UNIVERSITAT POLITÈCNICA DE VALÈNCIA

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Preservation and Improvement of Valencian Agro-diversity University Research Institute (COMAV)

Genome-wide association mapping of anthocyanin pigmentation traits in eggplant using a MAGIC population

Master's Thesis

Erasmus Mundus Master Programme in Plant Breeding - emPLANT +

AUTHOR: Diamante, Bryan Anthony

Tutor: Gramazio, Pietro

Cotutor: Plazas Ávila, María de la O

Experimental director: Luna Ramos, Jose

ACADEMIC YEAR: 2023/2024

GENOME-WIDE ASSOCIATION MAPPING OF ANTHOCYANIN PIGMENTATION TRAITS IN EGGPLANT USING A MAGIC POPULATION

BRYAN ANTHONY DIAMANTE
Author

PIETRO GRAMAZIO
Academic Tutor

MARIOLA PLAZAS ÁVILA
Academic Co-tutor

JOSE LUNA RAMOS
Experimental Director

ERASMUS MUNDUS MASTER IN PLANT BREEDING
Academic Course 2023-2024

Institute for the Conservation and Improvement of Valencian Agrodiversity
Universitat Politècnica de València
Valencia, Spain
July 2024

GENOME-WIDE ASSOCIATION MAPPING OF ANTHOCYANIN PIGMENTATION TRAITS IN EGGPLANT USING A MAGIC POPULATION

Bryan Anthony Diamante

Anthocyanins, the compounds responsible for the distinctive purple pigmentation in eggplant (*Solanum melongena*), are key targets in breeding due to their roles in plant defense, visual appeal, and potential health benefits. To uncover the genetic basis of anthocyanin pigmentation in eggplant, this study identified genomic regions and candidate genes involved in anthocyanin biosynthesis and regulation.

A genome-wide association study (GWAS) was conducted using an S5 multiparent advanced generation intercross (MAGIC) eggplant population, consisting of 308 lines derived from seven cultivated eggplants and a wild relative (*S. incanum*). The population was phenotyped for anthocyanin pigmentation intensity in the leaf vein (LVAntho), stem (SAntho), and fruit skin (FSAntho) in two different environments in Spain: Alcàsser, Valencia under field conditions, and Almería, Andalucia under plastic greenhouse conditions. Genotyping was performed using the skim whole-genome resequencing (SWGR) technique, resulting in 107,295 high-confidence genome-wide single nucleotide polymorphisms (SNPs).

The GWAS identified 10 lead SNPs within genomic regions significantly associated with anthocyanin pigmentation, distributed across chromosomes 1, 8, 9, and 10. Nearby these SNPs, three regulatory genes (*MYB113*, *bHLH*, *MYC1*) and a structural gene (*ANS*) putatively controlling anthocyanin pigmentation in eggplant were mapped. These candidate genes exhibited varying degrees ($0.05 \leq r^2 \leq 1.00$) of linkage disequilibrium (LD) with their associated SNPs.

This study enhances our understanding of the genetic basis of anthocyanin pigmentation in eggplant, and demonstrates the value of a diverse MAGIC population and dense SNP marker data in association mapping. The findings provide relevant insights for future eggplant breeding efforts aimed at improving anthocyanin content to meet market demands for purple-colored eggplants.

Keywords: Eggplant, anthocyanin, genome-wide association study (GWAS), multiparent advanced generation intercross (MAGIC) population, skim whole-genome resequencing (SWGR)
Academic Tutor: Pietro Gramazio
Academic Co-tutor: Mariola Plazas Ávila

MAPEO DE ASOCIACIÓN DEL GENOMA COMPLETO DE LOS CARACTERES DE PIGMENTACIÓN ANTOCIÁNICA EN BERENJENA UTILIZANDO UNA POBLACIÓN MAGIC

Bryan Anthony Diamante

Las antocianinas, los compuestos responsables de la pigmentación púrpura distintiva de la berenjena (*Solanum melongena*), son objetivos clave en la mejora debido a su papel en la defensa de la planta, el atractivo visual y los posibles beneficios para la salud. Para desentrañar la base genética de la pigmentación de antocianinas en la berenjena, en este estudio se identificaron regiones genómicas y genes candidatos involucrados en la biosíntesis y regulación de antocianinas.

Se realizó un estudio de asociación de todo el genoma (GWAS) utilizando una población de berenjenas intercruzadas de generación avanzada multiparental (MAGIC) S5, que consta de 308 líneas derivadas de siete berenjenas cultivadas y un pariente silvestre (*S. incanum*). La población fue fenotipada según la intensidad de la pigmentación antocianina en la vena de la hoja (LVAntho), el tallo (SAntho) y la piel del fruto (FSAntho) en dos ambientes diferentes en España: Alcàsser, Valencia en condiciones de campo, y Almería, Andalucía en condiciones de invernadero de plástico. El genotipado se realizó mediante la técnica de resecuenciación descremada del genoma completo (SWGR), lo que dio como resultado 107,295 polimorfismos de un solo nucleótido (SNP) de todo el genoma de alta confianza.

El GWAS identificó 10 SNPs principales dentro de regiones genómicas significativamente asociadas con la pigmentación antocianina, distribuidos en los cromosomas 1, 8, 9 y 10. Cerca de estos SNP, se mapearon tres genes reguladores (*MYB113*, *bHLH*, *MYC1*) y un gen structural (*ANS*) que supuestamente controla la pigmentación de antocianinas en la berenjena. Estos genes candidatos exhibieron diversos grados ($0.05 \leq r^2 \leq 1.00$) de desequilibrio de ligamiento (LD) con sus SNP asociados.

Este estudio mejora nuestra comprensión de la base genética de la pigmentación de antocianinas en la berenjena y demuestra el valor de una población MAGIC diversa y datos densos de marcadores SNP en el mapeo de asociaciones. Los hallazgos proporcionan información relevante para futuros esfuerzos en mejora de berenjena destinados a modificar el contenido de antocianinas para satisfacer las demandas del mercado de berenjenas de color púrpura.

Palabras clave: Berenjena, antocianina, estudio de asociación de genoma completo (GWAS), población intercruzada multiparental de generación avanzada (MAGIC), resecuenciación a baja cobertura

Tutor Académico: Pietro Gramazio
Cotutor Académico: Mariola Plazas Ávila

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (cont.)

ALIGNMENT WITH THE 2030 SUSTAINABLE DEVELOPMENT GOALS

Sustainable Development Goals (SDGs)	High	Medium	Low	Not Applicable
SDG 1: No Poverty			X	
SDG 2: Zero Hunger		X		
SDG 3: Good Health and Well-being	**X**			
SDG 4: Quality Education				X
SDG 5: Gender Equality				X
SDG 6: Clean Water and Sanitation			X	
SDG 7: Affordable and Clean Energy				X
SDG 8: Decent Work and Economic Growth		X		
SDG 9: Industry, Innovation, and Infrastructure		X		
SDG 10: Reduced Inequality				X
SDG 11: Sustainable Cities and Communities			X	
SDG 12: Responsible Consumption and Production	**X**			
SDG 13: Climate Action		X		
SDG 14: Life Below Water				X
SDG 15: Life on Land	**X**			
SDG 16: Peace, Justice, and Strong Institutions				X
SDG 17: Partnerships for the Goals			X	

SDG 3: Good Health and Well-being

This study contributes to understanding and enhancing the nutritional value of crops. Anthocyanins are antioxidants with potential health benefits, and breeding for higher content supports healthier diets and promotes well-being.

SDG 12: Responsible Consumption and Production

This study facilitates the development of eggplant varieties with anthocyanin-rich purple-colored fruits, promoting sustainable consumption and production. This supports efforts to ensure sustainable agricultural practices and reduce food waste.

SDG 15: Life on Land

This study equips breeding programs that enhance the genetic diversity and resilience of eggplant. By developing varieties with higher anthocyanin content, this study promotes sustainable agricultural practices, conserves genetic resources, and contributes to ecosystem health.

1. INTRODUCTION

1.1. Eggplant

1.1.1. Taxonomy

Eggplant (*Solanum melongena*, 2n = 2x = 24) is a warm-weather crop primarily grown in tropical and subtropical regions of the world (Taher et al., 2017). It is a berry-producing vegetable belonging to the extensive Solanaceae family, also known as the nightshade family (**Figure 1.1**; Taher et al., 2017). This family comprises approximately 3,000 species distributed among around 100 genera (Vorontsova and Knapp, 2012; Knapp et al., 2019). Of these genera, *Solanum* is the largest, encompassing roughly 1,500 species, including globally significant crops like potato and tomato (Frodin, 2004). The genus *Solanum* is exceptionally diverse and can be categorized into 13 clades, with eggplant belonging to the large and taxonomically challenging *Leptostemonum* clade (Knapp et al., 2019). This clade encompasses over 500 species found worldwide, many of which originated in the New World (Vorontsova and Knapp, 2012). The *Leptostemonum* clade, often referred to as the spiny solanums due to the presence of sharp epidermal prickles on stems and leaves, is recognized as the most diverse monophyletic group within the genus (Knapp et al., 2019).

1.1.2. Botany

Eggplant is woody and develops multiple branches in a roughly dichotomous branching pattern (Frary et al., 2007). The levels of anthocyanins, prickles, and hairiness on vegetative parts vary quantitatively (Frary et al., 2007). Inflorescences range from one to five andromonoecious cymes, although most modern cultivars display solitary hermaphrodite flowers (Frary et al., 2007). Flowers typically consist of five sepals, petals, and stamens; however, globose and round-fruited varieties often possess six, seven, or eight of these floral

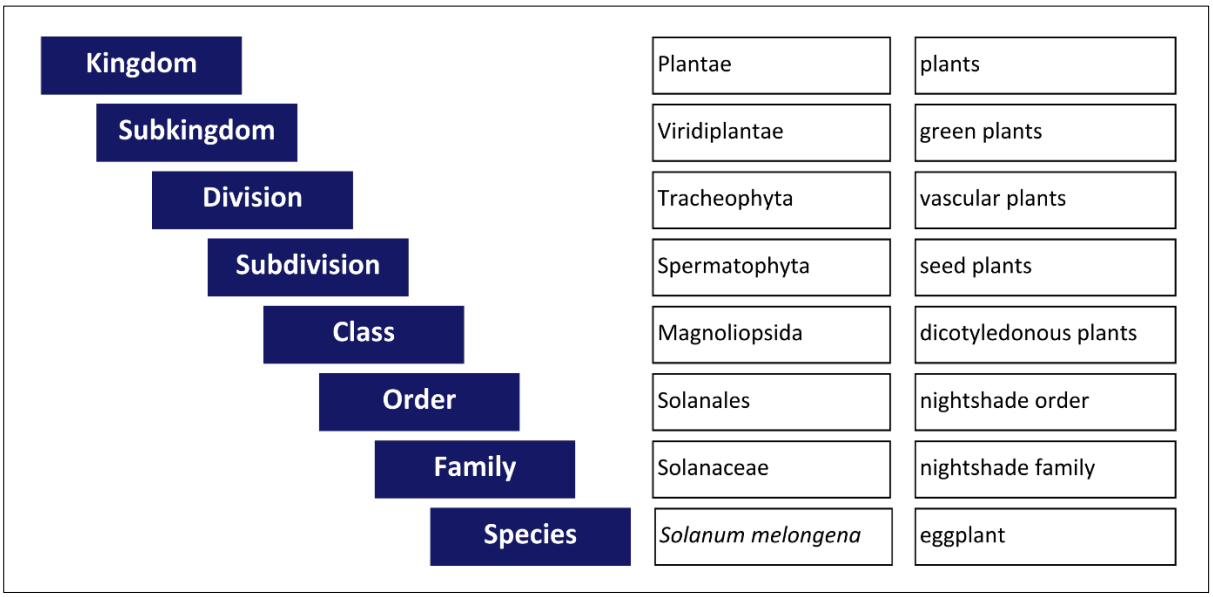

Figure 1.1. Taxonomic hierarchy of eggplant (*Solanum melongena*).

components (Frary et al., 2007). Eggplant is typically considered an autogamous species; however, in open fields and warm conditions, flowers attract insects, leading to an allogamy rate that can exceed 70% (Frary et al., 2007). The fruits are berries exhibiting a wide range of shapes (including round, intermediate, long, and snake-like) and sizes (from several grams to over a kilogram) (Frary et al., 2007). The presence or absence, as well as the distribution pattern, of two types of pigments, chlorophylls and anthocyanins, contribute to the wide diversity of fruit colors (Daunay et al., 2004).

1.1.3. Importance

Eggplants are widely produced and consumed around the world. The crop is cultivated on 1.89 million hectares globally, yielding a total production volume of 59.35 million tons (FAO, 2022). Its production is predominantly concentrated in Asia, which accounts for approximately 91% of the production area and 94% of the global production volume (FAO, 2022). The largest producers are China (64.57%), India (21.51%), Egypt (2.35%), Turkey (1.32%), and Indonesia (1.17%) (FAO, 2022). In terms of production area, the leading countries are China (43.19%), India (35.64%), Bangladesh (2.86%), Indonesia (2.66%), and Egypt (2.21%) (FAO, 2022). Apart from Egypt, significant eggplant-producing nations in Africa include Algeria, Ivory Coast, and Sudan (Solberg et al., 2022). In Europe, Italy and Spain are the primary eggplant producers, followed by Romania, Ukraine, Greece, France, and the Netherlands (Solberg et al., 2022). In the Americas, Mexico leads in eggplant production, with the USA as the next major producer (Solberg et al., 2022).

The crop is widely enjoyed across various cuisines and holds a significant place in the diets of many countries, notably India, Bangladesh, Southeast Asia, and the Middle East (Daunay and Lester, 1988). While eggplant is not known for being high in most health-related micronutrients, its exceptionally low-calorie content and rich array of vitamins, minerals, and bioactive compounds make it a healthy vegetable to consume (Raigón et al., 2008; Plazas et al., 2014; Docimo et al., 2016; Chapman, 2019). Eggplants exhibit a high oxygen radical absorbance capacity due to their abundance in phenolic compounds, particularly chlorogenic acid found in the fruit pulp and anthocyanins in the fruit peel (Cao et al., 1996; Mennella et al., 2012; Stommel et al., 2015). Both phenolic acids and anthocyanins possess numerous health-promoting properties (Plazas et al., 2013; Braga et al., 2016).

In addition to its economic and nutritional importance, eggplant provides a valuable system for exploring parallel and convergent evolution, alongside other solanaceous crops such as tomato and pepper (Chapman, 2019). The domestication of these solanaceous plants has resulted in enlarged fruit sizes and alterations in fruit color and shape (Chapman, 2019). Eggplant is also a model for understanding anthocyanin biosynthesis, as it exhibits significant variation in fruit colors and contributes to plant defense mechanisms (Daunay and Hazra, 2012; D'Amelia et al., 2018; Lv et al., 2019; Zhou et al., 2019).

1.1.4. Origin, domestication, and dissemination

Eggplant is considered native to the Indo-Chinese center of origin according to Vavilov (1951). However, the domestication history of eggplant has been the subject of prolonged debate, with studies struggling to pinpoint the precise location within the broad region of tropical Asia where domestication occurred (Meyer et al., 2012). While many studies propose India as the center of eggplant domestication, China also has an ancient written record of eggplant usage dating back approximately 2,000 years before the present (Bhaduri, 1951; Zeven and Zhukovsky, 1975; Martin and Rhodes, 1979; Lester and Hasan, 1991; Lester, 1998; Mace et al., 1999; Doganlar et al., 2002; Meyer et al., 2012). Conversely, some studies suggest that multiple independent domestications of eggplant from *S. insanum*, the confirmed eggplant wild progenitor naturally distributed in tropical Asia from Madagascar to the Philippines, occurred in multiple centers (Meyer et al., 2012; Knapp et al., 2013). However, the most recent study found no evidence for multiple domestications of eggplant (Page et al., 2019). Eggplant spread eastward to Japan in the 8th century (Taher et al., 2017) and later disseminated westward along the Silk Road into Western Asia, Europe, and Africa by Arab traders during the 14th century (Taher et al., 2017). Subsequently, it was introduced to the Americas soon after Europeans arrived there and later expanded into other parts of the world (Prohens et al., 2005).

1.1.5. Wild relatives

Eggplant is associated with numerous wild species that could serve as sources of variation for breeding programs, particularly traits relevant to climate change adaptation, and resistance to pests and diseases (Vorontsova et al., 2013; Rotino et al., 2014; Syfert et al., 2016). Although eggplant is believed to have originated in Asia, most of its wild relatives are from Africa (Weese and Bohs, 2010). The wild relatives of eggplant constitute one of the most diverse and complex groups in terms of their taxonomic and phylogenetic relationships (Vorontsova et al., 2013). Wild eggplants typically bear spiny, small, bitter, multi-seeded, and inedible fruits, while some varieties exhibit high levels of chlorogenic acid and other bioactive compounds, which may be of interest for human health (Meyer et al., 2015; Taher et al., 2017).

The wild relatives of eggplant represent one of the most diverse and complex groups in terms of their taxonomic and phylogenetic relationships (Vorontsova et al., 2013). Based on crossing and biosystematics data, nine wild species, in addition to *S. melongena*, comprise the eggplant complex, which encompasses the cultivated eggplant and its closest wild relatives (Knapp et al., 2013). Wild relatives can be classified based on their crossability with cultivated species into primary, secondary, and tertiary genepools (Harlan and de Wet, 1971).

The primary genepool of eggplant consists of cultivated eggplants and its wild ancestor *S. insanum*, which can be easily crossed to produce normal fertile hybrids (Plazas et al., 2016;

Ranil et al., 2017). The secondary genepool includes more than 40 wild relatives that can be crossed or are phylogenetically close to eggplant, but the success of the crosses and the viability or fertility of the resulting hybrids with eggplant may be reduced (Taher et al., 2017). For instance, some interspecific hybrids derived from secondary genepool are partly sterile or weak due to reproductive barriers such as *S. dasyphyllum*, *S. linnaeanum*, or *S. tomentosum* (Rotino et al., 2014; Kouassi et al., 2016). Additionally, secondary genepool includes *S. incanum*, a potent source of phenolic compounds and tolerant to some biotic and abiotic stresses, particularly drought (Knapp et al., 2013; Prohens et al., 2013; Syfert et al., 2016; Flores-Saavedra et al., 2024). The tertiary genepool includes more distantly related species, such as *S. torvum*, *S. elaeagnifolium*, and *S. sisymbriifolium*, which are utilized in breeding programs for their resistance features, although successful crossing requires specific breeding techniques (Kouassi et al., 2016; Plazas et al., 2016; Syfert et al., 2016; García-Fortea et al., 2019; Villanueva et al., 2021).

1.2. Plant anthocyanin

1.2.1. Description

Anthocyanins, a significant class of flavonoids, constitute a diverse group of plant secondary metabolites (Liu et al., 2018). These glycosylated polyphenolic compounds impart a range of colors such as orange, red, purple, and blue, distributed across various plant parts (Tanaka and Ohmiya, 2008). Their color variation, predominantly found in cell vacuoles, is influenced by the intravacuolar environment (Liu et al., 2018). Across plant species, anthocyanins typically derive from six widely distributed anthocyanidins: pelargonidin, cyanidin, delphinidin, peonidin, petunidin, and malvidin (Kong et al., 2003). In purple solanaceous tissues, delphinidin-based anthocyanins are particularly prevalent (Ichiyanagi et al., 2005).

In eggplant, anthocyanins accumulate mainly in flowers, leaves, stems, and fruit peels (Matsubara et al., 2005). Nasunin, a delphinidin-based anthocyanidin, is the primary component of the anthocyanin pigment found in eggplant peels (Kuroda and Wada, 1933; Sakamura et al., 1963). The level of anthocyanin in eggplant fruits peaks at the unripe stage and decreases as the fruits ripen, often disappearing entirely (Liu et al., 2018). This phenomenon is notable as eggplants commercially mature long before they reach physiological ripeness (Mennella et al., 2012).

1.2.2. Biosynthesis

The anthocyanin biosynthetic pathway is a well-characterized and highly conserved network across many plant species (**Figure 1.2**; Holton and Cornish, 1995; Tanaka and Ohmiya, 2008). This pathway begins with phenylalanine, which serves as the precursor for anthocyanin

biosynthesis (Chaves-Silva et al., 2019). Phenylalanine undergoes deamination by phenylalanine ammonia lyase (PAL) to produce trans-cinnamic acid, a substrate for cinnamic acid 4-hydroxylase (C4H) (Chaves-Silva et al., 2019). This leads to the formation of coumaric acid, which then enters the lignin branch. Coumaric acid is converted into coumaric acid coenzyme A by 4-coumaroyl CoA ligase (4CL) (Chaves-Silva et al., 2019). In the next step, chalcone synthase (CHS) catalyzes the combination of one molecule of coumaric acid and three molecules of malonyl-CoA to form tetrahydroxychalcone (Liu et al., 2018). This compound is subsequently transformed into the flavanone naringenin by chalcone isomerase (CHI) (Liu et al., 2018). Naringenin is hydroxylated by flavanone 3-hydroxylase (F3H) to produce dihydrokaempferol, a dihydroflavonol (Liu et al., 2018). Alternatively, dihydrokaempferol is hydroxylated by flavonoid 3'-hydroxylase (F3'H) or flavonoid 3',5'-hydroxylase (F3'5'H) to form two other dihydroflavonols: dihydroquercetin or dihydromyricetin, respectively (Liu et al., 2018). Dihydroflavonol 4-reductase (DFR) subsequently catalyzes the transformation of these dihydroflavonols into colorless anthocyanidins, which are then converted into colored anthocyanidins by anthocyanin synthase (ANS) (Chaves-Silva et al., 2019). Finally, these anthocyanidins can be further modified by uridine diphosphate (UDP)-glucose flavonoid-O-glycosyltransferase (UFGT) (UFGT), acetylase, and O-methyltransferase (OMT) to create pigments that are stable and water-soluble (Winkel-Shirley, 2001). The stabilized anthocyanins are transported to the vacuoles by glutathione transferase (GST) (Chaves-Silva et al., 2019).

