UNIVERSITAT POLITÈCNICA DE VALÈNCIA

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

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

Genome-wide association discovery of genomic regions for prickles-related traits using an eggplant MAGIC population

Master's Thesis

Master's Degree in Plant Genetics and Breeding

AUTHOR: Raymundo Martinez, Carlos Gonzalo

Tutor: Gramazio, Pietro

Cotutor: Plazas Ávila, María de la O

Experimental director: Luna Ramos, Jose

ACADEMIC YEAR: 2023/2024

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

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

Genome-wide association discovery of genomic regions for prickles-related traits using an eggplant MAGIC population

MASTER'S THESIS PRESENTED BY

CARLOS GONZALO RAYMUNDO MARTINEZ

ACADEMIC YEAR: 2022- 2024

TUTOR
DR. PIETRO GRAMAZIO

COTUTOR
MARÍA DE LA O ÁVILA

EXPERIMENTAL DIRECTOR
JOSE LUNA RAMOS

Valencia, September of 2024

Abstract

Prickles are a common defence mechanism in eggplant, nevertheless, their mechanism has been poorly studied. To identify genomic regions and candidate genes associated with prickles occurrence in eggplants a genome-wide association study (GWAS) was performed. To achieve high levels of association, a multi-parent advanced generation inter-cross (MAGIC) eggplant population was used. This population consisted of 312 S5 lines which were generated by crossing a wild relative (*Solanum incanum*) with seven others commonly cultivated eggplant varieties (*S. melongena*). These S5 lines were evaluated following a randomized complete block design with three blocks in two different locations: in an open field in Alcàsser and in a typical plastic greenhouse of Almería. The S5 lines were phenotyped for prickle presence in leaf, calyx and stem, as well as were high throughput genotyped following a skim whole genome resequencing approach (SWGR), yielding over 500 thousand highly confident single nucleotide polymorphisms (SNPs) that facilitate de assessment of the population's structure, evaluation of heterozygosity and the inference of haplotype blocks. Combining both genotypic and phenotypic data from the S5 MAGIC population, the GWAS analysis allowed the identification of genomic regions associated with prickle occurrence. A main region was found in chromosome 6. The results of this study will have practical implications for breeding programs that pursue the development of eggplant varieties without prickles. By shining light on the underlying genetic basis of prickles presence, this study provides a powerful tool for breeders to screen eggplant plantlets for prickles presence at a very early stage, saving important resources.

Keywords: Prickles, GWAS, MAGIC, Candidate genes, Eggplant.

Resumen

Las espinas son un mecanismo de defensa común en las plantas de berenjena, sin embargo, su mecanismo ha sido poco estudiado. Para este trabajo, se ha realizado un estudio de asociación del genoma completo (GWAS) para identificar regiones genómicas y genes candidatos asociados a la presencia de espinas en la berenjena común (*Solanum melongena*). Para conseguir altos niveles de asociación, se utilizó una población multiparental (MAGIC) de 312 líneas S5 que fueron generadas a partir de cruces entre un parental silvestre (*S. incanum*) y siete variedades de berenjena común. Estas líneas S5 se evaluaron siguiendo un diseño de bloques completos al azar y tres bloques en dos localizaciones diferentes; en campo abierto en Alcàsser y bajo invernadero de plástico en Almería. Las líneas S5 se fenotiparon para la presencia de espinas en hoja, tallo y cáliz. Además, se realizó un genotipado de alto rendimiento mediante secuenciación del genoma completo a baja cobertura, identificando más de 500 mil polimorfismos de un único nucleótido (SNPs) de alta confianza que ha facilitado el análisis de la estructura de la población, la evaluación de la heterocigosidad y la inferencia de los bloques haplotipos. Combinando ambos datos genotípicos y fenotípicos de la población S5 MAGIC, el análisis de GWAS nos permitió identificar una región genómica asociada a la presencia de espinas en el cromosoma 6. Los resultados de este estudio tienen importantes implicaciones prácticas para los programas de mejora que quieran desarrollar variedades de berenjena sin espinas. Al arrojar luz sobre el control genético detrás de la presencia de espinas, este estudio ofrece una herramienta muy poderosa para los mejoradores al poder evaluar plántulas de berenjena en una etapa muy temprana permitiendo ahorrar muchos recursos.

Palabras Clave: Espinas, GWAS, MAGIC, Genes candidatos, Berenjena.

Index

Figure and Table Index

Abbreviations

LOG: LONELY GUY

QTLs: Quantitative trait loci

GWAS: Genome-wide association study

SNPs: Single nucleotide Polymorphism

MAGIC: Multi-parent advanced generation inter-cross

QQ: Quantile quantile

LD: Linkage Disequilibrium

SWGR: Skim whole genome re-sequencing

PCA: Principal component analysis

TASSEL: Trait Analysis by Association, Evolution, and Linkage

MLM: Mixed Linear Model

SSD: Single Seed Descendant

PAD: Prickle presence in leaf's adaxial side

PAB: Prickle presence in leaf's abaxial side

PS: Prickle in Stem

PC: Prickles in Calyx

SILEX: Silica Matrix Extraction

FDR: False Discovery Rate

LOD: Limit of detection

MDS: Multidimensional scaling

PCs: Principal components

MAF: Minor Allele Frequency

NUDT19: Nudix Hydrolase 19

NADH: Nicotinamide adenine dinucleotide

CoA: Coenzyme A

RNA: Ribonucleic acid

NEC3: Carbonic anhydrase necarin-3

LYK3: LysM domain receptor-like kinase 3

ABC: ATP-Binding Cassette

ANNEX I. RELATIONSHIP OF THE PROJECT WITH THE SUSTAINABLE DEVELOPMENT GOALS OF THE 2030 AGENDA

Annex to the master's degree Final Project: Relationship of work with the Sustainable Development Goals of the 2030 Agenda

Degree of relationship of the project with the Sustainable Development Goals (SDGs).

Sustainable Development Goals		High	Middle	Low	Not appropriate
SDG 1.	**End of poverty.**	X			
SDG 2.	**Zero hunger.**	X			
SDG 3.	**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 non-polluting energy.**				X
SDG 8.	**Decent work and economic growth.**				X
SDG 9.	**Industry, innovation and infrastructures.**	X			
SDG 10.	**Reduction of inequalities**.				X
SDG 11.	**Sustainable cities and communities**.				X
SDG 12.	**Responsible production and consumption.**				X
SDG 13.	**Climate action.**				X
SDG 14.	**Underwater life.**				X
SDG 15.	**Life of terrestrial ecosystems.**				X
SDG 16.	**Peace, justice and solid institutions.**				X
SDG 17.	**Partnerships to achieve goals.**		X		

Description of the relations of the TFM with the SDGs with a higher degree of relationship.

This project aids in the end of poverty in the indirect way of providing insight into unknown plant gene regulation, which its knowledge can help in the future understand our crops in a more in depth way to then transfer this theoretical knowledge into practical knowledge in which would mean making plants better adapted to different environments which couldn't be used for crop farming before. What's more, at the same time zero hunger is another thing this project indirectly contributes to, since the increase in plant knowledge can translate to better crops.

Health and wellbeing are contributed by this project in the way that eggplants without prickles facilitates the work of the people that have to harvest them, since they don't have prickles. Also, the absence of prickles in the eggplants diminished greatly the number of punctures people suffer from handling them and makes them more desirable for consumers, and since eggplant is a highly nutritious vegetable, this would translate into a healthier society since people would eat more vegetables.

Last but not least, this project contributes to industry, innovation and infrastructure because it is part of an ongoing global effort by humans to understand plant genetics and make better use of the limited genetic resources on earth. Industry since a better understanding of plant genetics allows the rise of new genetic modification industries. Innovation since the research in plant genetics offers new ways of making crops have a higher nutrition or yield value. Infrastructure in the way that any modern and sustainable society should have a good genetic industry to propel new biotechnological advances that can feedback into the sustainability of the society itself.

1. Introducción

1.1 Eggplant (*Solanum melongena L.)*

1.1.1 The Plant

Eggplant (*Solanum melongena*, 2n = 2x = 24) is a widely cultivated vegetable around the world, known for its distinctive shape, wide variety of colours, and unique flavour. Originating in India, eggplant has been cultivated for millennia and has played a key role in many cultures' cuisines and dishes (Saha et al., 2023). Low in calories but rich in nutrients, eggplant is an excellent source of fibre, vitamins, minerals, such as vitamin C, vitamin K, folate, potassium, and magnesium. It also contains antioxidant compounds such as polyphenols and anthocyanins, which may help reduce the risk of chronic diseases and improve cardiovascular health, therefore, giving added value to eggplant cultivation (Saha et al., 2023).

1.1.2 Taxonomy

Solanum melongena L. (brinjal eggplant) is a member of a small monophyletic group that is mainly andromonoecious species in what englobes the large and diverse clade of *Solanum*. There are approximately 1300 species in the genus of *Solanum* L. in the nightshade family Solanaceae (Knapp et al., 2004). The family is composed of 101 genera, including other important horticultural and economic species such as *Nicotiana* L. and *Petunia* L. Of these genera *Solanum* is the largest genus, which includes well-known crops like potato and tomato, along with other minor crops. *Solanum* comprises around half of the species diversity of the family (Knapp et al., 2019) and can be categorized into 13 clades, with the monophyletic clade of *Leptostemonum* (Stern et al., 2011) being the biggest of them all, harbouring over 550 species distributed on all continents except Antarctica.

eggplant Classification	
Kingdom	*Plantae* - Plants
Subkingdom	*Tracheobionta* - Vascular plants
Superdivision	*Spermatophyta* - Seed plants
Division	*Magnoliophyta* - Flowering plants
Class	*Magnoliopsida* - Dicotyledons
Subclass	*Asteridae*
Order	*Solanales*
Family	*Solanaceae* Juss. - Potato family P
Genus	*Solanum* L. - nightshade P
Species	*Solanum melongena* L. - eggplant P

Figure 1.1 Taxonomic hierarchy of eggplant (S. melongena). Source: USDA (United States Department of Agriculture)

1.1.3 Botany

Eggplant is well known for its characteristic fruits, berries that exhibit a wide range of shapes and sizes and that usually turn brown in the presence of air when cut. The presence and absence of prickles in the plant contribute to the diversity of the family. Furthermore, the levels of prickles as well as levels of anthocyanin and hairiness on the vegetative parts vary quantitatively, contributing to the previously mentioned diversity. Eggplant is a non-woody type of plant that has many branches in a tough dichotomic branching pattern. When it comes to the inflorescence of the plant, they usually range from one to five andromonoecious inflorescences, although in modern-day cultivars they are usually hermaphrodite inflorescences. Flowers usually consist of five petals, sepals and stamens. Although eggplant is normally considered an autogamous species; it has been seen to be quite allogamous in open fields due to the presence of insects (Kumari et al., 2023; Rakha et al., 2021).

1.1.4 Eggplant wild relatives

Eggplant's wild relatives are very diverse and could offer a source of variation for breeding programs. This variation not only comes in the form of phenotypical variety but also physiological variation which expresses itself in a form of adaption to harsh climates and pathogen presence. Eggplant is thought to have originated in Asia, but other studies suggest that it originated from Africa since most of its wild relatives reside there (Aubriot

et al., 2018). This could be backed up by phylogenetic studies that further prove the presence of eggplant in Africa before spreading to Asia by humans; this is a common occurrence in which the domestication hotspot is commonly misinterpreted as the origin of the "artificial" diversity of eggplants in Asia. As any other wild relatives of a certain modern cultivar plant, their groups are the most diverse and complex when it comes to their taxonomic and phylogenetic relationships. These wild relatives are commonly prickly, small, bitter, multi-seeded and are obviously inedible; with some of the varieties containing high levels of bioactive compounds which may be of interest for human health or biochemistry (Knapp et al., 2013).

The eggplant complex is comprised of nine wild species in addition to *S. melongena*; these encompass the cultivated eggplant and its closest relatives (Knapp et al., 2013). There are various classification criteria for the wild relatives of eggplant; they can be classified based on their crossability with cultivated species, resulting in primary, secondary and tertiary gene pools. The primary gene pool is usually the wild relatives close to the common eggplant since it can cross easily, in this case *S. incanum*. In the second gene pool, the wild relatives are harder to cross with the cultivated species and include *S. linnaeanum*, *S. tomentosum*, among others (Kouassi et al., 2016; Rotino et al., 2014). Lastly, in the third gene pool, we can find wild relatives like *S. elaeagnifolium* and *S. torvum* (Kouassi et al., 2016) which are more distantly related species and therefore require specific breeding techniques to pull through the many reproductive barriers, although they offer interesting resistances to abiotic and biotic stresses alike.

1.1.5 Origin and domestication

Eggplant is considered to have originated in Asia; Indo-Chinese centre of origin (VAVILOV, 1951). As any other domestication history, the true origin of the eggplant is yet to be set in stone. Despite various studies, it has unable to pinpoint the exact location of the origin of the eggplant. Many studies propose that the centre of domestication is India, although China possesses various ancient records of eggplant usage dating back approximately 2000 years (Bhaduri, 1951; Daunay et al., 1998; Doganlar et al., 2002; Lester, 1991; Mace et al., 1999; F. W. Martin et al., 1981; Meyer et al., 2012; Zeven & de Wet, 1982). This also brings the problem of trusting written ancient record since the Chinese dynasty with reliable records was established later than the first mature Indian civilization that would record such things. Some studies suggest that there are multiple independent domestication events from *S. insanum*, which is the state-of-the-art eggplant wild

progenitor that is distributed along tropical Asia. From Asia, eggplant was distributed to other countries like Europe and Africa during the silk road from around the 2nd century BCE to the 14th century CE. It was later introduced to the rest of the continents like America, spreading all over the world.

