FINE MAPPING OF GENES DETERMINING VICINE-CONVICINE CONCENTRATION IN FABA BEAN

Henri Laugel

Master's thesis

Erasmus Mundus Master Programme in Plant Breeding (emPLANT)

Master of Sciences, Faculty of Agriculture and Forestry

University of Helsinki, Department of Agricultural Sciences

May 2022

ABSTRACT

UNIVERSITY OF HELSINKI

Faculty	Department
Faculty of Agriculture and Forestry	Department of Agricultural Sciences

Author
Henri Laugel

Title
FINE MAPPING OF GENES DETERMINING VICINE-CONVICINE CONCENTRATION IN FABA BEAN

Subject
Plant Breeding

Level	Month and year	Number of pages
M.Sc. Thesis	May 2022	47

Abstract

Faba bean (*Vicia faba* L.*)* is an annual herbaceous cool-season food legume widely cultivated worldwide, especially for its high seed protein content. However, its major limitation in being used as food and feed, is the presence of antinutritional factors in its seeds, especially vicine and convicine (VC), two related compounds, which may be harmful to livestock and G6PD-deficient humans. To remove VC, the most sustainable method is breeding for low-VC faba bean cultivars. To improve the efficiency and speed of breeding programs, breeders use marker-assisted selection (MAS). The identification of genes responsible for VC content allows the development of reliable DNA markers and a better understanding of the molecular basis of this trait. The major-effect QTL controlling VC content named "*VC1*", was identified in faba bean chromosome 1, and a few minor-effect QTLs were detected in previous studies. Hence, a total of 165 RILs from the cross Mélodie/2 (low-VC) x ILB 938/2 (high-VC) were genotyped and evaluated for VC content. Composite interval mapping was run on R/qtl software with accurate phenotypic data associated with a high-density SNP-based genetic map. Results revealed two minor-effect QTLs in addition to *VC1*. One was on chromosome 4 and had about 15% effect on convicine content. The other was on chromosome 5 and had 15% effect on vicine and total VC content. This research also reports candidate genes for the newly detected minor-effect QTLs through comparative genomics with the *Medicago truncatula* genome. Hypotheses were proposed on the role of these candidate genes on the VC biosynthetic pathway or transportation into the embryo beans for further testing.

Keywords
Faba bean, vicine, convicine, QTL, linkage mapping, candidate gene identification

Where deposited
Master's Programme in Agricultural Sciences, Department of Agricultural Sciences

Further information
Supervised by Dr. Frederick Stoddard and Dr. Hamid Khazaei

Table of contents

Abbreviations

AFLP	Amplified Fragment Length Polymorphism
ARPDP	5-Amino-6-Ribosylamino-2,3(1H,3H)-PyrimiDinedione 5'-Phosphate
BPL	Bean Pure Line
BNF	Biological Nitrogen Fixation
CIM	Composite Interval Mapping
cM	Centimorgan
CMS	Cytoplasmic Male Sterility
CTAB	CetylTrimethylAmmonium Bromide
DARPP	2,5-DiAmino-6-Ribosylamino-4(3H)-Pyrimidinone 5'-Phosphate
EST	Expressed Sequence Tag
FAIMS	Field Asymmetric waveform Ion Mobility Spectrometry
FIGS	Focused Identification of Germplasm Strategy
G6PD	Glucose-6-Phosphate Dehydrogenase
GWAS	Genome-Wide Association Study
HC	Hilum Color
HK	Haley-Knott regression method
HPLC	High-Performance Liquid Chromatography
ICARDA	International Center for Agricultural Research in Dry Areas
IM	Interval Mapping
INRAE	National Research Institute for Agriculture, Food and the Environment
KASP	Kompetitive Allele Specific PCR
LG	Linkage Group
MAS	Marker-Assisted Selection
MS	Mass Spectroscopy
Mt	*Medicago truncatula*
PCR	Polymerase Chain Reaction
PPR	PentatricoPeptide Repeat
PVE	Phenotypic Variation Explained
QTL	Quantitative Trait Locus
RIL	Recombinant Inbred Line

SNP Single-Nucleotide Polymorphism

SSD Single Seed Descent

VC Vicine and Convicine

1 INTRODUCTION

Agriculture faces many challenges, which are even more challenging due to global warming. The European production of grain legumes is deficient, representing only 30% of the supplemental protein source for animal feed (Neugschwandtner et al. 2015). Because of that, Europe is largely dependent on soybean imports from North and South America. To limit their environmental impact and to be more autonomous, it has become urgent for Europe to find alternative crops to reduce soybean imports and develop by itself its grain legume production. Faba bean (*Vicia faba* L.) seems to be an excellent solution. It is a cool-season grain legume crop cultivated essentially for its protein-rich seeds (Robinson et al. 2019). It is adapted to a wide range of climates, complementing soybean. Faba bean is grown where it is too cool for soybean, and soybean where it is too warm for faba bean. Within a cropping system, it diversifies the cropping rotation cycle, reducing diseases, pests and weeds in the field, and brings nitrogen through its ability to fix atmospheric nitrogen using its symbiosis with *Rhizobium* bacteria, lowering costs and environmental impacts related to fertilizers (Duc et al. 2015). Although it has many advantages, faba bean cultivation is still limited, mainly because of relatively little breeding efforts. It is susceptible to many abiotic and biotic stresses, and its yield remains unstable (Annicchiarico & Iannucci 2008). In addition, its seeds contain vicine and convicine (VC), two antinutritional compounds, making consumers and food-processing industries reluctant to use it.

One major-effect quantitative trait locus (QTL) on faba bean chromosome 1, "*VC1*", was identified for VC, and its implication in VC biosynthesis has been recently investigated (Björnsdotter et al. 2021). Genomic research in faba bean is still limited because no reference genome has been made publicly available yet, also explaining why breeding efforts are slower compared to the major cereals and oilseeds. The discovery and identification of *VC1* is accelerating the breeding of new low-VC faba bean cultivars worldwide. This is primarily due to the development of a high-throughput KASP (Kompetitive Allele Specific PCR) marker from the *VC1* gene (Khazaei et al. 2017), speeding up the selection through marker-assisted selection (MAS). However, unpublished data suggests that VC content is not monogenic. The presence of quantitative variation within both low-VC and high-VC faba beans may be explained by minor-effect QTLs.

The present thesis aims to map the potential minor-effect QTLs for VC. To do this, a high-density SNP-based genetic map was constructed to define the location of the potential QTLs and identify candidate genes to understand better the genetic basis of this trait. Once the genes responsible for VC content are identified, markers will be developed in order to assist breeders in a breeding process aiming to breed for VC-free faba bean cultivars.

2 LITERATURE REVIEW

2.1 Faba bean (*Vicia faba* L.)

Faba bean (*Vicia faba* L.) is a cool-season grain legume belonging to the Fabaceae family. It is a 1-2 m tall annual herbaceous plant with alternate and pinnate leaves, composed of 5-10 cm long leaflets fixed on a hollow square-shaped stem (Kirk 2004, Duc et al. 2015). Racemes grow at the leaf axils and produce from one to eight 3 cm long zygomorphic flowers with five white petals and a purple spot on both wing petals (Kirk 2004).

Faba bean is an annual crop cultivated for its very nutritious seeds, which have a high protein content (24-35% of seed dry matter), including globulin (79%), albumin (7%) and glutenin (6%) (Crépon et al. 2010, Longobardi et al. 2015). In addition, they contain many nutrients (K, Ca, Mg, Fe, and Zn) and bioactive compounds (polyphenols, carotenoids and carbohydrates) (Karkanis et al. 2018). Faba beans are mainly used as high-protein fodder or silage for farm animals or as a vegetable in fresh or cooked form (Kirk 2004, CABI 2022).

2.1.1 Center of origin and genetic resources

Faba bean is one of the first domesticated crops, and has been cultivated for thousands of years since the early Neolithic. Its place of origin and wild ancestor are still unknown (Kirk 2004, Duc et al. 2015). The main hypothesis supported by Cubero (1974) supposed that its center of origin is in the Near East in the Fertile Crescent, where its use as domesticated species started. From this point, faba bean would spread across four routes towards (1) Europe, (2) the Maghreb and Spain, (3) Ethiopia and (4) India. Later, Zong et al. (2009) demonstrated that the Chinese faba bean gene pool is isolated from the rest of the world, suggesting that China is a secondary center of genetic diversity.

Faba bean genetic resources are restricted to the cultivated form, as it is isolated reproductively from other species (Gnanasambandam et al. 2012). Nevertheless, it has a vast genetic diversity: 38,000 accessions are conserved worldwide within 43 national gene banks and the international center for agricultural research in dry areas (ICARDA), which gathers more than

10,000 of them (Duc et al. 2015). Several methods have been developed to make an efficient use of these genetic resources, including core collections to collect fewer accessions representing the genetic variation for a specific trait or focused identification of germplasm strategy (FIGS), see Khazaei et al. (2013), which is based on accession passport and collection site data.

Faba bean accessions are classified into four subspecies based on bean size: (1) var. *minor* with small but thick beans (0.4-0.6 g); (2) var. *equina* with medium-sized beans (0.6-1.0 g), known as the horse bean; (3) var. *major* with flat but large beans (1.0-3.0 g), known as the broad bean; and (4) var. *paucijuga*, a primitive form with small beans (0.3-0.4 g) (Kirk 2004, Khazaei et al. 2019). Faba bean can adapt to a variety of climates, especially cool temperatures, probably due to its early domestication and wide genetic diversity (Duc et al. 2015). Accessions are also classified according to their climatic adaptation and sowing date into winter and spring types (Karkanis et al. 2018). It is cultivated as a winter crop in mild-winter regions like the Mediterranean area, Southern China and Australia, and as a spring crop in hard-winter regions like continental Europe and Northern China (Redden et al. 2014).

2.1.2 Production

In 2020, among pulse crops (FAO 2015), faba bean had the highest mean yield (2.1 t/ha) (FAOSTAT 2022), the highest protein content (Robinson et al. 2019) and is the seventh most important production worldwide behind groundnut (*Arachis hypogaea* L.), common bean (*Phaseolus vulgaris* L.), chickpea (*Cicer arietinum* L.), pea (*Pisum sativum* L.), cowpea (*Vigna unguiculata* (L.) Walp.) and lentil (*Lens culinaris* Medik.) (FAOSTAT 2022). The world production was 5.67 Tg on 2.67 Mha. China is the leading producing country with 1.72 Tg and almost 1 Mha of area harvested. Ethiopia is the second most important producer with 1.07 Tg on 0.5 Mha. Together, both countries represent half of the overall production, respectively, 30% and 19% (FAOSTAT 2022).

Faba bean has many advantages in cropping systems. Like most legumes, it can create nodules from its symbiotic association with *Rhizobium* bacteria to fix atmospheric nitrogen through biological nitrogen fixation (BNF) (Sprent 2009, Duc et al. 2015, Karkanis et al. 2018). Neugschwandtner et al. (2015) concluded that faba bean let more nitrate residues in the soil and nitrogen residues on the surface than winter wheat, and measured an average N fixation of 219 kg

ha^{-1}. Faba bean is slotted into the crop rotation cycle in-between two main crops, in intermediate cropping, or in intercropping. Through the nitrogen input from BNF, it reduces nitrogen fertilization, increases soil fertility and improves the yield and seed protein concentration of the following cereal crop by exploiting the “break-crop effect” (Jensen et al. 2010, Karkanis et al. 2018, Klippenstein et al. 2022). By diversifying the rotation, it increases the biodiversity while reducing diseases, pests and weeds (Jensen et al. 2010). It is also used as green manure to improve soil fertility and structure through nitrogen and carbon input (Kirk 2004, Jensen et al. 2010, CABI 2022). In intercropping, high-biomass faba bean cultivars can be used as a bioenergy crop to produce biofuels (Jensen et al. 2010).

However, faba bean production remains limited; its total production area was halved in the last 60 years (FAOSTAT 2022). In some regions, farmers are reluctant to grow faba bean because of the high seed cost, particularly for genotypes with larger seeds (Duc et al. 2015), and its yield instability caused by the lack of sufficient breeding efforts to face abiotic and biotic stresses (Annicchiarico & Iannucci 2008). Food-processing industries and consumers are reluctant because of antinutritional compounds contained in its beans, especially vicine-convicine (VC) and tannins (Crépon et al. 2010).