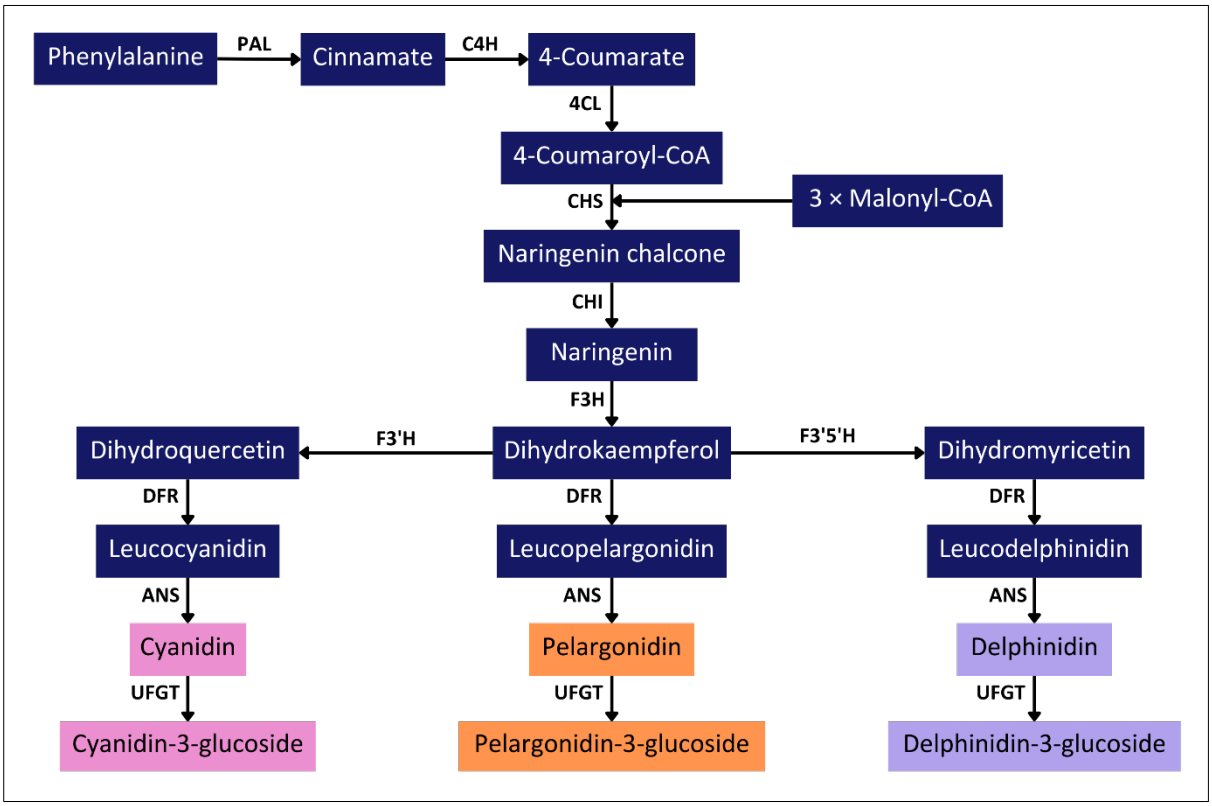

Figure 1.2. Anthocyanin biosynthetic pathway.

1.2.3. Regulation

The regulatory mechanism governing the initiation or cessation of anthocyanin biosynthesis in plants operates at the level of gene expression for both structural and regulatory genes involved in the biosynthetic pathway (Liu et al., 2018). Structural genes encode the enzymes responsible for each step of the biosynthetic pathway, whereas regulatory genes encode transcription factors that regulate the expression of these structural genes (Gonzali et al., 2009; Dubos et al., 2010).

Structural genes are classified into early biosynthetic genes (EBGs) and late biosynthetic genes (LBGs) (Dubos et al., 2010). EBGs – *CHS*, *CHI*, and *F3H* – are essential in the flavonoid pathway, playing a crucial role in synthesizing all downstream flavonoids (Liu et al., 2018). Generally, a consistent correlation between EBG expression levels and anthocyanin content is lacking, likely because EBG expression is necessary not only for anthocyanin production but also for synthesizing other flavonoids, such as flavonols and flavanones (Liu et al., 2018). In contrast, LBGs – *F3'H*, *F3'5'H*, *DFR*, *ANS*, and *UFGT* – are required for the biosynthesis of specific classes of flavonoids, including anthocyanins (Liu et al., 2018). The transcript levels of LBGs closely correlate with anthocyanin content and are markedly higher in pigmented tissues compared to non-pigmented ones, indicating that variations in LBG expression are a key factor influencing the quantitative differences in anthocyanin levels in solanaceous vegetables (Liu et al., 2018).

The regulatory complex governing the activity of structural genes within the anthocyanin biosynthetic pathway is known as the MYB-bHLH-WD40 (MBW) complex (Liu et al., 2018). The activation or repression function of the MBW complex is chiefly determined by MYB (Myeloblastosis) transcription factors, which bind to the promoters of structural genes along with common bHLH (basic Helix-Loop-Helix) and WD40 (40 amino acids terminating in tryptophan [W] and aspartic acid [D]) factors (Liu et al., 2018). The bHLH transcription factors play a crucial role in conferring specificity in identifying transcription factor binding sites within the promoters of target genes and subsequently initiating transcription (Montefiori et al., 2015). On the other hand, WD40 proteins serve as stable scaffolds, facilitating the assembly of the MBW complex by enabling the interaction between MYB and bHLH proteins (Liu et al., 2018). Within the MYB complex, the R2R3-MYB activator enhances the biosynthesis of anthocyanin, whereas R2R3-MYB and R3-MYB repressors negatively regulate the pathway (Albert et al., 2014; Liu et al., 2018). In addition to MYB repressors, microRNAs have also been discovered to downregulate anthocyanin biosynthesis at the post-transcriptional level (Liu et al., 2018).

1.2.4. Importance

Anthocyanins are responsible for various important traits in plants. Since they are present in the epidermis of flowers and fruits, anthocyanins provide visual cues to attract pollinators and seed dispersers (Tanaka et al., 2008; Rieseberg and Blackman, 2010). In addition to their colorful characteristics, they play a crucial role in plant defense mechanisms against various biotic and abiotic stresses (D'Amelia et al., 2018; Lv et al., 2019; Zhou et al., 2019). Functioning as photoprotective agents, anthocyanins protect the photosynthetic system of plants by absorbing excess light and scavenging free radicals (Guo et al., 2008). Additionally, they often accumulate in juvenile vegetative tissues and on the sun-exposed surfaces of fruits, providing protection against photoinhibition and photobleaching induced by light stress while maintaining photosynthetic efficiency at a substantial level (Steyn et al., 2002; Gould, 2003). The presence of pigmented anthocyanins can also decrease the incidence of insect infestation and pathogen attacks (Malone et al., 2009; Zhang et al., 2013). Furthermore, acting as antioxidants, anthocyanins prevent lipid peroxidation and maintain membrane integrity, thereby slowing cellular senescence and enhancing the postharvest quality of crops (Jiao et al., 2012). In eggplant, anthocyanin pigmentation in fruit peels is a key trait in breeding (Daunay and Hazra, 2012). Purple-colored eggplant fruits are highly demanded in many markets and developing dark purple eggplants, which result from the combination of anthocyanins and chlorophylls, is a primary goal in eggplant breeding programs (Li et al., 2018; Mangino et al., 2022).

1.3. Association mapping

1.3.1. Description

Association mapping is an approach that identifies associations between markers and traits through the study of natural or designed populations exhibiting a wide range of trait variations (Bergelson and Roux, 2010; Pasam and Sharma, 2014). Based on historic recombination, this method aims to pinpoint genomic regions and candidate genes linked to specific traits (Kulwal and Singh, 2021). The fundamental principle of association mapping relies on the co-inheritance of marker alleles in adjacent genomic regions within a population, driven by linkage disequilibrium (LD) and a strong correlation observed between allele variants and traits (Bush and Moore, 2012; Kulwal and Singh, 2021).

Association mapping is increasingly preferred over biparental linkage mapping due to its ability to harness natural diversity and explore functional variants across a broader range of germplasm, offering advantages such as enhanced mapping resolution (Gupta et al., 2005; Zhu et al., 2008; Gupta et al., 2013; Gupta et al., 2014; Gupta et al., 2019). The resolution achieved by association mapping in gene identification is notably higher because it can

densely genotype nucleotide-level variations and apply low LD-based association mapping in natural populations (Kulwal and Singh, 2021). Moreover, association mapping facilitates quicker identification of associations compared to biparental linkage mapping (Kulwal and Singh, 2021). Therefore, association mapping provides an alternative approach to identify marker-trait associations, utilizing genotypes with known or unknown ancestry that exhibit maximum genetic variability for the trait of interest (Gupta et al., 2014).

1.3.2. Genome-wide association study

Genome-wide association study (GWAS) is an association mapping approach aimed at identifying all genomic variants involved in controlling trait variation (**Figure 1.3**; Kulwal and Singh, 2021; Jain et al., 2024). Using appropriate statistical methods, this technique involves examining the genomes of numerous individuals to identify variants that occur more frequently in those with a specific phenotype compared to those with different phenotypes (Jain et al., 2024). Once these associated genomic variants are identified, they are typically used to locate nearby variants directly influencing the expression of the trait (Jain et al., 2024).

GWAS results are commonly visualized using a Manhattan plot, where single nucleotide polymorphisms (SNPs) with significant associations appear as prominent peaks (Pasam and Sharma, 2014). A Manhattan plot is a scatter plot showing negative logarithm p-values for SNP associations plotted against SNP positions (Pasam and Sharma, 2014). Longer peaks in the Manhattan plot indicate stronger associations of the surrounding genomic region with the trait (Jain et al., 2024). However, since GWAS analyzes many markers, the risk of detecting

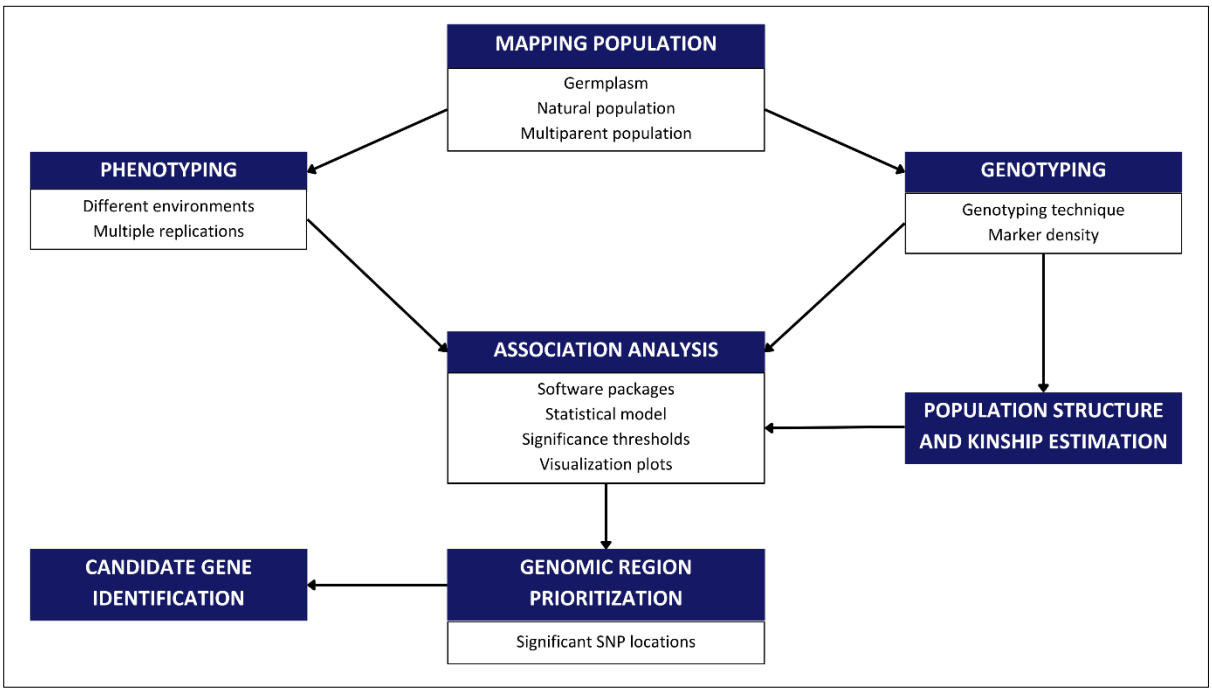

Figure 1.3. General pipeline for conducting genome-wide association study.

false positive associations increases (Bush and Moore, 2012; Brzyski et al., 2017). Typical approaches for correcting multiple testing include controlling the false discovery rate (FDR), which represents the expected proportion of false positives among all positive findings or employing the Bonferroni correction (Benjamini and Hochberg, 1995; Storey and Tibshirani, 2003). The Bonferroni correction involves dividing the chosen significance threshold by the total number of tests performed to establish the adjusted significance threshold (Tibbs-Cortes et al., 2021).

Furthermore, the quantile-quantile (QQ) plot serves as a valuable tool for assessing the reliability of GWAS results regarding marker-trait associations (Jain et al., 2024). This plot compares observed p-value distributions against expected distributions under the null hypothesis, where no association exists with the trait (Jain et al., 2024). In an ideal scenario, most points on the QQ plot should align along the diagonal line, indicating that tested SNPs are mostly unrelated to the trait (Jain et al., 2024). Deviations from this line, particularly points above the diagonal, suggest potential false positives due to factors such as population structure or familial relatedness (Jain et al., 2024). Such deviations suggest that the GWAS model inadequately corrects for these spurious associations, but they can be resolved by taking the population structure into account (Jain et al., 2024).

To conduct GWAS in plants, three key elements are necessary: (1) a mapping population phenotyped for the trait under study, (2) a genetic marker system for this population, and (3) analysis software capable of performing statistical analyses, including imputation of missing marker data, assessment of population structure, and marker-trait association analysis (Kulwal and Singh, 2021). Each of these components is discussed in the subsequent sections.

1.3.3. Mapping population

The success of association mapping studies depends significantly on the choice of population utilized (Gupta et al., 2014). Ideally, populations with diverse genotypes that encompass trait diversity are best suited for such studies (Kulwal and Singh, 2021). Depending on the population composition, association mapping studies are classified as broad-based or narrow-based (Islam et al., 2016). Broad-based studies encompass germplasm, landraces, cultivars, and natural populations, whereas narrow-based studies involve specially developed multiparent populations, such as the multiparent advanced generation intercross (MAGIC) population used in this study.

The MAGIC population consists of recombinant inbred lines (RILs) derived from a complex cross involving multiple parents (Singh and Singh, 2015). Its development includes multiple rounds of recombination, which enhances the precision and resolution of association mapping (Singh and Singh, 2015). Parental lines may be inbred lines, clones, or individuals selected based on their origin or intended use (Singh and Singh, 2015). One approach to creating a

MAGIC population involves initiating a complex cross with eight parental lines, forming four single crosses initially, followed by pairing these to generate two double crosses (Singh and Singh, 2015). The resulting double crosses are then crossed to produce the eight-parent complex cross (Singh and Singh, 2015). Using the single seed descent method manages this cross to produce the desired number of RILs, constituting the MAGIC population (Singh and Singh, 2015).

MAGIC populations can be utilized for association mapping studies. They can be developed at a suitable stage during the intermating process to achieve the desired mapping resolution (Singh and Singh, 2015). Created from multiple parents, these populations tend to segregate for numerous traits, contain multiple quantitative trait loci (QTLs) for each trait, and have more than two alleles for individual QTLs (Singh and Singh, 2015). They are ideal for constructing high-density maps and allow for the modeling of cytoplasmic effects (Singh and Singh, 2015). Furthermore, the parents of a MAGIC population can be selected to represent a significant portion of the variation found in the elite germplasm of a crop species (Singh and Singh, 2015). When based on elite parental lines with a combination of useful traits, these populations can be used directly or indirectly for variety development (Singh and Singh, 2015). Additionally, they can serve as training populations for genomic selection since, as compared to collections of breeding lines and cultivars, they lack the population structure (Singh and Singh, 2015). MAGIC populations also provide opportunities to study the interactions of genome segment introgressions and chromosomal recombination (Singh and Singh, 2015).

Accurate phenotyping of the mapping population is crucial for reliably identifying genomic regions controlling traits. Phenotyping methods should minimize errors in data generation and acquisition (Sahu et al., 2023). To enhance phenotyping accuracy, replicated tests across multiple locations and years are conducted, particularly when using inbred lines or clones for association mapping (Singh and Singh, 2015). Averaging data from these replicates reduces environmental variability and measurement errors, with the resulting average values used in association analyses (Singh and Singh, 2015). Specialized phenomics facilities worldwide provide substantial advantages for studying complex plant traits compared to traditional field phenotyping (Sahu et al., 2023).

1.3.4. Genetic marker system

Genotyping plays a crucial role in quantitative and population genetic studies, as well as in genomics-assisted breeding in crops (Adhikari et al., 2022). Over recent decades, advances in DNA sequencing technology have transformed these disciplines from being constrained by limited information to being abundant in data (Adhikari et al., 2022). As sequencing costs decrease and adoption increases, there is a growing emphasis on maximizing the utilization of these methods and technologies in breeding pipelines and genetic studies (Rasheed et al., 2017).

SNPs, the most prevalent type of sequence variation in eukaryotic genomes, are extensively used in GWAS due to their ability to be genotyped simultaneously in thousands of individuals, enabling efficient parallel analysis (Mammadov et al., 2012). SNPs represent genetic variations among individuals within a species at specific positions in their genomes (Singh and Singh, 2015). They are preferred in molecular genetics because they are compatible with high-throughput automated genotyping platforms and exhibit declining costs per data point (Pootakham, 2023). Rapid advancements in sequencing technologies and reduced sequencing costs have facilitated SNP discovery in various plant species, including non-model organisms with limited or no genomic resources (Pootakham, 2023).

In this study, skim whole-genome resequencing (SWGR) was employed to efficiently generate dense SNP markers in a cost-effective manner. SWGR provides a versatile, efficient, and dense marker system with unbiased coverage of the entire genome, along with more cost-effective sample preparation compared to other widely used methods for reduced genome representation genotyping (Huang et al., 2009; Rowan et al., 2015). While there is no established standard, SWGR is a sequencing approach that achieves less than 10X coverage of the genome (Malmberg et al., 2018). A notable advantage of SWGR is its ability to adjust marker density by varying sequencing depth, thereby allowing for proportional changes in sequencing costs (Malmberg et al., 2018). Consequently, selecting an optimal sequencing depth requires careful consideration of the balance between coverage and the number of samples (Malmberg et al., 2018).

1.3.5. Analysis software

Association analysis commences once both trait data and marker genotypic data for the population are available. These datasets undergo extensive preprocessing, including marker imputation and filtration, population structure assessment, and kinship estimation, before they are utilized for association mapping (Kulwal and Singh, 2021). To facilitate efficient association analysis, various software packages have been developed. The choice of software depends on the data volume and user expertise, with a preference for user-friendly tools capable of performing multiple tasks (Kulwal and Singh, 2021). Among the numerous software packages designed for association mapping, the Trait Analysis by Association, Evolution, and Linkage (TASSEL) software is widely utilized in plant systems by the plant breeding community (Bradbury et al., 2007).

TASSEL is a comprehensive software package used for evaluating trait associations, evolutionary patterns, and LD (Bradbury et al., 2007). It incorporates advanced statistical approaches for association mapping, including the General Linear Model (GLM) and the Mixed Linear Model (MLM) (Bradbury et al., 2007). The GLM method employs structured association analysis with a population structure (Q) matrix to minimize false associations (Singh and Singh, 2015). The Q matrix is derived using principal components analysis or multidimensional

scaling (MDS) (Singh and Singh, 2015). In contrast, the MLM method integrates both the kinship (K) matrix and the Q matrix in its model to further mitigate the risk of false-positive associations (Singh and Singh, 2015). The K matrix estimates the average relatedness between pairs of individuals or lines, which can be derived from pedigree information or genotype data from numerous unlinked markers across the entire genome of the organism (Singh and Singh, 2015).

1.3.6. Importance

GWAS has emerged to be a powerful method for genetic mapping in plants, with significant potential to expedite the exploitation of genetic diversity in improving crops (Tibbs-Cortes et al., 2021; Sahu et al., 2023). It has been conducted across various crops, resulting in the discovery of numerous genomic regions associated with diverse traits including yield characteristics (Li et al., 2019; Wang et al., 2019; Tsai et al., 2020; Ravelombola et al., 2021), morphological features (Huang et al., 2010; Bai et al., 2016; Yano et al., 2016; Contreras-Soto et al., 2017; Kim et al., 2021; Ye et al., 2021), and biotic and abiotic stress resistances (Gyawali et al., 2018; Ertiro et al., 2020; Joukhadar et al., 2020; Wu et al., 2020; Pang et al., 2021). Additionally, GWAS has been utilized to explore diverse phenotypes, such as the identification of genes associated with geographical divergence and adaptation during domestication (Chen et al., 2019), as well as the study of biochemical and molecular phenotypes encompassing amino acids, fatty acids, flavonoids, and nucleic acid metabolites (Chen et al., 2016; Arrones et al., 2022; Mangino et al., 2022). Moreover, GWAS plays a crucial role in uncovering novel associations with important traits and in identifying candidate genes for genetic engineering and genome editing aimed at crop improvement (Owens et al., 2014; Zhang et al., 2018).

