Genetics and Biochemistry of the Low Tannin Characteristic in *Vicia faba* L. and Development of a Molecular Marker for the *zt2* Gene

A Thesis Submitted to the College of

Graduate and Postdoctoral Studies

In Partial Fulfillment of the Requirements

For the Degree of Master of Science

In the Department of Plant Sciences

Saskatoon

By

Stefano Zanotto

PERMISSION TO USE

In presenting this thesis in partial fulfilment of the requirements for a postgraduate degree from the University of Saskatchewan, I agree that the libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.
Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Plant Sciences
51 Campus Dr.
University of Saskatchewan
Saskatoon, Saskatchewan (S7N 5A8)

ABSTRACT

Faba bean is a cool season grain legume which is grown in many areas of the world. It is used either as a food or feed and it has the highest concentration of protein in the seed compared to all the other pulses. The presence of condensed tannins (proanthocyanidins) in the testa of the seeds limits the use of faba bean both in the food and feed industries. Low tannin faba bean varieties are available for cultivation. The low tannin phenotype is controlled by two genes *zt1* and *zt2.* Low tannin plants also have completely white flowers. The presence of either of the genes in the recessive form is sufficient for determining the low tannin and white flower phenotype.

The main objective of this project was to develop and validate a molecular marker for the *zt2* gene. A set of 875 SNP markers physically anchored to the genome of *Medicago truncatula* was used to genotype a RIL population at the F6 segregating for *zt2.* One of the SNP markers, located 10.5 cM from *zt2,* successfully discriminated the genotypes carrying the recessive (*zt2*) allele from those carrying the dominant (*ZT2*) allele. The SNP marker was then converted to KASP assay format.

The second objective of this study was to test the available molecular tools for tagging the *zt1* and *zt2* genes for their efficacy in identifying the genotypes carrying the recessive alleles at both *zt1* and *zt2* loci. The $F_{2:3}$ population segregating for both low tannin genes were used for this purpose. The KASP marker for the *zt2* gene was again successful in distinguishing the recessive (*zt2*) allele from the dominant (*ZT2*) allele. For the *zt1* gene we used two SCAR markers, $SCG11_{1171}$ and $SCC5_{551}$ that are publicly available. Only $SCC5_{551}$ was partially effective in tagging the *zt1*

gene. This marker in combination with the *zt2* KASP marker may be used in future research aiming at the identification of the double recessive genotype *zt1zt2*.

The third objective of the research was to analyze the phenolic profiles of the seed coats and flowers of three faba bean genotypes, one wild type and two low tannin lines recessive for either *zt1* or *zt2.* For these analyses we used LC-MS techniques, covering over 100 known phenolic compounds. The results suggested a possible point of action of the *zt* genes in the flavonoid biosynthetic pathway. The results also showed clear differences in the phenolic profiles between the two low tannin lines both in the seed coats and the flowers. Furthermore, one unknown compound present in only one of the two low tannin lines was characterized, matching its MS/MS spectrum with that of a compound previously identified in another species. This and other compounds could be used as biochemical markers to distinguish the *zt1* from the *zt2* genotypes. In conclusion, this study enhanced our knowledge on the mode of action of the two low tannin genes *zt1* and *zt2,* linking the results from molecular analysis with those at the biochemical level. The study successfully developed a molecular marker for the *zt2* gene which could be used in combination with biochemical markers by plant breeders targeting the low tannin phenotype in their breeding programs.

ACKNOWLEDGMENT

I would like to thank my co-supervisors Drs. Albert Vandenberg and Bunyamin Tar'an for their guidance and support during my study and research experience at the University of Saskatchewan.

I also thank all the other members of my advisory committee, Dr. Robert Tyler, Dr. Eric Lamb, and in particular, Dr. Hamid Khazaei who led me through all the practical steps and taught me the fundamental knowledge to successfully complete this research project.

A special thank goes to Fatma Elzahraa Elessawy and Dr. Randy W. Purves for their patience in explaining to me the principles of LC-MS and their technical help in conducting the analysis and interpreting the data.

I also highly appreciate the help of the pulse breeding crew and the staff of the phytotron and field lab.

Finally, I acknowledge the financial support of the NORFAB project, the InnovationFund Denmark and the Agriculture Development Fund project support from the Saskatchewan Ministry of Agriculture that made the realization of this M.Sc. research possible.

DEDICATION

To all the members of my family, for their constant support

and their trust in me and my choices

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

(NH4)2SO4	Ammonium Sulfate
ACN	Acetonitrile
ANS	Anthocyanidin synthase
BC	Before Christ
bHLH	Basic helix loop helix
BSA	Bulked Segregant Analysis
C	Convicine
ACN	Acetonitrile
CFIA	Canadian Food Inspection Agency
CDC	Crop Development Center
cM	centiMorgans
CMSF	Core Mass Spectrometry Facility
CoA	Coenzyme A
DDMS	Data Dependent Mass Spectrometry
DFR	Dihydroflavonol reductase
FA	Formic acid
FC	Flower Colour
HC	Hilum Colour
HCL	Hydrochloric acid
ICARDA	International Center for Agricultural Research in Dry Areas
INRA	Institut National de la Recherche Agronomique
IS	Internal Standard
KASP	Kompetitive allele specific PCR

KCl	Potassium Chloride
LAR	Leucoanthocyanidin reductase
LC-MS	Liquid Chromatography-Mass Spectrometry
LG	Linkage groups
MBW	MYB-bHLH-WD40 (complex of transcription factors)
MgCl2	Magnesium Chloride
NaOH	Sodium Hydroxide
NPK	Nitrogen Phosphorus Potassium
PAs	Proanthocyanidins
PCR	Polymerase Chain reaction
PFP	Pentafluorophenyl
RAPD	Random Amplification of Polymorphic DNA
RBC	Red Blood Cell
RIL	Recombinant Inbred Line
SCAR	Sequence Characterized Amplified Region
SCC	Seed Coat Colour
SNP	Single Nucleotide Polymorphism
SRM	Selected Reaction Monitoring
SSD	Single Seed Descent
SSP	Stipule Spot Pigmentation
TT2	Transparent testa 2
TTG1	Transparent Testa Glabra 1
V	Vicine

CHAPTER 1 - INTRODUCTION

Faba bean (*Vicia faba* L.) is a globally important pulse crop. It is used either as food for humans or feed for animals. The crop has excellent nutritional value including high protein concentration (270-320 g kg^{-1} DM)(Duc et al., 1999). The adaptation to different environments is the most important characteristic that has enabled faba bean cultivation in many different areas of the world. Faba bean has the highest level of nitrogen fixation capability among the pulse crops, giving it an additional values both from the economic and ecological points of view (Köpke and Nemecek, 2010).

In Canada, faba bean is mainly grown in the provinces of Alberta, Saskatchewan and Manitoba. It is well adapted to the cooler and wetter prairie regions because it needs adequate soil moisture for full yield potential. In fact, the crop is highly sensitive to drought stress (Khan et al., 2010). Recently faba bean is gaining more interest as a substitute for pea and lentil in the moist regions of Saskatchewan, due to better ability to withstand water logging and good resistance to Aphanomyces root rot, to which pea and lentils are highly susceptible. It also has a much better standability compared to pea (Phelps, 2017).

From the nutritional point of view, faba bean can be divided into different groups depending on the seed content of vicine (V), convicine (C) and condensed tannins, which were demonstrated to be anti-nutritional components for monogastric animals (Crépon et al., 2010).

Condensed tannins (*proanthocyanidins*) are phenolic compounds that are concentrated in the seed coat. Faba bean accessions characterized by high proanthocyanidin content (wild types) in the seed coat also display coloured flowers (usually white with a large black spot in the wings), while the low tannin varieties have completely white flowers. The seed coat colour is also

affected by the presence of phenolic compounds. The presence of a black spot on the stipules, visible pre-flowering during the plant development, is related to the flower colour (Khazaei et al., 2014). The presence of black spots at the flower wings and condensed tannins in the seed coat, is controlled by the two independent recessive genes *zt1* and *zt2* (Picard, 1976). When a low tannin accession is crossed with a wild type, the segregation ratio of 3:1 for coloured versus white flowers in the F2 proves that the flower colour follows the Mendelian inheritance of a monogenic trait (Erith, 1930). The presence of either of the two genes in the homozygous recessive form results in low tannin and white flower phenotype.

Faba bean is a partially allogamous crop (Duc, 1997). This fact poses challenges to the production of tannin free seeds. In fact, if *zt1* and *zt2* plants are crossed, the seed of the progeny will have high tannin content and the plants will have wild type flower phenotype. Therefore, it is important to maintain careful isolation between low tannin lines containing the two different genes during the breeder seed multiplication process. For the same reason low and high tannin varieties must be cultivated at a distance of at least 1000 m in order to minimize outcrossing, which could result in seed lots with unacceptable seed purity (Canadian Seed Growers Association).

Up to now, most studies used low tannin accessions which contain one or the other *zt* gene in the recessive form (mainly *zt1*), and the double recessive genotype has never been obtained (Crépon et al., 2010). Furthermore, no efficient molecular marker is available to aid selection for the *zt2* gene.

This research had three objectives. The first was the identification of a molecular marker linked to the *zt2* gene. The second was to genotype plant materials derived from a cross between

lines carrying *zt1* and *zt2*, respectively, to identify the genotype recessive at both the *zt1* and *zt2* loci. The third objective was to perform LC-MS (Liquid Chromatography-Mass Spectrometry) analysis on the seed coats and flowers of three faba bean genotypes, a wild type, one *zt1* line and one *zt2* line, to characterize their phenolic profiles in these organs. The later analyses included not only condensed tannins, but also a broader spectrum of secondary metabolites in the category of phenolic compounds that are present in faba bean seed coats and flowers.

The specific objectives of this study were as follows:

1 To identify SNP (Single Nucleotide Polymorphism) marker(s) closely linked to the *zt2* gene and to validate the marker in a diverse panel of genotypes differing for their allelic state at the *zt2* locus.

2 To identify the genotype which carries both low tannin genes (*zt1zt2*) in the recessive form.

3 To analyze and compare the phenolic compounds in the seed coat and flower of two low tannin (Disco/2 [*zt2*] and CDC Snowdrop [*zt1*]) and a wild type (ILB 938/2) faba bean accessions using Liquid Chromatography Mass Spectrometry (LC-MS).

CHAPTER 2 - LITERATURE REVIEW

2.1 Faba bean origin and production

Faba bean (*Vicia faba* L.) sometimes referred to as broad bean, horse-bean and feed bean is a cool season grain legume. The center of origin and diffusion of faba bean is generally accepted to be the Near East, and the subspecies *V. faba paucijuga* which is presently found from Afghanistan to India is probably the primitive form of the cultivated species (Duc, 1997). *Vicia faba* is known to have been cultivated from early Neolithic times. Traces of faba bean were found in archeological deposits of Bronze, Stone and Iron Age in the Near East. There are two major types of faba bean which were distinguished by trait analysis, the small seeded form and the large seeded form. The first group is typical of south western Asia and is the most ancient in its origin, while the large seeded group developed in the Mediterranean basin which is probably a secondary center of origin. Traces of faba bean cultivation in central and southern Europe are dated to the 2nd to the 3^{d} Millennia BC (Duc et al. 2010). The introduction of faba bean to China is uncertain, but probably its cultivation started around 2100 years ago with the development of the silk road which connected China to the Middle East. The introduction of faba bean in South America occurred in the 15th century with the Spanish and Portuguese colonization (Duc et al., 2010).

Faba bean is an annual plant which requires cool conditions for best development. The root system is a tap root with secondary roots and bears nodules containing the nitrogen-fixing bacteria *Rhizobium leguminosarum*, bv. viciae. Faba bean is a partially allogamous crop with the estimate of natural outcrossing ranging from 2 to 84% and a mean of 32%. This variation depends on the activity of pollinators, the species and the geographic locations (Duc, 1997).

In 2016 the worldwide production of faba bean dry grains was 4.6 Mt over the cultivated areas of 2.4 million ha (FAOSTAT, 2016). The main world producer of faba bean was China with a production of 1.6 Mt from 806,900 ha of cultivated area (FAOSAT, 2016). However, this figure is likely to be underestimated because it did not account for the faba bean cultivated and consumed locally by subsistence farmers and for the large-seeded broad beans picked as green pods and consumed fresh (O'Sullivan and Angra, 2016). In Canada faba bean is mainly grown in the provinces of Alberta, Saskatchewan and Manitoba. In 2016 the total area cultivated with faba bean in the three provinces was 114.500 acres (46.340 ha) (Phelps, 2017). Commercial production in western Canada started in 1972, and since then, the area under production has fluctuated (Government of Saskatchewan, 2018).

2.2 Use and quality characteristics of the faba bean grain

Faba bean constitutes a basic food for many people living in the Mediterranean countries of North Africa, the near East and southern Europe. The crop is also widely used as a food in China and other Asian countries. The seeds are consumed fresh, frozen, dry or canned (Duc et al. 2010).

Recently faba bean has gained interest in the food industry for use as flour (Petitot et al. 2010) and protein isolate (Karaca et al. 2011). Because of its high concentration of protein in the seed compared to other pulses, faba bean has the potential to become a plant-based substitute for meat, similar to the use of soybean protein (Multari et al. 2015).

Because of its nutritional value, with protein and starch content among different genotypes ranging from 27 to 34% and 37 to 50% of seed dry matter, respectively (Duc et al., 1999), faba bean is also used as feed for animal nutrition, mainly as dry grain, mashed grain or pellets, and also as forage (Mikić et al., 2011).

Faba bean nutritive potential is partially limited both for human and animal diets because of some anti-nutritional factors present in the seeds. Vicine (V), convicine (C) and condensed tannins are the main antinutritional factors in faba bean. V and C (found in the cotyledons) are two inactive precursors of divicine and isouramil, redox compounds potentially toxic for humans carrying the widespread genetic deficiency of the erythrocyte (red blood cell) enzyme glucose-6-phosphate dehydrogenase (Crépon et al., 2010).

Condensed tannins are secondary metabolites present in the testa of seeds. These compounds create problems for protein, energy and starch digestion in monogastric animals (Van Der Poel et al. 1992; Duc et al. 1995; Crépon et al. 2010). Furthermore, they can negatively affect the absorption of several essential micronutrients, iron in particular (Multari et al. 2015).

For the animal feed industry, condensed tannins (proanthocyanidins) are the main concern for the introduction of faba bean into pig and poultry diets (Yuste et al. 1991, Gunawardena et al. 2014). Condensed tannins reduced true digestibility, increased the endogenous excretion of proteins, and reduced the digestible energy in pig diets. Similar effects were found for poultry, for which feed intake was also shown to be reduced by tannin rich faba beans (Vilariño et al., 2009). On the other hand faba bean was reported to be a possible substitute for soybean as a source of protein and energy in ruminant diets (Crépon et al., 2010).

Faba bean varieties with reduced content of proanthocyanidins in the seed coat are available, and they are known as low tannin faba bean. In research where low tannin genotypes and genotypes with normal tannin content were compared for their nutritional value, Duc et al. (1999) showed that the first group contain less fiber, and this reduction resulted in a higher concentration of protein and starch. Similar results were reported earlier by Helsper et al. (1993).

Therefore, the superior nutritional value of low tannin genotypes is not simply due to the absence of tannins, which directly have negative effects on monogastric digestion, but also to the intrinsic characteristics of the seeds with respect to nutrient concentration.

The presence of condensed tannins and other phenolic compounds in the seed coat also increased the cooking time of faba beans after dry storage (Nasar-Abbas et al. 2009), decreasing the marketability and consumption of dry faba bean as a human food.

2.3 Genetics of the *low tannin* phenotype and molecular breeding approach

Faba bean accessions with high condensed tannins content in the seed coat also display coloured flowers (normally with a large black spot in the wing), while the low-tannin varieties have completely white flowers. The white flower characteristic in faba bean is controlled by two recessive genes, *zt1* and *zt2,* which were described by Picard (1976). The presence of either gene in the homozygous recessive form is sufficient to produce the low tannin and white flower phenotype (Crofts et al., 1980).

One of the main goals of faba bean breeders is to enhance grain quality. During the past three decades plant breeders have developed low tannin varieties that express one of the two genes (O'sullivan and Angra, 2017). Duc et al. (1995) reported a broad variation in total phenolics, condensed tannins and tanning power (on precipitation of hemoglobin) among 16 faba bean genotypes chosen for different characteristics of their flower and seed coat colours.

The selection and screening of low tannin varieties is relatively easy (Gnanasambandam et al., 2012), because the trait is monogenic for either of the two genes, and plants that are homozygous recessive at either locus, produce white flowers and low tannin seeds. However,

crosses between both types of low tannin (both *zt* genes) produce F1 plants with high tannin and coloured flowers.

Therefore, it is important to maintain careful isolation during the seed multiplication process of low tannin varieties expressing either of the low tannin genes (Crofton et al. 2010). In Canada low and high tannin varieties and low tannin varieties with different low tannin genes must be cultivated at an isolation distance of at least 1000 m to reduce outcrossing by insect pollinators, to minimize the production of contaminated seed lots (Canadian Seed Growers Association).

Genes involved in the control of condensed tannins in the seed coat and pigmentation characteristics in the plant have been reported for other grain legumes. The *P* gene in common bean (*Phaseolus vulgaris* L.) is considered to be the controlling factor for the presence or absence of different flavonoids in the seed coat (Caldas and Blair 2009). The single recessive *tan* gene in lentil (*Lens culinaris*) controls the presence of condensed tannins in the seed coat with additional pleiotropic effects on plant pigmentation and seed coat thickness (Vaillancourt et al. 1986). In chickpea the presence of a single recessive gene determines the white flower characteristic and reduces condensed tannins in the seed coat (Gil and Cubero 1993), while in pea (*Pisum sativum* L.), the white flower characteristic is controlled by the *A* gene notably described by Gregor Mendel in 1865. The A gene was described in detail by Hellens et al. (2010) and it encodes a bHLH transcription factor which controls anthocyanin pigmentation in pea.

The identification of two complementary genes, *zt1* and *zt2*, which has apparently have the same function, *zt1* and *zt2*, in faba bean may be explained by the presence of repetitive DNA sequences that consist of different types of transposable elements in the genome of faba bean

(Cooper et al. 2017). Therefore, the two genes may be just a repetition of each other in the faba bean genome. A similar scenario was found in cotton (*Gossypium hirsutum* L.) by Lu et al. (2017), who characterized two homologous genes of *Arabidopsis thaliana* TT2 in tetraploid cotton. The genes were able to complement the mutant phenotype of *Arabidopsis* tt2, and induced proanthocyanidin accumulation. However, cotton is a polyploid species, therefore, the case of faba bean is unique among the diploid pulses.

The determination of which of the two alleles (*zt1* or *zt2*) is present in low tannin faba bean cultivars and gene bank accessions is necessary to enable breeders to choose the appropriate parents for hybridization (Torres et al., 2010). Therefore, to speed up and facilitate traditional breeding methods, the use of molecular tools such as DNA markers is a fundamental step towards the development of improved varieties. Useful DNA markers play a vital role in enhancing global food production, making selection more efficient, reliable and cost effective (Collard et al., 2005).

Faba bean is a diploid species with an exceptionally large genome compared to other grain legumes, with 13 Gb over 12 chromosomes (*n*=6 chromosomes). Because of these very large chromosomes, faba bean was used as a model species for plant cytogenetic studies in the 1970s and 1980s, but after the emergence of other plant model species such as *Arabidopsis thaliana* and *Medicago truncatula*, faba bean genome studies were progressively abandoned, with the result that now faba bean genetics and genomics resources are quite scarce. A complete current review of the genetics and genomics tools available in faba bean was reported by O'Sullivan and Angra (2017).

The availability of reliable molecular markers for the *zt1* and *zt2* genes will speed up the selection and development of low tannin faba bean varieties. Previous works to tag the *zt1* and *zt2* genes were done through the development of SCAR (Sequence Characterized Amplified Region) markers after the identification of RAPDs (Random Amplification of Polymorphic DNA) markers using BSA (Bulked Segregant Analysis) of an F2 population derived from a cross between a low tannin line and a wild type (Gutierrez et al. 2006; 2007). The markers were then validated to test their effectiveness in independent populations with diverse genetic backgrounds. For *zt1* the SCAR marker allowed the prediction of genotypes with 95% accuracy confirming its potential as a cost-effective breeding tool (Torres et al., 2010). For the SCAR marker linked to the *zt2* gene, the result of the validation process was not as reliable as for the *zt1* gene. The lower frequency of the *zt2* gene in low tannin faba bean germplasm is the main reason for the low success of the marker validation. A problem with this and other DNA markers related to different traits in faba bean is that they require a diversity of amplification, labelling and separation techniques and have not yet been systematically assessed across a set of germplasm (Cottage et al., 2012).

Most of the low tannin released cultivars contain the *zt1* gene. Duc et al. (1999) suggested that the improvement of faba bean nutritional value could result from wider use of the *zt2* gene. For these reason, a more reliable marker for *zt2* is needed to improve the precision of predictability in the plant breeding process.

Markers that allow quicker, cheaper and reproducible gene tagging and mapping can be developed through the use of Single Nucleotide Polymorphisms (SNPs).

Webb et al. (2016) mined and validated a large number of SNPs from the aligned transcriptomes of two contrasting inbred lines of faba bean. The SNP markers were validated

using the KASP assay analysis (Kompetitive Allele Specific PCR), that enables bi-allelic scoring of single nucleotide polymorphism (LGC Genomics©). Each SNP marker was linked to a single orthologue gene in the fully sequenced genome of the model species *Medicago truncatula* through BLAST analysis. The use of comparative genomics between different legume species can help to understand the conservation and differences in gene content and order among different taxa (Gnanasambandam et al., 2012). This technique has the potential to increase the knowledge of faba bean genome, which is the largest among all grain legume species.

The first gene-based genetic linkage map for faba bean was developed by Ellwood et al. (2008). This map was later enlarged and enriched with new SNP markers (Cottage et al. 2012; Webb et al. 2016) to yield the most updated version of the map called *Vicia faba* 2014 consensus map. The study by Webb et al., (2016) not only widened the marker coverage of faba bean genome thereby increasing the number of validated SNPs, but also for the first time a flower colour gene was mapped. The trait was mapped to chromosome two of faba bean in a region of highly conserved synteny with *Medicago truncatula* chromosome three. This gene was proven to be *ZT1*, the dominant version of *zt1.*

To date no one has reported the existence of a genotype recessive at both loci (*zt1zt2*). In previous literature it was suggested that the genotype may not be viable (Crépon et al., 2010), but this was never proven.

The *zt1zt2* genotype may have superior quality characteristics compared to genotypes recessive at a single locus, and the characterization of the double recessive genotype may also help to better understand the mode of action of the two complementary genes *zt1* and *zt2* in the phenolic compounds biosynthetic pathway.

2.4. Biochemistry and analysis of phenolic compounds

2.4.1 Phenolic compounds, classification and biosynthesis

Phenolic compounds are secondary metabolites that have one or more hydroxyl groups attached directly to an aromatic ring. The basic phenol structure is shown in Figure 2.1.

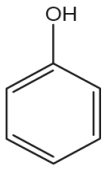


Figure 2.1. Structure of the phenol group

Phenolic compounds are a very large group of chemical compounds which can be classified in diverse ways. The classes of phenolic compounds that were considered in this research are: *phenolic acids* (which include *hydroxybenzoic acids* and *hydroxycinnamic acids*), *stilbenes*, *chalcones* and *flavonoids*. *Flavonoids* are a major class of phenolic compounds, they have a basic A-B- and C-ring structure, and they are normally represented as shown in Figure 2.2. This structure contains 15 carbon atoms arranged in a C6-C3-C6 structure (Vermeris and Nicholson, 2006).

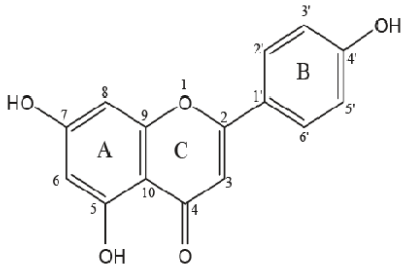


Figure2.2. Basic structure of the compounds belonging to the flavonoids class

The class of *flavonoids* is comprised of the following sub classes: *flavanones, isoflavones, flavones, dihydroflavonols, flavonols, leucoanthocyanidins, anthocyanidins, anthocyanins, flavan-3-ols and proanthocyanidins*.

Proanthocyanidins are referred to as condensed tannins, which are oligomeric or polymeric *flavonoids* consisting of *flavan-3-ol* units. Together with hydrolysable tannins, and complex tannins they are grouped as tannins, compounds which can bind and precipitate proteins (Vermeris and Nicholson, 2006). *Anthocyanins* and their aglycones *anthocyanidins* are plant pigments responsible for the coloration of different plant tissues (Tanaka et al. 2008; Truong et al. 2010).

Proanthocyanidins and *anthocyanins* share the upstream part the flavonoid pathway and are synthesized via different final branches. The pathway starts from *p*-coumaroyl Coenzyme A as a substrate. *P*-coumaroyl Coenzyme A is the end product of the phenylpropanoid pathway and is a substrate common to a number of phenylpropanoid compounds (Vermeris and Nicholson, 2006).

The flavonoid biosynthetic pathway has been deeply studied and most of the genes involved in the key passages of the pathway were characterized in model species (He et al. 2008). Furthermore, the ternary complex of transcription factors called MYB-bHLH-WD40 (MBW) is well recognized to play a key role in the regulation of the late steps of the pathway leading to the production of *anthocyanins* and *proanthocyanidins* (Albert et al. 2014; Li et al 2016).

2.4.2 Phenolic compounds in faba bean

Most of the work done so far to characterize the phenolic compounds in faba bean targeted the so called “tannins”, because of their interaction with protein.

Different studies have shown that tannins negatively affect protein digestion in mainly two ways. First they interact with protein to form tannin-protein complexes which tend to be insoluble, thus promoting protein precipitation; secondly, they block the activity of protein-digesting enzymes such as trypsin, with additional negative effects on enzymes like α-amylase and lipase on starch and lipids, respectively (Aw and Swanson, 1985; Butler et al., 1982; Yuste et al., 1992). Overall lipids and starch digestion is less affected because of the abundance of these enzymes in the digestive systems (Yuste et al., 1992). For protein digestion, tannins represent a big problem that results in limited use of faba bean in animal diets. Previous reports also showed a negative effect of tannins on voluntary feed intake by animals (Mangan, 1988). However, a limited PAs (*proanthocyanidins*) concentration is a positive characteristic in forage crops, because they can assist with avoidance of lethal pasture bloat in cattle by binding dietary proteins, thereby, slowing down their fermentation in the rumen. Positive effects of PAs on lactation, wool growth and live-weight gain were also reported for sheep (Aerts et al. 1999).

Considering their chemical structure, tannins can be divided in two different groups, termed condensed and hydrolysable tannins, although both can undergo hydrolytic processes in aqueous media. They differ in their component subunits and the bonding between them. Condensed tannins are polymers derived from *flavan-3-ols* while hydrolysable tannins are composed of gallic acid or ellagic acid esterified to a sugar moiety (Mole and Waterman, 1987). Different tannins react diversely with proteins depending on the spatial configuration of the molecules (Mangan, 1988). The basic structure of *flavan-3-ols* is shown in Figure 2.3**.**

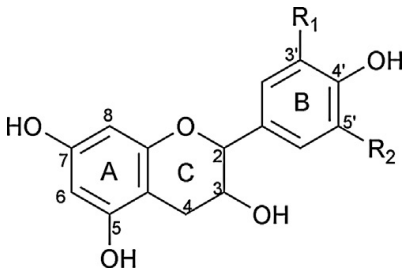


Figure 2.3. Basic structure of the flavan-3-ol unit.

Various techniques have been used to investigate the presence of tannins in plant tissues, some of them without adequate consideration of their limitations (Mangan et al. 1988). Methods reported for the quantification of condensed tannins in faba bean seed coats include the Bate-Smith (1953) vanillin-hydrochloric acid reagent, and the vanillin-sulfuric acid assay (Hesper et al. 1993). Another common method is the relative-astringency method of Bate-Smith (1973) which depends on precipitation of the whole protein by hemolyzed blood (haemanalysis), as measured by hemoglobin absorption at 578 nm wavelength and comparison with standard amounts of tannic acid, implemented for faba bean analysis by Duc et al. (1995). The vanillin-HCL method (Butler, 1982; Makkar and Becker 1993) was used for faba bean by Duc et al. (1999). The Folin–Ciocalteu reagent method as described by Singleton and Rossi (1965) for the quantification of the total phenolics content was used for faba bean by Chaieb et al. in 2011.

More recently, reports used LC-MS (Liquid Chromatography-Mass Spectrometry) to characterize various classes of phenolic compounds in different faba bean tissues. In fact, LC-MS has progressed to become a very powerful analytical tool for both quantitative and qualitative applications (El-Aneed et al. 2009). With this technique Nozzolillo et al. (1989) identified different *flavonoid* aglycones in the seed coats of various faba bean lines differing in seed coat colour. Bekkara et al. (1998) analyzed the phenolic exudation in the seed and root of two faba bean

cultivars, identifying catechin, *phenolic acids* derivatives and various *flavones* and *flavanones*. Merghem et al. (2004) characterized six *pronthocyanidins* in faba bean seed coats. More recently, Baginsky et al. (2013) identified and characterized different *procyanidins*, *prodelphinidins*, *flavonols* and *flavones* in immature faba bean seeds. In the later study the authors used tandem MS/MS spectra to assign unknown compounds into the most appropriate class of phenolic compounds based on their fragmentation pattern. In 2016, Šibul et al. studied the phenolic profile of different secondary metabolites present in the dry herb and root parts of different legumes and found high concentration of phenolic acids and flavonoids in faba bean leaves and roots.

To our knowledge there is no published study which connects the biochemical results of phenolic profiles analysis with the genetic control of these traits in faba bean. Furthermore, there are no information on how the two different low tannin genes *zt1* and *zt2* affect the phenolic compound biosynthetic pathway, nor on the differences in the phenolic profiles of low tannin faba bean accessions expressing the two low tannin genes.

Mirali et al. (2016) described the phenolic compound profile of lentil lines differing for their seed coat colour and were able to identify the point of action of the low tannin gene (*tan*) in the biochemical pathway of phenolic compounds.

The experiments described in chapter 5 of this thesis used a similar approach to obtain a detailed description of the phenolic compounds present in the seed coat of faba bean, for both the wild types and the low tannin lines. The overall goal was to gain sufficient knowledge to be able to propose a putative point of action in the *flavonoids* biosynthetic pathway for the two low tannin genes *zt1* and *zt2* in faba bean.

The research was designed to test the following hypotheses:

- 1. SNPs (Single Nucleotide Polymorphisms) closely linked to the *zt2* gene will be identified using the RIL (Recombinant Inbred Line) population derived from Disco/2 by ILB 938/2 cross.
- 2. KASP(s) marker(s) designed from the SNPs will be effective in discriminating faba bean genotypes containing the recessive allele *zt2* from those containing the dominant allele *ZT2*.
- 3. The double recessive genotype *zt1zt2* is viable and can be successfully identified with molecular markers.
- 4. The phenolic profiles in the seed coat and flowers of both low tannin lines (*zt1* and *zt2*) are different from those of the wild type.
- 5. There are differences in the phenolic profiles of the seed coat and flowers between the lines recessive either for *zt1* or *zt2.*

CHAPTER 3 - DEVELOPMENT OF MOLECULAR MARKER FOR THE *zt2* GENE

3.1 Abstract

Low tannin faba beans are increasingly requested by the feed and food industries for their superior nutritional quality and technical characteristics compared to varieties with normal tannin content. The low tannin phenotype is controlled by two recessive genes *zt1* and *zt2*, and the trait is linked to the absence of pigmentation in the flowers. In this work the $F_{5:6}$ derived individuals of a RIL population (FBR-8) were genotyped with a set of 190 SNP markers syntenic to the *Medicago truncatula* genome. The linkage map yielded 9 linkage groups (LG), and *zt2* was mapped on LG2 in a region collinear to Chromosome 3 of *Medicago truncatula.* The closest flanking SNP markers to *zt2* were then validated on a set of genotypes with diverse allelic states at the *zt2* locus and one of these SNP markers successfully distinguished faba bean genotypes carrying the recessive allele *zt2* from those with the dominant allele *ZT2.*

3.2 Introduction

Faba bean accessions with high condensed tannin content in the seed coat also display coloured flowers (normally with a large black spot in the wing petal), while the so-called low tannin varieties have completely white flowers. The white flower characteristic in faba bean is controlled by either one of two recessive genes, *zt1* and *zt2,* which were discovered by Picard (1976). The presence of either of the two genes in the homozygous recessive form is sufficient to produce the phenotype of low tannin and white flower (Crofts et al., 1980). These phenotypes are commonly referred to as low tannin.

The availability of reliable, low cost molecular markers for the genes *zt1* and *zt2* will speed up the selection and development of low tannin faba bean varieties. Previous research to tag the

zt1 and *zt2* genes was done through the development of SCAR (Sequence Characterized Amplified Region) markers after the identification of RAPDs (Random Amplification of Polymorphic DNA) markers using BSA (Bulked Segregant Analysis) of an F2 population derived from a cross between a low tannin line and a wild type (Gutierrez et al. 2006; 2008). The markers were then validated to test their effectiveness in independent populations with diverse genetic backgrounds. For *zt1* the SCAR marker allowed the prediction of genotypes with 95% accuracy, confirming its potential as a cost-effective breeding tool (Torres et al., 2010). In 2016, Webb et al. mapped the *ZT*1 gene in a genomic region of faba bean using a set of SNPs from the aligned transcriptomes of two contrasting inbred lines of faba bean and subsequently linked to a single orthologue gene in the genome of the model species *Medicago truncatula.*

For the SCAR marker linked to the *zt2* gene, the result of the validation process was not as strong and reliable in comparison to that for the *zt1* gene. The lower occurrence of the *zt2* gene in low tannin/white flower faba bean germplasm is the main reason for the low success of the marker validation. Therefore, a reliable marker for *zt2* is still needed.

3.3 Materials and Methods

3.3.1 *Plant material and phenotyping*

A recombinant inbred line (RIL) population of 174 $F_{5:6}$ derived individuals developed by single seed descent (SSD) was used for mapping the *zt2* gene. The population (FBR-8) was developed at the University of Helsinki (Khazaei et al., 2014) from the cross between Disco/2 a French line bred at INRA (Institut National de la Recherche Agronomique) and accession ILB 938/2 from the ICARDA (International Center for Agricultural Research in Dry Areas) germplasm collection. Disco/2 is a zero-tannin line with white flower homozygous for the recessive *zt2* gene

and ILB 938/2 is an Ecuadorian land race with spotted flower phenotype. Flower colour (Figure 3.1) is described as white (absence of pigments in the petals) or spotted (to denote typical wild-type petal markings which include a pronounced black spot on each wing petal and reddish-purple vein markings on the standard petal).

Figure 3.1. Coloured (left) and colourless (right) flower in faba bean.

At each generation during the population development, the plants were grown in pollinator-free environments to prevent outcrossing.

The RILs at the F6 generation were grown in a growth chamber in the Controlled Environment Facility of the College of Agriculture and Bioresources at the University of Saskatchewan, Saskatoon, Canada. Photoperiod/temperature was adjusted to 16h/22 °C day and 8h/18 °C night. The photon flux was set to 300 µmol m^{-2} s^{-1}. Plants were grown in 10 cm plastic pots filled with soilless growing mix Sunshine No. 3 (Sun-Gro Horticulture MA, USA). The plants were fertilized with NPK fertilizer (20-20-20) every two weeks and were watered manually to keep moisture conditions for optimal growth. *Amblyseius cucumeris* predatory mites (Biobest Group NV Westerlo, Belgium) were applied once a week to control thrips (*Heliothrips haemorrhoidalis).* A set of 19 genotypes (including the parents of the mapping population) differing for their allelic state at the *zt1* and *zt2* loci, called validation panel 1, was assembled to

validate the closest SNP markers for *zt2.* The validation panel was grown under the same controlled conditions as described above. Data regarding the origin of the genotypes of the validation panel are presented in Table 3.1.

A second validation process was performed on another set (validation panel 2) only for a KASP marker that was designed in our lab from the closest SNP marker selected after testing on validation panel 1. Marker analysis was done at LGC Genomics (Beverly, MA, USA). Validation panel 2 consisted of a set of F1 seeds and faba bean accessions. Details of validation panel 2 are presented in Supplementary Table 3.1**.**

Flower colour was visually recorded for all genotypes in the mapping population at each generation during development of the RIL population. The trait was scored as white vs spotted flower. Stipule spot pigmentation (SSP) and hilum colour (HC) data were also visually recorded at each generation and were included as additional morphological markers. They were scored as coloured vs uncoloured SSP and black vs uncoloured hilum respectively (Supplementary Figure 3.1 a and b).

Table 3.1. *Zt2* locus genotype, and phenotype of stipule spots, hilum colour, flower colour and seed coat, of 18 faba bean inbred lines used in validation panel 1.