2.1.3 Breeding

Breeding efforts for faba bean are relatively small, proportionally to its cultivated area (Duc et al. 2015). Breeding programs achieved to improve yield and lodging resistance (Hughes et al. 2020), and to develop cultivars with appropriate earliness to improve faba bean adaptability and resistance to abiotic stresses (Duc et al. 2015). Hence, faba bean average yield doubled in the last 60 years (FAOSTAT 2022). However, many breeding efforts are still needed for yield stability, resistance to biotic and abiotic stresses, and bean quality to face the challenges of climate change and mitigate the reluctance of farmers and food-processing industries.

Faba bean is a partially cross-pollinated crop. Its allogamy varies from 1 to 79% according to genotype, environment and population of pollinators (McVetty and Nugent-Rigby 1984, Gnanasambandam et al. 2012). Genotype determines cytoplasmic-nuclear male sterility (CMS), flower architecture and attractivity of pollinators through colors or scents (Duc et al. 2015). Faba bean has a mixed breeding system (Murphy-Bokern et al. 2017). It can be considered either a self-

pollinated or a cross-pollinated crop, depending on its outcrossing rate (McVetty and Nugent-Rigby 1984). As a self-pollinated crop, line breeding is performed by limiting cross-pollination by pollinators through isolation strategies, especially with the use of screen houses. However, the reproduction of a whole population without the help of pollinators might be very challenging (Gnanasambandam et al. 2012). As a cross-pollinated crop, faba bean is bred through recurrent mass selection, especially using the superior bean pure line (BPL) collection of ICARDA (Gnanasambandam et al. 2012). On the other hand, it was proposed to develop hybrid varieties using CMS, but it is limited by the high seed production costs. Nevertheless, the faba bean heterosis effect is used through the development of synthetic cultivars (Gnanasambandam et al. 2012, Adhikari et al. 2021).

2.1.4 Genomics

Faba bean is a diploid species with six chromosomes (2n = 12). It has a large genome size of 13,000 Mbp, making functional genomics long and costly in this species (Sato et al. 2010). There is no reference genome publicly available yet, making mapping and identification of genes of interest still limited. Molecular markers were developed based on synteny with other related species already sequenced. *Medicago truncatula* (*Mt*) and *Lotus japonica* were both developed as model legumes because they are diploid and short autogamous plants with small genomes (respectively 500 and 470 Mbp) (Sato et al. 2007), short life cycles, and the ability to be transformed (Rispail et al. 2010). Soybean and chickpea were also studied since there are economically important crops (Rispail et al. 2010). Kaur et al. (2014) compared the faba bean genome to chickpea, soybean, *Mt* and *L. japonicus* and obtained, 95%, 90%, 86% and 72% expressed sequence tags (EST) matches, respectively. Those four species showed a high collinearity with faba bean. Identifying single nucleotide polymorphism (SNP) from their whole genome sequences allows the development of SNP markers suitable for faba bean genotyping (Kaur et al. 2014).

2.2 Vicine and convicine

VC are located in the faba bean cotyledons and represent 1% of faba bean dry weight (Khamassi et al. 2013, Murphy-Bokern et al. 2017). They are related pyrimidine glycosides hydrolyzed into two oxidizing agents called divicine and isouramil by β-glucosidase (Khamassi et al.

2013, Björnsdotter et al. 2021). Griffiths and Ramsay (1992) speculated that they are produced because they have an antibiotic effect that protects faba bean against phytophagous pests during seed germination. They suggested that faba bean might have a higher VC content when grown in wet conditions since it is more favorable for the development of pests.

2.2.1 Favism

Their consumption causes favism, an acute hemolytic anemia, in humans affected by the glucose-6-phosphate dehydrogenase (G6PD) deficiency. VC cause oxidative stress by oxidizing intracellular glutathiones that deficient people cannot reduce back. The redox homeostasis is disrupted, resulting in an aggregation of red blood cells in vessels. The deficiency is present mainly in malaria-endemic countries, including the Mediterranean Basin, the Middle East, and North Africa. It is related to malaria as it reduces its severity. The malaria-specific parasite cannot reproduce in deficient people as the oxidized glutathiones are removed by their immune system (Khazaei et al. 2019). The gene controlling the deficiency is more frequently and severely expressed in men, since it is located on the X chromosome (Khazaei et al. 2019). Moreover, VC have an effect on the animals since they reduce the feeding efficiency in monogastric animals and are harmful to broiler chickens, laying hens, and pigs (Crépon et al. 2010, Khazaei et al. 2019).

2.2.2 Biosynthetic pathway

The VC biosynthetic pathway is not completely elucidated. Duc et al. (1989) discovered a low-VC accession, whose offspring have around 5% of the wild type VC content. The low-VC mutation was inherited in a single recessive Mendelian manner and was named "*vc*$^-$". Later Khazaei et al. (2017) developed a breeder-friendly kompetitive allele specific PCR (KASP) marker and reported candidate genes for *vc*$^-$. To elucidate the VC biosynthetic pathway, Vottonen (2018) elaborated on a list of candidate genes showing higher expression in both low-VC and high-VC genotypes of plants through gene expression analysis. Björnsdotter et al. (2021) used gene expression analysis and metabolite profiling, as well as QTL mapping and a comparative genomics approach with *Mt*, and found that the gene associated with the *vc*$^-$ locus is located on chromosome 1 between 1.22 and 1.43 centimorgans (cM), and encodes an isoform of 3,4-dihydroxy-2-butanone-4-phosphate-synthase/GTP cyclohydrolase II, an enzyme used in riboflavin biosynthesis. This

enzyme catalyzes the conversion of purine nucleoside triphosphate GTP into 2,5-diamino-6-ribosylamino-4(3H)-pyrimidinone 5'-phosphate (DARPP), converted into 5-amino-6-ribosylamino-2,3(1H,3H)-pyrimidinedione 5'-phosphate (ARPDP) through deamination. DARPP and ARPDP respectively lead to the production of vicine and convicine through hydrolysis, deamination and glucosylation (Björnsdotter et al. 2021). The causal gene for GTP cyclohydrolase II was named "*VC1*", and its recessive allele *vc*$^{-}$. The sequence of *vc*$^{-}$ contains a 2-nucleotide AT insertion, which causes a shift of the reading frame in the *VC1* region, preventing the production of GTP cyclohydrolase II, and resulting in low-VC content (Björnsdotter et al. 2021). On the other hand, through another gene expression analysis and metabolite profiling, Björnsdotter et al. (2021) demonstrated that VC are produced in the seed coat and then transported to the embryo, validating the fact that VC content is maternally determined.

However, the quantitative variation within both low and high-VC plants suggests that VC content is not monogenic. Puspitasari et al. (2022) investigated this quantitative variation through GWAS in 189 German winter faba bean lines. A total of 47 SNP markers with minor effect (between 0.37 and 5.61% of phenotypic variance explained) were associated with vicine and one with VC content.

2.2.3 Food-processing

Some efforts have been conducted to find an appropriate treatment to reduce VC content from faba bean seeds. This includes boiling, roasting (Cardador-Martínez et al. 2012) and microwave extraction (Ganzler and Salgó 1987), resulting in a significant decrease in VC content. Jamalian and Ghorbani (2005) tried several methods based on the soaking of seeds in different solutions and concluded that both continuous flow soaking in tap water and in acid solutions remove the whole VC content from seeds. The use of tap water is preferable since it is relatively not costly and has a lower effect on seed protein content. Vioque et al. (2012) experimented with alkaline extraction and acid precipitation, reducing the VC content by more than 99%. It is also possible to remove the VC content of faba bean flour through fermentation (Rizzello et al. 2016, Pulkkinen et al. 2019).

The food-processing sector is still looking for a cheap and energy-efficient method to reduce VC content (Khazaei et al. 2019). An efficient way would be to breed faba bean cultivars with low-

VC content, but this is challenging, especially because of time-consuming, inaccurate and expensive phenotyping methods (Khazaei et al. 2019). Hence, QTL mapping is used to determine the location of genes responsible for traits of interest and improve plant breeding efficiency by developing markers that can predict the plant phenotype through genotyping.

2.3 Hilum color

The faba bean hilum color (HC) is either black or colorless. Erith (1930) demonstrated that HC is maternally inherited and controlled by a single gene, the black color being dominant to colorless. Khazaei et al. (2015) located HC on chromosome 1 and 5.3 cM away from *vc*$^{-}$ gene. Since the initial donor of *vc*$^{-}$ found by Duc et al. (1989) has pale hilum, hilum color was used as a phenotypic marker to select for *vc*$^{-}$ in early breeding stages. However, this morphological marker is no longer valid, since the linkage between low-VC and colorless HC is broken (Khazaei et al. 2019).

The genetic basis underlying hilum color is still investigated. Warsame et al. (2019) identified a gene encoding dihydroflavonol 4-reductase (DFR) using GWAS analysis, whose results were supported by homozygosity mapping in three F2 populations and linkage mapping in a F6 RIL population. This enzyme plays a key role in anthocyanin biosynthesis in *Mt*.

3 RESEARCH OBJECTIVES

The aim of this study is to fine map genes responsible for VC content. The primary objective is to search for minor-effect QTLs responsible for the VC quantitative variation and identify candidate genes for each detected QTL. Using the same approach, the secondary objective is to confirm the studies of Björnsdotter et al. (2021) by checking for location and candidate genes for *VC1*.

For this purpose, a saturated linkage mapping was performed using Axiom array genotyping to map VC. The identification of new QTLs would allow us to understand better the molecular basis of VC content in faba bean and design markers to speed up the faba bean breeding process through MAS.

4 MATERIALS & METHODS

4.1 Plant material

The mapping population used was composed of 165 recombinant inbred lines (RILs), derived from a cross between Mélodie/2 and ILB 938/2, and advanced through SSD until the F8 generation (Khazaei et al. 2014, Gela et al. 2021). Mélodie/2 is a low-VC French cultivar with colorless hilum, developed by the French national research institute for agriculture, food and the environment (INRAE) from the initial *vc*$^{-}$ donor found by Duc et al. (1989). ILB 938/2 is a landrace originating from Ecuador and Colombia, with normal VC content and black hilum, derived from mass selection of ILB 438 (Khamassi et al. 2013).

4.2 Phenotyping

The VC phenotyping data were collected at F6 and F8 and adapted from Khazaei et al. (2017) and Purves et al. (2018). The F6 phenotyping data was measured by spectrophotometry (Sixdenier et al. 1996) and HPLC (high-performance liquid chromatography) (Khamasi et al. 2018), and F8 data by a highly selective mass spectrometry-based method (Purves et al. 2017). The 165 RILs were phenotyped at F6, and 106 RILs were phenotyped at F8. VC content was defined as the sum of vicine and convicine contents. Hilum color data was collected at F6 by Khazaei et al. (2017) based on visual observation.

To perform spectrophotometry, seeds were weighed and crushed. Then, it was incubated for 1h 15 min at 90°C, and diluted 10-fold. Absorbance was measured through spectroscopy, and converted into concentration (Khazaei et al. 2015). The values were expressed in % of seed dry weight and converted into mg/g. Mass spectroscopy was performed using an Agilent 1100 HPLC, associated with a high-field asymmetric waveform ion mobility spectrometry (FAIMS) system. Values were expressed in mg/g.

4.3 Genotyping

The genotypic data was generated by Gela et al. (2021) at F8. DNA from both parents and 165 RILs were extracted using the ionic detergent cetyltrimethylammonium bromide (CTAB) method, as described by Björnsdotter et al. (2021). A Quant-iT PicoGreen dsDNA Assay Kit (ThermoFisher Scientific, UK) was used to determine the DNA concentration of the samples, and their quality was assessed by agarose gel electrophoresis. Finally, the samples were genotyped using Axiom "Vfaba_v2" 60K array, developed from metatranscriptome data (O'Sullivan et al. 2019).

4.4 Linkage map construction

The linkage map was developed by Gela et al. (2021). The 35,363 SNP markers were filtered based on (1) polymorphism between parents, (2) segregation distortion using a chi-square (χ^2) test, and (3) missing data. The linkage map was built using both ASMap software (Taylor and Butler 2017) and MapDisto v.1.7.7.0.1 (Lorieux 2012). It included 4,089 markers, distributed in six linkage groups (LG), corresponding to the six chromosomes of faba bean. It spanned a distance of 1,229.5 cM.