2. OBJECTIVES

This study aims to perform GWAS on anthocyanin pigmentation traits in eggplant using a MAGIC population. Specifically, the objectives are to:

1. Identify SNPs associated with anthocyanin pigmentation in the leaf vein, stem, and fruit skin;
2. Locate candidate genes within genomic regions showing significant marker-trait associations; and
3. Evaluate LD patterns between significant SNPs and candidate genes.

3. MATERIALS AND METHODS

3.1. Plant materials

The S5 MAGIC eggplant population utilized in this study was created by crossing seven cultivated eggplants (*S. melongena*) named MM1597 (Parent A), DH ECAVI (Parent B), AN-S-26 (Parent D), H15 (Parent E), A0416 (Parent F), IVIA371 (Parent G), and ASI-S-1 (Parent H), along with a wild relative (*S. incanum*) named MM577 (Parent C) (**Figure 3.1**). The wild relative was selected for its drought resistance and high phenolic content (Knapp et al., 2013; Prohens et al., 2013). The performance of these founder parents has been previously assessed in morpho-agronomic and genetic diversity studies (Hurtado et al., 2014; Gramazio et al., 2017; Kaushik et al., 2018). The residual heterozygosity of the founder parents was determined to be 0.06% based on a previous genome resequencing study (Gramazio et al., 2019).

3.2. MAGIC population development

The S5 MAGIC eggplant population was developed by crossing the founder parents using a simple funnel method (**Figure 3.2**; Wang et al., 2017; Arrones et al., 2020). Initially, the eight founder parents (A to H) were crossed to produce simple hybrids (AB, CD, EF, and GH). These hybrids were then crossed in pairs (AB × CD and EF × GH) to generate double hybrids (ABCD and EFGH). To ensure complete genomic mixing and prevent selective mating, the double hybrids underwent chain pollination, where each individual served both as a female and male parent (Díez et al., 2002). The resulting quadruple hybrids (S0 recombinant lines) were self-fertilized through single seed descent to yield the S5 recombinant lines (or S5 MAGIC lines), which were subsequently genotyped and phenotyped in this study. In total, 308 S5 MAGIC eggplant lines were produced.

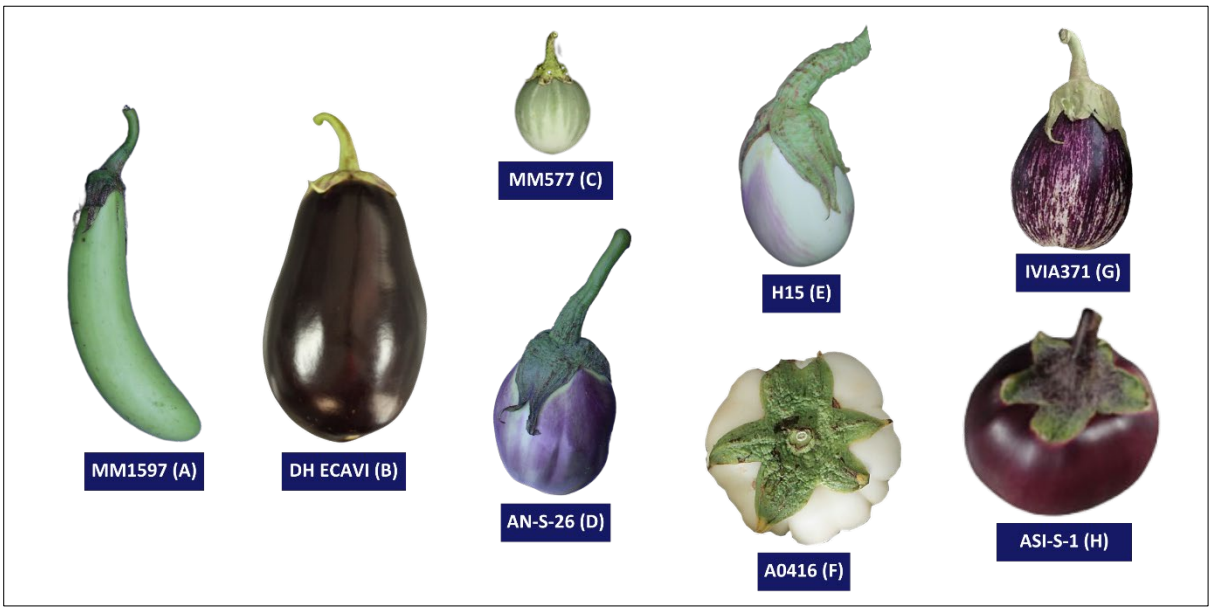

Figure 3.1. Founder parents of the S5 MAGIC eggplant population.

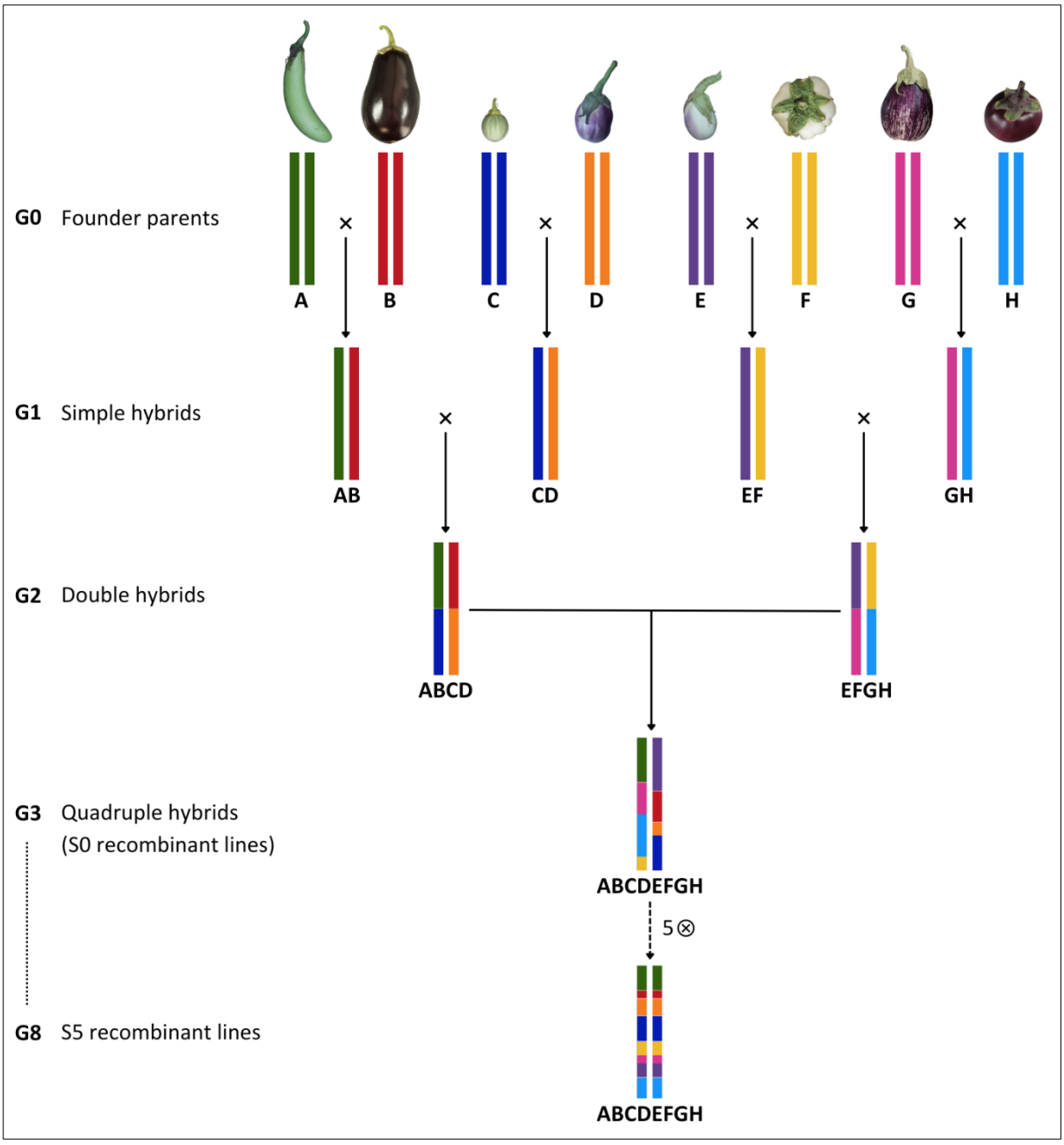

Figure 3.2. Simple funnel breeding design used to develop the S5 MAGIC eggplant population. The founder parents are MM1597 (A), DH ECAVI (B), MM577 (C), AN-S-26 (D), H15 (E), A0416 (F), IVIA371 (G), and ASI-S-1 (H).

3.3. Cultivation conditions and experimental design

Seeds of the 308 S5 MAGIC eggplant lines were germinated in petri dishes and subsequently transferred to seedling trays positioned in a climate-controlled chamber (Ranil et al., 2015). The chamber maintained a photoperiod of 16 hours of light at 25°C and 8 hours of darkness at 18°C. After acclimatization, the seedlings were transplanted to two locations in Spain: Alcàsser, Valencia (39°23'40"N, 0°26'55"W) in April 2023 under field conditions and Almería, Andalucia (36°49'14"N, 2°14'1"W) in September 2023 under plastic greenhouse conditions.

The plants were arranged in a randomized complete block design with three blocks, spaced 1.2 m apart between rows and 1 m apart within rows. Vertical strings were used for support, and manual pruning was employed to manage vegetative growth and flowering. Plants were fertigated through drip irrigation system, and phytosanitary treatments were administered as necessary.

3.4. Phenotyping

Phenotypic data were collected from 308 S5 MAGIC eggplant lines by scoring anthocyanin pigmentation level in the leaf vein (LVAntho), stem (SAntho), and fruit skin (FSAntho). In total, 1,622 plants were phenotyped for LVAntho, 1,643 for SAntho, and 1,575 for FSAntho. Phenotyping of leaves and stems for LVAntho and SAntho was conducted using a scale from 0 to 3, where 0 indicates absence, 1 indicates low presence, 2 indicates intermediate presence, and 3 indicates high presence of pigmentation. For fruit skin, FSAntho was recorded as either present (1) or absent (0). Figure 3.3 provides a visual representation of the differences in anthocyanin pigmentation scores. To reduce the influence of environmental factors and measurement errors on the traits under study, statistical averages were obtained. Specifically, medians were calculated for LVAntho and SAntho given the ordinal nature of the data, whereas mean was computed for FSAntho due to its binary classification. Table 3.1 presents the variability among the S5 MAGIC founders in terms of anthocyanin pigmentation.

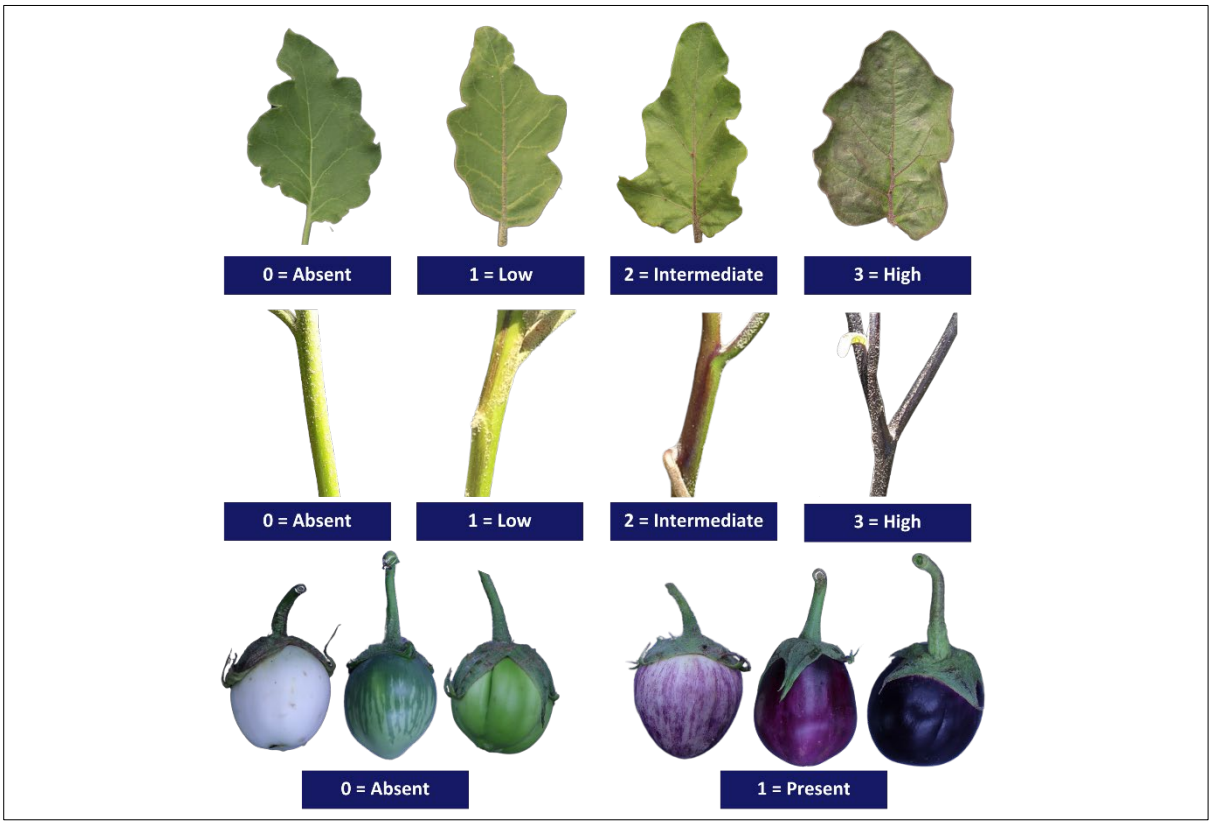

Figure 3.3. Anthocyanin pigmentation scoring in the leaf vein, stem, and fruit skin of the S5 MAGIC eggplant population.

Table 3.1. Anthocyanin pigmentation levels in the leaf vein (LVAntho), stem (SAntho), fruit skin (FSAntho) of the eight founder parents of the S5 MAGIC eggplant population.

Founder Parent Name	Founder Parent Code	Scientific Name	Origin	LVAntho	SAntho	FSAntho
MM1597	A	*S. melongena*	India	Absent	Absent	Absent
DH-ECAVI	B	*S. melongena*	Unknown	Low	Intermediate	Present
MM577	C	*S. incanum*	Israel	Absent	Absent	Absent
AN-S-26	D	*S. melongena*	Spain	Intermediate	High	Present
H15	E	*S. melongena*	Spain	Intermediate	High	Present
A0416	F	*S. melongena*	Unknown	Low	Low	Absent
IVIA371	G	*S. melongena*	Spain	Low	Intermediate	Present
ASI-S-1	H	*S. melongena*	China	High	High	Present

3.5. Genotyping

Leaf samples were collected from 308 S5 MAGIC eggplant lines. Genomic DNA was extracted from approximately 100 mg of young leaf using the Silica Matrix Extraction (SILEX) protocol (Vilanova et al., 2020). The quality and integrity of the extracted DNA were assessed via electrophoresis in 1% agarose gel, and spectrophotometry at absorbance ratios 260/230 and 260/280 using Nanodrop ND-1000 (Nanodrop Technologies, Wilmington, DE, USA). The concentration of the extracted DNA was determined using a Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, United States).

DNA samples were sequenced by SWGR with a read length of 150 bp, yielding approximately 3.9 Gb of high-quality paired-end reads per sample, corresponding to 3X coverage. These reads were aligned to the 67/3 eggplant reference genome version 3 (Barchi et al., 2019) using BWA-MEM (Li, 2013), followed by the removal of PCR duplicates with Picard version 1.119 (https://broadinstitute.github.io/picard/). Variant calling was conducted using Freebayes (Garrison and Marth, 2012), and biallelic SNPs were retained for further analysis using BCFtools version 1.13 (https://samtools.github.io/bcftools/bcftools.html). To mitigate false positives arising from low coverage sequencing, the data were compared to a gold standard generated from the 20X resequencing of the founder parents of the MAGIC population (Gramazio et al., 2019). Further filtering was performed: positions with more than 20% heterozygosity or monomorphic sites were excluded; genotypes supported by fewer than three reads were marked as missing; and loci with a minor allele frequency greater than 0.0031 were retained. Imputation of missing data was performed using Beagle version 22Jul22.46e (Browning and Browning, 2016), with the gold standard dataset serving as the reference panel, thereby ensuring high accuracy by retaining only original positions and excluding newly imputed ones.

The SWGR-generated SNPs were further filtered using TASSEL version 5.2.93 (Bradbury et al., 2007) to retain the most reliable ones (minor allele frequency > 0.15). In total, 107,295 SNPs were utilized for downstream analyses.

3.6. Population structure and kinship estimations, and heterozygosity assessment

MDS analysis was performed to investigate the genetic structure (Q) of the S5 MAGIC eggplant population using TASSEL version 5.2.93 (Bradbury et al., 2007). Additionally, genetic relatedness (K) and residual heterozygosity of the population were evaluated using the same software. The results were visually illustrated using ggplot2 and ggpubr packages in R (Wickham, 2016).

3.7. Genome-wide association mapping

Association mapping for anthocyanin pigmentation was conducted using the MLM of TASSEL version 5.2.93 (Bradbury et al., 2007), which incorporated the kinship (K) and population structure (Q) matrices as covariates. Multiple testing was corrected using the Bonferroni and FDR methods using the RAINBOWR package in R (Hamazaki and Iwata, 2020) to identify significant genomic regions at a significance level of 0.05. SNPs with a limit of detection (LOD) score (-log10[p-value]) exceeding the Bonferroni and FDR thresholds were considered significantly associated with anthocyanin pigmentation. Manhattan and QQ plots were generated using the qqman package in R (Turner, 2018) to visualize and interpret marker-trait associations and the distribution of p-values, respectively.

3.8. Candidate gene identification and linkage disequilibrium analysis

Anthocyanin-related genes located near the most significant SNP, also known as lead SNP, were identified using the 67/3 eggplant reference genome version 3 (Barchi et al., 2019) and considered as candidate genes controlling the assessed traits. LD patterns between the lead SNPs and the closest SNP to the candidate genes were assessed using the geneHapR package in R (Zhang et al., 2023b), with the correlation coefficient (r^2) used to measure pairwise LD.

4. RESULTS

4.1 Anthocyanin pigmentation

The S5 MAGIC eggplant population exhibited substantial phenotypic variability in LVAntho, SAntho, and FSAntho. Among the 1,622 S5 MAGIC individuals assessed for LVAntho, 620 (38%) exhibited no anthocyanin, 613 (38%) showed a low degree, 224 (14%) displayed intermediate levels, and 165 (10%) exhibited high levels of pigmentation. SAntho was evaluated in 1,643 individuals, with 372 (23%) showing no pigmentation, while 344 (21%), 649 (40%), and 278 (16%) individuals displayed low, intermediate, and high levels of anthocyanin, respectively. Lastly, among the 1,575 S5 MAGIC individuals evaluated for FSAntho, 989 (63%) lacked pigmentation, whereas 586 (37%) individuals showed presence of pigment. Phenotypic diversity of the S5 MAGIC eggplant population in terms of anthocyanin pigmentation in fruit is shown in **Figure 4.1**.

4.2. SNP marker distribution

Genotyping of 308 S5 MAGIC lines by SWGR coupled with imputation of missing marker data produced a total of 563,776 SNPs across the 12 chromosomes of eggplant (**Table 4.1**). Among these chromosomes, chromosome 9 had the fewest SNPs (18,080), while chromosome 1 had the most (71,385) (**Table 4.1**).

Subsequent filtering with TASSEL (minor allele frequency > 0.15) retained 107,295 markers, representing 19.03% of the total SNPs generated from SWGR (**Table 4.1**). Of the 12 eggplant

Figure 4.1. Phenotypic diversity of the S5 MAGIC eggplant population in terms of anthocyanin pigmentation in fruits.

chromosomes, chromosome 5 had the fewest SNPs (3,534), while chromosome 1 still had the most (17,354) (Table 4.1). These SNPs were then used for downstream analyses, including population structure and kinship estimations, residual heterozygosity assessment, and GWAS.

The SWGR-generated SNPs distributed across the eggplant genome had an overall average density of 492.52 SNPs/Mb, with chromosome 11 having the lowest density (457.71 SNPs/Mb) and chromosome 12 having the highest (522.88 SNPs/Mb) (Table 4.1). After filtering with TASSEL, the overall density was 96.33 SNPs/Mb, with chromosome 4 having the lowest density (43.54 SNPs/Mb) and chromosome 11 having the highest (185.90 SNPs/Mb) (Table 4.1). Figure 4.2 illustrates the genome-wide distribution of SNPs used in this study.

4.3. Population structure and residual heterozygosity

MDS analysis for assessing population structure, based on SNP marker data of 308 MAGIC eggplant lines, revealed that the first three principal components (PCs) explained 14.58% (PC1), 13.52% (PC2), and 8.94% (PC3) of the total variance (Figure 4.3). The 37.04% variance explained by these principal components indicates a noticeable level of genetic stratification within the population, as reflected by the moderate clustering observed in the MDS plots (Figure 4.4). Additionally, the S5 MAGIC eggplant population showed a low heterozygosity rate, with an average of 5.79%, according to SNP marker data (Figure 4.5). Of the 308 MAGIC eggplant lines, only 15 exhibited a heterozygosity rate exceeding 20%.

Table 4.1. Statistics of genome-wide SNPs generated from the skim whole-genome resequencing (SWGR) of the S5 MAGIC eggplant population.

