity found in naturally occurring accessions and the extreme simplicity of a diallelic system of RILs (Kover *et al.,* 2009). In this way, MAGIC populations are considered an emerging and powerful next-generation mapping resource within plant genetics, combining high-genetic recombination and diversity to dissect the structure of complex traits for improving breeding programs (Rakshit *et al.,* 2012; Mackay *et al.,* 2014; Huynh *et al.,* 2018; Ongom and Ejeta, 2018; Scott *et al.,* 2020). Despite being very similar to other multi-parent populations like the heterogeneous stock and Collaborative Cross populations used in mouse genetics, the MAGIC populations approach was first proposed for being used in crops in 2007 (Mackay and Powell, 2007; Cavanagh *et al.,* 2008).

Multiple founders in MAGIC populations enrich populations with higher allelic diversity when compared to those derived from typical bi-parental crosses, whereas multiple cycles of parental inter-crossing result in a set of rearranged genomes with a high level of fragmentation, thus giving greater opportunities for recombination and dramatically increasing the power of QTL detection (Bandillo *et al.,* 2013; Yamamoto *et al.,* 2014; Ongom and Ejeta, 2018). That makes them fit for both gene mapping and the effective generation of pre-breeding material. In this way, MAGIC populations provide an ideal platform and a common route for the discovery and high-resolution dissection of the QTLs and genes responsible for complex agronomic traits for crop improvement (Mackay and Powell, 2007; Cavanagh *et al.,* 2008).

MAGIC populations have been set up in many model species, thus demonstrating their power to detect polymorphisms underlying QTLs or genes related to complex traits of interest (Pascual *et al.,* 2015). Currently, MAGIC and MAGIC-like populations are available in a wide range of crop species including cereals, legumes, vegetables, fruit trees, and industrial crops, while many others are under development. Due to the wide genetic base of MAGIC populations, they are useful for QTL and gene discovery, the enhancement of breeding populations, and the development and release of new varieties (Huynh *et al.,* 2018). In this way, they facilitate investigation, not only in the genome architecture but also in its relationship with phenotypic traits and environment effects (Huang *et al.,* 2015).

3. Advantages and limitations of MAGIC populations

3.1. Advantages

MAGIC populations constitute a useful resource for genetic studies to the scientific community and, as indicated before, present clear advantages compared to classical bi-parental and germplasm populations. Thanks to their potential, they could be basically used for two different non-excluding purposes: (i) permanent immortal mapping populations for precise QTL locating and (ii) the development of new elite material to be directly released or to be included in breeding pipelines as pre-breeding materials (Bandillo *et al.,* 2013).

Regarding QTL detection, MAGIC populations are very useful genetic materials for linkage and association methodologies. The large number of accumulated recombinant events achieved over multiple rounds of inter-crossing and selfing improves QTL mapping accuracy, thus creating an opportunity to identify gene-trait associations with a greater resolution (Mackay and Powell, 2007; Huang *et al.,* 2012; Pascual *et al.,* 2015; Scott *et al.,* 2020) (Table 1). The increased recombination rate also facilitates further reductions in LD and provides little or no population structure that could lead to false-positive results (Valdar *et al.,* 2006; Mackay and Powell, 2007; Cavanagh *et al.,* 2008; Huang *et al.,* 2012; Sallam and Martsch, 2015).

The combination of multiple founders provides a higher genetic and phenotypic diversity within a single mapping population, thus increasing the number of QTLs that segregate in the cross and promoting novel rearrangements of alleles (Kover *et al.,* 2009; Bandillo *et al.,* 2013) (Table 1). An optimal founder selection allows for the targeting of multiple traits in a mixed background of parental genomes (Bandillo *et al.,* 2013; Pascual *et al.,* 2015). For instance, a QTL may not express in a single background of one founder but may be expressed in the MAGIC population mixture. MAGIC populations provide the opportunity to simultaneously assess the effect of multiple alleles at a locus and to study the interactions of cytoplasm effects, genome introgressions, and chromosomal recombination involving allelic diversity across the genome (Cavanagh *et al.,* 2008; Huang *et al.,* 2012; Bandillo *et al.,* 2013). Their potential for putative causal polymorphism identification underlying the QTLs, when coupled with available genome sequences, can be further enhanced to exploit current rapid advancements by the application of emerging high-throughput genotyping and phenomics platforms in the post-genomics era (Pascual *et al.,* 2015). These future ground-breaking approaches allow for the study of QTL functions and interactions across multiple traits within MAGIC populations (Mackay *et al.,* 2014; Gegas *et al.,* 2016).

Table 1. Advantages (+) and limitations (-) of linkage analysis (bi-parental populations), association mapping (GWAS) using germplasm populations, and multi-parent advanced generation inter-cross (MAGIC) for different characteristics of interest for their development and use. QTL: quantitative trait locus.

	Population type		
Characteristic	**Bi-parental**	**Germplasm**	**MAGIC**
Investment in time to be established	-	+	- -
Required population size	+	-	-
Genetic and phenotypic diversity	-	+ +	+
Suitability for coarse mapping	+	-	+
Suitability for fine mapping	-	+	+
QTLs resolution	-	+	+ +
Required marker density	+	-	-
Recombination rate	-	+ +	+
Low population sub-structure	+	-	+
Low LD	+	-	+

Regarding the potential use of the genetically diverse and immortal RILs that constitute the final MAGIC population, they provide a promising and useful germplasm to be exploited in breeding programs for the development of improved lines and hybrids or to be released as new cultivars (Bandillo *et al.,* 2013; Ongom and Ejeta, 2018). They can also be analyzed across a wide range of environments to increase the understanding of gene–environment interactions (G x E) and phenotypic plasticity, providing a permanent resource to study the basis of phenotypic traits (Pascual *et al.,* 2015; Diouf *et al.,* 2020; Scott *et al.,* 2020).

MAGIC populations represent an important pre-breeding resource, not only for the final lines already obtained but also for the possibility of developing new combinations. Previously identified QTLs and the novel ones detected in the final MAGIC could be synergically exploited to select the best MAGIC RILs as super trait-donor lines. In this way, they would serve as an advanced source for breeding programs to pyramidize novel combinations of QTLs. This strategy would be similar to the multi-parent advanced generation recurrent selection (MAGReS) approach proposed by the authors of (Huang *et al.,* 2015) but with some important differences. The MAGIC RIL sets selected for inter-crosses are more advanced and nearly homozygous (ideally F_6–F_8), and their selection can be targeted by using prior knowledge of bi-parental QTLs together with QTL haplotypes and predicted breeding values based on genome-background diversity (Huynh *et al.,* 2018). The resulting MAGReS lines will be fixed for positive haplotypes and will carry novel QTL combinations in other unknown loci conferring different unexpected traits. Therefore, selected MAGIC lines can be a valuable resource for genetic

improvement as super trait-donors that contribute to new QTL combinations (Huynh *et al.,* 2018).

3.2. Limitations

Though MAGIC populations overcome the main limitations of bi-parental and germplasm populations, they have some disadvantages linked to their development (Table 1). First and foremost, they need more investment in time and greater efforts to be developed due to their convoluted crossing scheme involving multiple founder lines and a large number of selfing generations to obtain an almost homozygous final MAGIC population (Mackay *et al.,* 2015; Pascual *et al.,* 2015). The crossing of selected parent lines and the following generations of inbreeding take several years due to the limitation of one-to-two generations per year of most of the crops, thus slowing down the development of the MAGIC population. The use of a DH population is the shortest way to produce fully homozygous populations, but they will incorporate recombination events that are only produced during the initial and advanced stages of inter-crossing. In addition, for many crops, this technique is strongly species- and genotype-dependent, requiring a prior establishment of an optimized and efficient protocol that is usually very laborious (Murovec and Bohanec, 2012). Another option is the so called "speed breeding" approach, which has been recently implemented in some cereal and model legume crops. "Speed breeding" greatly shortens generation time and accelerates breeding and research programs by growing plant populations under controlled photoperiod and temperature regimes (Watson *et al.,* 2018; Chiurugwi *et al.,* 2019). Controlled environments accelerate the development rate of plants and the harvesting and germination of immature seeds, thereby reducing generation time and increasing the number of plant generations obtained in one year. However, the facility costs required for controlled-environment growth chambers and the use of continuous supplemental lighting can make the cost of a project soar (Table 1). MAGIC populations that are highly powerful for fine genetic mapping also require a larger population size compared to a bi-parent population, which not only hinders population development but also limits phenotypic traits to evaluate. Therefore, population goals should be clearly defined before initiating population development since, for example, the identification of epistatic interactions requires a larger population size.

The presence of multiple founders involved in population development makes the process logistically challenging and labor-intensive (Ongom and Ejeta, 2018). Many molecular markers are required for the analysis of MAGIC populations due to their complex genomic structure conformation because of the multiple parental

admixture (Huang *et al.,* 2015) (Table 1). Knowledge of the population haplotype structure offers a more informative estimation of QTLs; however, due to the number of founder genomes, the direct observation of each parental contribution to a QTL remains a challenge (Huang and George, 2011; Sanneamnn *et al.,* 2015). The careful selection of founders is one of the most important design considerations because it ensures that the MAGIC population is going to be relevant as a long-term genetic resource panel (Huang *et al.,* 2015; Scott *et al.,* 2020). When wild species are in the panel of founder parents to broaden the genetic diversity, genetic and genomic incompatibility may appear. This is a handicap that may lead to a large reduction in the number of progenies derived from a specific MAGIC design. Usually, greater progeny reductions occur in early generations, since detrimental genes or combinations of such are manifested at the beginning, and these deficient lines leave no offspring. The subsequent occurrence of natural or artificial (either inadvertent or intentional) selection in the following generations is also extremely for the development of MAGIC populations. If genetic bottlenecks from a great reduction of lines occur in any generation, final lines will have family relationship that reduces the genetic diversity and creates population structure (Huang *et al.,* 2015). In addition, it is important to consider that the complexity of the funnel crossing schemes poses a potential intermating bias that can result in assortative mating instead of the assumed random mating (Ongom and Ejeta, 2018).

There are different MAGIC schemes including, among other approaches, four-, eight-, or 16-way recombinant inbred lines obtained by selfing (Broman *et al.,* 2018). With a higher number of founders, intermating becomes more and more unmanageable due to the number of individuals in each generation needed to achieve a complete admixture. Without the complete control of the reproduction, this complexity increases the chances of mating between individuals with similar genotypes, thus causing assortative mating and a deviation from random mating expectation (Templeton, 2006). Assortative mating introduces a generation of genetic subgroupings and distorts LD, causing wrong or spurious associations (Fisher, 1919; Fisher *et al.,* 1966; Wilson, 1978; Garnier-Géré and Chikhi, 2013).

When focusing on locations and mapping QTL frequency, bi-parental RILs regularly display ordinary or bi-modal distributions, whereas MAGIC populations show a continuous distribution that tends to deviate from their founders due to the unequal contribution of each parental genome in individual MAGIC lines (Li *et al.,* 2016). Broman (2005) suggested that the final MAGIC composition will directly depend on the cross-design selection and funnel structure chosen to construct the population. For this reason, before embarking on a MAGIC population development, it is important to make premeditated decisions on the purpose of the MAGIC population to be developed.

4. MAGIC development strategies

4.1. Cross-designs

MAGIC populations can be developed by different breeding designs that give rise to different population structures (Figure 3). Basically, the two main approaches to develop MAGIC populations are the following (Wang *et al.,* 2017):

(a) The "funnel" approach, in which the development of the MAGIC population starts with a funnel breeding scheme (Figure 3A). Typically, the funnel is created by several generations of inter-crossing among a number (n) of elite parental lines to obtain n/2 F_1 hybrids, which are subsequently inter-crossed in a set mating design to combine the genomes of all founders in the progeny lines.

(b) The "diallel" approach, in which the parents are crossed according to a diallel or half-diallel design, thus inter-crossing the parents in multiple funnels by a half-diallel mating system (two-way crosses), followed by the inter-crossing of the resulting F_1s until all the founders are represented in a single generation (Figure 3B). The diallel or half-diallel mating design is similar to the funnel approach, but it covers all possible two-way, four-way, and eight-way cross combinations. In the case of eight parental lines, a total of 28 two-way, 14 four-way, and seven eight-way crosses could be performed to develop the MAGIC population.

The "diallel" approach is the preferred cross-design used in the currently available MAGIC populations because it includes a higher diversity of allele combinations. However, in some cases, such as in tomato MAGIC populations, the "funnel" design has been chosen (Pascual *et al.,* 2015; Campanelli *et al.,* 2019). This could be due to the fact that in the MAGIC population developed by Pascual *et al*. (Pascual *et al.,* 2015), one of the founders was a variety with a small flower size, which makes hybridization difficult when used as a female parent. On the other hand, in the tomato MAGIC population developed by Campanelli *et al*. (Campanelli *et al.,* 2019) a wild species was used, thus hindering the multiple crossing process. Reducing the number of hybridizations needed by following the funnel design allows for the faster advancement of the development of the MAGIC populations.

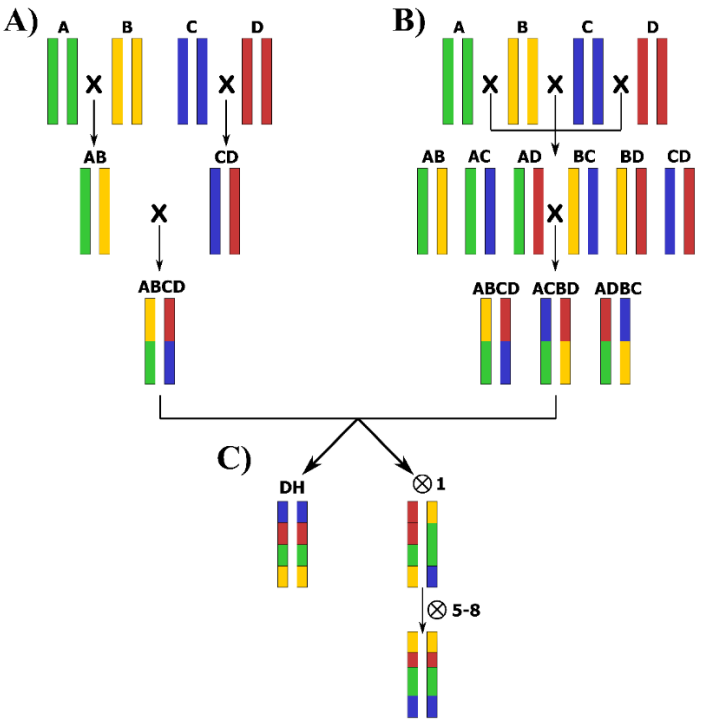


Figure 3. Cross-designs of a 4-way MAGIC population where the founders are A, B, C, and D: (A) "funnel" design; (B) "diallel" design; and (C) achievement of homozygous individuals by doubled haploids (DH) production, or by several rounds of selfing following the single-seed descent (SSD) method.

Once all founder genomes are admixed, some family inter-crossing can be performed to increase recombination and to finally achieve better mapping and QTL identification resolution, like in the barley and wheat MAGIC populations developed by Sannemann *et al*. (2015) and Stadlmeier *et al*. (2018), respectively. In any case, the final objective of both the funnel and diallel designs is the achievement of individuals with a high degree of homozygosis, which can be obtained by single-seed descent (SSD) or through double haploid production (Figure 3C). For instance, Sannemann *et al*. (2015) produced approximately 5000 barley MAGIC doubled haploid lines after eight-way inter-crossing. Through DH production via anther and microspore culture, they achieved completely homozygous individuals in a single step, significantly shortening the breeding program. Even though the doubled haploid production can be very fast in developing a MAGIC population when an optimal protocol is established, additional rounds of selfing generations after the admixing of genomes via hybridization can introduce novel recombination events,

although in a lower ratio than during the initial and advanced stages of inter-crossing (Huang *et al.,* 2015). When the development of highly homozygous lines is achieved through self-fertilization after hybridizations, five-to-eight generations of SSD are ideally recommended, resulting in RILs with complex pedigrees and with expected levels of genome heterozygosity between 3.125% and 0.319% for five and eight generations, respectively (Huang *et al.,* 2015). As a result, each RIL carries a mosaic of genome blocks theoretically contributed in equal proportions by all founders (Huang *et al.,* 2012; Rakshit *et al.,* 2012; Huang *et al.,* 2015). Blind SSD with no intended selection for preferable seed characteristics, plant type, or yield components was found to help create high diversity in morphological and agronomic traits in MAGIC populations (Huynh *et al.,* 2018).

Full-sib families obtained after several generations of selfing, which still have a moderate-to-high percentage of heterozygosity, can be used for coarse mapping and to develop NILs for the parts of the genome that still are in heterozygosis (Monforte and Tanksley, 2000). NILs are lines differing in a small genomic region or introgression fragment in an otherwise homogeneous genetic background, and they are usually obtained by repeated backcrossing to a recurrent parent. In this way by deriving NILs in advanced generations of selfing in MAGIC populations, there is no need of backcrossing for the precise location of the gene/QTL responsible for a trait of interest. Final lines could be further used as pre-breeding materials, or they can be directly released and registered as new varieties.

4.2. Founder selection

One of the most important decisions to take prior to the development of MAGIC populations is the selection of founders that are going to give rise to the final population. Ideally, founders should cover a broad phenotypic, genetic, and geographic diversity to exploit the potential of the population and to obtain a highly informative resource panel (Huang *et al.,* 2015) (Figure 4). Among MAGIC population founders, wild relative species or landraces coming from different origins that are very well adapted to their specific growing conditions could be candidates, in addition to commercial cultivars or elite breeding lines with desirable traits. The founders must be well characterized at the molecular and physiological levels to produce a practical resource. If the founders are also previously resequenced, it enables better design strategies to identify causal polymorphisms underlying QTLs, such as an efficient marker selection (Pascual *et al.,* 2015). It is important to not only ensure high diversity at the genetic and phenotypic level but also to carefully manage it. The selection of wild relatives to broaden genetic diversity could lead to incompatibilities, linkage drag due to a lack of recombination in some regions of the

genome, and the selection of certain phenotypic traits like a lack of seed dormancy or early flowering time. In these cases, the number of offspring families may suffer a reduction that could endanger the development of the MAGIC population (Huang *et al.,* 2015). Currently available MAGIC populations have generally been obtained using only intra-specific diversity, as in most cases their founders are landraces, improved and adapted breeding lines, or commercial cultivars. Though they are an interesting option to broaden population diversity, difficulties associated to the selection of wild species as founders make the development of inter-specific populations a challenge. However, some MAGIC populations incorporating one or more wild relative accessions have been developed or are under development (Campanelli *et al.,* 2019; Gramazio *et al.,* 2019).

An example of an inter-specific population already available is the tomato MAGIC population developed by Campanelli *et al.,* (2019). They selected a *S. cheesmaniae* accession as a founder due to a very large, interesting dataset of traits as biotic and abiotic stress tolerance, yield, and resilience (Nesbitt and Tanksley, 2002).

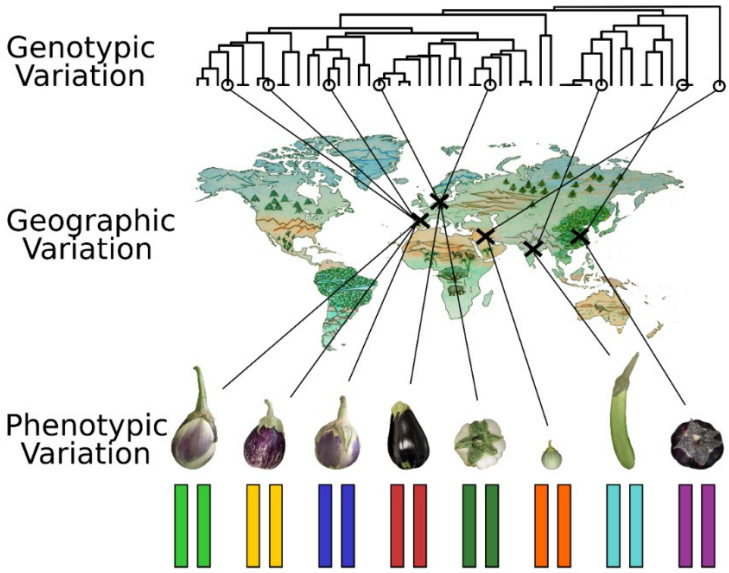


Figure 4. Maximizing founder selection based on phenotypic, genetic, and geographic diversity when using eggplant as a model.

4.3. Population sizes

Large population sizes are essential in each generation during the development of the MAGIC population to avoid genetic bottlenecks in some generations but also to have enough power for QTL identification. A mapping population of at least 50–250 individuals is generally required for coarse QTL mapping (Collard *et al.,* 2005; Jaganathan *et al.,* 2020). It is well-known that as population size increases, the power and resolution of QTL mapping also increases (Valdar *et al.,* 2006; Broman *et al.,* 2018). In addition, the larger the population size, the weaker the LD within and between chromosome blocks and the lower the number of genetic markers to "tag" a haplotype (Hill and Robertson, 1968; Ranc *et al.,* 2012; Visscher *et al.,* 2012). However, larger population sizes also imply higher difficulties to characterize and maintain a population. Certainly, there must be an equilibrium between population size and the efforts invested to establish the population.

An optimal population size strongly depends on genome size (Valdar *et al.,* 2006; Huang *et al.,* 2015; Huynh *et al.,* 2018) (Figure 5). Organisms with large genomes require an offspring of at least 500 individuals to provide a resolution power of the sub-centimorgan range, enough to detect singe QTLs that explain 5% of the phenotypic variability (Valdar *et al.,* 2006). Organisms with smaller genomes can afford to reduce their population size while maintaining the same resolution power. For instance, due to relatively small cowpea diploid genome (620 Mb), 305 MAGIC RILs are enough to cover its whole genome with fine resolution. A comparable level of physical resolution in an organism with a genome in the 5-Gb range (barley or diploid wheat) would require a population size of about 2500 RILs (Huynh *et al.,* 2018). MAGIC populations in plants with larger genomes should be larger than those with small genomes, but due to practical considerations and limitations, this is not always the case (Figure 5).

Figure 5. Population size of MAGIC populations developed compared to their genome size. Detailed information in Table 2.

Though a desired population size should be stablished before initiating MAGIC population development, on a practical level, maintaining it during generations is not very easy, and a certain percentage of losses of families during the process may occur, depending on the species and the founders used. In this way, as population progresses, the aimed population size could be affected by the presence of plants with poor development, fruit setting, or by the appearance of parthenocarpic fruits. In addition, if a wild species has been selected as a founder, prior drawbacks may increase. The largest losses of families take place in the early generations as selection for appropriate development, high fertility, and seed production takes place.

5. Analysis software for genetic gap construction and QTL mapping

The construction of genetic maps for marker–trait association analyses in MAGIC populations is challenging due to their complex cross-design. R/mpMap is one of the few pieces of software available for MAGIC map construction (Huang and George, 2011). This R package has been adapted to four-way and eight-way

MAGIC populations and has been used in the wheat and tomato MAGIC populations developed by Mackay *et al*. (Mackay *et al.,* 2014) and Pascual *et al*. (Pascual *et al.,* 2015), respectively. More recently, R/mpMap2 and magicMap R packages have been extended to other multi-parental populations (Shah *et al.,* 2019; Zheng *et al.,* 2019).

Regarding QTL mapping software, a wide range of statistical analysis tools for bi-parental crosses is available; however, this existing analysis software cannot be directly applied to multi-parent inbred line crosses (Lynch and Walsh, 1998; Doerge, 2002). While bi-parental analysis software only includes two allelic segregation patterns, multi-parent software must include all possible patterns according to the number of founders that are involved in the MAGIC population development. The higher complexity of multi-parent population designs requires a flexible and general framework for analyses capable of reconstructing founders' haplotype mosaics, to impute whole-genome genetic variants, and to handle multiple founder alleles and their population structures (Broman *et al.,* 2018). Therefore, different R packages started to appear due to the urgent need for analysis tools that are capable of handling the complexities associated with these advanced populations (Figure 6).

The first available QTL mapping software used for multi-patent population analysis was the R package HAPPY (Mott *et al.,* 2000), an interval mapping approach method based on founder probabilities. This software was used for the analysis of the Arabidopsis thaliana MAGIC, the first MAGIC population developed (Kover *et al.,* 2009) (Table 2). However, it was quickly replaced by R/qtl (Broman *et al.,* 2003) and R/mpMap (Huang and George, 2011). While R/qtl incorporated different functions to analyze more complex populations, R/mpMap was developed as a comprehensive set of methods for analyzing multi-parent designs with a greater flexibility in pedigree definition. Unlike R/qtl or R/HAPPY, R/mpMap was developed to accommodate linear mixed models to simultaneously assess genetic and environmental variation. Due to the clear advantages offered by R/mpMap, it is the most widespread and widely used software for MAGIC population analysis (Huang *et al.,* 2012; Pascual *et al.,* 2015; Sanneamnn *et al.,* 2015; Huynh *et al.,* 2018; Stadlmeier *et al.,* 2018; Shah *et al.,* 2019) (Table 2).

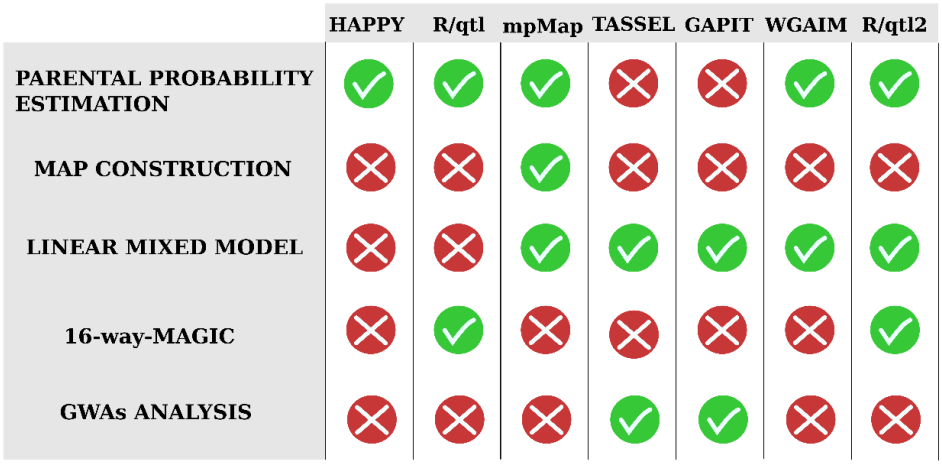


Figure 6. Analysis software used for QTL analysis in the available MAGIC populations, indicating the type of functionalities and analyses that can be performed in each of them.

However, MAGIC populations can also be directly analyzed for establishing marker–trait associations via GWAS (Mackay and Powell, 2007; Ogawa *et al.,* 2018; Sanneamnn *et al.,* 2018). The most used alternative is the Trait Analysis by aSSociation, Evolution and Linkage (TASSEL) software that uses phenotypic and genotypic data (Bradbury *et al.,* 2007). For controlling populations and family structures, this software implements general linear model and mixed linear model approaches. This useful software has been widely used for rice, cotton, wheat, sorghum, and *Brassica juncea* MAGIC population analyses (Bandillo *et al.,* 2013; Islam *et al.,* 2016; Milner *et al.,* 2016; Ongom and Ejeta, 2018; Yan *et al.,* 2020). Furthermore, some R packages are also available to conduct GWAS such as the so-called Genome Association and Prediction Integrated Tool (GAPIT) that implements advanced statistical methods, including mixed linear models (Lipka *et al.,* 2012). This genome association and prediction integrated tool has been used to analyze wheat, cotton, and maize MAGIC populations (Mackay *et al.,* 2014; Islam *et al.,* 2016; Anderson *et al.,* 2018).

Table 2. Current status of available MAGIC populations in the model species *Arabidopsis thaliana* and in crop plants, including the cross-design, founders, number of RILs in the MAGIC populations, target traits and software analysis used for QTL detection.

Crop	Design	Founders	Final RILs population	Target traits	QTLs analysis software	Reference
Model species						
A. thaliana	19-way, diallel	Natural accessions	1,026 S6	Germination date and bolting time	HAPPY	Kover et al. 2009
	8-way, diallel	Natural accessions	532 F5	Flowering time and leaf morphology	GenStat	Huang et al. 2011
Cereals						
Wheat	4-way, diallel	Cultivars	1,579 F6	Plant height and hectolitre weight	mpMap	Huang et al. 2012
	8-way, diallel	Cultivars	–	Plant height and hectolitre weight	mpMap	Huang et al. 2012
	8-way, diallel	Cultivars	1,091 F7	Awning	GAPIT	Mackay et al. 2014
	4-way, diallel	Commercial cultivars	1,458 F6:7	Coleoptile length and thickness and shoot length	WGAIM	Rebetzke et al. 2014
	60-way, NAM-like	Breeding lines	1,000 S4	Flowering time	In-house	Thépot et al. 2014
	4-way, diallel	Cultivars	> 338 F8	Plant height and grain yield	TASSEL	Milner et al. 2016
	8-way, diallel	Cultivars	2,125 F4	Plant height	GWAS	Sannemann et al. 2018
	8-way, funnel	Breeding lines	516 F6:8	Powdery mildew resistance	mpMap	Stadlmeier et al. 2018
	8-way, diallel	Elite lines and cultivars	> 3,000 S2:5	Number of recombination events	mpMap	Shah et al. 2019
Rice	8-way indica, diallel	Elite and modern cultivars	1,328 S7	Biotic/abiotic stress and grain quality	TASSEL	Bandillo et al. 2013
	8-way japonica, diallel	Elite and modern cultivars	500 S5	Biotic/abiotic stress and grain quality	TASSEL	Bandillo et al. 2013
	8-way MAGIC-plus, diallel	Elite and modern cultivars	S4 (in progress)	Biotic/abiotic stress and grain quality	TASSEL	Bandillo et al. 2013
	16-way MAGIC global, diallel	Elite and modern cultivars	–	Biotic/abiotic stress and grain quality	TASSEL	Bandillo et al. 2013
	12-way, funnel	Cultivars	1,600 S9	Plant height and heading date	TASSEL	Li et al. 2014
	8-way, diallel	Breeding lines	1,688 S5	Yield, plant height and heading date	TASSEL	Meng et al. 2016
	8-way, diallel	Cultivars	981 F6	Grain shape	GWAS	Ogawa et al. 2018
	4-way, diallel	Inbred lines	247 F7	Heading date	GWAS	Han et al. 2020

Table 2. Current status of available MAGIC populations in the model species *Arabidopsis thaliana* and in crop plants, including the cross-design, founders, number of RILs in the MAGIC populations, target traits and software analysis used for QTL detection.

Crop	Design	Founders	Final RILs population	Target traits	QTLs analysis software	Reference
Maize	8-way, diallel	Inbred lines	1,636 F6	Pollen shed, grain yield, and plant and ear height	QTLRel	Dell'Acqua et al. 2015
	4-way, funnel	Inbred lines	1,291 F4:5	Plant height, ear height, and flowering time	GAPIT	Anderson et al. 2018
Barley	8-way, funnel	Old landraces and a model cultivar	5,000 DH	Flowering time	mpMap	Sannemann et al. 2015
	32-way, funnel	Cultivars	324 F6	Climate and site-related agronomic adaptation	–	Bülow et al. 2016
Sorghum	29-way, diallel	Cultivars	~ 1,000 S7	Plant height	TASSEL	Ongom and Ejeta 2017
Oats	8-way, diallel	–	600 S6	–	–	Aberystwyth University (unpubl.)
Legumes						
Chickpea	8-way, diallel	Cultivars and breeding lines	~ 1,200 F6	Heat tolerance	–	Gaur et al. 2015
Faba bean	11-way, open pollination	Inbred lines	> 400 S9	Frost tolerance	–	Sallam and Martsch 2015
	4-way, funnel	Inbred lines	~ 1,000 F4	Flower colour and stipule spot pigmentation	–	Khazaei et al. 2017
Pigeonpea	8-way, diallel	Landraces and breeding lines	in progress	Resistance genes, maturing, and photoperiod	–	Saxena and Varshney 2018
Cowpea	8-way, diallel	Landraces and breeding lines	305 F8:10	Flowering, plant growth, seed size, and maturity	mpMap	Huynh et al. 2018
Soybean	8-way, funnel	Cultivars and exotic collections	764 F2:8	Yield under changing climatic conditions	–	Shivakumar et al. 2017
Groundnut	8-way, diallel	Breeding lines	~ 3,000 F6	Seed traits	–	Pandey et al. 2016
	8-way, diallel	Breeding lines	in progress	Aspergillus resistance and aflatoxin contamination	–	ICRISAT (unpublished)
	8-way, diallel	Breeding lines	in progress	Drought tolerance	–	ICRISAT (unpublished)
	–	Breeding and commercial lines	in progress	–	–	Tifton, Georgia, USA

Table 2. Current status of available MAGIC populations in the model species *Arabidopsis thaliana* and in crop plants, including the cross-design, founders, number of RILs in the MAGIC populations, target traits and software analysis used for QTL detection.

Crop	Design	Founders	Final RILs population	Target traits	QTLs analysis software	Reference
Vegetables and fruits						
Tomato	8-way, funnel	Cultivars and wild accessions	397 S3	Fruit weight	mpMap	Pascual et al. 2014
	8-way, funnel	Cultivars and wild accessions	400 F10	Resistance genes and fruit shape	–	Campanelli et al. 2019
	8-way, funnel	Cultivars and wild accessions	in progress	Morphoagronomic traits and resistance genes	–	Universitat Politècnica de València (unpubl.)
Strawberry	6-way, diallel	Cultivars	1,060 inter-cross	Fruit quality	–	Wada et al. 2012
Eggplant	8-way, funnel	Cultivars and wild accessions	in progress	Fruit traits	–	Universitat Politècnica de València (unpubl.)
Pepper	8-way, funnel	Landraces	in progress	Fruit traits	–	Universitat Politècnica de València (unpubl.)
Industrial and oil crops						
Cotton	12-way, funnel	Cultivars	1,500 F7	Fiber yield and resistance genes	–	Li et al. 2016
	11-way, diallel	Cultivars and a breeding line	> 550 S6	Fiber quality	TASSEL, GAPIT	Islam et al. 2016
Rapeseed	8-way	Elite cultivars	680 F6	Disease resistance, yield, plant architecture	–	Zhao et al. 2012
Chinese mustard	8-way, diallel	Breeding lines	408 F6	Quality traits (glucosinolate)	TASSEL	Yan et al. 2020

A whole-genome average interval mapping (WGAIM) approach was first proposed to accommodate all genotype information in a single model for bi-parental population analysis (Verbyla *et al.,* 2007), before being extended to MAGIC populations (Verbyla *et al.,* 2014). This analysis software was used in the wheat MAGIC population developed by Rebetzke *et al*. (2014) (Table 2). It is useful for QTL analysis with multiple alleles in a multi-environment or multi-trait data in a linear mixed model framework, which results in a greater understanding of traits or environments and the relationship between them. For this reason, they renamed this approach to multivariate multi-parent (MVMP) WGAIM (Verbyla *et al.,* 2014).

More recently, R/qtl2 software (Broman *et al.,* 2018) expanded the scope of the widely used R/qtl software package. This redesigned R package includes implementations of many different multi-parent population cross-designs and it is suited for high-dimensional genotype and phenotype data. Therefore, it includes numerous quality-control assessments as QTL genome scans by using the Haley–Knott regression method (Haley and Knott, 1992), linear mixed models to account for population structure, and best linear unbiased prediction (BLUP) based estimates of QTL effects.

6. An appraisal of MAGIC populations developed and evaluated

Despite the complexity and resources required in developing MAGIC populations, their number is continuously growing, and there are a number of MAGIC populations already available (Table 2). Currently, new ones are in progress—in model species, as well as in economically important crops, including cereals, legumes, and vegetables (Table 2). Each MAGIC population has been established with different purposes, showing clear differences in population designs features as well as in the way that they are assessed. In the end, all of them have been used in different studies as breeding materials sources, demonstrating their capability to identify QTLs, thus strongly reducing the list of candidate genes related to complex agronomic traits.

6.1. Model species

The first set of MAGIC lines was obtained in the model species Arabidopsis thaliana (Table 2). The 19-way MAGIC population was developed by Kover *et al*. (Kover *et al.,* 2009), and it served as a model and gave way to MAGIC development in other crops. With a final population size of at least 1026 S6 lines, it was used to

identify candidate genes for germination date and bolting time. In this way, the authors identified two QTLs for germination date and four QTLs for bolting time. In the case of bolting time, the four QTLs explained 63% of the total phenotypic variance, and they seemed to be linked to well-known genes that affect flowering time. Some years later, Huang *et al*. (Huang *et al.,* 2011) developed another eight-way MAGIC population to study flowering time and leaf morphology.

6.2. Cereals

The greatest number of MAGIC populations have been developed in cereals including wheat (nine), rice (eight), maize (two), barley (two), sorghum (one), and oats (in progress). Particularly, in wheat, there are different available MAGIC populations in both spring and winter wheat. Huang *et al*. (2012) developed the first two populations, one eight-way MAGIC and, previously, another four-way MAGIC population as an intermediate stage between bi-parental populations and the eight-parent population. Just like some novel MAGIC populations, the final RILs were used to identify QTL associations with traits of interest like plant height (Milner *et al.,* 2016; Sannemann *et al.,* 2018). A few years later, two different MAGIC populations were developed with about 1000–1500 F6:7 individuals to look for candidate genes involved in the control of awning and to assess QTLs for shoot length and coleoptile characteristics, respectively (Mackay *et al.,* 2014; Rebetzke *et al.,* 2014). More recently, Stadlmeier *et al*. (2018) constructed a reduced eight-way funnel MAGIC with 516 F6:8 RILs aimed at identifying QTLs for powdery mildew resistance.

In order to collect the broadest genotypic diversity present in rice, Bandillo *et al*. (2013) not only developed eight-way MAGIC populations—one in indica and another one in japonica subspecies—but also inter-crossed the indica and japonica base populations to increase the overall diversity; this is referred to as the "global MAGIC" population. With the development of these MAGIC populations, they aimed to detect major QTLs related to disease resistance and tolerance to abiotic stresses transmitted from the founder lines, such as blast resistance, bacterial blight resistance, salt tolerance, and submergence tolerance. They also planned to evaluate the final RILs in field trials to identify lines adapted to a range of production constraints in the major productor countries of Asia and Africa, particularly stress tolerance. Novel 12-way, eight-way, and four-way MAGIC populations aimed to identify QTLs related to heading dates (Li *et al.,* 2014; Meng *et al.,* 2016; Han *et al.,* 2020).

Two MAGIC populations have been developed in barley—eight-way and 32-way ones with population sizes of 5000 DH and 324 F6, respectively (Sannemann *et al.,* 2015; Bülow *et al.,* 2019). These populations have been used to study flowering time and climate and site-related agronomic adaptations. In the case of maize and sorghum, a study of the 1000–1500 final RILs was focused in traits related to plant architecture like plant height (Dell'Acqua *et al.,* 2015; Anderson *et al.,* 2018; Ongom and Ejeta, 2018). QTLs and candidate genes were suggested for two important and complex traits such as grain yield and flowering time (Dell'Acqua *et al.,* 2015).

6.3. Legumes

Many of the legume MAGIC populations have already been developed or are still in progress at International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) including those of chickpea, pigeonpea, and groundnut (Varshney, 2016). The highly recombinant chickpea MAGIC population with approximately 1200 F6 RILs allowed for the development of several heat-tolerant progenies, while the aim of the still in progress pigeonpea MAGIC is to look for resistance genes and early maturing and photoperiod intensive lines (Gaur *et al.,* 2012; Saxena and Varshney, 2017). Three additional groundnut eight-way MAGIC populations targeting different trait combinations are currently under development (Pandey *et al.,* 2016; Janila, 2017; Wang *et al.,* 2017; Choudhary *et al.,* 2019). While the first one focused on seed traits like fresh seed dormancy or oil content, the second MAGIC population was developed with Aspergillus resistance and aflatoxin contamination as targets, and the third one tries to dissect components of drought tolerance.

Two MAGIC populations have been developed in the faba bean. The 11-way faba bean MAGIC with less than 400 S9 final RILs aimed to find QTLs related to frost tolerance, while the four-way one with approximately 1000 F4 was focused on flower colour and stipule spot pigmentation (Sallam and Martsch, 2015; Khazaei *et al.,* 2018). There are also cowpea and soybean MAGIC populations, both focused on agronomic traits related to yield potential under changing climatic conditions (Huynh *et al.,* 2018; Shivakumar *et al.,* 2018).

6.4. Fruits and vegetables

Some MAGIC populations have already been developed in the tomato (two) and the strawberry (one). However, to our knowledge, there are others in progress in eggplants and peppers, as well as a new one in tomatoes, which is in the last stages

of development at Universitat Politècnica de València (Spain). The main goal of these populations is to dissect QTLs related to agronomic and fruit quality traits.

To widen the genetic diversity of the tomato, Pascual *et al*. (2015) developed the first tomato MAGIC population, selecting four accessions of cultivated tomato (*Solanum lycopersicum*) and four of weedy tomato (*S. lycopersicum* var. *cerasiforme*) as founders, including cherry tomato accessions, to study fruit weight. More recently, another tomato MAGIC population was developed by Campanelli *et al*. (2019) with the aim of developing organic tomato genotypes by participatory plant breeding (PPB). In this case, they constructed the first reported inter-specific MAGIC population by inter-crossing seven *S. lycopersicum* accessions and one wild accession of *S. cheesmaniae*. MAGIC lines were cultivated in different organic farms with different locations to carry out the PPB, thus showing significant phenotypic differences in development, productivity, and fruit colour. This variability was used to select families of tomatoes adapted to low input crop management, different environments, agricultural practices, and market conditions to be directly released as new varieties.

The six-way strawberry MAGIC population was used to study the major agronomic fruit traits, including flowering habit, fruit weight, fruit peel colour, and fruit firmness, as well as quality fruit traits such as soluble solid content or titratable acid (Wada *et al.,* 2017).

6.5. Industrial and oil crops

Li *et al*. (2016) developed an upland cotton MAGIC population with a total of 1500 F7 final RILs by inter-crossing 12 founder lines with the aim of increasing the intra-variety genetic diversity based on existing germplasm resources and improving fiber yield and quality. With the same objective, Islam *et al*. (2016) developed another 11-way cotton MAGIC population. In the case of oil seed crops, a rapeseed MAGIC population was developed with 680 F6 final RILs to study disease resistance, yield potential, plant architecture, and ecological adaptability (Zhao *et al.,* 2017). Recently, a Chinese mustard (*Brassica juncea*) MAGIC population was developed with a final population size of 408 F6 RILs mainly focused on glucosinolate-related traits to improve fruit quality (Yan *et al.,* 2020). Relevant QTLs were used to predict candidate genes associated with glucosinolate synthesis. QTL effects.

7. Future prospects

Currently available MAGIC populations including high phenotypic and genetic diversities have demonstrated their potential for QTL fine mapping. However, there is still a wide range of possibilities to be exploited.

7.1. Inter-specific MAGIC populations

It has been demonstrated in many crops that there has usually been a significant loss of genetic variability of the cultivated species because of domestication processes and plant genetic breeding programs (Smýkal *et al.,* 2018; Purugganan, 2019). This lack of genetic variation implies a higher risk of production losses and reduced yields when facing threats derived from different types of stress, and it also limits the exploitation to intra-specific variation (Mascher *et al.,* 2019), thus constraining the sources of variability for crop breeding programs. In this regard, with the aim of recovering part of this lost variability, geneticists and plant breeders have fallen back on related species and wild relatives as sources to broaden the genetic base of cultivated species (Prohens *et al.,* 2017). Including these wild species in multi-parent population schemes to get progenies from distant crosses opens new avenues for the exploitation of this variation.

MAGIC populations could be used as a way of including multiple genomic fragments or introgressions of one or several wild species into a cultivated background genome. Introducing a single wild relative among the eight founders of an inter-specific MAGIC population will include approximately 12.5% of the wild genome in the final RILs, similarly to a BC_2 bi-parental population. This approach not only allows for the study of foreign QTLs in a mixed background of the eight parents but also for the development of genetically characterized elite materials that can be easily incorporated in breeding programs. The novel eggplant and tomato populations that are still in progress at Universitat Politècnica de València are inter-specific MAGIC populations including wild accessions as founders. The aim here is to widen the genetic base of the crop and to take profit advantage of wild donor introgressions for relevant traits, including adaptation to climate change.

It is important to consider that on a practical level, inter-specific populations could present some drawbacks, e.g., crossing barriers, low hybrid fertility, sterility, or undesirable agronomic traits. In addition, linkage drag due to reduced recombination rate at introgressed fragments could be observed, and homozygosity

could not be reached in some genomic regions due to detrimental effects, resulting in negative selection (Prohens *et al.,* 2017). Some of these issues have been observed in the novel eggplant MAGIC population, in which an accession of the wild species *Solanum incanum* has been used as a female founder. This inter-specific crossing drags the maternal cytoplasmic background of the wild parent in some of the lines, and, on average, this reduces fertility, as has been observed in alloplasmic eggplant materials (Yoshimi *et al.,* 2013). It has been observed that these lines present a greater difficulty in fruit setting, as well as in the production of viable seeds, which suggests a lower viability of pollen. However, it also allows for the evaluation of the interactions of a wild cytoplasm with a mostly cultivated nuclear genome, thus opening new ways for the understanding of the effects of cytoplasm on the phenotype of important traits.

7.2. MAGIC-like approximations

New strategies and designs could be developed to generate multi-parent resources for plant species for which founder pure lines cannot be obtained or may take too much time to obtain. In this way, multi-parental populations have been already developed in some fruit trees like apple and strawberry, confirming the expected advantages of multi-population studies (Allard *et al.,* 2016; Wada *et al.,* 2017). The main problem of these species with QTL identification is the high heterozygosity and the appearance of false-positives due to relationships between individuals. In these cases, it is important to follow the pedigree to understand the kinship relationships for their analysis. In addition, considering marker data for common ancestors makes it possible to trace the source of favourable alleles.

Vegetative (or clonal or asexual) propagated populations consist of highly heterozygous clones that are genetically identical to their parents and can be conserved and utilized by continued vegetative reproduction for a long period of time (Allard, 1999). Most clonal species have the problem of inbreeding depression, but hybridization between different clones, or even the self-pollination of one clonal line, can produce seeds and therefore generate segregating clonal F_1 progenies (Zhang *et al.,* 2015). Taking a four-way MAGIC population as an example with four inbred lines A, B, C, and D as parents, the hybrid made between inbred lines A and B will be equivalent to the female parent of a clonal F_1 population after the female haploid building, and the hybrid made between inbred lines C and D will be equivalent to the male parent of a clonal F_1 population after the male haploid building (Zhang *et al.,* 2015). Then, a double cross (or four-way cross) will be made between the two hybrids—one used as female and the other one as male. Subsequent cycles of

hybridization between the clones can increase the admixing of the genomes and recombination, thus producing a pseudo-MAGIC population of clones that can be vegetatively propagated.

We also suggest the possibility of developing MAGIC-like populations in crops where selfing cannot be applied due to self-incompatibility, which makes pure lines impossible to obtain. This is a case similar to the so-called heterogeneous stocks and Collaborative Cross populations in mice (Mott *et al.,* 2000). In this case, as occurred for the generation of the mice CC lines, repeated generations of inbreeding through sibling mating can be applied to increase the degree of fixation.

7.3. Incorporation of MAGIC populations in breeding pipelines

Apart from the inherent value of MAGIC populations as experimental materials as mapping populations and for the detection of genes and QTLs, they may also represent an elite material for breeders (Huynh *et al.,* 2018; Zaw *et al.,* 2019). Given the nature of MAGIC populations, in which parents generally are complementary for traits of interest (Collard *et al.,* 2015; Scott *et al.,* 2020), new phenotypes will arise in the MAGIC population, and it may be possible to recover in the final MAGIC population lines that display improved characteristics that can be used as elite material for breeding programs or even directly as new cultivars developed through selection within MAGIC populations (Li *et al.,* 2013; Gaur *et al.,* 2019). In this respect, in a MAGIC population perfectly fixed in homozygosis, the frequency of specific genotypes in the final MAGIC population (as long as genes are unlinked) can be calculated as the product of the frequency of the desired alleles of each target locus among the founders. For example, the expected frequency of a genotype in a MAGIC population with eight founders in which the desired genotype corresponds to four unlinked genes with frequencies of 1/2, 1/4, 1/8, and 3/8 among the founders would be 3/512, and the minimum size for getting one individual with this genotype with a probability of 0.95 would be of 510 individuals. Therefore, given the large population sizes, it may be possible to find desired combinations of pyramided genes or QTLs among the MAGIC lines, even for multiple target loci, that may be of interest to be directly incorporated into the breeding pipelines by breeders. In this respect, as proposed by Scott *et al*. (Scott *et al.,* 2020), the availability of MAGIC population 'packages,' which integrate the MAGIC population material plus extensive phenotypic and genotypic characterization data, would greatly facilitate the incorporation of MAGIC materials in breeding pipelines.

MAGIC lines can be of particular interest as parents of hybrids. In this way, the evaluation of specific and general combining abilities (Malijan *et al.,* 2011) by

crossing MAGIC lines with testers may result in either the identification of MAGIC lines that give heterotic hybrids with already established elite lines or even in the establishment of heterotic groups within MAGIC populations. The availability of genotyping data in the MAGIC populations can also facilitate the use of genetic distances, or other genetic parameters, among lines as parameters to predict performance of hybrids (Alves *et al.,* 2019).

We also propose that MAGIC populations can also be used, mimicking the composite crosses evolutionary breeding strategy (Murphy *et al.,* 2013; Raggi *et al.,* 2017), to let natural selection act in highly diverse MAGIC populations—either after the final population has been obtained or during their development. Composite crosses that have been let evolve under cultivation conditions have proved valid ways to increase yields in cereals by exploiting adaptation to local environments (Ceccarelli and Grando, 2019; Masoni *et al.,* 2020). Though intentional selection is often avoided during the development of MAGIC populations (Huang *et al.,* 2015), artificial selection may contribute to increasing the frequencies of alleles of interest for specific loci, facilitating the incorporation of MAGIC materials in breeding programs. In this respect, Campanelli *et al*. (2019) developed a tomato MAGIC population in which participatory breeding was applied during the developmental phases of the MAGIC population. Additionally, given the interest in heterogeneous materials for organic agriculture (Willer *et al.,* 2020), subsets of MAGIC populations, or even the whole population, may be directly used for cultivation to obtain a highly diverse product of interest for specific markets demanding diversity.

Finally, as has been proposed for experimental populations with introgressions from crop wild relatives (Prohens *et al.,* 2017), the establishment of public–private partnerships (PPPs) for the development of MAGIC populations may greatly stimulate the use of MAGIC populations in breeding pipelines. In this way, developing MAGIC populations requires a long time, particularly for crop plants where only one or a few reproductive cycles can be achieved per year and a significant amount of resources are needed for their development (Mackay *et al.,* 2014; Pascual *et al.,* 2015), and PPPs may optimize the resources and expertise for the development of MAGIC populations; at the same time, breeders from private companies can spot interesting genotypes that can be followed in a pedigree. Several examples of successful PPPs in breeding exist (Lusser, 2014; Moore, 2015; Khush, 2017), and the nature of MAGIC populations, where a great diversity is usually present in founders, as well as the many new phenotypes and genetic combinations that arise during the development of the MAGIC population, is of particular interest to breeders and may greatly spur the use of MAGIC materials in breeding for the development of new cultivars.

8. Conclusions

MAGIC populations are recombinant inbred sets obtained after inter-crossing multiple parents. Although the development of MAGIC populations requires considerable efforts, theoretical and real studies show that MAGIC populations represent powerful tools for the detection of QTLs present in the set of parents, with considerable advantages over bi-parental and germplasm sets for the detection of QTLs. An important feature of MAGIC populations is that recombinant elite lines can be directly selected by breeders for being introduced in breeding programmes, or directly selected as new varieties. As the number of MAGIC populations is continuously growing, new tools for an efficient analysis of MAGIC populations have recently been developed. Also, new developments can extend the MAGIC approach to crops in which development of standard MAGIC populations are difficult to be obtained. We are confident that MAGIC populations will play an important role in addressing the formidable challenges faced by breeders in a scenario of climate change and the increased demand of plant products (Ray *et al.,* 2013; Hunter *et al.,* 2017; Ray *et al.,* 2019) by significantly contributing to the development of new generations of resilient, highly productive, and resource-efficient cultivars.

Data availability statement: The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/, PRJNA392603.

Author contributions: S.V., M.P., J.P., and P.G. conceived the idea and supervised the manuscript; A.A., S.V., M.P., G.M., and P.G. prepared a first draft of the manuscript; L.P., M.J.D., and J.P. reviewed and edited the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been funded by Ministerio de Ciencia, Innovación y Universidades, Agencia Estatal de Investigación and Fondo Europeo de Desarrollo Regional (grant RTI-2018-094592-B-100 from MCIU/AEI/FEDER, UE) and by 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). Andrea Arrones is grateful to Spanish Ministerio de Ciencia, Innovación y Universidades for a pre-doctoral (FPU18/01742) contract. Mariola Plazas is grateful to Generalitat Valenciana and Fondo Social Europeo for a post-doctoral grant (APOSTD/2018/014). Pietro Gramazio is grateful to Japan Society for the Promotion of Science for a post-doctoral grant (P19105, FY2019 JSPS Postdoctoral Fellowship for Research in Japan [Standard]).

Conflict of Interest: The authors declare no conflict of interest.

References:

Allard, A., Bink, M. C. A. M., Martinez, S., Kelner, J. J., Legave, J. M., Di Guardo, M., *et al*. (2016). Detecting QTLs and putative candidate genes involved in budbreak and flowering time in an apple multi-parental population. *J. Exp. Bot., 67*, 2875–2888, doi: 10.1093/jxb/erw130

Allard, R. W. (1999). Principles of plant breeding, 2nd ed., John Wiley & Sons: Hoboken, USA, 264. ISBN 0471023094

Alves, F. C., Granato, Í. S. C., Galli, G., Lyra, D. H., Fritsche-Neto, R., De Los Campos, G. (2019). Bayesian analysis and prediction of hybrid performance. *Plant Method., 15*, 1–18. doi: 10.1186/s13007-019-0388-x

Anderson, S. L., Mahan, A. L., Murray, S. C., and Klein, P. E. (2018). Four parent maize (FPM) population: effects of mating designs on linkage disequilibrium and mapping quantitative traits. *Plant Genome, 11*, 1–17. doi: 10.3835/plantgenome2017.11.0102

Bailey-Serres, J., Parker, J. E., Ainsworth, E. A., Oldroyd, G.E.D., and Schroeder, J. I. (2019). Genetic strategies for improving crop yields. *Nature, 575,* 109–118. doi: 10.1038/s41586-019-1679-0

Bandillo, N., Raghavan, C., Muyco, P. A., Sevilla, M. A. L., Lobina, I. T., Dilla-Ermita, C. J., *et al*. (2013). Multi-parent advanced generation inter-cross (MAGIC) populations in rice: Progress and potential for genetics research and breeding. *Rice, 6*, 1–15. doi: 10.1186/1939-8433-6-1

Beck, T., Shorter, T., and Brookes, A. J. (2020). GWAS central: a comprehensive resource for the discovery and comparison of genotype and phenotype data from genome-wide association studies. *Nucleic Acids Res., 48*, D933–D940. doi: 10.1093/nar/gkz895

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. doi: 10.1093/bioinformatics/btm308

Broman, K., Wu, H., Sen, Ś., and Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics, 19,* 889–890. doi: 10.1093/bioinformatics/btg112

Broman, K. W. (2005). The genomes of recombinant inbred lines. *Genetics, 169*, 1133–1146. doi: 10.1534/genetics.104.035212

Broman, K., Gatti, D., Simecek, P., Furlotte, N., Prins, P., Sen, Ś., *et al*. (2018). R/qtl2: software for mapping quantitative trait loci with high-dimensional data and multi-parent populations. *Genetics, 211*, 495–502. doi: 10.1101/414748

Bülow, L., Nachtigall, M., and Frese, L. (2019). A MAGIC population as an approach to the conservation and development of genetic diversity of winter barley for breeding purposes by on-farm management. *J. fur Kult., 71*, 286–298. doi: 10.5073/JfK.2019.11.02

Campanelli, G., Sestili, S., Acciarri, N., Montemurro, F., Palma, D., Leteo, F., *et al*. (2019). Multi-parental advances generation inter-cross population, to develop organic tomato genotypes by participatory plant breeding. *Agronomy, 9*, 119. doi: 10.3390/agronomy9030119

Cavanagh, C., Morell, M., Mackay, I., and Powell, W. (2008). From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. *Curr. Opin. Plant Biol., 11*, 215–221. doi: 10.1016/j.pbi.2008.01.002

Ceccarelli, S., and Grando, S. (2019). "From participatory to evolutionary plant breeding" in *Farmers and Plant Breeding: Current Approaches and Perspectives*, Westengem, O.T., Winge, T., Eds., Routledge: London, UK, 231–243. ISBN 9780429507335

Chiurugwi, T., Kemp, S., Powell, W., and Hickey, L. T. (2019). Speed breeding orphan crops. *Theor. Appl. Genet., 132*, 607–616. doi: 10.1007/s00122-018-3202-7

Choudhary, D., Agarwal, G., Wang, H., Pandey, M. K., Culbreath, A. K., Varshney, R. K., *et al*. (2019). Molecular markers and genomic resources for disease resistance in peanut-A review. *Legum. Res., 42*, 137–144. doi: 10.18805/LR-409

Clarke, J. H., Mithen, R., Brown, J. K. M., and Dean, C. (1995). QTL analysis of flowering time in *Arabidopsis thaliana*. *Mol. Gen. Genet., 248*, 278–286. doi: 10.1007/BF02191594

Collard, B. C. Y., Jahufer, M. Z. Z., Brouwer, J. B., and Pang, E. C. K. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica, 142*, 169–196. doi: 10.1007/s10681-005-1681-5

Dell'Acqua, M., Gatti, D. M., Pea, G., Cattonaro, F., Coppens, F., Magris, G., *et al*. (2015). Genetic properties of the MAGIC maize population: a new platform for high definition QTL mapping in *Zea mays*. *Genome Biol., 16*, 1–23. doi: 10.1186/s13059-015-0716-z

Diouf, I., Derivot, L., Koussevitzky, S., Carretero, Y., Bitton, F., Moreau, L., *et al*. (2020). Genetic basis of phenotypic plasticity and genotype x environment interaction in a multi-parental tomato population. *J. Ex. Bot*., eraa265. doi: 10.1093/ofid/ofy092/4987343

Doerge, R. W. (2002). Mapping and analysis of quantitative trait loci in experimental populations. *Nat. Rev. Genet., 3,* 43–52. doi: 10.1038/nrg703

Eberhart, S. (1972). Techniques and methods for more efficient population improvement in sorghum. In *Sorghum in Seventies*, Rao, N.G.P., Oxford And Ibh Publishing Co.: New Delhi, 197–213.

Falconer, D. S., and Mackay, T. F. C. (1996). Introduction to quantitative genetics, 4th ed., Longman: Harlow, UK.

