

Parallel Sequencing of Expressed Sequence Tags from Two Complementary DNA Libraries for High and Low Phosphorus Adaptation in Common Beans

Matthew W. Blair,* Andrea C. Fernandez, Fabio Pedraza, Monica C. Muñoz-Torres, Nuwan Sella Kapu, Kathleen Brown, and Jonathan P. Lynch

Abstract

Expressed sequence tags (ESTs) have proven useful for gene discovery in many crops. In this work, our objective was to construct complementary DNA (cDNA) libraries from root tissues of common beans (*Phaseolus vulgaris* L.) grown under low and high P hydroponic conditions and to conduct EST sequencing and comparative analyses of the libraries. Expressed sequence tag analysis of 3648 clones identified 2372 unigenes, of which 1591 were annotated as known genes while a total of 465 unigenes were not associated with any known gene. Unigenes with hits were categorized according to biological processes, molecular function, and cellular compartmentalization. Given the young tissue used to make the root libraries, genes for catalytic activity and binding were highly expressed. Comparisons with previous root EST sequencing and between the two libraries made here resulted in a set of genes to study further for differential gene expression and adaptation to low P, such as a 14 kDa proline-rich protein, a metalloproteinase, tonoplast intrinsic protein, adenosine triphosphate (ATP) citrate synthase, and cell proliferation genes expressed in the low P treated plants. Given that common beans are often grown on acid soils of the tropics and subtropics that are usually low in P these genes and the two parallel libraries will be useful for selection for better uptake of this essential macronutrient. The importance of EST generation for common bean root tissues under low P and other abiotic soil stresses is also discussed.

EXPRESSED SEQUENCE TAGS (ESTs) are the sequences of complementary DNA (cDNA) clones that represent transcribed genes (Hatey et al., 1998). These sequences are derived from cDNA libraries and collections of cDNA clones created from cloning messenger RNA (mRNA) through reverse transcription and ligation to a vector. Expressed sequence tags are considered to be part of the transcriptome and to represent to a greater or lesser degree the genes expressed in the genotypes and tissues at the time point at which the mRNA was extracted. Therefore, a large effort has gone into sequencing cDNA libraries from many different types of tissues under many different treatments in many different

M.W. Blair, A.C. Fernandez, F. Pedraza, and M.C. Muñoz-Torres, Bean Project and Biotechnology Unit, International Center for Tropical Agriculture (CIAT), A.A. 6713, Cali, Colombia; M.W. Blair, Dep. of Plant Breeding, Emerson Hall, Cornell Univ., Ithaca, NY 14853; A.C. Fernandez, current address: Department of Agronomy, Sugarbeet and Bean Research Unit, USDA-ARS, East Lansing, MI; F. Pedraza, current address: Sun Seeds, ND; M.C. Muñoz-Torres, current address: Dep. of Biology, Georgetown Univ., Washington, DC 20057; N.S. Kapu, current address: Forest Products Biotechnology/Bioenergy Group, Faculty of Forestry, The University of British Columbia, 2424 Main Mall, Forest Science Ctr., Vancouver, BC, Canada, V6T 1Z4; K. Brown and J.P. Lynch, Dep. of Horticulture, 102 Tyson Building, Pennsylvania State Univ., University Park, PA 16802. Received 19 Feb. 2011. *Corresponding author (mwb1@cornell.edu).

Abbreviations: ALV, Andean leaf library; ATP, adenosine triphosphate; cDNA, complementary DNA; CUGI, Clemson University Genomics Institute; dT, ?; EST, expressed sequence tag; GO, gene ontology; HP, high phosphate; LP, low phosphate; LVS, Mesoamerican leaf library; MIP, membrane intrinsic protein; NCBI, National Center for Biotechnology Information; nt, nucleotides; NOD, Mesoamerican nodule library; POD, Mesoamerican bean pod library; PSU, Pennsylvania State University; RNase, ribonuclease; RT-PCR, reverse transcription-polymerase chain reaction; RTS, Mesoamerican root library; SSH, suppression subtractive hybridization; TIP, tonoplast intrinsic protein; TF, transcription factor.

Published in The Plant Genome 4:204–217. Published 13 Sept. 2011.
doi: 10.3835/plantgenome2011.02.0093
© Crop Science Society of America
5585 Guilford Rd., Madison, WI 53711 USA
An open-access publication

All rights reserved. No part of this periodical may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Permission for printing and for reprinting the material contained herein has been obtained by the publisher.

genotypes or species. Expressed sequence tags make up a grand proportion of the GenBank nucleotide sequence collection (Benson et al., 2008).

Large numbers of cDNA libraries have been constructed for several crops with minimal sequencing of each library to determine the genes expressed in each of many different tissues. For instance, several projects were funded in the late 20th and early 21st centuries in the model species *Arabidopsis thaliana* (L.) Heynh. (mouse-ear cress) and *Medicago truncatula* Gaertn. (barrel clover) as well as the crops *Oryza sativa* L. (rice) and *Glycine max* (L.) Merr. (soybean) (for example, Vodkin et al., 2004). Since then more dispersed and disparate efforts have been made for EST generation from more specific tissues and more minor crops. Among the legumes, common bean (*Phaseolus vulgaris* L.) has a relatively small number of ESTs in the National Center for Biotechnology Information (NCBI) database (Benson et al., 2008) currently standing at around 80,000 sequences. Next generation sequencing of the transcriptome has not been published for this crop although it is a new alternative to individual cDNA clone-by-clone, EST sequencing (Varshney et al., 2009).

Libraries of cDNA clones are very useful for analysis of gene expression in a given crop under given growth conditions. These collections are physically stored so can they be returned to for probe use, resequencing, or other experimental purposes (Hatey et al., 1998). They often contain specific genes for the tissue being analyzed as well as housekeeping genes expressed in a wide range of tissues (Vodkin et al., 2004). The frequency of each type of gene can be measured by sequencing of nonnormalized libraries, and this has been the library type most frequently built for common bean. Expressed sequence tag sequencing of cDNA libraries provides information on the length of a gene as well as where translated open reading frames start and stop and where poly-A sites are located.

Common bean is among the most popular annual, short-season crops grown on earth due to its varying seed types and plant architectures, N-fixing capability, easy fit into different production systems, and good taste in various types of dishes with over 20 t grown globally per year (Broughton et al., 2003). As one of the grain legumes it is different from soybean or peanut (*Arachis hypogaea* L.) in that it is almost entirely consumed cooked and whole and is not processed for oil or feed. It is therefore an important pulse crop in the diets of the poor, for whom it provides the nutrients not available from cereals. Common bean like other pulses is complementary as a food source to crops such as rice, which have lower protein and are micronutrient poor. Common beans are primarily grown on small- to medium-scale farms in the tropics and subtropics on marginal or semi-marginal acid to neutral soils that are low in essential plant nutrients (Lynch, 2007).

Soil P is variable throughout the growing regions for common beans but tends to be low on volcanic-derived or highly weathered soils where common beans are often grown in the tropics or subtropics (Vance et al., 2003).

Adaptation to low P soils has been found by comparing the yield of various genotypes under high and low P availability (Yan et al., 1995). However, a series of adaptation mechanisms to low P have been suggested for legumes (Lynch and Brown, 2001; Wang et al., 2010). Root mechanisms such as organic acid exudation, increased root hair length and density, development of a greater amount of metabolically cheaper root classes, and perhaps greater activity of P uptake transporters are thought to be important (Lynch and Brown, 2001; Yan et al., 2004; Lynch, 2007). In addition, the use of mycorrhizal symbionts or development of shallower root systems have been suggested as ways for genotypes to adapt to low P soils, although efficiency in P use is also involved in getting the most yield from a limited P supply (Lynch and Beebe, 1995; Liao et al., 2001, 2004). The genes underlying these traits are poorly understood. The objectives of this research, therefore, were to create two cDNA libraries for root tissue from a common bean genotype exposed to low and high P growing conditions, to compare EST sequences from the libraries and detect differences in gene expression, and to identify unique genes as candidates for low P adaptation. In addition one of our goals was to compare these ESTs to previous root and nonroot EST libraries from Ramirez et al. (2005). Finally this EST sequencing effort was used to generate sequences for the genotype DOR364, which has been used as a low P susceptible parent for our principal genetic mapping population. The generation of these ESTs will help us to understand the reaction of a susceptible genotype to high versus low P status.

Materials and Methods

Plant Material and RNA Extraction

Plants of the genotype DOR364 were grown under high and low P conditions (P sufficiency or deficiency, respectively) for 5 d in a greenhouse at Pennsylvania State University (PSU) as follows. Seeds were sterilized with bleach, germinated, and grown in a nutrient solution with or without P. For both treatments, seedlings were grown in rolled up germination paper that was wrapped in foil and the bottom of the roll kept in nutrient solutions of 0 or 1 mmol P for the two treatments, respectively, which was replaced as used. Seedlings in the 0 mmol P solution had only the seed reserves as a P source. Tissues were collected from tap roots and from lateral and basal roots of plants grown under both conditions and frozen immediately in liquid N.

