

# *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. ORCID: <https://orcid.org/0000-0003-4889-056X> and Tosi, P. ORCID: <https://orcid.org/0000-0003-4171-6120> (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 <https://centaur.reading.ac.uk/80856/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

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

Publisher: American Chemical Society

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

[www.reading.ac.uk/centaur](http://www.reading.ac.uk/centaur)

**CentAUR**

Central Archive at the University of Reading

Reading's research outputs online

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

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

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

6 **Correspondence:** Ahmed Omar Warsame

7 Phone: +44(0)1183785473

8 Email: [A.OmarWarsame@pgr.reading.ac.uk](mailto:A.OmarWarsame@pgr.reading.ac.uk)

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26 **ABSTRACT**

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

38

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

41

42

43

44

45

46

47

48

49

50

## 51 INTRODUCTION

### 52 Faba bean production and utilization

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

66 Seeds of *Vf* contain on average about 27% protein<sup>9-11</sup> which provides affordable nutrition for  
67 millions of people around the world, hence its denomination as “the poor man’s meat”. While  
68 *Vf* has been traditionally utilized as dry grain for human consumption in developing  
69 countries, there is growing interest from food industries in developed countries to exploit its  
70 protein for the production of protein-rich vegan/vegetarian snacks,<sup>12</sup> the fortification of  
71 cereal-based food products such as bread and pasta without significantly affecting their  
72 structural and sensory quality<sup>13-14</sup>, or even the production of wholly *Vf*-based bread and  
73 pasta products<sup>15</sup>. *Vf* also represents as significant resource for agro-ecosystem sustainability  
74 and provision of feed for the growing global livestock inventory. Overall, the global  
75 production area for *Vf* has been increasing in the last two decades (Figure S1A) and a recent

76 meta-analysis of yield data from 39 legume species indicated that, in the right environment,  
77 *Vf* can be the highest yielding grain legume <sup>16</sup>. *Vf* also has a high capacity for biological  
78 nitrogen fixation, to the extent that the amount of N fixed by *Vf* alone was estimated to be  
79 comparable to that of soybean and pea combined <sup>17</sup>. For further details on the role of *Vf* on  
80 sustainable cropping systems, readers are referred to Jensen, et al. <sup>18</sup>, Köpke and Nemecek  
81 <sup>19</sup>. On the other hand, *Vf* is yet to be fully exploited as a feedstock for animal production due  
82 to presence of some anti-nutrients which limit its optimal inclusion ratio <sup>20-22</sup>. Removal of  
83 these anti-nutrients through the development of new low anti-nutrient cultivars or using  
84 simple processing techniques like fermentation <sup>13-14</sup> would make this crop a valuable protein  
85 resource for the animal production industry.

#### 86 **Faba bean as a sustainable global protein resource**

87 One of the greatest challenges in the 21<sup>st</sup> century is feeding the growing world population  
88 which it has been estimated may necessitate a 70% increase in food production by 2050 <sup>4</sup>.  
89 More than 30% of this increase has to be made via the production of protein-rich foods <sup>1</sup> to  
90 meet the expected rise in demands due to population growth, increased urbanization and  
91 improved incomes in many parts of the world <sup>1, 23-25</sup>. Protein is a critical nutrient required in  
92 large quantity by humans (~ 50 g protein per adult per day) to maintain normal body function  
93 <sup>26</sup> However, about one-third of the world population, mainly in Asia, Africa and Latin  
94 America, suffers from inadequate intake of proteins, vitamins and minerals <sup>27</sup>. On the other  
95 hand, in higher income countries, where daily animal-based protein intake is already high <sup>1</sup>,  
96 <sup>25</sup>, continued provision of nutritious feeds for the intensive animal production industry will  
97 pose a major challenge in the future. In particular, the livestock production sector in soybean  
98 non-producing countries will be burdened by the high price of imported soybean and soybean  
99 meal. For instance, EU countries have huge deficit in protein-rich feeds with nearly 70%  
100 being imported <sup>28</sup>. *Vf* is well-adapted to European climates, as testified by the high yields

101 recorded in this continent for this legume (Figure S1B), and it therefore has the potential to  
102 contribute to bridging the gap in animal feed self-sufficiency as part of the EU's policies to  
103 increase protein production from locally grown crops <sup>28</sup>. *Vf* is also a candidate crop to meet  
104 the protein demands of an emerging consumer category, particularly in developed economies,  
105 who are opting for animal meat free life style. For example, Statista <sup>29</sup> reported that 13% of  
106 European citizens would consider avoiding red meat while nearly 50% of the respondents in  
107 another study were willing to replace meat with other sources of proteins <sup>30</sup>.

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

#### 114 **Nutritional constraints to *Vf* utilization**

115 The main determinants of *Vf* utilization for human food and animal feed include: (i) protein  
116 concentration, (ii) protein quality, defined mainly by the content of sulfur-containing amino  
117 acids (S-AA) cysteine and methionine, and (iii) concentration of antinutrients in the seeds <sup>5</sup>.  
118 Protein concentration of *Vf*, although it can vary greatly between different genotypes (19-39  
119 %) <sup>31-33</sup>, is one of the highest among legumes. However, commercial varieties on the UK  
120 market contain about 27% protein on average, which is still far less than the protein density  
121 of soya meal, and so, further improvements in protein content is required in order for faba  
122 bean to displace imported soya in animal feed. The proportion of S-AA in the protein is  
123 another crucial quality criterion, particularly in animal feeding. However, like most plant  
124 proteins, *Vf* is poor in certain essential amino acids, namely methionine, cysteine and  
125 tryptophan <sup>5</sup>. Though relatively narrow, the genetic variation for the S-AA reported in *Vf*

126 indicates possibility of improving its nutritional quality. So far, the major breeding objectives  
127 for *Vf*, have been the reduction or removal of vicine and convicine (V-C) and tannins: V-C  
128 causes favism in humans and have deleterious effects on animals<sup>34-35</sup> while tannins lower  
129 protein digestibility<sup>10</sup>. Although these compounds can be removed by processing techniques  
130<sup>36-37</sup>, the most effective approach is probably removing them by breeding. This is now  
131 feasible with the availability of molecular markers closely linked to the V-C locus<sup>38</sup> and zero  
132 tannin gene (*zt-1*)<sup>39</sup>. Furthermore, the reduction of less significant antinutrients such as  
133 trypsin inhibitors, lectins and phytates would improve the nutritional value of *Vf* based feed  
134 products.

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



## 151 SEED STORAGE PROTEINS OF FABA BEAN

152 The major storage proteins of legumes are mainly enzymatically inactive proteins deposited  
153 in seed cotyledons which provide nutrients needed for seed germination and seedling growth  
154 and development<sup>42-43</sup>. Certain seed proteins in legumes including albumins and trypsin  
155 inhibitors, however, have been identified as antinutritional or allergenic agents and therefore  
156 are targeted for removal in breeding programs<sup>44</sup>. Seed storage proteins are classified  
157 according to the system developed by TB Osborne which is based on their solubility in  
158 different solvents<sup>43</sup>. Albumins and globulins are the major storage proteins of legumes and  
159 are soluble in water and saline solutions, respectively. Globulins alone constitute more than  
160 80% of total seed protein in *Vf*<sup>45</sup> and they are further classified based on their sedimentation  
161 coefficients into vicilin-type (7S) and legumin-type (11S)<sup>43</sup>. Both globulin proteins are found  
162 in nearly all legumes, but their denotations vary across species. For instance, globulins of *Vf*  
163 and pea are often referred as vicilin/convicilin and legumin while they are denoted as  
164 conglycinin and glycinin in soybean,  $\beta$  and  $\alpha$  conglutins in lupin, while phaseolin (a vicilin-  
165 like protein) is the only major globulin in common beans. Furthermore, decades of research  
166 on legume storage proteins have produced a sufficient database of annotated SDS-PAGE  
167 images of various species which facilitates faster identification of major globulin bands  
168 without the need for conducting tedious immunoblotting or HPLC procedures. When  
169 extracted under reducing conditions, the salt soluble fraction of legume seed proteins can be  
170 separated on SDS-PAGE into distinct bands which, based on their molecular weights, are  
171 identified as: convicilin ( $M_r \sim 60$  kDa), vicilin ( $M_r \sim 46-55$  kDa) and two major legumin  
172 subunits ( $M_r \sim 38-40$  and 23 kDa) (Table 1)

173 Legumin and vicilin share notable sequence and structural homology and are believed to  
174 originate from a common ancestral gene<sup>46</sup>. Mature legumin is hexameric with a mass of  
175 about 330 kDa<sup>45</sup> and is composed of two trimeric subunits (legumin A and B) while vicilin is

176 a trimeric protein formed by the assembly of three monomers (Figure 2). In contrast to  
177 legumin, vicilin lacks cysteine and is usually glycosylated in its C-terminus<sup>46</sup>. These  
178 structural variations may result in differences in the physiochemical properties of seed  
179 storage proteins which in turn determine their nutritional value and utilization. For instance,  
180 legumin and vicilin differ in their thermal properties<sup>47-48</sup>, affinity to bind flavor compounds  
181 under varying pH conditions<sup>49</sup> and emulsifying ability<sup>48</sup>. Therefore, from a breeding point of  
182 view, legumin/vicilin ratio could be manipulated to meet certain end-user requirements for  
183 protein functionality.