Fisher, S. R. A. (1919). The correlation between relatives on the supposition of mendelian inheritance. *Earth Environmental Sci. Transactions of the Royal Society of Edinburgh, 52,* 399–433. doi: 10.1017/S0080456800012163

Fisher, S. R. A., Moran, P. A. P., and Smith, C. A. B. (1966). Commentary on R. A. Fisher's paper on the correlation between relatives on the supposition of mendelian inheritance. Galton Laboratory, University College London. doi: 10.5555/19670102053

Garnier-Géré, P., and Chikhi, L. (2013). Population subdivision, Hardy-Weinberg equilibrium and the Wahlund effect. In *eLS*, John Wiley & Sons, Ltd: Chichester, UK. doi: 10.1002/9780470015902.a0005446.pub3

Gaur, P. M., Jukanti, A. K., and Varshney, R. K. (2012). Impact of genomic technologies on chickpea breeding strategies. *Agronomy, 2*, 199–221. doi: 10.3390/agronomy2030199

Gaur, P. M., Samineni, S., Thudi, M., Tripathi, S., Sajja, S. B., Jayalakshmi, V., *et al*. (2019). Integrated breeding approaches for improving drought and heat adaptation in chickpea (*Cicer arietinum* L.). *Plant Breed., 138*, 389–400. doi: 10.1111/pbr.12641

Gegas, V. C., Gay, A., Camargo, A., and Doonan, J. H. (2016). Challenges of crop phenomics in the post-genomic era. In *Phenomics*, Hancock, J. M., CRC Press: Boca Ratón, USA, 142–171. ISBN: 9781466590960

Geldermann, H. (1975). Investigations on inheritance of quantitative characters in animals by gene markers I. Methods. *Theor. Appl. Genet., 46*, 319–330. doi: 10.1007/BF00281673

Gramazio, P., Lerma, M. D., Villanueva, G., Vilanova, S., García-Fortea, E., Mangino, G., *et al*. (2019). Detection, molecular characterisation and aspects involving the transmission of tomato chlorotic dwarf viroid in eggplant. *Ann. Appl. Biol., 175*, 172–183. doi: 10.1111/aab.12527

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. doi: 10.1007/BF00224033

Haley, C. S., and Knott, S. A. (1992). A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity, 69,* 315–324. doi: 10.1038/hdy.1992.131

Hall, D., Tegström, C., and Ingvarsson, P. K. (2010). Using association mapping to dissect the genetic basis of complex traits in plants. *Briefings Funct. Genomics, 9,* 157–165. doi: 10.1093/bfgp/elp048

Han, Z., Hu, G., Liu, H., Liang, F., Yang, L., Zhao, H., *et al*. (2020). Bin-based genome-wide association analyses improve power and resolution in QTL mapping and identify favorable alleles from multiple parents in a four-way MAGIC rice population. *Theor. Appl. Genet., 133*, 59–71. doi: 10.1007/s00122-019-03440-y

Harlan, H. V., and Martini, M. L. (1929). A composite hybrid mixture. *Agron. J., 21*, 487–490. doi: 10.2134/agronj1929.00021962002100040014x

Hill, W. G., and Robertson, A. (1968). The effects of inbreeding at loci with heterozygote advantage. *Genetics, 60*, 615–628. doi: 10.1093/genetics/60.3.615

Holland, J. B. (2007). Genetic architecture of complex traits in plants. *Curr. Opin. Plant Biol., 10*, 156–161. doi: 10.1016/j.pbi.2007.01.003

Huang, B. E., and George, A. W. (2011). R/mpMap: a computational platform for the genetic analysis of multi-parent recombinant inbred lines. *Bioinformatics, 27*, 727–729. doi: 10.1093/bioinformatics/btq719

Huang, B. E., George, A. W., Forrest, K. L., Kilian, A., Hayden, M. J., Morell, M. K., *et al*. (2012). A multi-parent advanced generation inter-cross population for genetic analysis in wheat. *Plant Biotechnol. J., 10*, 826–839. doi: 10.1111/j.1467-7652.2012.00702.x

Huang, B. E., Verbyla, K. L., Verbyla, A. P., Raghavan, C., Singh, V. K., Gaur, P., *et al*. (2015). MAGIC populations in crops: current status and future prospects. Theor. Appl. *Genet., 128,* 999–1017. doi: 10.1007/s00122-015-2506-0

Huang, X., Paulo, M. J., Boer, M., Effgen, S., Keizer, P., Koornneef, M., *et al*. (2011). Analysis of natural allelic variation in *Arabidopsis* using a multi-parent recombinant inbred line population. *Proc. Natl. Acad. Sci.* U.S.A., *108,* 4488–4493. doi: 10.1073/pnas.1100465108

Hunter, M. C., Smith, R. G., Schipanski, M. E., Atwood, L. W., and Mortensen, D. A. (2017). Agriculture in 2050: recalibrating targets for sustainable intensification. *Bioscience*, *67*, 386–391. doi: 10.1093/biosci/bix010

Huynh, B. L., Ehlers, J. D., Huang, B. E., Muñoz-Amatriaín, M., Lonardi, S., Santos, J. R. P., *et al*. (2018). A multi-parent advanced generation inter-cross (MAGIC) population for genetic analysis and improvement of cowpea (*Vigna unguiculata* L. Walp.). *Plant J., 93,* 1129–1142. doi: 10.1111/tpj.13827

Islam, M. S., Thyssen, G. N., Jenkins, J. N., Zeng, L., Delhom, C. D., McCarty, J. C. (2016). A MAGIC population-based genome-wide association study reveals functional

association of *GhRBB1_A07* gene with superior fiber quality in cotton. BMC *Genomics, 17*, 1–17. doi: 10.1186/s12864-016-3249-2

Jaganathan, D., Bohra, A., Thudi, M., and Varshney, R. K. (2020). Fine mapping and gene cloning in the post-NGS era: advances and prospects. *Theor. Appl. Genet., 133*, 1791–1810. doi: 10.1007/s00122-020-03560-w

Janila, P. (2017). Develop MAGIC and bi-parental populations following SSD (2) Phenotyping of populations for target traits. *Agricultural Res. Knowledge*. doi: 20.500.11766/6640

Keurentjes, J. J. B., Bentsink, L., Alonso-Blanco, C., Hanhart, C. J., Vries, H. B. De, Effgen, S., *et al*. (2007). Development of a near-isogenic line population of *Arabidopsis thaliana* and comparison of mapping power with a recombinant inbred line population. *Genetics, 175*, 891–905. doi: 10.1534/genetics.106.066423

Khazaei, H., Stoddard, F. L., Purves, R. W., and Vandenberg, A. (2018). A multi-parent faba bean (*Vicia faba* L.) population for future genomic studies. *Plant Genet. Resour*., *16*, 419–423. doi: 10.1017/S1479262118000242

Khush, G. S. (2017). Public-private partnership in agricultural biotechnology. *Insights Global Challenges and Opportunities for the Century Ahead*, Reddy, V. D., Rao, K. V., Krishna, K. R., Eds., BS Publications: Hyderabad, India, 341. ISBN 9789352301867

Korte, A., and Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review self-fertilisation makes *Arabidopsis* particularly well suited to GWAS. *Plant Methods, 9*, 29. doi: 10.1186/1746-4811-9-29

Kover, P. X., Valdar, W., Trakalo, J., Scarcelli, N., Ehrenreich, I. M., Purugganan, M. D., *et al*. (2009). A multi-parent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana. PLoS Genet., 5,* e1000551. doi: 10.1371/journal.pgen.1000551

Kover, P. X., and Mott, R. (2012). Mapping the genetic basis of ecologically and evolutionarily relevant traits in *Arabidopsis thaliana*. *Curr. Opin. Plant Biol., 15,* 212–217. doi: 10.1016/j.pbi.2012.02.002

Kumar, J., Gupta, D. Sen, Gupta, S., Dubey, S., Gupta, P., and Kumar, S. (2017). Quantitative trait loci from identification to exploitation for crop improvement. *Plant Cell Rep., 36*, 1187–1213. doi: 10.1007/s00299-017-2127-y

Li, D. G., Li, Z. X., Hu, J. S., Lin, Z. X., and Li, X. F. (2016). Polymorphism analysis of multi-parent advanced generation inter-cross (MAGIC) populations of upland cotton developed in China. *Genet. Mol. Res., 15*. doi: 10.4238/gmr15048759

Li, X. F., Liu, Z. X., Lu, D. B., Liu, Y. Z., Mao, X. X., Li, Z. X., *et al*. (2013). Development and evaluation of multi-genotype varieties of rice derived from MAGIC lines. *Euphytica, 192*, 77–86. doi: 10.1007/s10681-013-0879-1

Li, Z., Ye, G., Yang, M., Liu, Z., Lu, D., Mao, X., *et al*. (2014). Genetic characterization of a multi-parent recombinant inbred line of rice population. *Res. Crop., 15*, 28–37. doi: 10.5958/j.2348-7542.15.1.004

Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J. (2012). GAPIT: Genome association and prediction integrated tool. *Bioinformatics, 28*, 2397–2399. doi: 10.1093/bioinformatics/bts444

Lusser, M. (2014). Workshop on public-private partnerships in plant breeding, *JRC Publications Repository, 68*. ISBN 9789279381669

Lynch, M., and Walsh, B. (1998). Genetics and analysis of quantitative traits, Oxford University Press: Oxford, UK, 980. ISBN 978-0878934812

Mackay, I., and Powell, W. (2007). Methods for linkage disequilibrium mapping in crops. *Trends Plant Sci., 12*, 57–63. doi: 10.1016/j.tplants.2006.12.001

Mackay, J., Bansept-Basler, P., Bentley, A. R., Cockram, J., Gosman, N., Greenland, A. J., *et al*. (2014). An eight-parent multi-parent advanced generation inter-cross population for winter-sown wheat: creation, properties, and validation. G3 Genes, Genomes, *Genet., 4,* 1603–1610. doi: 10.1534/g3.114.012963

Malijan, A., Sevilla, M., Bandillo, N., and Redona, E. (2011). Combining ability in a full diallel cross of eight founder lines of a heat tolerance magic population. *Philipp. J. Crop Sci., 36*, 213.

Mascher, M., Schreiber, M., Scholz, U., Graner, A., Reif, J.C., Stein, N. (2019). Genebank genomics bridges the gap between the conservation of crop diversity and plant breeding. *Nat. Genet., 51*, 1076–1081. doi: 10.1038/s41588-019-0443-6

Masoni, A., Calamai, A., Marini, L., Benedettelli, S., and Palchetti, E. (2020). Constitution of composite cross maize (*Zea mays* L.) populations selected for the semi-arid environment of South Madagascar. *Agronomy, 10*, 54. doi: 10.3390/agronomy10010054

Mei, J., Shao, C., Yang, R., Feng, Y., Gao, Y., Ding, Y., *et al*. (2020). Introgression and pyramiding of genetic loci from wild *Brassica oleracea* into *B. napus* for improving *Sclerotinia* resistance of rapeseed. *Theor. Appl. Genet., 133*, 1313–1319. doi: 10.1007/s00122-020-03552-w

Meng, L., Guo, L., Ponce, K., Zhao, X., and Ye, G. (2016). Characterization of three indica rice multi-parent advanced generation intercross (MAGIC) populations for quantitative trait loci identification. *Plant Genome, 9,* 1–14. doi: 10.3835/plantgenome2015.10.0109

Milner, S. G., Maccaferri, M., Huang, B. E., Mantovani, P., Massi, A., Frascaroli, E. (2016). A multi-parental cross population for mapping QTL for agronomic traits in durum wheat (*Triticum turgidum* ssp. durum). *Plant Biotechnol. J., 14*, 735–748. doi: 10.1111/pbi.12424

Mitchell-Olds, T. (2010). Complex-trait analysis in plants. *Genome Biol., 11*, 113. doi: 10.1186/gb-2010-11-4-113

Monforte, A. J., and Tanksley, S. D. (2000). Development of a set of near isogenic and backcross recombinant inbred lines containing most of the *Lycopersicon hirsutum* genome in a *L. esculentum* genetic background: A tool for gene mapping and gene discovery. *Genome, 43,* 803–813. doi: 10.1139/g00-043

Moore, G. (2015). Strategic pre-breeding for wheat improvement. *Nat. Plants, 1,* 10–12, doi: 10.1038/nplants.2015.18

Mott, R., Talbot, C. J., Turri, M. G., Collins, A. C., and Flint, J. (2000). A method for fine mapping quantitative trait loci in outbred animal stocks. *Proc. Natl. Acad. Sci.* U.S.A., *97,* 12649–12654. doi: 10.1073/pnas.230304397

Murovec, J., and Bohanec, B. (2012). Haploids and doubled haploids in plant breeding. In *Plant Breeding*, Abdurakhmonov, I., InTechOpen, 87–106. ISBN 9789533079325

Murphy, K. M., Carter, A. H., Jones, S. S. (2013). "Evolutionary breeding and climate change." in *Genomics and breeding for climate-resilient crops: concepts and strategies*, Chittaranjan, K., Eds., Springer: Berlin, Germany, 377–389. ISBN 9783642370458

Muthu, V., Abbai, R., Nallathambi, J., Rahman, H., Ramasamy, S., Kambale, R., *et al*. (2020). Pyramiding QTLs controlling tolerance against drought, salinity, and submergence in rice through marker assisted breeding. *PLoS One, 15*, 1–18. doi: 10.1371/journal.pone.0227421

Nesbitt, T. C., and Tanksley, S. D. (2002). Comparative sequencing in the genus *Lycopersicon*: implications for the evolution of fruit size in the domestication of cultivated tomatoes. *Genetics, 162*, 365–379. doi: 10.1093/genetics/162.1.365

Ogawa, D., Yamamoto, E., Ohtani, T., Kanno, N., Tsunematsu, H., Nonoue, Y. (2018). Haplotype-based allele mining in the Japan-MAGIC rice population. *Sci. Rep., 8*, 1–11. doi: 10.1038/s41598-018-22657-3

Ongom, P. O., and Ejeta, G. (2018). Mating design and genetic structure of a multi-parent advanced generation intercross (MAGIC) population of sorghum (*Sorghum bicolor* (L.) moench). *G3 Genes, Genomes, Genet., 8*, 331–341. doi: 10.1534/g3.117.300248

Pandey, M. K., Roorkiwal, M., Singh, V. K., Ramalingam, A., Kudapa, H., Thudi, M., *et al*. (2016). Emerging genomic tools for legume breeding: current status and future prospects. *Front. Plant Sci., 7,* 455. doi: 10.3389/fpls.2016.00455

Pascual, L., Desplat, N., Huang, B. E., Desgroux, A., Bruguier, L., Bouchet, J. P., *et al*. (2015). Potential of a tomato MAGIC population to decipher the genetic control of quantitative traits and detect causal variants in the resequencing era. *Plant Biotechnol. J.,* 13, 565–577. doi: 10.1111/pbi.12282

Pingali, P. L. (2012). Green revolution: impacts, limits, and the path ahead. *Proc. Natl. Acad. Sci.* U.S.A*., 109*, 12302–12308. doi: 10.1073/pnas.0912953109

Pradhan, P., Fischer, G., Van Velthuizen, H., Reusser, D. E., and Kropp, J. P. (2019). Closing yield gaps: how sustainable can we be? *PLoS One, 10*, e0129487. doi: 10.1371/journal.pone.0129487

Price, A. H. (2006). Believe it or not, QTLs are accurate! *Trends Plant Sci., 11*, 213–216. doi: 10.1016/j.tplants.2006.03.006

Prohens, J., Gramazio, P., Plazas, M., Dempewolf, H., Francisco, F., Marı, B.K., *et al*. (2017). Introgressiomics: a new approach for using crop wild relatives in breeding for adaptation to climate change. *Euphytica, 213*, 158. doi: 10.1007/s10681-017-1938-9

Purugganan, M. D. (2019). Evolutionary insights into the nature of plant domestication. *Curr. Biol., 29*, R705–R714. doi: 10.1016/j.cub.2019.05.053

Raggi, L., Ciancaleoni, S., Torricelli, R., Terzi, V., Ceccarelli, S., and Negri, V. (2017). Evolutionary breeding for sustainable agriculture: selection and multi-environmental evaluation of barley populations and lines. *F. Crop. Res., 204*, 76–88. doi: 10.1016/j.fcr.2017.01.011

Rakshit, S., Rakshit, A., and Patil, J. V. (2012). Multi-parent intercross populations in analysis of quantitative traits. *J. Genet., 91*, 111–117. doi: 10.1007/s12041-012-0144-8

Ranc, N., Muños, S., Xu, J., Paslier, M. C. Le, Chauveau, A., Bounon, R. (2012). Genome-wide association mapping in tomato (*Solanum lycopersicum*) is possible using genome admixture of *Solanum lycopersicum* var. *cerasiforme*. *G3 Genes, Genomes, Genet., 2*, 853–864. doi: 10.1534/g3.112.002667

Ray, D. K., Mueller, N. D., West, P. C., and Foley, J. A. (2013). Yield trends are insufficient to double global crop production by 2050. *PLoS One, 8*, e66428. doi: 10.1371/journal.pone.0066428

Ray, D. K., West, P. C., Clark, M., Gerber, J. S., Prishchepov, A. V., and Chatterjee, S. (2019). Climate change has likely already affected global food production. *PLoS One, 14*, 1–18. doi: 10.1371/journal.pone.0217148

Rebetzke, G. J., Verbyla, A. P., Verbyla, K. L., Morell, M. K., and Cavanagh, C. R. (2014). Use of a large multi-parent wheat mapping population in genomic dissection of coleoptile and seedling growth. *Plant Biotechnol. J., 12*, 219–230. doi: 10.1111/pbi.12130

Reddy, B. V. S., and Stenhouse, J. W. (1994). Sorghum improvement for semi-arid tropics region: Past, current and future research thrusts in Asia. *PKV Res. J., 18*, 155–169.

Sallam, A., and Martsch, R. (2015). Association mapping for frost tolerance using multi-parent advanced generation inter-cross (MAGIC) population in faba bean (*Vicia faba* L.). *Genetica, 143*, 501–514. doi: 10.1007/s10709-015-9848-z

Sannemann, W., Huang, B. E., Mathew, B., and Léon, J. (2015). Multi-parent advanced generation inter-cross in barley: high-resolution quantitative trait locus mapping for flowering time as a proof of concept. *Mol. Breed., 35*. doi: 10.1007/s11032-015-0284-7

Sannemann, W., Lisker, A., Maurer, A., Léon, J., Kazman, E., Cöster, H., *et al*. (2018). Adaptive selection of founder segments and epistatic control of plant height in the MAGIC winter wheat population WM-800. *BMC Genomics, 19*, 1–16. doi: 10.1186/s12864-018-4915-3

Saxena, R. K., and Varshney, R. K. (2017). "Whole-genome sequencing of pigeonpea: requirement, background history, current status and future prospects for crop improvement" in *The pigeonpea genome*, Varshney, R. K., Saxena, R. K., Jackson, A. A., Eds., Springer: Berlin, Germany. ISBN 9783319637952

Scott, M. F., Ladejobi, O., Amer, S., Bentley, A. R., Biernaskie, J., Boden, S. A., *et al*. (2020). Multi-parent populations in crops: a toolbox integrating genomics and genetic mapping with breeding. *Heredity*, 1–21. doi: 10.1038/s41437-020-0336-6

Sehgal, D., and Dreisigacker, S. (2019). Haplotypes-based genetic analysis: Benefits and challenges. *Vavilovskii Zhurnal Genet. Selektsii, 23*, 803–808. doi: 10.18699/VJ19.37-O

Shah, R., Huang, B. E., Whan, A., Newberry, M., Verbyla, K., Morell, M. K., *et al*. (2019). The complex genetic architecture of recombination and structural variation in wheat uncovered using a large 8-founder MAGIC population. *bioRxiv*. doi: 10.1101/594317

Shekhawat, N., Singh, K., Sharma, V., and Meghwal, D. R. (2020). MAGIC populations: usefulness to define genetic basis of complex crop traits. *Food Sci. Reports, 1*, 53–55.

Shivakumar, M., Kumawat, G., Gireesh, C., Ramesh, S. V., and Husain, S. M. (2018). Identification of unique characteristics of deception from facial expression. *Curr. Sci., 114*, 901–906. doi: 10.18520/cs/v114/i04/901-906

Smýkal, P., Nelson, M. N., Berger, J. D., and Von Wettberg, E. J. B. (2018). The impact of genetic changes during crop domestication. *Agronomy, 8*, 1–22. doi: 10.3390/agronomy8070119

Stadlmeier, M., Hartl, L., and Mohler, V. (2018). Usefulness of a multi-parent advanced generation intercross population with a greatly reduced mating design for genetic studies in winter wheat. *Front. Plant Sci., 9*, 1–12. doi: 10.3389/fpls.2018.01825

Suneson, C. A. (1956). An evolutionary plant breeding method. *Agron. J., 48*, 188–191. doi: 10.2134/agronj1956.00021962004800040012x

Templeton, A. R. (2006). Population genetics and microevolutionary theory, Wiley-Liss: New York, 705. ISBN 9780471409519

Thépot, S., Restoux, G., Goldringer, I., Hospital, F., Gouache, D., Mackay, I., *et al*. (2015). Eciently tracking selection in a multi-parental population: The case of earliness in wheat. *Genetics, 199*, 609–623. doi: 10.1534/genetics.114.169995

Tian, D., Wang, P., Tang, B., Teng, X., Li, C., Liu, X., *et al*. (2020). GWAS Atlas: a curated resource of genome-wide variant-trait associations in plants and animals. *Nucleic Acids Res.,* 48, D927–D932. doi: 10.1093/nar/gkz828

Valdar, W., Flint, J., and Mott, R. (2006). Simulating the collaborative cross: Power of quantitative trait loci detection and mapping resolution in large sets of recombinant inbred strains of mice. *Genetics, 172,* 1783–1797. doi: 10.1534/genetics.104.039313

Varshney, R. K. (2016). Exciting journey of 10 years from genomes to fields and markets: some success stories of genomics-assisted breeding in chickpea, pigeonpea and groundnut. *Plant Sci., 242*, 98–107. doi: 10.1016/j.plantsci.2015.09.009

Verbyla, A. P., Cullis, B. R., and Thompson, R. (2007). The analysis of QTL by simultaneous use of the full linkage map. *Theor. Appl. Genet., 116*, 95–111. doi: 10.1007/s00122-007-0650-x

Verbyla, A. P., Cavanagh, C. R., and Verbyla, K. L. (2014). Whole-genome analysis of multienvironment or multitrait QTL in MAGIC. *G3 Genes, Genomes, Genet., 4*, 1569–1584. doi: 10.1534/g3.114.012971

Visscher, P. M., Brown, M. A., McCarthy, M. I., and Yang, J. (2012). Five years of GWAS discovery. *Am. J. Hum. Genet., 90*, 7–24. doi: 10.1016/j.ajhg.2011.11.029

Wada, T., Oku, K., Nagano, S., Isobe, S., Suzuki, H., Mori, M., *et al*. (2017). Development and characterization of a strawberry MAGIC population derived from crosses with six strawberry cultivars. *Breed. Sci., 67*, 370–381. doi: 10.1270/jsbbs.17009

Wang, C., Fan, J., Long, L., Yang, J., Yu, X., Bao, B. W. (2016). Ecological response of rice multi-genotype variety in Dehong, Yunnan Province. *Yunnan Nong Ye Da Xue Xue Bao, 1*, 1–6. doi: 10.5555/20163075110

Wang, H., Guo, X., Pandey, M. K., Ji, X., Varshney, R. K., Nwosu, V., *et al*. (2017). History and impact of the international peanut genome initiative: the exciting journey toward peanut whole-genome sequencing. In *The Peanut Genome*, Varshney, R.K., Pandey, M.K., Puppala, N., Eds., Springer: Berlin, 117–133. ISBN 9783319639352

Watson, A., Ghosh, S., Williams, M. J., Cuddy, W. S., Simmonds, J., Rey, M. D., *et al*. (2018). Speed breeding is a powerful tool to accelerate crop research and breeding. *Nat. Plants, 4*, 23–29. doi: 10.1038/s41477-017-0083-8

Willer, H., Schlatter, B., Trávníček, J., Kemper, L., and Lernoud, J. (2020). The world of organic agriculture statistics and emerging trends 2020, Research Institute of Organic Agriculture (FiBL): Frick, Switzerland, 337. ISBN 9783037361184

Wilson, S. R. (1978). A note on assortative mating, linkage and genotypic frequencies. *Ann. Hum. Genet., 42*, 129–130. doi: 10.1111/j.1469-1809.1978.tb00937.x

Xu, Y., Li, P., Yang, Z., and Xu, C. (2017). Genetic mapping of quantitative trait loci in crops. *Crop J., 5*, 175–184. doi: 10.1016/j.cj.2016.06.003

Yamamoto, E., Iwata, H., Tanabata, T., Mizobuchi, R., Yonemaru, J. ichi, Yamamoto, T., *et al*. (2014). Effect of advanced intercrossing on genome structure and on the power to detect linked quantitative trait loci in a multi-parent population: A simulation study in rice. *BMC Genet., 15*, 1–17. doi: 10.1186/1471-2156-15-50

Yan, W., Zhao, H., Yu, K., Wang, T., Khattak, A. N., and Tian, E. (2020). Development of a multi-parent advanced generation intercross (MAGIC) population for genetic exploitation of complex traits in *Brassica juncea*: glucosinolate content as an example. *Plant Breed*., 1–11. doi: 10.1111/pbr.12820

Yoshimi, M., Kitamura, Y., Isshiki, S., Saito, T., Yasumoto, K., Terachi, T., *et al*. (2013). Variations in the structure and transcription of the mitochondrial atp and cox genes in wild *Solanum* species that induce male sterility in eggplant (*S. melongena*). *Theor. Appl. Genet., 126,* 1851–1859. doi: 10.1007/s00122-013-2097-6

Zaw, H., Raghavan, C., Pocsedio, A., Swamy, B. P. M., Jubay, M. L., Singh, R. K., *et al*. (2019). Exploring genetic architecture of grain yield and quality traits in a 16-way

indica by japonica rice MAGIC global population. *Sci. Rep., 9*, 1–11. doi: 10.1038/s41598-019-55357-7

Zhang, L., Li, H., Ding, J., Wu, J., and Wang, J. (2015). Quantitative trait locus mapping with background control in genetic populations of clonal F_1 and double cross. *J. Integr. Plant Biol., 57,* 1046–1062. doi: 10.1111/jipb.12361

Zhao, F., Zhao, H., Wang, X., and Li, X. (2017). Construction and application potential of MAGIC population on genetic breeding of rapeseed (*Brassica napus* L.). *ChineseJ. Oil Crop Sci., 39,* 145.

Zheng, C., Boer, M. P., and van Eeuwijk, F. A. (2019). Construction of genetic linkage maps in multi-parental populations. *Genetics, 212*, 1031–1044. doi: 10.1534/genetics.119.302229

OBJECTIVES

Highlighting inter-specific MAGIC populations as valuable pre-breeding resources.

In the present doctoral thesis, it is described the development of two inter-specific MAGIC populations in eggplant and tomato. We aimed at exploiting the huge phenotypic and genetic diversity of these next-generation pre-breeding resources, which is further enhanced by the introduction of wild founders. We intend to demonstrate de advantages of using MAGIC populations in breeding programs as powerful tools for the detection of genes/QTLs responsible for a trait of interest. Particularly, this thesis is focused in the study of fruit pigmentation traits involved not only in fruit appearance but also in fruit nutritional quality.

Therefore, we proposed the following specific main objectives:

1. Characterization of the first eggplant MAGIC (S3MEGGIC) population developed by the inter-cross of seven cultivated *S. melongena* and one wild relative *S. incanum* founders.

    1.1. Genotyping of the S3MEGGIC population using the 5k probes eggplant SPET platform.

    1.2. Evaluation of the population structure, the residual heterozygosity, and the founder contribution to the final population.

    1.3. Evaluation of the S3MEGGIC for anthocyanin presence in vegetative plant tissues (PA), fruit epidermis (FA), and light-insensitive anthocyanin pigmentation under the calyx (PUC), for candidate genes identification.

    1.4. Identification of the main responsible gene for the biosynthesis of chlorophylls in the eggplant fruit peel (FC).

    1.5. Understanding the difference between the fruit green netting (FN) and the FC traits, identification of the main responsible gene and elucidation of an improved gene annotation.

    1.6. Validation of FC and FN candidates and identification of new sequence variants in an eggplant germplasm collection for tracing the evolutionary changes that genes have undergone over the course of domestication.

2. Development of a tomato MAGIC (ToMAGIC) population developed by the inter-cross of four *S. lycopersicum* var. *cerasiforme* and four wild *S. pimpinellifolium* founders.

    2.1. Genotyping of the ToMAGIC population using the 12k probes tomato SPET panel.

    2.2. Evaluation of the population structure, the genetic relationship among final lines, and the founder contribution to the final population.

    2.3. Proof-of-concept of the for testing the potential of the ToMAGIC population for the high-precision fine mapping of a set of traits from different plant organs.

RESULTS

Deciphering a tiny fraction of the great complexity of genomes.

Chapter I:
Eggplant MAGIC population for multiple fruit traits breeding

Eggplant exists.

Research article

Newly developed MAGIC population allows identification of strong associations and candidate genes for anthocyanin pigmentation in eggplant

Giulio Mangino[1†], Andrea Arrones[1†], Mariola Plazas[2], Torsten Pook[3], Jaime Prohens[1], Pietro Gramazio[2*] and Santiago Vilanova[1*]

[1]Instituto de Conservación y Mejora de la Agrodiversidad Valenciana, Universitat Politècnica de València, Valencia, Spain

[2]Instituto de Biología Molecular y Celular de Plantas, Consejo Superior de Investigaciones Científicas-Universitat Politècnica de València, Valencia, Spain

[3]Animal Breeding and Genetics Group, Department of Animal Sciences, Center for Integrated Breeding Research, University of Göttingen, Göttingin, Germany

[*]Corresponding authors

[†]These authors have contributed equally to this work

Ph.D. candidate contribution

Together with G.M., A.A. had a main role in the following activities: performing the experiments, data collection, data analysis, data visualization, drafting manuscript, manuscript review and editing.

Citation: 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,* 847789. doi: 10.3389/fpls.2022.847789

Abstract

Multi-parent advanced generation inter-cross (MAGIC) populations facilitate the genetic dissection of complex quantitative traits in plants and are valuable breeding materials. We report the development of the first eggplant MAGIC population (S3 Magic EGGplant InCanum, S3MEGGIC; 8-way), constituted by the 420 S3 individuals developed from the intercrossing of seven cultivated eggplant (*Solanum melongena*) and one wild relative (*S. incanum*) parents. The S3MEGGIC recombinant population was genotyped with the eggplant 5k probes SPET platform and phenotyped for anthocyanin presence in vegetative plant tissues (PA) and fruit epidermis (FA), and for the light-insensitive anthocyanic pigmentation under the calyx (PUC). The 7,724 filtered high-confidence single-nucleotide polymorphisms (SNPs) confirmed a low residual heterozygosity (6.87%), a lack of genetic structure in the S3MEGGIC population, and no differentiation among subpopulations carrying a cultivated or wild cytoplasm. Inference of haplotype blocks of the nuclear genome revealed an unbalanced representation of the founder genomes, suggesting a cryptic selection in favour or against specific parental genomes. Genome-wide association study (GWAS) analysis for PA, FA, and PUC detected strong associations with two myeloblastosis (MYB) genes similar to *MYB113* involved in the anthocyanin biosynthesis pathway, and with a *COP1* gene which encodes for a photoregulatory protein and may be responsible for the PUC trait. Evidence was found of a duplication of an ancestral *MYB113* gene with a translocation from chromosome 10 to chromosome 1 compared with the tomato genome. Parental genotypes for the three genes were in agreement with the identification of the candidate genes performed in the S3MEGGIC population. Our new eggplant MAGIC population is the largest recombinant population in eggplant and is a powerful tool for eggplant genetics and breeding studies.

Keywords: multi-parent advanced generation inter-crosses (MAGIC), eggplant (*Solanum melongena* L.), *S. incanum*, anthocyanins, pigmentation under calyx (PUC), genome wide association study (GWAS), candidate genes, SPET (single primer enrichment technology)

1. Introduction

Multi-parent experimental populations are of great interest for the genetic dissection of quantitative traits and for the development of new recombinant materials for plant breeding (Huang *et al.,* 2015). Despite their complex management and resources requirement, multi-parent advanced generation inter-cross (MAGIC) populations represent powerful next-generation mapping tools by combining a high genetic diversity and recombination with a low population structure (Arrones *et al.,* 2020; Scott *et al.,* 2020). MAGIC populations are already available in model species, such as *Arabidopsis thaliana* and in several crops, such as cereals, pulses, and vegetables (Kover *et al.,* 2009; Bandillo *et al.,* 2013; Pascual *et al.,* 2015; Huynh *et al.,* 2018), and have demonstrated their power to dissect the structure of complex traits (Dell'Acqua *et al.,* 2015; Stadlmeier *et al.,* 2018).

Although the available MAGIC populations have become a useful resource for genetic studies and breeding, most of them have only exploited intra-specific variation. The incorporation of crop wild relatives (CWRs) as founders could be a way of including multiple wild genomic fragments or introgressions into cultivated background genomes (Arrones *et al.,* 2020). Apart from being of great interest for genetic analysis, inter-specific MAGIC populations can be useful for broadening the genetic base of crops and provide new variation for breeding multiple traits, including those related to adaption to climate change (Gramazio *et al.,* 2020a). However, so far, the potential of inter-specific MAGIC populations for plant breeding has largely remained unexploited (Arrones *et al.,* 2020).

Eggplant (*Solanum melongena* L.) is a major vegetable crop of increasing importance, ranking fifth in global production among vegetables (FAOSTAT, 2019). Despite its economic importance, eggplant has lagged behind other major crops and little effort has been made to develop immortal experimental populations and genetic and genomic tools (Gramazio *et al.,* 2018, 2019). So far, only one population of recombinant inbred lines (RIL) and one set of introgression lines (ILs) are publicly available (Lebeau *et al.,* 2013; Gramazio *et al.,* 2017a), while no multi-parent population has been developed, so far. Conversely, in other related Solanaceae crops, such as tomato, several experimental populations have been developed, including MAGIC populations, which have allowed great advances in the genetic dissection of traits of interest (Pascual *et al.,* 2015; Campanelli *et al.,* 2019). For this reason, the development of this type of population would represent a landmark in eggplant breeding.

Anthocyanins are responsible for a set of specific and relevant traits in eggplant which can be used as a model plant for other crops (Moglia *et al.,* 2020). Anthocyanins play a key role in plant defence mechanisms, and their synthesis and

accumulation may vary in response to specific biotic and abiotic stresses (D'Amelia *et al.,* 2018; Lv *et al.,* 2019; Zhou *et al.,* 2019). In addition, anthocyanins may prevent and alleviate human chronic diseases and provide health benefits (Toppino *et al.,* 2020). In eggplant, anthocyanins are responsible for the purple colour of the peel, one of the traits of greatest interest for eggplant breeding (Daunay and Hazra, 2012). Purple coloured eggplant fruits are the most demanded in various markets (Li *et al.,* 2018), and developing dark purple-coloured eggplants, which results from the combination of anthocyanins with chlorophylls, is a major objective in eggplant breeding programmes. Eggplants are variable for the presence of anthocyanins not only in fruits, but also in other plant organs and tissues such as hypocotyl, stem, leaves, leaf veins, prickles, flower calyx, or corolla (Toppino *et al.,* 2020). Anthocyanin biosynthesis has been widely studied in Solanaceae species (Van Eck *et al.,* 1994; Borovsky *et al.,* 2004; Gonzali *et al.,* 2009), but its genetic control in eggplant has not been fully clarified. The anthocyanin biosynthetic pathway is a very conserved network in many plant species, where many enzymes and regulatory transcription factors (TFs) are involved (Albert *et al.,* 2014).

The major anthocyanins in the eggplant fruit epidermis are delphinidin-3-p-coumaroylrutinoside-5-glucoside and delphinidin-3-rutinoside (Mennella *et al.,* 2012; Moglia *et al.,* 2020). Fruit anthocyanins (FA) are highly dependent on light (Jiang *et al.,* 2016), however, the genotypes that carry the pigmentation under the calyx (PUC) mutation are able to independently synthetise the anthocyanins in the fruit epidermis of the incidence of light (Tigchelaar *et al.,* 1968). Quantitative trait locus (QTL)-related studies using family based or genome-wide association (GWA) mapping approaches evidenced that chromosome 10 harbours most of the QTL/genes involved in anthocyanin formation, distribution, and accumulation (Doganlar *et al.,* 2002; Barchi *et al.,* 2012; Cericola *et al.,* 2014; Frary *et al.,* 2014; Toppino *et al.,* 2016, 2020; Wei *et al.,* 2020). The availability of high-quality eggplant genome sequences and transcriptomic data allowed the identification of putative candidate genes belonging to the myeloblastosis (MYB) family which controls the variation in anthocyanin content and fruit colour in eggplant (Docimo *et al.,* 2016; Li *et al.,* 2017, 2018, 2021; Xiao *et al.,* 2018; Moglia *et al.,* 2020; Shi *et al.,* 2021), highlighting their synteny with other Solanaceae. However, the genes underlying the PUC phenotype have not been identified, so far.

Here, we report on the first eggplant MAGIC population derived from an inter-specific cross of seven accessions of *S. melongena* and its wild relative *S. incanum* (Gramazio *et al.,* 2019). It represents the largest experimental population described, so far, in eggplant, with a similar population size to MAGIC populations in other solanaceous crops. The population has been genotyped by applying the Single Primer Enrichment Technology (SPET) to explore its genetic architecture and the contribution of founders to the final population, and phenotyped for the presence

of anthocyanins in the fruit epidermis and other plant organs and for the PUC trait. These traits were chosen due to their physiological, agronomic, and morphological relevance, high stability, and heritability. An association analysis has been performed to locate the genomic regions and to identify the candidate genes involved in the traits under study.

2. Materials and methods

2.1. Multi-parent advanced generation inter-cross population construction

The eggplant MAGIC population has been developed by intermating seven cultivated eggplants, i.e., MM1597 (A), DH ECAVI (B), AN-S-26 (D), H15 (E), A0416 (F), IVIA-371 (G) and ASI-S-1 (H), and the *S. incanum* accession MM577 (C) (Figure 1A). The wild relative founder was chosen for its tolerance to some biotic and abiotic stresses, mainly drought (Knapp *et al.,* 2013), and for showing a high phenolic content (Prohens *et al.,* 2013). The performance of the founders was comprehensively characterised in previous morphoagronomic and genetic diversity studies (Hurtado *et al.,* 2014; Gramazio *et al.,* 2017b; Kaushik *et al.,* 2018). In addition, their genomes had been resequenced (Gramazio *et al.,* 2019). The latter study highlighted that, in the founder parents, the residual heterozygosity was less than 0.06%.

In order to develop the eggplant S3 Magic EGGplant InCanum (S3MEGGIC) population, founder lines had been inter-crossed by following a simple "funnel" approach (Wang *et al.,* 2017; Arrones *et al.,* 2020; Figure 1A). The eight founders (A–H) had been pairwise inter-crossed to produce two-way or simple F_1 hybrids (AB, CD, EF, and GH), which were subsequently inter-crossed in pairs (AB × CD and EF × GH) to obtain two four-way or double hybrids (ABCD and EFGH). In order to achieve a complete admixture of all founder genomes and to avoid assortative mating, the double hybrids were inter-crossed following a chain pollination scheme, with each individual being used as female and male parents (Díez *et al.,* 2002; Supplementary Figure 1A).

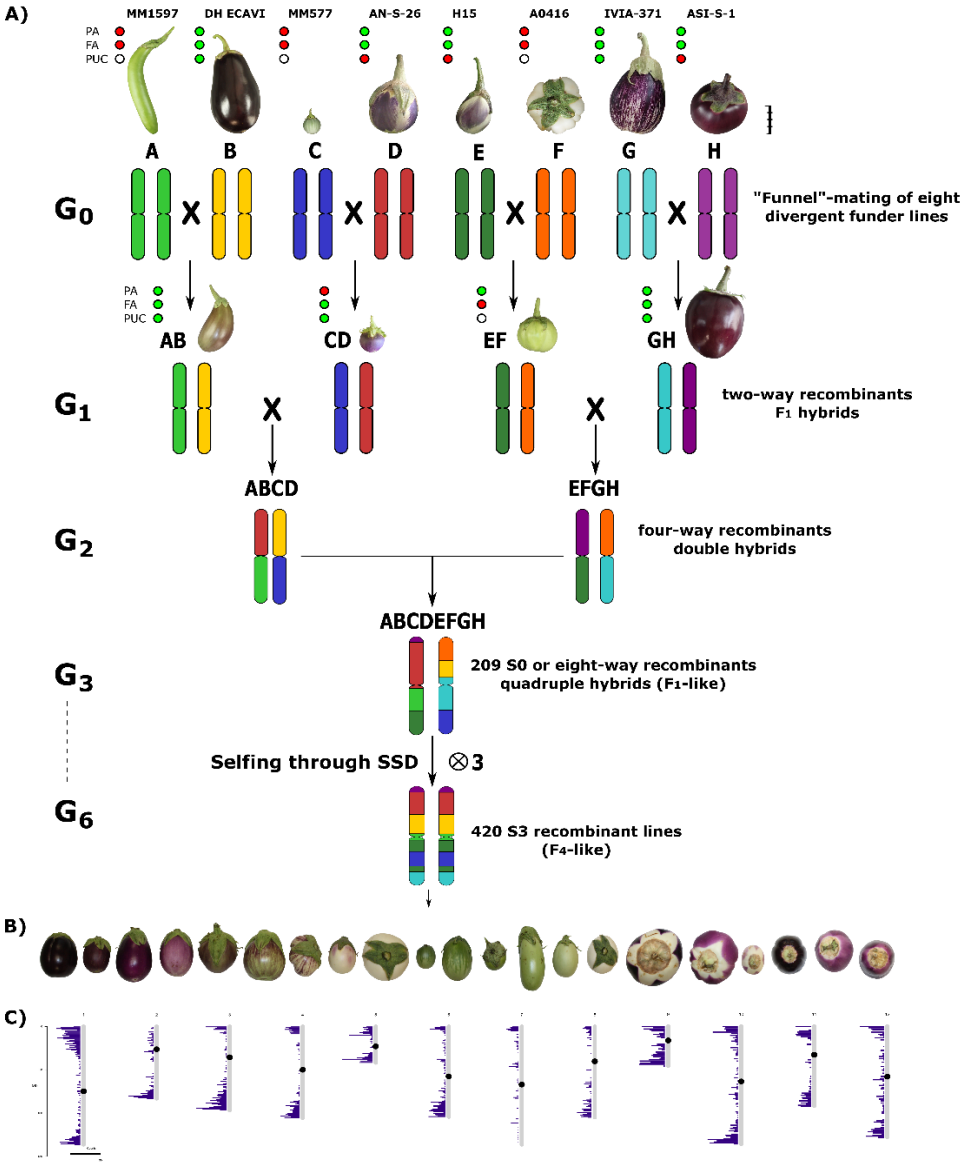


Figure 1. (A) The funnel breeding design, used across the six generations (G1–G6), to develop the 420 S3 individuals of the S3MEGGIC population. The eight parents, coded from A to H and each with a different colour to represent their genomic background, are represented above at a scale based on the real fruit size. Scale bar represents 5 cm. The four two-way hybrids obtained in the G1 generation (AB, CD, EF, and GH) are also represented at the same scale as the founders. Phenotyping of founders and two-way hybrids for absence (red) or presence (green) of anthocyanins in vegetative plant tissues (PA) and fruit epidermis (FA), or anthocyanic pigmentation under the calyx (PUC). White dots for PUC mean uncertainty for non-anthocyanin

fruits. (B) A representation of the phenotypic diversity of the S3 individuals found during the phenotyping. (C) Distribution of molecular markers across the chromosomes used for the genotyping.

All the obtained eight-way or quadruple hybrids (S0 generation) presented all the eight randomly shuffled genomes and only differed on the cytoplasm inherited from the maternal parent. The S0 progenies obtained using the double hybrid ABCD, as a female parent carried the cytoplasm of the wild *S. incanum* MM577, while those derived using the double hybrid EFGH as a female parent carried the cytoplasm of the cultivated *S. melongena* ASI-S-1. Subsequently, the S0 progenies were selfed for three generations by a single seed descent (SSD) to obtain the S3 segregating individuals that were phenotyped and genotyped in this study. To ensure the continuity of the S0 progenies and to accelerate the self-fertilisation process, four plants of each S0 progeny were germinated, and only the first two that set a viable seed were selected for the next generation (S1; Supplementary Figure 1B). From each of the two S0 selected plants, two S1 plants were germinated, and only the first setting fruit was selected for the S2 generation. The same was done for the S3 generation so that for each progeny, two plants were germinated and phenotyped. Despite this, only one was used for originating the next generation. On the other hand, in the S3 progenies, if two individuals displayed some phenotypic differences, both of them were included in the S3MEGGIC population.

2.2. Cultivation conditions

Seeds were germinated in Petri dishes, following the protocol developed by Ranil *et al.* (2015) and, subsequently, transferred to seedling trays in a climatic chamber under a photoperiod and temperature regime of 16 h light (25ºC) and 8 h dark (18ºC). After acclimatisation, plantlets were transplanted to 15 L pots and grown in a pollinator-free benched glasshouse of the Universitat Politècnica de València (UPV), Valencia, Spain (GPS coordinates: latitude, 39º 28’ 55’’ N; longitude, 0º 20’ 11’’ W; 7 m above sea level). Plants were spaced 1.2 m between rows and 1 m within the row, and fertirrigated using a drip irrigation system and trained with vertical strings. Pruning was done manually to regulate the vegetative growth and flowering. Phytosanitary treatments were performed when necessary. In order to shorten generation time of subsequent generations (S0–S3), plantlets were transplanted to individual thermoformed pots (1.3 L capacity) in a pollinator-free glasshouse, and selfings were stimulated by a mechanical vibration.

2.3. High-throughput genotyping

Young leaf tissue was sampled from 420 S3 individuals, the eight founders, and the four two-way hybrids. Genomic DNA was extracted using the silica matrix extraction (SILEX) extraction method (Vilanova *et al.,* 2020) and checked for quality and integrity by agarose electrophoresis and NanoDrop ratios (260/280 and 260/230), while its concentration was estimated with Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, United States). After dilution, the samples were sent to Identity Governance and Administration (IGA) Technology Services (IGATech, Udine, Italy) for library preparation and sequencing with NextSeq500 sequencer (150 paired-end) for a high-throughput genotyping using the SPET technology using the 5k probes eggplant SPET platform (Barchi *et al.,* 2019a). The latter comprises 5,093 probes, and was developed by filtering out the most informative and reliable polymorphisms (3,372 of them in coding sequences or CDS and 1,721 in introns and untranslated regions or UTR regions) from the set of over 12 million single-nucleotide polymorphisms (SNPs) identified among the MAGIC founders (Gramazio *et al.,* 2019).

Raw reads were demultiplexed and the adapters were removed using a standard Illumina pipeline and Cutadapt (Martin, 2011), while trimming was performed by ERNE (Del Fabbro *et al.,* 2013). Clean reads were mapped onto the eggplant reference genome "67/3" (Barchi *et al.,* 2019b) using BWA-MEM (Li, 2013) with default parameters, and only the uniquely aligned reads were selected for the variant calling performed with GATK 4.0 (DePristo *et al.,* 2011), following the best practice recommended by the Broad Institute (http://www.broadinstitute.org).

The SNPs identified by SPET were filtered using the Trait Analysis by Association, Evolution, and Linkage (TASSEL) software (ver. 5.0, Bradbury *et al.,* 2007) in order to retain the most reliable ones (minor allele frequency > 0.01, missing data < 10% and maximum marker heterozygosity < 70%). In addition, a linkage disequilibrium (LD) k-nearest neighbour genotype imputation method (LD KNNi) was performed to fill the missing calls or genotyping gaps.

2.4. Population structure, heterozygosity, and haplotype blocks inferring

A principal component analysis (PCA) was performed to assess the population structure of S3MEGGIC using the R package vcfR (Knaus and Grünwald, 2017) and the function glPCA of the Adegenet package (Jombart, 2008). Finally, the PCA was graphically plotted with ggplot2 (Wickham, 2016). An Analysis of Molecular

Variance (AMOVA) was performed to estimate the population differentiation according to the cytoplasm (cultivated vs. wild) of the individuals of the S3MEGGIC population by using the function poppr.amova of the poppr R package (Kamvar *et al.,* 2014). The residual heterozygosity and its distribution were evaluated with TASSEL software (ver. 5.0, Bradbury *et al.,* 2007). Parental contribution to S3MEGGIC individuals and haplotype blocks were estimated by using R-package HaploBlocker (Pook *et al.,* 2019).

2.5. Phenotyping and genome-wide association study

Phenotypic data were collected from the 420 S3 individuals grown during the 2019/2020 season. Three traits were screened using a binary classification (presence/absence), namely, anthocyanins presence in vegetative plant tissues (PA) and fruit epidermis (FA), and anthocyanic PUC (Supplementary Figure 2). The presence of PA was phenotyped when the purple coloration was observed in any vegetative plant parts such as in stem, branches, leaf veins, or prickles. For FA, anthocyanins were considered as present when the purple colouration was observed in the fruit epidermis regardless of their distribution (uniform, listed, etc.) or intensity. The PUC trait could only be phenotyped in anthocyanic fruits by removing the calyx and observing the presence of anthocyanins under it. FA and PUC traits were screened at the stage of commercial maturity (e.g., when the fruit was physiologically immature) which is the best stage for phenotyping these traits. Phenotypic characteristics of the eight founders and two-way hybrids are described in Figure 1A.

Using the phenotypic and genotypic data collected from the S3MEGGIC individuals, a Genome-Wide Association Study (GWAS) was performed for the selected traits using the TASSEL software (ver. 5.0, Bradbury *et al.,* 2007). For the association study, mixed linear model (MLM) analyses were conducted. The MLM analysis uses both fixed and random effects, which incorporates the kinship among the individuals. The multiple testing was corrected with the Bonferroni and the false discovery rate (FDR) methods (Holm, 1979; Benjamini and Hochberg, 1995) to identify candidate-associated regions at the significance level of 0.05 (Thissen *et al.,* 2002). SNPs with limit of detection (LOD) [–log10(p-value)] over these thresholds or cut off values were declared to be significantly associated with the anthocyanin presence. The R qqman package (Turner, 2018) was used to visualise the Manhattan plots and LDBlockShow (Dong *et al.,* 2020) to determine the LD and the plot haplotype block structure. The pattern of pairwise LD between SNPs was measured by LD correlation coefficient (r^2) by considering haplotype blocks with default r^2 values greater than 0.5 supported by the solid spine of LD method (Gabriel *et al.,*

2002; Barrett *et al.,* 2005). The LD was used to narrow down the genomic regions with significant associations. The genes underlying the associated regions were retrieved from the "67/3" eggplant reference genome (ver. 3) (Barchi *et al.,* 2019b). Genes were considered as potential candidates in controlling the assessed traits when carrying homozygous allelic variants classified as "high impact" according to SnpEff software v 4.2 prediction (Cingolani *et al.,* 2012) of the eight MAGIC founders (Gramazio *et al.,* 2019). The Integrative Genomics Viewer (IGV) tool was used for visual exploration of founder genome sequences to validate SnpEff results and to confirm the presence of the so-called "high impact" variants (Robinson *et al.,* 2020). In addition, eggplant and tomato syntenies were assessed by a BLASTx search of candidate genes sequences against the tomato genome (version SL4.0) in the Sol Genomics Network database (http://www.solgenomics.net).

3. Results

3.1. Multi-parent advanced generation inter-cross population construction

Seven accessions of eggplant and one of the wild relative *S. incanum* were selected as founder parents (A-H) for the construction of the eggplant MAGIC population (S3MEGGIC). Following a funnel breeding scheme (Figure 1A and Supplementary Figure 1), a total of 420 individuals of the MAGIC populations were obtained. First, founders were pairwise inter-crossed to produce four two-way hybrids (AB, CD, EF, and GH), which were subsequently inter-crossed in pairs to obtain four-way hybrids (ABCD and EFGH). One-hundred and forty-nine individuals of each of the two four-way hybrids were inter-crossed using a chain pollination scheme (Supplementary Figure 1A). Out of the theoretical maximum of 298 eight-way hybrid progenies (S0), seeds were obtained for 209 of them, of which 116 carried the *S. melongena* ASI-S-1 cytoplasm and 93 carried the *S. incanum* MM577 cytoplasm. Two plants per S0 progeny were used to advance the population, reaching 402 S1 progenies. These S1 progenies were advanced through single seed descend (SSD) to obtain the 391 S2 and 305 S3 MAGIC progenies. The final S3MEGGIC population was constituted by 420 S3 individuals, of which 348 individuals carried the cultivated cytoplasm and 72 the wild cytoplasm.

3.2. Single-primer enrichment technology genotyping

The genotyping of the 420 S3 MAGIC individuals, the eight founders, and the four two-way hybrids by the eggplant SPET platform yielded 22,146 SNPs. After filtering, 7,724 high-confidence SNPs were retained for the subsequent analysis, and the low percentage of missing calls (0.53%) was imputed. Filtered SNPs were distributed across the entire eggplant genome, although the distribution of SNPs varied within and among chromosomes (Table 1 and Figure 1C). Chromosome 9 had the highest average marker density after the SNP filtering, with 17.31 SNPs per Mb, while chromosome 7 had the lowest with an average of 2.30 SNPs per Mb. Generally, most of the SNPs were located in regions with a high gene density and decayed around the centromere (Figure 1C). The S3 MAGIC individuals exhibited a heterozygosity average of 6.87%, with only 15 individuals (3.57%) with a proportion of residual heterozygosity higher than 20% (Figure 2).

Table 1. Statistics of the genotyping using the eggplant SPET platform of the 420 S3MEGGIC population individuals using the '67/3' eggplant reference genome (Barchi *et al.,* 2019b).

Chr	Markers	Filtered markers	% Markers	% Filtered markers	Chr length (Mb)	Marker density (markers/Mb)	Filtered marker density (markers/Mb)
1	3,234	1,205	14.60	15.60	136.53	23.69	8.83
2	1,119	424	5.05	5.49	83.34	13.43	5.09
3	2,004	758	9.05	9.81	97.01	20.66	7.81
4	1,567	588	7.08	7.61	105.67	14.83	5.56
5	1,438	571	6.49	7.39	43.85	32.79	13.02
6	2,108	794	9.52	10.28	108.97	19.35	7.29
7	2,483	328	11.21	4.25	142.38	17.44	2.30
8	1,392	550	6.29	7.12	109.58	12.70	5.02
9	1,841	625	8.31	8.09	36.10	51.00	17.31
10	2,069	798	9.34	10.33	106.64	19.40	7.48
11	1,260	453	5.69	5.86	72.29	17.43	6.27
12	1,631	630	7.36	8.16	100.42	16.24	6.27
Total	22,146	7,724	100	100	1,142.78		
Average						21.58	7.69

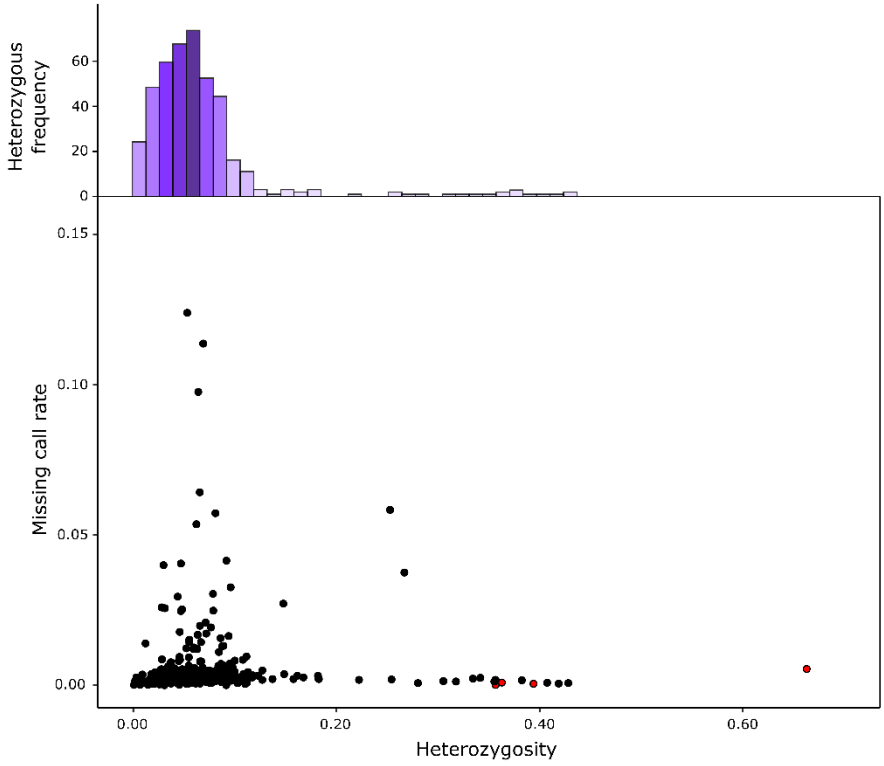


Figure 2. Heterozygosity and proportion of missing data from the S3MEGGIC population. The top histogram represents the observed heterozygosity which is skewed to the left (mean 6.87%; mode 5.02%). The lower graph reports the residual heterozygosity of S3 individuals as black dots compared against two-way hybrids represented as red dots.

3.3. Population structure

Population stratification, performed by PCA, indicated the absence of subgroups in the S3 individuals since no clear clustering was observed (Figure 3). The first two principal components (PCs) account for 5.15% (PC1) and 3.30% of the total variation, respectively, while the first 10 PCs explain together only 25.14% of the total variation, revealing the absence of genetic structure in the S3MEGGIC population. No differentiated clusters among individuals carrying the wild (*S. incanum* MM577) or the cultivated (*S. melongena* ASI-S-1) cytoplasm were observed. An Analysis of Molecular Variance (AMOVA) was also performed,

revealing that only 0.29% of the total sums of squares is accounted for the molecular variation among the *S. melongena* and *S. incanum* cytoplasm groups, thereby resulting in a very small phi-value of 0.0019, which indicates a low level of differentiation and supporting that no population structure exists.