It is notable that DOR364 is a widely grown Central American variety, developed as an advanced line by CIAT, and known to be sensitive to low P soils but to be resistant to bean golden yellow mosaic virus and to be commercially acceptable as a small red genotype (Beebe et al., 2006). The pedigree of this genotype is BAT1215 × (RAB166 × DOR125) and it has been used as a parent in a large number of crosses for countries of Central America and the Caribbean, having also been released as a variety in Cuba, Honduras, El Salvador, and Nicaragua.

Total RNA was extracted from the tissue of each treatment by grinding to a fine powder using a ribonuclease (RNase)-free treated mortar and pestle and more liquid N followed by addition of a TRIzol reagent (Invitrogen, Carlsbad, CA) extraction buffer and associated techniques for nucleic acid isolation following manufacturer guidelines to produce two RNA samples. These RNA pellets were resuspended in RNase-free water and quantified spectrophotometrically at Abs₂₆₀. RNA quality was checked through denaturing agarose gels (1.5%) containing formaldehyde, which were stained with ethidium bromide. Size selection between 1 and 3 kb in length was performed in the same gels but run in 1x Tris-acetate-ethylenediaminetetraacetic acid (EDTA) buffer. Poly-A mRNA was selected with Dynabeads (DynaL Corp., Raleigh NC).

Library Construction and Expressed Sequence Tag Sequencing

Two cDNA libraries were constructed from the mRNAs isolated as described above. The procedure followed was based on the UniZAP-cDNA synthesis kit (Stratagene, La Jolla, CA) as follows: mRNAs from both treatments were isolated by oligo dT cellulose and separately primed with hybrid oligo dT linker-primer oligonucleotides that contained an *Xho*I restriction site and that primed from the poly-A site of the mRNAs. StrataScript reverse transcriptase (Stratagene, Roche Applied Science, Indianapolis, IN) and 5-methyl deoxycytidine triphosphate (dCTP) were used to create single-stranded cDNAs. Following double-stranded cDNA synthesis with Polymerase I, *Eco*RI adaptors were ligated to the blunt ends and the samples were digested with *Xho*I. The results were cDNAs with an *Eco*RI sticky end on one side and an *Xho*I sticky end on the other side. The cDNA samples were then size-fractionated on an agarose gel, purified, and directionally ligated in the arms of the UniZAP XR vector. The resulting lambda phage libraries were packaged in Gigapack III Gold extracts ((Stratagene, Roche Applied Science) and transformed into *Escherichia coli* cell line XL-1-Blue MRF' for in vivo excision.

Cells were plated on 100 mg L⁻¹ Carbenicillin (EMD-4Biosciences, Merck KGaA, Darmstadt, Germany), 0.55 mmol isopropyl-β-d-thiogalactopyranoside (IPTG), and 80 μg mL⁻¹ X-Gal for blue-white screening. A total of 36,000 white (insert-containing) bacterial clone colonies were picked and arrayed into 96-well plates (luria broth, 4% glycerol, and Carbenicillin 100 mg L⁻¹) at the Clemson University Genomics Institute (CUGI) with automatic blue-white selection settings and standard parameters on a Q-bot robot (Genetix, New Milton, UK). Glycerol stocks were copied once into working and master copies of the libraries and stored at -80°C. Each of the libraries was spotted onto gridded, positively charged Amersham Hybond N+ nylon membranes (GE Healthcare, Piscataway, NJ) with six fields each with a 96-position double-replicate 4 by 4 pattern. Following CUGI's nomenclature system, the low P library was named Pv_DEe and the high P library was named Pv_DEd, where Pv stands for the

species name, D for the genotype, and E for the purpose of the libraries, which was to generate ESTs.

For the EST analysis, 3360 clones were mini-prepped through a modified alkaline lysis method according to CUGI protocols and sequenced from the 5' end of the clone using a T3 primer using an ABI3700 capillary sequencer (Applied Biosystems, Foster City, CA) at Clemson University. All sequences were screened for vector fragments to check for insert integrity. The sequences were also evaluated for Phred (Ewing et al., 1998) values and were trimmed using SeqTrim (Falgueras et al., 2011) to remove vector, adaptors, and poly-A sequences and low quality (Phred quality scores < 20) end sections of the sequence. Sequences of a length shorter than 100 bases after this trim process were omitted.

Data Assembly

Trimmed ESTs sequences were submitted to NCBI (GenBank numbers pending) and assembled using CAP3 (Huang and Madan, 1999) with default parameters (gap penalty factor, $N > 6$; segment pair score cutoff, $N > 40$; overlap length score cutoff, $N > 80$; etc.) to assemble ESTs into contigs. Annotation of all putative gene sequences derived from the resulting unigenes was performed using Blast2GO (Conesa and Götzt, 2008). Blastx (Altschul et al., 1997) was then used to define the function of each unigene by comparison to the nonredundant NCBI database. The cutoff E-value for sequence similarity was set at 10⁻³ and the top 20 Blastx hits were recorded as well as the distribution of top or most frequent E-values and species hits for the entire set of unigenes. Gene mapping or gene ontology (GO) terms were assigned based on the GO database also with Blast2GO. Analyzing EST sequences against the soybean database allowed further annotation of the genes and each one was classified according to molecular function, biological processes, and cellular component using Blast. After assembly, the two libraries were compared to find overlapping genes and genes with differential expression under the high and low P conditions.

Results

Library Construction and Expressed Sequence Tag Sequencing Results

Both libraries were successfully constructed from RNA extracted from seedling root tissue grown with low (no external) P (LP) or high P (HP) in the wick system for germination and growth at PSU. Due to the young age of the plants, seedlings had a tap root with young, developing lateral roots and 6 to 8 basal roots but no adventitious roots. Given the presence of seed reserves, lack of P availability had no significant effect on growth or dry weight at this plant stage as shown in previous studies as well (Bonser et al., 1996). Total RNA was carefully frozen and shipped to the CIAT for library construction. Given that a UniZAP phage vector (Stratagene) was used for both libraries it was a simple matter to construct the libraries in CIAT and to ship the in vivo excised *E. coli* cells to

CUGI for picking, library copying, mini-prepping, and sequencing. Because we were more interested in evaluating the LP library than the HP library, we sequenced 23 plates of the LP library (2208 clones) and 15 plates of the HP library (1440 clones) for a total of 3648 sequencing reactions from 38 plates of the two libraries.

Of the 3648 clones that were picked and used in sequencing reactions, 3346 successful sequences were obtained (a success rate of 91.7%) of which 2008 were from the LP library and 1338 were from the HP library (Table 1). SeqTrim (Falgueras et al., 2011) was used to filter out vector sequences and to determine sequence quality resulting in 1684 high quality sequences for the LP library and 1137 high quality sequences for the HP library.

Despite the differences in the number of high quality sequences, sequencing quality was proportionally similar in both the LP library (83.9%) and the HP library (84.5%). Any difference in sequencing quality might have been due to different plates being run at different times on the ABI automated capillary sequencer (Applied Biosystems). After the SeqTrim procedure, the average length of the HP library expressed sequence tags was 704 nucleotides (nt) while the average length of the LP expressed sequence tags was 659 nt with an overall average of 677.1 nt across both libraries.

Unigene Set Annotation

CAP3 (Huang and Madan, 1999) assembly resulted in a total of 948 unigenes (850 singletons and 98 contigs) for the ESTs from the HP library and 1424 unigenes (1292 singletons and 132 contigs) for the ESTs from the LP library (Table 1). Because the greater number of unigenes in the LP library was due to a higher proportion of singletons, the average length of the unigenes was lower in this library than in the HP library.

Blast2GO (Conesa and Götzt, 2008) and Blastx were very useful when annotating most unigenes from both libraries, and the vast majority of them in each library had significant sequence similarity to gene sequences from the nonredundant GenBank database (Benson et al., 2008). Among the HP library ESTs, 67.3% of the unigenes had a gene annotation and an assigned function while, for the LP library ESTs, 66.9% of the unigenes had such annotation and function determined.

Some unigenes from each library had a Blastx hit but no annotation, comprising 139 (14.7% of the total number of unigenes in the library) in the case of the HP library and 176 (12.4%) in the case of the LP library. In contrast, the number of unigenes with no hit was higher in the case of the LP library comprising 294 unigenes (19.9%) compared to 171 unigenes (18.0%) in the HP library. This difference could have been due to the shorter average length of the LP unigenes (685 nt) compared to the HP unigenes (734 nt). The size range of the unigenes from both libraries is shown in Supplemental Fig. S1 and ranged from very small ESTs or unigenes (100 to 200 nt in length with 100 nt being the minimum used to declare an EST) up to larger unigenes especially (1.1 to 1.2 kb).

However the overall average length of the cleaned and assembled unigenes was 705 nt and this was near the median for unigene length. Finally, the total number of unigenes with significant hits and/or annotation was 1906 (1591 with annotation and 315 with a hit but without annotation). The total number of unigenes with no hits was 465. The top hit of a unigene generally agreed with the top hits of the ESTs of that unigene. The E-value distributions for the ESTs in the HP library and in the LP library ranged from the minimum 10^{-3} to maximums in each case of 10^{-180} (Fig. 1) although the decline in frequency of hits (below 100) was seen at approximately 10^{-85} for the HP library to 10^{-105} for the LP library. Due to the higher number of ESTs in the second library compared to the first, it was not surprising that the maximum number of hits for a given 10^{-e} value was 340 for the LP library ESTs while it was 225 for the HP library ESTs. The maximum number of hits had significance values in the range of 10^{-10} to 10^{-20} for each library, which was not surprising given that we retrieved the top 20 hits per EST.

