

INVITED REVIEW

Legume genomics: understanding biology through DNA and RNA sequencing

Jamie A. O'Rourke^{1,2,†}, Yung-Tsi Bolon^{2,†}, Bruna Bucciarelli^{1,2} and Carroll P. Vance^{1,2,*}

¹United States Department of Agriculture, Agricultural Research Service, University of Minnesota, St. Paul, MN 55108, USA and

²Department of Agronomy and Plant Genetics, University of Minnesota, 1991 Upper Buford Circle, St. Paul, MN 55108, USA

* For correspondence. E-mail vance004@umn.edu

† These authors contributed equally to this work.

Received: 26 August 2013 Returned for revision: 9 January 2014 Accepted: 13 March 2014 Published electronically: 25 April 2014

• **Background** The legume family (Leguminosae) consists of approx. 17 000 species. A few of these species, including, but not limited to, *Phaseolus vulgaris*, *Cicer arietinum* and *Cajanus cajan*, are important dietary components, providing protein for approx. 300 million people worldwide. Additional species, including soybean (*Glycine max*) and alfalfa (*Medicago sativa*), are important crops utilized mainly in animal feed. In addition, legumes are important contributors to biological nitrogen, forming symbiotic relationships with rhizobia to fix atmospheric N₂ and providing up to 30 % of available nitrogen for the next season of crops. The application of high-throughput genomic technologies including genome sequencing projects, genome re-sequencing (DNA-seq) and transcriptome sequencing (RNA-seq) by the legume research community has provided major insights into genome evolution, genomic architecture and domestication.

• **Scope and Conclusions** This review presents an overview of the current state of legume genomics and explores the role that next-generation sequencing technologies play in advancing legume genomics. The adoption of next-generation sequencing and implementation of associated bioinformatic tools has allowed researchers to turn each species of interest into their own model organism. To illustrate the power of next-generation sequencing, an in-depth overview of the transcriptomes of both soybean and white lupin (*Lupinus albus*) is provided. The soybean transcriptome focuses on analysing seed development in two near-isogenic lines, examining the role of transporters, oil biosynthesis and nitrogen utilization. The white lupin transcriptome analysis examines how phosphate deficiency alters gene expression patterns, inducing the formation of cluster roots. Such studies illustrate the power of next-generation sequencing and bioinformatic analyses in elucidating the gene networks underlying biological processes.

Key words: Legume genomics, next-generation sequencing, NGS, bioinformatics, *Glycine max*, soybean, *Lupinus albus*, white lupin, transcriptome, genome re-sequencing, RNA-seq.

INTRODUCTION

With rapidly declining costs and increasing efficiencies, DNA and RNA high-throughput sequencing, via so-called next-generation technologies, are quickly becoming a mainstay in plant genetics, genomics and biochemistry (Chia and Ware, 2011; Ozsolak and Milos, 2011). Rather amazingly, multiple genomes have been sequenced for several species including: arabidopsis (Cao *et al.*, 2011; Gan *et al.*, 2011), rice (Huang *et al.*, 2013), maize (Lai *et al.*, 2010; Jiao *et al.*, 2012), soybean (Kim *et al.*, 2010; Lam *et al.*, 2010), *Medicago truncatula* (Branca *et al.*, 2011) and chickpea (Varshney *et al.*, 2013). Next-generation sequencing (NGS) of plant genomes is providing new insight into the genomic variation within and between species [single nucleotide polymorphisms (SNPs), presence–absence variations (PAVs), copy number variations (CNVs) and other types of structural variations (SVs)], the introgression of genetic elements such as quantitative trait loci (QTLs), the evolution of organs and pathways, and the status of epigenetic modifications (McMullen *et al.*, 2009; Atwell *et al.*, 2010; Huang *et al.*, 2010; Branca *et al.*, 2011; Schmitz and Zhang, 2011; Schneeberger and Weigel, 2011; Jiao *et al.*, 2012; Tomato Genome Consortium, 2012; Varshney *et al.*, 2013). Likewise, sequencing of plant transcriptomes (RNA-seq) is

providing atlases of gene expression, demonstrating alternative splicing events, defining QTL introgressions, characterizing small RNAs and identifying genes and pathways involved in acclimation to biotic and abiotic stresses (Wang *et al.*, 2009; Li *et al.*, 2010; Libault *et al.*, 2010; Severin *et al.*, 2010a, b; Zenoni *et al.*, 2010; Matas *et al.*, 2011; Yang *et al.*, 2011; Meyer *et al.*, 2012; O'Rourke *et al.*, 2013b; Atwood *et al.*, 2014). Next-generation sequencing approaches are replacing chip-based platforms for genomic studies. This change is placing extraordinary pressure on access to bioinformatics, computer speed and storage, and interpretation of data (Fig. 1). The bottleneck to solving long-standing issues in plant biology will be assembling and sifting through terabytes of data from these studies.

The legume family (Leguminosae) is second only to the grass family in economic and nutritional value (Graham and Vance, 2003). Approximately 17 000 species of legumes are found in nature. Many of these are important grain, pasture and agroforestry species. Grain and forage legumes are grown on some 200 Mha, or about 15 % of the Earth's arable surface. About 33 % of human's dietary nitrogen (N) needs are derived from legumes (maybe up to 60 % in developing countries). Moreover, upwards of 40 % of the world's cooking oil comes from soybean and peanut (Vance, 2001). In the past decade,

legumes have come to rival arabidopsis and maize as model systems for genomic studies. This has occurred not only because of their economic and nutritive value but also due to the fact that they, in symbiosis with soil bacteria collectively known as rhizobia, can provide their individual N needs through symbiotic nitrogen fixation (SNF). The following contribution will focus on addressing issues related to legume biology through the use of NGS. In so doing, references to other species may also be cited.

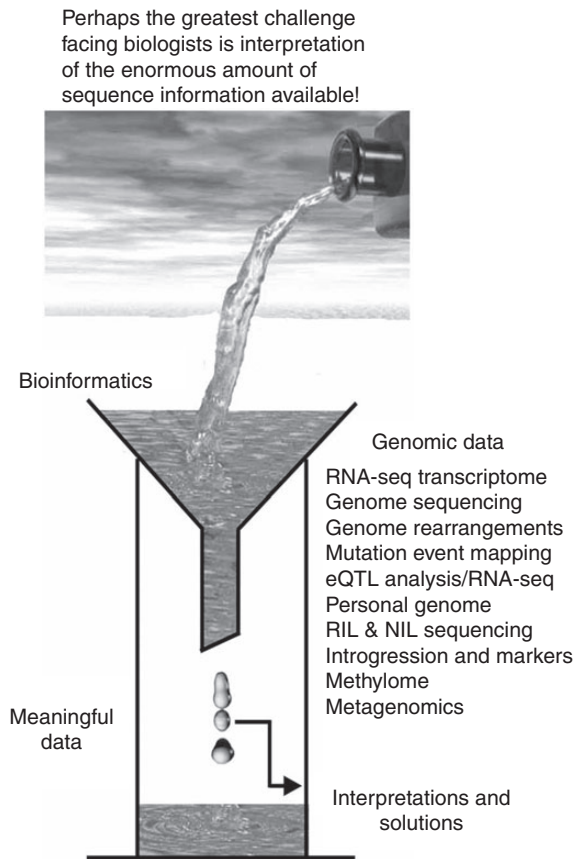


FIG. 1. The genomics bottleneck. Advances in legume genomics have been facilitated by next-generation sequencing technologies. Next-generation sequencing platforms have the power to provide unprecedented insight into biological, genetic, agronomic and academic questions posed by researchers. However, each experiment utilizing next-generation sequencing technology is bound by limitations including data storage, computational power and access to bioinformatic analysis and assistance. With these platforms, the distillation of biologically relevant information (and subsequent interpretation) is the rate-limiting step.

