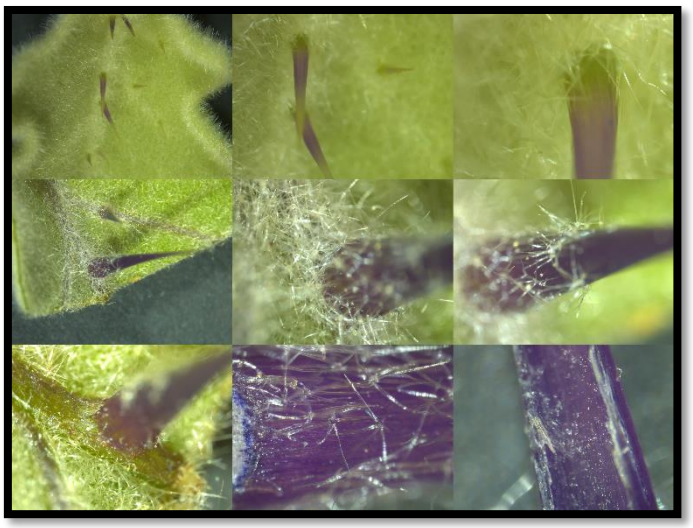
Universitat Politècnica de València

Escuela Técnica Superior de Ingeniería Agronómica y del Medio Natural (ETSIAMN)

Degree in Biotechnology

Genotyping and fine mapping of the prickle trait in a *S. melongena* introgression line



Final Bachelor Thesis

Academic year: 2019/2020

Sergio García Carbonell

Tutor: Santiago Vilanova Navarro

Experimental director: David Alonso Martín

Valencia, July 2020

Title: Genotyping and fine mapping of the prickle trait in a *S. melongena* introgression line

Abstract:

In this genotyping and fine mapping project of the prickle trait in a BC4 population of *a S. melongena* introgression line we will take as our starting point the data obtained in a previous tomato and eggplant genotyping project using Single Primer Enrichment Technology (SPET) in which it was observed that every single plant resulting from an advanced backcrossing between *S. melongena* and *S. insanum* showing introgressions in chromosome 6 had prickles in stem, leaves and calyx.

Thus, in this project, we aim to use around 600 BC4 seeds coming from two different fruits of the same BC3 plant to obtain two different populations: plants with prickles and plants without them. After performing DNA extractions from all the plants, specific SNP markers for the region of the chromosome 6 to which the prickle trait was associated are intended to be developed and this region will be gradually delimited by means of genotyping the plants using High Resolution melting (HRM) technology. Finally, we expect to obtain a series of candidate genes to be responsible for the prickle trait in eggplant so as to analyse them in detail and to be able to discard those that are not related to this trait. Then, we will focus the study in those genes that may be related to prickles.

A gene silencing experiment using CRISPR/Cas technology is also underway to precisely determine the gene controlling of the prickle trait in eggplant, which has not been identified yet.

Keywords: *S. melongena* ; introgression line ; prickle trait ; HRM ; genotyping ; fine mapping

Author: Sergio García Carbonell

Valencia, July 2020

Academic tutor: Santiago Vilanova Navarro, PhD

Experimental director: David Alonso Martín

Título: Genotipado y mapeo fino del carácter espinosidad en una línea de introgresión de *S. melongena*

Resumen:

En este proyecto de genotipado y mapeo fino del carácter espinosidad en una población BC4 de una línea de introgresión de *S. melongena* se partirá de la información obtenida en un proyecto previo de genotipado de tomate y berenjena mediante Single Primer Enrichment Technology (SPET) en el cual se observó que todas las plantas resultantes de los retrocruces avanzados entre *S. melongena* y *S. insanum* que presentaban introgresiones en el cromosoma 6 tenían espinas en tallo, hojas y cáliz.

Así pues, en este proyecto se propone utilizar unas 600 semillas BC4 provenientes de dos frutos distintos de una misma planta BC3 para obtener dos poblaciones distintas: plantas con espinas y plantas sin espinas. Después de realizar extracciones de ADN de todas las plantas se pretende desarrollar marcadores SNPs específicos para la región del cromosoma 6 a la que se asoció el carácter espinosidad y se irá acotando esta región genotipando las plantas mediante tecnología High Resolution Melting (HRM). Finalmente se espera obtener una serie de genes candidatos a ser responsables del carácter espinosidad en berenjena para analizarlos en detalle y poder descartar aquellos que no estén relacionados con este carácter. Luego centraremos el estudio en los genes que sí puedan estar relacionados con la espinosidad.

También está en marcha un experimento de silenciamiento génico mediante tecnología CRISPR/Cas de los genes candidatos para determinar de manera precisa el control genético del carácter espinosidad en berenjena, el cual no ha sido identificado hasta la fecha.

Palabras clave: *S. melongena* ; línea de introgresión ; caracter espinosidad ; HRM ; genotipado ; mapeo fino

Autor: Sergio García Carbonell

Valencia, July 2020

Tutor académico: Santiago Vilanova Navarro

Director experimental: David Alonso Martín

Acknowledgments:

A mi tutor, Santi, por ofrecerme participar en un proyecto tan interesante y ayudarme en todo lo que he necesitado.

A mis compañeros de laboratorio, por su amabilidad y el buen rollo que había siempre en el lab y, en especial, a David por haber estado conmigo en todo momento pese a estar metido en 158428 proyectos a la vez.

A mi familia, por todo el apoyo que me han dado durante todos estos años. En especial a mi madre, a mi madrina y a mi abuela que han sido y siempre serán mis pilares fundamentales

A Ele, por hacer mucho más llevaderas las horas y horas de redacción de este trabajo, y por alguna cosa más.

TABLE OF CONTENTS

FIGURES LIST

TABLES LIST

1. INTRODUCTION

1.1. IMPORTANCE OF CULTIVATED EGGPLANT

Common eggplant (*Solanum melongena* L., 2n = 2x = 24), also referred to as aubergine or brinjal, is part of the *Solanaceae* family, which includes many species of agricultural importance (Knapp et al., 2013). This is one of the most abundant genera in terms of plants with flowers since it includes more than 1400 species inhabiting all continents but the Antarctica and almost every habitat from desserts to mountainous regions (Knapp et al., 2013).

There are three main cultivated species of eggplant which are *Solanum melongena* L., *Solanum macrocarpon* L. (also called Gboma eggplant) and *Solanum aethiopicum* L. (known as the scarlett eggplant). The last two species are important primarily in Africa (Daunay and Hazra, 2012) whilst *Solanum melongena* is a major vegetable crop in Europe and Asia (Frary et al., 2007), especially in tropical and subtropical areas.

Cultivated eggplant is considered the third most important *Solanacea* crop only behind *Solanum lycopersicum* L. (tomato) and *Solanum tuberosum* (potato) according to Knapp et al., (2013) and, in terms of production, it is the fifth most important vegetable (FAOSTAT,2018), being Spain and Italy the main European producers (Figure 1). The consumption of eggplant is highly beneficial due to its high contents of fiber, micronutrients and bioactive compounds, which are mainly phenolics (Menella et al., 2012). It is for that reason a highly nutritious crop that constitutes an essential part of the Mediterranean diet and a very important vegetable in most Asiatic countries.

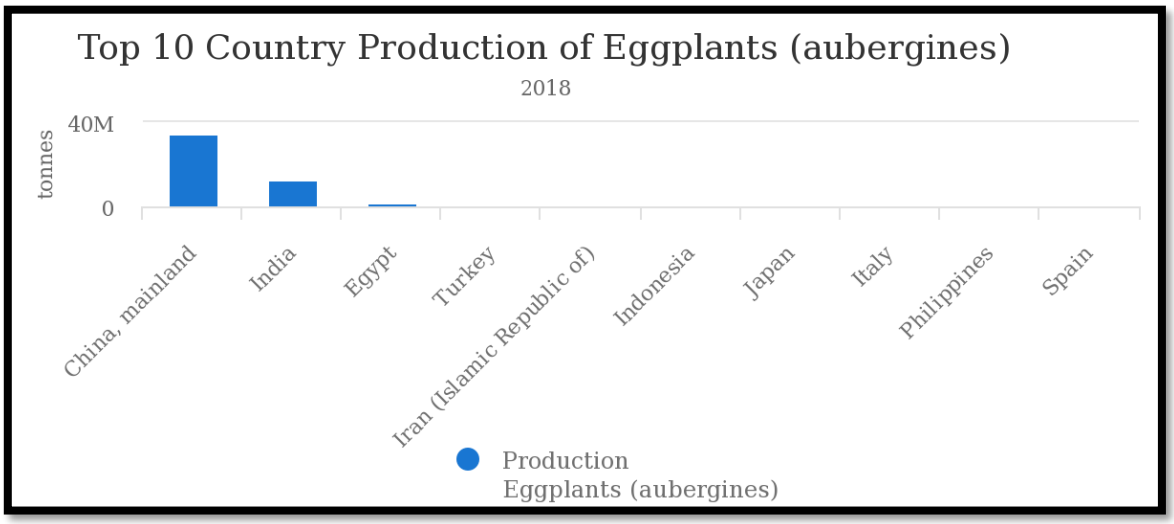


Figure 1. Top ten eggplant producing countries worldwide. Retrieved from *http://www.fao.org/faostat/en/#home.*

Despite the fact that common eggplant has a great phenotypic diversity, its genetic base is quite narrow (Muñoz-Falcón et al., 2009), mainly due to the genetic bottlenecks undergone during its domestication (Meyer *et al*., 2012).

Even though the common eggplant shares many of the breeding goals relevant to most crops such as yield and resistance to biotic and abiotic stresses it has some traits that are highly specific to eggplant and that are to be avoided at all cost as they are not interesting neither for the producers nor for the consumers, in particular fruit bitterness and leaf and calyx prickliness are the most important ones

(Portis et al., 2015), being this last one especially concerning since it has severe implications in what refers to the handling of the product.

1.2. PRICKLINESS IN EGGPLANT AND OTHER CROP SPECIES

S. melongena belongs to the clade of *Leptostemonum* which is known as the "spiny solanums" since most species possess sharp prickles on the stems and leaves (Weese and Bohs, 2007). This is the largest monophyletic group within the genus Solanum (Bohs, 2005) and it is characterized for having stellate trichomes and long attenuate anthers (Figure 2).

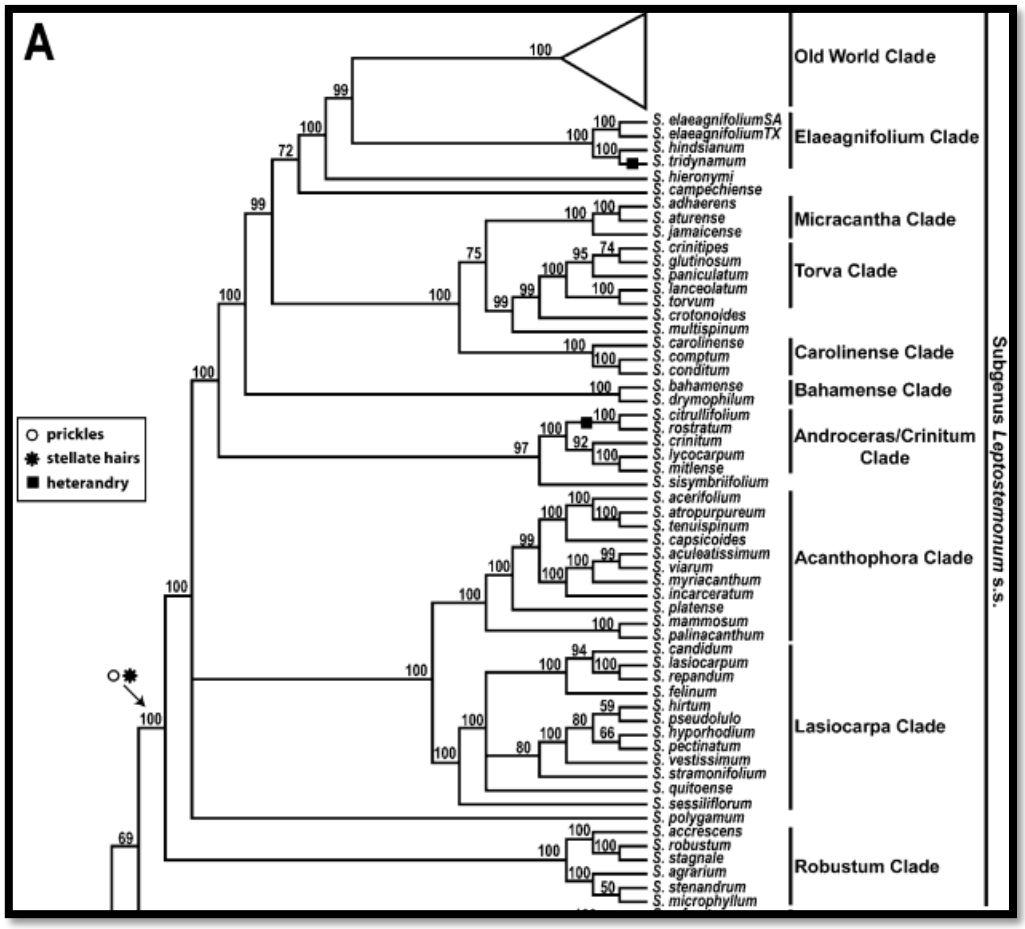


Figure 2. Phylogenetic tree of the *Leptostemonun* clade. Retrieved from Levin et al., (2006).

Prickles are defined as outgrowths of epidermal tissues and while most authors state that they are modified glandular trichomes (Kellog et al., 2011; Pandey et al., 2018) some others assure that they are modified leaves (Björkman and Anderson 1990) so there is still some uncertainty regarding its origin. Although they can be sometimes confused and used as synonyms, prickles, thorns and spines are not the same and they have very different morphologies. Prickles can be classified as modified epidermal cells whereas spines are usually referred to as modified leaves and thorns as modified branches or stems. According to this distinction, both thorns and spines have vascular bundles inside whilst prickles have not (Simpson, 2010). The main implication of this difference is that whereas thorns and spines are complex structures with different tissue layers, prickles are much simpler and easier to study.

Prickliness is an undesirable trait in eggplant and other crop species as prickles difficult the handling of the crop and may cause wounds to field workers and, when present in the calyx, to retailers and consumers. In addition, leaf and calyx prickles may cause damage to the fruits and deteriorate them what makes them an important source of rotting to take into consideration.

It has been commonly asserted that prickles are an adaptation to heavy grazing (Symon, 1986) and they are present in different varieties, landraces and wild species related to cultivated eggplant. Prickles can be found in several plant organs being stem, leaf, pedicle and calyx the most common ones.

Due to the fact that prickliness is a very unwanted trait as it makes it very difficult handling the products, the production of prickleless varieties is a prime objective in eggplant breeding projects. To achieve that and for an early selection of the most promising prickleless varieties through molecular marker-assisted selection (MAS) it is essential to determine the genetic control of this trait and to develop molecular markers linked to it.

Up to this date, several studies have determined that major quantitative trait loci (QTLs) related to prickliness are present in chromosome 6 (Frary et al., 2014; Gramazio et al., 2014, Portis et al., 2015). Nevertheless, no markers tightly linked to this gene/QTL have been reported.

The development of genotyped and phenotyped advanced backcross (AB) populations segregating for this trait has been of great utility to design markers and identify the gene/s controlling the character.

The characterization of this trait is an attractive objective in plant breeding since the great availability of crop wild relatives for the eggplant will allow to greatly improve this crop in terms of handling and production with even more precision and accuracy and in a much faster fashion once its genetic control is completely understood.