Inbred line	*Zt*2 allele genotype [a]	Stipule spot colour [c]	Hilum colour	Flower colour[b]	Seed coat colour	Origin / Source
Disco/2	*zt2*	colourless	colourless	white	beige	INRA, France
Disco/2 polled	*zt2*	colourless	colourless	white	beige	INRA, France [d]
ILB938/2	*ZT2*	colourless	black	spotted	green	Ecuador ICARDA [e]
1778(Crimson)	*ZT2*	spotted	black	red	beige	Sweden
346-10	*ZT2*	spotted	colourless	spotted	beige	U of S [f]
687-8	*ZT2*	spotted	black	spotted		UofS
AO1155	*ZT2*	colourless	colourless	white(*zt1*)	beige	AO, France
B2-22-11	*ZT2*	spotted	black	brown		
Aurora/2	*ZT2*	spotted	black	spotted	beige	Sweden
FB9-4	*ZT2*	spotted	black	spotted		UofS
L-43	*ZT2*	spotted	black	spotted	dark brown	

L-170	*ZT2*	spotted	black	spotted	beige	
L-175	*ZT2*	spotted	black	spotted		
NV153-1	*ZT2*	spotted	black	spotted		
P47-1B	*ZT2*	spotted	black	red	green	
Rinrei	*ZT2*	spotted	black	spotted		Japan
Snowbird	*ZT2*	colourless	colourless	white(*zt1*)	beige	Netherlands
CDCSnowdrop	*ZT2*	colourless	colourless	white(*zt1*)	beige	U of S
Gelber	*ZT2*	spotted	black	yellow		Sweden

[a] *ZT2*: wild-type (normal-tannin) allele; *zt2*: recessive allele conferring low tannin content.
[b] Flower colour is described as white (no pigments in the petals) or spotted (to denote typical wild-type petal markings which include a pronounced black spot on each wing petal and dark vein markings on the banner petal). Illustrated in Figure 3.1.
[c] SSP, Stipule spot pigmentation. [d] INRA, Institut National de la Recherche Agronomique. [e] ICARDA, International Centre for Agricultural Research in the Dry Areas, [f] University of Saskatchewan.

3.3.2 Statistical analysis

The observed phenotypic segregation ratios of the flower colour (FC), stipule spot pigmentation (SSP) and hilum colour (HC) were tested against the expected Mendelian genetic ratio using the χ^2 goodness-of-fit test. SAS software Version 9.4 (Copyright © 2002-2012 SAS Institute Inc., Cary, NC, USA) was used to perform the calculations.

3.3.3 Discovery and validation of the *zt*2 Marker

DNA was extracted from the leaf tissue (8 discs/genotype) of each individual of the F6 mapping population (Disco/2×ILB938/2) and from the parental lines and accessions of validation panel 1 using an LGC© plant sample collection kit (LGC Genomics) following the manufacturer instructions https://www.lgcgroup.com/plant-kit/. A set of 875 SNP markers, available from the Vf 2014 Consensus map constructed by Webb et al. (2016) was used to develop a linkage map and to identify the closest SNP marker to the *zt2* gene.

SNP genotyping was carried out using the KASPar™ (Kompetitive Allele Specific PCR) assay (KBioscience, UK) platform (Semagn et al. 2014) at LGC Genomics (Beverly, MA USA).

Primers for the KASP assay (two allele-specific forward primers and one common reverse primer) were designed using LGC Primer Picker software (LGC Genomics, Beverly, MA, USA) for the SNP that correctly discriminated the genotypes of validation panel 1. This KASP marker was then tested on validation panel 2 using DNA from cotyledons.

KASP reaction using cotyledon-derived DNA was performed using BioRad CFX384™ Real Time PCR System (Bio-Rad Laboratories, Inc) at the Pulse Crop Genotyping Lab (Department of Plant Sciences, University of Saskatchewan, Canada). Assay reaction volume was 10 µL of 50 ng/µL DNA, 2X KASP Reaction Mix, and 0.17 µM KASP Assay Mix (allele-specific primers, A1 and A2, and common primer, C1). Fluorescence was analyzed using CFX Manager Software. The PCR for the KASP marker assay was set at 95 °C for 15 min, followed by a 10 cycles touchdown of 94 °C for 20 s and 65 °C (Δ -0.8°C/cycle) for 2 min, then followed by up to 50 cycles of 94 °C for 20 s and 57 °C for 1 min.

Details on the KASPar principle, amplification of targeted region, fluorescence detection, and allele calling are available at http://www. kbioscience.co.uk/reagents/KASP_manual.pdf.

3.3.4 Linkage map

A linkage map was constructed using MapDisto v. 1.7.7.0.1 (Lorieux, 2012) with logarithm of odds (LOD) score of 3.0 and recombination fraction of 0.3. The map was generated using the procedure Automap. The map distance in centiMorgans (cM) was calculated using the Kosambi mapping function (Kosambi, 1943). The linkage map was plotted using Map Chart version 2.32 (Voorrips, 2002).

3.3.5 Candidate genes

Search for candidate genes orthologous to the flanking SNP markers for *zt2* in the fully sequenced genome of the model legume species *Medicago truncatula*, was done using Phytozome (https://phytozome.jgi.doe.gov/pz/portal.html). Sequences flanking the SNP markers were BLASTED against the *M. truncatula* genome (Mt4.0v1), and best hits for these sequences were considered depending on their E values. J Browse function (http://www.medicagogenome.org/jbrowse/) (Mt4.0v1), was then used to investigate the whole genomic region in between these SNPs markers for the presence of possible candidate genes.

3.4 Results and discussion

3.4.1 Segregation ratio

Data for segregation of coloured vs white flower colour (FC) at each generation of the RIL mapping population are presented in Table 3.2.

Table 3.2. Observed phenotypic segregation ratios, value of χ2 test and corresponding P value for flower colour (FC) coloured vs white among the RILs of the mapping population Disco/2 × ILB938/2 across the generations.

RILs	FC	FC		
Generation	Observed	Expected	χ2	*P*
F2	222:69	3:1	0.2577	0.6117
F3	182:99	5:3	0.6171	0.4321
F4	131:91	9:7	0.6867	0.4073
F5	107:98	17:15	0.0712	0.789625
F6	87:87	1:1	0.000	1

The phenotypic segregation ratios for coloured flowers vs white flowers across all the generations confirms that in the population Disco/2 × ILB938/2 flower colour segregated for a single gene (*zt2*) with the dominant allele conferring the spotted and the recessive allele the white flower characteristics, respectively.

3.4.2 Linkage map development

From the initial screening of 875 SNPs, 243 SNPs were polymorphic among the parental lines, and these were used to genotype all individuals of the mapping population and the accessions of validation panel 1.

After genotyping, the data of the SNP markers were manually cleaned using SNP viewer (http://results.lgcgenomics.com/software/snpviewer/). SNPs that had more than 25% missing data, had more than 20% heterozygosity, or were monomorphic, were eliminated from further analysis. This filtering step was done to rule out those SNPs which may compromise the quality of the linkage map. In particular, SNPs with missing data are those that gave no result for some genotypes regarding their allelic state. This may be due to errors during the KASP reaction. Heterozygous SNPs are those that present heterozygous calls in some genotypes. This is only partially expected in a RIL population at the F6, because all genotypes should be homozygous for one of the two parental alleles at most loci. The expected homozygosity after 6 generations of self-pollination is 98.5% (Singh and Singh, 2015). In our study we maintained a higher tolerance for heterozygous SNPs (compared to the theoretically expected 1.5%), to account for possible errors during the genotyping process, and we eliminated only the SNPs which were heterozygous in more than 20% of the genotypes. Monomorphic SNPs were those that showed no polymorphism among the genotypes of the mapping population. At the F6 all individuals of the

population should resemble the allelic state at each locus of either one of the two parents. Ideally, for all SNPs we should obtain a 1:1 ratio for frequency of the two parental alleles, and monomorphic SNPs were therefore eliminated.

The cutoff values described above for eliminating the rejected SNPs from our linkage analysis were chosen to maintain the highest possible SNPs density, but at the same time to eliminate those SNPs which would have negatively affected the accuracy of the linkage map.

A final set of 190 SNPs markers was used to construct the linkage map. Information about the SNP markers used to construct the linkage map are presented in Supplementary Table 3.2**.**

Finally, 4 genotypes were eliminated from our linkage analysis because they presented more than 25% missing data when all loci were considered. Therefore, the final set of genotypes used for the linkage analysis was reduced to 172.

To map the low tannin locus (*zt2*), we included the binary phenotypic score FC (white versus spotted flower) as an extra morphological marker into the linkage map. Furthermore, the seed hilum colour (HC) scored as black or colourless was also included as a morphological marker. Additional information about hilum colour (HC) and stipule spot pigmentation (SSP) phenotypes are presented in **Appendix 5**.

The map spanned 917.94 cM over 9 linkage groups. The lengths of the linkage groups ranged from 8.3 cM (LG6) to 178.2 cM (LG3). The full map is presented in Supplementary Figure 3.2**.**

The number of genetic linkage groups should be the same as chromosomes number, so for faba bean we would expect six linkage groups. However, the number of LGs often exceed the chromosome number, ranging from 7 to 48 in faba bean (Khazaei, 2014). Our map had 6 main

linkage groups and 3 small fragments. This mismatch between the LG number and chromosome number is mainly due to the co-ancestry of the two parents, and also to the marker development strategy. These have consequences for the density and proximity of the markers, and thus influence the number of LGs (Cheema and Dicks, 2009).

The trait flower colour (FC), corresponding to *zt2,* was mapped on linkage group 2 (LG2) which overlaps with Chromosome 3 in the 2014 *Vicia faba* consensus map (Webb et al., 2016). The portion of LG2 containing the FC locus is shown in Figure 3.2**.** This result proves at the molecular level that *zt1* and *zt2* are independent genes present on different chromosomes. The dominant allele *ZT1* was mapped on Chromosome 2 of the 2014 *Vicia faba* consensus map published by Webb et al. (2016). This confirms that the two *zt* loci are independent. This result was proven long ago at the genetic level (Picard et al. 1976), but this is the first time to our knowledge, that the two low tannin genes are tagged with molecular markers in a single physically anchored consensus linkage map.

The previous work for tagging *zt1* and *zt2* (Gutierrez at al. 2006, 2008) yielded SCAR markers linked to the two low tannin genes. However, these molecular markers were not assigned to a linkage map nor to specific chromosome locations. Furthermore, these markers were only partially validated in a diverse genetic background.

The closest SNP markers flanking the *zt2* locus are Vf_Mt1g072640_001 and Vf_Mt1g072740_001, both located three cM from *zt2* and Vf_Mt7g100500_001 located 10.5 cM from *zt2.* These two marker loci are found at 56.6 and 43.0 cM from the end of the linkage group, respectively.

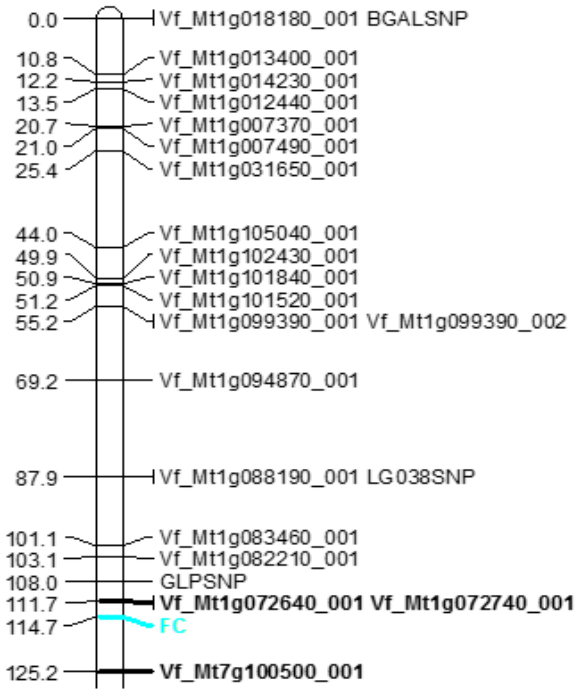


Figure 3.2. The flower colour locus (FC) in the linkage map is highlighted in blue. FC was mapped in linkage group 2, between the orthologous SNP markers Vf_Mt1g072640_001 and Vf_Mt1g072740_001 and SNP marker Vf_Mt7g100500_001.

3.4.3 Marker validation

The co-segregating markers Vf_Mt1g072640_001 and Vf_Mt1g072740_001 are the closest to the *zt2* gene, but the most effective molecular marker for gene tagging of *zt2* is the SNP flanking marker Vf_Mt7g100500_001 which is 10.5 cM from the FC locus in our map. In fact, only Vf_Mt7g100500_001 distinguished genotypes carrying the *zt2* allele in validation panel 1 with 100% accuracy. This is probably because the markers used to construct a linkage map are typical of the specific population used in the linkage study, and some of the polymorphisms may be lost in individuals of other accessions with different genetic backgrounds. This confirms the

importance of the validation step in the process of developing molecular markers (Collard et al. 2005).

KASP primers for Vf_Mt7g100500_001 SNP marker were then designed and used to perform a second validation step on a diverse set of faba bean accessions and some F1 seeds in validation panel 2 (Supplementary Table 3.1). The allele-specific primers for SNP Vf_Mt7g100500_001 were:

A1 = 5' GAAGGTGACCAAGTTCATGCTGAGTTAGCAGCTGCTACTAGAAAC-3', and A2= 5'-GAAGGTCGGAGTCAACGGATTGTGAGTTAGCAGCTGCTACTAGAAAT-3'. The conserved primer was C1 = 5'-CCTCCCTCGCCCAGAAGACATT-3'.

In this case the marker designed from the SNP Vf_Mt7g100500_001 correctly discriminated the *zt2* allele from the *ZT2* allele with 100% accuracy (Supplementary Table 3.1) matching the predicted *zt2* allele frequency for each of the 42 tested genotypes. The ease of design, and the effectiveness of this KASP marker confirms the potential marker systems based on SNPs for the construction of linkage maps, and rapid identification of crop cultivars (Agarwal et al. 2008).

A reliable, quick and easy to use KASP marker is an inexpensive tool for breeders who may wish to use *zt2* as a source of the white flower/low tannin characters for new faba bean varieties. Using KASP markers, the selection of these traits is possible at a very early stage of plant development, or even without the need to grow the plant, by simply extracting DNA from the cotyledons of the seed. It is notable that the partial out-crossing behavior of faba bean and the two genes controlling the low tannin/white flower phenotype are important challenges for plant breeders targeting these traits. Not only will the cross pollination between a low tannin variety

and a wild type lead to the loss of the low tannin trait in the progeny, but also the cross pollination of two low tannin varieties carrying the two different genes *zt1* and *zt2* will lead to the same result. For this reason, the availability of reliable molecular markers tagging both these genes will allow for a much faster and easy to perform quality checks of breeding material. For example, it will be possible to quickly assess the low tannin genotype of new accessions before their introduction into a breeding program, or to facilitate the inclusion of the low tannin genes in faba bean cultivars with improved nutritional value for human and animal consumption.

3.4.4 Candidate gene search

A reference genome for faba bean is still under development, and there is no information available yet regarding possible candidate genes for *zt2*. Recently, comparative genomics and synteny analysis have been implemented in faba bean, permitting the localization of QTL in regions with clear collinearity to fully sequenced model genomes (O'Sullivan and Angra, 2016).

The best hits on the *Medicago truncatula* genome for the BLASTED sequences of the SNP markers Vf_Mt1g072640_001, Vf_Mt1g072740_001 and Vf_Mt7g100500_001 are genes that express the following protein domains respectively: a PALE CRESS protein (PAC), a WD40 repeat protein, and a threonine dehydratase biosynthetic protein (http://www.medicagogenome.org/jbrowse/) (Mt4.0v1).

None of these proteins are expressed in flower tissue, nor are they reported to have a direct function in the biosynthetic pathway of phenolic compounds. WD40 repeat proteins were previously reported to be part of a complex system of transcription factors which are involved in the regulation of genes involved in the synthesis of anthocyanins in *Arabidopsis thaliana* (Walker

et al. 1999) and different legume species (Hellens et al. 2010, Mirali et al. 2016, Li et al. 2016) including faba bean (Webb et al. 2016). The latter study identified, through synteny, the *ZT1* gene (the dominant version of *zt1*) in faba bean as an ortholog of the Medicago WD-40 transcription factors, Transparent Testa Glabra 1 (TTG1). This may suggest that *ZT2* could also express a transcription factor involving one or more WD40 domains regulating the expression of genes in the flavonoid pathway. Following this hypothesis, we could infer that *zt2* is in fact a duplication of *zt1,* this being one case of repetitive DNA sequences reported in faba bean and other lineages that have 'giant genomes' (Cooper et al. 2017). However, genes encoding WD40 repeat proteins are very common in plant genomes and are involved in the regulation of expression of a very wide array of genes.

The region between the flanking SNP markers listed above spanned 5.27 Mb in the fully sequenced genome of *Medicago truncatula* (http://www.medicagogenome.org/) and contains about 700 genes for which the function of their expressed transcripts (mRNA) was characterized. This region of Chromosome 3 of faba bean is well conserved in Chromosome 1 of *Medicago truncatula* and a search for candidate genes characterized in the model species could be promising.

The next step for the identification of a candidate gene for *zt2* will be to fine map the genomic region where *zt2* is located with a higher number of SNPs markers to narrow down the search to a much shorter segment in the genome of *Medicago truncatula*.

3.5 Conclusions

The improvement of faba bean seed quality traits is one of the most crucial step towards greater adoption and use of the crop. From this perspective the low tannin characteristic,

together with low vicine-convicine (Khazaei et al. 2017) are the most critical traits to incorporate into new varieties.

In this work we developed and validated a reliable molecular marker for *zt2* in the inexpensive and easy to use KASP assay format. This marker successfully discriminated low tannin faba bean lines carrying *zt2* from those carrying *zt1* and wild type alleles of both genes.

CHAPTER 4- IDENTIFICATION OF THE DOUBLE RECESSIVE GENOTYPE *zt1zt2*

4.1 Abstract

Faba bean genotypes recessive for either of the two low tannin genes *zt1* or *zt2* present the same phenotype of white flower and low tannin in the seed coat. A genotype recessive for both genes may have a new phenotype, and improved nutritional quality compared to the genotypes recessive for a single low tannin gene. This study tested different molecular markers for their efficacy in tagging the two low tannin genes *zt1* and *zt2* in the $F_{2:3}$ derived individuals of a population segregating for the two low tannin genes. The KASP marker for *zt2* was effective in discriminating genotypes carrying the recessive allele *zt2zt2* from those carrying the dominant allele *ZT2ZT2* and the heterozygous *ZT2zt2.* The SCAR marker SCC_{5551} was partially successful in tagging the *zt1* allele.

4.2 Introduction

The white flower (low tannin) phenotype in faba bean is controlled by either of two independent recessive genes, *zt1* and *zt2*. So far no one has reported the existence of a genotype recessive at both loci (*zt1zt2*). In the previous literature it was suggested that the genotype may not be viable (Crépon et al., 2010), but this hypothesis was never proven.

The *zt1zt2* genotype may have superior quality characteristics compared to genotypes recessive at a single locus, and the characterization of the double recessive genotype may also help to better understand the modes of action of the two genes *zt1* and *zt2.*

Furthermore, faba bean is a partially allogamous crop. This fact poses some challenges to the production of tannin free seeds. In fact, if outcross occurs between two plants that display white flower, one recessive for *zt1* and the other for *zt2*, the progeny will have high tannin

content in the seed and coloured flowers. Therefore, during the seed multiplication process, it is important to maintain careful isolation between the low tannin lines expressing the two different genes. For the same reason, low and normal tannin varieties and low tannin varieties with different genes must be cultivated at a distance of no less than 1000 m to avoid outcrossing, as a method for minimizing the production of contaminated seeds.

Our goal is to use molecular techniques to identify genotypes with the double recessive *zt1* and *zt2* genes. We used a segregating population at the F2 created from a cross between *zt1* and *zt2* genotypes. The F1 had coloured flowers and the F2 population was first pre-screened based on the flower colour of the F2 plants, and only those with white flowers were kept. Then a test was done to examine if molecular markers tagging the two recessive genes *zt1* and *zt2* can be effectively used to identify the genotypes recessive at both loci.

4.3 Material and Methods

4.3.1 Plant Material and phenotyping

An F2 population developed at University of Saskatchewan in 2015 from a cross between Disco/2 (*zt2* line) and CDC Snowdrop (*zt1* line) was used for this research. In total, 867 F2 seeds derived from 8 different F1 plants were used for the analysis. The seeds of each F2 family were pre-screened based on their seed coat colour (SCC) and divided into three groups: 1) beige seed coat, 2) light brown seed coat and 3) brown seed coat. Figure 4.1 shows the phenotype of each group of seed coat colour. This prescreening was done to determine if there is any relationship between seed coat colour of F2 (F1-derived) seeds and flower colour of the F2 plants, therefore, to determine if the seed coat colour can be used as an indicator of the allelic state of the genotype at the low tannin loci, and to predict the flower colour.

Figure 4.1. Seed coat colour phenotypes of the F2 seeds. Brown, Light Brown, Beige (from left to right). Seeds of the population were prescreened to verify a possible relationship between seed coat colour and flower colour of the plant growing from the same seed.

The composition of each F1-derived F2 family is presented in Table 4.1 while, Figure 4.2 shows the outline of the F2 population development and the analysis to identify the double recessive low tannin plants.

Table 4.1. Composition of F2 population from Disco/2 x CDC Snowdrop cross for seed coat phenotype. The total number of seeds for each F2 family (A1 to A8) and for each seed coat colour group and their proportions of the total number of F2 individuals produced from each F1 plant are presented.

F2 Families	Beige	Light Brown	Brown	Total
A1	23	21	57	101
A2	30	27	70	127
A3	35	46	40	121
A4	39	31	48	118
A5	40	30	39	109
A6	30	21	51	102
A7	23	32	61	116
A8	44	29	52	125
Total	264	237	418	919
Proportion	0.287	0.257	0.454	

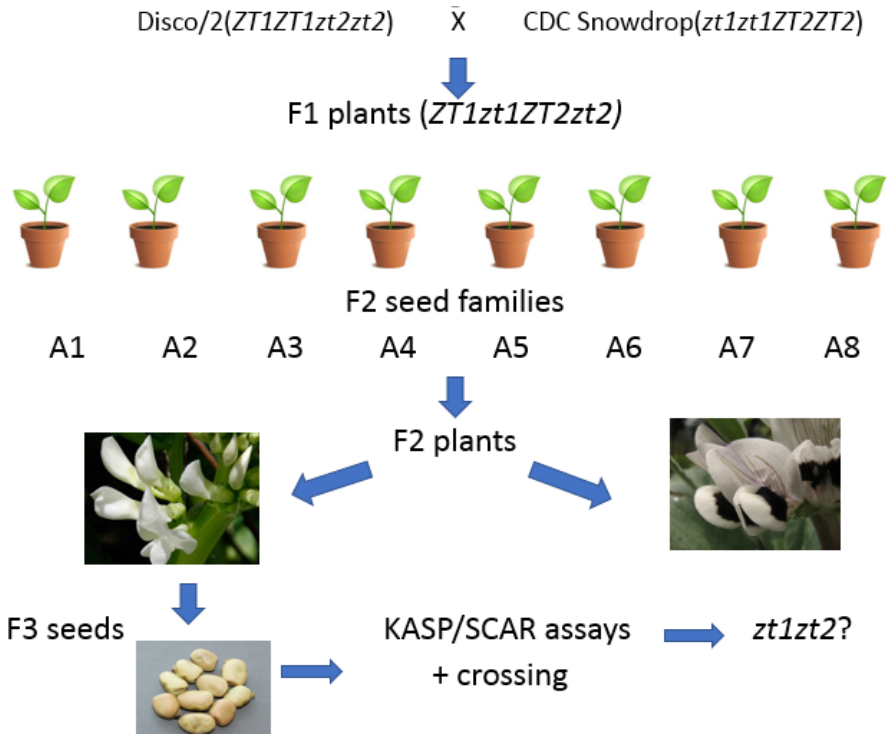


Figure 4.2. Development flow of the segregating population Disco/2 × CDC Snowdrop and subsequent analysis. Only F2 plants with white flowers were grown to maturity. F3 seeds were then genotyped with the new KASP marker for *zt2*. The *zt1zt2* genotypes could be potentially detected screening the *zt2* seeds with the SCAR SCC_{5551} and then crossing these plants with the parental lines for validation.

A validation panel composed of 12 accessions of known allelic state at the low tannin loci was used for the validation of the SCAR markers, together with replicates of the parental lines of the segregating population, Disco/2 and CDC Snowdrop. Accession ILB938/2, which was used in the other research components was also included in this analysis. Table 4.2 presents the information for all the lines used in this study.

The plants were grown in a growth chamber in the College of Agriculture and Bioresources Controlled Environment Facility of the University of Saskatchewan, Canada. Photoperiod was adjusted to 16h day and 8h night, the photon flux was set to 300 µmol m^{-2} s^{-1} and the temperature was 22 °C during the day and 18 °C at night. Plants were grown in 10 cm plastic pots filled with growing mix Sunshine No. 3 (Sun-Gro Horticulture MA, USA), fertilized with NPK

Table 4.2. Characteristics of faba bean accessions used as validation panel for the markers SCC5551 and SCG111171.

Genotype	Tannin genotype[a]	SSP colour[b] (*ssp1* genotype)	Hilum colour	Flower colour	Seed coat colour	Origin / Source
Disco/2	*zt2ZT1*	colourless (*SSP1, ssp2, SSP3*)	colourless	white	beige	INRA, France
CDC Snowdrop	*ZT2zt1*	colourless (*SSP1, SSP2, ssp3*)	colourless	white(*zt1*)	beige	CDC
ILB938/2	*ZT2ZT1*	colourless (*ssp1, SSP2, SSP3*)	black	spotted	green	Ecuadorian landrace, ICARDA
1778(crimson)	*ZT2ZT1*	Spotted	black	red		Sweden
346-10	*ZT2ZT1*	Spotted	colourless	spotted	beige	U of S
L-43	*ZT2ZT1*	Spotted	black	spotted	dark brown	
L-170	*ZT2ZT1*	Spotted	black	spotted	beige	
7070	*ZT2ZT1*	Spotted	spotted	spotted		
P47-1A1	*ZT2ZT1*	Spotted	black	red	green	
1250 bs s2	*ZT2zt1*	Colourless	colourless	white(*zt1*)	beige	
DI43170/540P1	*zt2Zt1*	Colourless	colourless	white(*zt2*)		
Gilber	*ZT2ZT1*	Spotted		yellow		
FBR-06-12	*ZT2ZT1*	Colourless		spotted		
FBR-06-21	*ZT2ZT1*	Colourless		spotted		
SnowdropP77-1	*ZT2zt1*	Colourless		white(*zt1*)	beige	

[a] *ZT2* and *ZT1*: wild-type (normal-tannin) alleles; *zt2* and *zt1*: recessive alleles conferring low tannin content.
[b] *SSP*, Stipule spot pigmentation. *SSP1, SSP2, SSP3* (spotted stipule), *ssp1, ssp2, ssp3*: recessive alleles conferring colourless stipule.

fertilizer (20-20-20) every two weeks and watered manually to keep optimal moisture conditions for plant growth. *Amblyseius cucumeris* predatory mites (Biobest Group NV Westerlo, Belgium) were applied once per week to control the thrips (*Heliothrips haemorrhoidalis*).

Stipule spot pigmentation (SSP) and flower colour (FC) data were visually recorded one week after emergence and at flowering, respectively.

4.3.2 DNA extraction and KASP assay for *zt*2

DNA was extracted from the cotyledon of each seed with the following steps. Briefly: the F3 seeds were scarified using side cutting pliers, then 10µL of distilled water was pipetted onto each exposed cotyledon to moisturize the surface. The surface was then scraped with a toothpick to collect some cotyledon tissue which was then deposited into the well of a 96-well plate PCR micro-plate. Each well was previously filled with 25µL H2O. Cotyledon tissue from each seed was deposited in a different well and each corresponding seed was placed into an Eppendorf tube plate, and then 40µL of 0.25M NaOH was added to each well. Plates were then covered with rubber lids and shaken on the plate-mixer at 1750 rpm for 30 s. After this, the plates were heated at 94°C for about 2 min, then rubber lids were removed and 60µL of 0.5M Tris-HCL (pH 8.0) was added to each well. The plates were again covered, and then shaken on plate mixer for 10 s at 1500 rpm. The plates were heated once again at 94°C for 2 min to denature the DNA and then placed on ice.

For the KASP assay 40µL of distilled water was pipetted into new 96 well plates, 10µL of stock DNA was pipetted into these wells for a 1:5 dilution.

2µL of the diluted DNA was then used in a 10 µL KASP reaction using a BioRad CFX384™ Real Time System (Bio-Rad Laboratories, Inc. Hercules, California, US) at the Pulse Crop

Genotyping Lab (Department of Plant Sciences University of Saskatchewan, Canada). Details about the *zt2* KASP marker used in this assay are described in Chapter 3.

4.3.3 *zt*1 SCAR assay

DNA was extracted from fresh leaf tissue collected from the plants one week after emergence using a Qiagen DNeasy Plant DNA Extraction Kit (Quiagen corporate, Venlo, Netherlands). The primers for SCAR markers SCG_{111171} and SCC_{5551} (Gutierrez et al. 2006) were obtained from IDT Integrated DNA Technologies (Integrated DNA Technologies, Inc.). The primers sequences are SCC_{5551}Fw. 5'-GATGACCGCCCATATAAGAGA-3' SCC5551Rev. 5'-GATGACCGCCGGGCAAAC3-3', SCG111171Fw. 5'-TGCCCGTCGTGACAGAAATA-3', SCG111171Rev. 5'-TGCCCGTCGTAGTAGGATA-3'.

SCAR-PCR amplifications were performed in a BioRad C1000™ thermal Cycler (Bio-Rad Laboratories, Inc, Hercules, California, US) at the Pulse Breeding Lab (Department of Plant Sciences, University of Saskatchewan, Canada) following the two protocols described by Gutierrez et al. (2006). Protocol one was carried out in a 25 µl reaction volume. Each PCR reaction contained 40–50 ng of genomic DNA, buffer (50 mM KCl, 75 mM Tris–HCl, 20 mM (NH4)2SO4), 1.5 mM MgCl2, 0.8 µM of each dNTP, 0.3 µM each specific primer and 1 unit of GenScrip® Taq polymerase (GenScript, Court Burlington, Ontario, CA). The thermal profile for PCR was an initial denaturation at 93 °C for 5 min followed by 35 cycles of 45 s denaturation at 93°C, 25 s annealing at 59°C, 2 min of extension at 72°C with a final extension of 10 min at 72°C. Protocol two consisted of a touchdown-PCR procedure to ensure specificity of the amplified product. The thermal profile varied from the previous PCR program in that the annealing temperature used in the first cycle was 70°C. The temperature was then dropped by 2°C per cycle until an optimal

annealing temperature of 60°C was reached. The remaining 25 cycles were performed at 60°C for 45 s. The reaction mixture was described previously, except that the concentration of the MgCl2 was adjusted to 1 mM. PCR products were then resolved in a 1.5% agarose gel for one hour at 100V, and then stained using Gelred™ and photographed using a Gel doc system™ imager (Alpha Innotec, Kasendorf, Germany).

4.3.4 Statistical analysis

The observed segregation ratios for the FC and SSP were tested against the expected ratios using the χ^2 test for goodness of fit. SAS software was used to perform the calculations SAS 9.4 (Copyright © 2002-2012 SAS Institute Inc., Cary, NC, USA). The independence test for verifying the relationship between SSC and FC was calculated using SAS Proc Freq.

4.4 Results and Discussion

4.4.1 Segregation ratios in the F2 population

Table 4.3 shows the results of the test of independence to verify if there is any relationship between the SCC of the (F1-derived) F2 seeds and the FC of the F2 plants grown from these seeds. The test was done considering either three or two seed coat colours. In the first case the three groups considered were beige, light brown, brown, while in the second case the light brown seeds were first grouped with the beige group and then with the brown group. This was done to account for the possible error of discriminating the light brown SCC as a separate category, since this intermediate SCC group was not clearly distinguishable from the beige and brown groups on a pure visual assessment. In all cases the χ^2 goodness of fit test suggested that the two characters are not linked by any relationship. *P* values > 0.05 shows that SCC cannot be used as a predictor of the flower colour of the plant since the seed coat and the plant grown from

the same seed have different genotypes. The seed coat of the F2 seed is maternally inherited, so it has the same genotype of the mother plant (F1 plant), while the cotyledons is an F2 tissue. All the F1 plants, hence the seed coat of the F2 seeds, should have the same genotype (heterozygous at all loci), *ZT1zt1Zt2zt2* when considering the *zt* loci. Therefore, factors other than genotype play a role in the differentiation of the seed coat colour which is observed in the F2 seeds. Possibly this is due to seed aging, Nasar-Abbas et al. (2009) reported that the SCC of faba bean changed from beige (initial colour) to brown and dark reddish-brown after 12 months storage and they found that increased temperature and longer storage time accelerated SCC darkening. Mirali et al. (2016) described similar results in lentil, with a progressive darkening of the green SCC depending on the storage conditions and duration.

Table 4.3. Results of the independence test between seed coat colour (F1-derived F2 seeds) and flower colour (F2 plants). *P* values >0.05 indicates there is no relationship between the two characters. The test was performed considering either three (beige, light brown, brown) or two (beige, brown) seed coat colours, row one and rows 2 and 3, respectively.

Category	Seed coat colour	χ^2	*P* value	*DF*	Relationship
1	beige-light brown-brown	0.5505	0.7594	2	no
2	beige (light brown /beige)-brown[a]	0.5457	0.4601	1	no
3	beige-brown (brown/ light brown) [b]	0.2350	0.6278	1	no

[a] ligh brown considered as beige

[b] ligh brown seeds considered as brown

At the F2, the ratio between spotted and white flowers was 9:7 as we expected for the segregation of two independent recessive genes (Table 4.4).

After flowering the plants showing spotted flowers were discarded. Only those with white flowers were kept until maturity. The seeds were harvested from each plant separately. A total of 334 plants with white flowers produced seeds.

Table 4.4. Goodness of fit of the observed segregation ratios for flower colour (FC) and stipule spot pigmentation (SSP) to the expected 9:7 ratio. The χ^2 test and the corresponding *P* values are given for the entire population and for the sub populations (F2 families).

F2 Family	FC Spotted	FC White	χ^2 (9:7)	*P* value	SSP Spotted	SSP Colourless	χ^2 (9:7)2	*P* value
Total	451	382	1.504	0.22	423	422	13.6679	0.0002**
A1	45	50	3.0451	0.081	43	57	7.134	0.0076**
A2	65	60	0.9175	0.3381	57	68	5.7611	0.0164*
A3	71	48	0.5636	0.4528	68	53	0.0001	0.9909
A4	64	50	0.0006	0.9812	60	55	0.7764	0.3782
A5	52	51	1.3908	0.2383	49	56	3.9185	0.0478*
A6	47	52	3.0978	0.0784	44	56	6.0978	0.0135*
A7	72	39	3.3475	0.0673	70	42	1.7778	0.1824
A8	35	32	0.438	0.5081	32	35	1.9619	0.1613

(*observed value is significantly different from expected value at *P*<0.05, ** observed value is strongly significantly different from expected value at P<0.01).

A single F3 seed randomly chosen from each F3 group of seeds (seeds collected from single F2 plant) was used for the KASP assay, performed with the newly developed KASP marker for the *zt2* gene. In this analysis 327 seeds were used, and seeds from seven plants were discarded because they did not reach optimal maturity. Control seeds from each of the two parental lines and negative controls were also included in the analysis.

The seeds discriminated as recessive for *zt1* by the KASP assay were then screened with two SCAR markers developed by Gutierrez et al. (2006). These seeds were grown until flowering, together with seeds of the validation panel and parental lines.

4.4.2 KASP and SCAR assays

The results of the KASP assay are shown in Figure 4.3. The analysis confirms the efficacy of the new *zt2* KASP marker in discriminating between genotypes homozygous for *zt2* (blue dots in the KASP plot) from those homozygous for *zt1*. This last group comprises both the genotypes that are homozygous dominant for allele b (yellow dots in the KASP plot), therefore dominant at the *ZT*2 locus (*ZT2ZT2*) and those that are heterozygous at the *ZT*2 locus (*ZT2zt2*) showed as green

dots in the KASP plot. The last group of genotypes, which were generated from a white flower parent are recessive for *zt1*. In fact, at the F2 at least one of the two genes must be recessive to have the white flower characteristic, and this allelic state will not change in the subsequent generations in a self-pollination mating system. Therefore, their allelic state at both loci at the F3 is *ZT2zt2 zt1zt1*. All the samples for the parental lines were correctly discriminated for their allelic state at the *zt2* locus, Disco/2 being *zt2* and CDC Snowdrop being *ZT2*.

Out of the total 327 F3 seeds, 257 were discriminated as *zt2zt2*, 44 as *zt1zt1* and 26 as heterozygous. The 26 heterozygous individuals (*ZT2zt2*) are counted as *zt1* for the reason explained above. Therefore, the total number of *zt1* individuals was 70.