GENOME SEQUENCING

To date, genome sequence information, either partial or complete, exists for 44 higher plants (41 at <http://phytozome.net>, accessed 22 January 2014; and three at Legume Information System, www.comparative-legumes.org, accessed 22 January 2014). Six of the currently sequenced genomes are legumes (Table 1); *Cajanus cajan*, pigeonpea (Varshney *et al.*, 2012); *Cicer arietinum*, chickpea (Varshney *et al.*, 2013); *Glycine max*, soybean (Schmutz *et al.*, 2010); *Lotus japonicus*, trefoil sp. (Sato *et al.*, 2008); *Medicago truncatula*, barrel medic (Young *et al.*, 2011); and *Phaseolus vulgaris*, common bean (www.phytozome.net). The estimated high confidence gene calls in these species range from 28 269 to 48 680 genes. The rapid advances in DNA sequencing can be illustrated in that it took some 8 years to sequence and annotate the *M. truncatula* genome (2003–2010), while the genomes of other species were sequenced in 2 years. Currently, the genomes of alfalfa (*Medicago sativa*), lupin (*Lupinus angustifolius*), peanut (*Arachis hypogaea*) and pea (*Pisum sativum*) are in the sequencing pipeline. It should be noted that a large collection of expressed sequence tags for a number of plant species are deposited at the Dana Farber Gene Index (<http://compbio.dfci.harvard.edu/tgi/plant.html>) and these have been invaluable in annotating legume genomes and making gene calls.

A synopsis of the genome sequence data acquired from the six currently sequenced legumes shows that they have both common and unique features. All six sequenced legumes belong to the Papilionoideae, which appears to have undergone a whole-genome duplication (WGD) event some 60 million years ago (MYA), and all appear to be derived from a common ancestor (Supplementary Data Fig. S1; Schmutz *et al.*, 2010; Young *et al.*, 2011; Varshney *et al.*, 2012, 2013]. Soybean underwent a second, probably allopolyploid, WGD event some 13 MYA (Schlueter *et al.*, 2004; Gill *et al.*, 2009; Cannon and Shoemaker, 2012). The WGD event 60 MYA currently appears to be confined to the Papilionoideae, as WGD signatures have not yet been identified in either the Mimosoid or Caesalpinoid subfamilies. A sizable portion (31–59 %) of Papilionoideae genomes are composed of transposable elements. The expansion of transposable elements may be related to the WGD which occurred some 60 MYA, while further expansion of transposable elements in soybean may have occurred at the second WGD event 13 MYA. Comparative analysis within the Papilionoideae subfamily indicates that there is striking synteny among euchromatic regions, with conserved blocks extending as far as a large portion of chromosome arms (Young *et al.*, 2011; Varshney

TABLE 1. Legume sequenced genomes

Legume	Genome (1N)	No. of genes*	No. of ESTs [†]	Reference
<i>Cajanus cajan</i> (pigeonpea)	11	48 680	25 640	Varshney <i>et al.</i> (2012)
<i>Cicer arietinum</i> (chickpea)	8	28 269	46 064	Varshney <i>et al.</i> (2013)
<i>Glycine max</i> (soybean)	20	46 430	1 530 030	Schmutz <i>et al.</i> (2010)
<i>Lotus japonicus</i> (trefoil)	6	30 799	243 067	Sato <i>et al.</i> (2008)
<i>Medicago truncatula</i> (barrel medic)	8	47 845	286 175	Young <i>et al.</i> (2011)
<i>Phaseolus vulgaris</i> (common bean)	11	31 600	149 769	In progress (www.phytozome.net)

*As of January 2014.

[†]According to the Dana Farber Gene Indices (compbio.dfci.harvard.edu/tgi/plant.html).

et al., 2012). For example, Mudge et al. (2005) evaluated a region on soybean chromosome 18 which correlated with nematode resistance and was co-linear with medicago for 33 of 35 genes. There appears to be a sizeable component of legume family-specific genes as well as species-specific genes. Varshney et al. (2012) identified from 455 to 3068 legume species-specific genes. A substantial number of the putative legume-specific genes are related to the development of N₂-fixing root nodules and to the transduction of biotic and abiotic signalling. Many of these gene families are clustered in the genome and are annotated as nodulins, nodule cysteine-rich (NCR) proteins, leucine-rich repeat (LRR) receptor-like proteins and transporters. Phylogenetic analysis of nodule-specific genes suggests that nodulation arose several times through co-opting of several ancestral genes involved in mycorrhizal colonization (Young et al., 2011).

Some whole-genome sequencing projects were also accompanied by transcriptome analyses via RNA-seq (Young et al., 2011; Varshney et al., 2012, 2013). RNA-seq data allowed for in-depth mining of splice variants, SNPs, simple sequence repeats (SSRs) and small interfering RNA (siRNA) species, the identification of sub-/neo-functionalized genes and transcript expression. Analysis of soybean, which has a high seed oil content, shows an overabundance of genes and transcripts related to lipid biosynthesis compared with other legume sequences (Schmutz et al., 2010). Similarly, in pigeonpea, which thrives in dry arid climates, there is an over-representation of genes related to drought tolerance (Varshney et al., 2012). Interestingly, there are 593 NCRs found in the *Medicago* genome, with >300 that show enhanced expression in root nodules (Fedorova et al., 2002; Young et al., 2011). The NCRs appear to be required for bacteroid development in mature nodules. A comparable NCR family is not apparent in soybean (Schmutz et al., 2010) or common bean (J.A. O'Rourke and C.P. Vance, unpubl. res.). Lack of NCRs in soybean and common bean suggests that other transcripts in nodules of these species must play a role in bacteroid differentiation and viability.

MULTIPLE GENOMES: RE-SEQUENCING

Re-sequencing the genomes of multiple distinct accessions within a single species can provide greater understanding of evolutionary history and domestication events (Huang et al., 2010; Cao et al., 2011; Jiao et al., 2012). In addition, re-sequencing can lay a foundation for genome-wide association studies (GWAS) and forces contributing to heterosis (Atwell et al., 2010; Lai et al., 2010; Lam et al., 2010; Branca et al., 2011). While the genome of cultivated soybean was sequenced by Schmutz et al. (2010), the whole genome of wild soybean *G. soja* Sieb and Zucc was sequenced by Kim et al. (2010). Their final *G. soja* mapped sequence length of 915 Mb covered 97.6% of the soybean genome to 43× coverage. They found that some 80% of the *G. soja* genome was duplicated, reflecting the two WGD events occurring at 60 and 13 MYA (Supplementary Data Fig. S1). As assessed by SNPs and PAVs, the genomes of *G. soja* differed from that of cultivated soybean by 0.3%; less than that between arabidopsis accessions (Ossowski et al., 2008). They estimated that while *G. max* and *G. soja* diverged approx. 287 000 years ago, soybean was domesticated only 6000–9000 years ago. The authors suggest that some 712 genes

representing 32.4 Mb of *G. max* were partially or completely absent in *G. soja*, suggesting that cultivated soybean acquired a number of genes during domestication that are not present in the wild species. This was further confirmed by Joshi et al. (2013) who identified 425 genes unique to *G. max* (completely absent in *G. soja*) including 12 genes involved in seed development and two involved in lipid metabolism.

