Faba bean : transcriptome analysis from etiolated seedling and developing seed coat of key cultivars for synthesis of proanthocyanidins, phytate, raffinose family oligosaccharides, vicine, and convicine

Ray, Heather; Bock, Cheryl; Georges, Fawzy

Publisher's version / Version de l'éditeur:
https://doi.org/10.3835/plantgenome2014.07.0028
The Plant Genome, 8, 1, pp. 1-11, 2015-03-13

Published March 13, 2015

ORIGINAL RESEARCH

Faba Bean: Transcriptome Analysis from Etiolated Seedling and Developing Seed Coat of Key Cultivars for Synthesis of Proanthocyanidins, Phytate, Raffinose Family Oligosaccharides, Vicine, and Convicine

Heather Ray, Cheryl Bock, and Fawzy Georges*

Abstract

Faba bean (*Vicia faba* L.) has been little examined from a genetic or genomic perspective despite its status as an established food and forage crop with some key pharmaceutical factors such as vicine and convicine (VC), which provoke severe haemolysis in genetically susceptible humans. We developed next-generation sequencing libraries to maximize information to elucidate the VC pathway or relevant markers as well as other genes of interest for the species. One selected cultivar, A01155, lacks synthesis of the favism-provoking factors, VC, and is low in tannin, while two cultivars, SSNS-1 and CDC Fatima, are wild-type for these factors. Tissues (5- to 6-d-old root and etiolated shoot and developing seed coat) were selected to maximize the utility and breadth of the gene expression profile. Approximately 1.2 × 10^6 expressed transcripts were sequenced and assembled into contigs. The synthetic pathways for phosphatidylinositol or phytate, the raffinose family oligosaccharides, and proanthocyanidin were examined and found to contain nearly a full complement of the synthetic genes for these pathways. A severe deficiency in anthocyanidin reductase expression was found in the low-tannin cultivar A01155. Approximately 5300 variants, including 234 variants specific to one of the three cultivars, were identified. Differences in expression and variants potentially related to VC synthesis were analyzed using strategies exploiting differences in expression between cultivars and tissues. These sequences should be of high utility for marker-assisted selection for the key traits vicine, convicine, and proanthocyanidin, and should contribute to the scant genetic maps available for this species.

Published in The Plant Genome 8
doi: 10.3835/plantgenome2014.07.0028

5585 Guilford Rd., Madison, WI 53711 USA
An open-access publication

CULTIVATED FABA BEAN (*Vicia faba* L.) is widely used as human food, especially in Europe, Northern Africa, and China. In view of its superior yield and feeding value over field pea (*Pisum sativum* L.) or other legumes, it is also widely used as animal feed for a variety of species, and it is one of the most effective nitrogen-fixing legumes (Bremer et al., 1988). As well, immature faba bean is a high-quality vegetable, supplying well-balanced protein and carbohydrate together with numerous antioxidants and essential vitamins. It is of particular value for lipid-restricted diets, as it contains less than 3% seed lipid, which is low in saturated fatty acids (unpublished results). As a legume suited to yield well in cooler conditions, it has potential for major expansion in Western Canada and elsewhere.

The importance of faba bean resides not only in their nutritional value but also in the fact that they contain medically important components including vicine and convicine (VC) (Ray and Georges, 2010). These pyrimidine derivatives constitute a rare class of antinutritional factors almost exclusive to faba bean, although also occurring in other *Vicia* species. The aglycones of

H. Ray, C. Bock, and F. Georges, Plant Biotechnology Institute, National Research Council, 110 Gymnasium Pl., Saskatoon, SK, Canada S7N 0W9; F. Georges, Dep. of Biochemistry, College of Medicine, Univ. of Saskatchewan, Saskatoon, SK, Canada S7N 0W0. This is NRCC publication no. 55994. Received 11 July 2014. Accepted 23 Dec. 2014. *Corresponding author (fawzy.georges@usask.ca).

Abbreviations: ANR, anthocyanidin reductase; DFR, dihydroflavonol 4-reductase; EST, expressed transcript; G6PD, glucose-6-phosphate dehydrogenase; HPLC, high-performance liquid chromatography; L-DOPA, L-3,4-dihydroxyphenylalanine; LDOX, leucoanthocyanidin dioxygenase; MIPS, L-*myo*-inositol 1-phosphate synthase; PCR, polymerase chain reaction; PLC, phospholipase C; PPO, polyphenol oxidase; RFO, raffinose family oligosaccharide; SNP, single nucleotide polymorphism; SSR, simple-sequence repeat; VC, vicine and convicine.

VC, divicine and isouramil, are the causative agents of a medical syndrome known as favism, an inherited disorder to which individuals with a variant glucose-6-phosphate dehydrogenase (G6PD) are susceptible. The G6PD deficiency, a recessive sex-linked trait, is one of the most common human enzyme deficiencies worldwide and about 400 million people are affected by this enzymopathy, which is associated with increased tolerance of malaria and appears to provide an overall survival advantage in areas with malarial mosquitoes. The synthetic pathway of these compounds has not been characterized, but the compounds are accumulated in substantial amounts in several of the tissues we have selected to examine. Identification of components of the synthetic pathway would be a critical step toward breeding lines free of these compounds. To date, other libraries do not contain sequence from the mutant cultivar.

An additional rare plant biochemical is found in faba bean, L-3,4-dihydroxyphenylalanine (L-DOPA), which is the precursor of the essential mammalian neurological factor dopamine and the major ingredient in medicines used to treat Parkinson's disease patients. Faba bean is the only significant crop that produces L-DOPA; it does so in relatively large amounts of up to 7% of the dry weight of pod tissue (Burbano et al., 1995). This species could, therefore, be suitable as a natural supplement ameliorating the symptoms of Parkinson's disease. The synthetic pathway in faba bean is not well characterized. An understanding of the synthetic pathway and presumed sequestration in this species is highly desirable, as it could lead to improved cultivars with higher levels of L-DOPA synthesis, among other benefits.

In addition, more common antinutritional factors such as phytates, seed coat condensed tannin or proanthocyanidins, and the sucrose galactosides including raffinose, stachyose, and verbascose are found. These factors are significant issues for consumption both in food and feed, although they are relatively low in faba bean.

This species has been among the orphan crops without sufficient acreage to suggest the utility of a major effort to analyze its genetics and without sufficient research to promote its extensive use or reduce its antinutritional factors. It has a very large diploid genome ($2n = 12$; estimated genome size = 13 gigabases), which has been little mapped or otherwise studied. Genetic maps, until recently, contained rather few markers (Torres et al., 2006). However, recent work used *Medicago truncatula* Gaertn. sequence to generate candidate primers, some of which amplified distinguishable fragments in faba bean and were successfully used to locate markers in a segregating population of recombinant inbred lines (Ellwood et al., 2008). These markers showed a substantial macrosynteny between faba bean, *M. truncatula*, and lentil (*Lens culinaris* Medik.), and in a few locations where higher marker density permitted the comparison, a significant degree of microsynteny as well (Ellwood et al., 2008; Cruz-Izquierdo et al., 2012). This work suggests that the genome is expanded throughout its length in at least a moderately uniform way. However, a higher proportion of multiple bands from the tested primers, relative to lentil (Ellwood et al., 2008), suggests that localized, small-scale duplications cannot, at this point, be excluded as a source of some of the great length of the faba bean genome. Updated maps (Satovic et al., 2013; Kaur et al., 2014) combining information from several sources confirm that the genome map is highly similar to that of *M. truncatula* with a fairly minimal number of chromosomal rearrangements and large blocks of both micro- and macrosynteny.