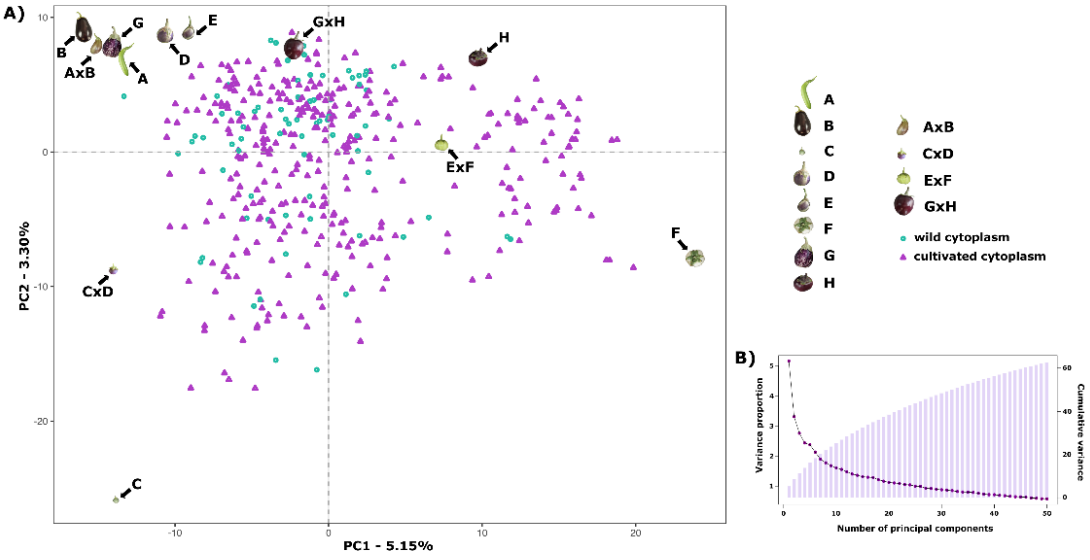


Figure 3. Results of a Principal Component Analysis (PCA) on the S3MEGGIC population. (A) PCA of the first two principal components (PCs), including S3 individuals, founder lines, and two-way hybrids. S3 individuals with wild cytoplasm are represented with blue dots while those ones with cultivated cytoplasm are represented with purple triangles. (B) Scree plot of the PCs (x-axis) and their contribution to variance (left y-axis); bar blot of the PCs (x-axis) and the cumulative proportion of variance explained (right y-axis).

The genome mosaics reconstruction of the S3 MAGIC individuals, in terms of the eight founder haplotypes, showed different haplotype block proportions depending on the genomic position for all chromosomes (Figure 4). The estimated contribution of some founders to the overall S3MEGGIC population differed from the expected value of 12.5%. Two of the founder genomes (A0416 and IVIA-371) had a high representation in the genome of the S3 individuals (32.6% and 23.6%, respectively), while two others (AN-S-26 and H15) had a small representation (0.3% in both cases). The wild founder, *S. incanum*, had an average haplotype representation of 5.8%.

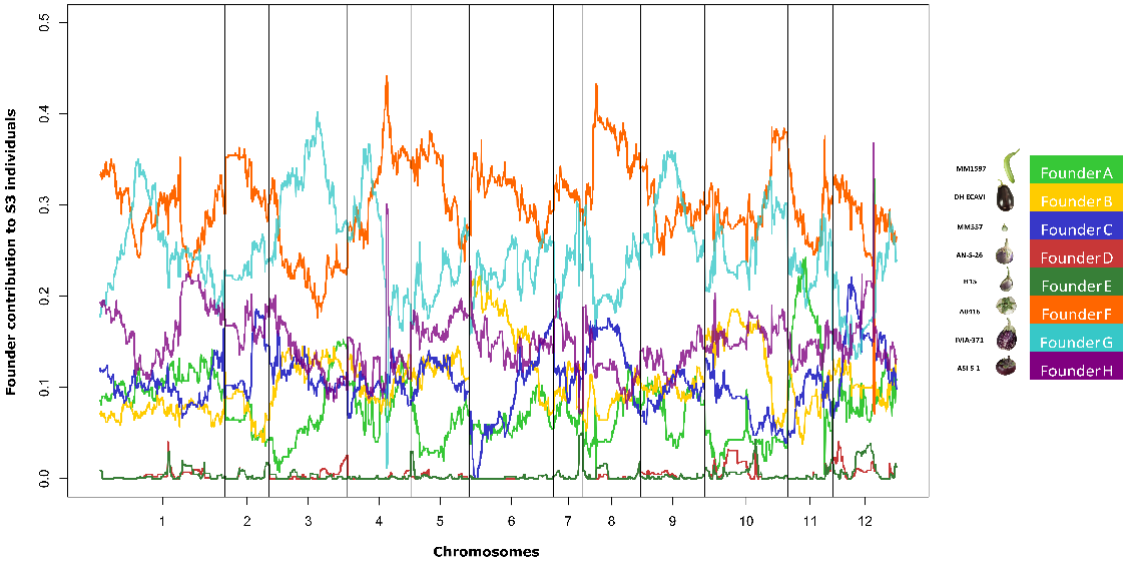


Figure 4. Genome-wide founder haplotype blocks assignment across the entire S3MEGGIC population. In the x-axis, 12 eggplant chromosomes are represented and, in the y-axis, the average percentage of founders' contribution for the 420 S3 individuals. In the legend, the colour code associated with each founder as in Figure 1.

3.4. Phenotypic variation among multi-parent advanced generation inter-cross individuals and association analysis

The screening for absence or presence of PA, FA, and PUC in purple fruits of the 420 S3 MAGIC individuals revealed a considerable variation (Figure 1B). Out of the 420 S3 individuals, 57.6% displayed PA and 37.5% FA. Among the individuals displaying FA, 64.3% had the PUC phenotype.

Given that no population structure was observed for the S3MEGGIC population, the phenotypic data, together with the genotypic information, were used for GWAS analysis (Figure 5). The GWAS was performed taking into account the kinship in an MLM leading to the identification of significant associations for the evaluated traits.

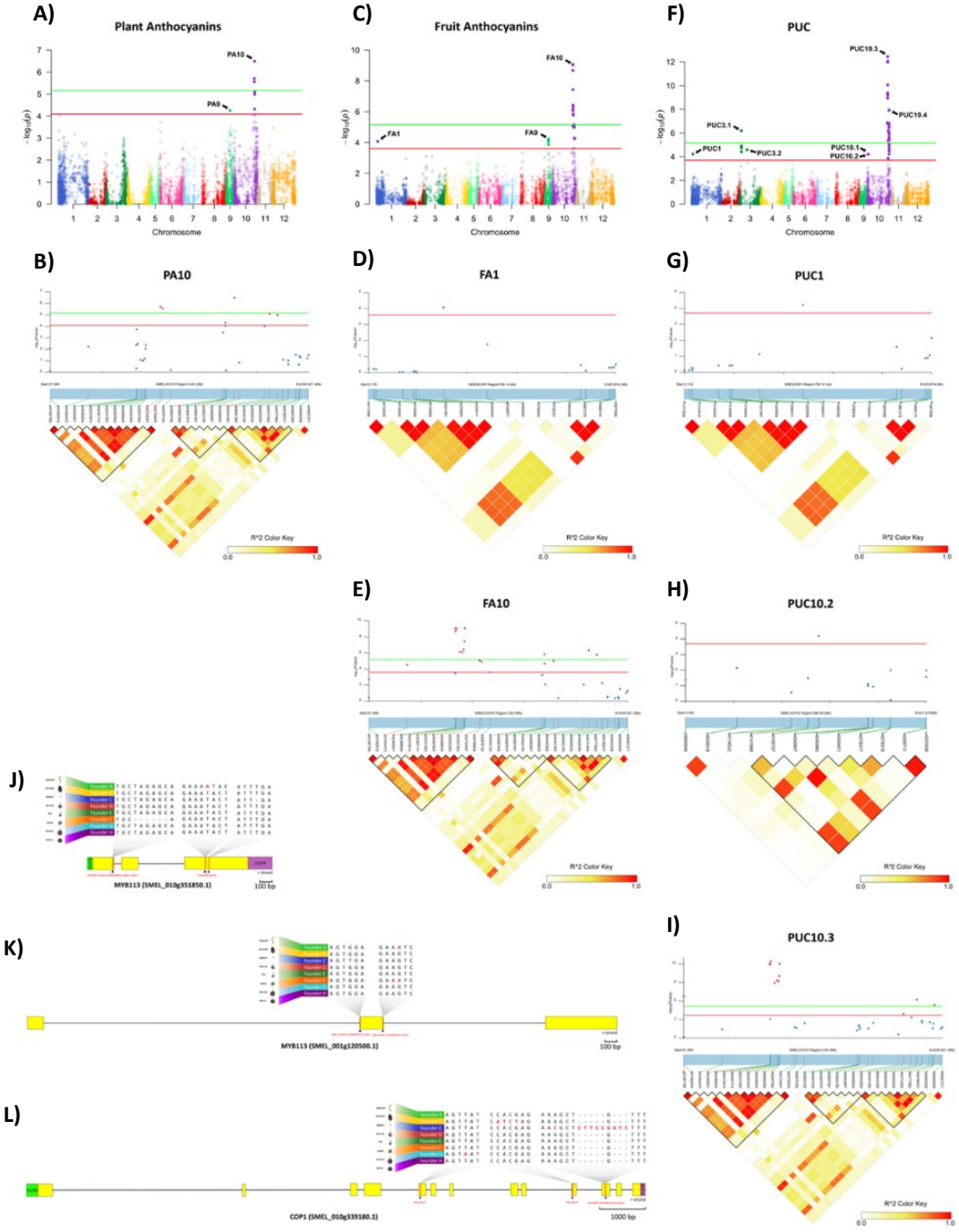


Figure 5. Genome-wide association mapping with linkage disequilibrium (LD) block heatmap of significant regions and potential candidate genes controlling PA, FA, and PUC in eggplant. (A,C,F) Manhattan plot for PA (A), FA (C), and PUC (F). Arrows

indicate the position of main peaks detected for each trait. The red and green horizontal lines represent, respectively, FDR and Bonferroni significance thresholds at p = 0.05. (B,D,E,G–I) Local Manhattan plot (top) and LD heatmap (bottom) surrounding the peaks PA10 (B), FA1 (D), FA10 (E), PUC1 (G), PUC10.2 (H) and PUC10.3 (I). Pairwise LD between SNPs is indicated as values of r^2 values: red indicates a value of 1 and white indicates 0. (J–L) Structure of candidate genes *MYB113* (SMEL_010g351850.1) (J), *MYB113* (SMEL_001g120500.1) (K), and *COP1* (SMEL_010g339180.1) (L) and effect of the high-impact variants detected in those of the eight founders.

3.5. Plant anthocyanins

The Manhattan plot for PA revealed two major peaks on chromosome 9 and 10 (PA9 and PA10), with seven significant SNPs over the FDR threshold (LOD > 4.09), and three of them over the Bonferroni threshold (LOD > 5.16) (Figure 5A). One SNP (LOD = 4.25) was mapped on PA9 region (between 17.0 and 17.4 Mb; Supplementary Figure 3A) and six SNPs, with an LOD between 4.31 and 6.48, were mapped on a PA10 region (between 91.08 and 94.81 Mb; Figure 5A).

3.6. Fruit anthocyanins

For FA, 22 significant SNPs above FDR threshold (LOD > 3.6), which included 11 SNPs above the Bonferroni threshold (LOD > 5.16), were plotted on three major peaks located on chromosomes 1, 9, and 10 (FA1, FA9, and FA10, respectively) (Figure 5C). One SNP (LOD = 4.07) was detected on FA1 region (between 5.11 and 5.88 Mb) in position 5,346,977 (Figure 5D). Five SNPs, with an LOD between 3.86 and 4.22, were located on FA9 region (between 16.2 and 17.4 Mb; Supplementary Figure 3B), which overlapped with the PA9 region (Supplementary Figure 3A). The other sixteen SNPs (LOD between 4.52 and 9.06) were detected on FA10 region between 91.08 and 94.81 Mb (Figure 5E), corresponding to the PA10 region, where significant associations were detected for the PA (Figure 5B).

3.7. Pigmentation under the calyx

Among the three traits evaluated, PUC was the one with the highest number of significant associations. The Manhattan plot for PUC revealed seven main peaks

on chromosomes 1 (PUC1), 3 (PUC3.1 and PUC3.2), and 10 (PUC10.1–10.4), with 36 SNPs over the FDR threshold (LOD > 3.7), and 23 of them are above the Bonferroni threshold (LOD > 5.16) (Figure 5F). One SNP with an LOD of 4.21 was located on the PUC1 region (between 5.11 and 5.88 Mb) in position 5,480,282 (Figure 5G), 133,305 bp away of the significant SNP on the FA1 region (Figure 5D). On the chromosome 3, four SNPs (LOD between 4.41 and 6.18) were detected on the PUC3.1 region (between 7.22 and 8.65 Mb; Supplementary Figure 3C) and one SNP (LOD = 4.58) on the PUC3.2 region (between 31.54 and 40.09 Mb; Supplementary Figure 3D). Thirty SNPs were mapped on chromosome 10 in four genomic regions. One of the SNPs (LOD = 4.2) was identified on the PUC10.1 region (between 2.08 and 2.67 Mb) in position 2,388,375 (Supplementary Figure 3E), and other one (LOD = 4.2) on the PUC10.2 region (between 3.94 and 4.34 Mb; Figure 5H). Another twelve SNPs were located on the PUC10.3 region, between 91.08 and 94.81 Mb (LOD between 3.86 and 12.44; Figure 5I), as observed for PA (Figure 5B) and FA (Figure 5D). The 16 remaining SNPs (LOD between 3.71 and 7.93) were found in the PUC10.4 region, between 98.25 and 100.55 Mb (Supplementary Figure 3F).

3.8. Candidate genes for anthocyanin biosynthesis

Based on the results of the GWAS analysis, putative candidate genes were identified close to or within the LD blocks defined in the genomic regions with significant associations (Supplementary Table 1). On chromosome 10, in the genomic region (91.08–94.81 Mb) associated with all the evaluated traits (PA10, FA10, and PUC10.3), a candidate gene was identified as similar to *MYB113* (SMEL_010g351850.1), a well-known regulatory transcription factor controlling anthocyanin synthesis in eggplant (Zhou *et al.,* 2019; Shi *et al.,* 2021). In addition, in the genomic region associated with FA and PUC (FA1 and PUC1) on chromosome 1 (5.11–5.88 Mb), another candidate gene was identified similar to *MYB113* (SMEL_001g120500.1). Variants that predicted high impact effects on protein function were annotated by SnpEff for both *MYB113* genes in the population founders (A, C, and F) that do not present anthocyanins in plants and fruits, as confirmed by aligning the founder gene sequences (Figures 5J,K). Specifically, for the founders A and C, single frameshift variants were identified in two different positions while a disruptive inframe deletion and a splice region variant were identified for the founder F in a third region of the eggplant gene SMEL_010g351850.1 (Figure 5J). For SMEL_001g120500.1, founders A and F exhibited the exact same splice donor, and the intron variant predicted as high impact, while in the founder C, a splice acceptor and intron variant were identified

(Figure 5K). Founder reconstruction of the S3 individuals and haplotype blocks were estimated for both candidate gene regions (chromosome 10: 91.08–94.81 Mb; chromosome 1: 5.11–5.88 Mb) (Supplementary Figures 4A,B). Founders that contribute most to anthocyanin-related traits in these regions are IVIA-371 (111 and 97 S3 individuals, respectively) and ASIS-1 (74 and 75 S3 individuals, respectively). Furthermore, a reciprocal best hit BLAST analysis of the SMEL_001g120500.1 onto the tomato genome indicated that this gene corresponded to the orthologue *SlANT1-like* with 70.99% identity. The eggplant SMEL_010g351850.1 corresponded to the tomato orthologue *SlAN2-like* with a 73.73% identity, which has been described as the best candidate gene for the anthocyanin fruit biosynthesis in tomato (Yan *et al.,* 2020). These results suggest that a single duplication event and a translocation of a fragment of at least 357 kb from chromosome 10 to 1 occurred during eggplant evolution (Figure 6).

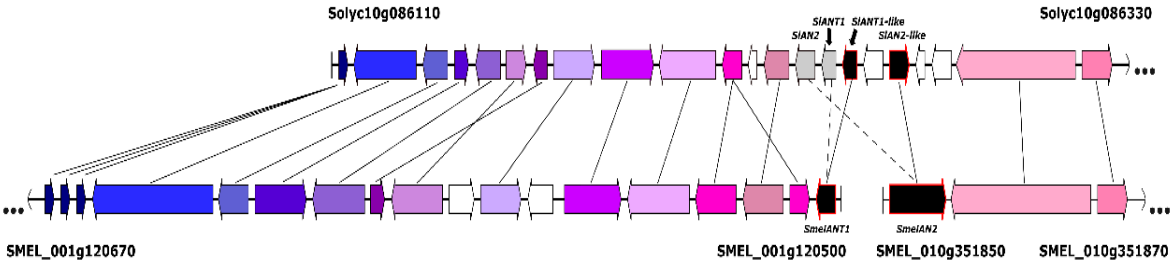


Figure 6. Tomato-eggplant microsynteny representation of a tomato region from chromosome 10 (Solyc10), eggplant region of chromosome 10 (SMEL_010), and a 357 kb fragment of eggplant chromosome 1 (SMEL_001), where candidate genes for anthocyanins synthesis are located.

A candidate gene that corresponds to a *COP1* (SMEL_010g339180.1), a gene that encodes for photo-regulatory proteins (Jiang *et al.,* 2016; Li *et al.,* 2018; He *et al.,* 2019; Naeem *et al.,* 2019), was reported in the genomic regions associated with PUC (PUC10.2) on chromosome 10 (3.94–4.34 Mb; Supplementary Table 1). High-effect variants were detected in COP1 gene in those founders that are able to synthesise anthocyanin under the calyx (B and G) and in the green wild species *S. incanum* (C) (Figure 5L). The C founder mutation was confirmed by the F_1 hybrid phenotype from the C x D inter-cross, *PUC* (anthocyanic pigmentation under the calyx) x *puc* (no anthocyanic pigmentation under the calyx), which showed the anthocyanin fruits with pigmentation under the calyx (Figure 1A). The wild founder presented multiple SNPs compared with the rest that were predicted to cause a frameshift and missense variants, while founders B and G exhibited variants that

were predicted to produce stop codons (Figure 5L). Founders with a higher haplotype representation contributing to PUC in this region (chromosome 10: 3.94–4.34 Mb) are IVIA-371 (113 S3 individuals) and DH_ECAVI (71 S3 individuals) (Supplementary Figure 4C).

Furthermore, we found two candidate genes annotated as similar to *BHLH*, basic helix loop helix protein A (SMEL_009g326640.1), and, similar to *SPA3*, protein SPA1-RELATED 3 (SMEL_010g338090.1), respectively (data not shown), close to LD blocks in the PA9 and FA9 (between 17,862,090 and 17,872,428 Mb), and PUC10.1 (between 3,128,678 and 3,143,068 Mb) regions. Although they were described as genes related to the anthocyanin biosynthetic pathway (Maier *et al.,* 2013; Li *et al.,* 2018; Shi *et al.,* 2021), no high-effects variants were identified in these genes for any of the founders.

4. Discussion

Multi-parent advanced generation inter-cross (MAGIC) populations are outstanding genetic materials for identifying gene-trait associations with high resolution (Arrones *et al.,* 2020; Scott *et al.,* 2020). The introduction of multiple founders with an increased genetic and phenotypic diversity, together with the multiple rounds of inter-crossing and selfing, increases the number of accumulated recombinant events and, thus, improves the mapping accuracy (Scott *et al.,* 2020). By introducing a wild relative as a founder parent, the genetic variability in the population increases, which is a key point for QTL identification (Gramazio *et al.,* 2020a). Here, we present the first eggplant S3MEGGIC population of which one of the founders was one accession of the close wild relative *S. incanum*.

Large population sizes are essential to increase the power and mapping resolution in MAGIC populations (Collard *et al.,* 2005; Valdar *et al.,* 2006; Jaganathan *et al.,* 2020). Following a simple "funnel" scheme design, the population was kept as large as possible to gather a large number of recombination events. However, a sharp reduction in the number of progenies was observed at the S0 generation, which might be related to the use of the wild species *S. incanum* (C) as a female founder and parent to obtain the simple (CD) and the double (ABCD) hybrids. This inter-specific crossing dragged the maternal cytoplasmic background of the wild parent, which might have caused a partial sterility and bias in subsequent generations. Some studies confirmed a strong effect of wild *Solanum* cytoplasm in the reduction in pollen fertility of alloplasmic lines (Khan *et al.,* 2020; Isshiki *et al.,* 2021). However, the PCA highlighted the absence of a population structure, also confirmed by the lack of genetically differentiated cytoplasmic groups.

The genotyping of the S3MEGGIC population was carried out with the 5k probes eggplant SPET platform with a well distributed marker density along all chromosomes (Barchi *et al.,* 2019a). This genotyping strategy has already been used in the analysis of bi-parental populations (Herrero *et al.,* 2020). In this study, we have verified that its use can be extended to multi-parent populations. The genotyping revealed a low heterozygosity in the S3MEGGIC individuals, similar to the expected value for an F5-like bi-parental inter-cross generation (6.25%). The contributions of each of the founder parents to the S3MEGGIC population revealed that some parents had a higher representation than others. Apart from drift effects, several biological reasons could potentially explain cryptic selection processes, causing the unbalanced representation of the genomes (Rockman and Kruglyak, 2008; Thépot *et al.,* 2015), including seed dormancy, delayed germination, precocity, reduced fertility, and parthenocarpy associated to some genomes which have already been reported in eggplant and other crops (Barchi *et al.,* 2010; Khan *et al.,* 2015; Prohens *et al.,* 2017). The rather limited contribution of the wild species *S. incanum* to the final S3 MAGIC individuals may have been caused by a selection pressure, as progenies bred from crosses involving the two different species tend to suffer from a reduced fertility and show a segregation distortion (Lefebvre *et al.,* 2002; Barchi *et al.,* 2010). In addition, *S. incanum* has a recalcitrant germination and a very erratic flowering and fruit set, which strongly depends on environmental conditions (Gisbert *et al.,* 2011; Mangino *et al.,* 2020, 2021). Other reasons for the segregation distortion could be the inability of the current genotyping density to efficiently distinguish between the founders that are genetically closer, like AN-S-26 and H15 genotypes. This phenomenon has already been observed in previous MAGIC populations (Dell'Acqua *et al.,* 2015). A deeper genotyping, resulting in a better haplotype reconstruction, might shed light on the mechanisms that have led to the unbalanced representation of the founder genomes in the S3MEGGIC population. Due to this low contribution, care should be taken in the future when analysing the traits that are present only in these founders.

The anthocyanin biosynthetic pathway is one of the most studied biochemical routes in plants due to its physiological importance. Major structural genes of this pathway are under the control of a regulatory complex, where myeloblastosis (MYB) TFs are recognised as main regulators alone or in complexes with other TFs (Ramsay and Glover, 2005; Kiferle *et al.,* 2015; Liu *et al.,* 2018). Activator or repressor MYB proteins directly and competitively bind the basic-helix-loop-helix (bHLH) via the amino terminus domain and can act as positive or negative transcriptional regulators in a tissue-specific mode to modulate the anthocyanin synthesis (Barchi *et al.,* 2019b; Moglia *et al.,* 2020).

In eggplant, anthocyanin-related MYB protein-encoding genes have been reported to be related to fruit peel colouration (Zhang *et al.,* 2014; Docimo *et al.,*

2016; Xiao *et al.,* 2018; Moglia *et al.,* 2020; Toppino *et al.,* 2020). In tomato, a cluster of four different MYB proteins was reported to be involved in the anthocyanin synthesis located on chromosome 10 and encoded by *SlAN2*, *SlANT1*, *SlANT1-like*, and *SlAN2-like* genes (Solyc10g086250, Solyc10g086260, Solyc10g086270, and Solyc10g086290, respectively). However, genetic associations to only two paralogue *MYB113* genes were detected in the S3MEGGIC population on chromosomes 1 and 10 (SMEL_001g120500.1 and SMEL_010g351850.1). The same occurs in potato, in which, only two anthocyanin genes have been identified, i.e., Sotub10g028550 and Sotub10g028540, both located on chromosome 10. Similarly, in pepper, in addition to CA10g11690 and CA10g11650, a third gene (CA10g11710) has been identified as an orthologue to *SlAN2-like* (Barchi *et al.,* 2019b). In eggplant, these two tomato orthologs were previously described as *SmelANT1* and *SmelAN2* (Docimo *et al.,* 2016; Barchi *et al.,* 2019b), corresponding to *SlANT1* and *SlAN2*. However, the reciprocal best hit BLAST analysis of the eggplant coding proteins showed a stronger homology, respectively, to tomato *SlANT1-like* and *SlAN2-like*. Although previous studies in eggplant indicated that the overexpression of *SmelANT1* accounts for constitutive upregulation of most anthocyanin biosynthetic genes (Zhang *et al.,* 2014; Shi *et al.,* 2021), we considered the *SmelAN2* as the best candidate gene for different reasons. Previous studies in eggplant have located the major anthocyanin-related QTLs on chromosome 10 (Barchi *et al.,* 2012; Cericola *et al.,* 2014; Toppino *et al.,* 2016), which are in agreement with the highest association signals for anthocyanin-related traits in our GWAS results. Furthermore, recent studies in tomato suggested that *SlAN2-like* functions as an activator to regulate biosynthesis genes, including *SlANT1-like*, and controls the accumulation of anthocyanins (Yan *et al.,* 2020). However, in the S3MEGGIC population, high-impact variants on protein function were found in both *MYB113* genes for non-anthocyanin fruits. These results could suggest that a duplication of function occurred during an eggplant evolution, and both genes may be required for anthocyanin synthesis.

Although it has been demonstrated that the activation of the anthocyanin biosynthetic pathway in eggplant is strongly regulated by light (Li *et al.,* 2018; Xiao *et al.,* 2018), the PUC mutation confers to some genotypes the ability to synthesise anthocyanins under the calyx regardless of incidence of light (Tigchelaar *et al.,* 1968). In this study, a candidate gene, related with light-dependent anthocyanin biosynthesis in fruits, was detected at the beginning of chromosome 10. The constitutive photomorphogenic1 (*COP1*) gene has been reported to be a regulatory TF responsible for mediating a light-regulated gene expression and development (Jiang *et al.,* 2016; Li *et al.,* 2018; He *et al.,* 2019; Naeem *et al.,* 2019). The *COP1* gene has been demonstrated to act as a light-inactivable repressor interacting with MYB TFs (Jiang *et al.,* 2016), and is considered a “molecular switch” in metabolic processes, which are stimulated by light. In presence of light, *COP1* expression is

inhibited, and its concentration decreases rapidly, promoting anthocyanins synthesis. Under dark conditions, *COP1* expression is induced and promotes the degradation of the photomorphogenesis-promoting TF and MYB inhibition by conforming the *COP1*/Suppressor of phya-105 (*SPA*) ubiquitin ligase complex (Maier *et al.,* 2013; Jiang *et al.,* 2016; Li *et al.,* 2018). Although PUC can only be observed in anthocyanic fruits, high-impact variants were also identified in the wild species *S. incanum COP1* gene (founder C, green fruit). The *PUC* allele is more common in eggplants from western countries, where European markets demand for more homogeneously pigmentation eggplants. On the other hand, in Asian eggplants, it is more common to find the *puc* phenotype, such as in the ASI-S-1 founder.

5. Conclusion

In conclusion, the S3MEGGIC population represents a landmark breeding material and a tool of great value, which allows the study and fine-mapping of complex traits due to (i) the highly phenotypically diverse founders; (ii) the large population size, being the largest developed eggplant experimental population so far; (iii) the high degree of homozygosity of the final individuals, which constitute a population of fixed "immortal" lines nearly homozygous at each locus; and (iv) the tailored genotyping SPET platform used for the genetic analysis of the population, which has been developed from the whole genome sequencing (WGS) of the founders and allows the comparison with the genotyped materials with the same setup (Gramazio *et al.,* 2020b). In addition, the S3MEGGIC population has demonstrated its potential usefulness for association studies, allowing the establishment of marker associations to anthocyanin-related genes and identification of candidate genes for plant and fruit anthocyanins, including the first identification of a candidate gene for an economically relevant breeding trait in eggplant such as PUC.

Data availability statement: The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/, PRJNA392603.

Author contributions: S.V., P.G., and J.P. conceived the idea and supervised the manuscript. G.M., A.A., M.P., and P.G. performed the field trials. G.M. and A.A. prepared a first draft of the manuscript. All other authors reviewed and edited the manuscript.

Funding: This work was supported by the Ministerio de Ciencia, Innovación y Universidades, Agencia Estatal de Investigación and Fondo Europeo de Desarrollo

Regional (grant RTI2018-094592-B-I00 from MCIU/AEI/FEDER, UE) and European Union's Horizon, 2020 Research and Innovation Programme under grant agreement no. 677379 (G2P-SOL project: Linking genetic resources, genomes and phenotypes of Solanaceous crops). A.A. is grateful to Spanish Ministerio de Ciencia, Innovación y Universidades for a pre-doctoral (FPU18/01742) contract. M.P. is grateful to Spanish Ministerio de Ciencia e Innovación for a post-doctoral grant (IJC2019-039091-I/AEI/10.13039/501100011033). P.G. is grateful to Spanish Ministerio de Ciencia e Innovación for a post-doctoral grant (FJC2019-038921-I/AEI/10.13039/501100011033). Funding for open access charge: Universitat Politècnica de València.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Supplementary material: The supplementary material for this article can be found online at:

https://www.frontiersin.org/articles/10.3389/fpls.2022.847789/full#supplementary-material

Supplementary Figure 1. (A) Chain pollination scheme of the four-way hybrids followed to obtain the eight-way hybrids. (B) For each S0 progeny, four plants were germinated, selecting for the next generation (S1) only the first two that set fruits with viable seed. For subsequent generations, two plants were germinated and only the first one that set fruit was selected for the next generation.

Supplementary Figure 2. Phenotyping of the S3MEGGIC population for presence or absence of PA, FA, and PUC.

Supplementary Figure 3. Local Manhattan plot (top) and LD heatmap (bottom) surrounding the peaks PA9 (A), FA9 (B), PUC3.1 (C), PUC3.2 (D), PUC10.1 (E), and PUC10.4 (F). The red and green horizontal lines represent, respectively, FDR and Bonferroni significance thresholds. Pairwise LD between SNPs is indicated as values of r^2 values: red indicates a value of 1 and white indicates 0.

Supplementary Figure 4. Founder haplotype blocks representation predicted for each of the S3 individuals for the three anthocyanin-related candidate gene regions: (A) *MYB113* (SMEL_001g120500.1) on chromosome 1 between 5.11 and 5.88 Mb identified by the FA1 and PUC1 associations; (B) *COP1* (SMEL_010g339180.1.01) on chromosome 10 between 3.94 and 4.34 Mb identified by the PUC10.2 association; and (C) *MYB113* (SMEL_010g351850.1) on chromosome 10 between 91.08 and 94.81 Mb identified by the PA10, FA10, and PUC10.3 associations.

Supplementary Table 1. List of candidate genes for plant anthocyanins (PA), fruit anthocyanins (FA) and anthocyanin pigmentation under the calyx (PUC).

References:

Albert, N. W., Davies, K. M., Lewis, D. H., Zhang, H., Montefiori, M., Brendolise, C., *et al*. (2014). A conserved network of transcriptional activators and repressors regulates anthocyanin pigmentation in Eudicots. *Plant Cell*, *26,* 962-980. doi: 10.1105/tpc.113.122069

Arrones, A., Vilanova, S., Plazas, M., Mangino, G., Pascual, L., Díez, M. J., *et al*. (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. doi: 10.3390/biology9080229

Bandillo, N., Raghavan, C., Muyco, P. A., Sevilla, M. A. L., Lobina, I. T., Dilla-Ermita, C. J., *et al*. (2013). Multi-parent advanced generation inter-cross (MAGIC) populations in rice: Progress and potential for genetics research and breeding. *Rice*, *6*, 1-15. doi: 10.1186/1939-8433-6-11

Barchi, L., Acquadro, A., Alonso, D., Aprea, G., Bassolino, L., Demurtas, O., *et al.* (2019a). Single Primer Enrichment Technology (SPET) for high-throughput genotyping in tomato and eggplant germplasm. *Front. Plant Sci.*, *10*, 1005. doi: 10.3389/fpls.2019.01005

Barchi, L., Lanteri, S., Portis, E., Stàgel, A., Valè, G., Toppino, L., *et al*. (2010). Segregation distortion and linkage analysis in eggplant (*Solanum melongena* L.). *Genome*, *53*, 805-815. doi: 10.1139/g10-073

Barchi, L., Lanteri, S., Portis, E., Valè, G., Volante, A., Pulcini, L., *et al*. (2012). A RAD tag derived marker based eggplant linkage map and the location of QTLs determining anthocyanin pigmentation. *PLoS One*, *7*, e43740. doi: 10.1371/journal.pone.0043740

Barchi, L., Pietrella, M., Venturini, L., Minio, A., Toppino, L., Acquadro, A., *et al*. (2019b). A chromosome-anchored eggplant genome sequence reveals key events in Solanaceae evolution. *Sci. Rep.*, *9*, 1-13. doi: 10.1038/s41598-019-47985-w

Barrett, J. C., Fry, B., Maller, J., and Daly, M. J. (2005). Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics*, *21*, 263-265. doi: 10.1093/bioinformatics/bth457

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. (Series B Methodological), 57*, 289-300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Borovsky, Y., Oren-Shamir, M., Ovadia, R., De Jong, W., and Paran, I. (2004). The *A* locus that controls anthocyanin accumulation in pepper encodes a *MYB* transcription factor homologous to *Anthocyanin2* of *Petunia*. *Theor. Appl. Genet.*, *109*, 23-29. doi: 10.1007/s00122-004-1625-9

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. doi: 10.1093/bioinformatics/btm308

Campanelli, G., Sestili, S., Acciarri, N., Montemurro, F., Palma, D., Leteo, F., and Beretta, M. (2019). Multi-parental advances generation inter-cross population, to develop organic tomato genotypes by participatory plant breeding. *Agronomy*, *9*, 119. doi: 10.3390/agronomy9030119

Cericola, F., Portis, E., Lanteri, S., Toppino, L., Barchi, L., Acciarri, N., *et al*. (2014). Linkage disequilibrium and genome-wide association analysis for anthocyanin pigmentation and fruit color in eggplant. *BMC Genomics*, *15*, 1-15. doi: 10.1186/1471-2164-15-896

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., *et al*. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*, *6*, 80-92. doi: 10.4161/fly.19695

Collard, B. C. Y., Jahufer, M. Z. Z., Brouwer, J. B., and Pang, E. C. K. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and markerassisted selection for crop improvement: the basic concepts. *Euphytica, 142*, 169–196. doi: 10.1007/s10681-005-1681-5

D'Amelia, V., Aversano, R., Chiaiese, P., and Carputo, D. (2018). The antioxidant properties of plant flavonoids: their exploitation by molecular plant breeding. *Phytochem. Rev.*, *17*, 611-625. doi: 10.1007/s11101-018-9568-y

Daunay, M. C., and Hazra, P. (2012). "Eggplant" in *Handbook of Vegetables* (Peter, K.V., and Hazra, P., eds), pp 257-322. Houston, TX: Studium Press.

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. doi: 10.1371/journal.pone.0085024

Dell'Acqua, M., Gatti, D.M., Pea, G., Cattonaro, F., Coppens, F., Magris, G., *et al*. (2015). Genetic properties of the MAGIC maize population: A new platform for high-definition QTL mapping in *Zea mays*. *Genome Biol.*, *16*, 1-23. doi: 10.1186/s13059-015-0716-z

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 nextgeneration DNA sequencing data. *Nat. Genet*., *43*, 491–498. doi: 10.1038/ng.806

Díez, M. J., Picó, B., and Nuez, F. (2002). Cucurbit Genetic Resources in Europe: *Ad Hoc Meeting held in Adana, Turkey, 19 January 2002.* Rome: International Plant Genetic Resources Institute. doi: 10.5555/20033188287

Docimo, T., Francese, G., Ruggiero, A., Batelli, G., De Palma, M., Bassolino, L., *et al*. (2016). Phenylpropanoids accumulation in eggplant fruit: Characterization of biosynthetic genes and regulation by a MYB transcription factor. *Front. Plant Sci*., 6, 1233. doi: 10.3389/fpls.2015.01233

Doganlar, S., Frary, A., Daunay, M. C., Lester, R. N., and Tanksley, S. D. (2002). Conservation of gene function in the Solanaceae as revealed by comparative mapping of domestication traits in eggplant. *Genetics*, *161,* 1713-1726. doi:10.1093/genetics/161.4.1713

Dong, S. S., He, W. M., Ji, J. J., Zhang, C., Guo, Y., Yang, T. L. (2020). LDBlockShow: a fast and convenient tool for visualizing linkage disequilibrium and haplotype blocks based on variant call format files. *Briefings in Bioinformatics,* bbaa227. doi: 10.1093/bib/bbaa227

FAOSTAT. Available online at: http://www.faostat.fao.org (Accessed February 20, 2020).

Frary, A., Frary, A., Daunay, M. C., Huvenaars, K., Mank, R., and Doğanlar, S. (2014). QTL hotspots in eggplant (*Solanum melongena*) detected with a high-resolution map and CIM analysis. *Euphytica, 197*, 211-228. doi: 10.1007/s10681-013-1060-6

Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B., *et al*. (2002). The structure of haplotype blocks in the human genome. *Science*, *296*, 2225-2229. doi: 10.1126/science.1069424

Gisbert, C., Prohens, J., and Nuez, F. (2011). Treatments for improving seed germination in eggplant and related species. *Acta Horticult., 898,* 45–52. doi: 10.17660/ActaHortic.2011.898.4

Gonzali, S., Mazzucato, A., and Perata, P. (2009). Purple as a tomato: towards high anthocyanin tomatoes. *Trends Plant Sci.*, *14*, 237-241. doi: 10.1016/j.tplants.2009.02.001

Gramazio, P., Pereira-Dias, L., Vilanova, S., Prohens, J., Soler, S., Esteras, J., *et al*. (2020a). 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 inter-specific tomato MAGIC population. *Hortic. Res.*, *7*, 174. doi: 10.1038/s41438-020-00395-w

Gramazio, P., Jaén-Molina, R., Vilanova, S., Prohens, J., Marrero, Á., Caujapé-Castells, J., *et al*. (2020b). 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). *Front. Plant Sci.*, *11*, 757. doi: 10.3389/fpls.2020.00757

Gramazio, P., Prohens, J., Borràs, D., Plazas, M., Herraiz, F. J., and Vilanova, S. (2017a). 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. doi: 10.1007/s10681-017-2057-3

Gramazio, P., Prohens, J., Plazas, M., Mangino, G., Herraiz, F. J., and Vilanova, S. (2017b). 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,* 1477. doi: 10.3389/fpls.2017.01477

Gramazio, P., Prohens, J., Plazas, M., Mangino, G., Herraiz, F. J., García-Fortea, E., *et al*. (2018). Genomic tools for the enhancement of vegetable crops: A case in eggplant. *Not. Bot. Horti Agrobot. Cluj-Napoca*, *46*, 1-13. doi: 10.15835/nbha46110936

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. doi: 10.3389/fpls.2019.01220

He, Y., Chen, H., Zhou, L., Liu, Y., and Chen, H. (2019). Comparative transcription analysis of photosensitive and non-photosensitive eggplants to identify genes involved in dark regulated anthocyanin synthesis. *BMC Genomics*, *20*, 1-14. doi:10.1186/s12864-019-6023-4

Herrero, J., Santika, B., Herrán, A., Erika, P., Sarimana, U., Wendra, F., *et al.* (2020). Construction of a high density linkage map in Oil Palm using SPET markers. *Sci. Rep., 10*, 1–9. doi: 10.1038/s41598-020-67118-y

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*, 65-70.

Huang, B. E., Verbyla, K. L., Verbyla, A. P., Raghavan, C., Singh, V. K., Gaur, P., *et al*. (2015). MAGIC populations in crops: current status and future prospects. *Theor. Appl. Genet.*, *128,* 999-1017. doi: 10.1007/s00122-015-2506-0

Hurtado, M., Vilanova, S., Plazas, M., Gramazio, P., Andújar, I., Herraiz, F. J., *et al.* (2014). Enhancing conservation and use of local vegetable landraces: the Almagro eggplant (*Solanum melongena* L.) case study. *Genet. Resourc. Crop Evol., 61*, 787–795. doi: 10.1007/s10722-013-0073-2

Huynh, B. L., Ehlers, J. D., Huang, B. E., Muñoz-Amatriaín, M., Lonardi, S., Santos, J. R. P., *et al*. (2018). A multi-parent advanced generation inter-cross (MAGIC) population for genetic analysis and improvement of cowpea (*Vigna unguiculata* L. Walp.). *Plant J., 93,* 1129-1142. doi: 10.1111/tpj.13827

Isshiki, S., Nakamura, I., Ureshino, K., and Khan, M. M. R. (2021). Pollen fertility differences in the progenies obtained from a cross between eggplant (*Solanum melongena* L.) as a seed parent and eggplant cytoplasmic substitution lines as pollen parents. *Austral. J. Crop Sci., 15,* 233–237. doi: 10.21475/ajcs.21.15.02.p2785

Jaganathan, D., Bohra, A., Thudi, M., and Varshney, R. K. (2020). Fine mapping and gene cloning in the post-NGS era: advances and prospects. *Theor. Appl. Genet., 133*, 1791–1810. doi: 10.1007/s00122-020-03560-w

Jiang, M., Ren, L., Lian, H., Liu, Y., and Chen, H. (2016). Novel insight into the mechanism underlying light-controlled anthocyanin accumulation in eggplant (*Solanum melongena* L.). *Plant Sci.*, *249*, 46-58. doi: 10.1016/j.plantsci.2016.04.001

Jombart, T. (2008). Adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics, 24*, 1403–1405. doi: 10.1093/bioinformatics/btn129

Kamvar, Z. N., Tabima, J. F., and Gr¨unwald, N. J. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, 1–14. doi: 10.7717/peerj.281

Kaushik, P., Plazas, M., Prohens, J., Vilanova, S., and Gramazio, P. (2018). Diallel genetic analysis for multiple traits in eggplant and assessment of genetic distances for predicting hybrids performance. *PLoS One, 13,* e0199943. doi: 10.1371/journal.pone.0199943

Khan, M. M. R., Arita, T., Iwayoshi, M., Ogura-Tsujita, Y., and Isshiki, S. (2020). Development of the functional male sterile line of eggplant utilizing the cytoplasm of *Solanum* kurzii by way of the amphidiploid. *Environ. Control Biol., 58,* 79–83. doi: 10.2525/ecb.58.79

Khan, M. M. R., Hasnunnahar, M., Iwayoshi, M., Ogura-Tsujita, Y., and Isshiki, S. (2015). Pollen degeneration in three functional male-sterile lines of eggplant with the wild *Solanum* cytoplasms. *Horticult. Environ. Biotechnol*., *56*, 350–357. doi: 10.1007/s13580-015-0015-3

Kiferle, C., Fantini, E., Bassolino, L., Povero, G., Spelt, C., Buti, S., *et al*. (2015). Tomato R2R3-MYB proteins *SlANT1* and *SlAN2*: Same protein activity, different roles. *PLoS One*, *10*, e0136365. doi: 10.1371/journal.pone.0136365

Knapp, S., Vorontsova, M. S., and 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. doi: 10.1371/journal.pone.0057039

Knaus, B. J., and Grünwald, N. J. (2017). VCFR: a package to manipulate and visualize variant call format data in R. *Mol. Ecol. Resourc., 17*, 44–53. doi: 10.1111/1755-0998.12549

Kover, P. X., Valdar, W., Trakalo, J., Scarcelli, N., Ehrenreich, I. M., Purugganan, M. D., *et al*. (2009). A multi-parent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. *PLoS Genet.*, *5*, e1000551. doi: 10.1371/journal.pgen.1000551

Lebeau, A., Gouy, M., Daunay, M. C., Wicker, E., Chiroleu, F., Prior, P., *et al*. (2013). Genetic mapping of a major dominant gene for resistance to *Ralstonia solanacearum* in eggplant. *Theor. Appl. Genet*., *126,* 143-158. doi: 10.1007/s00122-012-1969-5

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

Li, J., He, Y. J., Zhou, L., Liu, Y., Jiang, M., Ren, L., and Chen, H. (2018). Transcriptome profiling of genes related to light-induced anthocyanin biosynthesis in eggplant (*Solanum melongena* L.) before purple color becomes evident. *BMC Genomics*, *19*, 1-12. doi: 10.1186/s12864-018-4587-z

Li, J., Ren, L., Gao, Z., Jiang, M., Liu, Y., Zhou, L., *et al*. (2017). Combined transcriptomic and proteomic analysis constructs a new model for light-induced anthocyanin biosynthesis in eggplant (*Solanum melongena* L.). *Plant Cell Environ.*, *40*, 3069-3087. doi: 10.1111/pce.13074

Li, L., He, Y., Ge, H., Liu, Y., and Chen, H. (2021). Functional characterization of *SmMYB86*, a negative regulator of anthocyanin biosynthesis in eggplant (*Solanum melongena* L.). *Plant Sci.*, *302,* 110696. doi: 10.1016/j.plantsci.2020.110696

Liu, Y., Tikunov, Y., Schouten, R. E., Marcelis, L. F. M., Visser, R. G. F., and Bovy, A. (2018). Anthocyanin biosynthesis and degradation mechanisms in Solanaceous vegetables: A review. *Front. Chem.*, *6*, 52. doi: 10.3389/fchem.2018.00052

Lv, L. L., Feng, X. F., Li, W., Li, K. (2019). High temperature reduces peel color in eggplant (*Solanum melongena*) as revealed by RNA-seq analysis. *Genome*, *62*, 503-512. doi: 10.1139/gen-2019-0021

Maier, A., Schrader, A., Kokkelink, L., Falke, C., Welter, B., Iniesto, E., *et al*. (2013). Light and the E3 ubiquitin ligase *COP1/SPA* control the protein stability of the MYB transcription factors *PAP1* and *PAP2* involved in anthocyanin accumulation in *Arabidopsis*. *Plant J.*, *74*, 638-651. doi: 10.1111/tpj.12153

Mangino, G., Plazas, M., Vilanova, S., Prohens, J., and 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*, 467. doi: 10.3390/agronomy10040467

Mangino, G., Vilanova, S., Plazas, M., Prohens, J., and Gramazio, P. (2021). Fruit shape morphometric analysis and QTL detection in a set of eggplant introgression lines. *Sci. Horticult., 282*, 110006. doi: 10.1016/j.scienta.2021.110006

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J., 17*, 10–12. doi: 10.1089/cmb.2017.0096

Mennella, G., Lo Scalzo, R., Fibiani, M., D'Alessandro, A., Francese, G., Toppino, L., *et al*. (2012). Chemical and bioactive quality traits during fruit ripening in eggplant (*S. melongena* L.) and allied species. *J. Agric. Food Chem.*, *60*, 11821-11831. doi: 10.1021/jf3037424

Moglia, A., Francesco, E. F., Sergio, I., Alessandra, G., Milani, A. M., Cinzia, C., *et al*. (2020). Identification of a new R3 MYB type repressor and functional characterization of the members of the MBW transcriptional complex involved in anthocyanin biosynthesis in eggplant (*S. melongena* L.). *PLoS One*, *15*, e0232986. doi: 10.1371/journal.pone.0232986

Naeem, M., Muqarab, R., and Waseem, M. (2019). The *Solanum melongena COP1* delays fruit ripening and influences ethylene signaling in tomato. *J. Plant Physiol.*, *240*, 152997. doi: 10.1016/j.jplph.2019.152997

Pascual, L., Desplat, N., Huang, B. E., Desgroux, A., Bruguier, L., Bouchet, J. P., *et al*. (2015). Potential of a tomato MAGIC population to decipher the genetic control of quantitative traits and detect causal variants in the resequencing era. *Plant Biotechnol. J., 13*, 565-577. doi: 10.1111/pbi.12282

Pook, T., Schlather, M., De Los Campos, G., Mayer, M., Carolin Schoen, C., and Simianer, H. (2019). HaploBlocker: creation of subgroup-specific haplotype blocks and libraries. *Genetics, 212,* 1045–1061. doi: 10.1534/genetics.119.302283

Prohens, J., Gramazio, P., Plazas, M., Dempewolf, H., Kilian, B., Díez, M. J., *et al.* (2017). Introgressiomics: a new approach for using crop wild relatives in breeding for adaptation to climate change. *Euphytica, 213*, 158. doi: 10.1007/s10681-017-1938-9

Prohens, J., Whitaker, B. D., Plazas, M., Vilanova, S., Hurtado, M., Blasco, M., *et al.* (2013). Genetic diversity in morphological characters and phenolic acids content resulting from an inter-specific cross between eggplant, *Solanum melongena*, and its wild ancestor (*S. incanum*). *Ann. Appl. Biol., 162*, 242–257. doi: 10.1111/aab.12017

Ramsay, N. A. and Glover, B. J. (2005). MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. *Trends Plant Sci.*, *10*, 63-70. doi: 10.1016/j.tplants.2004.12.011

Ranil, R. H. G., Niran, H. M. L., Plazas, M., Fonseka, R. M., Fonseka, H. H., Vilanova, S., *et al.* (2015). Improving seed germination of the eggplant rootstock *Solanum torvum* by testing multiple factors using an orthogonal array design. *Sci. Horticult., 193*, 174–181. doi: 10.1016/j.scienta.2015.07.030

Robinson, J. T., Thorvaldsdóttir, H., Turner, D., and Mesirov, J. P. (2020). igv.js: An embeddable JavaScript implementation of the Integrative Genomics Viewer (IGV). *bioRxiv*, 075499. doi: 10.1093/bioinformatics/btac830

Rockman, M. V., and Kruglyak, L. (2008). Breeding designs for recombinant inbred advanced intercross lines. *Genetics*, *179,* 1069–1078. doi: 10.1534/genetics.107.083873

Scott, M. F., Ladejobi, O., Amer, S., Bentley, A. R., Biernaskie, J., Boden, S. A., *et al*. (2020). Multi-parent populations in crops: a toolbox integrating genomics and genetic mapping with breeding. *Heredity*, *125*, 396-416. doi: 10.1038/s41437-020-0336-6

Shi, S., Liu, Y., He, Y., Li, L., Li, D., and Chen, H. (2021). R2R3-MYB transcription factor *SmMYB75* promotes anthocyanin biosynthesis in eggplant (*Solanum melongena* L.). *Sci. Hortic.*, *282*, 110020. doi: 10.1016/j.scienta.2021.110020

Stadlmeier, M., Hartl, L., and Mohler, V. (2018). Usefulness of a multi-parent advanced generation intercross population with a greatly reduced mating design for genetic studies in winter wheat. *Front. Plant Sci.*, *9,* 1825. doi: 10.3389/fpls.2018.01825

Thépot, S., Restoux, G., Goldringer, I., Hospital, F., Gouache, D., Mackay, I., *et al.* (2015). Efficiently tracking selection in a multi-parental population: the case of earliness in wheat. *Genetics, 199,* 609–623. doi: 10.1534/genetics.114.169995

Thissen, D., Steinberg, L., Kuang, D. (2002). Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. *J. Educ. Behav. Stat.*, *27*, 77-83. doi: 10.3102/10769986027001077

Tigchelaar, E. C., Janick, J., and Erickson, H. T. (1968). The genetics of anthocyanin coloration in eggplant (*Solanum melongena* L.). *Genetics*, *60,* 475-491. doi:10.1093/genetics/60.3.475

Toppino, L., Barchi, L., Lo Scalzo, R., Palazzolo, E., Francese, G., Fibiani, M., *et al*. (2016). Mapping quantitative trait loci affecting biochemical and morphological fruit properties in eggplant (*Solanum melongena* L.). *Front. Plant Sci.*, *7*, 256. doi: 10.3389/fpls.2016.00256

Toppino, L., Barchi, L., Mercati, F., Acciarri, N., Perrone, D., Martina, M., *et al*. (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. doi: 10.3390/genes11070745

Turner, S. D. (2018). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *J. Open Source Softw.*, *3*, 731. doi: 10.21105/joss.00731

Valdar, W., Flint, J., and Mott, R. (2006). Simulating the collaborative cross: power of quantitative trait loci detection and mapping resolution in large sets of recombinant inbred strains of mice. *Genetics, 172,* 1783–1797. doi: 10.1534/genetics.104.039313

Van Eck, H. J., Jacobs, J. M. E., Van Den Berg, P. M. M. M., Stiekema, W. J., and Jacobsen, E. (1994) The inheritance of anthocyanin pigmentation in potato (*Solatium tuberosum* L.) and mapping of tuber skin colour loci using RFLPs. *Heredity*, *73*, 410-421. doi: 10.1038/hdy.1994.189

Vilanova, S., Alonso, D., Gramazio, P., Plazas, M., García-Fortea, E., Ferrante, P., *et al.* (2020). SILEX: a fast and inexpensive high-qualityDNA extraction method suitable for multiple sequencing platforms and recalcitrant plant species. *Plant Methods, 16*, 110. doi: 10.1186/s13007-020-00652-y

Wang, H., Guo, X., Pandey, M. K., Ji, X., Varshney, R. K., Nwosu, V*., et al.* (2017). "History and impact of the international peanut genome initiative: the exciting journey toward peanut whole-genome sequencing," in *The Peanut Genome*, eds R. Varshney, M. Pandey, and N. Puppala (New York, NY: Springer.), 117–134. doi: 10.1007/978-3-319-63935-2_8

Wei, Q., Wang, W., Hu, T., Hu, H., Wang, J., and Bao, C. (2020). Construction of a SNP-based genetic map using SLAF-Seq and QTL analysis of morphological traits in eggplant. *Front. Genet., 11,* 178. doi: 10.3389/fgene.2020.00178

Wickham, H. (2016). ggplot2 Elegant Graphics for Data Analysis (Use R!). New York, NY: Springer.

Xiao, X. O., Li, K., Feng, X. F., and Jin, H. (2018). Transcriptome analyses reveal anthocyanins biosynthesis in eggplant. *PeerJ Prepr*., *6*, e27289v1. doi: 10.1186/s12870-019-1960-2

Yan, S., Chen, N., Huang, Z., Li, D., Zhi, J., Yu, B., *et al*. (2020). Anthocyanin Fruit encodes an R2R3-MYB transcription factor, *SlAN2-like*, activating the transcription of *SlMYBATV* to fine-tune anthocyanin content in tomato fruit. *New Phytol.*, *225*, 2048-2063. doi: 10.1111/nph.16272

Zhang, Y., Hu, Z., Chu, G., Huang, C., Tian, S., Zhao, Z., and Chen, G. (2014). Anthocyanin accumulation and molecular analysis of anthocyanin biosynthesis-associated genes in eggplant (*Solanum melongena* L.). *J. Agric. Food Chem.*, *62*, 2906-2912. doi: 10.1021/jf404574c

Zhou, L., He, Y., Li, J., Liu, Y., and Chen, H. (2019). CBFs function in anthocyanin biosynthesis by interacting with *MYB113* in eggplant (*Solanum melongena* L.). *Plant Cell Physiol.*, *61,* 416-426. doi: 10.1093/pcp/pcz209

Research article

Mutations in the *SmAPRR2* transcription factor suppressing chlorophyll pigmentation in the eggplant fruit peel are key drivers of a diversified colour palette

Andrea Arrones[1], Giulio Mangino[1], David Alonso[1], Mariola Plazas[1], Jaime Prohens[1], Ezio Portis[2], Lorenzo Barchi[2], Giovanni Giuliano[3], Santiago Vilanova[1]*, Pietro Gramazio[4]*

[1]Instituto de Conservación y Mejora de la Agrodiversidad Valenciana, Universitat Politècnica de València, Camino de Vera 14, 46022 Valencia, Spain

[2]DISAFA, Plant Genetics and Breeding, University of Turin, Largo P. Braccini 2, 10095 Grugliasco, Italia

[3]Agenzia Nazionale Per Le Nuove Tecnologie, L'energia e Lo Sviluppo Economico Sostenibile (ENEA), Casaccia Research Centre, Via Anguillarese 301, 00123 Rome, Italia

[4]Instituto de Biología Molecular y Celular de Plantas, Consejo Superior de Investigaciones Científicas-Universitat Politècnica de València, Camino de Vera 14, 46022 Valencia, Spain

*Corresponding authors

Ph.D. candidate contribution

A.A. had a main role in the following activities: performing the experiments, data collection, data analysis, data visualization, drafting manuscript, manuscript review and editing.

Citation: 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,* 1025951. doi: 10.3389/fpls.2022.1025951

Abstract

Understanding the mechanisms by which chlorophylls are synthesized in the eggplant (*Solanum melongena*) fruit peel is of great relevance for eggplant breeding. A multi-parent advanced generation inter-cross (MAGIC) population and a germplasm collection have been screened for green pigmentation in the fruit peel and used to identify candidate genes for this trait. A genome-wide association study (GWAS) performed with 420 MAGIC individuals revealed a major association on chromosome 8 close to a gene similar to *APRR2*. Two variants in *SmAPRR2*, predicted as having a high impact effect, were associated with the absence of fruit chlorophyll pigmentation in the MAGIC population, and a large deletion of 5.27 kb was found in two reference genomes of accessions without chlorophyll in the fruit peel. The validation of the candidate gene *SmAPRR2* was performed by its sequencing in a set of MAGIC individuals and through its de novo assembly in 277 accessions from the G2P-SOL eggplant core collection. Two additional mutations in *SmAPRR2* associated with the lack of chlorophyll were identified in the core collection set. The phylogenetic analysis of *APRR2* reveals orthology within Solanaceae and suggests that specialization of *APRR2-like* genes occurred independently in Cucurbitaceae and Solanaceae. A strong geographical differentiation was observed in the frequency of predominant mutations in *SmAPRR2*, resulting in a lack of fruit chlorophyll pigmentation and suggesting that this phenotype may have arisen and been selected independently several times. This study represents the first identification of a major gene for fruit chlorophyll pigmentation in the eggplant fruit.

Keywords: fruit peel chlorophyll pigmentation, eggplant (*Solanum melongena*), *SmAPRR2*, multi-parent advanced generation inter-cross (MAGIC) population, fruit colour diversification, genome-wide association study (GWAS)

1. Introduction

Chlorophylls are the most widely distributed and important natural pigments because of their essential role in photosynthesis (Gross, 2012; Jia *et al.,* 2020). They are tetrapyrrole compounds, as are two other important plant cofactors, heme and phytochrome (Senge *et al.,* 2014). Chlorophylls are also important photosensitizers, producing Reactive Oxygen Species (ROS) under saturating light intensities (Foyer, 2018). Their metabolism impacts the assembly of photosynthetic machineries but also influences processes such as programmed cell death, the 'stay-green' phenomenon, and chloroplast–nucleus communication (Tanaka and Tanaka, 2006). To optimize photosynthesis and cope with alterations in the light environment, plants have evolved a complex and highly regulated process of chloroplast development (Dubreuil *et al.,* 2018). Given the increasing demand for plant products, as important food sources in the human diet and for many other industrial uses (Pérez-Gálvez *et al.,* 2020), photosynthesis and chloroplast biogenesis have received intensive investigation for their positive involvement in crop nutritional quality (Llorente *et al.,* 2020).

Leaves are the major organs of photosynthesis for most plants, where light energy is absorbed and received by chlorophyll molecules (Ospina Calvo *et al.,* 2017). Green fruits also contain functional chloroplasts with an important photosynthetic activity that affects the fruit growth, development, and composition, leading to the accumulation of metabolites associated with nutritional quality (Ospina Calvo *et al.,* 2017; Jia *et al.,* 2020). Therefore, enhancing fruit chloroplast activity and accumulation can result in mature fleshy fruit with higher nutritional values (Powell *et al.,* 2012; Pan *et al.,* 2013; Jia *et al.,* 2020).