Further re-sequencing of soybean was conducted by Lam et al. (2010). They sequenced the genomes of 17 wild (*G. soja*) and 14 cultivated (*G. max*) soybean lines. Phylogenetic analysis confirmed that wild and cultivated soybean originated from a common ancestor. The total number of SNPs was much higher in wild soybean as compared with the cultivated lines. A set of 205 614 SNPs was identified which could be useful for QTL mapping and GWAS. Cultivated soybean had a relatively homogeneous genetic background. However, some cultivated lines showed evidence of introgressions from wild soybean. As compared with rice and arabidopsis, both cultivated and wild soybean exhibited higher linkage disequilibrium (LD), decaying at 150 and 75 kb, respectively. Diversity analysis of LD blocks showed that there was a lower SNP ratio in long LD blocks compared with the whole genome. The identification of high LD in soybean along with a high number of SNPs led the authors to conclude that marker-assisted selection (MAS) would be the best approach to breeding for improvement. In addition, since wild and cultivated soybeans are cross-fertile, it is important that wild soybean be protected in order to provide new alleles for cultivated genomes.

Branca et al. (2011) used whole-genome re-sequencing of 26 diverse *M. truncatula* lines to lay a foundation for GWAS and evaluate LD. They found that *M. truncatula* was three times more diverse than soybean (some 3 million SNPs). Four clusters of genes had greater replacement site diversity as compared with other gene families: Toll interleukin repeats (TIRs), LRRs, nucleotide binding apoptosis (NB_ARCs) and NCRs. Similar to arabidopsis, LD decayed rapidly, much more so than in soybean. The authors concluded that whole-genome re-sequencing may be more efficient than a tagged SNP approach for GWAS.

Varshney et al. (2013) re-sequenced whole or partial genomes of 90 chickpea genotypes including 60 improved lines, 25 landraces and five wild genera (*Cicer*) accessions. They identified 4.4 million variants [SNPs and insertions/deletions (Indels)]. The cultivated genotypes displayed limited genetic diversity as compared with the wild landraces. They noted that the breeding for agronomic and phenotypic traits resulted in a loss of genetic diversity through the use of small sets of superior genotypes, which creates genetic bottlenecks. Six genomic regions ranging from 50 to 200 kb carried approx. 122 genes that appear to be candidates for selection in modern breeding efforts. Several genes involved in disease resistance were among those identified as candidates for selection. Extended synteny was detected between chickpea and other sequenced legumes, with the largest number of extended blocks observed for *M. truncatula*.

Genomic re-sequencing of plants with stable mutant phenotypes provides a powerful method for mapping mutational events in order to link a phenotype with a causal mutation. This type of forward genetics has proven a major genetic tool in determining gene function. In arabidopsis, Belfield et al. (2012) determined that fast neutron (FN) mutagenesis induced smaller genomic deletions and more single base pair substitutions than previously believed, highlighting the limitations of

chip-based technologies. In *P. vulgaris*, we (O'Rourke *et al.*, 2013b) applied NGS technologies to identify three types of SV in an FN population: (1) intercultivar SV, i.e. large sequence segments that are missing from all mutants and the wild-type plant, but are present in the reference genome; (2) intracultivar SV, i.e. sequences exhibiting differences among individuals within the mutant population; and (3) SV specific to a single mutant individual, potentially generated by FN irradiation (Fig. 2). Using this methodology, we identified >17 000 SNPs and 2700 Indels within gene-coding regions potentially impacting the mutant phenotypes. Additionally, large-scale deletions (>1 kb) were identified within gene regions for three of the mutant individuals. The utility of this type of analysis has driven the development of various analysis pipelines designed to identify deletions resulting in mutant phenotypes (Huang *et al.*, 2009; Zuryn *et al.*, 2010; Wang *et al.*, 2011; Leshchiner *et al.*, 2012).

Next-generation technologies are also proving important tools in traditional plant breeding programmes. Combining the power of NGS and QTL maps allows researchers to identify causal genes underlying traits of interest (Edwards *et al.*, 2013; Atwood *et al.*, 2014). Additionally, genotype by sequencing (GBS) provides a rapid and low-cost method for genotyping populations, allowing large-scale genomic selection in breeding programmes (Poland and Rife, 2012). The identification of SNPs and Indels coupled with simultaneous genotyping unique to the

breeding lines of interest facilitates high-throughput genotyping of progeny to select for traits of interest and may reduce yield drag by improving progeny selection (Varala *et al.*, 2011). GBS technologies have been successfully applied in soybean and common bean (Hyten *et al.*, 2010; Varala *et al.*, 2011; Sonah *et al.*, 2013).

As the cost of NGS continues to decrease, genome re-sequencing will become even more common. Currently, multiple reference lines of soybean, common bean, lotus and *Medicago* are being sequenced in addition to WGS analysis of mutant individuals and GBS of mapping populations. These data will provide further insight into evolution, genome architecture, Indels and gene copy number. However, mining useful information from multiple genome sequences will require even greater investment in bioinformatics due to the bottleneck in analysing large data sets (Fig. 1).

LEGUME FUNCTIONAL BIOLOGY THROUGH THE TRANSCRIPTOME

Evolution of plant transcriptomes

Transcriptome analyses of developing tissues and organs, along with that of plants exposed to various biotic and abiotic stresses, can provide insight and understanding of plant genes regulating