Chromosome	SNPs[1]	Filtered SNPs[2]	SNPs (%)	Filtered SNPs (%)	Chromosome Length (Mb)[3]	SNP Density (SNPs/Mb)	Filtered SNP Density (SNPs/Mb)
1	71,385	17,354	12.66	16.17	136.53	522.84	127.10
2	41,342	8,125	7.33	7.57	83.34	496.06	97.49
3	46,055	6,252	8.17	5.83	97.01	474.72	64.44
4	49,578	4,601	8.79	4.29	105.67	469.17	43.54
5	21,810	3,534	3.87	3.29	43.85	497.34	80.59
6	54,647	11,621	9.69	10.83	108.97	501.48	106.64
7	69,346	8,886	12.30	8.28	142.38	487.04	62.41
8	52,426	5,599	9.30	5.22	109.58	478.45	51.10
9	18,080	3,688	3.21	3.44	36.10	500.82	102.16
10	53,511	10,892	9.49	10.15	106.64	501.77	102.13
11	33,088	13,439	5.87	12.53	72.29	457.71	185.90
12	52,508	13,304	9.31	12.40	100.42	522.88	132.48
Total	**563,776**	**107,295**	**100.00**	**100.00**	**1,142.80**		
Mean						**492.52**	**96.33**

[1]SNPs generated from the SWGR coupled with missing marker data imputation
[2]SNPs retained after filtration (minor allele frequency > 0.15) using TASSEL
[3]Chromosome length based on 67/3 eggplant refence genome version 3 (Barchi et al., 2019)

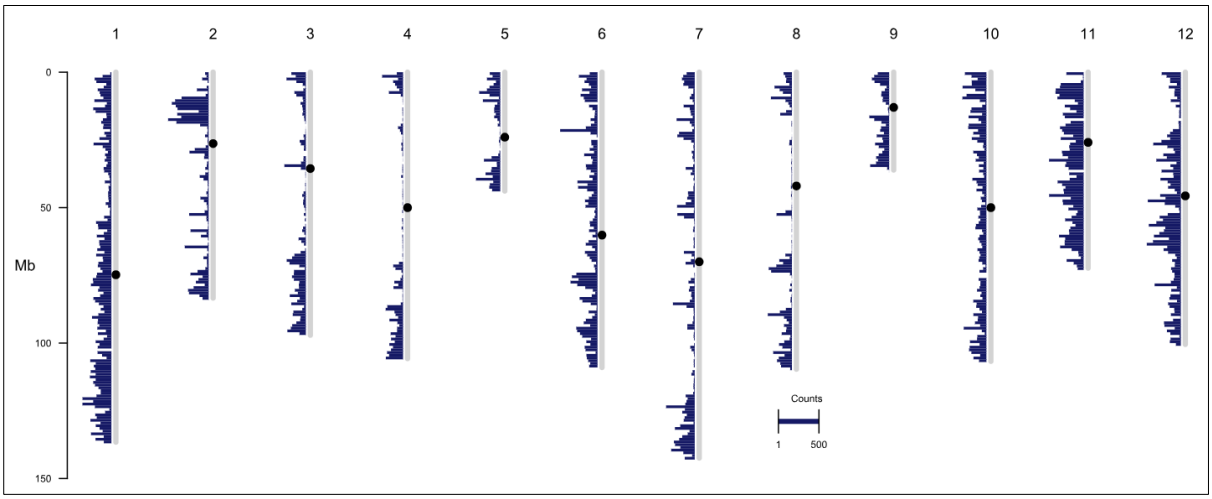

Figure 4.2. Genome-wide distribution of 107,295 SNPs used in this study.

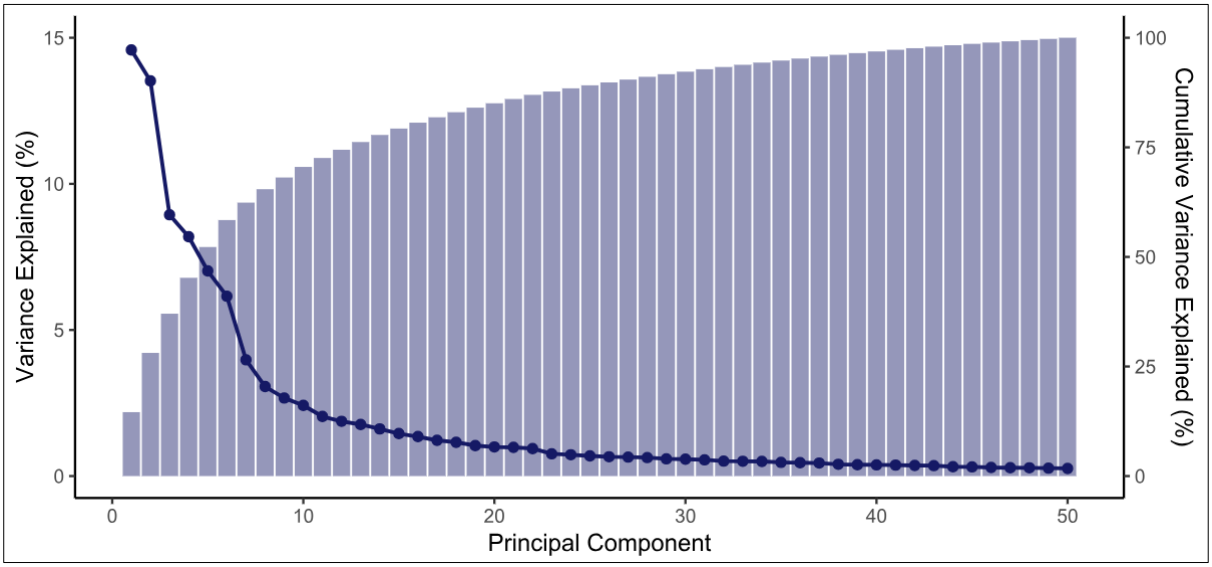

Figure 4.3. Scree plot of the variance (%, primary y-axis) and bar plot of the cumulative variance (%, secondary y-axis) explained by 50 principal components (x-axis).

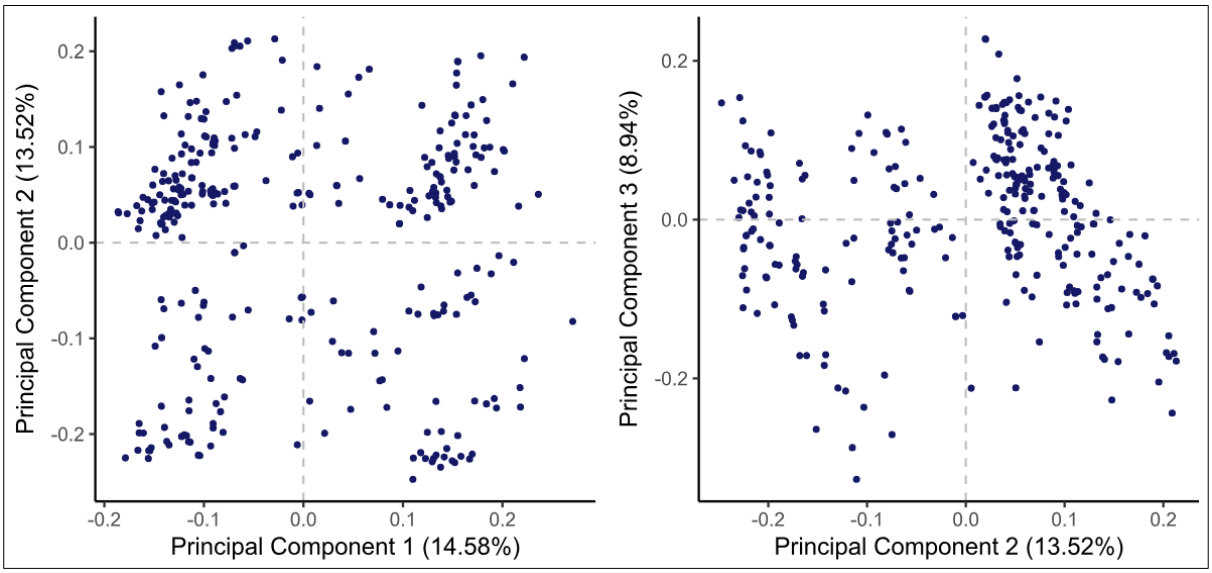

Figure 4.4. Multidimensional scaling plots for the 308 S5 MAGIC eggplant lines based on PC1 vs. PC2 (left) and PC2 vs. PC3 (right).

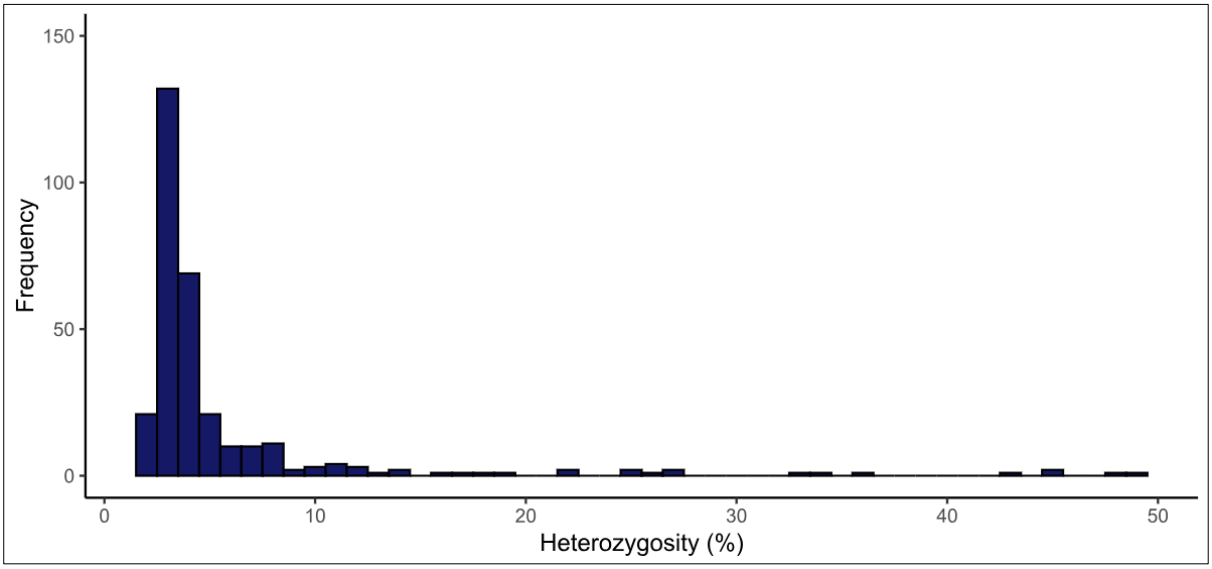

Figure 4.5. Residual heterozygosity rate (%) of the 308 S5 MAGIC eggplant lines, with an average of 5.79%, according to SNP marker data.

4.4. Marker-trait associations

The GWAS identified multiple SNP markers associated with anthocyanin pigmentation in eggplant, evidenced by peaks in the Manhattan plots surpassing the Bonferroni and FDR significance thresholds (Figures 4.6 to 4.8).

For LVAntho, SNP marker peaks exceeding the FDR significance threshold (LOD > 3.95) were identified on chromosomes 1, 8, 9, 10, and 12 (Figure 4.6). Notably, peaks on chromosomes 1 and 10 surpassed the more stringent Bonferroni significance threshold (LOD > 6.33) (Figure 4.6). The QQ plot indicated that the observed p-values closely followed the diagonal line in the lower range, suggesting reliable p-values with minimal inflation (Figure 4.6).

For SAntho, significant peaks above the FDR threshold (LOD > 4.22) were observed on chromosomes 1, 8, 9, and 10, with only the peak on chromosome 10 exceeding the Bonferroni threshold (LOD > 6.33) (Figure 4.7). The QQ plot showed slight deviation from the diagonal line, particularly at lower p-values, indicating minimal false-positive associations between SNP markers and the trait (Figure 4.7).

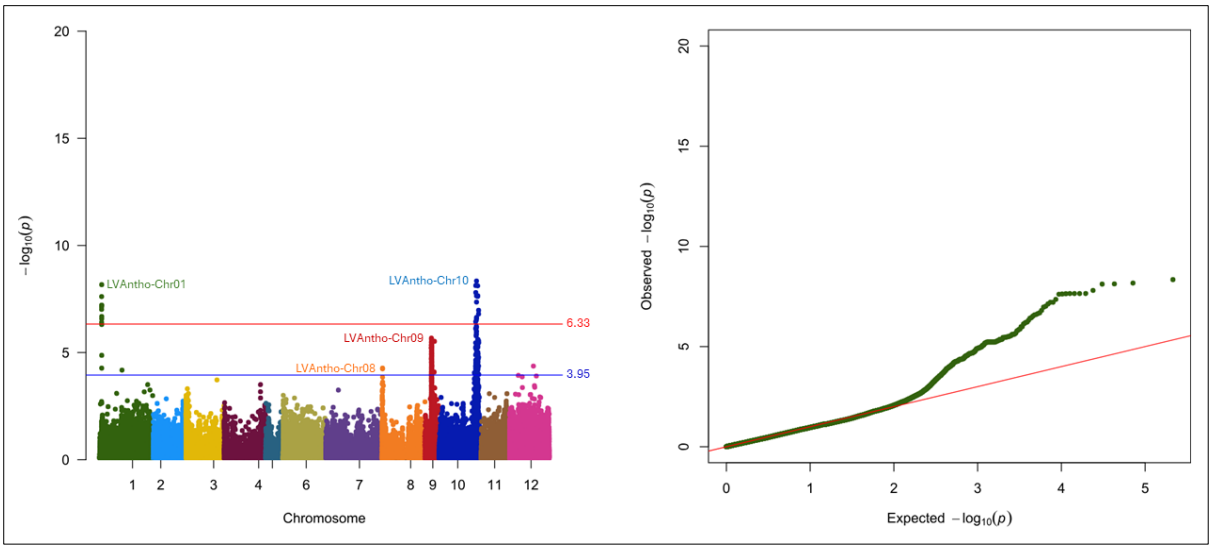

Figure 4.6. Manhattan (left) and QQ (right) plots for the genome-wide association mapping of LVAntho. The labelled SNPs were used to locate neighboring anthocyanin-related genes. Red and blue horizontal lines indicate Bonferroni and FDR significance thresholds, respectively.

For FSAntho, SNP markers surpassing the FDR significance threshold (LOD > 2.93) were detected across all chromosomes (Figure 4.8). However, only peaks on chromosomes 1 and 10 exceeded the more stringent Bonferroni significance threshold (LOD > 6.33) (Figure 4.8). The QQ plot exhibited significant deviations from the diagonal line, suggesting a high number of false-positive associations identified by the GWAS model (Figure 4.8).

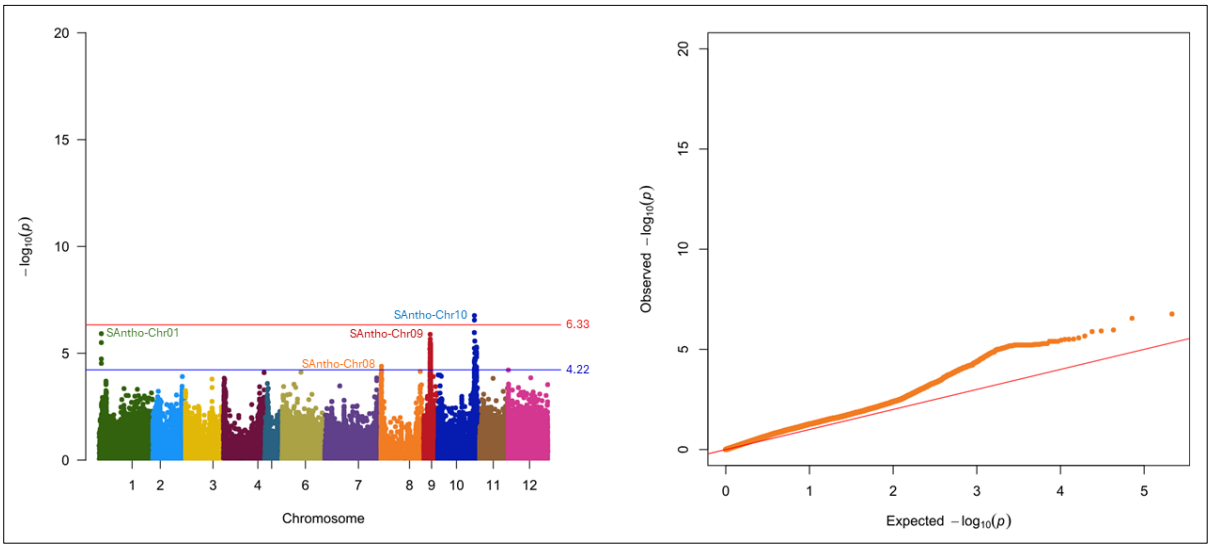

Figure 4.7. Manhattan (left) and QQ (right) plots for the genome-wide association mapping of SAntho. The labelled SNPs were used to locate neighboring anthocyanin-related genes. Red and blue horizontal lines indicate Bonferroni and FDR significance thresholds, respectively.

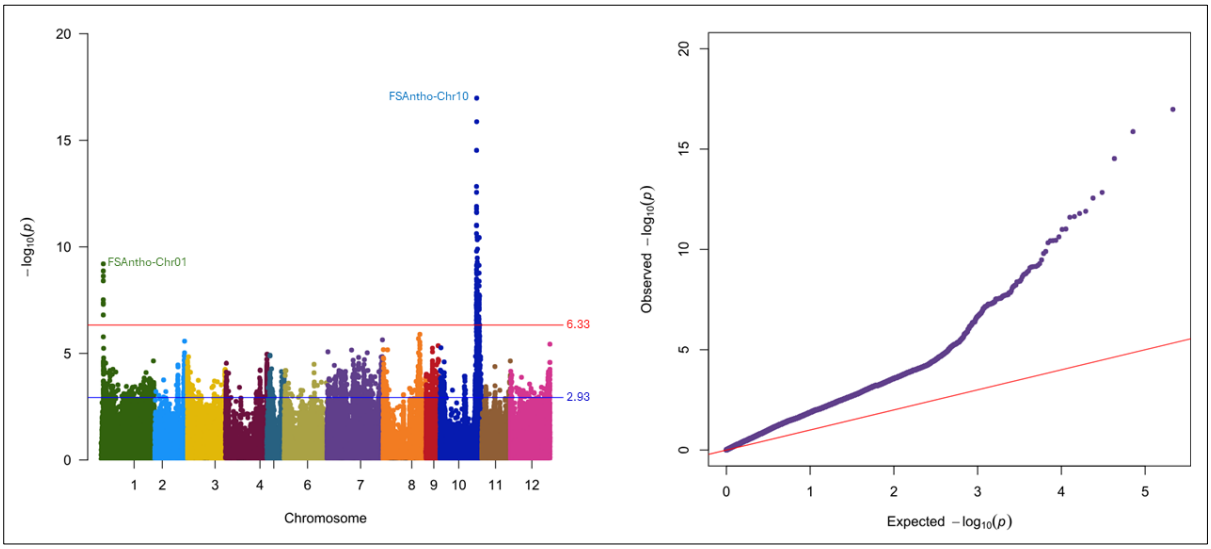

Figure 4.8. Manhattan (left) and QQ (right) plots for the genome-wide association mapping of FSAntho. The labelled SNPs were used to locate neighboring anthocyanin-related genes. Red and blue horizontal lines indicate Bonferroni and FDR significance thresholds, respectively.

Lead SNPs exceeding the FDR significance threshold for LVAntho (LOD > 3.95) and SAntho (LOD > 4.22) were utilized for identifying candidate genes, as suggested by QQ plots (**Figures 4.6** to **4.7**). However, for FSAntho, only lead SNPs surpassing the more stringent Bonferroni significance threshold (LOD > 6.33) were considered for candidate gene identification due to the high number of false-positive associations suggested by the QQ plot (**Figure 4.8**). Additionally, the sole significant SNP on chromosome 12 associated with LVAntho was excluded from candidate gene identification due to lack of nearby supportive SNPs (**Figure 4.6**).

4.5. Candidate genes and linkage disequilibrium patterns

Candidate genes located near the lead SNPs associated with anthocyanin pigmentation were mapped (**Table 4.2**). On chromosome 1, a gene similar to *MYB113* (SMEL_001g120500.1.01) was found in proximity to the three lead SNPs associated with the studied traits (SAntho-Chr01, FSAntho-Chr01, LVAntho-Chr01). On chromosome 8, the two lead SNPs associated with LVAntho (LVAntho-Chr08) and SAntho (SAntho-Chr08) were situated near a gene similar to *MYC1* (SMEL_008g298540.1.01). Additionally, on chromosome 9, the two lead SNPs associated with SAntho (SAntho-Chr09) and LVAntho (LVAntho-Chr09) were mapped close to genes similar to *bHLH69* (SMEL_009g326410.1.01) and *bHLH* (SMEL_009g326640.1.01). On chromosome 10, another gene similar to *MYB113* (SMEL_010g351850.1.01) was identified near a lead SNP associated with both SAntho (SAntho-Chr10) and FSAntho (SAntho-Chr10). Lastly, the lead SNP associated with LVAntho (LVAntho-Chr10) on chromosome 10 was positioned near two genes similar to *ANS* (SMEL_010g352310.1.01, SMEL_010g352330.1.01).