Top species hits for each of the libraries in general were with soybean (Fig. 2), as expected given the close phylogenetic relationship with common bean, both legumes being members of the tropical clade. Other top hits were with predicted genes of dicotyledenous species with full genome sequences such as grape (*Vitis vinifera* L.), poplar (*Populus balsamifera* L. subsp. *trichocarpa* (Torr. & A. Gray ex Hook.) Brayshaw), and castor bean (*Ricinus communis* L.). The next most common top hits were with genes from other legumes either common bean itself or *Medicago truncatula*, *Lotus corniculatus* L. var. *japonicus* Regel [syn. *Lotus japonicus* (Regel) K. Larsen], *Cicer arietinum* L., and various *Vigna* species.

Blast2GO (Conesa and Götzt, 2008) was also useful in mapping of genes to a GO term, with only 52 unigenes from either library not being mapped. This success rate was possible because of the four different mapping functions used to obtain GO information. The NCBI identified GeneNames in the species-specific entries of the GeneProduct table from the GO database. Gene identifiers were then used to retrieve UniProt IDs from the Protein Information Resource (Altschul et al., 1997), a nonredundant protein database that includes PDS, UniProt, Swiss-Prot, TrEMBL, and TAIR. For both the HP and LP library the majority of ESTs had similarity to named genes in the UniProtKB (nonredundant) database and second in the TAIR (*Arabidopsis thaliana* (L.) Heynh.) database (Supplemental Fig. S2). The highest evidence code for the Blastx hits and for the individual sequences (Fig. 3) was through inferred electronic annotation and second by direct annotation. Finally, the number of gene ontologies found during the EST annotation per sequences was 2 to 4 for up to 100 genes in the HP library and 150 genes in the LP library (Fig. 4) indicating protein domains were sometimes identified that were shared by several different ESTs.

Categorization of the unigenes by their GO terms and associated information was done for biological

processes and molecular function at the second level of complexity and for cellular compartmentalization at the third level of complexity according to Blast2GO (Table 2). This was useful in distinguishing the predominant biological and molecular role of each unigene and in which category they were most likely to be located for either library and to note any differences in unigene categorization between the HP and LP libraries, which might lead to the identification of candidate genes.

In terms of biological categorization a large number of unigenes were involved in “general cellular processes” (28.7 and 29.9% in the HP and LP libraries, respectively) and “metabolic processes” (27.6 and 28.0%), which was not surprising given that the root tissues sampled came from actively growing seedling. Within metabolic processes the most commonly found terms were binding (40.4% average) and catalytic activities (33.8%) with “structural molecular activity” only close to a tenth of the genes (9.3%) categorized in this function. Again, this was to be expected for actively growing roots, although the amount of binding activity was surprising. Meanwhile, in “cellular compartmentalization” most gene products had a predicted location in the cytoplasm (or cell part with 41.5 and 40.4% for HP and LP libraries, respectively) with a large amount in different types of organelles (up to 26.2 and 26.8% for the two libraries, respectively). Under “biological regulation” there were two categories (cell proliferation and pigmentation) that were only expressed under LP conditions, while under “molecular function” there was one metallochaperone found in LP but not under HP conditions.

Overall, there was high and significant correlation between the gene categorization in each library for biological processes ($r = 0.998$, $p = 0.001$), molecular function ($r = 0.999$, $p = 0.001$), and cellular compartmentalization ($r = 0.999$, $p = 0.001$). This reflected the similarity in developmental stage for each cDNA library and in growing conditions of the seedlings for each of the two treatments with the exception of P availability.

Comparison with Other Bean Expressed Sequence Tag Libraries

After performing the functional analysis described above, we were interested in comparing all the unigenes from the HP ($n = 948$) and LP ($n = 1424$) libraries with all the ESTs assembled from the libraries of Ramírez et al. (2005) using the same protocol of assembly we had used in the present work (Table 3). We found a total of 15,333 unigenes from across the five sublibraries of Ramírez et al. (2005) with average cleaned read lengths ranging from 730 nt for the Andean leaf library (ALV, made from G19833) to 541 nt for the Mesoamerican root library (RTS, made from Negro Jamapa) and a total of 9566 unigenes with the most individual sequences from the RTS library and the least from the Mesoamerican leaf library (LVS). The number of contiged ESTs ranged from 1363 for the shorter RTS sequences to 2109 for the longer Mesoamerican nodule library (NOD) clones, indicating redundancy in the latter

Table 1. Raw and high quality (HQ) sequences, percent successful sequencing, and HQ sequencing as well as singletons, contigs, and unigenes found in and across the high and low P root expressed sequence tag (EST) libraries.

EST numbers	High P	Low P	Total
Plates sequenced [†]	15	23	38
Raw sequences	1440	2208	3648
Successful sequences	1338	2008	3344
Percent success	99.6	99.6	91.7%
HQ sequences	1137	1684	2821
Percent HQ	84.5	83.9	84.4
Average length	704 nt [‡]	659 nt	677.1 nt
Singletons	850	1292	2142
Contigs [§]	98	132	230
Unigenes [§] (total)	948	1424	2372
Average length	734 nt	685 nt	704.6 nt
Gene annotation [¶]	638	953	1591
Hit but no annotation [¶]	139	176	315
No hit [§]	171	294	465

[†]Assumes 96-well plates sequenced on ABI3700 automated sequencer (Applied Biosystems, Foster City, CA).

[‡]nt, nucleotides.

[§]According to CAP3 assembly (Huang and Madan, 1999) and default parameters.

[¶]According to Blast2GO annotation (Conesa and Göt, 2008) and Blastx searches.

library (data not shown). The number of singletons ranged from 1825 for RTS to 1163 for LVS while the number of contigs ranged from 255 for LVS to 477 for RTS.

In library-to-library comparison, we found that the new libraries from DOR364 shared the most hits with the Mesoamerican RTS library made from Negro Jamapa root tissue submitted to low P soil conditions (319 sequences for LP and 227 for HP). Comparisons with the other libraries showed that the least number of hits to the HP or LP unigenes was with the ALV library, followed by a few more hits with the LVS library, followed by the NOD library and finally the Mesoamerican bean pod library (POD) (Table 3). As expected given its source tissue (roots), stress condition (low P) and gene pool association (Mesoamerican), the RTS library from Negro Jamapa had the highest percentage of unique and single hits with both the HP and LP library of all the library comparisons. Up to 24% of the genes shared a hit between these libraries while only 10 to 18% of the genes shared a hit with the ALV or LVS libraries. Surprisingly, the HP and LP root libraries shared 20 to 22% of genes with the POD library.

In this study we found a total of 24 genes expressed in the HP library and 36 in the LP library that were categorized by molecular function as transcriptional regulators. We used these to compare to the study of transcription factors that were differentially expressed with reverse transcription-polymerase chain reaction (RT-PCR) analysis in nodules of Negro Jamapa plants grown in high and low P evaluated by Hernández et al. (2007) of 372 different genes from the unigene set