1.3. UTILIZATION OF CROP WILD RELATIVES (CWRs) FOR PLANT BREEDING

CWRs are plant species that have an indirect use derived from its relatively close genetic relationship to a crop (Maxted, 2006). A more common definition was given by Hajjar and Hodgkin (2007) as they stated that CWRs are a group of plants that contain the crop's progenitor as well as other species that are phylogenetically closer or farer related than the former.

The main use of CWRs is to improve crops performance as they offer a great source of genetic diversity to be exploited in plant breeding to develop, for instance, better adapted crops to harsh environmental conditions (Hajjar and Hodgkin, 2007). Their usefulness relies on their relatedness, since the genetic variations and the broad pool of potentially useful genes they offer can be easily used for plant breeding (Tanksley and McCouch, 1997).

Another important characteristic of CWRs is that they allow the study of QTLs and candidate genes with much more accuracy than with other kinds of populations resulting from interspecific crosses.

CWRs have been used in plant breeding for more than 60 years improving crop varieties by means of introducing favourable traits. The only major disadvantage these resources may present is the introduction of undesirable agronomic traits and this has limited their usefulness in interspecific

cultivars. To avoid this issue, it is essential to carry out previous pre-breeding activities so as to reduce the undesired genetic background of the donor genome as much as possible (Longin and Reif, 2014).

CWRs utilization as genetic resources in eggplant breeding has been limited to provide better adaptation to biotic and abiotic stresses related to climate change (Rotino et al., 2014). However, CWRs remain largely unexploited in eggplant breeding in many different fields mainly due to the lack of research so at the present time efforts should be addressed towards the development of new varieties through plant breeding that take advantage of the genetic diversity provided by CWRs.

To this end, advanced backcross (ABs) and introgression lines (ILs) are powerful and available tools to develop pre-breeding materials containing genetic resources from CWRs in the genetic background of the conventional cultivated crop plants. Therefore, the development of ABs and ILs in the eggplant genepool would allow to improve the use of CWRs in eggplant breeding (Prohens et al., 2017).

1.4. ADVANCED BACKCROSS (AB) AND INTROGRESSION LINES (ILs)

AB and IL populations are important tools that allow to speed up genetic and genomic studies and they can easily improve the development of breeding programs on almost any crop by means of dissecting the genetics of agronomic traits. They can provide breeders with a lot of relevant information and, on top of it, they can be directly introduced into breeding pipelines (Zamir, 2001).

AB populations are obtained through repeated backcrossing of an initial interspecific hybrid derived from the cross between a recurrent and a donor parent. They contain one or more introgressions that can be fixed or not from the donor parent in the genetic background of the recurrent parent (Fulton et al., 1997). Thus, introgressions are genomic fragments derived from a donor parent, which is commonly a CWR, that are integrated into the genome of the recurrent parent, the cultivated crop species.

The big advantage that AB material represents with regard to other populations is that with every new round of backcrossing the proportion of the donor parent genome is reduced by 50%. Therefore, it provides a more powerful tool in terms of precise mapping of genes and QTLs than other materials such as F2, BC1 or recombinant inbred lines (RILs) and they constitute an excellent material for plant breeding (Tanksley and Nelson, 1996).

Moreover, IL populations are a special type of ABs that have single and fixed introgressed fragments (Eshed and Zamir, 1995) in which every single line contains a different and specific genomic fragment inside the genetic background of the recipient cultivated parent. All together they represent the entire genome from the donor parent with overlapping fragments (Zamir, 2001; Eduardo et al., 2005).

After several rounds of backcrossing, the introgressed fragments are fixed through self-pollination and that makes them immortal lines (Zamir, 2001), hence they are even more suitable than ABs to be directly introduced into a breeding pipeline. As a matter of fact, it has been proved by several studies that ILs show a higher efficiency in QTL estimation and fine mapping than other types of segregating populations (Eshed and Zamir, 1995; Alonso-Blanco et al., 2006; Yin et al., 2016) since the differences that can arise between the phenotype of the IL and the recurrent parent can only be due to the genes included within the introgressed fragment of the donor genome (Zamir, 2001).

IL populations can be further improved becoming sub-ILs, which are ILs with even shorter introgressed fragments. They can be achieved by performing more rounds of backcrossing and self-pollination,

therefore shortening the introgressions. Sub-ILs allow to increase the resolution of linkage maps and to find new markers that are even more closely associated to the studied genes or QTLs (Sacco et al., 2013).

Currently, the availability of IL materials is still limited although in the last decade great advances have been made. In the case of eggplant, the first AB and IL population that cover the whole genome was developed recently (Gramazio, 2017) and these new resources will allow to confirm the presence of existing QTLs in a faster and more precisely way together with the advantages that provide marker-assisted selection (MAS) techniques.

1.5. MARKER-ASSISTED SELECTION (MAS)

MAS is known in plant breeding as a very useful technique for selecting phenotypes of interest based on molecular markers. It involves the establishment of a tight linkage between the molecular marker and the chromosomal location of the gene/QTL associated to the trait of interest (Talukdar, 2013). Therefore, it is essential that the DNA markers are closely linked to the gene(s) governing the trait of interest that is going to be selected, that is, the favourable phenotype (Collard and Mackill, 2008). The identification of genes and QTLs governing the characters of interest usually relies on genetic linkage analysis and on the construction of linkage maps (Collard et al., 2005). Genetic linkage analysis is based on the differences caused during meiosis by the genetic recombination process (Tanksley, 1993) and the construction of linkage maps depends on the detected markers inside a population. The distances between markers and genes or QTLs can be measured in a physical magnitude but they are usually evaluated as genetic distances using centiMorgan (cM) as the measuring unit, since it is more representative. These distances depict the recombination frequency between two markers or points in the genome (Peters et al., 2003). Linkage maps are key in order to detect and analyse QTLs (Collard and Mackill, 2008).

Thus, MAS can greatly improve the speed and accuracy of the selection during AB and IL development due to the enormous potential of DNA markers for conventional plant breeding (Collard and Mackill, 2008). A lot of different genetic markers have been used for MAS and plant breeding in the past although nowadays DNA or molecular markers are the most common and used ones (Collard et al., 2005). In the early days of IL development amplified fragment length polymorphism markers (AFLPs), conserved ortholog set II markers (COSIIs) and simple sequence repeat markers (SSRs) were the most utilized albeit they have been replaced in their vast majority by single nucleotide polymorphism (SNPs) markers in the last few years due to the fact that these are easier to develop and much more common than the rest of variations.

Thanks to the next generation sequencing (NGS) techniques, the number of detected SNPs in most species have greatly increased. With the vast number of identified SNPs available it is now possible to use this information combined with high-throughput genotyping technologies in order to screen and select the plants for the next generation during the construction of an AB or IL population with much more efficiency (Ganal et al., 2014). Beyond that, due to the fact that sequencing costs are rapidly declining, new approaches and better genotyping platforms are being developed like genotyping-by-sequencing (GBS) or Single Primer Enrichment Technology (SPET). The former is able to develop novel SNPs in segregating populations at the same time that it genotypes the plants that constitute the population (Elshire et al., 2011). This platform presents several advantages such as the reduced cost and its time-effectiveness, as well as the increased marker density (Sukumaran and Yu, 2014). The

latter is a targeting sequencing method capable of discovering polymorphisms de novo and it greatly reduces the ascertainment bias.

1.6. GENOTYPING METHODOLOGIES

Genotyping is the process of determining the DNA sequence at a specific position within a gene or an individual. This process can determine sequence variations by means of comparing a DNA sequence to others from different samples or to a reference sequence. These sequence variations (the genotype) can be used as markers in linkage and association analysis to identify and determine genes relevant to specific traits of interest.

Genotyping was stablished as a field of study on its own by the time molecular markers were developed since previously there were only morphological and biochemical markers which had many limitations and drawbacks compared to DNA markers since only molecular markers can be analysed in the early stages of crop development, can be extracted from almost any tissue and usually show very high levels of polymorphism (Paterson et al., 1991). Therefore, DNA markers allowed to surpass the limitations that other types of markers had permitting the creation of molecular profiles for each species and individuals independent from growth conditions (Morell et al., 1995). Nevertheless, for a molecular marker to be useful it needs to meet some requirements like being polymorphic, having co-dominance inheritance, having frequent and uniform distribution throughout the genome and many more. Since there is no marker that fulfils all the requirements several different genotyping techniques have been developed based on different kinds of molecular markers. All of them have some advantages and drawbacks although in the last years SNP markers have stablished themselves as the most utilized ones.

1.6.1. SINGLE NUCLEOTIDE POLYMORPHISM (SNP)

SNPs are point mutations of single nucleotides at specific positions along the genome and they are the most common type of sequence variant. Typically, SNP markers have two different alleles and thus they are biallelic markers (Kruglyak, 1997). The first genotyping technology based on these molecular markers was proposed by Lander (1996) who stated that sequence polymorphisms derived from SNPs could include single base insertions, deletions, transitions and transversions, being the transitions the most common ones (Zhao and Boerwinkle, 2002).

Nowadays, SNP markers have stablished themselves as one of the preferred genotyping technologies due to the fact that they are abundant and very genetically stable throughout the genome. Apart from that, genotyping at a nucleotide level has a great advantage since it removes a lot of noise associated to other methodologies. They also present other advantages if compared to other technologies like their abundance along the genome, their high accuracy and repeatability and, the most important one, the allowance for implementing high-throughput genotyping (Tsuchihashi and Dracopoli, 2002). This technique also presents a major disadvantage which is the extremely low level of information that can provide a single SNP due to its low polymorphism unless It is associated with an important phenotype but it can be solved by means of analysing many of them simultaneously (Werner et al., 2002).

Therefore, SNP genotyping technologies are currently the most utilized ones and so many different platforms based on them have arisen. These methodologies can be classified according to their throughput depending on the number of samples and SNPs they can analyse at the same time. Low and medium throughput methodologies can only analyse a few hundreds of SNPs in one or a few

samples whereas high throughput technologies can handle more than 10 000 SNPs simultaneously and more than 100 samples in one run.

1.6.1.1. LOW-MEDIUM THROUGHPUT SNP GENOTYPING TECHNOLOGIES

Low-medium throughput genotyping technologies include those techniques that can only analyse a few samples at the same time and/or those that are not able to detect a large number of differences in a DNA sequence in one run. There are many platforms that allows this type of analysis but we have chosen for our experiment High resolution melting (HRM) since it is a very reliable technology.

1.6.1.1.1. High Resolution Melting (HRM)

HRM is a genotyping technique developed by Idaho Technology and the university of Utah in the 1990s consisting on the quantitative analysis of the melt curves of DNA fragments, usually following PCR amplification. It is based in the ability of quantitative PCR (qPCR) to detect polymorphisms analysing DNA melting curves and it represents a huge amelioration due to its great sensitivity since it is capable of detecting differences regarding to a single nucleotide (Willhelm and Pingoud, 2003).

The process is quite simple and it begins with a PCR in which the region of DNA that is being studied is amplified in the presence of a fluorescent dye capable of binding to double-stranded DNA (dsDNA) but not to single-stranded DNA (ssDNA). Once the PCR has finished, the DNA, which at this moment is mostly dsDNA is slowly denatured by means of increasing the temperature and thus, the fluorescent dye is gradually released. Each melting curve will be unique since it depends on the length of the DNA sequence, its composition and the complementarity between the chains (Wittwer et al., 2003).

The big advantage of HRM is its fast identification of variants in the region of interest that is amplified without the need of sequencing (Wittwer, 2009). For that reason, it has been widely used for mapping genes and analysing molecular markers which could be used in MAS and cultivars development. Besides that, another asset HRM has is the ease with which heterozygous and homozygous genotypes can be differentiated. The most notable shortcoming of HRM is the fact that you can detect differences between DNA sequences but you cannot identify neither the number and the position of those differences unless you sequence them (Chateigner-Boutin and Small, 2007). Another small drawback HRM has is that it loses sensitivity when the fragment size exceeds 200 bp. Also, concentration and the quality of DNA of all the studied samples has to be highly similar for the technique to be reliable and to prevent misclassifying the samples' melting curves (Croxford et al., 2008).

1.6.1.2. HIGH THROUGHPUT SNP GENOTYPING TECHNOLOGIES

High throughput genotyping technologies, on the other hand, will include those genotyping methodologies that are able to cover great portions of the genome in one analysis and/or those techniques that allow the analysis of hundreds of samples simultaneously. Here we will discuss GBS, which is one of the most utilized platforms for genome-wide association studies (GWAS), and SPET, another technology which was recently developed by NuGEN® and that is based on amplicon sequencing by NGS.

1.6.1.2.1. Genotyping by Sequencing (GBS)

GBS was first described by Elshire et al., (2011) and it is a modern genotyping technique based on next generation sequencing (NGS) capable of generating large-scale genomic data. The big improvement

this technology implements is that it allows the discovery of novel SNP markers without requiring previous information about the genome of interest (Scheben et al., 2017). It is a very fast and cost-effective method that represents one of the best high-throughput genotyping approaches at the present time (Essubalew et al., 2019). GBS has been employed in plant breeding in order to build high density genetic maps and to identify QTLs related to many traits of interest in all kinds cultivated crops and not only in model species (He et al., 2014) since it can work even in the absence of a reference genome sequence and due to the fact that it can develop novel SNPs in segregating populations at the same time it genotypes the plants of these populations (Elshire et al., 2011). The big difference this method has in comparison to other genotyping techniques based on sequencing is that in GBS one or more enzymes are used to cut up the DNA in shorter fragments before sequencing and thus, the complexity of the process is greatly reduced.

GBS represents a huge advancement in what refers to genotyping technologies since it reduces the PCR purification steps and the costs of the process (Bhat et al., 2016). Furthermore, GBS is a tool specifically designed to perform GWAS and MAS programs and so it is a very powerful resource in the field of plant breeding.

Despite the above-mentioned advantages of GBS it also has some limitations such as the analysis of polyploid genomes or the amount of missing data when performed at low coverage. It also has the same disadvantage all methods that use restriction enzymes to reduce genome complexity have which are the possible mutations at restriction sites.

1.6.1.2.2. Single Primer Enrichment Technology (SPET)

SPET, developed by IGATech, Udine, Italy, in collaboration with NuGEN®, represents an alternative to GBS based on target sequencing and genotyping methods with the additional feature of being able to discover de novo polymorphisms. SPET greatly reduces ascertainment bias and it allows higher multiplexing levels than PCR-based methods (Barchi et al., 2019a).

This robust, affordable high-throughput platform has a high degree of cross-transferability between crops and CWRs and thus it has the capability to make ILs development much more efficient in terms of time and cost by enhancing the selection process.

In order to work in an appropriate way SPET requires a priori genomic data and the detection of SNPs for the probe design. Then the probes are engineered in a way that they are adjacent to a region that contains a SNP and so they are able to detect that SNP together with others that may be present in the surrounding area (Barchi et al., 2019a).

Unlike GBS which is an open system with no ascertainment bias that detects polymorphisms in a random fashion, SPET is an array-based method that depends on an original catalogue of target SNPs and is capable of effectively targeting specific genomic regions. This stands out an important difference since SPET technology allows the detection of valuable markers in complex and repetitive genomes in a more efficient way than other genotyping technologies (Scaglione et al., 2019).

This platform allows the realization of cost-effective targeted genotyping projects like array-based methods at the same time that it detects new polymorphisms in the same way random complexity reduction systems such as GBS do (Scaglione et al., 2019). It is, therefore, a genotyping methodology strongly recommended to undergo large-scale genotyping projects in almost any species.

