# Association analyses to genetically study reproduction and seed quality features of faba bean (*Vicia faba* L.)

**Winda Puspitasari**

Georg-August-Universität Göttingen

# Association analyses to genetically study reproduction and seed quality features of faba bean (*Vicia faba* L.)

Dissertation
to obtain the Ph.D. degree
in the International Ph.D. Program for Agricultural Sciences in Goettingen (IPAG)
at the Faculty of Agricultural Sciences,
Georg-August-University Göttingen, Germany

**Winda Puspitasari**
born in Jakarta, Indonesia

Göttingen, May 2017

**D7**

1. Name of supervisor: Apl. Prof. Dr. Wolfgang Link
2. Name of co-supervisor: Dr. Christian Möller
3. Name of third examiner: Dr. Cathrin Westphal

Date of dissertation: 28.06.2017

To my dear family

# Table of Contents

**List of Tables**

**List of Figures**

# General Introduction

## Introduction

Faba bean (*Vicia faba* L.) is one of the most important food and feed legumes in the world. Due to its seed protein content, faba bean provides a valuable source of protein for food especially in Mediterranean countries and China. In developed countries, faba bean is mainly used as feed, mainly using seeds, but at times even as straw or silage. The main faba bean producers are China (1.4 Mt), Ethiopia (0.8 Mt), European countries (0.7 Mt; mainly France, U.K, Germany and Italy), Australia (0.3 Mt) and Morocco (0.2 Mt; FAO, 2014).

Faba bean is one of the oldest domesticated crops in the world. Numerous studies try to identify its origin, and all of the findings lead to southwestern Asia as the principal center origin of faba bean (Duc *et al.*, 2010). This was reinforced by the latest finding from archaeological excavation in north-west Syria, which found faba seeds dated to the late $10^{th}$ millennium BP, and in southern Levant, which suggested that the faba bean domestication started as early as the $11^{th}$ millennium BP (Tanno and Wilcox, 2006; Caracuta *et al.*, 2015). The wild ancestor of *Vicia faba* has not yet been identified until now so part of the potential diversity is probably lost or at least cannot be traced. Many attempts have been carried out to cross *Vicia faba* with other *Vicia* species, but this has never been a success due to ovules which stopped to develop or due to aborted embryos (Duc *et al.*, 2010; Wijaya, 2003).

Taxonomically, faba bean belongs to family *Leguminosae* or commonly named*Fabaceae* as well. The family is divided by three sub families and faba bean is included in sub family *Papilionoideae*, together with soybean, pea and chickpea. Faba bean is a member of genus *Vicia*, together with 230 other species which are commonly known as vetches. Considering the seed size, faba bean is subdivided into var. *minor* with small seeds, var. *equina* with medium seeds and var. *major* with large broad flat seeds. The ancient *Vicia faba paucijuga* which has very small round seeds is considered to be one of the subspecies of *Vicia faba*. According to Cubero (1974), *Vicia faba* has four subspecies, namely *minor, equine, major* and *paucijuga*.

Faba bean is annual crop which is sown either in autumn (winter type) or in spring (spring type). Faba bean will grow and develop best in relatively cool conditions, it is even particularly susceptible to high temperature. Winter type cultivars have a photoperiodic

response but not a significant vernalization requirement for flowering (Evans, 1959). They usually have higher number of tillers than spring ones. Faba bean has a taproot rooting system with secondary roots. The roots bear nodules containing the nitrogen-fixing bacteria *Rhizobium leguminosarum*. With this feature, faba bean delivers a benefit to the environment due to its ability to fix nitrogen from atmosphere. In addition, it also has a benefit in agroecosystems, either in crop rotation or intercropping systems in intensive cereal-dominated situations (Köpke and Nemecek, 2010).

Faba bean is an entomophilous plant. The flowers are usually visited by pollinating insects, such as honey bees (*Apis mellifera*), bumble bees (*Bombus* sp.) and solitary bees(Stoddard and Bond, 1987). Both self and cross fertilization can occur in the same plant. Cross fertilization fully depends on pollinator activity while self-fertilization occurs by pollinators or by spontaneous selfing. The rate of cross-fertilization is varying from about 45 – 60% and depends on genetic and environmental factors. Self fertilization if happening without pollinators or without external mechanic stimulus in faba bean is showing the so-called autofertility (Drayner, 1959). The degree of such autofertility varies among genotypes. Previous reports showed that autofertile flowers have fewer and shorter papillae on stigma, longer anther styles which show a nearly rectangular angle to the ovary, greater pollen grain number and early release of exudates from apparently more fragile stigma cuticle than in autosterile flowers (Kambal *et al.*, 1976; Chen, 2009).

Once the seed is set, either by selfing or by crossing, the amount and quality of the seed mass is decisive. Faba bean seeds supply protein-rich feed stuff and provide a valuable composition with a useful balance of carbohydrate, fibre, micronutrient and phytochemicals (Crepon *et al.*, 2010; Yahia *et al.*, 2013; Pasricha *et al.*, 2014). However, faba bean contains anti-nutritive compounds which limit the use in feed and food system and have health impact for human and several animals, such as tannins, vicine and convicine. Tannins were considered to be a main factor of reducing faba bean protein digestibility. Vicine and convicine gained special attention related to human nutrition. Hydrolysis of vicine and convicine produces the aglycones divicine and isouramil, which cause the oxidation of glutathione in red blood cell. This condition can be harmful for humans who cannot regenerate glutathione above normal rate due to genetic deficiency of G6PD (Glucose-6-Phosphate-Dehydrogenase) activity (so-called favism, Baker *et al.*, 1984; Mehta *et al.*, 2000). A single QTL (Quantitative Trait Locus) for vicine-convicine content was identified in

chromosome 1 for the*vc-* gene which reduces the vicine and convicine level by 10 until 20 fold (Khazaei *et al*., 2015). However, further studies are required to know QTL which control the heritable variation of vicine and convicine content in the wild type (normal vicine and convicine content) of *Vicia faba*.

Different from other *Vicia* with 14 chromosomes (2n = 2x = 14), *Vicia faba* is diploid with 12 chromosomes (2n = 2x = 12). Its genome size is about 13,000 Mb. The high DNA content is distributed across only these six chromosome pairs; there is one very large (about 18 µm length) metacentric pair with satellite and five similar (approx. 7-9 µm length) acrocentric pairs (Link *et al*., 2008). The metacentric chromosome probably originated from remote fusion of two telocentric chromosomes (Fuchs *et al.*, 1998). Due to the large size of the chromosomes and due to the ease of handling them (chromosome from root tip meristems is easily stained), faba bean has been a perfect choice for cytogenetic analysis. Several phenomena of DNA were observed for the first time in this species.

With the introduction of molecular tools for faba bean breeding, significant efforts have been made in the last two decades to understand the genetics and genomics of faba bean. Several kinds of molecular markers are available recently which increase the knowledge of genetic diversity and have facilitated genome analyses, and contribute to the exploitation of genetic variation. Restriction fragment length polymorphism (RFLP) has been first employed by Van de Ven *et al.* (1991) in faba bean as a first step to create a linkage map. With the introduction of the PCR (polymerase chain reaction), more effective and efficient techniques were developed. The random amplified polymorphic DNA (RAPD) technique is one of those PCR-based methods that have become widely used in the development of molecular maps for QTL identification. Arbaoui *et al.*(2008) have constructed a map of RAPD for QTL detection of frost tolerance. Amplified fragment length polymorphism (AFLP) is another PCR based marker that has high reproducibility and reliability. Genetic linkage maps and homology study of backcross families of faba bean using AFLP have been reported by Ali (2015). Simple sequence repeats (SSRs or microsatellites) based on di, tri or tetra nucleotide repeats in DNA sequencesis still widely used. Intron-targeted amplified polymorphic (ITAP) markers mapped in *Medicago truncatula*, soybean and lupine were used to develop the first gene-based genetic map of faba bean. Map construction resulting in six linkage group comprising 552 loci generated from 235 faba bean-derived EST-SSRs (expressed sequence tag-SSRs) was reported by El-Rodeny *et*

*al.*(2014). 551 SNPs (single nucleotide polymorphism) and 71 SSRs were employed by Kaur *et al.* (2014), which exhibited 12 linkage groups. Recently, a consensus linkage map which covers a full set of six linkage groups which could even be assigned to the physical chromosomes was reported by Webb *et al.* (2016). The map used SNP markers which were developed by Ellwood *et al.* (2008) and were converted to KASP (Kompetitive allele-specific PCR) by Cottage *et al.* (2012).

## Objectives

The objectives of the first chapter of the present study are to genetically study and quantify level and variation of autofertility in specific winter faba bean breeding germplasm and to identify QTL for autofertility and related traits. Hence, the first part's focus is on fertilization and thus genesis of seeds. The second chapter's focus is on the quality of seeds. It aims todevelop a NIRS-based so-called calibration for vicine-convicine content in faba bean seeds, to study heritability and genetic variation of vicine-convicine content in faba bean, to identify QTL that are responsible for vicine-convicine variation in vicine-convicine-containing (wild-type) faba bean genotypes and to verify whether the mutant allele for low vicine-convicine in faba bean ("vc-"; Duc, 1989) is allelic to a QTL for the variation in vicine-convicine-containing materials. A genome wide association analysis was employed for QTLs investigation.

## References


Ali, M.B., 2015. Genetic linkage maps and homology study of backcross families of German faba bean (*Vicia faba* L.). *Journal of Crop Improvement*, *29*(4), pp.474-490.

Arbaoui, M., Balko, C. and Link, W., 2008. Study of faba bean (*Vicia faba* L.) winter-hardiness and development of screening methods. *Field Crops Research*, *106*(1), pp.60-67.

Baker, M.A., Bosia, A., Pescarmona, G., Turrini, F. and Arese, P., 1984. Mechanism of action of divicine in a cell-free system and in glucose-6-phosphate dehydrogenase-deficient red cells. *Toxicologic Pathology*, *12*(4), pp.331-336.

Caracuta, V., Barzilai, O., Khalaily, H., Milevski, I., Paz, Y., Vardi, J., Regev, L. and Boaretto, E., 2015. The onset of faba bean farming in the Southern Levant. *Scientific Reports*, *5*, pp.1-5.

Chen, W., 2009. Pollination, fertilization and floral traits co-segregating with autofertility in faba bean. *Journal of New Seeds*, *10*(1), pp.14-30.

Cottage, A., Gostkiewicz, K., Thomas, J.E., Borrows, R., Torres, A.M. and O'Sullivan, D.M., 2012. Heterozygosity and diversity analysis using mapped single nucelotide polymorphisms in a faba bean inbreeding programme. *Molecular Breeding*, *30*(4), pp.1799-1809.

Crépon, K., Marget, P., Peyronnet, C., Carrouée, B., Arese, P. and Duc, G., 2010. Nutritional value of faba bean (*Vicia faba* L.) seeds for feed and food. *Field Crops Research*, *115*(3), pp.329-339.

Cubero, J.I., 1974. On the evolution of *Vicia faba* L. *Theoretical and Applied Genetics*, *45*(2), pp.47-51.

Drayner, J.M., 1959. Self-and cross-fertility in field beans (*Vicia faba* Linn.). *The Journal of Agricultural Science*, *53*(3), pp.387-403.

Duc, G., Sixdenier, G., Lila, M. and Furstoss, V., 1989. Search of genetic variability for vicine and convicine content in *Vicia faba* L.: a first report of a gene which codes for nearly zero-vicine and zero-convicine contents. In *1. International Workshop on'Antinutritional Factors (ANF) in Legume Seeds', Wageningen (Netherlands), 23-25 Nov 1988*. Pudoc.

Duc, G., Bao, S., Baum, M., Redden, B., Sadiki, M., Suso, M.J., Vishniakova, M. and Zong, X., 2010. Diversity maintenance and use of *Vicia faba* L. genetic resources. *Field Crops Research*, *115*(3), pp.270-278.

El-Rodeny, W., Kimura, M., Hirakawa, H., Sabah, A., Shirasawa, K., Sato, S., Tabata, S., Sasamoto, S., Watanabe, A., Kawashima, K. and Kato, M., 2014. Development of EST-SSR markers and construction of a linkage map in faba bean (*Vicia faba*). *Breeding Science*, *64*(3), pp.252-263.

Ellwood, S.R., Phan, H.T., Jordan, M., Hane, J., Torres, A.M., Avila, C.M., Cruz-Izquierdo, S. and Oliver, R.P., 2008. Construction of a comparative genetic map in faba bean (*Vicia faba* L.); conservation of genome structure with Lens culinaris. *BMC Genomics*, *9*(1), p.380.

Evans, L.T., 1959. Environmental control of flowering in *Vicia faba* L. *Annals of Botany*, *23*(4), pp.521-546.

FAOSTAT, 2014. Food and Agriculture Organization. Website: http://faostat.fao.org/. Visited at 7th of March, 2017.

Fuchs, J., Strehl, S., Brandes, A., Schweizer, D. and Schubert, I., 1998. Molecular-cytogenetic characterization of the *Vicia faba* genome-heterochromatin differentiation, replication patterns and sequence localization. *Chromosome Research*, *6*(3), pp.219-230.

Kambal, A.E., Bond, D.A. and Toynbee-Clarke, G., 1976. A study on the pollination mechanism in field beans (*Vicia faba* L.). *The Journal of Agricultural Science*, *87*(03), pp.519-526.

Kaur, S., Cogan, N.O., Forster, J.W. and Paull, J.G., 2014. Assessment of genetic diversity in faba bean based on single nucleotide polymorphism. *Diversity*, *6*(1), pp.88-101.

Khazaei, H., O'Sullivan, D.M., Jones, H., Pitts, N., Sillanpää, M.J., Pärssinen, P., Manninen, O. and Stoddard, F.L., 2015. Flanking SNP markers for vicine–convicine concentration in faba bean (*Vicia faba* L.). *Molecular Breeding*, *35*(1), p.38.

Köpke, U. and Nemecek, T., 2010. Ecological services of faba bean. *Field Crops Research*, *115*(3), pp.217-233.

Link, W., Hanafy, M., Malenica, N., Jelenić, S. and Jacobsen, H.J., 2008. Faba bean. *Compendium of Transgenic Crop Plants*, pp.71-88.

Mehta, A., Mason, P.J. and Vulliamy, T.J., 2000. Glucose-6-phosphate dehydrogenase deficiency. *Best Practice & Research Clinical Haematology*, *13*(1), pp.21-38.

Pasricha, V., Satpathy, G. and Gupta, R.K., 2014. Phytochemical & antioxidant activity of underutilized legume *Vicia faba* seeds and formulation of its fortified biscuits. *Journal of Pharmacognosy and Phytochemistry*, *3*(2), pp.75-80.

Stoddard, F.L. and Bond, D.A., 1987. The pollination requirements of the faba bean. *Bee World*, *68*(3), pp.144-152.

Tanno, K.I. and Willcox, G., 2006. The origins of cultivation of *Cicer arietinum* L. and *Vicia faba* L.: early finds from Tell el-Kerkh, north-west Syria, late 10th millennium BP. *Vegetation History and Archaeobotany*, *15*(3), pp.197-204.

Van de Ven, W.T.G., Waugh, R., Duncan, N., Ramsay, G., Dow, N. and Powell, W., 1991. Development of a genetic linkage map in *Vicia faba* using molecular and biochemical techniques. *Aspects of Applied Biology, 27*, pp.49-53.

Webb, A., Cottage, A., Wood, T., Khamassi, K., Hobbs, D., Gostkiewicz, K., White, M., Khazaei, H., Ali, M., Street, D. and Duc, G., 2016. A SNP-based consensus genetic map for synteny-based trait targeting in faba bean (*Vicia faba* L.). *Plant Biotechnology Journal*, *14*(1), pp.177-185.

Wijaya, A., 2003. Towards interspecific hybridization in *Vicia faba* L. PhD thesis, University of Göttingen.

Yahia, Y., Elfalleh, W., Tlili, N., Hannachi, H., Loumerem, M. and Ferchichi, A., 2013. Phytochemical contents and antioxidant activities of some Tunisian faba bean populations. *Romanian Agricultural Research*, *30*, pp.65-74.

# Chapter I. Association analysis of reproductive features in faba bean (*Vicia faba* L.)


## Abstract

Faba bean is partially allogamous plant which allows both self and cross fertilization. Self fertilization which occurs without pollinators or without external mechanic stimulus in faba bean is showing its so-called autofertility. The degree of such autofertility varies among genotypes. The objectives of the present study are to genetically study and quantify level and variation of autofertility in a specific winter faba bean breeding germplasm and to identify QTL which are connected with autofertility and related traits. A genome wide association analysis was employed for QTL investigation. The main genetic materials used in this study involved 200 inbred lines, named Q-set, which consisted of 189 lines of A-set (inbred lines for association study), seven further winter bean lines and four further spring bean lines. The A-set was derived from the so-called Göttingen Winter Bean Population (GWBP). The experiment was conducted in so-called bee-proof isolation houses in 2013, 2014 and 2015. Treatments of "tripped" and "un-tripped" were applied to the faba bean flowers during flowering time. Association analysis between DNA-markers and phenotypic expression of traits was carried out using TASSEL version 3.0. A total of 2018 polymorphic markers were used consisting of 189 SNP (Single Nucleotide Polymorphism) and 1829 AFLP (Amplified Fragment Length Polymorphism). To assess autofertility, the study focused on rate of fertilization, potential pod filling and actual pod filling, especially in the un-tripped treatment. Rate of fertilization of un-tripped treatment was low, with maximum of 37.14%, and high in heritability. Tripping obviously increases the mean values of the three aspects of autofertility. Higher heritability of rate of fertilization in tripped than in un-tripped treatment indicated marked genetic differences for the reaction of the genotypes to tripping. Intensive tripping that has been carried out in 2015 confirmed the result and showed that none of these genotypes showed 100% of rate of fertilization. Winter faba bean has a different, lower level of autofertility than spring beans. Our study resulted in several putative DNA-markers which are significantly related to several of the agronomic traits in faba bean. Nevertheless, no significant marker was found associated to the autofertility-related traits; this finding is further discussed.

## Introduction

Faba bean (*Vicia faba* L.) is one of the most important legume crops due to some advantages as food, feed and its ecological role and services (N-symbiosis and positive impact on crop rotation). However, it has been known that potential yield in faba bean is variable. Insufficiency of pollination can be a major constraint to the potential yield in faba bean, due to lack of autofertility or lack of pollinator activity (Köpke and Nemecek, 2010; Stoddard and Bond, 1987).

The reproductive mode of faba bean is partial allogamy, which means both self and cross fertilization occurs. Cross fertilization fully depends on pollinator activity while self-fertilization can occur by pollinators or by spontaneous selfing. Honey bees (*Apis mellifera*), bumble bees (*Bombus* sp.) and solitary bees visit the flowers and their foraging activity contributes to flower fertilization (Stoddard and Bond, 1987). The contribution of crop pollination is not only demonstrated by dominant species (as above mentioned), but also rare and more specialised pollinator species (Marzinzig *et al.*, under submission). The rate of cross-fertilization is varying from about 45 – 60%, depending on genetic and environmental factors. Such figures depend as well on the actual method of estimation (Hanna and Lawes, 1967; Link *et al.*, 1994; Gasim *et al.*, 2004).

The ability of some faba beans to self-fertilize without pollinators and without external mechanic stimulus is the so-called autofertility (Drayner, 1959). The mechanical stimulus on the stigma which helps fertilization is called tripping. This activity is usually carried out by visiting pollinators and can be imitated manually. The degree of autofertility varies among genotypes; outcrossed plants (F1 hybrids) usually are more autofertile than inbred plants. F1 hybrids showed a superiority of autofertility of more than 100% over the parents (Link, 1990).

The flower of faba bean is complete in having all reproduction organs, but in most genotypes spontaneous self-fertilization nevertheless is incomplete. Some reports showed that autofertile flower have fewer and shorter papillae on stigma, longer anther styles which show a nearly rectangular angle to the ovary, great pollen grain number and early release of exudates from an apparently more fragile stigma cuticle than in highly autosterile flowers (Kambal *et al.*, 1976; Chen, 2009).

An attempt to produce obligate autogamy faba bean has been carried out by breeding for autofertility together with a closed-flower character. A mutation of simple monogenic inheritance induces a high frequency of tightly closed flowers (Poulsen, 1977). The outcrossing rates were reduced, ranging from 5.1 to 22.9% in the selected lines carrying this trait (Knudsen and Poulsen, 1983), but the autofertility level tended to be low.

In spite of this reproductive mode, testing of inbred lines in faba bean breeding nevertheless is a very important step. Seed production of inbred lines may suffer from contamination by cross-pollen if the production was conducted in an open field. To avoid such contamination, seed production shall be carried out in pollinator-excluding cages to enforce self-fertilization. Without pollinator visits, faba bean will produce a very low pod set; hence, manual tripping is needed which is time consuming work and is relatively expensive.

Previous researchers used different parameters to define autofertility, such as (always in un-tripped conditions) number of seeds per flower, pods per plant and seeds per plant (Drayner, 1959; Rowlands, 1964). The study that has been conducted by Stoddard (1986) gave other complex parameters to describe autofertility, with incidence and effectiveness of pollination, incidence of fertlilization in flowers and ovules, and index of fertilization. Moreover, Link (1990) simplified autofertility as seed containing pods per standardized number of flowers (two flowers per inflorescences). This standardization came from his previous study, showing that with such reduced number of flowers all fertilized flowers develop to pods.

Winter faba beans provide some advantages in various aspects. Winter beans are sown earlier, i.e. already in autumn, which allows better use of moisture in spring and better use of residual nitrate in the soil. Winter beans tend to flower and mature earlier than spring beans. This advance of development will bring the advantages of escape from dry phases in June or July and a partly escape from *Sitona* weevil and aphid attacks (Link *et al.* 2010). An important, specific feature of winter faba bean is their capacity to have more and more synchronous tillers than spring beans. The yield potential of winter bean is higher than that of spring beans with, reportedly about 14% in north area of Germany and about 47% in three locations in UK (Herzog and Geisler, 1991; Link *et al.* 2010). Nevertheless, winter faba bean needs some survival traits to escape from winter kills. Survival-related traits of winter faba bean were extensively discussed by Link *et al.* (2010).

However, with all specific features of winter faba bean, little is known about its autofertility performance. Therefore, the objectives of the recent study are to genetically study and quantify level and variation of autofertility in a specific winter faba bean breeding germplasm; and to identify QTLs which are related to autofertility and related traits. A genome wide association analysis was employed for QTLs investigation.

## Materials and Methods

### *Genetic materials*

Genetic materials used in this study involved 200 inbred lines, named Q-set, which consisted of 189 lines (A-set; highly homozygous inbred lines for association study), seven further winter bean lines and four further spring bean lines. The A-set was derived from the so-called Göttingen Winter Bean Population (GWBP). This GWBP was developed in Göttingen starting in 1989 from initially mixing of 11 founder winter bean inbred lines (Hiverna/1-1, Webo/1-1, Wibo/1-1, Côte d'Or/1-1-3, L79/79/1, L977/88/S1wn, L979/S1/1/1sn, Bourdon/1-5, Arrisot/1-1, Banner 1-1, Bulldog 1-4). After nine generations of natural open-pollinated reproduction, 400 lines were pure bred via single seed descent from 400 initial, randomly-taken individuals. A total of 189 lines of these materials were genotyped for the current study; the same genotypes were analyzed for frost and drought features by Ali *et al.*(2016).

### *DNA markers*

A total of 2018 polymorphic markers were used consisting of 189 SNP (Single Nucleotide Polymorphism) and 1829 AFLP (Amplified Fragment Length Polymorphism) markers to study the association between markers and phenotypic expression. After filtering the markers with minor allele frequency of 5%, a total of 1322 markers remained, consisting of 175 SNP and 1147 AFLP markers. Among all 1322 markers, the average LD ($r^2$) was 0.0077 (Ali *et al.*, 2016).

### *Phenotyping of reproductive features*

The seeds were sown in plastic pots containing local compost soil and sand (3:1), one seed per pot on January-February in 2013, 2014 and 2015. The germination was allowed under temperature 5-10 $^{o}$C in green house. After two months, the plants were moved to bee-proof isolation houses each covering an area of 3.5 x 7.0 $m^2$. Every plant was tied with a bamboo stick for standing assistance.

There were two main treatments in the study, tripped and un-tripped. Each treatment had two replicates (one replicate was one plant). All plants (in tripped and un-

tripped treatment) were standardized as can be seen in Table 1. Each plant was topped three nodes after (above) the last standardized inflorescence was reached.

Table 1. Plant standardization in the tripped and un-tripped treatment.

| Year | Type | Number of tillers left | Number of inflorescences left | Number of flowers |
|---|---|---|---|---|
| 2013 | 1 | 3 | 10 per tiller | 2 per inflorescence |
| 2014 | 2 | 2 | 8 in the 1st tiller and 4 in the 2nd tiller | |
| 2015* | 1 | 3 | 10 per tiller | |
| | 2 | 2 | 8 in the 1st tiller and 4 in the 2nd tiller | |
| | 3 | 2 | No reduction of inflorescence | |

* The treatment in 2015 was un-tripped only

The tripped treatment was intended to always allow a nearly complete pods set. Because of the realized incomplete pods set in 2013 and 2014 in the tripped treatment, then in 2015 tripping was only carried out using a smaller number (N=58) inbred lines of the Q-set and in the type of standardization 2. In 2015, tripping was of maximum care and intensity, to verify whether such intensive tripping (more frequent, more carefully conducted, by better trained persons) could indeed bring pod set nearer to complete.