It should be noticed that the ratio between *zt2* and *zt1* genotypes, as discriminated by the KASP assay, exceeds the expected ratio of 3:2 at the F3 generation (ratio that considers only the progeny of plants with white flower at F2 because those with spotted flower were discarded). This ratio takes into consideration that the *zt2* group also encompasses the double recessive genotypes *zt1zt2* which cannot be distinguished from the genotypes that are recessive only for *zt2* by the KASP assay, since both groups are homozygous for allele 1. However, the single F3 seeds produced by single F2 plants, used for the KASP assay, were chosen randomly, this being the most likely explanation for the higher than expected occurrence of genotypes recessive for *zt2*.

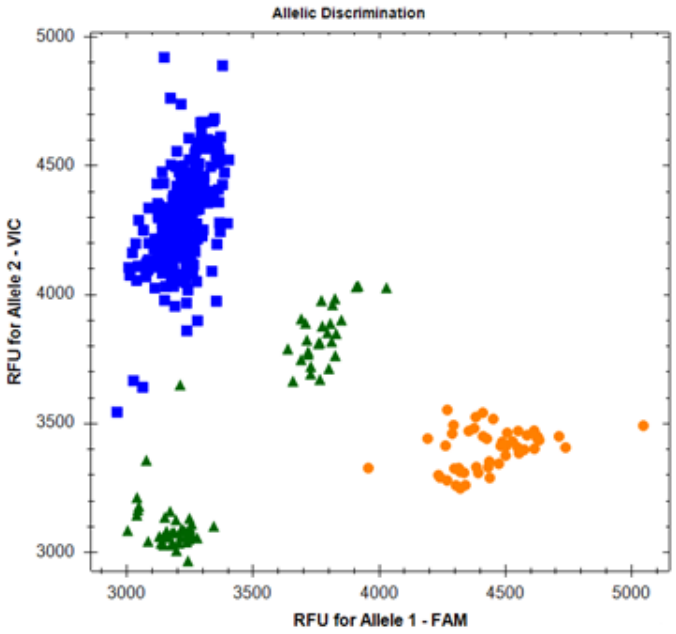


Figure 4.3. Results of the KASP assay on F3 seeds Disco/2×CDC Snowdrop. Blue dots represent genotypes homozygous for allele 2 (*zt2*), while yellow dots are the genotypes homozygous for allele 1 (*ZT2*). The green dots in the center of the figure are the genotypes heterozygous for the *zt2* gene (*zt2ZT2*), while the green dots in the bottom left are the negative controls.

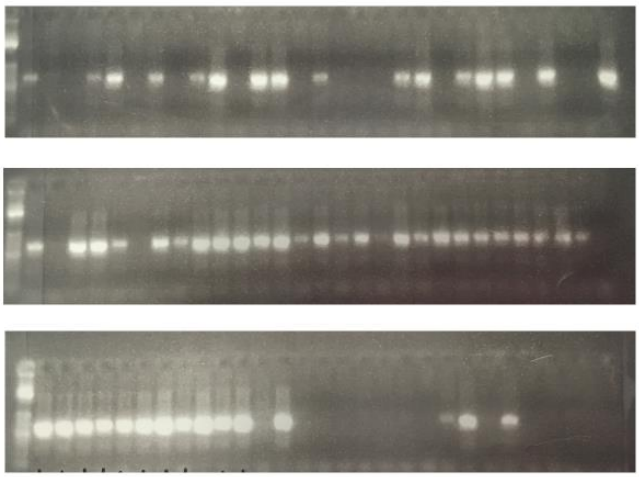
The SCAR marker SCC_{5551}, tested on the *zt1* seeds discriminated by the KASP assay, and on seeds of the validation panel resulted in the 77% and 68% prediction accuracy on the two groups, respectively. This was lower than what was reported by Gutierrez et al. (2006) of 95% accuracy of the same SCAR marker in detecting *zt1* genotypes from those dominant at the *ZT1* locus. All the *zt1* genotypes showed the expected white flower/uncoloured stipule phenotype. This was to be expected, since no plants with white flowers and spotted stipule were detected at the F2, therefore, at least one of the recessive genes was present for both characters and consequently inherited by the F3 plants because of the self-pollination at F2.

The results for the SCG_{111171} SCAR marker showed 54% and 46% prediction accuracy for *zt1* in the KASP discriminated seeds and validation panel, respectively. Therefore, we confirmed

the much lower reliability of SCG_{111171} compared to SCC_{5551} as already reported by Gutierrez et al. (2006).

Figure 4.4 shows the Agarose gels with the segregation of SCARs SCC_{5551} (a) and SCG_{111171} (b) in the *zt1* plants, parental lines and genotypes of validation panel.

a)



b)

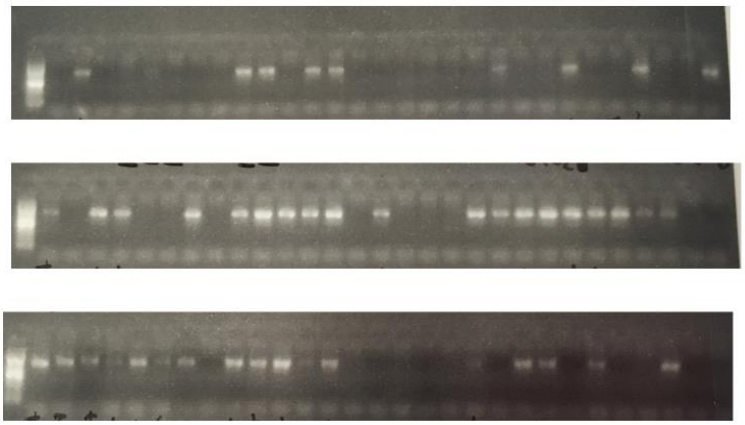


Figure 4.4. Agarose gels showing the segregation of SCARs SCC_{5551} (a) and SCG_{111171} (b). Bright bands indicate presence of the PCR products.

4.4.3 Future goal

As a follow-up step towards the identification of the double recessive genotype *zt1zt2* we propose to use the SCAR marker SCC_{5551} to detect the genotypes which carry *zt1* in the recessive form among the group of genotypes discriminated as *zt2* by the KASP assay. As explained above, this group should include the double recessive genotypes *zt1zt2.* This will require growing the plants and extracting DNA from the leaf tissue, since the quality of DNA from the cotyledons is not of sufficient quality for performing the SCAR assay. Finally, the F3 plants which are positively discriminated as *zt1* would have to be crossed with both parental lines, Disco/2 which carries *zt2* and CDC Snowdrop which carries *zt1.* If both crosses will result in F4 plants with white flowers, the F3 progenitor is truly the double recessive genotype.

4.5 Conclusions

This study implemented molecular techniques for detecting the double recessive low tannin genotype in a faba bean segregating population for the *zt1* and *zt2* genes. It also proved the superiority of SNP markers in the KASP assay compared to the old generation of markers such as the SCARs for screening the plants carrying the two different low tannin genes. The KASP marker system is much faster, more accurate and is less affected by low quality DNA for the analysis. Therefore, the development of a KASP marker for *zt1* would be a desirable tool for breeding low tannin faba bean.

CHAPTER 5- BIOCHEMICAL ANALYSIS OF THE LOW TANNIN CHARACTERISTIC IN FABA BEAN (*VICIA FABA L.*)

5.1 Abstract

Low tannin faba bean genotypes are distinguished from normal tannin genotypes by the absence of pigmentation in the flowers and reduced content of condensed tannins in the seed coats. The phenotype is controlled by either of the two recessive genes *zt1* and *zt2*. In this study we used liquid chromatography mass spectrometry (LC-MS) to study the phenolic profiles of the seed coats and flowers of 3 faba bean genotypes with different allelic state at *zt1* and *zt2;* ILB 938/2 with (dominant alleles for both genes), Disco/2 and CDC Snowdrop, *zt2* and *zt1* genotypes respectively. The results indicated clear differences, as expected, in the phenolic profiles of ILB 938/2 from those in the low tannin lines in both flowers and seed coats. Most importantly, distinct differences were also identified between the phenolic profiles of *zt1* and *zt2* genotypes, particularly evident in the flowers. These results proved that the two low tannin genes have different modes of action in the phenolic compounds biosynthetic pathway.

5.2 Introduction

Tannin is a general term to describe a vast group of polyphenols present in soluble extracts of plants. Their interaction with proteins and other biological macromolecules, especially their interference with enzymatic degradative processes, is one of the main reasons why they have been studied.

Based on their chemical structures, tannins can be divided in two different groups, condensed and hydrolysable tannins (Mole and Waterman, 1987). Condensed tannins (*proanthocyanidins*) have been studied in more detail than any other phenolic compounds in faba bean, because of their interference with protein digestion in monogastric mammals. They are

polymers derived from *flavan-3-ols* units that are synthesized through the *flavonoids* pathway, which is a branch of the more complex phenolic compounds biosynthetic pathway (He et al. 2008).

Condensed tannins, however, represent just a small group of secondary metabolites. To date not much research has been done to further characterize and differentiate the phenolic compounds present in the seed coat of faba bean or to characterize their genetic control.

LC-MS (Liquid Chromatography-Mass Spectrometry) was used by Mirali et al. (2014, 2016) to characterize the phenolic profiles of the seed coat of lentils lines differing in seed coat colour, based on the phenotypes resulting from the effects of two major genes. In particular, the study used Single Reaction Monitoring (SRM), to target specific phenolic compounds. In this method, each of the considered phenolic compounds was quantified relative to a related internal standard (IS).

The same technique was applied using an updated method (Elessawy et al. manuscript in preparation) to quantify and characterize the phenolic compounds present in the seed coat and flowers of three faba bean lines, a wild type and two low tannin lines, recessive for *zt1* and *zt2,* respectively. The purpose of this analysis was to better understand and to potentially determine the point of action of the two genes *zt1* and *zt2* in the phenolic biosynthetic pathway. Furthermore, the connection of these results with those at the molecular level found in chapter 3, indicated possible candidate genes for *zt2*.

5.3 Material and methods

The method for identification and quantification of phenolic compounds was developed and optimized by Mirali et al. (2014, 2016) and further expanded by Elessawy et al. (manuscript

in preparation). The method employs liquid chromatography (LC) combined with selective reaction monitoring (SRM) mass spectrometry on a triple quadrupole mass spectrometer. SRM isolates the parent ion of a selected phenolic compound in the first quadrupole of the mass spectrometer, the second quadrupole is used as a collision cell to fragment the ion, and the third to detect a specific intense product ion. Henceforth, this method is referred to as the targeted method. A second method based on accurate mass was also used for identification purposes. The method uses the same LC gradient and column, however, it uses a Q-Exactive mass spectrometer to accurately (to four decimal places) determine the m/z values of compounds and their product ions (the triple quadrupole is only accurate to 1 decimal place). With accurate mass data, software can be used to propose the chemical composition of unknowns and, in some cases, putatively identify unknowns. This method is referred to as the untargeted method.

5.3.1 Plant materials

Biochemistry studies of the seed coats and flowers involved three faba bean genotypes. ILB 938/2 is an Ecuadorian genotype with dominant alleles at both low tannin loci, with spotted flowers and condensed tannins in the seed coat. Disco/2 is a French cultivar which is recessive for *zt2* with white flowers and low condensed tannin content in the seed coat. The third genotype is CDC Snowdrop, a low tannin cultivar developed at the Crop Development Centre, University of Saskatchewan. It is recessive for *zt1* and, like Disco/2, has white flowers and low condensed tannin content in the seed coat.

Disco/2 and ILB 938/2 were used as parents for the F6 RIL population (FBR-8) used to map the *zt2* gene described in chapter 3.

All seeds and flowers used in this study were harvested from plants grown in a single growth chamber in the College of Agriculture and Bioresources Controlled Environment Facility of the University of Saskatchewan, Canada. Photoperiod was adjusted to 16h day and 8h night, the temperature was 22 °C during the day and 18 °C at night and the photon flux was set to 300 µmol m^{-2} s^{-1}.

5.3.2 Sample preparation

The seed coat was manually separated from the cotyledons of each seed with the help of a small wire cutter. The seed coat tissue of 10 seeds in total, harvested from three different single plants (3-4 seeds per plant) of the same accession were pooled. Using 100 mg of seed coat each, four different analytical replicates for each accession were accurately weighed and filled into Sarstedt 2mL microtubes. The samples were then freeze-dried for 12 h after being kept at -80 °C for one hour. Two ¼ inch ceramic spheres (MP Bio, cat. No. 6540-412) were added to each tube and the samples were then pulverized using a Mini-Beadbeater-16 (BioSpec Products, Inc. Bartlesville, OK, USA) for two repetitions lasting one min each. 1000 µL of the extraction solvent was then added to each sample. The extraction solvent was composed of acetone: water (70:30 v/v) containing the internal standards (IS), which were added to account for changes in the matrix among the genotypes and to enable quantification.

Samples were then shaken for one hour on an Eppendorf mixer at a speed of 1400 rpm at 23 °C. The tubes were centrifuged at 13,000 rpm for 15 min. An aliquot of 200 µL of supernatant was dried down with a Speed Vac (LABCONCO, Kansas City, USA). Dried samples were then reconstituted in 200 µL 90:10 water: methanol. Blank samples were prepared in a similar way using two empty Sarstedt 2 mL microtubes.

The flower samples were prepared following the same protocol as described for the seed coats, using fresh flowers harvested from blooming plants. Four different replicates for each accession were prepared using 100 mg of freeze dried flower tissue for each sample. The flowers were harvested from different plants of the same accession (approximately 10 flowers from each plant) and then pooled together (within each accession).

5.3.3 Liquid Chromatography

The following chromatographic conditions were applied for the analysis of phenolic compounds. For the targeted method, the LC hardware was an Agilent 1290 UPLC equipped with a G4226A autosampler, a G4220A binary pump, a G1316 TCC, and a G4212 DAD detector. For the untargeted method, a Dionex 3000 LC was used. For both methods, the column was an InfinityLab poroshell 120 pentafluorophenyl (PFP) (100 mm × 2.1 mm id), with 2.7 µm particle size (Agilent, Santa Clara, CA, USA). The mobile phases were water:formic acid (FA) (99:1, v/v) as solvent A and water:acetonitrile (ACN):FA (9:90:1, v/v/v), as solvent B. The mobile phase was delivered at a flow rate of 0.35 mL/min. The gradient used is described in Mirali et al. (2016) as shown in Table 5.1. The injection volume for each sample was 5 µL.

Table 5.1. Mobile phase gradient used in this experiment, where solvent A and B were water: formic acid (99;1, v/v) and water: acetonitrile: formic acid (9:90:1, v/v/v), respectively.

Time (min)	A%	B%	Flow (mL/min)
0	99	1	0.35
1	99	1	0.35
21	59	41	0.35
24	40	60	0.35
24.1	20	80	0.35
26	20	80	0.35
26.1	99	1	0.35
30	99	1	0.35

5.3.4 The targeted SRM method

Table 5.2 presents the complete list of phenolic compounds analyzed in this study, with their retention time, optimized molecular and fragment ions, and collision energy. The method, initially developed by Mirali et al. (2014), was expanded and optimized to include different pulse crops (Elessawy et al., manuscript in preparation), and implemented in faba bean in this study. All the compounds are assigned to the following phenolic classes, *anthocyanins*, *chalcones*, *flavonols*, *flavones*, *flavan-3-ols*, *flavanones*, *flavanonols*, *dihydroflavonols*, *stilbenes*, *hydroxybenzoic acids, hydroxycinnamic acids, isoflavones* and *proanthocyanidins*. The standards were purchased from the companies listed in Supplementary Table 5.1. The two compounds with an asterisk (*) in Table 5.2 (catechin-3-glucoside and kaempferol dirutinoside) were included even though the standards were not commercially available because they were previously identified as major components in lentil seed coats (Mirali et al. 2016).

Peak areas of each analyte were integrated with Thermo Xcalibur 2.2 software (Thermo Fisher Scientific, Mississauga, Ontario, CA) and were normalized to the peak area of a related internal standard (IS). The values were compared with a calibration curve for the analyte and reported per g of dry sample. The analysis was done at the Core Mass Spectrometry Facility (CMFS) at the UofS, Saskatoon, CA.

Table 5.2. SRM parameters and retention times of the polyphenolic compounds analyzed in this study**.** In bold = internal standards; * =compounds without standard but previously identified, In *Italics*= compounds classes.

Compound name	Polarity	Mol. Ion (m/z)	Fragment ion (m/z)	Collision energy	Retention time
Anthocyanins					
Malvidin-3-O-glucoside (oenin)	NEG	491	313	35	7.71
Pelargonin	NEG	593	431	16	5.65
Pelargonidin 3-O-glucoside	NEG	431	269	21	6.52
Peonidin-3,5-di-O-glucoside	NEG	623	299	34	6.23
Peonidin-3-O-glucoside	NEG	461	299	23	7.14

Cyanidin-3,5-di-O-glucoside	NEG	609	447	21	5.12
Cyanidin-3-O-rhamnoside	NEG	431	285	21	7.30
Cyanidin 3-O-glucoside	NEG	447	284	26	6.06
Delphinidin-3,5-di-O-glucoside	NEG	625	299	44	4.53
Delphinidin-3-O-rhamnoside	NEG	447	300	26	6.56
Delphinidin-3-β-D-Glucoside	NEG	463	300	25	5.57
Malividin-3,5-di-O-glucoside	NEG	653	329	31	6.75
Malvidin-3-O-galactoside	NEG	491	313	35	7.34
Chalcones					
Phloretin	NEG	273	167	18	19.10
Xanthohumol	POS	355	179	23	25.43
4-Hydroxy-6-methylcoumarin	NEG	175	131	20	14.86
Flavonols					
Kaempferol-3-O-rhamnoside	NEG	431	285	21	14.60
Fisetin	NEG	285	163	18	14.84
Isorhamnetin	NEG	315	300	22	19.23
Kaempferol	NEG	285	187	30	19.00
Kaempferol 3-O-robinoside-7-O-rhamnoside	NEG	739	593	28	11.06
Kaempferol 3-O-rutinoside-4'-glucoside	NEG	755	593	23	10.53
Kaempferol-3-O-D-galactoside	NEG	447	255	40	12.90
Kaempferol-3-O-glucoside	NEG	447	285	25	13.17
Kaempferol-3-O-rutinoside	NEG	593	285	33	12.87
Kaempferol-7-O-glucoside	NEG	447	285	25	13.50
Kaempferol-7-O-neohesperidoside	NEG	593	285	33	13.43
Kaempferol-di-rutinoside*	NEG	901	755	30	9.18
Tiliroside	NEG	593	285	35	17.17
Myricetin	NEG	317	151	24	14.87
Myricetin-3-O-rhamnoside	NEG	463	316	27	12.12
Quercetin	NEG	301	151	22	17.05
Quercetin-3,4'-di-O-glucoside	NEG	625	463	22	10.72
Quercetin-3-O-galactoside	NEG	463	300	27	12.25
Quercetin-3-O-glucoside (Isoquercetrin)	NEG	463	300	27	12.33
Quercetin-3-O-rhamnoside (Quercitrin)	NEG	447	300	27	13.45
Quercetin-3-O-rutinoside(Rutin)	NEG	609	300	36	11.95
Quercetin-4′-O-glucoside (Spiraeoside)	NEG	463	301	20	14.35
Quercetin-d3 (IS)	NEG	304	151	22	17.03
Flavones					
5,7-Dimethoxyflavone	POS	283	239	31	21.62
Apigenin	NEG	269	117	39	19.11
Apigenin-7-O-glucoside	NEG	431	268	34	14.00
Apigenin-7-O-neohesperidoside	NEG	577	269	33	13.74
Apigenin-7-O-rutinoside	NEG	577	269	33	13.51
Vitexin	NEG	431	311	25	11.43
Vitexin-2'-O-rhamnoside	NEG	577	293	42	11.15
Chrysin	POS	255	153	34	21.70
Diosmetin	NEG	299	284	23	19.29
Diosmetin-7-O-rutinoside	NEG	607	299	28	13.82
Flavone	POS	223	77	36	20.85
Luteolin	NEG	285	133	35	17.39
Luteolin-3',7-di-O-glucoside	NEG	609	285	41	11.55
Luteolin-4'-O-glucoside	NEG	447	285	25	14.68
Luteolin-7-O-glucoside	NEG	447	285	25	12.89
Luteolin-7-O-rutinoside	NEG	593	285	33	12.49
Luteolin-8'-O-glucoside	NEG	447	327	22	10.79
Tangeretin	POS	373	343	31	22.12
Flavan-3-ols					
(+)-Catechin	NEG	289	203	21	7.55
(-)-catechin gallate	NEG	441	289	19	12.64

catechin-glucoside*	NEG	451	137	21	6.70
(-)-Epicatechin	NEG	289	203	21	8.77
(-)-Epicatechin gallate	NEG	441	169	22	12.02
(-)-Epigallocatechin	NEG	305	125	24	7.27
(-)-Epigallocatechin gallate	NEG	457	169	18	10.20
(-)-Gallocatechin	NEG	305	125	24	5.63
±-Catechin-2,3,4-$^{13}C_3$ (IS)	NEG	292	206	21	7.55
Flavanones					
Eriocitrin	NEG	595	287	27	11.35
Eriodictyol	NEG	287	151	16	15.98
Flavanone	POS	225	121	21	22.36
Hesperetin	NEG	301	164	25	18.34
Hesperetin-7-O-rutinoside	NEG	609	301	27	13.07
Isosakuranetin	POS	287	153	26	21.86
Narigenin-7-O-rutinoside	NEG	579	271	25	12.47
Naringenin	NEG	271	151	19	17.88
Dihydroflavonols					
Dihydromyricetin	NEG	319	193	15	9.47
Dihydrokaempferol	NEG	287	125	21	13.14
Taxifolin (dihydroquercetin)	NEG	303	125	23	11.52
Stilbenes					
Resveratrol	NEG	227	143	25	14.94
Resveratrol-3-β-mono-D-glucoside	NEG	389	227	21	11.67
Resveratrol-(4-hydroxyphenyl-$^{13}C_6$) (IS)	NEG	233	149	27	14.93
Hydroxybenzoic acids					
3,4-Dihydroxybenzoic acid	NEG	153	109	16	4.42
4-amino salicylic acid	NEG	152	108	16	4.37
4-hydroxybenzoic acid	NEG	137	93	16	5.74
Gallic acid	NEG	169	125	16	2.86
Salicin	NEG	285	123	15	3.56
Syringic acid	NEG	197	182	15	8.19
Vanillic acid	NEG	167	108	20	7.13
Vanillic acid-4-β-D-glucoside	NEG	329	167	15	4.54
Vanillin	NEG	151	136	15	8.25
Vanillin-(ring-$^{13}C_6$) (IS)	NEG	157	142	15	8.22
4-hydroxybenzoic acid - $^{13}C_7$ (IS)	NEG	144	99	18	5.75
Hydroxycinnamic acids					
Caffeic acid	NEG	179	135	17	7.86
Chlorogenic acid	NEG	353	191	20	8.03
Ferulic acid (trans)	NEG	193	134	18	10.61
p-Coumaric acid (trans)	NEG	163	119	17	9.48
trans-3-hydroxycinnamic acid	NEG	163	119	17	10.38
Ferulic acid-d3 (IS)	NEG	196	134	19	10.61
Isoflavones					
Genistein	NEG	269	133	31	18.13
Prunetin	POS	285	241	44	22.43
Proanthocyanidins					
Procyanidin A2	NEG	575	285	27	11.76
Procyanidin B1	NEG	577	289	26	6.79
Procyanidin B2	NEG	577	289	26	8.05
Procyanidin B3	NEG	577	289	26	7.29
Procyanidin C1	NEG	865	407	39	9.53

5.3.5 The untargeted method

To identify unknown compounds, higher resolution, accurate mass data were needed that could not be obtained from the triple quadrupole mass spectrometer at the CMSF at the University of Saskatchewan. Therefore, this analysis was done in collaboration with the Canadian Food Inspection Agency (CFIA) at the Western Laboratory Network (Saskatoon, Saskatchewan, CAN).

A Q-Exactive mass spectrometer (Thermo Fisher Scientific, San Jose, CA, USA), was used to acquire full scan data using a mass resolution of 140,000 (@ m/z 200) in both positive and negative mode using a mass range of 140-2100 m/z. A QC (quality control) sample, which contains equal amounts of all samples, was injected every 8 runs to enable relative quantification. Conversely, identification samples, which contain all the samples within a genotype (one each for ILB 938/2, Disco/2 and CDC Snowdrop), were injected at the start and end of each acquisition.

Both full scan data and data dependent MS/MS spectra (DDMS) were acquired on the identification samples to help to identify them. To enable sufficient data points across an LC peak, the mass resolution of the full scan analysis was reduced to 70,000 (@ m/z 200) so that MS/MS could be carried out on the top 7 peaks from each scan using a stepped collision energy (15, 35, 55 eV). The MS/MS analysis used an exclusion list (m/z values) of known impurities and a dynamic exclusion time of 3 s.

For the analysis of the data, untargeted workflows were developed by adapting existing workflows from Compound Discoverer 2.1 software (Thermo Fisher Scientific, Mississauga, Ontario, CAN). The accurate mass data from the full scans restricts the number of possible molecular formula for each m/z value. When used along with the isotopic distribution in full scan

mode, for ions with m/z < 500, typically only a small number (e.g., 1-3) of chemical formulas are possible. Note that as the m/z increases, the number of possibilities typically also increases. Since numerous structural isomers can exist, MS/MS spectra are useful to help narrow down the possible structures. The unknown MS/MS spectra obtained from the analysis were compared with libraries of MS/MS spectra of standards and MS/MS spectra of known unknowns (compounds that have been putatively identified but for which no commercial standard exists). The latter comprised mainly compounds in the class of *proanthocyanidins*. Several libraries were used for the analysis including libraries developed by Dr. Purves, at the Core Mass Spectrometry Facility at the University of Saskatchewan (Purves, R. W. manuscript in preparation), libraries proprietary to the Thermo software, and also libraries available in public data bases and literature. These libraries were used to find matches with previously reported compounds for which MS/MS spectra were characterized.

For the detection of unknown compounds, we analyzed only the data from the flower tissue, and we focused on the differences between Disco/2 and CDC Snowdrop and between ILB 938/2 and Disco/2. These two specific comparisons were done to detect potential new compounds which can help to differentiate the two low tannin lines and the parents of the mapping population used for mapping the *zt2* gene (Disco/2 and ILB 938/2), respectively. The flower tissue was chosen primarily because of the novelty of this approach, but also to test the utility of this LC-MS workflow in the detection of unknown compounds.

5.3.6 Statistical analysis

Statistical analysis was done using SAS software Version 9.4 (Copyright © 2002-2012 SAS Institute Inc., Cary, NC, USA). Analysis of variance was used to determine significant

differences among genotypes for the values of μg/g of sample for each of the phenolic compounds. Tukey's test was used for comparing the values of μg/g samples (95 % confidence level).

5.4 Results and Discussion

5.4.1 Results for known compounds (targeted method)

5.4.1.1 Seed coat

Within the different classes of polyphenols, some compounds were not identified in the seed coats of any of the three faba bean genotypes.

From the *anthocyanins* class only cyanidin-3,5-di-O-glucoside and delphinidin-3,5-di-O-glucoside were identified. For the *flavones*, 5,7-dimethoxyflavone, apigenin, apigenin-7-O-glucoside, apigenin-7-O-neohesperidoside, apigenin-7-O-rutinoside, diosmetin, diosmetin-7-O-rutinoside, flavone and vitexin-2'-O-rhamnoside were not identified in any of the three genotypes. Among the *flavonols,* kaempferol-3-O-D-galactoside, kaempferol-di-rutinoside and tiliroside were also not detected. For the *flavan-3-ols* this was the case for (-)-catechin gallate and (-)-epicatechin gallate, whereas flavanone, hesperetin, isosakuranetin and narigenin-7-O-rutinoside were not detected from the *flavanones*. For the *hydroxybenzoic acids*, 4-amino salicylic acid, syringic acid, vanillic acid were not detected, whereas chlorogenic acid and p-coumaric acid (trans) were absent from the *hydroxycinnamic acids*. Finally, none of the *isoflavones* and *chalcones* were present. For the *dihydroflavonols, stilbenes* and *proanthocyanidins* all the compounds were present in at least one genotype.

Table 5.3 lists all the compounds identified in the seed coat together with the results of Tukey's test among comparisons of the three genotypes ILB 938/2, Disco/2 and CDC Snowdrop.

Table 5.3. Concentration of phenolic compounds in µg/g of sample detected in the seed coat of three faba bean genotypes ILB 938/2, Disco/2 and CDC Snowdrop. The values are means of 4 samples. Tukey's test was used for the multi treatment comparisons. Means of each compound with different letters are significantly different at p <0.05. SEM= standard error of the mean. The calculated P values are presented. Highlighted in green are the compounds which have significantly different concentrations between wild type and *zt* lines. Highlighted in blue are the compounds with significantly different concentration between the two *zt* lines.

COMPOUND NAME	Concentration (µg/g)			SEM	P Value
	ILB938/2 *ZT1ZT2*	DISCO/2 *ZT1zt2*	CDC SNOWDROP *zt1ZT2*		
Cyanidin-3,5-di-O-glucoside	4.9000 a	0.1125 b	0.06500 b	0.3009	<.0001
Delphinidin-3,5-di-O-glucoside	0.09250 a	0.09000 a	0.0925 a	0.0020	0.6224
Dihydromyricetin	3.505 a	1.915 b	0.7975 c	0.1387	<.0001
Dihydrokaempferol	0.02605 a	0.01162 b	0.0133 b	0.0024	0.0040
Taxifolin (dihydroquercetin)	0.2691 a	1.2367 b	0.1985 b	0.0509	<.0001
Vitexin	0.08195 a	0.0039 b	0.0037 b	0.0024	<.0001
Chrysin	0.01045 a	0.007125 a	0.005425 a	0.0014	0.0750
Luteolin	0.0952 a	0.02208 b	0.06345 c	0.0037	<.0001
Luteolin-3',7-di-O-glucoside	0.01838 a	0.01353 a	0.0119 a	0.0018	0.0702
Luteolin-4'-O-glucoside	0.1172 a	0.001425 b	0.002525 b	0.0014	<.0001
Luteolin-7-O-glucoside	0.01255 a	0.003 b	0.002925 b	0.0006	<.0001
Luteolin-7-O-rutinoside	0.009475 a	0.036 b	0.02922 b	0.0034	0.0010
Luteolin-8-C-glucoside	0.03275 a	0.00295 b	0.0023 b	0.0019	<.0001
Tangeretin	0.0012 a	0.002 a	0.00115 a	0.0005	0.3801
Kaempferol-3-O-rhamnoside	0.01938 a	0.0036 b	0.0039 b	0.0020	0.0005
Isorhamnetin	0.1614 a	0.9229 b	0.9449 b	0.0493	<.0001
Kaempferol	0.16 a	0.1202 ab	0.1067 b	0.0118	<.0001
Kaempferol 3-O-robinoside-7-O-rhamnoside (robinin)	3.1775 a	3.635 a	3.145 a	0.3008	0.4670
Kaempferol 3-O-rutinoside-4'-glucoside	1.47 a	0.47 b	0.4125 b	0.1705	0.0028
Kaempferol-3-O-glucoside	0.07475 a	0.04725 a	0.1945 a	0.0374	0.0469
Kaempferol-3-O-rutinoside	0.5395 a	0.3486 ab	0.1062 b	0.0800	0.0127
Kaempferol-7-O-glucoside	0.03585 a	0.01668 b	0.01363 b	0.004023	0.0072
Kaempferol-7-O-neohesperidoside	0.01065 a	0.00845 a	0.009375 a	0.0008	0.1793
Myricetin	1.7720 a	1.9587 a	2.2097 a	0.1176	0.0756
Myricetin-3-O-rhamnoside	4.4707 a	0.006775 b	0.02833 b	0.1436	<.0001
Quercetin	1.0418 a	2.3262 b	2.942 c	0.1433	<.0001
Quercetin-3,4'-di-O-glucoside	0.0425 a	0.09 a	0.195 b	0.0260	0.0071
Quercetin-3-O-galactoside	0.06953 a	0.819 b	0.3467 a	0.0695	0.0007
Quercetin-3-O-glucoside (Isoquercetrin)	0.1374 a	2.3584 b	2.6592 b	0.1699	<.0001
Quercetin-3-O-rhamnoside (Quercitrin)	0.3089 a	0.0067 b	0.006425 b	0.0139	<.0001
Quercetin-3-O-rutinoside(Rutin)	0.3079 a	15.0792 b	7.3583 c	0.5009	<.0001
Quercetin-4'-O-glucoside (Spiraeoside)	1.9323 a	0.8558 a	0.5865 a	0.4067	0.0965
(+)-Catechin	49.8225 a	0.3025 b	0.0625 b	0.2313	<.0001

catechin-glucoside	31.9125 a	31.7575 a	23.9325 a	4.6058	0.4115
(-)-Epicatechin	185.98 a	0.4775 b	0.035 b	3.8286	<.0001
(-)-Epigallocatechin	4.6100a	0.1350b	0.1267b	0.5460	0.0010
(−)-Gallocatechin	28.95 a	0.0375 b	0.0225 b	0.1850	<.0001
Eriocitrin	0.005799 a	0.01 ab	0.01735 b	0.0058	0.0523
Eriodictyol	0.0759 ab	0.1278 a	0.02432 b	0.0226	0.0310
Hesperetin-7-O-rutinoside	0.06803 a	0.3307 b	0.1849 c	0.0232	<.0001
Naringenin	0.0672 a	0.01853 b	0.01268 b	0.0035	<.0001
3,4-Dihydroxybenzoic acid	1.2000 a	1.7625 b	1.1200 a	0.0698	0.0002
4-hydroxybenzoic acid	3.97 ab	8.285 a	1.5675 b	1.5589	0.0387
Gallic acid	0.955 a	0.1725 b	0.21 b	0.0329	<.0001
Salicin	0.8775 a	0.5575 ab	0.3475 b	0.0985	0.0128
Vanillic acid-4-β-D-glucoside	6.0700 a	3.9125 b	2.0575 c	0.2286	<.0001
Vanillin	0.3825 a	0.605 a	0.385 a	0.0838	0.1535
Caffeic acid	0.7975 a	10.9425 b	3.5075 c	0.2514	<.0001
Ferulic acid (trans)	0.7975 a	24.4 b	0.9825 a	0.5276	<.0001
trans-3-hydroxycinnamic acid	0.4099 a	7.4758 b	0.4288 a	0.1690	<.0001
Procyanidin A2	0.0807 a	0.03335 b	0.03368 b	0.0057	0.0003
Procyanidin B1	88.33 a	0.02 b	0.02 b	0.7797	<.0001
Procyanidin B2	33.1 a	0.03 b	0.0225 b	0.1651	<.0001
Procyanidin B3	219.46 a	0.0525 b	0.03 b	2.5597	<.0001
Procyanidin C1	87.8625 a	0.0875 b	0.09 b	1.3921	<.0001
Resveratrol	0.1075 a	0.0448 a	0.04527 a	0.0204	0.0924
Resveratrol-3-β-mono-D-glucoside (Polydatin)	4.5758 a	0.009425 b	0.00755 b	0.0513	<.0001
4-Hydroxy-6-methylcoumarin	0.0528 a	0.08593 a	0.06118 a	0.0119	0.1776

The compounds with the highest concentration in the seed coat tissue that were also significantly different among the three genotypes are shown in Figure 5.1. The compounds belonging to the class of *proanthocyanidins* (condensed tannins) were present only in ILB 938/2. In the class of *flavan-3-ols*, (-)-epicatechin, (+)-catechin, (-)-epigallocatechin, (-)-gallocatechin were also found mainly in the wild type with minor traces in the low tannin lines. Resveratrol-3-β-mono-D-glucoside (polydatin) as well as cyanidin-3,5-di-O-glucoside were present only in ILB 938/2. The situation varied when considering the sub-classes of *flavonols* and *dihydroflavonols*. Among the *flavonols,* kaempferol 3-O-rutinoside-4'-glucoside, quercetin-3-O-rhamnoside (quercitrin) and myricetin-3-O-rhamnoside were present in significantly higher concentration in

ILB 938/2, however quercetin, quercetin-3-O-glucoside (isoquercetin), isorhamnetin and quercetin-3-O-rutinoside (rutin) were significantly more concentrated in the low tannin lines compared with the wild type concentrations. In the *dihydroflavonols* group, dihydromyricetin was present in significantly higher concentration in ILB 938/2. Taxifolin (dihydroquercetin) was present in significantly higher concentration in Disco/2 compared with both ILB 938/2 and CDC Snowdrop.