In the past two years, second-generation sequencing techniques have moved faba bean, as with many other species, from a paucity of sequence information to a relative abundance. Until recently almost all publicly available faba bean sequence was from an expressed transcript (EST) library we prepared from developing embryo of the garden variety 'Windsor' (Ray and Georges, 2010). This library has since been mined to identify simple-sequence repeats (SSR) (Akash and Myers, 2012). As well, sequencing from several tissues of two cultivars, used primarily to develop SSRs, has generated about 18,000 identified genes, equivalent to roughly 70% of a minimal plant genome [i.e., that suggested by *Arabidopsis thaliana* (L.) Heynh., *Brachypodium distachyon* (L.) P. Beauv., and *Oryza sativa* L. sequencing] (Kaur et al., 2012), and has been used to identify quantitative trait loci for Ascochyta blight resistance (Kaur et al., 2014). In addition, the mitochondrial genome has been sequenced (Negruk, 2013).

Analysis of the synthetic pathways for the compounds of medical significance, such as VC, is of high importance. It would permit identification of mutants and single nucleotide polymorphisms (SNPs) for key genes and thereby facilitate breeding lines with lower amounts of the compounds indicated. Breeding of low or zero lines would be exceptionally valuable in allowing the consumption of faba bean without concerns about these compounds. This would significantly increase the acceptability of the species as food, by removing the major, and almost the only, objection to its increased consumption. For these key pathways, essentially no information is available.

Expressed transcript libraries are particularly advantageous in species with very large genomes that have a reasonable degree of synteny to more established genomes. We wished to locate molecular markers in several significant cultivars for faba bean breeding and, particularly, to find markers relevant to mapping and potentially identifying the gene involved in the low VC mutation, vc^-. For this purpose, we developed EST libraries from 6-d-old root and shoot tissue of three cultivars of faba bean: the wild-type Canadian cultivars SSNS-1 and CDC Fatima and the low-VC line A01155, and from developing seed coat tissue of CDC Fatima and A01155. Expression comparisons between tissues and varieties were performed. We identified and compared variants between the varieties to select candidate markers for VC. In addition, we analyzed several genes of the phytate

pathway, the proanthocyanidin pathway, and the raffinose family oligosaccharides (RFOs) synthetic pathway. The sequence data, expression analysis, and identification of variants are expected to be of use in developing faba bean genomics, in marker-assisted breeding, and for analyzing biochemical pathways.

Materials and Methods

Plant Material

A01155 is a cultivar identified as being very low in VC (*vc*$^-$; less than 5% of wild-type concentration; Duc et al., 1989). The *vc* allele appears to be a single gene that is maternally expressed (Duc et al., 1989). A01155 is a white-flowered cultivar with light-colored seeds and is considered low in tannin (Oomah et al., 2011). CDC Fatima is an established cultivar developed for use in the prairie provinces of Canada, which has dominated the limited acreage grown, while SSNS-1 (a small-seeded line) is an advanced breeding line developed for feed (A. Vandenberg, personal communication, 2009). Both CDC Fatima and SSNS-1 have wild-type flowers, medium brown seeds, and are considered wild-type in tannin and VC levels (Oomah et al., 2011).

Library Preparation

Seeds were started in soil and in dark conditions for 5 or 6 d. They were then quickly washed, dissected away from the cotyledon, snap-frozen, ground in liquid nitrogen, and RNA prepared using Plant RNeasy kit (Qiagen). The mRNA was then purified from total RNA by extracting twice through one oligo-dT cellulose column, from PolyA Purist kit (Ambion, AM1916). Purifying the RNA through the column twice gave 1.0 to 1.2% yield, the optimal range for generation of high quality cDNA sequence. For seed coat RNA, plants were grown to seed development in growth chambers (16 h day at 20°C, night at 15°C; light ~300 μmol m^{-2} s^{-1}) in a soil mixture with a medium level of fertilizer. Seed coat tissue from developing seeds of CDC Fatima and A01155 was removed from developing seeds 9 to 10 mm long, snap-frozen, ground, and RNA similarly prepared. While the cultivars have slightly different development rates, and seeds within a pod also develop at different rates, we attempted to collect samples of uniform stage and condition.

cDNA libraries were constructed according to the cDNA rapid library preparation method manual (Roche Applied Science, 2009) using 400 ng mRNA each. Root and shoot libraries were multiplexed with individual adapters. Following library preparation, the libraries were pooled in equimolar concentrations and 3 copies per bead were used for emulsion polymerase chain reaction (PCR). A half plate using GS FLX Titanium chemistry on the GS Sequencer (Roche Applied Science) was prepared, using 2×10^6 beads. The maximum mRNA fragment length was cut off at 600 bp; approximately two-thirds of the sequences were between 350 and 550 bp in length. Seed coat libraries were prepared as for seedling libraries, but using a smaller amount of RNA; one-eight plate of each was sequenced.

Bioinformatics Analysis

Preprocessing included repeat filtering, low-complexity masking, and polyA removal using custom scripts developed by the Bioinformatics Section, Plant Biotechnology Institute, Saskatoon. Initial assembly was performed with Mira version 3.0.0 (Abekas Inc., 2011); this assembly was used to analyze expression of antinutritional genes. Assemblies of each library were performed using CLC-Bio Bioinformatics package, version 9.3, (http://www.clcbio.com) with stringent conditions (100% sequence identity, 0.3 overlap, and maximal mismatch cost). An additional assembly of all sequences together used stringency of 99% to allow homologs from different cultivars to coassemble. Other analysis-driven assemblies were made using CLC-Bio with stringency of 99 to 100% sequence identity. In some cases, the unmapped reads were restored to the assemblies to provide unigene lists.

Expression analysis was performed using the RNA-Seq function of CLC-Bio followed by spreadsheet analysis. Sequences from all libraries were coassembled at high stringency (0.3 overlap, 99% identity). Reads from each library were read successively to this assembly, again at high stringency (0.4 overlap, 99% identity), and compared and analyzed as detailed in Results and Discussion.

Variant analysis was performed using CLC-Bio. For variant analysis, pooled sequences from all tissues of each cultivar were grouped and assembled with stringent conditions of 0.3 overlap and 100% identity and unassembled sequences restored to provide a list of unigenes. The unigene list of each cultivar was in turn used as reference for the other two, using CLC-Bio functions of reads-to-reference with stringent parameters including 0.4 overlap with 99% identity, followed by quality-based variant detection at similarly stringent parameters. A minimum read depth of two was required. Insertions, deletions, and ambiguous calls were removed, as were sequences that were expressed in an embryo library of Windsor (Ray and Georges, 2010). Various filtering strategies were then employed as described in Results and Discussion.

Phytate and Sugar Analysis

Phytate and sugar were analyzed by high-performance liquid chromatography (HPLC) as described in Bock et al. (2009). For phytate, eight individual, mature seeds per cultivar from the same seed lots used to grow plant material were used; for sugars, six were used. Two technical replicates were used for each sample.