Eggplant (*Solanum melongena* L.) is a widely grown crop, ranking third in global production among Solanaceae crops (FAOSTAT, 2020). It is characterized by a high diversity of commercial fruit colours, which depend mainly on the presence or absence of two pigments: anthocyanins and chlorophylls (Daunay *et al.,* 2004; Page *et al.,* 2019). In eggplant, anthocyanins are responsible for the purple colour of fruit peel, one of the traits of greatest interest in eggplant breeding (Daunay and Hazra, 2012). Chlorophylls confer green colour to fruit peel that is visible when anthocyanins are absent or present in small concentrations. If anthocyanins are present, chlorophylls contribute to a darker background and reinforce the fruit darkness (Daunay *et al.,* 2004). Purple-coloured eggplants are the most demanded in many markets (Li *et al.,* 2018a), and developing dark purple-coloured eggplants, which result from the combination of anthocyanins with chlorophylls, is a major objective in eggplant breeding programs.

Candidate genes have been proposed for fruit anthocyanin synthesis, distribution, and accumulation (Moglia *et al.,* 2020; Toppino *et al.,* 2020; Mangino *et al.,* 2022; Yang *et al.,* 2022). However, causative genes for the presence of fruit chlorophylls in the eggplant fruit peel have not been elucidated yet. Although there are already some studies of QTLs in eggplant, none of them is focused on this trait. Up to now, it has only been observed that green fruit peel colour is dominant over the white one and the monogenic dominance of the green flesh over the white one, suggesting the dominance of chlorophyll presence over its absence (Daunay *et al.,* 2004). Other studies reported that wild eggplant populations mainly exhibit fruits with green or greenish colour, whereas the white ones are more typical of some cultivated and weedy eggplants from southern India, conferring them the "egg-plant" or "vegetable egg" common name (Davidar *et al.,* 2015; Mutegi *et al.,* 2015). A single major QTL at the top of chromosome 8 was reported for a related trait, the presence of a green ring in the flesh next to the skin, and a member of the ferredoxin gene family was suggested as the best candidate due to its involvement in chlorophyll production (Portis *et al.,* 2014).

The presence of chlorophylls in the eggplant fruit peel, not only for their influence on fruit colour but also for their potential effect on fruit composition, makes their study interesting for eggplant breeding. In this respect, the use of experimental and germplasm populations has been of great relevance for mapping quantitative trait loci (QTLs) and identifying candidate genes related to agronomic traits of interest (Gramazio *et al.,* 2018). Here, for the first time, we identify a major candidate gene controlling the fruit peel chlorophyll biosynthesis in eggplant by using a multi-parent advanced generation inter-cross (MAGIC) population and a core germplasm collection for gene validation.

2. Materials and methods

2.1. Plant materials

A total of 420 S3 individuals from the eggplant S3MEGGIC multi-parent advanced generation inter-cross (MAGIC) population developed by Mangino *et al*. (2022) were used to identify candidate genes associated with the chlorophyll pigmentation in the fruit peel. The S3MEGGIC population, which is the first and so far the only MAGIC population in eggplant, was obtained by inter-crossing one wild *S. incanum* accession and seven *S. melongena* accessions. Two out of the seven S3MEGGIC *S. melongena* founders (A0416 and IVIA-371; founders F and G, respectively) do not have chlorophyll in the fruit peel, resulting in a population of

S3 individuals segregating for this trait. In addition, genomic sequences of 277 *S. melongena* accessions, available from the eggplant germplasm core collection established in the framework of the G2P-SOL project (http://www.g2p-sol.eu/G2P-SOL-gateway.html) and phenotyped for fruit peel colour, were interrogated for the candidate gene identified for controlling chlorophyll pigmentation. This core collection includes accessions used for developing the first eggplant pan-genome (Barchi *et al.,* 2021).

2.2. Fruit chlorophyll phenotyping

The presence of chlorophyll distributed all over the peel of the fruit was screened in the 420 S3MEGGIC individuals and 277 core collection accessions using a binary classification (presence/absence). Fruits were phenotyped at the stage of commercial maturity (e.g., when the fruit is still physiologically immature). When the fruit peel had no anthocyanins or they were distributed irregularly, green pigmentation was easily phenotyped with the naked eye. However, when anthocyanins were distributed all over the fruit peel, the presence of chlorophylls was determined by observing the stylar scar on the distal part of the fruit (Daunay *et al.,* 2004).

2.3. S3MEGGIC genome-wide association study (GWAS) and haplotype diversity

The genotyping data of the 420 S3MEGGIC individuals, assessed by the Single Primer Enrichment Technology (SPET) eggplant probes available (Barchi *et al.,* 2019a), was retrieved from Mangino *et al*. (2022). Combining both phenotypic and genotypic data, a Genome-Wide Association Study (GWAS) was performed using the Trait Analysis by aSSociation, Evolution and Linkage (TASSEL) software (ver. 5.0, Bradbury *et al.,* 2007). For the association study, mixed linear model (MLM) analyses were conducted. The multiple testing was corrected with the Bonferroni and the false discovery rate (FDR) methods (Holm, 1979; Benjanmini and Hochberg, 1995) at the significance level of 0.05 (Thissen *et al.,* 2002). The R qqman (Turner, 2014) and LDBlockShow (Dong *et al.,* 2021) packages were used for the Manhattan plots visualization and linkage disequilibrium (LD) determination for haplotype block structure plotting, respectively. LD correlation coefficient (r^2) was used for the pattern of pairwise LD between SNPs measurement, considering haplotype blocks for r^2 values greater than 0.5 and supported by the solid spine of LD method (Gabriel *et al.,* 2002; Barrett *et al.,* 2005). Candidate genes found in the

most significant regions were first retrieved from the 67/3 eggplant reference genome (ver. 3) (Barchi *et al.,* 2019b). Candidate gene allelic variants were assessed by SnpEff software prediction (ver. 4.2, Cingolani *et al.,* 2012) on the eight S3MEGGIC founders' resequencing data (Gramazio *et al.,* 2019). Integrative Genomics Viewer (IGV) tool was used for the visual exploration of founder genome sequences to validate SnpEff results (Robinson *et al.,* 2020). Founder haplotypes were estimated for the regions with significant associations using the SNPs with the highest LODs. In addition, a comparative analysis of founder haplotype diversity across the 420 S3MEGGIC individuals was performed by combining genotypic and phenotypic data.

2.4. SmAPRR2 sequence analysis

Allelic variants of the eggplant ARABIDOPSIS PSEUDO RESPONSE REGULATOR2 (*SmAPRR2*) gene were interrogated in the S3MEGGIC population. Of the total 420 S3 individuals, 20 individuals presenting fruit peel chlorophyll (including all combinations of founder haplotypes with chlorophyll presence) and 60 without fruit chlorophyll (30 associated with the A0416 founder haplotype and 30 with the IVIA-371 founder haplotype) were selected to confirm the involvement of the high-impact variants in *SmAPRR2* identified in the resequencing data of the founders. Genomic DNA of the selected individuals was extracted using the SILEX extraction method (Vilanova *et al.,* 2020) and its quality and integrity were checked by agarose electrophoresis and Nanodrop ND-1000 spectrophotometer ratios 260/280 and 260/230 (Thermo Fisher Scientific, Waltham, MA, United States). DNA concentration was estimated with Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, United States) and adjusted to 50 ng/µL for PCR amplification. In order to genotype the *SmAPRR2* candidate allelic variants in the selected S3MEGGIC individuals, specific primers were designed based on the "67/3" reference genome sequence and founders' resequencing data (Supplementary Table 2) and the resulting amplifications were visualized by 1% agarose gel electrophoresis. For long amplification fragments (>5 kb), TaKaRa LA Taq® DNA Polymerase (Takara Bio Inc., Shiga, Japan), which is optimized for long-range PCR up to >15 kb fragments, was used. PCR amplicons were purified with ammonium acetate and sequenced by Sanger sequencing (DNA Sequencing Service, IBMCP-UPV, Valencia, Spain).

To compare the *SmAPRR2* gene structure, the nucleotide sequence was retrieved by a BLASTx search (e-value cut-off of $1e^{-5}$) against different eggplant reference genomes showing fruit peel chlorophyll, such as the Nakate-Shinkuro and the GUIQIE-1 (Hirakawa *et al.,* 2014; Li *et al.,* 2021), and others showing absence

of fruit peel chlorophyll, such as the 67/3 (ver. 3) and the HQ-1315 (Barchi *et al.,* 2019b; Wei *et al.,* 2020). A conservative domain analysis was performed by assessing the NCBI conserved domain server (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). To gain insight into the relationship of *APRR2-like* proteins in different species, a phylogenetic analysis was also performed using the predicted protein sequences from *Arabidopsis thaliana*, Cucurbitaceae and Solanaceae families. Protein alignments were performed using MEGA 11.0.10 software (http://megasoftware.net/) and the dendrogram was constructed using the IQtree web server (http://iqtree.cibiv.univie.ac.at/) via maximum likelihood method with default settings.

2.5. Germplasm collection validation

To validate the *SmAPRR2* allelic forms, we examined the GUIQIE-1 eggplant reference genome (Li *et al.,* 2021), since GUIQIE-1 fruits have fruit peel chlorophyll and its assembly is much more complete than the Nakate-Shinkuro reference genome (Hirakawa *et al.,* 2014). Furthermore, whole-genome resequencing data of the 277 eggplant G2P-SOL germplasm core collection were interrogated (unpublished data). After trimming with SOAPnuke software (Chen *et al.,* 2018) with filter parameters: "-l 20 -q 0.5 -n 0.03 -A0.28", reads were aligned against GUIQIE-1 genome using the Burrows-Wheeler Aligner (BWA) program (v. 0.7.17–r1188, Li and Durbin, 2009) and the 'mem' command with default parameters. Subsequently, reads mapping in the genomic region of *SmAPRR2* were extracted with samtools and de novo assembled using Megahit (v.1.2.9, Li *et al.,* 2015) with default parameters. The sequencing raw data of the newly sequenced accessions are available at NCBI SRA (BioProject ID PRJNA837769).

A multiple sequence alignment of the assembled sequences was then performed using the MAFFT program (ver. 7, Katoh and Hiroyuki, 2008) and the results were visualized in the Jalview alignment editor (ver. 2, Waterhouse *et al.,* 2009). The resulting data were used to associate the different variants identified in the *SmAPRR2* gene for the G2P-SOL core collection and the origin of the accessions.

3. Results

3.1. Phenotypic variation and association analysis

MAGIC population founders are contrasting for fruit peel chlorophyll (FC), showing six out of eight founders the presence of FC, namely MM1597, DH ECAVI, MM577, AN-S-26, H15, and ASI-S-1 (founders A, B, C, D, E, and H, respectively), and two of them, A0416 and IVIA-371 (founders F and G, respectively), the absence of FC (Figure 1). The screening for FC segregation among the 420 S3MEGGIC individuals revealed considerable variation for this trait. The dominance of the FC presence was confirmed by the phenotype of the simple hybrids obtained by crossing founders with and without FC for developing the S3MEGGIC population (ExF and GxH hybrids) (Figure 1). No maternal effect was observed on FC in hybrids.

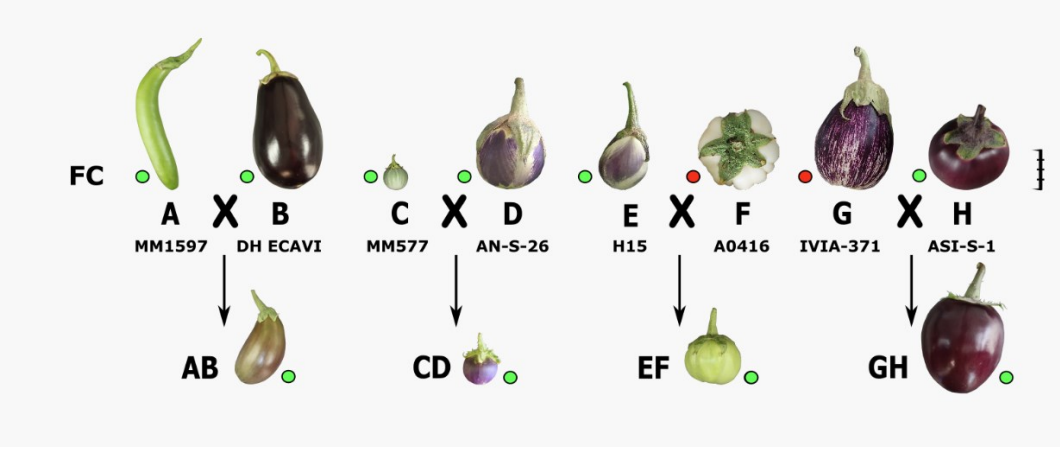


Figure 1. The eight founders, coded from A to H, from the S3MEGGIC population and the four simple hybrids obtained from their inter-cross (AB, CD, EF and GH) are represented at a scale based on the real fruit size. The scale bar represents 5 cm. Phenotyping for FC is represented by green and red dots, respectively.

GWAS analysis combining the genotypic and phenotypic data of the S3MEGGIC population led to the identification of significant associations for FC presence. The Manhattan plot revealed one major peak on chromosome 8, plus another one with a lower significance value on chromosome 4 (Figure 2). For the major peak on chromosome 8, 12 significant SNPs over the FDR threshold (LOD > 3.45) were identified, eight of them being over the Bonferroni threshold (LOD >

5.16) in the genomic region between 103.22 and 104.98 Mb and reaching the LOD of 34.88. On chromosome 4 there were eight significant SNPs over the FDR threshold (LOD > 3.45) in the region between 3.23 and 6.35 Mb, but none of them was significant over the Bonferroni threshold.

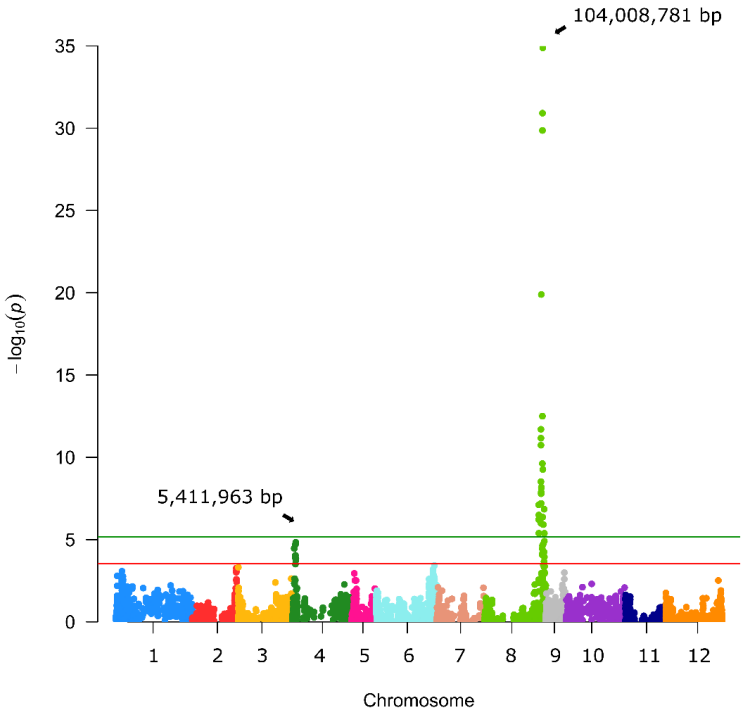


Figure 2. Manhattan plot for fruit chlorophyll. Arrows indicate the genome position of the highest peaks detected for the fruit peel chlorophyll trait. The red and green horizontal lines represent, respectively, FDR and Bonferroni significance thresholds at p = 0.05.

For the main candidate region on chromosome 8, a comparative analysis of the founder haplotype diversity was performed in the S3MEGGIC population (Figure 3). The haplotype prediction of the 156 S3 lines with the lack of FC (i.e., having A0416 or IVIA-371 founder haplotypes) was successful in 91.03% of the cases. In this way, 50.00% of the cases (78 of 156) corresponded to the A0416 haplotype and 41.03% (64 of 156) to the IVIA-371 haplotype. In 6.41% of cases (10 of 156), the absence of FC was associated with the wild *S. incanum* MM577 founder haplotype, which presents some degree of heterozygosity that could have interfered

with the association analysis. The remaining 2.46% (4 of 156) was evenly distributed between MM1597 and ASI-S-1 haplotypes, which could be the result of inaccurate phenotyping due to the presence of anthocyanins or of a small stylar scar, making phenotyping difficult.

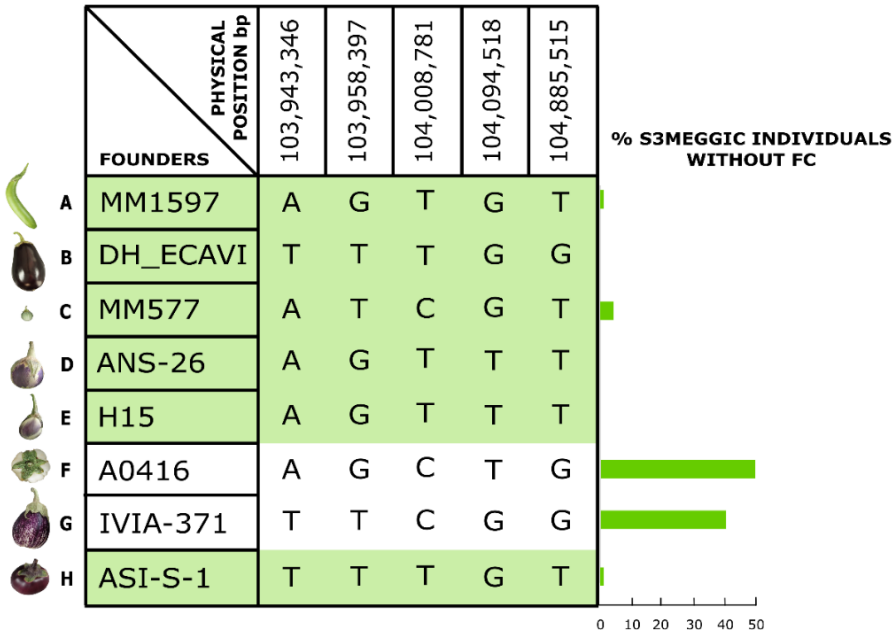


Figure 3. On the left, S3MEGGIC founder haplotypes at the most significant genomic region of chromosome 8; on the right, a histogram representing the percentage of S3MEGGIC individuals without FC with the same founder haplotype in that region.

3.2. Candidate genes for chlorophyll biosynthesis

Based on the results of the GWAS analysis, putative candidate genes were identified close to or within LD blocks defined in the genomic regions with significant associations (Supplementary Figure 1, Supplementary Table 1). From the S3MEGGIC founders' resequencing data, variants with high impact effects on protein structure and function were annotated by SnpEff for all the candidate genes. Under the major GWAS peak of chromosome 8, in the genomic region of 104,008,667 – 104,012,138 bp, a candidate gene was identified as similar to ARABIDOPSIS PSEUDO RESPONSE REGULATOR2 (*APRR2*,

SMEL_008g315370.1), which has been described as a pigment accumulation regulator and chloroplast development promotor in several solanaceous and cucurbitaceous crops (Pan *et al.,* 2013; Liu *et al.,* 2016; Oren *et al.,* 2019). This gene presented high impact variants in those S3MEGGIC founders that do not present FC (founders F and G), which was confirmed by aligning all founder gene sequences retrieved and visualized in IGV. Specifically, founders F and G exhibited small deletions predicted to cause a frameshift variant identified as high impact: TG instead of TCTCCG in the 14,010,718 bp position on exon 3 and AT instead of ACT in the 104,011,578 bp position on exon 6, respectively (Figure 4A).

Same procedure was performed for the minor peak on chromosome 4. All the candidate genes close to or within the LD block including the genomic region with significant association were annotated by SnpEff for each of the S3MEGGIC founders. In the genomic region of 5,457,658 – 5,461,306 bp, a candidate gene was identified as similar to *GLK2* (SMEL_004g203570.1), which has been described as a positive regulator of chloroplast development and pigment accumulation in tomato fruit (Powell *et al.,* 2012; Nguyen *et al.,* 2014). Although according to the *SmGLK2* annotation it might fit with the regulation of chlorophyll biosynthesis, no high-effect variants were predicted by SnpEff in this gene for any of the founders without FC.

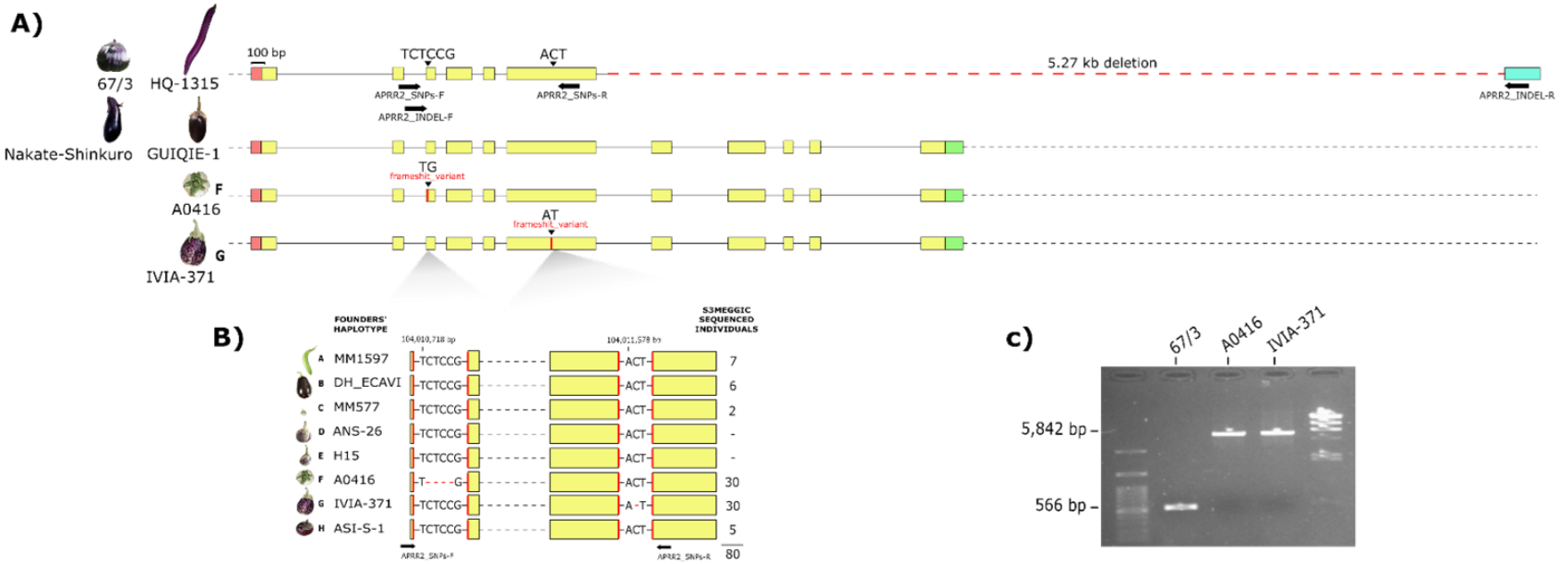


Figure 4. (A) *SmAPRR2* gene structure (5'-UTR in red, exons in yellow, 3'-UTR in green and 3'-UTR annotated in the 67/3 eggplant reference genome in blue) for 67/3, HQ-1315, Nakate-Shinkuro, GUIQIE-1 reference genomes and A0416 and IVIA-371 (S3MEGGIC founders F and G, respectively). Structural variations are indicated with black arrowheads and primers hybridization positions with horizontal black arrows. (B) The 80 S3MEGGIC selected individuals for sequencing based on their haplotype and their respective *SmAPRR2* gene sequence between APRR2_SNPs primers. (C) Electrophoresis gel of the PCR amplification using APRR2_INDEL primers for 67/3 and S3MEGGIC founders F and G genomes.

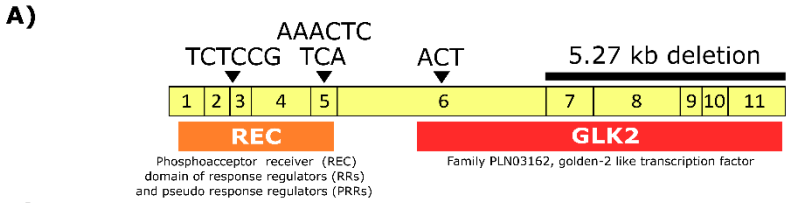
3.3. Variants on SmAPRR2 sequence

To confirm experimentally the high impact variants predicted by SnpEff for S3MEGGIC founders F and G in the *SmAPRR2* gene sequence, 80 S3MEGGIC individuals were selected based on their phenotype and haplotype (20 individuals with FC, 30 with A0416 haplotype and 30 with IVIA-371 haplotype). The *SmAPRR2* gene sequence of the selected individuals was amplified by PCR using the primers combination APRR2_SNPs, which amplified a region of 1,363 bp where the two high impact variants were located (Figure 4A, Supplementary Table 2). The Sanger sequencing of the amplicon confirmed the presence of specific allelic variants according to their phenotype (Figure 4B). The T(CTCC)G deletion was found in each of the 30 individuals with A0416 haplotype and the A(C)T deletion was found in each of the 30 individuals with IVIA-371 haplotype. As expected, no other variants were identified in the 20 remaining individuals that presented FC.

To further investigate the *SmAPRR2* gene structure, its nucleotide sequence was retrieved from the reference genome 67/3 (ver. 3, Barchi *et al.,* 2019b), which corresponds to an accession without FC (SMEL_008g315370.1), presenting a 3,471-bp gene comprising 6 exons. The same gene structure was observed also in the HQ-1315 eggplant reference genome (Wei *et al.,* 2020), which corresponds to another accession without FC (Smechr0802018). However, in the Nakate-Shinkuro (Sme2.5_00446.1_g00003.1) and the GUIQIE-1 eggplant reference genomes (Hirakawa *et al.,* 2014; Li *et al.,* 2021), both of which exhibit FC, a larger gene size of 4.794 bp with 11 exons was found (Figure 4A). Comparing the four reference genomes for the *SmAPRR2* gene structure, a large deletion of 5.27 kb was identified on 67/3 and HQ-1315 accessions causing the loss of exons 7 to 11. This large deletion did not correspond with the two high-effect variants observed in the S3MEGGIC population, where the larger gene size of 11 exons was the only *SmAPRR2* gene structure present in the population. The presence of the large deletion in the 67/3 accession was confirmed by PCR, using the APRR2_INDEL primers, comparing its amplicon size with those of founders F and G (Figure 4C, Supplementary Table 2). While founders F and G exhibited an amplicon size of 5,842 bp, the 67/3 accession amplified a shorter fragment of only 566 bp, which was in agreement with the in-silico prediction.

The large deletion present in the 67/3 and HQ-1315 accessions could be responsible for the absence of FC in these genotypes, since they did not present the same high-effect variants found on founders F and G. By analysing the full-length amino acid sequence of *SmAPRR2* in the NCBI conserved domain server (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi), the 132 truncated residues

at the C-terminal region caused the loss of a functional domain (184 amino acids) that corresponded to a golden-2 like transcription factor domain (Figure 5A).



B)

Genotype:	Exon 3	Exon 5	Exon 6	Exons 7-11	Phenotype:	G2P-SOL core collection accessions:
	TCTCCG	TCA ... AAACTC	ACT	complete	FC presence	198
67/3	TCTCCG	TCA ... AAACTC	ACT	deletion	FC absence	21
A0416	T- - - -G	TCA ... AAACTC	ACT	complete	FC absence	26
IVIA-371	TCTCCG	TCA ... AAACTC	A-T	complete	FC absence	21
	TCTCCG	TAA ... AAACTC	ACT	complete	FC absence	3
	TCTCCG	TCA ... A- - - -C	ACT	complete	FC absence	5
	TCTCCG	TCA ... AAACTC	ACT	complete	FC absence	3
						277

Figure 5. (A) Conservative domain analysis of *SmAPRR2* full-length amino acid gene sequence, including the coding DNA sequence (CDS) indicating each of the exons as yellow boxes, the position of the different variants identified as causative for fruit chlorophyll absence indicated with black arrowheads, the large 5.27 kb deletion and the conserved domains REC and GLK2 as orange and red boxes, respectively. (B) Gene sequence variants identified from the assembly and alignment of 277 accessions from the eggplant G2P-SOL germplasm core collection indicating the genotype, phenotype and the number of accessions carrying each variant.

To investigate the *APRR2* gene structure in relevant species in the Solanaceae and Cucurbitaceae families, a phylogenetic analysis was performed revealing orthology among *APRR2-like* proteins within families, although no clear relationship was found between the Cucurbitaceae and Solanaceae genes (Figure 6 and Supplementary Figure 2).

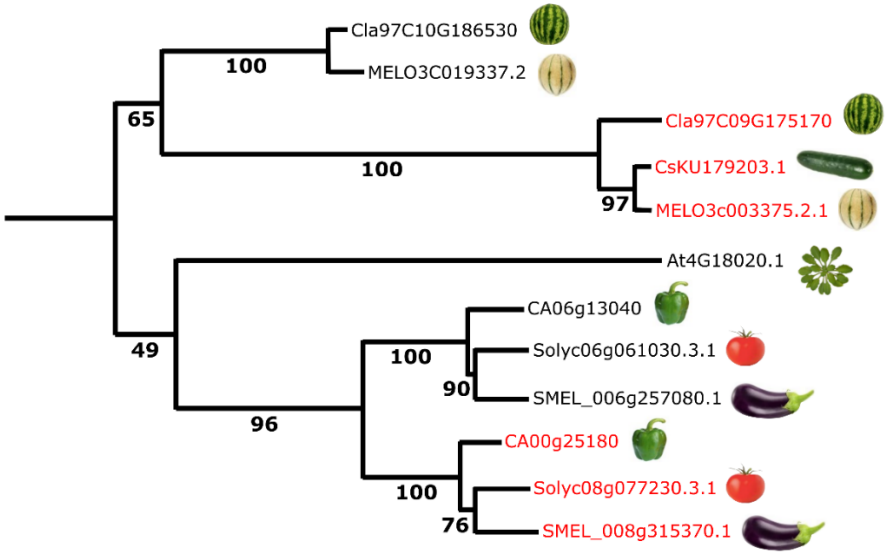


Figure 6. Maximum likelihood dendrogram of *APRR2-like* proteins from *Arabidopsis thaliana*, Cucurbitaceae and Solanaceae species. The red-coloured proteins control fruit chlorophyll pigmentation (Pan *et al.,* 2013; Liu *et al.,* 2016; Oren *et al.,* 2019; Jeong *et al.,* 2020; this paper). Cla, *Citrullus lanatus* (watermelon); MELO, *Cucumis melo* (melon); Cs, *Cucumis sativus* (cucumber); At, *Arabidopsis thaliana*; CA, *Capsicum annuum* (bell pepper); Solyc, *Solanum lycopersicum* (tomato); SMEL, *S. melongena* (eggplant). Gene/IDs are from Cucurbitgenomics.org (Cucurbitaceae), Solgenomics.net (Solanaceae), Arabidopsis.org (*Arabidopsis*) and NCBI (cucumber).

3.4. Gene validation in a germplasm collection

The *SmAPRR2* gene sequence was also analysed in a highly diverse germplasm population to validate the hypothesis of *SmAPRR2* as the best candidate gene for FC biosynthesis. The assembly and alignment of 277 accessions from the eggplant G2P-SOL germplasm core collection allowed their classification according to the candidate variants causative for FC absence (Figure 5B, Supplementary Table 3). Among the 79 accessions without FC, 26.58% (21 of 79) presented the 5.27 kb deletion, the same as the 67/3 and HQ-1315 reference genomes, 32.91% (26 of 79) presented the A0416 (S3MEGGIC population founder F) deletion, and 26.58% (21 of 79) presented the IVIA-371 (S3MEGGIC population founder G) deletion. The remaining 11 accessions exhibited two new high impact variants: three presented a C→A SNP in the 104,011,141 bp position on exon 5, resulting in a premature stop

codon, while 5 presented a small A(AACT)C deletion in the 104,011,149 bp position on exon 5, resulting in a frameshift mutation. The remaining 3 accessions without FC had none of the above variants or any apparent high-effect mutation in the sequence *SmAPRR2*. This discrepancy might be caused by phenotyping errors. As expected, for the 198 accessions with FC none of the five variants were observed in the *SmAPRR2* gene sequence.

Furthermore, we checked if there was an association between the geographical origins of the G2P-SOL germplasm core collection accessions and the different variants identified in the *SmAPRR2* gene (Figure 7). It was observed that the 5.27 kb deletion predominated in accessions from Italy (nine of 21, 42.86%), followed by Philippines (four of 21, 19.05%), while the rest of the accessions where mainly distributed among France, China, and India (two of 21 each, 9.52%). The S3MEGGIC founder F (A0416) indel was mainly observed in Thailand and Laos, where 73.08% (19 of 26) and 15.38% (four of 26) of the accessions originated, respectively. Finally, the S3MEGGIC founder G (IVIA-371) indel was distributed mainly between Turkey and Spain with 33.33% (seven of 21) of the accessions having this mutation present in each country, followed by India with a 19.05% (four of 21).

In addition, each high impact variation seemed to be mostly associated with a specific phenotype. The presence of homogeneous fruit anthocyanins with no pigmentation under the calyx was observed for the 5.27 kb deletion in 66.67% (14 of 21) of cases, while completely white fruits or the presence of green netting in a white background was found for the S3MEGGIC founder F (A0416) variant in 88.46% (23 of 26) of cases, and the presence of stripped distribution of anthocyanins on a white background for the S3MEGGIC founder G (IVIA-371) indel was found in 66.67% (14 of 21) of cases. Regarding the two new high impact variants identified, the low number of accessions did not allow to identify a region of predominant distribution, although the A(AACT)C deletion was exclusively located in India and Sri Lanka.

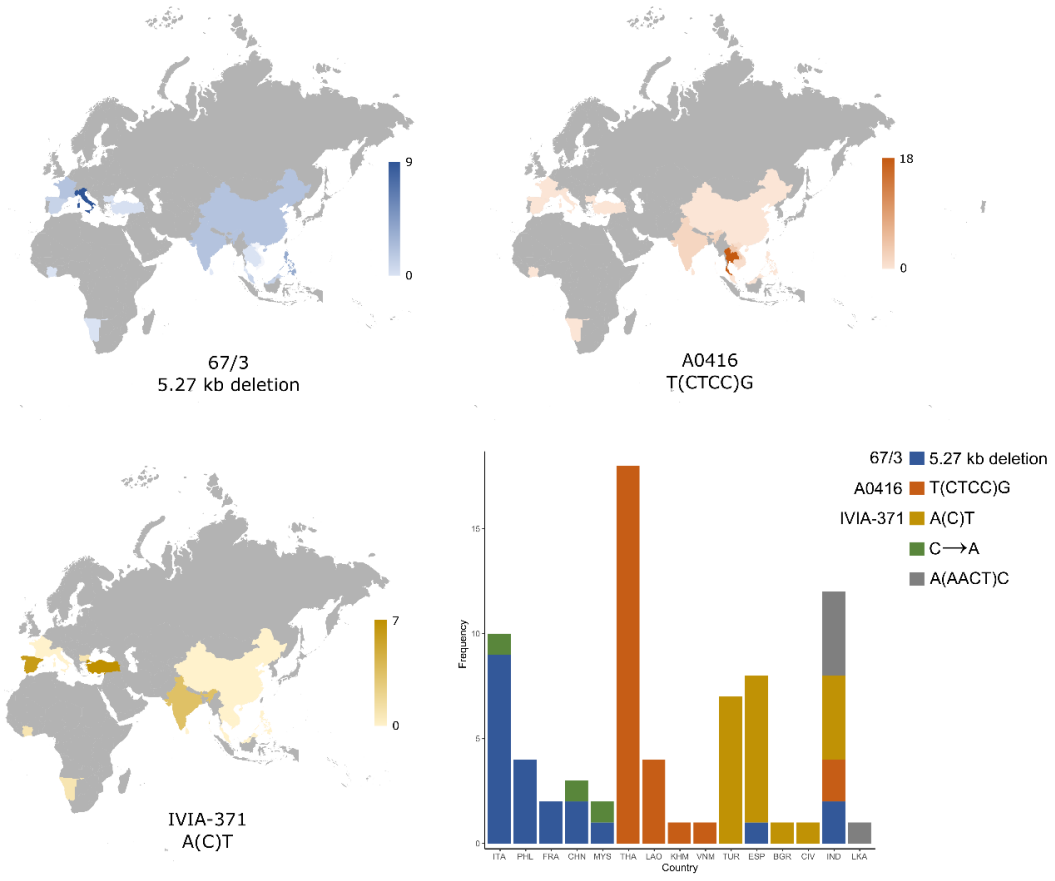


Figure 7. Distribution of the three most abundant high effect variants identified in the *SmAPRR2* gene sequence according to the country of origin of the 277 accessions from the eggplant G2P-SOL germplasm core collection. The numbers on the scale next to each map indicate the numbers of accessions from each country carrying the mutation. Bottom right, a stacked column chart comprising the frequency of the different *SmAPRR2* gene variants across the countries of origin: Italy (ITA), Philippines (PHL), France (FRA), China (CHN), Malaysia (MYS), Thailand (THA), Laos (LAO), Cambodia (KHM), Vietnam (VNM), Turkey (TUR), Spain (ESP), Bulgaria (BGR), Ivory Coast (CIV), India (IND), and Sri Lanka (LKA).

4. Discussion

Among fruit quality traits, peel colour is of primary importance because pigments that confer colour are not only associated with visual preferences, but also

with nutritional, health, and flavour values (Brand *et al.,* 2014). Eggplant varieties and landraces display different colours at commercial maturity, resulting from a combination of different patterns of presence or absence of anthocyanins and chlorophylls. Although chlorophylls in the peel of fleshy fruits might have a considerable effect on the quality composition of eggplant, as has been demonstrated for tomato (Li *et al.,* 2018b), responsible genes have not been elucidated yet.

To dissect this important trait for eggplant genetics and breeding, in this study we have combined classical forward genetics techniques with advanced association and sequencing analyses, using both experimental populations and germplasm materials to identify and validate a candidate gene controlling the synthesis of fruit peel chlorophyll. Up to now, the dissection of complex traits in eggplant has been usually performed using germplasm panels (Portis *et al.,* 2015) or temporary mapping populations, such as F_2 and BC_1 (Barchi *et al.,* 2012; Miyatake *et al.,* 2016; Pandiyaraj *et al.,* 2019; Bhanushree *et al.,* 2020). These populations have the advantage of being easy and fast to develop but are not as powerful as the immortal ones to identify strong associations, candidate genes or causative SNPs, and accumulate fewer genetic recombination events, limiting the resolution for QTL detection (Arrones *et al.,* 2020). In eggplant, this has led to the identification of very few candidate genes controlling traits of interest, lagging it behind other major crops where multiple bi-parental and multi-parent experimental populations have been developed and available for many years already (Gramazio *et al.,* 2018). Luckily, in the last few years, new advanced populations have been developed in eggplant, including recombinant inbred lines (RILs) (Lebeau *et al.,* 2013; Mishra *et al.,* 2020; Toppino *et al.,* 2020; Sulli *et al.,* 2021), and one set of introgression lines (ILs) (Gramazio *et al.,* 2017). These populations have allowed the identification of many and highly relevant QTLs (Mangino *et al.,* 2020; Mangino *et al.,* 2021; Toppino *et al.,* 2020; Rosa-Martínez *et al.,* 2022). Eventually, the first eggplant multi-parent advanced generation inter-cross (MAGIC) population has also been developed, allowing the identification of candidate genes for anthocyanin pigmentation in eggplant fruits with greater accuracy and statistical robustness (Mangino *et al.,* 2022).

In this study, this new MAGIC population, called S3MEGGIC, allowed the identification by GWAS analysis of a strong association (LOD = 34.88) for fruit chlorophyll in the peel and the identification of several potential candidate genes beneath or close to the association peak. Furthermore, the whole-genome resequencing of the founders coupled with the high-throughput genotyping of the 420 S3MEGGIC individuals through the identification of "high-impact" variants and haplotype predictions provided strong evidence that ARABIDOPSIS PSEUDO RESPONSE REGULATOR2 (*APRR2*) is the candidate gene for controlling this important trait. In tomato, the overexpression of *SlAPRR2* in transgenic plants

resulted in an increased plastid number, area, and pigment content, enhancing chlorophyll levels in immature unripe fruits (Pan *et al.,* 2013). Through the development of two F_2 and one BC_1 mapping populations in cucumber, *CsAPRR2* was identified as responsible for the green colour of immature fruits (Jiao *et al.,* 2017). In the case of melon and watermelon, the development of different bi-parental segregating populations and a GWAS panel allowed the association of *CmAPRR2* and *ClAPRR2* as causative genes regulating fruit rind chlorophyll accumulation (Oren *et al.,* 2019). Similarly, in pepper, *CaAPRR2* was found to be strongly associated with chlorophyll pigment accumulation in immature fruit tissues through the analysis of an F_2 population and confirmed through virus-induced gene silencing (VIGS) (Jeong *et al.,* 2020). Null mutations on *APRR2* have been related in some way to alterations in abscisic acid (ABA) signalling related to plastid development causing reduced chloroplast density and chlorophyll content (Pan *et al.,* 2013; Oren *et al.,* 2019). Here, we present the first evidence of *APRR2* as the main actor in chlorophyll presence in the peel of eggplant fruits. Previous studies focused on the eggplant green ring, observed in the flesh next to the skin in an F_2 population, identified a major QTL on chromosome 8 linked to the marker *35002_PstI_L402* (Portis *et al.,* 2014). Although these authors identified a member of the ferredoxin gene family as the best candidate gene, the marker position is only 623 kb away from the *SmAPRR2* gene.

Thus, *APRR2* homologs have been shown to control fruit chlorophyll content in six different species: three Cucurbitaceae (melon, watermelon and cucumber) and three Solanaceae (tomato, pepper and eggplant). It is noteworthy that the three Cucurbitaceae genes are orthologs of each other, as are the three Solanaceae ones. In contrast, no clear evolutionary relation could be found between the Cucurbitaceae and Solanaceae genes. Thus, the most likely scenario is that the specialization of these *APRR2-like* genes in controlling fruit chlorophyll content occurred early during the evolution of Cucurbitaceae and Solanaceae, and that the two events were independent from each other.

We also reported that the *SmAPRR2* gene sequence for the 67/3 and HQ-1315 eggplant reference genomes, both without fruit chlorophyll, presented a large deletion event of 5.27 kb, disrupting the chlorophyll synthesis in the fruit peel. However, a larger gene size of 4.794 bp with 11 exons was found in the Nakate-Shinkuro and GUIQIE-1 eggplant reference genomes, both with FC (Hirakawa *et al.,* 2014; Li *et al.,* 2021). The availability of different reference genomes contrasting for several traits, like in this study for fruit chlorophyll, fosters their dissection and highlights the need for developing newer, more phenotypically diverse, and accurate reference genomes (Li *et al.,* 2021). The Nakate-Shinkuro and GUIQIE-1 *SmAPRR2* gene structure is largely in agreement with the ones reported in other species, such as the tomato gene *Solyc08g077230* with a 5,831 bp gene size and 11 exons, the

cucumber *Csa3G049490* with 5,864 bp and 13 exons, the melon *Melo3C003375* with approximately 6,949 bp and 10 exons, the watermelon *Cla97C09G175170* with 7,206 bp and 12 exons, or the pepper *Ca06g13040* with 4,428 bp and 11 exons (Solgenomics.net and Cucurbitgenomics.org). The deletion found in 67/3 and HQ-1315 was associated with 132 residues causing the loss of a functional domain, the golden-2 like transcription factor. A similar deletion in the cucumber *CsAPRR2* gene, encoded for a truncated 101-amino acid protein, was suggested to control the white immature peel fruit colour (Liu *et al.,* 2016). A subcellular localization analysis revealed that the loss of the *CsAPRR2* golden-2 like transcription factor, which acts as a nuclear localization signal domain, caused the impossibility for the protein to enter the nucleus and perform its function (Jiao *et al.,* 2017).

The availability of a large number of accessions, which included a broad range of eggplant genetic diversity, of the eggplant G2P-SOL germplasm core collection made possible the in-silico validation of the hypothesis of *SmAPRR2* as the best candidate gene for fruit peel chlorophyll. Due to the high recalcitrance of in vitro regeneration after *Agrobacterium*-mediated transformation of eggplant (García-Fortea *et al.,* 2020; Mir *et al.,* 2021), routinely reverse genetics gene validation techniques, such as over-expression or knock-out, are still the major bottleneck in this crop. Therefore, the screening of large germplasm collection remains the best choice to indirectly validate gene functions. In fact, despite the explosion of CRISPR/Cas studies in many crops, the first and only eggplant CRISPR plant so far has been obtained by the authors of this study (Maioli *et al.,* 2020), reflecting the inevitable need for alternative indirect validations until new protocols overcome the recalcitrance.

Combining all this information, we highlighted that the two small high-impact variants identified in the S3MEGGIC population, the large deletion of the 67/3 and HQ-1315 reference genomes plus the two high-impact disruptive variants identified in the core collection are responsible for the lack of chlorophyll in the fruit. These five disruptive mutations indicate that the lack of chlorophyll in the eggplant fruit peel has arisen and been selected independently several times, suggesting a prominent role of this trait in eggplant domestication and diversification. Furthermore, we observed that mutated *SmAPRR2* gene variants displayed a different and specific geographical distribution, being also associated with a specific phenotype. This distribution might be related to the domestication flows of eggplant together with local preferences. It has been suggested that two main different domestication events occurred in eggplant, one in India and one in China, since both regions have a high diversity of landraces and populations of putatively wild eggplants (Meyer *et al.,* 2012). Here, we observed that India concentrated the greatest diversity for high impact gene variants for the *SmAPRR2*, including accessions with almost all the identified gene variants. Meyer *et al*. (2012) also

proposed that landraces originated from India would have spread west to western Europe, possibly carried by Arab traders, while landraces arising in China spread to the northeast and southeast Asia. This hypothesis is in agreement with the high frequencies observed of the 5.27 kb deletion present in 67/3 and HQ-1315 and the indel present in the S3MEGGIC founder G (IVIA-371) in Europe, where purple eggplants are popular, and the indel of the S3MEGGIC founder F (A0416) to southeast Asia, where white eggplants are most consumed. This may also suggest that the mutations causing the lack of chlorophyll in gene *SmAPRR2* predate the spread of eggplant to other regions outside its region of origin. In this respect, individuals with white fruits from wild populations of eggplant have been described in India and in other parts of Southeast Asia (Davidar *et al.,* 2015; Page *et al.,* 2019).

Although no clear evidence was found in our population, the Golden 2-like MYB (*GLK2*) gene, which was identified in the genome region where a minor GWAS peak was identified on chromosome 4, is described as a widely conserved transcription factor that positively regulates fruit chlorophylls in different species. Some reports suggest a role of GLK transcription factors in regulating chlorophyll levels in *Arabidopsis*, tomato, and pepper (Fitter *et al.,* 2002; Powell *et al.,* 2012; Brand *et al.,* 2014). The potential tomato ortholog is the *u* (*uniform ripening*) gene affecting shoulder colour in unripe fruits (Powell *et al.,* 2012). In eggplant, a QTL in chromosome 4 has been related to the reticulated pattern of chlorophylls in the fruit, also called fruit chlorophyll netting or variegation (Daunay *et al.,* 2004; Frary *et al.,* 2014). Although it seems to be independent of the uniform distribution of chlorophyll in the eggplant fruit peel, its presence in fruits could lead to phenotyping errors, which might account for the small discrepancies between the genotype and phenotype that we observed during the experimental procedures.

In conclusion, using different eggplant experimental materials, a MAGIC population and a germplasm core collection, combined with advanced association and bioinformatics analyses and classical genetics tools, we found that *SmAPRR2* is a key gene in the eggplant fruit peel chlorophyll biosynthesis. Our finding is also supported by its conserved function in regulating fruit green pigmentation of different vegetable crops. The dissection of the genetics of this important trait will be extremely useful to foster future breeding programs focused not only on specific market demand based on visual preferences, but also on developing new varieties with improved nutritional quality and resilient against the upcoming environmental challenges due to the implication of chlorophylls in stress biology of plants. The identification of the eggplant *SmAPRR2* represents a landmark for eggplant breeding for fruit colour and other related quality traits.

Data availability statement: The data presented in the study are deposited in the NCBI SRA repository, accession number PRJNA837769.

Author contributions: JP, SV, and PG conceived the idea and supervised the study. AA, GM, DA, MP, and PG performed the field trials. AA, GM, SV, and PG performed the analysis of the S3MEGGIC population and the APRR2 gene structure. LB, EP, and GG performed the analyses of the G2P-SOL core collection and the phylogenetic analysis. AA and PG prepared a first draft of the manuscript and the rest of the authors reviewed and edited the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been funded by grants RTI-2018–094592-BI00 and PID2021-128148OB-I00 funded by MCIN/AEI/10.13039/501100011033/and by "ERDF A way of making Europe", and by European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 677379 (G2P-SOL project: Linking genetic resources, genomes and phenotypes of Solanaceous crops). AA is grateful to Spanish Ministerio de Ciencia, Innovación y Universidades for a predoctoral (FPU18/01742) contract. PG is grateful to Spanish Ministerio de Ciencia e Innovación for a post-doctoral grant (FJC2019-038921-I/AEI/10.13039/501100011033). Funding for open access charge: Universitat Politècnica de València.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material: The supplementary material for this article can be found online at:

https://www.frontiersin.org/articles/10.3389/fpls.2022.1025951/full#supplementary-material

Supplementary Figure 1. Local Manhattan plot and LD heatmap of the candidate peaks on chromosomes 8 and 4, top and bottom figures, respectively. The red and green horizontal lines represent, respectively, FDR and Bonferroni significance thresholds at p=0.05. In red are represented the points that exceed these thresholds, while in blue are the points that do not. Pairwise LD between SNPs is indicated as r2 values: red indicates a value of 1 and white indicates 0.

Supplementary Figure 2. MEGA alignment of *APRR2-like* protein sequences from relevant species of the Solanaceae and Cucurbitaceae families and *Arabidopsis thaliana*.

Supplementary Table 1. List of candidate genes for chlorophyll pigmentation in the eggplant fruit peel under the identified GWAS peaks on chromosomes 4 and 8.

Supplementary Table 2. List of primers used in this study.

Supplementary Table 3. Gene sequence variants identified in the 277 accessions from the eggplant G2P-SOL germplasm core collection indicating the G2P-SOL ID, the variant present in each accession and the fruit peel chlorophyll phenotype presence (YES) or absence (NO).

References:

Arrones, A., Vilanova, S., Plazas, M., Mangino, G., Pascual, L., Díez, M. J., *et al*. (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*, 229. doi: 10.3390/biology9080229

Barchi, L., Lanteri, S., Portis, E., Valè, G., Volante, A., Pulcini, L., *et al*. (2012). A RAD tag derived marker based eggplant linkage map and the location of QTLs determining anthocyanin pigmentation. *PloS One, 7*, e43740. doi: 10.1371/journal.pone.0043740

Barchi, L., Acquardo, A., Alonso, D., Aprea, G., Bassolino, L., Demurtas, O., *et al*. (2019a). Single primer enrichment technology (SPET) for high-throughput genotyping in tomato and eggplant germplasm. *Front. Plant Sci., 10*, 1005. doi: 10.3389/fpls.2019.01005

Barchi, L., Pietrella, M., Venturini, L., Minio, A., Toppino, L., Acquardo, A., *et al*. (2019b). A chromosome-anchored eggplant genome sequence reveals key events in solanaceae evolution. *Sci. Rep., 9*, 1–13. doi: 10.1038/s41598-019-47985-w

Barchi, L., Rabanus-Wallace, M. T., Prohens, J., Toppino, L., Padmarasu, S., Portis, E., *et al*. (2021). Improved genome assembly and pan-genome provide key insights into eggplant domestication and breeding. *Plant J., 107,* 579–596. doi: 10.1111/tpj.15313

Barrett, J. C., Fry, B., Maller, J., and Daly, M. J. (2005). Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics, 21*, 263–265. doi: 10.1093/bioinformatics/bth457

Benjanmini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B (Methodolcol), 57*, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Bhanushree, N., Saha, P., Tomar, B. S., Gopala Krishnan, S., Gurung, B., Ghoshal, C., *et al*. (2020). Single marker analysis and mapping of QTLs governing fruit weight in eggplant (*Solanum melongena* L.). *Proc. Natl. Acad. Sci. India Section B: Biol. Sci., 90*, 769–775. doi: 10.1007/s40011-019-01149-y

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. doi: 10.1093/bioinformatics/btm308

Brand, A., Borovsky, Y., Hill, T., Rahman, K. A. A., Bellalou, A., Van Deynze, A., *et al*. (2014). *CaGLK2* regulates natural variation of chlorophyll content and fruit color in pepper fruit. *Theor. Appl. Genet., 127,* 2139–2148. doi: 10.1007/s00122-014-2367-y

Chen, Y., Chen, Y., Shi, C., Huang, Z., Zhang, Y., Li, S., *et al*. (2018). SOAPnuke: a MapReduce acceleration-supported software for integrated QualityC and preprocessing of high-throughput sequencing data. *Gigascience, 7*, gix120. doi: 10.1093/gigascience/gix120

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., *et al*. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of drosophila melanogaster strain W1118; iso-2; iso-3. *Fly, 6,* 80–92. doi: 10.4161/fly.19695

Daunay, M. C., Aubert, S., Frary, A., Doganlar, S., Lester, R. N., Barendse, G., *et al.* (2004). "Eggplant (*Solanum melongena*) fruit colour: pigments, measurements and genetics" in *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., and Hazra, P. (2012). "Eggplant" in *Handbook of vegetables*, eds K. V. Peter and P. Hazra. (Houston, TX: Studium Press), 257–322.

Davidar, P., Snow, A. A., Rajkumar,M., Pasquet, R., Daunay, M. C., andMutegi, E. (2015). The potential for crop to wild hybridization in eggplant (*Solanum melongena*; solanaceae) in southern India. *Am. J. Bot., 102*, 129–139. doi: 10.3732/ajb.1400404

Dong, S. S., He, W. M., Ji, J. J., Zhang, C., Guo, Y., and Yang, T. L. (2021). LDBlockShow: A fast and convenient tool for visualizing linkage disequilibrium and haplotype blocks based on variant call format files. *Briefings Bioinf., 22*, bbaa227. doi: 10.1093/bib/bbaa227

Dubreuil, C., Jin, X., Barajas-López, J. D., Hewitt, T. C., Tanz, S. K., Dobrenel, T., *et al*. (2018). Establishment of photosynthesis through chloroplast development is controlled by two distinct regulatory phases. *Plant Physiol., 176*, 1199–1214. doi: 10.1104/pp.17.00435

FAOSTAT. (2020). *FAOSTAT*. Available at: http://www.faostat.fao.org (Accessed April 5, 2022).

Fitter, D. W., Martin, D. J., Copley, M. J., Scotland, R. W., and Langdale, J. A. (2002). GLK gene pairs regulate chloroplast development in diverse plant species. *Plant J., 31*, 713–727. doi: 10.1046/j.1365-313X.2002.01390.x

Foyer, C. H. (2018). Reactive oxygen species, oxidative signaling and the regulation of photosynthesis. *Environ. Exp. Bot., 154,* 134–142. doi: 10.1016/j.envexpbot.2018.05.003

Frary, A., Frary, A., Daunay, M. C., Huvenaars, K., Mank, R., and Doğanlar, S. (2014). QTL hotspots in eggplant (*Solanum melongena*) detected with a high-resolution map and CIM analysis. *Euphytica*, *197*, 211–228. doi: 10.1007/s10681-013-1060-6

Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B., *et al*. (2002). The structure of haplotype blocks in the human genome. *Science*, *296*, 2225–2229. doi: 10.1126/science.1069424

García-Fortea, E., Lluch-Ruiz, A., Pineda-Chaza, B. J., García-Pérez, A., Bracho-Gil, J. P., Plazas, M., *et al*. (2020). A highly efficient organogenesis protocol based on zeatin riboside for in vitro regeneration of eggplant. *BMC Plant Biol., 20*, 1–16. doi: 10.1186/s12870-019-2215-y

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*, 1477. doi: 10.3389/fpls.2017.01477

Gramazio, P., Prohens, J., Plazas, M., Mangino, G., Herraiz, F. J., García-Fortea, E., *et al*. (2018) Genomic tools for the enhancement of vegetable crops: A case in eggplant. *Notulae Botanicae Horti Agrobotanici Cluj-Napoca, 46*, 1–13. doi: 10.15835/nbha46110936

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. doi: 10.3389/fpls.2019.01220

Gross, J. (2012). "Pigments in vegetables," in *Chlorophylls and carotenoids* (Berlin-Heidelberg: Springer-Verlag). ISBN 978146158428

Hirakawa, H., Shirasawa, K., Miyatake, K., Nunome, T., Negoro, S., Ohyama, A., *et al*. (2014). Draft genome sequence of eggplant (*Solanum melongena* L.): The representative *Solanum* species indigenous to the old world. *DNA Res*., *21*, 649–660. doi: 10.1093/dnares/dsu027

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian J. Stat., 6*, 65–70.