1.1.6 Economic importance

Eggplant is cultivated approximately on 1.89 hectares globally, making it a widely produced and consumed crop around the world. It yields a total production volume of 60 million tons (*FAOSTAT*, 2022). This production is predominantly concentrated in Asia and represents 94% of the global production volume (*FAOSTAT*, 2022). The largest producers are China (64.57%), India (21.51%), Egypt (2.35%), Turkey (1.32%), and Indonesia (1.17%) (*FAOSTAT*, 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%) (*FAOSTAT*, 2022).

The crop is widely used in various cuisines like, India, Bangladesh, the Middle East and Southeast Asia. Eggplant is a low-calorie food and can contain an assortment of vitamins, minerals and bioactive compounds, making it highly nutritious and healthy as a vegetable to consume in a well-balanced diet. Eggplants have phenolic compounds which are known to partake in the reduction of free radical effects which are correlated with aging and other health issues.

Apart from its economic and nutritional importance, eggplant is an important asset in assessing the physiology, genetics and evolution of other solanaceous crops that are commonly used, such as tomato or potatoes. This can give us insight into how they behave to different stresses and how their physiology tries to adapt to said challenges.

1.2 Plant prickles

Prickle is a sharp protrusion that covering plant and fruit calyxes of eggplant are considered undesirable agronomic traits, for they bring troubles and create additional costs for farmers. However, little is known about its regulatory genes and molecular mechanism of morphogenesis (S. Li et al., 2024).

1.2.1 Description

Prickle as an outgrowth organ is considered a bulwark of plant defences against biotic and abiotic stresses (Bagella et al., 2019). However, they are usually considered undesirable agronomic traits, because they will hurt the farmers engaging in agricultural

operations and scratch the skin of other eggplants during harvest and transportation, which brings additional labour and costs to field management and sales. Hence, creating prickle-free eggplant cultivars is one of the goals of the eggplant breeding programs (Alam & Salimullah, 2021). While it is hard to reach this goal sometimes, since prickles could be brought back accompanied by the application of some agricultural technologies. The integration of wild relatives' excellent genes into modern commercial eggplant cultivars by traditional crossbreeding methods will be accompanied by the introduction of prickles, and removing the prickles requires multi-generation backcross, which will cost many years (Hurtado et al., 2014).

1.2.2 Prickle formation

The molecular regulatory mechanisms of prickle development are not so clear for lacking ideal model plants and their complex genetic mechanism. For example, the inheritance of prickles in roses does not conform to Mendelian's laws (Arun Kumar et al., 2014; Kellogg et al., 2011; Khadgi & Weber, 2020; Qian et al., 2021). Prior studies deduced that prickles were modified of lignified glandular based on the comparison of morphological structure and announced that prickles originated from protodermal cells as glandular trichomes did (Arun Kumar et al., 2014; Kellogg et al., 2011; Khadgi & Weber, 2020). Based on this acknowledgement, comparative transcription methods were employed to mine some homologous genes related to glandular trichome development, which were considered candidate genes for the regulation of prickle initiation in eggplant (M. Chen et al., 2024; Qian et al., 2021). However, recent studies showed that genes related to trichome development had little effect on the development of prickle in Rosa chinensis, suggesting that the regulatory network of prickle development may be different from that of trichome (M. Chen et al., 2024). In addition, some opinions argued that the prickles were not specialized from the glandular trichome, for the originated cells and divergent morphogenesis process (M. Chen et al., 2024; Zhang et al., 2021a). Suggesting that is hard to say the homologous genes regulating trichome development possess the function related to prickle development.

1.2.3 Genetic Regulation

Eggplant has a high-quality genome sequenced and completed, which is conducive to fine mapping of the candidate genes (Wei et al., 2020). In the past 20 years, scientists have detected many QTLs associated with the prickle traits (leaf prickle, stem prickle, calyx prickle, and petiole prickle) on chromosomes 1, 2, 3, 5, 6, 7, 8, 9, and 12 by constructing

the interspecific and intraspecific hybrid F2 populations. It was found that one prickle trait could be regulated by multiple QTLs, and some QTLs could regulate more than one prickle trait (Frary et al., 2014; Portis et al., 2014; Qian et al., 2021). Among the many QTLs, one hotspot QTL located on chromosome 6 was found to be associated with all prickle traits (Frary et al., 2014).

Notably, a major QTL related to calyx prickle development was located within a 7 kb interval on chromosome 12, identifying a gene encoding a WUSCHEL-related homeobox-like protein as the candidate (Qian et al., 2021). Another QTL on chromosome 6 responsible for prickle presence or absence was fine-mapped to an interval of 133-kb, leading to the development of a PAV marker to assist in breeding programs (Miyatake et al., 2020). Furthermore, this QTL was narrowed down to a 28.3 kb region, identifying SmARF18, an auxin response factor, as a potential candidate due to a non-synonymous SNP (S. Li et al., 2023). Alas, neither of these candidate genes has undergone functional validation to confirm their role in prickle formation. However, a recent study found that a Cytokinin biosynthetic gene called *LONELY GUY* (*LOG*) in chromosome 6, that when having a splice-site mutation, was the cause of prickle loss in eggplant and across the *Solanum* (Satterlee et al., 2024).

1.2.4 Importance

Despite growing interest, functional studies on prickle-regulating genes are notably scarce. This lack of definitive research contributes to the ongoing ambiguity surrounding the genetic basis of prickle formation. Therefore, identifying and analysing the key genes responsible for prickle formation is essential for shedding light on the mechanisms underlying their morphogenesis (Zhang et al., 2024a).

1.3 Association Mapping

Much interest has been shown in using association mapping to identify genes responsible for the quantitative variation of complex traits with agricultural and evolutionary importance. Recent advances in genomic technology and the development of robust statistical analysis methods make association mapping appealing and affordable for plant research programs. The phenotypic variation of many complex traits of agri-cultural or evolutionary importance is influenced by multiple quantitative trait loci (QTLs), their interaction, the environment, and the interaction between QTL and the environment.

Linkage analysis and association mapping are the two most used tools for dissecting complex traits (Krishna et al., 2023).

1.3.1 Description

Association mapping is an approach that involves examining the marker-trait associations that can be attributed to the strength of linkage disequilibrium between markers and functional polymorphisms across a set of diverse germplasm (Zhu et al., 2008). This method aims to locate genomic regions and candidate genes associated with specific traits. The basic principle involved in association mapping is the co-inheritance of alleles of SNPs in adjoining regions in a population due to Linkage disequilibrium (LD) and a strong correlation between allele variants and traits in a natural population (Bush & Moore, 2012; J. A. Martin & Wang, 2011).

As a new alternative to traditional linkage analysis, association mapping offers various advantages like; increased mapping resolution, reduced research time and greater allele number (Zhu et al., 2008). Since its introduction to plants (Juliana et al., 2018), association mapping has continued to gain favourability in genetic research because of advances in high throughput genomic technologies, interests in identifying novel and superior alleles, and improvements in statistical methods (Zhu et al., 2008).

1.3.2 Genome-wide association study

A genome-wide association study (GWAS) is an approach used for association mapping, designed to identify all the genomic variants that contribute to trait variation. In typical GWAS, phenotype and genotype data are collected for a large sample of assembled individuals such as a diversity panel (Tibbs Cortes et al., 2021), using appropriate statistical methods to detect variants that are more frequently found in those displaying a specific phenotype compared to those with different phenotypes (Jain et al., 2024).

The Manhattan plot is a common method for visualizing GWAS results, where SNPs with the most significant associations appear as prominent peaks (Pasam & Sharma, 2014). This scatter plot displays the negative logarithm of p-values (y-axis) for SNP associations against the SNP positions (x-axis) (Pasam & Sharma, 2014). Longer peaks on the Manhattan plot indicate that the surrounding genomic region is strongly associated with the trait (Jain et al., 2024). A major problem of this naïve approach is its high rate of false positives, which occur when a finding is declared significant even though it is not actually true. In part, this is because testing each SNP in the dataset results in thousands or even millions

of statistical tests (Tibbs Cortes et al., 2021). Common methods for correcting for multiple testing include controlling the false discovery rate (FDR) and the Bonferroni correction (Benjamini & Hochberg, 1995; Storey & Tibshirani, 2003). The Bonferroni correction adjusts the significance threshold by dividing it by the total number of tests performed (Tibbs Cortes et al., 2021).

Additionally, the quantile-quantile (QQ) plot is a useful tool for assessing the robustness of GWAS results for marker-trait associations (Jain et al., 2024). This plot compares the observed distribution of p-values to their expected distribution under the null hypothesis, where no association with the trait exists (Jain et al., 2024). Since most tested SNPs are likely unrelated to the trait, most points on the QQ plot should fall along the diagonal line (Jain et al., 2024). Deviations from this line, especially points above the diagonal line, suggest potential false positives due to factors like population structure and familial relatedness (Jain et al., 2024). These deviations indicate that the GWAS model may not adequately correct for these consequential associations, but this can be addressed by accounting for population structure (Jain et al., 2024).

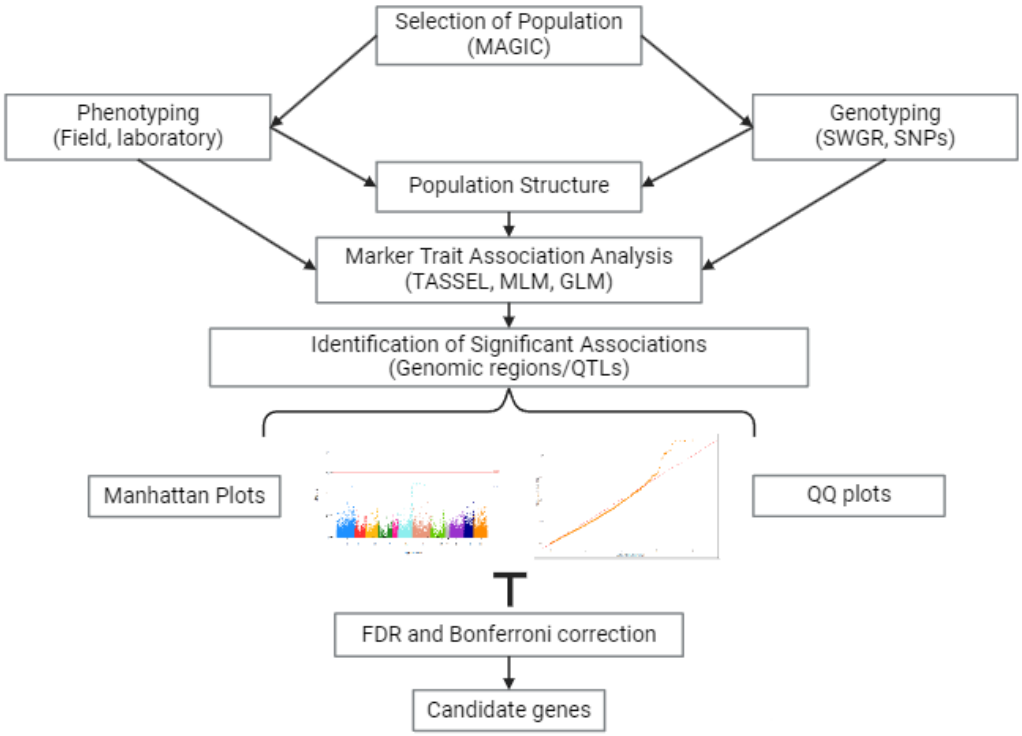

Figure 2.2 General pipeline for conducting genome-wide association study. Source: Pasam and Sharma (2014)

1.3.3 Mapping population

The choice of germplasm is critical to the success of association analysis (Breseghello & Sorrells, 2006; Flint-Garcia et al., 2003; Yu et al., 2005). Genetic diversity, the extent of genome-wide LD, and relatedness within the population determine the mapping resolution, marker density, statistical methods, and mapping power. Generally, plant populations amenable to association studies can be classified into one of five groups (Yu et al., 2005). The first group is an ideal sample with subtle population structure and familial relatedness, the second a multi-family sample, the third a sample with population structure, the fourth a sample with both population structure and familial relationships, and the fifth group would be a sample with severe population structure and familial relationships. Due to local adaptation, selection, and breeding history in many plant species, many populations for association mapping would fall into category four. Alternatively, we can classify populations according to the source of materials, germplasm bank collections, synthetic populations, and elite germplasm (Breseghello & Sorrells, 2006). In this study, a multiparent advanced generation intercross (MAGIC) population was utilized. That would fall in the second group.

1.3.4 Genetic Markers

In recent decades, progress in DNA sequencing technology has transformed these disciplines from being constrained by limited information to becoming abundant in data production (Adhikari et al., 2022). Genotyping is crucial for quantitative and population genetic studies, especially in genomic-assisted breeding in crops (Adhikari et al., 2022). As sequencing costs decline and its usage increases, there is a growing focus on maximizing the utilization of these methods and technologies in breeding pipelines and genetic studies as well as their efficiency and reliability (Rasheed et al., 2017).

Single-nucleotide polymorphisms as genetic markers have been used in several crops for mapping of genes for a variety of traits including agronomic, quality, biotic, and abiotic stress resistance. The major advantage of using SNPs is that these are present in abundance uniformly across the genome, even in species with a narrow genetic base (Kulwal & Singh, 2021).

Skim whole genome re-sequencing resequencing (SWGR) was employed to efficiently generate dense SNP markers in a cost-effective manner in this study. WGR is well suited for genotyping biparental cross populations with complex, small to moderate-sized

genomes and provides the lowest cost per marker data point. These approaches are generally better suited for de novo applications and more cost-effective when genotyping populations with large genomes or high heterozygosity (Scheben et al., 2017).

1.3.5 Software analysis

Association analysis can begin once both trait data and marker genotypic data for the population have been acquired. This data undergoes extensive preprocessing for several parameters; marker filtration, principal component analysis (PCA), kinship, and population structure, before being used for association mapping methods like MLM or GLM. Different software packages are required to carry out all these activities and perform the analysis with sufficient power to detect significant candidate regions. In recent years, several software packages have been developed to perform association analysis in an efficient manner (Kulwal & Singh, 2021). From the many software packages available for association mapping, in this study, Trait Analysis by Association, Evolution, and Linkage (TASSEL) software was used since it is a widely used tool in plant systems by the plant breeding community (Bradbury et al., 2007).