Results and Discussion

Library Characterization

Eight libraries containing a total of 1.2 million EST sequences were developed and sequenced using 454 sequencing technology (454 Life Sciences Corp.). The sequence quality was high, with the great majority of

sequences having high PHRED scores (Ewing et al., 1998). Over 90% of the unigenes were tentatively identifiable with a characterized gene. There were relatively few sequences related to photosynthesis, as etiolated seedling and developing seed coat tissues were used. A high degree of complexity was found in all libraries. The last new sequences entered in each library are approximately 80% likely to match a sequence already present, indicating that much more sequence richness exists in the libraries than we have uncovered to date. The genus *Vicia* appeared most closely related to *Cicer*, *Pisum*, and *Medicago*, with median E-values of 10^{-152}, less than 10^{-152}, and 10^{-149}, respectively. *Medicago truncatula* was the most closely related species with an established genome.

Each library was separately assembled at high stringency, resulting in over 340,000 unigenes among the eight libraries (summarized in Table 1; unigenes from each library appear in Supplemental Table S1–S8). Read lengths averaged approximately 370, while average contig lengths ranged from 620 to 820 among the libraries.

To quantify the summed faba bean transcriptome, we also assembled sequences from all the libraries together with stringent parameters. This assembly resulted in 32,400 contigs with an N_{50} of 1143 and includes 35% of the total sequence length, while the unigenes numbered 190,119 (Supplemental Table S9). The total sequence length of 8×10^7 bp (both assembled and nonassembled unique sequences) is a minimal estimate of transcriptome size. If an average transcript size of 2500 bp is assumed, this gives an estimate of about 32,000 transcribed genes. If numbers of transcripts follows the expected logarithmic function, at approximately 80% coverage, this suggests that in these tissues, about 64,000 transcribed genes would account for approximately 96% of all transcripts. This estimate is probably high for these tissues, as an unknown proportion of sequences that should have assembled may have failed to do so (due to sequencing or assembly errors), but presumably, many more genes would be expressed only in other tissues.

For the best-studied plant genomes, approximate numbers of expressed genes have been identified. For example, *A. thaliana* expresses at least 20,000 genes under different conditions (Schmid et al., 2005), while *M. truncatula* expresses about 40,000 (Benedito et al., 2008). Faba bean appears likely to express at least as many genes. A subset of each library was BLASTed and examined; most genes (over 90%) had homologs in genomes or transcriptomes from comparable tissues of legumes or other plants, and almost no homologs in clearly different phyla (results not shown).

Analysis of Expression of Genes for Antinutritional Factors

Early seedling is expected to express a broad cross section of genes necessary for general metabolism, while developing seed coat expresses numerous genes related to embryo development and factors that protect the wild-type seed from numerous stresses but may not be desired in the cultivated plant. We analyzed transcription of genes involved in synthesis of several factors of economic importance in the mature seed, such as phytate, proanthocyanidin (condensed tannin), and RFOs.

Table 1. Faba bean (*Vicia faba* L.) transcribed sequence libraries prepared by 454 sequencing. Each library assembled separately to provide number of contigs and its unmapped reads added to give number of unigenes.

Library name	Tissue	Cultivar	No. expressed transcripts	No. contigs	No. unigenes
VFARO1NG	6-d root	A01155	240,675	15,954	50,806
VFARO2NG	6-d root	SSNS-1	118,998	10,101	40,709
VFARO3NG	6-d root	CDC Fatima	134,626	11,913	41,650
VFASH1NG	6-d shoot	A01155	154,990	12,206	41,179
VFASH2NG	6-d shoot	SSNS-1	127,224	10,542	40,105
VFASH3NG	6-d shoot	CDC Fatima	93,127	9,324	32,366
VFASC1NG	Seed coat	A01155	193,335	13,649	49,898
VFASC2NG	Seed coat	CDC Fatima	184,906	10,539	46,612
Total			1,247,881	94,228	343,325

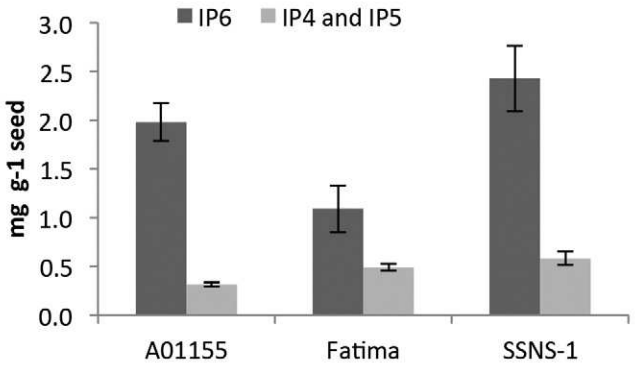

Figure 1. Phytate concentration in mature faba bean (*Vicia faba* L.) seed from cultivars AO1155, CDC Fatima, and SSNS-1.

Phytate and Phytate-Related Genes and Metabolites

Phytate (*myo*-inositol-1,2,3,4,5,6-hexa*kis*phosphate) is a common seed storage component of legume grains. Although it is important as a source of P during germination and early growth, phytate is indigestible by monogastric animals and can become a major source of phosphate pollution; it is also considered antinutritional due to its strong ability to chelate micronutrients, reducing their bioavailability.

Analysis of phytate in seeds of the faba bean cultivars used in this study, using a highly sensitive and specific HPLC procedure on individual seeds of field-grown material from western Canada, showed concentrations of 1.2 to 2.6 g kg^{-1} (Fig. 1). Phytate in mature seeds was highest in SSNS-1 and lowest in CDC Fatima. The very small seed size of SSNS-1 makes this cultivar highly enriched in this antinutritional component relative to the other varieties. For comparison, we analyzed 11 varieties of Egyptian provenance (courtesy of Dr. A.

Table 2. Expression, as numbers of transcripts, in faba bean (*Vicia faba* L.) seedling root and shoot and developing seed coat of selected genes involved in the metabolic pathways of anthocyanin or proanthocyanidin, raffinose family oligosaccharides, phosphatidylinositol-related, and polyphenol oxidase.

	Shoot			Root			Seed coat	
Gene[†]	A01155	CDC Fatima	SSNS-1	A01155	CDC Fatima	SSNS-1	A01155	CDC Fatima
Anthocyanin or condensed tannin pathway								
Flavanone 3-hydroxylase	30	9	12	30	9	12	975	709
Dihydroflavonol reductase	10	9	11	44	5	26	24	184
Leucoanthocyanidin dioxygenase	23	17	23	65	37	27	220	69
Anthocyanidin reductase	0	0	0	0	0	0	2	839
Leucoanthocyanidin reductase	0	1	0	19	8	0	2	0
tt12 (transporter)	22	28	29	12	13	16	10	13
tt10 (laccase-like PPO)	0	0	0	0	0	0	0	12
Transparent testa (other)	0	0	0	0	1	2	4	4
Raffinose family oligosaccharide pathway[‡]								
Raffinose synthase	4	1	7	19	4	11	0	0
Stachyose synthase	1	3	8	2	3	0	1	0
Galactinol synthase	0	0	0	1	0	0	0	19
Galactinol-sucrose galactosyltransferase	4	2	8	19	1	11	0	2
α-galactosidase	28	24	18	76	31	37	95	64
Phosphatidylinositol or phytate pathway								
MIPS	4	2	2	5	3	40	461	141
PIS	4	4	4	11	4	3	8	6
PI3K	10	14	15	12	9	10	4	2
PI4K	45	29	45	54	33	29	11	5
PI4P-5K	60	37	63	138	48	56	223	191
PI3P-5K	1	1	0	1	1	1	1	0
Inositol tetrakisphosphate 1-kinase	4	3	7	0	5	3	3	2
Inositolpentakisphosphate 2-kinase	10	13	3	0	12	14	12	15
Inositol transporter	31	13	11	15	7	30	24	10
Phospholipase C	39	23	41	37	28	27	63	102
Phospholipase D	38	86	44	150	80	62	142	88
DAGK	37	32	25	71	46	37	65	39
L-DOPA pathway, presumed								
Polyphenol oxidase	110	92	215	65	33	37	123	32
Total number of contigs[§]	43,229	32,635	40,035	60,058	39,262	38,784	55,739	43,905

[†] DAGK, diacylglycerol kinase; L-DOPA, L-3,4-dihydroxyphenylalanine; MIPS, *myo*-inositol phosphate synthase; PIS, phosphatidylinositol synthase; PI3K, phosphatidylinositol 3-kinase; PI3P-5K, phosphatidylinositol 3-phosphate 5-kinase; PI4K, phosphatidylinositol 4-kinase; PI4P-5K, phosphatidylinositol 4-phosphate 5-kinase; PPO, polyphenol oxidase.