4.5 QTL analysis

QTL mapping was performed through composite interval mapping (CIM) method in R/qtl v.1.50 software using vicine, convicine and total VC data from both F6 and F8 datasets. Hidden Markov model technology was used to calculate the conditional genotype probabilities through the calc.genoprob function and to simulate genotypes from the marker data through multiple imputations using the sim.geno function. The LOD significance threshold was set at 3. The effect plots were generated using the plot.pxg function. The phenotypic variance explained (PVE), also noted R^2, were estimated using the Equation 1. The effect due to QTL interactions was calculated using the addint function.

$$R^2 = 1 - 10^{-2LOD/n} \quad (1)$$

In which

n = population size

LOD = LOD score associated with the SNP marker

4.6 Candidate gene approach

Candidate gene identification is challenging in faba bean since no faba bean reference genome is publicly available yet. Moreover, other model legumes do not produce VC and thus should not have the specific genes faba bean has to produce those two compounds. Nevertheless, the candidate gene approach may give information about the biosynthetic pathways of VC, and help understanding the molecular basis of their production. As the markers used were developed mainly from *Mt*, this species was used for comparative genomics, and candidate genes were identified through BLAST sequence similarity searches of SNP markers located in the QTL intervals on *Mt* (Mt4.0v1) reference genome in Phytozome v13 (Goodstein et al. 2011). An e value threshold of $1e^{-6}$ was applied (The National Library of Medicine 2020). The QTL intervals for each detected QTL were calculated through the Bayesian credible interval method with 0.95 probability coverage based on each CIM result. If one QTL was detected several times, the QTL interval chosen for this QTL was the shortest one.

5 RESULTS

5.1 Phenotypic analysis of the parental lines and RIL population

The vicine, convicine and total VC content of Mélodie/2 were much lower than ILB 938/2 (Table 1). There was a wide range of values within the RIL mapping population at F6 and F8. They went beyond those found in the parents (Table 1).

Table 1 Vicine, convicine and total VC content of parental lines, and descriptive statistics of the RIL population for F6 and F8 generations.

Population	Number of individuals	Pyrimidine glycoside concentration (in mg/g)					
		F6			F8		
		Vicine	Convicine	Total VC	Vicine	Convicine	Total VC
Mélodie/2	15	0.00	0.40	0.40	0.29	0.02	0.31
ILB 938/2	15	3.30	3.80	7.10	5.30	3.66	8.95
RIL population	165 (F6); 106 (F8)	1.57	2.85	4.42	3.36	1.73	5.08
SD		1.56	2.45	3.88	2.39	1.52	3.76
Min		0.02	0.12	0.21	0.21	0.03	0.26
Max		7.24	8.36	14.34	7.55	5.89	13.08

The frequency distribution of the individuals from the RIL mapping population for vicine, convicine and total VC content showed segregation following a bimodal distribution at both generations (Figure 1).

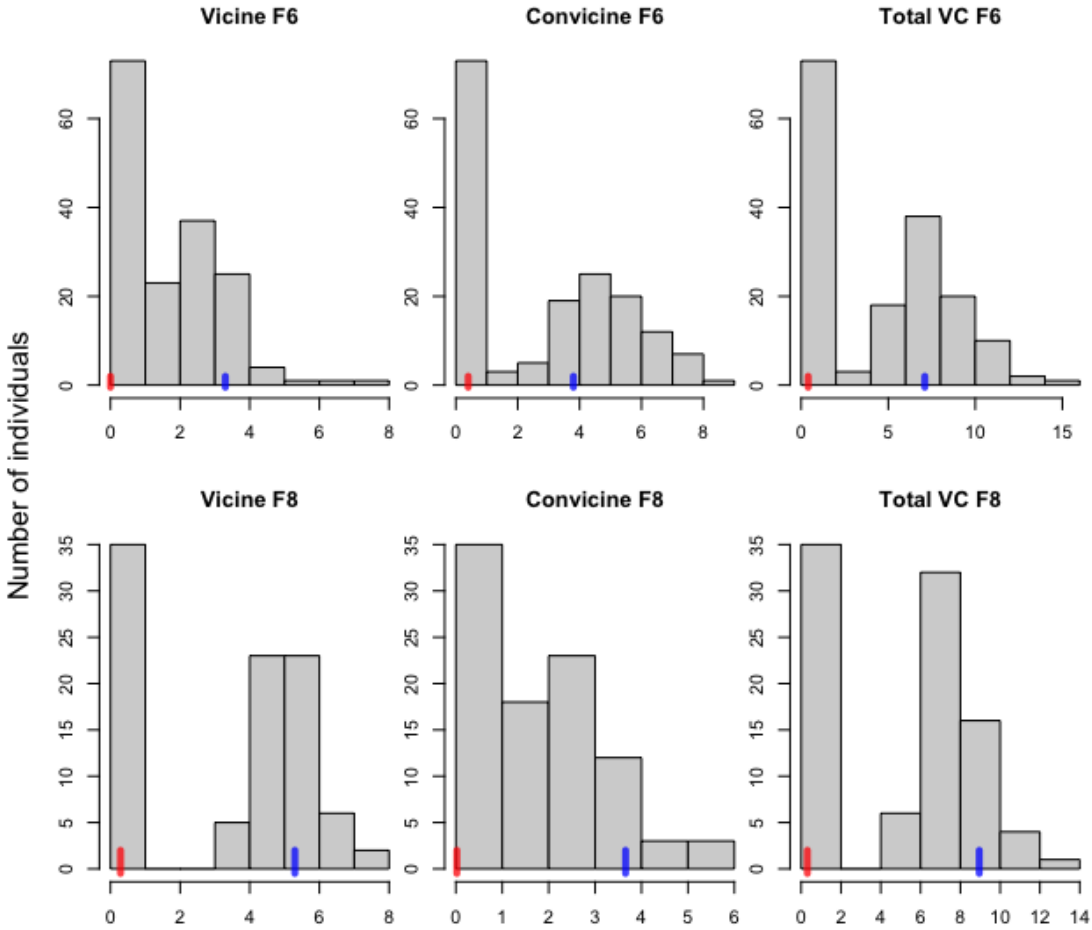

Figure 1 Frequency distribution of vicine, convicine and total VC content (in mg/g) for the RIL population at both F6 and F8 generations (in red, the value of Mélodie/2; in blue, the value of ILB 938/2).

Positive correlations were found between vicine and convicine contents in both generations (r = 0.864 in F6, r = 0.835 in F8) and between the total VC contents of both generations (r = 0.776).

5.2 QTL analysis

The CIM displayed six significant association signals at the beginning of chromosome 1 using vicine, convicine and total VC content at F6 and F8 generations (Figure 2), two association signals at the end of chromosome 4 for only convicine content at both F6 and F8 generations (Figure 2), and two association signals at the middle of chromosome 5 for only vicine and total VC content at the F8 generation (Figure 2b).

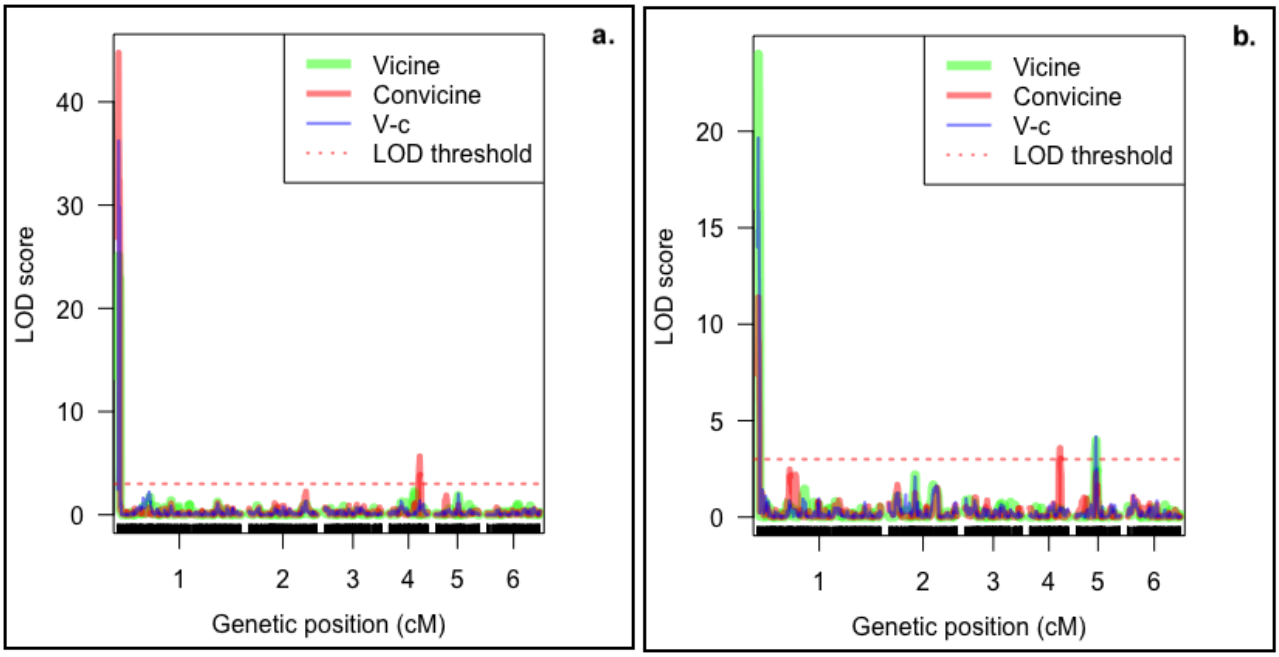

Figure 2 LOD plots obtained from CIM for vicine, convicine and total VC content at both **a.** F6 and **b.** F8 generations.

From those signals, three QTLs associated with VC content were detected, namely *VC1*, "*qC4*" and "*qVC5*". *VC1* was strongly linked to a marker located at 3.39 cM on chromosome 1 with high PVE (Table 2, Figure 3a). *qC4* was linked to 2 different markers both located at 98 cM on chromosome 4 (Table 2). *qVC5* was linked to 2 different markers as well, both located at 63 cM on chromosome 5 (Table 2). Markers linked to *qC4* and *qVC5* have moderated PVE (Figure 3b and 3c).

Table 2 Highest-LOD score markers linked to *VC1*, *qC4* and *qVC5* QTLs detected through CIM using vicine, convicine and total VC data at both F6 and F8 generations, with their genetic position, associated LOD score and PVE (R^2).

QTL	Gen	Data	DNA marker	Chr	Genetic position (in cM)	LOD score	R^2 (in %)
VC1	F6	Vicine	Affx-308495425	1	3.39	25.09	50.36
		Convicine	Affx-308495425	1	3.39	44.77	71.34
		Total VC	Affx-308495425	1	3.39	36.71	64.11
	F8	Vicine	Affx-308495425	1	3.39	23.20	63.51
		Convicine	Affx-308495425	1	3.39	17.50	53.25
		Total VC	Affx-308495425	1	3.39	20.43	58.83
qC4	F6	Convicine	Affx-309127928	4	98.32	5.67	14.63
	F8		Affx-1003939353	4	98.01	4.25	16.87
qVC5	F8	Vicine	Affx-1003954954	5	63.88	3.97	15.85
		Total VC	Affx-310743200	5	63.27	4.32	16.56

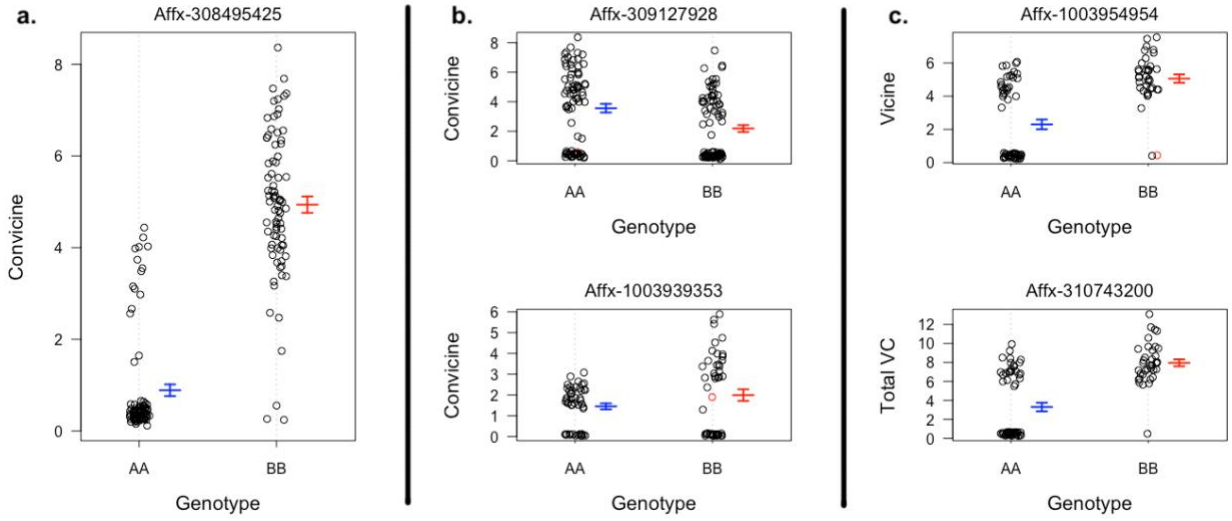

Figure 3 Plots showing the effect of the highest-LOD markers for each QTL detected by CIM (from left to right, **a.** the effect of the marker linked to *VC1* on convicine content in F6-dataset, **b.** the effect of both markers linked to *qC4*, and **c.** the effect of both markers linked to *qVC5*).