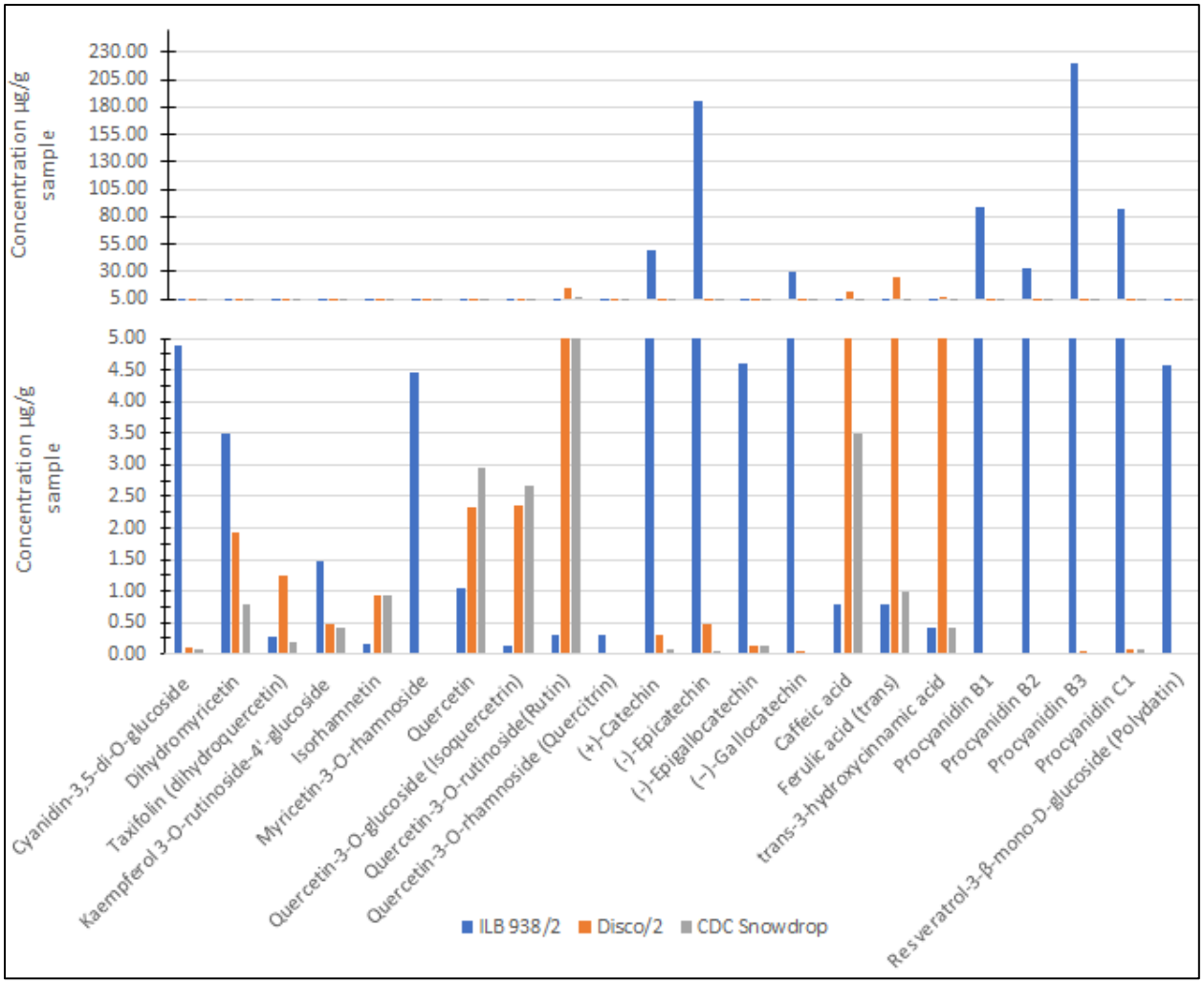


Figure 5.1. Concentration of the phenolic compounds in faba bean seed coat that exhibit the greatest differences among the three genotypes ILB 938/2, Disco/2 and CDC Snowdrop using the targeted LC-SRM method.

An interesting pattern is observed for the *hydroxycinnamic acids*, where caffeic acid, ferulic acid (trans) and trans-3-hydroxycinnamic acid are all present at consistently higher concentrations in Disco/2 compared with that in both ILB 938/2 and CDC Snowdrop, and CDC Snowdrop had significantly higher concentrations of only caffeic acid compared with ILB 938/2.

The main differences in the concentration of the phenolic compounds between the wild type and the two low tannin lines were for the *proanthocyanidins* and their precursors, as well as the flavan-3-ols, which were almost absent in both Disco/2 and CDC Snowdrop. This suggests that the two genes *zt1* and *zt2* act at a point in the phenolic compounds pathway that is preceding the synthesis of *flavan-3-ols* and *anthocyanidins* which are the aglycones of the *anthocyanins*. The presence of cyanidin-3,5-di-O-glucoside (*anthocyanins* class) only in the wild type supports this hypothesis.

The more heterogeneous situation for the classes of *flavonols* and *dihydroflavonols* may be an indication of some shifts induced by the two low tannin genes at earlier stages in the pathway. Unfortunately, for this method of analysis, we did not have standards for the compounds in the class of *leucoanthocyanidins,* which are the precursors of both *flavan-3-ols* and of the coloured *anthocyanins*. Consequently, we were unable to gain further insight one of the key parts of the pathway that leads to the production of *proanthocyanidins* (condensed tannins). These results, however, helped us to locate the action of *zt1* and *zt2* to a precise point within the complex phenolic compounds biosynthetic pathway. Figure 5.2 shows the general *flavonoids* (comprised of the classes of *flavonols*, *flavanones, isoflavones, flavones, flavan-3-ols, and anthocyanins*) biosynthetic pathway in plants. A probable point of action of the low tannin genes *zt1* and *zt2,* compared to the wild type, is likely in the steps between *dihydroflavonols* and *flavan-*

3-ols. This hypothesis suggests that, the *zt1* and *zt2* alleles may represent mutations in the DNA sequences coding for the genes DFR (Dihydroflavonol reductase), LAR (Leucoanthocyanidin reductase) or ANS (Anthocyanidin synthase), or possibly for transcription factors that regulate their gene expression.

This last hypothesis is derived from the results of Webb *et al.* (2016), who found that the *zt1* allele is responsible for the deletion of the entire 5' end of a faba bean gene syntenic in *Medicago truncatula*. The gene (MtWD40-1) was previously characterized in the model species and found to regulate the production of *anthocyanins*.

The results also showed some defined differences between Disco/2 and CDC Snowdrop in regard to phenolic profiles of the seed coat. In particular, the three *hydroxycinnamic acids,* caffeic acid, ferulic acid (trans) and trans-3-hydroxycinnamic acid, are clearly distinguished between Disco/2 and CDC Snowdrop. For taxifolin, and dihydromyricetin, the two genotypes had significantly different concentrations, again with a higher concentration of these compounds observed in Disco/2. These results are not detailed enough to suggest a possible different mode of action of the two genes in the phenolic compounds pathway, but they may be useful to define biochemical markers for the differentiation of low tannin accessions recessive for *zt1* from those recessive for *zt2*.

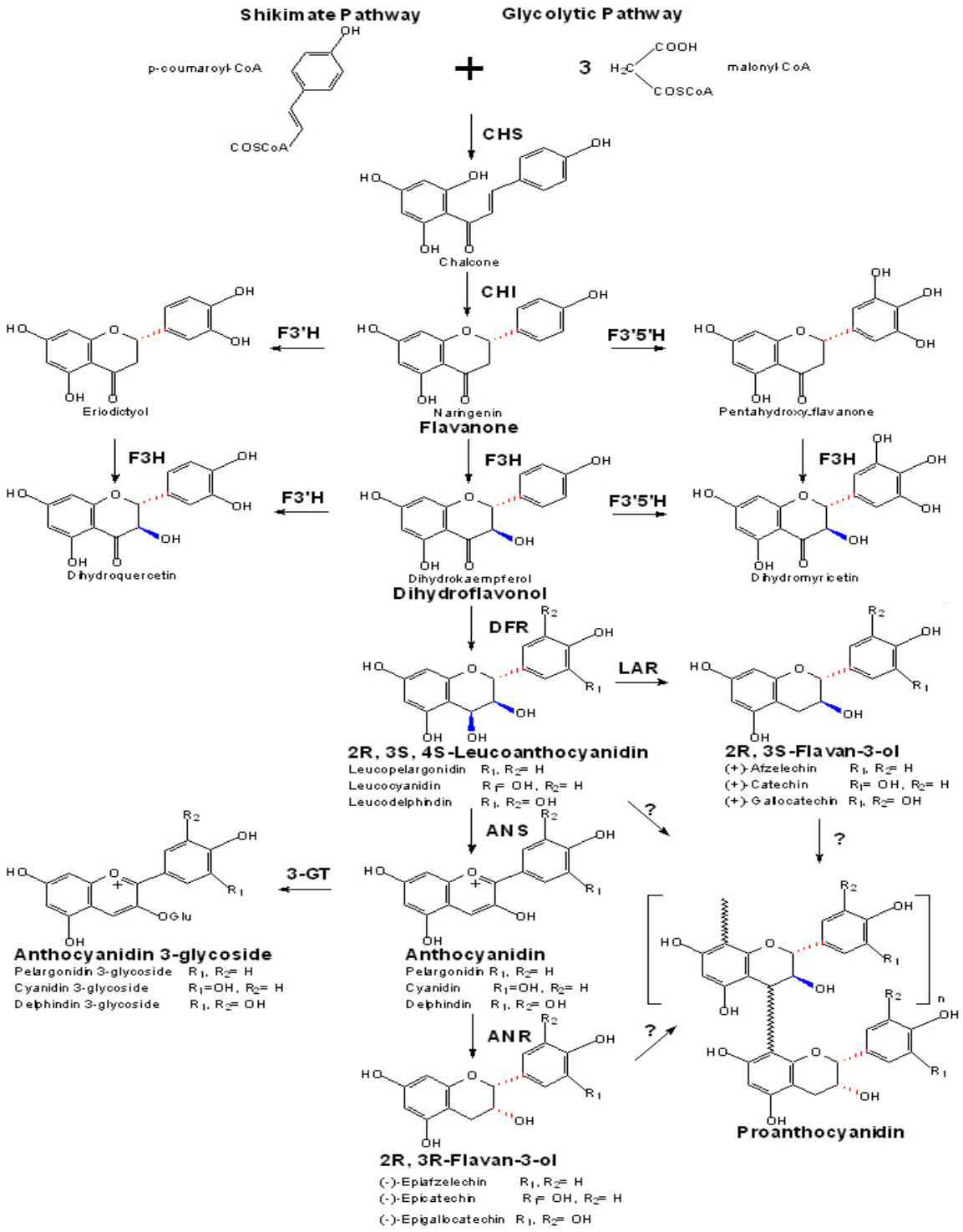


Figure 5.2. General *flavonoid* biosynthetic pathway which leads to the synthesis of *proanthocyanidins*, the possible point of action of *zt1* and *zt2* genes is between the synthesis of *dihydroflavonols* and *anthocyanidins/flavan-3-ols*. (from He *et al.,* 2008)

5.4.1.2 Flowers

The phenolic compounds identified in the floral tissue and their mean values among the three faba bean genotypes, ILB 938/2, Disco/2 and CDC Snowdrop, are listed in Table 5.4.

Table 5.4. Concentration of phenolic compounds in µg/g of sample detected in the flowers of three faba bean genotypes ILB 938/2, Disco/2 and CDC Snowdrop. The values are the mean of 4 samples. Tukey's test was used for the multi treatment comparisons. Means of each compound with different letters are significantly different at p <0.05. SEM= standard error of the mean. The calculated P values are presented. Highlighted in green are the compounds which have significantly different concentrations between wild type and *zt* lines. Highlighted in blue are the compounds with significantly different concentration between the two *zt* lines.

COMPOUND NAME	Concentration (µg/g)			SEM	P Value
	ILB938/2 *ZT1ZT2*	DISCO/2 *ZT1zt2*	CDC SNOWDROP *zt1ZT2*		
Cyanidin-3,5-di-O-glucoside	0.3933 a	0.8185 a	0.3715 a	0.2187	0.1874
Cyanidin-3-O-rhamnoside	0.05013 a	0.01677 b	0.0165 b	0.0044	0.0014
Delphinidin-3,5-di-O-glucoside	0.2695 a	0.2678 a	0.266 a	0.0018	0.4081
Delphinidin-3-O-rhamnoside	0.2771 a	0.0237 b	0.025 b	0.0415	0.0095
Delphinidin-3-β-D-Glucoside	0.02443 a	0.02285 a	0.0223 a	0.0006	0.0633
Malvidin-3-O-glucoside	1.5969 a	0.02217 b	0.02203 b	0.2228	0.0009
Pelargonin	0.02265 a	0.02193 b	0.02177 b	0.0001	0.0111
Dihydrokaempferol	0.07423 a	0.06578 a	0.02136 a	0.0214	0.8186
Taxifolin (dihydroquercetin)	0.0337 a	0.05653 a	0.02153 a	0.0088	0.0748
Apigenin	0.6134 a	0.6784 a	0.1315 a	0.2561	0.3046
Apigenin-7-O-glucoside	0.02003 a	0.03982 a	0.02588 a	0.0105	0.4263
Apigenin-7-O-rutinoside	1.3102 a	7.9608 ab	18.1831 b	2.6319	0.0045
Vitexin	0.3663 a	0.0031 b	0.0041 b	0.0788	0.0135
Diosmetin	0.042 a	0.09528 a	0.04033 a	0.0327	0.4352
Luteolin	1.3213 a	0.6336 a	0.3401 a	0.3965	0.2519
Luteolin-3',7-di-O-glucoside	5.8243 a	4.2915 b	0.116 c	0.1150	<.0001
Luteolin-4'-O-glucoside	0.1233 a	0.0329 b	0.0272 b	0.0173	0.0057
Luteolin-7-O-glucoside	0.2654 a	0.1156 a	1.0854 b	0.0533	<.0001
Luteolin-8-C-glucoside	0.04885 a	0.0034 b	0.004075 b	0.0084	0.0060
Kaempferol-3-O-rhamnoside	1.0316 a	0.00253 b	0.003 b	0.1314	0.0005
Isorhamnetin	0.07405 a	1.2842 b	0.0628 a	0.1522	0.0004
Kaempferol	4.4899 ab	3.0784 a	5.4642 b	0.4718	0.0182
Kaempferol 3-O-robinoside-7-O-rhamnoside (robinin)	2879.34 a	2619.24 a	2148.65 a	306.3900	0.2821

Kaempferol 3-O-rutinoside-4'-glucoside	2.1200 a	12.187 ab	25.2722 b	3.9834	0.0085
Kaempferol-3-O-D-galactoside	3.8454 a	1.2438 a	16.7685 b	0.9715	<.0001
Kaempferol-3-O-glucoside	4.5537 a	7.8773 a	92.3048 b	5.4279	<.0001
Kaempferol-3-O-rutinoside	0.9728 a	3.4127 b	4.8755 b	0.5772	0.0032
Kaempferol-7-O-glucoside	0.288 a	0.1584 b	NF	0.0116	0.0002
Myricetin	0.1244 a	0.1257 a	0.1232 a	0.0012	0.3707
Myricetin-3-O-rhamnoside	0.211 a	0.00515 b	0.002525 b	0.0114	<.0001
Quercetin	1.0421 a	2.6188 b	0.8348 a	0.2713	0.0023
Quercetin-3,4'-di-O-glucoside	0.01363 a	0.03212ab	0.04915 b	0.0066	0.0130
Quercetin-3-O-galactoside	0.945 a	1.6206 ab	2.5797 b	0.3790	0.0401
Quercetin-3-O-glucoside (Isoquercetrin)	0.9728 a	3.4127 b	4.8755 b	0.5772	0.0032
Quercetin-3-O-rhamnoside (Quercitrin)	1.8251 a	0.00418 b	0.002925 b	0.0937	<.0001
Quercetin-3-O-rutinoside(Rutin)	160.18 a	603.84 b	259.33 a	26.2977	<.0001
Quercetin-4'-O-glucoside (Spiraeoside)	0.2569 a	0.3652 a	0.2569 a	0.0428	0.1192
Tiliroside	0.01725 a	0.0321 ab	0.0421 b	0.0059	0.0386
(+)-Catechin	1.7763 a	0.013 a	0.0094 a	0.5060	0.0551
(-)-Epicatechin	0.02845 a	0.013 a	0.01305 a	0.0051	0.0916
(−)-Gallocatechin	0.0216 a	0.02055 a	0.01975 a	0.0004	0.1999
Eriodictyol	0.08942 a	0.1333 a	0.0823 a	0.0371	0.5926
Hesperetin	0.005700 a	0.01315 a	0.008025 a	0.0042	0.4674
Naringenin	0.2726 a	0.3438 a	0.2461 a	0.0553	0.4641
3,4-Dihydroxybenzoic acid	0.4861 a	0.3389 b	0.319 b	0.0336	0.0126
4-amino salicylic acid	0.2149 a	0.2503 a	0.2832 a	0.0259	0.2293
4-hydroxybenzoic acid	14.0294 a	9.0067 a	12.8945 a	1.2331	0.0457
Gallic acid	0.0449 a	0.0539 a	0.03632 a	0.0084	0.3757
Syringic acid	0.1413 a	0.1018 a	NF	0.0276	0.6155
Vanillic acid-4-β-D-glucoside	4.6914 a	4.274 a	3.8855 a	0.4786	0.5177
Vanillin	0.1846 a	0.1624 a	0.165 a	0.0493	0.9372
Caffeic acid	1.5021 a	0.7298 b	1.0349 b	0.0873	0.0005
Chlorogenic acid	0.02485 a	NF	0.02173 a	0.0012	0.0437
Procyanidin A2	0.08463 a	0.09133 a	0.02837 b	0.0084	0.0084
Procyanidin B1	0.0371 a	0.0209 a	0.02075 a	0.0053	0.1312
Procyanidin B2	0.4833 a	0.5357 a	0.5403 a	0.0675	0.7950
Procyanidin B3	1.5975 a	0.02 b	0.02 b	0.2227	0.0009
Resveratrol-3-β-mono-D-glucoside (Polydatin)	0.1713 a	0.1601 a	0.1336 a	0.0240	0.5446
Genistein	0.01783 a	0.02673 a	0.0138 a	0.0056	0.2445
Phloretin	0.004175 a	0.00285 a	0.009275 b	0.0006	<.0001

The list of compounds with the highest concentrations that were significantly different among the three genotypes in the floral tissue are shown in Figure 5.3.

A few compounds were found only in ILB 938/2, including delphinidin-3-O-rhamnoside and, malvidin-3-O-glucoside from the *anthocyanins* class, kaempferol-3-O-rhamnoside and quercetin-3-O-rhamnoside (quercitrin) from the *flavonols* class, and (+)-catechin and procyanidin B3 from *flavan-3-ols* and *proanthocyanidins* classes, respectively. Luteolin-3',7-di-O-glucoside and caffeic acid were also present in higher concentration in ILB 938/2 compared with Disco/2 and CDC Snowdrop. The compounds apigenin-7-O-rutinoside, luteolin-7-O-glucoside, kaempferol 3-O-rutinoside-4'-glucoside, kaempferol-3-O-D-galactoside and kaempferol-3-O-glucoside were present in significantly higher amounts in CDC Snowdrop compared with both Disco/2 and ILB 938/2, while kaempferol-3-O-rutinoside, quercetin-3-O-galactoside and quercetin-3-O-glucoside (isoquercetrin) distinguished both low tannin lines from ILB 938/2, being present at higher concentrations in the two low tannin genotypes. Two compounds, shown in Figure 5.4, were present in exceptionally higher amounts in the flowers compared to all the other quantified compounds. Kaempferol 3-O-robinoside-7-O-rhamnoside (robinin) had the highest concentration among all the compounds included in the analysis, but did not distinguish the three genotypes, whereas quercetin-3-O-rutinoside (rutin), which also had a high concentration, was significantly higher in concentration in the flower tissue of Disco/2 than in ILB 938/2 and CDC Snowdrop. This last compound, together with quercetin (Figure 5.4) were both quantified in the highest amount in Disco/2 and separated this genotype from the other two.

These results from the flower tissue analysis are in line with those found for the seed coats when considering a putative point of action of the two low tannin genes in the phenolic compound biosynthetic pathway. The identification of (+)-catechin, delphinidin-3-O-rhamnoside, malvidin-3-O-glucoside, and proanthocyanidin B3 only in the flower tissue of ILB 938/2 clearly

showed that the pathway leading to the accumulation of *anthocyanins* (colored compounds) and *proanthocyanidins* (condensed tannins) is blocked at a stage preceding the synthesis of *flavan-3-ols* in Disco/2 and CDC Snowdrop.

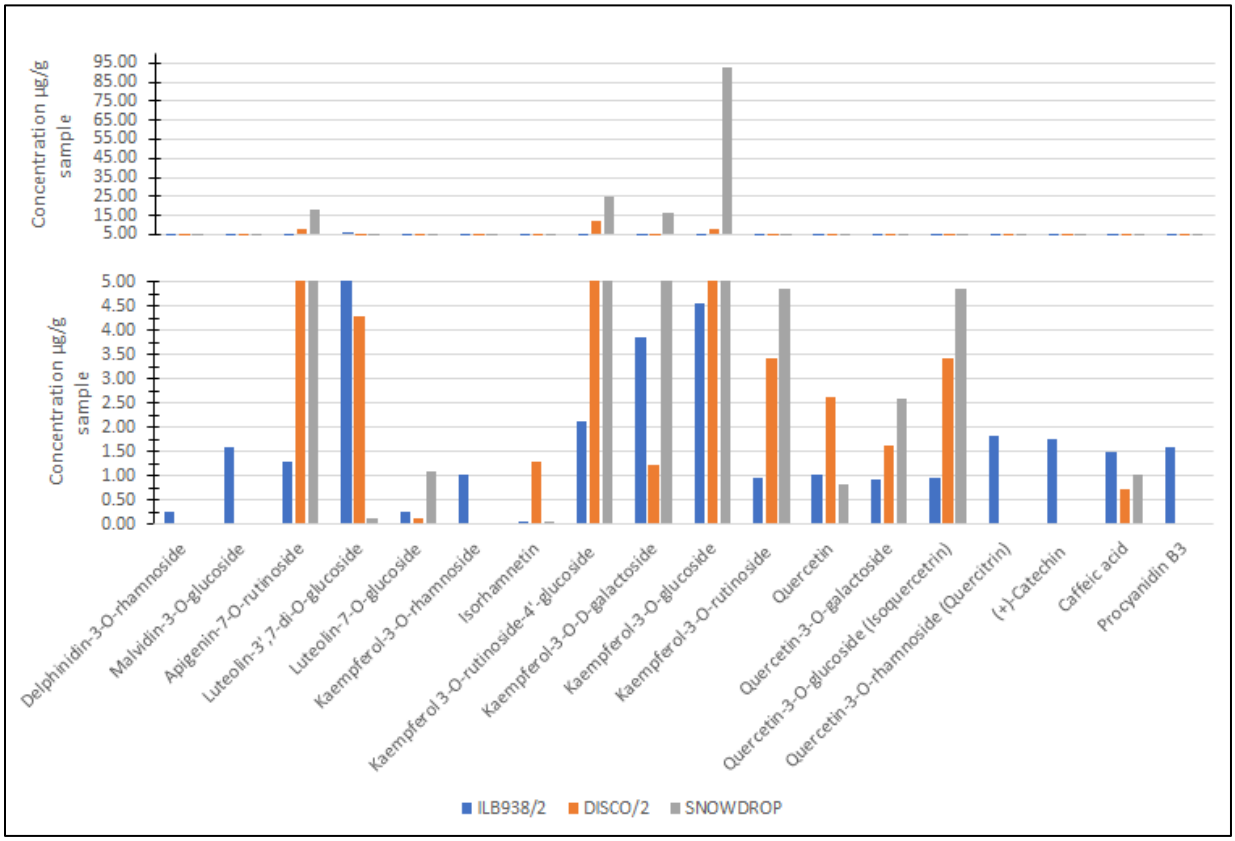


Figure 5.3. Concentration of the phenolic compounds in faba bean flowers that exhibit the greatest differences among the three genotypes ILB 938/2, Disco/2 and CDC Snowdrop using the targeted LC-SRM method.

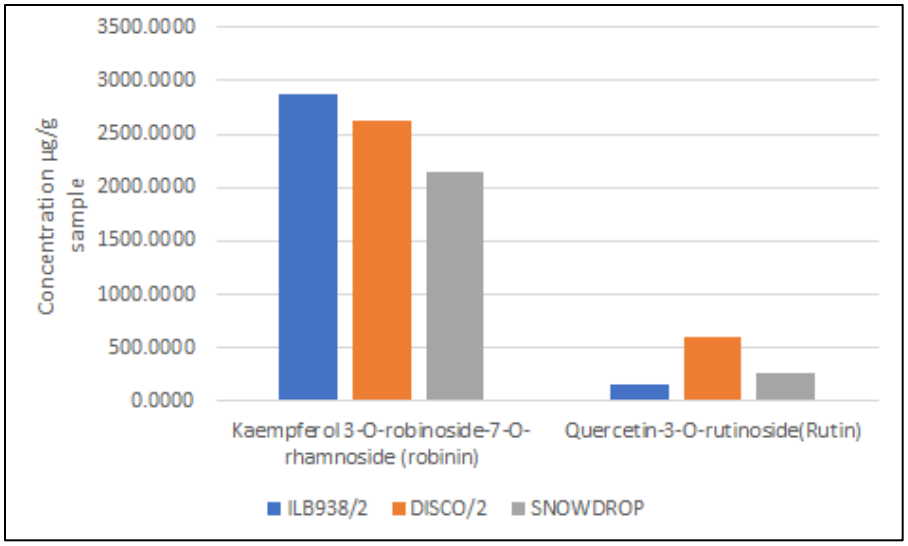


Figure 5.4. Concentration of the phenolic compounds with exceptionally high content in the flower tissue of Disco/2, ILB 938/2 and CDC Snowdrop, using the targeted LC-SRM method.

Compounds in the *dihydroflavonols* class were not significantly different in their concentrations among the three genotypes, but the presence of (+)-catechin only in ILB 938/2 supports the findings in the seed coat that *zt1* and *zt2* act in one of the intermediate steps between the *dihydroflavonols* and *flavan-3-ols/anthocyanins* biosynthesis. Once again, we can only infer this, since we do not have standards for the *leucoanthocyanidins,* which are the precursors of both *flavan-3-ols* and *anthocyanidins* as shown in Figure 5.2.

It is noteworthy that *proanthocyanidins* and *flavan-3-ols* were present only in low concentrations in the flowers of the wild type compared with the seed coat, while the *anthocyanins* had the opposite trend. This suggests that the accumulation of the phenolic compounds is tissue specific, and in particular the complex *proanthocyanidins* polymers and their precursors *flavan-3-ol*s are accumulated mainly in the seed coat. Furthermore, some *flavonols* were highly concentrated only in the flowers.

Clear differences exist in the polyphenols profiles in the flowers of Disco/2 and CDC Snowdrop, with some compounds present in very high concentrations in one of the two low tannin lines. This is an important result that could lead towards detection of biochemical markers. In fact, flower tissue is easier to study compared with seed coat as it can be harvested at an earlier stage, presents some visible differences (spotted vs white) and takes less time for the sample preparation in LC-MS studies. Therefore, a deeper analysis of the phenolic profile in the flower tissue could yield effective biochemical markers with applications in future studies. These can be useful indicators to better differentiate the mode of action of the two genes *zt1* and *zt2*. Furthermore, they can be matched with molecular markers and become additional tools to distinguish faba bean genotypes recessive for *zt2* or *zt1*.

5.4.2 Detection of unknown compounds in the flowers

The results of the analyses for the detection of unknown compounds showed clear differences in the phenolic profiles of the flowers of the three genotypes ILB 938/2, Disco/2 and CDC Snowdrop.

Figure 5.5 shows a volcano plot where each dot represents a compound detected by the LC-MS analysis in the flower tissue of Disco/2 and CDC Snowdrop. The red and green areas highlight the compounds that are significantly different in their content between Disco/2 and CDC Snowdrop at a *P* value of 0.05 and with a fold change of at least Log2=1. In the red and green areas are the compounds with the higher concentration in Disco/2 and CDC Snowdrop, respectively.

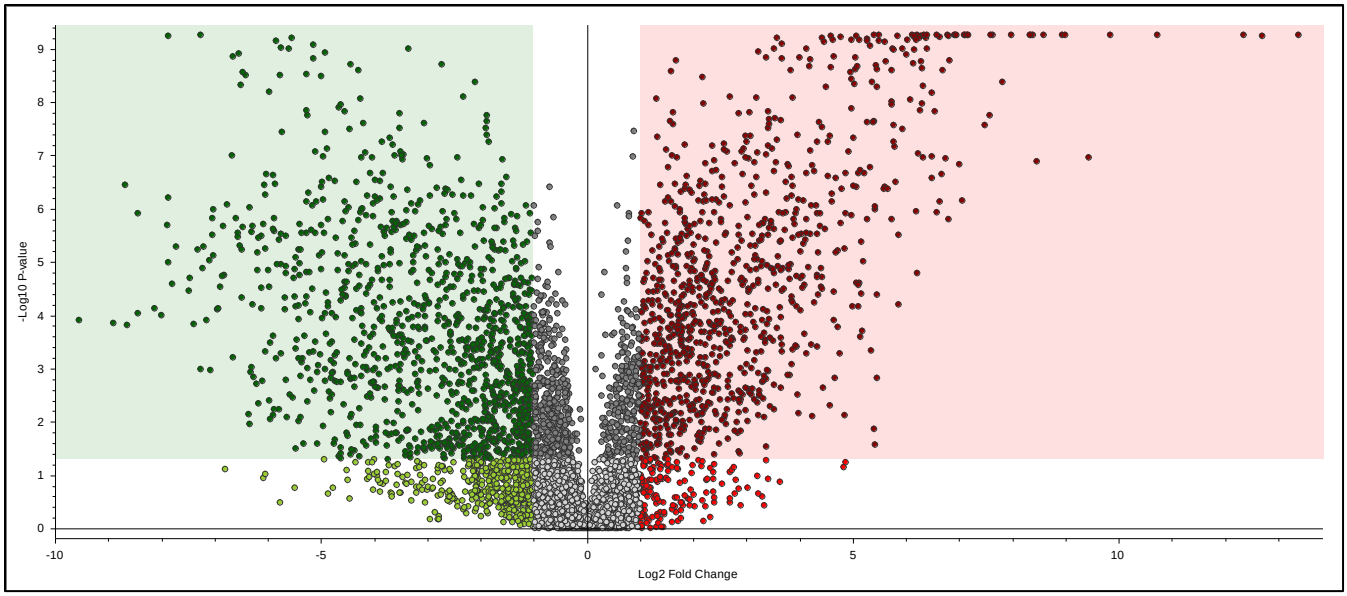


Figure 5.5. Volcano plot of phenolic compounds in the flower of Disco/2 vs CDC Snowdrop. The highlighted area includes all the compounds that are significantly different in their concentration (*P*< 0.05) between Disco/2 and CDC Snowdrop for a value of at least Log2=1 fold change. In red are the compounds with higher content in Disco/2 vs CDC Snowdrop and in green the compounds with higher content in CDC Snowdrop vs Disco/2.

The compounds corresponding to the dots in the far right and far left of the graph shown in Figure 5.5, with a cut off value of Log2 fold change=7.5, indicates a difference of at least 180-fold in the content of the selected compounds between the two genotypes. This selection yielded 33 compounds, all of them unknown to the reference libraries, and one compound was classified as similar to rutin (quercetin 3-O-rutinoside) by the software, based on the search results from the reference libraries. The majority of the compounds presented MS/MS spectra that are typical of compounds of the class of *flavonols*, with quercetin and kaempferol as the main fragment ions and various secondary fragment ions containing additional groups.

We were able to putatively identify some unknown compounds when comparing the MS/MS data generated in our analysis with those of previously reported compounds in the literature. One of these compounds likely corresponds to isorhamnentin-3-O-rutinoside, and it

was found in the medicinal herb Pollen Typhae by Chen *et al.* (2015). Figure 5.6 shows the MS/MS spectra associated with the unknown compound and the two intensity vs. time plots for Disco/2 and CDC Snowdrop, respectively (the peak elutes @ 13.6 min). The loss of a fragment ion with molecular weight 308 is consistent with the loss of a rutinose unit (dimeric sugar), while the fragment ions at m/z 315, 300, 271, are consistent with those of isorharmnetin.

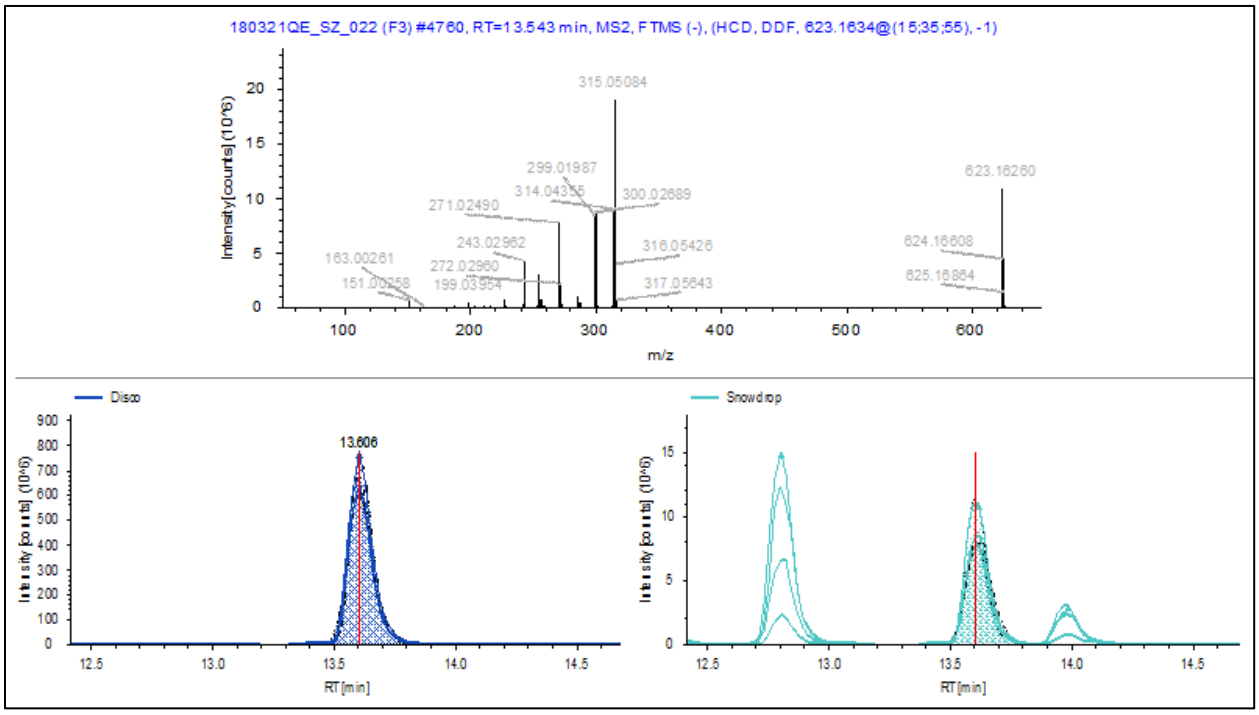


Figure 5.6. Detection of unknown compound in the flower of Disco/2 vs CDC Snowdrop faba bean. The graph at the top is the MS/MS spectrum, of a compound with m/z 623 and main fragment ions at m/z 315, 300 and 271. The two graphs at the bottom show the intensity of the ion as a function of retention time (compound elutes @ 13.6 min) in Disco/2 and CDC Snowdrop, respectively.

When comparing the phenolic profiles of the flower of Disco/2 with ILB 938/2, the results for the unknown compounds with a cut off value of Log2 fold changes of 7.5 yielded 67 compounds, but none of these compounds were identified by the software libraries. Most of the compounds with higher content in ILB 938/2 than Disco/2 present elevated molecular weights, which suggests the presence of complex oligomeric compounds in the class of *proanthocyanidins*

(condensed tannins). These spectra were absent from the group of compounds that differentiated Disco/2 from CDC Snowdrop. However, the MS/MS spectra of these compounds do not correspond with those of previously characterized *proanthocyanidins* and most of them appear to be modified forms of a single basic *flavonol* unit with molecular weight 578.

Figure 5.7 shows the spectra of a compound with m/z 866, and its MS/MS fragmentation pattern containing fragment ions with m/z 577, 431 and 285. A molecular ion with m/z 866 can correspond to a proanthocyanidin trimer (Montero *et al.* 2013), and this may be expected for a compound which differentiates the wild type ILB 938/2 from the *zt2* low tannin line.