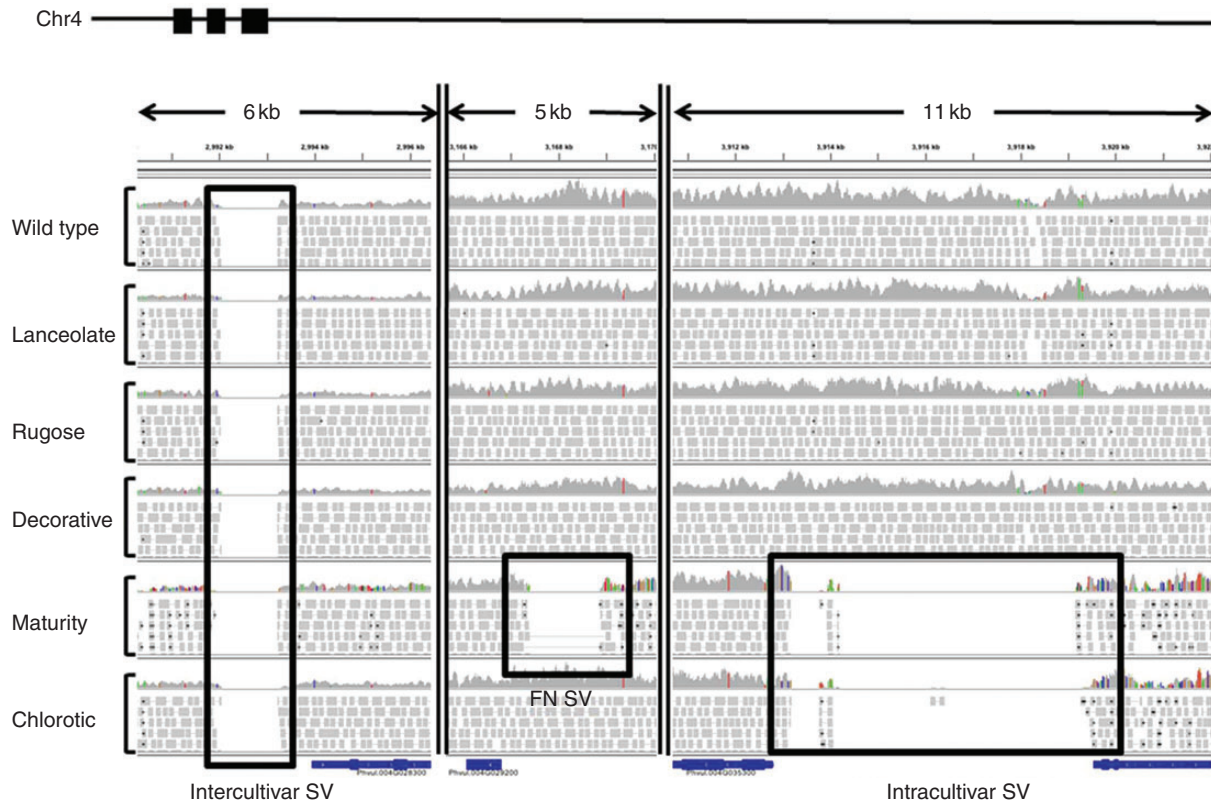


FIG. 2. Structural variations in *Phaseolus vulgaris* fast neutron (FN) mutants. DNA-seq data from six Red Hawk individuals aligned to the common bean reference genome (GI19833) available at www.phytozome.net and visualized using IGV (integrative genomics viewer; Robinson *et al.*, 2011). These regions of structural variation (SV) were all identified on chromosome 4 from regions highlighted by filled black boxes on the chromosome at the top of the figure. For each individual, a histogram plot illustrates the read depth, while individual reads are aligned below. Each type of SV is highlighted by a black box. Intercultivar SV: regions present in the reference genome but missing in all Red Hawk individuals, representing differences between common bean cultivars. FN SV: sequences missing in a single individual, but present in all other lines, probably resulting from FN mutagenesis. This type of SV is most probably responsible for the mutant phenotype. Intracultivar SV: regions of genomic heterogeneity within the Red Hawk cultivar. Deletions identified in two or more Red Hawk individuals illustrating the residual heterogeneity in the FN mutant population.

many processes (Z. Wang *et al.*, 2010a; Mochida and Shinozaki, 2011). Evaluation of plant transcriptomes has undergone a phenomenal expansion in the past 10 years. Several reviews have described the methods and approaches from sequencing individual transcripts [i.e. cDNAs via expressed sequenced tags (ESTs) and serial analyses of gene expression (SAGE)] to massive parallel signature sequencing (MPSS). Upon the sequencing of thousands of ESTs and genome sequencing of many plant species, chip-based technologies involving microarrays emerged as the dominant platform and approach to transcript expression. However, the aforementioned approaches to transcript studies have major limitations including: (1) sample size and normalization; (2) speed; (3) sensitivity; (4) cost; and (5) limited dynamic range (Wang *et al.*, 2009; L. Wang *et al.*, 2010; Ozsolak and Milos, 2011; Schneeberger and Weigel, 2011; Garg and Jain, 2013). Unprecedented advances in high-throughput next-generation RNA-seq have resulted in a paradigm shift in transcriptome studies (Lister *et al.*, 2008; Mochida and Shinozaki, 2011; Garg and Jain, 2013). In addition, the availability of powerful assembly software alleviates the requirement for a sequenced genome (Garg *et al.*, 2011; Kalavacharia *et al.*, 2011; Yang *et al.*, 2011; Zhang *et al.*, 2012; O'Rourke *et al.*, 2013b). RNA-seq has been used in several studies involving plants, including but not limited to: maize leaf development (Li *et al.*, 2010); rice transcriptome complexity (Zhang *et al.*, 2010); rubber plant, *Hevea brasiliensis* (Xia *et al.*, 2011); panicum development (Meyer *et al.*, 2012); tomato tissue and metabolism (Matas *et al.*, 2011); and grape (*Vitis vinifera*) berry development (Zenoni *et al.*, 2010).

Although RNA-seq data have been collected for several legume species, a large focus has been on developing genetic markers in chickpea (Garg *et al.*, 2011); lentil, *Lens culinaris* (Kaur *et al.*, 2011); peanut, *A. hypogaea* (Zhang *et al.*, 2012); alfalfa, *M. sativa* (Han *et al.*, 2011); and common bean, *P. vulgaris* (Kalavacharia *et al.*, 2011). Studies using RNA-seq to understand development, acclimation to stress, composition and N fixation have been limited to soybean (Bolon *et al.*, 2010; Libault *et al.*, 2010; Severin *et al.*, 2010a, b; Atwood *et al.*, 2014); alfalfa, *M. sativa* (Yang *et al.*, 2011); lupin, *L. albus* (O'Rourke *et al.*, 2013b); and *M. truncatula* (Boscari *et al.*, 2013). It is also noteworthy that RNA-seq has been applied to studies of legume microRNAs (miRNAs) in common bean (Paez *et al.*, 2012), *M. truncatula* (Lelandais-Briere *et al.*, 2009; Simon *et al.*, 2009) and soybean (Song *et al.*, 2011; Turner *et al.*, 2012). In the sections to follow, we will provide an overview of RNA-seq studies our laboratory has conducted to understand legume biology.

The soybean transcriptome

Soybean is a highly valued crop because it accumulates large amounts of oil, protein and seeds, and because, in symbiosis with soil rhizobia bacteria, it forms N₂-fixing root nodules. Soybean seed contains approx. 40% protein and 20% oil (Ohlrogge and Kuo, 1984; Bolon *et al.*, 2010). Soybean seed protein is a primary food for both livestock and humans. Soybean oil is the most abundant cooking oil on earth (Wilcox, 2004). Symbiotic N₂ fixation by soybean greatly reduces the need for added N fertilizer. To understand the genes involved in seed development and root nodule N₂ fixation, we evaluated the transcriptomes of

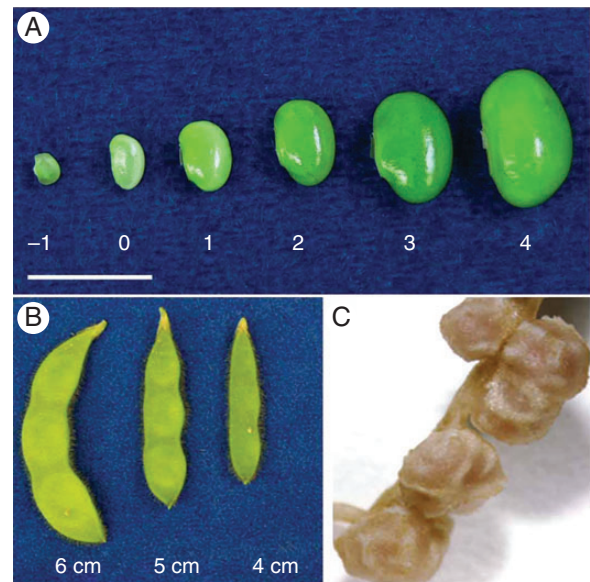


FIG. 3. Soybean tissues with RNA-seq transcriptome data. Total RNA was isolated from leaves, flowers, pods, roots, nodules and seven stages of seed development of soybean near-isogenic lines LoPro and HiPro. (A) Seed stages -1 to 4. (B) Pod sizes for corresponding seed stages from stage -1 to stage 2. (C) Nodule tissues attached to soybean roots.

seeds and root nodules. To complement these studies, we also analysed the transcriptomes of other soybean tissues (Fig. 3). We isolated total RNA from leaves, flowers, pods, roots, nodules and seven stages of seed development of two near-isogenic lines (NILs). One line segregated for high protein (HiPro; LD04-15154) and the other for low protein (LoPro; LD04-15146) (Seboldt *et al.*, 2000; Nichols *et al.*, 2006; Bolon *et al.*, 2010). These data were used to develop a soybean gene expression atlas (Severin *et al.*, 2010b) and to analyse SNPs between the HiPro and LoPro NILs (Bolon *et al.*, 2010; Bolon and Vance, 2012; Bolon *et al.*, 2013). The close genetic identity of the two NILs was confirmed by identifying only 387 SNPs between the two lines. The 1000 most highly expressed transcripts in each tissue (Combined HiPro and LoPro) are shown in Supplementary Data File S1. It should be noted that Libault *et al.* (2010) have also developed a soybean gene expression atlas based upon RNA-seq data derived from 14 tissues. They focused primarily on roots and nodules from a single genotype of *G. max*, Williams 82.