Jeong, H. B., Jang, S. J., Kang, M. Y., Kim, S., Kwon, J. K., and Kang, B. C. (2020). Candidate gene analysis reveals that the fruit color locus *C1* corresponds to *PRR2* in pepper (*Capsicum frutescens*). *Front. Plant Sci., 11*, 399. doi: 10.3389/fpls.2020.00399

Jia, T., Cheng, Y., Khan, I., Zhao, X., Gu, T., and Hu, X. (2020). Progress on understanding transcriptional regulation of chloroplast development in fleshy fruit. *Int. J. Mol. Sci., 21*, 6951. doi: 10.3390/ijms21186951

Jiao, J., Liu, H., Liu, J., Cui, M., Xu, J., Meng, H., *et al*. (2017). Identification and functional characterization of *APRR2* controlling green immature fruit color in cucumber (*Cucumis sativus* L.). *Plant Growth Regul*., *83*, 233–243. doi: 10.1007/s10725-017-0304-1

Katoh, K., and Hiroyuki, T. (2008). Recent developments in the MAFFT multiple sequence alignment program. *Briefings Bioinf., 9*, 286–298. doi: 10.1093/bib/bbn013

Lebeau, A., Gouy, M., Daunay, M. C., Wicker, E., Chiroleu, F., Prior, P., *et al*. (2013). Genetic mapping of a major dominant gene for resistance to *Ralstonia solanacearum* in eggplant. *Theor. Appl. Genet., 126,* 143–158. doi: 10.1007/s00122-012-1969-5

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with burrows–wheeler transform. *Bioinformatics, 25*, 1754–1760. doi: 10.1093/bioinformatics/btp324

Li, D., Lium, C. M., Luo, R., Sadakane, K., and Lam, T. W. (2015). MEGAHIT: An ultra-fast single-node solution for Large and complex metagenomics assembly *Via* succinct de bruijn graph. *Bioinformatics, 31*, 1674–1676. doi: 10.1093/bioinformatics/btv033

Li, J., He, Y. J., Zhou, L., Liu, Y., Jiang, M., Ren, L., *et al*. (2018a). Transcriptome profiling of genes related to light-induced anthocyanin biosynthesis in eggplant (*Solanum melongena* L.) before purple color becomes evident. *BMC Genomics, 19,* 1–12. doi: 10.1186/s12864-018-4587-z

Li, F., Song, X., Wu, L., Chen, H., Liang, Y., and Zhang, Y. (2018b). Heredities on fruit color and pigment content between green and purple fruits in tomato. *Scientia Hortic., 235*, 391–396. doi: 10.1016/j.scienta.2018.03.030

Li, D., Quian, J., Li, W., Yu, N., Gan, G., Jiang, Y., *et al*. (2021). A high-quality genome assembly of the eggplant provides insights into the molecular basis of disease resistance and chlorogenic acid synthesis. *Mol. Ecol. Resour., 21*, 1274–1286. doi: 10.1111/1755-0998.13321

Liu, H., Jiao, J., Liang, X., Liu, J., Meng, H., Chen, S., *et al*. (2016). Map-based cloning, identification and characterization of the w gene controlling white immature fruit color in cucumber (*Cucumis sativus* L.). *Theor. Appl. Genet., 129,* 1247–1256. doi: 10.1007/s00122-016-2700-8

Llorente, B., Torres-Montilla, S., Morelli, L., Florez-Sarasa, I., Matus, J. T., Ezquerro, M., *et al*. (2020). Synthetic conversion of leaf chloroplasts into carotenoid-rich plastids reveals mechanistic basis of natural chromoplast development. *Proc. Natl Acad. Sci. U.S.A., 117*, 21796–21803. doi: 10.1073/pnas.2004405117

Maioli, A., Gianoglio, S., Moglia, A., Acquadro, A., Valentino, D., Milani, A. M., *et al*. (2020). Simultaneous CRISPR/Cas9 editing of three PPO genes reduces fruit flesh browning in *Solanum melongena* L. *Front. Plant Sci., 11*, 607161. doi: 10.3389/fpls.2020.607161

Mangino, G., Plazas, M., Vilanova, S., Prohens, J., and 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, 467. doi: 10.3390/agronomy10040467

Mangino, G., Vilanova, S., Plazas, M., Prohens, J., and Gramazio, P. (2021). Fruit shape morphometric analysis and QTL detection in a set of eggplant introgression lines. *Sci. Horticult., 282*, 110006. doi: 10.1016/j.scienta.2021.110006

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*, 847789. doi: 10.3389/fpls.2022.847789

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. doi: 10.1016/j.ympev.2012.02.006

Mir, R., Calabuig-Serna, A., and Seguí-Simarro, J. M. (2021). Doubled haploids in eggplant. *Biology, 10,* 685. doi: 10.3390/biology10070685

Mishra, P., Tiwari, S. K., Kashyap, S. P., Tiwari, K. N., Singh, M., and 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.). *J. Plant Biochem. Biotechnol., 29*, 427–438. doi: 10.1007/s13562-020-00549-w

Miyatake, K., Saito, T., Negoro, S., Yamaguchi, H., Nunome, T., Ohyama, A., *et al*. (2016). Detailed mapping of a resistance locus against fusarium wilt in cultivated eggplant (*Solanum melongena*). *Theor. Appl. Genet., 129*, 357–367. doi: 10.1007/s00122-015-2632-8

Moglia, A., Francesco, E. F., Sergio, I., Alessandra, G., Milani, A. M., Cinzia, C., *et al*. (2020). Identification of a new R3 MYB type repressor and functional characterization of the members of the MBW transcriptional complex involved in anthocyanin biosynthesis in eggplant (*S. melongena* L.). *PloS One*, *15*, e0232986. doi: 10.1371/journal.pone.0232986

Mutegi, E., Snow, A. A., Rajkumar, M., Pasquet, R., Ponniah, H., Daunay, M. C., *et al*. (2015). Genetic diversity and population structure of Wild/Weedy eggplant (*Solanum insanum*, solanaceae) in southern India: Implications for conservation. *Am. J. Bot., 102*, 140–148. doi: 10.3732/ajb.1400403

Nguyen, C. V., Vrebalov, J. T., Gapper, N. E., Zheng, Y., Zhong, S., Fei, Z., *et al*. (2014). Tomato GOLDEN2-LIKE transcription factors reveal molecular gradients that function during fruit development and ripening. *Plant Cell*, *26,* 585–601. doi: 10.1105/tpc.113.118794

Oren, E., Tzuri, G., Vexler, L., Dafna, A., Meir, A., Faigenboim, A., *et al*. (2019). The multi-allelic *APRR2* gene is associated with fruit pigment accumulation in melon and watermelon. *J. Exp. Bot., 70,* 3781–3794. doi: 10.1093/jxb/erz182

Ospina-Calvo, B., Parapugna, T. L., and Lagorio, M. G. (2017). Variability in chlorophyll fluorescence spectra of eggplant fruit grown under different light environments: A case study. *Photochem. Photobiol. Sci., 16*, 711–720. doi: 10.1039/C6PP00475J

Page, A. M.L., Daunay, M. C., Aubriot, X., and Chapman., M. A. (2019). "The eggplant genome" in *Domestication of eggplants: A phenotypic and genomic insight* (Switzerland: Springer Nature). ISBN: 9783319992075

Pandiyaraj, P., Singh, T. H., Madhavi Reddy, K., Sadashiva, A. T., Gopalakrishnan, C., Reddy, A. C., *et al*. (2019). Molecular markers linked to bacterial wilt (*Ralstonia solanacearum*) resistance gene loci in eggplant (*Solanum melongena* L.). *Crop Prot., 124*, 104822. doi: 10.1016/j.cropro.2019.05.016

Pan, Y, Bradley, G., Pyke, K., Ball, G., Lu, C., Fray, R., *et al*. (2013). Network inference analysis identifies an *APRR2-like* gene linked to pigment accumulation in tomato and pepper fruits. *Plant Physiol., 161*, 1476–1485. doi: 10.1104/pp.112.212654

Pérez-Gálvez, A., Viera, I., and Roca, M. (2020). Carotenoids and chlorophylls as antioxidants. *Antioxidants, 9*, 505. doi: 10.3390/antiox9060505

Portis, E., Barchi, L., Toppino, L., Lanteri, S., Acciarri, N., Felicioni, N., *et al*. (2014). QTL mapping in eggplant reveals clusters of yield-related loci and orthology with the tomato genome. *PloS One, 9*, e89499. doi: 10.1371/journal.pone.0089499

Portis, E., Cericola, F., Barchi, L., Toppino, L., Acciarri, N., Pulcini, L., *et al*. (2015). Association mapping for fruit, plant and leaf morphology traits in eggplant. *PloS One, 10*, e0135200. doi: 10.1371/journal.pone.0135200

Powell, A. L. T., Nguyen, C. V., Hill, T., Cheng, K. L., Figueroa-Balderas, R., Aktas, H., *et al*. (2012). Uniform ripening encodes a golden 2-like transcription factor regulating tomato fruit chloroplast development. *Science, 336*, 1711–1715. doi: 10.1126/science.1222218

Robinson, J., Thorvaldsdóttir, H., Turner, D., and Mesirov, J. (2020). Igv.Js: An embeddable JavaScript implementation of the integrative genomics viewer (IGV). *bioRxiv, 5,* 075499. doi: 10.1101/2020.05.03.075499

Rosa-Martínez, E., Adalid-Martínez, A. M., García-Martínez, M. D., Mangino, G., Raigón, M. D., Plazas, M., *et al*. (2022). Fruit composition of eggplant lines with introgressions from the wild relative *S. incanum*: Interest for breeding and safety for consumption. *Agronomy*, *12,* 266. doi: 10.3390/agronomy12020266

Senge, M. O., Ryan, A. A., Letchford, K. A., MacGowan, S. A., and Mielke, T. (2014). Chlorophylls, symmetry, chirality, and photosynthesis. *Symmetry*, *6*, 781–843. doi: 10.3390/sym6030781

Sulli, M., Barchi, L., Toppino, L., Diretto, G., Sala, T., Lanteri, S., *et al*. (2021). An eggplant recombinant inbred population allows the discovery of metabolic QTLs controlling fruit nutritional quality. *Front. Plant Sci., 12*, 614. doi: 10.3389/fpls.2021.638195

Tanaka, A., and Tanaka, R. (2006). Chlorophyll metabolism. *Curr. Opin. Plant Biol., 9*, 248–255. doi: 10.1016/j.pbi.2006.03.011

Thissen, D., Steinberg, L., and Kuang, D. (2002). Quick and easy implementation of the benjamini-hochberg procedure for controlling the false positive rate in multiple comparisons. *J. Educ. Behav. Stat., 27*, 77–83. doi: 10.3102/10769986027001077

Toppino, L., Barchi, L., Mercati, F., Acciarri, N., Perrone, D., Martina, M., *et al*. (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*, 745. doi: 10.3390/genes11070745

Turner, S. D. (2014). Qqman: An r package for visualizing GWAS results using q-q and Manhattan plots. *bioRxiv, 10*, 005165. doi: 10.1101/005165

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. doi: 10.1186/s13007-020-00652-y

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009). Jalview version 2-a multiple sequence alignment Editor and analysis workbench. *Bioinformatics*, *25*, 1189–1191. doi: 10.1093/bioinformatics/btp033

Wei, Q., Wang, J., Wang, W., Hu, T., Hu, H., and Bao, C. (2020). A high-quality chromosome-level genome assembly reveals genetics for important traits in eggplant. *Horticulture Res., 7*, 153. doi: 10.1038/s41438-020-00391-0

Yang, G., Li, L.,Wei,M., Li, J., andYang, F. (2022). *SmMYB113* is a key transcription factor responsible for compositional variation of anthocyanin and color diversity among eggplant peels. *Front. Plant Sci., 13*, 843996. doi: 10.3389/fpls.2022.843996

Research article

The irregular fruit green netting: An eggplant domestication trait controlled by the *SmGLK2* gene with implications in fruit colour diversification

Andrea Arrones[1], Silvia Manrique[1], Joaquin Gomis-Cebolla[1], Virginia Baraja-Fonseca[1], Mariola Plazas[1], Jaime Prohens[1], Ezio Portis[2], Lorenzo Barchi[2], Giovanni Giuliano[3], Pietro Gramazio[1]*, Santiago Vilanova[1]*

[1]Instituto de Conservación y Mejora de la Agrodiversidad Valenciana, Universitat Politècnica de València, Camino de Vera 14, 46022 Valencia, Spain

[2]DISAFA, Plant Genetics and Breeding, University of Turin, Largo P. Braccini 2, 10095 Grugliasco, Italia

[3]Agenzia Nazionale Per Le Nuove Tecnologie, L'energia e Lo Sviluppo Economico Sostenibile (ENEA), Casaccia Research Centre, Via Anguillarese 301, 00123 Rome, Italia

*Corresponding authors

Ph.D. candidate contribution

A.A. had a main role in the following activities: performing the experiments, field trials, data collection, data analysis, data visualization, drafting manuscript, manuscript review and editing.

Citation: Arrones, A., Manrique, S., Gomis-Cebolla, J., Baraja-Fonseca, V., Plazas, M., Prohens, J., *et al*. (2023). The irregular fruit green netting: An eggplant domestication trait controlled by the *SmGLK2* gene with implications in fruit colour diversification. *bioRxiv,* 2023-06. doi: 10.1101/2023.06.28.546667.

Abstract

The distribution of chlorophylls in eggplant (*Solanum melongena*) peel exhibits either a uniform pattern or an irregular green netting. The latter, manifested as a gradient of dark green netting intensified in the proximal part of the fruit on a pale green background, is common in wild relatives and some eggplant landraces. Despite the selection of uniform chlorophylls during domestication, the netting pattern contributes to a greater diversity of fruit colours. Here, we have used over 2,300 individuals from different populations, including a multi-parental MAGIC population for candidate genomic region identification, an F_2 population for BSA-Seq, and advanced backcrosses for edges-to-core fine mapping, to identify *SmGLK2* gene as responsible for the irregular netting in eggplant fruits. We have also analysed the gene sequence of 178 *S. melongena* accessions and 22 wild relative species for tracing the evolutionary changes that the gene has undergone during domestication. Three different mutations were identified leading to the absence of netting. The main causative indel induces a premature stop codon disrupting the protein conformation and function, which was confirmed by Western blotting analysis and confocal microscopy observations. *SmGLK2* has a major role in regulating chlorophyll biosynthesis in eggplant fruit peel, and therefore in eggplant fruit photosynthesis.

Keywords: eggplant (*Solanum melongena*), *SmGLK2*, fruit green netting, irregular chlorophyll pattern, domestication trait, fruit colour diversification

1. Introduction

Variation in the presence or absence of chlorophylls in the eggplant (*Solanum melongena* L.) fruit peel contributes to the wide range of fruit colours in this species (Arrones *et al.,* 2022). Furthermore, the distribution of chlorophylls can be uniform or irregular, the latter being referred to as netting, reticulation, or variegation (Tigchelaar, 1968; Daunay *et al.,* 2004). The presence of the fruit green netting (FN) trait is commonly present in eggplant wild species and landraces. Cultivated eggplant differs morphologically and physiologically from its wild ancestors. Indeed, wild eggplants produce small, bitter, green-netted fruits, while domesticated eggplants are characterized by desirable agricultural traits which were directly and indirectly selected and fixed during domestication (Frary and Doğanlar, 2003; Fuller, 2007). These include big and non-bitter fruits, with a diversification of immature fruit peel colour, resulting from the combination of chlorophyll and anthocyanin pigments (Taher *et al.,* 2017; Page *et al.,* 2019; Page and Chapman, 2021). While numerous studies focus solely on the presence or absence of chlorophylls in the eggplant fruit peel, it is noteworthy that reticulated fruits often exhibit a gradient in netting density in a pale green background. This gradient often leads to a deep green colour in the proximal portion of the fruit, and gradually diminishes towards the distal end. Thus, FN contributes to further increase the colour diversity of eggplant, resulting in an irregular green or dark purple colour depending on the concurrent absence or presence of anthocyanins.

The gradient of expression of this trait is reminiscent of the green shoulder trait described in tomato, which is controlled by a single dominant gene known as the *Uniform rippening* (*U*) locus (MacArthur, 1934; Tanksley, 1992; Grandillo and Tanksley, 1996; Powell *et al.,* 2012; Nguyen *et al.,* 2014). Tomato green shoulder has been counter-selected by breeders, to improve fruit colour uniformity and facilitate maturity stage determination and harvesting (Powell *et al.,* 2012). Since tomato and eggplant display many significant similarities in their domestication syndromes (Chapman, 2019), a similar selection mechanism may have occurred in eggplant since immature fruits contain higher levels of chlorophyll in the peel.

In this study, we identified *SmGLK2* as the candidate gene controlling eggplant FN by using different experimental and germplasm populations, combining wild and cultivated genepools: (i) identification of candidate genomic region in a multi-parent advanced generation inter-cross (MAGIC) population; (ii) validation through bulked segregant analysis by sequencing (BSA-Seq) in a contrasting F_2 population; (iii) narrowing down the genomic region by fine-mapping in two advanced backcross (AB) sets; (iv) search for causative allelic variants in the candidate gene sequence by the resequencing of accessions contrasting for the

presence (FN) or absence (fn) of netting in the fruit peel; (v) tracing the changes that the gene has undergone through the domestication process in a set of accessions from a diverse germplasm collection; (vi) evaluation of the effect of the main mutation identified at the protein level by Western blotting analysis; and (vii) observation of chlorophyll autofluorescence emission from different tissues by confocal microscopy. The identification of *SmGLK2* gene as responsible for the eggplant FN trait is useful to understand the genetic basis of fruit colouration in eggplant as well as for eggplant breeding.

2. Materials and methods

2.1. Plant materials

Multi-parent advanced generation inter-cross (MAGIC) population

A total of 420 individuals from the eggplant S3MEGGIC population developed by Mangino *et al*. (2022) were used to identify candidate genomic regions associated with the FN trait. This population was previously used for identifying *SmAPRR2* as the causative gene for uniform chlorophyll pigmentation in the eggplant fruit peel (Arrones *et al.,* 2022). The S3MEGGIC population was obtained by inter-crossing seven *Solanum melongena* accession and one wild *S. incanum* INC accession (Figure 1A). Only the wild INC founder displayed the FN trait, while none of the seven *S. melongena* displayed this trait.

F_2 population

The white-fruited accession *S. melongena* MEL1 and the green netting-fruited *S. insanum* INS were used as parents to develop an F_2 population of 120 individuals segregating for the FN trait (Figure 1B). This population was used for the validation of the candidate genomic region. These parents were selected for the easier visualisation of the FN in the absence of anthocyanins (Supplementary Figure 1).

Advanced Backcross individuals and selfings

Advanced backcross (AB) individuals from two programs of development of inter-specific introgression lines (ILs) populations were used to narrow down the candidate genomic region responsible for the FN trait. One ILs population was developed by crossing one white *S. melongena* accession MEL1 and one accession of the wild *S. dasyphyllum* DAS, while the other ILs population was developed from the inter-specific cross between one *S. melongena* MEL5, which presents fruit anthocyanin (Mangino *et al.,* 2020), and the wild *S. insanum* INS (Plazas *et al.,* 2020;

Figure 1C). The two wild parents (DAS and INS) presented FN phenotype while none of the *S. melongena* (MEL1 and MEL5) had chlorophyll pigmentation in the fruit. For this study, AB individuals from the advanced stages of the ILs population development displaying the FN phenotype were selected. These AB individuals were used to confirm the FN candidate genomic region by locating the introgressed wild fragment by the single primer enrichment technology (SPET) high-throughput genotyping platform (Barchi *et al.,* 2019a). The AB individuals with the shortest introgressed fragment were selfed to find recombinants in their progeny and further narrow down the candidate region.

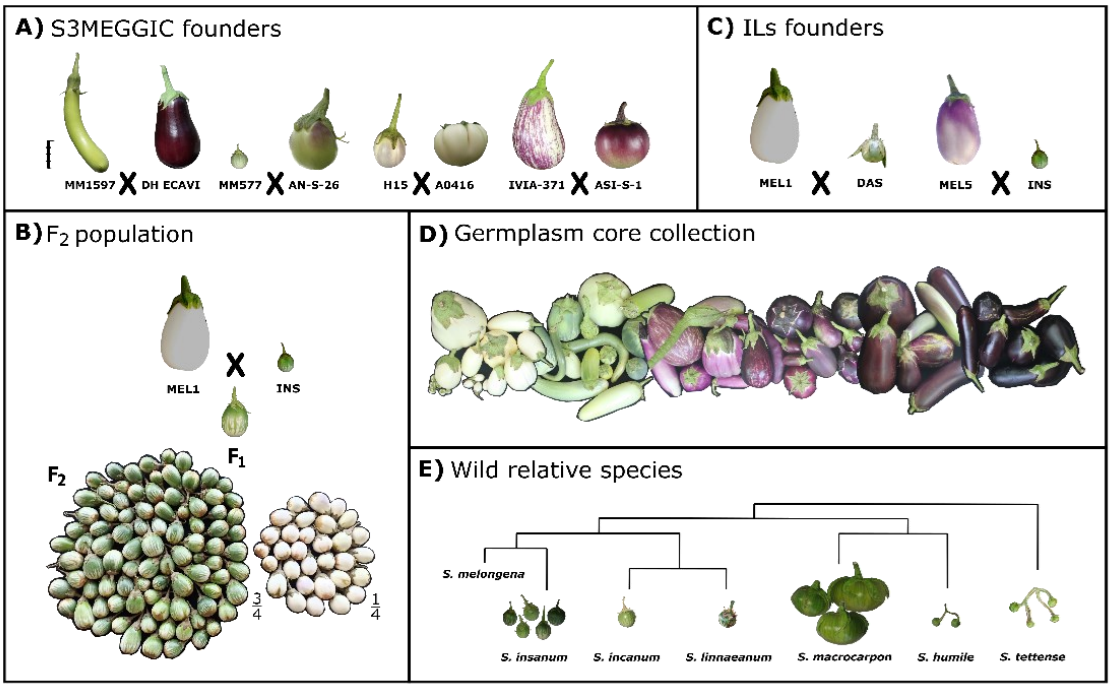


Figure 1. Plant materials used in the study. (A) The eight founders of the S3MEGGIC population developed by Mangino *et al*. (2022) that were inter-crosses in pairs to obtain the following generations. The scale bar represents 5 cm. (B) Parents, F_1 and F_2 segregating population for the fruit green netting (FN) trait. (C) Parents for the two inter-specific introgression lines (ILs) populations: *Solanum melongena* (MEL1 and MEL5), *S. dasyphillum* (DAS), and *S. insanum* (INS). (D) The variability among the selected accessions from the eggplant G2P-SOL germplasm core collection. (E) Wild relative species of the common *S. melongena* represented in a phylogenetic tree, including *S. incanum*, *S. insanum*, *S. linneanum*, *S. macrocarpon*, *S. humile*, and *S. tettense*.

Germplasm core collection and wild relative species

Genomic sequences of 178 *S. melongena* accessions (76 FN and 102 fn) available from the eggplant germplasm core collection established in the framework of the G2P-SOL project (http://www.g2p-sol.eu/G2P-SOL-gateway.html), were interrogated for the proposed candidate gene to find causative variants of the fn phenotype (Figure 1D). This core collection includes accessions used for developing the first eggplant pan-genome (Barchi *et al.,* 2021). In addition, 22 wild relative species, including *S. insanum*, *S. incanum*, *S. linneanum*, *S. macrocarpon*, *S. humile*, and *S. tettense*, all with FN phenotype, were selected to confirm the putative candidate gene structure (Figure 1E).

2.2. Cultivation conditions

Seeds from the studied individuals were germinated in Petri dishes, following the protocol developed by Ranil *et al*. (2015). They were subsequently transferred to seedling trays in a climatic chamber under photoperiod and temperature regime of 16 h light (25 °C) and 8 h dark (18 °C). After acclimatization, plantlets were grown either in a pollinator-free benched glasshouse or an open field plot at the UPV campus in Valencia, Spain (GPS coordinates: latitude, 39° 28′ 55″ N; longitude, 0° 20′ 11″ W; 7 m above sea level). Plants were spaced 1.2 m between rows and 1.0 m within the row, fertirrigated using a drip irrigation system and trained with vertical strings (greenhouse) or bamboo canes (open field). Pruning was done manually to regulate vegetative growth and flowering. Phytosanitary treatments were performed when necessary. To accelerate fruit setting, inflorescences of plants grown in the greenhouse were vibrated with a mechanical vibrator.

2.3. F_2 population development for BSA-Seq analysis

The complete F_2 population obtained by the inter-specific cross between MEL1 and INS parents was phenotyped for the FN trait and a chi-square (χ^2) test was performed to assess the goodness-of-fit to a 3:1 segregation model. This population was used for a bulked segregant analysis by sequencing (BSA-Seq). As the recommended size of each pool population is considered to be around 0.25 (Huang *et al.,* 2022), 30 individuals were selected from each of the FN and fn pools.

Young leaves were harvested from both parents and the 60 F_2 individuals selected by their different FN phenotypes. Genomic DNA was extracted using the

silica matrix (SILEX) protocol (Vilanova *et al.,* 2020) and checked for quality and integrity by agarose electrophoresis and NanoDrop (Thermo Fisher Scientific, Waltham, MA, USA) ratios (260/280 and 260/230), while its concentration was estimated with Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA). Pools were made by mixing an equal amount of DNA from each selected F_2 individual (20 ng/µl) and shipped to Novogene (Novogene Europe, Cambridge, United Kingdom) where genomic libraries (PE 150, insert 350 bp) were constructed and sequenced.

All raw reads were trimmed using the fastq-mcf tool from the Ea-utils package (Aronesty, 2013) with "q 30 -l 50" parameters, and the overall quality was checked using FastQC v.0.12.1 (Andrews *et al.,* 2010). Clean reads were mapped against the 67/3 v.3 eggplant reference genome (Barchi *et al.,* 2019b) using the BWA-MEM v.0.7.17–r1188 with default parameters (Li and Durbin, 2009). The ΔSNP-index was estimated using the QTL-seq software v.2.2.3 (Takagi *et al.,* 2013) from BAM files with "n1 30 -n2 30 -D 90 -d 5" options. The calculated ΔSNP-index were imported into R v.4.3.1 for the final plot.

2.4. Fine-mapping of the AB introgressed fragments

ILs parents (MEL1, DAS, MEL5, and INS) and the selected AB individuals were high-throughput genotyped using the 5k eggplant SPET platform (Barchi *et al.,* 2019a). The SNPs identified were filtered using the TASSEL software to retain the most reliable ones (minor allele frequency > 0.01, missing data < 20% and maximum marker heterozygosity < 80%) and the graphical visualization of the genotypes was performed by using Flapjack software (Milne *et al.,* 2010).

To further narrow down the candidate region, AB individuals with the shortest wild introgressed fragment were selfed and the progeny was genotyped by High Resolution Melting (HRM) on a LightCycler 480 Real-Time PCR (Roche Diagnostics, Meylan, France) to identify recombinants with shorter introgressions. To design primers pairs spanning the region and detect recombination breakpoints with higher precision, the genome of the four ILs parents was resequenced (PE150, insert 300 bp, 24 Gb) at the Beijing Genomics Institute (BGI Genomics, Hong Kong, China). Trimming, quality check, mapping and SNP calling were performed against the 67/3 v.3 eggplant reference genome (Barchi *et al.,* 2019b) as in Gramazio *et al*. (2019). Integrative Genomics Viewer (IGV) tool v.2.15.2 was used for the visual exploration and detection of variants among parental genome sequences (Robinson *et al.,* 2023). Primers were designed to detect differential indel variants, which are

listed in Supplementary Table 1. The informative recombinants were grown in a pollinator-free benched glasshouse and phenotyped for FN.

2.5. Candidate genes for FN and allelic variants

The genomic region narrowed down by fine mapping was explored for candidate genes, retrieved from the functional annotation of the 67/3 v.3 reference genome (Barchi *et al.,* 2019b). Simultaneously, the whole genome resequencing data of the eight founders of the S3MEGGIC populations (Gramazio *et al.,* 2019) and the four parents of the two ILs populations were interrogated for "high" impact allelic variants predicted by SnpEff (Cingolani *et al.,* 2012) that were contrasting between FN and fn. The nucleotide sequence for the best candidate gene, the eggplant Golden 2-like MYB (*SmGLK2*), was retrieved by a BLASTx search (e-value cut-off of 1e-5) against the HQ-1315 and GUIQIE-1 eggplant reference genomes (Wei *et al.,* 2020; Li *et al.,* 2021). Additionally, transcriptomic data from *S. melongena* 67/3 (SRR3884608), *S. incanum* INC (SRR2289250), and *S. insanum* MM0686 accession (SRR8736646) were downloaded from the NCBI SRA database. Transcripts were mapped against the *SmGLK2* gene sequence of the 67/3 v.3 reference genome using RNA STAR (Dobin and Gingeras, 2015) and checked using the IGV tool. Gene sequence of the tomato ortholog (*SlGLK2*, Solyc10g008160.3) was also analysed to elucidate the *SmGLK2* gene structure. A conservative domain analysis was performed by assessing the NCBI conserved domain server (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi).

Allelic variants of the *SmGLK2* gene were also interrogated in whole-genome resequencing data of 178 *S. melongena* accessions from the eggplant G2P-SOL germplasm core collection and 22 wild relative species (Gramazio *et al*., 2019, Barchi *et al.,* 2021 and unpublished data) for which FN phenotyping was available. For this purpose, raw reads were trimmed by SOAPnuke software (Chen *et al.,* 2018) with filter parameters "-l 20 -q 0.5 -n 0.03 -A 0.28" and aligned against de 67/3 v.3 eggplant reference genome (Barchi *et al.,* 2019b) using the BWA-MEM program v.0.7.17–r1188 (Li and Durbin, 2009). Subsequently, the clean reads mapped in the genomic region of *SmGLK2* were extracted with samtools (Danecek *et al.,* 2021) and de novo assembled using Megahit v.1.2.9 with default parameters (Li *et al.,* 2015). The raw data of the assembled *SmGLK2* of the G2P-SOL collection is available at NCBI SRA (BioProject IDs PRJNA649091, PRJNA392603 and PRJNA977872). A multiple sequence alignment of the assembled sequences was then performed using the MAFFT program v.7 (Katoh and Hiroyuki, 2008) and the results were visualized in the Jalview alignment editor v.2.11.2.6 (Waterhouse *et al.,* 2009). The resulting

alignments were used to identify the candidate variants in the *SmGLK2* gene for the G2P-SOL core collection and wild accessions.

2.6. Protein extraction and immunoblot analysis

Fruit peel tissues from eggplant accessions INS and MEL1 were harvested and ground with liquid nitrogen. Total proteins were extracted with extraction buffer [Tris-HCl 50 mM (pH 8), MgCl2 10 mM, glycerol 5%, Triton X-100 0.1%, β-mercaptoethanol 14mM]. Samples were centrifuged twice at 13,000 rpm for 15 min to remove cell debris and recover the cleanest supernatant containing proteins. Fourteen mg of protein were denatured at 95 ºC for 5 min in reducing loading buffer and separated by 12% sodium dodecyl sulfatepolyacrylamide gel electrophoresis (SDS-PAGE; Sambrook and Russell, 2006). Proteins were transferred (30V 1h 4 ºC) to polyvinylidene fluoride membranes (Immobilon®-E PVDF Membrane) and blocked overnight at 4 ºC with 5% non-fat dry milk (Sveltesse®, Nestlé) and gentle shacking. To detect the gene product of *SmGLK2*, membranes were incubated with an anti-sera peptide based polyclonal rabbit GLK2-antibody (1:3,000) and a peroxidase-conjugated anti-rabbit antibody (1:10,000) (Merck, Darmstadt, Germany) for 1 h at room temperature in TBST [Tris-HCl 20 mM (pH 7.6), NaCl 20 mM, Tween20 0.1% v/v] supplemented with 2% non-fat dry milk. The anti-sera peptide based polyclonal rabbit GLK2-antibodies were raised against the *SmGLK2* N-terminal sequence (synthetic peptide VSPPLSYTNENENY, 5–18 aa; and NMKSKSKEAKKSSG, 70–83 aa) (Davids Biotechnologie GmbH, Regensburg, Germany). Peroxidase activity was developed in ECL Plus Western blotting detection reagents (Amersham, Buckinghamshire, England) and the chemiluminescence signal was captured with the LAS-1000 imaging system (Fujifilm, Tokio, Japan).

2.7. Confocal microscopy

To determine differences in chloroplast number and structure between the netting and the uniform distribution of chlorophylls in the fruit peel as a result of *SmGLK2* and *SmAPRR2* expression, respectively, FN and fn tissues were analysed. Exploiting chlorophyll autofluorescence and emission spectra, confocal microscopy was used for monitoring chloroplast presence. Following the Confocal Laser Scanning Microscopy (CSLM) protocol developed by D'Andrea *et al*. (2014), thin sections of the pericarp were cut with a vegetable peeler or a single-edge razor blade. Samples were mounted on a microscope slide with a drop of water with the skin

facing the coverslip. Samples were imaged in an AxioObserver 780 (Zeiss) confocal microscope using a 488 nm argon laser and a water-immersion 40x objective, and chlorophyll emission was recorded between 634 and 723 nm. For each sample, a stack of 10 nm was taken, taking one optical section every 1 µm, as well as a lambda scan between 500 and 700 nm with a bandwidth of 8 nm and a step size of 5 nm to verify that no other pigments were emitting in this range.

3. Results

3.1. BSA-Seq for the FN trait

Through a GWAS in the S3MEGGIC population (Mangino *et al.,* 2022), Arrones *et al*. (2022) identified a significant association peak for chlorophyll content in the eggplant fruit peel on chromosome 8 which colocalized with the *SmAPRR2* gene. They also identified a minor but significant peak on chromosome 4 in the genomic region between 3.23 and 6.35 Mb (Figure 2A). However, no clear candidate gene was identified since the SnpEff software did not predict high-effect variants contrasting for the S3MEGGIC founders with and without chlorophylls in the fruit peel. The phenotyping of this population for fruit chlorophyll pigmentation was performed using a binary classification (presence/absence) without discriminating between uniform distribution or netting. This fact, together with previous studies relating this genomic region with the FN in eggplant (Tigchelaar, 1968; Doganlar *et al.,* 2002; Frary *et al.,* 2014), led us to hypothesize a possible relationship of this peak with the FN trait.

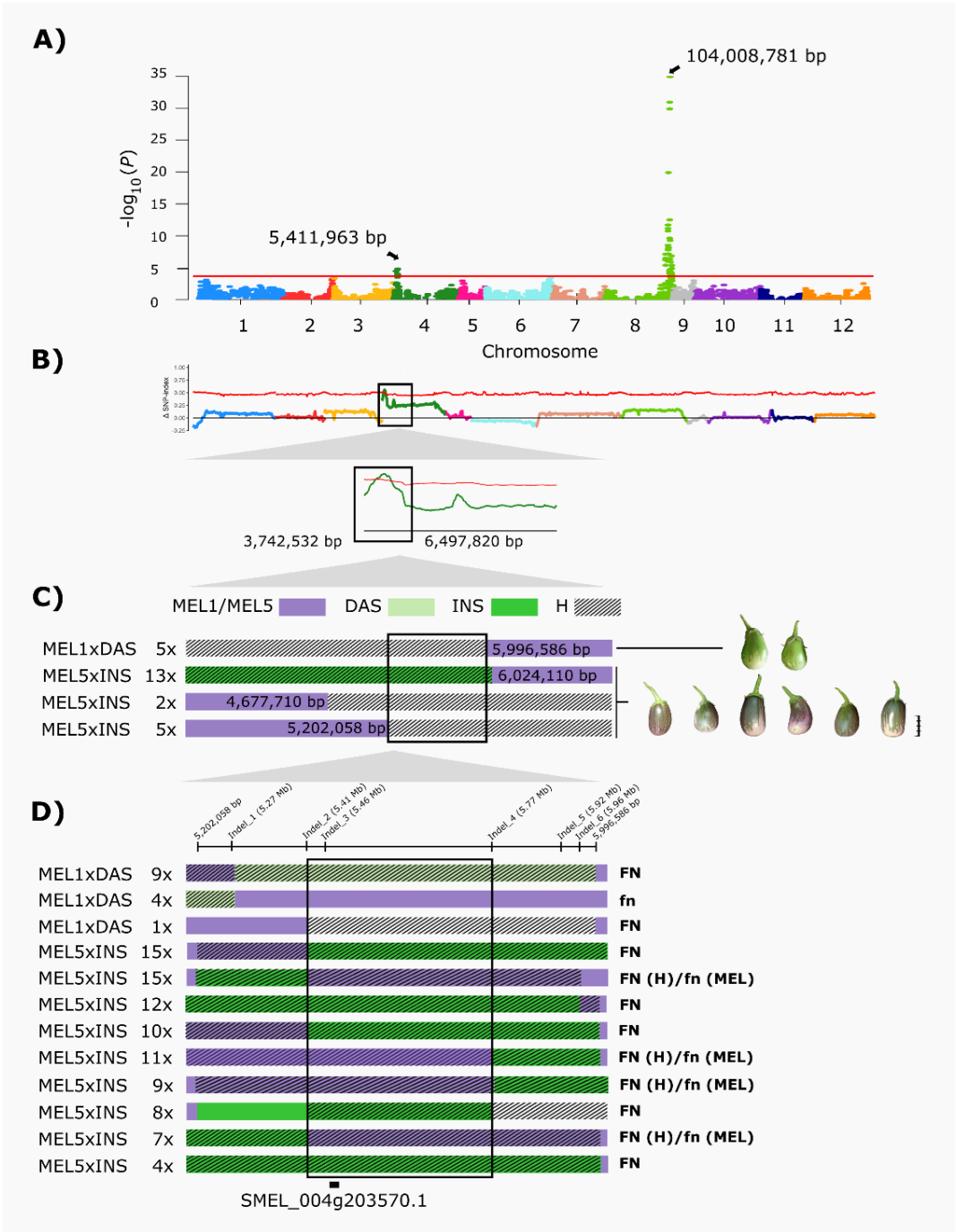


Figure 2. (A) Manhattan plot for fruit chlorophyll from the GWAS analysis in the S3MEGGIC population (Arrones *et al.,* 2022). Arrows indicate the genome position of the significant highest peaks. The horizontal red line represents the FDR

significance threshold at p = 0.05. (B) BSA-Seq results from the green netting-fruited and non-netting-fruited pools. The association significance threshold (ΔSNP-index) is represented with a red line. The black square delimits the significant candidate region. (C) Representation of the 25 selected advanced backcross (AB) individuals from two IL populations, MEL1 x DAS (5 accessions) and MEL5 x INS (20 accessions), covering the candidate region on chromosome 4. In purple the two *S. melongena* parents (MEL1 and MEL5) background, in green the introgressions coming from *S. dasyphyllum* (DAS) or *S. insanum* (INS), and with stripes the heterozygotes (H). On the right, the phenotype of the fruits obtained from the selected AB at a scale based on the real fruit size, all of them presenting the fruit green netting (FN). Scale bar represents 5 cm. (D) Representation of the 105 recombinant individuals coming from the 1,761 individuals generated by selfing of the AB individuals with the introgressed fragment in heterozygosis. The genomic indel positions used for the edges-to-core fine-mapping are specified on the top. The phenotype of the different individuals is indicated on the right. The position of the *SmGLK2* gene (SMEL_004g203570.1) is shown.

To confirm the candidate genomic region for the FN, an F_2 population was developed by crossing MEL1 (fn) and INS (FN) parents. The FN trait displayed classical patterns of Mendelian inheritance of a single dominant gene in the F_1 and F_2 populations. Among the 120 F_2 individuals, 74.17% (89 out of 120) were FN while 25.83% (31 out of 120) were fn (Figure 1C). Chi-square ($\chi 2$) analysis was in agreement with the 3:1 segregation with a value of 0.044 (d.f. = 1.0, $P < 0.05$). BSA-Seq was performed using 30 individuals from each of the FN and fn pools from the F_2 segregating population. The DNA pools sequencing resulted in an average of 117 million 150 bp raw reads yielded per bulk with a mean coverage of 33X. About 9,436,870 high-quality SNPs differentiating between parental lines were identified. Of these, 2,192,129 SNPs were selected by QTLseq software to compute the ΔSNP-index with a 99% confidence interval identifying a unique significant region of 2.76 Mb on chromosome 4, between 3,742,532 bp and 6,497,820 bp, which was consistent with the results obtained in the S3MEGGIC population (Figure 2B, Supplementary Table 2).

3.2. Fine-mapping of the candidate genomic region

Since two AB populations segregating for the FN trait were available, both were used for further inspecting the candidate genomic region identified by GWAS and BSA-seq analyses. Specifically, five AB individuals from the MEL1 × DAS set and 20 AB individuals from the MEL5 × INS were selected, all showing the FN phenotype. All of the 25 AB individuals presented a wild DAS or INS introgression

in a *S. melongena* background (MEL1 and MEL5, respectively) at the beginning of chromosome 4. The size and the physical location of the wild introgressed fragments varied according to the recombination and the selection made in the previous generation. The genotyping of these AB individuals by the eggplant SPET platform allowed to narrow down the candidate region to 0.80 Mb, between 5,202,058 bp and 5,996,586 bp (Figure 2C). There were approximately 48 putative candidate genes under this narrowed region (Supplementary Table 3).

To further delimit this region, 1,761 individuals were generated by selfing the AB individuals with the shortest wild introgressed fragment and screened following the edges-to-core approach. As a result, 105 recombinant individuals were identified (Supplementary Table 4) and based on their genotype and phenotype, the region of interest was reduced to 0.36 Mb, between 5,412,659 bp and 5,774,878 bp of chromosome 4 (Figure 2D). The number of putative candidate genes under this region was reduced to 14 genes (highlighted in Supplementary Table 3).

3.3. Identification of a responsible candidate gene

The 14 candidate genes annotated according to the 67/3 v.3 reference genome (Barchi *et al.,* 2019b) were interrogated for "high" impact allelic variants following SnpEff analysis, but none of the genes carrying such variants were consistent with the contrasting phenotypes. However, we decided to further interrogate the Golden 2-like MYB gene (*SmGLK2*, SMEL_004g203570.1, 5,457,658 – 5,461,306 bp) which was within the confidence interval, since it had previously been described as a positive regulator of chloroplast development and pigment accumulation in other Solanaceae species (Powell *et al.,* 2012; Brand *et al*, 2014; Nguyen *et al.,* 2014).

Although no high-effect variants were predicted by SnpEff in the coding sequence of *SmGLK2*, some intronic variants were shared by the FN S3MEGGIC founders and ILs parents (INC, DAS, and INS), compared to fn ones. Studying the gene structure in the other available eggplant reference genomes HQ-1315 and GUIQIE-1 (Wei *et al.,* 2020; Li *et al.,* 2021), it was observed that the *SmGLK2* gene model was split into two different genes (Smechr0400347-8 and EGP26827-8, respectively) between the second and third exon (Figure 3A). Furthermore, slight differences between *SmGLK2* gene structure compared to its tomato ortholog were identified (*SlGLK2*, Solyc10g008160.3). The *SlGLK2* 5'-UTR was divided into three different exons, the latter two coinciding with the two first exons of the *SmGLK2*. In addition, one extra exon was observed between the eggplant second and third exons (Figure 3A). Then, different transcriptomic data were analysed to reveal whether a gene mis-annotation existed in the eggplant reference genomes. When analysing the

67/3, INC and MM0686 transcriptomes against the SmGLK2 gene sequence of the 67/3 v.3 reference genome (Barchi *et al.,* 2019b), some transcripts also aligned with the intronic regions between the second and third exon (Figure 3B), suggesting an incorrect annotation in the available eggplant reference genomes, all of them developed from fn accessions. We therefore propose the presence of an extra exon between the second and third exon for the 67/3 v.3 reference genome and the merging of the two annotated genes in the HQ-1315 and GUIQIE-1 eggplant reference genomes into a single one (Wei *et al.,* 2020; Li *et al.,* 2021; Figure 3C).

3.4. Identification of causative allelic variants

In the suggested additional exon, a frameshift mutation was identified when comparing FN accessions against the 67/3 v.3 eggplant reference genome (Barchi *et al.,* 2019b), which was derived from an fn accession. This mutation consisted of a small insertion A(T)G in the 5,459,673 bp position (Figure 3C). In the presence of this insertion, a full-length GLK2 protein is generated, while in its absence, a premature stop codon leads to a truncated GLK2 protein (Figure 3D and 3E). Analysing the resequencing data, all FN parents (the wild INC, DAS, and INS) presented the T insertion, while the fn parents were identical to the 67/3 reference. The absence or presence of this insertion was also confirmed in the 67/3 (fn), INC (FN), and MM0686 (FN) transcripts.

To further study the *SmGLK2* allelic variants responsible for the loss of the FN trait over the course of domestication, a diverse germplasm collection was analysed. The aim was to trace the changes that the gene had undergone from wild to cultivated species. Therefore, we selected 178 *S. melongena* accessions from the eggplant G2P-SOL germplasm core collection, 76 FN and 102 fn, and 22 wild relative species, all of which showed the FN phenotype (Supplementary Table 5). All 76 FN accessions from the G2P-SOL core collection presented the T insertion. However, only 97 out of 102 fn accessions showed the absence of the T insertion, similar to 67/3. The remaining five accessions exhibited a different haplotype, with two new frameshift mutations: a 68 bp deletion on exon 1 from 5,457,779 – 5,457,847 bp and a 41 bp deletion on exon 5 from 5,461,264 – 5,461,305 bp. The 22 wild relative species analysed also presented the T insertion, indicating a complete and fully functional *GLK2* gene.

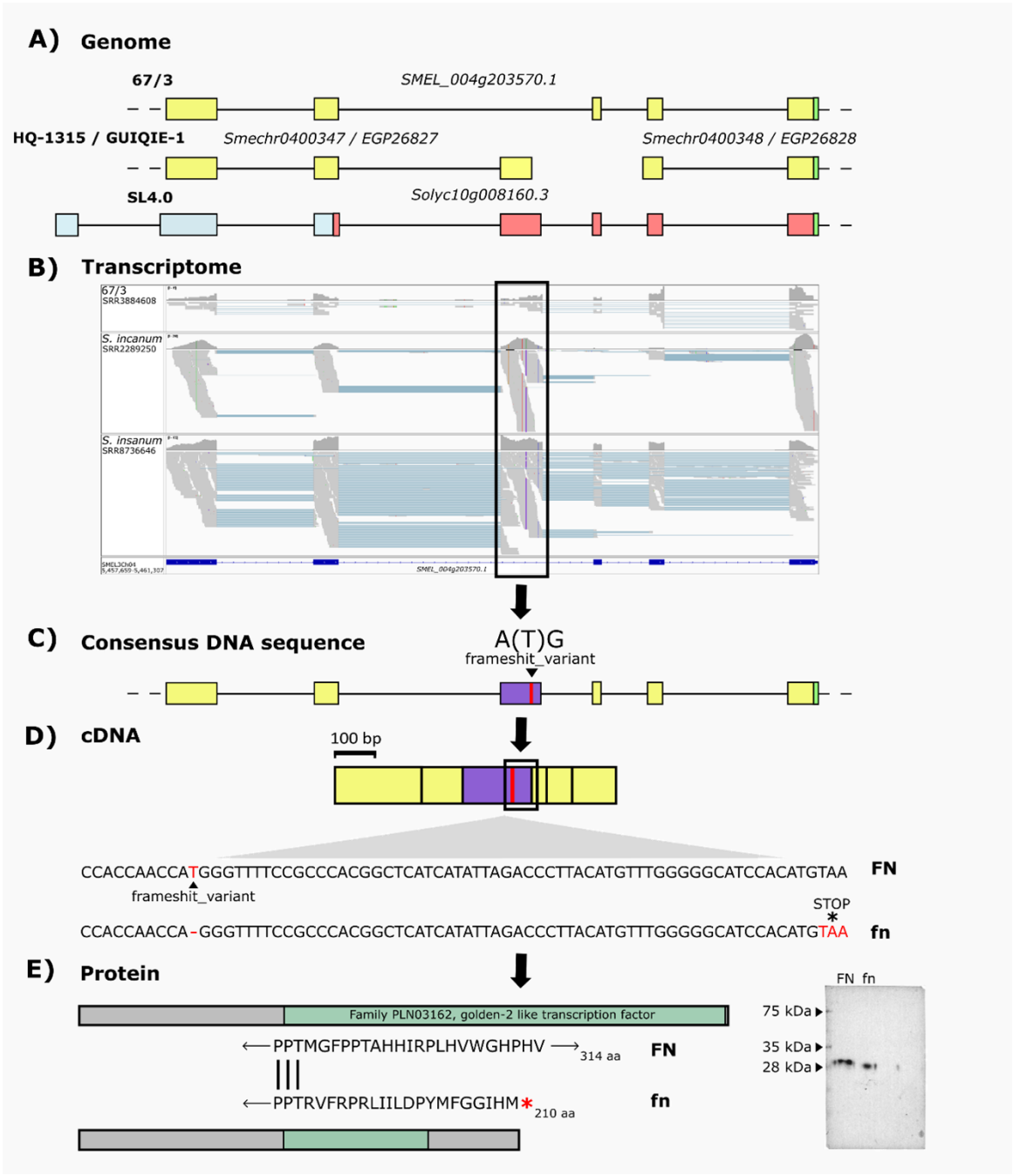


Figure 3. From the genomic sequence of the *SmGLK2* gene to protein structure. (A) *SmGLK2* annotated gene structure in 67/3, HQ-1315, and GUIQIE-1 eggplant reference genomes and SL4.0 tomato genome (5'-UTR in blue, eggplant annotated exons in yellow, tomato annotated exons in red, and 3'-UTR in green). (B) The *S.*

melongena 67/3 (SRR3884608), *S. incanum* INC (SRR2289250), and *S. insanum* MM0686 accession (SRR8736646) transcripts aligned against the *SmGLK2* gene sequence from the 67/3 v.3 eggplant reference genome and visualized in the IGV tool (Robinson *et al.,* 2023). The candidate structural variation identified is indicated with a purple line. (C) The suggested *SmGLK2* gene structural annotation as a consensus of previous information with the extra proposed exon in purple. The identified structural variation in the gene sequence is indicated with a black arrowhead and a red line. (D) Comparison of mRNA sequences for presence (FN) or absence (fn) of the fruit green netting trait. The premature stop codon downstream of the indel is indicated with an asterisk. (E) Comparison of the protein structure and sequence around the indel site for FN and fn. The golden-2 like transcription factor domain is indicated in green. On the right, a Western blot showing the difference in apparent molecular mass of the GLK2 protein from FN and fn tissues.

3.5. Differences in GLK2 protein structure

Analysis of the full-length amino acid sequence of *SmGLK2* in the NCBI conserved domain server, confirmed that the premature stop codon generated by the absence of the T insertion was translated into a truncated protein in fn accessions. Precisely, 148 aa of the golden-2-like transcription factor domain at the C-terminal region were lost. Western blot analysis with a polyclonal rabbit GLK2 antibody showed a differential molecular mass banding pattern of the target protein for INS (FN) and MEL1 (fn) samples. A difference in band migration was observed, with the apparent molecular masses of full-length and truncated versions of GLK2 being of approximately 30.5 kDa for FN and 28.75 kDa for fn, respectively (Figure 3E).

3.6. Cytological observation

Confocal microscopy was used to determine differences in chloroplast number and structure in FN and fn fruits. Fruit peel from areas displaying netting (FN) and uniform green coloration (fn) exhibited important differences (Figure 4). In the FN sample coming from the proximal part of a fruit with dark green netting, chloroplasts showed higher chlorophyll content. Instead, in the fn sample chloroplasts showed low levels of chlorophylls. Overall, our results indicate that the presence of a full-length GLK2 protein leads to the accumulation of higher amounts of chlorophylls, which results in the dark-green netted pattern in eggplant fruits.

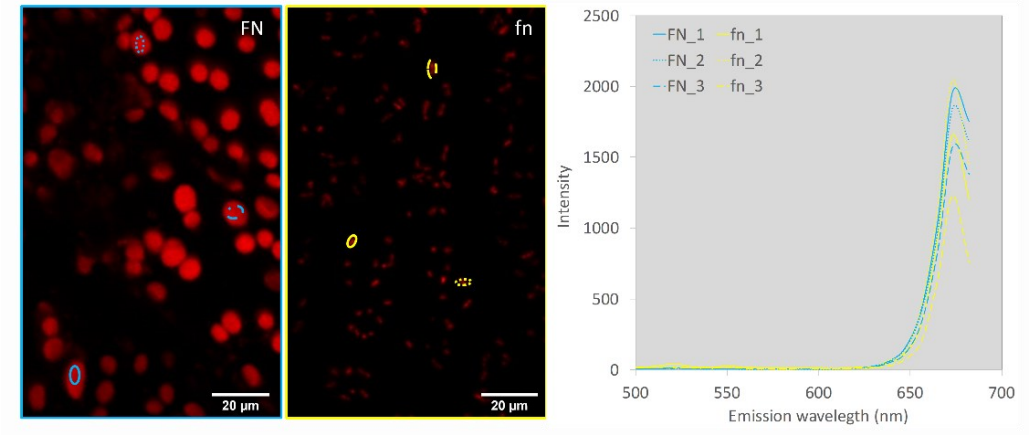


Figure 4. Confocal images of pericarp samples showing presence (FN) or absence (fn) of the fruit green netting trait. The FN sample comes from the proximal part of a fruit showing a dark green netting (*SmGLK2* expression), while the fn sample comes from a fruit with uniform distribution of chlorophylls (*SmAPRR2* expression). On the right, fluorescence emission spectra from FN and fn samples after excitation at 488 nm. Fluorescence intensity is represented relative to the total fluorescence of the sample.

4. Discussion

Eggplant domestication resulted in a diversification of fruit colour at the physiologically unripe stage, characteristic of commercial maturity. However, in recent times, modern breeding has resulted in a predominance of purple and dark purple colours, especially in Western markets (Taher *et al.,* 2017; Page *et al.,* 2019; Page and Chapman, 2021). Consumers demand new vegetable products with different and unusual aesthetic characteristics, broadening the consumers' options and fostering vegetable consumption (Di Gioia *et al.,* 2020). In this study, we have successfully harnessed different plant materials, combining germplasm and classical bi-parental (F_2) populations with advanced bi- (ILs) and multi-parent (MAGIC) populations, for the identification of the genes underlying FN in eggplant. This is interesting as it increases the diversity of the eggplant fruit colour palette.

Research on the genetics of FN in eggplants, also known as green reticulation or variegation, has been limited. Daunay *et al*. (2004) associated FN with a trait under monogenic dominant control following an F_2 proportion of 3:1. We confirmed the dominant hypothesis by the development of the F_1 hybrid and F_2 population from an FN × fn cross. In the published studies on this trait, *S. linneanum* was used as a

parental donor since FN is commonly present in this wild species (Doganlar *et al.,* 2002; Daunay *et al.,* 2004; Frary *et al.,* 2014). Although a few commercial varieties display the FN trait, it has been counter-selected during domestication because the green colour was erroneously related to immaturity (Page *et al.,* 2019). As occurs with other domestication traits, fn is recessive to the dominant allele of the wild species (Doganlar *et al.,* 2002; Frary *et al.,* 2014). This supports the paradigm that domestication usually involves the loss of gene function or regulation (Lester and Hasan, 1991).

In our previous study, through a GWAS in the S3MEGGIC population, we identified significant associations for the presence of chlorophyll pigmentation in the eggplant fruit peel (Arrones *et al.,* 2022). The Manhattan plot revealed one major peak on chromosome 8, which resulted in the identification of a gene identified as similar to ARABIDOPSIS PSEUDO RESPONSE REGULATOR2 (*APRR2*, SMEL_008g315370.1). The *SmAPRR2* gene was suggested as the best candidate gene for uniform fruit chlorophyll pigmentation. The GWAS also revealed a minor peak, although significant, at the beginning of chromosome 4. This region was rejected as SnpEff did not predict contrasting high-effect variants among the S3MEGGIC founders. Since previous studies identified a QTL related to FN on chromosome 4 explaining 67-78% of the variation (Doganlar *et al.,* 2002; Daunay *et al.,* 2004; Frary *et al.,* 2014), we decided to further investigate this genomic region. Moreover, the minor peak could have been obtained by phenotyping imprecisions as no discrimination between netting or uniform distribution of chlorophylls in the Arrones *et al*. (2022) study. These results together with a BSA-Seq of an F_2 population segregating for the trait and the fine mapping of two ABs populations, allowed us to confirm the candidate genomic region for FN and to narrow it down to 0.36 Mb. None of the 14 genes under this region presented variants consistent with the FN and fn phenotypes. However, one of these genes annotated as similar to Golden 2-like MYB (*SmGLK2*, SMEL_004g203570.1) was studied as the best candidate. This gene belongs to the widely conserved GARP family of MYB transcription factors whose functions are reported to be involved in chloroplast development in different species such as *Physcomitrella patens* (moss), *Zea mays* (maize), and *Arabidopsis thaliana* (Hall *et al.,* 1998; Rossini *et al.,* 2001; Fitter *et al.,* 2002; Yasumura *et al.,* 2005).

Through different complementary analyses, we identified an erroneous annotation of the *SmGLK2* in the available reference genomes, all of them developed from accessions with fn phenotype (Barchi *et al.,* 2019b; Wei *et al.,* 2020; Li *et al.,* 2021). An improved annotation of the *SmGLK2* gene was achieved thanks to the availability of resequencing and transcriptomic data. As a result, we found an extra coding exon where we identified a small insertion. The complete sequence of the *SmGLK2* gene is given by the presence of this insertion, while its absence results in

a frameshift mutation and a premature stop codon in the mRNA sequence. This was the reason for the fn phenotypes in our populations. This could also explain the HQ-1315 and GUIQIE-1 eggplant reference genomes annotation of *SmGLK2* as two different genes (Wei *et al.,* 2020; Li *et al.,* 2021). Furthermore, we verified the effects of this mutation at the protein level by Western blotting analysis. The presence of the small insertion resulted in a complete and fully functional *SmGLK2* gene product, while its absence resulted in a truncated golden-2-like transcription factor domain leading to a protein of smaller molecular weight.

Reverse genetics approaches, such as over-expression of the candidate genes or knock-out by CRISPR/Cas are the optimal gene validation techniques. However, due to the recalcitrance of eggplant to genetic transformation (García-Fortea *et al.,* 2020; Mir *et al.,* 2021), there is a need for alternative indirect validation methods in this crop until new, genotype-independent genetic transformation protocols become available. Validation through diverse experimental populations and the screening of large germplasm collections can provide a good alternative (Arrones *et al.,* 2022). For this reason, a set of *S. melongena* accessions from the eggplant G2P-SOL germplasm core collection and wild species, including *S. insanum, S. incanum, S. linneanum, S. macrocarpon, S. humile,* and *S. tettense*, were analysed. This methodology also allows for tracing the evolutionary changes that the gene has undergone. The complete *SmGLK2* gene sequence was confirmed in the wild relative species, all of them with FN phenotype. The small insertion identified was validated in the G2P-SOL germplasm core collection, contrasting genotyping and phenotyping data. Besides, two additional, rare frameshift mutations responsible for the fn phenotype were identified. This indicates that the lack of the FN trait has arisen and been selected independently several times during eggplant domestication and diversification.

All these results together provide strong evidence of *SmGLK2* gene as the gene controlling the FN trait. These results are also supported by previous studies performed in other Solanaceae species. In pepper, *CaGLK2* has also been described as a major gene controlling chlorophyll content and chloroplast compartment size in immature fruit (Brand *et al*, 2014). Three null mutations resulting in the appearance of premature stop codons and leading to truncated proteins were identified as responsible for a reduction in the chlorophyll content of pepper fruits. In tomato, *SlGLK2* has been related to the green shoulder trait, also referred to as the *Uniform ripening* (*U*) locus, which is very similar to FN in eggplant (MacArthur, 1934; Tanksley, 1992; Grandillo and Tanksley, 1996; Powell *et al.,* 2012). It has been demonstrated that the overexpression of *SlGLK2* increases chloroplast number and size, producing homogeneously dark-green fruits with enhanced nutritional quality (Powell *et al.,* 2012; Nguyen *et al.,* 2014). An A insertion causing a frameshift and a premature stop codon encoding for a truncated protein was identified as responsible

of the *u* phenotype (Powell *et al.,* 2012). Moreover, Nadakuduti *et al*. (2014) suggested that *SlGLK2* and *SlAPRR2* act independently as key transcription factors to directly activate genes involved in fruit chloroplast development in tomato. Likewise, the results obtained by confocal imaging of chlorophyll autofluorescence also support different roles for *SmGLK2* and *SmAPRR2* leading to two independent traits. Ultrastructural analysis of chloroplasts in netted areas of tomato fruits showed more developed chloroplasts, with more grana in the stroma than those of non-netted tomatoes (Powell *et al.,* 2012; Nguyen *et al.,* 2014). This is coherent with our observation of higher amounts of chlorophylls in eggplant netted areas. Overall, these results indicate that *SmGLK2* expression leads to more developed chloroplasts with higher chlorophyll accumulation. This may also be the reason why the green colour coming from the netting is persistent and remains throughout ripening, until turning yellow when over ripe, while chlorophylls from uniform pigmentation fade rapidly and give rise to yellow pigments (Page and Chapman*,* 2021).

By using multiple in silico and in vivo methodologies we have found that *SmGLK2* gene is responsible for the FN trait, with a 100% association between disruptive mutations in this gene and fn phenotypes. The eggplant FN has important implications for fruit visual quality particularly because eggplant is commercialized when it is still physiologically immature, and the expression of the trait is very intense in the proximal part of the fruit. The rescue of this domestication trait for future eggplant breeding programmes could be interesting to widen the diversity of the fruit colour palette. In addition, due to a higher concentration of functional chloroplasts and chlorophyll accumulation, *SmGLK2* could also be a valuable target for enhancing fruit nutritional quality

Data availability statement: The data presented in the study are deposited in the NCBI SRA repository, BioProject IDs PRJNA649091, PRJNA392603 and PRJNA977872.

