Seed storage proteins of faba bean (Vicia faba L): current status and prospects for genetic improvement

Article

Accepted Version

Warsame, A. O., O'Sullivan, D. M. and Tosi, P. (2018) Seed storage proteins of faba bean (Vicia faba L): current status and prospects for genetic improvement. Journal of Agricultural and Food Chemistry, 66 (48). pp. 12617-12626. ISSN 0021-8561 doi: https://doi.org/10.1021/acs.jafc.8b04992 Available at http://centaur.reading.ac.uk/80856/

To link to this article DOI: http://dx.doi.org/10.1021/acs.jafc.8b04992

Publisher: American Chemical Society

Seed storage proteins of faba bean (*Vicia faba*): current status and prospects for genetic improvement

Ahmed O. Warsame*, Donal M. O'Sullivan and Paola Tosi

School of Agriculture, Policy and Development, University of Reading, Reading, RG6 6AR, United Kingdom

Correspondence: Ahmed Omar Warsame

Phone: +44(0)1183785473

Email: A.OmarWarsame@pgr.reading.ac.uk

ABSTRACT

Faba bean (*Vicia faba*, L.) is one of the foremost candidate crops for simultaneously increasing both sustainability and global supply of plant protein. Its seeds contain about 27% proteins of which more than 80% -consist of globulin storage proteins (vicilin and legumin). For optimum utilization for human and animal nutrition, both protein content and quality have to be improved. Though initial investigations on the hereditability of these traits indicated possibility for genetic improvement, little has been achieved so far partly due to lack of genetic information coupled with the complex relationship between protein content and grain yield. This review reports on the current knowledge on faba bean seed storage proteins; their structure, composition and genetic control and highlights key areas for further improvement of the content and composition of faba bean seed storage proteins on the basis of recent advances in faba bean genome knowledge and genetic tools.

Key words: *Vicia faba*; sustainability; storage proteins; legumin and vicilin; genetic improvement

INTRODUCTION

Faba bean production and utilization

Nearly 60% of the global protein supply for human nutrition is sourced from plants [1-2] and about one third of this originates from grain legumes of the Fabaceae family [3]. Besides their nutritional significance, legume crops ability to fix atmospheric nitrogen via rhizobial symbiosis makes them invaluable components of sustainable crop production systems [4]. Faba bean (*Vicia faba*, hereafter *Vf*), also known as fava bean, broad bean, horse bean or field bean [5] is one of the world's oldest legume crops, its cultivation dating back to the 10th millennium BC [6-7]. From its origin in the Near East, *Vf* spread to the rest of the globe [7] and is currently cultivated in nearly 70 countries over the world (Figure 1A), occupying about 2.2 million ha that produce nearly 4 million tons annually [8]. China is the leading *Vf* producer with 36% of the global output, followed by Ethiopia (20%), Australia (10%) and United Kingdom (6%) (Figure 1B). The wide geographical distribution of *Vf* implies not only a great adaptation to diverse environmental conditions, but also suitability for diverse end uses and trade across continents.

Seeds of *Vf* contain on average about 27% protein [9-11] which provides affordable nutrition for millions of people around the world, hence its denomination as "the poor man's meat". While *Vf* has been traditionally utilized as dry grain for human consumption in developing countries, there is growing interest from food industries in developed countries to exploit its protein for the production of protein-rich vegan/vegetarian snacks, [12], the fortification of cereal-based food products such as bread and pasta without significantly affecting their structural and sensory quality [13-14] , or even the production of wholly *Vf*-based bread and pasta products [15]. *Vf* also represents as significant resource for agro-ecosystem sustainability and provision of feed for the growing global livestock inventory. Overall, the global production area for *Vf* has been increasing in the last two decades (Figure S1A) and a recent meta-analysis of yield data from 39 legume species indicated that, in the right environment, *Vf* can be the highest yielding grain legume [16]. *Vf* also has a high capacity for biological nitrogen fixation, to the extent that the amount of N fixed by *Vf* alone was estimated to be comparable to that of soybean and pea combined [17]. For further details on the role of *Vf* on sustainable cropping systems, readers are referred to Jensen, et al. [18,] Köpke and Nemecek [19].On the other hand, *Vf* is yet to be fully exploited as a feedstock for animal production due to presence of some anti-nutrients which limit its optimal inclusion ratio [20-22]. Removal of these anti-nutrients through the development of new low anti-nutrient cultivars or using simple processing techniques like fermentation [13-14] would make this crop a valuable protein resource for the animal production industry.

Faba bean as a sustainable global protein resource

One of the greatest challenges in the 21st century is feeding the growing world population which it has been estimated may necessitate a 70% increase in food production by 2050 [4]. More than 30% of this increase has to be made via the production of protein-rich foods [1] to meet the expected rise in demands due to population growth, increased urbanization and improved incomes in many parts of the world [1, 23-25]. Protein is a critical nutrient required in large quantity by humans (~ 50 g protein per adult per day) to maintain normal body function [26] However,t about one-third of the world population, mainly in Asia, Africa and Latin America, suffers from inadequate intake of proteins, vitamins and minerals [27]. On the other hand, in higher income countries, where daily animal-based protein intake is already high [1, 25], continued provision of nutritious feeds for the intensive animal production industry will pose a major challenge in the future. In particular, the livestock production sector in soybean non-producing countries will be burdened by the high price of imported soybean and soybean meal. For instance, EU countries have huge deficit in protein-rich feeds with nearly 70% being imported [28]. *Vf* is well-adapted to European climates, as testified by the high yields recorded in this continent for this legume (Figure S1B), and it therefore has the potential to contribute to bridging the gap in animal feed self-sufficiency as part of the EU's policies to increase protein production from locally grown crops [28]. *Vf* is also a candidate crop to meet the protein demands of an emerging consumer category, particularly in developed economies, who are opting for animal meat free life style. For example, Statista [29] reported that 13% of European citizens would consider avoiding red meat while nearly 50% of the respondents in another study were willing to replace meat with other sources of proteins [30].

Considering the projected impact of climate change on global crop production, meeting the nutritional requirements of the current and future generations would necessitate increased exploitation of the global genetic and natural resources for protein production systems based increasingly on biological nitrogen fixation. In this context, the fact that *Vf* is a high-yielding protein-rich crop with superior N fixation capability makes it a candidate crop for supporting increased protein production while maintaining sustainability of crop production systems.

Nutritional constraints to *Vf* utilization

The main determinants of *Vf* utilization for human food and animal feed include: (i) protein concentration, (ii) protein quality, defined mainly by the content of sulfur-containing amino acids (S-AA) cysteine and methionine, and (iii) concentration of antinutrients in the seeds [5]. Protein concentration of *Vf*, although it can vary greatly between different genotypes (19-39 %) [31-33], is one of the highest among legumes. However, commercial varieties on the UK market contain about 27% protein on average, which is still far less than the protein density of soya meal, and so, further improvements in protein content is required in order for faba bean to displace imported soya in animal feed. The proportion of S-AA in the protein is another crucial quality criterion, particularly in animal feeding. However, like most plant proteins, *Vf* is poor in certain essential amino acids, namely methionine, cysteine and tryptophan [5]. Though relatively narrow, the genetic variation for the S-AA reported in *Vf* indicates possibility of improving its nutritional quality. So far, the major breeding objectives for *Vf*, have been the reduction or removal of vicine and convicine (V-C) and tannins: V-C causes favism in humans and have deleterious effects on animals [34-35] while tannins lower protein digestibility [10]. Although these compounds can be removed by processing techniques [36-37], the most effective approach is probably removing them by breeding. This is now feasible with the availability of molecular markers closely linked to the V-C locus [38] and zero tannin gene (zt-1) [39]. Furthermore, the reduction of less significant antinutrients such as trypsin inhibitors, lectins and phytates would improve the nutritional value of *Vf* based feed products.

Understanding the genetic basis of the above limiting factors is a prerequisite for the development of new cultivars with desirable agronomic and nutritional attributes. Unfortunately, while scientific interest in *Vf* was high during 1970's and 1980's , when it became the model species for studying plant cytogenetics and stomatal regulation, *Vf* can now be considered an orphan crop [40]. For instance, less than 5% of the publications on legumes in the years 2004–2013 referred to *Vf* [9]. This is further reflected by the scarcity of information on the genetics of many important traits including protein content and quality, for which not a single QTL (Quantitative Trait Loci) has been reported, compared to 160 QTLs from 35 independent studies on soybean protein content [41]. In this context, in order for future work to proceed on a sound basis, we felt there was a need to marry the earlier biochemical literature, where the main species of storage protein were separated and classified, with the later genomic literature, which is replete with unannotated storage protein sequences and implicit map locations. The remainder of this review is devoted to a synthesis of the literature on *Vf* seed storage proteins, covering sequence, structure, composition and genetic basis for their synthesis and accumulation as well as taking a forward look at how this synthesis might be exploited in future research aiming to increase protein content and/or quality.