Annotations of the soybean genome (Schmutz *et al.*, 2010) predicted 46 430 high-confidence gene models and 19 780 lower confidence gene models. We found 90.4% (41 975) high-confidence gene models and 7176 lower confidence gene models to be transcriptionally active. Hierarchical clustering analyses of transcripts between tissues and development suggested three groupings of tissues; underground (roots and nodules), seed related and aerial tissues (flower, leaves and pods) (Severin *et al.*, 2010b). Tissue-specific expression was evaluated by combining counts from each genotype and then selecting transcripts with an RPKM (reads per kilobase per million) ≥ 1 , according to a modified RPKM formula (Severin *et al.*, 2010b). Tissue-specific transcripts in the combined NILs were identified including: seed = 81; leaf = 95; flower = 716; pods and shell

pods = 138; root = 727; and nodule = 418 (Supplementary Data File S2).

Due to the capacity of soybean to fix N₂ symbiotically, we determined whether genes other than the typical nodulin gene families and leghaemoglobins were highly expressed in nodule tissues. A novel PLAC8 (placenta-specific) transcript was highly expressed. This membrane protein possesses a cysteine-rich region and has been implicated in regulating cell number (Libault and Stacey, 2010). A highly expressed cytochrome P450 was also noted. Annotation for this gene suggests involvement in either flavonoid or abscisic acid metabolism. An embryo-specific protein with a lipoxygenase-related annotation and at least two remorin-like transcripts were abundantly expressed in nodules. Remorins are membrane proteins associated with lipid rafts. Toth *et al.* (2012) has recently shown that *L. japonicus* remorin 1 is localized to infected nodule cells, and overexpression of this gene increases nodule numbers. The diverse array of putative functions for soybean nodule-specific transcripts make this data set a valuable addition to root nodule biology.

Because of our interest in transporters, we queried the combined NIL RNA-seq data for transporter function. We found a total of 1733 transporters showing expression in one or more tissues (Supplementary Data File S3). Fifteen of the 20 most highly accumulated transporter gene transcripts in the NIL RNA-seq seed data were found to be aquaporin transporters. The remaining highly expressed transporter genes were annotated as magnesium, copper, nucleotide or sugar-related transporter genes in the seed (Bolon and Vance, 2012). Genes encoding sulfate, ABC, oligopeptide, triose phosphate and sugar transporters were all highly expressed in nodules (Supplementary Data Files S1, S3). The sulfate transporter 1 (SST1) gene has been shown to be critical for effective N₂ fixation (Krusell *et al.*, 2005). Sugar import into nodules has been shown to be crucial to maintain effective N₂ fixation (Vance and Heichel, 1981). Within the NIL RNA-seq data set, some 5–7 sugar transporters appear to be fairly nodule specific (Supplementary Data File S1). Root tissues, in general, displayed the highest expression of transporter genes, which is consistent with the role of roots in acquiring nutrients necessary for growth.

Seed formation proceeds through three stages: pre-storage, transition and storage maturation (Weber *et al.*, 2005). We performed transcriptome analyses on developing seed stages: –1 seed (<25 mg); +1 seed (approx. 50 mg); +2 seed (approx. 100 mg); +3 seed (approx. 200 mg); and +4 seed (approx. 300 mg) (Fig. 3). These stages correspond to those described by Weber *et al.* (2005), Bolon *et al.* (2010) and Severin *et al.* (2010b). A comparison of –1 seed with the +4 seed gene expression, visualized using MapMan metabolic pathways (Thimm *et al.*, 2004), shows that the greatest changes in transcript abundance occurred as increases in transcript abundance of lipid, primary amino acid metabolism and 1-carbon (C1) metabolism genes (Fig. 4). The greatest decreases in transcript abundance were in cell wall, nucleotide, fermentation, ascorbate and secondary metabolism. These changes in transcript abundance were not unexpected based on metabolic events occurring at each stage of seed development.

Because seed development and filling involves partitioning of N from leaves into pods and seeds, we thought it important to evaluate the expression of N assimilation enzyme gene

transcripts from pod formation through seed filling. Transcripts related to 11 gene families were surveyed (Fig. 5; Supplementary Data File S4). Figure 5A shows the gene expression patterns for NADH- and Fd-glutamate synthase (GOGAT). In early pods and seeds, Fd-GOGAT transcripts (Glyma03g28410 and Glyma19g31120) predominate, while later in seed development NADH-GOGAT (Glyma14g32500 and Glyma19g16450) is the dominant form. Expression of Fd-GOGAT in pods and early seed is consistent with the observation that green photosynthesizing tissues utilize Fd-GOGAT for N assimilation while in non-photosynthetic tissues, NADH-GOGAT is more important to N assimilation. Enzymes involved in allantoin metabolism (Fig. 5B; Supplementary Data File S4) appear most highly expressed during early pod set and seed formation, as evidenced by the expression of Glyma13g31430, allantoinase. Expression of genes involved in ammonium transport and ureide degradation (including allantoinase, allantoate amidohydrolase and urease) in early pod set and seed filling reflects the transport and subsequent metabolism of ureides in pods and seeds. Our data support the interpretation that pods and seeds participate in the metabolism of both ammonia and ureides, providing N for protein deposition in seeds.

As mentioned previously, soybean seed value is determined by the oil and protein composition. Using tissues from ten different stages of reproductive development, transcripts encoding enzymes related to oil biosynthesis were evaluated. Our analysis of this RNA-seq data set provides a comprehensive overview of oil metabolism throughout soybean seed development and a foundation for modelling gene networks involved in seed oil deposition. Transcripts encoding enzymes of 11 gene families involved in oil biosynthesis were identified, including 15 members of the fatty acid desaturase gene family (Supplementary Data File S5). Three patterns of lipid gene expression were noted based on the stage of seed development: (1) high expression in early pod and seed; (2) uniform expression throughout development; and (3) high expression during seed filling. Early pod seed expression was exemplified by glycerol-3-phosphate transferase and phospholipid diacylglycerol acyltransferase gene expression. Examples of genes with transcripts expressed throughout seed development include triacylglycerol lipase and diacylglycerol acyltransferase.

Unexpectedly, our data show that FAD2–2B and FAD2–2C are expressed early in seed formation. In contrast, and as expected, transcripts for FAD2–1B and FAD2–1A were overabundant during later seed development as oil bodies are deposited. In a previous study of seed filling, Severin *et al.* (2010b) noted 143 genes annotated by gene ontology (GO) analysis as nutrient reservoir. Included among these 143 genes are oleosins, sucrose-binding proteins, glycinins and lipoxygenases (Supplementary Data Files S1, S2), consistent with seed transcriptome analysis in the soybean NIL pair (Bolon *et al.*, 2010; Bolon and Vance, 2012). Proteomic analyses on the developing soybean seed (Hajduch *et al.*, 2005) also provide support for the presence of gene transcripts associated with metabolism, protein destination and storage.