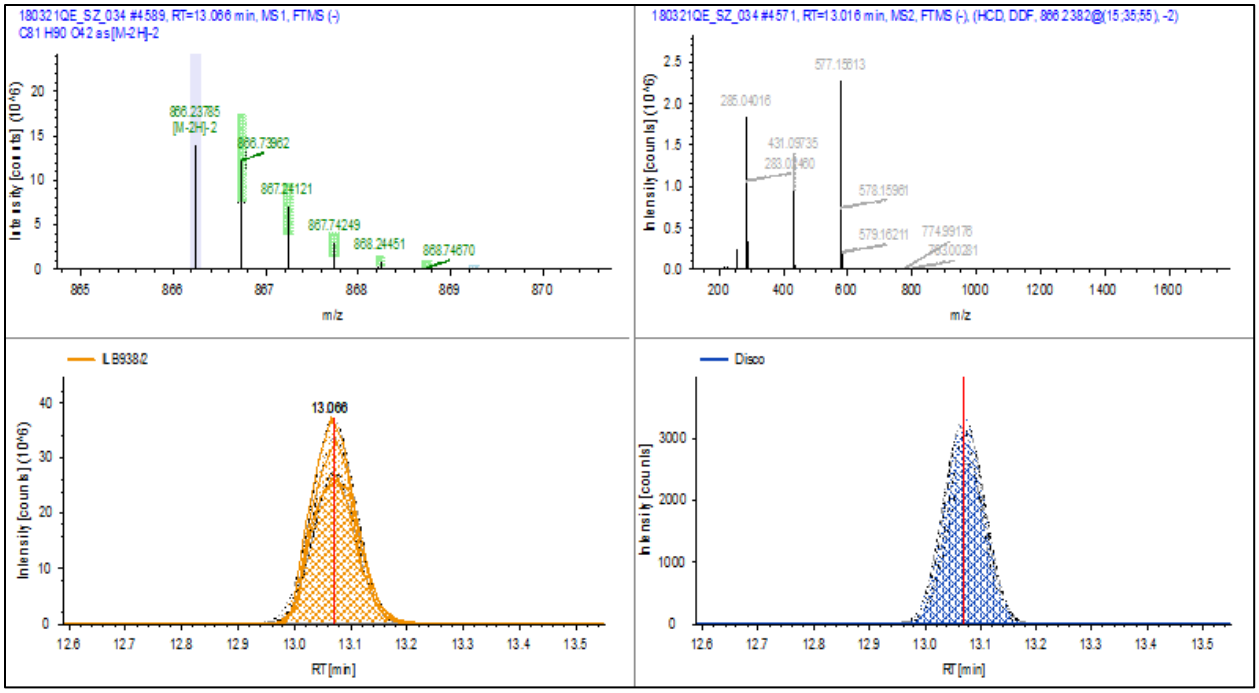


Figure 5.7. Detection of an unknown compound in the flowers of ILB 938/2 vs Disco/2 faba bean. The graph at the top left is the MS spectrum, of a compound with m/z 866. The graph at the top right is the MS/MS spectrum for the same ion with main fragment ions @ m/577, 431 and 285 (m/z). The two graphs at the bottom show the intensity of the ion as a function of retention time (compound elutes @ 13.066 min) in Disco/2 and CDC Snowdrop, respectively.

Nevertheless, this early assessment is clearly misleading when we take into consideration the fragmentation pattern. In fact, the MS/MS spectra containing the fragment ions m/z 577, 431

and 285 is not consistent with a proanthocyanidin trimer, and instead is consistent with that of a kaempferol with 2 rhamnose units added. The ion @ m/z 577 possibly corresponds to Kaempferol-3,7-O-dirhamnoside (kaempferitrin), which was previously characterized in studies of the medicinal properties of the inflorescences of the tree *Tilia americana* (Aguirre-Hernández *et al.* 2010). The molecular structure of Kaempferol-3,7-O-dirhamnoside is shown in Figure 5.8. The loss of each rhamnose group (146 m/z units) from the m/z 577 ion of kaempferol-3,7-O-dirhamnoside is consistent with the fragmentation in Figure 5.7 and the large fragment ion with m/z 285 (and smaller ion at m/z 255) are consistent with kaempferol. This basic unit and fragmentation pattern were common to many of the other compounds which differentiate ILB 938/2 from Disco/2. The same compound was found in very low amounts in CDC Snowdrop (similar value to Disco/2), therefore, it is synthesized mainly in ILB 938/2.

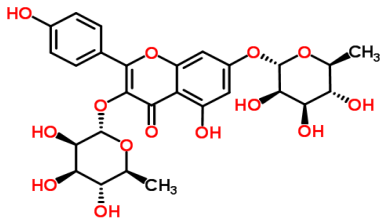


Figure 5.8. Molecular structure of Kaempferol-3,7-O-dirhamnoside, possible basic *flavonol* unit present in various compounds found in the flowers of ILB 938/2.

However, the loss of a 289 unit from the original compound with m/z 866 is not clearly explainable, and further research in the literature may be helpful to characterize this sub unit and therefore the structure of the ion with m/z 866.

The high occurrence of this *flavonol* unit with molecular weight 578 and its polymers and glycosylated forms in the flowers of the wild type ILB 938/2 seems to be an important pattern in the accumulation of phenolic compounds in the wild type/spotted flower compared with the white flower types. One possible explanation is that the low tannin genes *zt1* and *zt2* block the synthesis or accumulation of this compound in the flowers or shift the pathway towards the accumulation of other compounds in the *flavonols* class. Kaempferol-3,7-O-dirhamnoside may be a precursor of colored compounds in the wild type flowers of faba bean, but this hypothesis needs to be verified with further research.

These results are consistent with what was found for the analysis of the flowers in the targeted method in that they suggest that the *zt1* and *zt2* genes act on the phenolic compound biosynthetic pathway favoring the accumulation of different *flavonols* compared with those accumulated in the wild type. Furthermore, compounds in the classes of *flavan-3-ols* and *proanthocyanidins* are not present in the flowers of the wild type, whereas they are abundant in the seed coat, demonstrating tissue specific accumulation. On the other hand, we found a high concentration of unknown species with a *flavonol* basic unit (MW 578) in the flowers of ILB 938/2. This research also confirmed the significance and usefulness of flower tissue for LC-MS studies targeting the quantification and characterization of phenolic compounds.

In this study, all genotypes were grown in the same controlled conditions, as such the observed differences in the concentration of phenolic compounds among varieties are likely to primarily reflect the genotypic variation. Further studies that consider different years with different locations are needed to quantify the influence of environmental conditions on the phenolic composition of seed coats and flowers.

5.5 Conclusions

The study of the phenolic profiles of the three faba bean genotypes Disco/2, CDC Snowdrop and ILB 938/2 using LC-MS techniques identified significant differences in the phenolic compounds present in the seed coats and flowers of the three genotypes. The wild type was distinguishable from the two low tannin lines by the presence of compounds in the class of *flavan-3-ols, anthocyanins* and *proanthocyanidins,* which were absent from Disco/2 and CDC Snowdrop in the seed coat. In the flowers, the differences between *zt* lines and wild type were mainly due to the accumulation of compounds with a basic *flavonol* unit only detected in the wild type.

The results from the seed coat analysis suggest that the two genes *zt1* and *zt2* act in the phenolic compounds biosynthetic pathway between the synthesis of *dihydroflavonols* and *anthocyanins*. In the flower tissue results, *zt1* and *zt2* favor the accumulation of different *flavonol* species compared to the wild type.

Overall this analysis highlighted the differences between the phenolic profiles of Disco/2 and CDC Snowdrop both in the flower and seed coat tissues proving that the two genes have different modes of action in the phenolic compounds biosynthetic pathway.

CHAPTER 6- GENERAL DISCUSSION

6.1 Discussion

The low tannin phenotype in faba bean is controlled by two independent recessive genes, *zt1* and *zt2*, and genotypes that are homozygous recessive for either of the two genes display completely white flowers. Wild type flowers develop black spots on the wing flowers and purple-reddish veins in the banner petals. The low tannin trait is desired by plant breeders and food and feed industries because of the improved nutritional quality related to the trait. Condensed tannins in the seed coat, typical of genotypes with dominant alleles at either *zt* loci, are antinutritional compounds for monogastric animals that result in decreased digestion of protein. Furthermore, low tannin genotypes also have thinner seed coats, which is associated with higher protein fraction in the seed and faster cooking time due to reduced time for imbibition. These characteristics make low tannin faba bean increasingly requested by a food industry that is rapidly developing interest in plant-based foods. Despite the importance of the low tannin trait, very little is known about the phenolic compounds which differentiate low tannin from wild type genotypes. Most of the previous research focused on the quantification of total phenolics and tanning power, and only a few studies characterized some of the various polyphenolic compounds, which are often improperly categorized as tannins.

The purpose of this research was to develop and validate a molecular marker for the *zt2* gene and to use molecular techniques to identify the genotype which carries both low tannin genes in the recessive form *zt1zt2*. It also included an analysis of the phenolic compounds profile of the seed coats and flowers of three faba bean genotypes, one which had normal tannin

content in the seed coat and spotted flowers, and both of the two low tannin genotypes, one recessive for *zt1* and the other recessive for *zt2*.

The first objective was to develop a molecular marker linked to the *zt2* gene, using a RIL population at the F6. A set of SNP markers anchored to the fully sequenced genome of *Medicago truncatula* was used for this purpose. Two homologous SNP markers were located three cM from the *zt2* gene, and a third SNP marker was found 10.5 cM from the *zt2* gene. The results of the linkage study showed that the *zt2* gene is located in a genomic region of faba bean with conserved synteny with *Medicago truncatula* chromosome 1, which corresponds to chromosome 3 in the *Vicia faba* consensus map (Webb et al. 2016). This confirms at the molecular level that the two low tannin genes are located in two different chromosomes, since *zt1* was mapped in chromosome 2 of faba bean by Webb et al. (2016). It also affirms the potential of comparative genomics and synteny analysis using model or closely related species to create new molecular tools for an "orphan" crop like faba bean for which the genetic and genomic resources are still limited (Torres et al. 2010, Cruz-Izquerdo et al. 2012, Gnanasambandam et al. 2012, O'Sullivan and Angra 2016).

The three closest SNP markers were validated in a panel of diverse genotypes with different allelic state at the *zt2* locus. One of the markers successfully discriminated the genotypes with 100% accuracy, distinguishing those that are recessive for *zt2* from those carrying the dominant allele *ZT2*. A KASP marker was then designed for this SNP marker and was successfully validated in a second validation test. Markers in the KASP format are fast and easy to use and have been successfully used in faba bean for identifying QTLs for aschochyta blight

resistance (Kaur et al. 2014) and screening for low vicine-convicine concentration (Khazaei et al. 2017).

SNP markers are robust, bi-allelic and co-dominant in nature, highly abundant and amenable to high throughput genotyping, providing the ideal system for high-density genetic mapping (Kaur et al. 2014). Various highly saturated genetic maps were developed using SNPs in faba bean (Cottage et al. 2012, Kaur et al. 2014, Khazaei et al. 2014, Webb et al. 2016), and an ultra-high density SNP map for faba bean, available to use for the research community, will be developed in the near future (O'Sullivan and Angra 2016). The new KASP marker for the *zt2* gene is an important molecular tool for the screening of breeding material, and eventually for the wider introduction of this gene in faba bean varieties. In fact, most of the low tannin varieties now available for cultivation carry the *zt1* gene, and the incorporation of *zt2* in new varieties may lead to superior nutritional quality due to a significant decrease in fiber content in comparison to *zt1* types (Duc et al. 1999). Furthermore, effective molecular markers linked to both *zt* genes could be used for gene pyramiding for the selection of faba bean cultivars with improved nutritional value for human and animal consumption (Gnanasambandam et al. 2012). Considering the partial cross-pollination habit of faba bean, out-crossing of a low tannin variety and a wild type, or two low tannin varieties recessive for the two different genes will result in the production of visually distinguishable contaminated seed in the following generation. A variety recessive for both the low tannin genes could potentially reduce this problem and maintain for longer time the low tannin trait even if outcross with a normal tannin variety will happen. If a double recessive variety (*zt1zt1zt2zt2*) will experience some out-crossing with wild type pollen, some of the plants in the following generation will be heterozygous at both loci (genotype

ZT1zt1ZT2zt2). However, if in the following generations the variety can be maintained in isolation from further contamination, the plants within the population will cross pollinate among each-other. Eventually, both the undesired dominant alleles (*ZT1* and *ZT2*) will be reduced in frequency. Furthermore, genotypes of a variety recessive for both *zt1* and *zt2* will not lose the low tannin phenotype in the following generation if they will be pollinated by *zt1* or *zt2* genotypes. On the contrary, this will happen if the cross-pollination occurs between the two low tannin varieties recessive for only one of the two low tannin genes, one for *zt1* and the other for *zt2.*

On the other hand, plant breeders may be interested in eliminating one of the two low tannin genes from their breeding programs. For instance, if one of the two genes is found to be associated to better nutritional or agronomic characteristics. Having low tannin varieties with only one of the two recessive genes will also avoid the problem of cross-pollination between low tannin genotypes resulting in coloured flowers. Plant breeders will benefit from the availability of reliable molecular markers for the two genes *zt1* and *zt2* in order to screen their germplasm and select in favor of only one of the two genes. However, more research is needed to characterize all the three low tannin genotypes (*zt1, zt2* and *zt1zt2*) for nutritional, agronomic and possibly other characteristics.

The second objective of this research project was the identification of the genotype recessive for both low tannin genes *zt1* and *zt2* in a population originating from the initial cross between CDC Snowdrop (*zt1*) and Disco/2 (*zt2*). Molecular markers in the form of a SCAR marker for the *zt1* gene and the new KASP marker (described in Chapter 3) for the *zt2* gene were used to genotype the $F_{2:3}$ derived individuals of the segregating population. The KASP marker for the

zt2 gene was successful in distinguishing genotypes carrying the recessive allele (*zt2*), while SCAR marker SCC_{5551} is partially effective in tagging the *zt1* gene, and this could be used in combination with the *zt2* gene KASP marker to identify genotypes recessive at both loci. However, a more reliable marker in the form of KASP assay for tagging the *zt1* gene would be an even more desirable tool for plant breeder. Webb et al. (2016) mapped the *zt1* gene using a population at F2 segregating for flower colour. The genomic region where the gene was mapped was colinear with a portion of *Medicago truncatula* Chromosome 3. A candidate gene was found in that genomic region of the model species, corresponding to the *Medicago* WD-40 transcription factor TTG1 (Transparent Testa Glabra 1) which has a role in the regulation of *anthocyanins* biosynthesis. After cloning and sequencing of the coding region of the gene in the genomes of the parental lines of the mapping population, the study proved that a deletion in the 5' end of the sequence was responsible for the loss of flower colour in the white flower parent. Nonetheless, the two SNP markers flanking the flower colour (*zt1*) locus were not able to discriminate faba bean genotypes which are recessive for *zt1* from those that carry the dominant allele *ZT1* in a diversified validation panel. In this case, the two markers were polymorphic only between the parents of the mapping population used in that study, this being one of the main problems for the conversion of SNPs found in a bi-parental population, in valid markers to be used across genotypes with different genetic backgrounds (Mammadov et al. 2012).

The development and implementation of effective molecular tools are particularly useful for distinguishing genotypes recessive for both low tannin genes, since they both present the same phenotype (white flower). Projects for the sequencing the faba bean genome are under way (NORFAB project), and in the near future a complete reference genome for faba bean will

be available to use by scientists and breeders. It will soon be possible to fine map and to identify candidate genes for both low tannin genes. The molecular markers currently available for tagging *zt1* and *zt2* are still limited by margin of errors, since none of them are located within the gene sequences. Therefore, the availability of other phenotypical markers linked to the phenolic composition of the seed coats and flowers are useful for studying the mode of action of the two genes, and also as biochemical markers for the screening of germplasm.

Qualitative and quantitative LC-MS analysis of the phenolic profiles in the seed coats and flowers of faba bean lines with contrasting allelic states at the *zt* genes were helpful to enhance our knowledge on this topic. This is the first time that an elevated number of known chemicals species (over 100 analyzed compounds) belonging to various sub classes of phenolic compounds was studied in faba bean. Furthermore, some unknown compounds were characterized for the first time in the flowers of faba bean. The results indicated a possible point of action for the two *zt* genes in the *flavonoids* biosynthetic pathway, and more interestingly, showed clear differences between them. Some compounds clearly distinguished the profiles of *zt1* from those of *zt2* both in the flowers and the seed coats. An example is isorhamnentin-3-O-rutinoside which was present in the flowers of Disco/2 but not in the flowers of CDC Snowdrop. This compound could be effectively used as biochemical marker for germplasm screening to discriminate *zt1* genotypes from *zt2* genotypes. From this perspective, biochemical markers in the flowers identified with LC-MS analysis, represent a powerful alternative to the faster, but less precise, molecular markers which are currently available.

Flowers are abundant in the plant, can be harvested at the time of early flowering, and the samples are easier to prepare compared to seed coat tissues, since they do not require the

time consuming dehulling step. Furthermore, LC-MS analysis could be also used in the future for generating phenotypic data from RIL populations segregating for flower colour. This data, combined with higher SNPs coverage of the faba bean genome could be used to narrow down the search for candidate genes for *zt1* and *zt2*. For example, Albert et al. (2014) used the flowers of *Petunia hybrida* to study the network of transcriptional activators and repressors which controls anthocyanin pigmentation.

6.2 Future work

This study was successful in identifying and validating a molecular marker for the *zt2* gene. The logical follow-up is to fine map the population Disco/2 × ILB 938/2 with a much higher number of SNP markers, which will be soon available through our collaborations (O'Sullivan and Angra, 2016, NORFAB project). Ideally, it will be possible to find a candidate gene for *zt2* and to identify a polymorphism within the coding region of the gene which can be translated in a highly precise molecular marker.

Another approach would be to clone and amplify the sequences of the most probable candidate genes found in the genomic region *Medicago truncatula* that is collinear to the faba bean region where *zt2* was mapped. The amplification of this genes through PCR in faba bean lines with contrasting allelic state at the *zt2* locus, and the following alignment of these sequences should yield candidate SNPs within the gene sequences. This approach was successfully used by Khazaei et al. (2017) to identify SNP markers for low vicine-convicine. In this case, the SNPs were identified starting from candidate EST sequences associated with the v-c synthetic pathway, reported by Ray et al. (2015).

The genomic region within the closest flanking markers for *zt2* comprises several genes with characterized function in the model species *Medicago truncatula*. Among these genes, several code for WD40 repeat proteins which were involved in the regulation of *anthocyanins* biosynthesis in different species (Walker et al. 1999, Pang et al. 2009, Li et al. 2015, Webb et al. 2016).

The other priority for the establishment of molecular tools for the study and selection of the low tannin trait in faba bean, is the development of a strong and reliable molecular marker for *zt1*. The availability of reliable KASP markers for both of the low tannin genes, will also allow an easy identification of the double recessive genotype, *zt1zt2.* Then the molecular results could be combined with LC-MS analysis of the seed coats of this genotype. These results will very likely help to distinguish the genotype recessive at both loci from those recessive only for *zt1* or *zt2*, and possibly find out improved nutritional characteristics related to the presence of both recessive genes.

High resolution LC/MS analyses were an effective tool for the detection of unknown phenolic compounds. However, the process was slow and time consuming, since it had to be partially done manually, comparing the MS/MS spectra originated from our analysis with those already published in the literature. The availability of a software system with the capability of conducting rapid, precise and automatic search in a large and updated MS/MS data base is a desirable resource which will allow the analysis of the large amount of data such as those generated from RIL populations.

Polyphenols are not only antinutritional compounds, and they have many beneficial effects for human health such as antioxidant activity and anti-disease properties (Martin and

Appel, 2010). From this perspective the identification of new phenolic compounds, their absolute quantification in the various seed fractions, the characterization of their biological activity and understanding of their genetic control at the genomic level, could lead to the breeding of faba bean varieties with targeted phenolic profiles. Hart et al. (2015) found that different polyphenols present in the seed coats of common bean either promote or decreased the uptake of iron by Caco-2 cells. This suggests that other positive nutritional or agronomic characteristics may be linked to the presence or absence of specific phenolics.

Nowadays, one of the main challenges for agriculture is to attain the so-called nutritional security (Keatinge et al. 2011) intended as the guarantee for every human being to access a fully nutritious diet. In this perspective, the research leading to the enhancement of the content and availability of micronutrients in staple foods such as faba bean, is becoming a priority. And a deeper study of secondary metabolites such as polyphenols will become even more significant.

The development of nutritionally improved faba bean varieties, also represents a new opportunity to increase the utilization of a crop which has not been exploited yet in the global scale. Increased use of faba bean in appropriate agricultural ecosystems will provide new opportunities for environmental, ecological and economic benefits.

LITERATURE CITED

Aerts, R. J., Barry, T. N., McNabb, W. C. (1999). Polyphenols and agriculture: beneficial effects of proanthocyanidins in forages. Agriculture, Ecosystems and Environment. 75: 1–12.

Agarwal, M., Shrivastava, N., Padh. H., (2008). Advances in molecular marker techniques and their applications in plant sciences. Plant Cell. Rep. 27: 617–631.

Aguirre-Hernández, E. A., González-Trujanoa, M. E., Martínez, A. L. Morenoa, J., Kitec, G., Terrazas, T., Soto-Hernández, M., (2010). HPLC/MS analysis and anxiolytic-like effect of quercetin and kaempferol flavonoids from *Tilia americana* var. *mexicana*. (2010). Journal of Ethnopharmacology. 127: 91–97.

Albert, N.W., Davies, K.M., Lewis, D.H., Zhang, H., Montefiori, M., Brendolise, C., Boase, M.R., Ngo, H., Jameson, P.E., Schwinna, K.E., (2014). A conserved network of transcriptional activators and repressors regulates anthocyanin pigmentation in eudicots. Plant Cell 26, 962–980. Aw, T. L., swanson, B. G. (1985). Influence of Tannin on Phaseolus vulgaris Protein Digestibility and Quality. Journal of food science. 50: 67-71.

Baginsky, C., Peña-Neira, A., Ca´ ceres, A., Herna´ndez, T., Estrella, I., Morales, H., Pertuze´, R. (2013). Phenolic compound composition in immature seeds of fava bean (Vicia faba L.) varieties cultivated in Chile. Journal of Food Composition and Analysis. 31: 1–6.

Bate-Smith, E. C. (1953). Colour Reactions of Flowers attributed to (a) Flavanols and (b) Carotenoid Oxides. *Journal of Experimental Botany.* 10: 1-9.

Bate-Smith, E.C. (1973). HAEMANALYSIS OF TANNINS: THE CONCEPT OF RELATIVE ASTRINGENCY. Phytochemistry. 12: 907-912.

Bekkara, F., Jay, M., Viricel, M. R., Rome, S. (1998). Distribution of phenolic compounds within seed and seedlings of two *Vicia faba* cvs differing in their seed tannin content, and study of their seed and root phenolic exudations. *Plant and Soil.* 203:27–36.

Butler, L. G., Price, M. L., Brotherton, J. E., (1982). Vanillin Assay for Proanthocyanidins (Condensed Tannins): Modification of the Solvent for Estimation of the Degree of Polymerization. *J. Agric. Food Chem.* 30: 1087-1089.

Caldas, V. G., Blair, M. W. (2009). Inheritance of seed condensed tannins and their relationship with seed-coat colour and pattern genes in common bean (Phaseolus vulgaris L.). Theor. Appl. Genet (2009), 119: 131–142.

Canadian Seed Growers' Association. (2018). FOUNDATION, REGISTERED AND CERTIFIED PRODUCTION OF BEAN, CHICKPEA, FABABEAN, LENTIL, LUPIN, PEA, AND SOYBEAN. CIRCULAR 6 / Rev.02.01-2018. [online] Available at: http://seedgrowers.ca/tag/fababean/ [Accessed 07/04/18].

Chaieb, N., González, J. L., López-Mesas, M., Bouslama, M., Valiente, M. (2011). Polyphenols content and antioxidant capacity of thirteen faba bean (Vicia faba L.) genotypes cultivated in Tunisia. Food Research International. 44: 970–977.

Cheema, J., Dicks, J., Computational approaches and software tools for genetic linkage map estimation in plants. (2010). Briefings in Bioinformatics. 6: 595-608.

Chen, Y., Yu, H., Wu, H., Pan, Y., Wang, K., Jin, Y., Zhang, C. (2015). Characterization and Quantification by LC-MS/MS of the Chemical Components of the Heating Products of the Flavonoids Extract in Pollen Typhae for Transformation Rule Exploration. Molecules. 20: 18352-18366.

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

Cooper, J. W., Wilson, M. H., Derks, M.F.L., Smit, S., Kunert, K. J. Cullis, C., Foyer, C. (2017). Enhancing faba bean (*Vicia faba* L.) genome resources. Journal of Experimental Botany. 68: 1941–1953.

Cottage, A., Gostkiewicz, K., Thomas, J. E., Borrows, R., Torres, A. M., O'Sullivan, D. M. (2012). Heterozygosity and diversity analysis using mapped single nucelotide polymorphisms in a faba bean inbreeding program. Mol Breeding. 30: 1799–1809.

Crėpon, K., Marget, P., Peyronnet, P., Carroue´e, B., Arese, P., Duc, G. (2010). Nutritional value of faba bean (Vicia faba L.) seeds for feed and food. Field Crops Research. 115: 329–339.

Crofton, G. R. A., Bond, D. A., Duc, G., (2000). Potential seed multiplication problems arising from the existence of two genes for the absence of tannin in *Vicia faba* L. Plant Var. Seeds. 13: 131-139.

Crofts, H. J., Eveus, L. E. and McVetty, P. B. E., (1980). Inheritance, characterization and selection of tannin-free faba beans (Vicia faba L.). Can. J. Plant Sci. 60: 1135-1140.

Cruz-Izquierdo, S., Avila, C. M., Satovic, Z., Palomino, C., Gutierrez, N., Ellwood, S. R., Phan, H. T. T., Cubero, J. I., Torres, A. M. (2012). Comparative genomics to bridge Vicia faba with model and closely-related legume species: stability of QTLs for flowering and yield-related traits. Theor. Appl. Genet. 125: 1767–1782.

Duc, G., Brun, R. M., Jay, M., (1995). Genetic variation in tannin-related characters of faba-bean *seeds (Vicia faba* L.) and their relationship to seed-coat colour. Plant Breeding. 114: 272-274.

Duc, G. (1997). Faba bean (*Vicia faba* L.). Field Crops Research. 53: 99-109.

Duc, G., Margaret, P., Esnault, R., Le Guen, J., Bastianelli, D. (1999). Genetic variability for feeding value of faba bean seeds (*Vicia faba*): Comparative chemical composition of isogenics involving zero-tannin and zero-vicine genes. Journal of Agricultural Science. 133: 185-196.

Duc, g., Bao, S., Baum, M., Redden, B., Sadiki, M., Suso, M. J., Vishniakova, M., Zong, X. (2010). Diversity maintenance and use of Vicia faba L. genetic resources. Field Crops Research. 115: 270–278.

El-Aneed, A., Cohen, A., Banoud, J., (2009). Mass Spectrometry, Review of the Basics: Electrospray, MALDI, and Commonly Used Mass Analyzers. *Applied Spectroscopy Reviews.* 44: 210–230.

Ellwood, S. R., Phan, H. T, Jordan, M., Hane, J., Torres, A. M., Avila, C. M., Cruz-Izquierdo, Olivier, R. P. (2008). Construction of a comparative genetic map in faba bean (*Vicia faba* L.); conservation of genome structure with *Lens culinaris. BMC Genomics.* 380.

Erith, A. G. (1930). The inheritance of colour, size, form of seeds and of flower colour in Vicia faba L. Genetica. 13: 477–510.

FAOSTAT, (2018). 2016 World Production Statistics for Faba bean. Food and Agricultural Organization of the United Nations, Statistics Division. [online]. Available at: http://www.fao.org/ [Accessed 07/04/18].

Gil, J., Cubero, I. (1993). Inheritance of Seed Coat Thickness in Chickpea (*Cicer arietinum* L.) and its Evolutionary Implications. Plant Breeding. 111: 257—260.

Gnanasambandam, A., Paull, J., Torres, A., Kaur, S., Leonforte, T., Li, H., Zong, S., Yang, T., Materne, M. (2012). Impact of Molecular Technologies on Faba Bean (*Vicia faba* L.) Breeding Strategies. 2: 132-166

Government of Saskatchewan. (2018). Faba bean: Overview. [online]. Available at: https://www.saskatchewan.ca/business/agriculture-natural-resources [Accessed 07/04018].

Gunawardena, C. K., Zijlstra, R. T., Beltranena, E. (2014). Characterization of the nutritional value of air-classified protein and starch fractions of field pea and zero-tannin faba bean in grower pigs. J. Anim. Sci. 88: 660–670.

Gutierrez, N., Avila, C. M., Rodriguez-Suarez, C., Moreno, M.T., Torres, A. M. (2006). Development of SCAR markers linked to a gene controlling absence of tannins in faba bean. Mol. Breeding. 19: 305–314.

Gutierrez, N., Avila, C. M., Moreno, M.T., Torres, A. M. (2008). Development of SCAR markers linked to *zt-2*, one of the genes controlling absence of tannins in faba bean. Australian Journal of Agricultural Research. 59: 62–68.

Hart, J. J., Tako, E., Kochian, L. V., Glahn, R. P. (2015). Identification of Black Bean (Phaseolus vulgaris L.) Polyphenols That Inhibit and Promote Iron Uptake by Caco‐2 Cells. J. Agric. Food Chem. 63: 5950−5956.

He, F., Pan, Q., Shi, Y., Duan, C. (2008). Biosynthesis and Genetic Regulation of Proanthocyanidins in Plants. Molecules. 13: 2674-2703.

Hellens, R.P., Moreau, C., Lin-Wang, K., Schwinn, K.E., Thomson, S.J., Fiers, M.W.E.J., Frew, T.J., Murray, S.R., Hofer, J.M.I., Jacobs, J.M.E., Davies, K.M., Allan, A.C., Bendahmane, A., Coyne, C.J., Timmerman-Vaughan, G.M., Ellis, T.H.N., (2010). Identification of Mendel's white flower character. PLoS ONE. 5: e13230.

Helsper, J. P. F. G., Hoogendijk, J. M., van Norel, A., Burger-Meyer, K. (1993). Antinutritional Factors in Faba Beans *(Vicia faba* L.) As Affected by Breeding toward the Absence of Condensed Tannins. J. Agric. Food Chem. 41: 1058-1061.

Kaetingue, J. D. H., Yang, R. Y., Huges, J. d'A., Easdown, W. J., Holmer, R. (2011). The importance of vegetables in ensuring both food and nutritional security in attainment of the Millennium Development Goals. Food Sec. 3: 491–501.

Karaca, A. C., Low, N., Nickerson, M. (2011). Emulsifying properties of chickpea, faba bean, lentil and pea proteins produced by isoelectric precipitation and salt extraction. Food Research International. 44: 2742–2750.

Kaur, S., Kimber, R. B. E., Cogan, N. O. I., Materne, M., Forster, J. W. Paull, G. J. (2014). SNP discovery and high-density genetic mapping in faba bean (Viciafaba L.) permits identification of QTLs for ascochyta blight resistance. Plant Science. 218: 47-55.

Khan, H.R., Paull, J.G. Siddique, K.H.M., Stoddard, F.L. (2010). Faba bean breeding for drought-affected environments: A physiological and agronomic perspective. Field Crops Research 115: 279–286.

Khazaei, H., O'Sullivan, D. M., Sillanpää, M. J., Stoddard, F.L. (2014). Use of synteny to identify candidate genes underlying QTL controlling stomatal traits in faba bean (Vicia faba L.). Theor Appl Genet. 127: 2371–2385.

Khazaei, H., O'Sullivan, D. M., Sillanpä, M. J., Stoddard, F. L. (2014). Genetic analysis reveals a novel locus in Vicia faba decoupling pigmentation in the flower from that in the extra-floral nectaries. Mol. Breeding. 34: 1507–1513.

Khazaei, H. (2014). Leaf traits associated with drought adaptation in faba bean (*Vicia faba* L.). Doctoral Thesis.

Khazaei, H., Purves, R. W., Song, M., Stonehouse, R., Bett, K. E., Stoddard, F. L., Vandemberg, A. (2017). Development and validation of a robust, breeder-friendly molecular marker for the vc-locus in faba bean. Mol. Breeding. 37: 140.

Köpke, U., Nemecek, T., (2010). Ecological services of faba bean. Field Crops Research. 115: 217–233.

Kosambi, d. d. (1943). The estimation of map distances from recombination values. Annals of human genetics (1943).

Li, P., Chen, B., Zhang, G., Chen, L., Dong, Q., Wen, J., Mysore, K.S., Zhao, J., (2016). Regulation of anthocyanin and proanthocyanidin biosynthesis by *Medicago truncatula* bHLH transcription factor MtTT8. New Phytol. 3: 905-921.

Lorieux, M., (2012). MapDisto: fast and efficient computation of genetic linkage maps. Mol Breeding. 30: 1231–1235.

Lu, N., Marissa Roldan, M., Dixon, A. (2017). Characterization of two TT2-type MYB transcription factors regulating proanthocyanidin biosynthesis in tetraploid cotton, Gossypium hirsutum. Planta
DOI 10.1007/s00425-017-2682-z

Mammodav, J., Aggarwal, R., Buyyarapu, R., Kumpatla, S. (2012). SNP Markers and Their Impact on Plant Breeding. International Journal of Plant Genomics. Article ID 728398.

Mangan, J. L., (1988). NUTRITIONAL EFFECTS OF TANNINS IN ANIMAL FEEDS. Nutrition Research Reviews. 1: 209-231.

Martin, K.R., Appel, C.L., (2010). Polyphenols as dietary supplements: A double-edged sword. Nutr. Diet. 2: 1-12.

Makkar, H. P. S., Becker, K. (1993). Vanillic-hcl method for condensed tannins: effect of organic solvents used for extraction of tannins. Journal of Chemical Ecology. 19: 613-621.

Mendel, J.G., (1865). "Versuche über Pflanzenhybriden", Verhandlungen des naturforschenden Vereines in Brünn, Bd. IV für das Jahr. Abhandlungen, 3-47.

Merghem, R., Jay, M., Brun, N., Voirin, B., (2004). Qualitative Analysis and HPLC Isolation and Identification of Procyanidins from *Vicia faba. Phytochem. Anal*. 15: 95–99.

Mirali, M., Ambrose, S. J., Wood, S. J. Vandenberg, A., Purves, R. W. (2014). Development of a fast extraction method and optimization of liquid chromatography–mass spectrometry for the analysis of phenolic compounds in lentil seed coats. Journal of Chromatography B. 969: 149–161.

Mirali, M., Purves, R. W., Stonehouse, R. S., Bett, K., Vandemberg, A. (2016). Genetics and Biochemistry of Zero-Tannin Lentils. PLoS ONE 11(10): e0164624.

Mikić, A., Mihailovic´, V., Cupina, B., Dordević, V., Milić, D., Duc, G., Stoddard, F. L., Lejeune-Hėnaut, I. Marget, P., Hanocq, E. (2011). Achievements in breeding autumn-sown annual legumes for temperate regions with emphasis on the continental Balkans. Euphytica.180: 57–67.

Mole, S., Waterman, P. G. (1987) A critical analysis of techniques for measuring tannins in ecological studies. Oecologia. 72: 137-147.

Montero, L., Herrero, M., Prodanov, M., Ibáñez, E., Cifuentes, A. (2013). Characterization of grape seed procyanidins by comprehensive two-dimensional hydrophilic interaction × reversed phase liquid chromatography coupled to diode array detection and tandem mass spectrometry. Anal Bioanal Chem. 405: 4627–4638.

Multari, S., Stewart, D., Russell, W. R. (2015). Potential of Fava Bean as Future Protein Supply to Partially Replace Meat Intake in the Human Diet. Comprehensive Reviews in Food Science and Food Safety. 14: 511-522.

Nasar-Abbas, S.M., Siddique, K.H.M., Plummer, J.A., White, P.F., Harris, P.F., Dods, K., D'Antuono, M. (2009). Faba bean (Vicia faba L.) seeds darken rapidly and phenolic content falls when stored at higher temperature, moisture and light intensity. Food Science and Technology 42: 1703–1711.

NORFAB: Protein for the Northern Hemisphere. Research project, (2016-2021).

Nozzolillo, C., Ricciardi, L., Lattanzio, V. (1998). Flavonoid constituents of seed coats of *Vicia* faba (Fabaceae) in relation to genetic control of their colour. CAN. J. BOT. 67: 1600-1604.

O'Sullivan, D. M., Angra, D. (2016). Advances in Faba Bean Genetics and Genomics. Frontiers in Genetics. 7: 150.

Pang, Y., Wenger, J. P., Saathoff, K., Peel, G. J., Wen, J., Huhman, D., Allen S. N., Tang, Y., Cheng, X., Tadage, M., Ratet, P., Mysore, K. S., Summer, L. W., Marks, M. D., Dixon, R. A. (2009). A WD40 Repeat Protein from Medicago truncatula Is Necessary for Tissue-Specific Anthocyanin and Proanthocyanidin Biosynthesis But Not for Trichome Development. Plant Physiology. 151: 1114–1129

Phelps, S., (2017). Faba bean agronomy. [online]. Available at http://saskpulse.com/files/general/170217_ Faba_Bean_Agronomy.pdf. [Accessed 07/04/18].

Petitot, M., Boyer, L., Minier, C., Micard. V. (2010). Fortification of pasta with split pea and faba bean flours: Pasta processing and quality evaluation. Food Research International 43: 634–641.