LD analyses revealed varying levels of association between the lead SNPs and candidate genes, as depicted by the correlation-based LD patterns (**Figures 4.9** to **4.11**; **Table 4.2**). On chromosome 1, all lead SNPs associated with the three traits (SAntho-Chr01, FSAntho-Chr01, LVAntho-Chr01) exhibited low degrees of LD ($0.05 \leq r^2 \leq 0.07$) with the candidate gene *MYB113*, suggesting independent inheritance between the candidate gene and the SNPs. In contrast, LVAntho-Chr08 and SAntho-Chr08 showed very strong ($r^2 = 0.98$) and strong ($r^2 = 0.76$) LD, respectively, with the candidate gene *MYC1* on chromosome 8. No haplotypes were observed between the genomic locations of lead SNPs on chromosome 9 (SAntho-Chr09, LVAntho-Chr09) and the candidate genes (*bHLH69* and *bHLH*) due to significant distances between them, thus LD patterns were not discernible. SAntho-Chr10 and FSAntho-Chr10, which represent the same SNPs located at the 92,853,636th base pair of chromosome 10, exhibited a moderate level of LD ($r^2 = 0.48$) with the candidate gene *MYB113*. Interestingly, LVAntho-Chr10 showed complete LD ($r^2 = 1.00$) with two genes corresponding to *ANS*, despite being separated by over a million bases.

Table 4.2. Candidate genes identified near the lead SNPs associated with anthocyanin pigmentation in eggplant and their degree of linkage disequilibrium (LD).

Chr	Lead SNP Position	Lead SNP Name[1]	Candidate Gene[2]			LD[3]
			Name	Identifier	Position	
1	5,399,462	SAntho-Chr01	*MYB113*	SMEL_001g120500	5,386,038 - 5,387,495	0.05
	5,422,728	FSAntho-Chr01				0.05
	5,652,065	LVAntho-Chr01				0.07
8	1,559,003	LVAntho-Chr08	*MYC1*	SMEL_008g298540	1,587,564 - 1,589,468	0.98
	1,583,879	SAntho-Chr08				0.76
9	16,245,752	SAntho-Chr09	*bHLH69*[4]	SMEL_009g326410	17,464,629 - 17,472,956	
	16,415,919	LVAntho-Chr09	*bHLH*[4]	SMEL_009g326640	17,862,102 - 17,872,412	
10	92,853,636	SAntho-Chr10 FSAntho-Chr10	*MYB113*	SMEL_010g351850	92,326,476 - 93,331,202	0.48
	94,671,634	LVAntho-Chr10	*ANS*	SMEL_010g352310	93,595,482 - 93,603,287	1.00
			ANS	SMEL_010g352330	93,627,835 - 93,636,310	

[1]Lead SNP names designated for simple identification in this study
[2]Candidate genes identified using the 67/3 eggplant reference genome version 3 (Barchi et al., 2019)
[3]LD based on correlation coefficient (r^2) between the lead SNP and the closest SNP to the candidate gene
[4]Candidate genes with no support from LD analysis due to large distance from the lead SNPs

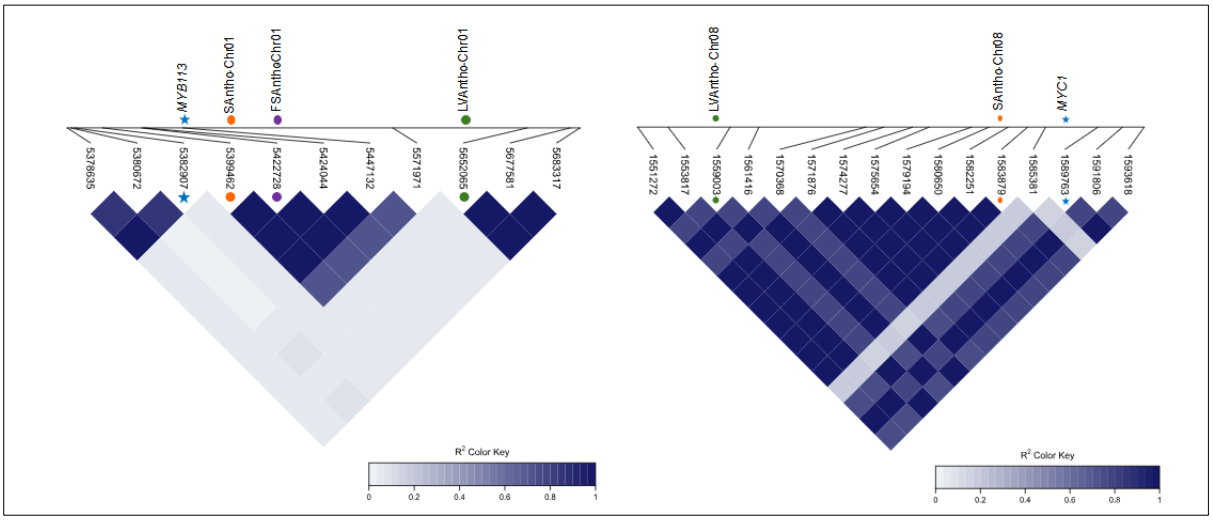

Figure 4.9. Linkage disequilibrium patterns between the lead SNPs on chromosome 1 (LVAntho-Chr01, SAntho-Chr01, FSAntho-Chr01) and the closest SNP (position 5,382,907) to *MYB113*-like gene (left); and between the lead SNPs on chromosome 8 (LVAntho-Chr08, SAntho-Chr08) and the closest SNP (position 1,589,763) to *MYC1*-like gene (right).

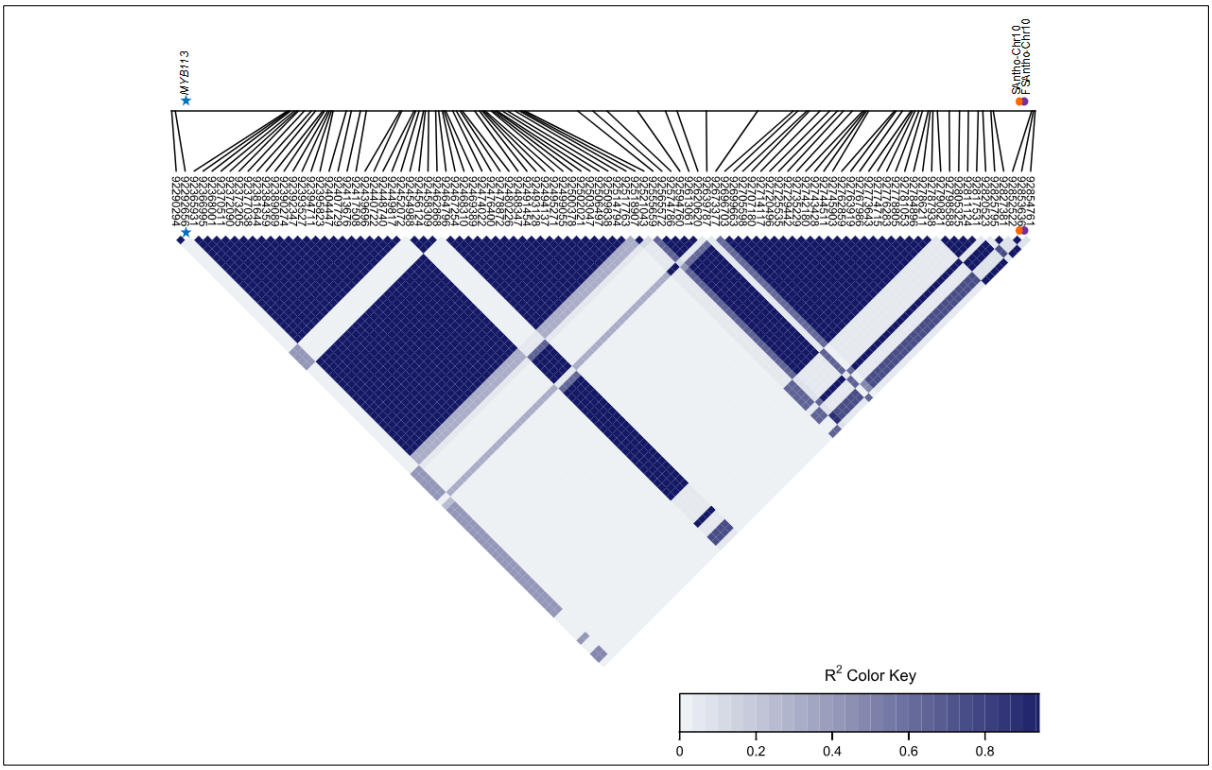

Figure 4.10. Linkage disequilibrium patterns between the lead SNPs on chromosome 10 (SAntho-Chr10, FSAntho-Chr10) and the closest SNP (position 92,292,656) to *MYB113*-like gene.

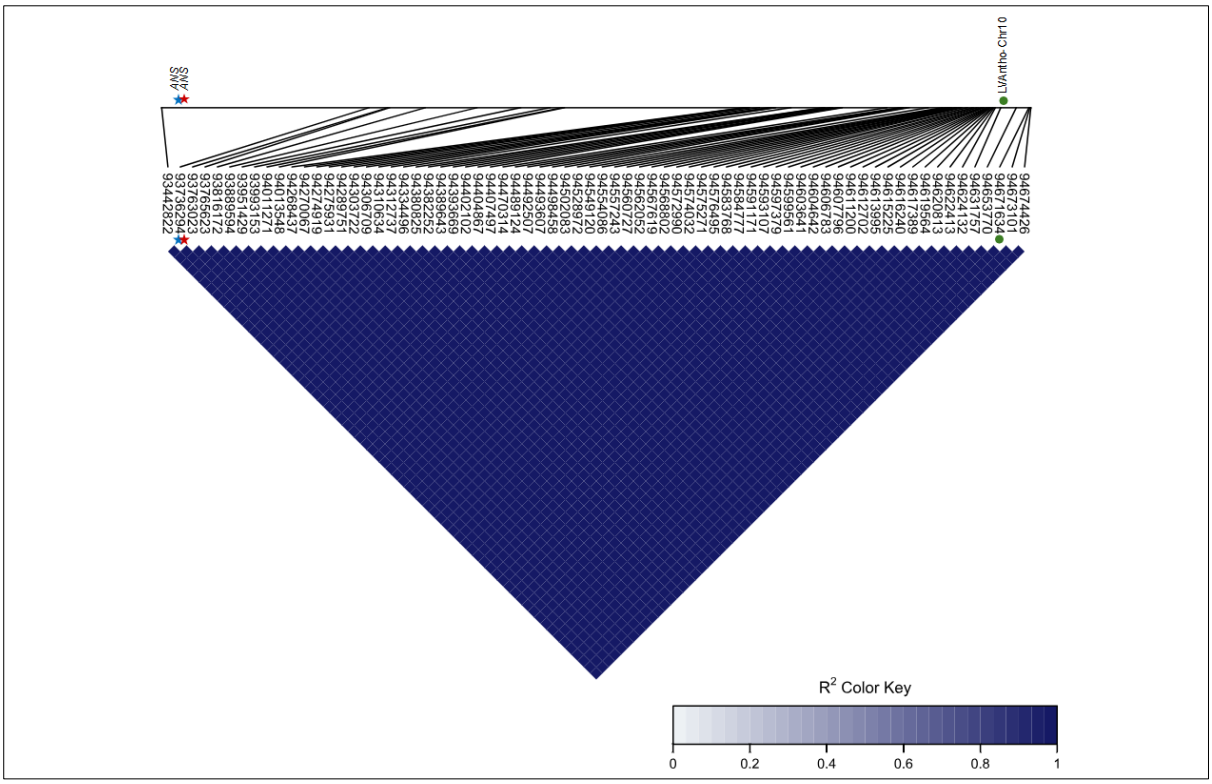

Figure 4.11. Linkage disequilibrium patterns between the lead SNP on chromosome 10 (LVAntho-Chr10) and the closest SNP (position 93,736,294) to two *ANS*-like genes.

5. DISCUSSION

Anthocyanins are essential compounds responsible for the purple coloration in various parts of eggplant, including its leaves, stem, flowers, and fruits. Acting as antioxidants, anthocyanins enhance the postharvest quality of crops by maintaining membrane integrity (Jiao et al., 2012). They also play a significant role in reducing insect infestation and pathogen attacks (Malone et al., 2009; Zhang et al., 2013). Given the high market demand for purple-colored eggplants (Li et al., 2018), understanding the genetic control of anthocyanin biosynthesis is crucial for eggplant breeding programs. To investigate the genetic basis of anthocyanin pigmentation in eggplant, particularly in the leaf vein, stem, and fruit skin, a GWAS was conducted using a MAGIC population.

5.1. Multiparent mapping population

The MAGIC population utilized in this study underwent five generations of self-fertilization, resulting in an S5 population (**Figure 3.2**), which followed the S3 population used in GWAS for anthocyanin-related traits by Mangino et al. (2022) and chlorophyll pigmentation by Arrones et al. (2022). As compared with germplasm panels and biparental populations, MAGIC populations provide superior genetic material for identifying gene-trait associations due to their enhanced mapping resolution (Arrones et al., 2020; Scott et al., 2020). The S5 MAGIC eggplant population used in this study was developed using eight founders with high degree of genetic and phenotypic diversity (Hurtado et al., 2014; Gramazio et al., 2017; Kaushik et al., 2018; Gramazio et al., 2019), including a wild relative of eggplant, *S. incanum*. The incorporation of multiple founders, along with multiple rounds of intercrossing and self-fertilization, leads to a higher number of recombinant events, thereby improving mapping accuracy (Scott et al., 2020). The inclusion of a wild relative improves genetic variability within the population, which is crucial for identifying QTLs controlling the trait of interest (Gramazio et al., 2020). Furthermore, the use of 308 S5 MAGIC eggplant lines in this study is essential, as larger population sizes are important for increasing the power and mapping resolution of GWAS (Collard et al., 2005; Valdar et al., 2006; Jaganathan et al., 2020). Using a simple funnel method to intercross the diverse MAGIC founders (**Figure 3.2**), a substantial phenotypic diversity was observed among the S5 MAGIC individuals, displaying varying levels of LVAntho, SAntho, and FSAntho (**Figure 4.1**).

5.2. Genome-wide marker system

In addition to the mapping population, the success of GWAS is significantly influenced by the genomic coverage of SNP markers (Haghi et al., 2022). SNP genotyping offers the benefit of delivering comprehensive whole-genome data at an affordable cost, enabling high-resolution mapping with a substantial number of SNPs throughout the genome (Seo et al., 2020). In this study, SWGR was employed to generate a well-distributed set of SNPs across the genomes of 308 S5 MAGIC eggplant lines, yielding a total of 107,295 SNPs with an average density of 96.33 SNPs/Mb (**Figure 4.2**; **Table 4.1**). This represents nearly a fourteenfold increase in the number

of SNPs and more than a twelvefold increase in SNP density compared to those produced by single primer enrichment technology in the similar studies conducted by Mangino et al. (2022) and Arrones et al. (2022) using the S3 MAGIC eggplant population. The substantial quantity of the genome-wide SNP data produced by the SWGR platform was particularly enhanced by the imputation of missing genotype data. This imputation was accurately conducted using the gold standard dataset generated from the genome resequencing of the eight MAGIC founders with 20X coverage performed by Gramazio et al. (2019). Imputing missing genotypic data is a common practice in genetic studies, including association mapping, to increase statistical power and refine mapping resolutions (Balding, 2006; Asimit and Zeggini, 2010; Marchini and Howie, 2010). The high-quality SNP data generated by SWGR revealed low residual heterozygosity rate in the 308 S5 MAGIC eggplant lines, averaging 5.79% (**Figure 4.5**). This low heterozygosity rate can be attributed to the self-pollinating nature of eggplants (Frary et al., 2007) and the five generations of self-fertilization employed to develop the population (**Figure 3.2**). Out of the 308 S5 MAGIC eggplant lines, only 15 showed a heterozygosity rate exceeding 20% (**Figure 4.5**), which is the same result found in the earlier studies by Mangino et al. (2022) and Arrones et al. (2022) using the S3 MAGIC population.

5.3. Association mapping model

While the use of a diverse multiparent population and well-distributed SNP marker data significantly improves the efficacy of GWAS, the statistical model employed is equally crucial. These models are used to calculate significant associations between genetic markers and trait of interest based on p-value, and variance components for the trait are calculated using analysis of variance to work out allelic effects (Kulwal and Singh, 2021). In this study, MLM was utilized, incorporating both population structure effects (Q) and kinship effects (K) to control for false marker-trait associations (Pritchard et al., 2000). Within the MLM framework, the Q matrix is fitted as a fixed effect while the K matrix serves as the variance-covariance matrix of the random effect (Pasam and Sharma, 2014). The justification for employing MLM in this study was supported by the MDS analysis using the SNP marker data, which indicated a moderate population stratification among the S5 MAGIC eggplant population (**Figure 4.4**). The first three principal components explained a cumulative variance of 37.04% (**Figure 4.3**), highlighting the need to account for population structure to avoid spurious associations. Since population structure is the major confounding factor for association mapping studies, the capacity of MLM to statistically account for population structure is essential for correcting the inflation of small genetic effects and controlling bias caused by this stratification (Devlin and Roeder, 1999; Pasam and Sharma, 2014; Liu et al., 2016a; Alseekh et al., 2021). Numerous studies employing various models such as GLM and MLM consistently favor the latter, demonstrating its superior performance across different scenarios (Zhao et al., 2007; Kang et al., 2008; Stich and Melchinger, 2009; Pasam et al., 2012).

5.4. Marker-trait associations

GWAS has become a powerful method for identifying genomic regions that control traits in crop species. By integrating phenotype data on eggplant anthocyanin pigmentation with SNP marker data, this study identified multiple marker-trait associations, represented as peaks of significant SNPs in Manhattan plots (**Figures 4.6** to **4.8**). Longer peaks in the Manhattan plot indicate stronger associations of nearby genomic regions with the trait (Kulwal and Singh, 2021). However, establishing significance thresholds is essential to distinguish true associations from false discoveries (Jain et al., 2024). This is especially important in GWAS, which involves independently testing numerous markers for association, thereby increasing the risk of false positive associations (Jain et al., 2024). In this study, significance thresholds were determined using Bonferroni and FDR correction methods to account for multiple testing. While SNPs surpassing these thresholds are suggested to be significantly associated with the trait, QQ plots served as a diagnostic tool to filter these significant associations. QQ plots display the observed distribution of p-values against their expected values under the null hypothesis of no association (Jain et al., 2024). Since most SNPs tested are not associated with the trait, the QQ plot should show observed p-values aligning closely with the expected line, particularly in the lower regions where there are generally no marker-trait associations (Jain et al., 2024). Depending on the deviations visualized in the QQ plots, either the Bonferroni or FDR significance threshold was selected to determine reliable associations between SNPs and the trait. In this study, QQ plots for LVAntho and SAntho showed that the observed p-values closely followed the diagonal line in the lower region, indicating good control over false positives (**Figures 4.6** to **4.7**). In contrast, the QQ plot for FSAntho showed considerable deviations, suggesting possible confounding factors that were not fully accounted for by the GWAS model used (**Figure 4.8**). As a result, the significant SNP peaks on chromosomes 1 and 10 associated with all three traits studied and those on chromosomes 8 and 9 for both LVAntho and SAntho were considered the most reliable association signals (**Figures 4.6** to **4.8**).

5.5. Candidate genes and linkage disequilibrium patterns

The anthocyanin biosynthetic pathway is a well-characterized and highly conserved network across many plant species (Holton and Cornish, 1995; Tanaka and Ohmiya, 2008). This pathway has been extensively studied in solanaceous plants (Chaim et al., 2003; Borovsky et al., 2004; De Jong et al., 2004; Bovy et al., 2007; Gonzali et al., 2009). It involves the interplay of multiple structural and regulatory genes essential for the synthesis and accumulation of anthocyanins (Liu et al., 2018). To enhance our understanding of anthocyanin biosynthesis and regulation, this study aims to identify these key genes in eggplant.

By focusing on the most reliable peaks identified by GWAS with the aid of Bonferroni and FDR significance thresholds as well as of QQ plots, lead SNPs within these peaks were located and

candidate genes near these lead SNPs were mapped (**Table 4.2**). In this study, a total of 10 most reliable SNP peaks distributed on chromosomes 1, 8, 9, and 10 of eggplant were identified: four were associated with LVAntho (LVAntho-Chr01, LVAntho-Chr08, LVAntho-Chr09, LVAntho-Chr10) and SAntho (SAntho-Chr01, SAntho-Chr08, SAntho-Chr09, SAntho-Chr10), and two were associated with FSAntho (FSAntho-Chr01, FSAntho-Chr10) (**Figures 4.6** to **4.8**; **Table 4.2**). In the vicinity of these lead SNPs, four different genes – *MYB113*, *bHLH*, *MYC1*, and *ANS* – were mapped, each showing varying degrees of LD with the respective lead SNPs they are associated with (**Table 4.2**).

Firstly, this study successfully identified a gene corresponding the *MYB113* in *Arabidopsis thaliana* near three lead SNPs on chromosome 1 (LVAntho-Chr01, SAntho-Chr10, FSAntho-Chr10) and two lead SNPs on chromosome 10 (SAntho-Chr10, FSAntho-Chr10) (**Table 4.2**). MYB113 belongs to the R2R3-MYB class within the MYB family, which is one of the largest transcription factor families in plants (Dubos et al., 2010). R2R3-MYB transcription factors serve as primary regulators that activate the expression of genes involved in anthocyanin biosynthesis (Stracke et al., 2004), with MYB113 specifically known for its role in regulating anthocyanin synthesis in eggplant. Known alternatively as *SmMYB1*, *SmMYB113* interacts with *CBF* (C-repeat binding factors) genes to regulate the expression of structural anthocyanin synthesis genes such as *SmCHS* and *SmDFR* in eggplant (Liu et al., 2016b; Zhou et al., 2020; Zhang et al., 2023). Studies involving eggplant transformation have shown that *SmMYB113* enhances anthocyanin biosynthesis in calluses (Shi et al., 2021) as well as of peels and pulps (Yang et al., 2022). Beyond eggplant, the regulation of anthocyanin synthesis by R2R3-MYB transcription factors has been observed in tomato (Kiferle et al., 2015), apple (Takos et al., 2006), potato (Liu et al., 2016b), bayberry (Niu et al., 2010), and tobacco (Li et al., 2017). The mapping of *MYB113* to chromosomes 1 and 10 in this study is consistent with findings from a previous association mapping study on eggplant anthocyanin pigmentation conducted by Mangino et al. (2022). Analysis revealed a low degree of LD ($0.05 \leq r^2 \leq 0.07$) between the lead SNPs on chromosome 1 and the candidate gene *MYB113*, while a moderate level of LD ($r^2 = 0.48$) was found between the lead SNPs on chromosome 10 and *MYB113* (**Figures 4.9** to **4.10**; **Table 4.2**). Although these levels of LD do not indicate strong associations between the lead SNPs and the candidate gene *MYB113*, existing literature provides evidence of the role of this gene in anthocyanin pigmentation in a wide range of plant species.