Ten primary traits were observed in this experiment:

1. Plant height (PH, in cm) is main stem height, measured from soil level until three nodes after the last standardized inflorescence.
2. Flowering time (FT) is number of days from sowing time until opening of flower of the second inflorescence (the first inflorescence was discarded and not counted throughout).
3. First flower position (FFP, in cm) is the position of first (earliest, deepest) counted inflorescence, observed in 2014 and 2015.
4. Number of flowers (NF) is the total actual number of flowers per plant after standardization.
5. Number of pods (NP) is the total number of seed-containing pods per plant at maturity.
6. Number of seeds (NS) is the total number of seeds per plant at maturity.
7. Seed yield (SY, in grams) is seed yield per plant.

8. Rate of fertilization (RF, in %) is ratio of number of pods to number of flowers, in percent ($RF = \frac{NP}{NF} x100\%$); per plant.
9. Potential pod filling (PPF, in %) is the ratio of actual number of seeds to the maximum possible number of seeds, in case of all flowers (NF) was transformed into pods $\left(PPF = \frac{NS}{NFx4} x100\%\right)$; per plant. Multiplier of 4 is taken as proxy for maximum number of seeds per pod.
10. Actual pod filling (APF, in %) is the ratio of actual number of seeds to maximum possible number seeds ($APF = \frac{NS}{NPx4} x100\%$); per plant. Multiplier of 4 is taken as proxy for maximum number of seeds per pod.APF shows whether pods show the maximum number of four seeds per pod or less.
11. Thousand kernel weight (TKW, in grams) is the estimated weight of 1000 seeds$\left(TKW = \frac{SY}{NS} x1000\right)$; per plant.

***Statistical analysis of phenotypic data***

In 2013 and 2014, the experiments were conducted in a randomized complete block design with two replicates for each treatment. Analysis of variance (ANOVA) of each treatment in two environments was carried out using PLABSTAT software (Utz, 2001) using the following model:

$$Y_{ijr} = \mu + G_i + E_j + R_r(E_j) + GE_{ij} + e_{ijr}$$

where$Y_{ijr}$= the observation of genotype i, environment j and block r; μ= general mean; $G_i$= effect of genotype i; $E_j$= effect of environment *j;* $R_r(E_j)$= effect of block r within environment j; $GE_{ij}$= interaction effect between genotype i and environment j; $e_{ijr}$= residual error term.

To see the treatment effect and its interactions with the genotypes and environments, an additional analysis of variance was conducted using the model:

$$Y_{ijkr} = \mu + G_i + E_j + T_k + R_r(E_j) + GE_{ij} + GT_{ik} + ET_{jk} + GET_{ijk} + e_{ijkr}$$

where$Y_{ijkr}$= the observation of genotype i, environment j, treatment k and block r; μ= general mean; $G_i$= effect of genotype i; $E_j$= effect of environment j; $T_k$= effect of treatment k;$R_r(E_j)$= effect of block r within environment j; $GE_{ij}$= interaction effect between genotype i and environment j; $GT_{ik}$= interaction effect between genotype i and treatment k; $ET_{jk}$=

interaction effect between environment j and treatment k; $GET_{ijk}$= interaction effect between genotype i, environment j and treatment k; $e_{ijkr}$= residual error term.

An analysis of variance of the un-tripped treatment was carried out as randomized complete block design with ten replicates (2013 (2 rep), 2014 (2 rep) and 2015 (6 rep)) using the following model:

$$Y_{ir} = \mu + G_i + R_r + GR_{ir} + e_{ir}$$

where$Y_{ir}$= the observation of genotype i and block r; μ= general mean; $G_i$= effect of genotype i; $R_r$= effect of block r; $GR_{ir}$= interaction effect between genotype i and block r; $e_{ir}$= residual error term.

Except the block effect ($R_r$), all other sources of variation were taken as fixed effects. Correlation coefficients were calculated for phenotypic correlation between traits using PLABSTAT software. Microsoft Office Excel 2008 was used for graphical display.

***Association analysis***

Association analysis between markers and phenotypic expression of traits was carried out using TASSEL version 3.0 (Bradbury *et al.*, 2007), based on the above-mentioned DNA-marker data of the Q-set inbred lines and on mean values of these inbred lines as resulting from the above-mentioned ANOVAs. The mixed linear model procedure of TASSEL was applied with an optimum level of compression and re-estimate of the variance component estimates of each marker. A kinship matrix was employed, which was developed by using the average genetic similarity among the 11 founder lines as a threshold (see Ali *et al.*, 2016 for details). A false discovery rate of 20% (FDR=0.20) was used to test the statistical significance of marker-trait associations (Benjamini and Hochberg, 1995). Phenotypic effects of the marker loci were calculated as differences between the means of inbred lines when grouped according to the marker classes. Positive value indicate that the specified marker allele is associated with an increase of the trait, while negative value indicates the marker allele is associated with a decrease of the trait. The phenotypic variance explained ($R^2$) by the significant makers was as well determined by TASSEL 3.0.

## Results

### *Genetic variation of reproductive features of un-tripped and tripped treatments*

Three aspects of autofertility were assessed in the recent experiment: rate of fertilization, potential pod filling and actual pod filling in un-tripped treatment. The mean values for these three traits were relatively low compared to the mean values after tripping (Table 2). Tripping obviously increased seeds and pods set. Rate of fertilization of un-tripped plants was on average only 6.69% and of tripped plants was on average 56.63%.

Table 2. Phenotypic results and analysis of variance of 189 lines of A-set of tripped and un-tripped treatment in two environments.

| Trait[§] | Min. | Max. | Mean | Var. cp. (G) | Var. cp. GxE | Heritability (%) |
|---|---|---|---|---|---|---|
| Un-tripped treatment | | | | | | |
| PH | 65.75 | 153.25 | 108.33 | 96.14** | 17.16* | 73.95 |
| NF | 16.75 | 42.00 | 32.32 | 25.66** | 28.65** | 77.48 |
| NP | 0.00 | 15.50 | 2.12 | 3.92** | 2.99** | 75.30 |
| NS | 0.00 | 32.50 | 4.17 | 17.79** | 15.22** | 65.95 |
| SY | 0.00 | 22.96 | 3.57 | 11.44** | 8.49** | 66.32 |
| RF | 0.00 | 34.74 | 6.69 | 29.45** | 13.84** | 66.42 |
| PPF | 0.00 | 21.56 | 3.33 | 8.89** | 5.88** | 56.65 |
| APF | 0.00 | 70.03 | 28.28 | 141.01** | 47.64 | 51.75 |
| Tripped treatment | | | | | | |
| PH | 58.75 | 141.25 | 109.05 | 92.53** | 3.28 | 73.09 |
| NF | 14.50 | 42.00 | 31.11 | 27.79** | 33.31** | 77.86 |
| NP | 6.84 | 24.00 | 16.31 | 7.31** | 3.78** | 70.82 |
| NS | 20.03 | 77.75 | 47.01 | 81.89** | 30.61** | 75.53 |
| SY | 9.73 | 53.28 | 32.35 | 33.29** | 10.10** | 74.91 |
| RF | 32.34 | 83.93 | 56.63 | 50.62** | 23.57** | 62.35 |
| PPF | 23.49 | 74.90 | 41.42 | 55.18** | 23.55** | 74.14 |
| APF | 45.83 | 91.60 | 72.38 | 60.24** | 11.45* | 75.27 |
| TKW[#] | 450.24 | 970.29 | 703.64 | 6944.65** | 1500.95** | 82.33 |

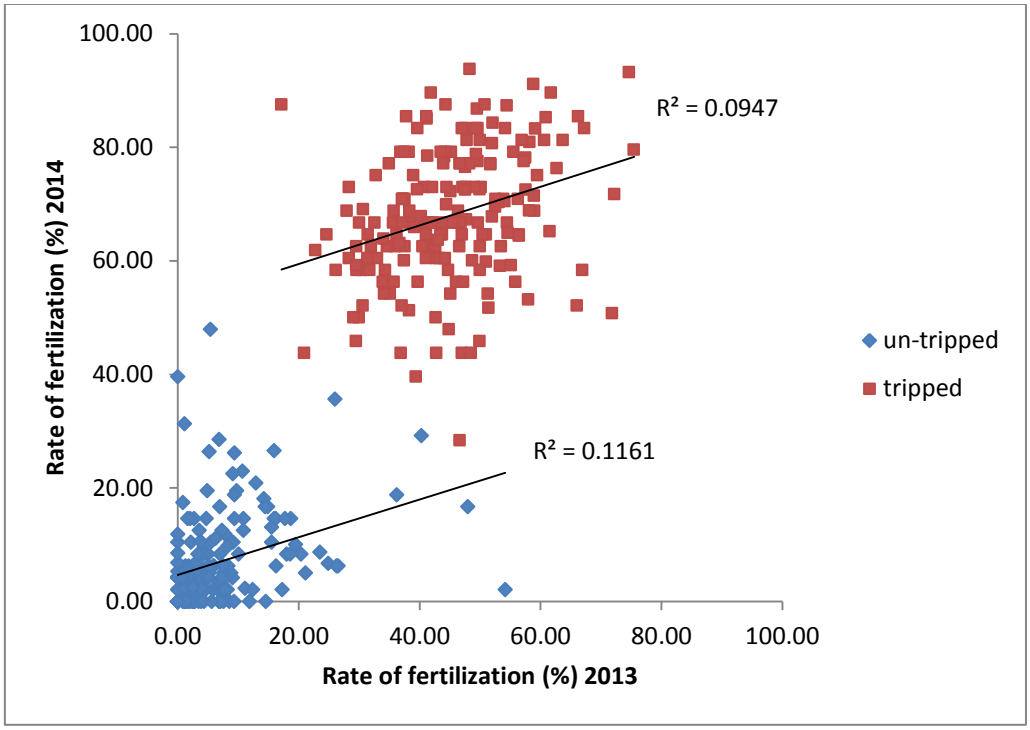
[§] Abbreviations are explained in materials and methods part

[#] TKW for tripped treatment only

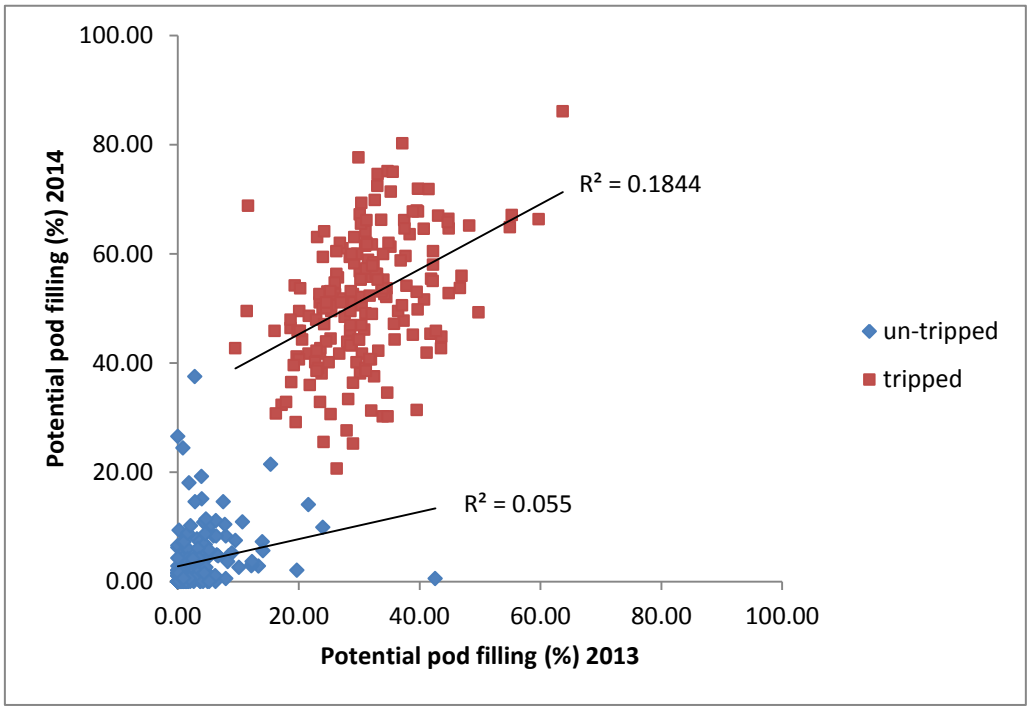
*,** Significant based on F-test for p= 0.05, 0.01, respectively

Analysis of variance showed that all reproductive features showed a highly significant variance ($p < 0.01$) for genotypes (G) of tripped as well as un-tripped treatments (Table 2). A similar tendency was also exhibited for variance of interaction of genotypes with environments (GxE), except for plant height in tripped and actual pod filling in un-tripped conditions (Table 2). Heritability of all traits was medium to high. Heritability of rate of fertilization in un-tripped treatment was higher than in tripped treatment, while of potential pod filling and actual pod filling were lower in the un-tripped treatment.

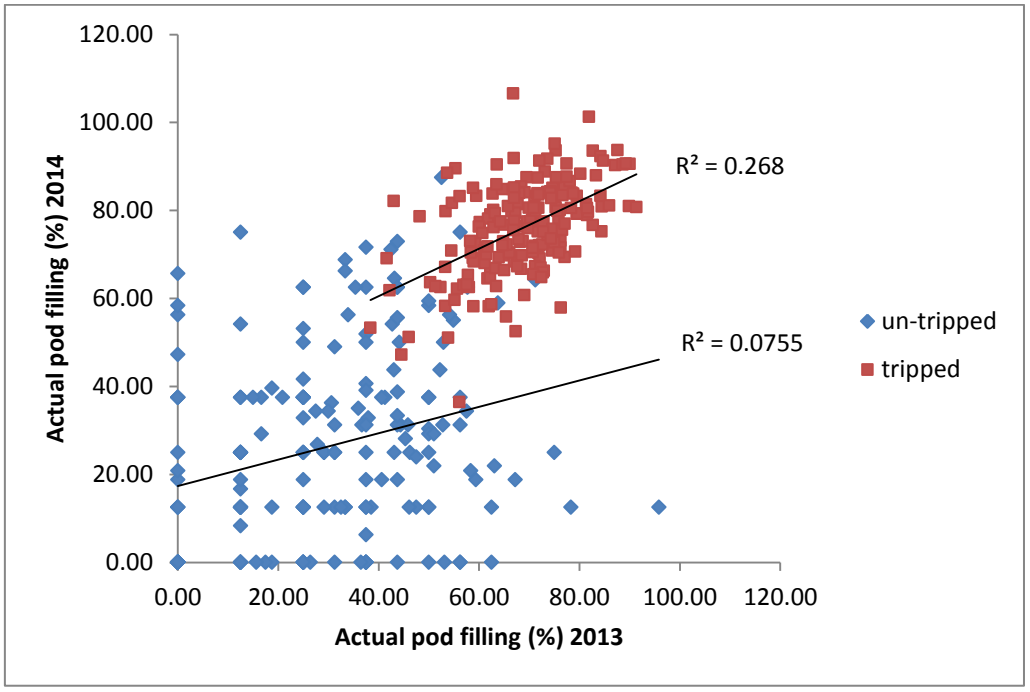
Coefficient of determination of rate of fertilization when comparing the un-tripped results in 2013 with those in 2014 was relatively low, with a significant coefficient of correlation ($r=0.341$, Figure 1). Rate of fertilization of tripped treatment in 2013 and 2014 was also significantly correlated, with coefficient of correlation of $r=0.308$. Potential pod filling in 2013 and 2014 of un-tripped and tripped treatment was significantly correlated (coefficient of correlation was $r=0.235$, $r=0.429$, respectively). Coefficient correlation of actual pod filling in 2013 and 2014 of un-tripped treatment was $r=0.275$ and of tripped $r=0.518$ and both were significant.




(a)




(b)




(c)

Figure 1. Rate of fertilization (a), potential pod filling (b) and actual pod filling (c) of 189 lines of A-set of tripped and un-tripped treatment in 2013 and 2014.

Analysis of variance of the two factors, genotypes and treatments, in 2013 and 2014 showed that rate of fertility, potential pod filling and actual pod filling varied with high significance ($p< 0.01$) due to genotypes (G), treatments (T) and interaction of genotypes - treatments (GxT) and interaction of genotypes with environments (GxE, Table 3). Heritability (means across the two treatments) of these traits was relatively high from 60.63 – 70.88%.

Table3. Analysis of variance of autofertility features of two factors (genotypes and treatments) of 189 lines of A-set across 2013 and 2014, including both treatments.

| Trait[§] | Var. cp. G | Var. cp. T | Var. cp. GxT | Var. cp. GxE | Heritability (%) |
|---|---|---|---|---|---|
| RF | 23.262** | 1247.059** | 30.170** | 10.543** | 65.66 |
| PPF | 17.110** | 725.735** | 27.771** | 7.383** | 70.88 |
| APF | 58.145** | 971.169** | 85.165** | 22.829** | 60.63 |

[§] Abbreviations are explained in materials and methods part
*,** Significant based on F-test for p= 0.05, 0.01, respectively

***Genetic variation of reproductive features of un-tripped treatment***

Faba bean showed a specific performance in un-tripped conditions in bee-proof house, markedly different from tripped treatment. Without tripping, the plants showed very low pod set, they had very many so-called false pods (empty pods), a stay-green habitus, they had additional and late occurring tillers, additional branches and even late, additional flowers and pods in unusual place (Figure 2).

The un-tripped treatment which was applied to the genotypes across 2013, 2014 and 2015 constantly gave a very low pod set, with rates of fertilization ranging from 0.23 - 37.14%, and with a heritability of 88.35% (Table 4). The same phenomenon of low range of values also can be seen for potential pod filling and actual pod filling. Analysis of variance showed that all reproductive features displayed a highly significant variance ($p< 0.01$) for genotypes.

Figure 2. Performance and strange features of un-tripped faba bean in bee-proof house.

Table 4. Phenotypic results and analysis of variance of 189 lines of A-set of un-tripped with ten of replicates (experiments across 2013, 2014 and 2015).

| Trait[§] | Min. | Max. | Mean | Var. cp. (G) | Heritability (%) |
|---|---|---|---|---|---|
| PH | 55.10 | 148.60 | 110.97 | 113.489** | 86.81 |
| NF | 16.60 | 44.80 | 32.49 | 28.84** | 85.24 |
| NP | 0.10 | 16.00 | 2.17 | 4.75** | 88.25 |
| NS | 0.10 | 27.20 | 4.16 | 19.57** | 84.24 |
| SY | 0.08 | 23.73 | 3.55 | 13.68** | 85.35 |
| RF | 0.23 | 37.14 | 6.70 | 40.74** | 88.35 |
| PPF | 0.08 | 22.09 | 3.22 | 11.02** | 83.99 |
| APF | 2.50 | 70.89 | 28.15 | 143.39** | 71.74 |

[§] Abbreviations are explained in materials and methods part

*,** Significant based on F-test for p= 0.05, 0.01, respectively

### *Faba bean performance after intensive tripping*

Intensive tripping was conducted only for 58 genotypes in 2015. The performance of three aspect of fertility then was compared to that of the same genotypes in 2013 and 2015. Rate of fertilization increased from 2013 to 2015 (Table 5). Such tendency occurred for potential pod filling and actual pod filling. Rate of fertilization, potential pod filling and actual pod filling were significantly different in 2013, 2014 and 2015 respectively (data was not shown). However, even intensive tripping could not give a near-to 100% result of these three traits.

Table 5. Faba bean performance of 58 lines of Q-set in comparison to intensive tripping in bee-proof house in 2015.

| Trait[§] | 2013 | | | 2014 | | | 2015 | | |
|---|---|---|---|---|---|---|---|---|---|
| | Min. | Max. | Mean | Min. | Max. | Mean | Min. | Max. | Mean |
| RF | 20.92 | 85.00 | 43.71 | 39.58 | 90.63 | 67.85 | 36.36 | 95.83 | 74.59 |
| PPF | 9.52 | 71.88 | 30.19 | 30.73 | 93.23 | 51.99 | 29.17 | 77.08 | 58.20 |
| APF | 41.67 | 102.50 | 68.96 | 52.50 | 112.24 | 76.79 | 57.95 | 97.22 | 78.25 |

[§] Abbreviations are explained in materials and methods part

### *Correlation of traits of tripped and un-tripped treatment*

Some agronomic traits did not significantly correlate topod set-relatedtraits (NP, NS, SY) in un-tripped treatment, while autofertility-related traits significantly correlated (Table 6). In the tripped treatment, rate of fertilization indeed correlated to agronomic traits as well as to pod set-related traits (Table 7). Potential pod filling showed the same tendency to rate of fertilization, but did not correlate to plant height and flowering time in tripped treatment. Actual pod filling showed a different tendency, it did not correlated to agronomic traits in both, un-tripped and tripped treatment.

Table 6. Correlation coefficient of traits of 189 lines of A-set in un-tripped treatment (experiment across 2013, 2014 and 2015).

| Trait§ | PH | NF | FT | FFP | NP | NS | SY | RF | PPF |
|---|---|---|---|---|---|---|---|---|---|
| NF | 0.144* | | | | | | | | |
| FT | 0.438** | -0.246** | | | | | | | |
| FFP | 0.735** | -0.295** | 0.757** | | | | | | |
| NP | 0.111 | 0.159* | -0.064 | -0.072 | | | | | |
| NS | 0.097 | 0.145* | -0.077 | -0.085 | 0.971** | | | | |
| SY | 0.086 | 0.154* | -0.113 | -0.113 | 0.948** | 0.983** | | | |
| RF | 0.059 | -0.021 | -0.042 | -0.034 | 0.961** | 0.946** | 0.927** | | |
| PPF | 0.047 | -0.017 | -0.053 | -0.052 | 0.922** | 0.965** | 0.950** | 0.969** | |
| APF | 0.025 | 0.094 | -0.047 | -0.073 | 0.679** | 0.728** | 0.746** | 0.701** | 0.742** |

§ Abbreviations are explained in materials and methods part

*,** Significant for p= 0.05, 0.01, respectively



Table 7. Correlation coefficient of traits of 189 lines of A-set in tripped treatment (experiment across 2013 and 2014).

| Trait§ | PH | NF | FT | FFP | NP | NS | SY | RF | PPF | APF |
|---|---|---|---|---|---|---|---|---|---|---|
| NF | 0.144* | | | | | | | | | |
| FT | 0.438** | -0.246** | | | | | | | | |
| FFP | 0.735** | -0.295** | 0.757** | | | | | | | |
| NP | 0.288** | 0.481** | 0.012 | 0.137 | | | | | | |
| NS | 0.238** | 0.385** | -0.013 | 0.117 | 0.831** | | | | | |
| SY | 0.360** | 0.348** | -0.008 | 0.233** | 0.730** | 0.836** | | | | |
| RF | 0.145* | -0.339** | 0.209** | 0.393** | 0.535** | 0.495** | 0.394** | | | |
| PPF | 0.089 | -0.297** | 0.136 | 0.295** | 0.388** | 0.676** | 0.514** | 0.805** | | |
| APF | 0.003 | -0.046 | -0.075 | -0.006 | -0.021 | 0.514** | 0.399** | 0.058 | 0.621** | |
| TKW | 0.217** | -0.131 | 0.031 | 0.224** | -0.252** | -0.394** | 0.149* | -0.206** | -0.353** | -0.287** |

§ Abbreviations are explained in materials and methods part

*,** Significant for p= 0.05, 0.01, respectively

### *Autofertility of winter and spring faba bean*

The experiment involved 189 winter beans (A-set) and four spring beans. The rate of fertilization, potential pod filling and actual pod filling of winter beans were significantly different from spring beans (Table 8). The three aspects of autofertiliy of un-tripped winter beans were lower than those of spring beans.

Table 8. Winter and spring beans in un-tripped treatment of 2013, 2014 and 2015.

| Trait | Mean value | | t-value |
|---|---|---|---|
| | Winter beans | Spring beans | |
| Rate of fertilization | 6.69 | 48.60 | 9.63** |
| Potential pod filling | 3.27 | 28.59 | 7.91** |
| Actual pod filling | 28.14 | 69.93 | 4.56** |

** Significant for p= 0.01

### *Association analysis of reproductive features*

Association analysis was carried out using the data from the entire series of un-tripped and tripped treatment and from means as resulting from the above mentioned ANOVA analyses. Looking on the 11 traits, only for five traits significant marker-trait associations were detected (cf. Table 9). However, the variation of the three traits of main interests (rate of fertilization, potential pod filling and actual pod filling) showed no significant association to any of the markers.

First flower position was associated with one AFLP and one SNP marker, with 9.76 and 9.22% of phenotypic variance explained per marker. Flowering time was associated to two SNP markers and one AFLP marker, with about 7 and 10% of phenotypic variance explained per marker. Plant height was associated to two AFLP markers with 9.26 and 5.39% of phenotypic variance explained per marker. Seed yield and thousand kernel weight were associated to one AFLP marker in tripped treatment. One SNP marker (Vf_Mt4g068010) and one AFLP marker (E32M58-384) was associated to two traits. Vf_Mt4g068010 was associated to both first flower position and flowering time with a similar amount of explained phenotypic variance for each trait (9.22% and 10.40% respectively). E32M58-384

was associated to flowering time and plant height with similar phenotypic variance for each trait (about 7.50%).