Picard, J., (1976). Aperçu sur l'hérédité du caractère absence de tanins dans les graines de féverole (*Vicia faba* L.). Ann. Amelior. Plantes. 26: 101-106.

Ray, H., Bock, C., Georges, F. (2015). Faba bean: transcriptome analysis from etiolated seedling and developing seed coat of key cultivars for synthesis of proanthocyanidins, phytate, raffinose family oligosaccharides, vicine, and convicine. Plant Genome (2015) 8, 1. https : //doi.org/138 35/plantgenome2014.07.002

Semagn, k., Babu, R., Hearne, S., Olsen, M., (2014). Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement. Mol. Breeding. 33: 1–14.

Sibul, F., Orćić, D., Vasić, M., Anaćkov, G., Nadpal, J., Savić, A., Mimica-Dukić, N. (2016). Phenolic profile, antioxidant and anti-inflammatory potential of herband root extracts of seven selected legumes. Industrial Crops and Products. 83: 641–653.

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

Singleton, V. L., Rossi, J. A. (1965). Colourimetry of Total Phenolics with Phosphomolybdic-Phosphotungstic Acid Reagents. Am. Jour. Enol. Vitic. 16: 144-158.

Tanaka, Y., Sasaki, N., Ohmiya, A., Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids. The plant journal. 54: 733-749.

Torres, A.M., Avila, C.M., Gutierrez, N., Palomino, C., Moreno, M.T., and Cubero, J.I. (2010). Marker-assisted selection in faba bean (*Viciafaba* L.). Field Crop Res. 115: 243–252.

Truong, V., Deighton, N., Thompson, R. T., Mcfeeters, R. F., Dean, L. O., Pecota, K.V., Yencho, G. C. (2010). Characterization of Anthocyanins and Anthocyanidins in Purple-Fleshed Sweetpotatoes by HPLC-DAD/ESI-MS/MS. J. Agric. Food Chem. 58: 404–410.

Vaillancourt, R., Slinkard, A. E. Reichert, R. D. (1985). THE INHERITANCE OF CONDENSED TANNIN CONCENTRATION IN LENTIL. Can. J. Plant Sci. 66: 241-246.

Van der Poel, A.F.B. Gravendeel, S., Van Kleef, D.J. Jansma, h.J.M., Kemp, B. (1992). Tannin-containing faba beans (*Vicia fabu* L.): effects of methods of processing on ileal digestibility of protein and starch for growing pigs. Animal Feed Science and Technology. 36: 205-214.

Vermeris, W., Nicholson, R., (2006). *Phenolic Compound Biochemistry*. Springer, Dordrecht, The Netherlands.

Vilariño, M., Métayer, J.P. Crépon, K. Duc, G. (2009). Effects of varying vicine, convicine and tannin contents of faba bean seeds (*Vicia faba* L.) on nutritional values for broiler chicken. Animal Feed Science and Technology. 150: 114–121

Voorrips, R. E., (2002). MapChart: Software for the Graphical Presentation of Linkage Maps and QTLs. *Journal of Heredity*. 93: 77–78.

Walker, A. R., Davison, P. A., Bolognesi-Winfield, A. C., James, M., Srinivasan, N., Blundell. T. L., Esch, J.J., Marks, M. D., Gray, J. C. (1999), The *TRANSPARENT TESTA GLABRA1* Locus, Which Regulates Trichome Differentiation and Anthocyanin Biosynthesis in Arabidopsis, Encodes a WD40 Repeat Protein. The Plant Cell. 11: 1337–1349.

Webb, A., Cottage, A., Wood, T., Khamassi, K., Hobbs, D., Gostkiewicz, K., White, M., Khazaei, H., Ali, M., Street, D., Duc, G., Stoddard, F.L., Maalouf. F., Ogbonnaya, C., Link, W., Thomas, J. and O'Sullivan, D. M., (2016). A SNP-based consensus genetic map for synteny-based trait targeting in faba bean (Vicia faba L.). Plant Biotechnology Journal. 14: 177–185.

Yuste, P., Longstaff, M., McCorquodale, C., (1992), The effect of proanthocyanidin-rich hulls and proanthocyanidin extracts from bean (Vicia faba L.) hulls on nutrient digestibility and digestive enzyme activities in young chicks. British Journal of Nutrition. 67: 57-65.

APPENDICES

Appendix 1.

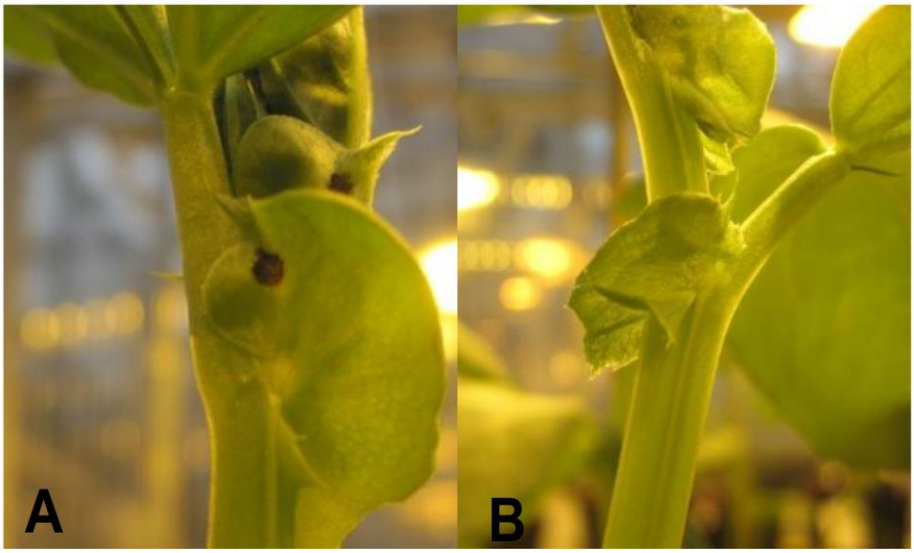
Supplementary table 3.1. F1 seeds and accessions used for the validation of KASP marker (validation panel 2). The crosses that originated the F1 seeds and faba bean accessions used as validation for the KASP marker designed from the SNP marker Vf_Mt7g100500_001, the number of seeds per entry, the known allelic state at *zt2* and the result of the KASP assay which proves 100% accuracy of the KASP marker.

Cross/Accession	Number of seeds	Predicted allele at *zt2*	Result of KASP assay
CDC Snowdrop × Disco/2	3	*Zt2zt2*	heterozigous *
GLORIA × DISCO/2	4	*Zt2zt2*	heterozigous
AF11212 ×DISCO/2	1	*Zt2zt2*	heterozigous
P47-1 ×DISCO/2	2	*Zt2zt2*	heterozigous
CDC Snowdrop ×TW	2	*Zt2zt2*	heterozigous
CDC Snowdrop ×A011550	1	*Zt2Zt2*	allele 1
AE11212	1	*Zt2Zt2*	allele 1
ILB 938/2	4	*Zt2Zt2*	allele 1
CDC Snowdrop	15	*Zt2Zt2*	allele 1
Disco/2	9	*zt2zt2*	allele 2

*Allele 1= *ZT2*, Allele 2= *zt2*, Heterozygous= *ZT2zt2*.

Appendix 2.

Supplementary Figure 3.1. a) Coloured (A) and colourless (B) stipule spot pigmentation in faba bean.



Supplementary Figure 3.1. b) Black (left) and colourless (right) seed hilum in faba bean.

Appendix 3.

Supplementary table 3.2. SNPs marker polymorphic between Disco/2 and ILB 938/2 used to construct the linkage map. (Excel file attached at the end of the document)

DNA \ Assay	SNP ID	Disco/2 allele	ILB938/2 allele	Sequence
AnMtS37SNP		C:C	T:T	
BGALSNP	1008056105	C:C	T:T	CGGTCGTCATTGGCC[C/T]GGATATRTTGCTCAT
CNGC4	1840314704	A:A	C:C	CTCTGTTTTTTCTTT[A/C]TTCACTTAACTATAC
CTPSNP		A:A	G:G	
GLIP089	1335262551	C:C	T:T	ATGCCCAATAGTTCT[C/T]ACTCATGAACATTTG
GLIP139SNP	1633844175	T:T	A:A	GGGGTTGATTTAGAT[T/A]AATTTTTATCACTTG
GLIP253SNP		C:C	T:T	
GLPSNP	1038965406	C:C	G:G	TAACACTCTTTTCAC[C/G]AAAGTATTGAACAAA
LG018SNP	1735570684	T:T	C:C	ACAAAATAATTTAAC[C/T]GAGAATCTTACCTTC
LG038SNP	1362883324	A:A	G:G	GCCTGTAAGACCCTC[A/G]TGCTTACTGAATATA
New_GLIP307SNP	1525766649	A:A	T:T	AAGTGGGAAGAAGTT[A/T]NTTTTTTTTTNNNGA
PGDH		T:T	A:A	
PRATSNP	1387244519	T:T	C:C	GACGGTGCTGGTGCA[C/T]ATTCTTTTTCTGATG
RBPC_0SNP	1845304182	G:G	A:A	GGGTGGWACWACCCA[A/G]TACACTGTCAACAAC
Vf_Mt1g007370_001	1724441631	T:T	C:C	AGGTAGGGTATATAGGTGTGCTCCAAGGAGCCGAGC GATTCGACAGTACAAGAGGACA[Y]GGATTCTCAACTT ATGTGCAATACTGGATAAGGAAGTCAATGTCAACAAT GGTGGCAAAATATGCTCGAGGAGTCACCGTTCCT
Vf_Mt1g007490_001	1225254442	T:T	C:C	AGGTTCAAGTAATTGGTCTTCTATGTAGAACTCCTAAA AAGGAAAGCAGATATGAATTGTTACGCCGAGTAGCAG CTGGTGG[Y]GGTTCGTTTAAAGGCGAGAATGACTTGA AGGTTCATATCCCTGGGGCAAATTTAAATGACATAGCT AACCAAGCTGATGATTTATT
Vf_Mt1g012440_001	1053155908	C:C	A:A	TTATCCAAGATGTCCTGAACCAGCCATTCTTCAAAATC GACGTGCTTAACAAGCTTGTGAAGGAGTGTGAGATGA TGTTAAGTATAATTTTCCCC[M]AAAACGGACCTCTAG GCCAATCCTTGTCAACCAGCGAGGTTTTTGAAGAAGTT GCGYGTGGGTCGACTGCATACGAAACTAAAGAGACGC TRGAGCATGT
Vf_Mt1g013400_001	1269276193	G:G	C:C	GCAAGAATAACCTCTTCGCACGCAATATTGCTATTGTT TCTTCTTCAACTTTTCGTTTTTTGGTTAAACAA[S]CTGT ATCTGAGAACCTTGAATATTCAACCTGTGATATGGGCT ACCTCAATTCTGTTTTGTCTTCTTCAAGTCAGGTTCATG CTGCTGATGATTCACCTGT
Vf_Mt1g014230_001	1246549645	T:T	A:A	TGCTTACTTGCCGTATGACCGAGAGTGGATTAAGCAA CAGACCTTTCAACATTTGAAGAAATTGGCTCATTGAT[W]GAACCCCAAAGTTTTAGCAAATAGATGTAAAGGAA ATCACCAAGGGAAGAACTTATTTGTTATGTGGGCTATC CTTTGATGGCTATGTTACTTAATATGA
Vf_Mt1g018180_001	1977365618	T:T	G:G	GAAAATCGAATCAAAATCGCGAAAGCCGGCGCGATTA AACCTTTAATTTCTCTAGTATC[K]TCGCAAGATCTTCAG CTGCAAGAATACGGCGTCACTGCGATTTTGAATCTCTC TTTATGCGACGAGAATAAAGAGCTTATAGCTTCCGCC GGAGCGATTA
Vf_Mt1g031650_001	1285035944	T:T	C:C	TAGACTTGAAACCATATGCTCAAGCTATGCAAATGAA GTTGGCCACTCTTCCAGATGGATGTGCAATCAAGAGG GTTCAAGCAAA[Y]AGGACTAATTGTCTTGCTGAAGAG AGTAGTTGTATTTGGAAAARTGGTAAGATCATTCAACA RATGATTCTAAGATTGAGAAATGTTGAGAGTGGGGAA T
Vf_Mt1g044570_001	1353373995	C:C	T:T	ATGCCGGCTACAGCAGGTAGGGTTCGCATGCCCGCCA ACAATAGGGTTCACAGTAGTGCAGCCCTTCAAACACA CGGCATATGGCARAGTGC[Y]ATTGGCTATGACCCTTAT GCACCCAATAAGGAAGACAAGGACACTTCTCAAAGTC

				AGCCAAGTACRGAACYTGATGCAGAGAATGCATATGC CAGCTTCC
Vf_Mt1g050730_001	1721968277	T:T	C:C	CAGCAGCATGATAATGAAATTTCAGACTTCCCTTCCAA ATTTGTWTTCTGGTGTCTGAGTGCGATTGAAAA[Y]GC ACTGCGACATGATGATGCCTACACTGAGGGAGAAGGC AACTCTTTCTTTCTTGACCCGTGGGGATTAGAATTTTCA AAGTGTTTTNCTACTGGTAAAGAT
Vf_Mt1g061800_001	1845254471	T:T	C:C	GGTATTCTCATTTAAGCTCCGTGGAGCTTATAATATGA TGGCAAAACTTCCGGAGGAAGTCTTGGAGAAAGGGG TTATATGTTCTTCTGCTGGAAATCA[Y]GCTCAAGGAGT TGCCTTGTCTGCCAAGAGATTGAATTGTAAYGCTGTTA TTGCTATGCCCGTTACCACGCCTGACATCAAG
Vf_Mt1g072640_001	1038938738	T:T	C:C	TCAATGATCTTTTGATTATGTATGATGGCTTTAATTTCG ACGAGTGGCTAAATAAATGTAAAAAGA[Y]CATGAATG ATACATTTCCAAGGGAGGATCCATATAGCATTCTTGTT CCACCTGGATTTGAGTCATTTGACATAGAT
Vf_Mt1g072740_001	1947290245	G:G	A:A	AGATCCCTTATGTTGTGAAGCTTAACCCGGATGAGGA TAAGCAGAATGTTTTGTTGGCTGGTATGAGTGATAAG AAGATTGTTCAGTGGGATATGAATAC[R]GGGCAGATA ACTCAAGAGTATGATCAGCATTTAGGGGCTGTGAATA CCATTACTTTTGTTGATAACAATAGGAGATTTGTTACTT CAAGTGATGACAAGT
Vf_Mt1g082210_001	1192300098	A:A	G:G	TCAACCGTGACTTGGCGGAGTTTCTTGATCGGAAAGTT CCGGCGAATGCAAACCCTGTGGATGGAGTGTTTTCTTT TGATGTGATTGTTGACAGGGAAAC[R]AATCTTCTTACT CGAATTTATCGTCCCGTTGAGGGAGATGAAAGTGTGA ACATTGTTGATCTTGAGAAGCCTGTTACCTCTGAGGTT
Vf_Mt1g083460_001	1957189487	C:C	T:T	AGGAAGGTAGAATTGCTGTAGTTGTGGGTGCTGTAAC TGATGATATCCG[Y]GTGTACGAAGTTCCAGCCATCAA AGTTGTAGCACTTAGGTTTACAGAGACTGCAAGGGCA AGGATTGAAAAGGCTGGAGGAGAATGCTTGACATTC GATCTC
Vf_Mt1g088190_001	1495498589	G:G	A:A	CAGGTGGACTTTTGTTCAAGATGAGTGATAGTAACAT GCAGTATGTCACATCCACT[R]CCTTCTTACTGGTCACA TATGCCAAATACTTAACCAAATCACATTCAGTTGTGAA TTGTGGTGGYACTACTGTAACACCAAAGAGACTCAGA ACTTTAGC
Vf_Mt1g094870_001	1346638090	A:A	G:G	ATGTTAAATCCGCTGAGATACCTGATGGACCAAATGG AACAAAATTTTTCGCTATAACCTTGTTAACTCCACTCAG TAAAGGTGAATCAATAACACTTGA[R]GTGCTTTACTTA TTGACACATTCTCTGGAACCTTTCCCAGTTGAAATTAG TCAATCAGAGTCGCAGTTGGTCTATTTCCGTGATAGTG CAATATTATTGT
Vf_Mt1g099390_001	1243268575	A:A	G:G	TTTCTTCAAACACTATCTCTCTTCGTCCTTCGAAACCAC ACTCGTTTCAACTTTCTCTAAATTGGAAGCTTCTGTTTT ACGATTCTACGTCGTTCACGCTTTGCAAACCAATCGGA GGGATAAGGTTGTTGAGTTTTTCACTGTTCATGGTCCT GATTTG
Vf_Mt1g099390_002	1414137428	A:A	T:T	CAAACACTATCTCTCTTCGTCCTTCGAAACCACACTCGT TTCAACTTTCTCTARATTGGAAGCTTCKGTTTTACGATT CTACGTCGTTCACGCTTTGCAA[W]CCAATCGGAGGGA TAAGGTTGTTGAGTTTTTCACTGTTCATGGTCCTGATTT GTTGCAGAGTTCTCAGGAGGATTGGACTCAATGGTTT G
Vf_Mt1g101520_001	1942911040	C:C	A:A	AGCAAGTTGTTGGATCGGGTTATTTCAGTTGAATTTGC TGCTAGAGATGACGATGTCAGG[M]GAAATGGACATA GTCCGGAGAGAGGCCGTGATCGTCAGCGTGACAGAA GCCATGATGGAAGGCGATCACCCAGTCCTTATCGTAG AGAAAGGGGTAGTCC

Vf_Mt1g101840_001	1007727826	A:A	G:G	TCAAGGGNAGTTTAGGTCCAGCTAAAGGTGAAGCTTA CCGAGATAATGATCGAATCTCGGTGAACGATTCGGTT CTGGCTGATTCWATTTGGAATTCTGG[R]CTTAATAAA ATGTTTTCTGATTTAAGAATTCGGGGGAAATCGGCTGT AGGTTTGAATCCGAATATTAGATTATATAGGT
Vf_Mt1g102430_001	1493657837	G:G	A:A	TGGGACTATTTCGACGTCGGTGTTCCCTACCCTACCGG ATTCCACCCCGATTCGTACTCCGATATCATTCGCTGGC AGAGTTTCGTCCGCC[R]ACGCAGACGAAACTCGCTAT TCTGCTTYGYCGGAGCGCCACGCCGCACTTTCCGCRAT GATTTCAGGGCAATTCTGTTGAATCAATGCCGTRACTC
Vf_Mt1g105040_001	1923110801	G:G	T:T	TGCAGTCCAGCTTTATGCCAAAGAGTGCAGTAAGCTC CTCCTAGATGTGCTTAAACGAGAACCTAATAAGAACA GTGACAATGAAGCGGTGGTGCCTGAT[K]ATACCGTTG CTCCTCGKGAATCTGTTTTTGATATATCAAAAGGCAAA AGGGCTTTCATTGAAGCAGAGGAAGCTCGTGAACTTT TGAG
Vf_Mt1g115740_001	1923110801	A:A	G:G	TGCAGTCCAGCTTTATGCCAAAGAGTGCAGTAAGCTC CTCCTAGATGTGCTTAAACGAGAACCTAATAAGAACA GTGACAATGAAGCGGTGGTGCCTGAT[K]ATACCGTTG CTCCTCGKGAATCTGTTTTTGATATATCAAAAGGCAAA AGGGCTTTCATTGAAGCAGAGGAAGCTCGTGAACTTT TGAG
Vf_Mt2g005590_001	1435014068	A:A	G:G	CTTTCTGCTGATTTTATTCACCAGAAACACCTTTGGGAC TGAATCTCCTCAACCAGCTTGCAGGGATGGTCAATGG TTTATCC[R]CTTAAKTTCTCTCTTCTATTTTGGCGTTTAT TCTTTGGTGTATTAACCACAAGAGTATTGACCGTCAGT ATTGCT
Vf_Mt2g005900_001	1770605960	A:A	T:T	GCTATAAAGGCTGTTGCCACTCCAAATCCAGCTGTGG AATTGCCATTAACCGCTGAAAATGTAGAAACTGTGCTT GATGAAATTCG[W]CCGTATCTYATTTCTGATGGTGGA AATGTTGCTCTACATGAAATCGATGGCAATGTTGTACG GTTGAAGTTACAAGGTGCTTGTGGATCGTGTCCGAGT T
Vf_Mt2g007220_001	1242748400	C:C	T:T	GAATTCACTGGGGACTATGATGCTCTTCAACCCTTTCC AGAGATTGATTTTGTTGACTTTTTCAACAGCCATCCCA AGCTGCTCAAGTTTGATAT[Y]CATGGAGCCATGTTTGC AGCACTATGCCAGAAGAACAGTCTAAAACATGT
Vf_Mt2g008610_001	1093137577	C:C	T:T	GATCTTCCTTCTTATCTGGTGGAAAGAGGAGCCAAACC TTGCTAGCTACTGCCATTGGAAGCAAAGTTGGTGTTGC AGCTGCTAGTCCTAAAAGACTTAT[Y]GTTGCTGCTGCT GCTTCTGCACCAAAGAAATCATGGATCCCTGGTGTTAG ATTCGGTGGAAACCTCGTCGACCCCGAATGGCTCGAY GGATCG
Vf_Mt2g009320_001	1815267363	A:A	T:T	TTAACCTAGACAAGATGTTTTGGTCTAAACCCTGTTCT CTGGCCCTKCCTCGTGACTCGCCGCTAAGAGTTGA[W] GAACCTGATTATCAGGGGGTTAAGCGATTCATGCTCA AACTCATGCTGTTTTATAGTAAGCAGAGCAAATCGATT CGAGGGGCTAATGTTGTTTAC
Vf_Mt2g010740_001	1182404129	T:T	C:C	AGGTACAATGTCAGAACAGAAATTGTCCATCAAATAA GCTTGTTTTTCCATTGATGGCACTACTAGCAGGAATAT TAGGCGGTGTTTTCGGANNNNAT[Y]GGAGGTGGAAT GCTTATAAGTCCACTTCTTCTTCAAGTCGGAATAGCTC CTGAGGTAA
Vf_Mt2g010880_001	1873029962	C:C	T:T	GGATGGACTTGGTAGCTGGTCAACACTTTTCCAAAGG CACCGCTTCCTCTTATCTGCTCTCGTGCTCCTTACTATT CTATGTACTGTTTATCTTTACTT[Y]GCTATCACCATAGG AAGTGGCTCCTGTTCTGGATTGAGTGGAGCACAGAAA GCTTCATGCCATATGGAGCTTGTCAAGGATTCTATTTC CAACGGCA

Vf_Mt2g021130_001	1939016644	G:G	A:A	TGAATACATTGGGGCCTCTGTCACTCTTTATCTGGACA TCCTGAACCTGTTCCTTTCCATCTTGAGGATTCTGAGG GAGGCAAACAATTAGTCATCCTTC[R]CAAATTGATCTT GACCAAAAATAAGTGAAACTACTTTGTTGTGTCGCATC TTTGTGATGCTGTAGTAAGCATTATGCAGATATACAAT GCTTCTCTAGTA
Vf_Mt2g026550_001	1231406030	C:C	T:T	ATCTGAGGATGATGTCAAAGTATATCCCGAGTTGTTG GACGATTTTGATTCTTATGACAAGAAATTTTCAAAAAG TTCCGACTGTTACCACCATAACAC[Y]GGGAAAATCAGC AAAATAGGAGTCACAGCCAAAGCAAGTCGAAGATCTG ACGACACCAAAACAGCTTGGAGGTAGAATGA
Vf_Mt2g030200_001	1198710232	T:T	C:C	GTGCAACAAGCAAGTGTGCCGAAAGTTGCTGTTTCCC CTTTAGAATTGGAAAATCCTACAGAGCCTCCATTGAAC TTACACAAGCCTAAAGAGCCTTA[Y]ACAGCAACAATTG TTTCTGTTGAAAGATTGGTAGGACCAAAGGCTCCTGG GGAGACTTGTCATATTGTGATCAATCATGATGGAAAT GTTCCTTACTGGG
Vf_Mt2g033460_001	1070986698	C:C	T:T	GAATAGTAGACACGGGTTCTTTTCTTCCGGATTTGCCC ATTCAGACGCAGATGG[Y]GAGGACCCATTTGGCGAAT CAAGGCGAATAGCTTGCAAAAGGTGTGGTGGCTTTCA TCTCTGGATACACACTAAGAAACAAAAGTCTCGAGCA AGATGGT
Vf_Mt2g033910_001	1323908729	T:T	C:C	ATTTTTCAGCAGACGCTTTTTGTTGCAAATTTAGGAGA TTCGCGTGTTGTTTTGGGAAAGAAAGTTGGGAATACG GGTGGGGTTGCGGCGATTCAG[Y]TGTCTACGGAACAT AATGCGAATCTTGAGGCTGTTAGACATGAACTTAAAG AGTTACACCCTCATGATCCTCAGATTGTTGTTCTCAAG CATGGGGTGTG
Vf_Mt2g040010_001	1539199136	A:A	G:G	ACCTCGTTGGAAACAATCGTTCGGAGTTTCCAAGATTC AATGTCTGTAACCAA[R]CAGCACAAATTCTGGGAAAC TCAACCGGTTGGTCAATACAAAGATGTTGGAGACTCA ACCTTACCTGAAGGACCGATTGAACCACCAA
Vf_Mt2g042440_001	1676954693	G:G	A:A	TCGGCATCTCTTCTCGAAGATGAGTGCTACAAGAAATA TGGCAAGACTGGCAA[R]TCATTCTATTATTCCCAAGTG GCAAGTACTATCAGATGGTTGACAACTACAAGTTC
Vf_Mt2g042720_001	1248250912	C:C	T:T	AAAACTTCTCATGGTACAATGYAGGTTCAGTACACTAT TTYGCAGACCCAATAACATTATTCATTGTTCAGCTAGC CTTAATGGG[Y]TGGGTTGAAGGTAGAAGATGGGCCG ATATYGTCAACCCGGGAAGYGTTGATATCGAGCCCAA GTTTCCAAATAGGCCTAACCCAAAACCAGATGTTGGTT
Vf_Mt2g069030_001	1636805800	G:G	A:A	GGGAAGAGGCGGGCCGGGCCTAACAGTTTCCCAAGG AAAAATATGGGTTGTTTATGGTTTCGCAGGAATGGAA GTCGATGACGTTCATTGCTTTGATCTG[R]CTCWTAAA ACATGGGCCCAAGTGGAAACAGGTGGGCTCAAACCA ACGGCCCGTAGTGTGTTTTCAACTTGTTTGACAGGGAA ACACATAATTGTATATGG
Vf_Mt2g082430_001	1450917991	T:T	C:C	TTGTTTTGGCAATTTCTGCATTGATCTGTTTTAGATATC ATAGGAAGAGATCACAGATAGGGAATAGTAGTTCTCG GAGGGCGGCTACGATACCGATTCG[Y]GCTAATGGTGC TGATTCTTGTACTATTTTGTCAGATTCAACCTTAGCTCC TGAATCTCCTGTTAGGTCTGGAAGAAATGGCATGAAT TTTTGGATTGATG
Vf_Mt2g086970_001	1751609764	G:G	T:T	ATCAACATCAGCAGCCGGGCCACAAATCTTGTGAAGC CTTCTTTTTTTATTCCTCCCCCTTCTTCTGCATCAATGAT GTTACCCCCTGTATCTTCATCCG[K]ACCTACTGCTCCTC YGCTTCATCCTACTGGTACAGTACAGCGTCCTTATGGT ACTCCAATGCTACAGCCATTTCCGCCTCCGACTCCACC ACCTTCACT
Vf_Mt2g087870_001	1995592314	T:T	C:C	GGTTGTTGCTGGTGGCTTATTGGGGTGCATTATAGCA TTTTTAATGAGGAAATAATGTTAAGATTCCATGCTATG TAAATGATAGTATTAATTAATCTAT[Y]GTTAAAACTGG GCGTGCATGAGGAAGAGGAGAGCATCATGGAAATGA CATAAAAGCTTGCATTTGTTGCAAGCATTCAAGCGAGT AAA

Vf_Mt2g100090_001	1490373562	T:T	A:A	TGTGGTGTTATGGAGTATTCATGAYCACATTGCAACTT TAGCTACAGAAGCAGGACC[W]GATGCTAAACAGGGA TCTAATGTTGGGGGAAGTGSTGAGAAAACTGCAGAAA GTCCTTCTGTTGGA
Vf_Mt3g008180_001	1708355009	T:T	C:C	TTGCTGCCATGCRATACGTAAAGGACGGAATCAAARC TGCTAGGCTAGTAATGCAACACACCCAACACACTTTGT TGGTTGGAGAGAAAGCGTCTGAATT[Y]GCAATTTCAA TGGGTCTTCCAGGACCAACAAACTTGAGTTCATCAGA ATCTATAGAGAAATGGACTGCATGGAAAA
Vf_Mt3g014740_001	1671711774	T:T	A:A	GGTATACAATTTCTGCTGCACGCGCAAKCTTCAYCGAC AATGATCTCGGATCCTTGTCTTTTGGAAAACTGGCAGA TTTTGTCATATTATCC[W]CKGATTCATGGAAAGATTTT GCTGAAACAGCATCTGCATCTGTAGAGGCAACATATG TTTCTGGAGTGAGAGCATATCCTTGA
Vf_Mt3g031280_001	1498361382	G:G	C:C	ATCCTACCTAAAGCAACTGCCATGCCAAATATTTGATC CAATCCCATTAGGCGAGGGTGCTGGTGTGGGATGGTT AAAAGATGTTGATTC[S]GTGGCAGTTCCTGAAGGTRC AATGCTRTATGAGCTGGTGAAGACTGGCATTGAACAA ACACACGCAGCCGTTGGAGTAGTAGTTCGTTTGAGGA AAGAGT
Vf_Mt3g031380_001	1757570916	G:G	A:A	TAGCTGTTGCAGGTGCTTTCCATACTAGTTTYATGGAA CCTGCTGTTTTAAGATTGGAAGCAGCATTGGCATCAAC GGACATTATAACCCCAAGAATACC[R]GTCATTTCCAAT GTAAATGCTCAGCCACACGCAGATCCTGACACAATWA AGAAGATATTGGCACGACA
Vf_Mt3g055920_001	1561441863	T:T	C:C	AGATAGTAAGTCTTTTGGAATGCACTACATGGTCGGG GGCTGGACCGGGACTAATGGATGCTGTTACTCGAGGT GCTATGCAGGCAGGAAAACCGGT[Y]GGTGGATTCAA GATAGGAAGAGAAGCTGGGGAATGGACAGCATCCAA CTTTCATCCATACTTACCACCTGAAAACTACCTTACCTG CAGGT
Vf_Mt3g060850_001	1742142305	G:G	A:A	AGATGCTTTGGGAGGTGACCTATTTTTGGTTAATAGG AGTTGAGGTCAGATATTTCTCTATATAGTAGTTGGCTT GTACATTTTGTCAACTAAAAGATCT[R]TCAGCTGAAAC CAAATTGACCCAAGATTATTGCCNGAGTTGAGGCATC ATCTAACATTTTCTTATAGTTCGATGTTTGTTGCCACTT GCCA
Vf_Mt3g061590_001	1437042113	T:T	C:C	AAGGTGGCAAGTTTTCCGGGGAAGGCTTTTGTGTTTT GGGGTATCCTCTACAAGTCGATTTTGCTCTTCAAAGAA ACAGGCAGAACATAGGTAGAAGGTA[Y]GTTGAAGTTT TCAGAAGTAAGAGACAGGAATACTACAAGGCAATTGC AAACGAGGTTTCAGACGCTCGAGGTGGTTCACCTCGT CGAAGTGCTCCCCG
Vf_Mt3g061640_001	1536109396	T:T	C:C	ATGATAAAGGACGGAATAATCGCCGGACGCGCTGTTC TTCTCGCCGGACAACCCGGAACCGGCAAAACCGCCAT CGCCATGGGGATGGCGAAGTCTCT[Y]GGCCTCGAAAC GCCGTTTGCTATGATCGCCGGAAGCGARCTCTTCTCGC TCGAGATGTCGAAGACTGAAGC
Vf_Mt3g062540_001	1031140295	G:G	A:A	TGGTGGACGGAGGAACTAGCTGTTCTCGACATCAAGT TCTTCGGGAGAGCCATCAGTGGTATGAAACAGCGCGG CGCCAAGTCTCCAACCATTGCAAGC[R]CACTGATTACC TACACAGAGCGATCTCTCCAAGAGCTAGTTCGGAATC ACTCCGGCAATGGAATCCGATCTTCAGGTTACCACGAT TCAGA
Vf_Mt3g065190_001	1639546187	G:G	A:A	CCATTCCTGCTCCAGTGCCACCCCGCCCCTTCGCTAAT GGCGTTGCCCCACCTCCTATTCCTGTCATCCACCCCCCA CCTCCTCAAGCCGCAGCCTTTCA[R]CCTATGCAGATGC CACCACCTGGACAACAAGTATGGCATCAGCAGCAGCA AGGTCAACCGATGATGCAACATGGAATGCCTCCTCCTC C

Vf_Mt3g070310_001	1036087037	G:G	A:A	ATTCCAAACCATCAAGAGATAGTGAGAAGGAGAAKG GATCTTCTCCTAAAATTATTGCAGAAAAGCTGGTAGCG TCTTTAGACTCAACAGTAAAATTATCY[R]GTCTGTCTAC CAGGGCTTGCAATGGTCATGTCAATTTTTATGCTTCTA GTAAACAAGTTTCCCAAAATGGAATTGTTCAAAATGCT GCAATTCC
Vf_Mt3g076590_001	1990709799	G:G	A:A	ACAGATACAAATGTGTTAAGGTCATTCTTGCTGGGCAT GATCATAAAGGTGGGTACTCCATTGATTCTCACGGGG TACACCATAGAGTTTTTGAAGCTGC[R]CTTGAATGTCC TCCTGGCTCAGATGCATTTGGAAGYGTTTATGTCTATG ATGACAGGATATCACTTGTAGGAACTGACAGAATGGC AAGTACAGACATGC
Vf_Mt3g077990_001	1465511448	C:C	A:A	AAGATTTGATGCTGTAAAGGACAATTAGCTCTAAACAT ATAATCAATCTGAT[M]CTGCGATTGCATCCGGAATTA AATAATACTTTCAATGGATCTTGACTTGGGTGGTAGTA ACTCCTGTCCGCCTCGACACAAATGGAGAAAAGTTGC ATA
Vf_Mt3g079200_001	1551530724	A:A	T:T	ATGGAGAAATTGAGTCTGTCAAAATGGTTCTTCAACG GGCTTGTGCCTTTGTAACCTATACAACAAGAGAAGGC GCAGAAAAGGCAGCAGAAGAACTCTC[W]AACAAATT GGTTATCAAGGGCCTTAGGCTAAAACTGATGTGGGGC AGGCCTCAGACAGCAAAACTTGAATCTGATGGTTCTG ATCAAGCAAGGCAG
Vf_Mt3g079210_001	1242012828	C:C	T:T	GGATGCTTGAATGCTCTAACAAATATGTCGGATTGCTT AACATTCGTGGAAGATGGGAGTAAATTGAGAAAGCC GGA[Y]AAAGGTTGTTGTCCTGAGTTAGCAGGGTTGGT TGATAGTAACCCTATTTGTTTGTGTCAGCTTTTAGG
Vf_Mt3g083770_001	1632005128	A:A	C:C	ACCAAGAAATCCTTCTGTAATACCTGCACAGAATTTGT CAGCAGCAACAAGTGCTTCACTTTCTTTTCCAGTTTCCT TAAGGCAACCAACATCAACAGGAC[M]CCTCCAGACAT CAGTGTCTTCAATGCCTCAACCYATGTCTGGTATATCA CCGTCACCAATAGCTAACCAGCCATCAACCTCTCAAGG TGTATCTTCTGGA
Vf_Mt3g083900_001	1564732983	C:C	T:T	GGTTTAATGCTCTAGAGCAGGTTGACGATGCATTGGC TGGTCTAAAAGTGATTGCATTTGA[Y]GTGAATTACATT AGGCTGTCTTTACAGACTTWTGTCCCAACGGTAGAGA GCATTTCATGCCTACAGAGAGTTGAAGATACCATTGAT GCATCTGTGCTGA
Vf_Mt3g084040_001	1128260313	G:G	C:C	GAAGGCTTAGCTCAGTGTTCAATCAAAATGGCTTTCTG CAACAAGCTTGGAAACCTCTTGAGGCAGGGTGCTACT CAGAGTA[S]CCATGCTCCTGTTTCATCCATGCTTAACTA TCTTCGCCACATGTCTTCAAGCAAGCTTTTCATTGGAG G
Vf_Mt3g084090_001	1700494419	C:C	T:T	AACGTTACAATAGCTACCTCAGAAAAGTTAACAGCAC AAAACTCCTTAACGCATCGTCCAAACTTCTCTTCCGAG CCACACTTTTAATAGCACTCGTCCT[Y]GTGTTCTTCTTC ACTTTCAATTACCCTCCACTCTCCGACTCCACCAACCAC CACTTCCACACCCATTCCCACTTCCTCGCCTCCGCATTC GGCGG
Vf_Mt3g085270_001	1215872866	G:G	T:T	CTACGTGGCAACAAAGCTTGTAAGCTCGCGCTTGAGA ATCTACCAGATGAAGTTTACGATACGGAGTGGGACCT TATTATGCTCGACGCGCCTAAAGG[K]TACTACGCGGA GGCTCCGGGACGYATGGCCGCCGTGTTCTCCGCTGCT GTTATGGCGAGGAACAGAAAGGGATCTGGTGTGACG CATGTTTTTCTGCATG
Vf_Mt3g085570_001	1998158399	A:A	G:G	AAGAATCACTGTGAGAGAAAAACTACTGTTATGGAAT TTGAATCAGACTGGATACAGTATAGAAAATCATCTGCT TATACTGCTCA[R]GTTATTCGTTCTGCAACATCTTTTAT GCCTTGGTCTTGTAGAATCCCAATCGGTGAAGGATTTT ATGAATTTCTTTAG
Vf_Mt3g086040_001	1953767441	A:A	G:G	GTGGTTGAAGCAAAGGGTAAAAAAGGAATGATGTAT CGCCAATACCAGCGCCGT[R]CTCCTGCTCCTTTACCCA AAGTCGAAGACGACGGTAACCCTAAGTTCGTCGTCTT