SEED STORAGE PROTEINS OF FABA BEAN

The major storage proteins of legumes are mainly enzymatically inactive proteins deposited in seed cotyledons which provide nutrients needed for seed germination and seedling growth and development [42-43]. Certain seed proteins in legumes including albumins and trypsin inhibitors, however, have been identified as antinutritional or allergenic agents and therefore are targeted for removal in breeding programs [44]. Seed storage proteins are classified according to the system developed by TB Osborne which is based on their solubility in different solvents [43]. Albumins and globulins are the major storage proteins of legumes and are soluble in water and saline solutions, respectively. Globulins alone constitute more than 80% of total seed protein in *Vf* [45] and they are further classified based on their sedimentation coefficients into vicilin-type (7S) and legumin-type (11S) [43]. Both globulin proteins are found in nearly all legumes, but their denotations vary across species. For instance, globulins of *Vf* and pea are often referred as vicilin/convicilin and legumin while they are denoted as conglycinin and glycinin in soybean, β and α conglutins in lupin, while phaseolin (a vicilin-like protein) is the only major globulin in common beans. Furthermore, decades of research on legume storage proteins have produced a sufficient database of annotated SDS-PAGE images of various species which facilitates faster identification of major globulin bands without the need for conducting tedious immunoblotting or HPLC procedures. When extracted under reducing conditions, the salt soluble fraction of legume seed proteins can be separated on SDS-PAGE into distinct bands which, based on their molecular weights, are identified as: convicilin (Mr >~60 kDa), vicilin (Mr=~46-55 kDa) and two major legumin subunits (Mr= ~38-40 and 23 kDa) (Table 1)

Legumin and vicilin share notable sequence and structural homology and are believed to originate from a common ancestral gene [46]. Mature legumin is hexameric with a mass of about 330 kDa [45] and is composed of two trimeric subunits (legumin A and B) while vicilin is a trimeric protein formed by the assembly of three monomers (Figure 2). In contrast to legumin, vicilin lacks cysteine and is usually glycosylated in its C-terminus [46]. These structural variations may result in differences in the physiochemical properties of seed storage proteins which in turn determine their nutritional value and utilization. For instance, legumin and vicilin differ in their thermal properties [47-48], affinity to bind flavor compounds under varying pH conditions [49]and emulsifying ability [48]. Therefore, from a breeding point of view, legumin/vicilin ratio could be manipulated to meet certain end-user requirements for protein functionality.

Structure and composition of *Vf* globulins

Legumin constitutes more than 50% of *Vf* globulins [45]. It is a hexameric protein with two major subunits - the α and β chains - which are connected by disulphide bonds. Under reducing conditions, these subunits form two bands of molecular weights of about 40 and 24 kDa, respectively (Figure 3). These subunits are also referred to as acidic and basic subunits or simply legumin A and B. Polypeptides of both legumins are highly homologous but notably distinguishable by the presence of more methionine residues in the peptide sequences of legumin A subunits [50]. *Vf* legumin A subunits appear to be more variable and show polymorphic bands between genotypes[51] as is also the case with Medicago legumin A [52]. On the other hand, vicilin-type proteins of *Vf* are trimeric [45] consisting predominantly of subunits of ~50 kDa while bands of ~66 kDa are referred as convicilin [42, 51]. The classification of 7S proteins into vicilin and convicilin was first coined in pea and has been accepted in many legumes including *Vf* (Table 1). Nonetheless, further investigation into their possible structural and functional differences have concluded that convicilin may be regarded as subunit of vicilin [53]. Such a denotation exists in soybean whereby subunits of 7S protein are categorized into α′ (~76 kDa), α (~72 kDa), and β (~53) kDa [54-55].

Regarding amino acid composition, nearly 50% of *Vf* seed protein is accounted for by just a few non-essential amino acids such as glutamic acid, aspartic acid, arginine and leucine while it is low in essential amino acids particularly S-AA (Figure 4). The concentration of S-AA is a critical determinant of the nutritional value of plant proteins destined for human consumption and animal feeding. In humans, dependence on poor quality proteins can result in reduced immunity and underdeveloped mental and physical capacity among young children [56]. Also, animal feeds deficient in critical amino acids can cost farmers in form of animal feed supplements of industrially synthesized S-AA [55].

The concentration of S-AA is strongly related to the relative proportions of S-AA rich proteins in the seeds. In *Vf* and other legumes, it is well accepted that legumins contain relatively higher S-AA compared to vicilin [42, 44, 57-58]. This is further confirmed by comparative analysis of coding sequences of vicilin and legumin subunits across legume species which clearly show that legumin subunits contain more residues of cysteine and methionine (Figure 4). This observation leads to the hypothesis that increasing the proportion of legumin subunits relative to vicilin would improve nutritional content of plant proteins. However, considering that vicilin is accumulated in legume seeds earlier than legumin [59-61], their ratios could be easily offset by the prevailing environmental conditions, e.g. soil nutritional status and onset of biotic and abiotic stresses during the plant growth, and in particular, during grain filling. In contrast to globulins, minor legume seed proteins such as elongation factor Tu, citrate synthase, albumin 2 (PA2), defensins 1 and 2 and Bowman–Birk inhibitors (BBI) contain higher S-AA [42, 62]. According to Krishnan, et al. [63], under higher N availability through fertilizer application or symbiotic fixation, S-AA containing proteins like Bowman-Birk protease inhibitor (BBI) were decreased in favour of β -subunits of β-conglycinins of soybeans. Similarly, ectopic overexpression of *VfAAP1* gene on *P. sativum* and *V. narbonensis* resulted in 30% increase in the globulin fraction but no significant effect on albumin, a S-AA rich protein [64]. Hence, it would appear that the negative correlation between high protein and S-AA content in *Vf* [11, 32, 65] may be the result of preferential accumulation of low nutritional quality protein fractions in higher protein lines.

Genetic control of globulins

Globulins are by far the most abundant seed proteins in legumes and, subsequently, their genetic control has been well investigated. In *Vf*, legumin subunit is encoded by relatively few genes which are classified as legumin A and B genes. A single legumin A gene has been located on the telomeric region of chromosome V of *Vf* [66]. It is not clear, however, whether the legumin A2 gene (*LegA2)* reported in pea [67] also exists in *Vf,* as no up to date information is available. Conversely, there are at least five transcribed genes (*LeB2, LeB3, LeB4, LeB6, LeB7*) for legumin B subunits[66, 68] , of which *LeB3* and *LeB4* have been mapped to chromosome II and III, respectively [66]. The vicilin coding gene [69] was also located on chromosome II, near the centromere [70-71]. While the documented number of genes for *Vf* globulins is relatively small, numerous legumin and vicilin minor subunits with various molecular masses and isoelectric points can be observed in 2D gel electrophoresis analysis [51], suggesting that *Vf* globulins undergo extensive post-translational processing. A similar occurrence has been found in other legumes including *Medicago truncatula* [72] *and Pisum sativum* [73].

There is considerable homology between *Vf* globulin subunits and those of other legumes (Table S1), and where genome sequences are available, it is now possible to classify and associate seed storage subunits to specific genome locations (Table S2). Considering the lack of genome sequence for *Vf*, this information is critical for synteny-based mapping of globulin genes and QTLs. For instance, in *M. truncatula,* several genomic regions coding for globulins have been mapped on chromosome I and VII [72] which are notably syntenic with *Vf* chromosome III and V [39-40] where legumin A and B genes were previously located, respectively [66].

Expression of globulin genes

Seed protein content can be thought of as the final output of a number of biochemical and physiological processes occurring throughout the crop life cycle, each of which are under the control of a regulatory network. Abundance of globulin proteins is regulated by a network of genes involving transcriptional regulation transport and post-translation modifications of storage proteins [72]. Among these are numerous seed specific genes which play profound regulatory roles in the synthesis and accumulation of seed storage proteins [72, 74]. Notably, seed specific transcription factors (TFs) such as *ABI5, LEC1, LEC2, ABI3, MYB#2, bHLH#1* and *FUS3* are key storage protein regulators [72, 75]. ABA insensitive 5 (*ABI5*) is expressed during seed filling stages in plants [75] and has been found at the center of the regulatory gene network for storage protein synthesis in *M. truncatula* [72]. Specifically, it is a major regulator for vicilin polypeptide abundance with *P. sativum abi5* mutants showing nearly 30% decrease in the abundance of vicilin-type globulin [72]. Similarly, *ABI3b* and LEAFY COTYLEDON-1 (*LEC*-1) homologs in soybean has been located at the hub of 118 genes related to seed protein content [74]. Given the microsynteny between *Vf* and the model crop *M. truncatula* [39], these findings will provide a reference for further discoveries in the genetics of *Vf* globulins.

Synthesis and accumulation of seed storage proteins

Globulins are synthesized in the endoplasmic reticulum (ER) sorted in the Golgi body and transported to the protein storage vacuole (PSV) by vesicles [72, 76]. During *Vf* seed development, a diphasic pattern of protein accumulation exists in which proteins synthesized during early developmental stages are only transitorily accumulated and subsequently degraded to sustain the growing embryo while proteins accumulated after heart stage (~12

DAP) are mainly destined for storage into cotyledons' protein bodies [77]. During the latter stage, globulin proteins show distinct expression patterns in which vicilin synthesis and accumulation precedes that of legumin and α chain polypeptides of legumin appear earlier than β chains [59]. Similar pattern of vicilin and legumin gene expression has also been reported in Medicago [78] and soybean [76].

The amount of protein accumulated during seed development can be attributed to various genetic and environmental factors acting on various plant processes ranging from nutrient uptake and transport, photosynthate production and remobilization to protein accumulation rate in the storage organs. However, there are strong indications that mechanisms underlying nitrogen (N) uptake, transport and assimilation could explain the variation in protein content more than any other factor. For instance, in pea, overexpression of the amino acid transporter gene amino acid permease (*AAP*), has been confirmed to play a critical role in increasing synthesis of seed storage proteins owing to increased leaf and pod phloem loading with free amino acids [79]. A similar mechanism could be attributed to the observed 2-3 times higher free amino acids in the cotyledons of high-protein (HP) *Vf* genotypes as compared to low-protein genotypes [80]. In rice, a major seed protein content QTL harboring the *OsAAP* gene was associated with higher uptake of amino acids and their distribution across plant tissues [81]. In addition, QTL for N-fixation have been linked to QTL for total N accumulation in common bean [82] and pea [83]. Also, improved capacity for N uptake can be a candidate trait to relax the yield-protein negative correlation. In fact, increased genetic capacity for N supply was associated with increased seed size in *Vf* [64] or seed number in pea [79]. These results should be taken into consideration when screening for high protein content in *Vf*.