During seed filling of dicot plants, particularly soybeans, the embryo enlarges as proteins, oil and carbohydrates are stored in the cotyledons (Le *et al.*, 2007; Agrawal *et al.*, 2008; Verdier and Thompson, 2008; Verdier *et al.*, 2008; Hajduch *et al.*, 2010). Several transcription factors act as major regulators

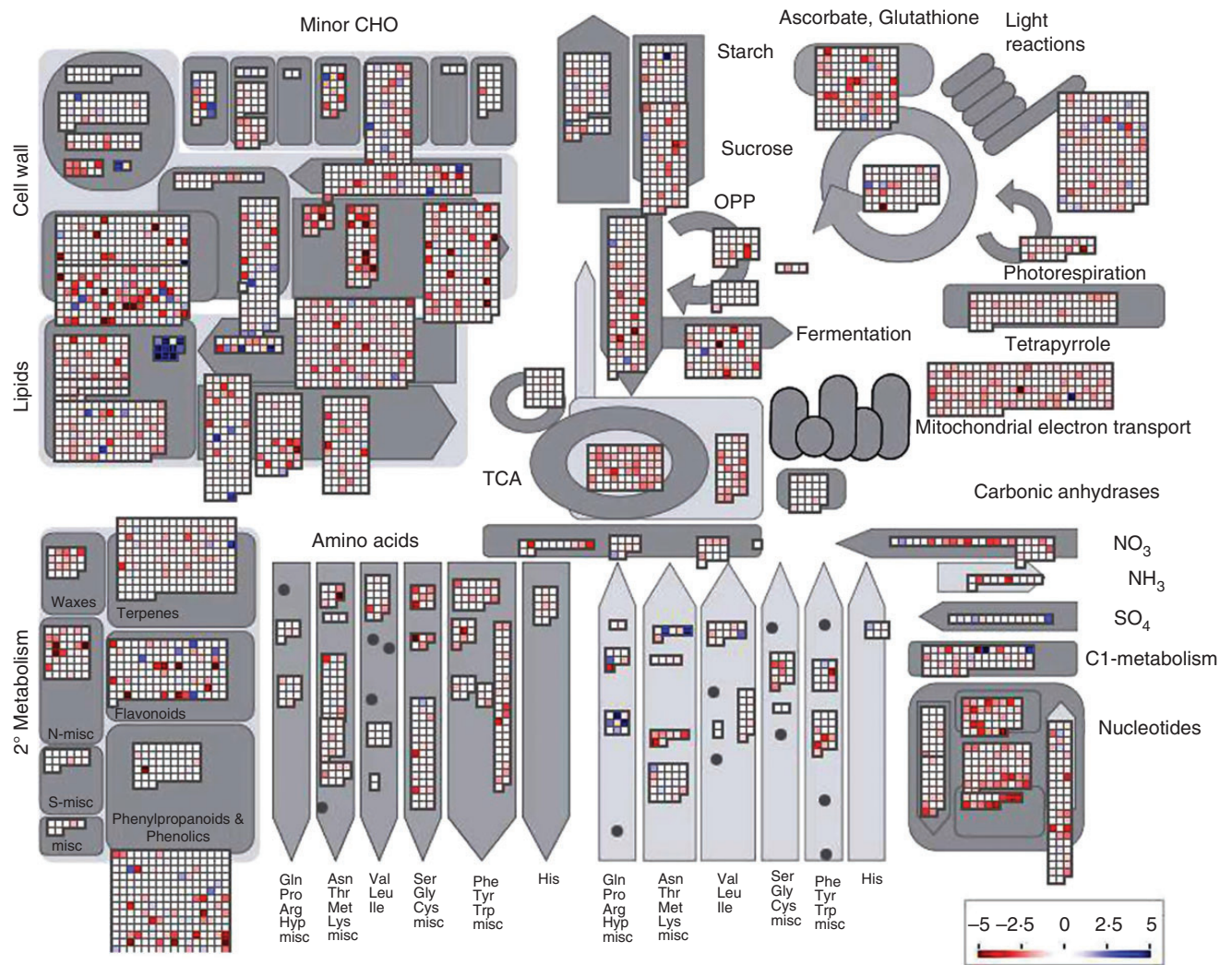


FIG. 4. Overview of soybean seed metabolism throughout seed development. A visualization of soybean metabolism gene transcript abundance changes is shown using MapMan. This map shows a comparison of -1 seed with $+4$ seed gene expression in general metabolic pathways. The greatest gene transcript abundance increases (blue) are seen in lipid, primary amino acid metabolism and C1-metabolism genes. The largest decreases in transcript abundance (red) are found in cell wall, nucleotide, fermentation, ascorbate and secondary metabolism genes.

of these processes (Verdier *et al.*, 2008; Le *et al.*, 2010), while an additional 900–1300 transcription factors have been associated with seed development (Benedito *et al.*, 2008; Le *et al.*, 2010). Our data set allowed us to profile transcription factors throughout seed formation and filling. Some 54 transcripts related to 12 transcription factor families were evaluated (Z. Wang *et al.*, 2010b; Supplementary Data File S6). Transcripts encoding AP2, VAL2, GL2, PKL, FIE and DOF4, were most highly expressed in early pod and embryo development. PKL is a crucial activator of embryo development while FIE suppresses premature development of the endosperm (Mosquna *et al.*, 2004; Li *et al.*, 2005). The transition from embryo to seedling development is regulated by VAL1 and 2 (Wang and Perry, 2013). Seed filling development is regulated by ABI3, FUS3 and WR11. These transcription factors appear to affect oil, carbohydrate and protein deposition during seed fill. The interaction of transcription factors throughout embryo and seed development is quite complex, with multiple transcription factors interacting at each stage.

The soybean NIL RNA-seq data set from 14 soybean tissues has significantly enhanced the options for understanding the soybean whole transcriptome. With this data set, along with those of Libault *et al.* (2010), legume researchers now have a stronger foundation for analysing gene networks in soybean. The data serve as a valuable resource for the legume community.

White lupin transcriptome: acclimation to phosphorus deficiency

Phosphorus (P_i) is a critical nutrient required for plant growth and development (Bielecki, 1973). While the P_i content is high in many soils, availability is often limited due to slow diffusion and high fixation rates. Moreover, in developing countries, the aged and weathered soils render P_i essentially unavailable (Bielecki, 1973; Vance *et al.*, 2003; Lambers *et al.*, 2008). Additionally, P_i fertilizer production is not sustainable, with inexpensive world supplies becoming limited by 2050–2080 (Cordell *et al.*, 2009; Gilbert, 2009).

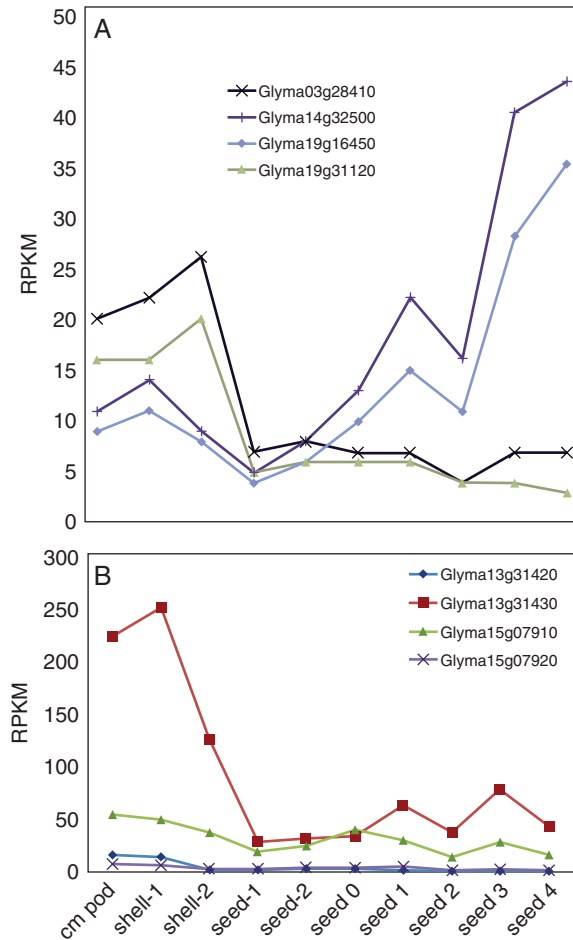


FIG. 5. Nitrogen assimilation enzyme gene transcripts from pod formation through to seed filling. (A) Gene transcript abundance patterns for NADH- and Fd-glutamate synthase (GOGAT). In early pods and seeds, Fd-GOGAT transcripts (Glyma03g28410 and Glyma19g31120) are more abundant, while later in seed development NADH-GOGAT (Glyma14g32500 and Glyma19g16450) transcripts are more highly expressed. (B) Allantoin metabolism gene transcript abundance (Supplementary Data File S5) is highest during early pod set and seed formation (Glyma13g31430, allantoinase). RPKM, reads per kilobase per million.