#### 184 **Structure and composition of *Vf* globulins**

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

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

208 The concentration of S-AA is strongly related to the relative proportions of S-AA rich  
209 proteins in the seeds. In *Vf* and other legumes, it is well accepted that legumins contain  
210 relatively higher S-AA compared to vicilin <sup>42, 44, 57-58</sup>. This is further confirmed by  
211 comparative analysis of coding sequences of vicilin and legumin subunits across legume  
212 species which clearly show that legumin subunits contain more residues of cysteine and  
213 methionine (Figure 4). This observation leads to the hypothesis that increasing the proportion  
214 of legumin subunits relative to vicilin would improve nutritional content of plant proteins.  
215 However, considering that vicilin is accumulated in legume seeds earlier than legumin <sup>59-61</sup>,  
216 their ratios could be easily offset by the prevailing environmental conditions, e.g. soil  
217 nutritional status and onset of biotic and abiotic stresses during the plant growth, and in  
218 particular, during grain filling. In contrast to globulins, minor legume seed proteins such as  
219 elongation factor Tu, citrate synthase, albumin 2 (PA2), defensins 1 and 2 and Bowman–Birk  
220 inhibitors (BBI) contain higher S-AA <sup>42, 62</sup>. According to Krishnan, et al. <sup>63</sup>, under higher N  
221 availability through fertilizer application or symbiotic fixation, S-AA containing proteins like  
222 Bowman-Birk protease inhibitor (BBI) were decreased in favour of  $\beta$ -subunits of  $\beta$ -  
223 conglycinins of soybeans. Similarly, ectopic overexpression of *VfAAP1* gene on *P. sativum*  
224 and *V. narbonensis* resulted in 30% increase in the globulin fraction but no significant effect

225 on albumin, a S-AA rich protein<sup>64</sup>. Hence, it would appear that the negative correlation  
226 between high protein and S-AA content in *Vf*<sup>11, 32, 65</sup> may be the result of preferential  
227 accumulation of low nutritional quality protein fractions in higher protein lines.

### 228 **Genetic control of globulins**

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

243 There is considerable homology between *Vf* globulin subunits and those of other legumes  
244 (Table S1), and where genome sequences are available, it is now possible to classify and  
245 associate seed storage subunits to specific genome locations (Table S2). Considering the lack  
246 of genome sequence for *Vf*, this information is critical for synteny-based mapping of globulin  
247 genes and QTLs. For instance, in *M. truncatula*, several genomic regions coding for globulins  
248 have been mapped on chromosome I and VII<sup>72</sup> which are notably syntenic with *Vf*

249 chromosome III and V<sup>39-40</sup> where legumin A and B genes were previously located,  
250 respectively<sup>66</sup>.

### 251 **Expression of globulin genes**

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

### 267 **Synthesis and accumulation of seed storage proteins**

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

273 DAP) are mainly destined for storage into cotyledons' protein bodies<sup>77</sup>. During the latter  
274 stage, globulin proteins show distinct expression patterns in which vicilin synthesis and  
275 accumulation precedes that of legumin and  $\alpha$  chain polypeptides of legumin appear earlier  
276 than  $\beta$  chains<sup>59</sup>. Similar pattern of vicilin and legumin gene expression has also been  
277 reported in Medicago<sup>78</sup> and soybean<sup>76</sup>.

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

## 295 **GENETIC IMPROVEMENT OF PROTEIN CONTENT AND QUALITY**

### 296 **Summary of the past work**

297 Several studies have focused on the genetic variation for protein content (**Table 2**) and to  
298 what extent protein content was correlated with yield of *Vf*. One study indicated that protein  
299 content was variable between and within varieties (n=33) with broad sense heritability of  
300 0.70 and no significant correlation with seed weight <sup>31</sup>. However, when larger set of  
301 germplasm (n=600) was screened, a clear negative relationship between seed weight and  
302 protein was detected although some large-seeded genotypes with above average protein  
303 content were also found <sup>65</sup>. Similarly, after four cycles of selection for protein content, Sjödin  
304 <sup>32</sup> concluded that protein content in *Vf* could be improved by selection but tended to  
305 negatively correlate with number of seeds per plant regardless of thousand seed weight.  
306 These early efforts also established the variability for S-AA content (**Table 3**) and nearly all  
307 investigations found a negative correlation between protein and S-AA content <sup>32, 65, 84</sup>. Under  
308 circumstances where desirable traits of interest are negatively correlated, deeper  
309 understanding of the genetic basis of the trade-offs between the traits and availability of  
310 appropriate tools to dissect and recombine them is crucial.

### 311 **Areas for future focus**

#### 312 *Uncoupling the negative yield-protein correlation*

313 Correlation between traits can arise due to gene linkage or pleiotropy <sup>85</sup>, with the latter being  
314 most common in plants, and its resolution requires deeper understanding of both traits.  
315 Therefore, several possible mechanisms have been investigated in various crops in order to  
316 unlock protein-yield association. It is hypothesized that the negative correlation between the  
317 two traits result when the high demand for N during seed filling stage coincides with decline  
318 in soil nutrients in the rhizosphere and nitrogen fixation, resulting in re-mobilization of  
319 nitrogen from leaves, which in turn shortens grain filling and reduces seed weights <sup>86</sup>. This is

320 in line with findings by Egle, et al.<sup>87</sup> who showed that majority of N accumulated during  
321 seed filling in barley was remobilized from leaves and stems, but that ongoing N uptake  
322 could also contribute. Furthermore, wheat genotypes with higher capability for post-anthesis  
323 N uptake deviate from grain-protein negative relationship<sup>88-89</sup> and selection for this trait has  
324 been therefore proposed as a possible criterion for simultaneous improvement of protein  
325 content and grain yield. The genetic basis of post-flowering N uptake is not yet fully  
326 understood either in cereals or in legumes but could be related to root structure and/or N  
327 transport capacity. For instance, pea genotypes with higher mineral nitrogen absorption and  
328 symbiotic nitrogen fixation have shown enhanced seed N content and yield<sup>83</sup>. Moreover,  
329 faster rate and relatively longer duration of N accumulation during seed development has  
330 been reported as a possible mechanism for combining high protein and large seed size in  
331 soybean<sup>90</sup>. The importance of N uptake capacity for protein content and yield was further  
332 demonstrated by Peng, et al.<sup>81</sup> who found major protein content QTL *qPCI* harboring a  
333 putative amino acid transporter gene (*OsAAP6*), which they proposed as candidate QTL for  
334 simultaneous selection for yield and protein content in rice. These areas of enquiry are  
335 amenable for further investigation and can potentially point to QTLs that can be used to  
336 improve protein content in *Vf* without significant yield reduction.

### 337 *Improving S-AA content by modifying legumin: vicilin ratio*

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



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

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

370 linkage map may offer an unprecedented opportunity to discover novel QTLs that could  
371 represent targets for improving nutritional quality.

### 372 *Exploiting mutagenesis approaches*

373 Large-scale mutagenesis using physical or chemical mutagenic agents is a well-established  
374 method of inducing novel variation to meet human requirements, but which is unlikely to be  
375 present in nature. This approach is all the more justified in the case of *Vf* where the primary  
376 gene pool lacks any known wild relatives. Indeed, several mutagenesis efforts have produced  
377 new sets of morphological phenotypes in *Vf*<sup>40, 97-98</sup>. However, no data is available on  
378 potential beneficial mutations in the seed composition of *Vf*. Although Sjödin<sup>97</sup> has reported  
379 to have identified some high protein content genotypes from a lot of seeds which had been  
380 mutagenized he could not ascertain whether the selected plants were genuine mutants or  
381 randomly isolated extremes in the original seed lot. There are several potential ways of  
382 exploiting induced mutations for improving protein content and/or quality. First, desirable  
383 mutations involving photosynthetic and N provision mechanisms can improve protein  
384 content. From ethyl methane sulfonate (EMS) mutagenized seeds, Duc<sup>98</sup> discovered a  
385 supernodulating line with 3-4 times higher number of nodules compared to the parental line.  
386 Considering the close relationship between N fixation and protein content, such a trait could  
387 be exploited in breeding programs. Secondly, knockdown/knockout or regulatory mutations  
388 leading to absence of major protein subunits such as vicilins can result in improved  
389 nutritional quality by increasing the ratio of S-AA rich subunits like legumin and albumins.  
390 Such mutations could be *cis*-linked to the structural loci themselves or *trans*-acting factors  
391 that would need to be mapped *de novo*. For instance, mutants of *PsABI5*, a major *trans*-acting  
392 regulator of vicilin abundance in pea, have shown an increased legumin abundance<sup>72</sup>.  
393 Thirdly, presence or absence of certain subunits can enable dissection of genetic control of  
394 individual protein subunits via a QTL mapping approach<sup>55</sup>. Lastly, it is possible via a reverse

395 genetic screen to select non-synonymous mutations that convert non-S-AA residues to S-AA  
396 residues in S-AA poor storage proteins such as vicilins, although, only a proportion of codons  
397 are available for single base changes that would result in this outcome. Moreover, the  
398 physico-chemical properties of cysteine (disulfide bridge-forming) and methionine  
399 (hydrophobic) may cause steric constraints<sup>99</sup>. However even a single well-placed additional  
400 methionine in each vicilin could give rise to a significant step up in S-AA levels and this  
401 approach is therefore worth trying. On a more practical level, full exploitation of mutagenesis  
402 for the above purposes requires high-throughput and cheap phenotyping methods to screen  
403 tens of thousands of plants for nutritional and agronomic traits.