GENETIC IMPROVEMENT OF PROTEIN CONTENT AND QUALITY

Summary of the past work

Several studies have focused on the genetic variation for protein content (**Table 2**) and to what extent protein content was correlated with yield of *Vf*. One study indicated that protein content was variable between and within varieties (n=33) with broad sense heritability of 0.70 and no significant correlation with seed weight [31]. However, when larger set of germplasm (n=600) was screened, a clear negative relationship between seed weight and protein was detected although some large-seeded genotypes with above average protein content were also found [65]. Similarly, after four cycles of selection for protein content, Sjödin [32] concluded that protein content in *Vf* could be improved by selection but tended to negatively correlate with number of seeds per plant regardless of thousand seed weight. These early efforts also established the variability for S-AA content (**Table 3**) and nearly all investigations found a negative correlation between protein and S-AA content [32, 65, 84]. Under circumstances where desirable traits of interest are negatively correlated, deeper understanding of the genetic basis of the trade-offs between the traits and availability of appropriate tools to dissect and recombine them is crucial.

Areas for future focus

Uncoupling the negative yield-protein correlation

Correlation between traits can arise due to gene linkage or pleiotropy [85],with the latter being most common in plants, and its resolution requires deeper understanding of both traits. Therefore, several possible mechanisms have been investigated in various crops in order to unlock protein-yield association. It is hypothesized that the negative correlation between the two traits result when the high demand for N during seed filling stage coincides with decline in soil nutrients in the rhizosphere and nitrogen fixation, resulting in re-mobilization of nitrogen from leaves, which in turn shortens grain filling and reduces seed weights [86]. This is in line with findings by Egle, et al. [87] who showed that majority of N accumulated during seed filling in barley was remobilized from leaves and stems, but that ongoing N uptake could also contribute. Furthermore, wheat genotypes with higher capability for post-anthesis N uptake deviate from grain-protein negative relationship [88-89] and selection for this trait has been therefore proposed as a possible criterion for simultaneous improvement of protein content and grain yield. The genetic basis of post-flowering N uptake is not yet fully understood either in cereals or in legumes but could be related to root structure and/or N transport capacity. For instance, pea genotypes with higher mineral nitrogen absorption and symbiotic nitrogen fixation have shown enhanced seed N content and yield [83]. Moreover, faster rate and relatively longer duration of N accumulation during seed development has been reported as a possible mechanism for combining high protein and large seed size in soybean [90]. The importance of N uptake capacity for protein content and yield was further demonstrated by Peng, et al. [81] who found major protein content QTL *qPC1* harboring a putative amino acid transporter gene (*OsAAP6),* which they proposed as candidate QTL for simultaneous selection for yield and protein content in rice. These areas of enquiry are amenable for further investigation and can potentially point to QTLs that can be used to improve protein content in *Vf* without significant yield reduction.

Improving S-AA content by modifying legumin: vicilin ratio

Considering difficulties in genetic improvement of limiting amino acids through conventional breeding approaches, several genetic engineering approaches have been attempted in various crops over recent decades. Detailed information on these strategies and results obtained can be found in Galili and Amir [56]. These included (i) overexpression of genes encoding proteins rich in the limiting amino acid, (ii) *in vitro* modification of genes encoding proteins of interest by adding more residues of the desired amino acid, (iii) introduction of genes coding for protein rich in the limiting amino acid from one species to another target food crop, or by

(iv) modification of biosynthetic and catabolic pathways to directly increase accumulation of target amino acid or indirectly by increasing accumulation of proteins containing the limiting amino acid. Yet, most of these attempts have not succeeded in producing new crop cultivars combining increased protein quality with desired agronomic traits. In rare cases where reasonable success was achieved, commercialization of the improved cultivars was hindered by legal restrictions on GMO release [56] and consumer resistance. Besides these challenges of consumer acceptability, the potential of transgenic approaches in *Vf* is limited by the inherently poor regenerating ability of *Vf* transgenics [91].

Alternative strategies include direct selection on QTL for S-AA content or indirectly by selecting for greater relative expression of protein subunits rich in S-AA rich subunits. To our knowledge, soybean is the only legume crop in which QTLs for individual S-AA has been mapped [92-93]. Though total seed content of the S-AA *per se* would be a good indicator, it may not be sufficient when considering as selection criteria, due to uncertainty about what percentage of the total S-AA detected is indeed imbedded in the main storage proteins. In *Vf* and other legumes, since it is observed that the legumin protein subunit have relatively higher S-AA content compared to vicilin [42, 44, 57-58], increasing legumin subunit in favor of vicilin would be expected to enhance the protein quality. In fact, the concept of manipulating legumin: vicilin (L/V) ratio to improve nutritional quality is not new in *Vf*. It was previously reported that variation in L/V ratio among varieties was consistent across years [94] and environments [95] and concluded that L/V ratio has genetic basis and could be used as a selection criteria to improve nutritional quality in *Vf* [94-95]. To our knowledge, since L/V ratio based approach was suggested as a practical breeding strategy for improving nutritional quality in soybean [57], only study has tried to map QTLs for L/V ratio and showed colocation between some QTLs for structural legumin and vicilin loci and L/V ratio [96]. The recent advances in *Vf* genetics tools such as development of 50 K SNP array and high-density linkage map may offer an unprecedented opportunity to discover novel QTLs that could represent targets for improving nutritional quality.

Exploiting mutagenesis approaches

Large-scale mutagenesis using physical or chemical mutagenic agents is a well-established method of inducing novel variation to meet human requirements, but which is unlikely to be present in nature. This approach is all the more justified in the case of *Vf* where the primary gene pool lacks any known wild relatives. Indeed, several mutagenesis efforts have produced new sets of morphological phenotypes in *Vf* [40, 97-98]. However, no data is available on potential beneficial mutations in the seed composition of *Vf*. Although Sjödin [97] has reported to have identified some high protein content genotypes from a lot of seeds which had been mutagenized he could not ascertain whether the selected plants were genuine mutants or randomly isolated extremes in the original seed lot. There are several potential ways of exploiting induced mutations for improving protein content and/or quality. First, desirable mutations involving photosynthetic and N provision mechanisms can improve protein content. From ethyl methane sulfonate (EMS) mutagenized seeds, Duc [98] discovered a supernodulating line with 3-4 times higher number of nodules compared to the parental line. Considering the close relationship between N fixation and protein content, such a trait could be exploited in breeding programs. Secondly, knockdown/knockout or regulatory mutations leading to absence of major protein subunits such as vicilins can result in improved nutritional quality by increasing the ratio of S-AA rich subunits like legumin and albumins. Such mutations could be *cis*-linked to the structural loci themselves or *trans*-acting factors that would need to be mapped *de novo*. For instance, mutants of *PsABI5*, a major *trans*-acting regulator of vicilin abundance in pea, have shown an increased legumin abundance [72]. Thirdly, presence or absence of certain subunits can enable dissection of genetic control of individual protein subunits via a QTL mapping approach [55]. Lastly, it is possible via a reverse genetic screen to select non-synonymous mutations that convert non-S-AA residues to S-AA residues in S-AA poor storage proteins such as vicilins, although, only a proportion of codons are available for single base changes that would result in this outcome. Moreover, the physico-chemical properties of cysteine (disulfide bridge-forming) and methionine (hydrophobic) may cause steric constraints [99]. However even a single well-placed additional methionine in each vicilin could give rise to a significant step up in S-AA levels and this approach is therefore worth trying. On a more practical level, full exploitation of mutagenesis for the above purposes requires high-throughput and cheap phenotyping methods to screen tens of thousands of plants for nutritional and agronomic traits.

In summary, *Vf* is one of the most important legumes crops with great potential to fulfil multiple nutritional and ecological services for the current and future generations. However, *Vf* can only play this role if it meets certain producer and end-user expectations which requires plant breeders and research community to address both agronomic and nutritional constraints simultaneously. In drawing together a synthesis of the literature on *Vf* seed protein content, contribution of different storage protein classes to overall abundance and to varying relative amounts of essential amino acids, globulin structure and globulin-encoding genes, we aim to provide an updated and comprehensive primer for researchers interested in the nutritional optimization of faba beans. We discuss a range of approaches by which protein content could be increased (without compromising yield) and protein quality ameliorated, some of which have successful precedent in related legume species. These include: high resolution mapping of protein, L:V ration and S-AA QTL using powerful modern quantitative genetics methods and genomics technologies; manipulation of known or still-to-be-discovered structural and regulatory genes by transformation and screening of mutant libraries to reveal novel structural and regulatory variants not found in nature. In parallel, as genome sequencing become cheaper and more genomic resources for *Vf* are accumulated, all the above should become ever more efficient, enhancing the prospects of increasing protein content and quality in this strategic crop.

ABBREVIATIONS USED

Vf, Vicia faba; S-AA, sulfur containing amino acid; V-C, vicine and convicine; QTL, quantitative trait loci; SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; HPLC, High Performance Liquid Chromatography; kDa, Kilo Dalton; N, nitrogen; GMO, Genetically Modified Organisms; EMS, Ethyl Methane Sulfonate (EMS)

ACKNOWLEDGEMENTS

The authors thank Islamic Development Bank (IDB) for the financial support during preparation of this manuscript.

SUPPORTING INFORMATION

Supplementary data including Figure S1 and Tables S1-S3 are provided in MS Word document.