1.7. FINE MAPPING

Fine mapping is the process by which a region of the genome that has been associated to a particular trait is analysed in detail in order to detect the genes or QTLs that are more likely to be causing the phenotype of interest (Schaid et al., 2018).

The fine mapping approach is usually applied when studying complex traits and it tends to follow a genotyping project in which the trait of interest is linked to a specific genomic region wide enough to contain hundreds of genes. So as to be able to study in detail the phenotype of interest it is important to determine which particular gene is causing it and, for that purpose, fine mapping is the best option since it is basically a more specific genotyping approach for which hundreds to thousands of markers for a particular region of the genome are designed in order to further shorten that genomic region. In the end, when a few candidate genes are identified other experiments are carried out to determine which one is responsible for the trait.

In order to efficiently fine map a character of interest it is strongly recommended to develop first an AB or an IL so as to create a suitable population for the process due to the fact that in those lines you can have a precise control over the introgressed genome regions that reduces considerably the work to be done and, especially, because by doing so more plants will have the chance of having undergone a recombination event that once analysed will allow the shortening of the region of interest (informative plants).

1.8. BACKGROUND OF THE PROJECT

All the above-mentioned methods and technologies have been applied in the last two decades in order to identify QTLs in eggplant and some advancements regarding prickliness have been made because of them. Doganlar et al., (2002) was the first to locate a major QTL related to prickliness in chromosome 6 and some minor ones in chromosomes 9 and 11 and he was followed by Frary et al., (2014) who also positioned a major QTL governing prickliness in chromosome 6 although he also found other QTLs related to the prickle trait in other chromosomes. The most relevant discovery was made by Gramazio et al., (2014) who found a candidate gene coding for prickliness on linkage group six at 100.1 cM.

It was not until recently that an AB population between *S. incanum* and *S. melongena* was developed (Gramazio et al., 2017) and with these new genetic resources and very similar methods Haller (2018) produced a new IL population between *S. insanum* and *S. melongena*. After having developed SNP markers throughout the whole genome and basing the study on all the existing information regarding QTLs in eggplant paying special attention to the ones related to prickliness, the 90 plants from the BC3 population developed by Haller (2018) were phenotyped and genotyped using SPET technology and it was checked that all 18 plants that showed an introgression on chromosome 6 across the region of about 100.1 cM had prickles on leaves, stem and calyx. Besides that, they were able to stablish the genomic region governing the prickle trait by around 104 and 107 Mb (Figure 3).

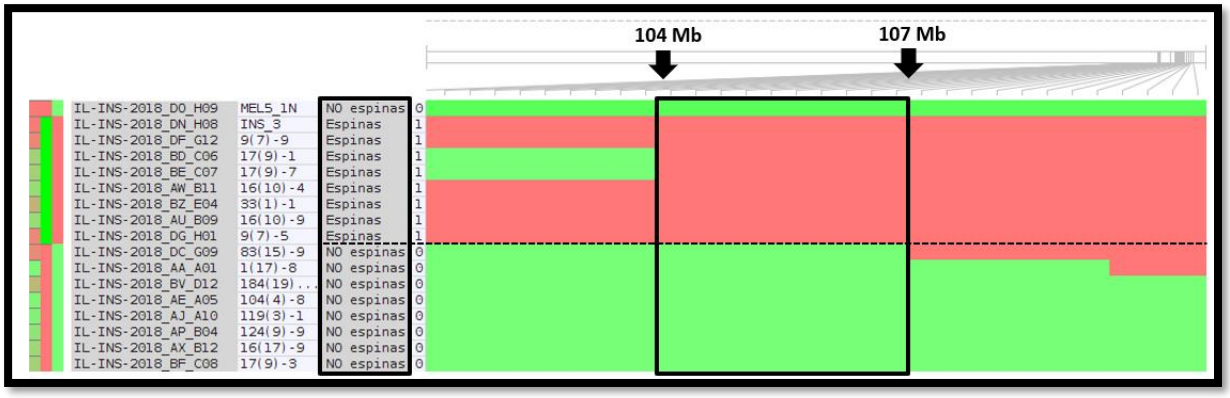


Figure 3. Results from the phenotyping and genotyping of a region of the chromosome 6 of prickly and prickleless BC3 plants. The green areas are the regions of the genome from *S. melongena* and the red ones are from *S. insanum*. As it can be observed, most prickly plants are red and most prickleless plants are green although there are two plants (17(9)-1 and 17(9)-7) that have prickles and have a smaller introgression than the rest and there is one plant (83(15)-9 that has no prickles and also has an introgression. Those are informative plants since they allow to reduce the candidate region from positions 104 Mb to 107 MB in chromosome 6 of *S. insanum.* Retrieved from Villanueva et al., (2018).

These results supported the conclusions of the study previously conducted by Frary et al., (2014) who also made similar findings in a F2 hybrid between eggplant and a wild prickly relative *S. linnaeamum*. In this case, the concerned QTL responsible for prickliness had been situated on chromosome 6 between 102.1 cM and 106.3 cM.

2. OBJECTIVES

For all of the above-mentioned, with the aim of further understanding prickliness in eggplant and in order to being able to use this knowledge in plant breeding programs, the following objectives for this project have been proposed:

1. To delimit and narrow the region of the chromosome 6 of *S. insanum* to which the prickle trait has been associated to in a BC4 population of a *S. melongena* introgression line.
2. To obtain a series of candidate genes to be responsible for the prickle trait in eggplant.
3. To obtain a batch of molecular markers closely related to the prickle trait in eggplant that will prove to be useful for selection in future research projects.
4. To pinpoint the gene controlling of prickliness in eggplant so as to determine the gene/s responsible for the trait.

3. MATERIALS AND METHODS

3.1. PLANT MATERIAL

In this study the utilized plant material was comprised of 304 BC4 plants derived from one fruit and 318 BC4 plants derived from another fruit of the same BC3 plant belonging to a developing IL population between the CWR *S. insanum* (INS1) and the cultivated eggplant *S. melongena* (MEL5).

The total 622 BC4 seeds from which the BC4 plants that we used for our experiment were germinated and grown were obtained in a previous project (Haller, 2018) and stored in the germplasm bank located in the Instituto Universitario de Conservación y Mejora de la Agrodiversidad Valenciana (COMAV). In this previous project, a BC3 population of a *S. melongena* × *S. insanum* IL composed of 30 plants was developed from the parental plants which passport data is represented in Table 1.

Table 1. Passport data of the parental plants utilized for the development of the IL population.

Genus	Species	Country	Province	Collection site	Accession name	Biological status of accession	Genetic pool
Solanum	*insanum*	Sri Lanka	Kandy	Peradeniya	INS1	Weedy	Primary
Solanum	*melongena*	Sri Lanka	-	-	MEL5	Cultivated	-

The selected donor parent for the development of the AB population, *S. insanum*, was a very prickly plant with small green and rounded fruits. It was collected as a weed in Peradeniya in Sri Lanka and cultivated in Spain. The recipient parent, *S. melongena*, is a cultivated eggplant from Sri Lanka that is prickleless and has dark purple and large fruits. It was a non-prickly plant with large dark purple fruits and anthocyanin pigmentation. Segregation for the prickle trait in leaves for both parentals and the hybrid are represented in Figure 4.

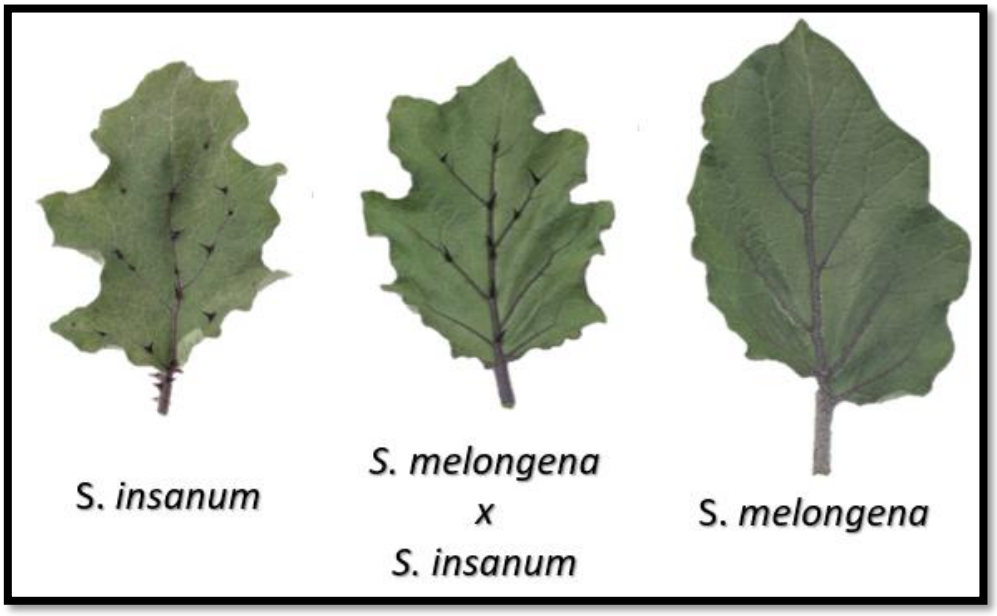


Figure 4. Leaf pictures of the parental plants and the hybrid resulting from their crossing to develop the AB population showing their differences in prickliness. Retrieved from Villanueva et al., (2018).

The BC3 plant selected in order to carry out this study was the one that less genetic background from *S. insanum* had while having an introgresssion in the region of interest on chromosome 6 and whose fruits had more seeds since we wanted to have a big sample size. The plant that fulfilled all the above-mentioned requirements was labelled 33(3)-1 in the greenhouse and it was a prickly plant with dark-green leaves and oval shaped purple fruits whose phenotype and genotyping data can be observed in Figure 5.

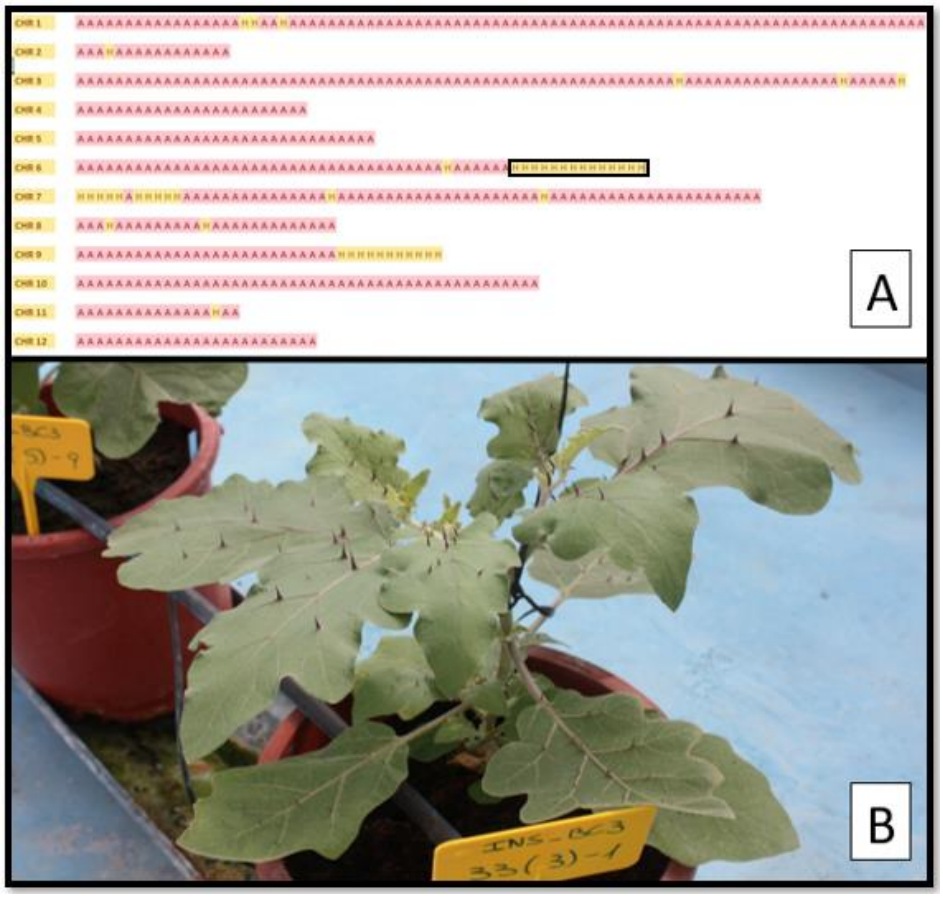


Figure 5. Genotyping data and phenotype of BC3 33(3)-1 plant. A Genotyping file of the BC3(33)3-1 plant. Pink squares with an "A" represent the areas of the genome homozygous for *S. melongena* and light orange squares with an "H" represent the hybrid regions. The target area in chromosome 6 is highlighted in black. **B** Close up picture of the 33(3)-1 BC3 plant from which the BC4 seeds were obtained showing the leaves morphology and prickliness.

After crossing it with the recurrent parent (the *S. melongena* parental) the seeds were collected and prepared for germination.

3.1.1. SEED GERMINATION AND GROWTH CONDITIONS

Seeds were germinated following the protocol designed by Ranil et al., (2015). As directed in this study, the seeds were put to soak in distilled water for 24 hours and the next day they were transferred to a 500ppm gibberellic acid solution in which they were soaked for another 24 hours. On the third day, seeds were transferred to petri dishes containing a cotton wool soaked in a solution of 1000ppm of potassium nitrate and they were covered by paper disks. Then, the petri dishes were incubated for 24 hours at 37ºC. On the fourth day the petri dishes were transferred to a climatic chamber and by the time the seeds had already germinated they were quickly put into seedling trays and maintained at

the climatic chamber. Following the protocol, the photoperiod and the temperature were set to 16 hours of light at 25ºC and 8 hours of darkness at 18ºC. About two weeks later, once plantlets were grown, they were transferred to small plastic pots and when the plants had developed their first 3-4 true leaves they were ready for DNA extraction. Since the DNA extractions were not immediate, plants were transplanted to 1.3L pots and 500g of commercial substrate N3 (Klasmann-Deilmann, Germany) were added to each one so as to allow for further development.

Plants were kept near the campus of the Polytechnic University of Valencia (UPV) in pollinator-free greenhouses belonging to COMAV (39º29'00.4''N 0º20'27.1''W) (Figure 6).

Figure 6. Picture of the BC4 plants in the COMAV pollinator-free greenhouses ready for samples to be collected.

3.2. EXPERIMENTAL DESIGN

In order to clarify the genetics of the prickle trait and to determine the gene(s) controlling it we developed our BC4 population and designed a set of 58 primers forward (F) and reverse (R) covering a fragment of the candidate region of chromosome 6 of less than 3Mb (from position 104005650 to 107184209 of the *S. melongena* reference genome "67/3" developed by Barchi et al., (2016)). Then we screened these markers in successive rounds by HRM and shortened the region of interest as much as possible so as to obtain a few candidate genes responsible for the prickle trait in eggplant and analyse them in detail.

Before starting the sample collection all 622 plants were phenotyped and a statistical analysis was done so as to confirm the suspected monogenic nature of the prickle trait. Samples were collected in order to extract genomic DNA to genotype by HRM. It is important to stand out that not all 622 samples could be analysed for all the selected SNPs since some extractions failed and by the time we were able to repeat them the plants had already die or the genetic material was very degraded. Finally, we combined the phenotyping and the genotyping files and after analysing the most informative plants

we were able to reduce the candidate region from hundreds of genes to a few of them. We then utilized the basic local alignment search tool (BLAST) from two different platforms in order to detect similarities between our candidate genes and genes from other species and we looked for information about them in several databases. The most promising ones will be tested in a future so as to confirm which one is responsible for the prickle trait in eggplant in an RT-PCR and a CRISPR knockout experiments.