Secondly, this study identified a gene similar to *bHLH69* in *A. thaliana* and *bHLH* in *Pisum sativum* near two lead SNPs, SAntho-Chr09 and LVAntho-Chr09 (**Table 4.2**). bHLH proteins are the second largest class of transcription factors in plants (Duan et al., 2021). It is a component of the MYB-bHLH-WD40 regulatory complex that governs the activity of structural genes in the anthocyanin biosynthetic pathway (Liu et al., 2018). While the regulatory function of MBW complex is primarily determined by MYB transcription factors, bHLH transcription factors play a crucial role in recognizing specific transcription factor binding sites in target gene promoters and activating transcription (Montefiori et al., 2015). In pepper and eggplant,

significantly higher transcript levels of *CabHLH* and *SmbHLH* have been observed in anthocyanin-pigmented fruits compared to non-pigmented ones (Stommel et al., 2009; Stommel and Dumm, 2015; Gisbert et al., 2016). The upregulation of these genes correlates positively with increased expression levels of structural genes and higher anthocyanin content (Liu et al., 2018). The function of bHLH transcription factors in flavonoid metabolism has also been investigated in other plants such as apple (Xie et al., 2017), gentian (Nakatsuka et al., 2008), and dendrobium (Li et al., 2017). In a related study by Mangino et al. (2022), the *bHLH* gene was also mapped to chromosome 9, consistent with the findings of the current study. Because the mapped *bHLH*-like genes and lead SNPs on chromosome 9 were significantly distant from each other, accurately assessing LD was not feasible in this study (**Table 4.2**). This observation aligns with findings by Mangino et al. (2022), who also located the *bHLH* gene outside of LD blocks. Nonetheless, prior research has highlighted the crucial roles of bHLH transcription factors in the regulation of anthocyanin biosynthesis in eggplant.

Thirdly, a gene similar to *MYC1* in *S. lycopersicum* was mapped near the lead SNPs LVAntho-Chr08 and SAntho-Chr08 (**Table 4.2**). MYC1 belongs to the bHLH protein family, containing the conserved bHLH domain signature, and is localized in the plant cell nucleus (Hichri et al., 2010). In grapevine, *VvMYC1* regulates anthocyanin synthesis, with transcript levels correlating with anthocyanin production in berry skins and seeds during development (Hichri et al., 2010). Transient expression of VvMYC1 revealed that it can interact with MYB-type transcription factors, such as VvMYBA1, VvMYBPA1, VvMYB5a, and VvMYB5b, to activate genes in the anthocyanin biosynthetic pathway and its own transcription (Kobayashi et al., 2002; Bogs et al., 2007; Walker et al., 2007; Deluc et al., 2008; Terrier et al., 2009; Hichri et al., 2010). A similar mechanism was observed in gerbera, where the interaction between GMYC1 and GMYB10 regulates anthocyanin biosynthesis by activating the promoter of a late anthocyanin biosynthetic gene, *PGDFR2* (Elomaa et al., 1998). In this study, LD analysis revealed that the candidate gene *MYC1* exhibited strong LD ($0.76 \leq r^2 \leq 0.98$) with the SNPs LVAntho-Chr08 and SAntho-Chr08, reinforcing the evidence that the candidate gene is associated with anthocyanin pigmentation (**Figure 4.9**; **Table 4.2**).

Lastly, this study identified two candidate genes near the lead SNP LVAntho-Chr10, corresponding to *ANS* of *A. thaliana* (**Table 4.2**). ANS, also known as leucoanthocyanidin dioxygenase (LDOX), is a key enzyme in the anthocyanin biosynthetic pathway that catalyzes the conversion of leucoanthocyanidins to anthocyanidins (Davies et al., 2003; Turnbull et al., 2004; Wellmann et al., 2006; Liu et al., 2018). It belongs to the 2-oxoglutarate-dependent (2ODD) oxygenase family, which utilizes molecular oxygen as a co-substrate and requires co-factors such as ascorbate, 2-oxoglutarate, and/or Fe^{2+} (Chua et al., 2008; Wang et al., 2019). In eggplant, the expression levels of *SmANS* are significantly higher in black or violet fruits compared to green or white fruits across all fruit developmental stages, including commercial ripeness (Stommel and Dumm, 2015; Gisbert et al., 2016). Similarly, in tomato and potato, *ANS* genes (*SlANS* and *StANS*, respectively) are highly expressed in anthocyanin-rich tissues

compared to non-pigmented controls (Sapir et al., 2008; André et al., 2009; Jung et al., 2009; Povero et al., 2011; Liu et al., 2015). As a structural gene, *ANS* is regulated by various transcription factors such as MYB, bHLH, and WD40-repeat proteins (Gonzalez et al., 2008). In carrot, for instance, the expression of *RsANS* correlates with *RsMYB1*, an ortholog of *AtMYB113* (Yi et al., 2018; Lim et al., 2016). Interestingly, in this study, while the two candidate *ANS* genes were mapped to chromosome 10, *MYB113* was found upstream, approximately 1.3 Mb away (**Table 4.2**). The spatial positioning of *MYB113* upstream of *ANS* may imply that *MYB113* directly regulates the expression of *ANS*, suggesting a regulatory cascade where *MYB113* activation leads to increased transcription of *ANS*, thereby promoting anthocyanin production. In contrast to a similar study by Mangino et al. (2022), which did not identify *ANS* genes as candidates for anthocyanin pigmentation, this study not only identified *ANS* but also found it to exhibit complete LD (r^2 = 1.00) with the lead SNP LVAntho-Chr10 (**Figure 4.11**, **Table 4.2**).

6. CONCLUSIONS

Anthocyanin pigmentation varies quantitatively across different parts of eggplant. Recognizing the importance of this trait in eggplant breeding, this study elucidates the genetic basis of anthocyanin biosynthesis and regulation by performing a GWAS. By integrating the anthocyanin pigmentation data gathered from the diverse S5 MAGIC eggplant population with genome-wide SNP marker data generated by the SWGR technique, this study highlights the (1) SNPs associated with anthocyanin pigmentation, (2) candidate genes putatively controlling anthocyanin pigmentation, and (3) LD patterns between these SNPs and candidate genes.

Firstly, numerous SNPs associated with anthocyanin pigmentation were found on chromosomes 1, 8, 9, and 10. Within the most reliable SNP peaks, 10 lead SNPs were identified: four were associated with LVAntho (LVAntho-Chr01, LVAntho-Chr08, LVAntho-Chr09, LVAntho-Chr10), four with SAntho (SAntho-Chr01, SAntho-Chr08, SAntho-Chr09, SAntho-Chr10), and two with FSAntho (FSAntho-Chr01, FSAntho-Chr10). These marker-trait associations were identified using MLM, which accounts for population structure and kinship in association mapping.

Secondly, candidate genes putatively controlling anthocyanin pigmentation were mapped near the lead SNPs. Using the eggplant reference genome sequence, three regulatory genes (*MYB113*, *bHLH*, *MYC1*) and a structural gene (*ANS*) involved in anthocyanin pigmentation were identified. Specifically, *MYB113* was found on chromosomes 1 and 10, *bHLH* on chromosome 9, *MYC1* on chromosome 8, and ANS on chromosome 10. Notably, the spatial positioning of *MYB113* upstream of *ANS* on chromosome 10, approximately 1.3 Mb apart, suggests a regulatory interaction that could potentially lead to increased anthocyanin pigmentation.

Lastly, correlation-based LD patterns between lead SNPs and candidate genes varied. *MYB113* showed low LD with lead SNPs on chromosome 1 (SAntho-Chr01, FSAntho-Chr01, LVAntho-Chr01) and moderate LD with the lead SNPs on chromosome 10 (SAntho-Chr10, FSAntho-Chr10). The considerable genomic distance between the lead SNPs on chromosome 9 (SAntho-Chr09, LVAntho-Chr09) and the two *bHLH* genes hindered LD assessment. However, the *bHLH* gene family member *MYC1* exhibited strong LD with lead SNPs on chromosome 8 (LVAntho-Chr08, SAntho-Chr08). Confirming its direct role in anthocyanin biosynthesis, *ANS* showed complete LD with the lead SNP on chromosome 10 (LVAntho-Chr10).

Despite extensive researches on anthocyanin biosynthesis and regulation in a wide range of plant species, this study provides additional evidence regarding the genetic basis of anthocyanin pigmentation in eggplant, while emphasizing the utility of a diverse MAGIC population and dense SNP marker data in association mapping. The findings of this study offer valuable insights for future eggplant breeding efforts aimed at enhancing anthocyanin content and meeting market demands for purple-colored eggplants.

7. BIBLIOGRAPHY

Adhikari, L., Shrestha, S., Wu, S., Crain, J., Gao, L., Evers, B., Wilson, D., Ju, Y., Koo, D. H., Hucl, P., Pozniak, C., Walkowiak, S., Wang, X., Wu, J., Glaubitz, J. C., DeHaan, L., Friebe, B., & Poland, J. (2022). A high-throughput skim-sequencing approach for genotyping, dosage estimation and identifying translocations. Scientific Reports, 12(1), 17583. https://doi.org/10.1038/s41598-022-19858-2

Albert, N. W., Davies, K. M., Lewis, D. H., Zhang, H., Montefiori, M., Brendolise, C., Boase, M. R., Ngo, H., Jameson, P. E., & Schwinn, K. E. (2014). A conserved network of transcriptional activators and repressors regulates anthocyanin pigmentation in eudicots. The Plant Cell, 26(3), 962-980. https://doi.org/10.1105/tpc.113.122069

Alseekh, S., Kostova, D., Bulut, M., & Fernie, A. R. (2021). Genome-wide association studies: assessing trait characteristics in model and crop plants. Cellular and Molecular Life Sciences, 78(15), 5743-5754. https://doi.org/10.1007/s00018-021-03868-w

André, C. M., Schafleitner, R., Legay, S., Lefèvre, I., Aliaga, C. A. A., Nomberto, G., Hoffmann, L., Hausman, J. F., Larondelle, Y., & Evers, D. (2009). Gene expression changes related to the production of phenolic compounds in potato tubers grown under drought stress. Phytochemistry, 70(9), 1107–1116. https://doi.org/10.1016/j.phytochem.2009.07.008

Arrones, A., Mangino, G., Alonso, D., Plazas, M., Prohens, J., Portis, E., Barchi, L., Giuliano, G., Vilanova, S., & Gramazio, P. (2022). Mutations in the SmAPRR2 transcription factor suppressing chlorophyll pigmentation in the eggplant fruit peel are key drivers of a diversified colour palette. Frontiers in Plant Science, 13, 1025951. https://doi.org/10.3389/fpls.2022.1025951

Arrones, A., Vilanova, S., Plazas, M., Mangino, G., Pascual, L., Díez, M. J., Prohens, J., & Gramazio, P. (2020). The dawn of the age of multi-parent MAGIC populations in plant breeding: Novel powerful next-generation resources for genetic analysis and selection of recombinant elite material. Biology (Basel), 9(8), Article 229. https://doi.org/10.3390/biology9080229

Asimit, J., & Zeggini, E. (2010). Rare variant association analysis methods for complex traits. Annual Review of Genetics, 44, 293–308.

Bai, X., Zhao, H., Huang, Y., Xie, W., Han, Z., Zhang, B., Guo, Z., Yang, L., Dong, H., Xue, W., Li, G., Hu, G., Hu, Y., & Xing, Y. (2016). Genome-wide association analysis reveals different genetic control in panicle architecture between Indica and Japonica rice. The Plant Genome, 9(2), 1–10.

Balding, D. J. (2006). A tutorial on statistical methods for population association studies. Nature Reviews Genetics, 7, 781–791.

Barchi, L., Pietrella, M., Venturini, L., Minio, A., Toppino, L., Acquadro, A., Andolfo, G., Aprea, G., Avanzato, C., Bassolino, L., Comino, C., Molin, A. D., Ferrarini, A., Maor, L. C., Portis, E., Reyes-Chin-Wo, S., Rinaldi, R., Sala, T., Scaglione, D., Sonawane, P., Tononi, P., Almekias-Siegl, E., Zago, E., Ercolano, M. R., Aharoni, A., Delledonne, M., Giuliano, G., Lanteri, S., & Rotino, G. L. (2019). A chromosome-anchored eggplant genome sequence reveals key events in Solanaceae evolution. Scientific Reports, 9(1), Article 11769. https://doi.org/10.1038/s41598-019-47985-w

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological), 57, 289–300. https://www.jstor.com/stable/2346101

Bergelson, J., & Roux, F. (2010). Towards identifying genes underlying ecologically relevant traits in Arabidopsis thaliana. Nature Reviews Genetics, 11(12), 867–879.

Bhaduri, P. N. (1951). Inter-relationship of the non-tuberiferous species of Solanum with some consideration on the origin of brinjal (*Solanum melongena* L.). Indian Journal of Genetics and Plant Breeding, 11, 75–82.

Bogs, J., Jaffé, F. W., Takos, A. M., Walker, A. R., & Robinson, S. P. (2007). The grapevine transcription factor VvMYBPA1 regulates proanthocyanidin synthesis during fruit development. Plant Physiology, 143(4), 1347-1361. https://doi.org/10.1104/pp.106.093203

Borovsky, Y., Oren-Shamir, M., Ovadia, R., De Jong, W., & Paran, I. (2004). The A locus that controls anthocyanin accumulation in pepper encodes a MYB transcription factor homologous to Anthocyanin2 of Petunia. Theoretical and Applied Genetics, 109(1), 23-29. https://doi.org/10.1007/s00122-004-1625-9

Bovy, A., Schijlen, E., & Hall, R. (2007). Metabolic engineering of flavonoids in tomato (*Solanum lycopersicum*): The potential for metabolomics. Metabolomics, 3(4), 399-412. https://doi.org/10.1007/s11306-007-0074-2

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics, 23, 2633–2635. https://doi.org/10.1093/bioinformatics/btm308

Braga, P. C., Lo Scalzo, R., dal Sasso, M., Lattuada, N., Greco, V., & Fibiani, M. (2016). Characterization and antioxidant activity of semi-purified extracts and pure delphinine-glycosides from eggplant peel (*Solanum melongena* L.) and allied species. Journal of Functional Foods, 20, 411–421. https://doi.org/10.1016/j.jff.2015.10.032

Browning, B. L., & Browning, S. R. (2016). Genotype imputation with millions of reference samples. American Journal of Human Genetics, 98(1), 116-126. https://doi.org/10.1016/j.ajhg.2015.11.020

Brzyski, D., Peterson, C. B., Sobczyk, P., Candès, E. J., Bogdan, M., & Sabatti, C. (2017). Controlling the rate of GWAS false discoveries. Genetics, 205, 61–75. https://doi.org/10.1534/genetics.116.193987

Bush, W. S., & Moore, J. H. (2012). Genome-wide association studies. PLoS Computational Biology, 8(12), e1002822. https://doi.org/10.1371/journal.pcbi.1002822

Cao, G., Sofic, E., & Prior, R. L. (1996). Antioxidant capacity of tea and common vegetables. Journal of Agricultural and Food Chemistry, 44, 3426–3431. https://doi.org/10.1021/jf9602535

Chaim, A. B. C., Borovsky, Y. B., De Jong, W. D. J., & Paran, I. P. (2003). Linkage of the A locus for the presence of anthocyanin and fs10.1, a major fruit-shape QTL in pepper. Theoretical and Applied Genetics, 106(6), 889-894. https://doi.org/10.1007/s00122-002-1132-9

Chapman, M. A. (2019). Introduction: The Importance of Eggplant. In M. Chapman (Ed.), The Eggplant Genome (Compendium of Plant Genomes). Springer. https://doi.org/10.1007/978-3-319-99208-2_1

Chaves-Silva, S., Santos, A. L. D., Chalfun-Júnior, A., Zhao, J., Peres, L. E. P., & Benedito, V. A. (2018). Understanding the genetic regulation of anthocyanin biosynthesis in plants – tools for breeding purple varieties of fruits and vegetables. Phytochemistry, 153, 11–27. https://doi.org/10.1016/j.phytochem.2018.05.005

Chen, E., Huang, X., Tian, Z., Wing, R. A., & Han, B. (2019). The genomics of *Oryza* species provides insights into rice domestication and heterosis. Annual Review of Plant Biology, 70, 639–665. https://doi.org/10.1146/annurev-arplant-050718-100202

Chen, W., Wang, W., Peng, M., Gong, L., Gao, Y., Wan, J., Wang, S., Shi, L., Zhou, B., Li, Z., Peng, X., Yang, C., Qu, L., Liu, X., & Luo, J. (2016). Comparative and parallel genome-wide association studies for metabolic and agronomic traits in cereals. Nature Communications, 7, 12767. https://doi.org/10.1038/ncomms12767

Chua, C. S., Biermann, D., Goo, K. S., & Sim, T. S. (2008). Elucidation of active site residues of *Arabidopsis thaliana* flavonol synthase provides a molecular platform for engineering flavonols. Phytochemistry, 69(1), 66–75. https://doi.org/10.1016/j.phytochem.2007.07.006

Collard, B. C. Y., Jahufer, M. Z. Z., Brouwer, J. B., & 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.

Contreras-Soto, R. I., Mora, F., de Oliveira, M. A. R., Higashi, W., Scapim, C. A., & Schuster, I. (2017). A genome-wide association study for agronomic traits in soybean using SNP markers and SNP based haplotype analysis. PLoS One, 12(2), e0171105. https://doi.org/10.1371/journal.pone.0171105

D'Amelia, V., Aversano, R., Chiaiese, P., & Carputo, D. (2018). The antioxidant properties of plant flavonoids: Their exploitation by molecular plant breeding. Phytochemical Reviews, 17, 611–625. https://doi.org/10.1007/s11101-018-9568-y

Daunay, M. C., & Aubert, S. (2004). Eggplant (*Solanum melongena*) fruit color: Pigments, measurements and genetics. In Proceedings of the 12th Eucarpia Meeting on Genetics and Breeding of Capsicum and Eggplant (pp. 108–116). Noordwijkerhout, The Netherlands.

Daunay, M. C., & Hazra, P. (2012). Eggplant. In K. V. Peter & P. Hazra (Eds.), Handbook of Vegetables (pp. 257–322). Studium Press.

Daunay, M. C., Lester, R. N., & Hazra, P. (1988). The usefulness of taxonomy for Solanaceae breeders, with special reference to the genus *Solanum* and to *Solanum melongena* L. (eggplant). Capsicum Newsletter, 7, 70–79.

Davies, K. M., Schwinn, K. E., Deroles, S. C., Manson, D. G., Lewis, D. H., Bloor, S. J., & Bradley, J. M. (2003). Enhancing anthocyanin production by altering competition for substrate between flavonol synthase and dihydroflavonol 4-reductase. Euphytica, 131(3), 259–268. https://doi.org/10.1023/A:1023990104158

De Jong, W. S., Eannetta, N. T., Jong, D. M., & Bodis, M. (2004). Candidate gene analysis of anthocyanin pigmentation loci in the Solanaceae. Theoretical and Applied Genetics, 108(3), 423-432. https://doi.org/10.1007/s00122-003-1455-1

Deluc, L., Bogs, J., Walker, A. R., Ferrier, T., Decendit, A., Merillon, J.-M., Robinson, S. P., & Barrieu, F. (2008). The transcription factor VvMYB5b contributes to the regulation of anthocyanin and proanthocyanidin biosynthesis in developing grape berries. Plant Physiology, 147(4), 2041-2053. https://doi.org/10.1104/pp.108.118919

Devlin, B., & Roeder, K. (1999). Genomic control for association studies. Biometrics, 55(4), 997-1004. https://doi.org/10.1111/j.0006-341X.1999.00997.x

Díez, M. J., Picó, B., & Nuez, F. (2002). Cucurbit genetic resources in Europe: Ad hoc meeting held in Adana, Turkey, 19 January 2002. Rome: International Plant Genetic Resources Institute.