CONFLICT OF INTEREST

The authors declare no competing financial interest.

REFERENCES

(1) Henchion, M.; Hayes, M.; Mullen, A.; Fenelon, M.; Tiwari, B., Future Protein Supply and Demand: Strategies and Factors Influencing a Sustainable Equilibrium. *Foods* **2017,** *6* (7), 53.

(2) Young, V. R.; Pellett, P. L., Plant proteins in relation to human protein and amino acid nutrition. *The American Journal of Clinical Nutrition* **1994,** *59* (5), 1203S-1212S.

(3) Smýkal, P.; Coyne, C. J.; Ambrose, M. J.; Maxted, N.; Schaefer, H.; Blair, M. W.; Berger, J.; Greene, S. L.; Nelson, M. N.; Besharat, N.; Vymyslický, T.; Toker, C.; Saxena, R. K.; Roorkiwal, M.; Pandey, M. K.; Hu, J.; Li, Y. H.; Wang, L. X.; Guo, Y.; Qiu, L. J.; Redden, R. J.; Varshney, R. K., Legume Crops Phylogeny and Genetic Diversity for Science and Breeding. *Crit. Rev. Plant Sci.* **2015,** *34* (1-3), 43-104.

(4) Foyer, C. H.; Lam, H. M.; Nguyen, H. T.; Siddique, K. H.; Varshney, R. K.; Colmer, T. D.; Cowling, W.; Bramley, H.; Mori, T. A.; Hodgson, J. M.; Cooper, J. W.; Miller, A. J.; Kunert, K.; Vorster, J.; Cullis, C.; Ozga, J. A.; Wahlqvist, M. L.; Liang, Y.; Shou, H.; Shi, K.; Yu, J.; Fodor, N.; Kaiser, B. N.; Wong, F. L.; Valliyodan, B.; Considine, M. J., Neglecting legumes has compromised human health and sustainable food production. *Nat Plants* **2016,** *2*, 16112.

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

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

(7) Cubero, J. I., On the evolution of Vicia faba L. *Theor. Appl. Genet.* **1974,** *45* (2), 47-51.

(8) FAOstat United Nations Organization for Food and Agriculture. (accessed 20 May).

(9) Duc, G.; Aleksić, J. M.; Marget, P.; Mikic, A.; Paull, J.; Redden, R. J.; Sass, O.; Stoddard, F. L.; Vandenberg, A.; Vishnyakova, M.; Torres, A. M., Faba Bean. In *Grain Legumes*, Ron, A. M. D., Ed. Springer Science+Business Media: New York, 2015; Vol. 10, pp 141-178.

(10) Makkar, H. P. S.; Becker, K.; Abel, H.; Pawelzik, E., Nutrient contents, rumen protein degradability and antinutritional factors in some colour- and white-flowering cultivars of Vicia faba beans. *J. Sci. Food Agric.* **1997,** *75* (4), 511-520.

(11) Schumacher, H.; Paulsen, H. M.; Gau, A. E.; Link, W.; Jurgens, H. U.; Sass, O.; Dieterich, R., Seed protein amino acid composition of important local grain legumes Lupinus angustifolius L., Lupinus luteus L., Pisum sativum L. and Vicia faba L. *Plant Breeding* **2011,** *130* (2), 156-164.

(12) Kaskinen, T.; Lähteenoja, S.; Sokero, M.; Suomela, I., Strategic Business Examples from Finland: The Growth of the Smartup Industry. In *Factor X: Challenges, Implementation Strategies and Examples for a Sustainable Use of Natural Resources*, Lehmann, H., Ed. Springer International Publishing: Cham, 2018; pp 325-333.

(13) Coda, R.; Varis, J.; Verni, M.; Rizzello, C. G.; Katina, K., Improvement of the protein quality of wheat bread through faba bean sourdough addition. *LWT - Food Science and Technology* **2017,** *82* (Supplement C), 296-302.

(14) Rizzello, C. G.; Verni, M.; Koivula, H.; Montemurro, M.; Seppa, L.; Kemell, M.; Katina, K.; Coda, R.; Gobbetti, M., Influence of fermented faba bean flour on the nutritional, technological and sensory quality of fortified pasta. *Food Funct.* **2017,** *8* (2), 860-871.

(15) VTT Gluten-free faba bean for bread and pasta. (accessed 25 July).

(16) Cernay, C.; Pelzer, E.; Makowski, D., A global experimental dataset for assessing grain legume production. *Scientific Data* **2016,** *3*, 160084.

(17) Baddeley, J.; Jones, S.; Topp, C.; Watson, C.; Helming, J.; Stoddard, F. Biological nitrogen fixation (BNF) by legume crops in Europe. www.legumefutures.de (accessed January 4).

(18) Jensen, E. S.; Peoples, M. B.; Hauggaard-Nielsen, H., Faba bean in cropping systems. *Field Crops Res.* **2010,** *115* (3), 203-216.

(19) Köpke, U.; Nemecek, T., Ecological services of faba bean. *Field Crops Res.* **2010,** *115* (3), 217-233.

(20) Perez-Maldonado, R. A.; Mannion, P. F.; Farrell, D. J., Optimum inclusion of field peas, faba beans, chick peas and sweet lupins in poultry diets. I. Chemical composition and layer experiments. *British Poultry Science* **1999,** *40* (5), 667-673.

(21) Koivunen, E.; Tuunainen, P.; Valkonen, E.; Rossow, L.; Valaja, J., Use of faba beans (Vicia faba L.) in diets of laying hens. *Agricultural And Food Science* **2014** *23*, 165–172.

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

(23) Kawashima, H.; Bazin, M. J.; Lynch, J. M., A modelling study of world protein supply and nitrogen fertilizer demand in the 21st century. *Environ. Conserv.* **2002,** *24* (1), 50-57.

(24) Speedy, A. W. In *Overview of world feed protein needs and supply*, FAO Animal Production and Health Proceedings (FAO), FAO: 2004.

(25) Chiari, N., Food Security. The Challenge of Nutrition in the New Century. *Relations. Beyond Anthropocentrism* **2017,** *5* (2), 145-156.

(26) WHO/FAO/UNU, Protein and amino acid requirements in human nutrition. *WHO Tech. Rep. Ser.* **2007,** (935), 1-265, back cover.

(27) Balyan, H. S.; Gupta, P. K.; Kumar, S.; Dhariwal, R.; Jaiswal, V.; Tyagi, S.; Agarwal, P.; Gahlaut, V.; Kumari, S., Genetic improvement of grain protein content and other health-related constituents of wheat grain. *Plant Breeding* **2013,** *132* (5), 446-457.

(28) de Visser, C. L. M.; Schreuder, R.; Stoddard, F., The EU's dependency on soya bean import for the animal feed industry and potential for EU produced alternatives. *OCL* **2014,** *21* (4), D407.

(29) Statista Meat consumption and vegetarianism in Europe - Statistics and Facts. https://www.statista.com/topics/3345/meat-consumption-and-vegetarianism-in-europe/ (accessed December 25).

(30) de Boer, J.; Aiking, H., Prospects for pro-environmental protein consumption in Europe: Cultural, culinary, economic and psychological factors. *Appetite* **2018,** *121*, 29-40.

(31) Griffiths, D. W.; Lawes, D. A., Variation in the crude protein content of field beans (Vicia faba L.) in relation to the possible improvement of the protein content of the crop. *Euphytica* **1978,** *27* (2), 487-495.

(32) Sjödin, J., Protein Quantity and Quality in Vicia Faba. In *Faba Bean Improvement: Proceedings of the Faba Bean Conference held in Cairo, Egypt, March 7–11, 1981*, Hawtin, G.; Webb, C., Eds. Springer Netherlands: Dordrecht, 1982; pp 319-331.

(33) Frauen, M.; Röbbelen, G.; Ebrneyer, E., Quantitative Measurement of Quality Determining Constituents in Seeds of Different Inbred Lines from A World Collection of *Vicia Faba*. In *Vicia baba: Agronomy, Physiology and Breeding*, Hebblethwaite, P. D.; Dawkins, T. C. K.; Heath, M. C.; Lockwood, G., Eds. Springer-Science+Business Media, B.V.: Brussels-Luxembourg, 1984; Vol. 10, pp 279-287.

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

(35) Yu, E.-M.; Zhang, H.-F.; Li, Z.-F.; Wang, G.-J.; Wu, H.-K.; Xie, J.; Yu, D.-G.; Xia, Y.; Zhang, K.; Gong, W.-B., Proteomic signature of muscle fibre hyperplasia in response to faba bean intake in grass carp. **2017,** *7*, 45950.

(36) Rizzello, C. G.; Losito, I.; Facchini, L.; Katina, K.; Palmisano, F.; Gobbetti, M.; Coda, R., Degradation of vicine, convicine and their aglycones during fermentation of faba bean flour. *Sci Rep* **2016,** *6*, 32452.

(37) Coda, R.; Melama, L.; Rizzello, C. G.; Curiel, J. A.; Sibakov, J.; Holopainen, U.; Pulkkinen, M.; Sozer, N., Effect of air classification and fermentation by Lactobacillus plantarum VTT E-133328 on faba bean (Vicia faba L.) flour nutritional properties. *Int. J. Food Microbiol.* **2015,** *193*, 34-42.

(38) Khazaei, H.; Purves, R. W.; Song, M.; Stonehouse, R.; Bett, K. E.; Stoddard, F. L.; Vandenberg, A., Development and validation of a robust, breeder-friendly molecular marker for the vc-locus in faba bean. *Mol. Breed.* **2017,** *37* (11), 140.