Table 9. Association analysis of 189 lines of A-set of mean values in un-tripped and tripped treatment.

| | Marker | LG* | Positions (cM)* | p-value | $R^2$ (%) | Effect** | Increase allele*** |
|---|---|---|---|---|---|---|---|
| First flower position | | | | | | | |
| 1 | E40M55-299 | 6 | 47.1 | $2.05 \times 10^{-5}$ | 9.76 | 6.94 | "1" |
| 2 | Vf_Mt4g068010 | 7 | 104.0 | $2.03 \times 10^{-4}$ | 9.22 | 6.45 | "G" |
| Flowering time | | | | | | | |
| 1 | Vf_Mt4g068010 | 7 | 104.0 | $7.58 \times 10^{-5}$ | 10.40 | 2.29 | "G" |
| 2 | Vf_Mt1g056180 | 5 | 153.3 | $1.46 \times 10^{-4}$ | 7.89 | 2.29 | "A" |
| 3 | E32M58-384 | 2 | 109.4 | $2.06 \times 10^{-4}$ | 7.46 | 1.96 | "1" |
| Plant height | | | | | | | |
| 1 | E36M56-229 | 4 | 168.6 | $1.50 \times 10^{-5}$ | 11.27 | 9.26 | "0" |
| 2 | E32M58-384 | 2 | 109.4 | $1.98 \times 10^{-4}$ | 7.50 | 5.39 | "1" |
| Seed yield of tripped treatment only | | | | | | | |
| 1 | E42M58-118 | 1 | 70.3 | $4.01 \times 10^{-5}$ | 9.06 | 9.03 | "0" |
| Thousand kernel weight of tripped treatment only | | | | | | | |
| 1 | E36M59-133 | 7 | 94.4 | $1.42 \times 10^{-5}$ | 9.96 | 73.69 | "0" |

* According to Welna (2014)

**Difference between the means of the two marker classes as calculated by TASSEL 3.0.

***To specify which homozygous marker class showed the higher average trait expression.

## Discussion

In the recent experiment, autofertility was examined in un-tripped treatment. For comparison purposes and as check treatment, the tripped treatment was carried out as well. Plant standardization, especially in number of flowers, was applied to maximize the opportunity of fertilized flowers develop to pods (irrespective whether spontaneously self-fertilized by the autofertile behaviour or assisted self-fertilized via tripping assistance). Faba bean has been known to have an excess number of flowers, which is a common feature of partial allogamy and outcrossing species to increase the reproductive success opportunity (Pattrick and Stoddard, 2010). One inflorescence has 3-8 flowers, Link (1990) reported that with three or more flowers per inflorescence, fertilized flowers can be easily aborted. So, he suggested that if only about one third of flowers were left at the plant as un-tripped flowers, then all flowers could develop to pods if autofertility was 100%; this statement was for spring faba beans (Link, 1990).

The assessment of autofertility focused on rate of fertilization, potential pod filling and actual pod filling. Rate of fertilization indicates the ratio of set pods to available number of flowers. This trait is simple but can give meaningful data. To some extent, the trait was the same as following the method of Link (1990). He reduced the number of flowers, used 12 inflorescences and used only one main tiller per plant due to analysing only spring faba beans in his experiment (they usually only have one tiller). In addition, to observe the capability of plants to set full seeds-containing pods in our experiment, potential pod filling was required to study. Actual pod filling give the information of the ability of setting fully filled pods. So this trait displays the variation between barely fertilized pods with only one seed set per pod to completely fertilized pods with the (typical) maximum of four seeds per pod.

Three types of standardization were applied in the recent study. The comparison of these was carried out in un-tripped treatment in 2015 with two replicates each. Mean values of rate of fertilization, potential pod filling and actual pod filling among the three types were not significantly different (Table 10, details are not shown). Therefore, there was no further analysis conducted for the different types of standardization in the three years of the experiment.

Table 10. Rate of fertilization, potential pod filling and actual pod filling of un-tripped treatment 2015 in different type of standardization.

| Trait | Standardization | | |
|---|---|---|---|
| | Type 1 | Type 2 | Type 3 |
| Rate of fertilization (%) | 6.40 | 6.49 | 7.20 |
| Potential pod filling (%) | 2.97 | 3.10 | 3.38 |
| Actual pod filling (%) | 30.35 | 24.34 | 29.49 |

The mean values of the three aspects of autofertility in un-tripped treatment were obviously lower than the means of the corresponding traits resulting from tripped treatment, with heritability being medium to high. Heritability of potential pod filling and actual pod filling of tripped treatment were surprisingly higher than that of un-tripped treatment. Higher heritability showed higher genetic variability, which indicates that the genotypes gave different responses to the tripping treatment. This was rather surprising, because the expectation was that the tripping treatment would cancel out the autofertility-caused differences between the inbred lines and thus very little variation and very low heritability should remain. Thus, genetic differences for the reaction to tripping are most plausible explanation for this finding.

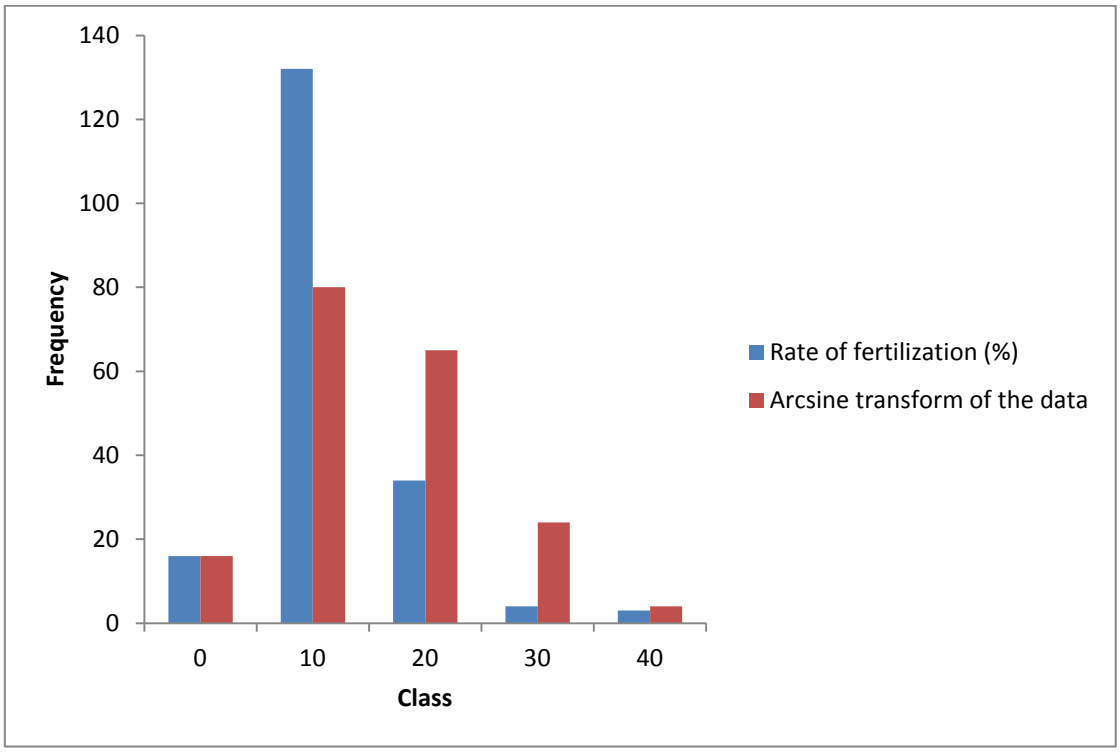



Figure 3. Rate of fertilization and its arcsine transformation data of un-tripped treatment.

To deal with the percentage data, with marked proportions being near to the lower limit of 0, especially in the un-tripped treatment, the arcsine square root transformation was conducted (McDonald, 2009; details are not presented). The results after this transformation of rate of fertilization in un-tripped treatment likely showed a more normal distribution than the original data (Figure 3). Most of the data is close to 0% in un-tripped treatment in 2013 and 2014, arcsine transformation disperses the data (Figure 4). In the contrary, rate of fertilization in tripped treatment was distributed from 40%-90%, then arcsine transformation compressed the data distribution. Analysis of variance of transformed rate of fertilization yielded similar result and conclusions as the presented results. In addition, association analysis was carried out with transformed data as well and gave similar results to the original ones. Arcsine transformation of potential pod filling and actual pod filling confirmed the presented results of ANOVA and association analysis as well. Due to the similarity of the results when based on original data and when based on their transformed version, the original data and their results were reported.

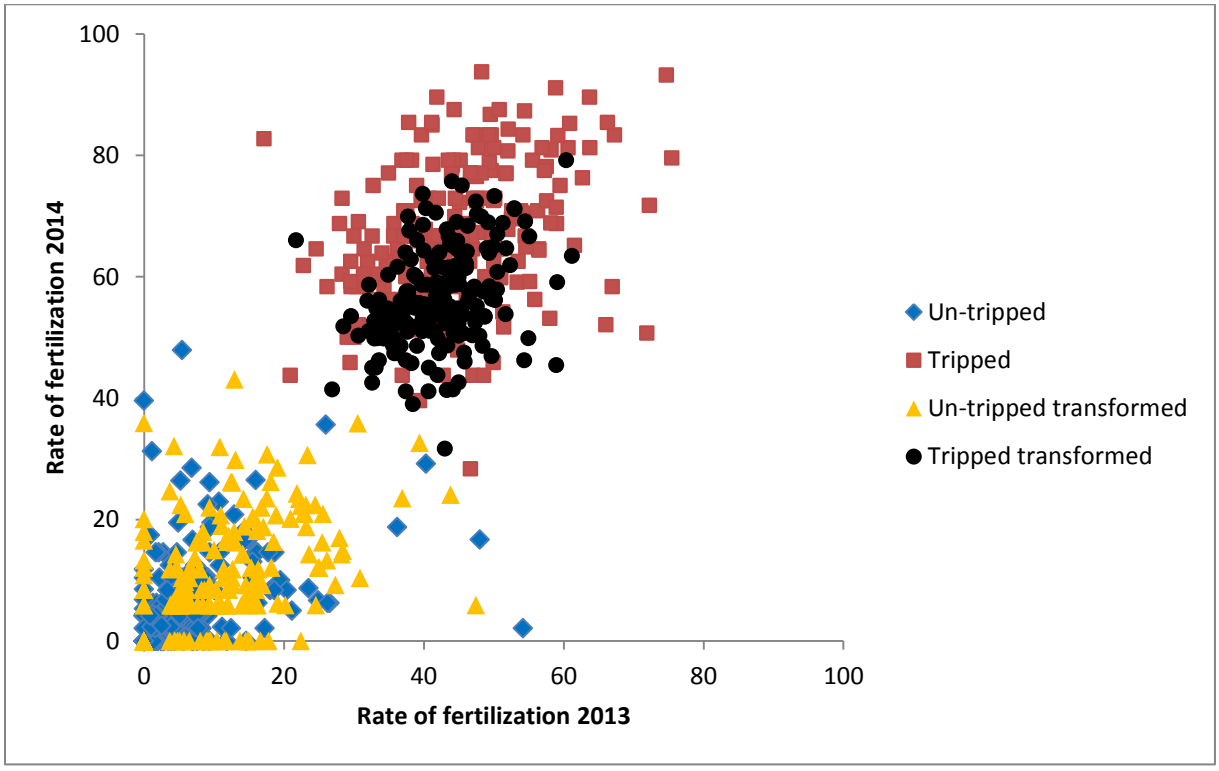



Figure 4. Rate of fertilization of original and transformed data of 189 lines of the A-set of tripped and un-tripped treatment in 2013 and 2014

In spite of the genotypes giving higher rate of fertilization with tripping on average, the significant interaction of genotypes-treatments showed that the genotypes gave different response to treatments. Significant interaction of genotypes-environments (different year) showed that genotypes also yielded different responses to the year-impact on rate of fertilization.

Intensive tripping was conducted in 2015 only using 58 genotypes to verify whether a relatively higher frequency and more care of tripping will give the expected maximum pod set (rate of fertilization near to 100%). The genotypes were selected to have a wide range of rate of fertilization (minimum to maximum). The results showed that rate of fertilization were highest in 2015, when intensive tripping was applied. Yet, a year-specific, environmental impact on this outcome cannot be ruled. In 2015, with intensive tripping, also an increased potential pod filling and actual pod filling was observed. Nevertheless, none of these genotypes showed 100% of flowers develop to pods. This supports the previous statement that genotypes give different and on average non-perfect response to tripping, corroborated by the high heritability in tripped treatment.

Correlation between traits is important to understand potential effects or genetic associations of one trait to another trait. No significant coefficient of correlation between the autofertility-related traits to plant height, flowering time and first flower position occurred in the un-tripped treatment. The autofertility-related traits highly correlated with the number of pods, number of seeds and seed yield per plant for obvious reasons.

Autofertility in the A-set of winter faba beans was very low compared to spring faba beans. This result is in accordance to that of Stoddard (1986) who also found that the autofertility of winter beans was lower than spring beans. Winter beans obviously tend to have a lower autofertility than spring beans. High level autofertility was observed from genotypes in Middle East countries which have spring type or even summer type of faba bean (Robertson and El-Sherbeeniy, 1995). Winter beans have special features, related to vernalization and hardening to survive frost events during winter. More research is needed to see how and why the winter-adapted behaviour should be associated with a lower level of autofertility.

Faba bean is naturally propagated with a mixed-breeding system which allows a certain heterozygosity level in the field. Heterozygosity obviously increases autofertility and decreases outcrossing rate (Drayner, 1959; Link, 1990). Similar results were shown in

unpublished data with lines from the A-set (Table 11). Six lines with different levels of autofertility were used and crossed to produce F1. Rate of fertilization, potential pod filling and actual pod filling of the F1 were much higher than those of the parental inbred lines.

Table 11. Rate of fertilization, potential pod filling and actual pod filling of six lines of A-set and F1 of crossing between the lines in un-tripped treatment (tested in the same environment of 2016; Brünjes, unpublished manuscript).

| Genotype | Trait[§] | | |
|---|---|---|---|
| | RF | PPF | APF |
| S_019 | 9.38 | 5.21 | 40.63 |
| S_035 | 0.00 | 0.00 | 0.00 |
| S_046 | 1.04 | 0.26 | 6.25 |
| S_085 | 50.00 | 19.79 | 40.23 |
| S_199 | 2.08 | 1.04 | 12.50 |
| WAB-EPFam-157 | 6.25 | 3.65 | 46.88 |
| S_019 x S_035 (F1) | 68.75 | 50.00 | 67.59 |
| S_046 x S_085 (F1) | 97.92 | 86.46 | 88.24 |
| S_199 x WAB-EPFam-157 (F1) | 32.99 | 19.01 | 62.25 |

[§] Abbreviations are explained in materials and methods part

Study of association between marker and trait showed significant associations for first flower position, flowering time, plant height, seed yield and thousand kernel weight with several markers. Most of the QTLs had minor effect and explained less than 10% of phenotypic variance. However, no significance association of an autofertility-related traits of the 189 lines of winter beans could be found, neither in un-tripped nor in the tripped treatment. Even though there were significant correlations between some agronomic traits with autofertility-related traits, nevertheless significant QTL could not be found. This could be due to a limited number of inbred lines used and due to the limited number of markers used. Further inbred lines would have offered a higher chance to result in significances of given marker effects, because of higher numbers of degrees of freedom when comparing the two groups of inbred lines (grouped according to markers). More markers would have offered a higher probability to find associated markers. This finding is in accordance to the previous reports of Ali *et al.* (2016) who observed drought and frost tolerance traits using

the same inbred lines and the same markers. With the very low average linkage disequilibrium amongpairs of markers in these inbred lines in A-set ($r^2$= 0.0077), a higher number of markers is needed to increase the probability of QTL detection.

Two markers were found to be associated with more than one trait. Vf_Mt4g068010 was associated to first flower position and flowering time. Allele G of this marker increase first flower position by 6.45 cm and increase flowering time by 2.29 days. AFLP marker E32M58-384 was associated to flowering time and plant height. DNA peak existence of this AFLP marker was significantly associated with an increase of both traits.

Marker of Vf_Mt1g056180 was significant for flowering time. This result is in accordance to the previous report of Sallam *et al.* (2016); they also found the same marker associated to days to flower and seed yield (in the same genetic material but with different experiments). This marker was also found to be associated with survival of a frost test and loss of leaf turgidity and loss of leaf colour due to a frost, hence with frost tolerance-related traits (Sallam and Martsch, 2015). The marker was located in 153.3 cM in linkage group 5 (Welna, 2014), which corresponds to a placement at 35.48 cM of chromosome 3 of *Vicia faba* (Webb *et al.*, 2016).

A major result from the present study waswinter faba bean has different, lower level of autofertility than spring beans. Therefore, to improve the level of autofertility, autofertile parents have to be included together with winter faba beans in breeding programs. Moreover, heterosis is still interesting to be exploited to increase the level of autofertility. Winter beans tend to flower early in the season then this topic has a special importance for yield gain.

## References


Ali, M., Welna, G.C., Sallam, A., Martsch, R., Balko, C., Gebser, B., Sass, O. and Link, W., 2016. Association analyses to genetically improve drought and freezing tolerance of faba bean (*Vicia faba*L.). *Crop Science*, *56*(3), pp.1036-1048.

Benjamini, Y. and Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, pp.289-300.

Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y. and Buckler, E.S., 2007. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, *23*(19), pp.2633-2635.

Chen, W., 2009. Pollination, fertilization and floral traits co-segregating with autofertility in faba bean. *Journal of New Seeds*, *10*(1), pp.14-30.

Drayner, J.M., 1959. Self-and cross-fertility in field beans (*Vicia faba* Linn.). *The Journal of Agricultural Science*, *53*(3), pp.387-403.

Gasim, S., Abel, S. and Link, W., 2004. Extent, variation and breeding impact of natural cross-fertilization in German winter faba beans using hilum colour as marker. *Euphytica*, *136*(2), pp.193-200.

Hanna, A.S. and Lawes, D.A., 1967. Studies on pollination and fertilization in the field bean (*Vicia faba* L.). *Annals of Applied Biology*, *59*(2), pp.289-295.

Herzog, H. and Geisler, G., 1991. Yield structure of winter faba beans grown in Northern Germany in dependence of different environments, seed rates, sowing rates and genotypes. *Journal of Agronomy and Crop Science*, *167*(3), pp.145-154.

Kambal, A.E., Bond, D.A. and Toynbee-Clarke, G., 1976. A study on the pollination mechanism in field beans (*Vicia faba* L.). *The Journal of Agricultural Science*, *87*(03), pp.519-526.

Knudsen, J.C.N. and Poulsen, M.H., 1983. Pollination studies on the closed-flower mutant of *Vicia faba* L. *Zeitschrift fur Pflanzenzüchtung, 9*, pp.106-115.

Köpke, U. and Nemecek, T., 2010. Ecological services of faba bean. *Field Crops Research*, *115*(3), pp.217-233.

Link, W., Ederer, W., Metz, P., Buiel, H. and Melchinger, A.E., 1994. Genotypic and environmental variation for degree of cross-fertilization in faba bean. *Crop Science*, *34*(4), pp.960-964.

Link, W., Balko, C. and Stoddard, F.L., 2010. Winter hardiness in faba bean: physiology and breeding. *Field Crops Research*, *115*(3), pp.287-296.

Link, W., 1990. Autofertility and rate of cross-fertilization: crucial characters for breeding synthetic varieties in faba beans (*Vicia faba* L.). *Theoretical and Applied Genetics*, *79*(5), pp.713-717.

McDonald, J.H., 2009. *Handbook of Biological Statistics* Vol. 2. Baltimore, MD: Sparky House Publishing.

Poulsen, M.H., 1977. Obligate autogamy in *Vicia faba* L. *The Journal of Agricultural Science*, *88*(1), pp.253-256.

Rowlands, D.G., 1964. Fertility studies in the broad bean (*Vicia faba* L.). *Heredity*, *19*(2), pp.271-277.

Sallam, A. and Martsch, R., 2015. Association mapping for frost tolerance using multi-parent advanced generation inter-cross (MAGIC) population in faba bean (*Vicia faba*. *Genetica*, *143*(4), pp.501-514.

Sallam, A., Dhanapal, A.P. and Liu, S., 2016. Association mapping of winter hardiness and yield traits in faba bean (*Vicia faba* L.). *Crop and Pasture Science*, *67*(1), pp.55-68.

Stoddard, F.L., 1986. Autofertility and bee visitation in contrasting genotypes of winter and spring faba beans (*Viciafaba* L.). *Plant Breeding, 97*, pp.171-182.

Stoddard, F.L. and Bond, D.A., 1987. The pollination requirements of the faba bean. *Bee World*, *68*(3), pp.144-152.

Utz, H.F., 2001. PLABSTAT: a computer program for statistical analysis of plant breeding experiments. *Institute for Plant Breeding, Seed Science and Population Genetics, University of Hohenheim, Stuttgart*.

Webb, A., Cottage, A., Wood, T., Khamassi, K., Hobbs, D., Gostkiewicz, K., White, M., Khazaei, H., Ali, M., Street, D. and Duc, G., Stoddard, F.L., Maalouf, F., Ogbonnaya, F.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*(1), pp.177-185.

Welna, G.C., 2014. Genetische Analysen für eine markergestützte Verbesserung der Trockenstresstoleranz von Winterackerbohnen. PhD thesis, University of Gottingen.

# Chapter II. Association analysis for vicine-convicine in faba bean (*Vicia faba* L.)


**Abstract**

Faba bean seeds supply a protein rich feedstuff and provide a nutritionally valuable composition. However, faba bean contains anti-nutritive compounds which limit the use in feed and food system and have health impact for humans and several animals, such as vicine and convicine. The aims of the study are to develop a NIRS-based so-called calibration for vicine-convicine content in faba bean seed, to study heritability and genetic variation of vicine-convicine content in faba bean, to identify QTLs that are responsible for vicine-convicine variation in vicine-convicine-containing (wild-type) faba bean genotypes and to verify whether the mutant allele for vicine-convicine in faba bean ("vc-"; Duc 1989) is allelic to a QTL for the variation in vicine-convicine-containing materials. A genome wide association analysis was employed for QTLs investigation. The main genetic materials used in this study involved 200 inbred lines, named Q-set, which consisted of 189 lines of A-set (inbred lines for association study), seven further winter bean lines and four further spring bean lines. The A-set was derived from the so-called Göttingen Winter Bean Population (GWBP). The experiment was carried out by HPLC and by NIR-spectrophotometry analysis of the faba bean seeds. We developed a NIRS calibration to allow for a NIRS-based prediction of seed vicine-convicine content. Association analysis between DNA-markers and phenotypic expression of traits was carried out using TASSEL version 3.0. A total of 2018 polymorphic markers were used consisting of 189 SNP (Single Nucleotide Polymorphism) and 1829 AFLP (Amplified Fragment Length Polymorphism). NIRS technology can be applied to predict vicine-convicine content in faba bean. A relatively well-performing calibration equation was produced and applied to analyse samples faba bean seeds across different replicates, treatments and years. Significant and large quantitative variations were found for vicine-convicine content with relatively high heritability. One AFLP-marker was significantly associated to the vicine-convicine variation in the genotypes and the probable position of the QTL was on chromosome number 5 of *Vicia faba*.

## Introduction

Faba bean (*Vicia faba* L.) is one of the most important grain legume crops in Europe which ranks third in area and production after soybean and pea (FAOSTAT, 2014). Faba bean seeds are protein rich and provide a valuable composition with a useful balance of carbohydrate, fibre, micronutrient and phytochemicals (Crepon *et al.*, 2010; Yahia *et al.*, 2013; Pasricha *et al.*, 2014).

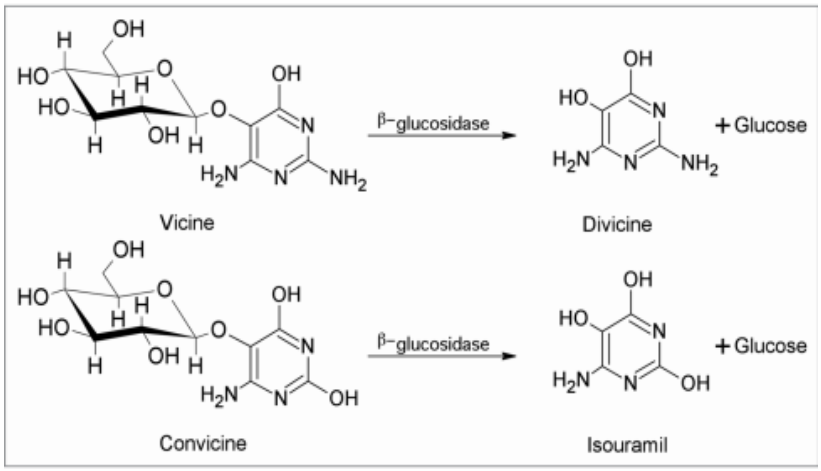



Figure 5.Hydrolisis of seed compounds vicine and convicine (Ray and George, 2010)

However, faba bean contains several anti-nutritive compounds which limit the use in feed and food system and have health impact for humans and several animals. Among these are the pyrimidine glucosides vicine and convicine. Hydrolysis of vicine and convicine produces the aglycones divicine and isouramil (Figure 5). The aglycones induce oxidative stress in red blood cell of human who have a genetic deficiency of G6PD (Glucose-6-Phosphate-Dehydrogenase) by oxidizing glutathione. In normal condition, glutathione is restored to its reduced, biologically active form by G6PD in the pentose phosphate pathway. Hence, reduced glutathione is essential to maintain the integrity of red blood cell by reducing reactive oxygen species. High rate of oxidized glutathione in G6PD deficiency cause acute haemolytic anaemia, the so-called favism (Baker *et al.*, 1984; Mehta *et al.*, 2000). About 400 million human individuals are G6PD-deficient, with prevalence in Mediterranian, Africa and Asia, a fact which makes the phenomenon the most common human deficiency in the world (Capellini and Fiorelli, 2008).

Among livestock species, laying hens are especially susceptible to vicine and convicine, which reduce their food consumption, body weight gain and egg weight (Marquardt *et al.*, 1976; Muduuli *et al.*, 1982). Recent studies on dietary vicine and convicine in laying hens showed a reduction on egg weight but no negative effect on egg production and egg quality parameters (Koivunen *et al.*, 2014; Lessire *et al.*, 2016).