[‡] Some GenBank sequences are identified as galactinol-sucrose galactosyltransferase and one of raffinose synthase or stachyose synthase.

[§] Numbers of contigs relate to assemblies prepared for this table only.

Sharkawy, Cairo University), as well as Canadian-grown seed of cultivar Windsor and found a similar range of phytate concentrations (results not shown). A second peak, believed to contain an intermediary *myo*-inositol polyphosphate (four or five phosphates) was routinely seen on HPLC traces of faba bean seed (Fig. 1). The cultivars used were previously analyzed for seed phytate by a different method (Oomah et al., 2011), where data suggest phytate concentrations of 9.5 to 15.1 g kg^{-1}; however, the colorimetric method employed is known to detect additional compounds. Using an HPLC analytical method, phytate concentrations of 4.2 to 5.5 g kg^{-1} were observed in two other faba bean cultivars (Goyoaga et al., 2011). In that study, phytate concentrations were also shown to be at similar levels in the cotyledon and embryo axis of dry, mature faba bean seed. During germination, phytate dropped rapidly in the germinating axis but stayed fairly constant in the cotyledon, while the *myo*-inositol 3,4,5,6-tetra*kis*-phosphate and *myo*-inositol 1,3,4,5,6-penta*kis*-phosphate amounts remained relatively constant until 9 d (Goyoaga et al., 2011).

We analyzed the expression of genes related to phytate synthesis, including phosphoinositides-related genes, and found a number of them to be expressed at low and moderately constant levels in seedling root and shoot and developing seed coat with relatively modest differences between varieties (Table 2). For example, the gene for L-*myo*-inositol 1-phosphate synthase (MIPS), the enzyme responsible for the de novo biosynthesis of *myo*-inositol 1-phosphate (initial substrate in the primary pathway of

phytate biosynthesis) was relatively little expressed in seedlings but showed a large increase in seed coat. This agrees with the observation that *myo*-inositol accumulates at high levels in developing seed coat of other crops, although with no corresponding phytate accumulation (Dong et al., 2013). Expression of the successive kinases that generate the various *myo*-inositol polyphosphates was observed in all tissues examined at a steady, low level (Table 2), while in embryo tissue of Windsor, we previously found few MIPS or kinase transcripts (Ray and Georges, 2010).

While phosphatidylinositol synthase was detected at low levels in all tissues, phospholipases C (PLC) were well expressed, as were phospholipases D and diacylglycerol kinase (Table 2). Phospholipase C and phosphatidylinositol-4-phosphate 5-kinase appeared somewhat induced in seed coat, possibly suggesting that in addition to its metabolic signaling the PLC route may contribute substantially to phytate synthesis through the generation of *myo*-inositol 1,4,5-*tris*phosphate, a halfway intermediate in phytate synthesis. Enzymes such as *myo*-inositol 3,4,5,6-tetra*kis*phosphate 1-kinase and *myo*-inositol 1,3,4,5,6-penta*kis*phosphate 2-kinase did not increase significantly in seed coat (Table 2), presumably signifying that they are most expressed in the embryo. No transcripts for these genes were detected in the Windsor early embryo library, suggesting that their synthesis must be confined to later stage embryo.

Proanthocyanidin Synthetic Genes

Condensed tannins (proanthocyanidins) are present in seed coat of many species. The presence of tannins is associated with a somewhat disagreeable bitter flavor, which makes tannin-free cultivars preferred in many crop species. There have been efforts to reduce tannin content in faba bean seed coat, and low-tannin cultivars are known (Oomah et al., 2011). Tannins are polymers of either 2,3 *cis*-flavan-3-ol or 2,3 *trans*-flavan-3-ol products of the anthocyanin pathway (Marles et al., 2003). The polymerization of flavanols may be followed by quinone formation, leading to cross-linking, which may contribute to the chemical resistance of the seed coat layer.

In seedling shoot and root, expression of genes of this pathway were generally similar in the wild-type and zero-tannin (zt^-) cultivars, while in the developing seed coat, several differences were evident. In the seedling, the genes common to the biosynthesis of anthocyanins and tannins, flavanone-3-hydroxylase and dihydroflavonol 4-reductase (DFR), were well expressed and very comparable between varieties. In developing seed coat, transcripts of both increased strongly with a much steeper increase for DFR in CDC Fatima. Expression of leucoanthocyanidin dioxygenase (LDOX), conversely, increased somewhat in A01155 seed coat. Anthocyanidin reductase (ANR), which synthesizes 2,3-*cis*-flavan-3-ols, was not expressed in seedlings. In developing seed coat, however, a very large increase was seen in CDC Fatima but not A01155. In addition, some transcripts of putative *tt10*, a laccase-like polyphenol oxidase which is implicated in quinone formation and cross-linking in tannin synthesis, were found in seed coat of CDC Fatima but absent from A01155 (Table 2).

In both cultivars, the parallel enzyme leucoanthocyanidin reductase, which synthesizes 2,3-*trans*-flavan-3-ols, appeared to be nearly absent; condensed tannin in faba bean therefore seems to be entirely through synthesis of 2,3-*cis*-flavan-3-ols. The transporter *tt12* differed relatively little between the two, and there were no conspicuous changes in homologs to known regulatory genes for tannin synthesis.

The observed changes in the pathway suggest the possibility of a single regulatory gene, interacting with seed-coat-expressed promoters for DFR, ANR, and possibly *tt10*, which differs between A01155 and the wild-type genotypes or, alternatively, that multiple mutations in DFR, ANR, and possibly *tt10* genes have been accumulated in A01155. The relative increase in LDOX in A01155 seed coat may suggest the presence of a feedback loop involving its promoter and a more distant product of the pathway.

Cultivar A01155 is white flowered and has a lighter-colored seed than CDC Fatima or SSNS-1 and had previously been considered to be low in tannins (zt^-) (Oomah et al., 2011). A relationship between white flower and tannin content has been observed in multiple species. However, use of an ethanol-HCl assay incorporating subtraction of polyvinylpyrrolidone-sequestered compounds from total phenolic content found very similar levels for total phenolics (30.9 and 31.8 mg equivalents of catechin for A01155 and CDC Fatima, respectively) and condensed tannin (4.35 and 4.33 mg equivalents of catechin for A01155 and CDC Fatima, respectively) (Oomah et al., 2011). These data suggest that for faba bean, the ethanol-HCl assay is not sufficiently accurate to determine tannin content. A PCR assay incorporating a variant, or quantifying DFR and ANR expression, may be more accurate.