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

420 the above should become ever more efficient, enhancing the prospects of increasing protein  
421 content and quality in this strategic crop.

#### 422 **ABBREVIATIONS USED**

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

#### 428 **ACKNOWLEDGEMENTS**

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

431

#### 432 **SUPPORTING INFORMATION**

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

#### 435 **CONFLICT OF INTEREST**

436 The authors declare no competing financial interest.

#### 437 **REFERENCES**

- 438 (1) Henchion, M.; Hayes, M.; Mullen, A.; Fenelon, M.; Tiwari, B., Future Protein Supply  
439 and Demand: Strategies and Factors Influencing a Sustainable Equilibrium. *Foods* **2017**, *6*  
440 (7), 53.
- 441 (2) Young, V. R.; Pellett, P. L., Plant proteins in relation to human protein and amino  
442 acid nutrition. *The American Journal of Clinical Nutrition* **1994**, *59* (5), 1203S-1212S.
- 443 (3) Smýkal, P.; Coyne, C. J.; Ambrose, M. J.; Maxted, N.; Schaefer, H.; Blair, M. W.;  
444 Berger, J.; Greene, S. L.; Nelson, M. N.; Besharat, N.; Vymyslický, T.; Toker, C.; Saxena, R.  
445 K.; Roorkiwal, M.; Pandey, M. K.; Hu, J.; Li, Y. H.; Wang, L. X.; Guo, Y.; Qiu, L. J.;  
446 Redden, R. J.; Varshney, R. K., Legume Crops Phylogeny and Genetic Diversity for Science  
447 and Breeding. *Crit. Rev. Plant Sci.* **2015**, *34* (1-3), 43-104.

- 448 (4) Foyer, C. H.; Lam, H. M.; Nguyen, H. T.; Siddique, K. H.; Varshney, R. K.; Colmer,  
449 T. D.; Cowling, W.; Bramley, H.; Mori, T. A.; Hodgson, J. M.; Cooper, J. W.; Miller, A. J.;  
450 Kunert, K.; Vorster, J.; Cullis, C.; Ozga, J. A.; Wahlqvist, M. L.; Liang, Y.; Shou, H.; Shi,  
451 K.; Yu, J.; Fodor, N.; Kaiser, B. N.; Wong, F. L.; Valliyodan, B.; Considine, M. J.,  
452 Neglecting legumes has compromised human health and sustainable food production. *Nat*  
453 *Plants* **2016**, *2*, 16112.
- 454 (5) Duc, G., Faba bean (*Vicia faba* L). *Field Crops Res.* **1997**, *53*, 99-109.
- 455 (6) Tanno, K.-i.; Willcox, G., The origins of cultivation of *Cicer arietinum* L. and *Vicia*  
456 *faba* L.: early finds from Tell el-Kerkh, north-west Syria, late 10th millennium b.p.  
457 *Vegetation History and Archaeobotany* **2006**, *15* (3), 197-204.
- 458 (7) Cubero, J. I., On the evolution of *Vicia faba* L. *Theor. Appl. Genet.* **1974**, *45* (2), 47-  
459 51.
- 460 (8) FAOstat United Nations Organization for Food and Agriculture. (accessed 20 May).
- 461 (9) Duc, G.; Aleksić, J. M.; Marget, P.; Mikic, A.; Paull, J.; Redden, R. J.; Sass, O.;  
462 Stoddard, F. L.; Vandenberg, A.; Vishnyakova, M.; Torres, A. M., Faba Bean. In *Grain*  
463 *Legumes*, Ron, A. M. D., Ed. Springer Science+Business Media: New York, 2015; Vol. 10,  
464 pp 141-178.
- 465 (10) Makkar, H. P. S.; Becker, K.; Abel, H.; Pawelzik, E., Nutrient contents, rumen protein  
466 degradability and antinutritional factors in some colour- and white-flowering cultivars of  
467 *Vicia faba* beans. *J. Sci. Food Agric.* **1997**, *75* (4), 511-520.
- 468 (11) Schumacher, H.; Paulsen, H. M.; Gau, A. E.; Link, W.; Jurgens, H. U.; Sass, O.;  
469 Dieterich, R., Seed protein amino acid composition of important local grain legumes *Lupinus*  
470 *angustifolius* L., *Lupinus luteus* L., *Pisum sativum* L. and *Vicia faba* L. *Plant Breeding* **2011**,  
471 *130* (2), 156-164.
- 472 (12) Kaskinen, T.; Lähteenoja, S.; Sokero, M.; Suomela, I., Strategic Business Examples  
473 from Finland: The Growth of the Startup Industry. In *Factor X: Challenges, Implementation*  
474 *Strategies and Examples for a Sustainable Use of Natural Resources*, Lehmann, H., Ed.  
475 Springer International Publishing: Cham, 2018; pp 325-333.
- 476 (13) Coda, R.; Varis, J.; Verni, M.; Rizzello, C. G.; Katina, K., Improvement of the protein  
477 quality of wheat bread through faba bean sourdough addition. *LWT - Food Science and*  
478 *Technology* **2017**, *82* (Supplement C), 296-302.
- 479 (14) Rizzello, C. G.; Verni, M.; Koivula, H.; Montemurro, M.; Seppa, L.; Kemell, M.;  
480 Katina, K.; Coda, R.; Gobetti, M., Influence of fermented faba bean flour on the nutritional,  
481 technological and sensory quality of fortified pasta. *Food Funct.* **2017**, *8* (2), 860-871.
- 482 (15) VTT Gluten-free faba bean for bread and pasta. (accessed 25 July).
- 483 (16) Cernay, C.; Pelzer, E.; Makowski, D., A global experimental dataset for assessing  
484 grain legume production. *Scientific Data* **2016**, *3*, 160084.
- 485 (17) Baddeley, J.; Jones, S.; Topp, C.; Watson, C.; Helming, J.; Stoddard, F. Biological  
486 nitrogen fixation (BNF) by legume crops in Europe. [www.legumefutures.de](http://www.legumefutures.de) (accessed  
487 January 4).
- 488 (18) Jensen, E. S.; Peoples, M. B.; Hauggaard-Nielsen, H., Faba bean in cropping systems.  
489 *Field Crops Res.* **2010**, *115* (3), 203-216.
- 490 (19) Köpke, U.; Nemecek, T., Ecological services of faba bean. *Field Crops Res.* **2010**,  
491 *115* (3), 217-233.
- 492 (20) Perez-Maldonado, R. A.; Mannion, P. F.; Farrell, D. J., Optimum inclusion of field  
493 peas, faba beans, chick peas and sweet lupins in poultry diets. I. Chemical composition and  
494 layer experiments. *British Poultry Science* **1999**, *40* (5), 667-673.
- 495 (21) Koivunen, E.; Tuunainen, P.; Valkonen, E.; Rossow, L.; Valaja, J., Use of faba beans  
496 (*Vicia faba* L.) in diets of laying hens. *Agricultural And Food Science* **2014** *23*, 165-172.