Moreover, the PVE of the interaction between *VC1* and *qC4* ranged from 0.29 to 2.57%, from 0 to 1.83% between *VC1* and *qVC5*, and from 0.23 to 0.31% between *qC4* and *qVC5*.

5.3 Candidate gene approach

The QTL interval for *VC1* included two markers at 3.39 cM. The QTL interval for *qC4* included 50 markers between 95.88 and 98.64 cM. The QTL interval for *qVC5* included six markers between 61.04 and 64.49 cM.

There was only one single candidate gene identified for *VC1* (Appendix 1). For *qC4*, 85 candidate genes were identified, of which 45 were located on chromosome 4 in *Mt*, 13 on chromosome 7, and the rest on chromosomes 2, 3 and 5 (Appendix 1). For *qVC5*, ten candidate genes were identified, of which five were located on chromosome 7 in *Mt*, three on chromosome 1, and the rest on chromosomes 2 and 3 (Appendix 1).

5.4 Hilum color

The CIM using binary hilum color data displayed a huge significant association signal at the beginning of chromosome 1 (Figure 4). The gene named "*qHC*" was located at 9.22 cM, linked to the Affx-308646555 marker (Table 3).

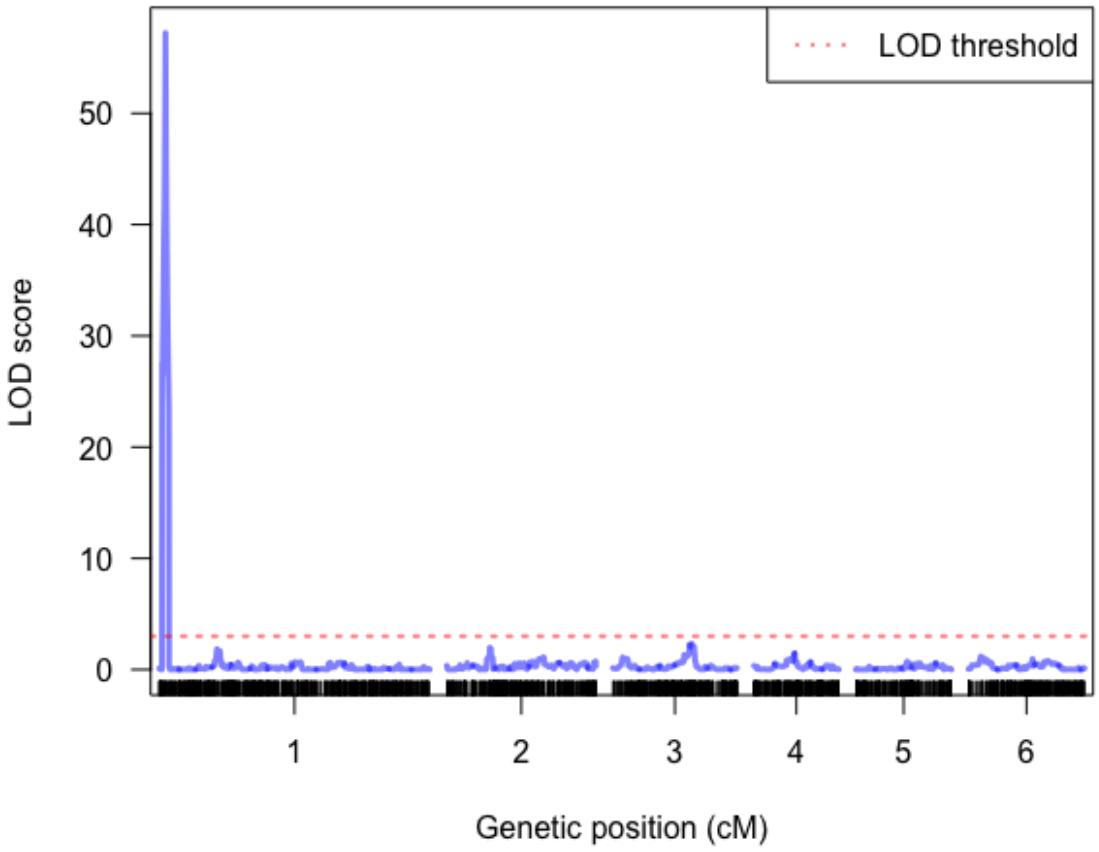

Figure 4 LOD plot obtained from CIM for hilum color at F6 generation.

Table 3 The highest-LOD score marker linked to the QTL detected through CIM using hilum color data at F6 generation, with its genetic position, associated LOD score and PVE (R^2).

QTL	Gen	Data	DNA marker	Chr	Genetic position (in cM)	LOD score	R^2 (in %)
qHC	F6	Hilum color	Affx-308646555	1	9.22	57.22	79.75

6 DISCUSSION

This study identified three QTLs responsible for VC content in faba bean, one associated with the major variation in VC, one associated with the quantitative variation, and a third one affecting only convicine. The major gene was expected from earlier studies, and there is only one report of a cause for the quantitative variation in VC (Puspitasari et al. 2022). The single gene controlling hilum color was also identified.

VC1 is the detected QTL with the highest PVE. It explained the major variation in VC (Table 2, Figure 3). It was located at the beginning of chromosome 1. It was noticed that *VC1* is not included in the QTL interval identified by Björnsdotter et al. (2021). In fact, it is 1.96 cM far from it. However, Björnsdotter et al. (2021) selected many markers within the 3.4-cM interval defined by Khazaei et al. (2015). The density of their genetic map was much higher at this specific location compared to the one used in this study, explaining the genetic distance observed between *VC1* in this study and in the one of Björnsdotter et al. (2021). Khazaei et al. (2017) and Björnsdotter et al. (2021) defined that Medtr2g009270 is the most probable gene for *VC1*. The QTL interval calculated in the present study included two markers, both at the same location (3.39 cM). The only candidate gene identified was Medtr2g009040. Since the genetic map used is less accurate than the one of Björnsdotter et al. (2021), it is less probable that this candidate gene corresponds to *VC1*. Moreover, it was noticed in the functional annotations of this *Mt* gene that Medtr2g009040 encodes a glycoside hydrolase. This enzyme hydrolyzes glycosidic bonds (Sieber et al. 2010). It could regulate the hydrolysis of the glucosidic bond, separating the glucosyl group from VC molecules and would result in the degradation of VC. However, low-VC content is conferred by the recessive allele of *VC1*, not the dominant one. In the case of Medtr2g009040, it would mean that ILB 938/2's allele would destroy VC molecules, which does not make sense. Thus, the candidate gene identified in this study does probably not correspond to *VC1*.

qC4 was located on chromosome 4. It explained some of the quantitative variation of convicine content (Table 2, Figure 3). Puspitasari et al. (2022) identified one single SNP marker (Affx-310027095) associated with VC content on chromosome 4. They used the same marker system to obtain their genotypic data as the one used for this study. The Affx-310027095 marker did not map on the linkage map of this study, but it is located between the Affx-310120776 marker and the Affx-

308653794, both of which mapped. Finally, it was observed that none of the SNP markers within the QTL interval identified for *qC4* is located between those two markers. It is thus concluded that *qC4* may not correspond to the QTL linked to the Affx-310027095 marker. The QTL interval identified for *qC4* was large, including many candidate genes (Appendix 1). Two markers were linked to *qC4*. From the sequence of the Affx-309127928 marker, two candidate genes were identified, one encodes an uncharacterized protein (Medtr5g009535) and another one encodes putatively a transmembrane protein (Medtr4g115740) (Appendix 1). From the sequence of the Affx-1003939353 marker, four candidate genes were identified, all of them encoding protein kinases (Medtr4g113100, Medtr5g075630, Medtr3g090665 and Medtr3g062500) (Appendix 1). *qC4* might correspond to Medtr4g115740 gene. It would regulate the transportation of convicine through the cell membranes. However, it is more probable that it corresponds to either Medtr4g113100, Medtr5g075630, Medtr3g090665 or Medtr3g062500, which could regulate the phosphorylation of the ARPDP molecule. Mélodie/2's allele would prevent convicine production by adding a phosphoryl group to the ARPDP. Finally, it was observed that many candidate genes are related to protein transportation. Thus, *qC4* might regulate the transportation of proteins related to convicine biosynthesis. This hypothesis is supported by the higher expression of protein-transport encoding genes in high-VC plants (Vottonen, 2018).

qVC5 was located on chromosome 5. It explained some of the quantitative variation of vicine and total VC content (Table 2, Figure 3). Puspitasari et al. (2022) identified two SNP markers (Affx-310628027 and Vf_Mt7g101170) associated with vicine content on chromosome 5. However, none of the markers within the QTL interval identified for *qVC5* corresponds to those markers. The QTL interval identified for *qVC5* is relatively short, including only five SNP markers, of which two were linked to *qVC5*. From the sequence of the Affx-1003954954 marker, one candidate gene was identified, encoding a myosin protein (Medtr7g099290) (Appendix 1). From the sequence of the Affx-310743200 marker, one candidate gene was identified, encoding a phosphatidylinositol N-acetylglucosaminyltransferase (Medtr7g099410) (Appendix 1). *qVC5* might correspond to this gene (Medtr7g099410). It would regulate the transfer of a phosphatidylinositol N-acetylglucosaminyl group into the compounds implicated in VC production (DARPP or ARPDP) or VC themselves, preventing VC production. However, it is more probable that *qVC5* corresponds to Medtr7g099290, since half of the candidate genes identified are also genes encoding myosin protein. Those proteins are responsible for cytoplasmic streaming (Madison and Nebenführ 2013). Thus, *qVC5* might

regulate the transportation of compounds implicated in VC production (DARPP or ARPDP) or VC themselves into the cells. This hypothesis is supported by the higher expression of myosin-encoding genes in high-VC plants (Vottonen, 2018).

The bimodal distribution of vicine, convicine and total VC data observed within the RIL mapping population (Figure 1) is explained by *VC1*. The individuals with the *vc*$^{-}$ allele have low-VC content, and the ones with *VC1* allele have normal VC content. The quantitative variation between both low-VC and normal VC content sets is explained by both other QTLs detected, *qC4* and *qVC5*. The strong positive correlation between vicine and convicine data within both datasets is explained by the fact that vicine and convicine are produced together during the riboflavin biosynthetic pathway. On the other hand, the positive correlation observed between VC content of both datasets is explained by the fact that the mapping population is a RIL population advanced by selfing. The individuals of the F6 and F8 populations have substantially the same genotypes, as the homozygosity within both of them is high. One limitation of this study is that there is only one replicate in both generations, but actually F6 and F8 could almost be considered as replicates. It would be relevant to run again the QTL analysis with the same RIL mapping population using phenotypic data of further generations.

Concerning hilum color, *qHC* on chromosome 1 was mapped 5.83 cM far from *VC1*. This result corresponds to the genetic distance observed by Khazaei et al. (2015) between the hilum color locus and *VC1* and confirms it. The QTL interval for hilum color included only the highest-LOD score marker it was linked to. The candidate gene identified for this marker is Medtr2g013080, located on *Mt* chromosome 2 from 3454986 to 3460192 bp, and described as a RING finger protein. This kind of gene produces proteins that play an important role in ubiquitination pathway. Ubiquination regulates many biological processes and results in the degradation of proteins or the alteration of protein-protein interactions (Guo and Tadi 2022). Medtr2g013080 might produce black color by affecting protein-protein interactions.