3.3. PHENOTYPING

The phenotyping process took place as soon as the plants had been transferred to the pollinator-free greenhouses and prickles became visible and it was quite simple since we only evaluated the prickle trait. For that purpose, we analysed the presence or absence of prickles (PR) in the stem, calyx and leaves of each plant and give them a score using a scale from 0 = absence of prickles to 1 = presence of prickles. We did not use a wider range of values due to the fact that all of the analysed plants either had plenty of sharp, pointy and big prickles in both the calyx and the stem as in the leaves or they had no prickles at all (Figure 7).

Figure 7. Segregation of the prickle trait in the BC4 population.

It is important to point out that the phenotyping process was carried out as soon as prickles appeared and became visible, about 15-20 days after germination, since afterwards and mainly due to stressing conditions prickleless plants could sometimes revert their phenotype and develop prickles so it is the only way of building a reliable phenotyping procedure.

3.4. STATISTICAL ANALYSIS

To analyse the obtained population in order to assess if the genetic control of the prickle trait was or not monogenic and, in order to confirm or reject our hypothesis “prickliness is monogenic and thus its segregation is mendelian (1:1)” (null hypothesis), we performed a Pearson’s Chi-square test (Pearson, 1900) using the phenotyping data of all the collected samples.

The chi-square test is an easy way of evaluating the deviations expected by chance if a hypothesis is true (Griffiths et al., 2000). In our case, we only had one degree of freedom (df) since the number of distinct phenotypic classes that we have is two (prickly and prickleless). For this kind of test, it has been established by convention that a probability value of less than 5% (α= 0.05) is to be taken as the cutoff line to reject a hypothesis. The χ^2 value associated to this probability for one degree of freedom is 3.841.

3.5. GENOTYPING

3.5.1. SAMPLE COLLECTION

In order to collect the samples, two or three young and healthy-looking leaves from each plant were wrapped in plastic bags and quickly put in liquid nitrogen to preserve them while in the greenhouse and back in the lab they were stored in the -75ºC freezer.

3.5.2. DNA EXTRACTION

In order to perform the DNA extraction for each sample two disks (6 mm of diameter) of frozen leaf tissue were put inside a 2 mL-Eppendorf tube with a crystal ball in it. Genomic DNA extractions were carried out following the protocol developed by Vilanova et al., (2020), a modified version of the standard CTAB method (Doyle and Doyle, 1987) that utilizes silica instead of ethanol precipitation so as to recover free of inhibitory compounds genomic DNA. After DNA was completely dried it was eluted in 100 µL of Invitrogen™ UltraPure water (Thermo Fischer Scientific, Waltham, USA) and stored at -20ºC until further use.

3.5.3. DNA QUANTIFICATION AND DILUTION

The integrity of the DNA and the overall quality of the extraction was check via electrophoresis on a 0.8% agarose gel in TAE buffer using GelRed® (Biotium, Fremont, CA USA) to stain DNA. The separation was carried out with a voltage of 80V in a 0.5X TAE buffer during 1h. The marker *Lambda*/*Hind*III (50ng) from Invitrogen™ was used to quantify the amount of DNA in each sample. The agarose gel was visualized with UV light in the GelDoc XR + System transilluminator (Bio-Rad, Hercules, CA USA) image analyser to examine the electrophoresis results.

To measure the DNA yield and concentration of all the samples the spectrophotometer NanoDrop™ ND-1000 (Thermo Scientific, Waltham, MA, USA) technology was applied and the A_{260}/A_{280} and A_{260}/A_{230} ratios were used as secondary test to assess the quality of the extraction since they measure protein and polysaccharide contamination respectively.

With the information provided by the spectrophotometer the samples were diluted to a final concentration of 50ng/µL so as to be used in the genotyping analysis.

3.5.4. PRIMERS DESIGN

For the primers design the Primer 3 Software (Koressaar et al., 2018) developed by Rozen and Skaletsky (2000) was used having into account that optimal primers must have a melting temperature (Tm) of around 58-60ºC and an average length of 20-25 bp. The amplicon target length was restricted to around 90 and 150 bp to ensure the sensitivity detection of the sequence variants since for the SNP analysis by HRM the shorter the fragment the better results you obtain.

In order to select which SNPs would be analysed so as to design and order the primers the resequenced version of the genome of *S. incanum* (Gramazio et al., 2019) was compared to the last version of the *S. melongena* reference genome (Barchi et al., 2019b) in order to find polymorphic SNPs since there was not available a draft assembly of the genome of *S. insanum* yet and it was assumed that *S. incanum* is a wild species phylogenetically very close to *S. insanum* that could serves as well so as to design the primers.

For the SNP genotyping of the region of interest we conducted two rounds of primers design. In total we developed 58 primers (F and R) for the chromosome 6 of *S. incanum* that were ordered to Macrogen, Inc.

3.5.5. SNP ANALYSIS BY HRM

Genotyping was conducted by PCR followed by HRM as recommended by Wittwer et al., (2003). We used the genotype of the recurrent parent (MEL5) as a control to easily detect differences between homozygote and heterozygote genotypes and so we included in all the analysis one sample of MEL5. It is important to take this measure since the extractions may have failed or may include some contaminants and it might happen that once the results are obtained it is impossible to assess which samples are heterozygous and which homozygous for each analysed SNP. By means of including a control sample in the experiment the complexity of the assessment is greatly reduced.

For the analysis the LightCycler® 480 System (Roche) was used and for the sample preparation 2 µL of the of 50 ng/µL dilution were mixed with 0.2 µL of the designed primers (both forward and reverse) diluted 1:10 to achieve a concentration of 10 pmol/µL, 1 µL of $MgCl_2$ 2.5 mM and with 5 µL of MasterMix qPCR No-ROX PyroTaq EvaGreen 5x for HRM 2 ml (500 rxn). The mix was completed with 1.6 µL of UltraPure water obtaining a final volume of 10 µL. In each round of HRM analysis two different SNP markers were evaluated for a total of 48 samples per SNP in a 96-well plate.
The reaction protocol that we followed consisted in a pre-incubation step at 95ºC for 15 minutes followed by a second amplification stage consisting of 55 amplification cycles at 95ºC for 10s (denaturation step), 60ºC for 15s (annealing step) and 72ºC for other 15s (elongation step). Then, a third melting step at 95ºC for 1 minute, 40ºC for another minute, 1 sec at 60ºC and rising the temperature at 0.02ºC/s until 95ºC- The process finalises with a cooling stage that lasts 10 minutes at 40ºC.
After the process is completed, the data is generated and can be analysed using the LightCycler 480 Software release 1.5.1.62 via “Tm calling” and “Gene scanning” analysis.

3.6. CANDIDATE GENES

3.6.1. BLAST SEARCH

We conducted a blast search so as to find similar genes in other related plant species. We used both the sequences for our genes provided by The Eggplant Genome Project database (*http://www.eggplantgenome.org/*) and by the Sol Genomics Network (Fernandez-Pozo et al., 2015) although the latter gave us much better results.

First, we used the NCBI BLAST tool (Altschul et al., 1990) so as to have a general idea about the function and relatedness to other species of our candidate genes. We applied the “nucleotide BLAST (nBLAST)” setting to compare only nucleotide sequences and we performed the search in the Reference RNA sequences (refseq_rna) database since this database only includes non-redundant and well-annotated reference sequences that are trustful and reliable. We also selected the program “Highly similar sequences” (megablast) to avoid dubious results. When these settings failed to give as any relevant result we tried the option nucleotide collection (nr/nt) that performs the search in a larger database with more sequences.

Afterwards, we decided to use the blast tool from the Sol Genomics Network (Fernandez-Pozo et al., 2015) since it makes use of a clade-oriented database that contains biological data for species in the

Solanaceae family and their close relatives and so it is much more specific and the results more orientated to our objectives. We applied the "nucleotide to nucleotide db (blastn)" program and we selected the Tomato Genome cDNA (ITAG release 4.0) and Tomato Genome CDS (ITAG release 4.0) databases due to the fact that they are the last updated versions of the tomato genome and thus they are more comprehensive and reliable than any other.

3.6.2. DATABASE RESEARCH

We first consulted the GenBank® database (Clark et al., 2016) for each candidate gene from which we could access the gene and protein related data. We then analysed all the available information regarding the orthologs of each candidate gene in the OrthoDB database (Kriventseva et al., 2019) which included functional categories, go terms and conserved domains. Then, we consulted the Gene Ontology Consortium (Ashburner et al., 2000) to further analyse the linked GO terms to each candidate gene, the integrative protein signature (Interpro) database (Hunter et al., 2009) and the Universal Protein Knowledgebase (UniProt) consortium (Apweiler et al., 2004) to find additional data regarding the predicted translated proteins derived from our genes of interest. Finally, we searched for information regarding protein interactions in the STRING (Mering et al., 2003) and for a specific candidate gene we consulted the CANTATAdb database (Szcześniak et al., 2016) that is specific for long non-coding RNAs (lncRNA) in plants.

4. RESULTS

4.1. STATISTICAL ANALYSIS

The phenotyping data used for the statistical analysis can be observed in Figure 8.

Figure 8. Results of the phenotyping process after having analysed all 622 plants. As it can be observed the distribution of the plants is almost 50/50 and thus we suspected that the control of the trait was strongly monogenic.

After having phenotyped all the plants from our population a Pearson's chi-square analysis was performed (Table 2).

When we analysed all plants for the fruit 1 we obtained 162 plants with prickles and 142 without them from the 152 expected in each category, for the fruit 2 we observed the exact same number of prickly and prickleless plants that we expected (159 in each class) and our χ^2 values for both fruits placed the probability value in more than 10% when our threshold to reject the hypothesis as it had been established by convention was less than 5%. We also did a global analysis considering all the plants together and the null hypothesis was still accepted.

That means that even if there are other minor QTLs involved in prickliness by determining this major one we should be able to effectively remove prickles from eggplant by knocking it off.

Table 2. Results of the Pearson chi-square analysis. The freedom degrees, the alpha level and the χ2 value associated to it are represented in the first file. We performed the analysis for each fruit independently (A and B) and then another one for the whole dataset (C). In each table the results from the chi-square test are shown in red and the acceptance or rejection of the null hypothesis in green.

Freedom degrees (n-1) = 1	**α = 0.05**	**χ2 = 3,841**	
Fruit 1 (304 plants)	Observed (O)	Expected (E)	**(O-E)^2/E**
With prickles	162	152	0,658
Without prickles	142	152	0,658
Chi square (χ^2)			**1,316**
Null hypothesis: monogenic control of the trait, mendelian segregation (1:1) =			**ACCEPTED**
A			

Fruit 2 (318 plants)	Observed (O)	Expected (E)	**(O-E)^2/E**
With prickles	159	159	0
Without prickles	159	159	0
Chi square (χ^2)			**0**
Null hypothesis: monogenic control of the trait, mendelian segregation (1:1) =			**ACCEPTED**
B			

Total (622 plants)	Observed (O)	Expected (E)	**(O-E)^2/E**
With prickles	321	311	0,322
Without prickles	301	311	0,322
Chi square (χ^2)			**0,644**
Null hypothesis: monogenic control of the trait, mendelian segregation (1:1) =			**ACCEPTED**
C			

4.2. PRIMERS DESIGN

We designed a total of 58 primers (Forward and Reverse) for the candidate region in chromosome 6 (from 104005650 bp to 107184209 bp) but some of them could not be used since they were too unspecific or they were too close to the SNP but did not include it and we had to modify them. Besides that, two primers failed in the SNP analysis by HRM and so, we performed a second round of primers design to change their sequence and repeat the analysis.

The sequences of all the utilized primers together with their position in the genome are summarized in Table 3.

Table 3. Sequences composition and size of all the designed primers and their position in the genome of *S. insanum*. Positions are measured in base pairs (bp) and the "nt" columns stand for nucleotide and they refer to the size of each primer (the values are in green).

SNP code	SMEL3Ch06 position (bp)	FORWARD (5'→3')	nt	REVERSE (5'→3')	nt
SNP_1	104005650	CCCTGATAGAAAATGTGTAAACAGC	25	CAACCAACTCAAATGACTATGC	22
SNP_2	104022822	CCCAGGCAACTGAAAAGAGTAG	22	TGTTCTCAAATTGGAAATATTGTTG	22
SNP_3	104498526	GCCCTCCAACTGTGATTTTC	20	TTAAGCTATACCCCACTGCTG	22
SNP_4	104616800	GGCGGAGAAATCAATGAAGC	20	GATAACTGTTGCGAATTTTGG	22
SNP_5	104676481	CTGAAGAAGAAGCAGATGATCC	22	CTTGTTCAGAGAGGTTCTACGG	22
SNP_6	104835522	GAAAGCCGGACTGTTGAGG	19	CTTAGACAATTGCTCGAAAACC	22
SNP_7	104890646	AGGGAGAGGGAGAAGTCAAGC	21	CTTCCTCTCTTTCCCCTTCG	22
SNP_8	104910233	TGAAGAAAGCTATCGCAGTGG	21	GCTTCTGAACACAACGAATGG	22
SNP_9	104938346	CCACCTCTGTCAATCATTTCG	21	TGATAATCTCACTTCTAGCTGATGC	22
SNP_10	105100846	GGGCTATTGCCTTACTTCTGG	21	GGAGGCAAAAATAAAGAGAATCAC	22
SNP_11	105258442	TCTTCAACCAAAGGAAAGAGG	21	TCTTGTTTGGAAGTAGGAAATGG	22
SNP_12	105338156	GCAGATAGGTGTTTAGAGACACAGG	25	TCTGTTGGAAGCCTACACTGC	22
SNP_13	105450271	GGCCCAGGCCTCCTATTG	18	AACACGGCCAGGTGATAATG	22
SNP_14	105545399	TGGGAGACTAGATATAAGAAAGAACTG	27	CCCATCTGCAGAGTACAAAGC	22
SNP_15	105546176	TCACCTGTGAATCAAAAGAAGC	22	CTGGCTGAAGATGTTTGTAAGC	22
SNP_16	105560836	TTTTTACTCTTCGGCCAACC	20	CCACTGGAGTCTCTGTGGAC	22
SNP_17	105772791	TTGGACTGATTTCTCCATTGC	21	TCTTCCTAGAGAAACTAAAAGAGAGG	22
SNP_18_19*	105787930	CAATCATCACAAAGTTTTTCAAGG	24	GGATATACGTCCTGATTCTACTTGG	22
SNP_20	106204000	CCAAATGCTTCTGATCATATTCG	23	TGGGAAATATTGTTGGTGTGG	22
SNP_21	106416198	CAATTTTTCCCAGGTTAATTCTTT	24	TTTCTGTTCCTAATACAAATTTCAACA	22
SNP_22	106557321	GCTGATGTTTTACAAAAATAAAAGCTG	27	GCGAGTCTTTTAGCCCTGTC	22
SNP_23	106736529	CTGTCATAATCTTGGGCAACC	21	AAATGGCATTAAACCAAGAACC	22
SNP_24	106758637	TGTTTCTTTCTAATTTCTTTTCTGC	25	ACTCAGGTCCACAAGTTAAGACC	22
SNP_26	107058802	TTGGAATTTCAGGTCGAAGC	20	GAATTGGTGGCTGGTGAGG	22
SNP_28	107075478	TTTCATTATCTTCAGCTGATGTCC	24	GGAAAAGATGCACGGATAGG	22
SNP_29	107075532	CCTATCCGTGCATCTTTTCC	20	AAATCATCATCCTCACTGTAGGG	22
SNP_31	107170521	TCCTCAATAATGCATAGAGATTCG	24	TCCAGTCCGTTTTAGCTTGG	22
SNP_32	107170597	TCCAAGCTAAAACGGACTGG	20	TTCTGATTCAATTCAGTAAAATGC	22
SNP_33	107184209	CCAAAACCCTGAAGAAATGC	20	TTTCTCCGGTGGTAGAGTGG	22

*: SNP_18 and SNP_19 were really close and so we decided to join them in a single primer.