Thus, efforts to improve P_i acquisition and utilization by plants are imperative.

White lupin (*L. albus*) forms cluster roots (Fig. 6) in response to P_i deficiency (Neumann and Romheld, 1999; Vance et al., 2003; Cheng et al., 2011; O'Rourke et al., 2013b). In P_i -deficient plants, upwards of 70 % of the root mass may occur as cluster roots. Cluster root formation is accompanied by striking biochemical changes including the release of enzymes and proteins, exudation of protons and organic acids, modified metabolism and altered hormone balance. The biochemical and molecular changes that occur during cluster root formation have made white lupin a model for studies of plant acclimation to P_i deficiency stress.

In an effort to better understand plant acclimation to P_i deficiency and the genes involved, we conducted a whole-genome transcript expression study in white lupin plants grown under P_i -sufficient and P_i -deficient conditions (O'Rourke et al., 2013b). The P_i -deficient root samples were composed primarily of

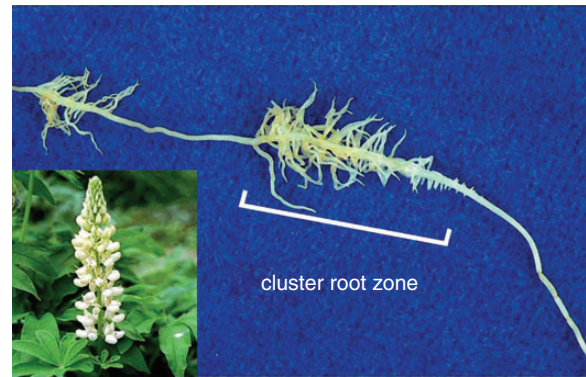


FIG. 6. White lupin cluster root phenotype. White lupin (*Lupinus albus*; inset image) cluster root formed when grown in P_i -deficient hydroponics for 14 d. A bracket spanning the region of closely spaced tertiary lateral rootlets delineates the cluster root region.

cluster roots. The transcripts were assembled *de novo* and used to build the first white lupin gene index (LAGI 1.0), which is available on the white lupin species information page at <http://comparative-legumes.org>. We project 50 734 transcriptionally active sequences representing 7.8 % of the genome. Considering all fully sequenced legumes, approx. 6 % of the genome is predicted to encode proteins (*M. truncatula*, 8.3 %; *L. japonicus*, 4.4 %; pigeonpea, 5.6 %; soybean, 5 %). The 1000 most highly expressed transcripts are provided in Supplementary Data File S7.

We detected 2128 transcripts that are differentially expressed due to P_i deficiency (Fig. 7; Supplementary Data File S8). In leaves, 1342 transcripts were differentially expressed due to P_i deficiency while 903 transcripts were differentially expressed in roots (some transcripts were differentially expressed in both leaves and root). Several functional classes of transcripts having increased expression under P_i -deficient conditions were over-represented in the transcripts of P_i -deficient cluster roots including glycolysis, C1-metabolism, glyoxylate related, cell wall, lipid metabolism, metal handling and transport (Fig. 8). Similar classes of transcripts were over-represented in P_i -deficient leaves. In comparison, classes of transcripts over-represented in P_i -sufficient leaf samples included photosynthesis, ATP synthesis and stress. Previous studies from our laboratory (Johnson et al., 1994, 1996) as well as others (Neumann and Romheld, 1999; Plaxon and Tran, 2011) have shown that P_i deficiency enhances glycolysis and organic acid metabolism. The lupin transcriptome reveals that other carbon pathways are also altered (Supplementary Figs S2, S3). Specifically, C1-metabolism, especially formate, glyoxylate and methionine metabolism, was significantly upregulated in P_i -deficient cluster roots. Acetyl-CoA metabolism was also upregulated in P_i -deficient cluster roots. Each of the above-mentioned new carbon metabolism pathways activated due to P_i deficiency contributes to carbon cycling through organic acids which can be exuded into the rhizosphere to increase the solubility of bound P_i (Supplementary Figure S3).

The increase in C1-metabolism in P_i -deficient roots also appears to contribute to the Yang cycle and ethylene biosynthesis (O'Rourke et al., 2013b). P_i deficiency is known to increase ethylene synthesis in roots (Gilbert et al., 2000; Ma et al., 2003; Zhang et al., 2003; Nagarajan and Smith, 2012; Niu et al., 2013). Ethylene interacts with auxin to alter root hair

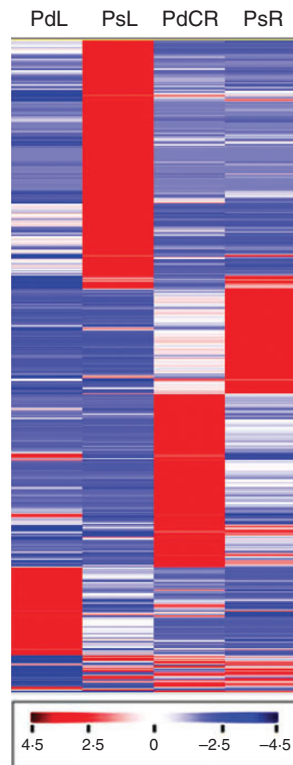


FIG. 7. Heatmap of genes differentially expressed due to P_i deficiency. A heatmap of 2128 transcripts in the LAGI 1.0 assembly differentially expressed due to P_i deficiency. To be considered differentially expressed, the transcript must have an RPKM ≥ 3 in at least one tissue, ≥ 2 -fold change between tissues and $P \leq 0.05$. A total of 1342 transcripts were differentially expressed between leaf samples and 903 between root samples due to P_i deficiency (117 transcripts were differentially expressed in both leaves and roots). Expression profiles are presented as Z-scores. Red indicates high expression, white indicates intermediate expression and blue indicates low expression. PdL, phosphate-deficient leaves; PsL, phosphate-sufficient leaves; PdCR, phosphate-deficient cluster roots; PsR, phosphate-sufficient normal roots.

and lateral root development, therefore impacting white lupin root architecture when subjected to P_i -deficient conditions.

The noticed effect of P_i deficiency on ethylene-related transcripts in cluster roots stimulated us to assess transcripts involved in the metabolism of additional plant hormones, since interactions between hormone synthesis and degradation are known to regulate root architecture and development. Plant hormones are also implicated in nutrient sensing and signalling (Torrey, 1976; Lopez-Bucio *et al.*, 2002; Rubio *et al.*, 2009; Chiou and Lin, 2011; Niu *et al.*, 2013). Transcripts involved in the biosynthesis and degradation of auxin, gibberellin (GA) and cytokinin all exhibited differential expression in P_i -deficient cluster roots (Supplementary Data File S9).