Of the tannin-associated genes that are more expressed in seed coat of CDC Fatima relative to A01155, ANR showed the clearest change in expression between A01155 and CDC Fatima. Contigs were assembled to provide a full-length gene with high homology to ANR of numerous species (KP006495, similar to ANR AY184243 of *M. truncatula*). Variants were detected in A01155 relative to CDC Fatima but will need to be confirmed (Supplemental Table S10). Dihydroflavonol reductase was also strongly induced in CDC Fatima seed coat relative to A01155, although there was some expression in seedling tissues. We identified two DFR sequences that were strongly upregulated in seed coat of CDC Fatima relative to A01155 and appear to be seed-coat-specific forms (DFR1, KP006492, and DFR2 KP006494, similar to DFR1, AY389346, and DFR2, AY389347 of *M. truncatula*, respectively) (Supplemental Table S10). We also examined the *tt10*-like transcripts detected only in CDC Fatima seed coat, which showed high homology to *tt10* laccase-15-like genes in soybean [*Glycine max* (L.) Merr.] and *M. truncatula* (KP006496, similar to XM_003602987 of the latter). No variants were detected (Supplemental Table S10). In what way these genes correspond with the zero-tannin trait of A01155 remains to be determined.

Raffinose Family Oligosaccharides and Related Genes and Metabolites

The RFO sugars are substantial components of most legume seeds, where they are considered an antinutritional factor because of their indigestibility in monogastric systems. Instead, they are degraded anaerobically by microflora of the intestine leading to buildup of intestinal gases, such as methane, and subsequent flatulence, making their reduction in edible seeds desirable (Bock et al., 2009; Polowick et al., 2009). However, RFOs are also essential to desiccation and proper seed maturation. Raffinose, the simplest member of the RFOs, is a trisaccharide that is produced by the transfer of an α-galactosyl residue from galactinol to a molecule of sucrose. Similarly, stachyose is the product of an α-galactosyl residue transfer from galactinol to raffinose. The sequential stacking of α-galactosyl residues in the growing chain is performed by a family of α-galactosyltransferases that are generally present in developing seeds of higher plants and more predominantly in pulse seeds (Peterbauer et al., 2002).

The faba bean seed mostly accumulates the higher oligosaccharides stachyose and verbascose, with smaller amounts of raffinose (Fig. 2). Adjugose, larger than verbascose by one galactosyl residue, has also been observed (Goyoaga et al., 2011). In our material, very small amounts of adjugose may have been present on HPLC traces, but in the absence of an adjugose standard we were unable to confirm its identity. The pathway components varied to some degree between the cultivars, with A01155 having higher amounts of free galactose and sucrose, but relatively little difference was seen in the RFOs. The Saskatchewan-grown cultivars differ from cultivars Brocal and Alameda (Goyoaga et al., 2011) in having relatively more stachyose and less verbascose. As well, the total RFOs were 0.30 to 0.36% in all three cultivars (Fig. 2) compared with about 0.45% in Brocal and Alameda (Goyoaga et al., 2011), a difference that might be due either to the genotype or growth location and temperature. The relatively higher galactose content in the *vc*⁻ cultivar, A01155, is of interest and may be viewed as a trade-off of one antinutritional component (favism-causing factors) for another (galactosemia-causing factor), since individuals who lack the ability to properly metabolize galactose may be prone to developing galactosemia, which is another serious medical disorder.

In the mature seed, RFOs are most concentrated in the embryo axis (Goyoaga et al., 2011). During germination, the sugars of this family are mobilized in order of size from the cotyledon and embryonic axis (shoots and root). The larger molecules—adjugose, verbascose, and stachyose—rapidly disappear, while raffinose gradually decreases but remains substantial at 9 d (Goyoaga et al., 2011). Meanwhile, fructose and glucose gradually increase while galactose, sucrose and raffinose gradually decrease.

We analyzed expression of genes related to RFO-sugar interconversions and metabolism in the germinating seed at 6 d (Table 2) and found strong expression of the α-galactosidases, which sequentially reduce the size

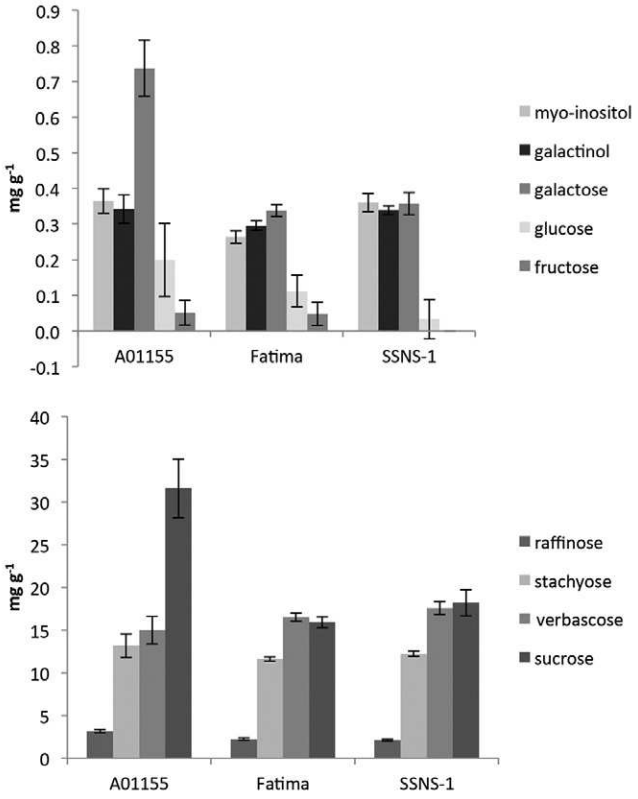

Figure 2. Sugar concentration in mature faba bean (*Vicia faba* L.) seed from cultivars AO1155, CDC Fatima, and SSNS-1. Top, minor sugars; bottom, major sugars.

of RFOs, although they were also well expressed in seed coat. Low numbers of stachyose synthase and galactinol-sucrose galactosyltransferase transcripts occurred in the seedling shoot and root, but galactinol synthase was almost completely absent. Analysis of expression in the developing seed coat also showed few transcripts of this group, except for an increase of galactinol synthase in CDC Fatima seed coat. The enzymes are expected to be found in larger amounts in embryo tissue, and examination of the library of early developing embryo of Windsor found a few transcripts of galactinol-sucrose galactosyltransferases (Ray and Georges, 2010). Analysis of later-stage embryo expression profiles might contribute a clearer picture of RFO metabolism in this tissue.

Biosynthesis of L-3,4-dihydroxyphenylalanine

L-3,4-dihydroxyphenylalanine accumulates in multiple tissues of faba bean, and genotypes lacking it have not been identified to date. The synthetic pathway to L-DOPA in faba bean involves tyrosine hydroxylation but is not yet well characterized (reviewed in Ray and Georges, 2010). A form of polyphenol oxidase (PPO), which hydroxylates tyrosine without proceeding to the formation of quinones, may be the enzyme involved. Transcripts of several PPO isoforms were found at substantial levels in all libraries (Table 2), although the forms presumably involved in L-DOPA synthesis have

not been distinguished. Tyrosine hydroxylases that are not PPO have been reported from rose moss (*Portulaca grandiflora* Hook.) (Yamamoto et al., 2001) and velvet bean [*Mucana pruriens* (L.) DC. var. *utils* (Wall. Ex Wight) Baker ex Burck] (Luthra and Singh, 2010); however, no corresponding sequences have been published. Comparable sequences might occur in faba bean, but no L-DOPA-free lines have been identified to facilitate their discovery. An understanding of the synthetic pathway and presumed sequestration in this species is highly desirable, as it could lead to improved cultivars and higher levels of L-DOPA, among other benefits.