4.3. SNP ANALYSIS BY HRM

From the total of 58 primers (F and R) that were designed we first analysed SNP 1 (at 104005650 bp) and SNP 22 (at 106557321 bp), the SNPs flanking the 2.55 Mb region suspected to contain the gene of interest. Since HRM is capable of detecting differences in a single nucleotide it was easy to detect which samples were homozygous and which were heterozygous for each SNP by comparing their melting curves to the parent plant that we used as a control (MEL5) (Figure 9).

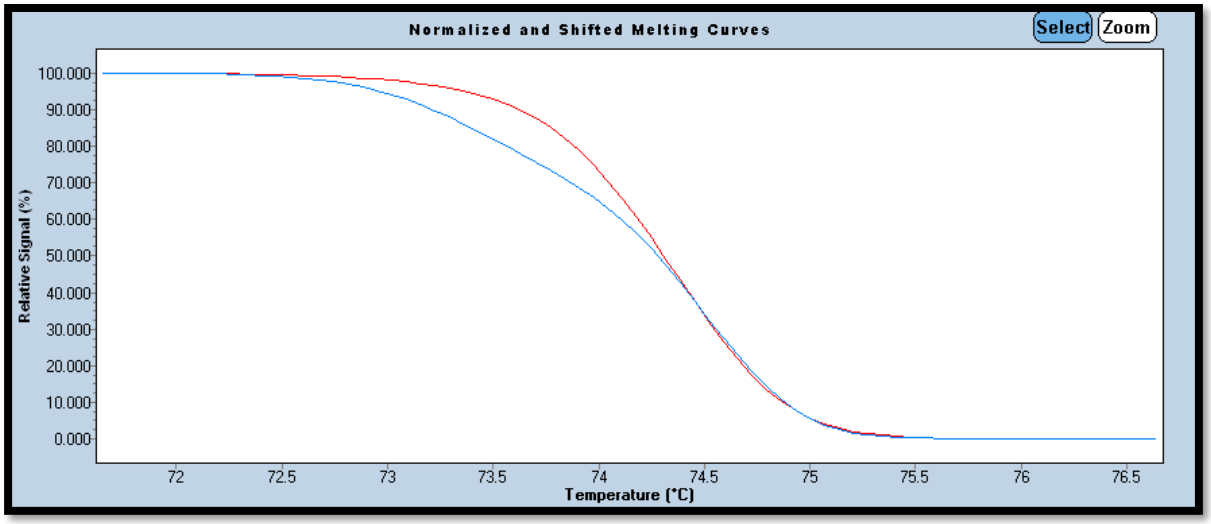


Figure 9. Image taken from the LightCycler 480 Software release 1.5.1.62 showing the differences in the melting curve between a homozygous sample and a heterozygous sample during the "Gene Scanning" analysis for SNPs 1 and 22. The "relative signal" axis shows the percentage of fluorescent dye that remains bound to the dsDNA and the "Temperature" axis shows the increasing temperature of the process. The red curve represents a homozygous sample and the blue one a heterozygous one.

For each round we selected those plants that were informative, that is, plants that were different for the two analysed SNPs (homozygous for one and heterozygous for the other) so as to pay them special attention in the analysis of the rest of SNPs. We did so because if the two flanking SNPs for the region of interest were either homozygous or heterozygous we inferred that there had been no recombination since it is very unlikely that a double recombination event takes place in such a small fragment.

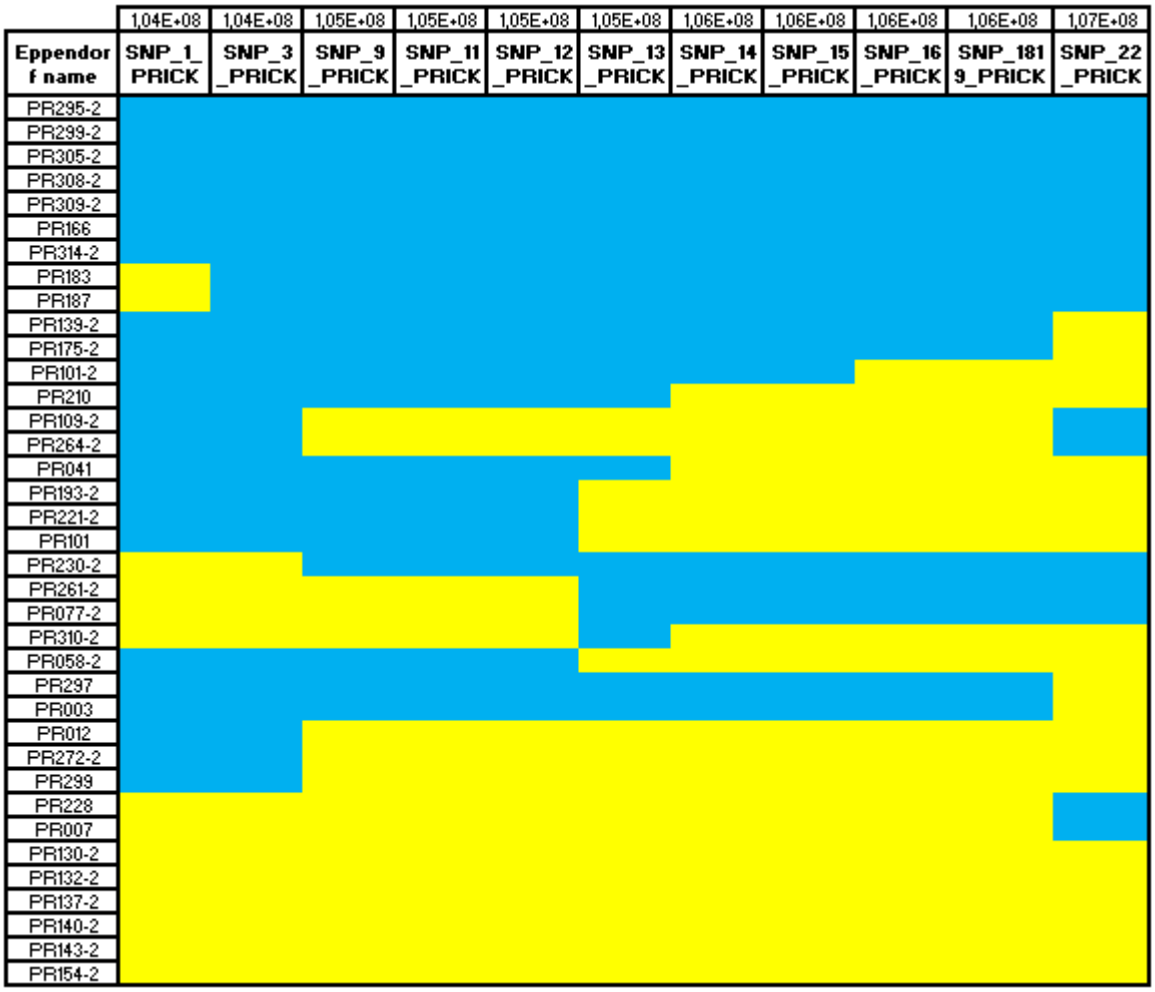
After selecting the plants that had been informative for the first round we checked SNP 3 (104498526 bp) and 15 (105546176bp) reducing the region from 2.5 Mb to 1.05 Mb. We also genotyped the non-informative plants from the first genotyping round in case they showed an introgression inside the region of interest. We continued analysing intern markers (SNPs 9, 11, 12, 16 and 18_19) and reducing the zone where prickliness coding gene could be located until we got a candidate region of 208.02 Kb in which only 16 genes were present.

Finally, we checked SNPs 13 and 14 but the analysis failed. After redesigning them, both of them worked fine and thus we could reduce the region of interest to 95129 bp (from position 105450271 to 10554399 in chromosome 6).

In the end we were not able to genotype the same number of plants that we phenotyped (622 samples) since some of them died and we could not extract more DNA to repeat the genotyping process for those whose results were dubious and we had to discard them. All the rare cases were annotated for further research in the future but we did not consider them in our analysis so in the end we had 544 plants in our genotyping file.

We have summarized the HRM analysis results in Figure 10. For more detailed information about the genotyping results see Appendix Table 1.

Figure 10. Extract from the genotyping results file. Blue cells represent homozygous markers and yellow cells heterozygous ones.



4.4. FINE MAPPING OF THE PRICKLE TRAIT

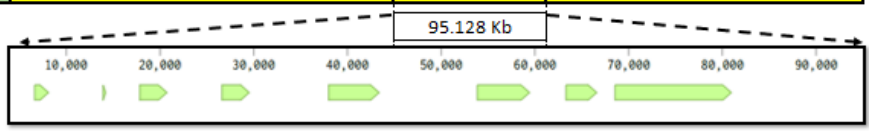
Once we had reviewed the results of the HRM analysis we merged them with the phenotyping results and began to consider all data together. First of all, we removed those plants that were not informative, that is, the samples that did not have prickles and were homozygous for all the SNPs and those that had prickles and were heterozygous for the whole region since that is what it was expected to happen after backcrossing our parent plant with the introgression in chromosome 6 with the recurrent parent. Then, we selected the informative samples, those that had prickles and were homozygous for any SNP and those that did not have prickles and were heterozygous for any SNP. These plants were interesting since they delimited the region where the gene coding for prickles could be located since if a plant had prickles and was homozygous for a particular SNP that region should not be hosting the gene of interest and the same thing happens if a plant showing no prickles was heterozygous for any SNP. We also selected for further studying plants that either were homozygous for flanking SNPs and had prickles or that were heterozygous and did not show prickliness since they could have undergone a double recombination phenomenon, which is very rare, or they could have reverted the phenotype due to stress or other similar factors affecting growth.

By studying all these plants, we were able to greatly shorten the region of interest (from position 105450271 bp to 105545399 bp in chromosome 6) until we only had a few candidate genes (Figure 11). Al that moment, we began to study the suspect genes so as to select the most probable ones for further research. The extended version of the genotype and phenotype file can be found in the Appendix Table 1.

Figure 11. Analysis of the informative samples and fine mapping of the prickle trait. In the first two rows the SNP numbers and their positions. In the first two columns the phenotype and the nº of plants. The first and the last plants are the parentals (MEL5 and INS1). Down below the legend and the final candidate region with the 8 candidate genes.

	SNPs	SNP_1 PRICK	SNP_3 PRICK	SNP_9 PRICK	SNP_11 PRICK	SNP_12 _PRICK	SNP_13 _PRICK	SNP_14 _PRICK	SNP_15 _PRICK	SNP_16 _PRICK	SNP_1819 _PRICK	SNP_22 _PRICK
phenotype	Position in chr. 6 (bp) / Nº of plants	104005650	104498526	104938346	105258442	105338156	105450271	105545399	105546176	105560836	105787930	106557321
NO PRICKLES	parental 1 (MEL5)	A	A	A	A	A	A	A	A	A	A	A
NO PRICKLES	23 plants	H	A	A	A	A	A	A	A	A	A	A
NO PRICKLES	1 plant	H	H	A	A	A	A	A	A	A	A	A
NO PRICKLES	4 plants	H	H	H	H	H	A	A	A	A	A	A
NO PRICKLES	9 plants	A	A	A	A	A	A	A	A	A	A	H
NO PRICKLES	1 plant	A	A	A	A	A	A	A	A	H	H	H
NO PRICKLES	2 plants	A	A	A	A	A	A	H	H	H	H	H
PRICKLES	2 plants	A	A	H	H	H	H	H	H	H	H	A
PRICKLES	1 plant	A	A	A	A	A	A	H	H	H	H	H
PRICKLES	4 plants	A	A	A	A	A	H	H	H	H	H	H
PRICKLES	2 plants	A	A	A	A	H	H	H	H	H	H	H
PRICKLES	4 plants	A	A	A	H	H	H	H	H	H	H	H
PRICKLES	20 plants	A	A	H	H	H	H	H	H	H	H	H
PRICKLES	7 plants	A	H	H	H	H	H	H	H	H	H	H
PRICKLES	1 plant	H	H	H	H	H	H	H	H	A	A	A
PRICKLES	2 plants	H	H	H	H	H	H	H	H	H	A	A
PRICKLES	12 plants	H	H	H	H	H	H	H	H	H	H	A
PRICKLES	parental 2 (INS1)	H	H	H	H	H	H	H	H	H	H	H

Leyend	
Homozygous	A
Heterozygous	H



4.5. CANDIDATE GENES

We ended up having 8 candidate genes that were located in the final delimited region of the chromosome 6 introgression that we could not further reduce (Figure 12) Their complete sequence, exact position and composition can be accessed in the Appendix Table 2. We began to research about them before conducting any other experiment in order to figure out which ones could be potentially related to prickliness and which could be rejected or studied the last.

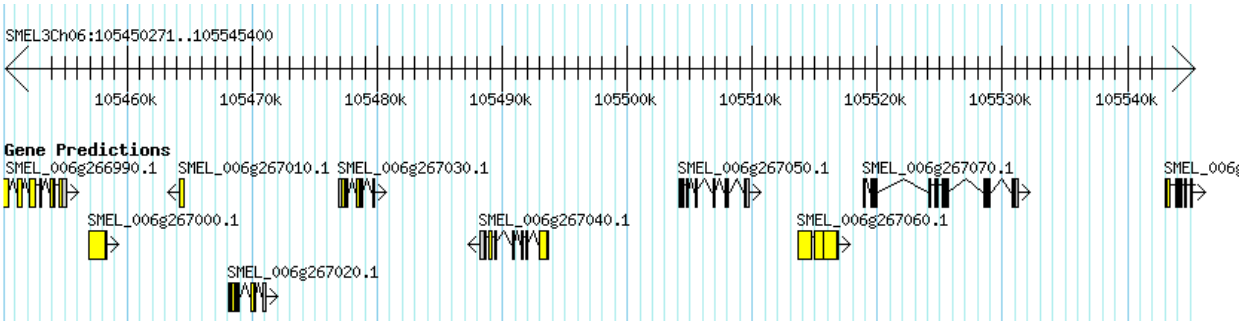


Figure 12. Image of the candidate region showing the 8 candidate genes in a genome browser of *S. melongena*. Retrieved from http://www.eggplantgenome.org/.

4.5.1. BLAST SEARCH

Before further proceeding with the research, we wanted to determine which of the candidate genes had more probabilities to be related to prickliness in order to analyse them first and so we first underwent two quick blast searches trying to identify orthologs for our candidate genes in other species.

For most genes the blast searches were very robust and the results were clear (similar genes in closely related species and very reliable information with high scores and low e-values) although for some of them the search resulted in uncharacterized genes with unknown functions or in many different genes with low scores. In those cases, we had to settle with detecting some conserved domains so as to analyse them and try to figure out the possible function of the gene. At the end, we were able to obtain some information from all the genes that we could use whilst discussing which of them were more likely to be responsible for prickliness in eggplant.