Author contributions: JP, PG, and SV conceived the idea and supervised the study. AA, MP, and PG performed the field trials. AA, VB-F, PG, and SV performed the analysis of the S3MEGGIC, F_2 and AB populations and the *GLK2* gene structure. EP, LB, and GG performed the analyses of the G2P-SOL core collection. AA and SM performed confocal and protein analysis. AA and PG prepared a first draft of the manuscript and the rest of the authors reviewed and edited the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been funded by grants 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

European Union Next Generation EU/ PRTR, CIPROM/2021/020 from Conselleria d'Innovació, Universitats, Ciència i Societat Digital (Generalitat Valenciana, Spain), and by European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 677379 (G2P-SOL project: Linking genetic resources, genomes and phenotypes of Solanaceous crops). AA is grateful to Spanish Ministerio de Ciencia, Innovación y Universidades for a predoctoral (FPU18/01742) contract. VB-F is grateful to Generalitat Valenciana for grant INVEST/2022/146, funded by European Union, Next Generation EU. SM would like to thank funding support from UPV and the Spanish Ministerio de Universidades under the program María Zambrano funded by the European Union Next Generation-EU. PG is grateful to Spanish Ministerio de Ciencia e Innovación for a post-doctoral grant (RYC2021–031,999-I) funded by MCIN/AEI /10.13039/501,100,011,033 and the European Union through NextGenerationEU/PRTR.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Supplementary material: The supplementary material for this article can be found online at:

https://www.biorxiv.org/content/10.1101/2023.06.28.546667v1.supplementary-material

Supplementary Figure 1. Appearance of eggplant fruits in presence (FN) or absence (fn) of the green netting and in combination with chlorophylls or/and anthocyanins.

Supplementary Table 1. List of primers used in this study.

Supplementary Table 2. Results obtained for the ΔSNP-index calculations on chromosome 4 using the QTL-seq software. Highlighted in yellow, the significant region between 3.74-6.50 Mb.

Supplementary Table 3. List of putative candidate genes for the eggplant fruit green netting (FN) under 5.20-5.99 Mb region on chromosome 4. Genes are divided into blocks according to the primers designed for the candidate region fine-mapping. Highlighted in yellow, the 14 candidate genes under the 5.41-5.77 Mb delimited region.

Supplementary Table 4. List of *Solanum melongena* accessions from the eggplant G2P-SOL germplasm core collection and wild relative species, indicating their phenotype for the fruit green netting presence (FN) or absence (fn), and their genotype for the identified allelic variants.

Supplementary Table 5. Genotype of the 1,761 accessions used for the edges-to-core fine-mapping approach, indicating for each indel position if the individual was heterozygous (H) or homozygous for *S. melongena* (MEL1 or MEL5), *S. dasyphyllum* (DAS), or *S. insanum* (INS).

References:

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data.

Aronesty, E. (2013). Comparison of sequencing utility programs. *The Open Bioinf. J., 7,* 1–8. doi: 10.2174/1875036201307010001

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*, 1025951. doi: 10.3389/fpls.2022.1025951

Barchi, L., Acquadro, A., Alonso, D., Aprea, G., Bassolino, L., Demurtas, O., *et al*. (2019a). Single Primer Enrichment Technology (SPET) for high-throughput genotyping in tomato and eggplant germplasm. *Front. Plant Sci., 10,* 1005. doi: 10.3389/fpls.2019.01005

Barchi, L., Pietrella, M., Venturini, L., Minio, A., Toppino, L., Acquadro, A., *et al*. (2019b). A chromosome-anchored eggplant genome sequence reveals key events in Solanaceae evolution. *Sci. Rep., 9,* 11769. doi: 10.1038/s41598-019-47985-w

Barchi, L., Rabanus-Wallace, M. T., Prohens, J., Toppino, L., Padmarasu, S., Portis, E., *et al*. (2021). Improved genome assembly and pan-genome provide key insights into eggplant domestication and breeding. *Plant J., 107*, 579–96. doi: 10.1111/tpj.15313

Brand, A., Borovsky, Y., Hill, T., Rahman, K. A. A., Bellalou, A., Van Deynze, A., and Paran, I. (2014). *CaGLK2* regulates natural variation of chlorophyll content and fruit color in pepper fruit. *Theor. Appl. Genet., 127,* 2139–48. doi: 10.1007/s00122-014-2367-y

Chapman, M. A. (2019). Introduction: The importance of eggplant. *The Eggplant Genome*, Springer, Cham., 1–10. doi:: 10.1007/978-3-319-99208-2_1

Chen, Y., Chen, Y., Shi, C., Huang, Z., Zhang, Y., Li, S., *et al*. (2018). SOAPnuke: A MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *Gigascience, 7,* 1–6. doi: 10.1093/gigascience/gix120

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., *et al*. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly, 6,* 80–92. doi: 10.4161/fly.19695

D'Andrea, L., Amenós, M., and Rodríguez-Concepción, M. (2014). Confocal laser scanning microscopy detection of chlorophylls and carotenoids in chloroplasts and chromoplasts of tomato fruit. *Plant Isoprenoids: Methods in Mol. Biol., 1153*, 227–32. doi: 10.1007/978-1-4939-0606-2_16

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., *et al*. (2021). Twelve years of SAMtools and BCFtools. *Gigascience, 10*. doi: 10.1093/gigascience/giab008

Daunay, M. C., Aubert, S., Frary, A., Doganlar, S., Lester, R. N., Barendse, G., *et al.* (2004). "Eggplant (*Solanum melongena*) fruit colour: pigments, measurements and genetics" in *Proceedings of the XIIth EUCARPIA meeting on genetics and breeding of capsicum*

and eggplant (Noordwijkerhout, The Netherlands: Plant Research International), 108–116.

Di Gioia, F., Tzortzakis, N., Rouphael, Y., Kyriacou, M. C., Sampaio, S. L., CFR Ferreira, I., and Petropoulos, S. A. (2020). Grown to be blue—antioxidant properties and health effects of colored vegetables. Part II: Leafy, fruit, and other vegetables. *Antioxidants, 9*. doi: 10.3390/antiox9020097

Dobin, A., and Gingeras, T. R. (2015). Mapping RNA-seq reads with STAR. *Curr. Protoc. Bioinf., 51,* 11.14.1-11.14.19. doi: 10.1002/0471250953.bi1114s51

Doganlar, S., Frary, A., Daunay, M. C., Lester, R. N., and Tanksley, S. D. (2002). Conservation of gene function in the Solanaceae as revealed by comparative mapping of domestication traits in eggplant. *Genetics, 161*: 1713–26. doi: 10.1093/genetics/161.4.1713

Fitter, D. W., Martin, D. J., Copley, M. J., Scotland, R. W., and Langdale, J. A. (2002). *GLK* gene pairs regulate chloroplast development in diverse plant species. *Plant J., 31,* 713–27. doi: 10.1046/j.1365-313X.2002.01390.x

Frary, A., Frary, A., Daunay, M. C., Huvenaars, K., Mank, R., and Doğanlar, S. (2014). QTL hotspots in eggplant (*Solanum melongena*) detected with a high resolution map and CIM analysis. *Euphytica, 197*, 211–28. doi: 10.1007/s10681-013-1060-6

Frary, A., and Doğanlar, S. (2003). Comparative genetics of crop plant domestication and evolution. *Turkish J. Agric. For., 27*, 59–69.

Fuller, D. Q. (2007). Contrasting patterns in crop domestication and domestication rates: Recent archaeobotanical insights from the old world. *Ann. Bot., 100*, 903–24. doi: 10.1093/aob/mcm048

García-Fortea, E., Lluch-Ruiz, A., Pineda-Chaza, B. J., García-Pérez, A., Bracho-Gil, J. P., Plazas, M., *et al*. (2020). A highly efficient organogenesis protocol based on zeatin riboside for in vitro regeneration of eggplant. *BMC Plant Biol., 20.* doi: 10.1186/s12870-019-2215-y

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*. doi: 10.3389/fpls.2019.01220

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*, 935–51. doi: 10.1007/BF00224033

Hall, L. N., Rossini, L., Cribb, L., and Langdale, J. A. (1998). *GOLDEN 2*: A novel transcriptional regulator of cellular differentiation in the maize leaf. *Plant Cell, 10*, 925–36. doi: 10.1105/tpc.10.6.925

Huang, L., Tang, W., and Wu, W. (2022). Optimization of BSA-seq experiment for QTL mapping. *G3 Genes, Genomes, Genet., 12.* doi: 10.1093/G3JOURNAL/JKAB370

Katoh, K., and Toh, H. (2008). Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinf., 9*, 286–98. doi: 10.1093/bib/bbn013

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. Solanaceae III: taxonomy, chemistry, evolution. *The Linnean Society of London*, London, UK, 369–387.

Li, D., Qian, J., Li, W., Yu, N., Gan, G., Jiang, Y., *et al*. (2021). A high-quality genome assembly of the eggplant provides insights into the molecular basis of disease resistance and chlorogenic acid synthesis. *Mol. Ecol. Resour., 21*, 1274–86. doi: 10.1111/1755-0998.13321

Li, D., Liu, C. M., Luo, R., Sadakane, K., and Lam, T. W. (2015). MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics, 31*, 1674–6. doi: 10.1093/bioinformatics/btv033

Li, H., Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics, 25*, 1754–60. doi: 10.1093/bioinformatics/btp324

MacArthur, J. W. (1934). Linkage groups in the tomato. *J. Genet., 29*, 123–33. doi: 10.1007/BF02981789

Mangino, G., Arrones, A., Plazas, M., Pook, T., Prohens, J., Gramazio, P., and Vilanova, S. (2022). Newly developed MAGIC population allows identification of strong associations and candidate genes for anthocyanin pigmentation in eggplant. *Front. Plant Sci., 13*. doi: 10.3389/fpls.2022.847789

Mangino, G., Plazas, M., Vilanova, S., Prohens, J., and 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*. doi: 10.3390/agronomy10040467

Milne, I., Shaw, P., Stephen, G., Bayer, M., Cardle, L., Thomas, W. T., *et al*. (2010). Flapjack-graphical genotype visualization. *Bioinformatics, 26,* 3133–4. doi: 10.1093/bioinformatics/btq580

Mir, R., Calabuig-Serna, A., Seguí-Simarro, J. M. (2021). Doubled haploids in eggplant. *Biology, 10*. doi: 10.3390/biology10070685

Nadakuduti, S. S., Holdsworth, W. L., Klein, C. L., and Barry, C. S. (2014). *KNOX* genes influence a gradient of fruit chloroplast development through regulation of *GOLDEN2-LIKE* expression in tomato. *Plant J., 78*, 1022–33. doi: 10.1111/tpj.12529

Nguyen, C. V., Vrebalov, J. T., Gapper, N. E., Zheng, Y., Zhong, S., Fei, Z., and Giovannoni, J. J. (2014). Tomato *GOLDEN2-LIKE* transcription factors reveal molecular gradients that function during fruit development and ripening. *Plant Cell, 26,* 585–601. doi: 10.1105/tpc.113.118794

Page, A. M., Daunay, M. C., Aubriot, X., and Chapman, M. A. (2019). Domestication of eggplants: A phenotypic and genomic insight. *The Eggplant Genome*, 193–212. doi: 10.1007/978-3-319-99208-2_12

Page, A. M. L., and Chapman, M. A. (2021). Identifying genomic regions targeted during eggplant domestication using transcriptome data. *J. Hered., 112*, 519–25. doi: 10.1093/jhered/esab035

Plazas, M., Gramazio, P., Vilanova, S., Kouassi, A. B., Fonseka, R. M., Rakha, M., *et al*. (2020). Introgression breeding from crop wild relatives in eggplant landraces for adaptation to climate change. *Crop Wild Relat., 32.*

Powell, A. L., Nguyen, C. V., Hill, T., Cheng, K. L., Figueroa-Balderas, R., Aktas, H., *et al*. (2012). *Uniform ripening* encodes a *golden 2-like* transcription factor regulating tomato fruit chloroplast development. *Science, 336*, 1708–11. doi: 10.1126/science.1221863

Ranil, R. H. G., Niran, H. M. L., Plazas, M., Fonseka, R. M., Fonseka, H. H., Vilanova, S., *et al*. (2015). Improving seed germination of the eggplant rootstock *Solanum torvum* by testing multiple factors using an orthogonal array design. *Sci. Hortic., 193*, 174–81. doi: 10.1016/j.scienta.2015.07.030

Robinson, J. T., Thorvaldsdóttir, H., Turner, D., and Mesirov, J. P. (2023). igv.js: an embeddable JavaScript implementation of the Integrative Genomics Viewer (IGV). *Bioinformatics, 39*, btac830. doi: 10.1101/2020.05.03.075499

Rossini, L., Cribb, L., Martin, D. J., and Langdale, J. A. (2001). The maize *Golden2* gene defines a novel class of transcriptional regulators in plants. *Plant Cell, 13*, 1231–44. doi: 10.1105/tpc.13.5.1231

Sambrook, J., and Russell, D.W. (2006). SDS-polyacrylamide gel electrophoresis of proteins. *Cold Spring Harbor Protoc*., 281–283. doi: 10.1101/pdb.prot4540

Taher, D., Solberg, S. Ø., Prohens, J., Chou, Y. Y., Rakha, M., and Wu, T. H. (2017). World vegetable center eggplant collection: Origin, composition, seed dissemination and utilization in breeding. *Front. Plant Sci., 8*, 1484. doi: 10.3389/fpls.2017.01484

Takagi, H., Abe, A., Yoshida, K., Kosugi, S., Natsume, S., Mitsuoka, C., *et al*. (2013). QTL-seq: Rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *Plant J., 74*, 174–83. doi: 10.1111/tpj.12105

Tanksley, S. D., Ganal, M. W., Prince, J. P., De Vicente, M. C., Bonierbale, M. W., Broun, P., *et al*. (1992). High density molecular linkage maps of the tomato and potato genomes. *Genetics, 132*, 1141–60. doi: 10.1093/genetics/132.4.1141

Tigchelaar, E. C., Janick, J., and Erickson, H. T. (1968). The genetics of anthocyanin coloration in eggplant (*Solanum melongena* L.). *Genetics, 60,* 475. doi: 10.1093/genetics/60.3.475

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. doi: 10.1186/s13007-020-00652-y

Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., and Barton, G. J. (2009). Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics, 25*, 1189–91. doi: 10.1093/bioinformatics/btp033

Wei, Q., Wang, J., Wang, W., Hu, T., Hu, H., and Bao, C. (2020). A high-quality chromosome-level genome assembly reveals genetics for important traits in eggplant. *Hortic. Res., 7*. doi: 10.1038/s41438-020-00391-0

Yasumura, Y., Moylan, E. C., and Langdale, J.A. (2005). A conserved transcription factor mediates nuclear control of organelle biogenesis in anciently diverged land plants. *Plant Cell, 17*, 1894–907. doi: 10.1105/tpc.105.033191

Chapter II:
Development of a novel inter-specific tomato MAGIC population

Seeking for the unexploited diversity.

Research article

A novel tomato inter-specific (*Solanum lycopersicum* var. *cerasiforme* and *S. pimpinellifolium*) MAGIC population facilitates trait association and candidate gene discovery in untapped exotic germplasm

Andrea Arrones[1†], Oussama Antar[1†], Leandro Pereira-Dias[1], Andrea Solana[1], Paola Ferrante[2], Giuseppe Aprea[2], Mariola Plazas[1], Jaime Prohens[1], María José Díez[1], Giovanni Giuliano[2], Pietro Gramazio[1], Santiago Vilanova[1*]

[1]Instituto de Conservación y Mejora de la Agrodiversidad Valenciana, Universitat Politècnica de València, Valencia, Spain

[2]Agenzia Nazionale Per Le Nuove Tecnologie, L'energia e Lo Sviluppo Economico Sostenibile (ENEA), Casaccia Research Centre, Rome, Italy

[*]Corresponding author

[†]These authors have contributed equally to this work

Ph.D. candidate contribution

A.A. had a main role in the following activities: performing the experiments, data collection, data analysis, data visualization, drafting manuscript, manuscript review and editing.

Citation: Arrones, A., Antar, O., Pereira-Dias, L., Solana, A., Ferrante, P., Aprea, G., *et al*. (2024). A novel tomato inter-specific (*Solanum lycopersicum* var. *cerasiforme* and *S. pimpinellifolium*) MAGIC population facilitates trait association and candidate gene discovery sheds light on tomato domestication gene history. *bioRxiv,* 2024-02. doi: 10.1101/2024.02.28.582481

Abstract

We developed a novel eight-way tomato multi-parental advanced generation inter-cross (MAGIC) population to improve the accessibility of the genetic resources of tomato relatives to geneticists and breeders. The inter-specific MAGIC population (ToMAGIC) was obtained by inter-crossing four accessions each of *Solanum lycopersicum* var. *cerasiforme* (SLC) and *S. pimpinellifolium* (SP), which respectively are the weedy relative and the ancestor of cultivated tomato. The eight exotic ToMAGIC founders were selected based on a representation of the genetic diversity and geographical distribution of the two taxa. The resulting MAGIC population comprises 354 lines which were genotyped using a new 12k tomato Single Primer Enrichment Technology (SPET) panel and yielded 6,488 high-quality SNPs. The genotyping data revealed a high degree of homozygosity (average 93.69%), an absence of genetic structure, and a balanced representation (11.62% to 14.16%) of the founder genomes. To evaluate the potential of the ToMAGIC population for tomato genetics and breeding, a proof-of-concept was conducted by phenotyping it for fruit size, plant pigmentation, leaf morphology, and earliness traits. Genome-wide association studies (GWAS) identified strong associations for the studied traits, pinpointing both previously identified and novel candidate genes near or within the linkage disequilibrium blocks. Domesticated alleles for fruit size were recessive and were found, at low frequencies, in wild/ancestral populations. Our findings demonstrate that the newly developed ToMAGIC population is a valuable resource for genetic research in tomato, offering significant potential for identifying new genes that govern key traits in tomato breeding. ToMAGIC lines displaying a pyramiding of traits of interest could have direct applicability for integration into breeding pipelines providing untapped variation for tomato breeding.

Keywords: tomato, *Solanum lycopersicum* var *cerasiforme*, *Solanum pimpinellifolium*, inter-specific multi-parent advanced generation inter-cross (MAGIC).

1. Introduction

Tomato (*Solanum lycopersicum* L.) is the most economically important vegetable crop and a model plant species, with an extensive pool of genetic tools and resources. The tomato research community has access to a wealth of genetic information for wild species, landraces, and modern cultivars, including high-quality genome sequences (Rothan *et al.,* 2019). Several databases compiling genomic, genetic, transcriptomic, phenotypic, and taxonomic information are available (Fei *et al.,* 2006, 2010; Bombarely *et al.,* 2010; Suresh *et al.,* 2014; Kudo *et al.,* 2017). Over decades, several tomato bi-parental populations have also been released including introgression lines (ILs), recombinant inbred lines (RILs), advanced backcrosses (ABs), among others (e.g., Eshed and Zamir, 1995; Paran *et al.,* 1995; Tanksley and Nelson, 1996; Lippman *et al.,* 2007; Salinas *et al.,* 2013; Fulop *et al.,* 2016).

In the genomics era, new multi-parental populations have been developed dramatically increasing mapping resolution (Scott *et al.,* 2020). Multi-parent advanced generation inter-cross (MAGIC) populations are powerful next-generation pre-breeding resources with increased diversity and high recombination rates, suitable for QTL mapping and candidate gene identification (Mackay and Powell, 2007; Cavanagh *et al.,* 2008; Arrones *et al.,* 2020; Scott *et al.,* 2020). In tomato, only two MAGIC populations have previously been released. The first one was a MAGIC population developed by crossing four large-fruited *S. lycopersicum* accessions with four cherry-type accessions of *S. l.* var. *cerasiforme* (Pascual *et al.,* 2015). Final lines were used to study fruit weight distribution in the population in different environments, identifying QTLs that colocalized with already cloned genes. Subsequently, Campanelli *et al.* (2019) developed a MAGIC population that included seven cultivated accessions of tomato and one of the wild *S. cheesmaniae* as founders. The *S. cheesmaniae* accession was selected for its biotic and abiotic stress tolerance, yield and resiliency (Nesbitt and Tanskley, 2002).

The development of MAGIC populations using wild species as founders represents a promising way to combine the potential of these experimental populations for QTL/gene mapping together with the exploitation of the large phenotypic and genetic variation from the wild donor introgressions. Here, we present a novel eight-way inter-specific tomato MAGIC population (ToMAGIC) obtained by using *S. l.* var. *cerasiforme* (SLC) and *S. pimpinellifolium* (SP) accessions as founders, which respectively are the closest relative and the ancestor of cultivated tomato (Peralta *et al.,* 2008). Cultivated tomato suffered strong genetic bottlenecks during domestication and breeding processes, resulting in low genetic diversity of tomato landraces and heirlooms (Blanca *et al.,* 2015). Based on previous morphological characterization and resequencing data availability, the eight selected

founders of the new ToMAGIC population represent a wide genetic and morphological variation, as well as differences in ecological adaptation (Blanca *et al.,* 2015; Gramazio *et al.,* 2020a). Founders are very diverse in terms of fruit, vegetative, and flowering traits but also their capacity of adaptation to different conditions, ranging from desert to tropical forest environments, and from sea level to over 1,500 m altitude. Therefore, one of the aims of this population is to recover Andean variability lost during the domestication process, by using a substantial proportion of the fully cross-compatible weedy and wild tomato diversity.

This ToMAGIC population may have a large potential to identify new genomic regions and candidate genes of interest in breeding, as well as to validate genes and QTLs already described in a genetic background other than that of cultivated tomato. In this way, another aim of this population is dissecting the control of different traits, including those involved in the early domestication of tomato (Frary and Doganlar, 2003). The introduction of exotic germplasm will be useful for shedding light on the genetics of agronomic and adaptation traits present in these materials, as well as for the selection of elite lines of interest for tomato breeding (Arrones *et al.,* 2020). In our work, the integration of high-throughput genotyping of the recombinant ToMAGIC population together with the phenotyping of specific traits across different plant parts has effectively demonstrated a proof-of-concept for the high-precision fine mapping of these traits. This approach has not only validated previously identified candidate genes for the traits studied in a SLC and SP genetic background, but also led to the discovery of new candidate genes, and the observation of additional phenotypic-causing variants, underscoring the great potential of the ToMAGIC population for tomato genetics and breeding.

2. Materials and methods

2.1. ToMAGIC founders

The inter-specific tomato MAGIC (ToMAGIC) population was developed through the inter-crossing of SLC and SP accessions. Founders consist of four weedy *S. l.* var. *cerasiforme*, i.e., BGV007931 (SLC1), LA2251 (SLC2), PI487625 (SLC3), and BGV006769 (SLC4), and four wild *S. pimpinellifolium,* i.e., BGV007145 (SP1), BGV006454 (SP2), BGV015382 (SP3), and BGV013720 (SP4). Their geographical origin, including geographical coordinates and altitude, and environmental parameters (mean temperature, temperature range, precipitation, etc.) are known (Martínez-Cuenca *et al.,* 2020). With respect to the Heinz 1706 SL4.0 reference genome (Hosmani *et al.,* 2019), the total variants identified in SLC accessions ranged

from 1.2 million in SLC2 to 1.9 million in SLC1, while in the SP accessions, they ranged from 3.1 million in SP4 to 4.8 million in SP3 (Gramazio *et al.,* 2020a). This set of variants was over 1,600-fold more abundant than the one used in the previous study of Blanca *et al.* (2015), where the eight founders were also genotyped.

2.2. ToMAGIC population development

Although low heterozygosity levels were observed for founders in previous studies (Blanca *et al.,* 2015), before starting with the ToMAGIC population cross-design, two generations of selfing of the founders were performed to ensure high homozygosity. To develop the ToMAGIC population, founder lines were inter-crossed by following a "funnel" approach including two extra generations of inter-crosses among the offspring of the double hybrid crosses. These extra steps were performed to increase recombination events among the genomes of the eight founders during the population development to achieve better mapping and QTL identification resolution (Arrones *et al.,* 2020). The first step in developing the MAGIC population consistent in crossing the SLC parents with the SP ones to produce inter-specific F_1 hybrids (SLC1 × SP2, SLC2 × SP1, SLC3 × SP4, and SLC4 × SP3). These F_1 hybrids were subsequently inter-crossed in pairs (SLC1 × SP2 with SLC2 × SP1, and SLC3 × SP4 with SLC4 × SP3) directly (') and reciprocally ('') to obtain four genetically segregating double hybrids (DHY1', DHY1'', DHY2', and DHY2''). In this way, genomes from both species were mixed since the beginning of the development of the MAGIC population. Then, DHY1' or DHY1'' individuals were crossed with DHY2' or DHY2'' individuals obtaining a set of materials coming from the first inter-cross generation (IC1), which were an admixture of the genomes of the eight founders. DHYs were crossed by following a chain pollination scheme, where each individual was used as a female and male parent of different crosses (Díez *et al.,* 2002; Mangino *et al.,* 2022). Initially, the first DHY1 line served as the male parent in an inter-cross with the first DHY2 line, which acted as the female parent. This pattern continued with the roles reversed: the first DHY2 line then served as the male parent in an inter-cross with the second DHY1 line as the female. This alternating pattern persisted until the final stage, where the last DHY2 line was used as the male parent in an inter-cross with the first DHY1 line, again acting as the female (Supplementary Figure 1). In the same way, individuals from the second inter-cross (IC2) generation were also inter-crossed following a chain pollination scheme. This step was repeated to obtain the individuals from the third inter-cross generation (IC3). Finally, progenies of the IC3 were selfed for five generations by single seed descent (SSD) to obtain the ToMAGIC recombinant inbred lines. To accelerate the obtention of the SSD generations, selfings were stimulated by

mechanical vibration and pruning was done manually, regulating vegetative growth and flowering. A set of 354 ToMAGIC lines were used in this study for phenotyping and genotyping.

Seeds from the 354 ToMAGIC lines were germinated in seedling trays with Humin-substrat N3 substrate (Klasmann-Deilmann, Germany) in a climatic chamber under a photoperiod and temperature regime of 16 h light (25 ºC) and 8 h dark (18 ºC). Plantlets were subsequently transplanted to individual thermoformed pots (1.3 l capacity) for acclimatisation and grown in a pollinator-free glasshouse of the Universitat Politècnica de València (UPV, Valencia, Spain). Plants were fertirrigated using a drip irrigation system and trained with vertical strings. Phytosanitary treatments against whiteflies and *Tuta absoluta* were performed when necessary.

2.3. High-throughput genotyping

Young leaf tissue was sampled from the 354 ToMAGIC lines. Genomic DNA was extracted using the SILEX extraction method (Vilanova *et al.,* 2020). DNA quality and integrity were checked by agarose electrophoresis and NanoDrop ratios (260/280 and 260/230), while its concentration was estimated using a fluorescent DNA intercalating agent (e.g., Quant-iT PicoGreen dsDNA Assay Kit, Thermo Fisher Cat. No. P7589) and a microplate reader (Thermo Fisher Scientific)). Samples were sent to IGATech company (Udine, Italy) for library preparation and sequencing (150 paired-end) for a high-throughput genotyping using a newly developed 12k probes tomato Single Primer Enrichment Technology (SPET) panel, which is considerably improved over the original 5k probes tomato set (Barchi *et al.*, 2019). The new SPET panel comprises 12,000 probes and was developed by selecting the most informative and reliable polymorphisms (of which ~11,500 within 100 nt of a gene and ~500 in intergenic regions) (Aprea *et al*., in preparation).

Cleaning of raw reads was performed using Fastp (Chen, 2023). Clean reads were mapped onto the tomato reference genome Heinz 1706 SL4.0 (Hosmani *et al.,* 2019) using BWA-MEM (Li, 2013) with default parameters, and only the uniquely aligned reads were selected for variant calling with GATK 4.0 (DePristo *et al.,* 2011), following the best practices recommended by the Broad Institute. The SNPs identified by the tomato SPET panel were first filtered by coverage ≥ 10 and quality GQ ≥ 20, removed the monomorphic sites using bcftools (Danecek *et al.,* 2021), and then filtered using the TASSEL software (ver. 5.0, Bradbury *et al.,* 2007) to retain the most reliable ones (minor allele frequency > 0.01, missing data < 0.1, and maximum marker heterozygosity < 0.7). In addition, a linkage disequilibrium (LD) k-nearest neighbour genotype imputation method (LD KNNi) was performed to fill

the missing calls or genotyping gaps (Troyanskaya *et al.,* 2001). Final marker density along chromosomes was represented using the R package chromPlot (Oróstica and Verdugo, 2016).

2.4. Population diversity analysis

A principal component analysis (PCA) was performed to assess the population structure of the MAGIC population. PCA scores were generated in TASSEL software (ver. 5.0, Bradbury *et al.,* 2007). For graphically plotting the final PCA results the R package ggplot2 was used (Wickham, 2016). A heat map of the kinship matrix to identify possible relationships between lines was generated with GAPIT software (v.3, Wang and Zhang, 2021). A dendrogram of the MAGIC population was generated using the neighbor-joining method (Saitou and Nei, 1987) and the graphical representation was displayed and edited using the iTOL v.4 software (Letunic and Bork, 2019) to evaluate the genetic similarities among ToMAGIC lines and founders. Parental contribution to the ToMAGIC lines and haplotype blocks was estimated by using the R package HaploBlocker (Pook *et al*., 2019).

2.5. ToMAGIC phenotyping

A proof-of-concept for testing the potential of the MAGIC population for GWAS analysis and detection of genomic regions associated with different types of traits was performed by phenotyping the eight parents and the 354 ToMAGIC lines for a set of traits from different plant organs. The traits evaluated included two related to fruit size (fruit locule number and fruit weight), one to plant pigmentation (plant anthocyanin), two to leaf morphology (lobing/serration and leaf complexity), and one to earliness (number of leaves below the first inflorescence). Tomato fruits evaluated for fruit weight and cut transversally for locule number counting. Presence of plant anthocyanin was observed in vegetative plant parts (stem, branches, leaf veins or leaf area) and scored in a range from 0 (slight presence, mainly on the stem) to 4 (strong presence in all plant parts). Leaf lobing/serration was scored in a range from 1 (lack of lobing/serration) to 7 (very serrated leaf). Leaf complexity was screened using a binary classification for pinnate (0) and bipinnate (1) compound leaves. The number of leaves below the first inflorescence was recorded by counting the leaves of the primary shoot when the first flower bud was visible. Pearson pair-wise coefficient of correlation (r) among traits was calculated, and their significance was assessed using a Bonferroni correction at the $p<0.05$ probability level

(Hochberg, 1988) using R packages psych (Revelle, 2007) and corrplot (Wei and Simko, 2017).

2.6. Genome-Wide Association Study (GWAS)

Using the genotypic and phenotypic data collected from the ToMAGIC lines, GWAS analyses were performed for the selected traits using the GAPIT software (v.3, Wang and Zhang, 2021). General linear model (GLM), mixed linear model (MLM), and BLINK analyses were conducted for the association study (Price *et al.,* 2006; Yu *et al.,* 2006; Huang *et al.,* 2019). Comparison of models was displayed in roundness Manhattan plots. The multiple testing was corrected with the Bonferroni and the false discovery rate (FDR) methods (Holm, 1979; Benjamini and Hochberg, 1995) with a significance level of 0.05 (Thissen *et al.,* 2002). Bonferroni threshold is defined as the -log10 of the desired overall alpha level ($\alpha = 0.05$) divided by the total number of SNPs. Therefore, it remains constant among the different association models. Meanwhile, FDR threshold values are retrieved by adjusting p-values to control the proportion of false positive. Thus, FDR thresholds vary among models and traits. SNPs with a limit of detection (LOD) score (calculated as -log10[p-value]) exceeding these specified thresholds or cut-off values in the three GWAS models were considered significantly associated with the traits under evaluation. Associations were considered significant if the same SNP exceeded the cut-off thresholds in at least two of the implemented models, indicating robustness. The top significant SNPs delimited the candidate genomic regions. All markers within these genomic regions were used to calculate the correlation coefficient (r^2). SNPs with default r^2 values greater than 0.5 were considered for haplotype block estimation. The R package geneHapR was used for haplotype statistics (Zhang *et al.,* 2023a). The genes underlying the haplotype blocks were retrieved from the Heinz 1706 SL4.0 tomato reference genome (Hosmani *et al.,* 2019). Candidate genes were assessed by SnpEff software v 4.2 prediction (Cingolani *et al.,* 2012) of the eight MAGIC founders' resequencing data (Gramazio *et al.,* 2019) in order to identify causative mutations contrasting for different phenotypes. The Integrative Genomics Viewer (IGV) tool was used for the visual exploration of founder genome sequences to validate SnpEff results (Robinson *et al.,* 2023). A conservative domain analysis was performed using the NCBI conserved domain server (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) to assess the predicted variants at the protein level. The BLASTp (e-value cut-off of $1e^{-10}$) alignment tool and EnsemblPlants browser were used to compare the homology of protein sequences encoded by genes belonging to the same gene family. Haplotype and phenotype boxplots and density plots were generated with the R package ggplot2

(Wickham, 2016). To assess the significance of differences among different haplotypes pairwise *t*-tests were performed. Density plots represent the density of the data at each value of x, allowing peak values to potentially exceed 1, particularly when data are densely concentrated around specific values.

3. Results

3.1. MAGIC population construction

In the first stage of MAGIC population development, SLC and SP accessions of different origins (Figure 1A) were inter-crossed pairwise (Figure 1B). These materials are native to different geographic regions of South and Central America, mainly from Ecuador and Northern Peru, and provide a representation of the Andean variability lost during the domestication process in Mesoamerica (Figure 1A). They were selected since they are considered genetic diversity reservoirs barely exploited in tomato breeding (Gramazio *et al.,* 2020a). They include a wide molecular variability and phenotypic diversity in plant and inflorescence architecture, leaf, flower, and fruit traits, together with resistance or tolerance (in some of the founders) to biotic and abiotic stresses (Blanca *et al.,* 2015), including water and salt stress adaptation (Martínez-Cuenca *et al.,* 2020). The eight founders have previously been characterized morphoagronomically and their genomes have been resequenced (Blanca *et al.,* 2015; Gramazio *et al.,* 2020a).

These weedy (SLC) and wild (SP) tomato species are cross-compatible (Gramazio *et al.,* 2020a), and thus the manual inter-cross was successfully performed. As a result of the inter-cross of the eight founders, the F_1 hybrids, and the DHY hybrids, 112 IC1 individuals were obtained. The subsequent inter-crossing following a chain pollination scheme resulted in the obtention of 232 IC2 and 481 IC3 individuals. The latter individuals were self-pollinated to produce 475 S1, 452 S2, 427 S3, 400 S4, and the final population of 354 S5 (ToMAGIC) lines (Figure 1B).

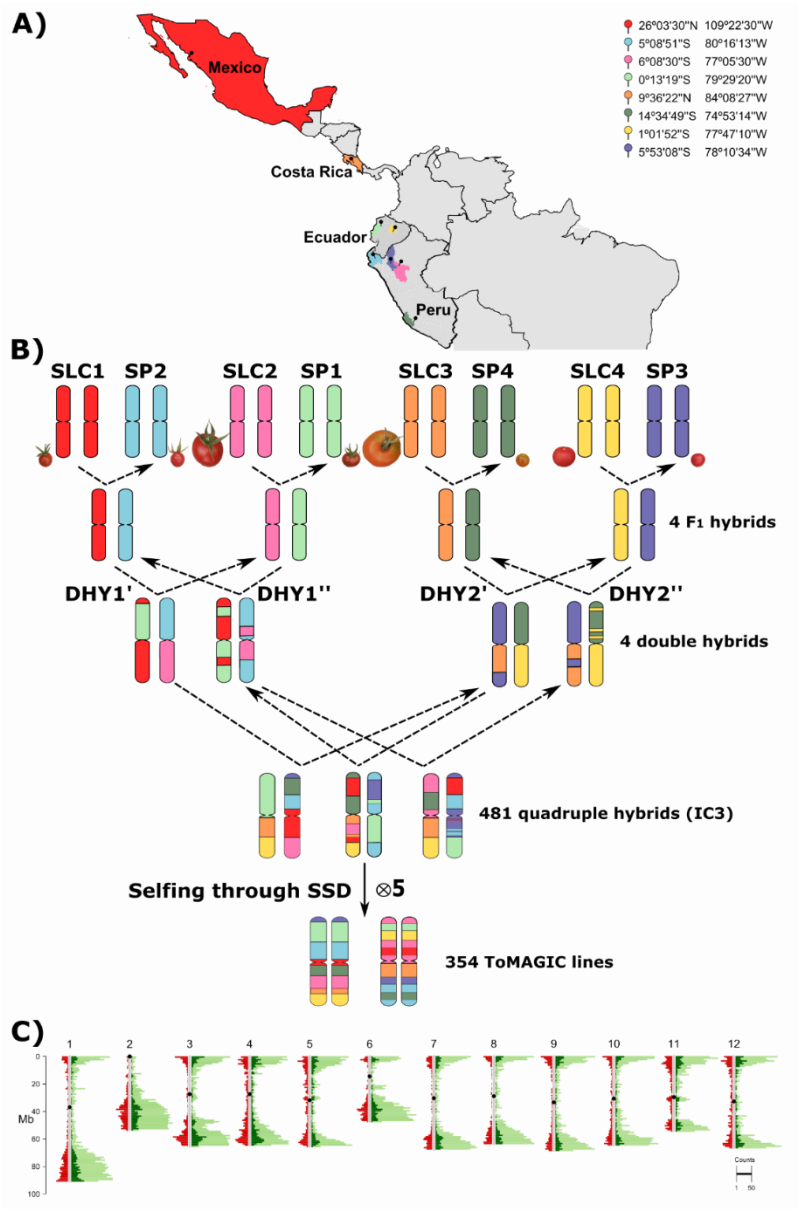


Figure 1. (A) Origin of the different SLC and SP founders selected for the ToMAGIC population development represented with the different colours code. (B) The funnel breeding design to develop the 354 ToMAGIC lines. The eight founders with a different colour to represent their genomic background, are represented at a scale based on the real fruit size. Arrows indicate the direction of the cross. (C) Distribution of the 6,488 filtered markers (in red), the Heinz 1706 SL4.0 annotated genes (in light green), and the genes covered by the filtered markers (in dark green) across the 12 tomato chromosomes.

3.2. Genotyping

A total of 4,268,587 SNPs were generated from the genotyping of the 354 ToMAGIC lines using the newly developed 12k probes tomato SPET panel. After filtering, 6,488 markers were retained for the subsequent GWAS analysis. A higher marker density was observed in gene-rich regions located outside pericentromeric regions (Figure 1C). However, the distribution of SNPs along the different tomato chromosomes was fairly uniform, with an average marker density of 8.51 per Mb (Table 1). This ratio is an average of the whole chromosome including the centromere where recombination and gene density are extremely low as observed in Figure 1C. Excluding centromeric regions, marker density in the euchromatic regions increased to 17.95 markers per Mb, which is equivalent to almost two markers per 100 kb. The filtered markers cover 16.91% of the total annotated genes. The residual heterozygosity of the ToMAGIC lines was on average 6.31%.

Table 1. Chromosome-wide distribution of the SNP positions used for the genome-wide association study (GWAS) in the tomato MAGIC population.

Chr	Markers	% Markers	Chr length (Mb)	Marker density (markers/Mb)	Marker density euchromatic regions (markers/Mb)	Genes	Covered genes
1	646	9.96	90.86	7.11	18.15	4,133	619
2	568	8.75	53.47	10.62	20.58	3,379	518
3	624	9.62	65.30	9.56	20.37	3,324	578
4	815	12.56	64.46	12.64	23.15	2,819	689
5	656	10.11	65.27	10.05	17.61	2,382	496
6	408	6.29	47.26	8.63	16.22	2,769	392
7	425	6.55	67.88	6.26	14.22	2,517	379
8	346	5.33	64.00	5.41	12.32	2,428	315
9	436	6.72	68.51	6.36	17.87	2,521	403
10	406	6.26	64.79	6.27	14.11	2,520	342
11	552	8.51	54.38	10.15	19.31	2,326	446
12	606	9.34	66.69	9.09	21.48	2,444	498
Total	6,488	100	772.87			33,562	5,675
Average		8.33		8.51	17.95		

3.3. Population structure

A lack of genetic structure in ToMAGIC population was supported by the Principal Component Analysis (PCA), in which no differentiated groups were observed (Figure 2A). The first two PCs accounted only for 3.40% of the genetic

variance, the first ten PCs 9.93%, and it required 41 PCs to explain 20% of the genetic variation, underscoring the weak population structure of the population. In addition, kinship coefficients between pairs of ToMAGIC lines varied from 0 to 1.32 (on a scale of 0 to 2), with 98.35% of the pairs with kinship values <0.5 (Figure 2B). These results revealed a low genetic relatedness among ToMAGIC lines.

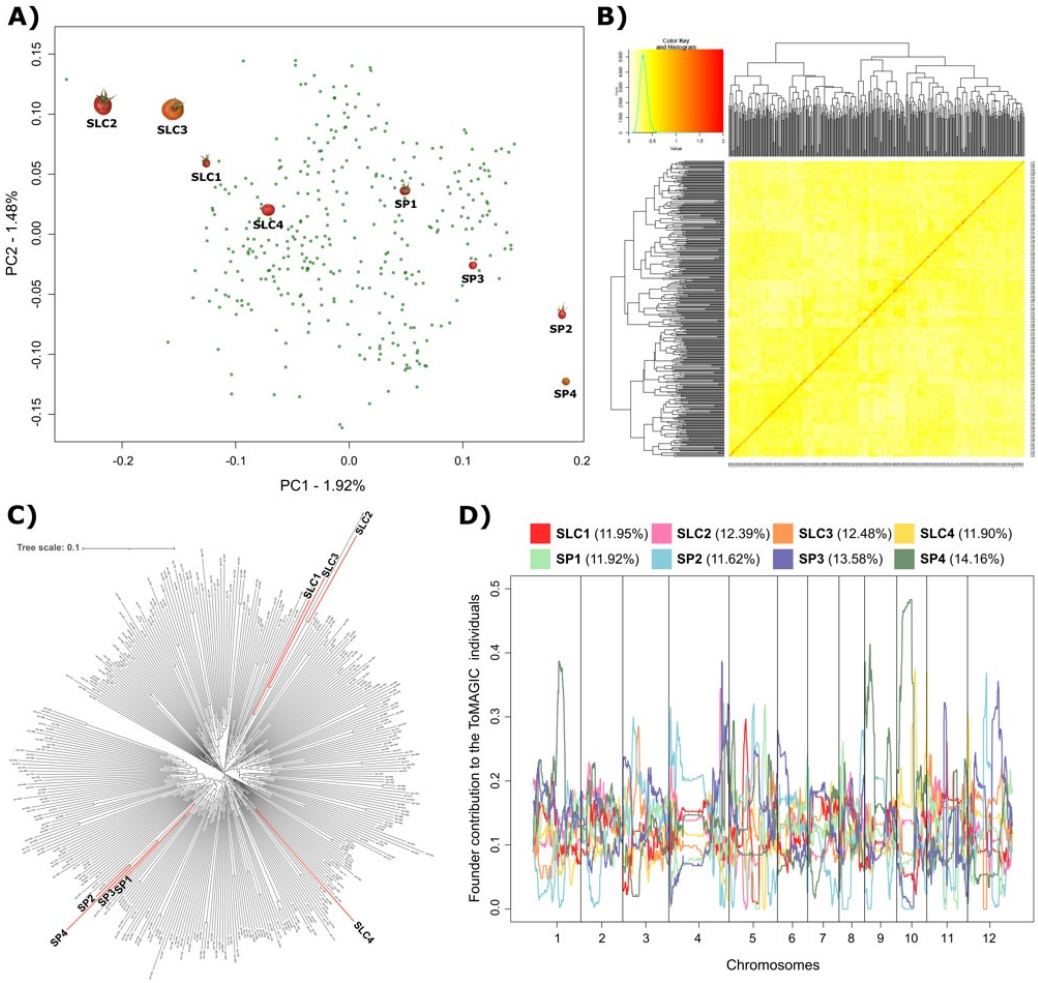


Figure 2. Population structure of the inter-specific ToMAGIC population. (A) Principal component analysis (PCA) plot of the first two PCs. (B) Heatmap plot of genetic relationship based on the kinship matrix. (C) Dendrogram indicating founders' locations with coloured red branches. (D) Genome-wide founder haplotype blocks assignment across the 12 tomato chromosomes (x-axis) as the average percentage of founders' contribution to the ToMAGIC lines (y-axis) with a different colour associated with each founder.

SLC founders were grouped close together, with negative values of the PC1, while SP founders had positive values for the PC1 (Figure 2A). A similar grouping was observed in the dendrogram of the MAGIC population and founders (Figure 2C). SLC2 and SLC3 are the closest accessions to cultivated tomato and plot in the first PCA quadrant with low values for the PC1 and high for the PC2. SLC4 is the closest to SP founders in the PCA (Figure 2A) and is separated from the rest of SLC founders in the dendrogram (Figure 2C). The estimated average contribution of each founder to the overall population was around the theoretically expected value of 12.50%, with the range varying from 11.62% for SP2 to 14.16% for SP4. However, the reconstruction of genome mosaics for the 354 ToMAGIC lines, considering the eight founder haplotypes, revealed different haplotype block proportions at different chromosomal positions (Figure 2D).

3.4. Phenotyping analysis

Phenotyping for locule number, fruit weight, plant anthocyanin pigmentation, leaf lobing/serration, leaf complexity, and number of leaves below the first inflorescence revealed a wide range of variation, including transgressive lines for some of the studied traits (Table 2, Supplementary Figure 2). For the locule number trait, the average for SP founders was 2 lobules, while the average for SLC was 2.75 and the range between 2 and 4. However, ToMAGIC lines with up to 5 and 6 locules were identified, although most of the lines only had 2 locules, resulting in an average value of 2.2. For the fruit weight, ToMAGIC lines showed an intermediate average (2.72 g) between the SP and SLC founders weight averages of 1.60 g and 4.97 g, respectively. However, the range of variation of the founders was greater (from 0.97 g to 11.59 g) than those of the ToMAGIC lines (0.44 to 7.01), and no lines were found with a higher weight than the heaviest founder (SLC3). For the plant anthocyanin pigmentation, the mean of SLC founders (0.50) was lower than that of the SP founders (1.25), mainly due to the high level of plant pigmentation of the SP4 founder. The range of variation was greater for the ToMAGIC lines (from 0 to 4) than for the founders (from 0 to 3). For the leaf lobing/serration, ToMAGIC lines showed an intermediate average (3.69 g) between the SP and SLC founders averages of 2.50 and 6, respectively. The ToMAGIC lines covered all the variation range found in the founders, from the lack of lobing/serration (1) to very serrated leaves (7). For the leaf complexity, ToMAGIC lines showed an intermediate average (0.26) between the SP (0) and SLC (0.50) founders. For the number of leaves below the first inflorescence, the SP founders had a slightly lower number (4.33) than SLC founders (6.66), while ToMAGIC lines had an average of 5.36 leaves. However, the range of variation was much larger for the ToMAGIC lines (from 4 to 10) than for the founders (from 4 to 7). Pearson pairwise correlations among the trait's evaluation

were conducted, and only a slight positive correlation (r = 0.3261; p=1.57e^{-7}) between leaf lobing/serration and leaf complexity, was observed (Figure 3B).

Table 2. Means and range values for SLC and SP founders, and ToMAGIC lines for the phenotypic traits evaluated.

Trait	SLC		SP		ToMAGIC lines	
	Average	Range	Average	Range	Average	Range
Locule number	2.75	2 – 4	2	2	2.20	2 – 6
Fruit weight (g)	4.97	1.61 – 11.59	1.60	0.97 – 2.89	2.72	0.44 – 7.01
Plant anthocyanin	0.50	0 – 1	1.25	0 – 3	0.94	0 – 4
Leaf lobing/serration	6	5 – 7	2.50	1 – 3	3.69	1 – 7
Leaf complexity	0.50	0 – 1	0	0	0.26	0 – 1
Number of leaves below the first inflorescence	6.66	6 – 7	4.33	4 – 5	5.36	4 – 10

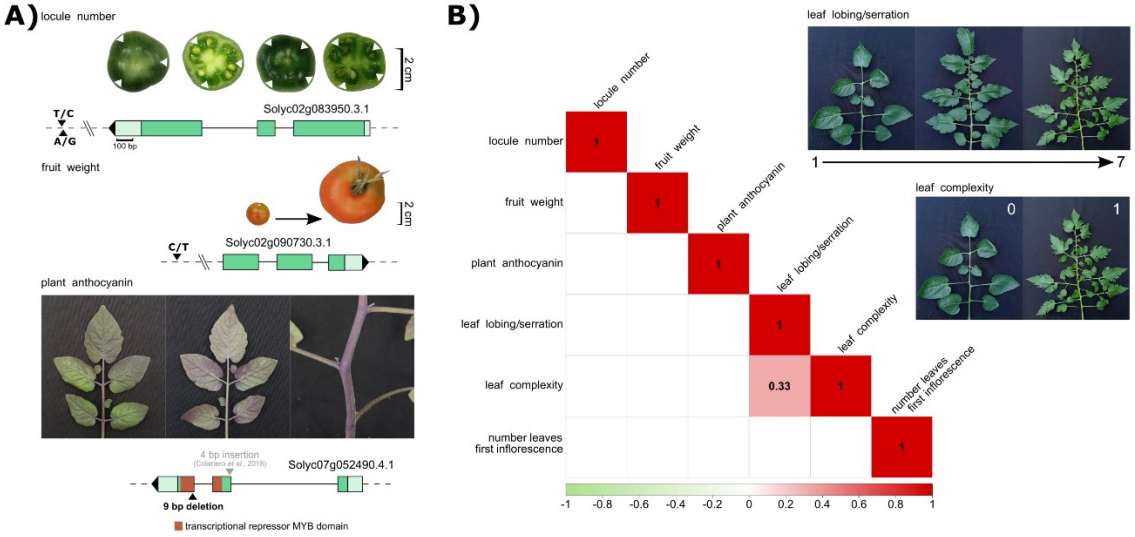


Figure 3. (A) A representation of different phenotypes for the locule number, fruit weight, and plant anthocyanin traits, together with the known genes controlling these traits indicating the phenotypic-causing variants. (B) Correlation analysis among all the studied traits showing a slight positive correlation between the leaf morphology traits, corresponding to leaf lobing/serration and leaf complexity. On the right, a representation of the phenotypic scores for both traits.

3.5. Fruit size

Locule number

The Manhattan plot for fruit locule number revealed one significant peak on chromosome 2 (Figure 4A, Table 3). For the GLM model, 28 SNPs were above the FDR threshold (LOD > 3.73), 20 of them over the Bonferroni threshold (LOD > 5.11) between 44.78 and 46.13 Mb. For the MLM model, 16 SNPs were above the FDR threshold (LOD > 4.19), nine of them over the Bonferroni threshold between a reduced region of 44.82 and 46.02 Mb (Figure 4B). For the BLINK model, a single SNP was above the FDR (LOD > 15.27) and Bonferroni thresholds at 45.87 Mb position. This association peak accounted for 26.84% of the total phenotypic variance of the locule number trait.

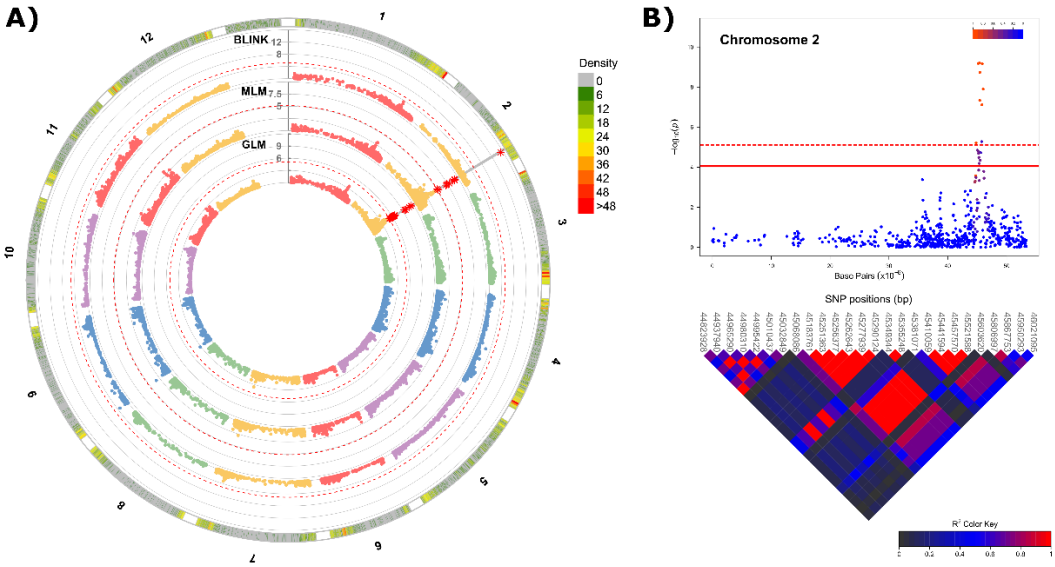


Figure 4. Genome-wide association results for the locule number trait. (A) Manhattan plots comparing GLM, MLM and BLINK models. The solid grey line indicates the common significant markers detected by two or more models. The red asterisks indicate the SNPs exceeding the Bonferroni threshold, represented as a dashed red line. (B) On the top, a chromosome-wise Manhattan plot with the top significant markers. Bonferroni and FDR thresholds are represented with red dashed and continuous lines, respectively. The colour from blue to red indicates r^2 from 0 to 1. On the bottom, heat map of pairwise linkage disequilibrium (LD). SNP positions under the significant region are indicated in bp. The colour from black to red indicates r^2 from 0 to 1.

In the genomic candidate region on chromosome 2, the *WUSCHEL* gene (Solyc02g083950.3.1, 45,191,157-45,192,582 bp) was identified (Table 3). *WUSCHEL* gene controls stem cell fate in the apical meristem directly affecting locule number during tomato fruit development (Barrero *et al.,* 2006; Muños *et al.,* 2011). The two multi-locular founders of the ToMAGIC population, SLC2 and SLC3, showed two SNPs immediately downstream of the *WUSCHEL* gene that were previously described as directly associated with an increased locule number (Muños *et al.,* 2011). Specifically, a T/C transition at 45,189,386 bp and a A/G transition at 45,189,392 bp are considered as the responsible SNPs for the locule number trait (Figure 3A). These two SNPs were in almost complete linkage disequilibrium, and they are considered as a unique haplotype.

Haplotype analyses were performed to associate the candidate genomic regions with the phenotypic effects. For the locule number, a significant difference was observed between the haplotype of the SLC2 and SLC3 founders, which are the ones showing more than 2 locules, and the rest of the haplotypes of the ToMAGIC founders according to pairwise *t*-test for multiple comparisons (Figure 5A). When generating the density plot, higher values were also associated with the SLC2 (at 3 locules) and SLC3 (at 4 locules) founder haplotype. The density curve of the rest of the founder haplotypes exceeds density values of 1, since fruits with only two locules predominate in the ToMAGIC population.

Table 3. Association analysis results for GLM, MLM, and BLINK models and list of candidate genes for locule number, fruit weight, plant anthocyanin, leaf lobing/serration, leaf complexity, and number of leaves below the first inflorescence.

Trait	GLM			MLM			BLINK			Candidate genes		
	Chromosome	Genomic region (Mb)	LOD	Chromosome	Genomic region (Mb)	LOD	Chromosome	Genomic region (Mb)	LOD	Abbreviation	Name	Position (bp)
Locule number	2	44.78 - 46.13	11.39	2	44.82 - 46.02	9.20	2	45.87	15.27	*WUSCHEL*	Solyc02g083950.3.1	45,191,157 - 45,192,582
Fruit weight	2	50.51 - 50.55	5.37	-	-	-	2	50.55	5.33	*FW2.2*	Solyc02g090730.3.1	50,292,691 - 50,293,481
Plant anthocyanin	7	8.38 - 61.70	15.28	7	59.97 - 60.88	12.42	7	60.44	21.14	*SlMYB-ATV*	Solyc07g052490.4.1	60,912,702 - 60,913,855
	2	27.13 - 33.38	5.14	-	-	-	2	33.38 - 46.91	7.64	*bHLH*	Solyc02g063430.4.1	33,546,773 - 33,549,186
Leaf lobing/serration	4	62.30 - 63.23	11.63	4	62.30 - 62.91	9.51	4	62.87	6.46	*AP3/DEF*	Solyc04g081000.3.1	63,032,681 - 63,036,255
										OVATE9	Solyc04g080210.1.1	62,437,899 - 62,438,699
										ANT	Solyc04g077490.3.1	60,418,478 - 60,421,941
Leaf complexity	4	62.49 - 62.73	5.84	4	62.49	5.38	4	62.49	8.93	*KNOTTED1*	Solyc04g077210.3.1	60,124,504 - 60,131,770
										IAA9	Solyc04g076850.3.1	59,750,087 - 59,755,552
Number of leaves below the first inflorescence	11	2.05 - 2.80	9.19	11	2.17 - 2.80	8.54	11	2.80	24.22	*FT1*	Solyc11g008640.1.1	2,854,837 - 2,857,237
										FT2	Solyc11g008650.1.1	2,866,945 - 2,867,166
										SP1	Solyc11g007880.1.1	2,135,303 - 2,135,602
										J	Solyc11g010570.2.1	3,671,232 - 3,676,350

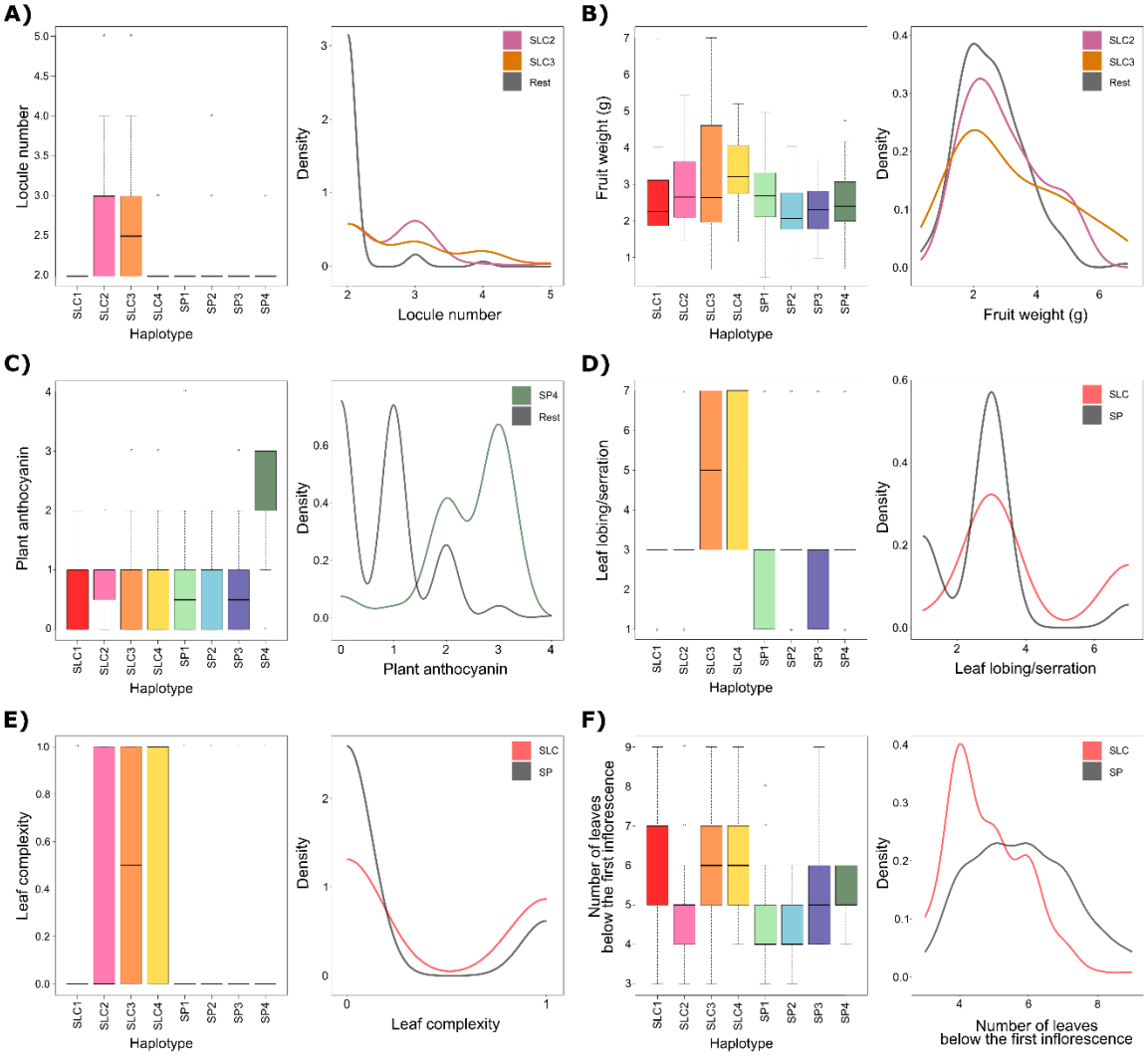


Figure 5. Haplotype analysis of the ToMAGIC lines for each of the MAGIC founders' haplotype in combination with phenotypic data. Boxplot and density plot distribution in the candidate genomic regions for: (A) locule number; (B) fruit weight; (C) plant anthocyanin on chromosome 7; (D) leaf lobing/serration; (E) leaf complexity; and (F) number of leaves below the first inflorescence. Boxplots represent the ToMAGIC lines phenotypes associated to the eight haplotypes with the different colours code for each founder and hollow dots correspond to outliers. Density plots represent the variation among groups that show significant differences.