(39) Webb, A.; Cottage, A.; Wood, T.; Khamassi, K.; Hobbs, D.; Gostkiewicz, K.; White, M.; Khazaei, H.; Ali, M.; Street, D.; Duc, G.; Stoddard, F. L.; Maalouf, F.; Ogbonnaya, F. C.; Link, W.; Thomas, J.; O'Sullivan, D. M., A SNP-based consensus genetic map for synteny-based trait targeting in faba bean (Vicia faba L.). *Plant Biotechnol. J.* **2016,** *14* (1), 177-85.

(40) O'Sullivan, D. M.; Angra, D., Advances in Faba Bean Genetics and Genomics. *Front Genet* **2016,** *7*, 150.

(41) Patil, G.; Mian, R.; Vuong, T.; Pantalone, V.; Song, Q.; Chen, P.; Shannon, G. J.; Carter, T. C.; Nguyen, H. T., Molecular mapping and genomics of soybean seed protein: a review and perspective for the future. *Theor. Appl. Genet.* **2017,** *130* (10), 1975-1991.

(42) Liu, Y.; Wu, X.; Hou, W.; Li, P.; Sha, W.; Tian, Y., Structure and function of seed storage proteins in faba bean (Vicia faba L.). *3 Biotech* **2017,** *7* (1), 74.

(43) Shewry, P. R.; Casey, R., Seed Proteins. In *Seed Proteins*, Shewry, P. R.; Casey, R., Eds. Springer Netherlands: Dordrecht, 1999; pp 1-10.

(44) Joshi, J.; Pandurangan, S.; Diapari, M.; Marsolais, F., Comparison of Gene Families: Seed Storage and Other Seed Proteins. In *The Common Bean Genome*, Pérez de la Vega, M.; Santalla, M.; Marsolais, F., Eds. Springer International Publishing: Cham, 2017; pp 201-217.

(45) Müntz, K.; Horstmann, C.; Schlesier, B., Vicia globulins. In *Seed Proteins*, Shewry, P. R.; Casey, R., Eds. Springer Netherlands: Dordrecht, 1999; pp 259-284.

(46) Kesari, P.; Sharma, A.; Katiki, M.; Kumar, P.; R Gurjar, B.; Tomar, S.; K Sharma, A.; Kumar, P., Structural, Functional and Evolutionary Aspects of Seed Globulins. *Protein and peptide letters* **2017,** *24* (3), 267-277.

(47) Meng, G. T.; Ma, C. Y., Thermal properties of Phaseolus angularis (red bean) globulin. *Food Chem.* **2001,** *73* (4), 453-460.

(48) Kimura, A.; Fukuda, T.; Zhang, M.; Motoyama, S.; Maruyama, N.; Utsumi, S., Comparison of Physicochemical Properties of 7S and 11S Globulins from Pea, Fava Bean, Cowpea, and French Bean with Those of Soybean—French Bean 7S Globulin Exhibits Excellent Properties. *J. Agric. Food Chem.* **2008,** *56* (21), 10273-10279.

(49) Heng, L.; van Koningsveld, G. A.; Gruppen, H.; van Boekel, M. A. J. S.; Vincken, J. P.; Roozen, J. P.; Voragen, A. G. J., Protein–flavour interactions in relation to development of novel protein foods. *Trends Food Sci. Technol.* **2004,** *15* (3), 217-224.

(50) Baumlein, H.; Wobus, U.; Pustell, J.; Kafatos, F. C., The legumin gene family: structure of a B type gene of Vicia faba and a possible legumin gene specific regulatory element. *Nucleic Acids Res.* **1986,** *14* (6), 2707-20.

(51) Tucci, M.; Capparelli, R.; Costa, A.; Rao, R., Molecular heterogeneity and genetics of Vicia faba seed storage proteins. *Theor. Appl. Genet.* **1991,** *81* (1), 50-58.

(52) Le Signor, C.; Gallardo, K.; Prosperi, J. M.; Salon, C.; Quillien, L.; Thompson, R.; Duc, G., Genetic diversity for seed protein composition in Medicago truncatula. *Plant Genetic Resources* **2005,** *3* (1), 59-71.

(53) O'Kane, F. E.; Happe, R. P.; Vereijken, J. M.; Gruppen, H.; van Boekel, M. A. J. S., Characterization of Pea Vicilin. 1. Denoting Convicilin as the α-Subunit of the Pisum Vicilin Family. *J. Agric. Food Chem.* **2004,** *52* (10), 3141-3148.

(54) Krishnan, H. B.; Natarajan, S. S.; Oehrle, N. W.; Garrett, W. M.; Darwish, O., Proteomic Analysis of Pigeonpea (Cajanus cajan) Seeds Reveals the Accumulation of Numerous Stress-Related Proteins. *J. Agric. Food Chem.* **2017,** *65* (23), 4572-4581.

(55) Boehm, J. D.; Nguyen, V.; Tashiro, R. M.; Anderson, D.; Shi, C.; Wu, X.; Woodrow, L.; Yu, K.; Cui, Y.; Li, Z., Genetic mapping and validation of the loci controlling 7S α′ and 11S A-type storage protein subunits in soybean [Glycine max (L.) Merr.]. *Theor. Appl. Genet.* **2017**, 1-13.

(56) Galili, G.; Amir, R., Fortifying plants with the essential amino acids lysine and methionine to improve nutritional quality. *Plant Biotechnol. J.* **2013,** *11* (2), 211-22.

(57) Kwanyuen, P.; Pantalone, V. R.; Burton, J. W.; Wilson, R. F., A new approach to genetic alteration of soybean protein composition and quality. *J. Am. Oil Chem. Soc.* **1997,** *74* (8), 983-987.

(58) Jackson, P.; Boulter, D.; Thurman, D. A., A comparison of some properties of vicilin and legumin isolated from seeds of *Pisum sativum*, *Vicia faba* and *Cicer arietinum*. *New Phytol.* **1969** *68* 25-33.

(59) De Pace, C.; Delre, V.; Mugnozza, G. T. S.; Maggini, E.; Cremonini, R.; Frediani, M.; Cionini, P. G., Legumin of Vicia faba major: accumulation in developing cotyledons, purification, mRNA characterization and chromosomal location of coding genes. *Theor. Appl. Genet.* **1991,** *83*, 17-23.

(60) Abirached-Darmency, M.; Dessaint, F.; Benlicha, E.; Schneider, C., Biogenesis of protein bodies during vicilin accumulation in Medicago truncatula immature seeds. *BMC Research Notes* **2012,** *5*, 409-409.

(61) Gallardo, K.; Le Signor, C.; Vandekerckhove, J.; Thompson, R. D.; Burstin, J., Proteomics of Medicago truncatula seed development establishes the time frame of diverse metabolic processes related to reserve accumulation. *Plant Physiol.* **2003,** *133* (2), 664-82.

(62) Rubio, L. A.; Perez, A.; Ruiz, R.; Guzman, M. A.; Aranda-Olmedo, I.; Clemente, A., Characterization of pea (Pisum sativum) seed protein fractions. *J. Sci. Food Agric.* **2014,** *94* (2), 280-7.

(63) Krishnan, H. B.; Bennett, J. O.; Kim, W.-S.; Krishnan, A. H.; Mawhinney, T. P., Nitrogen Lowers the Sulfur Amino Acid Content of Soybean (Glycine max [L.] Merr.) by Regulating the Accumulation of Bowman−Birk Protease Inhibitor. *J. Agric. Food Chem.* **2005,** *53* (16), 6347-6354.

(64) Rolletschek, H.; Hosein, F.; Miranda, M.; Heim, U.; Gotz, K. P.; Schlereth, A.; Borisjuk, L.; Saalbach, I.; Wobus, U.; Weber, H., Ectopic expression of an amino acid transporter (VfAAP1) in seeds of Vicia narbonensis and pea increases storage proteins. *Plant Physiol.* **2005,** *137* (4), 1236-49.

(65) Lafiandra, D.; Polignano, G. B.; Filippetti, A.; Porceddu, E., Genetic variability for protein content and S-aminoacids in broad-beans (Vicia faba L.). *Die Kulturpflanze* **1981,** *29* (1), 115-127.

(66) Fuchs, J.; Schubert, I., Localization of seed protein genes on metaphase chromosomes of Vicia faba via fluorescencein situ hybridization. *Chromosome Research* **1995,** *3* (2), 94-100.

(67) Rerie, W. G.; Whitecross, M.; Higgins, T. J. V., Developmental and environmental regulation of pea legumin genes in transgenic tobacco. *Molecular and General Genetics MGG* **1991,** *225* (1), 148-157.

(68) Heim, U.; Schubert, R.; Baumlein, H.; Wobus, U., The legumin gene family: structure and evolutionary implications of Vicia faba B-type genes and pseudogenes. *Plant Mol. Biol.* **1989,** *13* (6), 653-63.

(69) Weschke, W.; Bassüner, R.; Van Hai, N.; Czihal, A.; Baümlein, H.; Wobus, U., The structure of a Vicia faba vicilin gene. *Biochemie und Physiologie der Pflanzen* **1988,** *183* (2-3), 233-242.

(70) Jiri, M.; Winfriede, W.; Helmut, B.; Uta, P.; Andreas, H.; Ulrich, W.; Ingo, S., Localization of vicilin genes via polymerase chain reaction on microisolated field bean chromosomes. *The Plant Journal* **1993,** *3* (6), 883-886.

(71) Fuchs, J.; Joos, S.; Licheter, P.; Schubert, I., Localization of Vicilin Genes on Field Bean Chromosome II by Fluorescent in situ Hybridization. *J. Hered.* **1994,** *85* (6), 487-488.