Since *VC1* was included in a QTL interval ranging from two markers at the same location, both SNP markers might be used by breeders for MAS. As they are strongly linked with *VC1*, they give the QTL location very accurately. In the same way, the highest-LOD markers linked to *qC4* and

qVC5, might be used for MAS. However, they are less predictive. Further studies are needed to fine map those two QTLs.

Other less restrictive QTL mapping methods could be used, but might give other QTL locations and shorter QTL intervals. Using CIM, we avoided shortcomings from marker regression, standard interval mapping (IM), Haley-Knott (HK) or extended HK regression methods.

It should be noted that the Axiom "Vfaba_v2" 60K array used to obtain the genotypic data, allowed the building of a saturated genetic map that would be very efficient for fine mapping of *qC4* and *qVC5*, and for discovering new potential VC-related QTLs through other mapping population. New genotyping and phenotyping methods will without a doubt, significantly improve the precision of mapping in the near future.

7 CONCLUSIONS

The present study identified three QTLs responsible for VC content in faba bean. The first was *VC1*. Then, *qC4* affected only convicine content and was associated with a large number of 85 candidate genes. Finally, *qVC5* affected vicine and total VC content and was associated with a set of ten candidate genes. For all identified candidate genes, hypotheses on their influence on VC content production or transportation, based on the VC biosynthesis pathway described by Björnsdotter et al. (2021), were proposed for further testing.

A first reference genome for faba bean will soon be made publicly available. Using the knowledge already collected from previous studies and further sequencing efforts, scientists will be able to fine map and sequence genes for important traits in faba bean. This will improve our understanding of the molecular processes behind the VC production and transportation into the bean embryos and make us able to develop more predictive markers for genes of interest to perform MAS.

Acknowledgments

I would like to thank my supervisors: Dr. Frederick Stoddard, head of the Legume Science Research Group, for his guidance throughout the whole project and his help to improve the overall thesis quality thanks to his expert knowledge on faba bean, and Dr. Hamid Khazaei, Vegetable Breeder in the World Vegetable Center in Tainan City, for his help and advice to conduct the QTL mapping and the candidate gene approach through a constant monitoring.

I would like to thank the entire team I worked with, including Dr. Alan Schulman, head of the Institute of Biotechnology, Dr. Marko Jääskeläinen, Postdoctoral Researcher at the Institute of Biotechnology, and Dr. Kiflemariam Belachew, Postdoctoral Researcher in the Crop Science Research Group

I would like to thank the emPLANT Erasmus Mundus Master in Plant Breeding and more particularly Dr. Alicia Ayerdi-Gotor and Dr. Teemu Teeri, coordinators of the program at respectively UniLaSalle Beauvais and the University of Helsinki. They gave me the opportunity to learn and acquire skills through the best study environment.

I would like to thank my colleague Sarkal Jyakhwa, M.Sc. emPLANT student in plant breeding, for his comments and feedback on the mini-thesis and popular science assignments related to this thesis.

Finally, I would like to dedicate this achievement to my family and friends for their love and constant support.

8 LITERATURE CITED

Adhikari, K. N., Khazaei, H., Ghaouti, L., Maalouf, F., Vandenberg, A., Link, W., & O'Sullivan, D. M. (2021). Conventional and Molecular Breeding Tools for Accelerating Genetic Gain in Faba Bean (*Vicia Faba* L.). Frontiers in Plant Science, 12. https://www.frontiersin.org/article/10.3389/fpls.2021.744259

Annicchiarico, P., & Iannucci, A. (2008). Breeding Strategy for Faba Bean in Southern Europe based on Cultivar Responses across Climatically Contrasting Environments. Crop Science, 48, 983–991. https://doi.org/10.2135/cropsci2007.09.0501

Björnsdotter, E., Nadzieja, M., Chang, W., Escobar-Herrera, L., Mancinotti, D., Angra, D., Xia, X., Tacke, R., Khazaei, H., Crocoll, C., Vandenberg, A., Link, W., Stoddard, F. L., O'Sullivan, D. M., Stougaard, J., Schulman, A. H., Andersen, S. U., & Geu-Flores, F. (2021). VC1 catalyses a key step in the biosynthesis of vicine in faba bean. Nature Plants, 7(7), 923–931. https://doi.org/10.1038/s41477-021-00950-w

CABI (2022). *Vicia faba* (faba bean). CABI. https://www.cabi.org/isc/datasheet/56364#toPictures Visited 14.04.2022

Cardador-Martínez, A., Maya-Ocaña, K., Ortiz-Moreno, A., Herrera-Cabrera, B. E., Dávila-Ortiz, G., Múzquiz, M., Martín-Pedrosa, M., Burbano, C., Cuadrado, C., & Jiménez-Martínez, C. (2012). Effect of Roasting and Boiling on the Content of Vicine, Convicine and L-3,4-dihydroxyphenylalanine in *Vicia faba* L. Journal of Food Quality, 35(6), 419–428. https://doi.org/10.1111/jfq.12006

Crépon, K., Marget, P., Peyronnet, C., Carrouée, B., Arese, P., & Duc, G. (2010). Nutritional value of faba bean (*Vicia faba* L.) seeds for feed and food. Field Crops Research, 115(3), 329–339. https://doi.org/10.1016/j.fcr.2009.09.016

Cubero, J. I. (1974). On the evolution of *Vicia faba* L. Theoretical and Applied Genetics, 45(2), 47–51. https://doi.org/10.1007/BF00283475

Duc, G., Aleksić, J. M., Marget, P., Mikic, A., Paull, J., Redden, R. J., Sass, O., Stoddard, F. L., Vandenberg, A., Vishnyakova, M., & Torres, A. M. (2015). Faba Bean. In A. M. De Ron (Ed.), Grain Legumes (pp. 141–178). Springer. https://doi.org/10.1007/978-1-4939-2797-5_5

Duc, G., Sixdenier, G., Lila, M., & Furstoss, V. (1989). Search of genetic variability for vicine and convicine content in *Vicia faba* L.: A first report of a gene which codes for nearly zero-vicine and zero-convicine contents. 1. International Workshop on 'Antinutritional Factors (ANF) in

Legume Seeds', Wageningen (Netherlands), 23-25 Nov 1988. https://scholar.google.com/scholar_lookup?title=Search+of+genetic+variability+for+vicine+and+convicine+content+in+Vicia+faba+L.%3A+a+first+report+of+a+gene+which+codes+for+nearly+zero-vicine+and+zero-convicine+contents&author=Duc%2C+G.+%28Station+de+Genetique+et+Amelioration+des+Plantes+INRA%2C+Dijon+%28France%29%29&publication_year=1989

Erith, A. G. (1930). The inheritance of colour, size and form of seeds, and of flower colour in *Vicia Faba* L. Genetica, 12(4), 477–510. https://doi.org/10.1007/BF01486760

FAO (2015). What Are Pulses? https://www.fao.org/pulses-2016/news/news-detail/en/c/337107/ Visited 23.04.2022

FAOSTAT (2022). https://www.fao.org/faostat/en/#home Visited 10.03.2022

Ganzler, K., & Salgó, A. (1987). Microwave-extraction -a new method superseding traditional Soxhlet extraction. Zeitschrift Für Lebensmittel-Untersuchung Und Forschung, 184(4), 274–276. https://doi.org/10.1007/BF01027662

Gela, T. S., Bruce, M., Chang, W., Stoddard, F. L., Schulman, A. H., Vandenberg, A., & Khazaei, H. (2021). Genomic regions associated with chocolate spot (*Botrytis fabae* Sard.) resistance in faba bean (*Vicia faba* L.). bioRxiv. https://doi.org/10.1101/2021.11.22.469473

Gnanasambandam, A., Paull, J., Torres, A., Kaur, S., Leonforte, T., Li, H., Zong, X., Yang, T., & Materne, M. (2012). Impact of Molecular Technologies on Faba Bean (*Vicia faba* L.) Breeding Strategies. Agronomy, 2(3), 132–166. https://doi.org/10.3390/agronomy2030132

Goodstein, D., Shu, S., Howson, R., Neupane, R., Hayes, R., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., & Rokhsar, D. (2011). Phytozome: A comparative platform for green plant genomics. Nucleic Acids Research, 40, D1178-86. https://doi.org/10.1093/nar/gkr944

Griffiths, D. W., & Ramsay, G. (1992). The concentration of vicine and convicine in *Vicia faba* and some related species and their distribution within mature seeds. Journal of the Science of Food and Agriculture, 59(4), 463–468. https://doi.org/10.1002/jsfa.2740590406

Guo, H. J., & Tadi, P. (2022). Biochemistry, Ubiquitination. In StatPearls. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK556052/

Hughes, J., Khazaei, H., & Vandenberg, A. (2020). Genetics of Height and Branching in Faba Bean (*Vicia faba*). Agronomy, 10(8), 1191. https://doi.org/10.3390/agronomy10081191

Jamalian, J., & Ghorbani, M. (2005). Extraction of favism-inducing agents from whole seeds of faba bean (*Vicia faba* L var major). Journal of the Science of Food and Agriculture, 85(6), 1055–1060. https://doi.org/10.1002/jsfa.2075

Jensen, E. S., Peoples, M. B., & Hauggaard-Nielsen, H. (2010). Faba bean in cropping systems. Field Crops Research, 115, 203–216. https://doi.org/10.1016/j.fcr.2009.10.008

Karkanis, A., Ntatsi, G., Lepse, L., Fernández, J. A., Vågen, I. M., Rewald, B., Alsiņa, I., Kronberga, A., Balliu, A., Olle, M., Bodner, G., Dubova, L., Rosa, E., & Savvas, D. (2018). Faba Bean Cultivation – Revealing Novel Managing Practices for More Sustainable and Competitive European Cropping Systems. Frontiers in Plant Science, 9, 1115. https://www.frontiersin.org/article/10.3389/fpls.2018.01115

Kaur, S., Kimber, R. B. E., Cogan, N. O. I., Materne, M., Forster, J. W., & Paull, J. G. (2014). SNP discovery and high-density genetic mapping in faba bean (*Vicia faba* L.) permits identification of QTLs for ascochyta blight resistance. Plant Science, 217–218, 47–55. https://doi.org/10.1016/j.plantsci.2013.11.014

Khamassi, K., Jeddi, F. B., Hobbs, D., Irigoyen, J., Stoddard, F., O'Sullivan, D. M., & Jones, H. (2013). A baseline study of vicine–convicine levels in faba bean (*Vicia faba* L.) germplasm. Plant Genetic Resources, 11, 250–257. https://doi.org/10.1017/S1479262113000105

Khazaei, H. (2014). Leaf traits associated with drought adaptation in faba bean (*Vicia faba* L.). University of Helsinki, Finland.