Identification of Vicine and Convicine Candidate Genes via Expression Profiling and Single Nucleotide Polymorphism Analysis

Strategy

We chose material suitable for two strategic approaches to the isolation of the VC factor. The first sought to identify genes that are expressed in CDC Fatima and SSNS-1 but absent or rare in A01155 using a comparison of the sequences in each library to an assembly of all sequences, then comparing combinations of expression characteristics. The second approach sought SNPs in expressed genes of suitable characteristics that are different in the mutant cultivar A01155 compared with the wild-type cultivars CDC Fatima and SSNS-1. Seedling root and shoot tissues and seed coat tissues were considered suitable for such a strategy, as all are probable sites of VC synthesis (Ray and Georges, 2010).

Relative Expression Analysis

The embryo contains VC, as do numerous other tissues, but the VC trait has been found to be maternally expressed (Duc et al., 1989); it is synthesized elsewhere and imported into the embryo. No studies to date, to our knowledge, have examined the source of VC in the embryo. The maternal seed coat seemed a plausible synthetic site, as did the plant as a whole, which is represented by the 6-d-old root and shoot. This analysis was used to identify candidate genes involved in VC synthesis. If there is a difference in expression of a critical gene in VC synthesis, it might appear either as present or absent between A01155 and the *vc*$^+$ genotypes (wild-type in tannin and VC levels), or it might differ substantially in expression. The data were examined to identify sequences that met several criteria in terms of their expression: they were little or not expressed in A01155 but significantly more expressed in CDC Fatima and SSNS-1, a difference found in all three tissues; they were not expressed in embryo tissue, as examined in a previously prepared cDNA library from Windsor (Ray and Georges, 2010); they were not highly similar to known genes; and they appeared likely to be enzymes or regulatory factors. Regulatory genes with such a profile were of high interest, but very few were encountered with a suitable expression profile distinct from other weakly expressed genes. As the level of wild-type expression decreased, the probability of an expression profile being of interest by chance increased (signal-to-noise ratio decreased), which made it difficult to identify potential regulatory or other genes naturally expressed at a low level. While the libraries were less than saturated, it seemed reasonable to suppose that the synthesizing enzymes of compounds often accumulated to over 1% of dry weight should have a detectable level of the corresponding transcripts. However, if the gene affected by the *vc*$^-$ mutation is in fact regulatory, its transcript might not be detectable, although components of the synthetic pathway induced by it could potentially be found. As well, if the factor is a product of posttranscriptional modification, it might not be detectable.

Expression analysis was performed on early root and shoot tissues and developing seed coat using the RNA-Seq function of the CLC-Bio program and filtered to reduce lists of expressed sequences. As *vc*$^-$ is a maternally expressed trait (Duc et al., 1989) and is presumed not to be expressed in embryo (Ray and Georges, 2010), genes expressed in embryo of Windsor, which is wild-type for VC, were excluded.

Reads from each library were read successively to a highly stringent assembly of reads from all libraries placed together. Reads that did not map to this assembly were not considered further, as the probability of VC RNA being present at such a low frequency as 10^{-4} or less was presumed to be negligible. For each tissue separately, the cultivars were then compared and filtered using RNA-Seq (CLC-Bio); the full files for each tissue are appended as Supplementary Table S11–S13. Contigs or reads were identified where number of reads exceeded five in each wild-type library (CDC Fatima and SSNS-1), and at least a three-fold increase was found between A01155 and one of the wild-type libraries. This excluded some potentially valid but less-expressed candidates. As the search strategy did not distinguish whether both or only one of the wild-type cultivars had clearly stronger expression of a sequence than did A01155, the data were then exported to a spreadsheet and further filtering was performed, reducing the candidates to those with at least two-fold higher expression in both of CDC Fatima and SSNS-1 relative to A01155. The lists were compared and nine sequences present in all three tissues identified. These sequences were examined by BLASTx; after removal of those found to be clearly identifiable with a known function, a single candidate, 4684, remained (Table 3). This candidate appears to encode a near-full-length protein of unknown function, at least 143 amino acids in length, with low similarity to any widely conserved protein but short regions of amino acid similarity to little-characterized sequences of faba bean and field peas. It has major differences in expression between the cultivars. However, there is one transcript related to this sequence in the Windsor embryo library (frequency 5×10^{-4}).

Because the criteria were stringent, it is very possible that valid candidates were excluded. As well, the sites of synthesis of VC are not characterized in detail, and it is

Table 3. Faba Bean (*Vicia faba* L.) candidate sequences for involvement in vicine or convicine synthetic pathway as shown by reduced expression in 'A01155' (low-tannin, *vc*$^-$) tissues relative to higher expression in 'CDC Fatima' and 'SSNS-1' (*vc*$^+$) tissues. Contig identification number refers to assembly of all A01155 tissues together. Sequences shown in Supplemental Table S14.

Contig number	Number of transcripts†								Length	BLASTx best hit	E-value
	Shoot			Root			Seed coat				
	F	S	A	F	S	A	F	A			
									bp		
1899	10	1	0	8	6	0	99	0	512	Predicted: uncharacterized protein At5g39570-like isoform X2 (*Cicer arietinum* L.)	10^{-12}
299	18	4	0	21	6	0	30	2	449	14-3-3 protein [*Vigna angularis* (Willd.) Ohwi & H. Ohashi]	10^{-86}
412	15	31	4	16	12	1	32	0	1271	Predicted: UDP-arabinopyranose mutase 2-like (*Cicer arietinum* L.)	0
4518	24	18	3	12	19	3	18	3	370	Predicted: reticuline oxidase-like protein-like (*Cicer arietinum* L.)	10^{-71}
4684	19	54	0	7	7	0	183	0	429	Ovary protein induced by treatment with gibberellic acid (*Pisum sativum* L.)	0.17
4824	41	43	1	8	12	0	0	0	452	Hypothetical protein MTR_2g103170 (*Medicago truncatula* Gaertn.)	10^{-101}

† A, A01155; F, CDC Fatima; S, SSNSI.

Table 4. Variants detected using each cultivar (pooled sequence from all tissues) as reference. For variant detection, all unigenes of a cultivar were used as the reference (Supplementary Table S15–S17). Variants were filtered to include only single nucleotide polymorphisms at 100% frequency.

Cultivar used as reference	No. contigs in reference	Summed length of all contigs	No. unigenes in reference	No. variants in A01155	No. variants in CDC Fatima	No. variants in SSNS-1	No. variants in common
		bp					
A01155	25,373	21.6 M	103,679	–	1,647	666	80
CDC Fatima	21,720	16.8 M	93,289	2,144	–	560	72
SSNS-1	15,881	12.6 M	59,357	1,763	1,125	–	102

possible that the product is transported widely from its site of synthesis. Therefore we examined sequences that met the criteria in any two of the three tissues profiled; five additional sequences were identified by this means (Table 3; Supplementary Table S14). Two are completely uncharacterized, while one is of the 14-3-3 class. Proteins of the 14-3-3 group may interact with a wide variety of phosphorylated proteins to modify their function, thus having the potential to affect cell processes at a variety of levels. At this point, no candidate can be eliminated.

This strategy for VC candidate identification requires confirmation of the expression profile by a method such as quantitative real-time PCR, followed by biochemical analysis of the gene and enzyme product, both of which are outside the scope of the present paper. Under optimal conditions, it could be possible to identify an entire coregulated pathway by this method.