(72) Le Signor, C.; Aimé, D.; Bordat, A.; Belghazi, M.; Labas, V.; Gouzy, J.; Young, N. D.; Prosperi, J.-M.; Leprince, O.; Thompson, R. D.; Buitink, J.; Burstin, J.; Gallardo, K.,

Genome-wide association studies with proteomics data reveal genes important for synthesis, transport and packaging of globulins in legume seeds. *New Phytol.* **2017,** *214* (4), 1597-1613.

(73) Bourgeois, M.; Jacquin, F.; Savois, V.; Sommerer, N.; Labas, V.; Henry, C.; Burstin, J., Dissecting the proteome of pea mature seeds reveals the phenotypic plasticity of seed protein composition. *Proteomics* **2009,** *9* (2), 254-71.

(74) Zhaoming, Q.; Zhanguo, Z.; Zhongyu, W.; Jingyao, Y.; Hongtao, Q.; Xinrui, M.; Hongwei, J.; Dawei, X.; Zhengong, Y.; Rongsheng, Z.; Chunyan, L.; Wei, Y.; Zhenbang, H.; Xiaoxia, W.; Jun, L.; Qingshan, C., Meta□analysis and transcriptome profiling reveal hub genes for soybean seed storage composition during seed development. *Plant, Cell Environ.* **2018,** *0* (0), 1–19.

(75) Verdier, J.; Thompson, R. D., Transcriptional regulation of storage protein synthesis during dicotyledon seed filling. *Plant Cell Physiol.* **2008,** *49* (9), 1263-71.

(76) Mori, T.; Maruyama, N.; Nishizawa, K.; Higasa, T.; Yagasaki, K.; Ishimoto, M.; Utsumi, S., The composition of newly synthesized proteins in the endoplasmic reticulum determines the transport pathways of soybean seed storage proteins. *The Plant Journal* **2004,** *40* (2), 238-249.

(77) Panitz, R.; Borisjuk, L.; Manteuffel, R.; Wobus, U., Transient expression of storage-protein genes during early embryogenesis ofVicia faba: synthesis and metabolization of vicilin and legumin in the embryo, suspensor and endosperm. *Planta* **1995,** *196* (4), 765-774.

(78) Wang, X. D.; Song, Y.; Sheahan, M. B.; Garg, M. L.; Rose, R. J., From embryo sac to oil and protein bodies: embryo development in the model legume Medicago truncatula. *New Phytol.* **2012,** *193* (2), 327-38.

(79) Zhang, L.; Garneau, M. G.; Majumdar, R.; Grant, J.; Tegeder, M., Improvement of pea biomass and seed productivity by simultaneous increase of phloem and embryo loading with amino acids. *Plant J.* **2015,** *81* (1), 134-46.

(80) Golombek, S.; Rolletschek, H.; Wobus, U.; Weber, H., Control of storage protein accumulation during legume seed development. *J. Plant Physiol.* **2001,** *158* (4), 457-464.

(81) Peng, B.; Kong, H.; Li, Y.; Wang, L.; Zhong, M.; Sun, L.; Gao, G.; Zhang, Q.; Luo, L.; Wang, G.; Xie, W.; Chen, J.; Yao, W.; Peng, Y.; Lei, L.; Lian, X.; Xiao, J.; Xu, C.; Li, X.; He, Y., OsAAP6 functions as an important regulator of grain protein content and nutritional quality in rice. *Nature Communications* **2014,** *5*, 4847.

(82) Ramaekers, L.; Galeano, C. H.; Garzón, N.; Vanderleyden, J.; Blair, M. W., Identifying quantitative trait loci for symbiotic nitrogen fixation capacity and related traits in common bean. *Mol. Breed.* **2013,** *31* (1), 163-180.

(83) Bourion, V.; Rizvi, S. M. H.; Fournier, S.; Larambergue, H. d.; Galmiche, F.; Marget, P.; Duc, G.; Burstin, J., Genetic dissection of nitrogen nutrition in pea through a QTL approach of root, nodule, and shoot variability. *Theor. Appl. Genet.* **2010,** *121*, 71-86.

(84) Griffiths, D. W., An Assesment of the Potential for Improving the Nutritive Value of Field Beans (Vicia Faba)- A Progress Report. In *Vicia faba: Agronomy, Physiology and Breeding*, Hebblethwaite, P. D.; Dawkins, T. C. K.; Heath, M. C.; Lockwood, G., Eds. Springer-Science+Business Media, B.V. : Brussels-Luxembourg, 1984; Vol. 10, pp 271-278.

(85) Chen, Y.; Lübberstedt, T., Molecular basis of trait correlations. *Trends Plant Sci.* **2010,** *15* (8), 454-461.

(86) Munier-Jolain, N.; Larmure, A.; Salon, C., Determinism of carbon and nitrogen reserve accumulation in legume seeds. *C. R. Biol.* **2008,** *331* (10), 780-787.

(87) Egle, K.; Beschow, H.; Merbach, W., Nitrogen allocation in barley: Relationships between amino acid transport and storage protein synthesis during grain filling. *Canadian Journal of Plant Science* **2015,** *95* (3), 451-459.

(88) Bogard, M.; Allard, V.; Brancourt-Hulmel, M.; Heumez, E.; Machet, J.-M.; Jeuffroy, M.-H.; Gate, P.; Martre, P.; Le Gouis, J., Deviation from the grain protein concentration–

grain yield negative relationship is highly correlated to post-anthesis N uptake in winter wheat. *J. Exp. Bot.* **2010,** *61* (15), 4303-4312.

(89) Taulemesse, F.; Le Gouis, J.; Gouache, D.; Gibon, Y.; Allard, V., Bread Wheat (Triticum aestivum L.) Grain Protein Concentration Is Related to Early Post-Flowering Nitrate Uptake under Putative Control of Plant Satiety Level. *PLoS ONE* **2016,** *11* (2), e0149668.

(90) Poeta, F.; Ochogavia, A. C.; Permingeat, H. R.; Rotundo, J. L., Storage-Associated Genes and Reserves Accumulation in Soybean Cultivars Differing in Physiological Strategies for Attaining High Seed Protein Concentration. *Crop Sci.* **2017,** *57* (1), 427-436.

(91) Hanafy, M.; Pickardt, T.; Kiesecker, H.; Jacobsen, H.-J., Agrobacterium-mediated transformation of faba bean (Vicia faba L.) using embryo axes. *Euphytica* **2005,** *142* (3), 227-236.

(92) Wang, X.; Jiang, G.-L.; Song, Q.; Cregan, P. B.; Scott, R. A.; Zhang, J.; Yen, Y.; Brown, M., Quantitative trait locus analysis of seed sulfur-containing amino acids in two recombinant inbred line populations of soybean. *Euphytica* **2014,** *201* (2), 293-305.

(93) Warrington, C. V.; Abdel-Haleem, H.; Hyten, D. L.; Cregan, P. B.; Orf, J. H.; Killam, A. S.; Bajjalieh, N.; Li, Z.; Boerma, H. R., QTL for seed protein and amino acids in the Benning x Danbaekkong soybean population. *Theor. Appl. Genet.* **2015,** *128* (5), 839-50.

(94) Martensson, P., Variation in legumin : vicilin ratio between and within cultivars of Vicia faba L. var. minor. Martinus Nijhoff. World crops: production, utilization and description, volume 3.: The Hague, 1980; pp 159-172.

(95) Gatehouse, J.; Croy, R.; McIntosh, R.; Paul, C.; Boulter, D., Quantitative and qualitative variation in the storage proteins of material from the EEC joint field bean test. *Quantitative and qualitative variation in the storage proteins of material from the EEC joint field bean test.* **1980**, 173-188.

(96) Ma, Y.; Kan, G.; Zhang, X.; Wang, Y.; Zhang, W.; Du, H.; Yu, D., Quantitative Trait Loci (QTL) Mapping for Glycinin and beta-Conglycinin Contents in Soybean (Glycine max L. Merr.). *J. Agric. Food Chem.* **2016,** *64* (17), 3473-83.

(97) Sjödin, J., Induced morphological variation in Vicia faba L. *Hereditas* **1971,** *67* (2), 155-179.

(98) Duc, G., Mutagenesis of faba bean (Vicia faba L.) and the identification of five different genes controlling no nodulation, ineffective nodulation or supernodulation. *Euphytica* **1995,** *83* (2), 147-152.

(99) Brosnan, J. T.; Brosnan, M. E., The sulfur-containing amino acids: an overview. *J. Nutr.* **2006,** *136* (6 Suppl), 1636s-1640s.

(100) Utsumi, S.; Yokoyama, Z.-i.; Mori, T., Comparative Studies of Subunit Compositions of Legumins from Various Cultivars of Vicia faba L. Seeds. *Agric. Biol. Chem.* **1980,** *44* (3), 595-601.

(101) Fontes, E. P. B.; Moreira, M. A.; Davies, C. S.; Nielsen, N. C., Urea-elicited changes in relative electrophoretic mobility of certain glycinin and β-conglycinin subunits. *Plant Physiol.* **1984,** *76* (3), 840-842.

(102) Poysa, V.; Woodrow, L.; Yu, K., Effect of soy protein subunit composition on tofu quality. *Food Res. Int.* **2006,** *39* (3), 309-317.

(103) Mertens, C.; Dehon, L.; Bourgeois, A.; Verhaeghe-Cartrysse, C.; Blecker, C., Agronomical factors influencing the legumin/vicilin ratio in pea (Pisum sativum L.) seeds. *J. Sci. Food Agric.* **2012,** *92* (8), 1591-6.