4.5.1.1. SMEL_006g267000.1

For the gene SMEL_006g267000.1 NCBI blast results were fairly reliable since all the matches pointed out the same gene, a lysM domain receptor-like kinase 3 protein coding gene, with very high scores and low e-values in many related species such as *S. tuberosum* (potato*), S. pennellii* and *S. lycopersicum* (tomato). When we performed the search in the Sol Genomics Networtk database it gave as a single match with a receptor-like kinase mRNA in *S. lycopersicum* (Solyc06g075030.1).

4.5.1.2. SMEL_006g267010.1

When we analysed the gene SMEL_006g267010.1 we only had one match in both blast searches and it was an uncharacterized lncRNA from *S. lycopersicum* (Solyc06g075040.1).

4.5.1.3. SMEL_006g267020.1

Gene SMEL_006g267020.1 had a few matches in *C. annuum* (pepper) and *S. lycopersicum* and all of them suggested that it was a bifunctional monodehydroascorbate reductase and carbonic anhydrase nectarin-3-like protein coding gene. In the Sol Genomics Network database we also obtained one match indicating it was an alpha carbonic anhydrase 7 (Solyc06g075050.1).

4.5.1.4. SMEL_006g267030.1

The gene SMEL_006g267030.1 was also a bifunctional monodehydroascorbate reductase and carbonic anhydrase nectarin-3-like protein coding gene. In fact, some of the found entries coincided with the previous gene but SMEL_006g267030.1 had also matches with *S. pennellii* and *S. tuberosum*. The Sol Genomics Network search results were also similar to those of the previous candidate gene with a single match pointing out a carbonic alpha anhydrase in *S. lycopersicum* (Solyc06g075060.1).

4.5.1.5. SMEL_006g267040.1

For the gene SMEL_006g267040.1 we found strong evidence suggesting that it was a chloroplastic nudix hydrolase 19 protein coding gene since we had matches in similar species like *S. tuberosum*, *S. lycopersicum*, *S. pennellii* and even in *C. annuum*. Blast search in the Sol Genomics Network database provided a match with a NADH pyrophosphatase-like gene in *S. lycopersicum* that contained a NUDIX hydrolase domain (Solyc06g075080.2).

4.5.1.6 SMEL_006g267050.1

When we blast-searched the gene SMEL_006g267050.1 all the results indicated that it was a cytokine riboside 5'-monophosphate phosphoribohydrolase coding gene although there were differences in what refers to the specific form of the LOG family it was. In *S. tuberosum* it was described as a LOG4-like or LOG1-like gene whereas in *S. lycopersicum* it was defined as a LOG8 gene. In *S. pennellii* it was also stated that it was a LOG1-like gene. In the Sol Genomics Network database we found one match for *S. lycopersicum* that coincided but it did not specify anything about the LOG family (Solyc06g075090.2).

4.5.1.7. SMEL_006g267060.1

The BLAST search for the gene SMEL_006g267060.1 provided as a result an uncharacterized protein coding gene in *S. tuberosum*, *S. pennellii*, *S. lycopersicum* and many more species. The only thing all matches had in common was the presence of two conserved domains what made us suspect it was a gene related to the nucleotide-diphospho-sugar transferases family. The only result that was not and uncharacterized gene was in the model species *A. thaliana* in which it was defined as a glycosyltransferase family protein 2 coding gene. The Sol Genomics Network database also provided a match with a glycosyl transferase family 2 protein gene in *S. lycopersicum* (Solyc06g075100.2).

4.5.1.8. SMEL_006g267070.1

The last gene, SMEL_006g267070.1 was also an uncharacterized protein coding gene in the NCBI blast search for *S. tuberosum*, *S. lycopersicum*, *S. pennellii* and in *C. annuum*. In *S. tuberosum* one transcript variant had a conserved domain that related it to aminotransferases and this was the result with the highest score although the other match in *S. tuberosum* as well as the matches in *S. lycopersicum, S. pennellii* and *C. annuum* pointed out that it was a gene related to DUF707, a family of uncharacterized genes in *A. thaliana*. The Sol Genomics Network blast search also detected a match with the DUF707 gene in *S. lycopersicum* although the database identified it as a lysine ketoglutarate reductase trans-splicing protein (Solyc06g075110.2).

All the blast searches results are summarized in Table 4.

Table 4. Summary of the blast searches obtained results. Only the most likely function based on all the information available has been annotated.

Name	SMEL3Ch06 position	Most likely function
SMEL_006g267000.1	105456950..105458396	LysM domain receptor-like kinase 3
SMEL_006g267010.1	105464241..105464489	Long non-coding RNA
SMEL_006g267020.1	105468115..105471025	bifunctional monodehydroascobate reductase and carbonic anhydrase nectarin-3-like
SMEL_006g267030.1	105476869..105479798	bifunctional monodehydroascobate reductase and carbonic anhydrase nectarin-3-like
SMEL_006g267040.1	105488260..105493655	nudix hydrolase 19 chloroplastic
SMEL_006g267050.1	105504136..105509693	Cytokinin riboside 5'-monophosphate phosphribohydrolase
SMEL_006g267060.1	105513604..105516858	Gycosyl transferase family 2 protein
SMEL_006g267070.1	105518832..105531210	Lysine ketoglutarate reductase trans-splicing related 1

4.5.2. DATABASE RESEARCH

After having localized orthologs for our candidate genes in other species we wanted to know their most probable function as well as to detect their conserved domains to figure out if they could be related to prickliness in eggplant and to do so we consulted several databases to research about them. We annotated all the available information that we could find so as to gain an overview about the function of our genes and to be able to discuss the most probable candidate genes to be governing the prickle trait in eggplant.

4.5.2.1. SMEL_006g267000.1

The gene SMEL_006g267000.1 was in all likelihood a lysM domain receptor-like kinase 3 as we checked during the BLAST search and so we consulted the GenBank database entry for this gene in *S. lycopersicum, S. pennellii and S. tuberosum* and the OrthoDB database for the protein kinase domain so as to find that it was related to ATP binding (GO:0005524) and protein phosphorylation (GO:0006468) and that it included a conserved domain with serine-threonine/tyrosine-protein kinase catalytic activity that was also linked to the protein kinase activity term (GO:0004672). Regarding its function, we suppose that it was involved in signal transduction mechanisms and cell wall and membrane biogenesis since that are the main functions of the protein kinase domain. We also looked for further information in the Uniprot database an made similar findings. We confirmed that it was a transmembrane protein and, in a manually annotated and reviewed *Arabidopsis thaliana* entry, we found that this protein was involved in defence toward both abiotic and biotic stresses functioning as a cell surface receptor that might recognize microbe-derived N-acetylglucosamine (NAG)-containing ligands.

4.5.2.2. SMEL_006g267010.1

The gene SMEL_006g267010.1 was uncharacterized and we could not find any conserved domain in its sequence although when we found out that it was a lncRNA we searched in the CANTATA database and found that it was similar to an uncharacterized lncRNA from *S. tuberosum*.

4.5.2.3. SMEL_006g267020.1

In what refers to the SMEL_006g267020.1 gene we had confirmed in the BLAST search that coded a bifunctional monodehydroascorbate reductase and carbonic anhydrase nectarin 3-like protein and we analysed the OrthoDB for the alpha carbonic anhydrase domain to find out that it was related to the transport of inorganic ions and involved in the cell cycle control (cell division and chromosome partitioning). This domain was also related to the nitrogen metabolism and to the Go terms metal ion binding (GO:0046872), carbonate dehydratase activity (GO:0004089) and zinc ion binding (GO:0008270). In the Uniprot database we found that the function of this protein in *C. annuum* was the reversible hydration of carbon dioxide molecules. In *Nicotiana sylvestris* (wood tobacco) it had the same linked GO terms and it had more than 90% of similitude with another manually annotated and reviewed protein (bifunctional monodehydroascorbate reductase and carbonic anhydrase nectarin 3) in *Nicotiana langsdorffii x Nicotiana sanderae (Ornamental tobacco)* that had both carbonate dehydratase and monodehydroascorbate reductase activities that was involved in the regulation of nectar pH and that it was an extracellular or secreted protein. When we analysed if it had any relevant interactions with other proteins we found out that it was related to auxin signalling proteins (transport and signal transduction proteins).

4.5.2.4. SMEL_006g267030.1

SMEL_00g267030.1 was also a bifunctional monodehydroascorbate reductase and carbonic anhydrase nectarin 3-like protein and so the same information regarding the previous gene applies to this one.

4.5.2.5. SMEL_006g267040.1

For the SMEL_00g267040.1 gene we had confirmed that it was a chloroplastic nudix hydrolase 19 in all the related species to eggplant and so we investigate the nudix hydrolase in the OrthoDB to find that it had something to do with nucleotide and lipid transport and metabolism and with defence mechanisms. Nudix hydrolases also participate in the nicotinate and nicotinamide metabolism and are present in peroxisomes. They are associated to the hydrolase activity (GO:0016787), metal ion binding (GO:0046872), cytosol (GO:0005829) and chloroplast (GO:0009507) GO terms. Besides the NUDIX hydrolase domain, nudix hydrolases also have a NADH pyrophosphatase-like N terminal domain related to nucleotide transport and metabolism. In the Uniprot database we also found that, in *N. tabacum*, this protein was also associated to the NADH pyrophosphatase activity (GO:0035529), NAD catabolic process (GO:0019677), NADH metabolic process (GO:0006734) and NADP catabolic process (GO:0006742) GO terms and it was located in the cytosol whereas in *C. annuum* it was located in both the cytosol and chloroplasts. In a manually annotated and reviewed entry for this protein in *A. thaliana* that had a 50% identity its defined function was to mediate the hydrolysis of some nucleoside diphosphate derivatives with high affinity for NADPH and it was located in the chloroplasts.

4.5.2.6. SMEL_006g267050.1

The gene SMEL_006g267050.1 that had been associated to different forms of the cytokinin riboside 5'-monophosphate phosphoribohydrolase protein coding gene in *S. tuberosum*, *S. lycopersicum*, *S. pennellii* and *C. annuum* was related to signal transduction mechanisms, cell motility and vesicular trafficking, transport and secretion and linked to the GO terms hydrolase activity (GO:0016787) and cytokinin biosynthetic process (GO:0009691). Regarding the LOG family to which this protein belongs, in *S. lycopersicum* it was LOG8, characterized for having DNA recombination-mediator protein A conserved domain, in *S. tuberosum* one transcript variant was a LOG-4 like protein (same as in *C. annuum*) with a lysine decarboxylase conserved domain and other variant a LOG1-like protein (as in *S. tuberosum*) that had the same conserved domain of LOG8 (DNA recombination-mediator protein A domain). Further investigating in the Uniprot database we discovered that the LOG-1 like cytokinin riboside 5'-monophosphate phosphoribohydrolase in *N. tabacum* was a cytokinin-activating enzyme and a phosphoribohydrolase that converts inactive cytokinin nucleotides to the biologically active free-base forms. It was located in the cytosol and the nucleus. Looking for interactions in the STRING database we found that it might be related to zeatin biosynthesis, an adenine-derived cytokinin.

4.5.2.7. SMEL_006g267060.1

When we analysed the gene SMEL_00Gg267060.1 we could only find two conserved domains associated to the glycosyltransferase family (Carbohydrate transport and metabolism). We accessed the GenBank with *S. lycopersicum*, *S. tuberosum* and *S. pennellii* and found a relation between the uncharacterized gene and the nucleotide-diphospho-sugar transferases family. In the OrthoDB we discovered that this family was linked to the integral component membrane (GO:0016021) GO term. In the Uniprot database we could not find much more information and so we analysed the result from the BLAST search that matched with *A. thaliana* but we still could not find anything useful but the fact that it was a transmembrane protein. Looking for relevant interactions in the STRING database we discovered that it might be involved in the fructose and mannose metabolism.

4.5.2.8. SMEL_006g267070.1

For the gene SMEL_006g267070.1 we had another uncharacterized protein in all the eggplant related species. In *S. tuberosum*, one transcript variant was related to the aminotransferase-like domain and had a chromosome segregation ATPase conserved domain involved in the control of the cell cycle. The other variant for *S. tuberosum* as well as the matches with the other related species were characterized for having the DUF707 conserved domain. When we analyse it in the OrthoDB database we found out that this domain was involved in the carbohydrate transport and metabolism and linked to the integral component of membrane (GO:0016021) GO term. Whilst searching in the Uniport database we discovered that, in *S. lycopersicum*, the DUF707 domain was associated to a secretory carrier-associated membrane protein probably involved in membrane trafficking and linked to the protein transport (GO:0015031) GO term. This was a plasma membrane protein belonging to the SCAMP family. In *N. tabacum* and *N. silvestris* another uncharacterized protein containing this domain was related to the pseudourine synthase activity (GO:009982), RNA binding (GO:0003723) and pseudourine synthesis (GO:0001522) GO terms. This uncharacterized protein had a 50% similitude with a trimethylguanosine synthase in *A. thaliana*, a transmembrane protein located in the endosome and Golgi apparatus. When we analysed the STRING database it related the *S. lycopersicum* match (Lysine ketoglutarate reductase trans-splicing protein) with the cupin superfamily.

5. DISCUSSION

5.1. STATISTICAL ANALYSIS

The results of the Pearson's chi-square test showed that our null hypothesis (monogenic control of the prickle trait and mendelian 1:1 segregation) was very likely to be true as we had initially suspected. Furthermore, we checked that the results of the analysis were valid if we studied both fruits independently but also if we considered all data together. Finally, we checked that even if the stablished threshold for this kind of analysis is a 5% confidence level (χ^2 value < 3.841) our data is capable of supporting a confidence level up to 10% (χ^2 value < 2.706) what means that, if the hypothesis is true, deviations from the expectations are expected 10% of the time and so, our data is very solid and the hypothesis very likely to be true.

5.2. PRICKLINESS

Once we finally extracted all the available information related to our candidate genes we started to analyse them so as to assess their most probable function in eggplant and their possible relationship with the prickle trait. In order to do so we first studied in more detail prickliness in eggplant and other species.

We already discussed the difference between thorns, spines and prickles, that is, the absence of vascular bundles in prickles. The most likely origin of prickles is the outgrowth of epidermal cells since they are extensions of the cortex and the epidermis but we have not deepened into the molecular insights of prickle development mainly because there is almost no bibliography regarding this topic.

One of the most relevant studies on prickles was carried out by Kellog et al., (2011) on raspberries, blackberries and roses and, in that study, it was suggested that the structures that develop into prickles are early-stage glandular trichomes and that prickliness could be a modification or an extension of the glandular trichomes development. Based on these claims, Kellog and his colleagues further developed his theory stating that, since glandular trichomes are rich in secondary plant metabolites (Tattini et al.,

2000), they were most likely involved in prickle development as well as in acting as signals to the epidermal and cortical tissues maybe activating the rapid growth and lignification of prickles. This theory was supported by a previous research in which it was discovered that communication between the L1 and L2 embryonic tissue layers was essential for prickle development, suggesting that molecular signals between cortical cells and epidermal cells might play a key role in prickle development (Coyner et al., 2005).

Glandular trichomes are originated in the epidermis and arise from the L1 layer, besides that, prickle lignification occurs in a distal-proximal manner and the final size of the prickle is proportional to the size of the trichome head. With all this evidence, Kellog et al., (2011) proposed that the trichome head may send signal to the epidermal and/or cortical tissues provoking a cell proliferation process that results in prickle development.