The advantages of using TASSEL include the implementation of advanced statistical approaches for association mapping, such as the General Linear Model (GLM) and the Mixed Linear Model (MLM) (Bradbury et al., 2007).

1.3.6 Importance

GWAS methodology has advanced such that it is now a powerful tool for the analysis of simple traits under additive genetic scenarios, and for the dissection of more complex genetic architectures (Korte & Farlow, 2013). Genome-wide association studies have investigated agriculturally important traits in many major crop species, including maize (*Zea mays L.*), wheat (*Triticum aestivum L.*), rice (*Oryza sativa L.*), cotton (*Gossypium hirsutum L.),* and numerous other crops beyond the model plant species of Arabidopsis (Ersoz et al., 2007; Han & Huang, 2013; Sahito et al., 2024). Genome-wide association studies have identified genomic regions associated with many agronomic, physiological, and fitness traits including flowering time, plant height, kernel number, stress tolerance, and grain yield (Ersoz et al., 2007; Gupta et al., 2019)**.** Genome-wide association studies have also been used to study other types of phenotypes. Genome-wide association studies in rice have identified genes associated with geographical divergence and adaptation during domestication (E. Chen et al., 2019) as well as with biochemical and molecular

phenotypes including flavonoid, fatty acid, amino acid, and nucleic acid metabolites (W. Chen et al., 2016). These studies are used both to detect novel associations with valuable traits and to validate loci identified by other methods. Genome-wide association studies may be conducted as stand-alone investigations, as a component of gene cloning studies, or as the foundational step in marker-assisted selection, among other uses. In turn, exploiting this information accelerates crop breeding (Tibbs Cortes et al., 2021).

2. Objectives

This study aims to perform a genome-wide association study (GWAS) for prickle traits in eggplant by using a MAGIC S5 population by:

1. Identifying single nucleotide polymorphism (SNP) markers associated with prickle formation in the adaxial and abaxial sides of the leaf, stem and calyx.
2. Detect putative candidate genes within genomic regions that exhibit significant marker-trait associations.

3. Materials and Methods

3.1 Plant Materials

The S5 eggplant MAGIC population that was used in this study was generated by intercrossing seven eggplant cultivars: DH ECAVI, AN-S-26, A0416, MM597, AISI-S-1, MM1597, H15 and MM577; being the latter an eggplant wild relative (*S. incanum*). MM577 which came from an Israeli desertic area, was chosen for its drought resistance capabilities (Knapp et al., 2013). These parent lines were previously evaluated for their genetic diversity and morpho-agronomic characteristics and capabilities (Kaushik et al., 2018; Prohens et al., 2017; Vilanova et al., 2014).

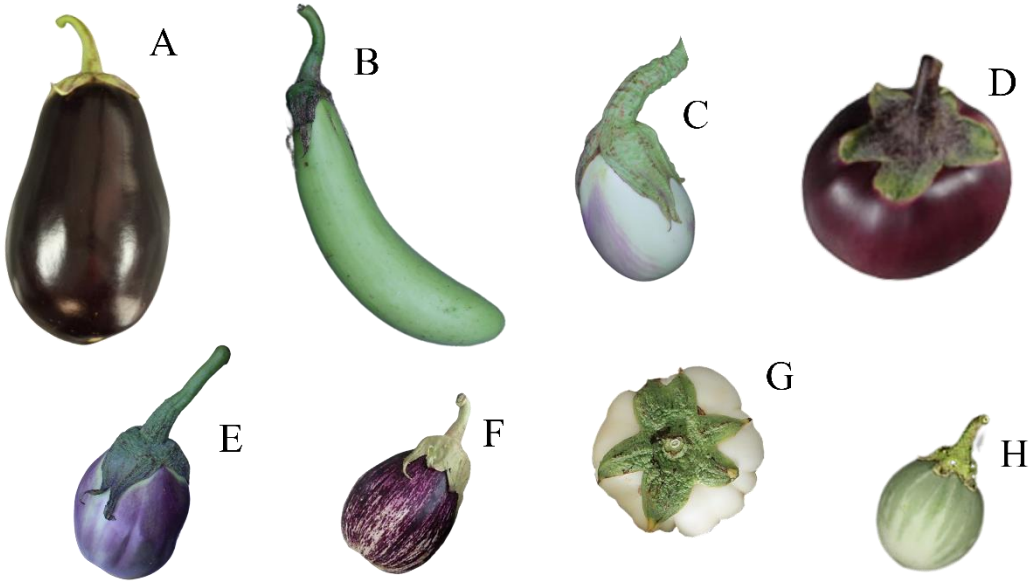

Figure 3.1 Founder parents of the S5 MAGIC eggplant population: DH ECAVI (A), MM1597 (B), H15 (C), ASI-S-1 (D), AN-S-26 (E), IVIA371 (F), A0416 (G) and MM577 (H).

3.2 MAGIC population development

MAGIC (Multiparent Advanced Generation Inter-Cross) populations are powerful tools in plant genetic research. These populations are composed of multiple parental lines that have been reciprocally crossed for several generations to create a high genetic diversity in a single final population. In the case of eggplant, a MAGIC population could include varieties selected for their different characteristics, such as yield, disease resistance, fruit size and colour, chlorophyll content, and prickles. Thanks to several generations of MAGIC eggplant populations and the phenotyping of both their fruits and their corresponding plants, it is possible to analyse and search for some major genes for characters such as prickles and anthocyanin levels to try to determine their genetic control or at least associate some regions with said genetic control that will allow us in the future to use markers for a selection of these characters in other improvement programs for other eggplant varieties with agro-economic value.

For the development of the S5 MAGIC population, a funnel method was used for the inter-crosses between the funding parents (Figure 3.2) (Arrones et al., 2020; Wang et al., 2017). The eight founder parents (A to H) were initially crossed in pairs to generate simple Hybrids (AB, CD, EF and GH). To be able to guarantee the complete genomic admixture and to avoid assortative mating, the double hybrids were subject to chain pollination; in

which each individual was used once as a female and as a male parent (Díez et al., 2002). Therefore, the quadruple hybrids produced (S0 recombinant lines) were then self-fertilized through single seed descendent (SSD) to generate the S5 recombinant lines (S5 MAGIC lines), which then were genotyped and phenotypes for this study. Therefore, in total, 326 S5 MAGIC lines were produced.

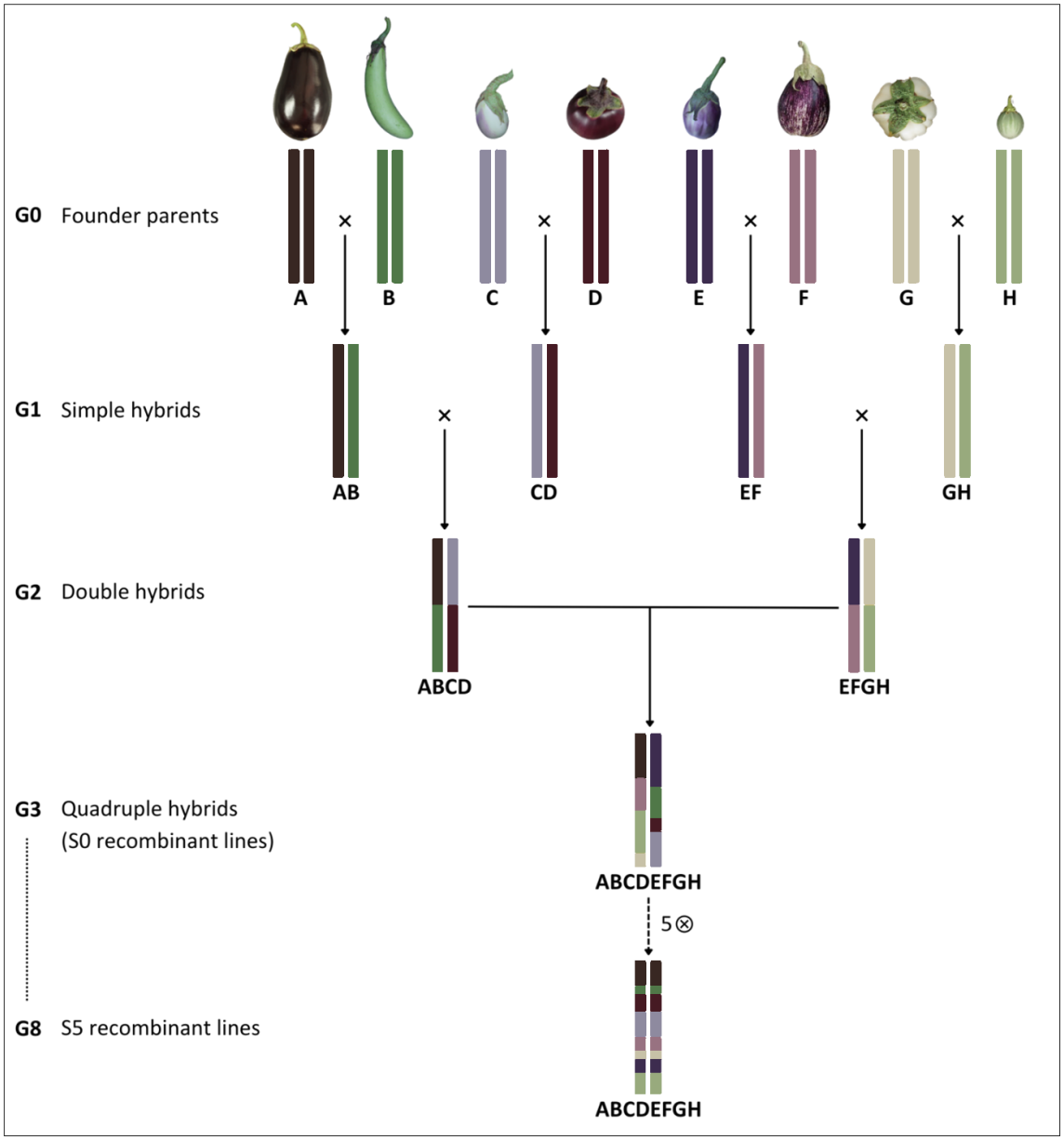

Figure 4.2 Simple funnel breeding design used to develop the S5 MAGIC eggplant population. The founder parents being DH ECAVI (A), MM1597 (B), H15 (C), ASI-S-1 (D), AN-S-26 (E), IVIA371 (F), A0416 (G) and MM577 (H).

3.3 Cultivation conditions and development

No specific protocol was followed. The seeds of the 326 S5 MAGIC lines were germinated directly in seedling trays which were placed in a climate-controlled chamber (Yuan et al., 2015). This chamber was maintained on a 16-hour photoperiod of light at 25ºC and 8 hours of darkness at 18ºC. After proper climatization, the seedlings were grown in two separate locations in Spain: Alcàsser. Valencia (39°23'40"N, 0°26'55"W) under open field conditions and Almería, Andalucía (36°49'14"N, 2°14'1"W) under greenhouse plastic conditions. The plants were arranged in spaces 1.2 meters apart between rows and 1 meter apart within rows. Plants were fertigated with drip irrigation system, and phytosanitary treatments were administered correspondingly. To manage vegetative growth and flowering manual pruning was preformed and vertical string were used for training.

3.4 Phenotyping

Phenotypic data from the 326 S5 MAGIC lines was recorded by scoring prickle presence in the leaf's adaxial side (PAD), leaf's abaxial side (PAB), stem (PS) and calyx (PC). In total 1,746 plants were phenotyped for PAD, 1,746 for PAB, 1,744 for PS and 1,705 for PC. The phenotyping of leaves, stem and calyx was performed by using scale from 0 to 10, where 0 indicates total absence of prickles, 2 indicates presence of one or two prickles, 3 indicates 3,4 or 5 prickles, 5 indicates 6 or 7 prickles, 7 indicates 8 or 9 prickles and finally 9 indicates 10 or more prickles present. Figure 3.3 represents a visual representation of the different prickle presence scores. Statistical averages were used to diminish the impact of environmental factors on the traits. Means/medians were calculated for all traits given the ordinal nature of the data. Table 3.1 gives us an overview of the phenotypical diversity present among the different founder parents.

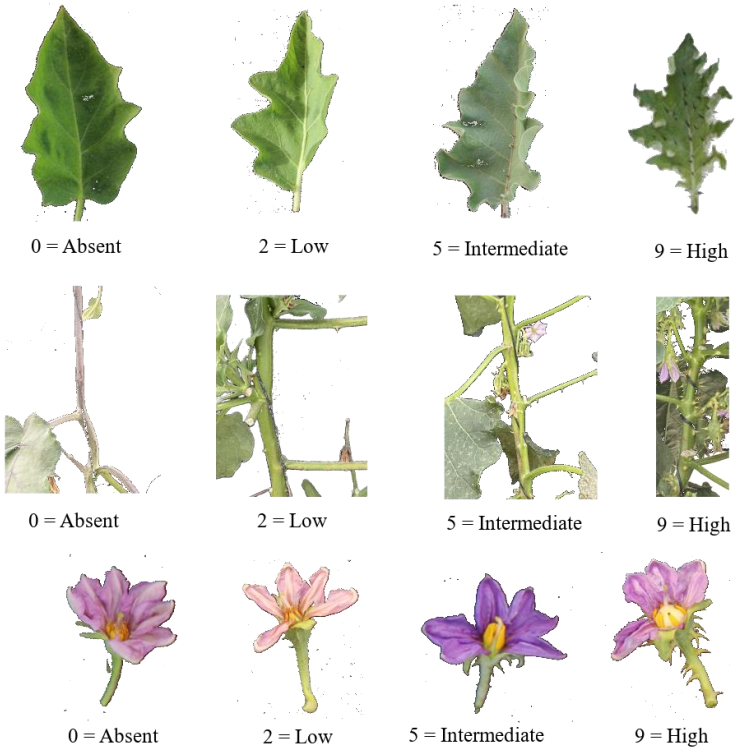

Figure 5.3 Prickle presence scoring in leaves, stem and calyx of the S5 MAGIC eggplant population.