(104) Ladjal E, Y.; Boudries, H.; Mohamed, C.; Romero, A., *Pea, Chickpea and Lentil Protein Isolates: Physicochemical Characterization and Emulsifying Properties*. 2015; Vol. 11, p 43-51.

(105) Schumacher, H.; Paulsen, H. M.; Gau, A. E., Phenotypical indicators for the selection of methionine enriched local legumes in plant breeding. *Agriculture and Forestry Research* **2009** *4*(59), 339-344.

FIGURE CAPTIONS

Figure 1. Global distribution of *Vf* cultivation (A) and the major producing countries (B). Data was sourced from FAOstats and distribution map was generated using Tableau Public 2018.1.

Figure 2. Predicted ribbon structures of *Vf* globulins. Vicilin (A) is trimeric consisting of 3 protomers (a=light blue, b= magenta and c= green) while legumin is hexameric consisting of legumin A (B) and legumin B (C). Spherical balls in legumin subunits represent disulfide bonds. The models were generated using SWISS-MODEL and processed with PyMOL software. Model description details are in Table S3.

Figure 3. 1D SDS-PAGE showing the major subunits of *Vf* globulins and the variation in protein band abundance among 11 inbred lines.

Figure 4. Amino acid composition (g/16 g N) of *Vf* seed protein (Makkar et al., 1997; Grela et al., 2017). It clearly shows the abundance of several amino acids and deficiency of the S-AA in *Vf* proteins.

Figure 5. Relative abundances of limiting amino acids within legumin and vicilin coding sequences of 7 legume species (Table S1). Annotated protein accessions were obtained from Uniprot and the amino acid residues were counted using “seqinr” package in R.

Table 1. Major globulin polypeptides of *Vf* and related species as annotated on SDS-PAGE

Species	**11S legumin-like (~kDa)**		**7S Vicilin-like (~kDa)**		**Ref.**
	α chain	**β chain**	**vicilin**	**Convicilin**	
Vicia faba	38	22-24	31-65		De Pace, et al. [59]
	38-47	..	..	64	Liu, et al. [42]
	40	20	..	..	Gatehouse, et al. [95]
	35-39	23-25	42-48	66	Tucci, et al. [51]
	36-51	19-23	..	..	Utsumi, et al. [100]
	40	23-24	54	~73	This study
Medicago truncatulla	36-46	23-24	46-47	60-92	Le Signor, et al. [52]
	42-46	23	46-47		Gallardo, et al. [61]
	38-41	..	47	70	Le Signor, et al. [72]
*Glycine max**	37	20	52-72		Fontes, et al. [101]
	37	20	52-72		Boehm, et al. [55]
	37	20	52-72		Poysa, et al. [102]
	37−44	17−22	53-76		Krishnan, et al. [54]

Pisum sativum	40-45	18-25	53	60-88	Bourgeois, et al. [73]
	40	24.8	47.2	67.2	Mertens, et al. [103]
	40	..	..	>70	Rubio, et al. [62]
	37	25	43-53	70	Ladjal E, et al. [104]

*7S subunits of *G.max* consist of α`, α and β polypeptides.

Table 2. Genetic variability in seed protein content in *Vf*

No. genotypes	**Protein content (%)**	**Reference**
33	22-38	Griffiths and Lawes [31]
600	19-34	Lafiandra, et al. [65]
125	22-36	Sjödin [32]
125	29-38	Frauen, et al. [33]
30	23-39	Griffiths [84]
12	26-30	Makkar, et al. [10]

Table 3. Genetic variability in sulfur-containing amino acids in *Vf* (g/16 g N)

No. genotypes	Methionine	Cysteine	Reference
111	0.6-1.0	1.0-1.5	Lafiandra, et al. [65*]
125	0.8-1.4	1.3-1.4	Sjödin [32*]
125	0.1-0.2	0.2-0.6	Frauen, et al. [33]
12	0.8-1.1	1.1-1.4	Makkar, et al. [10]
50	0.6 - 0.9	1.0 - 1.4	Schumacher, et al. [105]
46	0.6-0.9	0.9-1.2	Schumacher, et al. [11]

* S-AA reported as % protein

Figure 1


A
B
United States
India
Egypt 3%
Sudan 3%
Germany 3%
France 4%
United Kingdom 6%
Australia 10%
Ethiopia 20%
China 36%
Other countries 15%

Figure 2

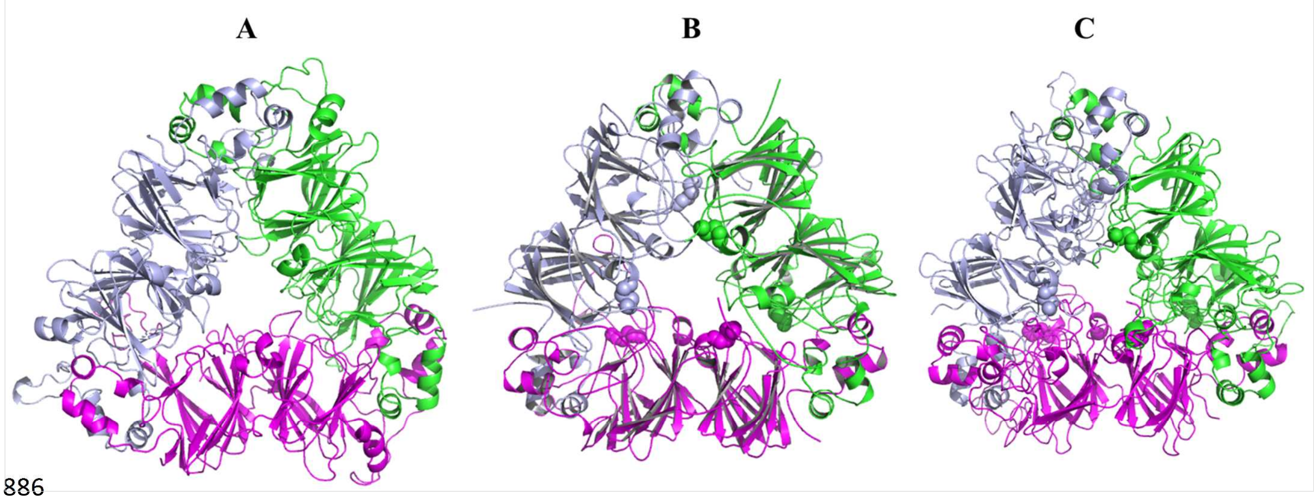


Figure 3

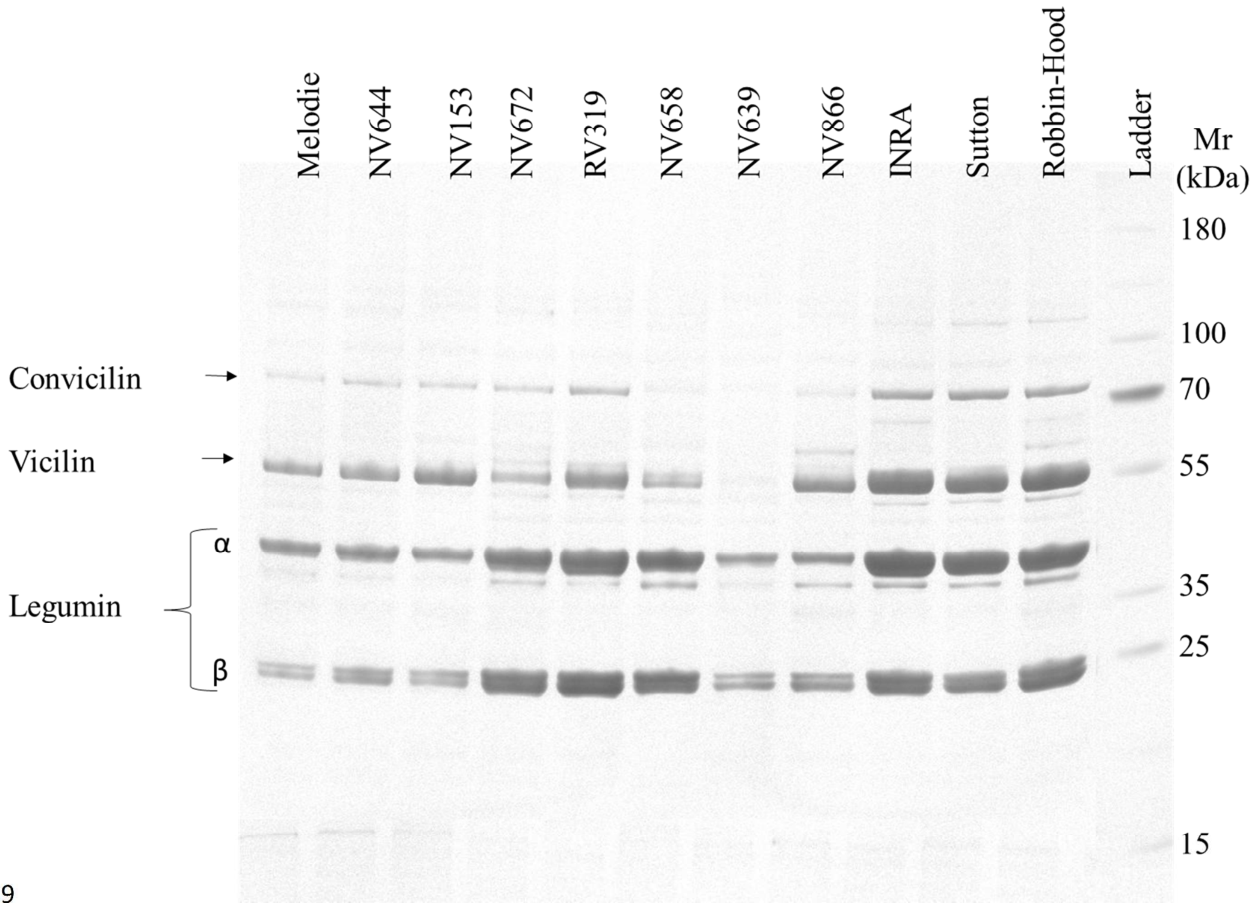

Melodie
NV644
NV153
NV672
RV319
NV658
NV639
NV866
INRA
Sutton
Robbin-Hood
Ladder
Mr (kDa)
180
100
70
55
35
25
15
Convicilin
Vicilin
Legumin
α
β