Fruit weight

The Manhattan plot for fruit weight also revealed one significant peak on chromosome 2, although only for GLM and BLINK models (Supplementary Figure 3A, Table 3). For the GLM model, eight SNPs were above the FDR threshold (LOD > 4.15), three of them over the Bonferroni threshold (LOD > 5.11) between 50.51 and 50.55 Mb (Supplementary Figure 3B). For the BLINK model, a single SNP was above the FDR (LOD = 5.33) and Bonferroni thresholds at 50.55 Mb position. This association peak explained 14.76% of the total phenotypic variance of the fruit weight trait.

Under the significant peak on chromosome 2, the well-known *FW2.2* gene (Solyc02g090730.3.1, 50,292,691-50,293,481 bp) was identified (Table 3). This gene is differentially expressed in floral development and controls carpel cell division (Frary *et al.,* 2000). The wild-type SNP was identified in all the ToMAGIC founders, except for founders SLC2 and SLC3, which have larger fruit weights (Blanca *et al.,* 2015). This SNP corresponds to a C/T change upstream of the 5' region of *FW2.2* gene at 50,292,019 bp (Figure 3A).

In the haplotype analysis, pairwise *t*-test revealed a significant difference between SLC2 and SLC3 on one side and SP founders from the other (Figure 5B). When generating the density plot, most of the lines are around 2 to 3 g since light fruits predominate in the ToMAGIC population with an average weight of 2.72 g (Table 2). Lines with weights greater than 3 show mostly SLC2 and SLC3 haplotypes.

3.6. Plant pigmentation

The Manhattan plot for plant anthocyanin revealed two significant peaks: one major peak on chromosome 7 and one minor but significant peak on chromosome 2 (Supplementary Figure 4A, Table 3). For the GLM model, 43 SNPs were above the FDR threshold (LOD > 3.33) on chromosome 7, 21 of them over the Bonferroni threshold (LOD > 5.11) between 8.38 and 61.70 Mb. On chromosome 2, 14 SNPs were above the FDR threshold, being only two of them over the Bonferroni threshold between 27.13 and 33.38 Mb. For the MLM model, only one association peak was identified on chromosome 7 with eight SNPs over the FDR threshold (LOD > 4.17), five of them over the Bonferroni threshold between a reduced region of 59.97 and 60.88 Mb (Supplementary Figure 4B). For the BLINK model, a single SNP was above the FDR (LOD > 5.67) and Bonferroni thresholds reaching a LOD of 21.14 on chromosome 7 at 60.44 Mb position. On chromosome 2, only two SNPs were above the FDR and Bonferroni thresholds at 33.38 and 46.91 Mb positions reaching

a LOD of 7.64 and 6.70, respectively. The association peak on chromosome 7 explained 15.14% of the total phenotypic variance of the plant anthocyanin trait, while the peak on chromosome 2 explained 4.68% of the phenotypic variance.

Under the major GWAS peak on chromosome 7, in the genomic region of 60,912,702-60,913,855 bp, a *MYB-like* transcription factor (*SlMYB-ATV,* Solyc07g052490.4.1) was identified (Table 3). The *SlMYB-ATV* (myeloblastosis-atroviolacea) gene has been described as a repressor of anthocyanin synthesis in vegetative tissues of tomato plants (Colanero *et al.,* 2018). However, we did not observe the previously described mutations in the gene sequence in our accessions. In contrast, a 9-bp in frame deletion at 60,912,903 bp position, deleting 3 amino acids in the transcriptional repressor MYB domain was identified in the SP4 founder, which is the unique founder showing anthocyanins in all plant parts (Figure 3A).

The same procedure was followed for the minor peak on chromosome 2. All the genes located near or within the LD block were assessed by SnpEff (Cingolani *et al.,* 2012) for all of the MAGIC founders. We found the *bHLH* TF (Solyc02g063430.4.1 between 33,546,773-33,549,186), which belongs to a family involved in the regulation of anthocyanin biosynthesis in plants (Petroni and Tonelli, 2011) (Table 3). However, no high-effect variants were predicted distinguishing between anthocyanin-containing and anthocyaninless founders.

In the haplotype analysis for chromosome 7, a significant difference was observed between the SP4 founder, which is the one showing increased levels of plant anthocyanins, and the rest of ToMAGIC founders according to pairwise *t*-test (Figure 5C). When generating the density plot, higher anthocyanin values were also associated with the SP4 founder haplotype.

3.7. Leaf morphology

Leaf lobing/serration

The Manhattan plot for leaf lobing/serration revealed one significant peak on chromosome 4 (Supplementary Figure 5A, Table 3). For the GLM model, 20 SNPs were above the FDR threshold (LOD > 3.86), ten of them over the Bonferroni threshold (LOD > 5.11) between 62.30 and 63.23 Mb. For the MLM model, nine SNPs were above the FDR threshold (LOD > 4.72), nine of them over the Bonferroni threshold between a reduced region of 62.30 and 62.91 Mb (Supplementary Figure 5B). For the BLINK model, a single SNP was above the FDR (LOD > 6.46) and Bonferroni thresholds at 62.87 Mb position. This association peak accounted for 53.84% of the total phenotypic variance of the leaf lobing/serration trait.

Different genes involved in the leaf shape were detected within the candidate genomic region on chromosome 4 identified in the GWAS for the leaf lobing/serration (Table 3). In order of proximity to the candidate region, we found the *APETALA3/DEFICIENS* or *AP3/DEF* gene (Solyc04g081000.3.1 between 63,032,681-63,036,255 bp), which has been described as a regulator of petal and sepal development (Quinet *et al.,* 2014), the ovate family protein 9 or *OVATE9* gene (Solyc04g080210.1.1 between 62,437,899-62,438,699 bp) which belongs to a family protein that regulates different plant organs shape, including cotyledons, leaves, and fruits (Snouffer *et al.*, 2020), and the *AP2-like* ethylene-responsive transcription factor *AINTEGUMENTA* or *ANT* gene (Solyc04g077490.3.1 between 60,418,478-60,421,941 bp), which plays a role as an auxin regulator in shoot and flower meristem maintenance, organ size and polarity, flower initiation, ovule development, floral organ identity, cell proliferation (Horstman *et al*., 2014). No high-effect variants were predicted by SnpEff in the coding sequence of these genes contrasting for the different founders' phenotypes.

Haplotype results revealed a significant difference between SLC and SP founders according to pairwise *t*-test (Figure 5D). Although the haplotypes density plot also did not show a bimodal distribution for SLC and SP founders, it showed a higher density for SP haplotypes in lines exhibiting lack of lobing/serration or moderate lobing values, and a slightly higher density for SLC haplotypes in the very serrated leaf values.

Leaf complexity

The Manhattan plot for leaf complexity revealed one significant peak on chromosome 4 (Supplementary Figure 6A, Table 3). For the GLM model, three SNPs were above the FDR threshold (LOD > 4.70), two of them over the Bonferroni threshold (LOD > 5.11) between 62.49 and 62.73 Mb (Supplementary Figure 6B). For the MLM model, a single SNP was above the FDR (LOD > 5.38) and Bonferroni thresholds at 62.49 Mb position. At the same position for the BLINK model, a single SNP was above the FDR (LOD > 4.95) and Bonferroni thresholds reaching a LOD of 8.93. This association peak accounted for 4.12% of the total phenotypic variance of the leaf complexity trait.

Two genes involved in the leaf complexity were detected within the candidate genomic region on chromosome 4 identified in the GWAS (Table 3). In order of proximity to the candidate region we found the *KNOTTED1* gene (Solyc 04g077210.3.1 between 60,124,504-60,131,770 bp) which is expressed during leaf development and affects leaf morphology altering leaf complexity (Shani *et al.,* 2009), and the *entire* or *INDOLE-3-ACETIC ACID9 IAA9* gene (Solyc04g076850.3.1 between 59,750,087-59,755,552 bp), which controls leaf morphology from compound to simple leaves (Zhang *et al.,* 2007). No high-effect

variants were predicted by SnpEff in the coding sequence of these genes for the founders with contrasting phenotypes.

Haplotype results revealed a significant difference between SLC and SP founders according to the pairwise *t*-test (Figure 5E). Although the haplotypes density plot did not show a bimodal distribution for SLC and SP founders, it showed a higher density for SP haplotypes in pinnate leaves, and a slightly higher density for SLC haplotypes in the bipinnate leaves.

3.8. Earliness

The Manhattan plot for the number of leaves below the first inflorescence revealed one significant peak on chromosome 11 (Supplementary Figure 7A, Table 3). For the GLM model, seven SNPs were above the FDR threshold (LOD > 4.30), four of them over the Bonferroni threshold (LOD > 5.11) between 2.05 and 2.80 Mb. For the MLM model, two SNPs were above the FDR (LOD > 5.53) and Bonferroni thresholds between a reduced region of 2.17 and 2.80 Mb (Supplementary Figure 7B). For the BLINK model, a single SNP was above the FDR (LOD > 4.95) and Bonferroni thresholds reaching a LOD of 24.22 at 2.80 Mb position. The association peak explained 5.52% of the total phenotypic variance of the number of leaves below the first inflorescence trait.

Different genes implicated in the flowering pathway were identified in the candidate genomic region on chromosome 11 proposed in the GWAS for the number of leaves below the first inflorescence (Table 3). In order of proximity to the candidate region we found two *FLOWERING LOCUS T* (*FT*) genes (*FT1* Solyc11g008640.1.1 between 2,854,837- 2,857,237 bp and *FT2* Solyc11g008650.1.1 between 2,866,945-2,867,166 bp), which have been described as mediating the onset of flowering and the floral transition in all angiosperms (Pin and Nilsson, 2012), the *SELF-PRUNING INTERACTING PROTEIN 1* or *SP1* gene (Solyc11g007880.1.1 between 2,135,303-2,135,602 bp), which is involved in a conserved signalling system that regulates flowering (Pnueli *et al.,* 2001), and the *JOINTLESS* or *J* gene (Solyc11g010570.2.1 between 3,671,232-3,676,350 bp), which plays a role in flowering promotion (Szymkowiak and Irish, 2006). The FT1 and FT2 proteins have respectively a 71.68% (124/173) and 87.69% (57/65) identity with the well-known *SINGLE-FLOWER TRUSS* (*SFT*, Solyc03g063100.2.1) gene product according to BLASTp alignment. While *FT1* is recognized as a paralogue of the *SFT* gene in EnsemblPlants, *FT2* seems to be a truncated pseudogene. Nevertheless, no clear variants were predicted by SnpEff in the coding sequence of these genes contrasting for the different founders' phenotypes.

Haplotype results did not differentiate between SLC and SP founders (Figure 5F). Pairwise *t*-test only revealed a significant difference between SLC1, SLC3, and SLC4 from SLC2, SP1, and SP2 founders, with SP3 and SP4 in intermediate positions. The haplotype density plot also did not show a bimodal distribution for SLC and SP founders. However, it showed a trend for lower number leaves below the first inflorescences for the SP haplotypes, while SLC haplotypes were distributed along a wide range of number of leaves below the first inflorescence.

4. Discussion

We present a novel inter-specific ToMAGIC population of 354 lines constructed by combining the genomes of SLC and SP founders. SLC accessions are phylogenetically positioned between SP and cultivated tomato (Blanca *et al.,* 2015, 2022). Therefore, founders were selected to exploit the wide diversity found in the tomato closest relatives taking advantage of their interbreeding compatibility (Peralta *et al.,* 2008). Previous resequencing of the selected founders allowed to significantly enhance recombination detection, haplotype prediction, and causal variants identification within the MAGIC population (Gramazio *et al.,* 2020a).

The MAGIC population was generated through a systematic "funnel" approach (Arrones *et al*., 2020) involving multiple rounds of inter-cross of the eight selected founders and five generations of selfing, totalling ten generations. The three inter-crossing generations from the two double hybrids and the blind SSD process ensured high levels of recombination, maintaining a high genetic and morphological diversity. The final population consisted of 354 ToMAGIC lines which was considered an appropriate population size to detect QTLs according to: (i) tomato genome size (Huynh *et al*., 2018), (ii) simulations of the power for detection of QTLs of an eight-way MAGIC population (Dell'Acqua *et al*., 2015), and (iii) population size of previously developed tomato MAGIC populations (Pascual *et al*., 2015; Campanelli *et al*., 2019). The ToMAGIC lines were genotyped by using a newly developed 12k probes tomato panel, based on SPET, which is a robust technology based on target SNPs, but also capable of discovering novel SNPs (Barchi *et al.,* 2019). Although SPET has been mostly used in the biomedical field, it has demonstrated its potential as a high-throughput and high-efficiency genotyping platform in *Solanum* species (Gramazio *et al.,* 2020b; Mangino *et al.,* 2022). In this study, more than 4 million SNPs were generated with the 12k probes tomato SPET panel. After stringent filtering, 6,488 were retained as markers, while in the previous tomato MAGIC population developed by Pascual *et al.* (2015), 1,486 markers obtained by a custom-made genotyping platform (Fluidigm 96.96 Dynamic Arrays, San Francisco, CA) were used for population analyses. In general, genetic diversity

within the phylogenetic groups of the tomato clade is relatively low, which is one of the main reasons of the reduction in the final number of SNPs. Genomic divergence is estimated as 0.6% between *S. pimpinellifolium* and cultivated tomato, whereas most of SNPs are distributed in gene-poor regions (Sato *et al*., 2012). The genotypic data revealed the absence of genetic structure, which is one of the advantages of MAGIC populations (Arrones *et al.,* 2020), and a balanced representation of the founder genomes. The average contribution of each founder to the overall population was around 12.50%, which is the expected value for a population developed from eight founders.

We have demonstrated the power of our ToMAGIC population for the fine mapping of traits of interest in tomato breeding. Specifically, GWAS analysis detected strong associations for all the traits evaluated using three different models (GLM, MLM, and BLINK), supporting the robustness of the associations detected (Price *et al.,* 2006; Yu *et al.,* 2006; Huang *et al.,* 2019). This population could also be used to validate candidate genomic regions or genes previously identified through selective sweeps.

The implementation of SLC and SP accessions as founders have introduced a wide genetic and phenotypic diversity in the ToMAGIC population (Blanca *et al.,* 2015; Gramazio *et al*., 2020a). Our proof-of-concept, focusing on a subset of traits from different plant parts has revealed a large phenotypic diversity in the ToMAGIC population, including transgressive lines to some of the founders for all traits except leaf morphology. Within the phenotypic diversity of the final population, wild alleles showed a dominant effect over domesticated alleles in most traits. For instance, ToMAGIC lines tend to produce small fruits and simpler leaves, more similar to SP than to cultivated tomato. This prevalent dominance of wild alleles has been previously observed during the development of other inter-specific populations (Semel *et al.,* 2006).

Large tomato fruit size is a typical domestication trait, controlled by at least five different genes (Pereira *et al.,* 2021). It is tempting to speculate that, similar to the non-shattering spike trait in cereals (Lin *et al.,* 2012), it negatively affects plant fitness in the wild, by reducing seed dispersal by small vertebrates. Drawing on this parallel, the most likely scenario is that recessive alleles for large fruit size in tomato and non-shattering spike in cereals were both pre-existing in wild/weedy populations, and that they were not completely counterselected due to their recessive nature. Under this scenario, human selection for higher harvestable biomass probably acted on the rare homozygous plants that appeared in these wild populations. Consistent with this hypothesis, the non-functional (domesticated) allele of the rice shattering gene *sh4* is found, at low frequency, in the wild ancestor *O. rufipogon* (Lin *et al.,* 2007).

Almost all wild tomato species produce bilocular small fruits, and therefore, locule number and fruit weight played a crucial role in the increase in fruit size during domestication (Alpert *et al*., 1995; Lippman and Tanksley, 2001; Barrero *et al.,* 2006). On one hand, as a result of the GWAS analysis for locule number, an associated genomic region was identified that colocalized with the *WUSCHEL* gene. Mutations on this gene have been necessary to increase locule number during domestication (Muños *et al.,* 2011). However, previous sequence analysis on this gene revealed that the diversity of this locus was drastically reduced in the cultivated species (Muños *et al.,* 2011; van der Knaap *et al.,* 2014). Only two SNPs have been identified in this gene responsible for the large-fruited phenotype, which are the same two SNPs that we have found in our population. On the other hand, the GWAS analysis for fruit weight revealed an associated genomic region on chromosome 2 between 50.51 and 50.55 Mb in the region where the *FW2.2* gene is located (Frary *et al.,* 2000). Similarly, but not as precisely as in our ToMAGIC populations, in the tomato MAGIC developed by Pascual *et al.* (2015) a peak with the highest LOD value between 46.35 and 47.49 Mb was also identified. The *FW2.2* gene is responsible for up to 30% of the fruit weight variation between large domesticated tomatoes and the small-fruited wild relatives (Nesbitt and Tanksley, 2001). All modern tomatoes contain the large-fruited allele for *FW2.2* (Blanca *et al.,* 2015; Beauchet *et al.,* 2021), which was also identified in the two large-fruited SLC ToMAGIC founders. Molecular evolutionary studies suggested that this allele originated in wild tomatoes long before the process of domestication (Nesbitt and Tanksley, 2002). Indeed, fruit weight was strongly selected in SLC in the Andean region of Ecuador and Northern Peru prior to the domestication of tomato in Mesoamerica (Blanca *et al.,* 2015).

Anthocyanins are the main responsible for purple pigmentation in tomato leaf veins, leaf tissues, and stem (Barrett *et al.,* 2010; Jaakola, 2013). Plant anthocyanins are more commonly present in wild tomato species, where they have a main protective function against UV-visible light and other stressful conditions such as cold temperature, pathogens, or drought (Gould, 2004; Olsen *et al*., 2009; Zhang *et al*., 2014). The GWAS results identified an associated genomic region which colocalized with the previously described *SlMYB-ATV* gene. Overexpression of the coding protein acts as an inhibitor of anthocyanin production by silencing key regulators of the biosynthesis pathway (Cao *et al*., 2017; Colanero *et al*., 2018). The *atv* mutation was described as a 4 bp insertion in the second exon which led to a frameshift variant resulting in a premature stop codon with a strong impact in the polypeptide. This mutation was identified as the causal agent of anthocyanin production in the vegetative part of the plant (Colanero *et al*., 2018). Here, a novel mutation in the "purple" SP4 founder was found. Specifically, a 9 bp deletion leading to a disruptive inframe deletion which directly affects the transcription repressor MYB domain was identified. This demonstrates the significance of the ToMAGIC

population as a reservoir of novel candidate genes and causative alleles. Interestingly, of the four SP founders, SP4 is the only one showing anthocyanin pigmentation as well as the one collected at the highest altitude (1,020 m) and lowest mean annual temperature (13°C), in agreement with the proposed role of anthocyanins as UV-sunscreens in cold temperatures (Martínez-Cuenca *et al.,* 2020).

Cultivated tomato leaf morphology has typical bipinnate compound leaves with moderately deep lobes, while there is a huge diversity of leaf morphology among wild tomato species (Zhang *et al.,* 2007; Kang *et al.,* 2010; Nakayuma *et al.,* 2023). Since leaf lobing/serration and leaf complexity traits are correlated, both traits have usually been studied together (Kang *et al.,* 2010). Actually, the GWAS results identified an associated genomic region on chromosome 4 around 62 Mb position for both traits, and candidate genes affecting both traits were identified within this genomic region. Although the *AP3/DEF* gene has mainly been related to petal and sepal development, other genes belonging to the same MADS box family are involved in tomato leaf development. Specifically, the *APETALA1/FRUITFULL* (*AP1/FUL*) MADS box genes are involved in the organogenic activity of the leaf margin and leaf complexity (Burko *et al.,* 2013). The *ANT* gene also belongs to a family of APETALA 2/ETHYLENE RESPONSE FACTOR (AP2/ERF) domain transcription factors which affects plant leaf shape and size by regulating cell proliferation (Horstman *et al.,* 2014). The *OVATE* gene was first identified in tomato as a key regulator of fruit shape (Wang *et al.,* 2016). However, expression of *OVATE* genes can also result in dwarf plants with shorter and thicker organs such as rounder leaves (Snouffer *et al.,* 2020). The tomato *KNOTTED1* promotes cytokinin biosynthesis which is directly related to cell proliferation (Nakayuma *et al.,* 2023), and different levels of cytokinins led to a broad spectrum in leaf complexity (Shani *et al.,* 2009; Shwartz *et al.,* 2016). This gene has a key role in the molecular mechanism behind leaf development and evolution and has been repeatedly exploited to generate natural variations in leaf shape (Ichihashi and Tsukaya, 2015). The *IAA9* gene is a transcriptional repressor in auxin signal transduction (Abe-Hara *et al.,* 2021). Tomato mutants for *IAA9* also showed altered leaf morphology with the compound leaf changing to a single leaf (Zhang *et al.,* 2007; Ueta *et al.,* 2017; Abe-Hara *et al.,* 2021). In this way, leaf development is mainly influenced by cell proliferation and different hormones as a result of the activity of a complex gene network (Nakayuma *et al.,* 2023). An accurate phenotyping of the ToMAGIC population for these traits has allowed to narrow down a genomic region that harbours a large number of genes related to leaf morphology. This genomic region could be further narrowed down by studying the segregation of the cross between two isolines to enable the identification of the responsible gene/s.

The existence of early-flowering alleles in wild species indicates the relevance of exploiting the genetic variation present in tomato wild relatives (Jiménez-Gómez

et al., 2007). Although the mechanisms controlling the transition from vegetative to reproductive growth are complex, several genes involved in flowering regulation are known (Meir *et al.,* 2021; Zhang *et al.,* 2023b). The number of leaves below the first inflorescence trait is a proxy for earliness in tomato (Honma *et al.,* 1963) and is easily scored and commonly assessed to evaluate the earliness in tomato (Jiménez-Gómez *et al.,* 2007; Nakano *et al.,* 2016; Silva *et al.,* 2019). The GWAS analysis for the number of leaves below the first inflorescence identified an association on chromosome 11, where several genes related to flowering time were found (two *FT* genes, *SP1*, and *J*). The most studied *FT* gene is the tomato ortholog *SINGLE-FLOWER TRUSS* (*SFT*) gene on chromosome 3, which encodes for florigen and induces flowering in day-neutral (Turck *et al.,* 2008; Meir *et al.,* 2021; Zhang *et al.,* 2023b). Here, we report the *FT1* gene on chromosome 11, a paralogue of the *SFT* gene which may also be involved in the flowering regulation. The *SP1* gene is a member of the *CETS* family of regulatory genes, together with *FT* genes, controlling flowering time (Pnueli *et al.,* 2001). However, they play an antagonistic role, since *SP1* delays flowering in tomato (Zhang *et al.,* 2023b). The *J* gene is involved in the same pathway as the *SFT* gene but with a small role in flowering promotion (Szymkowiak and Irish, 2006; Zhang *et al.,* 2023b). A better understanding of the mechanisms underlying the tomato flowering regulatory pathways will allow breeding to target more precise candidate genes for the induction of early flowering. Nevertheless, once again, the ToMAGIC population has led us to a genomic region directly involved in the transition to flowering, pointing to new candidate genes.

Overall, the genotyping results together with the large morphological variation observed in the new inter-specific SLC/SP tomato MAGIC population, as well as the appearance of transgressive phenotypes, indicate that recombination and variation were maximised in the final population. The ToMAGIC population size was suitable for an accurate association detection of well-known traits as a proof-of-concept to validate the efficiency of the population. The ToMAGIC population has demonstrated a high potential for the fine mapping of traits of interest from different plant parts. A novel mutation was identified in the *SlMYB-ATV* gene responsible of the anthocyanin pigmentation in vegetative tissues. Further transcriptional expression analysis of genes under the anthocyanin biosynthesis pathway and gene editing will be essential to elucidate the effect of this mutation. Candidate genes were proposed for leaf morphology and earliness related traits. Fine mapping and additional gene expression analysis could better elucidate the genetic control of these traits. Given the fact that the population contains representatives of the tomato ancestor (SP) and the primitive weedy forms (SLC) of tomato, it can also be a tool of great relevance for studying the genetic changes in the early stages of tomato domestication. It is also evident from our study that the derived ToMAGIC population or core collections developed from it can contribute to tomato genetics research and breeding programs. Currently, the ToMAGIC population is being

assessed for nitrogen use efficiency, drought tolerance, and resistance to different pathogens. Recombinant lines with combinations of traits of interest present in different founders can also be of direct interest to breeders or even for selection of small-fruited new cultivars.

Data availability statement: The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/, PRJNA616074 (founders resequencing data) and PRJNA1103671 (ToMAGIC lines genotyping).

Author contributions: SV, PG, MJD, and JP conceived the idea and supervised the manuscript; AA OA, LP-D, AS, and MJD performed the field trials. GA and GG in collaboration with TECAN Genomics designed the 12k SPET panel. All authors analysed the results. AA and OA prepared a first draft of the manuscript and the rest of authors reviewed and edited the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by grant PID2020-118627RB-I00 funded by MICIU/AEI/10.13039/501100011033, TED2021-129296B-I00 funded by MCIN/AEI/10.13039/501100011033 and European Union Europea NextGenerationEU/ PRTR, CIPROM/2021/020 funded by Conselleria d'Innovació, Universitats, Ciència i Societat Digital of the Generalitat Valenciana, the HARNESSTOM innovation action (grant agreement No. 101000716) funded by the European Commission H2020 Research and Innovation Programme, and the Horizon Europe PRO-GRACE project (grant agreement No. 10194738). Andrea Arrones is grateful to Spanish Ministerio de Ciencia, Innovación y Universidades for a predoctoral (FPU18/01742) contract. Oussama Antar is grateful to Conselleria d'Innovació, Ciència i Societat Digital of the Generalitat Valenciana for a pre-doctoral grant within the Santiago Grisolía program (CIGRIS/2021/113). Leandro Pereira-Dias is grateful to Universitat Politècnica de Valencia and the Spanish Ministerio de Universidades for a post-doctoral grant under the Margarita Salas funded by the European Union NextGenerationEU/PRTR. Andrea Solana is grateful to Spanish Ministerio de Ciencia e Innovación for a pre-doctoral grant (PRE2022-102368) funded by MCIN/AEI/10.13039/501100011033 and FSE+. 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. Funding for open access: Universitat Politècnica de València.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Supplementary material: The supplementary material for this article can be found online at:

https://www.biorxiv.org/content/10.1101/2024.02.28.582481v1.supplementary-material

Supplementary Figure 1. Chain pollination scheme of the double hybrids (DHY) followed during three inter-crossing generations (IC1, IC2, and IC3) to obtain the quadruple hybrids as a complete admixture of the genomes of the eight founders.

Supplementary Figure 2. Distribution of the phenotypic value of the studied traits in the ToMAGIC lines. Founder trait values are indicated with vertical lines.

Supplementary Figure 3. Genome-wide association results for the fruit weight trait. (A) Manhattan plots comparing GLM, MLM and BLINK models. (B) On the top, a chromosome-wise Manhattan plot with the top significant markers. Bonferroni threshold is represented with red dashed line. On the bottom, heat map of pairwise linkage disequilibrium (LD).

Supplementary Figure 4. Genome-wide association results for the plant anthocyanin trait. (A) Manhattan plots comparing GLM, MLM and BLINK models. (B) On the top, a chromosome-wise Manhattan plot with the top significant markers. Bonferroni and FDR thresholds are represented with red dashed and continuous lines, respectively. On the bottom, heat map of pairwise linkage disequilibrium (LD).

Supplementary Figure 5. Genome-wide association results for the leaf lobing/serration trait. (A) Manhattan plots comparing GLM, MLM and BLINK models. (B) On the top, a chromosome-wise Manhattan plot with the top significant markers. Bonferroni and FDR thresholds are represented with red dashed and continuous lines, respectively. On the bottom, heat map of pairwise linkage disequilibrium (LD).

Supplementary Figure 6. Genome-wide association results for the leaf complexity trait. (A) Manhattan plots comparing GLM, MLM and BLINK models. (B) On the top, a chromosome-wise Manhattan plot with the top significant markers. Bonferroni threshold is represented with red dashed line. On the bottom, heat map of pairwise linkage disequilibrium (LD).

Supplementary Figure 7. Genome-wide association results for the number of leaves below the first inflorescence trait. (A) Manhattan plots comparing GLM, MLM and BLINK models. (B) On the top, a chromosome-wise Manhattan plot with

the top significant markers. Bonferroni and FDR thresholds are represented with red dashed and continuous lines, respectively. On the bottom, heat map of pairwise linkage disequilibrium (LD).

References:

Abe-Hara, C., Yamada, K., Wada, N., *et al.* (2021). Effects of the *sliaa9* mutation on shoot elongation growth of tomato cultivars. *Front. Plant. Sci., 12,* 627832. doi: 10.3389/fpls.2021.627832

Alpert, K. B., Grandillo, S., and Tanksley, S. D. (1995). *fw 2.2*: a major QTL controlling fruit weight is common to both red- and green-fruited tomato species. *Theor. Appl. Genet., 91,* 994–1000. doi: 10.1007/BF00223911

Arrones, A., Vilanova, S., Plazas, M., Mangino, G., Pascual, L., Díez, M. J., *et al*. (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*, 229. doi: 10.3390/biology9080229

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. doi: 10.3389/fpls.2019.01005

Barrero, L. S., Cong, B., Wu, F., and Tanksley, S. D. (2006). Developmental characterization of the fasciated locus and mapping of *Arabidopsis* candidate genes involved in the control of floral meristem size and carpel number in tomato. *Genome, 49,* 991–1006. doi: 10.1139/g06-059

Barrett, D. M., Beaulieu, J. C., and Shewfelt, R. (2010). Color, flavor, texture, and nutritional quality of fresh-cut fruits and vegetables: desirable levels, instrumental and sensory measurement, and the effects of processing. *Critical Reviews in Food Sci. and Nutrition, 50,* 369–389. doi: 10.1080/10408391003626322

Beauchet, A., Gévaudant, F., Gonzalez, N., and Chevalier, C. (2021). In search of the still unknown function of *FW2.2/CELL NUMBER REGULATOR*, a major regulator of fruit size in tomato. *J. Exp. Bot., 72,* 5300–5311. doi: 10.1093/jxb/erab207

Benjanmini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc., 57*, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

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. doi: 10.1186/s12864-015-1444-1

Blanca, J., Sanchez-Matarredona, D., Ziarsolo, P., Montero-Pau, J., Van der Knaap, E., Díez, M. J., *et al*. (2022). Haplotype analyses reveal novel insights into tomato history and domestication driven by long-distance migrations and latitudinal adaptations. *Horticulture Research, 9*. doi: 10.1093/hr/uhac030

Bombarely, A., Menda, N., Tecle, I. Y., Buels, R. M., Strickler, S., Fischer-York, T., *et al*. (2010). The sol genomics network (solgenomics.net): Growing tomatoes using Perl. *Nucleic Acids Res., 39,* 1149–1155. doi: 10.1093/nar/gkq866

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. doi: 10.1093/bioinformatics/btm308

Burko, Y., Shleizer-Burko, S., Yanai, O., Shwartz, I., Zelnik, I. D., Jacob-Hirsch, J., *et al*. (2013). A role for *APETALA1/FRUITFULL* transcription factors in tomato leaf development. *Plant Cell*, *25*, 2070–2083. doi: 10.1105/tpc.113.113035

Campanelli, G., Sestili, S., Acciarri, N., Montemurro, F., Palma, D., Leteo, F., *et al*. (2019). Multi-parental advances generation inter-cross population, to develop organic tomato genotypes by participatory plant breeding. *Agronomy, 9,* 119. doi: 10.3390/agronomy9030119

Cao, K., Cui, L., Zhou, X., Ye, L., Zou, Z., and Deng, S. (2016). Four tomato *FLOWERING LOCUS T-like* proteins act antagonistically to regulate floral initiation. *Front. Plant. Sci., 6,* 1213. doi: 10.3389/fpls.2015.01213

Cavanagh, C., Morell, M., Mackay, I., and Powell, W. (2008). From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. *Curr. Opin. Plant Biol., 11,* 215–221. doi: 10.1016/j.pbi.2008.01.002

Chen, S. (2023). Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *iMeta,* e107. doi: 10.1002/imt2.107

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., *et al*. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain *w1118; iso-2; iso-3*. *Fly, 6,* 80–92. doi: 10.4161/fly.19695

Colanero, S., Perata, P., and Gonzali, S. (2018). The *atroviolacea* gene encodes an R3-MYB protein repressing anthocyanin synthesis in tomato plants. *Front. Plant. Sci., 9,* 830. doi: 10.3389/fpls.2018.00830

DePristo, M., 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,* 491–501. doi: 10.1038/ng.806.A

Dell'Acqua, M., Gatti, D. M., Pea, G., Cattonaro, F., Coppens, F., Magris, G., *et al.* (2015). Genetic properties of the MAGIC maize population: a new platform for high definition QTL mapping in *Zea mays*. *Genome Biol*., *16*, 1–23. doi: 10.1186/s13059-015-0716-z

Díez, M. J., Picó, B., and Nuez, F. (2002). Cucurbit Genetic Resources in Europe: Ad Hoc Meeting held in Adana, Turkey, 19 January 2002. Rome: International Plant Genetic Resources Institute.

Doganlar, S., Frary, A., Ku, H. M., and Tanksley, S. D. (2002). Mapping quantitative trait loci in inbred backcross lines of *Lycopersicon pimpinellifolium* (LA1589). *Genome, 45*, 1189–1202. doi: 10.1139/g02-091

Eshed, Y., and Zamir, D. (1995). An Introgression Line population of *Lycopersicum pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. *Genetics, 141,* 1147–1162.

Fei, Z., Joung, J. G., Tang, X., Zheng, Y., Huang, M., Lee, J. M., *et al*. (2010). Tomato functional genomics database: A comprehensive resource and analysis package for tomato functional genomics. *Nucleic Acids Res., 39,* 1156–1163. doi: 10.1093/nar/gkq991

Fei, Z., Tang, X., Alba, R., and Giovannoni, J. (2006). Tomato Expression Database (TED): a suite of data presentation and analysis tools. *Nucleic Acids Res., 34,* 766–770. doi: 10.1093/nar/gkj110

Frary, A., and Doganlar, S. (2003). Comparative genetics of crop plant domestication and evolution. *Turkish J. Agric. For., 27*, 59–69.

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. doi: 10.1126/science.289.5476.85

Fulop, D., Ranjan, A., Ofner, I., Covington, M. F., Chitwood, D. H., West, D., *et al*. (2016). A new advanced backcross tomato population enables high resolution leaf QTL mapping and gene identification. *G3 Genes, Genomes, Genet., 6,* 3169–3184. doi: 10.1534/g3.116.030536

Gould, K. S. (2004). Nature's Swiss army knife: the diverse protective roles of anthocyanins in leaves. *J. Biomed. Biotechnol., 5,* 314.

Gramazio, P., Jaén-Molina, R., Vilanova, S., Prohens, J., Marrero, Á., Caujapé-Castells, J., *et al*. (2020b). 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). *Front. Plant Sci., 11*, 757. doi: 10.3389/fpls.2020.00757

Gramazio, P., Pereira-Dias, L., Vilanova, S., Prohens, J., Soler, S., Esteras, J., *et al*. (2020a). 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 inter-specific tomato MAGIC population. *Hortic. Res., 7,* 174. doi: 10.1038/s41438-020-00395-w

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. doi: doi: 10.3389/fpls.2019.01220

Hochberg, Y. (1988). A sharper bonferroni procedure for multiple tests of significance. *Biometrika, 75*, 800–802. doi: 10.1093/biomet/75.4.800

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scand. J. Stat.*, *6,* 65–70.

Honma, S. (1963). Flowering and earliness in the tomato. *J. Hered., 54*, 212-218.

Horstman, A., Willemsen, V., Boutilier, K., and Heidstra, R. (2014). AINTEGUMENTA-LIKE proteins: Hubs in a plethora of networks. *Trends Plant Sci., 19*, 146–157. doi: 10.1016/j.tplants.2013.10.010

Hosmani, P. S., Flores-Gonzalez, M., Geest, H. van de, Maumus, F., Bakker, L. V, Schijlen, E., *et al*. (2019). An improved de novo assembly and annotation of the tomato reference genome using single-molecule sequencing, Hi-C proximity ligation and optical maps. *bioRxiv,* 767764. doi: 10.1101/767764

Huang, M., Liu, X., Zhou, Y., Summers, R. M., and Zhang, Z. (2019). BLINK: A package for the next level of genome-wide association studies with both individuals and markers in the millions. *Gigascience, 8,* 154. doi: 10.1093/gigascience/giy154

Jaakola, L. (2013). New insights into the regulation of anthocyanin biosynthesis in fruits. *Trends Plant Sci., 18*, 477–483. doi: 10.1016/j.tplants.2013.06.003

Jiménez-Gómez, J. M., Alonso-Blanco, C., Borja, A., Anastasio, G., Angosto, T., Lozano, R., *et al*. (2007). Quantitative genetic analysis of flowering time in tomato. *Genome, 50*, 303–315. doi: 10.1139/G07-009

Kang, J., and Sinha, N. R. (2010). Leaflet initiation is temporally and spatially separated in simple and complex tomato (*Solanum lycopersicum*) leaf mutants: A developmental analysis. *Botany, 88,* 710–724. doi: 10.1139/b10-051

Kudo, T., Kobayashi, M., Terashima, S., Katayama, M., Ozaki, S., Kanno, M., *et al*. (2017). TOMATOMICS: A web database for integrated omics information in tomato. *Plant Cell Physiol., 58,* e8(1–12). doi: 10.1093/pcp/pcw207

Letunic, I., and Bork, P. (2019). Interactive Tree of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res., 47*, 256–259. doi: 10.1093/nar/gkz239

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

Lin, Z., Griffith, M. E., Li, X., Zhu, Z., Tan, L., Fu, Y., *et al*. (2007). Origin of seed shattering in rice (*Oryza sativa* L.). *Planta, 226,* 11–20. doi: 10.1007/s00425-006-0460-4

Lin, Z., Li, X., Shannon, L. M., Yeh, C. T., Wang, M. L., Bai, G., *et al*. (2012). Parallel domestication of the *Shattering1* genes in cereals. *Nat. Genet., 44*, 720–724. doi: 10.1038/ng.2281

Lippman, Z., Semel, Y., and Zamir, D. (2007). An integrated view of quantitative trait variation using tomato inter-specific introgression lines. *Curr. Opin. Genet. Dev., 17*, 545–552. doi: 10.1016/j.gde.2007.07.007

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. doi: 10.1093/genetics/158.1.413

Mackay, I., and Powell, W. (2007). Methods for linkage disequilibrium mapping in crops. *Trends Plant Sci., 12,* 57–63. doi: 10.1016/j.tplants.2006.12.001

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*, 847789. doi: 10.3389/fpls.2022.847789

Martínez-Cuenca, M. R., Pereira-Dias, L., Soler, S., López-Serrano, L., Alonso, D., Calatayud, Á., *et al*. (2020). Adaptation to water and salt stresses of *Solanum pimpinellifolium* and *Solanum lycopersicum* var. *cerasiforme*. *Agronomy, 10,* 1169. doi: 10.3390/agronomy10081169

Meir, Z., Aviezer, I., Chongloi, G. L., Ben-Kiki, O., Bronstein, R., Mukamel, Z., *et al*. (2021). Dissection of floral transition by single-meristem transcriptomes at high temporal resolution. *Nat. Plants, 7,* 800–813. doi:10.1038/s41477-021-00936-8.

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. doi: 10.1104/pp.111.173997

Nakano, H., Kobayashi, N., Takahata, K., Mine, Y., and Sugiyama, N. (2016). Quantitative trait loci analysis of the time of floral initiation in tomato. *Sci. Hortic., 201*, 199–210. doi: 10.1016/j.scienta.2016.02.009

Nakayama, H., Ichihashi, Y., and Kimura, S. (2023). Diversity of tomato leaf form provides novel insights into breeding. *Breed. Sci., 73*, 76–85. doi: 10.1270/jsbbs.22061

Nesbitt, T. C., and Tanksley, S. D. (2001). *fw2.2* directly affects the size of developing tomato fruit, with secondary effects on fruit number and photosynthate distribution. *Plant Physiol., 127,* 575–583. doi: 10.1104/pp.010087

Nesbitt, T. C., and Tanksley, S. D. (2002). Comparative sequencing in the genus lycopersicon: Implications for the evolution of fruit size in the domestication of cultivated tomatoes. *Genetics, 162*, 365–379. doi: 10.1093/genetics/162.1.365

Ofner, I., Lashbrooke, J., Pleban, T., Aharoni, A., and Zamir, D. (2016). *Solanum pennellii* backcross inbred lines (BILs) link small genomic bins with tomato traits. *Plant J., 87*, 151–160. doi: 10.1111/tpj.13194

Olsen, K. M., Slimestad, R., Lea, U. S., Brede, C., Lovald, T., Ruoff, P., *et al*. (2009). Temperature and nitrogen effects on regulators and products of the flavonoid pathway: Experimental and kinetic model studies. *Plant, Cell Environ., 32*, 286–299. doi: 10.1111/j.1365-3040.2008.01920.x

Oróstica, K. Y., and Verdugo, R. A. (2016). Chromosome visualization tool: A whole genome viewer. *Int. J. Plant Genomics, 32*, 2366–2368. doi: 10.1155/2011/373875

Paran, I., Goldman, I., Tanksley, S. D., and Zamir, D. (1995). Recombinant inbred lines for genetic mapping in tomato. *Theor. Appl. Genet., 90,* 542–548. doi: 10.1007/BF00222001

Pascual, L., Desplat, N., Huang, B. E., Desgroux, A., Bruguier, L., Bouchet, J. P., *et al*. (2015). Potential of a tomato MAGIC population to decipher the genetic control of quantitative traits and detect causal variants in the resequencing era. *Plant Biotechnol. J., 13*, 565–577. doi: 10.1111/pbi.12282

Peralta, I. E., Spooner, D. M., Knapp, S. (2008). Taxonomy of wild tomatoes and their relatives (*Solanum* Sect. *Lycopersicoides*, Sect. *Juglandifolia*, Sect. *Lycopersicon*; Solanaceae). *Syst. Bot. Monogr., 84,* 1–186.

Pereira, L., Zhang, L., Sapkota, M., Ramos, A., Razifard, H., Caicedo, A. L., *et al*. (2021). Unraveling the genetics of tomato fruit weight during crop domestication and diversification. *Theor. Appl. Genet.*, *134*, 3363–3378. doi: 10.1007/s00122-021-03902-2

Petroni, K., and Tonelli, C. (2011). Recent advances on the regulation of anthocyanin synthesis in reproductive organs. *Plant Sci., 181,* 219–229. doi: 10.1016/j.plantsci.2011.05.009

Pin, P. A., and Nilsson, O. (2012). The multifaceted roles of *FLOWERING LOCUS T* in plant development. *Plant, Cell Environ., 35,* 1742–1755. doi: 10.1111/j.1365-3040.2012.02558.x

Pnueli, L., Gutfinger, T., Hareven, D., Ben-Naim, O., Ron, N., Adir, N., *et al*. (2001). Tomato SP-interacting proteins define a conserved signaling system that regulates shoot architecture and flowering. *Plant Cell, 13,* 2687–2702. doi: 10.1105/tpc.010293

Pook, T., Schlather, M., De Los Campos, G., Mayer, M., Carolin Schoen, C., and Simianer, H. (2019). Haploblocker: Creation of subgroup-specific haplotype blocks and libraries. *Genetics, 212,* 1045–1061. doi: 10.1534/genetics.119.302283

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet., 38,* 904–909. doi: 10.1038/ng1847

Quinet, M., Bataillel, G., Dobrev, P. I., Capel, C., Gómez, P., Capel, J.*, et al.* (2014). Transcriptional and hormonal regulation of petal and stamen development by *STAMENLESS*, the tomato (*Solanum lycopersicum* L.) orthologue to the B-class *APETALA3* gene. *J. Exp. Bot., 65,* 2243–2256. doi: 10.1093/jxb/eru089

Revelle, W. (2017). psych: Procedures for personality and phychological research. Northwestern University: Evanston, IL, USA.

Robinson, J., Thorvaldsdóttir, H., Turner, D., and Mesirov, J. (2023). igv.js: an embeddable JavaScript implementation of the Integrative Genomics Viewer (IGV). *Bioinformatics, 39,* btac830. doi: 10.1093/bioinformatics/btac830

Rothan, C., Diouf, I., and Causse, M. (2019). Trait discovery and editing in tomato. *Plant J., 97*, 73–90. doi: 10.1111/tpj.14152

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol., 4*, 406–425. doi: 10.1093/oxfordjournals.molbev.a040454

Salinas, M., Capel, C., Alba, J. M., Mora, B., Cuartero, J., Fernández-Muñoz, R., *et al*. (2013). Genetic mapping of two QTL from the wild tomato *Solanum pimpinellifolium* L. controlling resistance against two-spotted spider mite (*Tetranychus urticae* Koch). *Theor. Appl. Genet., 126,* 83–92. doi: 10.1007/s00122-012-1961-0

Scott, M. F., Ladejobi, O., Amer, S., Bentley, A. R., Biernaskie, J., Boden, S. A., *et al*. (2020). Multi-parent populations in crops: a toolbox integrating genomics and genetic mapping with breeding. *Heredity, 125,* 396–416. doi: 10.1038/s41437-020-0336-6

Semel, Y., Nissenbaum, J., Menda, N., and Zamir, D. (2006). Overdominant quantitative trait loci for yield and fitness in tomato. *Proc. Natl. Acad. Sci., 103*, 12981–12986. doi: 10.1073/pnas.0604635103

Shani, E., Burko, Y., Ben-Yaakov, L., Berger, Y., Amsellem, Z., Goldshmidt, A., *et al*. (2009). Stage-specific regulation of *Solanum lycopersicum* leaf maturation by class 1 KNOTTED1-LIKE HOMEOBOX Proteins. *Plant Cell, 21,* 3078–3092. doi: 10.1105/tpc.109.068148

Shwartz, I., Levy, M., Ori, N., and Bar, M. (2016). Hormones in tomato leaf development. Dev. Biol., 419, 132–142. doi: 10.1016/j.ydbio.2016.06.023

Silva, G. F. F., Silva, E. M., Correa, J. P. O., Vicente, M. H., Jiang, N., Notini, M. M., *et al*. (2019). Tomato floral induction and flower development are orchestrated by the interplay between gibberellin and two unrelated microRNA-controlled modules. *New Phytol., 221,* 1328–1344. doi: 10.1111/nph.15492

Snouffer, A., Kraus, C., and van der Knaap, E. (2020). The shape of things to come: ovate family proteins regulate plant organ shape. *Curr. Opin. Plant Biol., 53*, 98–105. doi: 10.1016/j.pbi.2019.10.005

Suresh, B. V., Roy, R., Sahu, K., Misra, G., and Chattopadhyay, D. (2014). Tomato genomic resources database: An integrated repository of useful tomato genomic information for basic and applied research. *PLoS One, 9*, e86387. doi: 10.1371/journal.pone.0086387

Szymkowiak, E. J., and Irish, E. E. (2006). *JOINTLESS* suppresses sympodial identity in inflorescence meristems of tomato. *Planta, 223*, 646–658. doi: 10.1007/s00425-005-0115-x

Tanksley, S. D., and Nelson, J. C. (1996). Advanced backcross QTL analysis: A method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. *Theor. Appl. Genet., 92*, 191–203. doi: 10.1007/BF00223376

Thissen, D., Steinberg, L., and Kuang, D. (2002). Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. *J. Educ. Behav. Stat., 27,* 77–83. doi: 10.3102/10769986027001077

Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., *et al*. (2001). Missing value estimation methods for DNA microarrays. *Bioinformatics, 17*, 520–525. doi: 10.1093/bioinformatics/17.6.520

Turck, F., Fornara, F., and Coupland, G. (2008). Regulation and identity of florigen: *Flowering Locus T* moves center stage. *Annu. Rev. Plant Biol., 59,* 573–594. doi: 10.1146/annurev.arplant.59.032607.092755

Ueta, R., Abe, C., Watanabe, T., Sugano, S. S., Ishihara, R., Ezura, H., *et al*. (2017). Rapid breeding of parthenocarpic tomato plants using CRISPR/Cas9. *Sci. Rep., 7,* 507. doi: 10.1038/s41598-017-00501-4

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. doi: 10.3389/fpls.2014.00227

Vilanova, S., Alonso, D., Gramazio, P., García-Fortea, E., Ferrante, P., Schmidt, M., *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. doi: 10.1186/s13007-020-00652-y

Wang, J., and Zhang, Z. (2021). GAPIT version 3: boosting power and accuracy for genomic association and prediction. *Genomics, Proteomics Bioinforma., 19,* 629–640. doi: 10.1016/j.gpb.2021.08.005

Wang, S., Chang, Y., and Ellis, B. (2016). Overview of OVATE FAMILY PROTEINS, a novel class of plant-specific growth regulators. *Front. Plant Sci., 7,* 417. doi: 10.3389/fpls.2016.00417

Wei, T., and Simko, V. (2017). R Package "Corrplot": Visualization of a Correlation Matrix. R Core Team: Vienna, Austria.

Wickham, H. (2009). ggplot2: Elegant graphics for data analysis. *Media, 35*, 10–1007.

Yu, J., Pressoir, G., Briggs, W. H., Bi, I. V., Yamasaki, M., Doebley, J. F., *et al*. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet., 38*, 203–208. doi: 10.1038/ng1702

Zhang, D., Ai, G., Ji, K., Huang, R., Chen, C., Yang, Z., *et al*. (2023). *EARLY FLOWERING* is a dominant gain-of-function allele of *FANTASTIC FOUR 1/2c* that promotes early flowering in tomato. *Plant Biotechnol. J*. doi: 10.1111/pbi.14217

Zhang, J., Chen, R., Xiao, J., Qian, C., Wang, T., Li, H., *et al*. (2007). A single-base deletion mutation in *SlIAA9* gene causes tomato (*Solanum lycopersicum*) *entire* mutant. *J. Plant Res., 120*, 671–678. doi: 10.1007/s10265-007-0109-9

Zhang, R., Jia, G., and Diao, X. (2023). geneHapR: an R package for gene haplotypic statistics and visualization. *BMC Bioinformatics, 24,* 199. doi: 10.1186/s12859-023-05318-9

Zhang, Y., Butelli, E., and Martin, C. (2014). Engineering anthocyanin biosynthesis in plants. *Curr. Opin. Plant Biol., 19,* 81–90. doi: 10.1016/j.pbi.2014.05.011

GENERAL DISCUSSION

Believe it or not, MAGIC populations are powerful breeding tools.

In the genomics era, there is an increasing trend towards the development of new multi-parental populations (Scott, *et al.,* 2020). Among them, the development of Multi-Parent Advanced Generation Inter-Cross (MAGIC) populations across several crops is experiencing a significant surge (Arrones *et al.,* 2020; Scott, *et al.,* 2020). However, developing MAGIC populations is not an easy job due to the complexity of the crossing scheme for incorporating multiple founders. Nevertheless, this scheme allows to increase genetic and phenotypic diversity, recombination rate, and mapping accuracy (Mackay and Powell, 2007; Cavanagh *et al.,* 2008). Given their great potential, new MAGIC populations are developed at an increasing pace (Samineni *et al.,* 2021; Escobar *et al.,* 2022; Kumar *et al.,* 2023; Yuan *et al.,* 2023).

Under the current climate change scenario, there is a real need of new crop varieties more resilient and efficient in the use of resources (Ray *et al.,* 2019). This situation highlights the CWRs as a powerful source of variation for breeding traits and their potential use in "introgressiomics" pre-breeding approaches (Prohens *et al.,* 2017). While integrating CWRs in breeding programs and experimental populations may be challenging, it represents a valuable approach for leveraging the available genetic resources.

The introduction of CWRs as founders of experimental populations opens new avenues for the exploitation of the wild variation. The development of inter-specific MAGIC populations gives rise to the generation of immortal individuals with multiple wild introgressions into an admixture background genome (Arrones *et al.,* 2020; Scott *et al.,* 2020). These individuals could be directly exploited not only for their intrinsic potential but also as donors carrying novel and unexpected QTL combinations (Huynh *et al.,* 2018). However, inter-specific populations could present some drawbacks, e.g., crossing barriers, low hybrid fertility, sterility, or undesirable agronomic traits. For this reason, in this doctoral thesis, we proposed the exploitation of the closest relative's diversity, which is easier and less time-consuming. The use of the closest relatives could avoid the implementation of specific techniques to overcome the reproductive barriers. In the eggplant MAGIC population, a *S. incanum* accession has been used as a founder, which is the closest wild relative to the cultivated eggplant, belonging to the secondary genepool (GP2). In the case of the tomato MAGIC population here presented, *S. lycopersicum* var. *cerasiforme* and *S. pimpinellifolium* accessions have been used as founders, which are the closest relative and the wild ancestor of the cultivated tomato, respectively.

The development of this doctoral thesis represents an important advance for eggplant and tomato breeding. With the development of the inter-specific eggplant MAGIC population, we aimed to reduce the gap in genetic and genomics tools available for this crop compared to other Solanaceae species (Gramazio *et al.,* 2018,

2019). The development of the inter-specific tomato MAGIC population complements the information already available with the new traits provided by the wild founders. This doctoral thesis not only entails a significant increase in the genetic diversity of both crops but also presents permanent resources for the scientific community for the elucidation of the genetic basis of traits of interest.

1. Characterization of the first eggplant MAGIC population for anthocyanin-related traits

Here we presented the first MAGIC population developed in eggplant (S3MEGGIC), being the largest recombinant population so far (Mangino *et al.,* 2022). The inter-specific S3MEGGIC was developed by inter-crossing seven *S. melongena* accessions and one wild *S. incanum*. During the first inter-crosses, the *S. incanum* accession was used as a female founder, and individuals derived from this cross dragged the maternal cytoplasmic background of the wild parent. Although subsequent generations suffered partial sterility (Khan *et al.,* 2020; Isshiki *et al.,* 2021), the PCA analysis revealed the absence of population structure since no differentiated cytoplasmic groups were observed.

The genotyping of the final 420 S3MEGGIC individuals using the 5k probes eggplant SPET platform demonstrated its potential as a high-throughput and high-efficiency genotyping platform extending the use of this technology to multi-parent populations (Barchi *et al.,* 2019). Combining targeted SNPs analysis with method simplicity, the SPET platform represents an interesting alternative to GBS or microarrays (Scheben *et al.,* 2017). Genotyping results revealed a low residual heterozygosity of the S3MEGGIC individuals (6.87%), similar to the expected value for an F5-like bi-parental inter-cross generation (6.25%). However, a cryptic selection process may have caused an unbalanced representation of the founder genomes (Rockman and Kruglyak, 2008; Thépot *et al.,* 2015). This could be explained by the inability of the genotyping density to efficiently distinguish between the founders that are genetically closer, such as founders D and E (Dell'Acqua *et al.,* 2015); or by the effect of different biological processes, such as seed dormancy, delayed germination, precocity, reduced fertility, and parthenocarpy associated to some genomes (Barchi *et al.,* 2010; Khan *et al.,* 2015; Prohens *et al.,* 2017). In fact, during the population construction, we observed a recalcitrant germination and an irregular flowering and fruit setting in individuals carrying the wild cytoplasm, which were strongly affected by environmental conditions.

In this doctoral thesis, the S3MEGGIC population was evaluated for different traits to identify the candidate genes involved and to widen the genetic knowledge of this crop. Firstly, the S3MEGGIC population was assessed for anthocyanin-

related traits, including anthocyanin presence in vegetative plant tissues (PA) and fruit epidermis (FA), and light-insensitive anthocyanin pigmentation under the calyx (PUC). Anthocyanins are responsible for the purple colour of the eggplant fruit peel and play a key role in plant defence mechanisms (Daunay and Hazra, 2012; Zhou *et al.,* 2019; Li *et al.,* 2021b). Developing dark uniform purple-coloured eggplants, which results from the combination of anthocyanins with chlorophylls all along the fruit, is a major objective in eggplant breeding programmes (Li *et al.,* 2018). For these reasons, together with the high stability and heritability, and its ease of score, we selected PA, FA, and PUC traits to identify the main responsible of the eggplant anthocyanin biosynthesis.