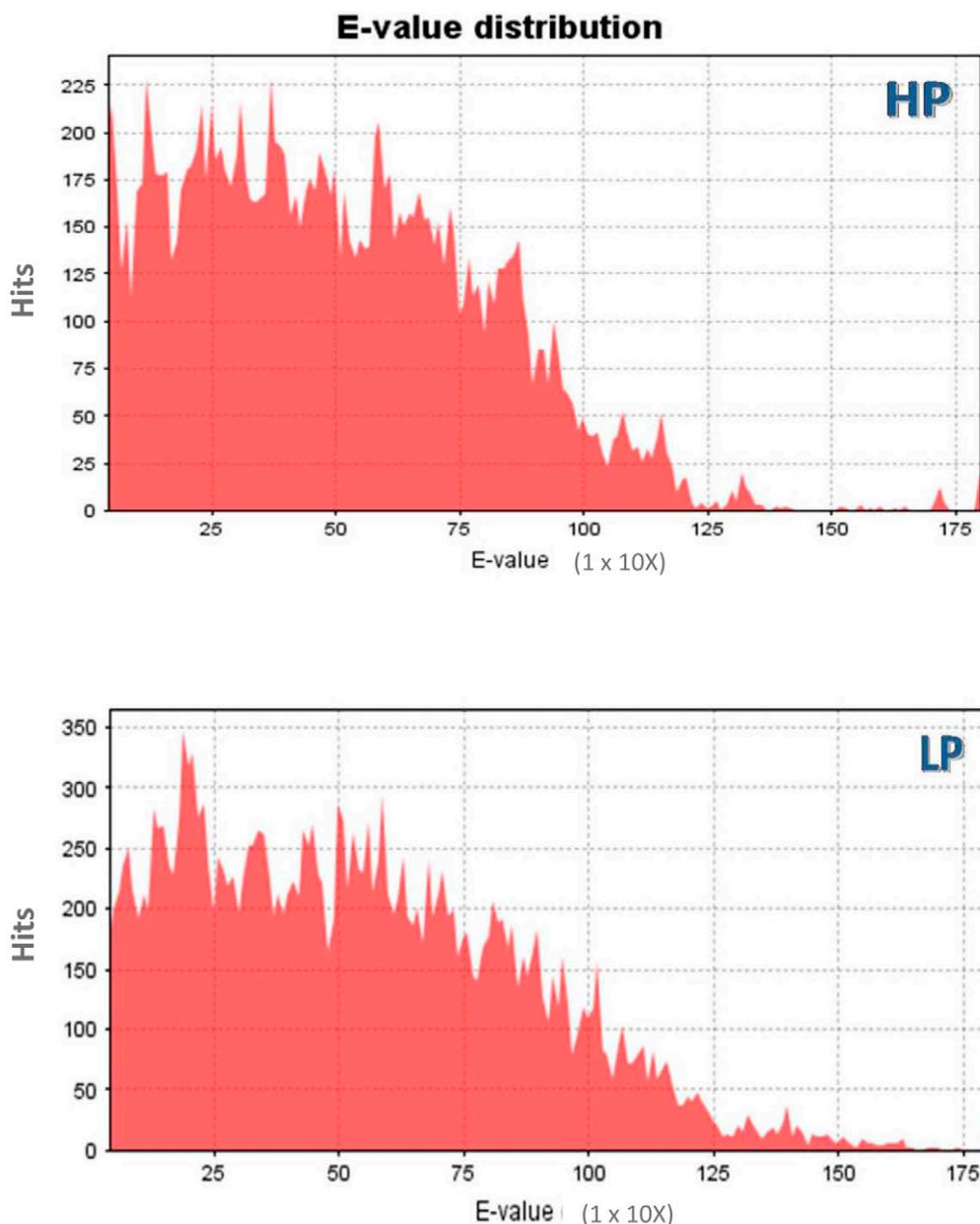


Figure 1. Distribution of expressed sequence tag Blastx hits in terms of E-value for the expressed sequence tags (ESTs) from the sequencing of the high P (HP) and low P (LP) complementary DNA (cDNA) libraries.

of Ramírez et al. (2005). We found some overlap in the same categories of transcription factors but cannot specify if they are the same exact genes or not because overlaps were partial between the sequences and many of the genes belonged to complex gene families such as the large number of AP2-domain containing transcription factors.

Differential Expression between the High and Low Phosphorus Libraries

Even more relevant than the comparisons with other bean EST libraries, we found that by comparing the HP and LP collections of ESTs we were able to identify significant differences in gene expression. These were most likely to represent additional valid differences in gene expression between HP and LP growth where all other

conditions were shared. In this analysis of the unigenes described above we found 163 genes that were shared under the two conditions (17.2% of the total in the case of the HP unigenes and 11.5% of the LP unigenes). This ratio was higher if we only considered those genes with annotation (25.6% and 17.2%, respectively, for the HP and LP unigenes).

Meanwhile, the numbers of unigenes that were unique to each of the libraries were 785 in the case of the HP library and 1261 in the case of the LP library considering all unigenes. However, this number was reduced to only 475 and 790 unigenes when considering those unigenes with full annotation and Blastx hits, which were those of most interest since they were the ones functionally assigned and mapped in gene ontology. It

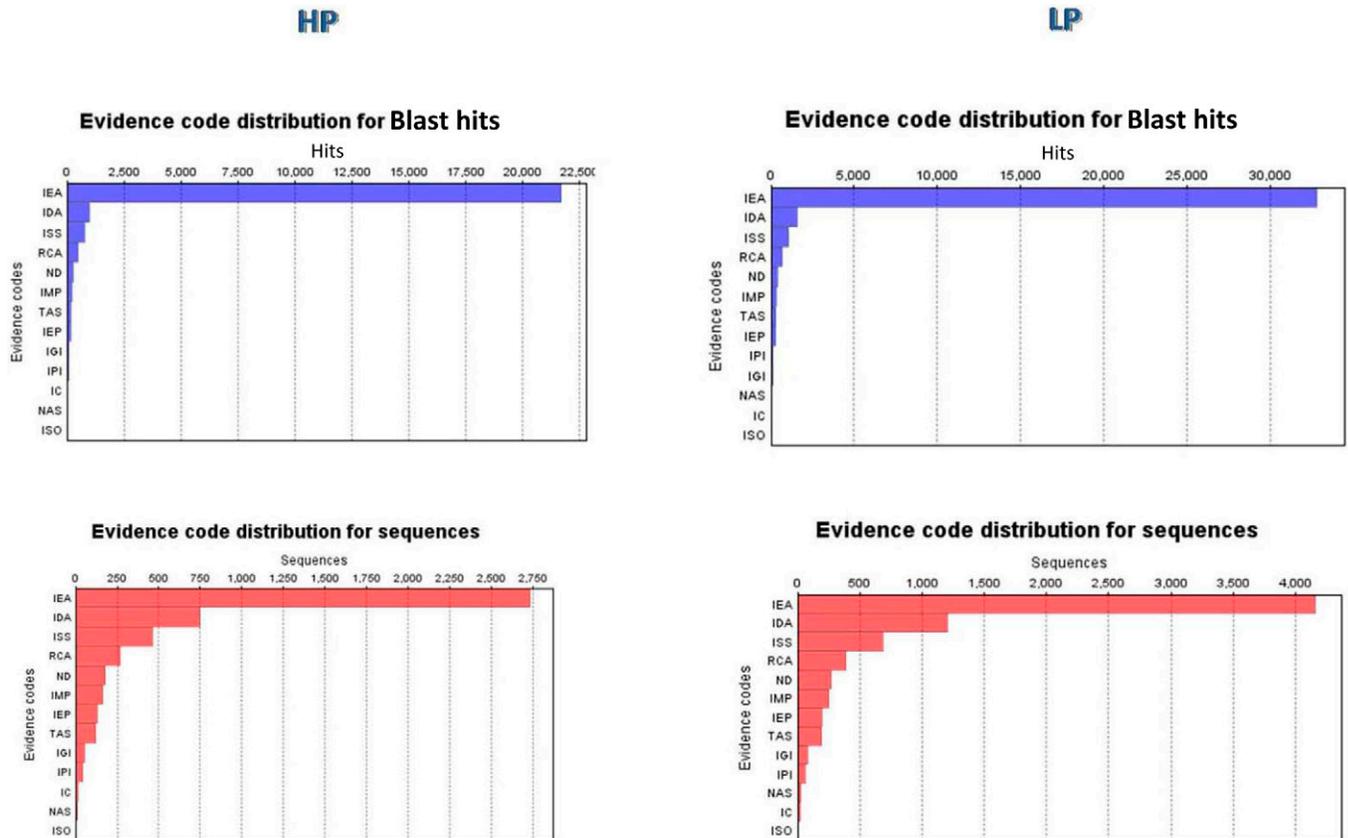


Figure 3. Gene ontology evidence codes for Blastx hits of individual expressed sequence tag (EST) sequences annotated for the high P (HP) and low P (LP) complementary DNA (cDNA) libraries.

is worth noting that we used a high threshold value for the determination of genes shared by the two libraries of 10^{-150} to ensure that they were true homologs and not just orthologous or paralogous members of a single gene family. At lower threshold, that is, 10^{-3} or 10^{-20} , a larger number of shared genes were found between the HP and LP libraries (data not shown). Following the initial analysis of common unigenes between the two libraries we were most interested in determining whether there were certain unigenes with a different level of expression between the HP and LP libraries. To do this we compared the number of ESTs within each of the unigenes, considering both the HP and LP versions of those homologs that were common between the conditions. The greatest differences in expression were found for a set of 34 unigenes (Table 4).

These differentially expressed unigenes ranged in differential expression from a factor of 1:6 to 1:24 (with correlation of delta factors) where a delta value of >5 was used as the threshold for determining differential expression among the sequenced clones of the two libraries so as to counteract the effect of the 50% greater number of clones sequenced from the LP than the HP collection. These differential genes were about evenly split between genes expressed more in HP conditions (22 unigenes) compared to LP condition (12 unigenes). For the remainder of the unigenes delta factors were <5 and for almost half of the shared genes no difference in

expression level ($\Delta = 0$) was observed. Interestingly among the group of 34 unigenes highlighted for their high differential expression there were many repeat hits suggesting that different parts of the same genes were captured in the EST collection or that various members of the same gene family with varying expression were important. Some of these unigenes included a set of 14 kDa proline-rich protein (one gene overexpressed in HP and a different one overexpressed in LP), a set of elongation factor proteins (all overexpressed in HP), and a set of pathogenesis-related proteins (one overexpressed in HP and another overexpressed in LP). Additionally, two types of tonoplast intrinsic proteins (TIPs) and a single type of cyclophilin gene were overexpressed only in the LP library. Lastly, some of the genes differentially expressed under either HP or LP conditions had no hits in the overall analysis and were not studied further.