Variants for Marker-Assisted Selection and Vicine and Convicine Candidate Identification

We identified variants, principally SNPs, among the three cultivars to determine the amount of variation present among them for two purposes. The first was to identify SNPs of potential utility for mapping and breeding, and the second was for the specific purpose of identifying candidate SNPs associated with VC accumulation.

For SNPs differing among the three cultivars but not associated with VC, we prepared a reference assembly from each variety. Pooled sequences from all tissues of each cultivar were grouped and contigs assembled, then nonassembling sequences (i.e., unique) were restored to the list, providing a list of unigenes. Each assembled sequence was in turn used as reference for the other two cultivars. The variant list contained 560 to 2144 SNPs between different pairwise combinations of cultivars, while smaller numbers (80–102) were common to any two cultivars (Table 4). The variants are listed in full in Supplementary Tables S15 through S17. Numbers of variants involving comparisons with SSNS-1 are lower, as there was no seed coat library for this variety. Occasionally, more than one variant is found in a transcript.

Cultivated faba bean has a relatively narrow genetic base (Torres et al., 2010), but sufficient variation remains among the three cultivars to provide useful numbers of

Table 5. Faba bean (*Vicia faba* L.) variants present in root, shoot, or seedcoat expressed transcripts of 'CDC Fatima' and 'SSNS-1' relative to 'A01155'. Variants were filtered to include only single nucleotide polymorphisms at 100% frequency. Sequences are shown in Supplementary Table S18.

Sequence	Location in sequence	Reference	Allele	Coverage in CDC Fatima	Coverage in SSNS-1	BLASTx results	E-value
GQNZ9BZ02I6GH0	277	A	G	8	26	Narbonin (*Medicago truncatula* Gaertn.)	10^{-11}
GQNZ9BZ02F6Q02	195	C	G	23	6	Uncharacterized protein LOC101495589 (*Cicer arietinum* L.)	10^{-28}
GQNZ9BZ02I3NLV	39	A	T	3	4	Hypothetical protein (*Acinetobacter baumannii*)	7.5
GQNZ9BZ02J012N	332	A	G	3	4	Unknown (*Medicago truncatula* Gaertn.)	10^{-36}
GQNZ9BZ02FU9BS	145	C	T	3	3	Predicted: Serine-rich adhesin for platelets-like (*Cicer arietinum* L.)	10^{-19}
Contig 158 (1..2228)	373	T	C	3	22	Polyadenylate-binding protein (*Medicago truncatula* Gaertn.)	10^{-76}
Contig 172 (1..1880)	1816	C	T	5	5	Putative adenosylhomocysteinase (*Trifolium pratense* L.)	10^{-33}
Contig 379 (1..1471)	827	C	T	3	3	None	
Contig 450 (1..1955)	1464	C	G	4	3	Endoglucanase 6-like (*Cicer arietinum* L.)	10^{-81}
Contig 500 (1..1479)	666	C	A	5	5	S-adenosylmethionine synthetase-2 (*Pisum sativum* L.)	10^{-90}
Contig 1375 (1..769)	374	C	T	10	16	None	
Contig 2174 (1..704)	641	T	A	5	5	Hypothetical protein MTR_3g086650 (*Medicago truncatula* Gaertn.)	10^{-06}
Contig 4254 (1..1865)	727	T	C	10	4	Protein-tyrosine phosphatase mitochondrial (*Medicago truncatula* Gaertn.)	10^{-75}
Contig 4672 (1..696)	256	C	T	3	4	Predicted: UDP-glucose flavonoid 3-O-glucosyltransferase 7-like (*Cicer arietinum* L.)	10^{-65}
Contig 5045 (1..2208)	1482	G	A	4	5	Stress-induced phosphoprotein (*Medicago truncatula* Gaertn.)	10^{-39}
Contig 5369 (1..1385)	341	A	G	3	17	Hypothetical protein MTR_2g076590 (*Medicago truncatula* Gaertn.)	10^{-78}
Contig 7796 (1..725)	658	G	A	3	3	Predicted: Abscisic stress-ripening protein 3-like (*Solanum lycopersicum* L.)	10^{-09}
Contig 11700 (1..264)	184	G	C	9	5	Predicted: Selenoprotein H-like (*Cicer arietinum* L.)	10^{-36}
Contig 13586 (1..355)	130	A	C	5	3	RNA-binding protein (*Medicago truncatula* Gaertn.)	0.03
Contig 14971 (1..228)	118	C	T	3	7	Predicted: Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex-like (*Cicer arietinum* L.)	10^{-19}
Contig 21734 (1..749)	38	A	G	5	3	None	

variants. CDC Fatima and SSNS-1 appeared more closely related to each other than to A01155 as shown by the smaller number of SNPs differing between them. These variants await confirmation; a high proportion of them would be expected to be functional and should be of use as markers for mapping involving these and other cultivars. However, confirming them is beyond the scope of this work. The number of variants detected represents a minimum estimate; greater sequence depth would increase the proportion of reliable variants between lines.

To identify SNPs potentially associated with the VC synthetic pathway, we sought SNPs that were present in both Fatima and SSNS-1 relative to A01155. We used filtering as provided in CLC-Bio program to screen SNPs for those that satisfied all criteria. Eighty SNPs were identified in both CDC Fatima and SSNS-1 when compared with A01155. These were individually examined and BLASTed. Those with a strong homology match in the Windsor embryo library were discarded. Of the remaining 21 SNP candidates in Table 5, five are based on single reads in the reference cultivar, while the rest are based on multiple reads. Three are completely uncharacterized, while several have potential to be directly related to VC synthesis, and several may be good markers but are most unlikely to be components of the VC pathway. Full sequences appear in Supplementary Table S18. Investigation of these SNP candidates would be of high interest but is beyond the scope of this paper. A sequence identified this way might be within a synthetic or regulatory component of the VC synthetic pathway. Alternatively, such a sequence might be a physically linked marker useful for breeding purposes but not a candidate in terms of function.

Conclusions

These preliminary data are released as a contribution to genetic and genomic research in faba bean, an undervalued crop with considerable potential as a cool-weather forage and food crop. It contains several antinutritional factors, some common to many species and some unique and understudied. We have had a particular interest in analyzing the synthetic pathway of the pyrimidine derivatives VC implicated in favism. We developed strategies using transcriptome sequencing from different cultivars and tissues to optimize the chances of finding either expression differences or variants related to genes in the biosynthetic pathway of VC, the nature of which is unknown, and identified candidate sequences using both strategies. We have also analyzed the pathways of better-known antinutritional factors and identified the deficiency in low-tannin lines.

However, current trends in research in Canada are not favorable for pursuit of this interest to fruition.

Therefore we are releasing the data that was generated for the purpose of identifying critical factors for VC synthesis in the hope that it may be useful to researchers elsewhere who are studying this crop. These critical antinutritional factors, when identified and removed from faba bean, should position this legume as a major rather than a minor crop.

NCBI Accessions

Library data have been submitted to the NCBI BioSample database as accessions SAMN02650916 to SAMN02650923. Individual sequences of the condensed tannin pathway were deposited at GenBank as accessions KP006492 (DFR1), KP006494 (DFR2), KP006495 (ANR), and KP006496 (*tt10*-like).

Supplemental Information Available

Supplemental tables 1 through 18 are available with this article.