				CATTCGCTCGGCTGATGTTTACCTTTGGTATCCTCTTAG TATTG
Vf_Mt3g086230_001	1066774494	G:G	A:A	TATTCATCTATGGCAGCAACAATGGCAATGATGATCAA CTCCAAGTGTCTAAACAC[R]GCCATACCAAAGTTGCAA AACCCAGCAACTAGCACTACTAGAGTAACAACATGCA AAACCACCACCGTTCAGAACCTACCAAAGGCTCT
Vf_Mt3g086600_001	1523220683	G:G	T:T	AGGTAGCCATTGAARAATACTTCRAGCCTCTTGACAAS GAGGCTGAGTTTTTAATGAAAATGCAACTARAAGGGG AAGAGAAAACTTCGGAGATGATGAT[K]AAAGCACTAC AGGAACAAGCTATGCGGCGACAAGCTGAAGCAGAAA AANGTACAAGTGTTCATCAAACAGAAAGTGTTGAAAC CAACCAGAAGGAA
Vf_Mt3g087030_001	1646314581	T:T	C:C	ATTTCAGCAATCCTTTTGATCTGTTTGAGACGCTATTTG AGGGCATGGGTAGTGGCAGCAGCAGAGGTTCTTGGA ATGGAGCAGTGGATGG[Y]GAAGATGAGTATTACAGT CTTATTTTGAATTTTAAAGAAGCTATTTTCGGAGTGGA AAAGGAGATAGAGATAAGGCGACTTGAGAATTGTGG AACTTGC
Vf_Mt3g087150_001	1499930808	T:T	A:A	AGGAAGCTGAAAAGGCAAACAGTGTTTCACCACTTGC CTTCTTAAGCGTTGTAGAGGTTCCTAGTCAGGCTAAAA GCGGAA[W]GTGTTTAGACACTCATATTAATTGTCAAA ACAACATTGATTTTCAAATGAAGRGCAAAGATATTTAT ACCCAGTGCATAATTGATATACCTACTGTGAATG
Vf_Mt3g087150_002	1334710648	G:G	A:A	TAAGCGTTGTAGAGGTTCCTAGTCAGGCTAAAAGCGG AAWGTGTTTAGACACTCATATTAATTGTCAAAACAAC ATTGATTTTCAAATGAAG[R]GCAAAGATATTTATACCC AGTGCATAATTGATATACCTACTGTGAATGGAAACTCA ATTTCCCCTGAATCCTATGAAGAGGGTGTTGAAAGTTT TAAAACT
Vf_Mt3g087760_001	1464559498	C:C	T:T	AGAGAGGTTTTTGTTGTGAATTGTATAATTGGTCACCA TTTGAGTCACAAGGAATTTGGTGTTTGTTTAAGCTTGA TGAAAAGCTTACTCTCTCGCGATTCTAC[Y]GATCCTTAT TTGATTTCGCAACTTGGGTATATTCAATTACAAACCGG TGACTTGGAAGGTGCGAAAGCTTCGTTTTTGAAAGCC GAAGTTGATGGGAAGA
Vf_Mt3g088610_001	1397002075	G:G	A:A	CTTCTTCAACCATTCCAAACACACCYTTAACCTTGAGTT CCCACATCTACGTAGATCTCAAATCACCACTGACATTA TCCAACTATCTCAGCTTCTCTTC[R]TCAAAGCGCCGAC AACCTCCAAGTCTTTTCACGGTCAGAGCTAGTGACTCA GATTTCGAAGCCGCGGTTGTTGCCGGTAA
Vf_Mt3g091460_001	1952231569	T:T	C:C	GGTGGGGTTGTGGAAGATGGGGTTGTTTTGGGGTGT TTGATGAAGGGTTATTTTCTCAAGGGGATGGAGAAGG AGGCTTTAGAGTGTTA[Y]CAAGAGGTGTTTGCTGAGG GGAAGAGGATGAGTGACATTGCTTATAACTCAGTGCT TGATGCTCTTACAAAGAACGGAAAGTTTGATGA
Vf_Mt3g092810_001	1994334164	G:G	A:A	GTTCAGGTGGTTACGNTTAGGGTTATTGGAAAAGACG GAAAAGTCCTGGTTGAATCGCATCAGGAATTGTCGGA TGGTAGAGTCAGGGAGCGGGGTAGACC[R]TTGTCGG AGAAAATGAAGCCCAACGAGGAACCGGAATCGGCTG CTGTTAGAGGGATTAAGGAAGAGCTTGGTTCTGT
Vf_Mt3g095660_001	1777884970	A:A	T:T	TTGCTAATTTTGGTTGCTGCAAGTTCGGGAAATGATCA TGAGGGAAATTCTGCAATCAGGAAATTTGTTATAGGT GCCCAAAC[W]AATGAAGGTCTTGAAAATAAAGATGT CTTCTATGGAACYGCTGGTTGTTTATATTCTATAAACCC ACTGTTTCATGTATTTCCACCTAC
Vf_Mt3g099130_001	1488599943	G:G	C:C	CTCTCCCAATGCTCCTTCCTCAAAYGGTGATGCAACTTC TGATAACTCCGGCCAACACGG[S]GATGGAGGCGCTTC CGGTGGTGGTACTGGTGGTGCTCTTGATGAAGCTCGA TTTCGGTATAAGAAGGCTGTTGCTGGTTTACGTAGTGT GCTTGTTGCTAT

Vf_Mt3g102180_001	1939937736	A:A	C:C	ATCGGAAAGCGGTGATTATCGGATTCTCCACCGGAAT CATTTTTCTAATACTTTCCTTGACATGTTTCGCGGTGGT GGTTAAGAAACAA[M]GAA**N**GAAAAATGGTAAAGGCT CTTCCGGTTCGAGCGTGATGGCTTCGGATGCAGCGGC AACGGCGGAAGCTGCGGTTGTGATGCAAATGGAGCA AGAGAG
Vf_Mt3g104310_001	1349771927	C:C	G:G	TCTCTCCACACAACCCTACCTTCTCTATCCACAACCTCA CCTACACCATATGCACCACCGCCTCCCCCATCGGCTTC CTCCCCCGCCGGCCTTTCAAAGAC[S]GGCGAGTACGT GATCTCAAAGGTTGATGATCTCTTGAATTGGGCTCGTC GCGGTTCCCTCTGGCCAATGACCTTTGGGC**Y**CGCTTGT TGCGCTGTTGAAA
Vf_Mt3g105550_001	1694925058	C:C	G:G	ATCAAGGAAAAGAAAGATCTATGCTTATTCCTAATTGT TTCTTTCGTATGGGTTGTTCCGCTGTTTTGCTCTCTAAT CGTCGTCG**T**GATTATAGCCGCGC[S]AA**R**TATCGTCTTG AACACATTGTTCGTACTCATAAAGGCGCTGATGACCGC AGTTTCAGGT
Vf_Mt3g109200_001	1076887548	C:C	A:A	TGAATGATTTCGTCAAGATGACAGATGAATTCATACAA CAGTTGGAATTGCCAGAAACAAAGCCAAGGGTTCAGA CTTATGAAACCTTCTATAAGCAAGA[M]AACGATAAGC CACTTCACCCTTGATGCTTTCAGCTTGT**Y**AAAATTCTGC ATTATATTCACTTTGCAGTATTGTATGAT
Vf_Mt3g109330_001	1527050489	T:T	C:C	CGAAGCTTTGTTGGCGGTTGATTCTCATGGCGCCTTCG AAGCTCTTCGTCTTCGCTCTTTCCGTTGCCCTAATTTTC GTCTTCGGTACTTCCGAAG[Y]CGACGTTTCCATCGATG ATTCCGATTCATCTTCTCTCAAGATCCAATTAGATCAAC TCAATACCAAGATCCAATTTCTCG
Vf_Mt3g114780_001	1220763894	T:T	G:G	ACGTGCCACGCGCCGAAGCTTAACGAAGTTTACGCTG AAATTTTTAGGGTTTTGAAACCGGGTGCTCTTTACGTT TCATATGAATGGGTCACAACGGA[K]AAATTTGTGTCA GAAAATTCCGAACATGTTGAGATTATTCAAGGGATTG AAAGAGGTGATGCTTTGCCTGGGCTTAGAAGCTACGC TGATATAGCTGAA
Vf_Mt3g115990_001	1124251109	C:C	G:G	TATGCTAAGAATGA**M**AAGTTGCGATATACTGTTCTTG CCAAGAAGATGTTTCTGGTGACGAATGATGG[S]CTTG TTAACTTCAAGATTAAGGGAGGATGTGACGTTGACCA AGATTTTAA
Vf_Mt3g116080_001	1383060710	T:T	C:C	ATAAACTCATTGAACTCTTCGAGTCATGGATGTCGAGA CAC**R**GTAAGATTTACGA**K**AGCATCGAGGAAAAACTAC TCCGGTTCGAAGTGTTTAAGGATAA[Y]CTAAAGCACA TTGATGATAGAAACAAGGCTGTTAGCAACTATTGGCTT GGTTTGAATGAGTTTGCTGATTTGAGTCACCAAGAATT CAAG
Vf_Mt3g118200_001	1787699891	C:C	T:T	GATACTGAGAGTAAATCTTCTGTTATCGACGAGGCTTA TAAACTTCAGATTCCCATTGTGGCTCTCGTCGATTCTG CTATGCCTCTTCATACTTT[Y]AGCCGCATTGCTTATCCT ATTCCTGTTAATCCCTCTGTTCAGTTTGTCTATCTGTTTT GTAATTTAATTACCAAAACATTGCTTCTCGAGAAGAAC AAC
Vf_Mt3g118210_001	1454487833	T:T	C:C	GGATGGGTGGATCCAAGACCAAGATCAGAAAGCTTA AGATTGTTAAAATTGATAAAGAGCTTAATGTTATCATG ATCAAAGGTGCTCTGCCTGGTAAGCC[Y]GGGAATCTA ATTCGGATAGCTCCGGCTAAGATTGTTGGCAAGAACA TTCCCAAGAATTA
Vf_Mt4g014430_001	1308139280	C:C	T:T	AGGAGTACAATGCAACTATAGAATTCCTCTGGGCTCCT CTTCTCGTTGAATCTAATTCCGA[Y]GATCCGGTGAATC ACCGACTAGATGAACGGATAATTCGTCCTGATTCAGTT ATGAAGCATGCATCACTGTGGGAGCATGCTGATATAC TTGTTTTCAA
Vf_Mt4g016930_001	1022895730	A:A	G:G	GATTCTCACAAAAGGTGATAGGTATTCTGGAAAAGGA AGGAGTAGATTATGAGAGTGTGGATGTGTTGGATGA AGATTATAACTATGGATTGAGGGAGAC[R]GTTAAAAA GTATAGTAACTGGCCTACATTCCCTCAGATATTTTTGA

				ATGGTGAACTTGTTGGTGGTTGTGATATATTGACATCC ATGCATGAAAAAGGTG
Vf_Mt4g019080_001	1683303615	A:A	G:G	GTGTTGAGTTTAGTGTTTCAAGGGATGGTAAAAATGA GATTTTGATTAGAAATGGTGTTTCAAGTAAAATCTATG TCACCAATTTGGTACTAACA[R]GTGGTGAAAAGATGC ATACTTGGATTGATTATGAAGCAAGTTCAATGAGGTTA GAAGTAAGATTGAGTAAAAATGGGAATTCAAAACCAT TTGATYCATT
Vf_Mt4g031820_001	1001364102	C:C	A:A	GGAGTGCACTGGATTGGATTTGGTTCACACCAAAAAG AATAGAGAAACGTTTGAAACAACAAGGTCTTAAAGGA AATTCATAT[M]GAATCATGGTTGGAGATATTAGAGAT ATGGTGAAAATGATCAAAGAAGCTAAATCTAAACYCA TGGACCCTTACTCTAATGATATTGCTCCTCGTGTTTT
Vf_Mt4g035200_001	1097145874	G:G	T:T	AAAATGGAAACAAGCATCTTTTCACAAAGTGGATATT GATGCTGATGAAATA[K]GTAATTTGAATGAAGATAAT TTGATCGGCCACGGTGGTACGGGAAAGGTTTATCGAG TCGCGTTGAAGAAAANCGGAATGGTAGTGGCCGTGA AGCA
Vf_Mt4g044470_001	1392104708	A:A	G:G	GATGTTGACACAACAATTGGAGCCTTTAGCTGAGCAA CAAATAGTAGGGATAATGGAGCTAAGACAATCATCAC ACCAAGCAGAAGATGCTCT[M]TCTCAAGGCCATGAAC ARCTCCACCATTGTATTGTCGACACCATTGCCGGAGGG CCGGTCATCGACGGTGTTCAACAGATGGTCGCTGCCA TGG
Vf_Mt4g053250_001	1392104708	C:C	A:A	GATGTTGACACAACAATTGGAGCCTTTAGCTGAGCAA CAAATAGTAGGGATAATGGAGCTAAGACAATCATCAC ACCAAGCAGAAGATGCTCT[M]TCTCAAGGCCATGAAC ARCTCCACCATTGTATTGTCGACACCATTGCCGGAGGG CCGGTCATCGACGGTGTTCAACAGATGGTCGCTGCCA TGG
Vf_Mt4g053270_001	1160147607	T:T	A:A	GAGATCTGTGCTGTGTGGTGGTTCACCAGCCATTTTCA TGTTCGCATATTGTATCTACTTCTATGCAAGATCAAGA ATGAGTGGTTTCTTACAACTATCT[W]TCTTCATTGGCT ACAAYGCATGCATATGCTATGCTTTCTTCTTGATATTTG GTGCCATCAGTTTCCGAGTTTCTTTGCTATTTGTTCGCC ATATATACC
Vf_Mt4g061070_001	1374920491	C:C	T:T	ATTGAGAATTCTTTAGGAGGGAGCATCCATAGGAATT ATGATCTTTTACTCAGGCACAGGCTACATATTGTAGGA GAAGTAAAATATGCAGTCCATCA[Y]TGTTTGATGGCC AACCATGGTATTAGACTTGAGAATTTGAAGCGCGTTCT TAGTCATCCTCAGG
Vf_Mt4g064820_001	1480510347	T:T	C:C	GTTGGGTTATTTTGTGGGAAGCTTTCAGATTACTACAG TTGGTGGATGGCGATTCATGTATGGATTTAGTGCTCCA GTTGC[Y]GTGTTAATGGGGCTCGGAATGTGGACTCTC CCCGCTTCTCCTCGGTGGTTGCTTCTCAAAGCAGTACA AGGAAAAGGTTCTTTTCAAGAATTGAAGGA
Vf_Mt4g068480_001	1092780421	G:G	A:A	GCCTGCTGTATCTATTGGCAATGTTGGTCAGCTAACAG CGGACCTTTTGGTTTCATCATTGGGTTCTGAGAAAGTT GGTTACTT[R]GACGATCCTTACGTTCTTCCATGTGTTG GCAATGATGCTTATGGACCTGTTCCTCAAGGAGATCTT GCACTTCCTCTTGAAG
Vf_Mt4g076870_001	1643554427	C:C	A:A	CTCCGTGCCTGAGCAAGCATTGTGTTTTCTTGCTGGTG CCAATTCTATCTTTGCTGGTGAAAAGCTTCTCACAACT GCTAACAATGATTTTGATACTGAT[M]AACTCATGTTTA AAGTTCTCGGTCTCCTTCCAAAAGCTCCAACCTTAGAC GAAGATGAAACTAGCGAGACAGAGAACTATAAAGAA GCTGCTTCTTCTAG

Vf_Mt4g083980_001	1975595778	T:T	C:C	TTTGTCGATGACGATGTTGTTGGAGGACGTCGTTCGA TCCATTGAACTATGGCTTCGTCTCTTGAAGAAGCCCCA ATCCCAAACCTACATCAACCCTAA[Y]CTCGAYCCTGTTT TATTGGTACCCGGCGTTGGTGGATCAATTCTTAATGCT GTTAATACCGACGGTAGCCAGGAACGTGTTTGGGTTC GGTTTCT
Vf_Mt4g086420_001	1632173798	G:G	A:A	ACTGTCAATTGTTTAGGAACTGTATTTCTGGACATATA TATGATTCTGGGCCACTAGATGTTACAAGTGATTTTGG CTTCCGTTTTTCTCTACACCCTTC[R]ATTGCAAAAGTGC CCGGCCCATCAAAGCTTGTTTCTTTGGTAGCAAAGTCT GTTGCATCTGGTTTAGATGCTTTATACTTGACTCGATTT GAATCCCAAT
Vf_Mt4g100760_001	1273448724	A:A	G:G	GCTGCTCAAAAGAGACCAGATATATTGGTTCAGGTGG GACTTGAGTACAGATACATGCCGCCTGTTGC[R]AAAC TGATAGAAATAGTGAACGGAGGAAGCCTTGGACATGT TAGAATGGTATCAATTCGGGAGCACCGGTTTCCTTT
Vf_Mt4g101130_001	1347090997	T:T	C:C	TCGTCATCGTCGATAACAAACCCGACAAGCTCCGCGG CGAGATGCTCGACCTTCAACACGCCGCTGCTTTCCTCC CCCGAGTCAAGATCAACTCTTCCGT[Y]GAATACTCCGT CACCGCCGGCTCTGATCTATGCATCGTTACCGCCGGTG CACGACAGATCGCTGRCGAGTCCAGGCTTAACCTTCTC CAGAGGAA
Vf_Mt4g124930_001	1238916002	C:C	T:T	AACAACAAAACATSTGATGTTCTTCCTAAACATGCTGT TTGCACATCAAACTTGAATCGATCAAGACCAATCCAAA TTCAATC[Y]GCCAAAAACACAAACGGGTTTCAAATGG ATTATGATATGAATGAAATTGACATGGAAAGTTTGGA AGAAGA
Vf_Mt4g132020_001	1253706856	G:G	A:A	CACTTTGCTACAACCGTCGCTAAATATCATGGCGAAGG GTCGCAAATTCACGGCGAACCG[R]AACGAGCGGTTGT TAGGAACAAGCTACATCCAAACCTCCACTGTTACCGTT ACTCATGAAGTAACGGATTTCAGGGAAGAAGATGTGT GGTC
Vf_Mt4g132300_001	1270881468	C:C	A:A	GCAATCGTAATCTCCAAAAACGATTAGATCGGTTTATT TGTCATTTACGTGATCTTAAGATAAGTAGTATTGACAT GATTGGGATGGATGAAATACCTGG[M]GTATCACACA TCAAAGAGAAGAAAGTGAGAAATGAGATCAAACAAC TGGTACTCCCTGTTTTGCCCAGTGATCATTATGGTCTCC TTTTGACAATTTCTAG
Vf_Mt5g005050_001	1352280043	T:T	C:C	ATTGTAATTGGGAATGGGAAAAACACGTTCCTTCTTTA CCTTCTTCAGCATATCCACTCCTCCGTTCCAATTTCAAA ACCCCTCTCTCTTCTACTCT[Y]CACTCCGTTTTCCACTCT TCTGATAACCTCACTTCAAAGCTCCTCATCAAGCTGAA TAATGGAGAATTTGTGGAGGCTGTGATAATGAGATAT GATACT
Vf_Mt5g005120_001	1713895145	T:T	C:C	TAAAATGTTGGGATCCTAGAGGTGCAAGCGGACAGG AGCGCACTCTTGTTGGGACATATCCACAGCCTGAGCG TGTTTACTCTCTATCTCTTGTTGGACA[Y]CGGCTTGTTG TAGCAACAGCTGGAAGACATGTTAATGTTTATGACTTG AGGAACATGTCTCTGCCTGAACAACGAAGGGAATCAT CATTGAAATATCAAA
Vf_Mt5g005990_001	1322296619	G:G	C:C	TGTGCAATTTTNCCGAAAGKCCTTTCYACAGGCAATTT TCCCTGAAGATCATTCTTTATTATTCATGTATTCTTGTT CAAATAGGTTAGA[S]AAGTCAATGTACTAGAAATGAG GTCAAGGCAGTATGCCCTAGAATTTTGTATATTAT
Vf_Mt5g007030_001	1451147422	T:T	C:C	TGGAATCTCGCAAAGGTGATGGTCAGTTGTCTGCATG CGAGGTTACCCTTTTGCATCACGAAGACCTGGGAATTC CAAAAGATGTTGCAAAGTTAGGAGT[Y]CGCCATGGAA TGTGGGGAGCTGTCAAGAAATTGCACTCTGGTATGAG AGCATACCAGAATGCTAGGAAAACAGATACTTCTTTGT CAAGATGTGCATTGA
Vf_Mt5g007360_001	1562850661	A:A	G:G	CTGAAGTTATTACCCATGAACAATGGAAGCTACAGAA GCCAGGGACAAAGAGTGTGATCACTAGTTGGGTAAA GAGTGCATTTCTTAGCGGGAAGA[R]CAATATGCAAGA

				ATCTGGTAATAAGAATCTTGAATTTGTTCTCAGCACAG CTCCAATATGGGATAGGTT
Vf_Mt5g016250_001	1554768933	T:T	A:A	AAATGGATCGCAAAAACGTTCTCTCCATCGCATTGGTC TGCATTGTCTTCGCCGGCGTCGGAGGTCAATC[W]CCC TCCTCAGCTCCGACAACATCTCCGGTAACAGTCGCAAC TCCCTCTGTTTCTCCMGTCGCAGCACCTTCCAAACCAA AATCACCAGCTCC
Vf_Mt5g016600_001	1005215713	A:A	G:G	GAATGGTCTGGGTTTGTGTTGAATTCGAGGTTGCTTTG GAAGGATGTTGATGATAAACCGGAGTGGATTAAGGA TCTTGACGCGTTGGAYGGGAAGRGT[R]AGGAGATAG AGAGTCCACTATCTTTGCTCAAGAGTGACACTGTGGTA GAACCACTAGGAAATTGTGGACATCATGTTTTGCTTTG GTGGTTGCGGGTTGA
Vf_Mt5g025470_001	1337048384	G:G	A:A	TCTAAGCATGTAACTGAACAAAGTGAGGCTTTGGAAA TTGAGAGGAAGAATGT[R]AGAGAGACTCTCATTTCTG CTTATGACAGGAACTTAGCTTCAAAGCTGAAAGCTGA GATTTCTAGCATGGAAGAGAACAA
Vf_Mt5g025960_001	1563431267	A:A	G:G	AGTCAGGAAAGGCAAACCTTGGACTAGAACTGCACAA GGTTTTGACAACTGAATCA[R]GCTGCGGAGAAGACAT CCTTAAATCTCTGAATTTGAAATCTGAACATATGGCTC TCGAGACCATTAATAAATTGGAAGCTGTTATATTCTCT TTGAAAGA
Vf_Mt5g027960_001	1079505606	T:T	C:C	AACTGCAAGCAACTAGAGAGCAGCGGATTGAAGAGC CCATCCTATACATAAAGATGCAAATTGCCTTATTTAAG CTTGAGCAAGGTGACCAAAAAGAATG[Y]AAGAATCTT CTTGAAAATGGAAAGAGTACACTTGACAGCATGACAG ATATTGATCCAACTGTGTATGCAAGCTATTATTGGGTA TCTTCTCAATATCACA
Vf_Mt5g033220_001	1340190785	G:G	T:T	ATCAATTTAAGAACACCTTTGAATTTGAAGCCAATATA CACAATTCCTTTGAAGATAAGTGAACAAGAGGAAA[K] TATCATGGAGGAAGAGTGTTCAACAACACCAAGAGG GGAAGGATCAAGAATTCCAACATGTTTGATATGTCCA CCAGCTCCRAGGAAGCGAAAA
Vf_Mt5g045630_001	1540464014	A:A	G:G	ATGCTGGCCACAGCAAAAGTGCAAGAGAACTTATGGA GAAATATTACATCGGTGAATTTGACACTTCGTCTACGA TTTCAACCAAGAAAGGTTATTACGA[R]AAGTTCACGA CGCAGCTGAAGAATAAACCATATATGGGTTTTTCCGTA GCTGTAGTTGGCATCTCTRTGGTGGTTGGTTTCTTATA CTTAAGGAAGAA
Vf_Mt5g055940_001	1675748413	C:C	T:T	ATCGAAACTCATGTTCAGCTTTCAACATTCACCAAAGC CTTTTGTAACTGCCCTTACAATTATGGCT[Y]ATTTCCGA ACAGTAGCATTTGTCCGGTTTGTATGGGTTTGCCTGGT GCTTTGCCTGTTTTGAACTCTAAGGTTATTGAGTTTGCT GTGAAATTGGGACT
Vf_Mt5g072620_001	1495073171	A:A	C:C	AGAAGGAGCTTCTGGATCCTGCGATTAAAGGGACTGT GAATGTGCTYACCGCGGCTAAGGAAGC[M]GGGGTGA AGCGCGTGGTTATTACGTCGTCTATCTCGGCCATTACT CCYAGTCCTGATTGGCCTTCTGATGTTGTCAAGAGAG
Vf_Mt5g074880_001	1737917974	A:A	G:G	AAGAAAGATGCGATTAGGAAAGCAGTACGCGTTYSGK ATCTGTTGGTGGA[R]GCTGTGAATGGCGGTGTTCAAG ATTCTTTCATTTACCAGAAACGAATTGAGCTTGAAATT CGTGCTCTTGCTG
Vf_Mt5g075540_001	1716503263	C:C	G:G	TTTGCCGATGGGTTTCCGAATATATTTATAAATAATGC AGARGAGCTCCGGGGTCAACATGTTGCCTTTTTGGCG TCGTTCAGCTCTCCGGCTCAAGT[S]TTTGAACAGCTGT CTGTCATATATGCACTCCCTCGTCTATTTGTTGCTTCCT TCACTTTGGTATTGCCTTTCTTCCCTACTGGATCATTTG A

Vf_Mt5g076080_001	1859968753	A:A	C:C	CACTCAAGCTGACTCRAATCATGTAATAAGCTCGAGTG TATGTAAATATGTTTTGGTCAGAGAGCCATCTTTGTCT GTTTCGYTGTA[M]TTTCGGGTTTCAGTTCTATGGAACA TTAATCTTTCCATAATCCATGGAGTAGCATATGTTAAT AATATAACTACTGGATGCATCTTGTTTCTGTTGCTTA
Vf_Mt5g077800_001	1611306364	G:G	A:A	AAGCAATCTTAAAGCTAATGCGGCTGGGTTAACCAAG CAGAACAATGCTG[R]GCCGGATATTCCMACTAAACGG CCTAAACCTGAGCAGCCTTCAGACTCACAGTTCAAAAT TCCTCAAAGTCCACAAGTGCCTAAACCTCAGTCATCAT CC
Vf_Mt5g078030_001	1144878766	C:C	G:G	TGCCTGACTGCGCGTCCTGCGTGGCACGCGCAGTCAC GCGGGCGGACGACATGTGTCGCAACACATGCGGAGG CACGGTCCAGCTCGACGGGTGTTTGGT[S]AAGTACGA TAATGCTACTTTTTTGGGTGTGGAGGATAAGAATGTTT TGTTGAAAAAGTGCGGGCCTTCCGTGGGCTATAATCC GGAAGCTATGGGCTC
Vf_Mt5g079730_001	1953733000	C:C	T:T	CAGGGTATCCATCATTCAAGGGTCAGGTTTTAAATGG ACAGCAACTCTGGGAA[Y]TAATTGAAGGCCTTGAAGC TAATGATTTATTGTACTACACTCACTTATTAACAGGT
Vf_Mt5g080450_001	1083222602	C:C	A:A	TTCGACCCAACTGCTAGAAGTGATGATGGAACATGCT TATACACATTTGAATAATCTAAGGATTCTCCTATTAGG TTAAGTTAAGGAGGAGAAGCTGAGA[M]TTTAGATAA TAAATCGAATAATTTTATCATCTTGTTAAAATATTGGAT TGTTGTTGT
Vf_Mt5g097360_001	1893913011	G:G	A:A	CCTCCACGACAGCAGTCTCGGGTTGGAGGATTGCAAA CATCTCTATCTCTGGTTCCTTTGGATCCGCGCTTGTCTC CTGAGGATCCTAGGTC[R]AATTCTGATAATCTTCGAGA ATCTCCTACKGAAAGTGCTAGTTCTAGAGAAACCTGGC CTACTGCTGATGCAATTGCTGCAAAGAAGATGGAGAA
Vf_Mt6g012630_001	1939301813	G:G	A:A	GTCCAACTGCAGGACCTCCGATTGTGGGGATCRCTCCT CTTGCTGGTTCCTATGGGAGATATGAAGGTGACACTG GTGACTGGCCGCGGTCTATGTACTC[R]GCTCCCAGGG ATGAAATTTCACCCATGGAGTTTACAATTCGATTGGTT TGTCCAACTGTAAACCTTGGGGGTGTAATAGGGAAAG GTGGTAGCATAATAAA
Vf_Mt6g071210_001	1483496102	T:T	C:C	GTATCTCTTGATGCWGCAAAGGTTATAAATGTAACTC CTCGATATGCTCGACTACGGGCTARTGGAAATGGAT[Y]GGCAAAAGGAGAAATTATTGGRTGGCAGAACATTGA ACTTATCAGTGATAGGCCACTTGAAACCATGTTGAAA GAATTTAAGCAATTGAAAGATGAGTAC
Vf_Mt6g071210_002	1483496102	A:A	G:G	GTATCTCTTGATGCWGCAAAGGTTATAAATGTAACTC CTCGATATGCTCGACTACGGGCTARTGGAAATGGAT[Y]GGCAAAAGGAGAAATTATTGGRTGGCAGAACATTGA ACTTATCAGTGATAGGCCACTTGAAACCATGTTGAAA GAATTTAAGCAATTGAAAGATGAGTAC
Vf_Mt6g071210_003	1230422291	G:G	A:A	GGTACTCAACTCCAGGCGGTTATTCTGCGAAGGCGGT CCATCCGATAGCCCTCGGGAAGGTTATGAGTATTGCA AAAATGATGAAAGCCGAGTTTRATA[R]TGAGAACTAT TCACTTTCTGCCATTGGTGGTGTTGAGACAGGTGGTG ATGCTGCTGAGTTTATTCTTCTTGGTGCAAATACTGTC CAGGTATGCA
Vf_Mt6g071280_001	1520153419	G:G	A:A	ACGCGCGTAACATAGCACACCGTGACATTAAACCTTC GAACATCCTCGTTAACAACAAAAACGAGGTTAAGATA GCAGATTTTGGTGTTAGTAA[R]TTGATGGGTCGCACG CTTGAAGCATGCAACTCTTACGTTGGCACGTGTGCTTA TATGAGTCCGGAACGGTTTGACCCGGAAGTTTATGGC GGGAACTATA
Vf_Mt6g082480_001	1934276177	C:C	G:G	AAAATATGAAAGCAGCACAAATTGTTATTTGCACCTTC CTTTTGGTGCTT[S]TGGTTAATGTTCAAGTGTCAATGG CAGTAACATGCAGTCCAGTGCAACTTAGCTCATGTGTG AGTGCAATCACTTCTTCAACTCCTCCATCTAAACTATGC

Vf_Mt7g013160_001	1262526466	T:T	G:G	TCATAAATAAGTCCATGTTCGATTATATTGATCGGTCG TCATATAGAGCTTTGGCATTTGATTACCTAGAAATATC CCCTGGCTACCCGTTTTTGTC[K]GGCATCCGGGCATGG ATCTCTTTACTATTCATGGAAGATGTAAAGGAAGATCT TATGGATGTGTTTGGGATCCAAATGGATTTCTGTGATG TGGCAAATT
Vf_Mt7g038120_001	1174825482	T:T	C:C	GGTGGTGGTGATTCGGTTGTTTCGGGTGATTTGGTTG AGAGGGTGAAAGATACATGGAAGAGTGTTGCGGAGA CTGCTTCTTATGC[Y]GGGGRAAARRWTWAAAGAAAC TTCTGATGATCTCACTCCATATGTTCAACAGTTGCTTGA TTCGCATCCTTATCTTAACATGGTGGTTGTTCCGGTTG GT
Vf_Mt7g076290_001	1522614325	T:T	C:C	ATTGCCTCCAAAACTCTTTCAAGTCAATCTGGCACTCA AGAAGATGCTATAGACGTTGATGATGACACGGTAGAT CTACTTGACGATTGTGATTCTCC[Y]GAGATTTGCGAAG AAGATGGGTGTAGTTTTTNGTTGACAGATGAAAATAT GGACCTTGTTGGAGCTGCTTATCCAGAAGAAGTGAGA AGGTTT
Vf_Mt7g080040_001	1028278339	C:C	T:T	CCACTGCAGAAGTGTGTGATGCAAACCCACAGTTAAT TTTAAGCGGGGAGCTTCGTGCCCTTCAGCCGATTTTCC AGATTTACGGTCGAAGACAGGTTTT[Y]TCCGGACCAA TAGTTACCTTAAAGGTGTTTGAGGACAATGTCTTAGTT CGCGAGTTTCTTGAAGAGAAAGGAAATGGAAGAGTG CTTGTTGTGGA
Vf_Mt7g084010_001	1886939595	G:G	A:A	CAAGTAGAGAAATGATAAGCCAAGCACATGCAGCAG CTTGGCGCTTGAGAAAGAGTTCAGCTGGTAAGGCTGG GGGTTTAGGACAGTTTGGCAATTGTAA[R]GTGTGGGG CAATGTAGACCCAGACAAAAGTGGACCTATCTCAGGC TCACCAGATATGAGTGATGTCAGGAAGAATTCAAATT ATCAAAAATGGGAAGTCA
Vf_Mt7g086700_001	1840921946	G:G	A:A	TAGGTTGCGGGGGAACTGGTAATATAATGTGTACAGA ATGCGGGGGACGAGGTCATCTGGGACCCAAATGATC A[R]AGTCTGACTTTTTAGAATTTGTTGTAYCACCAACA TGTAATGTTTTGGTTGTCATTGCTTGTAAGTCATAATC ATTTTCAATTTT
Vf_Mt7g088140_001	1655084987	C:C	T:T	TTGGACAGGCTTTGACAGCACTACTTGCATCTCAGGGT GCTACTTTGGGGCTGAGTGATTTATTGGGGAGTCTTG GAACAAGTAATGCTCTTAATCATGG[Y]GGTGTTCCAG CTGCAGGTCATGGAGTTCAGAGTGGGTACAGTGCTCA GCCTACTATAAGTCCATCAGTTATGGGAGCCTATGGG AATTCAGTACCRCAGG
Vf_Mt7g092450_001	1245979661	C:C	T:T	AGGAAGGTGCAGAGGAGGAGGATCCGGATAACAAAA ACGATGACATTGATGGTTTAGGAGAACCGATTGAAAC [Y]GAGGAAAAGGAAGATGATCAACCCGTYGAGCGTC CTCGAAGGACATCCAAGTATATGACTAAGTACGAGCG YGCTAGGATTCTCGGTACCCGCGCTCTTC
Vf_Mt7g093500_001	1896275254	C:C	T:T	AAGCCTACATCAACATTCGTAAGGGGAAGATCATCCC CGGCTACGAGGTGAATCTCACTGTTTCATGGGAAGGT GAAGCGAAAGACACCGAAGGTAATTC[Y]TTGATGAAA GTTAACGGTACTGTTGAAATCCCTTATATCTCCGATGA GAACGCGGATGAGGATCCTGATATTAGGGTTACGGTT GAAGATGA
Vf_Mt7g093640_001	1494862198	C:C	G:G	GGTGGAAGGGTGATTGTCAGCCATCCTCAAGGGAGA GAAATACTGAAACAGCAACGA[S]AACAGTACCCGGAA GTAGTAGTTTCTGATCTTCCTGACAAAGCACATTTACA AAGTGTTGCAGCTGCTTATTCCTTTRATGTGGCTGAGT TTGTGGATGA
Vf_Mt7g098440_001	1486389747	A:A	G:G	CTACAAGAAGACCTCGAGACGGAAGTTGTGGTGGAG GAAACTCCTACACAGGAGGCATATCAACCGGTGCATA ATGTTGCTCATACTATCCGTGAA[R]CCCCTGTCGCCCT TGTGGAAGAGTCTTTTGAGGAGCCTGCTAAGAAGACT TATGCATCTATT