Discussion

The principal achievement of this study was to create two root-based common bean cDNA libraries and to evaluate them through EST sequencing of over 3600 clones. The root libraries were of interest because they were from young tissue at the initial stages of exposure to low P stress, when they are responding to the stress but have not yet begun to display overall growth reduction (Bonser et al., 1996; Basu et al., 2007). This differentiates the LP library described here from the RTS library described

Annotation distribution

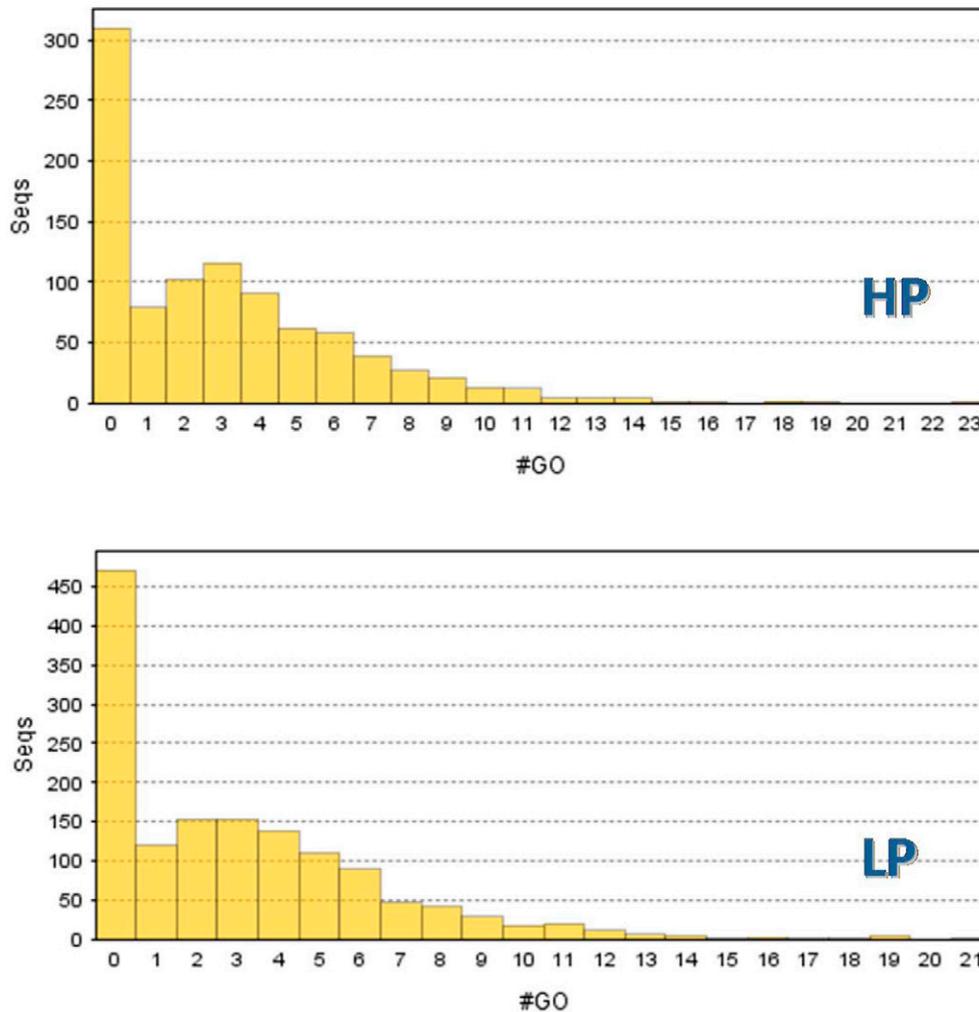


Figure 4. The number of gene ontologies for each expressed sequence tag (EST) sequences found during the EST sequence annotation differentiating the searches performed for the high P (HP) and low P (LP) library ESTs. Seqs, number of sequences; #GO, number of gene ontologies.

in Ramírez et al. (2005), which was made from older plants (21 days after planting), when P stress has caused multiple direct and indirect effects on growth and development. Expressed sequence tags differentially expressed in the DOR364 libraries would be expected to represent early responses to P stress, while those in the RTS library from Ramírez et al. (2005) would represent longer-term stress responses.

In addition, we used a newer, small red seeded advanced line from CIAT, named DOR364, which has been the subject of several studies of root traits associated with adaptation to LP (Bonser et al., 1996; Basu et al., 2007; Liao et al., 2001, 2004), while Ramírez et al. (2005) used an older, more variable, black bean variety from Mexico, Negro Jamapa 81, which has been the subject of several molecular studies (Hernández et al., 2007, 2009) but which has variable sources. DOR364 is also pertinent because it continues to be used as a variety in Central America and Cuba while Negro Jamapa has been replaced by more modern black beans in Mexico

and elsewhere. Furthermore DOR364 is the parent of an important genetic population (Blair et al., 2003).

Finally, in our study, we did not use rhizobial inoculation that was used in root material for the cDNA library construction in Ramírez et al. (2005), which could confound effects of low P on roots themselves. Also, unlike the previous effort, our HP library provided a control for the LP library given that it was produced from plants grown in exactly the same way as the LP plants except for addition of sufficient P. Notably the tissues of HP plants were harvested at exactly the same time as the LP plants. Root cDNA libraries are of interest because much of the plants' interaction with the soil environment is via genes expressed in the root system.

Our interest in comparing the Ramírez et al. (2005) libraries to our new LP and HP library set was that we used a similar approach of sequencing approximately three and a half thousand clones per library, assembling the ESTs, and designating unigenes per library. In addition, the five libraries of Ramírez et al. (2005) were of

Table 2. Differences in categorization of the unigenes from the high P (HP) and low P (LP) root expressed sequence tag (EST) libraries according to biological processes, molecular function, and cellular compartmentalization.

Gene Ontology	No. HP [†]	No. LP [†]	Total No.	Percent HP	Percent LP
Biological process[‡]					
Biological regulation	83	121	204	6.5	6.5
Carbon utilization	1	2	3	0.1	0.1
Cell proliferation	0	4	4	0.0	0.2
Cellular component organization	45	74	119	3.5	4.0
Cellular processes	365	555	920	28.7	29.9
Cellular component biogenesis	36	44	80	2.8	2.4
Developmental process	55	81	136	4.3	4.4
Death	3	4	7	0.2	0.2
Immune system process	5	6	11	0.4	0.3
Localization	95	117	212	7.5	6.3
Metabolic process	351	520	871	27.6	28.0
Multiorganism process	23	31	54	1.8	1.7
Multicellular organismal process	45	70	115	3.5	3.8
Nitrogen utilization	2	2	4	0.2	0.1
Pigmentation	0	1	1	0.0	0.1
Rhythmic processes	1	1	2	0.1	0.1
Reproduction	29	43	72	2.3	2.3
Response to stimulus	129	181	310	10.2	9.7
Sulfur utilization	2	2	4	0.2	0.1
Molecular function[‡]					
Antioxidant activity	16	12	28	1.9	1.0
Binding	337	504	841	40.1	40.6
Catalytic activity	278	426	704	33.1	34.3
Electron carrier activity	27	34	61	3.2	2.7
Enzyme regulator activity	10	11	21	1.2	0.9
Metallochaperone activity	0	1	1	0.0	0.1
Molecular transducer activity	7	8	15	0.8	0.6
Nutrient reservoir activity	2	3	5	0.2	0.2
Structural molecular activity	80	113	193	9.5	9.1
Translation regulator activity	12	28	40	1.4	2.3
Transporter activity	47	66	113	5.6	5.3
Transcription regulator activity	24	36	60	2.9	2.9
Cellular component[§]					
Apoplast	22	36	58	1.9	2.0
Cell part	490	710	1200	41.5	40.4
Extracellular region	1	1	2	0.1	0.1
Organelle part	148	216	364	12.5	12.3
Organelle – membrane bounded	309	470	779	26.2	26.8
Organelle – non-membrane bound	108	157	265	9.2	8.9
Protein complex	54	80	134	4.6	4.6
Protein–DNA complex	10	13	23	0.8	0.7
Vesicle	38	74	112	3.2	4.2

[†]Refers to number of unigenes in each case for HP or LP root libraries.

[‡]Performed at level 2 for greater details of biological process and molecular function.

[§]Performed at level 3 for greater details of cellular component placement.

interest because they were constructed from various tissues and from both Andean and Mesoamerican sources in the case of leaves. Finally, the potential to compare root ESTs from low P tissues was interesting as it showed a high similarity between genes expressed in our root library from DOR364 and the previous root library from Negro Jamapa. This was true of both HP and LP libraries, respectively, with up to 22 or 24% of the genes having a unique hit and 30 or 29% having a similar hit, showing that the sampling of mRNA was similar for one-third of the genes between the new root libraries and the previous RTS library. Comparison with libraries from Ramírez et al. (2005) showed 319 genes in common with our HP library and 227 in common with our LP library.

Apart from our study here and that of Ramírez et al. (2005), there are only a few other low P libraries in common bean or in other legumes. Suppression subtractive hybridization (SSH) libraries have been made, however, by Tian et al. (2007) for common bean roots and shoots and by Guo et al. (2008) for soybean. The library from Tian et al. (2007) contained too few entries for a detailed comparison with either the RTS library of Ramírez et al. (2005) or the presently constructed library. Meanwhile, Hernández et al. (2007) created an RT-PCR assay from transcription factors (TFs) discovered by Ramírez et al. (2005), which has proven useful for evaluating gene expression in common bean and which we used to compare with putative TFs discovered during our sequencing. In those comparisons we found only a small number of the same homologs using a strict cutoff threshold given the large number of gene-family members in most TF families. More specific parts of the root system such as root hairs during *Rhizobium* spp. infection and various stages of nodule development have also been studied by EST sequencing usually with differential display being a bottleneck to full analysis of the transcriptome of each of these tissues such as in the study by Tian et al. (2007).