Docimo, T., Francese, G., Ruggiero, A., Batelli, G., De Palma, M., Bassolino, L., Toppino, L., Rotino, G. L., Mennella, G., & Tucci, M. (2016). Phenylpropanoids accumulation in eggplant fruit: Characterization of biosynthetic genes and regulation by a MYB transcription factor. Frontiers in Plant Science, 6, 1233. https://doi.org/10.3389/fpls.2015.01233

Doganlar, S., Frary, A., Daunay, M. C., Lester, R. N., & Tanksley, S. D. (2002). A comparative genetic linkage map of eggplant (*Solanum melongena*) and its implications for genome evolution in the Solanaceae. Genetics, 161(4), 1697–1711. https://doi.org/10.1093/genetics/161.4.1697

Duan, Z., Tian, S., Yang, G., Wei, M., Li, J., & Yang, F. (2021). The basic helix-loop-helix transcription factor SmbHLH1 represses anthocyanin biosynthesis in eggplant. Frontiers in Plant Science, 12, 757936. https://doi.org/10.3389/fpls.2021.757936

Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C., & Lepiniec, L. (2010). MYB transcription factors in *Arabidopsis*. Trends in Plant Science, 15, 573–581. https://doi.org/10.1016/j.tplants.2010.06.005

[FAO] Food and Agriculture Organization. (2022). FAOSTAT production database. Retrieved from http://www.faostat.fao.org

Elomaa, P., Mehto, M., Kotilainen, M., Helariutta, Y., Nevalainen, L., & Teeri, T. H. (1998). A bHLH transcription factor mediates organ, region, and flower type-specific signals on dihydroflavonol-4-reductase (*dfr*) gene expression in the inflorescence of *Gerbera hybrida* (Asteraceae). The Plant Journal, 16(1), 93-99. https://doi.org/10.1046/j.1365-313x.1998.00269.x

Ertiro, B. T., Labuschagne, M., Olsen, M., Das, B., Prasanna, B. M., & Gowda, M. (2020). Genetic dissection of nitrogen use efficiency in tropical maize through genome-wide association and genomic prediction. Frontiers in Plant Science, 11, 474. https://doi.org/10.3389/fpls.2020.00474

Flores-Saavedra, M., Plazas, M., Vilanova, S., Prohens, J., & Gramazio, P. (2023). Induction of water stress in major *Solanum* crops: A review on methodologies and their application for identifying drought tolerant materials. Scientia Horticulturae, 318, 112105. https://doi.org/10.1016/j.scienta.2021.112105

Frary, A., Doganlar, S., & Daunay, M. C. (2007). Eggplant. In C. Kole (Ed.), Vegetables. Genome Mapping and Molecular Breeding in Plants (Vol. 5, pp. 179-213). Springer. https://doi.org/10.1007/978-3-540-34536-7_9

Frodin, D. G. (2004). History and concepts of big plant genera. Taxon, 53(3), 753–776. https://doi.org/10.2307/4135449

García-Fortea, E., Gramazio, P., Vilanova, S., Fita, A., Mangino, G., Villanueva, G., Arrones, A., Knapp, S., Prohens, J., & Plazas, M. (2019). First successful backcrossing towards eggplant (*Solanum melongena*) of a New World species, the silverleaf nightshade (*S. elaeagnifolium*), and characterization of interspecific hybrids and backcrosses. Scientia Horticulturae, 246, 563-573. https://doi.org/10.1016/j.scienta.2018.11.033

Garrison, E. P., & Marth, G. T. (2012). Haplotype-based variant detection from short-read sequencing. arXiv: Genomics. https://doi.org/10.48550/arXiv.1207.3907

Gisbert, C., Dumm, J. M., Prohens, J., Vilanova, S., & Stommel, J. R. (2016). A spontaneous eggplant (Solanum *melongena* L.) color mutant conditions anthocyanin-free fruit pigmentation. HortScience, 51(7), 793–798. https://doi.org/10.21273/HORTSCI10573-16

Gonzalez, A., Zhao, M., Leavitt, J. M., & Lloyd, A. M. (2008). Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/MYB transcriptional complex in *Arabidopsis* seedlings. The Plant Journal, 53(5), 814–827. https://doi.org/10.1111/j.1365-313X.2007.03369.x

Gonzali, S., Mazzucato, A., & Perata, P. (2009). Purple as a tomato: Towards high anthocyanin tomatoes. Trends in Plant Science, 14, 237–241. https://doi.org/10.1016/j.tplants.2009.02.001

Gould, K. S. (2003). Anthocyanins in leaves: Light attenuators or antioxidants? Functional Plant Biology, 30, 865–873. https://doi.org/10.1071/FP03118

Gramazio, P., Pereira-Dias, L., Vilanova, S., Prohens, J., Soler, S., Esteras, J., Garmendia, A., & Díez, M. J. (2020). Morphoagronomic characterization and whole-genome resequencing of eight highly diverse wild and weedy *S. pimpinellifolium* and *S. lycopersicum* var. *cerasiforme* accessions used for the first interspecific tomato MAGIC population. Horticulture Research, 7(1), 174. https://doi.org/10.1038/s41438-020-00395-w.

Gramazio, P., Prohens, J., Plazas, M., Mangino, G., Herraiz, F. J., & Vilanova, S. (2017). Development and genetic characterization of advanced backcross materials and an introgression line population of *Solanum incanum* in a *S. melongena* background. Frontiers in Plant Science, 8, Article 1477. https://doi.org/10.3389/fpls.2017.01477

Gramazio, P., Yan, H., Hasing, T., Vilanova, S., Prohens, J., & Bombarely, A. (2019). Whole-genome resequencing of seven eggplant (*Solanum melongena*) and one wild relative (*S. incanum*) accessions provides new insights and breeding tools for eggplant enhancement. Frontiers in Plant Science, 10, Article 1220. https://doi.org/10.3389/fpls.2019.01220

Guo, J., Han, W., & Wang, M. H. (2008). Ultraviolet and environmental stresses involved in the induction and regulation of anthocyanin biosynthesis: A review. African Journal of Biotechnology, 7, 4966–4972.

Gupta, P. K., Kulwal, P. L., & Jaiswal, V. (2014). Association mapping in crop plants: Opportunities and challenges. Advances in Genetics, 85, 109–147.

Gupta, P. K., Kulwal, P. L., & Jaiswal, V. (2019). Association mapping in plants in the post-GWAS genomics era. Advances in Genetics, 104, 75–154. https://doi.org/10.1016/bs.adgen.2018.12.001

Gupta, P. K., Kulwal, P. L., & Mir, R. R. (2013). QTL mapping: Methodology and applications in cereal breeding. In P. K. Gupta & R. K. Varshney (Eds.), Cereal Genomics II (pp. 275-318). Springer.

Gupta, P. K., Rustgi, S., & Kulwal, P. L. (2005). Linkage disequilibrium and association studies in higher plants: Present status and future prospects. Plant Molecular Biology, 57, 461–485.

Gyawali, S., Chao, S., Vaish, S. S., Singh, S. P., Rehman, S., Vishwakarma, S. R., & Verma, R. P. S. (2018). Genome wide association studies (GWAS) of spot blotch resistance at the seedling and the adult plant stages in a collection of spring barley. Molecular Breeding, 38, 62. https://doi.org/10.1007/s11032-018-0826-1

Haghi, R., Ahmadikhah, A., Fazeli, A., & Shariati, V. (2022). Candidate genes for anthocyanin pigmentation in rice stem revealed by GWAS and whole-genome resequencing. The Plant Genome, 15, e20224. https://doi.org/10.1002/tpg2.20224

Hamazaki, K., & Iwata, H. (2020). RAINBOW: Haplotype-based genome-wide association study using a novel SNP-set method. PLoS Computational Biology, 16(2), e1007663. https://doi.org/10.1371/journal.pcbi.1007663

Harlan, J. R., & de Wet, J. M. J. (1971). Toward a rational classification of cultivated plants. Taxon, 20, 509–517. https://doi.org/10.2307/1218252

Hichri, I., Heppel, S. C., Pillet, J., Léon, C., Czemmel, S., Delrot, S., Lauvergeat, V., & Bogs, J. (2010). The basic helix-loop-helix transcription factor MYC1 is involved in the regulation of the flavonoid biosynthesis pathway in grapevine. Molecular Plant, 3(3), 509-523. https://doi.org/10.1093/mp/ssp118

Holton, T. A., & Cornish, E. C. (1995). Genetics and biochemistry of anthocyanin biosynthesis. Plant Cell, 7, 1071–1083. https://doi.org/10.1105/tpc.7.7.1071

Huang, X., Feng, Q., Qian, Q., Zhao, Q., Wang, L., Wang, A., Guan, J., Fan, D., Weng, Q., Huang, T., Dong, G., Sang, T., & Han, B. (2009). High-throughput genotyping by whole-genome resequencing. Genome Research, 19(6), 1068–1076. https://doi.org/10.1101/gr.089516.108

Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., Li, C., Zhu, C., Lu, T., Zhang, Z., Li, M., Fan, D., Guo, Y., Wang, A., Wang, L., Deng, L., Li, W., Lu, Y., Weng, Q., Liu, K., Huang, T., Zhou, T., Jing, Y., Li, W., Lin, Z., Buckler, E. S., Qian, Q., Zhang, Q. F., Li, J., & Han, B. (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. Nature Genetics, 42(11), 961–967.

Hurtado, M., Vilanova, S., Plazas, M., Gramazio, P., Andújar, I., Herraiz, F. J., Castro, A., & Prohens, J. (2014). Enhancing conservation and use of local vegetable landraces: The Almagro eggplant (*Solanum melongena* L.) case study. Genetic Resources and Crop Evolution, 61, 787-795. https://doi.org/10.1007/s10722-013-0044-8

Ichiyanagi, T., Kashiwada, Y., Shida, Y., Ikeshiro, Y., Kaneyuki, T., & Konishi, T. (2005). Nasunin from eggplant consists of cis-trans isomers of delphinidin 3-[4-(p-coumaroyl)-L-rhamnosyl (1). Journal of Agricultural and Food Chemistry, 53(24), 9472–9477. https://doi.org/10.1021/jf051841y

Islam, M. S., Thyssen, G. N., Jenkins, J. N., Zeng, L., Delhom, C. D., McCarty, J. C., Deng, D. D., Hinchliffe, D. J., Jones, D. C., & Fang, D. D. (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), 903. https://doi.org/10.1186/s12864-016-3203-2

Jaganathan, D., Bohra, A., Thudi, M., & Varshney, R. K. (2020). Fine mapping and gene cloning in the post-NGS era: advances and prospects. Theoretical and Applied Genetics, 133, 1791-1810. https://doi.org/10.1007/s00122-020-03560-w

Jain, P., Dutta, B., & Sevanthi, A. M. (2024). Trait Based Association Mapping in Plants. In P. Anjoy, K. Kumar, G. Chandra, & K. Gaikwad (Eds.), Genomics Data Analysis for Crop Improvement. Springer Protocols Handbooks. https://doi.org/10.1007/978-981-99-6913-5_6

Jiao, Y., Jiang, Y., Zhai, W., & Yang, Z. (2012). Studies on antioxidant capacity of anthocyanin extract from purple sweet potato (*Ipomoea batatas* L.). African Journal of Biotechnology, 11, 7046–7054. https://doi.org/10.5897/AJB11.3859

Joukhadar, R., Hollaway, G., Shi, F., Kant, S., Forrest, K., Wong, D., Petkowski, J., Pasam, R., Tibbits, J., Bariana, H., Bansal, U., Spangenberg, G., Daetwyler, H., Gendall, T., & Hayden, M. (2020). Genome-wide association reveals a complex architecture for rust resistance in 2300 worldwide bread wheat accessions screened under various Australian conditions. Theoretical and Applied Genetics, 133(9), 2695–2712. https://doi.org/10.1007/s00122-020-03654-5

Jung, C. S., Griffiths, H. M., De Jong, D. M., Cheng, S., Bodis, M., Kim, T. S., & De Jong, W. S. (2009). The potato developer (D) locus encodes an R2R3 MYB transcription factor that regulates expression of multiple anthocyanin structural genes in tuber skin. Theoretical and Applied Genetics, 120(1), 45–57. https://doi.org/10.1007/s00122-009-1158-3

Kang, H. M., Zaitlen, N. A., Wade, C. M., Kirby, A., Heckerman, D., Daly, M. J., & Eskin, E. (2008). Efficient control of population structure in model organism association mapping. Genetics, 178(3), 1709–1723. https://doi.org/10.1534/genetics.107.080101

Kaushik, P., Plazas, M., Prohens, J., Vilanova, S., & Gramazio, P. (2018). Diallel genetic analysis for multiple traits in eggplant and assessment of genetic distances for predicting hybrids performance. PLoS One, 13(6), e0199943. https://doi.org/10.1371/journal.pone.0199943

Kiferle, C., Fantini, E., Bassolino, L., Povero, G., Spelt, C., Buti, S., Quattrocchio, F., Koes, R., & Perata, P. (2015). Tomato R2R3-MYB proteins SlANT1 and SlAN2: Same protein activity, different roles. PLoS ONE, 10(8), e0136365. https://doi.org/10.1371/journal.pone.0136365

Kim, S., Zhao, K., Jiang, R., Molitor, J., Borevitz, J. O., Nordborg, M., & Marjoram, P. (2006). Association mapping with single-feature polymorphisms. Genetics, 173, 1125–1133. https://doi.org/10.1534/genetics.105.054593

Knapp, S., Aubriot, X., & Prohens, J. (2019). Eggplant (*Solanum melongena* L.): Taxonomy and relationships. In M. A. Chapman (Ed.), The Eggplant Genome. Compendium of Plant Genomes. Springer. https://doi.org/10.1007/978-3-319-99208-2_2

Knapp, S., Vorontsova, M. S., & 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), e57039. https://doi.org/10.1371/journal.pone.0057039

Kobayashi, S., Ishimaru, M., Hiraoka, K., & Honda, C. (2002). *Myb*-related genes of the Kyoho grape (*Vitis labruscana*) regulate anthocyanin biosynthesis. Planta, 215(6), 924-933. https://doi.org/10.1007/s00425-002-0817-9

Kong, J. M., Chia, L. S., Goh, N. K., Chia, T. F., & Brouillard, R. (2003). Analysis and biological activities of anthocyanins. Phytochemistry, 64, 923–933. https://doi.org/10.1016/S0031-9422(03)00438-2

Kouassi, B., Prohens, J., Gramazio, P., Kouassi, A. B., Vilanova, S., Galán-Ávila, A., Herraiz, F. J., Seguí-Simarro, J. M., & Plazas, M. (2016). Development of backcross generations and new interspecific hybrid combinations for introgression breeding in eggplant (*Solanum melongena*). Scientia Horticulturae, 213, 199–207. https://doi.org/10.1016/j.scienta.2016.10.039

Kulwal, P. L., & Singh, R. (2021). Association Mapping in Plants. In P. Tripodi (Ed.), Crop Breeding. Methods in Molecular Biology, vol. 2264. Humana. https://doi.org/10.1007/978-1-0716-1201-9_8

Kuroda, C., & Wada, M. (1933). The colouring matter of eggplant (Nasu). Proceedings of the Imperial Academy (Japan), 9, 51–52.

Lester, R. N. (1998). Genetic resources of Capsicum and eggplant. In Proceedings of Eucarpia meeting on genetics and breeding of Capsicum and eggplant (pp. 25–30). Avignon, France.

Lester, R. N., & Hasan, S. M. (1991). Origin and domestication of the brinjal eggplant, *Solanum melongena*, from *S. incanum*, in Africa and Asia. In J. G. Hawkes, R. N. Lester, M. Nee, & N. Estrada (Eds.), Solanaceae III: Taxonomy, chemistry, evolution (pp. 369–387). Kew: Royal Botanic Gardens.

Li, C. H., Qiu, F., Ding, L., Huang, M. Z., Huang, S. R., Yang, G. S., & Wang, Y. (2017). Anthocyanin biosynthesis regulation of *DhMYB2* and *DhbHLH1* in *Dendrobium* hybrids petals. Plant Physiology and Biochemistry, 112, 335-345. https://doi.org/10.1016/j.plaphy.2016.09.018

Li, F., Wen, W., Liu, J., Zhang, Y., Cao, S., He, Z., Rasheed, A., Jin, H., Zhang, C., Yan, J., Zhang, P., Wan, Y., & Xia, X. (2019). Genetic architecture of grain yield in bread wheat based on genome-wide association studies. BMC Plant Biology, 19, 168. https://doi.org/10.1186/s12870-019-1750-4

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv: Genomics. https://doi.org/10.48550/arXiv.1303.3997

Li, J., He, Y. J., Zhou, L., Liu, Y., Jiang, M., Ren, L., & 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), 201. https://doi.org/10.1186/s12864-018-4587-z

Li, P., & Cheng, L. (2008). The shaded side of apple fruit becomes more sensitive to photoinhibition with fruit development. Physiologia Plantarum, 134, 282–292. https://doi.org/10.1111/j.1399-3054.2008.01131.x

Lim, S. H., Song, J. H., Kim, D. H., Kim, J. K., Lee, J. Y., Kim, Y. M., & Ha, S. H. (2016). Activation of anthocyanin biosynthesis by expression of the radish R2R3-MYB transcription factor gene *RsMYB1*. Plant Cell Reports, 35(3), 641–653. https://doi.org/10.1007/s00299-015-1909-3

Liu, X., Huang, M., Fan, B., Buckler, E. S., & Zhang, Z. (2016a). Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. PLoS Genetics, 12(2), e1005767. https://doi.org/10.1371/journal.pgen.1005767

Liu, L. Y., Lin-Wang, K., Deng, C., Warran, B., Wang, L., Yu, B., Yang, H., Wang, J., Espley, R. V., Zhang, J., Wang, D., & Allan, A. C. (2015). Comparative transcriptome analysis of white and purple potato to identify genes involved in anthocyanin biosynthesis. PLoS ONE, 10(6), e0129148. https://doi.org/10.1371/journal.pone.0129148

Liu, Y., Lin-Wang, K., Espley, R. V., Wang, L., Yang, H., Yu, B., Dare, A., Varkonyi-Gasic, E., Wang, J., & Zhang, J. (2016b). Functional diversification of the potato R2R3 MYB anthocyanin activators AN1, MYBA1, and MYB113 and their interaction with basic helix-loop-helix cofactors. Journal of Experimental Botany, 67(8), 2159-2176. https://doi.org/10.1093/jxb/erv552

Liu, Y., Tikunov, Y., Schouten, R. E., Marcelis, L. F. M., Visser, R. G. F., & Bovy, A. (2018). Anthocyanin biosynthesis and degradation mechanisms in Solanaceous vegetables: A review. Frontiers in Chemistry, 6, 52. https://doi.org/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. https://doi.org/10.1139/gen-2019-0021

Mace, E., Lester, R. N., & Gebhardt, C. (1999). AFLP analysis of genetic relationships among the cultivated eggplant, *Solanum melongena* L., and wild relatives (Solanaceae). Theoretical and Applied Genetics, 99(4), 626–633.

Malmberg, M. M., Barbulescu, D. M., Drayton, M. C., Shinozuka, M., Thakur, P., Ogaji, Y. O., Spangenberg, G. C., Daetwyler, H. D., & Cogan, N. O. I. (2018). Evaluation and recommendations for routine genotyping using skim whole genome re-sequencing in canola. Frontiers in Plant Science, 9, 1809. https://doi.org/10.3389/fpls.2018.01809

Malone, L. A., Barraclough, E. I., Lin-Wang, K., Stevenson, D. E., & Allan, A. C. (2009). Effects of red-leaved transgenic tobacco expressing a MYB transcription factor on two herbivorous insects, *Spodoptera litura* and *Helicoverpa armigera*. Entomologia Experimentalis et Applicata, 133(2), 117–127. https://doi.org/10.1111/j.1570-7458.2009.00910.x

Mammadov, J., Aggarwal, R., Buyyarapu, R., & Kumpatla, S. (2012). SNP markers and their impact on plant breeding. International Journal of Plant Genomics, 2012, 728398. https://doi.org/10.1155/2012/728398

Mangino, G., Arrones, A., Plazas, M., Pook, T., Prohens, J., Gramazio, P., & Vilanova, S. (2022). Newly developed MAGIC population allows identification of strong associations and candidate genes for anthocyanin pigmentation in eggplant. Frontiers in Plant Science, 13, 847789. https://doi.org/10.3389/fpls.2022.847789

Marchini, J., & Howie, B. (2010). Genotype imputation for genome-wide association studies. Nature Reviews Genetics, 11, 499–511.

Martin, F. W., & Rhodes, A. M. (1979). Subspecific grouping of eggplant cultivars. Euphytica, 28(3), 367–383.

Matsubara, K., Kaneyuki, T., Miyake, T., & Mori, M. (2005). Antiangiogenic activity of nasunin, an antioxidant anthocyanin, in eggplant peels. Journal of Agricultural and Food Chemistry, 53, 6272–6275. https://doi.org/10.1021/jf050796r

Mennella, G., Lo Scalzo, R., Fibiani, M., D'Alessandro, A., Francese, G., Toppino, L., Acciarri, N., de Almeida, A. E., & Rotino, G. L. (2012). Chemical and bioactive quality traits during fruit ripening in eggplant (*Solanum melongena* L.) and allied species. Journal of Agricultural and Food Chemistry, 60(47), 11821–11831. https://doi.org/10.1021/jf3037424

Meyer, R. S., Karol, K. G., Little, D. P., Nee, M. H., & Litt, A. (2012). Phylogeographic relationships among Asian eggplants and new perspectives on eggplant domestication. Molecular Phylogenetics and Evolution, 63(3), 685–701. https://doi.org/10.1016/j.ympev.2012.02.006

Meyer, R. S., Whitaker, B. D., Little, D. P., Wu, S.-B., Kennelly, E. J., Long, C.-L., & Litt, A. (2015). Parallel reductions in phenolic constituents resulting from the domestication of eggplant. Phytochemistry, 115, 194–206. https://doi.org/10.1016/j.phytochem.2015.02.006

Montefiori, M., Brendolise, C., Dare, A. P., Lin-Wang, K., Davies, K. M., Hellens, R. P., & Allan, A. C. (2015). In the Solanaceae, a hierarchy of bHLHs confer distinct target specificity to the anthocyanin regulatory complex. Journal of Experimental Botany, 66(5), 1427–1436. https://doi.org/10.1093/jxb/eru494

Nakatsuka, T., Haruta, K. S., Pitaksutheepong, C., Abe, Y., Kakizaki, Y., Yamamoto, K., & Nishihara, M. (2008). Identification and characterization of R2R3-MYB and bHLH transcription factors regulating anthocyanin biosynthesis in gentian flowers. Plant and Cell Physiology, 49(12), 1818-1829. https://doi.org/10.1093/pcp/pcn164

Niu, S. S., Xu, C. J., Zhang, W. S., Zhang, B., Li, X., Lin-Wang, K., Ferguson, I. B., Allan, A. C., & Chen, K. S. (2010). Coordinated regulation of anthocyanin biosynthesis in Chinese bayberry (*Myrica rubra*) fruit by a R2R3 MYB transcription factor. Planta, 231(4), 887-899. https://doi.org/10.1007/s00425-009-1096-3

Owens, B. F., Lipka, A. E., Magallanes-Lundback, M., Tiede, T., Diepenbrock, C. H., Kandianis, C. B., Kim, E., Cepela, J., Mateos-Hernandez, M., Buell, C. R., Buckler, E. S., DellaPenna, D., Gore, M. A., & Rocheford, T. (2014). A foundation for provitamin A biofortification of maize: Genome-wide association and genomic prediction models of carotenoid levels. Genetics, 198(4), 1699–1716.