Khazaei, H., O'Sullivan, D. M., Jones, H., Pitts, N., Sillanpää, M. J., Pärssinen, P., Manninen, O., & Stoddard, F. L. (2015). Flanking SNP markers for vicine–convicine concentration in faba bean (*Vicia faba* L.). Molecular Breeding, 35, 38. https://doi.org/10.1007/s11032-015-0214-8

Khazaei, H., O'Sullivan, D. M., Stoddard, F. L., Adhikari, K. N., Paull, J. G., Schulman, A. H., Andersen, S. U., & Vandenberg, A. (2021). Recent advances in faba bean genetic and genomic tools for crop improvement. Legume Science, 3, 75. https://doi.org/10.1002/leg3.75

Khazaei, H., O'Sullivan, D., Sillanpää, M., & Stoddard, F. (2014). Use of synteny to identify candidate genes underlying QTL controlling stomatal traits in faba bean (*Vicia faba* L.). Theoretical and Applied Genetics, 127, 2371-2385. https://doi.org/10.1007/s00122-014-2383-y

Khazaei, H., Purves, R. W., Hughes, J., Link, W., O'Sullivan, D. M., Schulman, A. H., Björnsdotter, E., Geu-Flores, F., Nadzieja, M., Andersen, S. U., Stougaard, J., Vandenberg, A., & Stoddard, F. L. (2019). Eliminating vicine and convicine, the main antinutritional factors restricting faba bean usage. Trends in Food Science & Technology, 91, 549–556. https://doi.org/10.1016/j.tifs.2019.07.051

Khazaei, H., Purves, R. W., Song, M., Stonehouse, R., Bett, K. E., Stoddard, F. L., & Vandenberg, A. (2017). Development and validation of a robust, breeder-friendly molecular marker for the *vc*$^{-}$ locus in faba bean. Molecular Breeding, 37(11), 140. https://doi.org/10.1007/s11032-017-0742-5

Khazaei, H., Street, K., Bari, A., Mackay, M., & Stoddard, F. L. (2013). The FIGS (Focused Identification of Germplasm Strategy) Approach Identifies Traits Related to Drought Adaptation in *Vicia faba* Genetic Resources. PLoS ONE, 8(5), e63107. https://doi.org/10.1371/journal.pone.0063107

Kirk, W. D. J. (2004). Faba bean: *Vicia faba*. Bee World, 85(3), 60–62. https://doi.org/10.1080/0005772X.2004.11099625

Klippenstein, S. R., Khazaei, H., Vandenberg, A., & Schoenau, J. (2022). Nitrogen and phosphorus uptake and nitrogen fixation estimation of faba bean in western Canada. Agronomy Journal, 114(1), 811–824. https://doi.org/10.1002/agj2.20945

Longobardi, F., Sacco, D., Casiello, G., Ventrella, A., & Sacco, A. (2015). Chemical Profile of the Carpino Broad Bean by Conventional and Innovative Physicochemical Analyses. https://doi.org/10.1111/jfq.12143

Lorieux, M. (2012). MapDisto: Fast and efficient computation of genetic linkage maps. Molecular Breeding, 30(2), 1231–1235. https://doi.org/10.1007/s11032-012-9706-y

Madison, S. L., & Nebenführ, A. (2013). Understanding myosin functions in plants: Are we there yet? Current Opinion in Plant Biology, 16(6), 710–717. https://doi.org/10.1016/j.pbi.2013.10.004

Manna, S. (2015). An overview of pentatricopeptide repeat proteins and their applications. Biochimie, 113, 93–99. https://doi.org/10.1016/j.biochi.2015.04.004

McVetty, P. B. E., & Nugent-Rigby, J. (1984). Natural cross pollination of faba beans (*Vicia faba* L.) grown in Manitoba. Canadian Journal of Plant Science, 64(1), 43–46. https://doi.org/10.4141/cjps84-007

Murphy-Bokern, D., Stoddard, F., & Watson, C. (2017). Legumes in cropping systems. In Legumes in Cropping Systems (p. 256). https://www.researchgate.net/publication/321668720_Legumes_in_cropping_systems

Neugschwandtner, R., Ziegler, K., Kriegner, S., Wagentristl, H., & Kaul, H.-P. (2015). Nitrogen yield and nitrogen fixation of winter faba beans. Acta Agriculturae Scandinavica, Section B — Soil & Plant Science, 65(7), 658–666. https://doi.org/10.1080/09064710.2015.1042028

O'Sullivan, D. M., Angra, D., Harvie, T., Tagkouli, V., & Warsame, A. (2019). A genetic toolbox for Vicia faba improvement. In International confer- ence on legume genetics and genomics, may 13–17, 2019. Dijon: France.

Pulkkinen, M., Coda, R., Lampi, A.-M., Varis, J., Katina, K., & Piironen, V. (2019). Possibilities of reducing amounts of vicine and convicine in faba bean suspensions and sourdoughs. European Food Research and Technology, 245(7), 1507–1518. https://doi.org/10.1007/s00217-019-03282-4

Purves, R. W., Khazaei, H., & Vandenberg, A. (2018). Toward a high-throughput method for determining vicine and convicine levels in faba bean seeds using flow injection analysis combined with tandem mass spectrometry. Food Chemistry, 256, 219–227. https://doi.org/10.1016/j.foodchem.2018.02.104

Purves, R. W., Zhang, H., Khazaei, H., & Vandenberg, A. (2017). Rapid analysis of medically relevant compounds in faba bean seeds using FAIMS and mass spectrometry. International Journal for Ion Mobility Spectrometry, 20(3–4), 125–135. https://doi.org/10.1007/s12127-017-0226-7

Puspitasari, W., Allemann, B., Angra, D., Appleyard, H., Ecke, W., Möllers, C., Nolte, T., Purves, R. W., Renner, C., Robertson-Shersby-Harvie, T., Tacke, R., Windhorst, A., Yaman, S., & Link, W. (2022). NIRS for vicine and convicine content of faba bean seed allowed GWAS to prepare for marker-assisted adjustment of seed quality of German winter faba beans. Journal of Cultivated Plants, 74(01–02), Article 01–02. https://doi.org/10.5073/JfK.2022.01-02.01

Redden, R., Paull, J., Zong, X., Sass, O., Yang, T., & Ling, L. (2014). Faba bean. In Broadening the Genetic Base of Grain Legumes (pp. 75–93). https://doi.org/10.1007/978-81-322-2023-7_4

Ricciardi, L., Filippetti, A., De Pace, C., & Marzano, C. F. (1985). Inheritance of seed coat colour in broad bean (*Vicia faba* L.). Euphytica, 34(1), 43–51. https://doi.org/10.1007/BF00022862

Rispail, N., Kaló, P., Kiss, G., Ellis, N., Gallardo, K., Thompson, R., Prats, E., Larrainzar, E., Ladrera, R., González, E., Arrese-Igor, C., Ferguson, B., Gresshoff, P., & Rubiales, D. (2010). Model legumes contribute to faba bean breeding. Field Crops Research, 115, 253–269. https://doi.org/10.1016/j.fcr.2009.03.014

Rizzello, C. G., Losito, I., Facchini, L., Katina, K., Palmisano, F., Gobbetti, M., & Coda, R. (2016). Degradation of vicine, convicine and their aglycones during fermentation of faba bean flour. Scientific Reports, 6(1), 32452. https://doi.org/10.1038/srep32452

Robinson, G. H. J., Balk, J., & Domoney, C. (2019). Improving pulse crops as a source of protein, starch and micronutrients. Nutrition Bulletin, 44(3), 202–215. https://doi.org/10.1111/nbu.12399

Sato, S., Isobe, S., & Tabata, S. (2010). Structural analyses of the genomes in legumes. Current Opinion in Plant Biology, 13(2), 146–152. https://doi.org/10.1016/j.pbi.2009.12.007

Sato, S., Nakamura, Y., Asamizu, E., Isobe, S., & Tabata, S. (2007). Genome Sequencing and Genome Resources in Model Legumes. Plant Physiology, 144(2), 588–593. https://doi.org/10.1104/pp.107.097493

Satovic, Z., Avila, C. M., Cruz-Izquierdo, S., Díaz-Ruíz, R., García-Ruíz, G. M., Palomino, C., Gutiérrez, N., Vitale, S., Ocaña-Moral, S., Gutiérrez, M. V., Cubero, J. I., & Torres, A. M. (2013). A reference consensus genetic map for molecular markers and economically important traits in faba bean (*Vicia faba* L.). BMC Genomics, 14(1), 932. https://doi.org/10.1186/1471-2164-14-932

Sieber, S. A., Böttcher, T., Staub, I., & Orth, R. (2010). Small Molecules as Versatile Tools for Activity-Based Protein Profiling Experiments. In H.-W. (Ben) Liu & L. Mander (Eds.), Comprehensive Natural Products II (pp. 629–674). Elsevier. https://doi.org/10.1016/B978-008045382-8.00159-3

Sixdenier, G., Cassecuelle, F., Guillaumin, L., & Duc, G. (1996). Rapid spectrophotometric method for reduction of vicine and convicine in faba bean seed. Faba Bean Information Service. https://scholar.google.com/scholar_lookup?title=Rapid+spectrophotometric+method+for+reduction+of+vicine+and+convicine+in+faba+bean+seed&author=Sixdenier%2C+G.&publication_year=1996

Sprent, J. I. (2009). Legume Nodulation: A Global Perspective (Wiley-Blackwell). https://www.wiley.com/en-us/Legume+Nodulation%3A+A+Global+Perspective-p-9781405181754

Taylor, J., & Butler, D. (2017). R Package ASMap: Efficient Genetic Linkage Map Construction and Diagnosis. Journal of Statistical Software, 79, 1–29. https://doi.org/10.18637/jss.v079.i06

The National Library of Medicine. (2020). BLAST Results: Expect Values, Part 1. https://www.youtube.com/watch?v=ZN3RrXAe0uM Visited 29.04.2022

Vioque, J., Alaiz, M., & Girón-Calle, J. (2012). Nutritional and functional properties of *Vicia faba* protein isolates and related fractions. Food Chemistry, 132(1), 67–72. https://doi.org/10.1016/j.foodchem.2011.10.033

Vottonen, L. (2018). Uncovering gene expression differences associated with high and low vicine and convicine content in faba bean (*Vicia faba*). MSc thesis, University of Helsinki, Finland.

Warsame, A. O., Angra, D., & O'Sullivan, D. (2019). Identification of a candidate gene controlling hilum colour in faba bean. Phenotypic diversity of Botrytis spp. populations and identification of resistance sources from landraces of faba bean in Morocco., Conference: 9th International Conference on Legume Genetics and Genomics, Dijon. https://www.researchgate.net/publication/357091695_Phenotypic_diversity_of_Botrytis_spp_populations_and_identification_of_resistance_sources_from_landraces_of_faba_bean_in_Morocco

Zong, X., Liu, X., Guan, J., Wang, S., Liu, Q., Paull, J., & Redden, R. (2009). Molecular variation among Chinese and global winter faba bean germplasm. TAG. Theoretical and Applied Genetics, 118, 971–978. https://doi.org/10.1007/s00122-008-0954-5

9 APPENDICES

9.1 Appendix 1: Candidate genes associated with the detected QTLs for vicine and convicine

Table 4 Candidate genes associated with *VC1*, *qC4* and *qVC5*, identified by BLAST sequence similarity searches with the sequences of the marker included in the QTL intervals, on Medicago truncatula (Mt4.0v1) genome in Phytozome v13 (Goodstein et al. 2011)

QTL	Gene ID	Position	e value	Description
VC1	Medtr2g009040	chr2:1738189..1743193 reverse	1.53e-18	PTHR37701:SF1 - METHYL-CPG-BINDING DOMAIN-CONTAINING PROTEIN 8
qC4	Medtr4g108870	chr4:45122731..45125978 forward	1.18e-13	PTHR30231 - DNA POLYMERASE III SUBUNIT EPSILON
qC4	Medtr4g108870	chr4:45122731..45125978 forward	5.70e-24	PTHR30231 - DNA POLYMERASE III SUBUNIT EPSILON
qC4	Medtr4g109130	chr4:45276042..45278237 reverse	5.70e-24	PTHR11673 - TRANSLATION INITIATION FACTOR 5A-RELATED
qC4	Medtr4g130917	chr4:54579115..54581914 reverse	3.37e-14	PTHR11673 - TRANSLATION INITIATION FACTOR 5A-RELATED
qC4	Medtr3g437680	chr3:12821899..12831667 forward	1.26e-19	PTHR11673 - TRANSLATION INITIATION FACTOR 5A-RELATED
qC4	Medtr5g042990	chr5:18905198..18907263 reverse	7.92e-16	PTHR11673 - TRANSLATION INITIATION FACTOR 5A-RELATED
qC4	Medtr4g109390	chr4:45426024..45434064 reverse	4.68e-25	PTHR11255:SF54 - DIACYLGLYCEROL KINASE 1
qC4	Medtr4g109550	chr4:45527726..45529540 forward	6.09e-11	PTHR10202:SF13 - PRESENILIN-2
qC4	Medtr4g109470	chr4:45477290..45481844 forward	1.53e-18	PTHR24298:SF104 - FLAVONOID 3'-MONOOXYGENASE
qC4	Medtr3g024520	chr3:7619577..7624416 reverse	1.43e-12	PTHR24298:SF104 - FLAVONOID 3'-MONOOXYGENASE
qC4	Medtr3g025260	chr3:7674966..7683298 forward	1.43e-12	PTHR24298:SF104 - FLAVONOID 3'-MONOOXYGENASE
qC4	Medtr4g109570	chr4:45540211..45542412 forward	2.59e-9	A0A072URP6
qC4	Medtr4g112540	chr4:46005092..46009932 reverse	2.42e-22	PF11833 - Protein of unknown function (DUF3353) (DUF3353)
qC4	Medtr4g111770	chr4:46184393..46185411 forward	1.10e-7	PTHR33203:SF4 - F27J15.22-RELATED
qC4	Medtr8g077000	chr8:32682880..32692655 reverse	1.26e-19	PTHR13097 - TRANSCRIPTION INITIATION FACTOR IIE, ALPHA SUBUNIT
qC4	Medtr4g113070	chr4:46444557..46454935 forward	4.68e-25	PTHR19241:SF280 - ABC TRANSPORTER G FAMILY MEMBER 32
qC4	Medtr4g011630	chr4:2881084..2890071 forward	1.75e-11	PTHR19241:SF219 - ABC TRANSPORTER G FAMILY MEMBER 38
qC4	Medtr4g011640	chr4:2892453..2903046 reverse	1.75e-11	PTHR19241:SF219 - ABC TRANSPORTER G FAMILY MEMBER 38