Another study carried out in *S. viarum* by Pandey et al., (2018) also suggested a crucial role of glandular trichomes in prickles development either by directly transforming into them or indirectly by signalling inductor factors. The latter is more likely since prickleless mutants also have glandular trichomes but they do not develop into prickles for any reason. In this study they performed some transcriptome analysis and discovered that in prickleless mutants several metabolites were down or up-regulated when compared to normal prickly plants such as ethylene and salicylic acid, both defence molecules also involved in plant development. They also found several pathogen-related protein and transcription factor families involved in plant growth up-regulated in the prickly epidermis and so they suspected that some of these defence-related genes had a potential role in prickle development maybe providing an optimal micro environment for epidermal tissues prickles outgrowth.

We analysed those statements and suggested that, if true, the gene governing prickliness in eggplant could be related to either cell signalling activities or to the production of some specific secondary metabolites. It could even be a transcription factor coding gene that regulate the expression of several other genes inducing prickles development. Our hypothesis is based on the idea that we have proved the trait is governed by a single gene/QTL and so it must be a very important gene maybe functioning as a regulatory element and thus controlling the expression of many other genes involved in the development of prickles. With all these theories considered and without ruling out any possible explanation we evaluated our candidate genes.

5.3. EVALUATION OF CANDIDATE GENES

5.3.1. SMEL_006g267000.1

According to what we discussed previously, the gene SMEL_006g267000.1 (lysM domain receptor-like kinase 3) is not a bad candidate gene since it has a protein kinase domain and therefore it might be involved in signal transduction mechanisms and cell wall or membrane biogenesis (Buendia et al., 2018). Besides that, in *A. thaliana* the lysM receptor kinase acts as a cell surface receptor involved in cell defence recognizing microbe-derived N-acetylglucosamine (NAG)-containing ligands (Wan et al., 2008) and since prickles constitute another form of defence against herbivores it is possible that this protein had evolved to control the development of prickles in the eggplant so we will further research in more depth this gene.

5.3.2. SMEL_006g267010.1

For SMEL_006g267010.1 we ended up finding that it was a lncRNA with an unknown function in *S. tuberosum* and since many lncRNAs act as genetic regulators up or down-regulating the expression of

hundreds to thousands of genes simultaneously (Liu et al., 2015) it is our most promising candidate gene.

We strongly suspect of this molecule to be responsible for prickliness in eggplant due to the fact that lncRNAs functional roles include gene regulation in plant stress responses (Budak et al., 2020) but more research will be required so as to fully characterize SMEL_006g267010.1 and completely understand its functions and interactions as up to the date they remain unknown.

5.3.3. SMEL_006g267020.1 and SMEL_006g267030.1

Both SMEL_006g267020.1 and SMEL_006g267030.1 are carbonic anhydrases that might be involved in the auxin signalling pathway among other things and since auxins are important phytohormones involved in many growth and developmental processes (Mignolli et al., 2017) we have to consider them as good candidate genes.

5.3.4. SMEL_006g267040.1

In what refers to gene SMEL_006g267040.1 it is not a good candidate gene since it seems to be involved in redox reactions and chloroplast processes (Decros et al., 2019) which are not related to prickliness at all.

5.3.5. SMEL_006g267050.1

SMEL_006g267050.1 is another potential candidate gene since its conserved domains are related to signal transduction mechanisms and intracellular trafficking and secretion as they seem to be involved in the cytokinin biosynthetic process and cytokinins (CKs) are a class of phytohormones involved in cell growth and differentiation (Šmehilová et al., 2016).

5.3.6. SMEL_006g267060.1

When we analysed SMEL_006g267060.1 we could only find a couple of conserved domains and they do not seem to be related to prickliness at first glance since glucosyltransferases participate in the biosynthesis of sugars (Vogt and Jones, 2000) but more research is needed so as to discard it.

5.3.7. SMEL_006g267070.1

Finally, SMEL_006g267070.1 does not seem to be related to prickle development neither since the conserved ATPase domain is involved in carbohydrates transport and the aminotransferase-like domain is found on transposases that do not seem to have anything to do with prickliness. Furthermore, lysine ketoglutarate reductase trans-splicing proteins are involved in the lysine metabolism which is an essential amino acid that constitutes an important building block for proteins and a precursor for glutamate among other functions (Galili, 2002) but none of them seems to be related to prickle development.

5.4. FUTURE PROSPECT

The final objective of this project was to determine the gene controlling prickliness in eggplant and we have not been able to fulfil it as we had to stop the laboratory activities due to the alarm state before achieving it. However, we have started out more experiments to confirm our hypothesis and we have

on mind some other tests that we would like to carry out so as to study in depth prickliness in eggplant and finally be able to pinpoint the gene controlling this trait.

5.4.1. REAL-TIME PCR (RT-PCR) ANALYSIS

We had already started this experiment by performing the RNA extractions for some selected plants from the genotyping project so as to check if we were able to detect differences in what refers to the RNA levels of our candidate genes in prickly and prickleless plants. The idea behind this experiment was the assumption that the candidate gene governing the prickle trait should be more expressed in prickly plants and by measuring the amount of RNA present in each selected sample we could assess those differences and check if they coincided with our most likely genes to be controlling prickliness.

The experiment itself is quite simple and very useful for our purpose, being the limiting and most challenging step the RNA extractions since they are by far more demanding than DNA extractions and, furthermore, when we started out the project we did not have a reliable and efficient RNA extraction protocol and we had to try several ones and to choose and modify one so as to tune it, which took us a lot of time.

5.4.2. CRISPR/CAS KNOCKOUT EXPERIMENT

Another test that was also on the verge of being carried out was a knockout experiment using CRISPR/Cas technology. We wanted to try to silence the most probable candidate genes one by one in several samples so as to check which one of them was provoking the prickly phenotype. This is a large test that can last several months when working with plants but we had already developed the guide RNAs (gRNAs) and synthesized the gRNA oligos before we had to stop all the activity in the lab and we were almost ready to develop the CRISPR library.

We will retake the experiment as soon as we can since we know it is a decisive test from which we will be able to state with a 100% confidence level which candidate gene was actually governing prickliness in eggplant.

5.5. PROJECT RELEVANCE

The main objectives of this project were to delimit the candidate region to which the prickle trait had been associated, to obtain a series of candidate genes for governing prickliness in eggplant, to select a collection of molecular markers closely related to the prickle trait so as to use them in future research projects and to determine the gene governing prickliness in eggplant.

The first two objective have been amply achieved since we began the project with a candidate region of about 2.5 Mb (from positions 104005650 to 106557321 bp) with hundreds of genes on it and we have shortened it to a final region of less than 100 Kb (from positions 105450271 to 105545399 bp) in which we have identified just 8 candidate genes). Furthermore, we have deeply analysed those genes and we have a couple of very promising options that will be further analysed first so as to finally pinpoint the genetics of prickliness in eggplant.

The third objective has also been accomplished since from the 58 primers that we developed to genotype the samples we know SNPs 13 and 14 are the closest flanking markers to the gene governing prickliness and so we can utilize them as markers closely linked to the prickle trait in eggplant in future projects.

The last objective is yet to be accomplished since we have a few candidate genes from which two are very promising but we have not performed any knock-out or over-expression experiment to test our hypothesis and determine the gene controlling the prickle trait in eggplant. However, we have already begun several experiments so as to fulfil this final objective.

Besides those objectives we also tested our suspicion that the prickle trait was monogenic and had a strong mendelian segregation after observing the its segregation once we phenotyped all the samples and that could only be possible if the gene or, at least, the major QTL governing the trait was located in the candidate region we were studying. The statistical analysis confirmed our hypothesis and so we analysed each candidate gene in detail knowing that if a single gene was responsible for prickliness in eggplant it should be a regulatory element controlling the expression of many other genes like a transcription factor or a non-coding RNA.

Finally, we could also extract some findings regarding the sample size after completing the project since previously to our study it had been estimated with density maps that the theoretical percentage of recombination for our region of about 3 Mb was 13% (about 4.3% per Mb). Since we wanted to have as many recombinants as possible in our population so as to have many informative samples that allowed us to shorten the candidate region we selected a sample size of around 600 plants. We expected to have around 80 recombinant plants based on the theoretical value but in the end, we obtained 95 recombinant plants so it is very likely that with a smaller sample size we would have been able to perform the study as well. The more plants the more recombination and thus the more you will be able to fine map the trait of interest but it is also important to take into consideration the available space and the time it takes to phenotype all the plants. In our case, we could afford a big enough sample size since in the pollinator-free greenhouses we had plenty of room for our plants and because the phenotyping process for the prickle trait is easy and fast and can be done in the early stages of the plant development but it is always interesting to optimize the sample size for research projects and we will take into consideration the information we have obtained in future experiments.

6. CONCLUSIONS

We can state the following conclusions after having fulfilled this project:

- Prickliness in eggplant is a monogenic character with a strong mendelian segregation.
- The major QTL governing prickliness in eggplant has been associated to a region of 95. 128 Kb in the chromosome 6.
- We have reduced the list of candidate genes for controlling the prickle trait in eggplant to 8 candidate genes from which three of them of them are very promising.
- We have confirmed the effectiveness of HRM as a method for fine mapping a trait of interest even in the absence of an accurate reference genome with great accuracy.
- We have not been able to determine the exact gene governing prickliness in eggplant with the analysis that have been carried out but two more experiments have been proposed and have already been initiated whose results will be crucial in order to pinpoint the gene controlling the prickle trait.

- A series of closely linked and flanking markers to the prickle trait in eggplant have been developed so as to be used in future research projects.
- The obtained results will allow us to further understand the genetics of prickliness in eggplant and to apply this knowledge in plant breeding programs in the future.

7. BIBLIOGRAPHY

ALONSO-BLANCO, C., KOORNNEEF, M., & VAN OOIJEN, J. W. (2006). QTL analysis. In *Arabidopsis Protocols* (pp. 79-99). Humana Press.

ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W. & LIPMAN, D. J. (1990) Basic local alignment search tool. *Journal of molecular biology*, *215*(3), 403-410.

ASHBURNER, M., BALL, C. A., BLAKE, J. A., BOTSTEIN, D., BUTLER, H., CHERRY, J. M., ... & HARRIS, M. A. (2000). Gene ontology: tool for the unification of biology. *Nature genetics*, *25*(1), 25-29.

APWEILER, R., BAIROCH, A., WU, C. H., BARKER, W. C., BOECKMANN, B., FERRO, S., ... & MARTIN, M. J. (2004). UniProt: the universal protein knowledgebase. *Nucleic acids research*, *32*(suppl_1), D115-D119.

BARCHI, L., DELLEDONNE, M., LANTERI, S., DAL MOLIN, A., MINIO, A., FERRARINI, A., ... & BASSOLINO, L. (2016). A high quality eggplant genome sequence: a new tool for the analysis of Solanaceae family evolution and for the molecular deciphering of complex traits. In *20th EUCARPIA General Congress* (pp. 122-122). Agroscope.

BARCHI, L., ACQUADRO, A., ALONSO, D., APREA, G., BASSOLINO, L., DEMURTAS, O. C., ... & SCAGLIONE, D. (2019a). Single Primer Enrichment Technology (SPET) for high-throughput genotyping in tomato and eggplant germplasm. *Frontiers in plant science*, *10*, 1005.

BARCHI, L., PIETRELLA, M., VENTURINI, L., MINIO, A., TOPPINO, L., ACQUADRO, A., ... & COMINO, C. (2019b). A chromosome-anchored eggplant genome sequence reveals key events in Solanaceae evolution. *Scientific reports*, *9*(1), 1-13.

BHAT, J. A., ALI, S., SALGOTRA, R. K., MIR, Z. A., DUTTA, S., JADON, V., ... & SINGH, G. P. (2016). Genomic selection in the era of next generation sequencing for complex traits in plant breeding. *Frontiers in genetics*, *7*, 221.

BJÖRKMAN, C., & ANDERSON, D. B. (1990). Trade-off among antiherbivore defences in a South American blackberry (Rubus bogotensis). *Oecologia*, *85*(2), 247-249.

BOHS, L. (2005) Major clades in Solanum based on ndhF sequences. In: Keating RC, Hollowell VC, Croat TB (eds) A festschrift for William G. D'Arcy: the legacy of a taxonomist, Monographs in systematic botany from the Missouri Botanical Garden, vol 104. Missouri Botanical Garden Press, St. Louis, pp 27–49

BUDAK, H., KAYA, S. B., & CAGIRICI, H. B. (2020). Long Non-coding RNA in Plants in the Era of Reference Sequences. *Frontiers in Plant Science*, *11*.

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*, *9*, 1531.

CHATEIGNER-BOUTIN, A. L., & SMALL, I. (2007). A rapid high-throughput method for the detection and quantification of RNA editing based on high-resolution melting of amplicons. *Nucleic acids research*, *35*(17), e114.

CLARK, K., KARSCH-MIZRACHI, I., LIPMAN, D. J., OSTELL, J., & SAYERS, E. W. (2016). GenBank. *Nucleic acids research*, *44*(D1), D67–D72.

COLLARD, B. C., JAHUFER, M. Z. Z., BROUWER, J. B., & PANG, E. C. K. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. *Euphytica*, *142*(1-2), 169-196.

COLLARD, B. C., & MACKILL, D. J. (2008). Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *363*(1491), 557–572.

COYNER, M. A., SKIRVIN, R. M., NORTON, M. A., & OTTERBACHER, A. G. (2005). Thornlessness in blackberries: A review. *Small fruits review*, *4*(2), 83-106.

CROXFORD, A. E., ROGERS, T., CALIGARI, P. D., & WILKINSON, M. J. (2008). High-resolution melt analysis to identify and map sequence-tagged site anchor points onto linkage maps: a white lupin (Lupinus albus) map as an exemplar. *New Phytologist*, *180*(3), 594-607.

DAUNAY, M. C., and HAZRA, P. (2012). "Eggplant," in Handbook of Vegetables, eds K. V. Peter and P. Hazra (Houston, TX: Studium Press), 257–322.

DECROS, G., BEAUVOIT, B., COLOMBIÉ, S., CABASSON, C., BERNILLON, S., ARRIVAULT, S., ... & GIBON, Y. (2019). Regulation of pyridine nucleotide metabolism during tomato fruit development through transcript and protein profiling. *Frontiers in Plant Science*, *10*, 1201.

DOGANLAR, S., FRARY, A., DAUNAY, M. C., LESTER, R. N., & TANKSLEY, S. D. (2002). Conservation of gene function in the Solanaceae as revealed by comparative mapping of domestication traits in eggplant. *Genetics*, *161*(4), 1713-1726.

Doyle, J.J., Doyle, J.L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin 19, 11-15.

EDUARDO, I., ARÚS, P., & MONFORTE, A. J. (2005). Development of a genomic library of near isogenic lines (NILs) in melon (Cucumis melo L.) from the exotic accession PI161375. *Theoretical and Applied Genetics*, *112*(1), 139-148.

ELSHIRE, R. J., GLAUBITZ, J. C., SUN, Q., POLAND, J. A., KAWAMOTO, K., BUCKLER, E. S., & MITCHELL, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS one*, *6*(5), e19379.

ESHED, Y., & ZAMIR, D. (1995). An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. *Genetics* 141(3), 1147–1162.

FAOSTAT (2018). FAOSTAT Production Databases. Visto el 13 de mayo de 2020, disponible online en: http://www.fao.org/faostat/en/#home

FERNANDEZ-POZO, N., MENDA, N., EDWARDS, J. D., SAHA, S., TECLE, I. Y., STRICKLER, S. R., ... & YAN, A. (2015). The Sol Genomics Network (SGN)—from genotype to phenotype to breeding. *Nucleic acids research*, *43*(D1), D1036-D1041.