Acknowledgments

We thank Jacek Novak, Larissa Ramsay, and Dustin Cram for their expert assistance with bioinformatic and data handling issues, and Janet Condie for the later stages of library preparation and 454 sequencing. Seeds were provided by A. Vandenberg, Crop Development Centre, University of Saskatchewan. This work was supported in part by a NAPGEN sequencing grant (*Na*tural *P*roducts *Gen*omics initiative). This is NRCC publication Number 55994.

References

Abekas Inc. 2011. User operation guide: Applicable to Mira software V3.0.0 and higher. Abekas, Inc., Menlo Park, CA.

Akash, M.W., and G.O. Myers. 2012. The development of faba bean expressed sequence tag–simple sequence repeats (EST-SSRs) and their validity in diversity analysis. Plant Breed. 131:522–530. doi:10.1111/j.1439-0523.2012.01969.x

Benedito, V.A., I. Torres-Jeres, J.D. Murray, A. Andriankaja, S. Allen, K. Kakar, and M.K. Udvardi. 2008. A gene expression atlas of the model legume *Medicago truncatula*. Plant J. 55:504–513. doi:10.1111/j.1365-313X.2008.03519.x

Bock, C., H. Ray, and F. Georges. 2009. Down-regulation of galactinol synthesis in oilseed *Brassica napus* leads to significant reduction of antinutritional oligosaccharides. Botany 87:597–603. doi:10.1139/B09-037

Bremer, E., D.A. Rennie, and R.J. Rennie. 1988. Dinitrogen fixation of lentil, field pea and fababean under dryland conditions. Can. J. Soil Sci. 68:553–562. doi:10.4141/cjss88-053

Burbano, C., C. Cuadrado, M. Muzquiz, and J.I. Cubero. 1995. Variation of fabaism-inducing factors (vicine, convicine and L-DOPA) during pod development in *Vicia faba* L. Plant Foods Hum. Nutr. 47:265–274. doi:10.1007/BF01088335

Cruz-Izquierdo, S., C.M. Avila, Z. Satovic, C. Palomino, N. Gutierrez, S.R. Ellwood, H.T.T. Phan, J.I. Cubero, and A.M. Torres. 2012. Comparative genomics to bridge *Vicia faba* with model and closely related legume species: Stability of QTLs for flowering and yield-related traits. Theor. Appl. Genet. 125:1767–1782. doi:10.1007/s00122-012-1952-1

Dong, J., W. Yan, C. Bock, K. Nokhrina, W. Keller, and F. Georges. 2013. Perturbing the metabolic dynamics of *myo*-inositol in developing *Brassica napus* seeds through in vivo methylation impacts its utilization as phytate precursor and affects downstream metabolic pathways. BMC Plant Biol. 13:84. doi:10.1186/1471-2229-13-84

Duc, G., G. Sixdenier, M. Lila, and V. Furtoss. 1989. Search of genetic variability for vicine and convicine content in *Vicia faba* L. A first report of a gene which codes for nearly zero-vicine and zero-convicine contents. In: J. Huisman, A.F.B. Van der Poel, and I.E. Liener, editors, Recent advances of research in antinutritional factors in legume seeds. Academic Publishers, Wageningen, NLD. p. 305–313.

Ellwood, S.R., H.T.T. Phan, M. Jordan, J. Hane, A.M. Torres, C.M. Avila, S. Cruz-Izquierdo, and R.P. Oliver. 2008. Construction of a comparative genetic map in faba bean (*Vicia faba* L.); conservation of genome structure with *Lens culinaris*. BMC Genomics 9:380. doi:10.1186/1471-2164-9-380

Ewing, B., Hillier, L., Wendl, M.C., and P. Green. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. 8:175–185 doi:10.1101/gr.8.3.175

Goyoaga, C., C. Burbano, C. Cuadrado, C. Romero, E. Guillamón, A. Varela, and M. Muzquiz. 2011. Content and distribution of protein, sugars and inositol phosphates during the germination and seedling growth of two cultivars of *Vicia faba*. J. Food Compos. Anal. 24:391–397. doi:10.1016/j.jfca.2010.11.002

Kaur, S., L.W. Pembleton, N.O. Cogan, K.W. Savin, T. Leonforte, J. Paull, M. Materne, and J.W. Forster. 2012. Transcriptome sequencing of field pea and faba bean for discovery and validation of SSR genetic markers. BMC Genomics 13:104. doi:10.1186/1471-2164-13-104

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

Luthra, P.M., and S. Singh. 2010. Identification and optimization of tyrosine hydroxylase activity in *Mucuna pruriens* DC. var. *utilis*. Planta 231:1361–1369. doi:10.1007/s00425-010-1140-y

Marles, M.A.S., H. Ray, and M.Y. Gruber. 2003. New perspectives on proanthocyanidin biochemistry and molecular regulation. Phytochemistry 64:367–383. doi:10.1016/S0031-9422(03)00377-7

Negruk, V. 2013. Mitochondrial genome sequence of the legume *Vicia faba*. Front. Plant Sci. 4:128. doi:10.3389/fpls.2013.00128

Oomah, B.D., G. Luc, C. Leprelle, J.C.G. Drover, J.E. Harrison, and M. Olson. 2011. Phenolics, phytic acid, and phytase in Canadian-grown low-tannin faba bean (*Vicia faba* L.) genotypes. J. Agric. Food Chem. 59:3763–3771. doi:10.1021/jf200338b

Peterbauer, T., J. Mucha, L. Mach, and A. Richter. 2002. Chain elongation of raffinose in pea seed: Isolation, characterization, and molecular cloning of a multifunctional enzyme catalyzing the synthesis of stachyose and verbascose. J. Biol. Chem. 277:194–200. doi:10.1074/jbc.M109734200

Polowick, P., D. Baliski, C. Bock, H. Ray, and F. Georges. 2009. Development and analysis of transgenic peas with reduced raffinose oligosaccharide content. Botany 87:526–532. doi:10.1139/B09-020

Ray, H., and F. Georges. 2010. A genomic approach to nutritional, pharmacological and genetic issues of faba bean (*Vicia faba*): Prospects for genetic modifications. GM Crops 1:99–106. doi:10.4161/gmcr.1.2.11891

Roche Applied Science. 2009. cDNA rapid library preparation method manual. GS FLX Titanium Series, October 2009 (Rev. Jan 2010). Roche Applied Science, Manheim, Germany.

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

Schmid, M., T.S. Davison, S.R. Henz, U.J. Pape, M. Demar, M. Vingron, and J.U. Lohmann. 2005. A gene expression map of *Arabidopsis thaliana* development. Nat. Genet. 37:501–506. doi:10.1038/ng1543

Torres, A.M., C.M. Avila, N. Gutierrez, C. Palomino, M.T. Moreno, and J.I. Cubero. 2010. Marker-assisted selection in faba bean (*Vicia faba* L.). Field Crops Res. 115:243–252. doi:10.1016/j.fcr.2008.12.002

Torres, A.M., B. Roman, C.M. Avila, Z. Satovic, D. Rubiales, J.C. Sillero, and M.T. Moreno. 2006. Faba bean breeding for resistance against biotic stresses: Towards application of marker technology. Euphytica 147:67–80. doi:10.1007/s10681-006-4057-6

Yamamoto, K., N. Kobayashi, K. Yoshitama, S. Teramoto, and A. Komamine. 2001. Isolation and purification of tyrosine hydroxylase from callus cultures of *Portulaca grandiflora*. Plant Cell Physiol. 42:969–975. doi:10.1093/pcp/pce125