QTL	Gene ID	Position	e value	Description
qC4	Medtr4g011620	chr4:2871766..2879824 forward	2.13e-10	PTHR19241:SF219 - ABC TRANSPORTER G FAMILY MEMBER 38
qC4	Medtr3g463680	chr3:25520344..25528512 reverse	3.37e-14	PTHR19241:SF169 - ABC TRANSPORTER G FAMILY MEMBER 31
qC4	Medtr3g107870	chr3:49768123..49775604 forward	2.13e-10	PTHR19241//PTHR19241:SF281 - ATP-BINDING CASSETTE TRANSPORTER
qC4	Medtr7g104100	chr7:42130712..42138118 reverse	1.75e-11	PTHR19241//PTHR19241:SF261 - ATP-BINDING CASSETTE TRANSPORTER
qC4	Medtr7g098740	chr7:39535662..39543685 reverse	2.13e-10	PTHR19241:SF219 - ABC TRANSPORTER G FAMILY MEMBER 38
qC4	Medtr7g098750	chr7:39544754..39553280 reverse	2.13e-10	PTHR19241:SF219 - ABC TRANSPORTER G FAMILY MEMBER 38
qC4	Medtr7g104110	chr7:42139606..42147066 reverse	2.13e-10	PTHR19241//PTHR19241:SF261 - ATP-BINDING CASSETTE TRANSPORTER
qC4	Medtr7g104130	chr7:42151341..42159530 reverse	2.13e-10	PTHR19241//PTHR19241:SF261 - ATP-BINDING CASSETTE TRANSPORTER
qC4	Medtr7g098370	chr7:39358638..39368630 forward	9.04e-9	PTHR19241:SF219 - ABC TRANSPORTER G FAMILY MEMBER 38
qC4	Medtr7g098800	chr7:39587082..39593988 reverse	9.04e-9	PTHR19241:SF219 - ABC TRANSPORTER G FAMILY MEMBER 38
qC4	Medtr7g098890	chr7:39619867..39626413 reverse	9.04e-9	PTHR19241:SF219 - ABC TRANSPORTER G FAMILY MEMBER 38
qC4	Medtr7g104150	chr7:42169306..42180259 reverse	3.16e-8	PTHR19241//PTHR19241:SF261 - ATP-BINDING CASSETTE TRANSPORTER
qC4	Medtr7g407080	chr7:1082475..1085876 reverse	1.10e-7	PTHR19241:SF219 - ABC TRANSPORTER G FAMILY MEMBER 38
qC4	Medtr7g098300	chr7:39332580..39341077 forward	1.10e-7	PTHR19241:SF219 - ABC TRANSPORTER G FAMILY MEMBER 38
qC4	Medtr7g098780	chr7:39572257..39579351 reverse	1.10e-7	PTHR19241:SF219 - ABC TRANSPORTER G FAMILY MEMBER 38
qC4	Medtr2g102640	chr2:44208768..44216038 reverse	2.13e-10	PTHR19241:SF219 - ABC TRANSPORTER G FAMILY MEMBER 38
qC4	Medtr2g102660	chr2:44220747..44227634 reverse	3.16e-8	PTHR19241:SF219 - ABC TRANSPORTER G FAMILY MEMBER 38
qC4	Medtr8g015780	chr8:5195935..5196622 forward	3.84e-7	3.6.3.27 - Phosphate-transporting ATPase / ABC phosphate transporter
qC4	Medtr4g113100	chr4:46475885..46479891 forward	7.92e-16	PF00069//PF00560//PF08263 - Protein kinase domain (Pkinase) // Leucine Rich Repeat (LRR_1) // Leucine rich repeat N-terminal domain (LRRNT_2)
qC4	Medtr5g075630	chr5:32169814..32173885 reverse	2.13e-10	PF00069//PF00560//PF08263//PF13855 - Protein kinase domain (Pkinase) // Leucine Rich Repeat (LRR_1) // Leucine rich repeat N-terminal domain (LRRNT_2) // Leucine rich repeat (LRR_8)

QTL	Gene ID	Position	e value	Description
qC4	Medtr3g090665	chr3:41157660..41160115 forward	2.59e-9	2.7.11.1 - Non-specific serine/threonine protein kinase / Threonine-specific protein kinase
qC4	Medtr3g062500	chr3:28221569..28226064 reverse	1.10e-7	PF00069//PF00560//PF08263 - Protein kinase domain (Pkinase) // Leucine Rich Repeat (LRR_1) // Leucine rich repeat N-terminal domain (LRRNT_2)
qC4	Medtr4g121913	chr4:50332972..50333173 reverse	1.75e-11	Uncharacterized protein
qC4	Medtr4g113600	chr4:46709490..46710604 forward	1.43e-12	Plant/MSJ11-3 protein, putative
qC4	Medtr4g116790	chr4:48342065..48345946 forward	2.42e-22	PTHR23500:SF12 - SUGAR TRANSPORT PROTEIN 7
qC4	Medtr4g116770	chr4:48329423..48333293 forward	2.95e-21	PF02990//PF07690 - Endomembrane protein 70 (EMP70) // Major Facilitator Superfamily (MFS_1)
qC4	Medtr0204s0040	scaffold0204:18086..20626 forward	2.95e-21	PTHR23500:SF12 - SUGAR TRANSPORT PROTEIN 7
qC4	Medtr0204s0020	scaffold0204:5185..8670 forward	4.10e-13	PTHR23500:SF12 - SUGAR TRANSPORT PROTEIN 7
qC4	Medtr5g094760	chr5:41413783..41416616 reverse	1.75e-11	PTHR23500:SF12 - SUGAR TRANSPORT PROTEIN 7
qC4	Medtr4g124080	chr4:51241102..51242712 forward	1.26e-19	PTHR11780//PTHR11780:SF8 - NADH-UBIQUINONE OXIDOREDUCTASE FLAVOPROTEIN 1 NDUFV1
qC4	Medtr4g116840	chr4:48353076..48368829 forward	1.26e-19	PTHR12136//PTHR12136:SF49 - STEROIDOGENIC ACUTE REGULATORY PROTEIN STAR
qC4	Medtr2g449780	chr2:21926213..21935419 forward	4.10e-13	PTHR12136//PTHR12136:SF49 - STEROIDOGENIC ACUTE REGULATORY PROTEIN STAR
qC4	Medtr4g117050	chr4:48490974..48494729 reverse	2.13e-10	PTHR11772:SF18 - ALUMINUM INDUCED PROTEIN WITH YGL AND LRDR MOTIFS
qC4	Medtr2g022520	chr2:7790361..7796721 forward	7.42e-10	PTHR11772:SF18 - ALUMINUM INDUCED PROTEIN WITH YGL AND LRDR MOTIFS
qC4	Medtr4g115920	chr4:47897676..47901019 reverse	7.92e-16	PTHR21015:SF25 - MONOGALACTOSYLDIACYLGLYCEROL SYNTHASE 2, CHLOROPLASTIC-RELATED
qC4	Medtr4g114690	chr4:47200289..47201853 forward	1.26e-19	PTHR26312:SF75 - F18B13.21 PROTEIN-RELATED
qC4	Medtr4g123070	chr4:50797780..50801047 reverse	1.10e-26	PTHR22912//PTHR22912:SF148 - DISULFIDE OXIDOREDUCTASE
qC4	Medtr4g117600	chr4:48786109..48787634 forward	1.26e-19	PTHR33294:SF3 - AWPM-19-LIKE PROTEIN

QTL	Gene ID	Position	e value	Description
qC4	Medtr4g120050	chr4:49765243..49767507 reverse	9.04e-9	Uncharacterized protein
qC4	Medtr4g115580	chr4:47734788..47737734 reverse	4.38e-19	PTHR20855//PTHR20855:SF32 - ADIPOR/PROGESTIN RECEPTOR-RELATED
qC4	Medtr4g125370	chr4:51981894..51982910 forward	1.86e-17	PTHR39113:SF1 - MEMBRANE LIPOPROTEIN-RELATED
qC4	Medtr4g118320	chr4:49019542..49021164 forward	2.42e-22	PTHR11570:SF1 - S-ADENOSYLMETHIONINE DECARBOXYLASE PROENZYME
qC4	Medtr2g020990	chr2:7080607..7082204 reverse	6.09e-11	PTHR11570:SF1 - S-ADENOSYLMETHIONINE DECARBOXYLASE PROENZYME
qC4	Medtr4g122070	chr4:50404200..50411541 reverse	9.65e-15	PTHR13484 - FIP1-LIKE 1 PROTEIN
qC4	Medtr4g102690	chr4:42564580..42567744 reverse	2.59e-9	
qC4	Medtr0204s0020	scaffold0204:5185..8670 forward	2.42e-22	PTHR23500:SF12 - SUGAR TRANSPORT PROTEIN 7
qC4	Medtr0204s0040	scaffold0204:18086..20626 forward	5.34e-18	PTHR23500:SF12 - SUGAR TRANSPORT PROTEIN 7
qC4	Medtr4g116790	chr4:48342065..48345946 forward	1.26e-19	PTHR23500:SF12 - SUGAR TRANSPORT PROTEIN 7
qC4	Medtr4g116770	chr4:48329423..48333293 forward	5.34e-18	PF02990//PF07690 - Endomembrane protein 70 (EMP70) // Major Facilitator Superfamily (MFS_1)
qC4	Medtr7g091690	chr7:36292790..36297111 reverse	7.92e-16	PTHR23500:SF12 - SUGAR TRANSPORT PROTEIN 7
qC4	Medtr5g094760	chr5:41413783..41416616 reverse	2.77e-15	PTHR23500:SF12 - SUGAR TRANSPORT PROTEIN 7
qC4	Medtr4g116420	chr4:48228379..48231422 forward	2.95e-21	PTHR13832//PTHR13832:SF299 - PROTEIN PHOSPHATASE 2C
qC4	Medtr4g123080	chr4:50809343..50811831 reverse	1.26e-19	PTHR13832:SF316 - PROTEIN PHOSPHATASE 2C 43-RELATED
qC4	Medtr4g115720	chr4:47820720..47821118 forward	7.62e-7	Transmembrane protein, putative
qC4	Medtr4g116770	chr4:48329423..48333293 forward	1.26e-19	PF02990//PF07690 - Endomembrane protein 70 (EMP70) // Major Facilitator Superfamily (MFS_1)
qC4	Medtr2g009110	chr2:1780751..1787840 forward	1.75e-11	PTHR12097 - SPLICING FACTOR 3B, SUBUNIT 1-RELATED
qC4	Medtr5g009535	chr5:2315861..2316484 forward	1.03e-20	Uncharacterized protein
qC4	Medtr4g115740	chr4:47825728..47830181 forward	4.38e-19	Transmembrane protein, putative