Table 6.1 Prickle presence in the leaves (PAB/PAD), stem (PS), calyx (PC) of the eight founder parents of the S5 MAGIC eggplant population.

Founder Name	Founder Code	Scientific Name	Origin	PAD/PAB	PS	PC
DH ECAVI	A	*S. melongena*	Unknown	Absent	Absent	Low
MM1597	B	*S. melongena*	India	Absent	Absent	Absent
H15	C	*S. melongena*	Spain	Absent	Absent	Low
ASI-S-1	D	*S. melongena*	China	Absent	Absent	Low
AN-S26	E	*S. melongena*	Spain	Absent	Low	Low
IVIA371	F	*S. melongena*	Spain	Absent	Absent	Low
A0416	G	*S. melongena*	Unknown	Absent	Absent	Intermediate
MM577	H	*S. incanum*	Israel	Low	Intermediate	Intermediate

3.5 Genotyping

Genomic DNA of 326 S5 MAGIC lines was extracted from approximately 100 mg of young leaves using the Silica Matrix Extraction (SILEX) protocol (Vilanova et al., 2020). The integrity and quality of the extracted DNA was checked via electrophoresis in 1% agarose gel and spectrophotometry at an absorbance of 260/230 and 260/280 using Nanodrop ND-1000 (Nanodrop Technologies, Wilmington, DE, USA). The concentration of the extracted DNA was determined by using a Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, United States of America).

The sequencing of the DNA samples was performed by skim whole genome resequencing (SWGR) with a read length of 150 bp, yielding approximately 3.0 GB of high-quality paired end reads per sample with a 3X coverage. Using BWA-MEM (H. Li, 2013) the alignment of the reads were done against the 67/3 eggplant reference genome version 3, to be continued by the removal of PCR duplicates with Picard version 1.119 (https://broadinstitute.github.io/picard). Variant calling was performed using Freebayes (Garrison & 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: loci with a minor allele frequency greater than 0.0031 were retained, positions with more than 20% heterozygosity or monomorphic sites were excluded and genotypes supported by fewer than three reads were marked as missing. Imputation of missing data was performed using Beagle (Browning & 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 the Trait Analysis by Association, Evolution, and Linkage (TASSEL) software version 5.2.93 (Bradbury et al., 2007) to retain the most reliable ones (minor allele frequency > 0.15) with the addition of thinning by position (10k) to reduce the computational load and avoid overfitting in analyses such as GWAS ensuring that the remaining markers are more independent and informative (Chang et al., 2018). In total, 15,176 SNP markers were utilized for downstream analyses.

3.6 Population structure and residual heterozygosity assessments

Mixed Linear Model (MLM) and Multidimensional scaling (MDS) were conducted to examine the genetic structure of the S5 MAGIC population using TASSEL software version 5.2.93 (Bradbury et al., 2007). Genetic relatedness and the residual heterozygosity of the population was also evaluated using TASSEL software version 5.2.93 (Bradbury et al., 2007).

3.7 Genome-wide association mapping

Association mapping for prickle presence was done using the mixed linear model (MLM) of TASSEL software version 5.2.93 (Bradbury et al., 2007), incorporating the kinship and population structure matrices as covariates. Testing was corrected using the Bonferroni

and False Discovery Rate (FDR) methods to identify candidates associated with significant genomic regions. SNPs with limit of detection (LOD) score (−log10(p-value)) surpassing the Bonferroni and FDR significance thresholds were considered significantly associated with prickle presence.

3.8 Candidate region identification

Prickle related genes positioned in the vicinity of the most significant SNP marker were identified by looking at the genomic map of the 67/3 reference genome version 3 (Portis et al., 2014) and were considered candidate genes for taking part in the traits studied.

4. Results

4.1 Prickle presence

The S5 MAGIC eggplant population showed phenotypical variability in prickle presence in PAB, PAD, PS and PC (Figure 4.1). Among the 1,690 S5 MAGIC individuals assessed for PAB, 1,481 (87,63%) didn't display prickles while only 209 (12,37%) did display prickles. For PAD, 1,690 individuals were evaluated, 1,337 (79,11%) presenting no prickles and 353 (20,89%) with prickles. PS with 1,688 individuals were also mostly with no prickles 1,299 (76,95%) than with prickles 389 (23,05%). In contrast, PC showed the opposite, since with 1650 individuals, only 441 (26,72%) showed no prickles, while 1,209 (73,27%) did show prickles. Phenotypic diversity of the S5 MAGIC eggplant population in leaves, stems and calyx is shown in Figure 4.2.

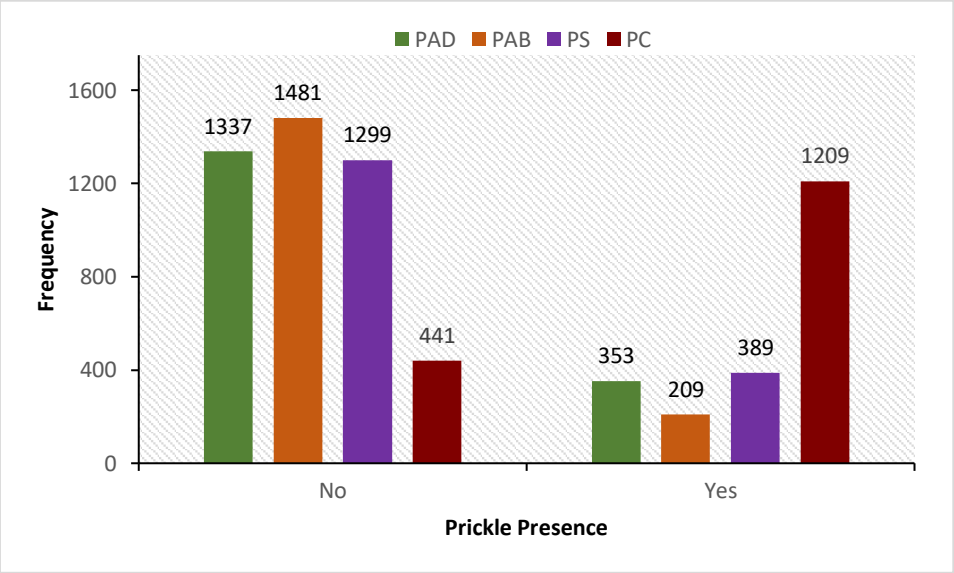

Figure 7.1 Phenotypic diversity of the S5 MAGIC population in terms of presence of prickles in the leaves (PAD/PAB), stem (PS) and calyx (PC). Instead of using the range from 0 to 10, a binomial approach was used in this scenario.

Figure 8.2 Phenotypic diversity of the S5 MAGIC eggplant population in terms of fruit morphology.

4.2 SNP marker distribution

326 S5 MAGIC lines were genotyped by SWGR coupled with imputation of missing marker data, producing a total of 563,783 SNPs across the 12 chromosomes of eggplant (Table 4.1). Among these chromosomes: chromosome 1 had the most with 71,387 SNPs while chromosome 9 had the fewest SNPs (18,080) (Table 4.1).

After filtering with TASSEL with a minor allele frequency of 0.15 and thinning by sites 10k (Chang et al., 2018) 15,176 markers were retained, representing 2.69% of the total SNPs generated from SWGR (Table 4.1). Of the 12 eggplant chromosomes, chromosome 9 had the lowest number of SNPs (525) and chromosome 1 still had the highest number of SNPs (2,380) (Table 4.1). These SNPs were then used for downstream analyses, including population structure and kinship estimation, residual heterozygosity assessment and GWAS.

Across the eggplant genome, the SWGR that had generated many SNPs, had an overall average density of 492,54 SNPs/Mb, with chromosome 11 having the lowest density (457.71 SNPs/Mb) and chromosome 12 having the highest density (522.89 SNPs/Mb). After filtering with TASSEL, the overall density diminished to 13.70 SNPs/Mb, in this case chromosome 4 having the lowest density (6.59 SNPs/Mb) and chromosome 11 having the highest (28.16 SNPs/Mb) (Table 4.1). Figure 4.3 illustrates the distribution of the SNPs used in this study across de Eggplant genome.

Table 9.1 Statistics of 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	71,387	2,380	12.66	15.68	136.53	522.87	17.43
2	41,343	1,034	7.33	6.81	83.34	496.08	12.41
3	46,055	893	8.17	5.88	97.01	474.74	9.21
4	49,579	696	8.79	4.59	105.67	469.19	6.59
5	21,810	539	3.87	3.55	43.85	497.38	12.29
6	54,647	1,648	9.69	10.86	108.97	501.49	15.12
7	69,347	1,285	12.30	8.47	142.38	487.06	9.03
8	52,428	811	9.30	5.34	109.58	478.44	7.40
9	18,080	525	3.21	3.46	36.1	500.83	14.54
10	53,511	1,593	9.49	10.50	106.64	501.79	14.94
11	33,088	2,036	5.87	13.42	72.29	457.71	28.16
12	52,509	1,736	9.31	11.44	100.42	522.89	17.29
Total	**563,784**	**15,176**	**100.00**	**100.00**	**1,142.80**		
Average						**492.54**	**13.70**

[1]SNPs derived from the SWGR coupled with missing marker data imputation
[2]SNPs generated after filtration (minor allele frequency > 0.15 and thinning by sites 10k) using TASSEL
[3]Chromosome length based on 67/3 eggplant refence genome version 3 (Barchi et al., 2019)

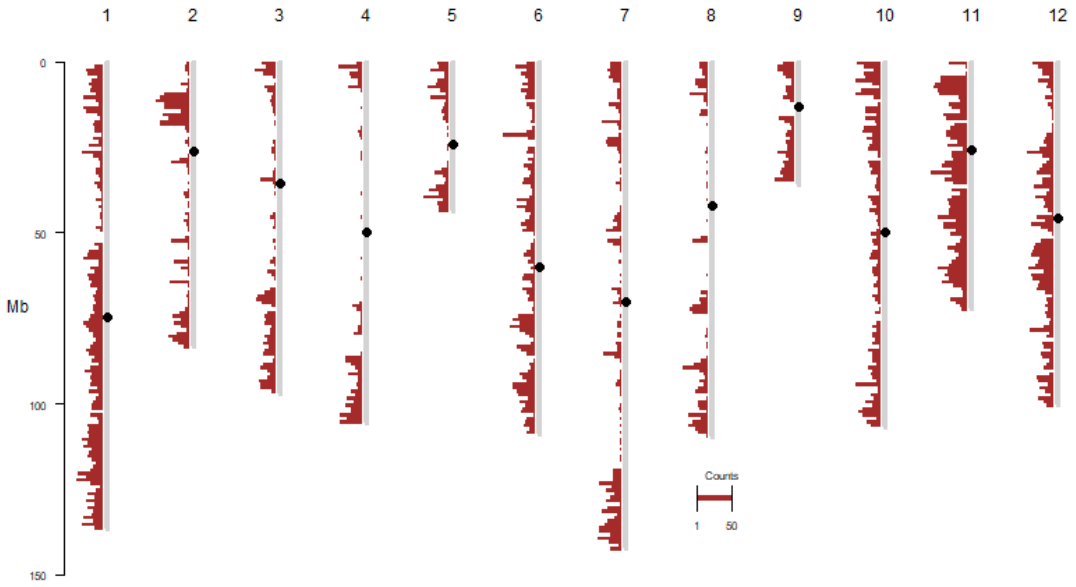

Figure 10.3 Genome-wide distribution of 15,176 SNPs used in this study.

4.3 Population structure and residual heterogeneity

Multidimensional scaling (MDS) analysis for assessing population structure, based on SNP marker data of 326 MAGIC eggplant lines, revealed that the first three principal components (PCs) explained 5% (PC1), 5% (PC2), and 5% (PC3) of the total variance (Figure 4.5). Of the 326 MAGIC eggplant lines, only 15 exhibited a heterozygosity rate above 20%. The 15% variance explained by these principal components indicates a somewhat 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 5.81%, according to SNP marker data (Figure 4.6).