Vicine was known as the first simple pyrimidine derivative found in nature. It was discovered for the first time in 1870 in vetch seeds (*Vicia sativa)* by Ritthausen. He also discovered convicine in 1881, a similar substance in *Vicia sativa*. However, the structures were remarkably only published for vicine in 1953 as 2,4-diamino-6-oxypyrimidine-5-(ß-D-glucopyranoside) and for convicine in 1968 as 2,4,5-trihydroxy-6-aminopyrimidine-5-(β-D-glucopyranoside, Bendich and Clements, 1953; Bien *et al.*, 1968). Due to the two compounds having similar structures, their isolation from faba bean as pure compounds require detail and efficient techniques.

Beside *Vicia faba* and *Vicia sativa*, vicine and convicine were also found in *Vicia bithynica* and *Vicia narbonensis* (Pitz *et al.*, 1980; Griffith and Ramsay, 1992). In addition, vicine was found and extracted from *Mamordica charantia* (*Cucurbitaceae* family) indicating that the compound is not unique to *Vicia* (Dutta *et al.*, 1981). The two compounds accumulate in cotyledon and seed coat. During pod development, vicine and convicine content increase throughout seed growth and gradually decline until a constant level in mature seed (Jamalian and Bassiri, 1978; Burbano *et al.*, 1995). Furthermore, vicine and convicine synthesis likely occurs in testa during seed development (Griffith and Ramsay, 1996). The role of vicine and convicine in faba bean is presumably as defence mechanism of plant against pathogens (Bjerg *et al.*, 1984; Rizello *et al.*, 2016).

Vicine and convicine can be extracted by soaking in acid or by cooking methods such as boiling, roasting and frying (Hussein *et al.*, 1986; Cardador-Martinez *et al.*, 2012). In addition, hydrolysis and fermentation of faba bean can be applied to remove vicine and convicine (McKay, 1992; Rizzello *et al.*, 2016). The treatment maybe feasible for preparation of human food but it is very expensive and laborious for preparation of animal feed. In addition, fresh or raw faba bean are commonly consumed as vegetable that does not need to have further treatment, which means that the risks still remains.

Some methods were developed to determine vicine-convicine level in faba bean, such as colorimetric method using UV spectrophotometry and high performance liquid

chromatography (HPLC; Khamassi *et al.*, 2013; Pulkkinen *et al.*, 2015; Ferhatoglu and Vandenberg, 2015). Both methods require chemical process including separation or extraction of targeted compounds from raw materials followed by detection using instruments.

In recent studies, the use of fast analytical techniques such as near infrared spectroscopy (NIRS) has shown many advantages compared to standard techniques. NIRS analysis is rapid, low cost and simple in preparation. In addition, in many cases intact (non destructed) samples can be analysed; moreover, NIRS analyses can simultaneously be carried out for several traits. With many advantages, NIRS has been widely used for qualitative and quantitative analysis in agriculture as well as in other subject (Font *et al.*, 2006; Ozaki *et al.,* 2006). However, NIRS analysis requires reference data which is usually produced by laboratory work. Therefore, a due combination between NIRS and laboratory work may produce result in a more efficient and less costly way than relying on laboratory work only. The NIRS has been successfully applied in evaluation of protein, starch and total polyphenol in faba bean (Wang *et al.* 2014).

The progress of molecular genetics research in faba bean has been rapid in the last decade, and that went in line with the availability of numerous sequences and marker data. The existence of an increasing number and density of molecular maps will remarkably support plant breeding research (O'Sullivan and Angra, 2016). In this study, we used a genome-wide association study (GWAS) to search for QTL of a desired trait, vicine-convicine content, in faba bean. Recently, GWAS has become a preferable method to identify genomic regions associated with complex quantitative traits. Compared to traditional bi-parental mapping, GWAS offers an advantage because rather than exploiting recombination opportunities within a family, it exploits historical recombination and natural genetic diversity (Zhu *et al.* 2008). GWAS uses and analyses the linkage disequilibrium (LD), therefore genetic relatedness within the set of genetic materials must be considered.

The finding of the mutant allele 'vc-' with the connected phenotypic expression of a very low vicine-convicine content (i.e. reduction to 10 until 20 fold) allows the reduction of vicine-convicine in faba bean through plant breeding (Duc, 1989). The allele was known to have a simple monogenic inheritance (Duc, 1997). The genetic locus of this low vicine-convicine gene is linked to the hilum colour locus with alleles for black versus colourless hilum with about 10.1 cM distance (Duc, 2004). Obviously, not all colourless hilum seeds are

low vicine-convicine. This locus was identified as single QTL for vicine-convicine in chromosome I of *Vicia faba* from a recombinant inbred line population of crossing between Mélodie/2 (low vicine–convicine, homozygous for 'vc-') and ILB938/2 (normal vicine–convicine, wild-type at this locus), flanked by DNA-Markers with 1.0 cM in one side and 2.6 cM in other side (Khazaei *et al.*, 2015).

It is unclear whether the very low vicine and convicine content as caused by the allele 'vc-' is necessary to avoid the manifestation of favism in susceptible humans. Moreover, it is unclear whether this very low level is necessary to avoid a reduction of performance in animals such as laying hence (if fed with a diet that includes faba bean). And it is unknown whether the homozygous 'vc-' genotypes may suffer from pleiotropic agronomic disadvantages. Hence, it is of interest to genetically analyse the smaller variation of vicine and convicine content in wild type faba beans.

The purposes of this study are to develop a NIRS-based so-called calibration for vicine-convicine content in faba bean seed, to study heritability and the genetic variation of vicine-convicine content in non-'vc-' (wild type) faba bean, to identify QTLs that are responsible for vicine-convicine variation in such vicine-convicine-containing faba bean genotypes and to verify whether the mutant allele for vicine-convicine in faba bean ("vc-"; Duc, 1989) is allelic to a QTL for the variation in vicine-convicine-containing materials.

.

## Materials and methods

### *Genetic materials*

Genetic materials used in this study involved 200 inbred lines, named Q-set, which consisted of 189 lines (A-set; highly homozygous inbred lines for association study), seven further winter bean lines and four further spring bean lines. The A-set was derived from the so-called Göttingen Winter Bean Population (GWBP). This GWBP was developed in Göttingen starting in 1989 from initially mixing of 11 founder winter bean inbred lines (Hiverna/1-1, Webo/1-1, Wibo/1-1, Côte d'Or/1-1-3, L79/79/1, L977/88/S1wn, L979/S1/1/1sn, Bourdon/1-5, Arrisot/1-1, Banner 1-1, Bulldog 1-4). After nine generations of natural open-pollinated reproduction, 400 lines were pure bred via single seed descent from 400 initial, randomly-taken individuals. A total of 189 lines of these materials were genotyped for the current study; the same genotypes were analyzed for frost and drought features by Ali *et al.*(2016).

Seeds of the Q-set lines were harvested from a field experiment (r=2, open field) in 2013 and from experiments in bee-proof isolation cages in 2013 (r=2), 2014 (r=2) and 2015 (r=6). Treatments (tripped and un-tripped) were applied in these bee-proof isolation cages during flowering time (see page 12-13).

For the development of a NIRS calibration, a total of 246 seed samples were used and named “Calibration Set”, consisting of 171 seed samples of 148 lines of the Q-set, 51 further seed samples of 49 spring bean inbred lines from the German breeding company NPZ Lembke, one sample each of nine inbred lines derived from a cross of Mélodie/7x Hiverna/2, samples of four cultivars of low vicine-convicine content (Disco, Divine, Mandoline, Tiffany) and samples of the 11 founder lines of the GWBP. Hence, a total number of 221 genotypes was involved, and one or several seed samples per genotype of Q-set were employed from different environment (see appendix 1).

### *DNA markers*

A total of 2018 polymorphic markers were used consisting of 189 SNP (Single Nucleotide Polymorphism) and 1829 AFLP (Amplified Fragment Length Polymorphism) markers to study the association between markers and phenotypic expression. After filtering the markers with minor allele frequency of 5%, a total of 1322 markers remained, consisting

of 175 SNP and 1147 AFLP markers. Among all 1322 markers, the average LD ($r^2$) was 0.0077 (Ali *et al.*, 2016).

### *Phenotyping of vicine-convicine*

Vicine-convicine content of faba bean seeds was determined by HPLC (High Performance Liquid Chromatography) and by spectrophotometry analysis. As the first step of analysis, faba bean seeds were ground using a Retsch milling machine. The resulting faba bean flour then was used for both analysis methods.

Laboratorium reference analysis of vicine-convicine content of faba bean samples was conducted using HPLC method as described by Khamassi *et al.* (2013) at NIAB (National Institute of Agricultural Botany, UK). The sum of total vicine-convicine of samples was determined by using both peak height and peak area for each vicine and convicine.

Spectrophotometry analysis was carried out by using near infrared (NIR) spectrophotometry, which was performed by near infrared scanning monochromator (NIRSystem model 6500, Foss NIRSystems Inc., MD, USA). Faba bean flour was placed in a standard ring cup (cuvette) and then scanned. All spectral data were recorded as logarithm of reciprocal of reflectance log (1/R) in wavelength range from 400 to 2498 nm at 2 nm intervals.

Calibration and validation of spectral data were carried out using WinISI II Project Manager v1.50 software. Samples in the calibration set were used to establish statistical relations between spectral data and laboratory reference values (HPLC data). The optimum calibration equation was developed in this study using the statistical approach of modified partial-least-square (MPLS) regression (Shenk and Westerhaus, 1991). Calibration equation was determined with several combinations of derivative, of the log 1/R data, derivative sizes (gap, the length in nm), segment length used in first smoothing and segment length used in second smoothing (Shenk *et al.*, 2008). The use of derivative spectra is to minimize both additive and multiplicative effects in the spectra (Rinnan *et al.*, 2009). In addition to derivatives, scatter correction using standard normal variate and detrending (SNVD) was applied for the calibration to minimize the differences in spectra related to particle size and path length of samples (Barnes *et al.*, 1989). Several so-called factors for calibration were analyzed and fixed as describe in Table 12.

Table 12. The factors and factor levels of calibration of vicine-convicine content

| Factors | Tested factor levels | Optimum factor levels |
|---|---|---|
| Scatter | SNV and detrend | SNV and detrend |
| Maximum number of terms | auto and max | 16 (max) |
| Cross validation groups (internal) | Max | 128 |
| Repeatability file | none, file-1, file-2, file-3 | None |
| Derivate | 1, 2 | 2 |
| Gap | 2, 3, 4, 8 | 3 |
| First smooth | 2, 3, 4, 8 | 3 |
| Second smooth | 1, 2, 3, 4 | 2 |
| Delete outliers | none and 1x delete | None |

The internal performance of the calibration was assessed based on standard error of calibration (SEC), coefficient of determination ($R^2$), standard error of cross-validation (SECV) and coefficient of determination in cross-validation ($R^2CV$) as given by WinISI.

A semi-external validation procedure was applied to evaluate the performance of the calibration. For this purpose, a total of 156 randomly selected flour samples of the A-set was used and divided into five subset groups of 31 or 32 samples each. The calibration performance of the semi-external validation was assessed by standard error of performance corrected for bias [SEP(C)], the coefficient of determination in validation ($R^2V$) as given by WinISI, the ratio performance deviation (RPD) [SD/SEP(C)] and the range-to-error ratio (RER) [Range/SEP(C)] (Williams and Sobering, 1996). The main focus was on maximizing the $R^2V$-value.

Table 13. The experiments of 200 lines of Q-set in 2013, 2014 and 2015

| Year | Environment | Treatment | Replicates |
|---|---|---|---|
| 2013 | Open field | Open pollinated | 2 rep |
| 2013 | Bee-proof isolation house | Self-fertilized, tripped | 2 rep |
| 2014 | Bee-proof isolation house | Self-fertilized, tripped | 2 rep |
| 2015 | Bee-proof isolation house | Self-fertilized, tripped | 1 rep |
| 2015 | Bee-proof isolation house | Self-fertilized, un-tripped | 6 rep |

After the process of finding the optimum NIRS calibration equation (further details in appendix 2), then named “ult.eqa” (ultimate equation), this “ult.eqa” was applied to predict 1450 seed samples of 200 inbred lines of Q-set across different experiments (Table 13). Not from all inbred lines from these experiments seed sample could be taken, due to frequently absent or low seed set in the un-tripped treatment of the experiment.

***Statistical analysis of phenotypic data***

- **The NIRS recording of samples**

All available ground samples were scanned to record NIR spectra and each sample was scanned twice, to establish two NIR-scanning replicates per flour sample. The NIRS recording was organised according alpha lattice randomization. Throughout, the NIRS scanning of ten samples was conducted in 20 minutes, as one incomplete block, with 10 or 15 such incomplete blocks in this design. A total of 14 such alpha lattices were established to recordspectra from all 1450 samples twice. Each recorded spectra was used to predict vicine-convicine. These predictions were further statistically analysed. The analysis of variance of the lattice randomizations was performed by PLABSTAT Software (Utz, 2001). The lattice-adjusted means of the predicted vicine-convicine content of the samples were employed to continue the statistical analyses (see below). Heritability (repeatability of the method, $h^2$) of the genotypes was calculated as genotypic variance per phenotypic variance.

- **Analysis of variance of the genotypes**

As first attempt, the statistical analysis of vicine-convicine content of the genotypes was carried out using the data from open field 2013, bee-proof isolation houses 2013 and 2014 (tripped treatment only). This data set was nearly complete, with only 8.33% of missing data. This analysis was conducted to estimate heritability. Environment and genotype were considered as random factors for this analysis. This ANOVA was performed by PLABSTAT software using the following model:

$$Y_{ijr} = \mu + G_i + E_j + R_r(E_j) + GE_{ij} + e_{ijr}$$

where $Y_{ijr}$= the observation of the *i*th genotype in the *j*th environment and in the *r*th block; μ= general mean; $G_i$= effect of genotype i; $E_j$= effect of environment *j;* $R_r(E_j)$= effect of

block r within environment j; $GE_{ij}$= interaction effect between genotype i and environment j; $e_{ijr}$= residual error term.

***Calculating BLUP (Best Linear Unbiased Prediction)***

The intention was to in addition include results from treatments (un-tripped) that yielded very incomplete data. This additional data set showed 48.77% of missing data due to insufficient seed set. When joining the above, complete data set with this additional, fragmentary data set, there were between 3 - 12 vicine-convicine results available from maximum 13 results per genotype. To acknowledge the thus resulting differences in precision, BLUP (Best Linear Unbiased Prediction) estimation was conducted. For this purpose, the above ANOVA (based on the smaller, nearly complete data set, with three environments) was repeated with this bigger data set in a more basic, more comprehensible version, following this linear model:

$$Y_{ir} = \mu + G_i + R_r + e_{ir}$$

where$Y_{ir}$= the observation of the *i*th genotype in the *r*th block; μ= general mean; $G_i$= effect of genotype i; $R_r$ = effect of block r; $e_{ijr}$= residual error term. No environment effects were included, and all their replicates were enumerated. From this, variance components for genotype variance and error variance were estimated and used for BLUP calculations.

Based on this basic, comprehensible linear model, PLABSTAT was exploited to calculate unbiased mean values of genotypes across the maximum of 13 results per genotype (open field 2013 (r=2), bee-proof isolation house in 2013 (tripped, r=2), 2014 (tripped, r=2) and 2015 (tripped, r=1 and un-tripped, r=6)). To calculate unbiased means, PLABSTAT substitutes missing values by fictitious values which are calculated iteratively minimizing the residual mean square (Yates, 1933; Healy and Westmacott, 1956; cited in Plabstat Manual). These unbiased mean values entered the BLUP estimation procedure as $x_i$-values. In the given approach, this procedure aims at maintaining the effects of genotypes and of replications unchanged, as estimated from the existing body of data, and yields mean values for the genotypes which include all available results per genotype. These unbiased means for genotypes are yet calculated without acknowledging the different numbers of existing data and of fictitious values on which they are based on. To acknowledge this, BLUP values were calculated.

$$BLUP_i = \mu + \{\left[\frac{var.comp(G)}{var.comp(G) + \left[\frac{1}{n}.var.comp(e)\right]}\right][x_i - \mu]\}$$

Where *var. comp (G)*= variance component of genotypes; *var. comp (e)*= error variance component; *n*= number of results per genotype, $x_i$= unbiased phenotypic value based on maximum of 12 replicates. These BLUP estimates of vicine-convicine content of the genotypes then entered the association analysis.

***Association analysis***

Association analysis between markers and phenotypic expression of traits was carried out using TASSEL version 3.0 (Bradbury *et al.*, 2007), based on the above-mentioned DNA-marker data of the Q-set inbred lines and on BLUP values of these inbred lines as resulting from the above-mentioned strategy. The mixed linear model procedure of TASSEL was applied with an optimum level of compression and re-estimate of the variance component estimates of each marker. A kinship matrix was employed, which was developed by using the average genetic similarity among the 11 founder lines as a threshold (see Ali *et al.*, 2016 for details). A false discovery rate of 20% (FDR=0.20) was used to test the statistical significance of marker-trait associations (Benjamini and Hochberg, 1995). Phenotypic effects of the marker loci were calculated as differences between the means of inbred lines when groups according to the marker classes. Positive value indicate that the specified marker allele is associated with an increase of the trait, while negative value indicates the marker allele is associated with a decrease of the trait. The phenotypic variance explained ($R^2$) by the significant makers was as well determined by TASSEL 3.0.

## Results

### *Performance of the ult.eqa calibration equation for vicine-convicine*

In developing NIRS calibration for vicine-convicine, the second derivative transformation of (2, 3, 3, 2) was applied to the raw data (Table 12; Appendix 2) and gave the best performance. Other combination of factor levels did not give better calibration and did not improve the results (details not shown). The seed samples of the Calibration-set covered a large variability of vicine-convicine content, including 37 samples of low, seven samples of medium and the rest of normal content of vicine-convicine (Figure 6). The coefficient of determination of calibration was 0.966 and the coefficient of determination of internal validation of WinISI software was 0.847 (Table 14).

Table 14. Statistical parameters as a result from WinISI II to describe the performance of the chosen calibration equation for analysing vicine-convicine of the calibration set.

| N | Mean | Standard deviation | Standard error of calibration | Coeff. of determination of calibration | Standard error of cross validation | Coeff. of determination of cross val. |
|---|---|---|---|---|---|---|
| 246 | 0.537 | 0.236 | 0.044 | 0.966 | 0.094 | 0.847 |

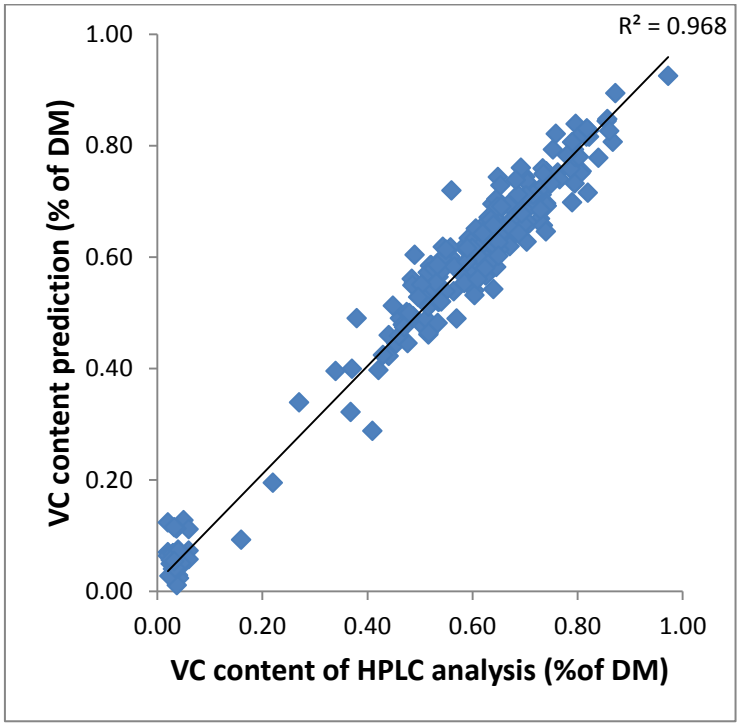



Figure 6. Scatter plot of Calibration-set of vicine-convicinecontent of HPLC analysis and its NIRS-based prediction using ult.eqa.

Semi-external validation of five validation subsets was carried out to verify the performance of the ult.eqa calibration equation. The coefficient of determination of the five subsets of validation was relatively high, varying from 0.864-0.903 (Table 15; see also Figure 7). The RPD ranged from 2.67-3.14 and the RER ranged from 9.63-14.0.

Table 15. Statistical parameters to describe the performance of the chosen calibration equation "VC-ult" when predicting the five validation subsets (semi-external validation).

| Validation subset | N | Mean | SD | Range | SEP(C) | $R^2$EV | RPD | RER |
|---|---|---|---|---|---|---|---|---|
| Val-1 | 32 | 0.647 | 0.096 | 0.488-0.893 | 0.036 | 0.864 | 2.67 | 11.25 |
| Val-2 | 31 | 0.640 | 0.114 | 0.444-0.925 | 0.040 | 0.885 | 2.85 | 12.03 |
| Val-3 | 31 | 0.644 | 0.113 | 0.321-0.826 | 0.036 | 0.903 | 3.14 | 14.03 |
| Val-4 | 31 | 0.633 | 0.106 | 0.459-0.847 | 0.034 | 0.899 | 3.12 | 11.41 |
| Val-5 | 31 | 0.639 | 0.109 | 0.478-0.844 | 0.038 | 0.877 | 2.87 | 9.63 |

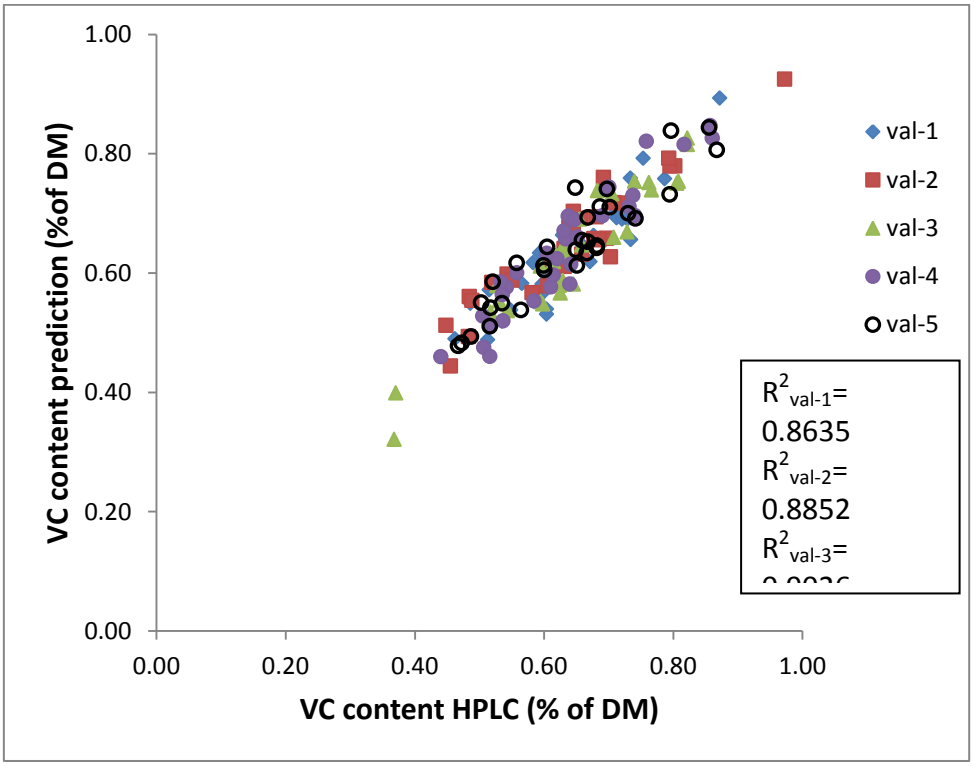



Figure 7. Scatter plot of five semi-external validation subsets

*Analysis of variance of NIRS recording of samples*

A number of 14 alpha lattices (time-based) were established and applied for the scanning process to record NIRS spectra. The analysis of variance of all lattices showed significant effect of the samples (Table 16). The heritability values of vicine-convicine of lattices were medium to high, the $h^2$ value here should be rather termed 'technical repeatability'. Lattices efficiency was relatively low, the highest efficiency is 126.1% compared to a randomized complete block design.

Table 16. Analysis of variance of NIRS recording of samples with alpha lattice design

| Group | Design | Var. Comp of Entry | CV (%) | $h^2$ (%) | Lattice Efficiency (%) |
|---|---|---|---|---|---|
| Lattice 1 | 10 x 10 | 0.0097** | 10.6 | 78.54 | 100.0 |
| Lattice 2 | 10 x 10 | 0.0112** | 12.3 | 79.43 | 101.0 |
| Lattice 3 | 10 x 10 | 0.0094** | 12.2 | 75.20 | 110.0 |
| Lattice 4 | 10 x 10 | 0.0055** | 13.2 | 60.77 | 100.0 |
| Lattice 5 | 10 x 10 | 0.0105** | 12.9 | 76.36 | 126.1 |
| Lattice 6 | 10 x 10 | 0.0088** | 12.0 | 73.95 | 111.6 |
| Lattice 7 | 10 x 10 | 0.0107** | 12.2 | 76.16 | 111.7 |
| Lattice 8 | 10 x 10 | 0.0118** | 12.5 | 77.89 | 110.2 |
| Lattice 9 | 10 x 10 | 0.0069** | 12.6 | 68.32 | 100.0 |
| Lattice 10 | 15 x 10 | 0.0216** | 14.8 | 84.54 | 100.0 |
| Lattice 11 | 10 x 10 | 0.0147** | 10.4 | 82.12 | 104.0 |
| Lattice 12 | 10 x 10 | 0.0103** | 13.6 | 67.76 | 113.1 |
| Lattice 13 | 10 x 10 | 0.0139** | 11.8 | 78.98 | 100.8 |
| Lattice 14 | 10 x 10 | 0.0128** | 11.8 | 73.78 | 101.6 |
| Mean | | | 12.4 | 75.27 | 106.4 |

*Genetic variation of vicine-convicine content*

The analysis of variance based on the lattice-adjusted means of the twice-scanned samples showed highly significant effects of the genotypes and the environments on the vicine-convicine content of the 189 genotypes of A-set (Table 17). Although the interaction of ExG also showed significance, the variance components of ExG was smaller than the variance component of genotypes and environments. Heritability of the NIRS-predicted

vicine-convicine content was quite high ($h^2$=0.788). Here, the term heritability applies because data from real environments were included.