QTL	Gene ID	Position	e value	Description
qC4	Medtr4g120130	chr4:49795591..49803884 reverse	2.42e-22	1.1.1.35//4.2.1.17//4.2.1.55//5.1.2.3//5.3.3.8 - 3-hydroxyacyl-CoA dehydrogenase / Beta-keto-reductase // Enoyl-CoA hydratase / Unsaturated acyl-CoA hydratase // 3-hydroxybutyryl-CoA dehydratase / Crotonase // Dodecenoyl-CoA isomerase / Dodecenoyl-CoA Delta-isomerase
qC4	Medtr2g436300	chr2:14109766..14119095 forward	4.10e-13	1.1.1.35//4.2.1.17//4.2.1.55//5.1.2.3//5.3.3.8 - 3-hydroxyacyl-CoA dehydrogenase / Beta-keto-reductase // Enoyl-CoA hydratase / Unsaturated acyl-CoA hydratase // 3-hydroxybutyryl-CoA dehydratase / Crotonase // Dodecenoyl-CoA isomerase / Dodecenoyl-CoA Delta-isomerase
qC4	Medtr5g068780	chr5:29134510..29137675 reverse	6.50e-17	PTHR11655//PTHR11655:SF18 - 60S/50S RIBOSOMAL PROTEIN L6/L9
qC4	Medtr3g093110	chr3:42566705..42568353 forward	9.04e-9	PTHR11655//PTHR11655:SF18 - 60S/50S RIBOSOMAL PROTEIN L6/L9
qC4	Medtr4g113150	chr4:46508372..46512527 reverse	4.10e-13	KOG0107 - Alternative splicing factor SRp20/9G8 (RRM superfamily)
qC4	Medtr4g122980	chr4:50738151..50743564 reverse	2.95e-21	PTHR10353:SF33 - BETA-GLUCOSIDASE-LIKE SFR2, CHLOROPLASTIC
qC4	Medtr4g116030	chr4:47969043..47978176 forward	1.53e-18	PTHR11662:SF243 - ANION TRANSPORTER 6, CHLOROPLASTIC-RELATED
qC4	Medtr4g125470	chr4:52029087..52033535 reverse	4.10e-13	PTHR22893:SF67 - 12-OXOPHYTODIENOATE REDUCTASE 3
qC4	Medtr4g125490	chr4:52051633..52053330 forward	2.13e-10	PF08879//PF08880 - WRC (WRC) // QLQ (QLQ)
qVC5	Medtr7g099830	chr7:40084556..40088849 reverse	5.70e-24	PTHR23180:SF293 - ADP-RIBOSYLATION FACTOR GTPASE-ACTIVATING PROTEIN AGD7
qVC5	Medtr1g069000	chr1:29556809..29560472 forward	5.00e-12	PTHR23180:SF293 - ADP-RIBOSYLATION FACTOR GTPASE-ACTIVATING PROTEIN AGD7
qVC5	Medtr7g099410	chr7:39870142..39875778 forward	1.53e-18	PTHR21726//PTHR21726:SF36 - PHOSPHATIDYLINOSITOL N-ACETYLGLUCOSAMINYLTRANSFERASE SUBUNIT P DOWN SYNDROME CRITICAL REGION PROTEIN 5 -RELATED
qVC5	Medtr7g099290	chr7:39804145..39817615 forward	1.03e-20	PTHR13140:SF256 - MYOSIN-14-RELATED
qVC5	Medtr7g115430	chr7:47692246..47704474 forward	4.68e-6	PTHR13140:SF390 - MYOSIN-11-RELATED

QTL	Gene ID	Position	e value	Description
qVC5	Medtr1g069290	chr1:29882617..29897859 reverse	4.38e-19	PF00063//PF00612//PF01843 - Myosin head (motor domain) (Myosin_head) // IQ calmodulin-binding motif (IQ) // DIL domain (DIL)
qVC5	Medtr1g015620	chr1:3948599..3971342 forward	6.09e-11	PTHR13140:SF304 - MYOSIN IA HEAVY CHAIN-RELATED
qVC5	Medtr3g105200	chr3:48491992..48513135 forward	1.75e-11	PTHR13140:SF304 - MYOSIN IA HEAVY CHAIN-RELATED
qVC5	Medtr2g062660	chr2:26463302..26475233 forward	6.09e-11	PTHR13140:SF270 - MYOSIN-12
qVC5	Medtr7g080110	chr7:30441901..30446792 forward	7.92e-16	PTHR14208 - BASIC LEUCINE ZIPPER AND W2 DOMAIN-CONTAINING PROTEIN

9.2 Appendix 2: QTL analysis for seed coat color

Seed coat can be either violet, spotted, brown, green, red, black or beige. Mélodie/2 has beige and ILB 938/2 has green seed coat. Data was collected at F7 by visual observation. Its color is determined by two multiallelic loci, which are unlinked, but sometimes epistatic (Ricciardi et al. 1985). The CIM displayed a huge significant signal associated with seed coat color at the middle of chromosome 4 (Figure 5). The gene named "*qSC*" was located at 63.86 cM, linked to the Affx-308344772 marker (Table 5).

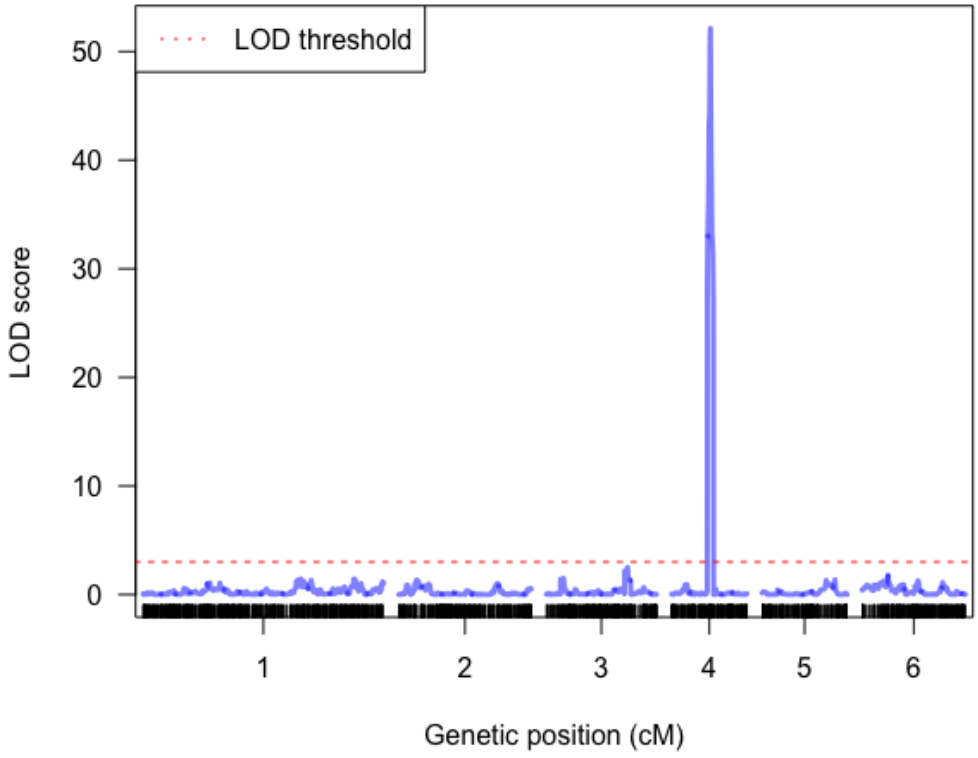

Figure 5 LOD plot obtained from CIM for green seed coat color

Table 5 The highest-LOD score marker linked to the QTL detected through CIM using seed coat color data at F7 generation, with its genetic position, associated LOD score and PVE (R^2).

QTL	Gen	Data	DNA marker	Chr	Genetic position (in cM)	LOD score	R^2 (in %)
qSC	F7	Seed coat color	Affx-308344772	4	63.86	50.18	75.56

The QTL interval included 15 markers all at 63.86 cM, and 30 candidate genes were reported (Table 6). From the sequence of the Affx-308344772 marker, one candidate gene was identified,

encoding pentatricopeptide repeat (PPR) proteins (Medtr8g074780) (Table 6). Those proteins regulate the gene expression at the RNA level (Manna 2015). Since seed coat color may be determined by epistatic effect, it is probable that in this mapping population, it is determined by the RNA silencing of the other causal gene. Overall, it was noticed that 24 out of the 30 candidate genes are located on *Mt* chromosome 8, the rest being on chromosomes 2, 5 and 7 (Table 6).

Table 6 Candidate genes associated with *qSC*, identified by BLAST sequence similarity searches with the sequences of the marker included in the QTL intervals, on Medicago truncatula (Mt4.0v1) genome in Phytozome v13 (Goodstein et al. 2011)

QTL	Gene ID	Position	e value	Description
qSC	Medtr8g074780	chr8:31615265..31617312 reverse	5.00e-12	PF01535//PF13041 - PPR repeat (PPR) // PPR repeat family (PPR_2)
qSC	Medtr8g076260	chr8:32297240..32301590 forward	4.10e-13	PTHR14360:SF1 - PROTEIN FMP32, MITOCHONDRIAL
qSC	Medtr5g021410	chr5:8267884..8272412 reverse	2.59e-9	PTHR14360:SF1 - PROTEIN FMP32, MITOCHONDRIAL
qSC	Medtr8g071100	chr8:30158574..30171725 forward	6.50e-17	PTHR10629:SF34 - DNA (CYTOSINE-5)-METHYLTRANSFERASE CMT2
qSC	Medtr7g047650	chr7:16695884..16696974 forward	6.50e-17	PTHR10629:SF34 - DNA (CYTOSINE-5)-METHYLTRANSFERASE CMT2
qSC	Medtr7g451840	chr7:17670720..17671210 forward	6.50e-17	PTHR10629:SF34 - DNA (CYTOSINE-5)-METHYLTRANSFERASE CMT2
qSC	Medtr8g073305	chr8:31082944..31085398 reverse	1.10e-26	Uncharacterized protein
qSC	Medtr8g071220	chr8:30217920..30223478 forward	4.68e-25	PTHR22870//PTHR22870:SF190 - REGULATOR OF CHROMOSOME CONDENSATION
qSC	Medtr8g019620	chr8:6946946..6951587 forward	2.13e-10	PTHR22870:SF166 - PROTEIN RCC2
qSC	Medtr2g041170	chr2:18039562..18048357 forward	2.77e-15	PTHR22870:SF166 - PROTEIN RCC2
qSC	Medtr8g006745	chr8:912883..914036 reverse	9.03e-28	PF01207 - Dihydrouridine synthase (Dus) (Dus)
qSC	Medtr8g067420	chr8:28143498..28151636 reverse	5.70e-24	tRNA-dihydrouridine(20) synthase (NAD(P)(+)) / tRNA-dihydrouridine synthase 2

QTL	Gene ID	Position	e value	Description
qSC	Medtr8g072360	chr8:30587927..30595373 forward	5.70e-24	PTHR11782:SF3 - APYRASE 7-RELATED
qSC	Medtr8g070510	chr8:29971762..29976008 forward	7.92e-16	PF00036//PF13499 - EF hand (EF-hand_1) // EF-hand domain pair (EF-hand_7)
qSC	Medtr5g030620	chr5:12993767..13000069 reverse	9.65e-15	Phosphoenolpyruvate carboxykinase (GTP) / Phosphopyruvate carboxylase
qSC	Medtr5g430730	chr5:13068096..13075503 reverse	9.65e-15	Phosphoenolpyruvate carboxykinase (GTP) / Phosphopyruvate carboxylase
qSC	Medtr8g067275	chr8:28135275..28142103 reverse	3.37e-14	Phosphoenolpyruvate carboxykinase (GTP) / Phosphopyruvate carboxylase
qSC	Medtr8g071270	chr8:30250058..30252447 reverse	2.95e-21	Uncharacterized protein
qSC	Medtr8g066097	chr8:27454118..27455284 forward	5.31e-24	GAF domain-like
qSC	Medtr8g066155	chr8:27481484..27484085 reverse	9.60e-21	Uncharacterized protein
qSC	Medtr8g066097	chr8:27454118..27455284 forward	1.17e-19	GAF domain-like
qSC	Medtr8g066097	chr8:27454118..27455284 forward	1.43e-18	Uncharacterized protein
qSC	Medtr8g066155	chr8:27481484..27484085 reverse	2.12e-16	Uncharacterized protein
qSC	Medtr8g066155	chr8:27481484..27484085 reverse	1.34e-12	Uncharacterized protein
qSC	Medtr8g066155	chr8:27481484..27484085 reverse	2.41e-9	Uncharacterized protein
qSC	Medtr8g066097	chr8:27454118..27455284 forward	1.03e-7	Uncharacterized protein
qSC	Medtr8g066155	chr8:27481484..27484085 reverse	1.03e-7	Uncharacterized protein
qSC	Medtr8g071180	chr8:30195076..30199339 reverse	1.53e-18	Cadmium-induced protein AS8, putative
qSC	Medtr8g071920	chr8:30347941..30355218 reverse	2.95e-21	PTHR21576:SF7 - MAJOR FACILITATOR PROTEIN-RELATED

QTL	Gene ID	Position	e value	Description
qSC	Medtr8g076620	chr8:32482155..32484876 forward	5.00e-12	PTHR31604:SF3 - PROTEIN SHI RELATED SEQUENCE 1-RELATED