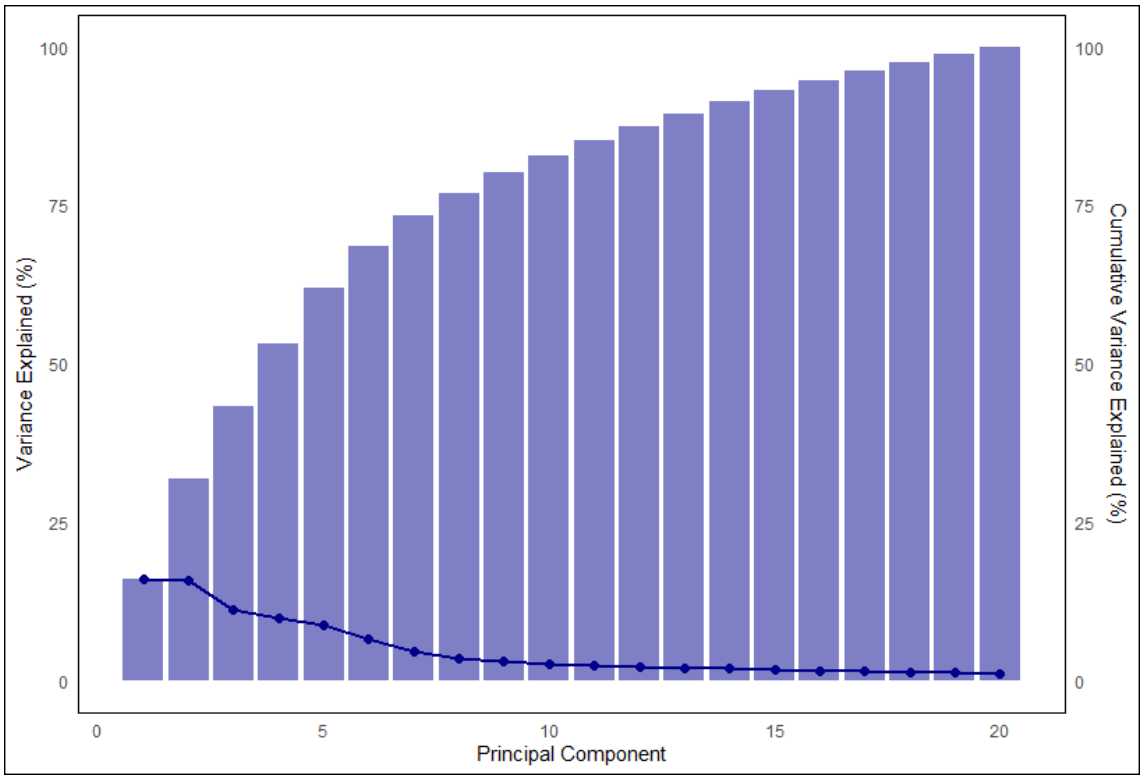

Figure 11.4 Scree plot of the variance (%, primary y-axis) and bar plot of the cumulative variance (%, secondary y-axis) explained by 50 principal components (x-axis)

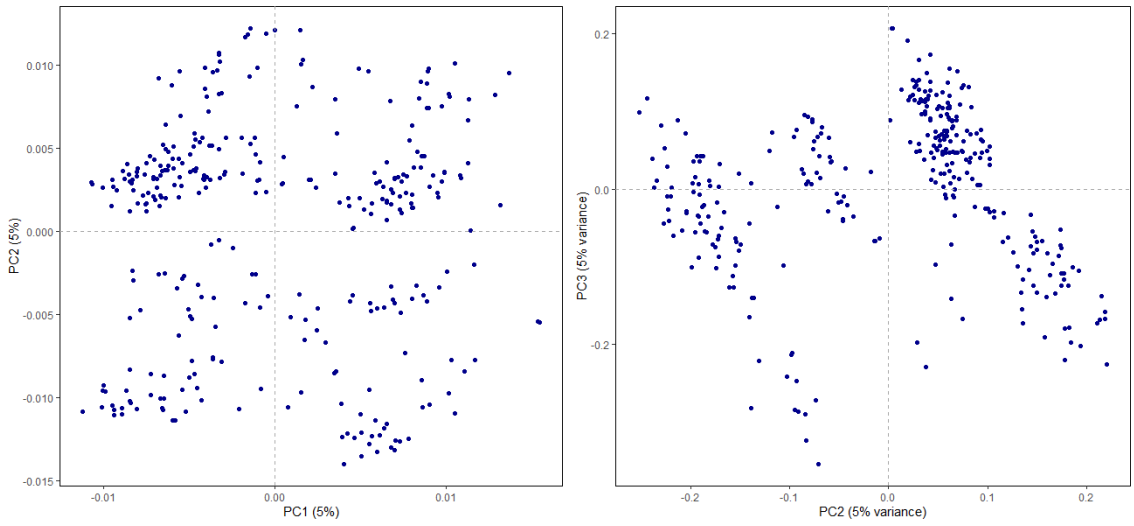

Figure 12.5 Multidimensional scaling plots for the 308 S5 MAGIC eggplant lines based on PC1 vs. PC2 (left) and PC2 vs. PC3 (right). PC1 accounted for 5% of the variance, PC2 accounted for 5%, and PC3 also accounted for 5%.

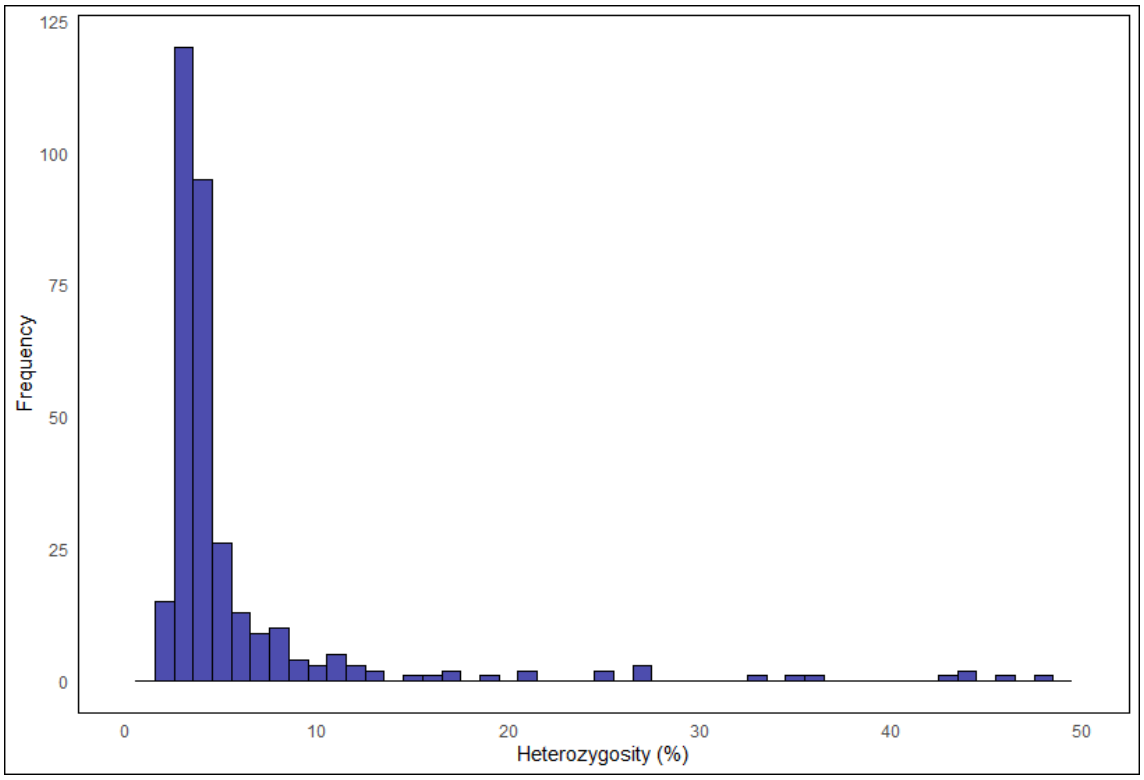

Figure 13.6 Residual heterozygosity rate (%) of the 326 S5 MAGIC eggplant lines, with an average of 5.81%, according to SNP marker data.

4.4 Marker-trait associations

Multiple SNP markers associated with prickle presence in eggplant were identified by the GWAS, this was proved by peaks in the Manhattan plots that surpassing the FDR and Bonferroni significance thresholds (Figures 4.7 to 4.9).

For PAB, SNP marker peaks exceeding the FDR significance threshold (LOD > 4.17) were not present, some lose SNPs were observed but were not part of a distinguishable peak, so they were dismissed. A peak on chromosome 6 surpassed the more stringent Bonferroni significance threshold (LOD > 6.08) (Figure 4.7). Reliable p-values with minimal inflation were suggested by the observed p-values that followed closely the diagonal line in the QQ plot (Figure 4.7).

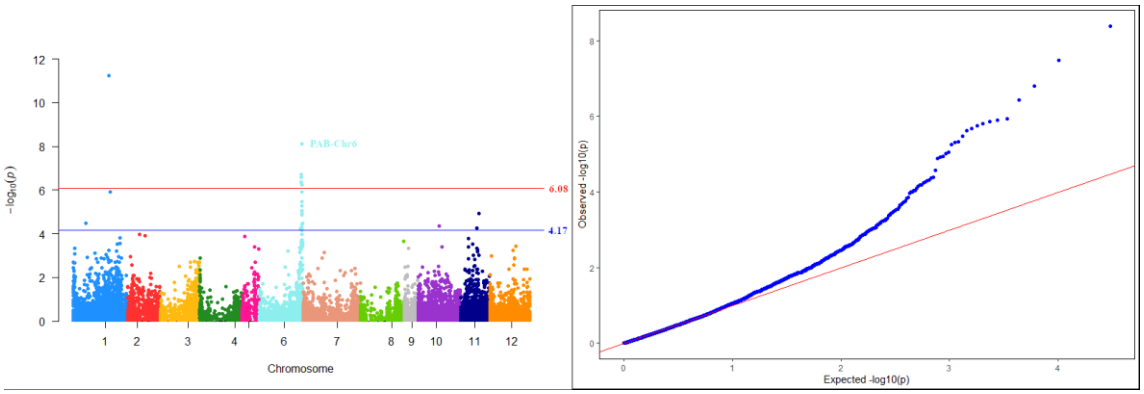

Figure 14.7 Manhattan (left) and QQ (right) plots for the genome-wide association mapping of PAB. The labelled SNPs were used to locate neighbouring prickle-related genes. Red and blue horizontal lines indicate Bonferroni and FDR significance thresholds, respectively.

Secondly, PAD had also no significant peaks above the FDR threshold (LOD > 3.95) with only the peak on chromosome 6 exceeding the Bonferroni threshold (LOD > 6.08) (Figure 4.8). There was a noticeable peak on chromosome 1, but as previously mentioned, the difficulty in being able to determine a clean peak ended up getting dismissed. Slight deviation from the diagonal line in the QQ plot, particularly at lower p-values, indicating minimal false-positive associations between SNP markers and the trait (Figure 4.8).

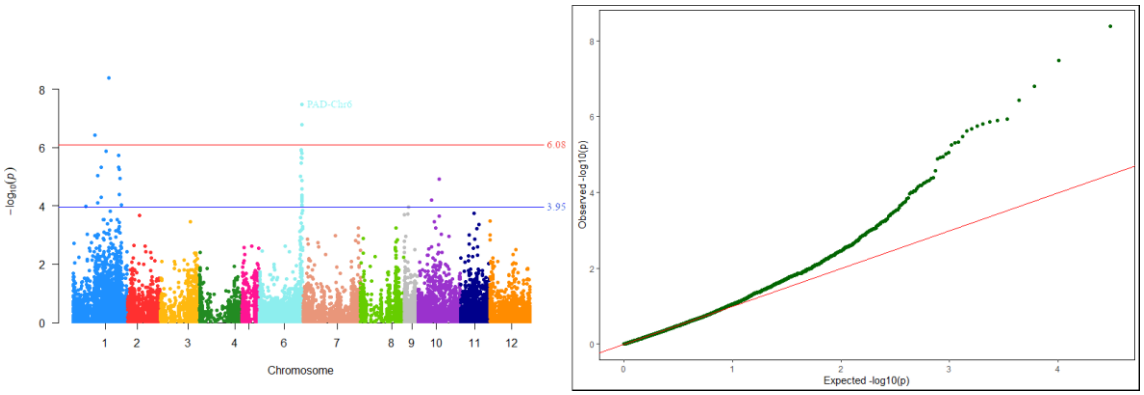

Figure 15.8 Manhattan (left) and QQ (right) plots for the genome-wide association mapping of PAD. The labelled SNPs were used to locate neighbouring prickle-related genes. Red and blue horizontal lines indicate Bonferroni and FDR significance thresholds, respectively.

Thirdly, for PC, no peak above the FDR threshold (LOD > 4.69) was observed on chromosome 6, with no other peaks exceeding the Bonferroni threshold either (LOD > 6.08) (Figure 4.9); this result is further discussed later on. 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 4.9).

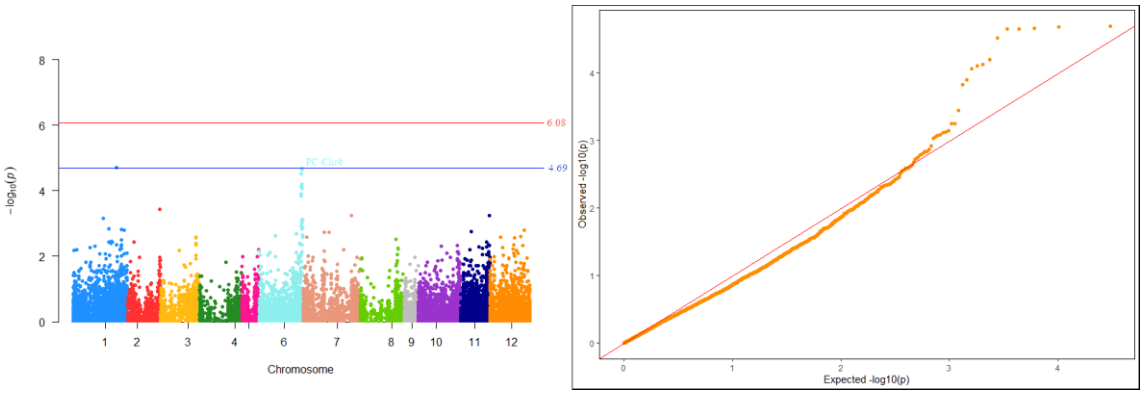

Figure 16.9 Manhattan (left) and QQ (right) plots for the genome-wide association mapping of PC. The labelled SNPs were used to locate neighbouring prickle-related genes. Red and blue horizontal lines indicate Bonferroni and FDR significance thresholds, respectively.

Lastly, PS peaks didn't manage to surpass the FDR threshold (LOD > 4.04) and only one peak from chromosome 6 exceeded the Bonferroni threshold (LOD > 6.08) (Figure 4.10). Once again, 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 4.10).