- 497 (22) Lessire, M.; Gallo, V.; Prato, M.; Akide-Ndunge, O.; Mandili, G.; Marget, P.; Arese,  
498 P.; Duc, G., Effects of faba beans with different concentrations of vicine and convicine on  
499 egg production, egg quality and red blood cells in laying hens. *Animal* **2016**, 1-9.
- 500 (23) Kawashima, H.; Bazin, M. J.; Lynch, J. M., A modelling study of world protein  
501 supply and nitrogen fertilizer demand in the 21st century. *Environ. Conserv.* **2002**, 24 (1), 50-  
502 57.
- 503 (24) Speedy, A. W. In *Overview of world feed protein needs and supply*, FAO Animal  
504 Production and Health Proceedings (FAO), FAO: 2004.
- 505 (25) Chiari, N., Food Security. The Challenge of Nutrition in the New Century. *Relations.*  
506 *Beyond Anthropocentrism* **2017**, 5 (2), 145-156.
- 507 (26) WHO/FAO/UNU, Protein and amino acid requirements in human nutrition. *WHO*  
508 *Tech. Rep. Ser.* **2007**, (935), 1-265, back cover.
- 509 (27) Balyan, H. S.; Gupta, P. K.; Kumar, S.; Dhariwal, R.; Jaiswal, V.; Tyagi, S.; Agarwal,  
510 P.; Gahlaut, V.; Kumari, S., Genetic improvement of grain protein content and other health-  
511 related constituents of wheat grain. *Plant Breeding* **2013**, 132 (5), 446-457.
- 512 (28) de Visser, C. L. M.; Schreuder, R.; Stoddard, F., The EU's dependency on soya bean  
513 import for the animal feed industry and potential for EU produced alternatives. *OCL* **2014**, 21  
514 (4), D407.
- 515 (29) Statista Meat consumption and vegetarianism in Europe - Statistics and Facts.  
516 <https://www.statista.com/topics/3345/meat-consumption-and-vegetarianism-in-europe/>  
517 (accessed December 25).
- 518 (30) de Boer, J.; Aiking, H., Prospects for pro-environmental protein consumption in  
519 Europe: Cultural, culinary, economic and psychological factors. *Appetite* **2018**, 121, 29-40.
- 520 (31) Griffiths, D. W.; Lawes, D. A., Variation in the crude protein content of field beans  
521 (*Vicia faba* L.) in relation to the possible improvement of the protein content of the crop.  
522 *Euphytica* **1978**, 27 (2), 487-495.
- 523 (32) Sjödin, J., Protein Quantity and Quality in *Vicia Faba*. In *Faba Bean Improvement:*  
524 *Proceedings of the Faba Bean Conference held in Cairo, Egypt, March 7-11, 1981*, Hawtin,  
525 G.; Webb, C., Eds. Springer Netherlands: Dordrecht, 1982; pp 319-331.
- 526 (33) Frauen, M.; Röbbelen, G.; Ebrmeyer, E., Quantitative Measurement of Quality  
527 Determining Constituents in Seeds of Different Inbred Lines from A World Collection of  
528 *Vicia Faba*. In *Vicia faba: Agronomy, Physiology and Breeding*, Hebblethwaite, P. D.;  
529 Dawkins, T. C. K.; Heath, M. C.; Lockwood, G., Eds. Springer-Science+Business Media,  
530 B.V.: Brussels-Luxembourg, 1984; Vol. 10, pp 279-287.
- 531 (34) Crépon, K.; Marget, P.; Peyronnet, C.; Carrouée, B.; Arese, P.; Duc, G., Nutritional  
532 value of faba bean (*Vicia faba* L.) seeds for feed and food. *Field Crops Res.* **2010**, 115 (3),  
533 329-339.
- 534 (35) Yu, E.-M.; Zhang, H.-F.; Li, Z.-F.; Wang, G.-J.; Wu, H.-K.; Xie, J.; Yu, D.-G.; Xia,  
535 Y.; Zhang, K.; Gong, W.-B., Proteomic signature of muscle fibre hyperplasia in response to  
536 faba bean intake in grass carp. **2017**, 7, 45950.
- 537 (36) Rizzello, C. G.; Losito, I.; Facchini, L.; Katina, K.; Palmisano, F.; Gobbetti, M.;  
538 Coda, R., Degradation of vicine, convicine and their aglycones during fermentation of faba  
539 bean flour. *Sci Rep* **2016**, 6, 32452.
- 540 (37) Coda, R.; Melama, L.; Rizzello, C. G.; Curiel, J. A.; Sibakov, J.; Holopainen, U.;  
541 Pulkkinen, M.; Sozer, N., Effect of air classification and fermentation by *Lactobacillus*  
542 *plantarum* VTT E-133328 on faba bean (*Vicia faba* L.) flour nutritional properties. *Int. J.*  
543 *Food Microbiol.* **2015**, 193, 34-42.
- 544 (38) Khazaei, H.; Purves, R. W.; Song, M.; Stonehouse, R.; Bett, K. E.; Stoddard, F. L.;  
545 Vandenberg, A., Development and validation of a robust, breeder-friendly molecular marker  
546 for the vc-locus in faba bean. *Mol. Breed.* **2017**, 37 (11), 140.

- 547 (39) Webb, A.; Cottage, A.; Wood, T.; Khamassi, K.; Hobbs, D.; Gostkiewicz, K.; White,  
548 M.; Khazaei, H.; Ali, M.; Street, D.; Duc, G.; Stoddard, F. L.; Maalouf, F.; Ogonnaya, F. C.;  
549 Link, W.; Thomas, J.; O'Sullivan, D. M., A SNP-based consensus genetic map for synteny-  
550 based trait targeting in faba bean (*Vicia faba* L.). *Plant Biotechnol. J.* **2016**, *14* (1), 177-85.
- 551 (40) O'Sullivan, D. M.; Angra, D., Advances in Faba Bean Genetics and Genomics. *Front*  
552 *Genet* **2016**, *7*, 150.
- 553 (41) Patil, G.; Mian, R.; Vuong, T.; Pantalone, V.; Song, Q.; Chen, P.; Shannon, G. J.;  
554 Carter, T. C.; Nguyen, H. T., Molecular mapping and genomics of soybean seed protein: a  
555 review and perspective for the future. *Theor. Appl. Genet.* **2017**, *130* (10), 1975-1991.
- 556 (42) Liu, Y.; Wu, X.; Hou, W.; Li, P.; Sha, W.; Tian, Y., Structure and function of seed  
557 storage proteins in faba bean (*Vicia faba* L.). *3 Biotech* **2017**, *7* (1), 74.
- 558 (43) Shewry, P. R.; Casey, R., Seed Proteins. In *Seed Proteins*, Shewry, P. R.; Casey, R.,  
559 Eds. Springer Netherlands: Dordrecht, 1999; pp 1-10.
- 560 (44) Joshi, J.; Pandurangan, S.; Diapari, M.; Marsolais, F., Comparison of Gene Families:  
561 Seed Storage and Other Seed Proteins. In *The Common Bean Genome*, Pérez de la Vega, M.;  
562 Santalla, M.; Marsolais, F., Eds. Springer International Publishing: Cham, 2017; pp 201-217.
- 563 (45) Müntz, K.; Horstmann, C.; Schlesier, B., Vicia globulins. In *Seed Proteins*, Shewry,  
564 P. R.; Casey, R., Eds. Springer Netherlands: Dordrecht, 1999; pp 259-284.
- 565 (46) Kesari, P.; Sharma, A.; Katiki, M.; Kumar, P.; R Gurjar, B.; Tomar, S.; K Sharma, A.;  
566 Kumar, P., Structural, Functional and Evolutionary Aspects of Seed Globulins. *Protein and*  
567 *peptide letters* **2017**, *24* (3), 267-277.
- 568 (47) Meng, G. T.; Ma, C. Y., Thermal properties of Phaseolus angularis (red bean)  
569 globulin. *Food Chem.* **2001**, *73* (4), 453-460.
- 570 (48) Kimura, A.; Fukuda, T.; Zhang, M.; Motoyama, S.; Maruyama, N.; Utsumi, S.,  
571 Comparison of Physicochemical Properties of 7S and 11S Globulins from Pea, Fava Bean,  
572 Cowpea, and French Bean with Those of Soybean—French Bean 7S Globulin Exhibits  
573 Excellent Properties. *J. Agric. Food Chem.* **2008**, *56* (21), 10273-10279.
- 574 (49) Heng, L.; van Koningsveld, G. A.; Gruppen, H.; van Boekel, M. A. J. S.; Vincken, J.  
575 P.; Roozen, J. P.; Voragen, A. G. J., Protein-flavour interactions in relation to development  
576 of novel protein foods. *Trends Food Sci. Technol.* **2004**, *15* (3), 217-224.
- 577 (50) Baumlein, H.; Wobus, U.; Pustell, J.; Kafatos, F. C., The legumin gene family:  
578 structure of a B type gene of *Vicia faba* and a possible legumin gene specific regulatory  
579 element. *Nucleic Acids Res.* **1986**, *14* (6), 2707-20.
- 580 (51) Tucci, M.; Capparelli, R.; Costa, A.; Rao, R., Molecular heterogeneity and genetics of  
581 *Vicia faba* seed storage proteins. *Theor. Appl. Genet.* **1991**, *81* (1), 50-58.
- 582 (52) Le Signor, C.; Gallardo, K.; Prospero, J. M.; Salon, C.; Quillien, L.; Thompson, R.;  
583 Duc, G., Genetic diversity for seed protein composition in *Medicago truncatula*. *Plant*  
584 *Genetic Resources* **2005**, *3* (1), 59-71.
- 585 (53) O'Kane, F. E.; Happe, R. P.; Vereijken, J. M.; Gruppen, H.; van Boekel, M. A. J. S.,  
586 Characterization of Pea Vicilin. 1. Denoting Convicilin as the  $\alpha$ -Subunit of the *Pisum* Vicilin  
587 Family. *J. Agric. Food Chem.* **2004**, *52* (10), 3141-3148.
- 588 (54) Krishnan, H. B.; Natarajan, S. S.; Oehrle, N. W.; Garrett, W. M.; Darwish, O.,  
589 Proteomic Analysis of Pigeonpea (*Cajanus cajan*) Seeds Reveals the Accumulation of  
590 Numerous Stress-Related Proteins. *J. Agric. Food Chem.* **2017**, *65* (23), 4572-4581.
- 591 (55) Boehm, J. D.; Nguyen, V.; Tashiro, R. M.; Anderson, D.; Shi, C.; Wu, X.; Woodrow,  
592 L.; Yu, K.; Cui, Y.; Li, Z., Genetic mapping and validation of the loci controlling 7S  $\alpha'$  and  
593 11S A-type storage protein subunits in soybean [*Glycine max* (L.) Merr.]. *Theor. Appl.*  
594 *Genet.* **2017**, 1-13.
- 595 (56) Galili, G.; Amir, R., Fortifying plants with the essential amino acids lysine and  
596 methionine to improve nutritional quality. *Plant Biotechnol. J.* **2013**, *11* (2), 211-22.