Table 17. Analysis of variance based on three environments (2013H, 2013F, 2014H) of the A-set of faba beans.

| Source | Degree of Freedom | Means Square | Variance component ($10^{-3}$) | F |
|---|---|---|---|---|
| Environment (E) | 2 | 1.3994 | 2.618 | 40.40** |
| Genotype (G) | 188 | 0.0446 | 0.772 | 4.71** |
| E x G | 349 | 0.0095 | 0.410 | 1.48** |
| Heritability ($h^2$): 0.788 | | | | |

Vicine-convicine content of 189 lines of A-set from three environments in 2013 and 2014 ranged about 0.50-0.90%, showed an average of 0.69% and showed a seemingly normal distribution (Figure 8).

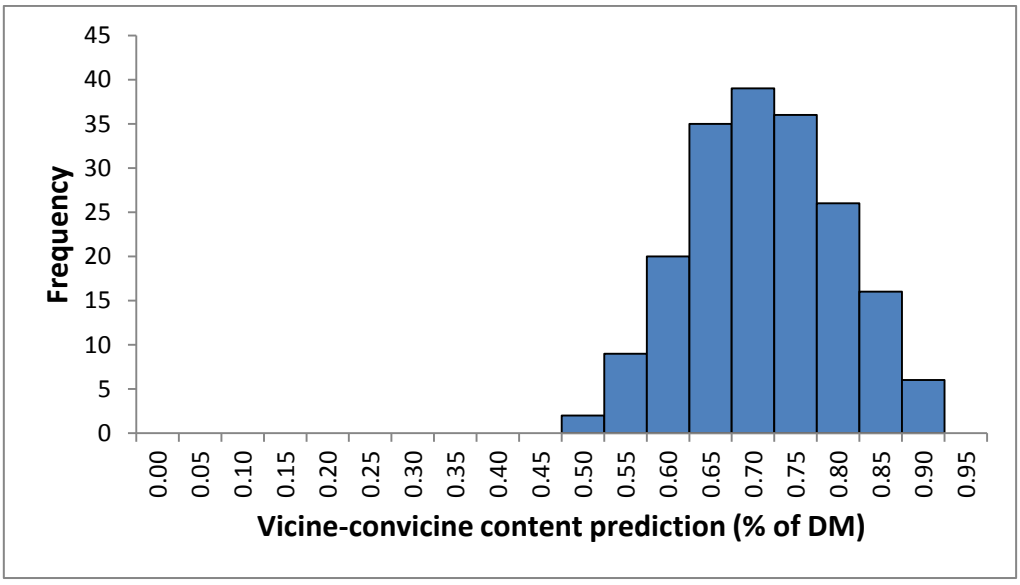



Figure 8. Frequency distribution plot of vicine-convicine content prediction of the A-set.

***BLUP of vicine-convicine content***

BLUP of vicine-convicine content were determined by considering the confidence level based on number of available data of genotypes. BLUP values have small difference compared to means of vicine-convicine content with coefficient of determination 0.997

(Figure 9). Increasing available data will reduce deviation between the 'real' data and BLUP values, which means the available data could be representative to the real data (Figure 10). Small number of available data causes bigger deviation of the real data and BLUP value, which makes BLUP value depends on general mean (0.69% of DM).

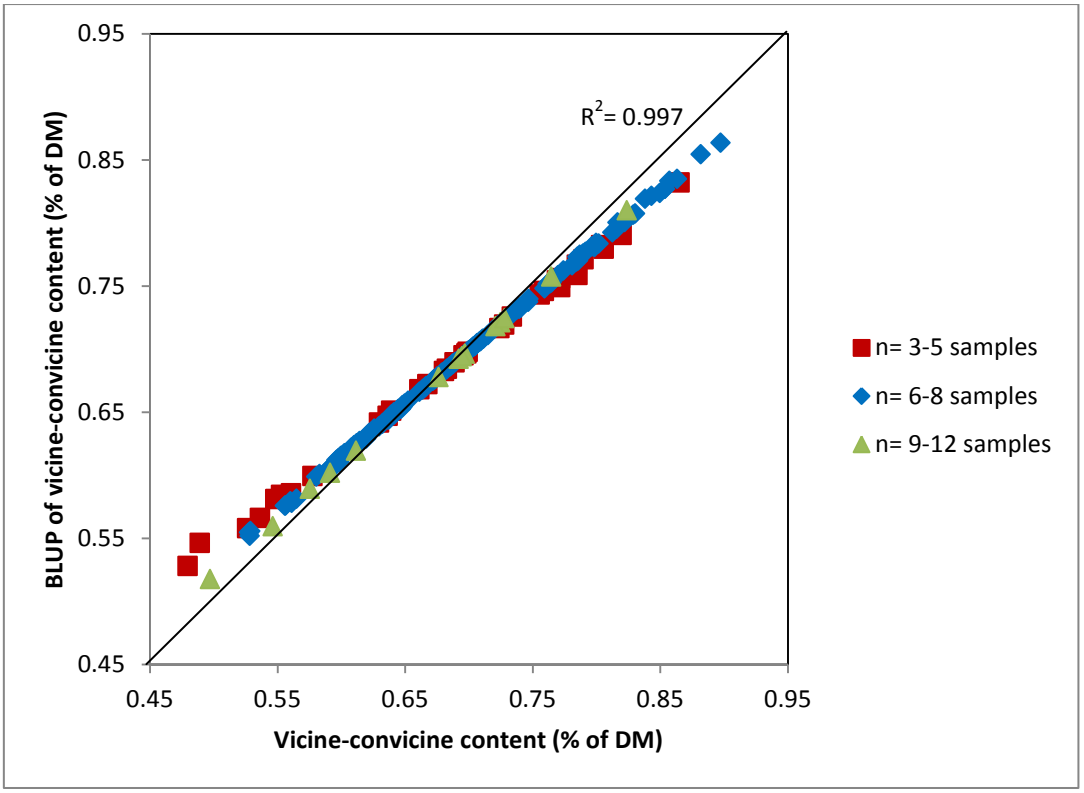



Figure 9. BLUP values of vicine-convcine content of 189 lines of A-set in comparison with 'real' data, entries grouped according to number of replicates per entry.

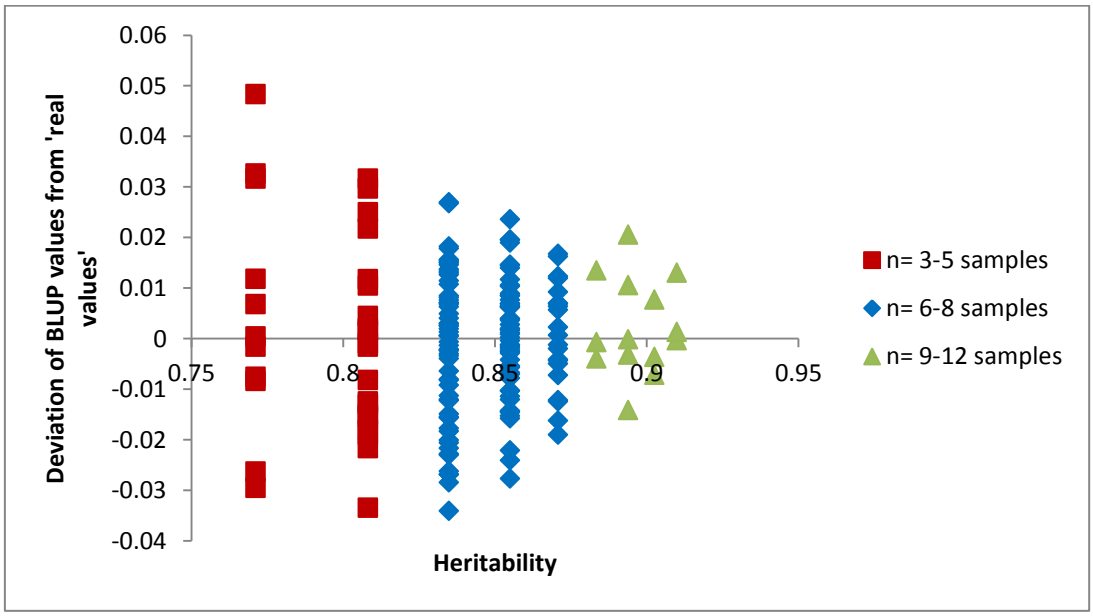


Figure 10. Heritabity and deviation of BLUP values from 'real values' with the number of existing data.

### *Association analysis*

The phenotypic, NIRS-estimated vicine-convicine content was associated to one AFLP marker (Table 18), E40M59-387, with 8.32% of phenotypic variance explained. The allele "0" (absence of peak) of the marker was associated with a higher vicine-convicine content. The marker was in unknown position based on the linkage map that was developed by Welna (2014). However, this marker was available in another linkage map (Ali, 2015). There it found a position based on a BCFam.2 population, in a linkage group 4, together with further 12 AFLP markers (Figure 11). This marker was 35.3 cM distant from common marker E40M59-467 with three markers in between. The common marker was located at position 59.9 cM in linkage group 6 (Welna, 2014) and corresponded to chromosome 5 on the consensus map of *Vicia faba* (Webb *et al.* 2016).

Table 18. Association analyses results for vicine-convicine content (minor allele frequency 5%; *n* = 189 inbred lines; mixed linear model, Kinship-matrix, FDR 20%).

| DNA marker | Linkage Group* | Position* | *p*-value | $R^2$ (%) | Increase allele** | Allele effect |
|---|---|---|---|---|---|---|
| E40M59-387 | - | - | $1.09 \times 10^{-4}$ | 8.32 | "0" | 0.066 |

* was included but not mapped in Welna (2014)

** to specify which homozygous marker class showed the higher trait expression

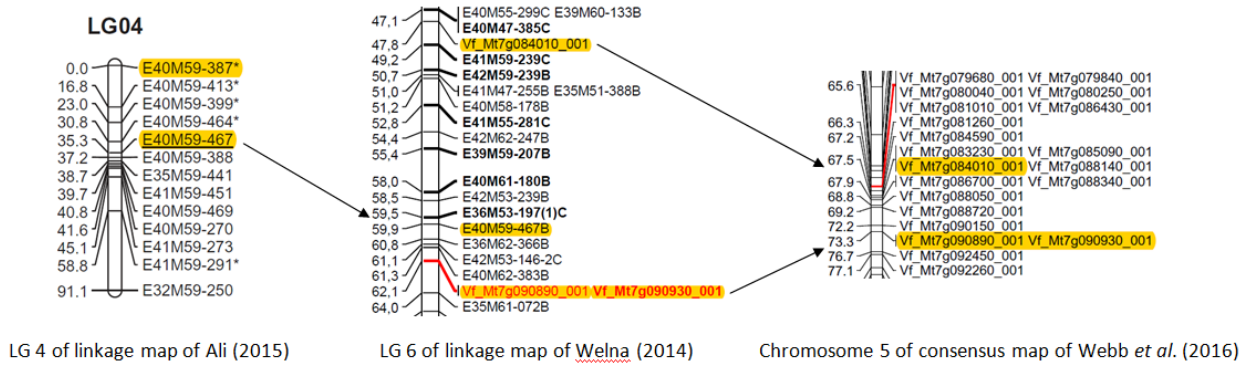



Figure 11. The scheme of position prediction of marker E40M59-387 in *Vicia faba.*

## Discussion

In the developed calibration equation for vicine-convicine, the second derivative transformation of (2, 3, 3, 2) was applied to the raw data and gave the best performance based on coefficient of determination of calibration and some parameters in semi-external validation. Derivatives of spectra have been usually used in spectroscopy due to greater amplitude than the raw spectra that can separate out peaks of overlapping band. Second derivatives can be very useful due to the fact that not only band intensity and peak location are maintained with the raw spectra, but also the increase of band resolution (Shenk *et al.* 2008).

We included a wide range of genetic materials of vicine-convicine content as calibration set. Several low vicine-convicine materials were included to complete the main genetic materials which have normal ('wild type', medium to high) vicine-convicine content. A relatively well-performing calibration equation was produced with a coefficient determination of calibration ($R^2$) and internal cross validation (1-VR) of 0.966 and 0.847 respectively. The coefficient of determination indicated good quantitative information of equation, 96.6% of vicine-convicine content variability in the calibration set were explained by the model.

Semi-external validation was conducted to verify the performance of calibration equation using only normal vicine-convicine content genetic materials. Coefficient of determination of five validation subsets showed relatively high value, which strengthen the reliability of calibration equation. Based on available literature, the usefulness of NIRS calibration frequently was estimated by using RPD and RER. The RPD values of the five semi-external validation subsets were 2.67-3.14, which ideally the ratio should be at least 3. Nevertheless, the RER values of the subset were 9.63-14.03, which ideally has minimum value of 10. Both parameters indicated that the calibration should be useful for screening purposes (Williams and Sobering, 1996; Ozaki *et al.*, 2006).

NIRS spectra recording of the samples was conducted in time-based alpha lattice randomization. Variability of vicine-convicine was significant among the samples and the proportion of genetic effect to vicine-convicine in the samples relatively high. Lattice efficiency of 14 lattices was relatively low, showing that the lattice-based randomizations

were not too efficient compared to randomized complete block randomizations of the NIRS spectra recording of samples.

Significant and large quantitative variations were found for vicine-convicine content among 189 lines of A-set in three environments in 2013 and 2014. The heritability was medium, 0.788, which showed relatively high proportion of genetic variance compared to phenotypic variance.

The intention of calculating BLUP was to in addition include results from treatments (un-tripped) that yielded very incomplete data, missing value was almost half of the un-tripped data. The calculation of BLUP values was depending on the number of existing data. A small number of the data (few existing replicates) will cause a BLUP value come closer up towards the general mean and in the extreme case (if no data available) then the BLUP value will be equal to the general mean. In contrary, a high number of available replicates will lead a BLUP value to be close to actual, direct mean value of these replicates, which makes the BLUP value more reliable. BLUP values then were used for association analysis, which yet gave similar result to the original data (details not presented here).

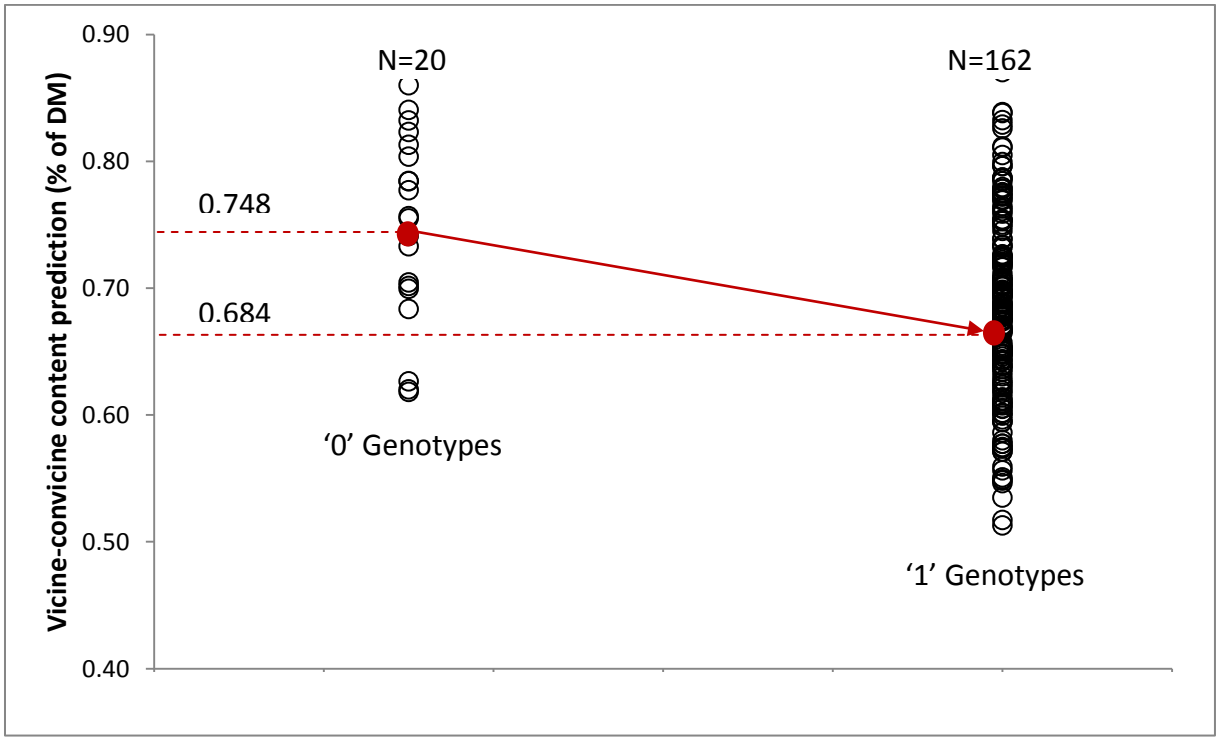



Figure 12. Allele effect of marker E40M59-387.

Association study is considered to be a powerful approach to reveal genomic polymorphism which has effect to quantitative traits. In the present study, we found one significant AFLP marker associated with vicine-convicine content, E40M59-387. The marker explained 8.32% of phenotypic variance, with its allele "0" increasing 0.066% the vicine-convicine content. A brief visualization of that allele effect of E40M59-387 is shown in Figure 12. To examine the reliability of this finding, association analysis was carried out again using general linear model (GLM) method in Tassel 3.0. The result showed that the same marker had the smallest p-value even though it was not significant with GLM.

The marker E40M59-387 was in unknown position in the linkage group of Welna (2014). The linkage group was developed by using 101 pure lines of biparental recombinant inbred lines (RILs), so-called M-set. However, that marker was available in a map of BCFam.2 population of Ali (2015), in his linkage group 4. BCFam.2 came from a cross between lines S_122 and S_253, which are part of the current A-set. This marker E40M59-387 was found in Ali (2015) in linkage group 4 together with further 12 AFLP markers. That linkage group spans 91.1 cM. Among its markers, only E40M59-467 was available in the M-set in linkage group 6 (Welna, 2014). Even though E40M59-387 has 35.3 cM distance to E40M59-467, there are 3 other markers placed between them, with the closest marker at 16.8 cM distance to E40M59-387 (Ali, 2015). 11 SNP markers were located on Welna's (2014) linkage group 6, together with E40M59-467 in the M-set. Three closest SNP markers to E40M59-467 are Vf_Mt7g090890_001, Vf_Mt7g090930_001 (both are 2.25 cM distant) and Vf_Mt7g084010_001 (12.07 cM distant). These 11 SNP markers are in the same chromosome number 5 of the consensus map of *Vicia faba* (Webb *et al.*, 2016).

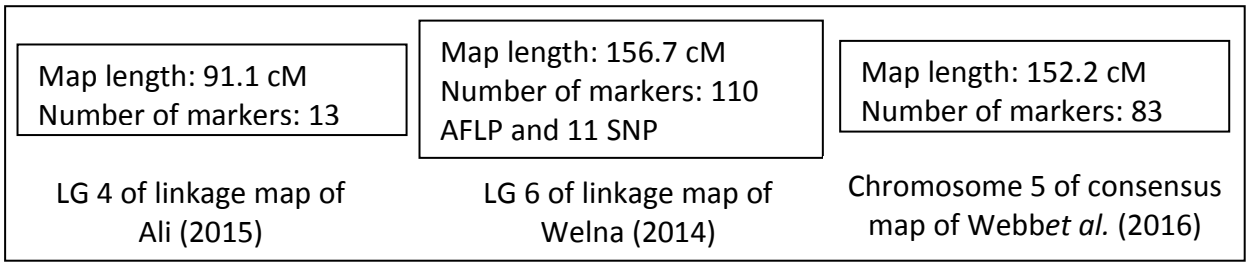



Figure 13. Map length and number of markers of linkage map of Ali (2015), Welna (2014) and consensus map (Webb *et al.* 2016).

It is very challenging to estimate the position of E40M59-387 in the consensus map due to limited available data. The common marker E40M59-467 is very likely located

between 67.5 and 73.3 cM of chromosome 5 as mapped by Webb *et al.* (2016). Taking the scales of the three different maps is equal and equally trust-worthy, then E40M59-387 is located at about 35.3 cM distance from E40M59-467 in unknown direction (upward or downward). It is clear that not only the scale among three maps may be different (Figure 13), but also the direction is not known. To have an estimation of the position of E40M59-387, the three linkage maps were merged with MergeMap (Wu *et al.*, 2008) using assumption of the same scale and direction (see Appendix 3). The estimation showed that E40M59-387 is located in 42.2 cM of the merge linkage map in approximate position between 36.06 cM and 48.65 cM in chromosome 5 of *Vicia faba*. The graph is made using MapChart 2.30 (Voorrips, 2002). We tentatively conclude that the AFLP marker E40M59-387 is very probably located at chromosome 5 of *Vicia faba*.

This finding shows that the QTL that is responsible for vicine-convicine variation in vicine-convicine-containing (wild-type) faba bean genotypes, especially in winter faba bean, is very probably different from the one in the previous study of Khazaei *et al.* (2015). In the previous study, a single QTL for vicine-convicine content was identified in chromosome 1, being the location of the *vc-* gene. The study used F5 recombinant inbred line from crossing of Mélodie/2 (low vicine-convicine) and ILB 938/2 (normal vicine-convicine).

However, further studies are still needed to identify tightly linked markers for vicine-convicine content loci in wild type faba bean genotypes. In addition, biosynthetic pathway of vicine-convicine is not known yet. The single major QTL (Khazaei *et al.*, 2015). which was found for vc- could lead to uncover biosynthetic genes for *vc-* locus. Nevertheless, a species model for faba bean genomic study, *Medicago truncatula*, does not contain vicine-convicine, so further study and analysis which involve other species probably needed to reveal vicine-convicine biosynthetic pathway.

## References


Ali, M.B., 2015. Genetic linkage maps and homology study of backcross families of German faba bean (*Vicia faba* L.). *Journal of Crop Improvement*, *29*(4), pp.474-490.

Ali, M., Welna, G.C., Sallam, A., Martsch, R., Balko, C., Gebser, B., Sass, O. and Link, W., 2016. Association analyses to genetically improve drought and freezing tolerance of faba bean (*Vicia faba* L.). *Crop Science*, *56*(3), pp.1036-1048.

Baker, M.A., Bosia, A., Pescarmona, G., Turrini, F. and Arese, P., 1984. Mechanism of action of divicine in a cell-free system and in glucose-6-phosphate dehydrogenase-deficient red cells. *Toxicologic Pathology*, *12*(4), pp.331-336.

Barnes, R.J., Dhanoa, M.S. and Lister, S.J., 1989. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra.*AppliedSpectroscopy*, *43*(5), pp.772-777.

Bendich, A. and Clements, G.C., 1953. A revision of the structural formulation of vicine and its pyrimidine aglucone, divicine. *Biochimica et Biophysica acta*, *12*(1), pp.462-477.

Benjamini, Y. and Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, pp.289-300.

Bien, S., Salemnik, G., Zamir, L. and Rosenblum, M., 1968. The structure of convicine. *Journal of the Chemical Society C: Organic*, pp.496-499.

Bjerg, B., Heide, M., Knudsen, J.C.N. and Soerensen, H., 1984. Inhibitory effects of convicine, vicine and dopa from *Vicia faba* on the in vitro growth rates of fungal pathogens/Hemmwirkung von Convicin, Vicin und Dopa aus *Vicia faba* auf in vitro Wachstumsraten von pilzlichen Krankheitserregern. *Zeitschrift für Pflanzenkrankheiten und Pflanzenschutz/Journal of Plant Diseases and Protection*, pp.483-487.

Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y. and Buckler, E.S., 2007. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, *23*(19), pp.2633-2635.

Burbano, C., Cuadrado, C., Muzquiz, M. and Cubero, J.I., 1995. Variation of favism-inducing factors (vicine, convicine and L-DOPA) during pod development in *Vicia faba* L. *Plant foods for Human Nutrition*, *47*(3), pp.265-274.

Cappellini, M.D. and Fiorelli, G., 2008. Glucose-6-phosphate dehydrogenase deficiency. *The Lancet*, *371*(9606), pp.64-74.

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

Crépon, K., Marget, P., Peyronnet, C., Carrouée, B., Arese, P. and Duc, G., 2010. Nutritional value of faba bean (*Vicia faba* L.) seeds for feed and food. *Field Crops Research*, *115*(3), pp.329-339.

Duc, G., Sixdenier, G., Lila, M. and Furstoss, V., 1989. Search of genetic variability for vicine and convicine content in *Vicia faba* L.: a first report of a gene which codes for nearly zero-vicine and zero-convicine contents. In *1. International Workshop on'Antinutritional Factors (ANF) in Legume Seeds', Wageningen (Netherlands), 23-25 Nov 1988*. Pudoc.

Duc, G., Marget, P., Page, D., Domoney, C. and Dijon Cedex, F., 2004. Facile breeding markers to lower contents of vicine and convicine in faba bean seeds and trypsin inhibitors in pea seeds. *Publication-European Association for Animal Production*, *110*, pp.281-286.

Duc, G., 1997. Faba bean (*Vicia faba* L.). *Field Crops Research*, *53*(1-3), pp.99-109.

Dutta, P.K., Chakravarty, A.K., Chowdhury, U.S. and Pakrashi, S.C., 1981. Vicine, a favism-inducing toxin from Momordica charantia Linn. seeds. *Indian J. Chem. B*, *20*, pp.669-671.