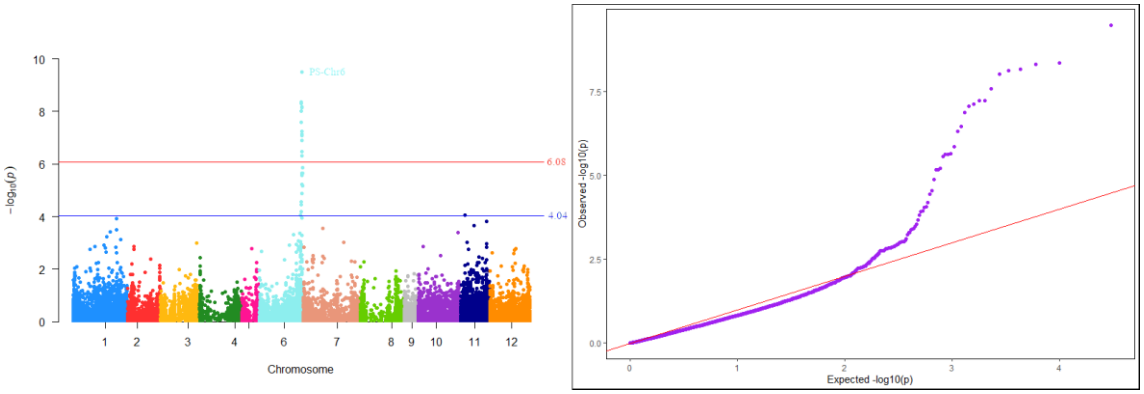

Figure 17.10 Manhattan (left) and QQ (right) plots for the genome-wide association mapping of PS. The labelled SNPs were used to locate neighbouring prickle-related genes. Red and blue horizontal lines indicate Bonferroni and FDR significance thresholds, respectively.

In this study SNPs exceeding the Bonferroni significance threshold (LOD > 6.08) were utilized to identify candidate genes, as suggested by QQ plots (Figures 4.7, 4.8 and 4.10). Additionally, any SNPs on chromosomes with a lack of nearby supportive SNPs to form a peak were excluded from candidate gene identification (Figure 4.7 to 4.10).

4.5 Candidate genes

In this study, candidate genes located near the lead SNP associated with prickle presence were recorded (Table 4.2). On chromosome 6, a gene similar to *LOG3* (SMEL_006g267050.1.01) was found in proximity to the lead SNP associated with three of the studied traits (PAB, PAD and PS). There were other genes located near the lead SNP in chromosome 6 like: *NUDT19*, *NEC3*, *LYK3* and *ABCG28*.

Table 18.2 Candidate genes identified near the lead SNPs associated with prickle presence in eggplant.

Chr	Lead SNP Position	Lead SNP Name[1]	Candidate Gene[2]		
			Name	Identifier	Position
6	105,464,056	PAB-Chr6 PAD-Chr6 PS-Chr6	*LOG3* *NUDT19* *NEC3* *LYK3* *ABCG28*	SMEL_006g267050.1.01 SMEL_006g267040.1.01 SMEL_006g267030.1.01 SMEL_006g267000.1.01 SMEL_006g266990.1.01	105,504,136 – 105,509,693 105,488,260 – 105,493,655 105,476,869 – 105,479,798 105,456,950 – 105,458,396 105,443,994 – 105,455,155
6	N/A[3]	PC-Chr6	N/A[3]	N/A[3]	N/A[3]

[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]Lead SNP below FDR threshold

5. Discussion

5.1 Multiparent mapping population

The inclusion of a wild relative enhances genetic variability within the population, which is essential for identifying quantitative trait loci (QTLs) linked to the trait of interest (Gramazio et al., 2020). Furthermore, the use of 326 S5 MAGIC eggplant lines in this research is crucial, as larger population sizes significantly boost the power and mapping resolution of GWAS (Collard et al., 2005; Jaganathan et al., 2020; Valdar et al., 2006). Furthermore, the MAGIC population used in this study underwent five generations of self-fertilization, resulting in the creation of an S5 population (Figure 4.2). This population followed the S3 population that was part of the Genome-Wide Association Study (GWAS) on prickle-related traits conducted by Mangino et al. (2022). Compared to germplasm panels and biparental populations, MAGIC populations provide superior genetic resources for identifying gene-trait associations due to their enhanced mapping resolution (Arrones et al., 2020; Scott et al., 2020). The S5 MAGIC eggplant population examined here was developed from eight founders known for their significant genetic and phenotypic diversity (Gramazio et al., 2017, 2019; Hurtado et al., 2014; Kaushik et al., 2018), which included a wild relative, S. incanum. The combination of multiple founders and several rounds of intercrossing and self-fertilization leads to an increased number of

recombinant events, thereby improving mapping accuracy (Scott et al., 2020). By applying a straightforward funnel method to intercross the diverse MAGIC founders, we observed considerable phenotypic diversity among the S5 MAGIC individuals, with varying levels of prickle presence (see Figure 4.2).

5.2 Genome-wide marker system

In this study, SWGR was employed to generate a well-distributed set of SNPs across the genomes of 326 S5 MAGIC eggplant lines, yielding a total of 15,176 SNPs with an average density of 13,70 SNPs/Mb (Figure 4.3; Table 4.1). This said, apart from the mapping population, the success of GWAS is significantly influenced by the genomic coverage of SNP markers (Ballesta et al., 2020). 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 (Rasheed et al., 2024). Imputing missing genotypic data is a widespread practice in genetic studies, including association mapping, to increase statistical power and refine mapping resolutions (Asimit & Zeggini, 2010; Balding, 2006; Marchini & Howie, 2010). So, 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). The high-quality SNP data generated by SWGR revealed low residual heterozygosity rate in the 326 S5 MAGIC eggplant lines, averaging 5.81% (Figure 4.6). This low heterozygosity rate can be caused by the self-pollinating nature of eggplants (Frary et al., 2007) and the five generations of self-fertilization employed to develop the population. Out of the 326 S5 MAGIC eggplant lines, only 15 showed a heterozygosity rate exceeding 20% (Figure 4.6), 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

The use of a multiparent population combined with well-distributed SNP marker data significantly improves the effectiveness of Genome-Wide Association Studies (GWAS). However, the selection of the statistical model is also crucial. These models help identify

meaningful associations between genetic markers and the traits of interest by employing p-values, while variance components for the traits are obtained through analysis of variance to assess allelic effects (Kulwal & Singh, 2021). In this study, a Mixed Linear Model (MLM) was utilized, which incorporated both population structure and kinship effects to reduce the likelihood of false associations between markers and traits (Pritchard et al., 2000). Numerous studies employing various models, including Generalized Linear Models (GLM) and MLM, consistently show a preference for the latter, demonstrating its superior effectiveness across different contexts (Bradbury et al., 2007; Kang et al., 2008; Pasam et al., 2012; Stich & Melchinger, 2009). The decision to use MLM was supported by Multidimensional Scaling (MDS) analysis of the SNP marker data, which indicated a somewhat level of population stratification within the S5 MAGIC eggplant population (Figure 4.4). Since population structure is a major confounding factor in association mapping studies, MLM's ability to statistically adjust for this structure is essential for correcting the inflation of minor genetic effects and addressing biases caused by stratification (Alseekh et al., 2021; Devlin & Roeder, 1999; Liu et al., 2016; Pasam & Sharma, 2014). The first three principal components explained a cumulative variance of 15% (Figure 4.5), emphasizing the importance of accounting for population structure to avoid misleading associations.

5.4 Marker-trait associations

This study integrated phenotype data on prickle presence in eggplant with single nucleotide polymorphism (SNP) marker data, resulting in the discovery of multiple marker-trait associations, represented as peaks of significant SNPs in Manhattan plots (Figures 4.7 to 4.10). The height of these peaks in the Manhattan plot indicates the strength of the association between neighbouring genomic regions and the trait (Kulwal & Singh, 2021). It is essential to set significant thresholds to distinguish true associations from false positives (Jain et al., 2024). This is particularly important in GWAS, which involves independently testing a large number of markers for associations, increasing the risk of false positive findings (Jain et al., 2024). Significant thresholds were determined using false discovery rate (FDR) and Bonferroni correction methods to tackle the issue of multiple testing. SNPs that surpass these thresholds are deemed to have a significant association with the trait, while quantile-quantile (QQ) plots were utilized more of like as a diagnostic tool to validate these significant associations. Additionally, QQ plots compare the observed distribution of p-values to their expected values under the null

hypothesis of no association (Jain et al., 2024). Since most tested SNPs are not associated with the trait, the QQ plot is anticipated to show observed p-values closely matching the expected line, particularly in the lower regions where marker-trait associations are generally absent (Jain et al., 2024). Based on the deviations noted in the QQ plots, either the FDR or Bonferroni significance threshold was selected to identify reliable associations between SNPs and the trait.

Although QQ plots for PAB, PAD, PC and PS showed that the observed p-values closely followed the diagonal line in the lower region, indicating good control over false positives (Figures 4.7 to 4.10), the Manhattan plots seemed to show some irregular peaks muffled by lose SNPs. This is likely due to the fact that a very restrictive filtering was used during the TASSEL analyses; MAF having a minor impact than thinning by sites over the number of SNPs left over after filtering. As a result, the only significant SNP peak on chromosomes 6 was associated with all four traits studied for PAB, PAD, PC and PS was considered the most reliable association signal (Figures 4.7 to 4.10).

5.5 Candidate regions

The development of prickles in plants is governed by a complex and conserved genetic network that has been studied across various species. Similar to the anthocyanin biosynthetic pathway, the regulation of prickle formation involves a coordinated interaction of multiple structural and regulatory genes. Research on prickle development has been particularly focused on solanaceous plants, where significant advances have been made in identifying the key genetic factors involved (Zhang et al., 2021, 2024). For this purpose, this study aims to identify the equivalent key genes in eggplant.

By focusing on the most reliable peaks identified by GWAS with the aid of the Bonferroni and FDR thresholds as well as of QQ plots, lead SNPs within these peaks were located and candidate genes in the vicinity of these lead SNPs were mapped (Table 4.2). In this study, only one reliable SNP peak was distinguishable on chromosome 6 of eggplant was identified: it was associated with three out of the four traits; PAB, PAD and PS (PAB-Chr6, PAD-Chr6, PS-Chr6) (Figures 4.7, 4.8 and 4.10). In the vicinity of this lead SNP, five different genes – *LOG3*, *NUDT19*, *NEC3*, *LYK3* and *ABCG28* were recorded (Table 4.2).

In the case of PC-Chr6, the peak did not reach both thresholds, although this is not erroneous data (Kaler & Purcell, 2019). This is because this SNP is responsible for the absence of prickles, therefore, since there was a mayor frequency of plants with prickles than without prickles for the PC trait (Table 4.2), this affected the peak's significance; although this can also be due to a variety of factors like a MAF (Low Minor Allele Frequency), population structure and environmental factors and polygenic nature of the trait (Cai et al., 2013; Kaler & Purcell, 2019).

Most importantly, the study successfully identified a gene corresponding to the *LOG3 Arabidopsis thaliana* like gene near the SNP on chromosome 6 (PAB-chr6, PAD-Chr6 and PS-Chr6). *LOG3* belongs to the *LOG* gene family, which encodes enzymes responsible for converting inactive cytokinin nucleotides into their active forms (Zhang et al., 2024b). Specifically, *LOG3* functions as a cytokinin riboside 5'-monophosphate phosphoribohydrolase, which catalyses the final step in the direct activation of cytokinins. This process involves the removal of the phosphate group from cytokinin nucleotides (such as cytokinin riboside 5'-monophosphate) to produce the free-base form of cytokinins, which are the biologically active molecules (L. Chen et al., 2022). The association of *LOG3* to affecting prickle presence on chromosome 6 in this study is consistent with findings from a previous association mapping study on eggplant prickle presence conducted by Satterlee et al. (2024).

In the other hand, this study also identified a gene similar to *NUDT19* from *A. thaliana* near the lead SNP, PAB-Chr6, PAD-Chr6 and PS-Chr6 (Table 4.2). Nudix Hydrolase 19 (*NUDT19*) in *Arabidopsis thaliana* is a chloroplast-localized enzyme that belongs to the Nudix hydrolase family. The Nudix hydrolases are characterized by their ability to catalyse the hydrolysis of a wide range of nucleotide diphosphate derivatives, which includes nucleotide sugars, nicotinamide adenine dinucleotide (NADH), coenzyme A (CoA) derivatives, and capped ribonucleic acid (RNA) structures (McLennan, 2006). Prickle development in plants, including eggplants, could be influenced by stress responses. Given that NUDT19 is involved in hydrolysing nucleotide derivatives (McLennan, 2006), which play a role in signalling and stress responses, it's possible that *NUDT19* or its homologs could indirectly influence prickle development under certain conditions. Stress conditions, like drought or pathogen attack, might alter nucleotide

levels and signalling pathways, leading to increased prickle formation as a defensive mechanism (Praveen et al., 2023).

Thirdly, a gene similar to *NEC3* in *Nicotiana langsdorffii* x *Nicotiana sanderae* was found near the lead SNP; PAB-Chr6, PAD-Chr6 and PS-Chr6 (Table 4.2). Bifunctional monodehydroascorbate reductase and carbonic anhydrase nectarin-3 (*NEC3*) belong to the *NEC* gene family, this family is associated with nectar production and secretion in plants and includes genes that encode proteins with various enzymatic activities, such as carbonic anhydrase and monodehydroascorbate reductase. These enzymes contribute to the protective and physiological roles in nectar, helping the plant defend against oxidative stress and maintain nectar quality (Silva et al., 2018). Prickle development is typically associated with structural and regulatory genes that control epidermal cell differentiation, and these are distinct from the genes involved in nectar production (Bagella et al., 2019; S. Li et al., 2024). However, the multifunctional nature of some plant proteins means that while NEC3 itself isn't directly linked to prickle formation, broader plant stress response pathways (in which NEC3 may be involved) could indirectly influence various plant traits, though this would be speculative without specific studies.