The study of genes expressed under low P conditions is important because this abiotic stress affects more than half of the bean production area in Africa and Latin America and usually affects subsistence farmers who grow beans on more marginal, hillside soils (Lynch, 2007). Yet this is not an intractable problem. Breeding evidence and genetic studies have shown that beans have a capacity for adaptation to low P soils via different mechanisms, including root hair proliferation and exudation of organic acids to capture inorganic P (Liao et al., 2004) or the production of long shallower basal roots that search more of the topsoil for available P (Lynch and Brown, 2001; Yan et al., 2004). Beans also vary in their production of adventitious roots to capture soil P later in the life of the plant when demand is highest (Ochoa et al., 2006) and potentially in mycorrhizal symbiosis. Finally aerenchyma can be used to construct roots with lower metabolic cost (Lynch and Ho, 2005; Nielsen et al., 2001; Postma and Lynch, 2011). However, to date, no cDNA libraries have been constructed from adventitious roots, aerenchyma tissue, or mycorrhizal-infected common bean root tissue.

Table 3. Comparison of high P (HP) and low P (LP) root derived expressed sequence tags (ESTs) with the EST libraries of Ramírez et al. (2005).

Library used for the comparison with HP and LP libraries	No. of ESTs	No. of Unigenes	HP library			LP library				
			ESTs with unique hit	Percent	ESTs with hit	Percent	ESTs with unique hit	Percent	ESTs with hit	Percent
ALV (Andean leaf library)	2555	1684	96	10	117	12	164	12	212	15
LVS (Mesoamerican leaf library)	2704	1418	153	16	171	18	205	14	254	18
NOD (Mesoamerican nodule library)	3901	2247	162	17	188	20	277	19	330	23
POD (Mesoamerican bean pod library)	2985	1945	213	22	260	27	284	20	366	26
RTS (Mesoamerican root library)	3188	2302	227	24	286	30	319	22	409	29

Table 4. Unigenes with the most different gene expression between the high P (HP) and low P (LP) root libraries.

Query identification	Description	EST ¹ no.	Subject identification	Description	EST no.	Δ	Percent identity
LP_Contig36	Elongation factor 1-	7	HP_Contig51	Elongation factor 1-	2	5	85.7
LP_Contig6	tonoplast intrinsic protein	8	HP_Contig15	Tonoplast intrinsic protein	3	5	97.2
LP3b09	Elongation factor 1-	1	HP_Contig61	Elongation factor 1-	6	5	98.3
LP5c08	Elongation factor 1-alpha	1	HP_Contig61	Elongation factor 1-	6	5	84.7
LP6f10	14 kDa proline-rich protein	1	HP_Contig72	Extensin-like protein	6	5	85.6
PV_DEd0001b_b1_D07	Elongation factor 1-	1	HP_Contig81	Elongation factor 1-	6	5	98.9
PV_DEd0001b_b1_H04	Elongation factor 1-	1	HP_Contig81	Elongation factor 1-	6	5	98.9
PV_DEd0001b_b1_H04	Elongation factor 1-	1	HP_Contig61	Elongation factor 1-	6	5	87.9
PV_DEd0002c_b1_A01	Elongation factor-1 alpha	1	HP_Contig81	Elongation factor 1-	6	5	98.2
PV_DEd0002d_b1_C08	Elongation factor-1 alpha	1	HP_Contig81	Elongation factor 1-	6	5	79.1
PV_DEd0002d_b1_F10.	Elongation factor 1-	1	HP_Contig81	Elongation factor 1-	6	5	99.3
PV_DEd0002d_b1_F10	Elongation factor 1-	1	HP_Contig61	Elongation factor 1-	6	5	86.9
LP_Contig36	Elongation factor 1-	7	HP1e06	No hits	1	6	97.8
LP_Contig2	No hits	10	HP_Contig29	No hits	3	7	99.4
LP_Contig51	No hits	4	HP_Contig98	No hits	11	7	93.9
LP_Contig6	Tonoplast intrinsic protein	8	HP3f07	Tonoplast intrinsic protein	1	7	93.9
LP_Contig6	Tonoplast intrinsic protein	8	HP1b03	Tonoplast intrinsic protein	1	7	86.5
LP_Contig76	No hits	8	HP1f04	No hits	1	7	93.9
LP_Contig132	Cyclophilin	10	HP2b07	cyclophilin	1	9	96.0
LP_Contig39	No hits	2	HP_Contig98	No hits	11	9	100.0
LP1c01	No hits	1	HP_Contig32	No hits	11	10	97.9
LP7b02.	No hits	1	HP_Contig32	No hits	11	10	99.7
LP4e01	14 kDa proline-rich protein	1	HP_Contig58	14 kDa proline-rich protein	12	11	92.1
LP6g05	14 kDa proline-rich protein	1	HP_Contig58	14 kDa proline-rich protein	12	11	94.5
PV_DEd0001a_F0	14 kDa proline-rich protein	1	HP_Contig58	14 kDa proline-rich protein	12	11	97.6
PV_DEd0001c_C10	Proline-rich protein	1	HP_Contig58	14 kDa proline-rich protein	12	11	97.4
LP_Contig81	Pathogenesis-related protein	13	HP4g08	Pathogenesis-related protein	1	12	98.2
LP_Contig81	Pathogenesis-related protein	13	HP2d01	Pathogenesis-related protein	1	12	91.0
LP_Contig85	14 kDa proline-rich protein	24	HP_Contig58	14 kDa proline-rich protein	12	12	99.7
LP4b12.g1.ab1	Pathogenesis related protein	1	HP_Contig24	Pathogenesis-related protein	15	14	99.0
PV_DEd0002c_F11	Pathogenesis-related protein	1	HP_Contig24	Pathogenesis-related protein	15	14	80.4
PV_DEd0003b_B12	Pathogenesis related protein	1	HP_Contig24	Pathogenesis-related protein	15	14	88.2
LP_Contig85	14 kDa proline-rich protein	24	HP1b11	14 kDa proline-rich protein	1	23	97.1
LP_Contig85	14 kDa proline-rich protein	24	HP2e01	14 kDa proline-rich protein	1	23	92.6

¹EST, expressed sequence tag.

It is worth highlighting that we purposefully sequenced more ESTs from the low P library than from the high P library to find the genes that were expressed under low P conditions in a susceptible genotype such as

DOR364. We sampled both lateral and basal roots when doing the tissue collection and RNA extraction; therefore, most parts of the root system were sampled including root hairs and taproots. Because of this generalized

sampling procedure, the same one followed by Ramírez et al. (2005), we are not able to distinguish if a specific EST is from one of these subtissues; however, we can determine that expression of the genes was from young tissues as our source plants were seedlings compared to the larger root system of the previous study. In terms of other nutrients, the supplies were normal for growth including micronutrients and we hoped to also capture genes that might be important for mineral uptake such as those involved in iron homeostasis, which has been studied for the DOR364 genotype (Blair et al., 2010).

In terms of those genes we had an interest in for micronutrient uptake, we did an initial screening of genes related to the iron uptake pathway such as iron reductase (Blair et al., 2010), which might affect mineral because DOR364 in addition to being sensitive to low P conditions is also a low accumulator of seed iron and poor in iron reductase activity (Blair et al., 2009, 2010). The libraries, therefore, can be useful for the discovery of various root expressed genes and/or could be used in further EST sequencing or gene discovery. We conducted initial testing of high density filters for each of the libraries and used them in hybridization experiments to find ferritin, nicotianamine synthase, and nicotianamine transferase homologs with success.

The DOR364 libraries and the same hybridization-based search could be used for adding to sequence information for other genes of interest such as P transporters and transcription factors involved in P stress (Hernández et al., 2007; Valdéz-López et al., 2008). Phosphatases and many other categories of genes involved in uptake including PIDS-4 and PvPS2 reported by Tian et al. (2007) during P starvation in the accession G19833 could also be important. At least three of the differentially expressed P-starvation genes found by Tian et al. (2007) were also present in our EST collections (data not shown) although E-values were low due to the small sequence lengths of the SSH fragments from these authors. It would also be interesting to capture the DOR364 alleles of genes involved in acid exudation (Yan et al., 2004), genes that are involved in P uptake such as P transporters (M. Harrison, personal communication, 2010), or other genes responsible for root architecture traits associated with P efficiency (Lynch, 2007).

Another interesting result of this study, was that our *in silico* analysis itself resulted in a set of genes that would be worthy of study for adaptation to low P growth conditions. The comparison of differences in EST sequence frequency between the HP and LP libraries showed that the most interesting putatively differentially expressed genes including a large set of 14 kDa proline-rich proteins, some of which were overexpressed in HP and some in LP conditions. It has been shown that the 14 kDa proline-rich protein genes are specifically expressed in the cortical ground meristem of root apex of common bean and that expression declines as the analysis moves spatially further from the root tip (Choi et al., 1996). Although the role of the 14 kDa proline-rich proteins is not precisely understood they

are important in areas of high cell division and differentiation in roots and are thought to be induced by P starvation in beans and across various other species of legumes (Graham et al., 2006; Hernández et al., 2007). It was interesting that some of the mRNAs for these proteins were specific to each of the P treatments confirming that there may be a small, differentially expressed gene family encoding these proteins.