- 597 (57) Kwanyuen, P.; Pantalone, V. R.; Burton, J. W.; Wilson, R. F., A new approach to  
598 genetic alteration of soybean protein composition and quality. *J. Am. Oil Chem. Soc.* **1997**, *74*  
599 (8), 983-987.
- 600 (58) Jackson, P.; Boulter, D.; Thurman, D. A., A comparison of some properties of vicilin  
601 and legumin isolated from seeds of *Pisum sativum*, *Vicia faba* and *Cicer arietinum*. *New*  
602 *Phytol.* **1969** *68* 25-33.
- 603 (59) De Pace, C.; Delre, V.; Mugnozsa, G. T. S.; Maggini, E.; Cremonini, R.; Frediani, M.;  
604 Cionini, P. G., Legumin of *Vicia faba* major: accumulation in developing cotyledons,  
605 purification, mRNA characterization and chromosomal location of coding genes. *Theor.*  
606 *Appl. Genet.* **1991**, *83*, 17-23.
- 607 (60) Abirached-Darmency, M.; Dessaint, F.; Benlicha, E.; Schneider, C., Biogenesis of  
608 protein bodies during vicilin accumulation in *Medicago truncatula* immature seeds. *BMC*  
609 *Research Notes* **2012**, *5*, 409-409.
- 610 (61) Gallardo, K.; Le Signor, C.; Vandekerckhove, J.; Thompson, R. D.; Burstin, J.,  
611 Proteomics of *Medicago truncatula* seed development establishes the time frame of diverse  
612 metabolic processes related to reserve accumulation. *Plant Physiol.* **2003**, *133* (2), 664-82.
- 613 (62) Rubio, L. A.; Perez, A.; Ruiz, R.; Guzman, M. A.; Aranda-Olmedo, I.; Clemente, A.,  
614 Characterization of pea (*Pisum sativum*) seed protein fractions. *J. Sci. Food Agric.* **2014**, *94*  
615 (2), 280-7.
- 616 (63) Krishnan, H. B.; Bennett, J. O.; Kim, W.-S.; Krishnan, A. H.; Mawhinney, T. P.,  
617 Nitrogen Lowers the Sulfur Amino Acid Content of Soybean (*Glycine max* [L.] Merr.) by  
618 Regulating the Accumulation of Bowman-Birk Protease Inhibitor. *J. Agric. Food Chem.*  
619 **2005**, *53* (16), 6347-6354.
- 620 (64) Rolletschek, H.; Hosein, F.; Miranda, M.; Heim, U.; Gotz, K. P.; Schlereth, A.;  
621 Borisjuk, L.; Saalbach, I.; Wobus, U.; Weber, H., Ectopic expression of an amino acid  
622 transporter (VfAAP1) in seeds of *Vicia narbonensis* and pea increases storage proteins. *Plant*  
623 *Physiol.* **2005**, *137* (4), 1236-49.
- 624 (65) Lafiandra, D.; Polignano, G. B.; Filippetti, A.; Porceddu, E., Genetic variability for  
625 protein content and S-aminoacids in broad-beans (*Vicia faba* L.). *Die Kulturpflanze* **1981**, *29*  
626 (1), 115-127.
- 627 (66) Fuchs, J.; Schubert, I., Localization of seed protein genes on metaphase chromosomes  
628 of *Vicia faba* via fluorescence in situ hybridization. *Chromosome Research* **1995**, *3* (2), 94-  
629 100.
- 630 (67) Rerie, W. G.; Whitecross, M.; Higgins, T. J. V., Developmental and environmental  
631 regulation of pea legumin genes in transgenic tobacco. *Molecular and General Genetics*  
632 *MGG* **1991**, *225* (1), 148-157.
- 633 (68) Heim, U.; Schubert, R.; Baumlein, H.; Wobus, U., The legumin gene family: structure  
634 and evolutionary implications of *Vicia faba* B-type genes and pseudogenes. *Plant Mol. Biol.*  
635 **1989**, *13* (6), 653-63.
- 636 (69) Weschke, W.; Bassüner, R.; Van Hai, N.; Czihal, A.; Baumlein, H.; Wobus, U., The  
637 structure of a *Vicia faba* vicilin gene. *Biochemie und Physiologie der Pflanzen* **1988**, *183* (2-  
638 3), 233-242.
- 639 (70) Jiri, M.; Winfriede, W.; Helmut, B.; Uta, P.; Andreas, H.; Ulrich, W.; Ingo, S.,  
640 Localization of vicilin genes via polymerase chain reaction on microisolated field bean  
641 chromosomes. *The Plant Journal* **1993**, *3* (6), 883-886.
- 642 (71) Fuchs, J.; Joos, S.; Licheter, P.; Schubert, I., Localization of Vicilin Genes on Field  
643 Bean Chromosome II by Fluorescent in situ Hybridization. *J. Hered.* **1994**, *85* (6), 487-488.
- 644 (72) Le Signor, C.; Aimé, D.; Bordat, A.; Belghazi, M.; Labas, V.; Gouzy, J.; Young, N.  
645 D.; Prospero, J.-M.; Leprince, O.; Thompson, R. D.; Buitink, J.; Burstin, J.; Gallardo, K.,