This study also identified a gene similar to *LYK3* in *A. thaliana* that was found near the lead SNP, PAB-Chr6, PAD-Chr6 and PS-Chr6 (Table 4.2). LysM domain receptor-like kinase 3 (*LYK3*) belongs to the LysM receptor-like kinase (LysM-RLK) family in plants. This family is known for its role in plant-microbe interactions, particularly in recognizing chitin and peptidoglycan, which are key components of fungal cell walls and bacterial cell walls, respectively (Buendia et al., 2018). LYK3 is especially important in the symbiotic relationship between legumes and rhizobia, where it plays a crucial role in the signalling pathway that leads to the formation of nitrogen-fixing nodules. Regarding its relation to prickle presence in eggplant, there is no established or direct connection between LYK3 and prickle formation (Buendia et al., 2018). As mentioned previously, prickle development is typically governed by different sets of genes, often related to epidermal cell differentiation and secondary metabolite pathways (Kellogg et al., 2011; S. Li et al., 2023). LYK3's primary role in symbiosis and pathogen recognition makes it unlikely to be directly involved in the genetic regulation of prickle formation in eggplant or other solanaceous plants.

Lastly, this study identified a gene similar to *ABCG28* in *A. thaliana* that was found near the lead SNP; PAB-Chr6, PAD-Chr6 and PS-Chr6 (Table 4.2). ABC transporter G family member 28 (*ABCG28*) belongs to the ABC (ATP-Binding Cassette) transporter superfamily, specifically the G subfamily. These transporters are membrane proteins involved in the translocation of various molecules across cellular membranes, using energy derived from ATP hydrolysis (Devi et al., 2024). ABCG transporters are known for their roles in transporting lipids, phytohormones, and secondary metabolites(Devi et al., 2024). Alas, there is no direct evidence or established link between *ABCG28* gene, and the development of prickles as consulted in previous literature. Furthermore, as previously mentioned prickle formation is more likely related to genes involved in epidermal cell differentiation and possibly those that regulate the plant's response to biotic and abiotic stresses, rather than to ABC transporters, which are generally more involved in the transport of substances across membranes; specifically regulating ROS status, being mostly expressed in the mature pollen grains and pollen tubes (Do et al., 2019).

6 Conclusions

Prickle presence varies quantitatively across different parts of the eggplant. So, recognizing the importance of this trait in eggplant breeding, this study tries to uncover the peak of the iceberg of the genetic basis of prickle presence regulation by performing a GWAS. By integrating the prickle presence data gathered form the diverse S5 MAGIC eggplant population with genome-wide SNP marker data generated by the SWGR technique, this study can: (1) highlight the SNP associated with prickle presence and (2) candidate genes putatively controlling prickle presence.

Firstly, only one SNP associated with prickle presence was found on chromosome 6. This highly reliable SNP peak was identified to be associated with PAB, PAD and PS (PAB-Chr6, PAD-Chr6 and PS-Chr6), with PC being excluded due to population structure and strict filtering affecting its significance. These marker-trait associations were identified using MLM, which accounts for population structure and kinship in association mapping.

Secondly, candidate genes putatively controlling prickle presence were found near the lead SNP. Using the eggplant reference genome, two regulatory genes (*LOG3* and *NUDT19*) that had been previously recorded in literature and had some interaction with the formation of pickles were found. Both genes being in on chromosome 6, may suggest a regulatory hotspot for prickle presence in eggplant and other *solanum* species.

Even though extensive research on prickle presence and regulation in a wide range of plant species has been made, this study provides little but still relevant additional evidence regarding the genetic basis of prickle presence in eggplant, while emphasizing the utilization of a diverse MAGIC population and dense SNP marker data in association mapping. The findings of this study offer valuable knowledge for future eggplant breeding efforts aimed at removing prickles and meeting current and future market demands for prick-less 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 2022 12:1*, *12*(1), 1–12. https://doi.org/10.1038/s41598-022-19858-2

Alam, I., & Salimullah, M. (2021). Genetic Engineering of Eggplant (Solanum melongena L.): Progress, Controversy and Potential. *Horticulturae 2021, Vol. 7, Page 78*, *7*(4), 78. https://doi.org/10.3390/HORTICULTURAE7040078

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 2020, Vol. 9, Page 229*, *9*(8), 229. https://doi.org/10.3390/BIOLOGY9080229

Arun Kumar, M. B., Dadlani, M., Kumar, R., & Jacob, S. R. (2014). Identification and validation of informative SSR markers suitable for ensuring the genetic purity of brinjal (Solanum melongena L.) hybrid seeds. *Scientia Horticulturae*, *171*, 95–100. https://doi.org/10.1016/J.SCIENTA.2014.03.034

Asimit, J., & Zeggini, E. (2010). Rare variant association analysis methods for complex traits. *Annual Review of Genetics*, *44*(Volume 44, 2010), 293–308. https://doi.org/10.1146/ANNUREV-GENET-102209-163421/CITE/REFWORKS

Aubriot, X., Knapp, S., Syfert, M. M., Poczai, P., & Buerki, S. (2018). Shedding new light on the origin and spread of the brinjal eggplant (Solanum melongena L.) and its wild relatives. *American Journal of Botany*, *105*(7), 1175–1187. https://doi.org/10.1002/AJB2.1133

Bagella, S., Filigheddu, R., Benesperi, R., Giordani, P., Minuto, L., Viciani, D., Caria, M. C., Pisanu, S., & Casazza, G. (2019). Thorn, spine and prickle patterns in the Italian flora. *Plant Biosystems - An International Journal Dealing with All Aspects of Plant Biology*, *153*(1), 118–133. https://doi.org/10.1080/11263504.2018.1474961

Balding, D. J. (2006). A tutorial on statistical methods for population association studies. *Nature Reviews Genetics 2006 7:10*, *7*(10), 781–791. https://doi.org/10.1038/nrg1916

Ballesta, P., Bush, D., Silva, F. F., & Mora, F. (2020). Genomic predictions using low-density SNP markers, pedigree and GWAS information: A case study with the non-model species Eucalyptus cladocalyx. *Plants*, *9*(1). https://doi.org/10.3390/plants9010099

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*(1), 289–300. https://doi.org/10.1111/J.2517-6161.1995.TB02031.X

Bhaduri, S. N. (1951). Influence of the numbers of Rhizobium supplied on the subsequent nodulation of the legume host plant. *Annals of Botany*, *15*(2), 209–218. https://doi.org/10.1093/oxfordjournals.aob.a083276

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*(19), 2633–2635. https://doi.org/10.1093/BIOINFORMATICS/BTM308

Breseghello, F., & Sorrells, M. E. (2006). Association Analysis as a Strategy for Improvement of Quantitative Traits in Plants. *Crop Science*, *46*(3), 1323–1330. https://doi.org/10.2135/CROPSCI2005.09-0305

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

Buendia, L., Girardin, A., Wang, T., Cottret, L., & Lefebvre, B. (2018). LysM receptor-like kinase and lysM receptor-like protein families: An update on phylogeny and functional characterization. *Frontiers in Plant Science*, *871*, 409605. https://doi.org/10.3389/FPLS.2018.01531/BIBTEX

Bush, W. S., & Moore, J. H. (2012). Chapter 11: Genome-Wide Association Studies. *PLOS Computational Biology*, *8*(12), e1002822. https://doi.org/10.1371/JOURNAL.PCBI.1002822

Cai, M., Dai, H., Qiu, Y., Zhao, Y., Zhang, R., Chu, M., Dai, J., Hu, Z., Shen, H., & Chen, F. (2013). SNP Set Association Analysis for Genome-Wide Association Studies. *PLOS ONE*, *8*(5), e62495. https://doi.org/10.1371/JOURNAL.PONE.0062495

Chang, L. Y., Toghiani, S., Ling, A., Aggrey, S. E., & Rekaya, R. (2018). High density marker panels, SNPs prioritizing and accuracy of genomic selection. *BMC Genetics*, *19*(1). https://doi.org/10.1186/s12863-017-0595-2

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*(Volume 70, 2019), 639–665. https://doi.org/10.1146/ANNUREV-ARPLANT-050718-100320/CITE/REFWORKS

Chen, L., Jameson, G. B., Guo, Y., Song, J., & Jameson, P. E. (2022). The LONELY GUY gene family: from mosses to wheat, the key to the formation of active cytokinins in plants. *Plant Biotechnology Journal*, *20*(4), 625–645. https://doi.org/10.1111/PBI.13783

Chen, M., Li, Z., He, X., Zhang, Z., Wang, D., Cui, L., Xie, M., Zhao, Z., Sun, Q., Wang, D., Dai, J., & Gong, D. (2024). Comparative transcriptome analysis reveals genes involved in trichome

development and metabolism in tobacco. *BMC Plant Biology*, *24*(1), 1–17. https://doi.org/10.1186/S12870-024-05265-4/FIGURES/9

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 2016 7:1*, *7*(1), 1–10. https://doi.org/10.1038/ncomms12767

Daunay, M. C., Lester, R. N., Dalmon, A., Ferri, M., Kapilima, W., Poveda Aguilar, M. M., & Jullian, E. (1998). *The use of wild genetic resources for eggplant (Solanum melongena breeding. 2. Crossability and fertility of interspecific hybrids*.

Devi, R., Goyal, P., Verma, B., Hussain, S., Chowdhary, F., Arora, P., & Gupta, S. (2024). A transcriptome-wide identification of ATP-binding cassette (ABC) transporters revealed participation of ABCB subfamily in abiotic stress management of Glycyrrhiza glabra L. *BMC Genomics 2024 25:1*, *25*(1), 1–20. https://doi.org/10.1186/S12864-024-10227-Z

Díez, M. J., Picó, B., & Nuez, F. (2002). *Cucurbit genetic resources in Europe. Ad hoc meeting held in Adana, Turkey, 19 January 2002.*

Do, T. H. T., Choi, H., Palmgren, M., Martinoia, E., Hwang, J. U., & Lee, Y. (2019). Arabidopsis ABCG28 is required for the apical accumulation of reactive oxygen species in growing pollen tubes. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(25), 12540–12549. https://doi.org/10.1073/PNAS.1902010116

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

Ersoz, E. S., Yu, J., & Buckler, E. S. (2007). Applications of linkage disequilibrium and association mapping in crop plants. In *Genomics-Assisted Crop Improvement: Vol. 1: Genomics Approaches and Platforms* (pp. 97–119). Springer.

FAOSTAT. (2022). https://www.fao.org/faostat/en/#home

Flint-Garcia, S. A., Thornsberry, J. M., & Edward IV, S. B. (2003). Structure of Linkage Disequilibrium in Plants. *Annual Review of Plant Biology*, *54*(Volume 54, 2003), 357–374. https://doi.org/10.1146/ANNUREV.ARPLANT.54.031902.134907/CITE/REFWORKS

Frary, A., Doganlar, S., & Daunay, M. C. (2007). Eggplant. In *Vegetables* (pp. 287–313). Springer.

Frary, A., Frary, A., Daunay, M. C., Huvenaars, K., Mank, R., & Doğanlar, S. (2014). QTL hotspots in eggplant (Solanum melongena) detected with a high resolution map and CIM analysis. *Euphytica*, *197*(2), 211–228. https://doi.org/10.1007/S10681-013-1060-6/TABLES/2

Garrison, E., & Marth, G. (2012). *Haplotype-based variant detection from short-read sequencing*. https://arxiv.org/abs/1207.3907v2

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*, 480262. https://doi.org/10.3389/FPLS.2019.01220/BIBTEX

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

Han, B., & Huang, X. (2013). Sequencing-based genome-wide association study in rice. *Current Opinion in Plant Biology*, *16*(2), 133–138. https://doi.org/10.1016/J.PBI.2013.03.006

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*(4), 787–795. https://doi.org/10.1007/S10722-013-0073-2/TABLES/1

Jain, P., Dutta, B., & Sevanthi, A. M. (2024). *Trait Based Association Mapping in Plants*. 159–190. https://doi.org/10.1007/978-981-99-6913-5_6

Juliana, P., Singh, R. P., Poland, J., Mondal, S., Crossa, J., Montesinos-López, O. A., Dreisigacker, S., Pérez-Rodríguez, P., Huerta-Espino, J., Crespo-Herrera, L., & Govindan, V. (2018). Prospects and Challenges of Applied Genomic Selection—A New Paradigm in Breeding for Grain Yield in Bread Wheat. *The Plant Genome*, *11*(3), 180017. https://doi.org/10.3835/PLANTGENOME2018.03.0017

Kaler, A. S., & Purcell, L. C. (2019). Estimation of a significance threshold for genome-wide association studies. *BMC Genomics*, *20*(1), 1–8. https://doi.org/10.1186/S12864-019-5992-7/TABLES/3

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

Kellogg, A. A., Branaman, T. J., Jones, N. M., Little, C. Z., & Swanson, J. D. (2011). Morphological studies of developing Rubus prickles suggest that they are modified glandular trichomes. *Https://Doi.Org/10.1139/B11-008*, *89*(4), 217–226. https://doi.org/10.1139/B11-008

Khadgi, A., & Weber, C. A. (2020). RNA-Seq Analysis of Prickled and Prickle-Free Epidermis Provides Insight into the Genetics of Prickle Development in Red Raspberry (Rubus ideaus L.). *Agronomy 2020, Vol. 10, Page 1904*, *10*(12), 1904. https://doi.org/10.3390/AGRONOMY10121904

Knapp, S., Aubriot, X., & Prohens, J. (2019). *Eggplant (Solanum melongena L.): Taxonomy and Relationships* (pp. 11–22). https://doi.org/10.1007/978-3-319-99208-2_2

Knapp, S., Chase, M. W., & Clarkson, J. J. (2004). Nomenclatural changes and a new sectional classification in Nicotiana (Solanaceae). *Taxon*, *53*(1), 73–82. https://doi.org/10.2307/4135490

Knapp, S., Vorontsova, M. S., & Prohens, J. (2013a). 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