FAOSTAT, 2014. Food and Agriculture Organization. Website: http://faostat.fao.org/. Visited at 7th of March, 2017.

Ferhatoglu, Y. and Vandenberg, A., 2015. Re-evaluating TiCl4 and UV assays for detection of vicine and convicine in high-throughput screening of immature and mature seeds of faba bean. *African Journal of Plant Science*, *9*(12), pp.476-489.

Font, R., del Río-Celestino, M. and de Haro-Bailón, A., 2006. The use of near-infrared spectroscopy (NIRS) in the study of seed quality components in plant breeding programs. *Industrial Crops and Products*, *24*(3), pp.307-313.

Griffiths, D.W. and Ramsay, G., 1992. The concentration of vicine and convicine in *Vicia faba* and some related species and their distribution within mature seeds. *Journal of the Science of Food and Agriculture*, *59*(4), pp.463-468.

Griffiths, D.W. and Ramsay, G., 1996. The distribution of pyrimidinone glucosides in developing seedlings of *Vicia faba* and *Vicia narbonensis*. *Journal of the Science of Food and Agriculture*, *72*(4), pp.469-475.

Hussein, L., Motawei, H., Nassib, A. and Khalil, S., 1986. The complete elimination of vicine and convicine from the faba beans by combinations of genetic selection and processing techniques. *Plant Foods for Human Nutrition*, *36*(3), pp.231-242.

Jamalian, J. and Bassiri, A., 1978. Variation in vicine concentration during pod development in broad bean (*Vicia faba* L.). *Journal of Agriultural and Food Chemistry*, *26*(6), pp.1454-1456.

Koivunen, E., Tuunainen, P., Valkonen, E., Rossow, L. and Valaja, J., 2014. Use of faba beans (*Vicia faba* L.) in diets of laying hens. *Agricultural and Food Science*, *23*(3), pp.165-172.

Khamassi, K., Jeddi, F.B., Hobbs, D., Irigoyen, J., Stoddard, F., O'Sullivan, D.M. and Jones, H., 2013. A baseline study of vicine–convicine levels in faba bean (*Vicia faba* L.) germplasm. *Plant Genetic Resources*, *11*(03), pp.250-257.

Khazaei, H., O'Sullivan, D.M., Jones, H., Pitts, N., Sillanpää, M.J., Pärssinen, P., Manninen, O. and Stoddard, F.L., 2015. Flanking SNP markers for vicine–convicine concentration in faba bean (*Vicia faba* L.). *Molecular Breeding*, *35*(1), p.38.

Lessire, M., Gallo, V., Prato, M., Akide-Ndunge, O., Mandili, G., Marget, P., Arese, P. and Duc, G., 2016. Effects of faba beans with different concentrations of vicine and convicine on egg production, egg quality and red blood cells in laying hens. *Animal*, pp.1-9.

Marquardt, R.R., Campbell, L.D. and Ward, T., 1976. Studies with chicks on the growth depressing factor (s) in faba beans (*Vicia faba* L. var. minor). *The Journal of Nutrition*, *106*(2), pp.275-284.

McKay, A.M., 1992. Hydrolysis of vicine and convicine from faba beans by microbial β-glucosidase enzymes. *Journal of Applied Bacteriology*, *72*(6), pp.475-478.

Mehta, A., Mason, P.J. and Vulliamy, T.J., 2000. Glucose-6-phosphate dehydrogenase deficiency. *Best Practice & Research Clinical Haematology*, *13*(1), pp.21-38.

Muduuli, D.S., Marquardt, R.R. and Guenter, W., 1982. Effect of dietary vicine and vitamin E supplementation on the productive performance of growing and laying chickens. *British Journal of Nutrition*, *47*(01), pp.53-60.

O'Sullivan, D.M. and Angra, D., 2016. Advances in faba bean genetics and genomics. *Frontiers in Genetics*, *7*.

Ozaki, Y., McClure, W.F. and Christy, A.A. eds., 2006. *Near-infrared spectroscopy in food science and technology*. John Wiley & Sons.

Pasricha, V., Satpathy, G. and Gupta, R.K., 2014. Phytochemical & Antioxidant activity of underutilized legume *Vicia faba* seeds and formulation of its fortified biscuits. *Journal of Pharmacognosy and Phytochemistry*, *3*(2), pp.75-80.

Pitz, W.J., Sosulski, F.W. and Hogge, L.R., 1980. Occurrence of vicine and convicine in seeds of some Vicia species and other pulses. *Canadian Institute of Food Science and Technology Journal, 13(1),* pp.35-39.

Pulkkinen, M., Gautam, M., Lampi, A.M., Ollilainen, V., Stoddard, F., Sontag-Strohm, T., Salovaara, H. and Piironen, V., 2015. Determination of vicine and convicine from faba bean with an optimized high-performance liquid chromatographic method. *Food Research International*, *76*, pp.168-177.

Ray, H. and Georges, F., 2010. A genomic approach to nutritional, pharmacological and genetic issues of faba bean (*Vicia faba*): prospects for genetic modifications. *GM crops*, *1*(2), pp.99-106.

Rinnan, Å., van den Berg, F. and Engelsen, S.B., 2009. Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends in Analytical Chemistry*, *28*(10), pp.1201-1222.

Rizzello, C.G., Losito, I., Facchini, L., Katina, K., Palmisano, F., Gobbetti, M. and Coda, R., 2016. Degradation of vicine, convicine and their aglycones during fermentation of faba bean flour. *Scientific Reports*, *6*, pp.1-11.

Shenk, J.S. and Westerhaus, M.O., 1991. Populations structuring of near infrared spectra and modified partial least squares regression. *Crop Science*, *31*(6), pp.1548-1555.

Shenk, J.S., Workman Jr, J.J. and Westerhaus, M.O., 2008. 17Application of NIR Spectroscopy to Agricultural Products.

Utz, H.F., 2001. PLABSTAT: a computer program for statistical analysis of plant breeding experiments. *Institute for Plant Breeding, Seed Science and Population Genetics, University of Hohenheim, Stuttgart*.

Voorrips, R.E., 2002. MapChart: software for the graphical presentation of linkage maps and QTLs. *Journal of heredity*, *93*(1), pp.77-78.

Wang, J., Liu, H. and Ren, G., 2014. Near-infrared spectroscopy (NIRS) evaluation and regional analysis of Chinese faba bean (*Vicia faba* L.). *The Crop Journal*, *2*(1), pp.28-37.

Webb, A., Cottage, A., Wood, T., Khamassi, K., Hobbs, D., Gostkiewicz, K., White, M., Khazaei, H., Ali, M., Street, D. and Duc, G., Stoddard, F.L., Maalouf, F., Ogbonnaya, F.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*(1), pp.177-185.

Welna, G.C., 2014. Genetische Analysen für eine markergestützte Verbesserung der Trockenstresstoleranz von Winterackerbohnen. PhD thesis, University of Gottingen.

Williams, P.C. and Sobering, D.C., 1996. How do we do it: a brief summary of the methods we use in developing near infrared calibrations. *Near infrared spectroscopy: The future Waves*, pp.185-188.

Wu, Y., Close, T.J. and Lonardi, S., 2008, August. On the accurate construction of consensus genetic maps. In *Comput Syst Bioinformatics Conf* (Vol. 7, pp. 285-296).

Yahia, Y., Elfalleh, W., Tlili, N., Hannachi, H., Loumerem, M. and Ferchichi, A., 2013. Phytochemical contents and antioxidant activities of some Tunisian faba bean populations. *Romanian Agricultural Research*, *30*, pp.65-74.

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

# General Discussion

In the present thesis, reproductive features (autofertility) and quality features (vicine-convicine content of seed) were studied in 189 inbred lines of winter faba bean. These lines were derived from a winter faba bean population, which was initially created by mixing 11 founder lines from Germany, French and UK. With the various genetic backgrounds of these 11 winter bean founder lines, it is expected that the 189 lines represent a winter bean germplasm that hold promise for Western and Central Europen environments, especially referring to winter conditions. This germplas pool is currently the major germplasm to breed winter faba beans for Germany.

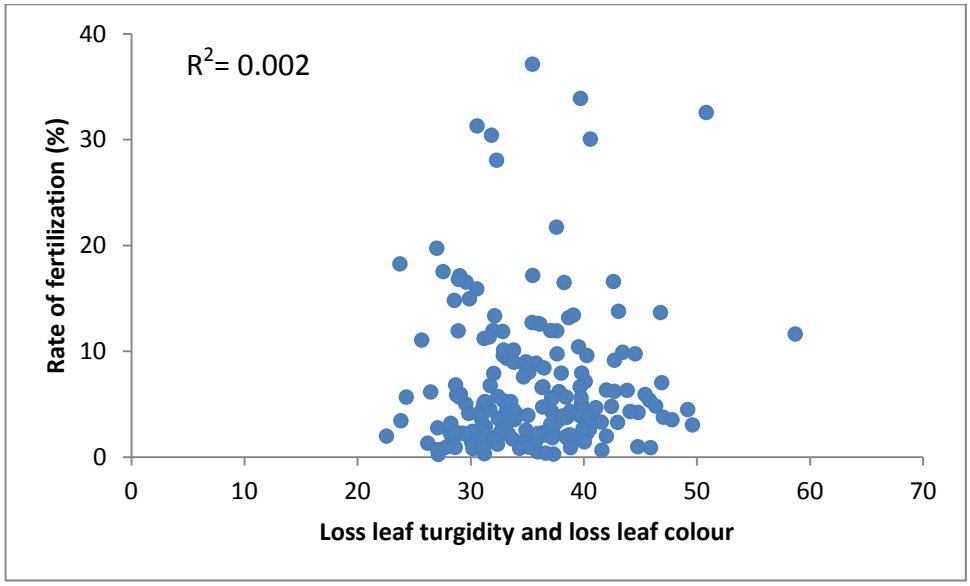



Figure 14. Correlation of LT+LC and rate of fertilization of 189 lines of A-set.

Autofertility is a specific term in faba bean, to emphasize the ability to spontaneously self fertilize, this means, without any external mechanical support. The level of autofertility is genetically variable, and the present chapter 1 is focussed on autofertility in these winter faba bean lines. To assess autofertility, the study focused on rate of fertilization, potential pod filling and actual pod filling in un-tripped treatment. Sallam (2014) and Ali *et al.* (216) studied the very same genetic materials for frost and drought tolerance. Loss of leaf turgidity plus loss of leaf colour (LT+LC) was one of the most important capability to survive their freezing temperature treatments. The correlation between LT+LC and the rate of fertilization of the 189 lines of A-set was not significant (Figure 14); frost tolerance does not correlate to

autofertility. Without marked and significant correlation, especially with the correlation not being negative, there is large opportunity to uncover frost tolerant genotypes with a higher level of autofertility. However, the present study showed that the autofertility level of 189 lines of A-set is lower than that of spring beans used. Hence, if the correlation analyis was conducted with a genetic material consisting of both types, winter beans and spring beans, then surely a clear correlation would occur between higher frost-tolerance and lower autofertility.

Although there are clear facts showing that faba bean seeds give benefits as human food and animal feed, the vicine-convicine compound in seeds is limited its uses. Vicine and convicine are pyrimidine glycosides compounds which are found in relatively significant amount in faba bean seeds. The present experiment used 189 lines which have normal (wild-type) vicine-convicine content. It was observed to vary from 0.45-0.90% (using the NIRS-based prediction approach). This range certainly differs from the very low content of vicine-convicine, being near to zero, as exhibited by Mélodie-7 (the HPLC results were 0.02%, Melodie-7 was also used in the experiment). The correlation between LT+LC and vicine-convicine content of 189 lines of A-set is as well not significant (Figure 15) which showed, again, frost tolerance does not correlate to vicine-convicine content.

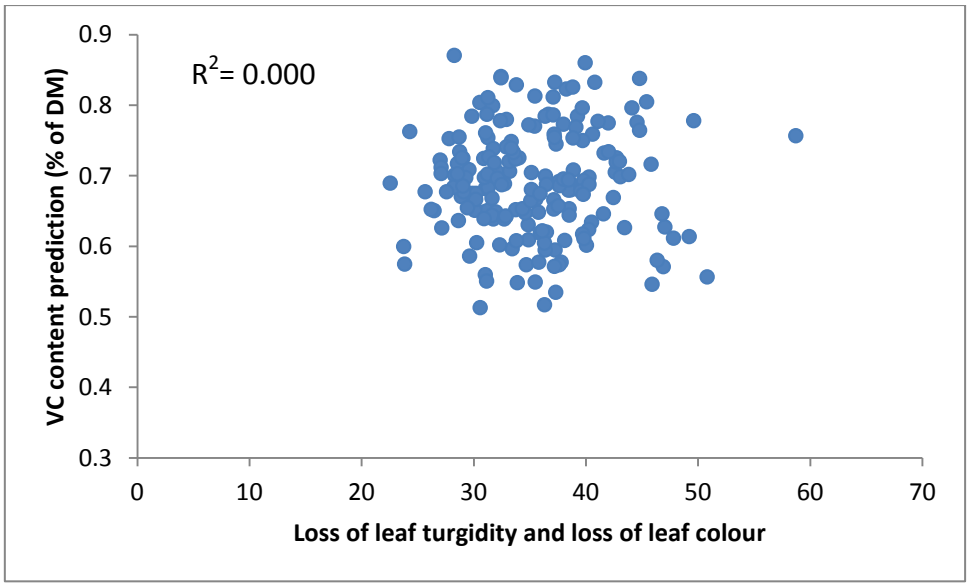



Figure 15. Correlation of LT+LC and vicine-convicine content of 189 lines of A-set.

The genetic materials used in the present study were developed almost thirty years ago by using the 11 founder winter bean lines. These genetic materials were already used for

several purposes especially related to freezing tolerance. The present study is the first autofertility observation and vicine-convicine analysis using winter faba beans adapted for Central European conditions. The A-set tends to have low rate of fertilization in un-tripped treatment. This is in accordance to the previous study of Stoddard (1986) which showed the lower autofertility level of UK winter faba bean than spring beans. With the above-shown very low frost tolerance-autofertility correlation, wider genetic materials of winter faba bean in breeding program could probably allow to improve its level of autofertility. More or less the same is true for vicine-convicine content. The A-set lines contain normal vicine-convicine content. Due to its unfavourable effect, finding a faba bean with a lowered vicine-convicine, even in winter faba bean, has special importance.

The genetic map which was used in the present study spans 1633.2 cM with 1159 marker loci in seven major linkage groups (Welna, 2014). The map size was comparable and reasonable when compared with other maps size developed by Ellwood *et al.* (2008), by Kaur *et al.* (2014) and by Webb *et al.* (2016). The map of Welna (2014) comprised a high number of AFLP markers together with 111 SNP markers. Each major linkage group comprised 9 to 22 SNP which can be associated with the six chromosomes of *Vicia faba* via SNP markers which are common with the consensus map of Webb *et al.* (2016; Welna, 2014; Ali *et al.*, 2016).

A genome wide association study was undertaken here, to identify QTLs that are responsible for both autofertility-related traits and vicine-convicine content. In the study of autofertility, nine markers were found associated with several agronomic traits which explained little of the phenotypic variation. However, no significance association of autofertility-related traits of the 189 lines of winter beans could be found. One AFLP marker was significantly associated to vicine-convicine content with a rather limited quantity of explained variation. Ali *et al.* (2016) when using the same genetic material and marker did, similarly, not find associated markers for all their studied traits, only could detect six significance association out of eight observed traits. As discussed by them, here again, a major reason for not detecting associations between markers and trait may be the very low linkage disequilibrium in the A-set material. Among all possible marker pairs of 175 SNP and 1147 AFLP markers, the average LD ($r^2$) was very low, about 0.0077 (Ali *et al.*, 2016).

Further studies are required to detect and mark QTLs for autofertility and for vicine-convicine in wild-type germplasm, and furthermore to identify the genes which are

responsible for genetic variation in autofertility level and in vicine-convicine in faba bean. With the recent acceleration in research and progress of faba bean sequence and marker data sets (O'Sullivan and Angra, 2016), there has been a marked increase in density and utility of gene-based genetic maps. Moreover, this could lead to better anchor genetic maps to the well-established sequenced legume, *Medicago truncatula* (i.e., to better exploit synteny between *Vicia faba* and *Medicago truncatula*). One of the latest attempts to discover a gene in fababeans focussed on the gene ZT-1 which controls the flower pigmentation and seed coat tannin. The gene was mapped to chromosome 2 *ofVicia faba* and by synteny to chromosome 3 of *Medicago truncatula (*O'Sullivan and Angra, 2016*)*. This finding revealed a logical biological process in the form of a transparent testa transcription factor which had previously studied in *Medicago truncatula* to determine flower colour (Webb *et al.*, 2016, O'Sullivan and Angra, 2016). A similar situation could exist in vicine-convicine. The VC-locus with alleles VC- and vc-, the latter causes very low vicine-convicine content in faba bean seeds, mapped to chromosome 1 of *Vicia faba* and via synteny to chromosome 2 of *Medicago truncatula* (Khazaei *et al.*, 2015). This should be a stimulus to further develop high-resolution mapping of that locus. A similar strategy could be applied in further studies to find the more important wild-type vicine-convicine QTLs. With the initial identification and location these QTLs, SNP mining in syntenic intervals could be conducted to further develop high-resolution mapping and high precicion localization of the QTLs.

## References


Ali, M., Welna, G.C., Sallam, A., Martsch, R., Balko, C., Gebser, B., Sass, O. and Link, W., 2016. Association analyses to genetically improve drought and freezing tolerance of faba bean (*Vicia faba* L.). *Crop Science*, *56*(3), pp.1036-1048.

Khazaei, H., O'Sullivan, D.M., Jones, H., Pitts, N., Sillanpää, M.J., Pärssinen, P., Manninen, O. and Stoddard, F.L., 2015. Flanking SNP markers for vicine–convicine concentration in faba bean (*Vicia faba* L.). *Molecular Breeding*, *35*(1), p.38.

O'Sullivan, D.M. and Angra, D., 2016. Advances in faba bean genetics and genomics. *Frontiers in Genetics*, *7*.

Sallam, A., 2014. Detailed genetic approach to improve frost tolerance of German winter faba beans. PhD thesis, University of Gottingen.

Webb, A., Cottage, A., Wood, T., Khamassi, K., Hobbs, D., Gostkiewicz, K., White, M., Khazaei, H., Ali, M., Street, D. and Duc, G., Stoddard, F.L., Maalouf, F., Ogbonnaya, F.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*(1), pp.177-185.

Welna, G.C., 2014. Genetische Analysen für eine markergestützte Verbesserung der Trockenstresstoleranz von Winterackerbohnen. PhD thesis, University of Gottingen.

# Summary

Faba bean (*Vicia faba* L.) is an allogamous plant which allows both self and cross fertilization. Self fertilization which occurs without pollinators or without any other external mechanic stimulus in faba bean is the so-called autofertility. The degree of such autofertility varies among genotypes. Once the seed is set, either by selfing or by crossing, the amount and quality of the seed mass is decisive. Faba bean seeds supply rich-protein content feedstuff and provide a valuable nutrition composition. However, faba bean contains anti-nutritive compounds such as vicine and convicine which limit its use in feed and food systems and have health impact for human. The objectives of the first chapter of the present study are to genetically study and quantify level and variation of autofertility in a specific winter faba bean breeding germplasm and to identify QTL for autofertility and related traits. Hence, the first part's focus is on fertilization and thus genesis of seed. The second chapter's focus is on the quality of seed. It aims to develop a NIRS-based so-called calibration for the vicine-convicine content in faba bean seed, to study heritability and genetic variation of vicine-convicine content in faba bean, to identify QTLs that are responsible for vicine-convicine variation in vicine-convicine-containing (wild-type) faba bean genotypes and to verify whether the mutant allele for vicine-convicine in faba bean ("vc-") is allelic to a QTL for the variation in vicine-convicine-containing materials.

A number of field and laboratory experiments were carried out to genetically study features of reproduction and of seed quality of faba bean. The main genetic materials used in this study involved 200 inbred lines, named Q-set, which consisted of 189 lines of A-set (inbred lines for association study), seven further winter bean lines and four further spring bean lines. The A-set was derived from the so-called Göttingen Winter Bean Population (GWBP). The experiments to study features of reproduction were conducted in so-called bee-proof isolation houses in 2013, 2014 and 2015. Treatments of "tripped" and "un-tripped" were applied to the faba bean flowers during flowering time. The experiment to study seed quality (vicine-convicine content) was carried out by HPLC and by NIR-spectrophotometry analysis of the faba bean seeds harvested from the above-mentioned experiments. We developed a NIRS calibration to allow for a NIRS-based prediction of seed vicine-convicine content. Genome-wide association analyses between DNA-markers and

phenotypic expression of traits was carried out using TASSEL version 3.0. A total of 2018 polymorphic markers were used consisting of 189 SNP (Single Nucleotide Polymorphism) and 1829 AFLP (Amplified Fragment Length Polymorphism).

To assess autofertility, the study focused on rate of fertilization, potential pod filling and actual pod filling, especially in un-tripped treatment. Rate of fertilization of un-tripped treatment was low, with maximum of 37.14%, and high in heritability. Tripping obviously increased the mean values of the three aspects of autofertility. Higher heritability of rate of fertilization in tripped than in un-tripped treatment indicated genetic differences for the reaction of tripping. Intensive tripping that has been carried out confirmed the result and showed that none of these genotypes showed a 100% result for rate of fertilization. A major result from the present study was that winter faba bean has a different, lower level of autofertility than spring beans. NIRS technology can be applied to predict vicine-convicine content in faba bean. A relatively well-performing calibration equation was produced and applied to analyse samples of faba bean seeds across different replicates, treatments and years. Significant and large quantitative variations were found for vicine-convicine content with relatively high heritability. Our study resulted in several putative DNA-markers which are significantly related to several of the agronomic features in faba bean as well as to vicine-convicine content. One AFLP-marker was significantly associated to the vicine-convicine variation in the genotypes and, by carefully inspecting three different linkage maps and the syntenic relationship to *Medicago truncatula*, the very probable position of that QTL was determined on chromosome number 5 of *Vicia faba*.

The presented findings are a further step forward in research and breeding of highly fertile European winter faba beans with improved seed quality.

# Zusammenfassung

Die Ackerbohne (*Viciafaba*) ist eine partiell allogame Pflanze, die zugleich Selbst- als auch Fremdbefruchtung erlaubt. Selbstbefruchtung, welche ohne Bestäuber oder ohne andere externe mechanische Stimuli geschieht, ist die sogenannte Autofertilität. Der Grad dieser Autofertilität variiert zwischen Genotypen. Sobald Samen angesetzt sind, ausSelbstung oder aus Kreuzung, sind die Menge und die Qualität der Samenmasse entscheidend. Ackerbohnen-Samen sind ein proteinreicherRohstoff und bieten eine ernährungsrelevantwertvolle Zusammensetzung. Ackerbohnen enthalten allerdings anti-nutritive Inhaltsstoffe wie Vicin und Convicin, die die Nutzung in Futter oder Nahrungsmitteln begrenzen und für Menschen gesundheitlich bedeutsamsind.

Die Ziele des ersten Kapitels dieser Studie sind, das Niveau und die Variation der Autofertilität und verwandter Merkmale in einem spezifischen Winterackerbohnen-Genpool züchterisch zu studieren und QTL für Autofertilität und verwandter Merkmale zu identifizieren. Entsprechend fokussiert das erste Kapitel auf die Befruchtung und also auf die Entstehung von Samen. Das zweite Kapitel betrachtet die Qualität von Samen. Es zielt auf die Entwicklung einer NIRS-basierten sogenannten Kalibration für den Vicin und Convicin-Gehalt in Ackerbohnen-Samen, auf die Identifizierung von QTL die für die Variation von Vicin und Convicin in Vicin- und Convicin-enthaltenden Ackerbohnen (Wild-Typ), und auf die Prüfung der Frage, ob das Mutanten-Allel „vc-“ der Ackerbohne kosegregiert mit einem Vicin-Convicin-QTL in Vicin und Convicin-enthaltendem Material.

Mehrere Feld- und Laborversuche wurden durchgeführt, um züchterisch Reproduktionsmerkmale und Samenqualität der Ackerbohne zu studieren. Das hauptsächliche genetische Material, welches in der hier vorliegenden Studie benutzt wurde, umfasst 200 Inzuchtlinien, genannt Q-Satz, der aus 189 Linien des A-Satzes besteht (Inzuchtlinien für Assoziationsanalyse), sieben weitere Winterbohnen-Inzuchtlinien und vier Sommerbohnen-Linien. Der A-Satz war aus der sogenannten Göttinger Winterackerbohnen-Population (GWBP) entwickelt worden. Die Versuche zum Studium der Reproduktion wurden in sogenannten bienensicheren Isolierhäusern durchgeführt, in den Sommern 2013, 2014 und 2015. Als Behandlung wurden die Ackerbohnenblüten während der Blütezeit „getrippt“ oder „ungetrippt“ gelassen. Die Versuche zur Analyse der Samenqualität (Vicin- und Convicin-Gehalt) wurden mit HPLC und NIR-spektroskopischer Analyse der

Ackerbohnen-Samen durchgeführt; diese waren von den oben genannten Versuchen geerntet worden. Es wurde eine NIRS Kalibration entwickelt, um eine NIRS-basierte Vorhersage des Samen-Vicin- und Convicin-Gehaltes zu erlauben. Genomweite Assoziationsanalysen zwischen DNS-Markern und der phänotypischen Merkmalsausprägung wurden mittels TASSEL (Version 3.0) durchgeführt. Es wurden insgesamt 2018 polymorphe Marker benutzt, die aus 189 SNP (Single Nucleotide Polymorphism) und aus 1829 AFLP (Amplified Fragment LengthPolymorphism) bestanden.