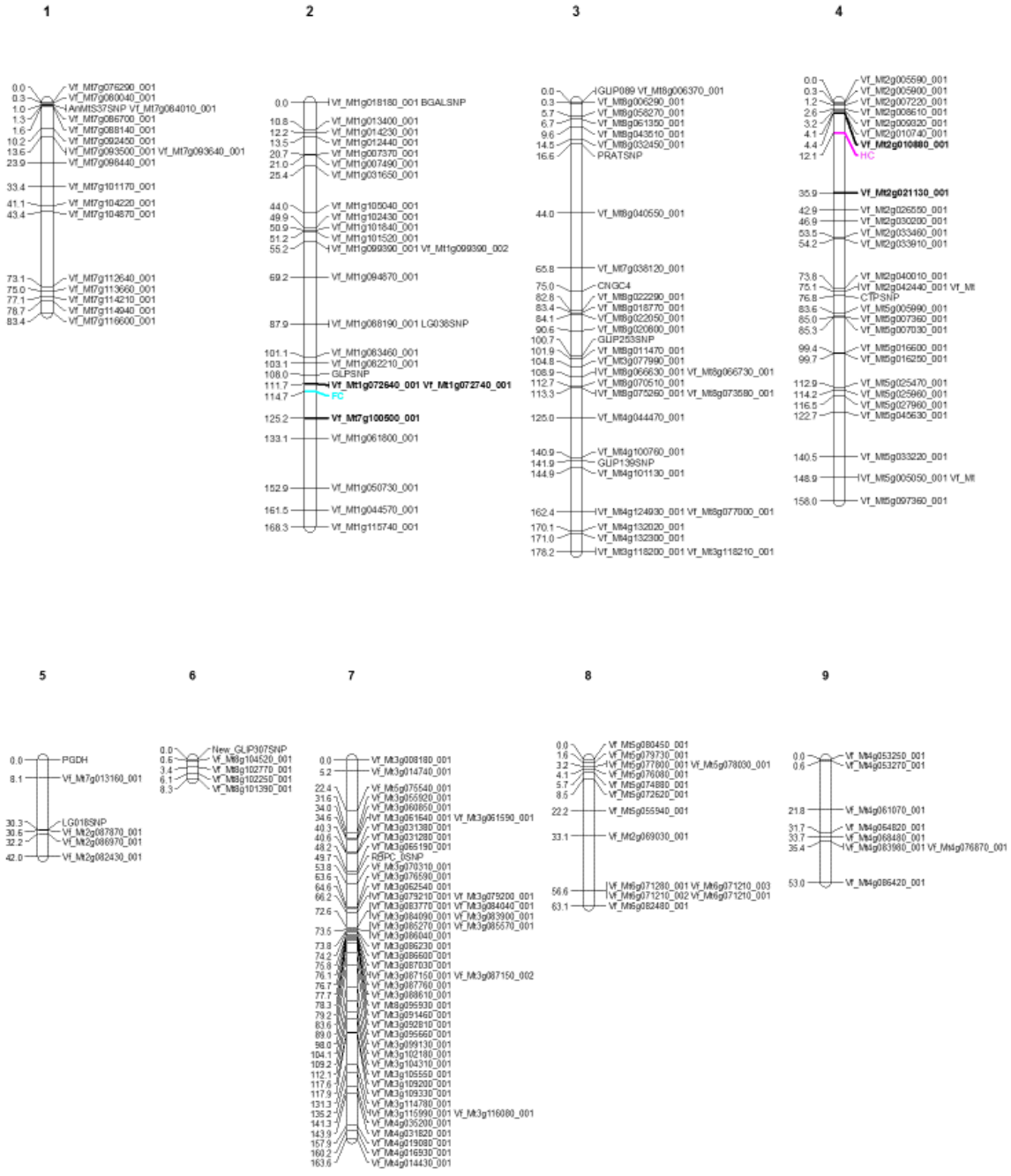
Vf_Mt7g100500_001	1587642015	T:T	C:C	TCTAAAAATGGAGACTCTGAACACATTGCCGCACAGA CGTTCCCTTTCCGTGAGTTAGCAGCTGCTACTAGAAA[Y]TTTAGAGCAGAATGTCTTCTGGGCGAGGGAGGCTTT GGCAGAGTGTACAAAGGGCATTTGGAA
Vf_Mt7g101170_001	1149845815	C:C	T:T	CTTCTCAAGTATGTTACAACSGTGAAGAACTAGAGTCC TYGATGATATATTGGTGCTGCGAGCTCTCAAATTATGC TGAATAT[Y]GTCTTGTGTTTTCTGGTGACTCTWGAGTC TGGCGCAGATAAGATTCTCGTTTAGGTTATTACGCATA GTACTTTGCTTCCATATATTCAACATTA
Vf_Mt7g104220_001	1076102851	A:A	G:G	TAATCAAACAATCACAAACGTAATTCGCATCAAAGAAT TAGGAGTCAAAAAAGTGATTATATCAAATCTTCAACCA GTTGGATGTCTACCTTC[R]GTRACAGCTTCATCCTCATT CAAACAATGCAACGAAACATCGAATAACCTCCTCGTA GCYTACCATAACACGCTCTTAACCGAAGCTGTGACAAA ATTGA
Vf_Mt7g104870_001	1949482251	A:A	G:G	GGAAGAAGGGTTCGCACTGTGATTCAAGTTGTCCGAT CTGACGGTGTACTTGTCATTGGAAAATCACCCGATGA GAAGCAGATTGTTGTGAAAAC[R]TCACCTGATTCACCC CTACCAGCTCCTCATACACCTTTTGTTGAAGTCATCGGT ATTGTTAGCAGTGACAAGTCCATYAAAGCTGAGTTATG GACCAACT
Vf_Mt7g112640_001	1993252933	T:T	C:C	ATTCTTTGGCGTGGCCAGGAGAAGATTATGAGCGACA TTCTTTCTCGGGAAGA[Y]CGGGAAGAGGGTCCTGCTG GTGTTAAAGTTGATAATAGATTTTAGTTCCTCTTTTGG ATACA
Vf_Mt7g113660_001	1375965655	C:C	T:T	GTGTTGAAATTAGCTTTGAGCATTCATCTTGATGTTCT ATTTATTTTTCCTAGGAATGAAGTCAATTATCAAGAGA TCTGATTAATAAATAGTATTGAAG[Y]ACATGTATCTCT TYGTAACATTTTTGTGCATTTTGAAATTCTAGACAAGCC AATATACTGTGGTAGCTCACACTATTGT
Vf_Mt7g114210_001	1571258715	T:T	A:A	GTTCATTTGTGGTCCGTATAAAAGCTTGATAAATTTGT ACTACTAYTGYG[W]GGGTATCATTAACATTGTTCTGAT CAATCATTTGGTGTCTGCAACTGTGCTCTTTGCCCCAA ATTTGCTGGAGTGGAGGAATGTTTCTTTATCATATTTA
Vf_Mt7g114940_001	1476837292	G:G	A:A	AGGCTTTTCATCACAACTACTTTCAATAATTGAAGCAC GTGATATGGAAGAAAACCAACAACTTGAGAGGAGTTT GCATTCTGAAGATG[R]CGAGTCCCTCTCATTACCTGGA AGAAGAAGTGGGAATGAGCAATGGCGCAAATYCAGA TCTGAGATTGGTTCTGATAACCTAAGTTCATCTGATTG CCCA
Vf_Mt7g116600_001	1047741133	C:C	T:T	AATCTGGAGCTATCAAACATTTGCTTGGCCTTCTAGAC GAGGGACATACATTTGCCATGAAAGATGCTGCGTCAG CCATATTTAACCTATGTCTAGTCCA[Y]GATAATAAAGA GAGGACCGTGCGAGAAGGGGCGGTTAAGGTCATTTT GAACAAGATCRTGAACCACATTCTAGTTGACGAGCTG TTGGCTATACTGGCAC
Vf_Mt8g006290_001	1950735554	T:T	G:G	TCGTAATGAGACACGGCGATCGCATCGACAACTTCGA TCCGTTATGGATATCCACCGCGCAACGACCGTGGGAC CCACC[K]CTCGTTGAAGAGGGTCGGGTTCGGTCCTTCT GTACGGGTCGGAAGTTTCGCAATCTTTTCAAATATCAG CTTCATCGTGT
Vf_Mt8g006370_001	1291053027	C:C	T:T	GAAGACTTGACATATGAAAGGCGGAGAGAATTCTTAG AGAAGCATCATCTGCCAGAAGAACTCCCCATTGTTTCC TTCCGCAC[Y]GAAGCTGGCATTTCACCTGCTGTTTTAG CTACATTATCTCATGTTGCTCACGCAGAGCTACCCCTA GTTGCTTCAGCTGGAGAAAGTACAAAACTCCCTG
Vf_Mt8g011470_001	1449560922	A:A	G:G	AGAAAAGAGGGTTGTGTTTCTTTCCAAGTTGCTTGTTT CCAACTTGCATTGCTTGAGACAAGAAAAATTTGGGAG ATT[R]ATGTATATAGAAGTGATTAAATGGAGACATTAT TGATTAGGAGATAGATATTCTCTTGGATCATATTTATT TGAAAGTATATTTAAGGACCAAGAAGATT

Vf_Mt8g018770_001	1890078776	A:A	C:C	AGGTAATTGCCATAACTGTTGAAGATGAAGAGGACAT TGCAAAGTTTAAAGATTACAAACCTTCAGCATCTGAAT CTAGTGACCC[M]CCTGCCAAAGAAACACCTGCCCCAC CTCCCCCAAAGAAAGAGGTGGCAGAGGAGCCTGCAC GAGAACCTGAGCCAAAGGCTTCCAAGCCTAGTGCACC AC
Vf_Mt8g020800_001	1694454354	G:G	A:A	AGGAAATGGCTCAGAAAGCTCTTGAATGTCCTTGTATT GCTGACCTTAGAA[R]TGGTGCTTGTGGGTCTCAWTTC TCAGAGGCTTTCCTTTGTTWTATCAAGAGTACTTCTGA AGAAAAG
Vf_Mt8g022050_001	1367370002	A:A	T:T	AGGCAACACAGCTTGATCTCCATCGTCATTTCTATACC CTTGGAGCTGGTGTGATAGAGGAGGTGCGTGTCCAAC GTGATAAAGGATTTGGGTTTGTGAGGTTTAGTACTCA CGCGGAAGCAGCTCTGGCAATTCAGATGGGAAATG
Vf_Mt8g022290_001	1534230610	T:T	C:C	GAGGGAATCTGAAGCAAAAGCATTGGAAGGAGGTGG CGGAGATCGTTAACGGGAGAGAGGATTATTTGAAGG CGCCGAAAACTGATATCCAGTGYAAGAA[Y]CGRATCG ATACGRTTAAGAAGAAGTATAAATCGGAGAAGGCTAA GATCG
Vf_Mt8g032450_001	1578510453	T:T	C:C	GCTCCCGTTCCTGAAAACTGCAACACCGATGGCTTCTT TGATTCTTTCCTTCGGCATTTCATCAAAGTCGATCAAAT [Y]CAACCTGGTCGARTCTCTTGTACTCTTGTCGCCAAA CCACCTCTCTGT
Vf_Mt8g040550_001	1736692999	G:G	A:A	AATACTGCTCGCCTGCTGGCATTGCTCATGCACAGGG GTTTTAAGTTTTCCATTACCAACTCCATTTTGTGGCAGA CGGCTGAACGATCACTGATCTGGA[R]GCAGTCTCAAG AACAGCAAAATGTTTACCCTCATTGCCCCTACAAAGCA AGGGATATGA
Vf_Mt8g043510_001	1251854208	T:T	C:C	GAAATGGTGTTCATCCCAATGCTGATAAGATGTTGAA GCCGAGTTTTGGGGAGCCTGGACAGTGTTTTCCATTG AAAGGCAGCAG[Y]GGGTTTGTTCAGATTAAGTTGCGC ACCGAAATTATTCCTGAGGCTATTACCTTGGAGCATGT AGCAAAG
Vf_Mt8g058270_001	1460062632	G:G	A:A	TAGTTGGTCCTGCTGCATCAGATAACTGAGAGGATGA ATTTCTGCTTTTATGCTTTGCAGCTATCTCCTTGGGGTC CTTCAACAGTAGTTTTATTAGTTC[R]GATGCAACTACA CTACCACAAGCCTTATTATCATTCATCTGAGAATCTATC AGAGGATTAGTTTCCCTATCAACCTTCTTACATAGCCA GTATATAGATG
Vf_Mt8g061350_001	1434045736	C:C	A:A	TCAATTGAGAAGATGATTGGAGAGGCATACTCTGTAA TTGACAAAGCAGTGAAAGTGGGCACATTGCATAGGAA CACYGGAGCACATCGCAAGTCT[M]GGCTTGCCAGAA GAAAGAGAGCAATAGAAATTCATCATGGTTGGTATAC TCCTGTTMCTCAA
Vf_Mt8g066630_001	1001396667	G:G	C:C	GAACTCAACTCCCTTCTCCGATCGCTTGGTTTAAAGCC AAGCGCTGAACAGTTGGAAGGATTCATTCAGCGAGCA GACACGAACAACAACGG[S]CTGATTGAATTCTCGGAG TTCGTTGCTCTGGTGGCGCCGGAGTTGTTGCCGGCGA AGTCGCCGTATACGGAGGAGCAACTTCGGCAGCTGTT CAGGATGT
Vf_Mt8g066730_001	1162249213	C:C	T:T	GGATCCAATCGAGCGAAAGCGATAGGGAAGATGTTCT TCAAGACGAAACTGTGCTGCAAGTTTCGCGTGGGGAC TTGTCCTTACATAACGAA[Y]TGTAACTTTGCTCACAGT GTTGAGGAGCTTCGTCGGCCGCCGGCGAATTGGCAG GAGATTGTGGCGGCGCATGAGGAAGAGAGGAACGTG CTTGAGCAGC
Vf_Mt8g070510_001	1532181446	A:A	G:G	TGGACTGGGATAAGAATGGAATGGTGAACTTCAAGG AGTTTCTTTTTGCATTTACACGGTGGGTTGGAATTGAT GAAGGTGAGGATGAGTAAAATGCATA[R]GCCTGAAT CATGTCATCTAAATTGTCTTTGGCTCTATCTTCTGCTAA TAGTMTGTAGCTTTATGCTTCAACCAAAACAAAAANG GCTTTTCCAGTTTTC

Vf_Mt8g073580_001	1694807519	A:A	G:G	ACGAAGGAATATACGACAATGATTATCAAGGAAGCCC TCCTCCTTCCAGGGTCAAGAGGTTCAGTGATAAAGATC AAATGGAATACACTGAGATATTAAA[R]ATGCGTGGAC TTCCATTCAATGTGACTAAATCTCAAATTATTGARTTTT TTAAAGATTACAAACTGATAGAAGATAGGGTGCACAT TGCATGTCGCCCAG
Vf_Mt8g075260_001	1694807519	C:C	T:T	ACGAAGGAATATACGACAATGATTATCAAGGAAGCCC TCCTCCTTCCAGGGTCAAGAGGTTCAGTGATAAAGATC AAATGGAATACACTGAGATATTAAA[R]ATGCGTGGAC TTCCATTCAATGTGACTAAATCTCAAATTATTGARTTTT TTAAAGATTACAAACTGATAGAAGATAGGGTGCACAT TGCATGTCGCCCAG
Vf_Mt8g077000_001	1241799286	C:C	T:T	ACAATGGGTTAGAGAAGAAGACTTGGCAAAGGACCT CAAATTGCATACAAAGCAACTTCGTCGAACT[Y]TGCG GTTCTTTGAAGAAGAAAAGATTATCACTCGAGATCATA GGAGAGAGACAG
Vf_Mt8g095930_001	1102100487	C:C	T:T	TCCTACTAGTTTCAATGCTCCGGTTTYGAGTTCAATGG ATTTATCGTCCCC[Y]GGTAAATCGGACGGAATGTCGG TTTCGATGGACGAGACAATGTCGTMTTGTRATTCTTTC AAGAGTCCTGATATTGAGTATGTGGATAACAGTGATG T
Vf_Mt8g101390_001	1033557310	G:G	A:A	ATGTCAATGAAGGGTTTAGTGGAGGTGAACGAAAAC GTAATGAAATCTTGCAGCTTGCGGTTTTGGGGGCAGA TTTGGCAATTTTGGATGAAATTGATTCAGGATTGGATG TTGATGCGCTCAGGGATGTGGCAAGTGCGGTTAATAA AATTCTTACCCCAGAGAATTCTTTGTTGATGATTACTCA CTATCGACGAATTC
Vf_Mt8g102250_001	1235755982	T:T	C:C	CCACACTCACCTTCTGGTGACACAGTTCTCATAACCAA AGATGATCTTGAAGGCTATCATAGCACATCTGATACAA GTATGGAGCTGGCTATCGTG[Y]CTAAAGTGCCAAAGA CTAAGCCCATGATAGCACGACTATCTTGCCTTTTTGCA TCCTTAAAAGTATCAGGTAGTAGTGTTCCAATGACAAG
Vf_Mt8g102770_001	1293816900	A:A	G:G	TATTTAAGAGGCTTTTGGTTTTTGATCTGTCGTTCAAC GAAATTGCGTCGCTACATGGTGTTTCTAGGGTTTGTAA TACATTGAAGGAGCTCTATGTATC[R]AAAAATGAAGT TACTAAAATCGAGGAGATTGAGCATTTTCATCTGTTGC AGATTTTGGAACTYGGCTCGAATAAATTACGGGT
Vf_Mt8g104520_001	1126269050	C:C	T:T	ATAGGTTTCATAGCTACTGGAATGTTATTACCACTTATT GCTGTCATAATATGTCGAAATAAAG[Y]GAAAATATGG AAGGTTATATTGACACAATTCAGCAAAATAACAAGGA ATAATCAACATATAGAAGCCTTTCTTGAGAGCCAAGGT CCTTTGAATCTAAAA

Appendix 4.

Supplementary figure 3.2. Linkage map for the RIL population Disco/2×ILB 938/2, the map spanned 917.64 cM over 9 linkage groups.



Appendix 5. Stipule spot pigmentation (SSP) and hilum colour (HC).

Stipule spot pigmentation (SSP)

During this M.Sc. research project data regarding stipule spot pigmentation, defined as presence of a black spot on the stipules, visible at early stage during the plant growth, were recorded in two different populations segregating for this trait. In particular, in the population Disco/2×ILB 938/2, used for mapping *zt2* (Research component 1) and in the population Disco/2×CDC Snowdrop used in research component 2. Despite the study of the genetics of SSP was not a goal of this M.Sc. project here I present some data and possible novel discovery related to it.

From previous studies by Khazaei et al. (2014) we know that there are 3 different genes which determine the presence or absence of a black spot in the stipule at the base of each leave in faba bean.

Two genes *ssp3* and *ssp2*, were thought to be linked to the two low tannin/white flower genes *zt1zt2.* In their recessive forms *ssp3* and *ssp2* determine the absence of the black spot in the stipule. The third gene *ssp1* decouples the stipule spot pigmentation from that of the flower, because it was found only in accessions such as ILB 938/2 which has spotted (wild type) flower and uncoloured stipule. All information in Khazaei et al. (2014).

The segregation ratio for SSP observed in the population Disco/2(ssp2) × ILB 938/2(ssp1) respects the expected mendelian ratio for a trait controlled by two independent recessive genes. Results are presented in Table 1.

Table 1. Observed segregation ratios, value of χ2 test and corresponding P value for stipule spot pigmentation (SSP)spotted versus colourless, black and uncoloured seed hilum colour (HC).

Generation	SSP				HC			
	Observed	**(expected)**	**χ^2**	***P***	**Observed**	**(expected)**	**χ^2**	***P***
F2	163:128	9:7	0.006	0.9353		(3:1) [b]		
F3	110:179	25:39	0.1215	0.7275	142:90	5:3	0.1655	0.6841
F4	64:158	81:175	0.8115	0.3677	117:88	9:7	0.0564	0.8122
F5	51:155					17:15 [b]		
F6	40:135				88:87	1:1	0.0057	0.940

[b] Data are missing for this trait and generation.

However, the ratio for the stipule spot pigmentation (SSP) character in the population Disco/2 × CDC Snowdrop at F2 did not meet the ratio of 9:7, and a new phenotype, spotted flower and uncoloured SSP, appeared at this generation. This was unexpected because we thought that SSP and flower colour (FC) characters segregated together in this population, being the result of linked genes. In fact, both parents of our population present white flowers and uncoloured stipule, but for the action of two different genes at both loci. Disco/2 carrying *ssp2* and *zt2* and CDC Snowdrop carrying *zt1* and *ssp3*. Because of this genetic configuration of the two parents we should have obtained the same ratio of 9:7 between spotted and uncoloured stipule observed for the flower colour and only two phenotypes, being spotted flower spotted stipule (*ZTSSP*) and white flower uncoloured stipule (*ztssp*).

On the other hand, the presence of the new phenotype spotted flower uncoloured stipule (*ZTssp*), suggests that the linkage between one of the two couples of genes was broken. I suggest

that this was the linkage between the genes *zt1* and *ssp3* or their dominant alleles *ZT1* and *SSP3*. The linkage between *zt2* and *ssp2* was demonstrated to be strong and stable (Khazaei et al. 2014) with a series of crosses involving accessions recessive for *zt2* and *ssp2* with wild types and ILB 938/2 which is source of *ssp1*. A study regarding crosses between sources of *zt1 ssp3* and wild types, or *zt2 ssp2* specifically targeting the genetics of SSP has not been published yet, so this drove me to the conclusion that the linkage between *zt1* and *ssp3* could be somehow broken in the population Disco/2 × CDC Snowdrop.

However, the study of the genetics of the stipule spot pigmentation was not a goal of this M.Sc. project, therefore we didn't continue with further experiments. Table 2 presents the results of the χ^2 goodness of fit test for the F2 generation of the population Disco/2 × CDC Snowdrop.

Table 2. Observed segregation ratios for the characters Flower colour (FC) and Stipule Spot Pigmentation (SSP). Values of χ^2 test and corresponding P values are given for the entire population and for the sub populations (F2 families).

F2 FAMILY	FC Spotted	FC White	χ^2 (9:7)	*P* value	SSP Spotted	SSP Colourless	χ^2 (9:7)2	P value3
Total	451	382	1.504	0.22	423	422	13.6679	0.0002**
A1 [a]	45	50	3.0451	0.081	43	57	7.134	0.0076**
A2	65	60	0.9175	0.3381	57	68	5.7611	0.0164*
A3	71	48	0.5636	0.4528	68	53	0.0001	0.9909
A4	64	50	0.0006	0.9812	60	55	0.7764	0.3782
A5	52	51	1.3908	0.2383	49	56	3.9185	0.0478*
A6	47	52	3.0978	0.0784	44	56	6.0978	0.0135*
A7	72	39	3.3475	0.0673	70	42	1.7778	0.1824
A8	35	32	0.438	0.5081	32	35	1.9619	0.1613

(*observed value is significantly different from expected value P<0.05, ** observed value is strongly significantly different from expected value P<0.01).

[a] See Research component 2 for explanation about F2 families

It is noteworthy that in the population Disco/2×ILB 938/2 the phenotype colourless stipule/spotted flower is present since it is typical of ILB 938/2 (*ssp1ZT1ZT2*) which decouples the stipule spot pigmentation from that of the flower.

Hilum colour (HC)

The trait HC was recorded as an additional morphological marker in the RIL population Disco/2×ILB 938/2.

At F6 the segregation ratio for the trait (black vs uncoloured) was 1:1 (Table 1), proving that this character is controlled by a single gene with the dominant allele conferring the black hilum and the recessive allele the colourless hilum (see segregation ratios for all the generations).

Hilum colour (HC) was mapped in linkage group number 2 of the linkage map constructed to map *zt2* (Figure 1) and in Chromosome 1 of the 2014 *Vicia faba* consensus map.

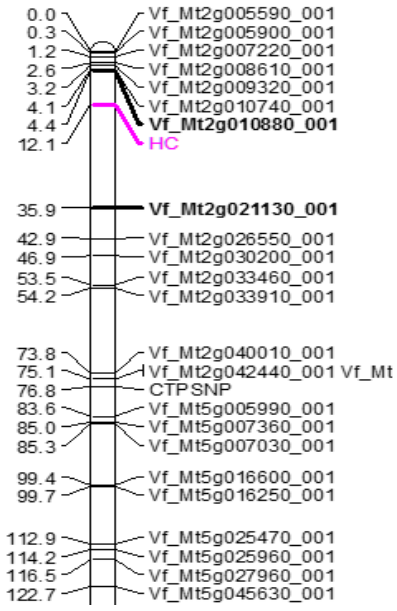


Figure 1. Hilum colour locus (HC) in the linkage map

This result is consistent with that reported by Khazaei et al. (2015) who mapped HC in Chromosome 1 in the same map using the biparental RIL population Melodie x ILB938/2 (khazaei et al 2015). The same SNP marker Vf_Mt2g010880_001 which is the closest to HC in our map, was found to be the closest SNP marker to the HC locus in the study by Khazaei et al. (2015).

Appendix 6.

Supplementary table 5.1. List of standards for the phenolic compounds analyzed with the targeted method.

Name	Company	Formula	Mw
3,4-dihydroxybenzoic acid	Sigma-Aldrich	$C_7H_6O_4$	154.12
4-Aminosalicylic acid	Sigma-Aldrich	$H_2NC_6H_3$-2-$(OH)CO_2H$	153.14
4-hydroxybenzoic acid	Sigma-Aldrich	$C_7H_6O_3$	138.12
Gallic acid	Sigma-Aldrich	$C_7H_6O_5$	170.12
Salicin	Sigma-Aldrich	$C_{13}H_{18}O_7$	286.28
Syringic acid	Extrasynthese	$C_9H_{10}O_5$	198.18
Vanillic acid	Sigma-Aldrich	$C_8H_8O_4$	168.15
Vanillic acid 4-β-D-glucoside	Sigma-Aldrich	$C_{14}H_{18}O_9$	330.29
Vanillin	Sigma-Aldrich	$C_8H_8O_3$	152.15
Tannic acid	Sigma-Aldrich	$C_{76}H_{52}O_{46}$	1701.2
Vanillin-(ring-13C6)	Sigma-Aldrich	$^{13}C_6C_2H_8O_3$	158.1
4-hydroxybenzoic acid- $^{13}C_7$	Sigma-Aldrich	$^{13}C_6H_4{}^{13}CO_2H$	145.07
Caffeic acid	Sigma-Aldrich	$C_9H_8O_4$	180.16
Chlorogenic acid	Sigma-Aldrich	$C_{16}H_{18}O_9$	354.31
trans-ferulic acid	Sigma-Aldrich	$C_{10}H_{10}O_4$	194.18
trans-p-coumaric acid	Sigma-Aldrich	$C_9H_8O_3$	164.16
trans-3-hydroxycinnamic acid	Sigma-Aldrich	$C_9H_8O_3$	164.16
Ferulic acid-d3	TRC	$C_{10}D_3H_7O_4$	197.2
4-Hydroxy-6-methylcoumarin	Sigma-Aldrich	$C_{10}H_8O_3$	176.17
Phloretin	Sigma-Aldrich	$C_{15}H_{14}O_5$	274.27
Xanthohumol	Extrasynthese	$C_{21}H_{22}O_5$	354.4
5,7-dimethoxyflavone	Sigma-Aldrich	$C_{17}H_{14}O_4$	282.29
Apigenin	Extrasynthese	$C_{15}H_{10}O_5$	270.25
Apigenin-7-O-glucoside	Extrasynthese	$C_{21}H_{20}O_{10}$	432.38
Apigenin-7-O-neohesperidoside	Extrasynthese	$C_{27}H_{30}O_{14}$	578.52
Apigenin-7-O-rutinoside (Isorhoifolin)	Extrasynthese	$C_{27}H_{30}O_{14}$	578.53
Apigenin-8-C-glucoside	Extrasynthese	$C_{21}H_{20}O_{10}$	432.38
5,7-dihydroxyflavone	Extrasynthese	$C_{15}H_{10}O_4$	254.24
Diosmetin	Sigma-Aldrich	$C_{16}H_{12}O_6$	300.26
Diosmetin-7-rutinoside	Sigma-Aldrich	$C_{28}H_{32}O_{15}$	608.56

Flavone	Sigma-Aldrich	$C_{15}H_{10}O_2$	222.24
Luteolin	Extrasynthese	$C_{15}H_{10}O_6$	286.25
Luteolin-3',7-di-O-glucoside	Extrasynthese	$C_{27}H_{30}O_{16}$	610.53
Luteolin-4'-O-glucoside	Extrasynthese	$C_{21}H_{20}O_{11}$	448.38
Luteolin-7-O-glucoside	Extrasynthese	$C_{21}H_{20}O_{11}$	448.38
Luteolin-7-rutinoside	Sigma-Aldrich	$C_{27}H_{30}O_{15}$	594.52
Luteolin-8'-C-glucoside	Extrasynthese	$C_{21}H_{20}O_{11}$	448.38
Tangeretin	Extrasynthese	$C_{20}H_{20}O_7$	372.38
Vitexin-2''-O-rhamnoside	Extrasynthese	$C_{27}H_{30}O_{14}$	578.53
Kaempferol-3-O-rhamnoside	Sigma-Aldrich	$C_{21}H_{20}O_{10}$	432.38
3,3',4',7-tetrahydroxyflavone	Extrasynthese	$C_{15}H_{10}O_6$	286.25
Isorhamnetin	Sigma-Aldrich	$C_{16}H_{12}O_7$	316.26
Kaempferol	Sigma-Aldrich	$C_{15}H_{10}O_6$	286.24
Kaempferol-3-O-robinoside-7-O-rhamnoside	Sigma-Aldrich	$C_{33}H_{40}O_{19}$	740.66
Kaempferol-3-rutinoside-4'-glucoside	Sigma-Aldrich	$C_{33}H_{40}O_{20}$	756.66
Kaempferol-3-O-D-galactoside	Sigma-Aldrich	$C_{21}H_{20}O_{11}$	448.38
Kaempferol-3-O-glucoside (Astragalin)	Extrasynthese	$C_{21}H_{20}O_{11}$	448.38
Kaempferol-3-*O*-β-rutinoside	Sigma-Aldrich	$C_{27}H_{30}O_{15}$	594.52
Kaempferol-7-O-glucoside	Extrasynthese	$C_{21}H_{20}O_{11}$	448.38
Kaempferol-7-O-neohesperidoside	Extrasynthese	$C_{27}H_{30}O_{15}$	594.53
Myricetin	Sigma-Aldrich	$C_{15}H_{10}O_8$	318.24
Myricetin-3-O-rhamnoside	Extrasynthese	$C_{21}H_{20}O_{12}$	464.38
Quercetin dihydrate	Extrasynthese	$C_{15}H_{10}O_7.2H_2O$	338.27
Quercetin-3,4'-di-O-glucoside	Extrasynthese	$C_{27}H_{30}O_{17}$	626.53
Quercetin-3-O-galactoside	Extrasynthese	$C_{21}H_{20}O_{12}$	464.38
Quercetin-3-O-glucopyranoside (Isoquercitrin)	Extrasynthese	$C_{21}H_{20}O_{12}$	464.38
Quercetin-3-O-rhamnoside	Extrasynthese	$C_{21}H_{20}O_{11}$	448.38
Quercetin-3-O-rutinoside	Extrasynthese	$C_{27}H_{30}O_{16}$	610.53
Quercetin-4'-O-glucoside	Sigma-Aldrich	$C_{21}H_{20}O_{12}$	464.38
kaempferol-3-O-glucoside-6-E''-coumaroyl	Extrasynthese	$C_{30}H_{26}O_{13}$	594.53
Quercetin-d3 (major)	TRC	$C_{15}D_3H_7O_7$	305.25
(+)-catechin	Extrasynthese	$C_{15}H_{14}O_6$	290.27
(-)-catechin gallate	Extrasynthese	$C_{22}H_{18}O_{10}$	442.38
(-)-Epicatechin	Extrasynthese	$C_{15}H_{14}O_6$	290.27
(-)-Epicatechin gallate	Extrasynthese	$C_{22}H_{18}O_{10}$	442.37
(-)-Epigallocatechin	Extrasynthese	$C_{15}H_{14}O_7$	306.28
(-)-Epigallocatechin gallate	Extrasynthese	$C_{22}H_{18}O_{11}$	458.37
(−)-Gallocatechin	Sigma-Aldrich	$C_{15}H_{14}O_7$	306.27
±-Catechin-2,3,4-$^{13}C_3$	Sigma-Aldrich	13C3C12H14O6	293.25
eriodictoyl-7-rutinoside	Extrasynthese	$C_{27}H_{32}O_{15}$	596.55
Eriodictyol	Extrasynthese	$C_{15}H_{12}O_6$	288.25
Flavanone	Sigma-Aldrich	$C_{15}H_{12}O_2$	224.25
Hesperetin	Sigma-Aldrich	$C_{16}H_{14}O_6$	302.29
Hesperetin-7-rutinoside	Sigma-Aldrich	$C_{28}H_{34}O_{15}$	610.57
Isosakuranetin	Extrasynthese	$C_{16}H_{14}O_5$	286.29
Narigenin-7-rutinoside	Sigma-Aldrich	$C_{27}H_{32}O_{14}$	580.53

Naringenin	Extrasynthese	$C_{15}H_{12}O_5$	272.25
Dihydromyricetin	Sigma-Aldrich	$C_{15}H_{12}O_8$	320.26
Dihydrokaempferol	Sigma-Aldrich	$C_{15}H_{12}O_6$	288.25
Taxifolin	Sigma-Aldrich	$C_{15}H_{12}O_7$	304.25
Genistein	Sigma-Aldrich	$C_{15}H_{10}O_5$	270.24
Prunetin	Sigma-Aldrich	$C_{16}H_{12}O_5$	284.26
callistephin chloride	Extrasynthese	$C_{21}H_{21}ClO_{10}$	468.84
Cyanidin-3,5-di-O-glucoside chloride	Extrasynthese	$C_{27}H_{31}O_{16}Cl$	646.95
Cyanidin-3-O-rhamnoside	Extrasynthese	$C_{21}H_{21}O_{10}Cl$	468.84
Delphinidin-3,5-di-O-glucoside chloride	Extrasynthese	$C_{27}H_{31}O_{17Cl}$	662.98
Delphinidin-3-O-rhamnoside chloride	Extrasynthese	$C_{21}H_{21}O_{11}Cl$	484.84
Delphinidin 3-β-D-Glucoside	Santa Cruz Biotechnology Inc	$C_{21}H_{21}ClO_{12}$	500.84
Kuromanin chloride	Extrasynthese	$C_{21}H_{21}O_{11}Cl$	484.82
Malvidin-3,5-di-O-glucoside chloride	Extrasynthese	$C_{29}H_{35}O_{17}Cl$	691.01
Malvidin-3-galactoside chloride	Sigma-Aldrich	$C_{23}H_{25}ClO_{12}$	528.89
Malvidin-3-O-glucoside (rhamnoside?) chloride	Extrasynthese	$C_{23}H_{25}O_{12}Cl$	528.89
Pelargonin chloride	Extrasynthese	$C_{27}H_{31}O_{15}Cl$	630.97
peonidin-3-5-di-O-glucoside chloride	Extrasynthese	$C_{28}H_{33}O_{16}Cl$	661
peonidin-3-O-glucoside chloride	Extrasynthese	$C_{22}H_{23}O_{11}Cl$	498.85
Petunidin chloride	Extrasynthese	$C_{16}H_{13}O_7Cl$	352.79
Resveratrol	Sigma-Aldrich	$C_{14}H_{12}O_3$	228.24
Resveratrol 3-beta-mono-D-glucoside	Santa Cruz Biotechnology Inc	$C_{20}H_{22}O_8$	390.39
Resveratrol-(4-hydroxyphenyl-13C6)	Sigma-Aldrich	$^{13}C_6C_8H_{12}O_3$	234.2
Procyanidin A2	Extrasynthese	$C_{30}H_{26}O_{12}$	576.52
Procyanidin B1	Extrasynthese	$C_{30}H_{26}O_{12}$	578.52
Procyanidin B2	Extrasynthese	$C_{30}H_{26}O_{12}$	578.52
Procyanidin B3	Adooq	$C_{30}H_{26}O_{12}$	578.52
Procyanidin C1	Sigma-Aldrich	$C_{45}H_{38}O_{18}$	866.77