FRARY, A., DOGANLAR, S., & DAUNAY, M. C. (2007). Eggplant. In *Vegetables* (pp. 287-313). Springer, Berlin, Heidelberg.

FRARY, A., FRARY, A., DAUNAY, M. C., HUVENAARS, K., MANK, R., & DOGANLAR, S. (2014). QTL hotspots in eggplant (Solanum melongena) detected with a high resolution map and CIM analysis. *Euphytica*, *197*(2), 211-228.

FULTON, T. M., BECK-BUNN, T., EMMATY, D., ESHED, Y., LOPEZ, J., PETIARD, V., ... & TANKSLEY, S. D. (1997). QTL analysis of an advanced backcross of Lycopersicon peruvianum to the cultivated tomato and comparisons with QTLs found in other wild species. *Theoretical and Applied Genetics*, *95*(5-6), 881-894.

GALILI, G. (2002). New insights into the regulation and functional significance of lysine metabolism in plants. *Annual* Review *of Plant Biology*, *53*(1), 27-43.

GANAL, M. W., WIESEKE, R., LUERSSEN, H., DURSTEWITZ, G., GRANER, E. M., PLIESKE, J., & POLLEY, A. (2014). High-throughput SNP profiling of genetic resources in crop plants using genotyping arrays. In *Genomics of Plant Genetic Resources* (pp. 113-130). Springer, Dordrecht.

GRAMAZIO, P.; PROHENS, J.; PLAZAS, M.; ANDÚJAR, I.; HERRAIZ, F.J.; CASTILLO, E., … & VILANOVA, S. (2014). Location of chlorogenic acid biosynthesis pathway and polyphenol oxidase genes in a new interspecific anchored linkage map of eggplant. BMC Plant Biology, 1*4*(1), 350.

GRAMAZIO, P., PROHENS, J., PLAZAS, M., MANGINO, G., HERRAIZ, F. J., & VILANOVA, S. (2017). Development and genetic characterization of advanced backcross materials and an introgression line population of Solanum incanum in a S. melongena background. *Frontiers in plant science*, *8*, 1477.

GRAMAZIO, P., YAN, H., HASING, T., VILANOVA, S., PROHENS, J., & BOMBARELY, A. (2019). Whole-Genome resequencing of seven eggplant (Solanum melongena) and one wild relative (S. incanum) accessions provides new insights and breeding tools for eggplant enhancement. *Frontiers in plant science*, *10*, 1220.

GRIFFITHS, A. J., MILLER, J. H., SUZUKI, D. T., LEWONTIN, R. C., & GELBART, W. M. (2000). Mendel's experiments. In *An Introduction to Genetic Analysis. 7th edition*. WH Freeman. Chi-square test.

HAJJAR, R., & HODGKIN, T. (2007). The use of wild relatives in crop improvement: a survey of developments over the last 20 years. *Euphytica*, *156*(1-2), 1-13.

HE, J., ZHAO, X., LAROCHE, A., LU, Z. X., LIU, H., & LI, Z. (2014). Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Frontiers in plant science*, *5*, 484.

HUNTER, S., APWEILER, R., ATTWOOD, T. K., BAIROCH, A., BATEMAN, A., BINNS, D., ... & FINN, R. D. (2009). InterPro: the integrative protein signature database. *Nucleic acids research*, *37*(suppl_1), D211-D215.

KAUSHIK, P., PROHENS, J., VILANOVA, S., GRAMAZIO, P., & PLAZAS, M. (2016). Phenotyping of eggplant wild relatives and interspecific hybrids with conventional and phenomics descriptors provides insight for their potential utilization in breeding. *Frontiers in Plant Science*, *7*, 677.

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. *Botany*, *89*(4), 217-226.

KNAPP, S., VORONTSOVA, M. S., & PROHENS, J. (2013). Wild relatives of the eggplant (Solanum melongena L.: Solanaceae): new understanding of species names in a complex group. *PloS one*, *8*(2), e57039.

KÕRESSAAR, T., LEPAMETS, M., KAPLINSKI, L., RAIME, K., ANDRESON, R., & REMM, M. (2018). Primer3_masker: integrating masking of template sequence with primer design software. *Bioinformatics*, *34*(11), 1937-1938.

KRIVENTSEVA, E. V., KUZNETSOV, D., TEGENFELDT, F., MANNI, M., DIAS, R., SIMÃO, F. A., & ZDOBNOV, E. M. (2019). OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic acids research*, *47*(D1), D807-D811.

KRUGLYAK, L. (1997). The use of a genetic map of biallelic markers in linkage studies. *Nature genetics*, *17*(1), 21-24.

LANDER, E. S. (1996). The new genomics: global views of biology. *Science*, *274*(5287), 536-539.

LEVIN, R. A., MYERS, N. R., & BOHS, L. (2006). Phylogenetic relationships among the "spiny solanums" (Solanum subgenus Leptostemonum, Solanaceae). *American journal of botany*, *93*(1), 157-169.

LIU, J., WANG, H., & CHUA, N. H. (2015). Long noncoding RNA transcriptome of plants. *Plant biotechnology journal*, *13*(3), 319-328.

LONGIN, C. F. H., & REIF, J. C. (2014). Redesigning the exploitation of wheat genetic resources. *Trends in Plant Science*, *19*(10), 631-636.

MAXTED, N., FORD-LLOYD, B. V., JURY, S., KELL, S., & SCHOLTEN, M. (2006). Towards a definition of a crop wild relative. *Biodiversity & Conservation*, *15*(8), 2673-2685.

MENNELLA, G.; LO SCALZO, R.; FIBIANI, M.; D'ALESSANDRO, A.; FRANCESE, G.; TOPPINO, L.; ROTINO, G. L. (2012). Chemical and Bioactive Quality Traits During Fruit Ripening in Eggplant (S. melongena L.) and Allied Species. *Journal of Agricultural and Food Chemistry*, 60(47), 11821–11831.

MERING, C. V., HUYNEN, M., JAEGGI, D., SCHMIDT, S., BORK, P., & SNEL, B. (2003). STRING: a database of predicted functional associations between proteins. *Nucleic acids research*, *31*(1), 258-261.

MEYER, R. S.; KAROL, K. G.; LITTLE, D. P.; NEE, M. H.; y LITT, A. (2012). Phylogeographic relationships among Asian eggplants and new perspectives on eggplant domestication. *Molecular Phylogenetics and Evolution*, 63(3), 685–701

HALLER, M. (2018). Molecular tools to advance introgression breeding in eggplant (Master's thesis). Department of Environmental System Sciences of ETH Zurich Institute of Agricultural Sciences. Zurich, Germany.

MIGNOLLI, F., MARIOTTI, L., PICCIARELLI, P., & VIDOZ, M. L. (2017). Differential auxin transport and accumulation in the stem base lead to profuse adventitious root primordia formation in the aerial roots (aer) mutant of tomato (Solanum lycopersicum L.). *Journal of plant physiology*, *213*, 55-65.

MORELL, M. K., PEAKALL, R., APPELS, R., PRESTON, L. R., & LLOYD, H. L. (1995). DNA profiling techniques for plant variety identification. *Australian Journal of Experimental Agriculture*, *35*(6), 807-819.

MUÑOZ-FALCÓN, J. E., PROHENS, J., VILANOVA, S., & NUEZ, F. (2009). Diversity in commercial varieties and landraces of black eggplants and implications for broadening the breeders' gene pool. *Annals of Applied Biology*, *154*(3), 453-465.

PANDEY, S., GOEL, R., BHARDDWAJ, A., ASIF, M. H., SAWANT, S. V., & MISRA, P. (2018). Transcriptome analysis provides insight into prickle development and its link to defense and secondary metabolism in Solanum viarum Dunal. *Scientific reports*, *8*(1), 1-12.

PATERSON, A. H., TANKSLEY, S. D., & SORRELS, M. E. (1991). DNA markers in plant improvement. In *Advances in agronomy* (Vol. 46, pp. 39-90). Academic Press.

PEARSON, K. (1900). X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *50*(302), 157-175.

PETERS, J. L., CNUDDE, F., & GERATS, T. (2003). Forward genetics and map-based cloning approaches. *Trends in plant science*, *8*(10), 484-491.

PORTIS E, CERICOLA F, BARCHI L, TOPPINO L, ACCIARRI N, PULCINI L., … & ROTINO, G. L. (2015) Association Mapping for Fruit, Plant and Leaf Morphology Traits in Eggplant. PLoS ONE 10(8): e0135200.

PROHENS, J., GRAMAZIO, P., PLAZAS, M., DEMPEWOLF, H., KILIAN, B., DÍEZ, M. J., … & KNAPP, S. (2017). Introgressiomics: a new approach for using crop wild relatives in breeding for adaptation to climate change. *Euphytica* 213:158.

RANIL, R. H. G., NIRAN, H. M. L., PLAZAS, M., FONSEKA, R. M., FONSEKA, H. H., VILANOVA, S., ... & PROHENS, J. (2015). Improving seed germination of the eggplant rootstock Solanum torvum by testing multiple factors using an orthogonal array design. *Scientia Horticulturae*, *193*, 174-181.

ROTINO, G. L., SALA, T., & TOPPINO, L. (2014). Eggplant. In *Alien Gene Transfer in Crop Plants, Volume 2* (pp. 381-409). Springer, New York, NY.

ROZEN, S., & SKALETSKY, H. (2000). Primer3 on the WWW for general users and for biologist programmers. In *Bioinformatics methods and protocols* (pp. 365-386). Humana Press, Totowa, NJ.

SACCO, A., DI MATTEO, A., LOMBARDI, N., TROTTA, N., PUNZO, B., MARI, A., & BARONE, A. (2013). Quantitative trait loci pyramiding for fruit quality traits in tomato. *Molecular Breeding*, 31(1), 217-222.

SCAGLIONE, D., PINOSIO, S., MARRONI, F., DI CENTA, E., FORNASIERO, A., MAGRIS, G., ... & MORGANTE, M. (2019). Single primer enrichment technology as a tool for massive genotyping: a benchmark on black poplar and maize. *Annals of botany*, *124*(4), 543-551.

SCHAID, D. J., CHEN, W., & LARSON, N. B. (2018). From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nature reviews. Genetics*, *19*(8), 491–504.

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.

ŠMEHILOVÁ, M., DOBRůSKOVÁ, J., NOVÁK, O., TAKÁC, T., & GALUSZKA, P. (2016). Cytokinin-specific glycosyltransferases possess different roles in cytokinin homeostasis maintenance. *Frontiers in plant science*, *7*, 1264.

GETACHEW, E., HERMINE, N. B., BELL, J. M., WOSENE, G. (2017). Genotyping by Sequencing for Plant Breeding- A Review. Advances in Biotechnology & Microbiology journal, 15(4).

SIMPSON, M. G. (2010). "Plant Morphology". In: *Plant Systematics, 2nd. edition*. Elsevier Academic Press. Chapter 9.

SUKUMARAN, S., & YU, J. (2014). Association mapping of genetic resources: achievements and future perspectives. In *Genomics of plant genetic resources* (pp. 207-235). Springer, Dordrecht.

SYMON, D. (1986). A Survey of Solanum Prickles and Marsupial Herbivory in Australia. *Annals of the Missouri Botanical Garden,73*(4), 745-754.

SZCZEŚNIAK, M. W., ROSIKIEWICZ, W., & MAKALOWSKA, I. (2016). CANTATAdb: A Collection of Plant Long Non-Coding RNAs. *Plant & cell physiology*, *57*(1), e8.

TALUKDAR, D. (2013). Leguminosae. In S. Maloy & K. Hughes (Eds.), Brenner's Encyclopedia of Genetics (Second Edition) (pp. 212-216). Academic Press.

TANKSLEY, S. D. (1993). Mapping polygenes. *Annual review of genetics*, *27*(1), 205-233.

TANKSLEY, S. D., & NELSON, J. C. (1996). Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. *Theoretical and Applied Genetics*, *92*(2), 191-203.

TANKSLEY, S. D., & MCCOUCH, S. R. (1997). Seed banks and molecular maps: Unlocking genetic potential from the wild. *Science,* 277(5329), 1063-1066.

TATTINI, M., GRAVANO, E., PINELLI, P., MULINACCI, N., & ROMANI, A. (2000). Flavonoids accumulate in leaves and glandular trichomes of Phillyrea latifolia exposed to excess solar radiation. *New phytologist*, *148*(1), 69-77.

TSUCHIHASHI, Z., & DRACOPOLI, N. C. (2002). Progress in high throughput SNP genotyping methods. *The pharmacogenomics journal*, *2*(2), 103-110.

VILANOVA, S., ALONSO, D., GRAMAZIO, P., PLAZAS, M., FERRANTE, P., SCHMIDT, M., DIEZ-NICLÓS, M. J., USADEL, B., GIULANO, G., PROHENS, J. (2020). SILEX: A fast and inexpensive high-quality DNA extraction method suitable for multiple sequencing platforms and recalcitrant plant species.

VILLANUEVA, G., ALONSO, D., GARCÍA-FORTEA, E., MANGINO, G., VILANOVA, S., PROHENS, J., PLAZAS, M., GRAMAZIO, P. (diciembre, 2018). Mapeo del carácter espinosidad en berenjena (*Solanum melongena*). I Congreso de Jóvenes Investigadores en Ciencias Agroalimentarias, Almería, España.

VOGT, T., & JONES, P. (2000). Glycosyltransferases in plant natural product synthesis: characterization of a supergene family. *Trends in plant science*, *5*(9), 380-386.

WAN, J., ZHANG, X. C., NEECE, D., RAMONELL, K. M., CLOUGH, S., KIM, S. Y., ... & STACEY, G. (2008). A LysM receptor-like kinase plays a critical role in chitin signaling and fungal resistance in Arabidopsis. *The Plant Cell*, *20*(2), 471-481.

WEESE, T. L. & BOHS, L. (2007). A Three-Gene Phylogeny of the Genus Solanum (Solanaceae). *Systematic Botany*, 32(2), 445–463.

WERNER, M., SYCH, M., HERBON, N., ILLIG, T., KÖNIG, I. R., & WJST, M. (2002). Large-scale determination of SNP allele frequencies in DNA pools using MALDI-TOF mass spectrometry. *Human mutation*, *20*(1), 57-64.

WILLHELM, J., & PINGOUD, A. (2003). Real-time polymerase chain reaction. *Chembiochem*, *4*(11), 1120-1128.

WITTWER, C. T., REED, G. H., GUNDRY, C. N., VANDERSTEEN, J. G., & PRYOR, R. J. (2003). High-resolution genotyping by amplicon melting analysis using LCGreen. *Clinical chemistry*, *49*(6), 853-860.

WITTWER, C. T. (2009). High-resolution DNA melting analysis: advancements and limitations. *Human mutation*, *30*(6), 857-859.

YIN, X., STRUIK, P. C., GU, J., & WANG, H. (2016). Modelling QTL-trait-crop relationships: past experiences and future prospects. In *Crop systems biology* (pp. 193-218). Springer, Cham.

ZAMIR, D. (2001). Improving plant breeding with exotic genetic libraries. *Nature reviews genetics*, *2*(12), 983-989.

ZHAO, Z., & BOERWINKLE, E. (2002). Neighboring-nucleotide effects on single nucleotide polymorphisms: a study of 2.6 million polymorphisms across the human genome. *Genome research*, *12*(11), 1679-1686.