Additionally, we found a set of differentially expressed translation elongation factor proteins, which are related to new cell generation; these are of particular interest because cell division, ethylene responsiveness, and meristem activity are increased as part of the low P response (Ma et al., 2003). Meanwhile, other interesting genes in LP conditions included those for two types of TIPs, which are members of major cell membrane intrinsic protein (MIPs) family that might be involved in vacuole accumulation of water and small molecules (Juah et al., 1999; Johanson et al., 2001; Takahashi et al., 2004). Graham et al. (2006) also reported a single TIP gene as an important gene for P starvation. Hernández et al. (2007) found a differentially expressed aquaporin, which is also a member of the MIP family of genes, but did not classify it into a specific type. A total of 33 of the 126 differentially up- or downregulated genes identified by Hernández et al. (2007) were found in our EST collection. Finally, a single type of cyclophin gene was also overexpressed only in the LP library in this study but not in the study of Hernández et al. (2007).

An additional set of putatively differentially expressed genes was found by observing those unigenes found only in the LP library but not in the HP library. These included cell proliferation and pigmentation category genes in the case of the biological processes and genes with metallochaperone activity in the case of molecular function. For this latter category, we analyzed one EST from the LP library (LP1.c05.g1) identified as a metalloproteinase by comparisons to similar soybean genes. The metalloproteinase family is known to be part of a large and diverse family of peptidases (Rawlings and Barrett, 1993). The role of this gene in adaptation to low P conditions is uncertain; however, peptidases and proteases of various sorts are expressed under stress conditions such as drought, although this would be the first report in beans of expression under low P stress. If we find more peptidases that are differentially expressed, the MEROPS database (Rawlings et al., 2010) would be a useful location to compare these genes. Meanwhile, the gene from the pigmentation category was the clone PV_DEd002a_A11 and was related to a group of soybean-annotated genes for adenosine triphosphate (ATP) citrate synthase, which is involved in producing citrate.

Roots often exude citrate as an efficient way to mobilize P or other elements in the soil, which are then taken up by specific transporters (Neumann and Martinola, 2002). Therefore it was not surprising to find this gene in the LP library and involved in low P tolerance and not expressed in the HP library. Finally, the four genes in the

cell proliferation category, all expansins, were of interest because this finding might shed light on root elongation under low P stress, which is known to be related to the cell wall formation and ethylene signal transduction pathways (Ma et al., 2003). Expansins are also involved in cell growth and proliferation and in lateral and tap root growth in soybean (Guo et al., 2008). These might be related through an interconnected pathway with the elongation factors that were found differentially expressed in HP conditions. Querying the soybean genome with each candidate EST proved helpful in establishing whether they shared any sequence similarity with well-studied soybean genes.

All of the genes of interest described above could be used in RT-PCR, northern blot, or macroarray studies to confirm their differential expression in differentially treated root tissue with HP and LP conditions. Of interest also would be time course of parallel gene expression under the two treatments. Our attempts to use the macroarrays from Ramírez et al. (2005) failed based on RNA quality after shipment between laboratories, but RT-PCR is a simpler technique to carry out in house and has the advantage of creating stable cDNA immediately as part of the assay technique. Unfortunately there has been no high-throughput effort to create microarrays for common bean but this study shows that such a project would be valuable especially in the context of root tissue gene analysis.

Conclusions

In common bean there have been few cDNA libraries made from root tissues with only one library made specifically for roots under low P conditions before this study (Ramírez et al., 2005). Unlike the root EST library from that study our libraries were from younger plants grown with high P or low P conditions equivalent to germination in sufficient or highly deficient soils, respectively. In addition to adding significantly to the number of sequences of root ESTs we also diversified the number of cDNA libraries creating one for HP conditions and one for LP conditions both from the currently used and released variety DOR364. Just as the previous library was useful for analyzing differentially expressed genes through hybridizations and for discovering TFs that are involved in response to low P soils using RT-PCR (Hernández et al., 2007), our study identified a set of putatively differentially expressed genes for low P adaptation based on different EST frequencies in the HP and LP libraries.

Some of our differentially expressed genes matched those found in a cross-species comparison with *Arabidopsis thaliana*, *Glycine max*, *Medicago truncatula*, and *Lotus albus* Janka (unresolved) conducted by Graham et al. (2006) and in the macroarray experiments of Hernández et al. (2007). The advantage of our study was that we created two similar libraries from roots harvested at the same age. The only difference between them was the P treatment, and the seedlings were uninoculated with *Rhizobia* spp. and in very early stages of P-stress

response, so the candidate genes we identified are likely to be involved in development of adaptive processes of roots to P deficiency.

In summary, the present EST sequences derived from the two new cDNA libraries complement the sequences from Ramírez et al. (2005) and the analyses of Graham et al. (2006) and Hernández et al. (2007). The research community working on root responses to P stress will be able to make use of the new root ESTs so as to have better characterized root expressed genes. It was notable that we found differences between the libraries both in transcription factors as well as in adaptive genes for abiotic stress tolerance more generally (14 kDa proteins, expansins, and MIP and TIP genes). Further sequencing of the library might uncover differential transcription factors expressed under high versus low P since these genes are often less abundantly expressed and their corresponding mRNAs are rare. All of this work is aimed at finding genes that control important root traits such as basal root growth angle (Lynch and Brown, 2001), root acid exudation (Yan et al., 2004), or P uptake efficiency (Beebe et al., 2006) to the sufficient degree to convert them into molecular markers to assist in breeding projects.

Supplemental Information Available

Supplemental material is available free of charge at <http://www.crops.org/publications/tpg>.

Acknowledgments

Sequencing was carried out by M. Atkins, M. Palmer, and C. Saski. We appreciate the support provided by CIAT (D. Cortes and J. Tohme) and by CUGI (R. Wing, J. Tomkins, and D. Main) for access to computational platforms and some supplies needed during the course of this work. We are especially grateful to the USAID linkage grant program and the Generation Challenge Program for funding this research.

References

- Altschul, S.F., T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D.J. Lipman 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucl. Acids Res.* 25 (17):3389–3402. doi:10.1093/nar/25.17.3389
- Basu, P., Y.J. Zhang, J.P. Lynch, and K.M. Brown. 2007. Ethylene modulates genetic, positional, and nutritional regulation of root plagio-gravitropism. *Funct. Plant Biol.* 34:41–51. doi:10.1071/FP06209
- Beebe, S.E., M. Rojas, X. Yan, M.W. Blair, F. Pedraza, F. Muñoz, J. Tohme, and J.P. Lynch. 2006. Quantitative trait loci for root architecture traits correlated with phosphorus acquisition in common bean. *Crop Sci.* 46:413–423. doi:10.2135/cropsci2005.0226
- Benson, D.A., I. Karsch-Mizrachi, D.J. Lipman, J. Ostell, and D.L. Wheeler. 2008. GenBank. *Nucleic Acids Res.* 36:25–30. doi:10.1093/nar/gkm929
- Blair, M.W., C. Astudillo, M. Grusak, R. Graham, and S. Beebe. 2009. Inheritance of seed iron and zinc content in common bean (*Phaseolus vulgaris* L.). *Mol. Breed.* 23:197–207. doi:10.1007/s11032-008-9225-z
- Blair, M.W., S.J.B. Knewton, C. Astudillo, C.M. Li, A.C. Fernandez, and M.A. Grusak. 2010. Variation and inheritance of iron reductase activity in the roots of common bean (*Phaseolus vulgaris* L.) and association with seed iron accumulation QTL. *BMC Plant Biol.* 10:215. doi:10.1186/1471-2229-10-215
- Blair, M.W., F. Pedraza, H.F. Buendía, E. Gaitán-Solís, S.E. Beebe, P. Gepts, and J. Tohme. 2003. Development of a genome-wide anchored microsatellite map for common bean (*Phaseolus vulgaris* L.). *Theor Appl Genet* 107: 1362–1374.