Figure 4

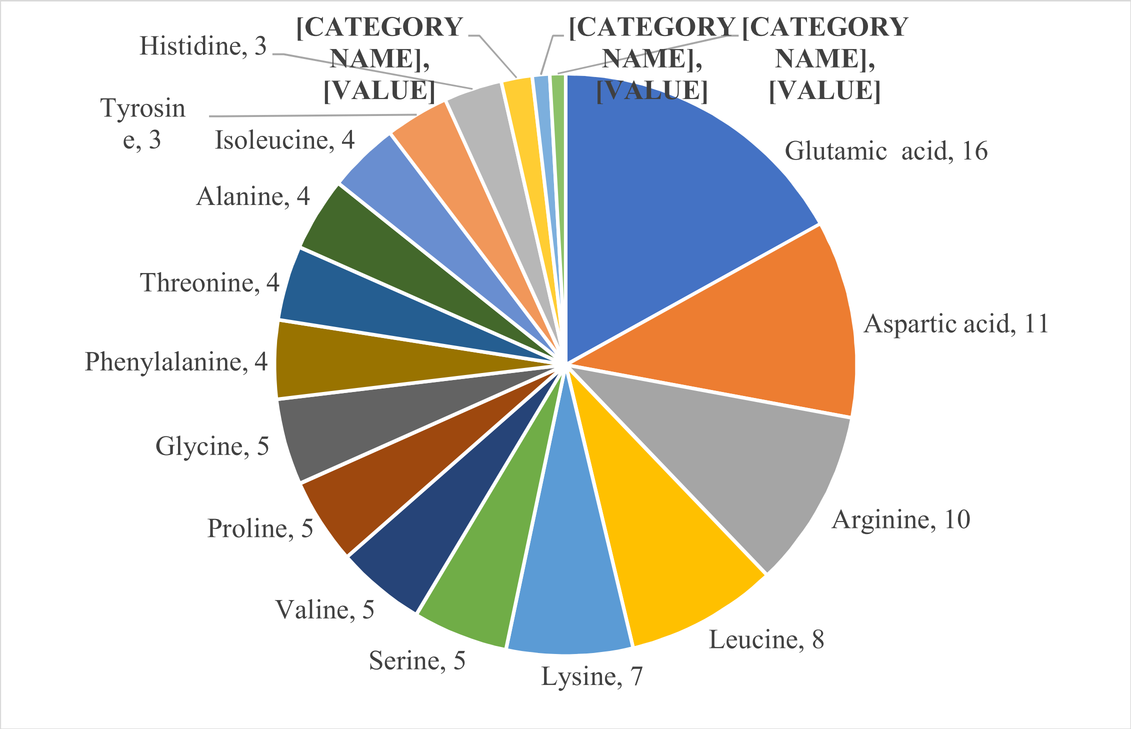
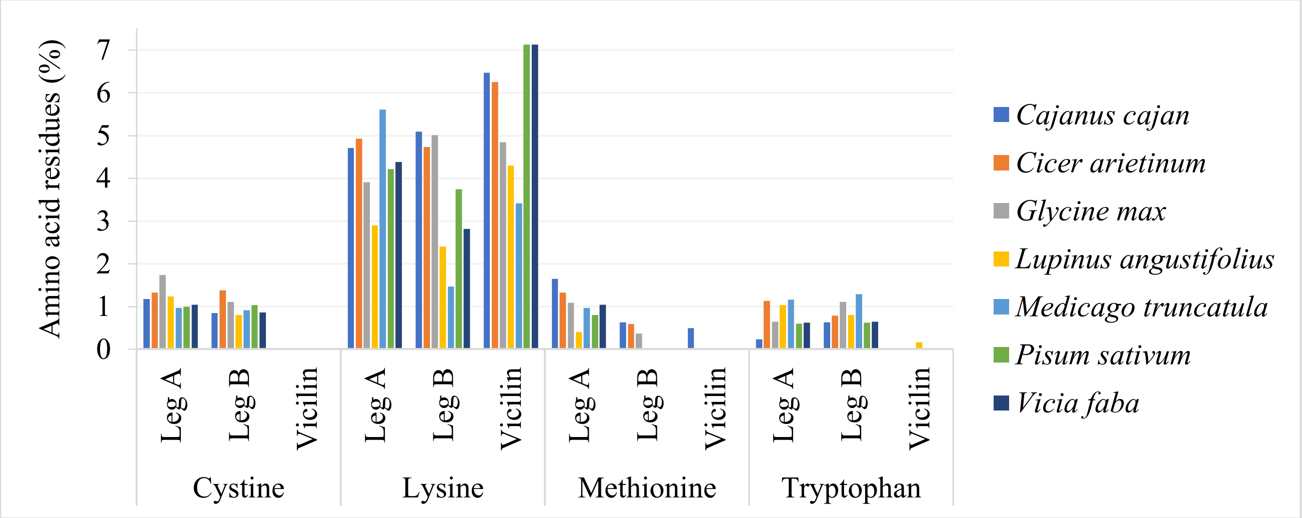
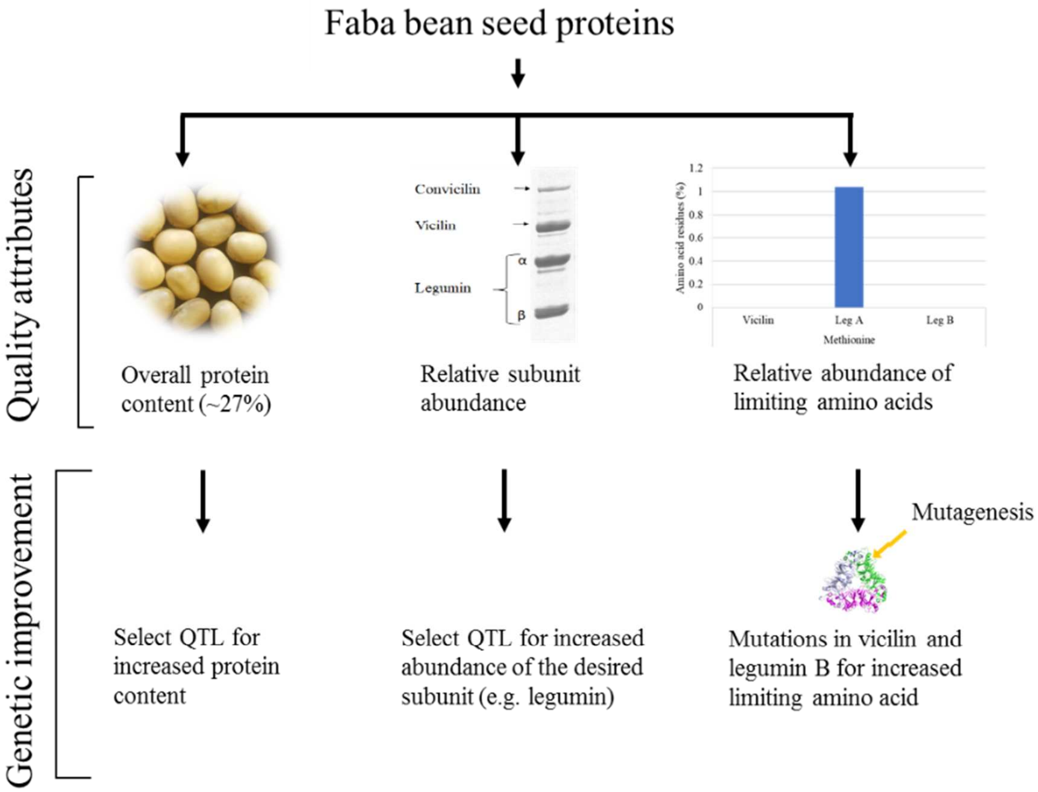

[CATEGORY NAME], [VALUE]
[CATEGORY NAME], [VALUE]
[CATEGORY NAME], [VALUE]
Histidine, 3
Tyrosine, 3
Isoleucine, 4
Alanine, 4
Threonine, 4
Phenylalanine, 4
Glycine, 5
Proline, 5
Valine, 5
Serine, 5
Lysine, 7
Leucine, 8
Arginine, 10
Aspartic acid, 11
Glutamic acid, 16

Figure 5


Amino acid residues (%)
0
1
2
3
4
5
6
7
Leg A
Leg B
Vicilin
Cystine
Lysine
Methionine
Tryptophan
Cajanus cajan
Cicer arietinum
Glycine max
Lupinus angustifolius
Medicago truncatula
Pisum sativum
Vicia faba

TOC Graphic


Faba bean seed proteins
Quality attributes
Overall protein content (~27%)
Convicilin
Vicilin
Legumin
α
β
Relative subunit abundance
1.2
1
0.8
0.6
0.4
0.2
0
Amino acid residues (%)
Vicilin
Leg A
Leg B
Methionine
Relative abundance of limiting amino acids
Genetic improvement
Mutagenesis
Select QTL for increased protein content
Select QTL for increased abundance of the desired subunit (e.g. legumin)
Mutations in vicilin and legumin B for increased limiting amino acid