Page, A., Gibson, J., Meyer, R. S., & Chapman, M. A. (2019). Eggplant domestication: Pervasive gene flow, feralization, and transcriptomic divergence. Molecular Biology and Evolution, 36(7), 1359–1372. https://doi.org/10.1093/molbev/msz062

Pang, Y., Wu, Y., Liu, C., Li, W., St. Amand, P., Bernardo, A., Wang, D., Dong, L., Yuan, X., Zhang, H., Zhao, M., Li, L., Wang, L., He, F., Liang, Y., Yan, Q., Lu, Y., Su, Y., Jiang, H., Wu, J., Li, A., Kong, L., Bai, G., & Liu, S. (2021). High-resolution genome-wide association study and genomic prediction for disease resistance and cold tolerance in wheat. Theoretical and Applied Genetics, 134, 2857–2873.

Pasam, R. K., & Sharma, R. (2014). Association Mapping: A New Paradigm for Dissection of Complex Traits in Crops. In P. B., K., Bandopadhyay, R., Suravajhala, P. (Eds.), Agricultural Bioinformatics. Springer. https://doi.org/10.1007/978-81-322-1880-7_1

Pasam, R., Sharma, R., Malosetti, M., van Eeuwijk, F., Haseneyer, G., Kilian, B., & Graner, A. (2012). Genome-wide association studies for agronomical traits in a world-wide spring barley collection. BMC Plant Biology, 12(1), 16. https://doi.org/10.1186/1471-2229-12-16

Plazas, M., Andújar, I., Vilanova, S., Hurtado, M., Gramazio, P., Herraiz, F. J., & Prohens, J. (2013). Breeding for chlorogenic acid content in eggplant: Interest and prospects. Notulae Botanicae Horti Agrobotanici Cluj-Napoca, 41(1), 26–35. https://doi.org/10.15835/nbha4119036

Plazas, M., Prohens, J., Cuñat AN, Vilanova, S., Gramazio, P., Herraiz, F. J., Andújar, I. (2014). Reducing capacity, chlorogenic acid content and biological activity in a collection of scarlet (*Solanum aethiopicum*) and Gboma (*S. macrocarpon*) eggplants. International Journal of Molecular Sciences, 15(10), 17221-17241. https://doi.org/10.3390/ijms151017221

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

Pootakham, W. (2023). Genotyping by Sequencing (GBS) for Genome-Wide SNP Identification in Plants. In Y. Shavrukov (Ed.), Plant Genotyping. Springer. https://doi.org/10.1007/978-1-0716-3024-2_1

Povero, G., Gonzali, S., Bassolino, L., Mazzucato, A., & Perata, P. (2011). Transcriptional analysis in high-anthocyanin tomatoes reveals synergistic effect of *Aft* and *atv* genes. Journal of Plant Physiology, 168(3), 270–279. https://doi.org/10.1016/j.jplph.2010.07.022

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. Genetics, 155(2), 945-959. https://doi.org/10.1093/genetics/155.2.945

Prohens, J., Blanca, J. M., & Nuez, F. (2005). Morphological and molecular variation in a collection of eggplant from a secondary center of diversity: Implications for conservation and breeding. Journal of the American Society for Horticultural Science, 130(1), 54–63.

Prohens, J., Whitaker, B. D., Plazas, M., Vilanova, S., Hurtado, M., Blasco, M., Gramazio, P., & Stommel, J. R. (2013). Genetic diversity in morphological characters and phenolic acids content resulting from an interspecific cross between eggplant, *Solanum melongena*, and its wild ancestor (*S. incanum*). Annals of Applied Biology, 162, 242–257. https://doi.org/10.1111/aab.12017

Raigón, M. D., Prohens, J., Muñoz-Falcón, J. E., & Nuez, F. (2008). Comparison of eggplant landraces and commercial varieties for fruit content of phenolics, minerals, dry matter and protein. Journal of Food Composition and Analysis, 21, 370–376. https://doi.org/10.1016/j.jfca.2008.03.006

Ranil, R. H. G., Niran, H. M. L., Plazas, M., Fonseka, R. M., Fonseka, H. H., Vilanova, S., Andújar, I., Gramazio, P., Fita, A., & Prohens, J. (2015). Improving seed germination of the eggplant rootstock *Solanum torvum* by testing multiple factors using an orthogonal array design. Scientia Horticulturae, 193, 174-181. https://doi.org/10.1016/j.scienta.2015.06.003

Ranil, R. H. G., Prohens, J., Aubriot, X., Niran, H. M. L., Plazas, M., Fonseka, R. M., Vilanova, S., Fonseka, H. H., Gramazio, P., & Knapp, S. (2017). *Solanum insanum* L. (subgenus Leptostemonum Bitter, Solanaceae), the neglected wild progenitor of eggplant (*S. melongena* L.): A review of taxonomy, characteristics and uses aimed at its enhancement for improved eggplant breeding. Genetic Resources and Crop Evolution. https://doi.org/10.1007/s10722-016-0467-z

Rasheed, A., Hao, Y., Xia, X., Khan, A., Xu, Y., Varshney, R. K., & He, Z. (2017). Crop Breeding Chips and Genotyping Platforms: Progress, Challenges, and Perspectives. Molecular Plant, 10(8), 1047-1064. https://doi.org/10.1016/j.molp.2017.06.008

Ravelombola, W., Qin, J., Shi, A., Song, Q., Yuan, J., Wang, F., Chen, P., Yan, L., Feng, Y., Zhao, T., Meng, Y., Guan, K., Yang, C., & Zhang, M. (2021). Genome-wide association study and genomic selection for yield and related traits in soybean. PLOS ONE, 16(8), e0255761.

Rieseberg, L. H., & Blackman, B. K. (2010). Speciation genes in plants. Annals of Botany, 106(3), 439–455. https://doi.org/10.1093/aob/mcq126

Rotino, G. L., Sala, T., & Toppino, L. (2014). Eggplant. In A. Pratap & J. Kumar (Eds.), Alien Gene Transfer in Crop Plants, Vol. 2 (pp. 381–409). Springer. https://doi.org/10.1007/978-1-4614-9572-7_16

Rowan, B. A., Patel, V., Weigel, D., & Schneeberger, K. (2015). Rapid and inexpensive whole-genome genotyping-by-sequencing for crossover localization and fine-scale genetic mapping. G3, 5, 385–398. https://doi.org/10.1534/g3.114.016501

Sahu, T. K., Singh, M., Kalia, S., & Singh, A. K. (2023). Genome-Wide Association Study (GWAS): Concept and Methodology for Gene Mapping in Plants. In A. Raina, M. R. Wani, R. A. Laskar, N. Tomlekova, & S. Khan (Eds.), Advanced Crop Improvement, Volume 2. Springer. https://doi.org/10.1007/978-3-031-26669-0_17

Sakamura, S., Watanabe, S., & Obata, Y. (1963). The structure of the major anthocyanin in eggplant. Agricultural and Biological Chemistry, 27, 663–665.

Sapir, M., Oren-Shamir, M., Ovadia, R., Reuveni, M., Evenor, D., Tadmor, Y., Nahon, S., Shlomo, H., Chen, L., Meir, A., & Levin, I. (2008). Molecular aspects of anthocyanin fruit tomato in relation to high pigment-1. Journal of Heredity, 99(3), 292–303. https://doi.org/10.1093/jhered/esm128

Scott, M. F., Ladejobi, O., Amer, S., Bentley, A. R., Biernaskie, J., Boden, S. A., Clark, M., Dell'Acqua, M., Dixon, L. E., Filippi, C. V., Fradgley, N., Gardner, K. A., Mackay, I. J., O'Sullivan, D., Percival-Alwyn, L., Roorkiwal, M., Singh, R. K., Thudi, M., Varshney, R. K., Venturini, L., Whan, A., Cockram, J., & Mott, R. (2020). Multi-parent populations in crops: a toolbox integrating genomics and genetic mapping with breeding. Heredity (Edinb), 125(6), 396-416. https://doi.org/10.1038/s41437-020-0336-6.

Seo, J., Lee, G., Jin, Z., Kim, B., Chin, J. H., & Koh, H.-J. (2020). Development and application of indica–japonica SNP assays using the Fluidigm platform for rice genetic analysis and molecular breeding. Molecular Breeding, 40, 39. https://doi.org/10.1007/s11032-020-01123-x

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

Singh, B. D., & Singh, A. K. (2015). Marker-Assisted Plant Breeding: Principles and Practices. Springer New Delhi.

Solberg, S., van Zonneveld, M., Rakha, M. T., Taher, D. I., Prohens, J., Jarret, R., van Dooijeweert, W., & Giovannini, P. (2022). Global strategy for the conservation and use of eggplants. Global Crop Diversity Trust. Bonn, Germany.

Steyn, W. J., Wand, S. J. E., Holcroft, D. M., & Jacobs, G. (2002). Anthocyanins in vegetative tissues: A proposed unified function in photoprotection. New Phytologist, 155, 349–361. https://doi.org/10.1046/j.1469-8137.2002.00482.x

Stich, B., & Melchinger, A. E. (2009). Comparison of mixed-model approaches for association mapping in rapeseed, potato, sugar beet, maize, and Arabidopsis. BMC Genomics, 10, 94. https://doi.org/10.1186/1471-2164-10-94

Stommel, J. R., & Dumm, J. M. (2015). Coordinated regulation of biosynthetic and regulatory genes coincides with anthocyanin accumulation in developing eggplant fruit. Journal of the American Society for Horticultural Science, 140(2), 129–135. https://doi.org/10.21273/JASHS.140.2.129

Stommel, J. R., Lightbourn, G. J., Winkel, B. S., & Griesbach, R. J. (2009). Transcription factor families regulate the anthocyanin biosynthetic pathway in *Capsicum annuum*. Journal of the American Society for Horticultural Science, 134(3), 244-251.

Stommel, J. R., Whitaker, B. D., Haynes, K. G., & Prohens, J. (2015). Genotype × environment interactions in eggplant for fruit phenolic acid content. Euphytica, 205, 823–836. https://doi.org/10.1007/s10681-015-1415-2

Storey, J. D., & Tibshirani, R. (2003). Statistical significance for genome wide studies. Proceedings of the National Academy of Sciences, 100, 9440–9445. https://doi.org/10.1073/pnas.1530509100

Stracke, R., Werber, M., & Weisshaar, B. (2001). The R2R3-MYB gene family in *Arabidopsis thaliana*. Current Opinion in Plant Biology, 4(5), 447-456. https://doi.org/10.1016/S1369-5266(00)00199-0

Syfert, M. M., Castañeda-Álvarez, N. P., Khoury, C. K., Särkinen, T., Sosa, C. C., Achicanoy, H. A., Bernau, V., Prohens, J., Daunay, M. C., & Knapp, S. (2016). Crop wild relatives of the brinjal eggplant (*Solanum melongena*): Poorly represented in genebanks and many species at risk of extinction. American Journal of Botany, 103(4), 635–651. https://doi.org/10.3732/ajb.1500539

Taher, D., Solberg, S.Ø., Prohens, J., Chou, Y., Rakha, M., & Wu, T. (2017). World Vegetable Center eggplant collection: Origin, composition, seed dissemination and utilization in breeding. Frontiers in Plant Science, 8, 1484. https://doi.org/10.3389/fpls.2017.01484

Takos, A. M., Jaffe, F. W., Jacob, S. R., Bogs, J., Robinson, S. P., & Walker, A. R. (2006). Light-induced expression of a *MYB* gene regulates anthocyanin biosynthesis in red apples. Plant Physiology, 142(4), 1216-1232. https://doi.org/10.1104/pp.106.088104

Tanaka, Y., & Ohmiya, A. (2008). Seeing is believing: Engineering anthocyanin and carotenoid biosynthetic pathways. Current Opinion in Biotechnology, 19, 190–197. https://doi.org/10.1016/j.copbio.2008.02.015

Tanaka, Y., Sasaki, N., & Ohmiya, A. (2008). Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids. Plant Journal, 54, 733–749. https://doi.org/10.1111/j.1365-313X.2008.03447.x

Terrier, N., Torregrosa, L., Ageorges, A., Vialet, S., Verries, C., Cheynier, V., & Romieu, C. (2009). Ectopic expression of *VvMybPA2* promotes proanthocyanidin biosynthesis in Vitis vinifera L. and suggests additional targets in the pathway. Plant Physiology, 149(2), 1028-1041. https://doi.org/10.1104/pp.108.132852

Tibbs-Cortes, L., Zhang, Z., & Yu, J. (2021). Status and prospects of genome-wide association studies in plants. The Plant Genome, 14(1), e20077.

Tsai, H.-Y., Janss, L. L., Andersen, J. R., Orabi, J., Jensen, J. D., Jahoor, A., & Jensen, J. (2020). Genomic prediction and GWAS of yield, quality and disease-related traits in spring barley and winter wheat. Scientific Reports, 10, 3347. https://doi.org/10.1038/s41598-020-60203-2

Turnbull, J. J., Nakajima, J., Welford, R. W., Yamazaki, M., Saito, K., & Schofield, C. J. (2004). Mechanistic studies on three 2-oxoglutarate-dependent oxygenases of flavonoid biosynthesis: anthocyanidin synthase, flavonol synthase, and flavanone 3beta-hydroxylase. Journal of Biological Chemistry, 279(2), 1206–1216. https://doi.org/10.1074/jbc.M309228200

Turner, S. D. (2018). qqman: An R package for visualizing GWAS results using Q-Q and Manhattan plots. Journal of Open Source Software, 3(31), 731. https://doi.org/10.21105/joss.00731

Valdar, W., Flint, J., & 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. https://doi.org/10.1534/genetics.104.039313

Vavilov, N. I. (1951). The origin, variation, immunity and breeding of cultivated plants. New York, NY: Chronica Botanica.

Vilanova, S., Alonso, D., Gramazio, P., Plazas, M., García-Fortea, E., Ferrante, P., Schmidt, M., Díez, M. J., Usadel, B., & Giuliano, G. (2020). SILEX: A fast and inexpensive high-quality DNA extraction method suitable for multiple sequencing platforms and recalcitrant plant species. Plant Methods, 16, Article 110. https://doi.org/10.1186/s13007-020-00652-y

Villanueva, G., Rosa-Martínez, E., Şahin, A., García-Fortea, E., Plazas, M., Prohens, J., & Vilanova, S. (2021). Evaluation of advanced backcrosses of eggplant with *Solanum elaeagnifolium* introgressions under low N conditions. Agronomy, 11(9), 1770. https://doi.org/10.3390/agronomy11091770

Vorontsova, M. S., & Knapp, S. (2012). A new species of *Solanum* (Solanaceae) from South Africa related to the cultivated eggplant. PhytoKeys, 8, 1–11. https://doi.org/10.3897/phytokeys.8.2462

Vorontsova, M. S., Stern, S., Bohs, L., & Knapp, S. (2013). African spiny *Solanum* subgenus *Leptostemonum* (Solanaceae): A thorny phylogenetic tangle. Botanical Journal of the Linnean Society, 173, 176–193. https://doi.org/10.1111/boj.12053

Walker, A. R., Lee, E., Bogs, J., McDavid, D. A. J., Thomas, M. R., & Robinson, S. P. (2007). White grapes arose through the mutation of two similar and adjacent regulatory genes. The Plant Journal, 49(5), 772-785. https://doi.org/10.1111/j.1365-313X.2006.02997.x

Wang, J., Yan, C., Li, Y., Li, C., Zhao, X., Yuan, C., Sun, Q., & Shan, S. (2019). GWAS discovery of candidate genes for yield-related traits in peanut and support from earlier QTL mapping studies. Genes (Basel), 10(10), 803.

Wang, H., Guo, X., Pandey, M. K., Ji, X., Varshney, R. K., Nwosu, V., & Guo, B. (2017). History and impact of the international peanut genome initiative: The exciting journey toward peanut whole-genome sequencing. In R. Varshney, M. Pandey, & N. Puppala (Eds.), The Peanut Genome (pp. 117-134). New York, NY: Springer. https://doi.org/10.1007/978-3-319-63935-2_8

Wang, Z., Wang, S., Wu, M., Yang, H., Qin, Y., Zhao, X., & Wang, Z. (2019). Evolutionary and functional analyses of the 2-oxoglutarate-dependent dioxygenase genes involved in the flavonoid biosynthesis pathway in tobacco. Planta, 249(2), 543–561. https://doi.org/10.1007/s00425-018-3019-2

Weese, T., & Bohs, L. (2010). Eggplant origins: Out of Africa, into the Orient. Taxon, 59, 49–56. https://doi.org/10.2307/27757050

Wellmann, F., Griesser, M., Schwab, W., Martens, S., Eisenreich, W., Matern, U., & Lukacin, R. (2006). Anthocyanidin synthase from *Gerbera hybrida* catalyzes the conversion of (+)-catechin to cyanidin and a novel procyanidin. FEBS Letters, 580(6), 1642–1648. https://doi.org/10.1016/j.febslet.2006.02.004

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

Winkel-Shirley, B. (2001). Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. Plant Physiology, 126, 485–493.

Wu, Y., Zhou, Z., Dong, C., Chen, J., Ding, J., Zhang, X., Mu, C., Chen, Y., Li, X., Li, H., Han, Y., Wang, R., Sun, X., Li, J., Dai, X., Song, W., Chen, W., & Wu, J. (2020). Linkage mapping and genome-wide association study reveals conservative QTL and candidate genes for Fusarium rot resistance in maize. BMC Genomics, 21(1), 357.

Xie, X. B., Zhao, J., Hao, Y. J., Fang, C. B., & Wang, Y. (2017). The ectopic expression of apple *MYB1* and *bHLH3* differentially activates anthocyanin biosynthesis in tobacco. Plant

Cell, Tissue and Organ Culture, 131(1), 183-194. https://doi.org/10.1007/s11240-017-1272-0

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

Yano, K., Yamamoto, E., Aya, K., Takeuchi, H., Lo, P. C., Hu, L., Yamasaki, M., Yoshida, S., Kitano, H., Hirano, K., & Matsuoka, M. (2016). Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. Nature Genetics, 48(8), 927–934.

Ye, J., Wang, X., Wang, W., Yu, H., Guo, A., Li, C., Sun, P., Wang, X., Li, H., Ouyang, B., Zhang, J., Zhang, Y., Han, H., Giovannoni, J. J., Fei, Z., & Ye, Z. (2021). Genome-wide association study reveals the genetic architecture of 27 agronomic traits in tomato. Plant Physiology, 186(4), 2078–2092.

Yi, G., Kim, J. S., Park, J. E., Shin, H., Yu, S. H., Park, S., & Huh, J. H. (2018). MYB1 transcription factor is a candidate responsible for red root skin in radish (*Raphanus sativus* L.). PLoS ONE, 13(9), e0204241. https://doi.org/10.1371/journal.pone.0204241

Zeven, A. C., & Zhukovsky, P. (1975). Dictionary of cultivated plants and their centres of diversity. Wageningen, The Netherlands: Centre for Agricultural Publishing and Documentation. Retrieved from https://edepot.wur.nl/318076

Zhang, L., Duan, Z., Ma, S., Sun, S., Sun, M., Xiao, Y., Ni, N., Irfan, M., Chen, L., & Sun, Y. (2023a). SlMYB7, an AtMYB4-like R2R3-MYB transcription factor, inhibits anthocyanin accumulation in *Solanum lycopersicum* fruits. Journal of Agricultural and Food Chemistry, 71(6), 18758-18768. https://doi.org/10.1021/acs.jafc.3c00547

Zhang, R., Jia, G., & Diao, X. (2023b). geneHapR: An R package for gene haplotypic statistics and visualization. BMC Bioinformatics, 24(1), 199. https://doi.org/10.1186/s12859-023-05318-9

Zhang, Y., Butelli, E., De Stefano, R., Schoonbeek, H. J., Magusin, A., Pagliarani, C., Wellner, N., Hill, L., Orzaez, D., Granell, A., Jones, J. D., & Martin, C. (2013). Anthocyanins double the shelf life of tomatoes by delaying overripening and reducing susceptibility to gray mold. Current Biology, 23(12), 1094-100. https://doi.org/10.1016/j.cub.2013.04.072

Zhang, Y., Massel, K., Godwin, I. D., & Gao, C. (2018). Applications and potential of genome editing in crop improvement. Genome Biology, 19, 210.

Zhao, K., Aranzana, M. J., Kim, S., Lister, C., Shindo, C., Tang, C., Toomajian, C., Zheng, H., Dean, C., Marjoram, P., & Nordborg, M. (2007). An *Arabidopsis* example of association mapping in structured samples. PLoS Genetics, 3(1). https://doi.org/10.1371/journal.pgen.0030001

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

Zhu, C., Gore, M., Buckler, E. S., & Yu, J. (2008). Status and prospects of association mapping in plants. Plant Genome, 1, 5–20.