- Bonser, A.M., J. Lynch, and S. Snapp. 1996. Effect of phosphorus deficiency on growth angle of basal roots in *Phaseolus vulgaris*. *New Phytol.* 132:281–288. doi:10.1111/j.1469-8137.1996.tb01847.x
- Broughton, W.J., G. Hernández, M.W. Blair, S. Beebe, P. Gepts, and J. Vanderleyden. 2003. Beans (*Phaseolus* spp.) – Model food legumes. *Plant Soil* 252:55–128. doi:10.1023/A:1024146710611
- Choi, D.W., J.Y. Song, Y.M. Kwon, and S.G. Kim. 1996. Characterization of a cDNA encoding a proline-rich 14 kDa protein in developing cortical cells of the roots of bean (*Phaseolus vulgaris*) seedlings. *Plant Mol. Biol.* 30:973–982. doi:10.1007/BF00020808
- Conesa, A., and S. Götz. 2008. Blast2GO: A comprehensive suite for functional analysis in plant genomics. *Int. J. Plant Genomics* 2008:619832. doi:10.1155/2008/619832
- Ewing, B., L. Hillier, M.C. Wendl, and P. Green. 1998. Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. *Genome Res.* 8:175–185.
- Falgueras, J., A. Lara, N. Fernández-Pozo, F. Cantón, G. Pérez-Trabado, and G. Claros. 2011. SeqTrim: A high-throughput pipeline for pre-processing any type of sequence read. *BMC Bioinformatics* 11:38. doi:10.1186/1471-2105-11-38
- Graham, M.A., M. Ramírez, V. Valdes-Lopez, M. Lara, M. Tesfaye, C.P. Vance, and G. Hernández. 2006. Identification of candidate phosphorus stress induced genes in *Phaseolus vulgaris* through clustering analysis across several plant species. *Func. Plant Biol.* 33:789–797.
- Guo, W., L. Zhang, J. Zhao, H. Liao, C. Zhuang, and X. Yan. 2008. Identification of temporally and spatially phosphate-starvation responsive genes in *Glycine max*. *Plant Sci.* 175:574–584. doi:10.1016/j.plantsci.2008.06.007
- Hatey, F., G. Tosser-Klopp, K. Cloucard-Martinato, P. Mulsant, and F. Gasser. 1998. Expressed sequence tags for genes: A review. *Genet. Sel. Evol.* 30:521–524. doi:10.1186/1297-9686-30-6-521
- Hernández, G., M. Ramírez, O. Valdés-López, M. Tesfaye, M.A. Graham, T. Czechowski, A. Schlereth, M. Wandrey, A. Erban, C. Cheung, H.C. Wu, M. Lara, C.D. Town, J. Kopka, M.K. Udvardi, and C.P. Vance. 2007. Phosphorus stress in common bean: Root transcript and metabolic responses. *Plant Physiol.* 144:752–767. doi:10.1104/pp.107.096958
- Hernández, G., O. Valdes-Lopez, M. Ramirez, N. Goffard, G. Weiller, R. Aparicio-Fabre, S.I. Fuentes, A. Erban, J. Kopka, and M.K. Udvardi. 2009. Global changes in the transcript and metabolic profiles during symbiotic nitrogen fixation in phosphorus-stressed common bean plants. *Plant Physiol.* 151:1221–1238. doi:10.1104/pp.109.143842
- Huang, X., and A. Madan. 1999. CAP3: A DNA sequence assembly program. *Genome Res.* 9:868–877. doi:10.1101/gr.9.9.868
- Johanson, U., M. Karlsson, I. Hohanesson, S. Gustavsson, S. Sjövall, L. Frayse, A.R. Weig, and P. Kjellbom. 2001. The complete set of genes encoding major intrinsic proteins in Arabidopsis provides a framework for a new nomenclature for major intrinsic proteins in plants. *Plant Physiol.* 126:1358–1369. doi:10.1104/pp.126.4.1358
- Juah, G.-Y., T.E. Phillips, and J.C. Rogers. 1999. Tonoplast intrinsic protein isoforms as markers for vacuolar functions. *Plant Cell* 11:1867–1882. doi:10.1105/tpc.11.10.1867
- Liao, H., G. Rubio, X. Yan, A. Cao, K.M. Brown, and J.P. Lynch. 2001. Effect of phosphorus availability on basal root shallowness in common bean. *Plant Soil* 232:69–70. doi:10.1023/A:1010381919003
- Liao, H., X. Yan, G. Rubio, S.E. Beebe, M.W. Blair, and J.P. Lynch. 2004. Genetic mapping of basal root gravitropism and phosphorus acquisition efficiency in common bean. *Funct. Plant Biol.* 31:959–970. doi:10.1071/FP03255
- Lynch, J.P. 2007. Roots of the second green revolution. *Aust. J. Bot.* 55:493–512. doi:10.1071/BT06118
- Lynch, J.P., and S.E. Beebe. 1995. Adaptation of beans (*Phaseolus vulgaris* L.) to low phosphorus availability. *HortScience* 30:1165–1171.
- Lynch, J.P., and K.M. Brown. 2001. Topsoil foraging – An architectural adaptation to low phosphorus availability. *Plant Soil* 237:225–237. doi:10.1023/A:1013324727040
- Lynch, J.P., and M.D. Ho. 2005. Rhizoeconomics: Carbon costs of phosphorus acquisition. *Plant Soil* 269:45–56. doi:10.1007/s11104-004-1096-4
- Ma, Z., T.I. Baskin, K.P. Brown, and J.P. Lynch. 2003. Regulation of root elongation under phosphorus starvation involves changes in ethylene responsiveness. *Plant Physiol.* 131:1381–1390. doi:10.1104/pp.012161
- Neumann, G., and E. Martinola. 2002. Cluster roots: An underground adaptation for survival in extreme environments. *Trends Plant Sci.* 7:1–14. doi:10.1016/S1360-1385(02)02241-0
- Nielsen, K.L., A. Eshel, and J.P. Lynch. 2001. The effect of phosphorus availability on the carbon economy of contrasting common bean (*Phaseolus vulgaris* L.) genotypes. *J. Exp. Bot.* 52:329–339. doi:10.1093/jexbot/52.355.329
- Ochoa, I.E., M.W. Blair, and J.P. Lynch. 2006. QTL analysis of adventitious root formation in common bean (*Phaseolus vulgaris* L.) under contrasting phosphorus availability. *Crop Sci.* 46:1609–1621. doi:10.2135/cropsci2005.12-0446
- Postma, J.A., and J.P. Lynch. 2011. Theoretical evidence for the functional benefit of root cortical aerenchyma in soils with low phosphorus availability. *Ann. Bot. (London)* (in press).
- Ramírez, M., M.A. Graham, L. Blanco-López, S. Silvente, A. Medrano-Soto, M.W. Blair, G. Hernández, C.P. Vance, and M. Lara. 2005. Sequencing and analysis of common bean ESTs: Building a foundation for functional genomics. *Plant Physiol.* 137:1211–1227. doi:10.1104/pp.104.054999
- Rawlings, N.D., and A.J. Barrett. 1993. Evolutionary families of peptidases. *Biochem. J.* 290:205–218.
- Rawlings, N.D., A.J. Barrett, and A. Bateman. 2010. MEROPS: The peptidase database. *Nucleic Acids Res.* 38:D227–233. doi:10.1093/nar/gkp971
- Takahashi, H., M. Rai, T. Kitagawa, S. Morita, T. Masemura, and K. Tanaka. 2004. Differential localization of tonoplast intrinsic proteins on the cellular membrane of the protein body seed type II and the aluerone grain rice seed. *Biosci. Biotechnol. Biochem.* 68:1728–1736. doi:10.1271/bbb.68.1728
- Tian, J., P. Venkatachalam, H. Liao, X. Yan, and K. Raghothama. 2007. Molecular cloning and characterization of phosphorus starvation responsive genes in common bean (*Phaseolus vulgaris* L.). *Planta* 227:151–165. doi:10.1007/s00425-007-0603-2
- Valdés-López, O., C. Arenas-Huertero, M. Ramírez, L. Girard, F. Sánchez, C.P. Vance, J.L. Reyes, and G. Hernández. 2008. Essential role of MYB transcription factor: PvPHR1 and microRNA: PvmiR399 in phosphorus-deficiency signaling in common bean roots. *Plant Cell Environ.* 31:1834–1843. doi:10.1111/j.1365-3040.2008.01883.x
- Vance, C.P., C. Uhde-Stone, and D.L. Allan. 2003. Phosphorus acquisition and use: Critical adaptations by plants for securing a non-renewable resource. *New Phytol.* 157:423–447. doi:10.1046/j.1469-8137.2003.00695.x
- Varshney, R.K., T.J. Close, N.K. Singh, D.A. Hoisington, and D.R. Cook. 2009. Orphan legume crops enter the genomics era! *Curr. Opin. Plant Biol.* 12:202–210. doi:10.1016/j.pbi.2008.12.004
- Vodkin, L.O., A. Khanna, R. Shealy, S.J. Clough, D.O. Gonzalez, R. Philip, G. Zabala, F. Thibaud-Nissen, M. Sidarous, M.W. Ström-vik, C.E. Schmidt, E. Retzel, J. Erpelting, R.C. Shoemaker, A.M. Rodriguez-Huete, J.C. Polacco, V. Coryell, P. Keim, G. Gong, L. Liu, J. Pardinias, and P. Schweitzer. 2004. Microarrays for global expression constructed with a low redundancy set of 27,500 sequenced cDNAs representing an array of developmental stages and physiological conditions of the soybean plant. *BMC Genomics* 5:73. doi:10.1186/1471-2164-5-73
- Wang, X., X. Yan, and H. Liao. 2010. Genetic improvement of phosphorus efficiency in soybean: A radical approach. *Ann. Bot. (London)* 106:215–222. doi:10.1093/aob/mcq029
- Yan, X., S.E. Beebe, and J.P. Lynch. 1995. Genetic variation for phosphorus efficiency in common bean in contrasting soil types II. Yield response. *Crop Sci.* 35:1094–1099. doi:10.2135/cropsci1995.0011183X003500040029x
- Yan, X., H. Liao, S.E. Beebe, M.W. Blair, and J.P. Lynch. 2004. QTL mapping of root hairs and acid exudation traits and their relationship to phosphorus uptake in common bean. *Plant Soil* 265:17–29. doi:10.1007/s11104-005-0693-1