Um die Autofertilität zu bestimmen, konzentrierte sich die Studie auf die Befruchtungsrate, das Hülsenfüllungs-Potential und die realisierte Hülsenfüllung, besonders in der ungetrippten Behandlung. Die Befruchtungsrate in der ungetrippten Behandlung war niedrig, mit einem Maximum von 37,14%, und einer hohen Erblichkeit. Das Trippen erhöhte offensichtlich den Mittelwert der drei Autofertilitäts-Aspekte. Die höhere Erblichkeit der Befruchtungsrate in der getrippten Behandlung, im Vergleich zu ungetrippt, ist ein Indiz für genetische Unterschiede in der Reaktion auf das Trippen. Intensives Trippen, welches durchgeführt wurde, bestätigte die Ergebnisse und zeigte, dass keiner dieser Genotypen eine 100% Befruchtung (Autofertilität) zeigte. Ein wesentliches Ergebnis der vorliegenden Studie war, dass Winterackerbohnen eine andere, niedrigere Autofertilität haben als Sommerackerbohnen.

NIRS Technologie kann zur Vorhersage des Vicin- und Convicin-Gehaltes von Ackerbohnen eingesetzt werden. Eine relativ gute Kalibrationsgleichung wurde entwickelt und zur Analyse von Ackerbohnen-Samenproben aus verschiedenen Wiederholungen, Behandlungen und Jahren eingesetzt. Signifikante und große quantitative Unterschiede wurden mit hoher Erblichkeit für Vicin und Convicin-Gehalt gefunden. Diese Studie ergab mehrere mutmaßliche DNS-Marker, die signfikant zu mehreren der agronomischen Merkmale der Ackerbohnen und zum Vicin und Convicin-Gehalt in Beziehung stehen. Ein AFLP-Marker war signifikant mit dem Vicin und Convicin-Gehalt assoziiert; mittels sorgfältiger Durchsicht von drei verschiedenen Kopplungskarten und der syntänischen Beziehung zu *Medicagotruncatula* wurde die sehr wahrscheinliche Position dieses QTL bestimmt, als auf dem Chromosom 5 von *Viciafaba* liegend.

Die vorgestellten Befunde sind ein weiterer Schritt voran in Forschung und Züchtung von hoch fertilen europärchen Winterackerbohnen mit verbesserter Samenqualität.

# Rangkuman

Kacang faba (*Vicia faba* L.) merupakan tanaman alogami yang dapat melakukan pembuahan sendiri maupun pembuahan silang. Pembuahan sendiri terjadi tanpa adanya bantuan polinator maupun rangsangan mekanik eksternal, yang pada kacang faba dikenal dengan istilah autofertilitas. Level autofertilitas bervariasi di antara genotipe. Ketika pembuahan terjadi, baik pembuahan sendiri maupun silang, jumlah dan kualitas biji sangat menentukan. Biji kacang faba kaya akan protein dan mengandung komposisi nutrisi lain yang bernilai tinggi. Namun demikian, kacang faba mengandung senyawa anti nutrisi, seperti vicin dan convicin, yang membatasi pemanfaatannya sebagai pangan dan pakan serta memiliki dampak kesehatan bagi manusia. Tujuan dari penelitian ini pada bab pertama adalah untuk mempelajari secara genetik dan mengukur level dan variasi autofertilitas pada kacang faba winter yang spesifik dan untuk mengidentifikasi QTL untuk autofertilitas dan karakter terkait. Jadi fokus bab pertama adalah pada pembuahan dan asal biji. Fokus bab kedua adalah pada kualitas biji yang dihasilkan. Penelitian bab kedua bertujuan untuk mengembangkan kalibrasi berbasis NIRS untuk kandungan vicin-convicin pada biji kacang faba, mempelajari heritabilitas dan variasi genetik kandungan vicin-convicin, mengidentifikasi QTL yang berperan untuk variasi genotipe kacang faba yang mengandung vicin-convicin (tipe liar) dan memverifikasi apakah alel mutan yang terdapat pada vicin-convicin alelik dengan QTL pada material yang mengadung vicin-convicin.

Sejumlah eksperimen dilakukan di lapang dan laboratorium untuk mempelajari secara genetik karakter reproduksi dan kualitas biji kacang faba. Materi genetik utama yang digunakan pada penelitian ini melibatkan 200 galur murni, bernama set-Q, yang terdiri dari 189 galur set-A (galur murni untuk studi asosiasi), tujuh galur kacang faba winter dan empat galur kacang faba spring. Set-A berasal dari Göttingen Winter Bean Population (GWBP). Studi fitur reproduksi dilakukan pada rumah isolasi bebas lebah pada 2013, 2014 dan 2015. Perlakuan “tripped” dan “un-tripped” diterapkan pada bunga faba selama musim berbunga. Studi kualitas biji (kandungan vicin-convicin) dilakukan dengan menggunakan analisa HPLC dan spektrofotometri NIR pada biji yang dihasilkan pada penelitian sebelumnya. Kami mengembangkan kalibarsi NIRS untuk memprediksi kandungan vicin-convicin berbasis NIRS. Analisis asosiasi seluruh genom (GWAS) antara penanda DNA dengan karakter fenotipik

dilakukan dengan menggunakan TASSEL version 3.0. Sebanyak 2018 penanda polimorfik digunakan yang terdiri dari 189 SNP (Single Nucleotide Polymorphism) dan 1829 AFLP (Amplified Fragment Length Polymorphism).

Untuk menentukan autofertilitas, penelitian difokuskan pada karakter prosentase pembuahan, potensi pengisian polong dan pengisian polong sebenarnya, terutama pada perlakuan 'un-tripped'. Prosentase pembuahan pada perlakuan 'un-tripped' rendah, dengan nilai maksimum 37,14% dan heritabiltasnya tinggi. Tripping secara nyata meningkatkan nilai rata-rata ketiga aspek autofertilitas tersebut. Heritabilitas prosentase pembuahan pada perlakuan 'tripped' lebih tinggi daripada 'un-tripped' yang mengindikasikan perbedaan reaksi tripping merupakan faktor genetik. Tripping yang intens juga mengkonfirmasi hasil tersebut dan menunjukkan bahwa tidak satupun genotype menghasilkan prosentase pembuahan 100%. Hasil penting dari penelitian ini adalah kacang faba winter memiliki level autofertilitas yang berbeda dan lebih rendah dibandingkan dengan kacang faba spring. Teknologi NIRS dapat diterapkan untuk memprediksi kandungan vicin-convicin pada kacang faba. Diperoleh persamaan kalibrasi yang baik dan dapat diterapkan untuk menganalisa contoh biji kacang faba yang dihasilkan pada pengulangan, perlakuan dan tahun yang berbeda. Diperoleh variasi kandungan vicin-convicin yang lebar dan signifikan dengan nilai heritabilitas yang cukup tinggi. Penelitian ini juga menghasilkan beberapa penanda DNA putatif yang secara signifikan terasosiasi dengan beberapa karakter agronomi dan juga kandungan vicin-convicin. Satu penanda AFLP berasosiasi signifikan terhadap variasi vicin-convicin pada genotype, dan dengan menganalisa lebih jauh tiga peta keterpautan yang berbeda dan hubungan sinteni dengan *Medicago truncatula*, posisi QTL tersebut sangat mungkin berada pada kromosom 5 *Vicia faba*.

Penemuan ini merupakan langkah awal untuk penelitian dan pemuliaan kacang faba winter Eropa yang tinggi fertilitas dengan kualitas biji yang baik.

# Appendix 1

**Description of 246 samples of Calibration-set**

| No | Source | Pedigree | Sample Identification |
|---|---|---|---|
| 1 | Green Foil House 2013 | S_002-1-1-3 | A-set |
| 2 | Green Foil House 2013 | S_005-1-1-1 | A-set |
| 3 | Green Foil House 2013 | S_016-1-1-3-1 | A-set |
| 4 | Green Foil House 2013 | S_020-1-2-6 | A-set |
| 5 | Green Foil House 2013 | S_034-1-2 weisse blüte,wn-1 | A-set |
| 6 | Green Foil House 2013 | S_046-1-1-1-5 | A-set |
| 7 | Green Foil House 2013 | S_052-1-1-1-3 | A-set |
| 8 | Green Foil House 2013 | S_054-1-3-1-3 | A-set |
| 9 | Green Foil House 2013 | S_064-1-3-1-1 | A-set |
| 10 | Green Foil House 2013 | S_066-1-1-1-4 | A-set |
| 11 | Green Foil House 2013 | S_079-1-2-2-5 | A-set |
| 12 | Green Foil House 2013 | S_085-1-1-1 | A-set |
| 13 | Green Foil House 2013 | S_102-1-1-4 | A-set |
| 14 | Green Foil House 2013 | S_104-1-1-1-5 | A-set |
| 15 | Green Foil House 2013 | S_108-1-1-1 | A-set |
| 16 | Green Foil House 2013 | S_123-1-1-4 | A-set |
| 17 | Green Foil House 2013 | S_126-1-1-1 | A-set |
| 18 | Green Foil House 2013 | S_131-1-1-4 | A-set |
| 19 | Green Foil House 2013 | S_133-1-1-1 | A-set |
| 20 | Green Foil House 2013 | S_142-1-1-2 | A-set |
| 21 | Green Foil House 2013 | S_153-1-1-1-2 | A-set |
| 22 | Green Foil House 2013 | S_158-1-1-1-3 | A-set |
| 23 | Green Foil House 2013 | S_167-2-5 | A-set |
| 24 | Green Foil House 2013 | S_186-1-1-3 | A-set |
| 25 | Green Foil House 2013 | S_209-2-1 | A-set |
| 26 | Green Foil House 2013 | S_226-1-1-1-1 | A-set |
| 27 | Green Foil House 2013 | S_233-1-2-1-2 | A-set |
| 28 | Green Foil House 2013 | S_242-1-6 | A-set |
| 29 | Green Foil House 2013 | S_265-1-1-1-5 | A-set |
| 30 | Green Foil House 2013 | S_281-1-1-3 | A-set |
| 31 | Green Foil House 2013 | S_295-1-1-1-16 | A-set |
| 32 | Green Foil House 2013 | S_298-1-1-1-1 | A-set |
| 33 | Green Foil House 2013 | S_300-1-3-1-4 | A-set |
| 34 | Green Foil House 2013 | S_309-2-4 | A-set |
| 35 | Green Foil House 2013 | S_322-1-1-4 | A-set |
| 36 | Green Foil House 2013 | WAB-EP02-Fam/S1_157-1-2-4-3-1-1-10 | A-set |
| 37 | Green Foil House 2013 | S_028-1-3-1-1-5 | A-set |
| 38 | Green Foil House 2013 | S_065-1-1-1 | A-set |

Continuation of table

| No | Source | Pedigree | Sample Identification |
|---|---|---|---|
| 39 | Green Foil House 2013 | S_083-1-1-1-6 | A-set |
| 40 | Green Foil House 2013 | S_097-1-1-1-3 | A-set |
| 41 | Green Foil House 2013 | S_218-2-4 | A-set |
| 42 | Green Foil House 2013 | S_236-1-1-2 | A-set |
| 43 | Green Foil House 2013 | S_241-1-2 | A-set |
| 44 | Green Foil House 2013 | S_269-1-1 | A-set |
| 45 | Green Foil House 2013 | S_290-1-1-1 | A-set |
| 46 | Green Foil House 2013 | S_291-1-1-1 | A-set |
| 47 | Green Foil House 2013 | S_329-1-1-4 | A-set |
| 48 | Green Foil House 2013 | Hiverna/2-5 EP1-1-8-1-3-3-9 | Winter lines |
| 49 | Green Foil House 2013 | Limbo-7-2 | Spring lines |
| 50 | Green Foil House 2014 | S_4-1-6 | A-set |
| 51 | Green Foil House 2014 | S_010-1-1-1-1 | A-set |
| 52 | Green Foil House 2014 | S_012-1-1-1 | A-set |
| 53 | Green Foil House 2014 | S_013-2-2 | A-set |
| 54 | Green Foil House 2014 | S_022-1-1-1-1 | A-set |
| 55 | Green Foil House 2014 | S_043-1-1-2 | A-set |
| 56 | Green Foil House 2014 | S_048-3-6 | A-set |
| 57 | Green Foil House 2014 | S_069-2-9 | A-set |
| 58 | Green Foil House 2014 | S_076-1-1-2 | A-set |
| 59 | Green Foil House 2014 | S_077-1-1-3 | A-set |
| 60 | Green Foil House 2014 | S_081-1-24 | A-set |
| 61 | Green Foil House 2014 | S_082-2-2-1-1-4 | A-set |
| 62 | Green Foil House 2014 | S_084-2-7 | A-set |
| 63 | Green Foil House 2014 | S_100-1-1-1 | A-set |
| 64 | Green Foil House 2014 | S_115-1-1-2 | A-set |
| 65 | Green Foil House 2014 | S_116-1-1-1-3 | A-set |
| 66 | Green Foil House 2014 | S_125-1-2 | A-set |
| 67 | Green Foil House 2014 | S_129-1-2-4 | A-set |
| 68 | Green Foil House 2014 | S_150-1-2-1-3 | A-set |
| 69 | Green Foil House 2014 | S_160-1-1-1-2 | A-set |
| 70 | Green Foil House 2014 | S_161-2-1 | A-set |
| 71 | Green Foil House 2014 | S_162-1-1-2-4 | A-set |
| 72 | Green Foil House 2014 | S_163-1-1 | A-set |
| 73 | Green Foil House 2014 | S_165-1-1-2 | A-set |
| 74 | Green Foil House 2014 | S_166-1-1-2 | A-set |
| 75 | Green Foil House 2014 | S_168-1-1-3 | A-set |
| 76 | Green Foil House 2014 | S_169-1-1-4 | A-set |
| 77 | Green Foil House 2014 | S_177-1-1-2 | A-set |
| 78 | Green Foil House 2014 | S_182-1-1-3 | A-set |
| 79 | Green Foil House 2014 | S_189-1-1-2-3 | A-set |
| 80 | Green Foil House 2014 | S_190-1-1-5 | A-set |

Continuation of table

| No | Source | Pedigree | Sample Identification |
|---|---|---|---|
| 81 | Green Foil House 2014 | S_192-1-1-2 | A-set |
| 82 | Green Foil House 2014 | S_195-1-1-2 | A-set |
| 83 | Green Foil House 2014 | S_197-1-1-2-5-9 | A-set |
| 84 | Green Foil House 2014 | S_198-1-1-1-1 | A-set |
| 85 | Green Foil House 2014 | S_201-1-1-1-3 | A-set |
| 86 | Green Foil House 2014 | S_221-1-1-2-4-8 | A-set |
| 87 | Green Foil House 2014 | S_231-1-1-1-1 | A-set |
| 88 | Green Foil House 2014 | S_252-1-1-1-11 | A-set |
| 89 | Green Foil House 2014 | S_308-1-1-1-2 | A-set |
| 90 | Green Foil House 2014 | S_330-1-1-2 | A-set |
| 91 | Green Foil House 2014 | WAB98_98-4-1-1-1 | A-set |
| 92 | Green Foil House 2014 | Melodie-7-1 | Spring lines |
| 93 | Green Foil House 2014 | S_045-1-1-1 | A-set |
| 94 | Green Foil House 2014 | S_062-2-14 | A-set |
| 95 | Green Foil House 2014 | S_238-1-1-3 | A-set |
| 96 | Green Foil House 2014 | S_275-1-1-2 | A-set |
| 97 | Green Foil House 2014 | WAB_EP98_21-2-1 EP4-1-1-2-3 | A-set |
| 98 | Green Foil House 2014 | Hedin/2-3-3 | Spring lines |
| 99 | Green Foil House 2014 | Minica-5-5 | Spring lines |
| 100 | NPZ | C-141 | NPZ lines |
| 101 | NPZ | C-142 | NPZ lines |
| 102 | NPZ | C-143 | NPZ lines |
| 103 | NPZ | C-144 | NPZ lines |
| 104 | NPZ | C-145 | NPZ lines |
| 105 | NPZ | C-146 | NPZ lines |
| 106 | NPZ | C-147 | NPZ lines |
| 107 | NPZ | C-148 | NPZ lines |
| 108 | NPZ | C-149 | NPZ lines |
| 109 | NPZ | C-150 | NPZ lines |
| 110 | NPZ | C-151 | NPZ lines |
| 111 | NPZ | C-152 | NPZ lines |
| 112 | NPZ | C-153 | NPZ lines |
| 113 | NPZ | C-154 | NPZ lines |
| 114 | NPZ | C-155 | NPZ lines |
| 115 | NPZ | C-156 | NPZ lines |
| 116 | NPZ | C-157 | NPZ lines |
| 117 | NPZ | C-158 | NPZ lines |
| 118 | NPZ | C-159 | NPZ lines |
| 119 | NPZ | C-160 | NPZ lines |
| 120 | NPZ | C-161 | NPZ lines |
| 121 | NPZ | C-162 | NPZ lines |
| 122 | NPZ | C-163 | NPZ lines |

Continuation of table

| No | Source | Pedigree | Sample Identification |
|---|---|---|---|
| 123 | NPZ | C-164 | NPZ lines |
| 124 | NPZ | C-165 | NPZ lines |
| 125 | NPZ | C-166 | NPZ lines |
| 126 | NPZ | C-167 | NPZ lines |
| 127 | NPZ | C-168 | NPZ lines |
| 128 | NPZ | C-169 | NPZ lines |
| 129 | NPZ | C-170 | NPZ lines |
| 130 | NPZ | C-171 | NPZ lines |
| 131 | NPZ | C-172 | NPZ lines |
| 132 | NPZ | C-173 | NPZ lines |
| 133 | NPZ | C-174 | NPZ lines |
| 134 | NPZ | C-176 | NPZ lines |
| 135 | NPZ | C-177 | NPZ lines |
| 136 | NPZ | C-178 | NPZ lines |
| 137 | NPZ | C-179 | NPZ lines |
| 138 | NPZ | C-180 | NPZ lines |
| 139 | NPZ | C-181 | NPZ lines |
| 140 | NPZ | C-182 | NPZ lines |
| 141 | NPZ | C-183 | NPZ lines |
| 142 | NPZ | C-184 | NPZ lines |
| 143 | NPZ | C-185 | NPZ lines |
| 144 | NPZ | C-186 | NPZ lines |
| 145 | NPZ | C-187 | NPZ lines |
| 146 | NPZ | C-188 | NPZ lines |
| 147 | NPZ | C-189 | NPZ lines |
| 148 | NPZ | C-190 | NPZ lines |
| 149 | Isolation cage 2013 | [(MelodiexHiv)xHiverna]-2wn | Crosses |
| 150 | Isolation cage 2013 | [(MelodiexHiv)xHiverna]-3wn | Crosses |
| 151 | FOH 2006 | MelodiexHiverna EP 2 | Crosses |
| 152 | FOH 2006 | MelodiexHiverna EP 3 | Crosses |
| 153 | FOH 2006 | MelodiexHiverna EP 4 | Crosses |
| 154 | FOH 2006 | MelodiexHiverna EP 8 | Crosses |
| 155 | FOH 2006 | MelodiexHiverna EP 15 | Crosses |
| 156 | FOH 2006 | MelodiexHiverna EP 18 | Crosses |
| 157 | FOH 2006 | MelodiexHiverna EP 20 | Crosses |
| 158 | NPZ | C-181 | NPZ |
| 159 | NPZ | C-187 | NPZ |
| 160 | Green Foil House 2014 | S_163-1-1 | A-set |
| 161 | Green Foil House 2014 | S_168-1-1-3 | A-set |
| 162 | Green Foil House 2013 | S_218-2-4 | A-set |
| 163 | Green Foil House 2013 | S_241-1-2 | A-set |
| 164 | Green Foil House 2013 | Melodie-7-1 | Spring lines |

Continuation of table

| No | Source | Pedigree | Sample Identification |
|---|---|---|---|
| 165 | Green Foil House 2014 | Melodie-7-1 | Spring lines |
| 166 | FOH 2012 | Melodie-7-5 | Spring lines |
| 167 | Green Foil House 2015 | S_003-1-1-1-1-6 | A-set |
| 168 | Green Foil House 2015 | S_009-1-1-4-2 | A-set |
| 169 | Green Foil House 2015 | S_025-1-12-5 | A-set |
| 170 | Green Foil House 2015 | S_055-1-3-1-5-5 | A-set |
| 171 | Green Foil House 2015 | S_070-1-1-4-2 | A-set |
| 172 | Green Foil House 2015 | S_119-1-1-1-2-6 | A-set |
| 173 | Green Foil House 2015 | S_163-1-1-1 | A-set |
| 174 | Green Foil House 2015 | S_181-1-1-1-5 | A-set |
| 175 | Green Foil House 2015 | S_185-1-1-2-2 | A-set |
| 176 | Green Foil House 2015 | S_235-1-1-2-4-5 | A-set |
| 177 | Green Foil House 2015 | S_238-1-1-3-5 | A-set |
| 178 | Green Foil House 2015 | S_240-1-1-2-5-5 | A-set |
| 179 | Green Foil House 2015 | S_246-1-1-1-3-1 | A-set |
| 180 | Green Foil House 2015 | S_254-2-2-15-19-6 | A-set |
| 181 | Green Foil House 2015 | S_259-1-1-2-2 | A-set |
| 182 | Green Foil House 2015 | S_264-1-1-1-6-5 | A-set |
| 183 | Green Foil House 2015 | S_268-1-25-5 | A-set |
| 184 | Green Foil House 2015 | S_272-1-3-1-2-4 | A-set |
| 185 | Green Foil House 2015 | S_289-1-1-1-4-1 | A-set |
| 186 | Green Foil House 2015 | S_299-1-8-1 | A-set |
| 187 | Green Foil House 2015 | 29H(Ascochyta-resistent)-1-3-21-4 | Winter lines |
| 188 | Green Foil House 2015 | Melodie-7-1-4 | Spring lines |
| 189 | Green Foil House 2015 | S_008-1-1-1-7 | A-set |
| 190 | Green Foil House 2015 | S_029-1-1-1-3-7 | A-set |
| 191 | Green Foil House 2015 | S_035-1-1-2-3-7 | A-set |
| 192 | Green Foil House 2015 | S_038-1-1-1-3-8-7 | A-set |
| 193 | Green Foil House 2015 | S_067-2-3-7 | A-set |
| 194 | Green Foil House 2015 | S_120-1-1-1-7 | A-set |
| 195 | Green Foil House 2015 | S_145-1-2-4-7 | A-set |
| 196 | Green Foil House 2015 | S_147-1-1-3-7 | A-set |
| 197 | Green Foil House 2015 | S_151-1-1-1-1-7 | A-set |
| 198 | Green Foil House 2015 | S_172-1-1-1-1-7 | A-set |
| 199 | Green Foil House 2015 | S_174-1-1-3-7 | A-set |
| 200 | Green Foil House 2015 | S_175-1-1-9-7 | A-set |
| 201 | Green Foil House 2015 | S_176-1-1-2-7 | A-set |
| 202 | Green Foil House 2015 | S_196-1-1-1-2-7 | A-set |
| 203 | Green Foil House 2015 | S_199-1-3-1-5-7 | A-set |
| 204 | Green Foil House 2015 | S_202-1-1-3-7 | A-set |
| 205 | Green Foil House 2015 | S_232-1-1-1-16-6-7 | A-set |
| 206 | Green Foil House 2015 | S_243-1-1-3-7 | A-set |

Continuation of table

| No | Source | Pedigree | Sample Identification |
|---|---|---|---|
| 207 | Green Foil House 2015 | S_258-1-3-4-7 | A-set |
| 208 | Green Foil House 2015 | S_271-1-2-1-7-7 | A-set |
| 209 | Green Foil House 2015 | S_274-2-3-7 | A-set |
| 210 | Green Foil House 2015 | S_280-1-3-1-5-7 | A-set |
| 211 | Green Foil House 2015 | S_282-1-1-1-4-7 | A-set |
| 212 | Green Foil House 2015 | S_284-1-1-3-7 | A-set |
| 213 | Green Foil House 2015 | S_287-1-3-7 | A-set |
| 214 | Green Foil House 2015 | S_303-1-8-7 | A-set |
| 215 | Green Foil House 2015 | Côte d`Or x BPL…)-95-4-1-1-15-7 | Winter lines |
| 216 | Green Foil House 2015 | S_019-1-1-1-2-7 | A-set |
| 217 | Green Foil House 2013 | S_085-1-1-1 | A-set |
| 218 | Green Foil House 2013 | S_281-1-1-3 | A-set |
| 219 | Green Foil House 2014 | S_069-2-9 | A-set |
| 220 | Green Foil House 2014 | S_163-1-1 | A-set |
| 221 | Green Foil House 2013 | S_085-1-1-1 | A-set |
| 222 | Green Foil House 2013 | S_281-1-1-3 | A-set |
| 223 | Green Foil House 2014 | S_069-2-9 | A-set |
| 224 | Green Foil House 2014 | S_163-1-1 | A-set |
| 225 | 2010 | Disco | Low VC cultivar |
| 226 | 2014 | Devine | Low VC cultivar |
| 227 | 2012 | Mandoline | Low VC cultivar |
| 228 | Green Foil House 2015 | Tiffany | Low VC cultivar |
| 229 | Open Field 2013 | S_002-1-1 | A-set |
| 230 | Open Field 2013 | S_177-1-1 | A-set |
| 231 | Open Field 2013 | S_115-1-1 | A-set |
| 232 | Open Field 2013 | WAB-EP02-Fam/S1_159-1-2-4-1-1-3-1-3 | A-set |
| 233 | Open Field 2013 | S_070-1-1 | A-set |
| 234 | Open Field 2013 | S_196-1-1-1 | A-set |
| 235 | Open Field 2013 | S_160-1-1-1 | A-set |
| 236 | Green Foil House 2014 | Bulldog 1-4-3-5-1 | Founder lines |
| 237 | Green Foil House 2014 | L79/79/1-4-1-1-3-2 | Founder lines |
| 238 | Green Foil House 2014 | L977/88/S1wn-5-1 | Founder lines |
| 239 | Green Foil House 2014 | L979/S1/1/1sn-10-1-1-4-1 | Founder lines |
| 240 | Green Foil House 2014 | Bourdon/1-5-1-1-1-2 | Founder lines |
| 241 | Green Foil House 2014 | Arrisot/1-1-1-1-4-1 | Founder lines |
| 242 | Green Foil House 2014 | Banner/1-1-1-1-4-1 | Founder lines |
| 243 | Green Foil House 2014 | Wibo/1-1-1 | Founder lines |
| 244 | Green Foil House 2014 | Webo/1-1-1 -10-12 | Founder lines |
| 245 | Green Foil House 2014 | Hiverna/1-1-2 EP3-2-4 | Founder lines |
| 246 | Green Foil House 2014 | CôteD`Or/1-1-3-1-2-1-1-2-2 -3-22 | Founder lines |