- 646 Genome-wide association studies with proteomics data reveal genes important for synthesis,  
647 transport and packaging of globulins in legume seeds. *New Phytol.* **2017**, *214* (4), 1597-1613.  
648 (73) Bourgeois, M.; Jacquin, F.; Savoie, V.; Sommerer, N.; Labas, V.; Henry, C.; Burstin,  
649 J., Dissecting the proteome of pea mature seeds reveals the phenotypic plasticity of seed  
650 protein composition. *Proteomics* **2009**, *9* (2), 254-71.  
651 (74) Zhaoming, Q.; Zhanguo, Z.; Zhongyu, W.; Jingyao, Y.; Hongtao, Q.; Xinrui, M.;  
652 Hongwei, J.; Dawei, X.; Zhengong, Y.; Rongsheng, Z.; Chunyan, L.; Wei, Y.; Zhenbang, H.;  
653 Xiaoxia, W.; Jun, L.; Qingshan, C., Meta-analysis and transcriptome profiling reveal hub  
654 genes for soybean seed storage composition during seed development. *Plant, Cell Environ.*  
655 **2018**, *0* (0), 1-19.  
656 (75) Verdier, J.; Thompson, R. D., Transcriptional regulation of storage protein synthesis  
657 during dicotyledon seed filling. *Plant Cell Physiol.* **2008**, *49* (9), 1263-71.  
658 (76) Mori, T.; Maruyama, N.; Nishizawa, K.; Higasa, T.; Yagasaki, K.; Ishimoto, M.;  
659 Utsumi, S., The composition of newly synthesized proteins in the endoplasmic reticulum  
660 determines the transport pathways of soybean seed storage proteins. *The Plant Journal* **2004**,  
661 *40* (2), 238-249.  
662 (77) Panitz, R.; Borisjuk, L.; Manteuffel, R.; Wobus, U., Transient expression of storage-  
663 protein genes during early embryogenesis of *Vicia faba*: synthesis and metabolization of  
664 vicilin and legumin in the embryo, suspensor and endosperm. *Planta* **1995**, *196* (4), 765-774.  
665 (78) Wang, X. D.; Song, Y.; Sheahan, M. B.; Garg, M. L.; Rose, R. J., From embryo sac to  
666 oil and protein bodies: embryo development in the model legume *Medicago truncatula*. *New*  
667 *Phytol.* **2012**, *193* (2), 327-38.  
668 (79) Zhang, L.; Garneau, M. G.; Majumdar, R.; Grant, J.; Tegeder, M., Improvement of  
669 pea biomass and seed productivity by simultaneous increase of phloem and embryo loading  
670 with amino acids. *Plant J.* **2015**, *81* (1), 134-46.  
671 (80) Golombek, S.; Rolletschek, H.; Wobus, U.; Weber, H., Control of storage protein  
672 accumulation during legume seed development. *J. Plant Physiol.* **2001**, *158* (4), 457-464.  
673 (81) Peng, B.; Kong, H.; Li, Y.; Wang, L.; Zhong, M.; Sun, L.; Gao, G.; Zhang, Q.; Luo,  
674 L.; Wang, G.; Xie, W.; Chen, J.; Yao, W.; Peng, Y.; Lei, L.; Lian, X.; Xiao, J.; Xu, C.; Li, X.;  
675 He, Y., OsAAP6 functions as an important regulator of grain protein content and nutritional  
676 quality in rice. *Nature Communications* **2014**, *5*, 4847.  
677 (82) Ramaekers, L.; Galeano, C. H.; Garzón, N.; Vanderleyden, J.; Blair, M. W.,  
678 Identifying quantitative trait loci for symbiotic nitrogen fixation capacity and related traits in  
679 common bean. *Mol. Breed.* **2013**, *31* (1), 163-180.  
680 (83) Bourion, V.; Rizvi, S. M. H.; Fournier, S.; Larambergue, H. d.; Galmiche, F.; Marget,  
681 P.; Duc, G.; Burstin, J., Genetic dissection of nitrogen nutrition in pea through a QTL  
682 approach of root, nodule, and shoot variability. *Theor. Appl. Genet.* **2010**, *121*, 71-86.  
683 (84) Griffiths, D. W., An Assessment of the Potential for Improving the Nutritive Value of  
684 Field Beans (*Vicia faba*)- A Progress Report. In *Vicia faba: Agronomy, Physiology and*  
685 *Breeding*, Hebblethwaite, P. D.; Dawkins, T. C. K.; Heath, M. C.; Lockwood, G., Eds.  
686 Springer-Science+Business Media, B.V. : Brussels-Luxembourg, 1984; Vol. 10, pp 271-278.  
687 (85) Chen, Y.; Lübberstedt, T., Molecular basis of trait correlations. *Trends Plant Sci.*  
688 **2010**, *15* (8), 454-461.  
689 (86) Munier-Jolain, N.; Larmure, A.; Salon, C., Determinism of carbon and nitrogen  
690 reserve accumulation in legume seeds. *C. R. Biol.* **2008**, *331* (10), 780-787.  
691 (87) Egle, K.; Beschow, H.; Merbach, W., Nitrogen allocation in barley: Relationships  
692 between amino acid transport and storage protein synthesis during grain filling. *Canadian*  
693 *Journal of Plant Science* **2015**, *95* (3), 451-459.  
694 (88) Bogard, M.; Allard, V.; Brancourt-Hulmel, M.; Heumez, E.; Machet, J.-M.; Jeuffroy,  
695 M.-H.; Gate, P.; Martre, P.; Le Gouis, J., Deviation from the grain protein concentration-

- 696 grain yield negative relationship is highly correlated to post-anthesis N uptake in winter  
697 wheat. *J. Exp. Bot.* **2010**, *61* (15), 4303-4312.
- 698 (89) Taulemesse, F.; Le Gouis, J.; Gouache, D.; Gibon, Y.; Allard, V., Bread Wheat  
699 (*Triticum aestivum* L.) Grain Protein Concentration Is Related to Early Post-Flowering  
700 Nitrate Uptake under Putative Control of Plant Satiety Level. *PLoS ONE* **2016**, *11* (2),  
701 e0149668.
- 702 (90) Poeta, F.; Ochogavia, A. C.; Permingeat, H. R.; Rotundo, J. L., Storage-Associated  
703 Genes and Reserves Accumulation in Soybean Cultivars Differing in Physiological Strategies  
704 for Attaining High Seed Protein Concentration. *Crop Sci.* **2017**, *57* (1), 427-436.
- 705 (91) Hanafy, M.; Pickardt, T.; Kiesecker, H.; Jacobsen, H.-J., Agrobacterium-mediated  
706 transformation of faba bean (*Vicia faba* L.) using embryo axes. *Euphytica* **2005**, *142* (3), 227-  
707 236.
- 708 (92) Wang, X.; Jiang, G.-L.; Song, Q.; Cregan, P. B.; Scott, R. A.; Zhang, J.; Yen, Y.;  
709 Brown, M., Quantitative trait locus analysis of seed sulfur-containing amino acids in two  
710 recombinant inbred line populations of soybean. *Euphytica* **2014**, *201* (2), 293-305.
- 711 (93) Warrington, C. V.; Abdel-Haleem, H.; Hyten, D. L.; Cregan, P. B.; Orf, J. H.; Killam,  
712 A. S.; Bajjalieh, N.; Li, Z.; Boerma, H. R., QTL for seed protein and amino acids in the  
713 Benning x Danbaekkong soybean population. *Theor. Appl. Genet.* **2015**, *128* (5), 839-50.
- 714 (94) Martensson, P., Variation in legumin : vicilin ratio between and within cultivars of  
715 *Vicia faba* L. var. minor. Martinus Nijhoff. World crops: production, utilization and  
716 description, volume 3.: The Hague, 1980; pp 159-172.
- 717 (95) Gatehouse, J.; Croy, R.; McIntosh, R.; Paul, C.; Boulter, D., Quantitative and  
718 qualitative variation in the storage proteins of material from the EEC joint field bean test.  
719 *Quantitative and qualitative variation in the storage proteins of material from the EEC joint*  
720 *field bean test.* **1980**, 173-188.
- 721 (96) Ma, Y.; Kan, G.; Zhang, X.; Wang, Y.; Zhang, W.; Du, H.; Yu, D., Quantitative Trait  
722 Loci (QTL) Mapping for Glycinin and beta-Conglycinin Contents in Soybean (*Glycine max*  
723 *L. Merr.*). *J. Agric. Food Chem.* **2016**, *64* (17), 3473-83.
- 724 (97) Sjödin, J., Induced morphological variation in *Vicia faba* L. *Hereditas* **1971**, *67* (2),  
725 155-179.
- 726 (98) Duc, G., Mutagenesis of faba bean (*Vicia faba* L.) and the identification of five  
727 different genes controlling no nodulation, ineffective nodulation or supernodulation.  
728 *Euphytica* **1995**, *83* (2), 147-152.
- 729 (99) Brosnan, J. T.; Brosnan, M. E., The sulfur-containing amino acids: an overview. *J.*  
730 *Nutr.* **2006**, *136* (6 Suppl), 1636s-1640s.
- 731 (100) Utsumi, S.; Yokoyama, Z.-i.; Mori, T., Comparative Studies of Subunit Compositions  
732 of Legumins from Various Cultivars of *Vicia faba* L. Seeds. *Agric. Biol. Chem.* **1980**, *44* (3),  
733 595-601.
- 734 (101) Fontes, E. P. B.; Moreira, M. A.; Davies, C. S.; Nielsen, N. C., Urea-elicited changes  
735 in relative electrophoretic mobility of certain glycinin and  $\beta$ -conglycinin subunits. *Plant*  
736 *Physiol.* **1984**, *76* (3), 840-842.
- 737 (102) Poysa, V.; Woodrow, L.; Yu, K., Effect of soy protein subunit composition on tofu  
738 quality. *Food Res. Int.* **2006**, *39* (3), 309-317.
- 739 (103) Mertens, C.; Dehon, L.; Bourgeois, A.; Verhaeghe-Cartryse, C.; Blecker, C.,  
740 Agronomical factors influencing the legumin/vicilin ratio in pea (*Pisum sativum* L.) seeds. *J.*  
741 *Sci. Food Agric.* **2012**, *92* (8), 1591-6.
- 742 (104) Ladjal E, Y.; Boudries, H.; Mohamed, C.; Romero, A., *Pea, Chickpea and Lentil*  
743 *Protein Isolates: Physicochemical Characterization and Emulsifying Properties.* 2015; Vol.  
744 11, p 43-51.

745 (105) Schumacher, H.; Paulsen, H. M.; Gau, A. E., Phenotypical indicators for the selection  
746 of methionine enriched local legumes in plant breeding. *Agriculture and Forestry Research*  
747 **2009** 4(59), 339-344.

748

749

750

751

752

753

754

755

756

757

758

759

760

761

762

763

764

## 765 **FIGURE CAPTIONS**

766 Figure 1. Global distribution of *Vf* cultivation (A) and the major producing countries (B).  
767 Data was sourced from FAOstats and distribution map was generated using Tableau Public  
768 2018.1.