Knapp, S., Vorontsova, M. S., & Prohens, J. (2013b). 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

Korte, A., & Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: A review. *Plant Methods*, *9*(1), 1–9. https://doi.org/10.1186/1746-4811-9-29/FIGURES/4

Kouassi, B., Prohens, J., Gramazio, P., Kouassi, A. B., Vilanova, S., Galán-Ávila, A., Herraiz, F. J., Kouassi, A., 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

Krishna, T. P. A., Veeramuthu, D., Maharajan, T., & Soosaimanickam, M. (2023). The Era of Plant Breeding: Conventional Breeding to Genomics-assisted Breeding for Crop Improvement. *Current Genomics*, *24*(1), 24. https://doi.org/10.2174/1389202924666230517115912

Kulwal, P. L., & Singh, R. (2021). Association Mapping in Plants. *Methods in Molecular Biology*, *2264*, 105–117. https://doi.org/10.1007/978-1-0716-1201-9_8/FIGURES/1

Kumari, R., Akhtar, S., Solankey, S. S., & Aakanksha. (2023). Variation in the Eggplant (Solanum melongena) Genotypes for Health-Promoting Bioactive Compounds and Agro-Morphological Traits. *Agricultural Research*, *12*(2), 143–153. https://doi.org/10.1007/s40003-022-00643-6

Lester. (1991). Origin and domestication of the brinjal eggplant, Solanum melongena, from S. incanum, in Africa and Asia. In *Solanaceae III: taxonomy, chemistry, evolution.* (pp. 369–387.). The Royal Botanic Gardens. https://worldveg.tind.io/record/17406

Li, H. (2013). *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM*. https://arxiv.org/abs/1303.3997v2

Li, S., He, Y., Li, D., Shi, S., Wang, Y., Tang, X., Ge, H., Liu, Y., & Chen, H. (2023). *Fine Mapping an AUXIN RESPONSE FACTOR, SmARF18, As a Candidate Gene of the PRICKLE LOCUS that Controls Prickle Absence/Presence on Various Organs in Eggplant (Solanum Melongena L.)*. https://doi.org/10.2139/SSRN.4374662

Li, S., He, Y., Li, D., Shi, S., Wang, Y., Tang, X., Ge, H., Liu, Y., & Chen, H. (2024). Fine mapping an AUXIN RESPONSE FACTOR, SmARF18, as a candidate gene of the PRICKLE LOCUS that controls prickle absence/presence on various organs in eggplant (Solanum melongena L.). *Scientia Horticulturae*, *327*. https://doi.org/10.1016/j.scienta.2024.112874

Mace, E. S., Lester, R. N., & Gebhardt, C. G. (1999). *AFLP analysis of genetic relationships among the cultivated eggplant, Solanum melongena L., and wild relatives (Solanaceae)*. Theoretical and Applied Genetics; Springer. https://doi.org/10.1007/S001220051277/METRICS

Marchini, J., & Howie, B. (2010). Genotype imputation for genome-wide association studies. *Nature Reviews Genetics 2010 11:7*, *11*(7), 499–511. https://doi.org/10.1038/nrg2796

Martin, F. W., Rhodes, A. M., Ortiz, M., & Díaz, F. (1981). Variation in Okra. *Euphytica*, *30*(3), 697–705. https://doi.org/10.1007/BF00038798/METRICS

Martin, J. A., & Wang, Z. (2011). Next-generation transcriptome assembly. *Nature Reviews Genetics 2011 12:10*, *12*(10), 671–682. https://doi.org/10.1038/nrg3068

McLennan, A. G. (2006). The Nudix hydrolase superfamily. *Cellular and Molecular Life Sciences*, *63*(2), 123–143. https://doi.org/10.1007/S00018-005-5386-7/METRICS

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

Miyatake, K., Saito, T., Nunome, T., Yamaguchi, H., Negoro, S., Ohyama, A., Wu, J., Katayose, Y., & Fukuoka, H. (2020). Fine mapping of a major locus representing the lack of prickles in eggplant revealed the availability of a 0.5-kb insertion/deletion for marker-assisted selection. *Breeding Science*, *70*(4), 438–448. https://doi.org/10.1270/JSBBS.20004

Pasam, R. K., & Sharma, R. (2014). Association mapping: A new paradigm for dissection of complex traits in crops. *Agricultural Bioinformatics*, *9788132218807*, 1–20. https://doi.org/10.1007/978-81-322-1880-7_1/TABLES/1

Portis, E., Barchi, L., Toppino, L., Lanteri, S., Acciarri, N., Felicioni, N., Fusari, F., Barbierato, V., Cericola, F., Valè, G., & Rotino, G. L. (2014). QTL Mapping in Eggplant Reveals Clusters of Yield-Related Loci and Orthology with the Tomato Genome. *PLOS ONE*, *9*(2), e89499. https://doi.org/10.1371/JOURNAL.PONE.0089499

Praveen, A., Dubey, S., Singh, S., & Sharma, V. K. (2023). Abiotic stress tolerance in plants: a fascinating action of defense mechanisms. *3 Biotech 2023 13:3*, *13*(3), 1–21. https://doi.org/10.1007/S13205-023-03519-W

Prohens, J., Gramazio, P., Plazas, M., Dempewolf, H., Kilian, B., Díez, M. J., Fita, A., Herraiz, F. J., Rodríguez-Burruezo, A., Soler, S., Knapp, S., & Vilanova, S. (2017). Introgressiomics: a new approach for using crop wild relatives in breeding for adaptation to climate change. *Euphytica*, *213*(7), 1–19. https://doi.org/10.1007/S10681-017-1938-9/METRICS

Qian, Z., Zhang, B., Chen, H., Lu, L., Duan, M., Zhou, J., Cui, Y., & Li, D. (2021). Identification of Quantitative Trait Loci Controlling the Development of Prickles in Eggplant by Genome Re-sequencing Analysis. *Frontiers in Plant Science*, *12*. https://doi.org/10.3389/fpls.2021.731079

Rakha, M., Prohens, J., Taher, D., Wu, T. hor, & Solberg, S. O. (2021). Eggplant (solanum melongena, S. aethiopicum and S. macrocarpon) breeding. *Advances in Plant Breeding Strategies: Vegetable Crops: Volume 9: Fruits and Young Shoots*, 163–203. https://doi.org/10.1007/978-3-030-66961-4_5/FIGURES/8

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

Rasheed, A., Xia, X., He, Z., & Rasheed, A. (2024). *Evolution in the Genotyping Platforms for Plant Breeding*. 65–78. https://doi.org/10.1007/978-981-99-4673-0_4

Rotino, G. L., Sala, T., & Toppino, L. (2014). Eggplant. *Alien Gene Transfer in Crop Plants*, *2*, 381–409. https://doi.org/10.1007/978-1-4614-9572-7_16/FIGURES/3

Saha, P., Singh, J., Bhanushree, N., Harisha, S. M., Tomar, B. S., & Rathinasabapathi, B. (2023). Eggplant (Solanum melongena L.) Nutritional and Health Promoting Phytochemicals. *Compendium of Crop Genome Designing for Nutraceuticals*, 1–32. https://doi.org/10.1007/978-981-19-3627-2_53-1

Sahito, J. H. ;, Zhang, H. ;, Gishkori, Z. G. N. ;, Ma, C. ;, Wang, Z. ;, Ding, D. ;, Zhang, X. ;, Tang, J., Martínez-Gómez, P., Hussain Sahito, J., Zhang, H., Ghulam, Z., Gishkori, N., Ma, C., Wang, Z., Ding, D., Zhang, X., & Tang, J. (2024). Advancements and Prospects of Genome-Wide Association Studies (GWAS) in Maize. *International Journal of Molecular Sciences 2024, Vol. 25, Page 1918*, *25*(3), 1918. https://doi.org/10.3390/IJMS25031918

Satterlee, J. W., Alonso, D., Gramazio, P., Jenike, K. M., He, J., Arrones, A., Villanueva, G., Plazas, M., Ramakrishnan, S., Benoit, M., Gentile, I., Hendelman, A., Shohat, H., Fitzgerald, B., Robitaille, G. M., Green, Y., Swartwood, K., Passalacqua, M. J., Gagnon, E., … Lippman, Z. B. (2024). Convergent evolution of plant prickles by repeated gene co-option over deep time. *Science*, *385*(6708). https://doi.org/10.1126/science.ado1663

Scheben, A., Batley, J., & Edwards, D. (2017). Genotyping-by-sequencing approaches to characterize crop genomes: choosing the right tool for the right application. *Plant Biotechnology Journal*, *15*(2), 149–161. https://doi.org/10.1111/PBI.12645

Silva, F. A., Guirgis, A., & Thornburg, R. (2018). Nectar analysis throughout the genus Nicotiana suggests conserved mechanisms of nectar production and biochemical action. *Frontiers in Plant Science*, *9*, 377503. https://doi.org/10.3389/FPLS.2018.01100/BIBTEX

Stern, S., Agra, M. de F., & Bohs, L. (2011). Molecular delimitation of clades within New World species of the “spiny solanums” (Solanum subg. Leptostemonum). *TAXON*, *60*(5), 1429–1441. https://doi.org/https://doi.org/10.1002/tax.605018

Storey, J. D., & Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences*, *100*(16), 9440–9445. https://doi.org/10.1073/PNAS.1530509100

Tibbs Cortes, L., Zhang, Z., & Yu, J. (2021). Status and prospects of genome-wide association studies in plants. *The Plant Genome*, *14*(1), e20077. https://doi.org/10.1002/TPG2.20077

VAVILOV, N. I. (1951). The Origin, Variation, Immunity and Breeding of Cultivated Plants. *Soil Science*, *72*(6). https://journals.lww.com/soilsci/fulltext/1951/12000/the_origin,_variation,_immunity_and_breeding_of.18.aspx

Vilanova, S., Alonso, D., Gramazio, P., Plazas, M., García-Fortea, E., Ferrante, P., Schmidt, M., Díez, M. J., Usadel, B., Giuliano, G., & Prohens, J. (2020). SILEX: A fast and inexpensive high-quality DNA extraction method suitable for multiple sequencing platforms and recalcitrant plant species. *Plant Methods*, *16*(1), 1–11. https://doi.org/10.1186/S13007-020-00652-Y/FIGURES/5

Vilanova, S., Hurtado, M., Cardona, A., Plazas, M., Gramazio, P., Herraiz, F. J., Andújar, I., & Prohens, J. (2014). Genetic diversity and relationships in local varieties of eggplant from different cultivar groups as assessed by genomic SSR markers. *Notulae Botanicae Horti Agrobotanici Cluj-Napoca*, *42*(1), 59–65. https://doi.org/10.15835/NBHA4219414

Wang, H., Guo, X., Pandey, M. K., Ji, X., Varshney, R. K., Nwosu, V., Guo, B., Wang, H., Guo, Á. X., Ji, Á. X., Guo, Á. B., Guo, X., Pandey, M. K., & Varshney, Á. R. K. (2017). *History and Impact of the International Peanut Genome Initiative: The Exciting Journey Toward Peanut Whole-Genome Sequencing*. 117–133. https://doi.org/10.1007/978-3-319-63935-2_8

Wei, Q., Wang, J., Wang, W., Hu, T., Hu, H., & Bao, C. (2020). A high-quality chromosome-level genome assembly reveals genetics for important traits in eggplant. *Horticulture Research*, *7*(1), 153. https://doi.org/10.1038/S41438-020-00391-0/41990331/41438_2020_ARTICLE_391.PDF

Yu, J., Pressoir, G., Briggs, W. H., Bi, I. V., Yamasaki, M., Doebley, J. F., McMullen, M. D., Gaut, B. S., Nielsen, D. M., Holland, J. B., Kresovich, S., & Buckler, E. S. (2005). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics 2005 38:2*, *38*(2), 203–208. https://doi.org/10.1038/ng1702

Yuan, Y., Liang, Y., Gao, L., Sun, R., Zheng, Y., & Li, D. (2015). Functional heterologous expression of a lysophosphatidic acid acyltransferase from coconut (Cocos nucifera L.) endosperm in Saccharomyces cerevisiae and Nicotiana tabacum. *Scientia Horticulturae*, *192*, 224–230. https://doi.org/10.1016/J.SCIENTA.2015.06.003

Zeven, A., & de Wet, J. (1982). *Dictionary of cultivated plants and their regions of diversity*. http://edepot.wur.nl/350203

Zhang, L., Sun, H., Xu, T., Shi, T., Li, Z., & Hou, W. (2021a). Comparative transcriptome analysis reveals key genes and pathways involved in prickle development in eggplant. *Genes*, *12*(3), 1–16. https://doi.org/10.3390/genes12030341

Zhang, L., Sun, H., Xu, T., Shi, T., Li, Z., & Hou, W. (2021b). Comparative Transcriptome Analysis Reveals Key Genes and Pathways Involved in Prickle Development in Eggplant. *Genes 2021, Vol. 12, Page 341*, *12*(3), 341. https://doi.org/10.3390/GENES12030341

Zhang, L., Zhang, R., Yan, P., Zeng, L., Zhao, W., Feng, H., Mu, R., & Hou, W. (2024a). PE (Prickly Eggplant) encoding a cytokinin-activating enzyme responsible for the formation of prickles in eggplant . *Horticulture Research*. https://doi.org/10.1093/hr/uhae134

Zhang, L., Zhang, R., Yan, P., Zeng, L., Zhao, W., Feng, H., Mu, R., & Hou, W. (2024b). PE (Prickly Eggplant) encoding a cytokinin-activating enzyme responsible for the formation of prickles in eggplant. *Horticulture Research*, *11*(7). https://doi.org/10.1093/HR/UHAE134

Zhu, C., Gore, M., Buckler, E. S., & Yu, J. (2008). Status and Prospects of Association Mapping in Plants. *The Plant Genome*, *1*(1). https://doi.org/10.3835/PLANTGENOME2008.02.0089