The *MYB113* gene was identified as the best candidate gene for PA and FA since it is a well-known regulatory transcription factor controlling anthocyanin synthesis in eggplant (Zhou *et al.,* 2019; Shi *et al.,* 2021; Yang *et al.,* 2022). The *COP1* gene was found the best candidate for PUC since it is involved in light-regulated gene expression and development (Jiang *et al.,* 2016; Li *et al.,* 2018; He *et al.,* 2019; Naeem *et al.,* 2019). Variants that predicted high-impact effects on protein function were identified for founders that do not present PA or FA, and for founders able to synthesise PUC. Further analysis evidenced a duplication of an ancestral *MYB113* gene with a translocation from chromosome 10 to chromosome 1 compared with the tomato reference genome (Barchi *et al.,* 2019; Hosmani *et al.,* 2019). These results could suggest a duplication event occurred during eggplant evolution and may both genes required for anthocyanin synthesis. However, other available reference genomes pointed to a genomic reorganization and both genes colocalized in chromosome 10 (Wei *et al.,* 2020b; Li *et al.,* 2021a). A new version of the 67/3 eggplant reference genome developed by Barchi *et al.* (2019), which is currently under development, will shed light on the eggplant genome organization to trace the evolutionary changes of the *MYB113* gene. Recently, other candidate genes have been proposed for non-photosensitive and less-photosensitive anthocyanin biosynthesis (He *et al.,* 2022; Luo *et al.,* 2023). In any case, the candidate genomic region consistently maps to the beginning of chromosome 10. Further studies should be performed to clarify the molecular mechanisms underlying the light-dependent anthocyanin biosynthesis.

With this study, the S3MEGGIC population showcased its potential utility for association studies, enabling the establishment of marker associations with anthocyanin-related genes and the identification of candidate genes for PA and FA. Furthermore, it allowed the first identification of a candidate gene for an economically relevant breeding trait in eggplant such as PUC.

2. Identification of the main responsible gene for the biosynthesis of chlorophylls in the eggplant fruit peel

Subsequently, the S3MEGGIC population was screened for eggplant fruit green pigmentation (FC) (Arrones *et al.,* 2022). Chlorophylls confer a green colour to eggplant fruit peel, contributing to a darker background and reinforcing fruit darkness when anthocyanins are present (Daunay *et al.,* 2004). Not only for the chlorophylls influence on fruit colour and appearance, but also for their potential effect on fruit composition since they are involved in photosynthesis, makes their study interesting for eggplant breeding.

The *APRR2* gene was identified as the best candidate for FC since it had previously been described as a chloroplast development promotor in several solanaceous and cucurbitaceous crops (Pan *et al.,* 2013; Liu *et al.,* 2016; Oren *et al.,* 2019). Variants that predicted high-impact effects on protein function were predicted for founders that do not present FC, but also an additional deletion was identified when retrieving the gene sequence in two eggplant reference genomes developed from accessions without FC. Two different variants were observed when validating our hypothesis in a set of accessions from the G2P-SOL eggplant core collection. In conclusion, five disruptive mutations were identified in the *APRR2* gene sequence explaining the lack of fruit chlorophyll pigmentation.

The phylogenetic analysis of *APRR2* gene revealed orthology within Solanaceae and suggests that specialization of *APRR2-like* genes occurred independently in Cucurbitaceae and Solanaceae since no clear evolutionary relation was observed. Furthermore, a strong geographical differentiation was observed in the frequency of the identified variants in the eggplant *APRR2* gene which could be related to the domestication flows of eggplant together with local preferences (Meyer *et al.,* 2012; Davidar *et al.,* 2015; Page *et al.,* 2019). This indicates that this phenotype may have arisen and been selected independently several times and suggests a prominent role of this trait in eggplant domestication and diversification.

This study represented the first evidence of *APRR2* as the main actor involved in eggplant FC. These results were validated through the use of different experimental materials combined with advanced association and bioinformatic analysis and classical genetic tools. Our finding was also supported by the conserved function of the *APRR2* gene in regulating fruit green pigmentation of different vegetable crops (Pan *et al.,* 2013; Jiao *et al.,* 2017; Oren *et al.,* 2019; Jeong *et al.,* 2020). In addition, subsequently published articles confirmed that the eggplant *APRR2* gene regulates the green/white colour of the fruit peel (Fang *et al.,* 2023; Lv *et al.,* 2023). The dissection of the genetics of the FC trait will be extremely useful to foster future breeding programs focused not only on specific market demands based on visual preferences but also on developing new crop varieties with enhanced

nutritional quality and increased resilience to forthcoming environmental challenges, due to the role of chlorophylls in stress biology of plants.

3. Identification of the main responsible gene for the irregular fruit green netting pattern in the eggplant fruit peel

The GWAS performed for the irregular fruit green netting pattern in the eggplant fruit peel (FN) revealed a minor but significant peak on chromosome 4. Previous studies associated the same region to a chlorophylls-related trait referred to as netting, reticulation, or variegation (Tigchelaar, 1968; Doganlar *et al.,* 2002; Daunay *et al.,* 2004; Frary *et al.,* 2014). The S3MEGGIC phenotyping for fruit chlorophyll pigmentation was performed using binary classification (presence/absence) (Arrones *et al.,* 2022). Unfortunately, we later realised that the distribution of chlorophylls can be uniform or irregular, the latter being named as fruit green netting. This evidence led us to hypothesize a possible relationship of the association peak on chromosome 4 and the FN trait.

The FN trait is manifested as a gradient of dark green netting intensified in the proximal part of the fruit on a pale green background, reminiscent of the green shoulder trait described in tomato (Powell *et al.,* 2012; Nguyen *et al.,* 2014). This trait is common in wild relatives and some eggplant landraces, while is rarely found in commercial varieties. The introduction of the *S. incanum* founder has provided the S3MEGGIC population with wild traits, such as the FN (Arrones *et al.,* 2023). The recovery of this trait lost during domestication contributes to a further increase in the colour diversity of eggplant.

The *GLK2* gene was identified as the best candidate for FN since it had previously been described as a positive regulator of chloroplast development and pigment accumulation in other Solanaceae species (Powell *et al.,* 2012; Brand *et al*, 2014; Nguyen *et al.,* 2014). We used over 2,300 individuals from different populations, including the S3MEGGIC population for candidate genomic region identification, an F_2 population for BSA-Seq, and advanced backcrosses for edges-to-core fine mapping. A better annotation of the *GLK2* gene sequence allowed us to identify a single bp insertion leading to a premature stop codon. The synthesis of a truncated GLK2 protein was confirmed by Western blotting analysis. The gene sequence was also analysed in a set of *S. melongena* accessions from the G2P-SOL eggplant core collection and wild relative accessions. Two new frameshift mutations were identified explaining the absence of FN. Differences in chloroplast number and structure between FN (irregular chlorophylls pattern) and FC (uniform chlorophylls distribution) samples were observed by confocal microscopy. More developed

chloroplasts with higher chlorophylls content were observed in FN samples coming from the proximal part of a fruit with dark green netting.

By using multiple *in silico* and *in vivo* methodologies we found that the *GLK2* gene is responsible for the eggplant FN trait. This study allowed us to better understand the difference between both chlorophyll-related traits, as well as to explore the different layers of pigmentation in the eggplant fruit peel. The rescue of this wild trait could be interesting for future eggplant visual quality breeding programmes, particularly to widen the diversity of the fruit colour palette. In addition, the *GLK2* gene could also be a valuable target for enhancing nutritional quality due to a higher concentration of functional chloroplasts and chlorophyll accumulation, and therefore an increased eggplant fruit photosynthesis (Blanke, 1989).

4. Development of a tomato MAGIC population by the inter-cross of four *S. lycopersicum* var. *cerasiforme* and four wild *S. pimpinellifolium* founders

The inter-specific tomato MAGIC (ToMAGIC) population represented a novel source of genetic and phenotypic diversity due to the combination of *S. lycopersicum* var. *cerasiforme* (SLC) and *S. pimpinellifolium* (SP) genomes (Arrones *et al.,* 2024). The eight ToMAGIC founders were selected based on a representation of the genetic diversity and geographical distribution of the two taxa, recovering the Andean variability lost during domestication (Blanca *et al.,* 2015; Gramazio *et al.,* 2020).

The genotyping of the final 354 ToMAGIC lines using the newly developed 12k probes tomato SPET panel generated more than 4 million SNPs due to the increased genetic variation of the population given by the use of SLC and SP founders. The 6,488 markers retained after filtering were distributed almost uniformly along the chromosomes, although with a lower density in the pericentromeric regions. The genotypic data revealed the absence of genetic structure avoiding false-positive results, which is one of the advantages of MAGIC populations (Arrones *et al.,* 2020). At least 41 PCs are required to explain only a 20% of the genetic variation, underscoring the weak population structure within the ToMAGIC population. Kinship analysis also revealed a low genetic relatedness among ToMAGIC lines. This reinforces the absence of population structure and indicates that variation was maximized in the final lines (Huang *et al.,* 2015; Arrones *et al.,* 2020). Genotyping results also revealed a balanced representation of the founder genomes to the final population (Pook *et al.,* 2019). The average contribution of each founder to the overall population was around 12.50%, which is the theoretical expected value for a population developed from eight founders.

However, different haplotype block proportions were identified depending on the genomic position in all chromosomes.

In this doctoral thesis, the ToMAGIC population was tested for a proof-of-concept analysis where the final lines were phenotyped for specific traits across different plant parts. One of the aims of this population was to dissect the control of different traits, including those involved in the early domestication of tomato (Frary and Doganlar, 2003). Specifically, the ToMAGIC lines were phenotyped for different traits including locule number, fruit weight, plant anthocyanin, leaf lobing/serration, leaf complexity, and number of leaves below the first inflorescence. The population exhibited a large phenotypic diversity including transgressive lines, with a dominant effect of wild alleles over domesticated alleles (Semel *et al.,* 2006). The finding of strong associations for the *WUSCHEL* and *FW2.2* genes with respectively the locule number and fruit weight traits (Frary *et al.,* 2000; Muños *et al.,* 2011), together with the previously described mutations found in the large-fruited accessions, allowed to validate the usefulness and precision of our population. The identification of a novel mutation in the *MYB-ATV* gene resulting in plant anthocyanin pigmentation highlighted the huge diversity of the population, including novel phenotypic-causing variants (Colanero *et al.,* 2018). The pinpointing of candidate genomic regions harbouring a large number of genes related to leaf morphology and earliness demonstrated the potential of the population for mapping and suggesting new candidate genes.

In conclusion, the ToMAGIC population represents a landmark breeding material for future genetics and breeding studies in tomato due to the high degree of genetic and phenotypic diversity provided by the founders, the huge genotyping data generated, the absence of population structure, and the balanced representation of the founder genomes, making it very useful for fine-mapping of complex traits. This approach validated previously identified candidate genes but also led to the discovery of new candidates and the observation of novel phenotypic-causing variants. The introduction of SLC and SP founders can also be a tool of great relevance for studying the genetic changes in the early stages of tomato domestication. ToMAGIC lines pyramiding multiple traits of interest could be directly integrated into breeding pipelines providing unexploited variation for tomato breeding.

5. Concluding remarks and future perspectives

Developing MAGIC populations using CWRs as founders is a promising strategy for harnessing the potential of these experimental populations for QTL/gene mapping together with the exploitation of the large phenotypic and genetic variation

offered by wild introgressions. These populations in combination with high-throughput phenotyping and genotyping techniques result in the generation of new elite plant materials including a proportion of the lost ancestral diversity. This constitutes the cornerstone on which the studies complied in this doctoral thesis were developed.

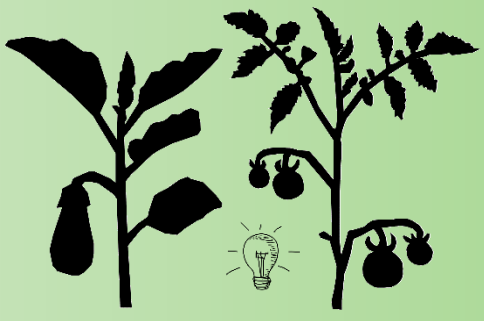
With the development of two inter-specific MAGIC populations in two major Solanaceae crops, eggplant and tomato, we have demonstrated their power to dissect the genetic control of different traits of interest. In the S3MEGGIC population clear candidates have been proposed for anthocyanin and chlorophyll pigmentation in the fruit peel. The next step would be their functional validation with reverse genetics approaches, such as the over-expression of the candidate genes or their knock-out by CRISPR/Cas. The emergent CRISPR/Cas technology allows gene editing by introducing point mutations in one or multiple target sequences simultaneously (Doudna and Charpentier, 2014). Although in tomato genome editing is feasible and several traits have already been edited (Tiwari *et al.,* 2023), only one example has been reported so far in eggplant (Maioli *et al.,* 2020). Great efforts are being made to optimize an eggplant transformation protocol. However, due to its recalcitrance to genetic transformation (García-Fortea *et al.,* 2020; Mir *et al.,* 2021), there is a need for alternative indirect validation methods. Here, different methodologies have been used such as the study of synteny with other species, validation on a set of accessions from the eggplant core collection, Western blotting analysis, and confocal microscopy observations.

The low level of heterozygosity of the final individuals is translated into permanent immortal mapping populations of great relevance for the scientific community (Arrones *et al.,* 2020; Scott *et al.,* 2020). This doctoral thesis has mainly been focused on the study of fruit pigment biosynthesis, the identification of responsible genes and their causative polymorphisms. However, both MAGIC populations present a plethora of traits that can be dissected. Wild founders have been selected due to their tolerance to some biotic and abiotic stresses (Knapp *et al*. 2013; Blanca *et al.,* 2015). Currently, the S3MEGGIC population has been advanced to the S5 generation (S5MEGGIC) and the final 326 lines have been resequenced 3X. The final population has been phenotyped for plant, fruit, and root traits, and a set of S5MEGGIC lines are being evaluated under water stress conditions. The ToMAGIC population has also been phenotyped for several plant and inflorescence architecture, leaf, flower, and fruit traits. Furthermore, founders and a set of ToMAGIC lines are being assessed for tolerance to water deficit, nitrogen use efficiency, *Fusarium oxysporum*, *Verticillium dahliae*, *Clavibacter michiganensis*, and *Xanthomonas campestris*. The identification of interesting alleles in a multi-environment or multi-trait background could lead to the origination of elite materials.

Overall, this thesis provides useful phenomic and genomic information, together with advanced materials with relevant implications for eggplant and tomato genetics and breeding. The inter-specific eggplant and tomato MAGIC populations have demonstrated the utility of CWRs in the development of experimental populations and contribute to broadening the genetic base of both crops.

GENERAL CONCLUSIONS

It is all over but the shouting.



1. The first eggplant MAGIC (S3MEGGIC) population developed by inter-crossing seven *S. melongena* and one *S. incanum* accessions represents the largest recombinant population developed so far in eggplant. The genotyping of the S3MEGGIC individuals with the 5k probes eggplant SPET platform revealed a lack of genetic structure and low residual heterozygosity.

2. The phenotyping and genotyping of the inter-specific S3MEGGIC population demonstrated its potential utility for association studies, enabling the establishment of marker associations with anthocyanin-related traits. Candidate genes were proposed for anthocyanin presence in vegetative plant tissues and fruit epidermis, and light-insensitive anthocyanin pigmentation under the calyx. Causal variants explaining the absence of anthocyanin biosynthesis were identified.

3. The S3MEGGIC population was also phenotyped for fruit chlorophyll pigmentation. A clear candidate gene (*Sm*APRR2) was identified and different causal variants were found explaining the absence of fruit chlorophylls in the population, in the available reference genomes, and in a set of accessions from the eggplant G2P-SOL core collection. A strong geographical differentiation was observed in the frequency of the identified variants which seemed to be related to domestication flows and local preferences.

4. The introduction of the wild *S. incanum* as one of the S3MEGGIC founders provided the final lines with the wild fruit green netting trait. The main responsible gene controlling this trait (*Sm*GLK2) was identified and different causal variants explaining the absence of the irregular chlorophylls pattern were found by analysing over 2,300 individuals from different populations. The main causal variant led to a disruption of the protein conformation and function, which was confirmed by Western blotting analysis.

5. Understanding the difference between uniform chlorophyll distribution and green fruit netting allowed us to explore the different layers of pigmentation of the eggplant fruit peel. Differences in chloroplast number and structure from tissue samples coming from both tissues were analysed by confocal microscopy observations.

6. The identification of the candidate genes involved in the eggplant fruit coloration will be extremely useful to widen the diversity of the fruit colour palette, and to foster future breeding programs focused on the development of new crop varieties with enhanced nutritional quality due to the antioxidant properties of the anthocyanins and the photosynthetic activity of the chlorophylls.

7. With the development of the inter-specific tomato MAGIC (ToMAGIC) population developed by inter-crossing four *S. l.* var. *cerasiforme* and four *S. pimpinellifolium* accessions we aimed to exploit the Andean variability lost during the domestication process by using a substantial proportion of the fully cross-compatible weedy and wild tomato diversity.

8. The genotyping of the ToMAGIC lines with the newly developed 12k probes tomato SPET panel generated more than 4 million SNPs due to the increased genetic diversity included by the *S. l.* var. *cerasiforme* and *S. pimpinellifolium* founders. The genotyping results revealed the absence of population structure and a balanced representation of the founder genomes.

9. Our proof-of-concept focusing on a subset of traits from different plant parts has revealed a large phenotypic diversity in the ToMAGIC population, which allowed to validate previously described candidate genes, propose new candidate genes, and identify novel phenotypic-causing variants for different traits of interest related to tomato domestication.

10. The development of two inter-specific MAGIC populations in two major Solanaceae crops, eggplant and tomato, using CWRs as founders has showcased the potential of these experimental populations for including a huge phenotypic and genetic diversity and for the fine mapping of traits of interest. These immortal populations could be directly exploited as elite materials or be used as super-trait donors carrying novel QTL combinations in a multi-trait background.

GENERAL REFERENCES

Piece of cake, see you tomorrow at eight.

Arrones, A., Antar, O., Pereira-Dias, L., Solana, A., Ferrante, P., Aprea, G., Plazas, M., Prohens, J., Díez, M. J., Giuliano, G., Gramazio, P., and Vilanova, S. (2024). Harnessing tomato genetic resources: development of an inter-specific (*Solanum lycopersicum* var. *cerasiforme* and *S. pimpinellifolium*) tomato MAGIC population for enhanced trait mapping and candidate gene identification. *bioRxiv*.

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. doi: 10.3389/fpls.2022.1025951

Arrones, A., Manrique, S., Baraja-Fonseca, V., Plazas, M., Prohens, J., Portis, E., *et al*. (2023). The irregular fruit green netting: An eggplant domestication trait controlled by the *SmGLK2* gene with implications in fruit colour diversification. *bioRxiv*, 2023-06. doi: 10.1101/2023.06.28.546667

Arrones, A., Vilanova, S., Plazas, M., Mangino, G., Pascual, L., Díez, M. J., *et al*. (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*, 229. doi: 10.3390/biology9080229

Aubriot, X., and Daunay, M. C. (2019). "The eggplant genome" in *Domestication of eggplants: Eggplant and relatives: from exploring their diversity and phylogenetic relationship to conservation challenges* (Switzerland: Springer Nature).

Barchi, L., Lanteri, S., Portis, E., Stàgel, A., Valè, G., Toppino, L., *et al*. (2010). Segregation distortion and linkage analysis in eggplant (*Solanum melongena* L.). *Genome*, *53*, 805-815. doi: 10.1139/g10-073

Barchi, L., Pietrella, M., Venturini, L., Minio, A., Toppino, L., Acquardo, A., *et al*. (2019). A chromosome-anchored eggplant genome sequence reveals key events in solanaceae evolution. *Sci. Rep., 9*(1), 1–13. doi: 10.1038/s41598-019-47985-w

Bauchet, G., and Causse, M. (2012). "Genetic diversity in plants" in *Genetic diversity in tomato (Solanum lycopersicum) and its wild relatives, 8*, 134–162.

Blanca, J., Cañizares, J., Cordero, L. Pascual, L., Díez, 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. doi: 10.1371/journal.pone.0048198

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*(1), 1–19. doi: 10.1186/s12864-015-1444-1

Blanca, J., Sanchez-Matarredona, D., Ziarsolo, P., Montero-Pau, J., Van der Knaap, E., Díez, M. J., *et al*. (2022). Haplotype analyses reveal novel insights into tomato history and domestication driven by long-distance migrations and latitudinal adaptations. *Horticulture Research, 9*. doi: 10.1093/hr/uhac030

Blanke, M. M., and Lenz, F. (1989). Fruit photosynthesis. *Plant, Cell Environment, 12,* 31–46. doi: 10.1111/j.1365-3040.1989.tb01914.x

Bombarely, A., Menda, N., Tecle, I. Y., Buels, R. M., Strickler, S., Fischer-York, T*et al*. (2010). The Sol Genomics Network (solgenomics. net): growing tomatoes using Perl. *Nucleic Acids Res., 39*(suppl_1), D1149-D1155. doi: 10.1093/nar/gkq866

Boopathi, N. M. (2020). "Genetic mapping and marker assisted selection: Basics, practice and benefits, second edition" in *Genetic Mapping and Marker Assisted Selection: Basics, Practice and Benefits, Second Edition*. doi: 10.1007/978-981-15-2949-8

Brand, A., Borovsky, Y., Hill, T., Rahman, K. A. A., Bellalou, A., Van Deynze, A., and Paran, I. (2014). *CaGLK2* regulates natural variation of chlorophyll content and fruit

color in pepper fruit. *Theor. Appl. Genet., 127,* 2139–48. doi: 10.1007/s00122-014-2367-y

Campanelli, G., Sestili, S., Acciarri, N., Montemurro, F., Palma, D., Leteo, F., *et al*. (2019). Multi-parental advances generation inter-cross population, to develop organic tomato genotypes by participatory plant breeding. *Agronomy*, *9*(3), 119. doi: 10.3390/agronomy9030119

Causse, M., Desplat, N., Pascual, L., Le Paslier, M. C., Sauvage, C., Bauchet, *et al*. (2013). Whole genome resequencing in tomato reveals variation associated with introgression and breeding events. *BMC Genomics*, *14*(1). doi: 10.1186/1471-2164-14-791

Causse, M., Friguet, C., Coiret, C., Lépicier, M., Navez, B., Lee, M., *et al*. (2010). Consumer preferences for fresh tomato at the European scale: a common segmentation on taste and firmness. *J. Food Sci.*, *75*(9), S531–S541. doi: 10.1111/j.1750-3841.2010.01841.x

Cavanagh, C., Morell, M., Mackay, I., and Powell, W. (2008). From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. *Curr. Opin. Plant Biol., 11*, 215–221. doi: 10.1016/j.pbi.2008.01.002

Chakrabarti, M., Zhang, N., Sauvage, C., Mu~nos, S., Blanca, J., Cañizares, J., Diez, M.J., *et al*. (2013). A cytochrome P450 regulates a domestication trait in cultivated tomato. *Proc. Nat.l Acad. Sci*. *USA*, *110*, 17125–17130. doi: 10.1073/pnas.1307313110

Cockram, J., and 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*. Cham: Springer, 109–138. doi: 10.1007/10_2017_48

Colanero, S., Perata, P., and Gonzali, S. (2018). The *atroviolacea* gene encodes an R3-MYB protein repressing anthocyanin synthesis in tomato plants. *Front. Plant. Sci., 9,* 830. doi: 10.3389/fpls.2018.00830

Chapman, M. A. (2019). "The eggplant genome" in *Introduction: the importance of eggplant* (Switzerland: Springer Nature).

Daunay, M. C., Aubert, S., Frary, A., Doganlar, S., Lester, R. N., Barendse, G., *et al.* (2004). "Eggplant (*Solanum melongena*) fruit colour: pigments, measurements and genetics" in *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., and Hazra, P. (2012). "Eggplant," in *Handbook of Vegetables*, eds K. V. Peter and P. Hazra (Houston, TX: Studium Press), 257–322.

Davidar, P., Snow, A. A., Rajkumar, M., Pasquet, R., Daunay, M. -C., and Mutegi, E. (2015). The potential for crop to wild hybridization in eggplant (*Solanum melongena*; solanaceae) in southern India. *Am. J. Bot., 102*(1), 129–139. doi: 10.3732/ajb.1400404

Dell'Acqua, M., Gatti, D. M., Pea, G., Cattonaro, F., Coppens, F., Magris, G., *et al*. (2015). Genetic properties of the MAGIC maize population: a new platform for high definition QTL mapping in *Zea mays*. *Genome Biol., 16*, 1–23. doi: 10.1186/s13059-015-0716-z

Dempewolf, H., Hodgins, K. A., Rummell, S. E., Ellstrand, N. C., and Rieseberg, L. H. (2012). Reproductive isolation during domestication. *Plant Cell*, *24*(7), 2710–2717. doi: 10.1105/tpc.112.100115

Doerge, R. W. (2002). Mapping and analysis of quantitative trait loci in experimental populations. *Nat. Rev. Genet., 3*, 43–52. doi: 10.1038/nrg703

Doganlar, S., Frary, A., Ku, H. M., and Tanksley, S. D. (2002). Mapping quantitative trait loci in inbred backcross lines of *Lycopersicon pimpinellifolium* (LA1589). *Genome*, *45*(6), 1189–1202. doi: 10.1139/g02-091

Doudna, J. A., and Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. *Science, 346*, 1258096. doi: 10.1126/science.1258096

Escobar, E., Oladzad, A., Simons, K., Miklas, P., Lee, R. K., Schroder, S., *et al*. (2022). New genomic regions associated with white mold resistance in dry bean using a MAGIC population. *Plant Genome, 15*(1), e20190. doi: 10.1002/tpg2.20190

Eshed, Y., and Zamir, D. (1995). An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. *Genet.*, *141*(3), 1147–1162. doi: 10.1093/genetics/141.3.1147

Fang, H., Wang, P., Wang, W., Peng, J., Zheng, J., Zhu, G., *et al*. (2023). Fine mapping and identification of *SmAPRR2* regulating rind color in eggplant (*Solanum melongena* L.). *Int. J. Mol. Sci., 24*, 3059. doi: 10.3390/ijms24043059

FAOSTAT. Available online at: http://www.fao.org (Accessed May 18, 2023).

Flores-Saavedra, M., Gramazio, P., Vilanova, S., Mircea, D.M., Ruiz-González, M.X., Vicente, O., Prohens, J., Plazas, M. (2024). Evaluation of a set of eggplant (*Solanum melongena*) lines with introgressions of *S. incanum* under water stress conditions. *bioRxiv*, 2023-05. doi: 10.1101/2023.05.04.53937

Frary, A., Frary, A., Daunay, M.C., Huvenaars, K., Mank, R., and Doğanlar, S. (2014). QTL hotspots in eggplant (*Solanum melongena*) detected with a high-resolution map and CIM analysis. *Euphytica*, *197*, 211–228.

Frary, A., and Doganlar, S. (2003). Comparative genetics of crop plant domestication and evolution. *Turkish J. Agric. For., 27*, 59–69.

Frary, A., Nesbitt, T. C., Frary, A., Grandillo, S., Knaap, E. V. D., Cong, B., *et al*. (2000). *fw2.2*: a quantitative trait locus key to the evolution of tomato fruit size. *Science, 289*(5476), 85-88. doi: 10.1126/science.289.5476.85

Figàs, M. R., Prohens, J., Raigón, M. D., Fita, A., García-Martínez, M. D., Casanova, C., *et al*. (2015). Characterization of composition traits related to organoleptic and functional quality for the differentiation, selection and enhancement of local varieties of tomato from different cultivar groups. *Food Chem.*, *187*, 517–524. doi: 10.1016/j.foodchem.2015.04.083

Fu, Y.-B., and Dong, Y.-B. (2015). Genetic erosion under modern plant breeding: case studies in Canadian crop gene pools. *Genet. Diversity Erosion Plants*, 89–104. doi: 10.1007/978-3-319-25637-5_4

Fulop, D., Ranjan, A., Ofner, I., Covington, M. F., Chitwood, D. H., West, D., *et al*. (2016). A new advanced backcross tomato population enables high resolution leaf QTL mapping and gene identification. *G3 Genes, Genomes, Genet., 6,* 3169–3184. doi: 10.1534/g3.116.030536

García-Fortea, E., Lluch-Ruiz, A., Pineda-Chaza, B. J., García-Pérez, A., Bracho-Gil, J. P., Plazas, M., *et al*. (2020). A highly efficient organogenesis protocol based on zeatin riboside for in vitro regeneration of eggplant. *BMC Plant Biol., 20.* doi: 10.1186/s12870-019-2215-y

Gascuel, Q., Diretto, G., Monforte, A. J., Fortes, A. M., and Granell, A. (2017). Use of natural diversity and biotechnology to increase the quality and nutritional content of tomato and grape. *Front. Plant Sci.*, *8*, 652. doi: 10.3389/fpls.2017.00652

Gasparini, K., dos Reis Moreira, J., Peres, L. E. P., and Zsögön, A. (2021). De novo domestication of wild species to create crops with increased resilience and nutritional value. *Curr. Opin. Plant Biol.*, *60*, 102006. doi: 10.1016/j.pbi.2021.102006

Gonzali, S., and Perata, P. (2021). Fruit colour and novel mechanisms of genetic regulation of pigment production in tomato fruits. *Horticulturae*, *7*(8), 259. doi: 10.3390/horticulturae7080259

Gramazio, P., Alonso, D., Arrones, A., Villanueva, G., Plazas, M., Toppino. L., *et al*. (2023). Conventional and new genetic resources for an eggplant Green Revolution. *J. Ex. Bot.,* erad260. doi: 10.1093/jxb/erad260

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 inter-specific tomato MAGIC population. *Hortic. Res., 7,* 174.

Gramazio, P., Prohens, J., Plazas, M., Mangino, G., Herraiz, F.J., García-Fortea, E., and Vilanova, S. (2018). Genomic tools for the enhancement of vegetable crops: A case in eggplant. *Not. Bot. Horti. Agrobot. Cluj-Napoca, 46*, 1-13. doi: 10.15835/nbha46110936

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*, 1477. doi: 10.3389/fpls.2017.01477

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. doi: 10.3389/fpls.2019.01220

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

Grandillo, S., Termolino, P., and van der Knaap, E. (2013). Molecular mapping of complex traits in tomato. *Genet., Genomics, Breed. Tomato*, 150–227.

Gürbüz, N., Uluişik, S., Frary, A., Frary, A., and Doğanlar, S. (2018). Health benefits and bioactive compounds of eggplant. *Food Chem.*, *268*(May), 602–610. doi: 10.1016/j.foodchem.2018.06.093

Hajjar, R., and Hodgkin, T. (2007). The use of wild relatives in crop improvement: A survey of developments over the last 20 years. *Euphytica*, *156*(1–2), 1–13. doi: 10.1007/s10681-007-9363-0

Harlan, J. R., and de Wet, J. M. J. (1971). Toward a rational classification of cultivated plants. *Taxon*, *20*, 509–517. doi: 10.2307/1218252

He, Y., Chen, H., Zhou, L., Liu, Y., and Chen, H. (2019). Comparative transcription analysis of photosensitive and non-photosensitive eggplants to identify genes involved in dark regulated anthocyanin synthesis. *BMC Genomics, 20,* 678. doi: 10.1186/s12864-019-6023-4

He, Y., Li, S., Dong, Y., Zhang, X., Li, D., Liu, Y., and Chen, H. (2022). Fine mapping and characterization of the dominant gene SmFTSH10 conferring non-photosensitivity in eggplant (*Solanum melongena* L.). Theoretical and Applied Genetics, 135(7), 2187-2196.

Hirakawa, H., Shirasawa, K., Miyatake, K., Nunome, T., Negoro, S., Ohyama, A., *et al*. (2014). Draft genome sequence of eggplant (*Solanum melongena* L.): the representative *Solanum* species indigenous to the old world. *DNA Res.*, *21*(6), 649–660. doi: 10.1093/dnares/dsu027

Hosmani, P.S., Flores-Gonzalez, M., van de Geest, H., *et al*. (2019). An improved de novo assembly and annotation of the tomato reference genome using single-molecule sequencing, Hi-C proximity ligation and optical maps. *bioRxiv,* 767764. doi: 10.1101/767764

Huang, B. E., Verbyla, K. L., Verbyla, A. P., Raghavan, C., Singh, V. K., Gaur, P., *et al.* (2015). MAGIC populations in crops: current status and future prospects. *Theor. Appl. Genet., 128,* 999–1017. doi: 10.1007/s00122-015-2506-0

Huynh, B. L., Ehlers, J. D., Huang, B. E., Muñoz-Amatriaín, M., Lonardi, S., Santos, J. R. P., *et al*. (2018). A multi-parent advanced generation inter-cross (MAGIC) population for genetic analysis and improvement of cowpea (*Vigna unguiculata* L. Walp.). *Plant J., 93,* 1129-1142. doi: 10.1111/tpj.13827

Isshiki, S., Nakamura, I., Ureshino, K., and Khan, M. M. R. (2021). Pollen fertility differences in the progenies obtained from a cross between eggplant (*Solanum melongena* L.) as a seed parent and eggplant cytoplasmic substitution lines as pollen parents. *Austral. J. Crop Sci., 15,* 233–237. doi: 10.21475/ajcs.21.15.02.p2785

Jeong, H. -B., Jang, S. -J., Kang, M. -Y., Kim, S., Kwon, J. -K., and Kang, B. -C. (2020). Candidate gene analysis reveals that the fruit color locus C1 corresponds to *PRR2* in pepper (*Capsicum frutescens*). *Front. Plant Sci., 11*, 399. doi: 10.3389/fpls.2020.00399

Jiang, M., Ren, L., Lian, H., Liu, Y., and Chen, H. (2016). Novel insight into the mechanism underlying light-controlled anthocyanin accumulation in eggplant (*Solanum melongena* L.). *Plant Sci., 249,* 46–58. doi: 10.1016/j.plantsci.2016.04.001

Jiao, J., Liu, H., Liu, J., Cui, M., Xu, J., Meng, H., *et al.* (2017). Identification and functional characterization of *APRR2* controlling green immature fruit color in cucumber (*Cucumis sativus* l.). *Plant Growth Regul., 83*(2), 233–243. doi: 10.1007/s10725-017-0304-1

Kersey, P. J., Collemare, J., Cockel, C., Das, D., Dulloo, E. M., Kelly, L. J., *et al*. (2020). Selecting for useful properties of plants and fungi – Novel approaches, opportunities, and challenges. *Plants People Planet*, *2*(5), 409–420. doi: 10.1002/ppp3.10136

Kevei, Z., King, R. C., Mohareb, F., Sergeant, M. J., Awan, S. Z., and Thompson, A. J. (2015). Resequencing at ≥40-fold depth of the parental genomes of a *Solanum lycopersicum* × *S. pimpinellifolium* recombinant inbred line population and characterization of frame-shift indels that are highly likely to perturb protein function. *G3: Genes, Genomes, Genet.*, *5*(5), 971–981. doi: 10.1534/g3.114.016121

Khan, M. M. R., Arita, T., Iwayoshi, M., Ogura-Tsujita, Y., and Isshiki, S. (2020). Development of the functional male sterile line of eggplant utilizing the cytoplasm of *Solanum* kurzii by way of the amphidiploid. *Environ. Control Biol., 58,* 79–83. doi: 10.2525/ecb.58.79

Khan, M. M. R., Hasnunnahar, M., Iwayoshi, M., Ogura-Tsujita, Y., and Isshiki, S. (2015). Pollen degeneration in three functional male-sterile lines of eggplant with the wild *Solanum* cytoplasms. *Horticult. Environ. Biotechnol*., *56*, 350–357. doi: 10.1007/s13580-015-0015-3

Knapp, S., and Peralta, I. (2016). *The Tomato (Solanum lycopersicum L., Solanaceae) and Its Botanical Relatives,* Springer, 7–21.

Knapp, S. (2002). Tobacco to tomatoes: A phylogenetic perspective on fruit diversity in the Solanaceae. *J. Ex. Bot., 53*(377), 2001–2022. doi: 10.1093/jxb/erf068

Knapp, S., Vorontsova, M. S., and Prohens, J. (2013). Wild relatives of the eggplant (*Solanum melongena* L.: Solanaceae): new understanding of species names in a complex group. *PLoS One*, *8*(2). doi: 10.1371/journal.pone.0057039

Knapp, S., Aubriot, X., and Prohens, J. (2019). "The eggplant genome" in *Eggplant (Solanum melongena L.): taxonomy and relationships* (Switzerland: Springer Nature).

Kouassi, B., Prohens, J., Gramazio, P., Kouassi, A. B., Vilanova, S., Galán-Ávila, A., *et al*. (2016). Development of backcross generations and new inter-specific hybrid

combinations for introgression breeding in eggplant (*Solanum melongena*). *Sci. Horticulturae*, *213*, 199–207, doi: 10.1016/j.scienta.2016.10.039

Kumar, J., Gupta, D. Sen, Gupta, S., Dubey, S., Gupta, P., and Kumar, S. (2017). Quantitative trait loci from identification to exploitation for crop improvement. *Plant Cell Rep.*, *36*(8), 1187–1213. doi: 10.1007/s00299-017-2127-y

Kumar, N., Boatwright, J. L., Brenton, Z. W., Sapkota, S., Ballén-Taborda, C., Myers, M. T., *et al*. (2023). Development and characterization of a sorghum multi-parent advanced generation intercross (MAGIC) population for capturing diversity among seed parent gene pool. *G3: Genes, Genomes, Genet., 13*(4), jkad037. doi: 10.1093/g3journal/jkad037

Lebeau, A., Gouy, M., Daunay, M. C., Wicker, E., Chiroleu, F., Prior, P., *et al*. (2013). Genetic mapping of a major dominant gene for resistance to *Ralstonia solanacearum* in eggplant. *Theor. Appl. Genet.*, *126*(1), 143–158. doi: 10.1007/s00122-012-1969-5

Lester, R. N., and Hasan, S. M. Z. (1991). Origin and domestication of the brinjal egg-plant, *Solanum melogngena*, from *S. incanum*, in Africa and Asia (No. BOOK). *The Royal Botanic Gardens.*

Li, D., Qian, J., Li, W., Yu, N., Gan, G., Jiang, *et al*. (2021a). A high-quality genome assembly of the eggplant provides insights into the molecular basis of disease resistance and chlorogenic acid synthesis. *Mol. Ecol. Resour.*, *21*(4), 1274–1286. doi: 10.1111/1755-0998.13321

Li, J., He, Y. J., Zhou, L., Liu, Y., Jiang, M., Ren, L., *et al.* (2018). Transcriptome profiling of genes related to light-induced anthocyanin biosynthesis in eggplant (*Solanum melongena* L.) before purple color becomes evident. *BMC Genomics, 19*, 201. doi: 10.1186/s12864-018-4587-z

Li, L., He, Y., Ge, H., Liu, Y., and Chen, H. (2021b). Functional characterization of SmMYB86, a negative regulator of anthocyanin biosynthesis in eggplant (*Solanum melongena* L.). *Plant Sci., 302,* 110696. doi: 10.1016/j.plantsci.2020.110696

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*(11), 1220–1226. doi: 10.1038/ng.3117

Lippman, Z., Semel, Y., and Zamir, D. (2007). An integrated view of quantitative trait variation using tomato inter-specific introgression lines. *Curr. Opin. Genet. Dev., 17*, 545–552. doi: 10.1016/j.gde.2007.07.007

Liu, H., Jiao, J., Liang, X., Liu, J., Meng, H., Chen, S., *et al.* (2016). Map-based cloning, identification and characterization of the *w* gene controlling white immature fruit color in cucumber (*Cucumis sativus* L.). *Theor. Appl. Genet., 129*(7), 1247–1256. doi: 10.1007/s00122-016-2700-8

Luo, L., Molthoff, J., Li, Q., Liu, Y., Luo, S., Li, N., *et al*. (2023). Identification of candidate genes associated with less-photosensitive anthocyanin phenotype using an EMS mutant (*pind*) in eggplant (*Solanum melongena* L.). *Front. Plant Sci., 14.* doi: 10.3389/fpls.2023.1282661

Lv, Z., Jin, Q., Li, Z., Li, T., Wang, Y., You, Q., *et al*. (2023). Fine mapping and candidate gene analysis of the *Gv1* locus controlling green-peel color in eggplant (*Solanum melongena* L.). *Horticulturae, 9*, 888. doi: 10.3390/horticulturae9080888

Mennella, G., Rotino, G. L., Fibiani, M., D'Alessandro, A., Francese, G., Toppino, L., *et al.* (2010). Characterization of health-related compounds in eggplant (*Solanum melongena* L.) lines derived from introgression of allied species. *J. Agricult. Food Chem., 58*(13), 7597-7603. doi: 10.1021/jf101004z

Mackay, I., and Powell, W. (2007). Methods for linkage disequilibrium mapping in crops. *Trends Plant Sci., 12*, 57–63. doi: 10.1016/j.tplants.2006.12.001

Maioli, A., Gianoglio, S., Moglia, A., Acquadro, A., Valentino, D., Milani, A. M., *et al*. (2020). Simultaneous CRISPR/Cas9 editing of three PPO genes reduces fruit flesh browning in *Solanum melongena* L. *Front. Plant Sci., 11*, 607161. doi: 10.3389/fpls.2020.607161

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*, 847789. doi: 10.3389/fpls.2022.847789

Mangino, G., Plazas, M., Vilanova, S., Prohens, J., and 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*(4), 467. doi: 10.3390/agronomy10040467

Mangino, G., Vilanova, S., Plazas, M., Prohens, J., and Gramazio, P. (2021). Fruit shape morphometric analysis and QTL detection in a set of eggplant introgression lines. *Sci. Horticult., 282*, 110006. doi: 10.1016/j.scienta.2021.110006

Massaretto, I. L., Albaladejo, I., Purgatto, E., Flores, F. B., Plasencia, F., Egea-Fernández, J. M., *et al*. (2018). Recovering tomato landraces to simultaneously improve fruit yield and nutritional quality against salt stress. *Front. Plant Sci.*, *871*(November). doi: 10.3389/fpls.2018.01778

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*(3), 685–701. doi: 10.1016/j.ympev.2012.02.006

Meyer, R. S., and Purugganan, M. D. (2013). Evolution of crop species: Genetics of domestication and diversification. *Nat. Rev. Genet.*, *14*(12), 840–852. doi: 10.1038/nrg3605

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. doi: 10.1007/BF00226743

Mir, R., Calabuig-Serna, A., Seguí-Simarro, J. M. (2021). Doubled haploids in eggplant. *Biology, 10*. doi: 10.3390/biology10070685

Mishra, P., Tiwari, S. K., Kashyap, S. P., Tiwari, K. N., Singh, M., and 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.). *J. Plant Biochem. Biotechnol, 29*, 427–438. doi: 10.1007/s13562-020-00549-w

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. doi: 10.1104/pp.111.173997

Naeem, M., Muqarab, R., and Waseem, M. (2019). The *Solanum melongena* COP1 delays fruit ripening and influences ethylene signaling in tomato. *J. Plant Physiol., 240,* 152997. doi: 10.1016/j.jplph.2019.152997

Nguyen, C. V., Vrebalov, J. T., Gapper, N. E., Zheng, Y., Zhong, S., Fei, Z., *et al*. (2014). Tomato *GOLDEN2-LIKE* transcription factors reveal molecular gradients that function during fruit development and ripening. *Plant Cell*, *26*(2), 585–601. doi: 10.1105/tpc.113.118794

Ofner, I., Lashbrooke, J., Pleban, T., Aharoni, A., and Zamir, D. (2016). *Solanum pennellii* backcross inbred lines (BILs) link small genomic bins with tomato traits. *Plant J., 87*, 151–160. doi: 10.1111/tpj.13194

Oladosu, Y., Rafii, M. Y., Arolu, F., Chukwu, S. C., Salisu, M. A., Olaniyan, B. A., *et al*. (2021). Genetic diversity and utilization of cultivated eggplant germplasm in varietal improvement. *Plants*, *10*(8), 1714. doi: 10.3390/plants10081714

Oren, E., Tzuri, G., Vexler, L., Dafna, A., Meir, A., Faigenboim, A., *et al.* (2019). The multi-allelic *APRR2* gene is associated with fruit pigment accumulation in melon and watermelon. *J. Exp. Bot., 70*(15), 3781–3794. doi: 10.1093/jxb/erz182

Page, A. M.L., Daunay, M. C., Aubriot, X., and Chapman., M. A. (2019). "The eggplant genome" in *Domestication of eggplants: A phenotypic and genomic insight* (Switzerland: Springer Nature).

Page, A. M. L., and Chapman, M. A. (2021). Identifying genomic regions targeted during eggplant domestication using transcriptome data. *J. Heredity*, *112*(6), 519–525. doi: 10.1093/jhered/esab035

Pan, Y, Bradley, G., Pyke, K., Ball, G., Lu, C., Fray, R., *et al.* (2013). Network inference analysis identifies an *APRR2-like* gene linked to pigment accumulation in tomato and pepper fruits. *Plant Physiol., 161(*3), 1476–1485. doi: 10.1104/pp.112.212654

Paran, I., Goldman, I., Tanksley, S. D., and Zamir, D. (1995). Recombinant inbred lines for genetic mapping in tomato. *Theor. Appl. Genet.*, *90*(3–4), 542–548. doi: 10.1007/BF00222001

Pascual, L., Desplat, N., Huang, B. E., Desgroux, A., Bruguier, L., Bouchet, J. P., *et al*. (2015). Potential of a tomato MAGIC population to decipher the genetic control of quantitative traits and detect causal variants in the resequencing era. *Plant Biotechnol. J.*, *13*(4), 565–577. doi: 10.1111/pbi.12282

Peralta, I. E., and Spooner, D. M. (2001). Granule-bound starch synthase (GBSSI) gene phylogeny of wild tomatoes (*Solanum* L. section *Lycopersicon* [Mill.] Wettst. subsection *Lycopersicon*). *American J. Bot.*, *88*(10), 1888–1902. doi: 10.2307/3558365

Peralta, I. E., and Spooner, D. M. (2007). "Genetic improvement of solanaceous crops" in *History, origin and early cultivation of tomato (Solanaceae)*, *2*, 1–27.

Plazas, M., López-Gresa, M. P., Vilanova, S., Torres, C., Hurtado, M., Gramazio, P., *et al*. (2013). Diversity and relationships in key traits for functional and apparent quality in a collection of eggplant: Fruit phenolics content, antioxidant activity, polyphenol oxidase activity, and browning. *J. Agricult. Food Chem.*, *61*(37), 8871–8879. doi: 10.1021/jf402429k

Plazas, M., Vilanova, S., Gramazio, P., Rodríguez-Burruezo, A., Fita, A., Herraiz, F. J., *et al*. (2016). Inter-specific hybridization between eggplant and wild relatives from different genepools. *J. American Society Horticult. Sci.*, *141*(1), 34–44. doi: 10.21273/JASHS.141.1.34

Pook, T., Schlather, M., De Los Campos, G., Mayer, M., Schoen, C. C., and Simianer, H. (2019) Haploblocker: Creation of subgroup-specific haplotype blocks and libraries. *Genetics, 212,* 1045–61.

Powell, A. L., Nguyen, C. V., Hill, T., Cheng, K. L., Figueroa-Balderas, R., Aktas, H., *et al*. (2012). *Uniform ripening* encodes a *golden 2-like* transcription factor regulating tomato fruit chloroplast development. *Science, 336*, 1708–11. doi: 10.1126/science.1221863

Prohens, J., Gramazio, P., Plazas, M., Dempewolf, H., Francisco, F., Marı, B. K., *et al*. (2017). Introgressiomics: a new approach for using crop wild relatives in breeding for adaptation to climate change. *Euphytica*, *213*(7), 1–19. doi: 10.1007/s10681-017-1938-9

Qian, Z., Zhang, B., Chen, H., Lu, L., Duan, M., Zhou, J., *et al*. (2021). Identification of quantitative trait loci controlling the development of prickles in eggplant by genome re-sequencing analysis. *Fornt. Plant Sci.*, *12*(September), 1–13. doi: 10.3389/fpls.2021.731079

Raigón, M. D., Prohens, J., Muñoz-Falcón, J. E., and Nuez, F. (2008). Comparison of eggplant landraces and commercial varieties for fruit content of phenolics, minerals, dry matter and protein. *J. Food Composition Analysis*, *21*(5), 370–376. doi: 10.1016/j.jfca.2008.03.006

Ranjan, A., Ichihashi, Y., and Sinha, N. R. (2012). The tomato genome: implications for plant breeding, genomics and evolution. *Genome Biol., 13*(8), 1-8. doi: 10.1186/gb-2012-13-8-167

Ray, D. K., West, P. C., Clark, M., Gerber, J. S., Prishchepov, A. V., and Chatterjee, S. (2019). Climate change has likely already affected global food production. *PLoS One, 14*, 1–18. doi: 10.1371/journal.pone.0217148

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*, 1118–32. doi: 10.1093/molbev/msz297

Rockman, M. V., and Kruglyak, L. (2008). Breeding designs for recombinant inbred advanced intercross lines. *Genetics*, *179,* 1069–1078. doi: 10.1534/genetics.107.083873

Rosa-Martínez, E., Adalid-Martínez, A. M., García-Martínez, M. D., Mangino, G., Raigón, M. D., Plazas, *et al*. (2022). Fruit composition of eggplant lines with introgressions from the wild relative *S. incanum*: interest for breeding and safety for consumption. *Agronomy*, *12*(2), 266. doi: 10.3390/agronomy12020266

Rosa-Martínez, E., Villanueva, G., Şahin, A., Gramazio, P., García-Martínez, M. D., Raigón, M. D., *et al*. (2023). Characterization and QTL identification in eggplant introgression lines under two N fertilization levels. *Horticultural Plant Journal*. doi: doi.org/10.1016/j.hpj.2022.08.003

Saliba-Colombani, V., Causse, M., Langlois, D., Philouze, J., and Buret, M. (2001). Genetic analysis of organoleptic quality in fresh market tomato. 1. Mapping QTLs for physical and chemical traits. *Theor. Appl.* Genet, 102, 259-272. doi: 10.1007/s001220051643

Salinas, M., Capel, C., Alba, J. M., Mora, B., Cuartero, J., Fernández-Muñoz, R., *et al*. (2013). Genetic mapping of two QTL from the wild tomato *Solanum pimpinellifolium* L. controlling resistance against two-spotted spider mite (*Tetranychus urticae* Koch). *Theor. Appl. Genet., 126,* 83–92. doi: 10.1007/s00122-012-1961-0

Samineni, S., Sajja, S. B., Mondal, B., Chand, U., Thudi, M., Varshney, R. K., *et al*. (2021). MAGIC lines in chickpea: Development and exploitation of genetic diversity. *Euphytica, 217*(7), 137. doi: 10.1007/s10681-021-02874-0

Sato, S., Tabata, S., Hirakawa, H., Asamizu, E., Shirasawa, K., Isobe, S., *et al*. (2012). The tomato genome sequence provides insights into fleshy fruit evolution. *Nature*, *485*(7400), 635–641. doi: 10.1038/nature11119

Schauer, N., Zamir, D., and Fernie, A. R. (2005). Metabolic profiling of leaves and fruit of wild species tomato: a survey of the *Solanum lycopersicum* complex. *J. Ex. Bot.*, *56*(410), 297–307. doi: 10.1093/jxb/eri057

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*(2), 149–161. doi: 10.1111/pbi.12645

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. doi: 10.3389/fpls.2019.01606

Scott, M. F., Ladejobi, O., Amer, S., Bentley, A. R., Biernaskie, J., Boden, *et al*. (2020). Multi-parent populations in crops: a toolbox integrating genomics and genetic mapping with breeding. *Heredity*, 1–21. doi: 10.1038/s41437-020-0336-6

Semel, Y., Nissenbaum, J., Menda, N., and Zamir, D. (2006). Overdominant quantitative trait loci for yield and fitness in tomato. *Proc. Natl. Acad. Sci., 103*, 12981–12986. doi: 10.1073/pnas.0604635103

Shi, S., Liu, Y., He, Y., Li, L., Li, D., and Chen, H. (2021). R2R3-MYB transcription factor *SmMYB75* promotes anthocyanin biosynthesis in eggplant (*Solanum melongena* L.). *Sci. Hortic.*, *282*, 110020. doi: 10.1016/j.scienta.2021.110020

Sim, S. C., Durstewitz, G., Plieske, J., Wieseke, R., Ganal, M. W., Van Deynze, A., *et al*. (2012). Development of a large SNP genotyping array and generation of high-density genetic maps in tomato. *PloS one, 7*(7), e40563. doi: 10.1371/journal.pone.0040563

Smýkal, P., Nelson, M. N., Berger, J. D., and Von Wettberg, E. J. B. (2018). The impact of genetic changes during crop domestication. *Agronomy*, *8*(7), 1–22. doi: 10.3390/agronomy8070119

Sulli, M., Barchi, L., Toppino, L., Diretto, G., Sala, T., Lanteri, S., *et al*. (2021). An eggplant recombinant inbred population allows the discovery of metabolic QTLs controlling fruit nutritional quality. *Front. Plant Sci., 12*, 638195. doi: 10.3389/fpls.2021.638195

Suresh, B. V., Roy, R., Sahu, K., Misra, G., and Chattopadhyay, D. (2014). Tomato genomic resources database: an integrated repository of useful tomato genomic information for basic and applied research. *PloS one, 9*(1), e86387. doi: 10.1371/journal.pone.0086387

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. *American J. Bot.*, *103*(4), 635–651. doi: 10.3732/ajb.1500539

Tanksley, S. D., and Nelson, J. C. (1996). Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. *Theor. Appl. Genet.*, *92*, 191–203. doi: 10.1007/BF00223376

Tassone, M. R., Bagnaresi, P., Desiderio, F., Bassolino, L., Barchi, L., Florio, F. E., *et al*. (2022). A genomic BSAseq approach for the characterization of QTLs underlying resistance to *Fusarium oxysporum* in eggplant. *Cells, 11*(16), 2548. doi: 10.3390/cells11162548

The 100 Tomato Genome Sequencing Consortium. (2014). Exploring genetic variation in the tomato (*Solanum* section *Lycopersicon*) clade by whole-genome sequencing. *Plant J., 80*(1), 136–148. doi: 10.1111/tpj.12616

Thépot, S., Restoux, G., Goldringer, I., Hospital, F., Gouache, D., Mackay, I., *et al*. (2015). Eciently tracking selection in a multi-parental population: The case of earliness in wheat. *Genetics, 199*, 609–623. doi: 101534/genetics.114.169995

Tigchelaar, E.C., Janick, J., and Erickson, H.T. (1968). The genetics of anthocyanin coloration in eggplant (*Solanum melongena* L.). *Genetics, 60,* 475. doi: 10.1093/genetics/60.3.475

Tiwari, J. K., Singh, A. K., and Behera, T. K. (2023). CRISPR/Cas genome editing in tomato improvement: Advances and applications. *Front. Plant Sci.*, *14*, 1121209. doi: 10.3389/fpls.2023.1121209

Toppino, L., Barchi, L., Mercati, F., Acciarri, N., Perrone, D., Martina, M., *et al*. (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*(7), 745. doi: 10.3390/genes11070745

Toppino, L., Ribolzi, S., Shaaf, S., Bassolino, L., Carletti, G., Fadda, S., *et al*. (2018). Development of an introgression lines population and genetic mapping of novel traits linked to key breeding traits in eggplant. In *Proceedings of the 62th SIGA Congress* Verona, Italy, vol. 25, p. 28. ISBN 978-88-904570-8-1

Toppino, L., Valè, G., and Rotino, G. L. (2008). 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. *Mol. Breeding*, *22*, 237–250. doi: 10.1007/s11032-008-9170-x

van de Wouw, M., Kik, C., Van Hintum, T., Van Treuren, R., and Visser, B. (2010). Genetic erosion in crops: Concept, research results and challenges. *Plant Genet. Res.: Characterisation and Utilisation*, *8*(1), 1–15. doi: 10.1017/S1479262109990062

van Rengs, W. M. J., Schmidt, M. H. W., Effgen, S., Le, D. B., Wang, Y., Zaidan, M. W. A. M., *et al*. (2022). A chromosome scale tomato genome built from complementary PacBio and Nanopore sequences alone reveals extensive linkage drag during breeding. *Plant J.*, 572–588. doi: 10.1111/tpj.15690

Villanueva, G., Plazas, M., Gramazio, P., Moya, R. D., Prohens, J., and Vilanova, S. (2023). Evaluation of three sets of advanced backcrosses of eggplant with wild relatives from different genepools under low N fertilization conditions. *Horticulture research, 10,* uhad141. doi: 10.1093/hr/uhad141

Villanueva, G., Rosa-Martínez, E., Şahin, A., García-Fortea, E., Plazas, M., Prohens, J., *et al*. (2021). Evaluation of advanced backcrosses of eggplant with *Solanum elaeagnifolium* introgressions under low N conditions. *Agronomy, 11*, 1770. doi: 10.3390/agronomy11091770

Vorontsova, M. S., Stern, S., Bohs, L., and Knapp, S. (2013). African spiny *Solanum* (subgenus *leptostemonum*, *solanaceae*): A thorny phylogenetic tangle. *Bot. J. Linnean Society*, *173*(2), 176–193. doi: 10.1111/boj.12053

Voss-Fels, K., and Snowdon, R. J. (2016). Understanding and utilizing crop genome diversity via high-resolution genotyping. *Plant Biotechnol. J.*, *14*(4), 1086–1094. doi: 10.1111/pbi.12456

Wambugu, P. W., Ndjiondjop, M. N., and Henry, R. J. (2018). Role of genomics in promoting the utilization of plant genetic resources in genebanks. *Briefings Funct. Genomics*, *17*(3), 198–206. doi: 10.1093/bfgp/ely014

Wei, Q., Wang, J., Wang, W., Hu, T., Hu, H., and Bao, C. (2020b). A high-quality chromosome-level genome assembly reveals genetics for important traits in eggplant. *Horticult. Res.*, *7*(1). doi: 10.1038/s41438-020-00391-0

Wei, Q., Wang, W., Hu, T., Hu, H., Wang, J., and Bao, C. (2020a). Construction of a SNP-based genetic map using SLAF-Seq and QTL analysis of morphological traits in eggplant. *Front. Genet.*, *11*, 178. doi:10.3389/fgene.2020.00178

Yang, G., Li, L., Wei, M., Li, J., and Yang, F. (2022). *SmMYB113* is a key transcription factor responsible for compositional variation of anthocyanin and color diversity among eggplant peels. *Front. Plant Sci.*, *13*, 843996. doi: 10.3389/fpls.2022.843996

Yuan, G., Sun, K., Yu, W., Jiang, Z., Jiang, C., Liu, D., *et al*. (2023). Development of a MAGIC population and high-resolution quantitative trait mapping for nicotine content in tobacco. *Front. Plant Sci.*, *13,* 1086950. doi: 10.3389/fpls.2022.1086950

Zheng, C., Boer, M. P., and van Eeuwijk, F. A. (2014). A general modeling framework for genome ancestral origins in multi-parental populations. *Genetics, 198*, 87–101. doi: 10.1534/genetics.114.163006

Zhou, L., He, Y., Li, J., Liu, Y., and Chen, H. (2019). CBFs function in anthocyanin biosynthesis by interacting with MYB113 in eggplant (*Solanum melongena* L.). *Plant Cell Physiol. 61,* 416–426. doi: 10.1093/pcp/pcz209

Zuriaga, E., Blanca, J., and Nuez, F. (2009). Classification and phylogenetic relationships in *Solanum* section *Lycopersicon* based on AFLP and two nuclear gene sequences. *Genet. Res. Crop Evol.*, *56*(5), 663–678. doi: 10.1007/s10722-008-9392-0