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

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

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

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

782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800

**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.
	$\alpha$ chain	$\beta$ chain	vicilin	Convicilin	
<i>Vicia faba</i>	38	22-24	31-65		De Pace, et al. <sup>59</sup>
	38-47	..	..	64	Liu, et al. <sup>42</sup>
	40	20	..	..	Gatehouse, et al. <sup>95</sup>
	35-39	23-25	42-48	66	Tucci, et al. <sup>51</sup>
	36-51	19-23	..	..	Utsumi, et al. <sup>100</sup>
	40	23-24	54	~73	This study
<i>Medicago truncatulla</i>	36-46	23-24	46-47	60-92	Le Signor, et al. <sup>52</sup>
	42-46	23	46-47		Gallardo, et al. <sup>61</sup>
	38-41	..	47	70	Le Signor, et al. <sup>72</sup>
<i>Glycine max</i> *	37	20	52-72		Fontes, et al. <sup>101</sup>
	37	20	52-72		Boehm, et al. <sup>55</sup>
	37	20	52-72		Poysa, et al. <sup>102</sup>
	37-44	17-22	53-76		Krishnan, et al. <sup>54</sup>

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

\*7S subunits of *G.max* consist of  $\alpha'$ ,  $\alpha$  and  $\beta$  polypeptides.