# Appendix 2

**Vicine and convicine content of the 200 lines, including the 189 lines of A-set, as used for GWAS analyses (NIRS-predicted with ult.eqa)**

| No | Pedigree | Vicine-convicine content | Sample identification |
|---|---|---|---|
| 1 | S_002-1-1-3-1 | 0.5738 | A-set |
| 2 | S_003-1-1-1-1-1 | 0.6515 | A-set |
| 3 | S_4-1-6-1 | 0.6726 | A-set |
| 4 | S_005-1-1-1-1 | 0.7205 | A-set |
| 5 | S_008-1-1-1-1 | 0.5525 | A-set |
| 6 | S_009-1-1-4-1 | 0.7174 | A-set |
| 7 | S_010-1-1-1-1-1 | 0.6989 | A-set |
| 8 | S_012-1-1-1-1 | 0.7119 | A-set |
| 9 | S_013-2-2-1 | 0.5703 | A-set |
| 10 | S_015-1-1-1-2-1 | 0.6156 | A-set |
| 11 | S_016-1-1-3-1-1 | 0.7292 | A-set |
| 12 | S_019-1-1-1-2-1 | 0.7044 | A-set |
| 13 | S_020-1-2-6-1 | 0.7799 | A-set |
| 14 | S_021-2-1-1 | 0.5174 | A-set |
| 15 | S_022-1-1-1-1-1 | 0.6487 | A-set |
| 16 | S_025-1-12-1 | 0.5713 | A-set |
| 17 | S_027-1-1-1-1 | 0.6386 | A-set |
| 18 | S_028-1-3-1-1-5-1 | 0.7134 | A-set |
| 19 | S_029-1-1-1-3-1 | 0.6608 | A-set |
| 20 | S_030-2-2-1 | 0.5975 | A-set |
| 21 | S_033-2-12-1 | 0.6875 | A-set |
| 22 | S_034-1-2 weisse blüte,wn-1-1 | 0.4691 | A-set |
| 23 | S_035-1-1-2-3-1 | 0.5179 | A-set |
| 24 | S_036-1-2-5-1 | 0.6887 | A-set |
| 25 | S_038-1-1-1-3-8-1 | 0.7606 | A-set |
| 26 | S_039-1-1-4-1 | 0.7746 | A-set |
| 27 | S_040-1-1-1-2-1 | 0.8395 | A-set |
| 28 | S_043-1-1-2-1 | 0.6931 | A-set |
| 29 | S_045-1-1-1-1 | 0.8071 | A-set |
| 30 | S_046-1-1-1-5-1 | 0.6954 | A-set |
| 31 | S_048-3-6-1 | 0.6294 | A-set |
| 32 | S_050-2-12-1 | 0.7955 | A-set |
| 33 | S_052-1-1-1-3-1 | 0.683 | A-set |
| 34 | S_054-1-3-1-3-1 | 0.7637 | A-set |
| 35 | S_055-1-3-1-5-1 | 0.662 | A-set |
| 36 | S_059-1-2-2-4-1 | 0.5999 | A-set |

Continuation of table

| No | Pedigree | Vicine-convicine content | Sample identification |
|---|---|---|---|
| 37 | S_060-1-7-1 | 0.7655 | A-set |
| 38 | S_062-2-14-1 | 0.7754 | A-set |
| 39 | S_064-1-3-1-1-1 | 0.6433 | A-set |
| 40 | S_065-1-1-1-1 | 0.756 | A-set |
| 41 | S_066-1-1-1-4-1 | 0.6814 | A-set |
| 42 | S_067-2-3-1 | 0.6191 | A-set |
| 43 | S_069-2-9-1 | 0.6345 | A-set |
| 44 | S_070-1-1-4-1 | 0.6862 | A-set |
| 45 | S_072-1-1-4-1 | 0.853 | A-set |
| 46 | S_076-1-1-2-1 | 0.5899 | A-set |
| 47 | S_077-1-1-3-1 | 0.7388 | A-set |
| 48 | S_079-1-2-2-5-1 | 0.8434 | A-set |
| 49 | S_081-1-24-1 | 0.6379 | A-set |
| 50 | S_082-2-2-1-1-4-1 | 0.731 | A-set |
| 51 | S_083-1-1-1-6-1 | 0.6571 | A-set |
| 52 | S_084-2-7-1 | 0.5424 | A-set |
| 53 | S_085-1-1-1-1 | 0.7545 | A-set |
| 54 | S_093-1-1-1-1,3-1 | 0.6206 | A-set |
| 55 | S_097-1-1-1-3-1 | 0.5235 | A-set |
| 56 | S_100-1-1-1-1 | 0.7324 | A-set |
| 57 | S_102-1-1-4-1 | 0.5727 | A-set |
| 58 | S_104-1-1-1-5-1 | 0.6926 | A-set |
| 59 | S_106-1-1-2-1-1 | 0.6915 | A-set |
| 60 | S_108-1-1-1-1 | 0.7607 | A-set |
| 61 | S_111-1-1-1-2-1 | 0.6562 | A-set |
| 62 | S_115-1-1-2-1 | 0.6124 | A-set |
| 63 | S_116-1-1-1-3-1 | 0.6295 | A-set |
| 64 | S_119-1-1-1-2-1 | 0.5791 | A-set |
| 65 | S_120-1-1-1-1 | 0.7445 | A-set |
| 66 | S_122-1-1-4-2-6-9-1 | 0.7334 | A-set |
| 67 | S_123-1-1-4-1 | 0.6317 | A-set |
| 68 | S_125-1-2-1 | 0.6158 | A-set |
| 69 | S_126-1-1-1-1 | 0.7385 | A-set |
| 70 | S_129-1-2-4-1 | 0.7882 | A-set |
| 71 | S_131-1-1-4-1 | 0.725 | A-set |
| 72 | S_132-1-1-3-1 | 0.8055 | A-set |
| 73 | S_133-1-1-1-1 | 0.6574 | A-set |
| 74 | S_134-1-2-1-2-1 | 0.749 | A-set |
| 75 | S_142-1-1-2-1 | 0.6661 | A-set |
| 76 | S_145-1-2-4-1 | 0.6452 | A-set |
| 77 | S_147-1-1-3-1 | 0.6337 | A-set |

Continuation of table

| No | Pedigree | Vicine-convicine content | Sample identification |
|---|---|---|---|
| 78 | S_150-1-2-1-3-1 | 0.7194 | A-set |
| 79 | S_151-1-1-1-1-1 | 0.6814 | A-set |
| 80 | S_153-1-1-1-2-1 | 0.604 | A-set |
| 81 | S_158-1-1-1-3-1 | 0.6345 | A-set |
| 82 | S_160-1-1-1-2-1 | 0.8039 | A-set |
| 83 | S_161-2-1-1 | 0.5874 | A-set |
| 84 | S_162-1-1-2-4-1 | 0.6605 | A-set |
| 85 | S_163-1-1-1 | 0.5986 | A-set |
| 86 | S_165-1-1-2-1 | 0.6801 | A-set |
| 87 | S_166-1-1-2-1 | 0.6116 | A-set |
| 88 | S_167-2-5-1 | 0.7114 | A-set |
| 89 | S_168-1-1-3-1 | 0.7294 | A-set |
| 90 | S_169-1-1-4-1 | 0.6935 | A-set |
| 91 | S_170-1-1-2-1 | 0.6653 | A-set |
| 92 | S_172-1-1-1-1-1 | 0.5202 | A-set |
| 93 | S_173-1-1-3-1 | 0.6508 | A-set |
| 94 | S_174-1-1-3-1 | 0.6286 | A-set |
| 95 | S_175-1-1-9-1 | 0.6879 | A-set |
| 96 | S_176-1-1-2-1 | 0.6119 | A-set |
| 97 | S_177-1-1-2-1 | 0.5264 | A-set |
| 98 | S_181-1-1-1-1 | 0.5587 | A-set |
| 99 | S_182-1-1-3-1 | 0.6555 | A-set |
| 100 | S_185-1-1-2-1 | 0.5927 | A-set |
| 101 | S_186-1-1-3-1 | 0.8377 | A-set |
| 102 | S_189-1-1-2-3-1 | 0.7363 | A-set |
| 103 | S_190-1-1-5-1 | 0.7612 | A-set |
| 104 | S_191-1-3-5-5-1 | 0.7234 | A-set |
| 105 | S_192-1-1-2-1 | 0.538 | A-set |
| 106 | S_194-1-1-2-1 | 0.6312 | A-set |
| 107 | S_195-1-1-2-1 | 0.5978 | A-set |
| 108 | S_196-1-1-1-2-1 | 0.5828 | A-set |
| 109 | S_197-1-1-2-5-9-1 | 0.8108 | A-set |
| 110 | S_198-1-1-1-1-1 | 0.7136 | A-set |
| 111 | S_199-1-3-1-5-1 | 0.7655 | A-set |
| 112 | S_201-1-1-1-3-1 | 0.6377 | A-set |
| 113 | S_202-1-1-3-1 | 0.6363 | A-set |
| 114 | S_209-2-1-1 | 0.6874 | A-set |
| 115 | S_210-1-1-1-3-1 | 0.7007 | A-set |
| 116 | S_213-1-1-1-2-1-1 | 0.6259 | A-set |
| 117 | S_217-1-1-2-4-1 | 0.6576 | A-set |
| 118 | S_218-2-4-1 | 0.7707 | A-set |

Continuation of table

| No | Pedigree | Vicine-convicine content | Sample identification |
|---|---|---|---|
| 119 | S_220-1-1-6-1 | 0.7751 | A-set |
| 120 | S_221-1-1-2-4-8-1 | 0.6723 | A-set |
| 121 | S_226-1-1-1-1-1 | 0.7482 | A-set |
| 122 | S_227-1-1-1-8-8-1 | 0.6945 | A-set |
| 123 | S_231-1-1-1-1-1 | 0.7173 | A-set |
| 124 | S_232-1-1-1-16-6-1 | 0.7089 | A-set |
| 125 | S_233-1-2-1-2-1 | 0.586 | A-set |
| 126 | S_235-1-1-2-4-1 | 0.7979 | A-set |
| 127 | S_236-1-1-2-1 | 0.5526 | A-set |
| 128 | S_238-1-1-3-1 | 0.6055 | A-set |
| 129 | S_240-1-1-2-5-1 | 0.5539 | A-set |
| 130 | S_241-1-2-1 | 0.6652 | A-set |
| 131 | S_242-1-6-1 | 0.8022 | A-set |
| 132 | S_243-1-1-3-1 | 0.6427 | A-set |
| 133 | S_245-1-3-1 | 0.6197 | A-set |
| 134 | S_246-1-1-1-3-1 | 0.5759 | A-set |
| 135 | S_249-1-1-2-4-1 | 0.5891 | A-set |
| 136 | S_252-1-1-1-11-1 | 0.8553 | A-set |
| 137 | S_253-1-1-4-14-1 | 0.6888 | A-set |
| 138 | S_254-2-2-15-19-1 | 0.7351 | A-set |
| 139 | S_258-1-3-4-1 | 0.6295 | A-set |
| 140 | S_259-1-1-2-1 | 0.6277 | A-set |
| 141 | S_264-1-1-1-6-1 | 0.6313 | A-set |
| 142 | S_265-1-1-1-5-1 | 0.6509 | A-set |
| 143 | S_267-2-5-1 | 0.669 | A-set |
| 144 | S_268-1-25-1 | 0.6974 | A-set |
| 145 | S_269-1-1-1 | 0.6852 | A-set |
| 146 | S_271-1-2-1-7-1 | 0.6318 | A-set |
| 147 | S_272-1-3-1-2-1 | 0.5447 | A-set |
| 148 | S_274-2-3-1 | 0.5474 | A-set |
| 149 | S_275-1-1-2-1 | 0.714 | A-set |
| 150 | S_277-1-1-4-1 | 0.7548 | A-set |
| 151 | S_279-2-1-2-1 | 0.7728 | A-set |
| 152 | S_280-1-3-1-5-1 | 0.7476 | A-set |
| 153 | S_281-1-1-3-1 | 0.8788 | A-set |
| 154 | S_282-1-1-1-4-1 | 0.7276 | A-set |
| 155 | S_284-1-1-3-1 | 0.7696 | A-set |
| 156 | S_285-2-1-1 | 0.6423 | A-set |
| 157 | S_286-1-1-4-1 | 0.5505 | A-set |
| 158 | S_287-1-3-1 | 0.6946 | A-set |
| 159 | S_289-1-1-1-4-1 | 0.6728 | A-set |

Continuation of table

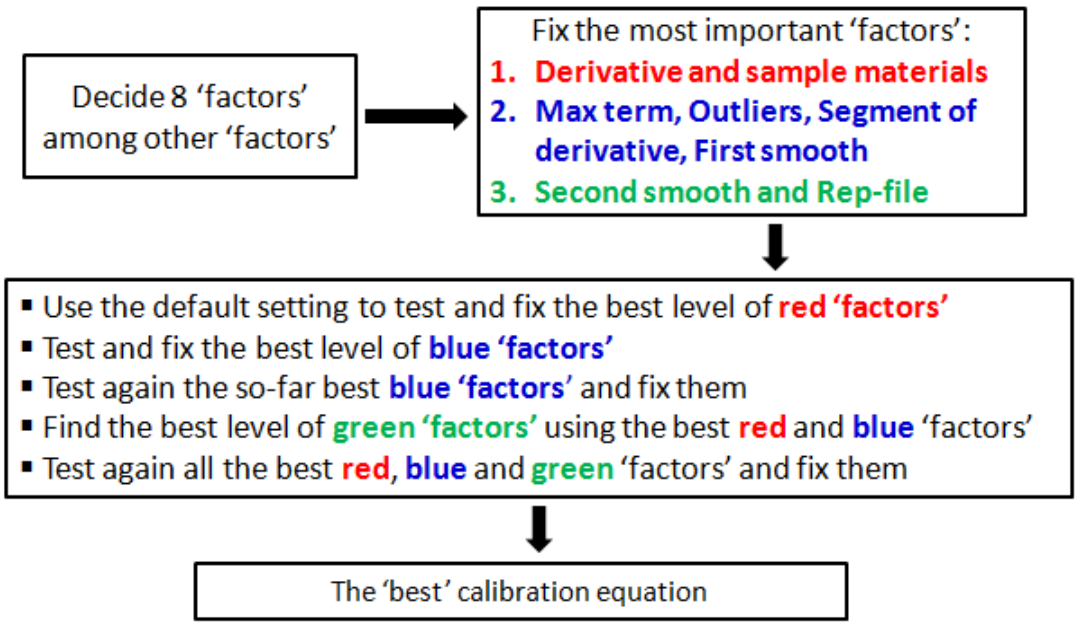
| No | Pedigree | Vicine-convicine content | Sample identification |
|---|---|---|---|
| 160 | S_290-1-1-1-1 | 0.504 | A-set |
| 161 | S_291-1-1-1-1 | 0.6774 | A-set |
| 162 | S_295-1-1-1-16-1 | 0.5601 | A-set |
| 163 | S_298-1-1-1-1-1 | 0.6425 | A-set |
| 164 | S_299-1-8-1 | 0.6626 | A-set |
| 165 | S_300-1-3-1-4-1 | 0.7916 | A-set |
| 166 | S_301-1-1-1-5-1 | 0.8872 | A-set |
| 167 | S_302-1-2-1-2-1 | 0.6102 | A-set |
| 168 | S_303-1-8-1 | 0.7584 | A-set |
| 169 | S_304-1-3-1-1-1 | 0.5161 | A-set |
| 170 | S_307-1-3-1 | 0.7938 | A-set |
| 171 | S_308-1-1-1-2-1 | 0.6518 | A-set |
| 172 | S_309-2-4-1 | 0.6949 | A-set |
| 173 | S_310-1-2-1-1-1 | 0.6775 | A-set |
| 174 | S_312-1-5-1 | 0.8154 | A-set |
| 175 | S_314-1-1-2-1 | 0.6935 | A-set |
| 176 | S_315-1-4-1 | 0.6394 | A-set |
| 177 | S_319-1-1-2-3-1 | 0.6624 | A-set |
| 178 | S_322-1-1-4-1 | 0.8119 | A-set |
| 179 | S_326-1-1-4-19-1 | 0.677 | A-set |
| 180 | S_328-1-1-1-2-1 | 0.8187 | A-set |
| 181 | S_329-1-1-4-1 | 0.774 | A-set |
| 182 | S_330-1-1-2-1 | 0.7773 | A-set |
| 183 | S_331-1-1-5-1 | 0.7537 | A-set |
| 184 | Cote d`Or/1-1 x BPL4628/1521.1)-18-3-1-1-4-3-1 | 0.8151 | Winter lines |
| 185 | Côte d`Or x BPL…)-95-4-1-1-15-1 | 0.6784 | Winter lines |
| 186 | CôteD`Or/1-1-3-1-2-1-1-2-2 -3-22-1 | 0.7935 | Winter lines |
| 187 | Hiverna/2-5 EP1-1-8-1-3-3-9-1 | 0.7453 | Winter lines |
| 188 | Hiverna/1-1-2 EP3-2-4-1 | 0.6147 | Winter lines |
| 189 | WAB_EP98_21-2-1 EP4-1-1-2-3-1 | 0.6017 | A-set |
| 190 | WAB_EP98_98-3-1 EP4-1-2-7-1 | 0.7743 | A-set |
| 191 | WAB98_98-4-1-1-1-1 | 0.6946 | A-set |
| 192 | WAB_EP98_267-11-1 -7-1 | 0.8435 | A-set |
| 193 | WAB-EP02-Fam/S1_157-1-2-4-3-1-1-10-1 | 0.844 | A-set |
| 194 | WAB-EP02-Fam/S1_159-1-2-4-1-1-3-1-3-6-1 | 0.6848 | A-set |
| 195 | Webo/1-1-1 -10-12-1 | 0.5504 | Winter lines |
| 196 | 29H(Ascochyta-resistent)-1-3-21-1 | 0.6969 | Winter lines |
| 197 | Limbo-7-2-1 | 0.5958 | Spring lines |
| 198 | Melodie-7-1-1 | 0.1159 | Spring lines |
| 199 | Hedin/2-3-3-1 | 0.5749 | Spring lines |
| 200 | Minica-5-5-1 | 0.4445 | Spring lines |

# Appendix 3

**Strategic filtering to find the 'best' calibration**

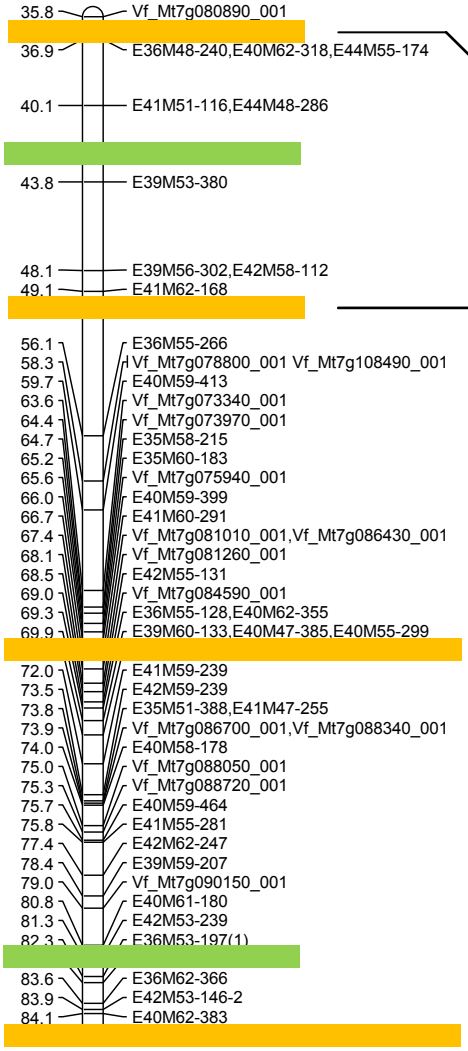
In order to obtain the 'best' calibration equation of vicine convicine in winISI II Project Manager v1.5 software, several steps and factors were considered.

- Determine samples as Calibration-set. 251 samples with HPLC and spectral data were available. After filtering outliers, 246 samples were remained.
- Decide which factors included in the developing of calibration equation. Several factors were studied, i.e. Derivatives, Sample materials, Max term, Outliers, Segment of Derivative (Gap), First smooth, Second smooth and Rep-file.
- Establish the rank of importance. Each factor was tested using the level of software default and other levels, evaluation was carried out according to coefficient of determination of calibration equation and coefficient of calibration of semi-external validation. The most powerful factor was decided by comparing the range of best coefficient of determination and the worst coefficient of determination in each factor, the biggest range was the first rank and so on. The resulting ranking was:
    1. Derivative and Sample materials.
    2. Max term, Outliers, Gap and First smooth.
    3. Second smooth and Repeatability file.
- Determine optimum factors level according the Table 12, the details are in the diagram below. Evaluation has been carried out in each level of factor to find the best level of each factors.
- The 'best' calibration equation was obtained.

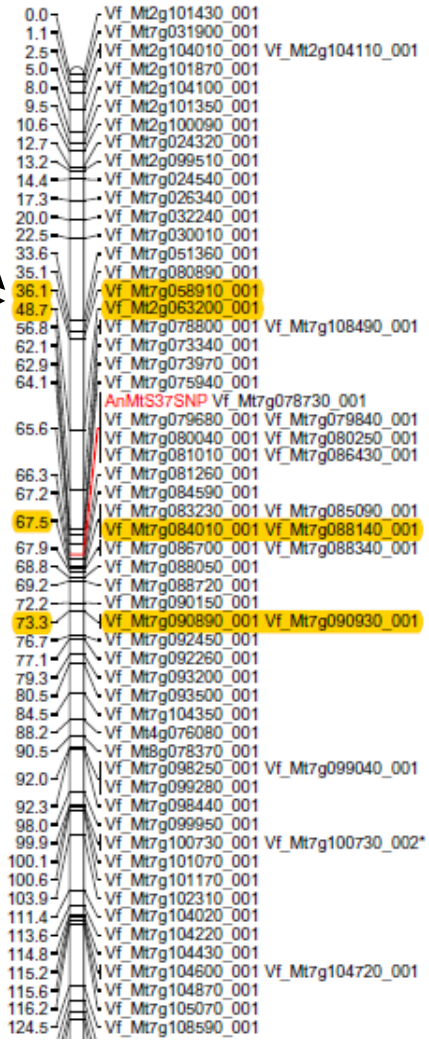

Decide 8 'factors' among other 'factors'
Fix the most important 'factors':
1. Derivative and sample materials
2. Max term, Outliers, Segment of derivative, First smooth
3. Second smooth and Rep-file
Use the default setting to test and fix the best level of red 'factors'
Test and fix the best level of blue 'factors'
Test again the so-far best blue 'factors' and fix them
Find the best level of green 'factors' using the best red and blue 'factors'
Test again all the best red, blue and green 'factors' and fix them
The 'best' calibration equation

# Appendix 4

**Estimation of localization of E40M59-387 of the merge linkage map in the consensus map of *Vicia faba***




Merge linkage map of the three linkage maps




Chromosome 5 of *Vicia faba* (Webb et al., 2016)

# Acknowledgement

All praises to Allah for giving me the opportunity to complete this PhD work.

I am very grateful to Prof. Dr. Wolfgang Link for his very valuable supervision and excellent guidance throughout the study period. I thank Wolfgang for the very pleasant conversations we had and his advices and help throughout these four years.

My special thanks are addressed to Dr. Christian Möllers for his willingness to take the role of co-referee and to Dr. Catrin Westphal for kindly consenting to be a member of the examination committee.

I highly appreciate the Ministry of Research, Technology and Higher Education of Indonesia for the financial support.

I would like to thank to all staff members in the Division of Plant Breeding, Georg-August-Universität Göttingen; Prof. Dr. H. Becker, PD Dr. Wolfgang Ecke, Dr. Christian Möllers, Dr. Antje Schierholt, Dr. Bernd Horneburg and Dr. Sabine von Witzke-Ehbrecht for their kindness, help and the valuable discussions.

I really thank to Regina Martsch for the excellent work in the preparation of experiments and for her kindness and the warm relationship we have.

I am thankful to all of my colleagues for a lot of nice moments as well as their kindness, support and valuable help.

I will never forget the kindness and generousity of the Indonesian community members in Göttingen who have given me their full support, valuable advice and help.

I coudnot find suitable word to express grateful feeling to my bestfriend and lovely husband Emiel Afrans Titho for his never ending support, valuable help and patience for the achievement during PhD study. I am very grateful withmy three lovelydaughters, Aisyah, Aliya and Afiqah, for colouring my life day by day.

Finally, my warmest gratitudeto my parents for love, support and pray for the achievement of the PhD work.