801

802

803

804

805

806

807

808

809

810

811

812

813

814

815

816

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

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

817

818

819

820

821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838

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

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

\* S-AA reported as % protein

839  
840  
841

842

843

844

845

846

847

848

849

850

851

852

853

854

855

856

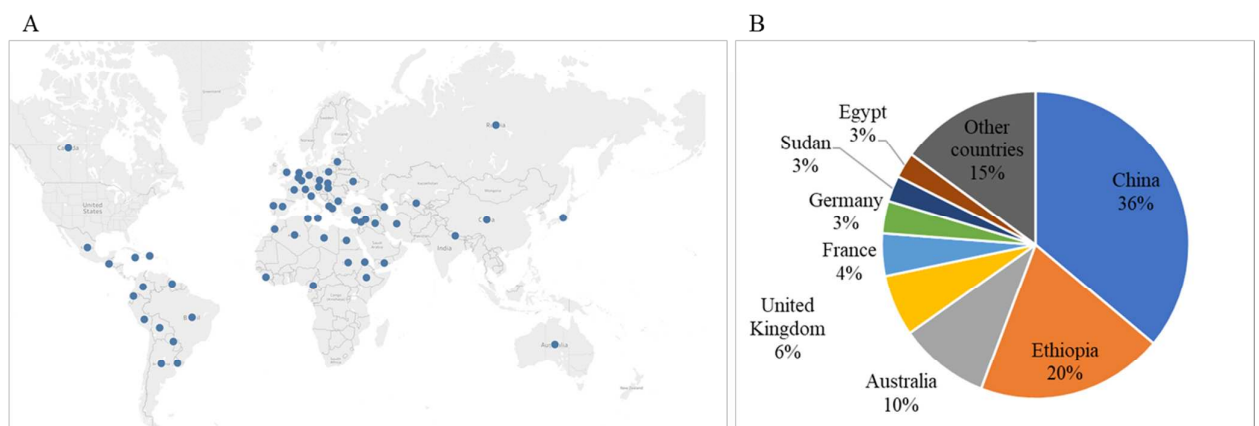
857

858

859

860

861 Figure 1



862

863

864

865

866

867

868

869

870

871

872

873

874

875

876

877

878

879

880

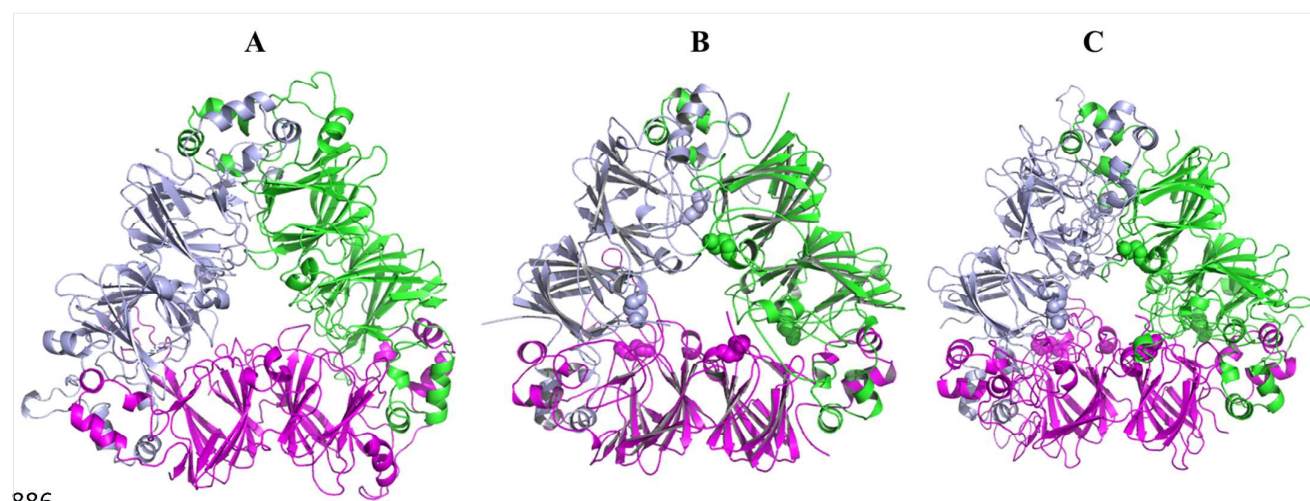
881

882

883

884 Figure 2

885



886



887

888

889

890

891

892

893

894

895

896

897

898

899

900

901

902

903

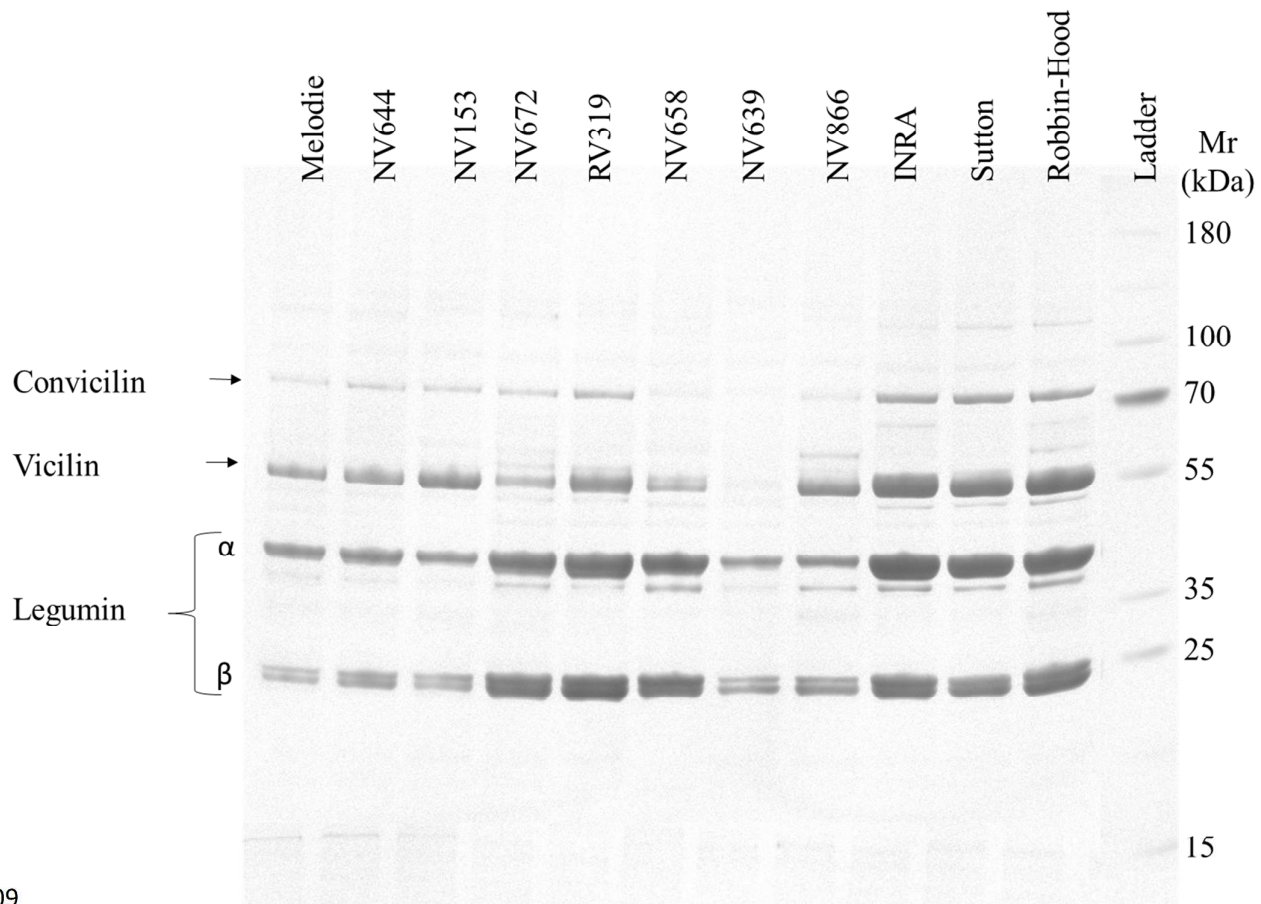
904

905

906 Figure 3

907

908



909

910

911

912

913

914

915

916

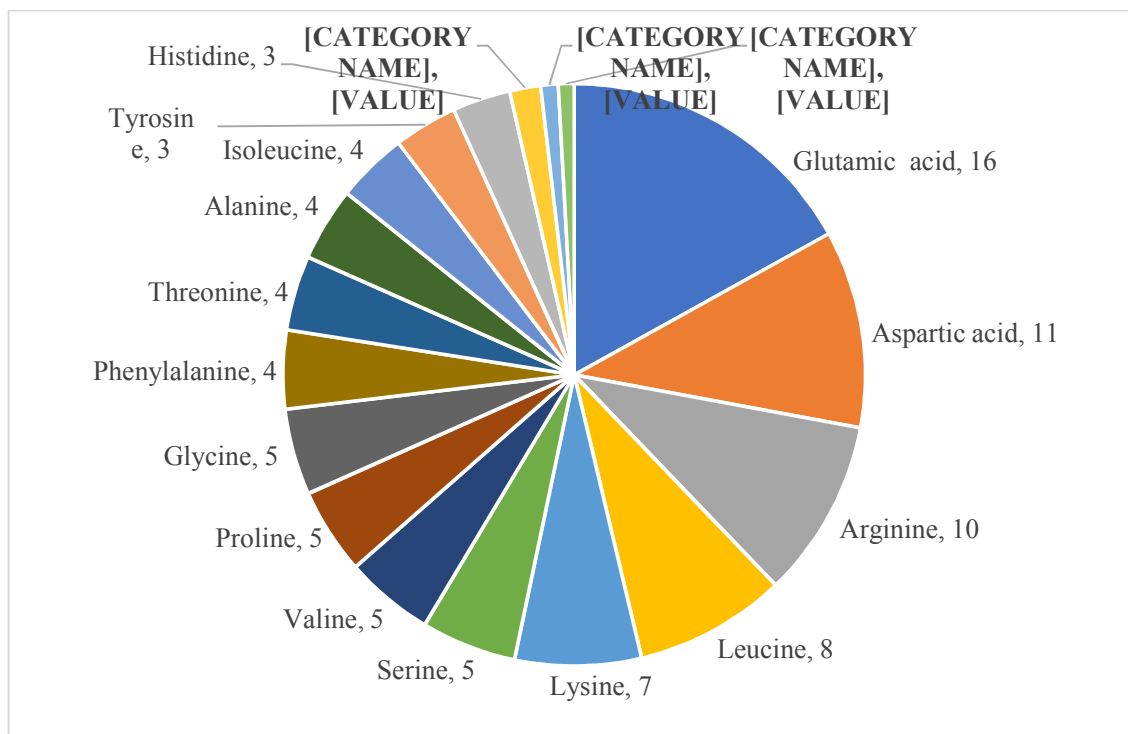
917

918

919

920

921 Figure 4



922

923

924

925

926

927

928

929

930

931

932

933

934

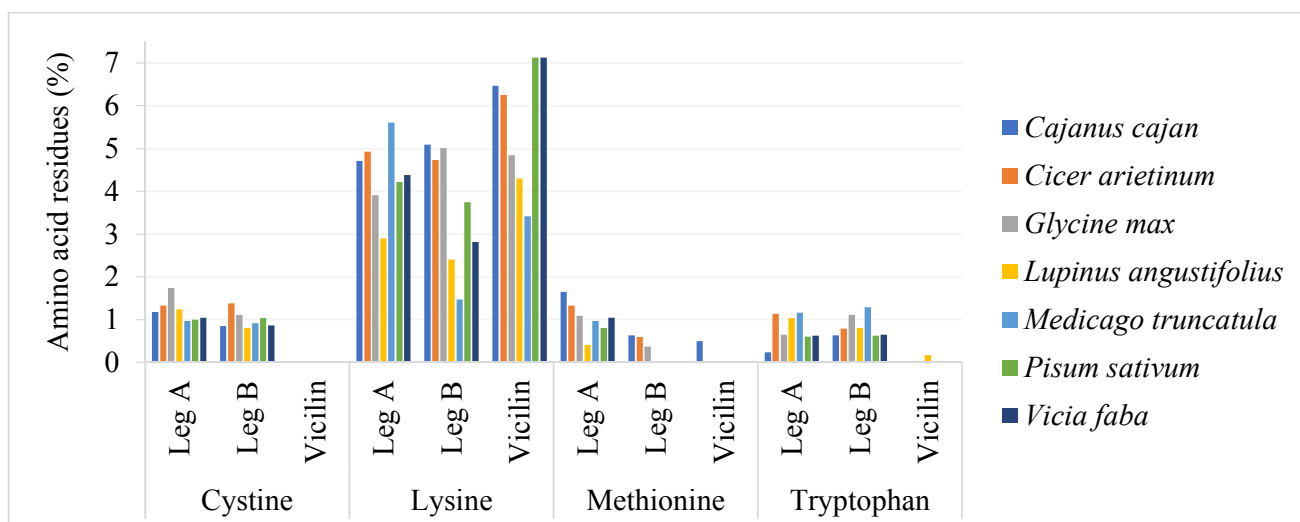
935

936

937

938

939 Figure 5



940

941

942

943

944

945

946

947

948

949

950

951

952

953

954

955

956

957

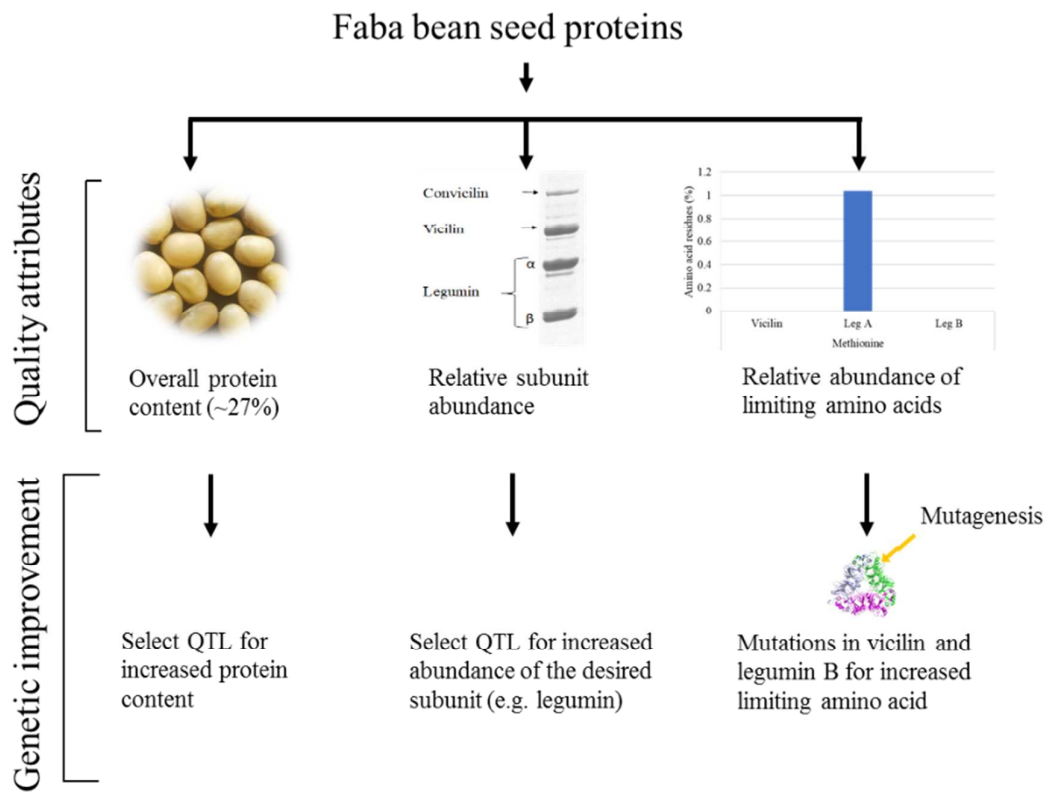
958

959

960

961

TOC Graphic



962

963

964

965

966

967

968

969