Auxin is a well known effector of lateral root development and density (Torrey, 1976; Casimiro *et al.*, 2001). Previous findings from our lab (Gilbert *et al.*, 2000; Cheng *et al.*, 2011) showed an important role for auxin in signalling cluster root development under P_i deficiency. The addition of exogenous auxin stimulated cluster root development, even under P_i -sufficient conditions. Furthermore, the inhibition of auxin transport blocked cluster root formation. The RNA-seq data further support a role for auxin homeostasis in cluster root formation. We found transcripts

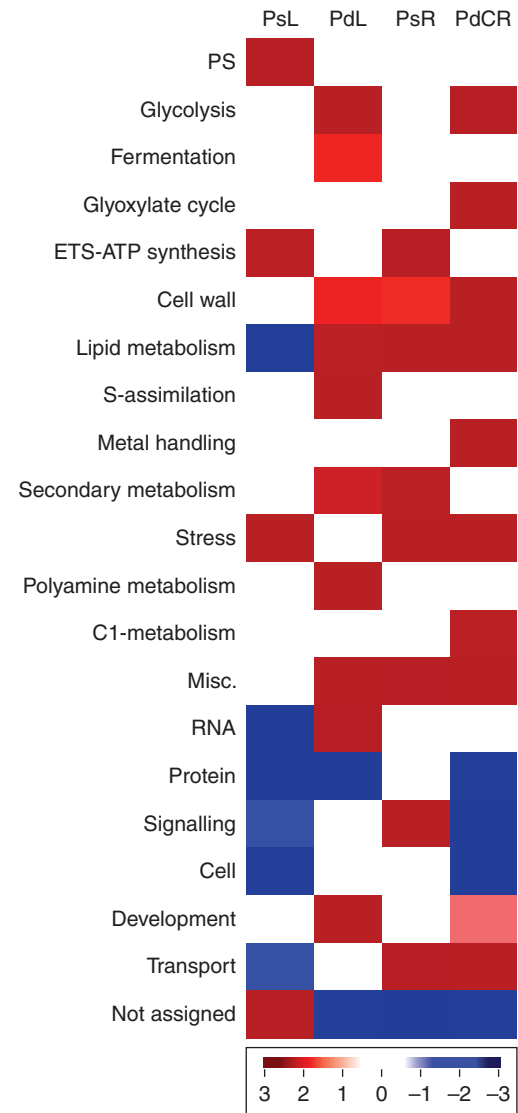


FIG. 8. Gene categories with altered expression due to P_i deficiency in white lupin. The identification of over-represented MapMan categories (Thim *et al.*, 2004) among genes differentially expressed due to P_i deficiency relative to the LAGI 1.0 transcriptome (Fisher test and Benjamin–Hochberg correction, Z-value cut-off = 1). MapMan categories over- (red) or under- (blue) represented in P_i -sufficient leaves (PsL), P_i -deficient leaves (PdL), P_i -sufficient roots (PsR) and P_i -deficient cluster roots (PdCR).

related to auxin homeostasis to be more highly expressed in P_i deficiency-induced cluster roots than in other treatments (Supplementary Data File S9). P_i deficiency results in enhanced sensitivity to auxin and increased auxin concentration in P_i -deficient roots (Lopez-Bucio *et al.*, 2002; Al-Ghazi *et al.*, 2003; Cheng *et al.*, 2011).

Cytokinin applications have been reported to inhibit root growth and suppress lateral root formation (Li *et al.*, 2006; Werner *et al.*, 2010; Cui *et al.*, 2011). Moreover, P_i deficiency is known to reduce cytokinin activity in roots (Lo *et al.*, 2008; Werner *et al.*, 2010). Additional evidence for a role for cytokinin in lateral root formation has been demonstrated by the overexpression of cytokinin oxidase which results in reduced cytokinin levels and increased lateral root formation (Lo *et al.*, 2008;

Werner *et al.*, 2010). Our RNA-seq data revealed that P_i deficiency increases the expression of transcripts encoding cytokinin oxidase, a gene regulating cytokinin degradation. We treated P_i-deficient plants with a synthetic cytokinin, benzyladenine (BA), which resulted in an inhibition of cluster root formation in P_i-deficient plants. Thus, the increased expression of cytokinin oxidase in P_i-deficient cluster roots may be related to maintaining cytokinin homeostasis at levels that promote cluster root development during P_i deficiency.

Similarly, we noted that P_i deficiency appeared to stimulate the expression of transcripts involved in GA biosynthesis. To test the effect of GA on cluster root formation we added either GA₃ or paclobutrazol, an inhibitor of GA biosynthesis, to white lupin plants. Altering GA availability produced striking phenotypes. Additional GA₃ reduced the density of lateral rootlets within the cluster root zone of P_i-deficient plants, while the application of paclobutrazol greatly stimulated cluster root formation in both P_i-sufficient and P_i-deficient plants (O'Rourke *et al.*, 2013b). Similarly, silencing GA oxidase in *Populus* increased root GA concentration and decreased lateral root numbers, while in *Arabidopsis* exogenous GA application had little effect on lateral root formation (Jiang *et al.*, 2007). The lupin transcriptome analysis also revealed that the GA receptor, GID16, and phytochrome-interacting factor (PIF1) transcripts, which modulate GA signalling through DELLA proteins, had increased transcript abundance under P_i-deficient conditions. It appears that GA plays an important role in cluster root formation as evidenced by GA transcript expression and chemical modulation of GA availability. Our data suggest that P_i deficiency-induced cluster root formation is modulated by either increased GA production or increased sensitivity of roots to GA.

The expression of transporter genes is required for acclimation to nutrient deficiencies. Of the 2128 transcripts differentially expressed in white lupin in response to P_i deficiency, 110 were annotated as transporters (Supplementary Data File S10). In P_i-deficient cluster roots, 11 transporter families representing 42 transcripts had enhanced expression when compared with P_i-sufficient roots. The most abundant transporter families were found to be phosphate, membrane H⁺, ATPases, divalent cation and multidrug and toxin efflux (MATE) transporters.

In comparison, eight transporter families, represented by 22 transcripts, exhibited decreased expression in P_i-deficient cluster roots, including lipid, sugar and sulfate transporters. Few transporter families in leaves were impacted by low P_i. Increased transporter gene expression during nutrient-deficient conditions has been noted for *Arabidopsis* (Misson *et al.*, 2005), common bean (Hernandez *et al.*, 2007), tomato (Liu *et al.*, 1998) and rice (Li *et al.*, 2010). Wang *et al.* (2002) reported an interaction among P_i, K and Fe transporters in tomatoes exposed to various nutrient stresses, suggesting that multiple transporters are coordinately regulated irrespective of the particular nutrient deficiency. Our RNA-seq data support such an interpretation.

Previous studies have shown miRNAs (miR399 and miR395) directly regulate sulfate and phosphate deficiency responses (Sunkar and Zhu, 2004; Sunkar, 2010). Additional studies have investigated the post-transcriptional regulation by miRNAs in response to nutritional, biotic and abiotic stresses (Jones-Rhoades and Bartel, 2004; Sunkar and Zhu, 2004; Fujii *et al.*, 2005; Bari *et al.*, 2006; Chiou *et al.*, 2006; Hsieh *et al.*, 2009; Zhu *et al.*, 2009; Sunkar, 2010; Valdes-Lopez *et al.*, 2010). However, sequencing of miRNAs requires special library construction and was thus beyond the scope of our study.

A major goal for crop production is to develop lines with improved acclimation to P_i deficiency. Identification of genes shared in common between species that respond to P_i deficiency may be useful for MAS. We identified 12 genes commonly expressed due to P_i deficiency in three species; white lupin (Uhde-Stone *et al.*, 2003; O'Rourke *et al.*, 2013b), potato (Hammond *et al.*, 2011) and *Arabidopsis* (Misson *et al.*, 2005; Morcuende *et al.*, 2007; Thibaud *et al.*, 2010) (Table 2). These genes, among others reported by Hammond *et al.* (2011), may serve as molecular markers for crop improvement of P_i stress acclimation. In recent years, genes regulating P_i acclimation QTLs have been identified. In *Arabidopsis*, the LOW PHOSPHATE ROOTS (LPR) QTL was identified as a multicopper oxidase (Svistoonoff *et al.*, 2007). Similarly, the PUP1 QTL in rice was shown to be a protein kinase and designated as phosphate starvation tolerance 1 (PSTOL1). Shi *et al.* (2013) identified monogalactosyl diacyl glycerol synthase (MGDG) synthase and a glucose-6-phosphate transporter (G-6-PT) as controlling P_i

TABLE 2. Candidate genes for phosphorus acclimation marker-assisted selection

LAGI 1.0 ID	PsR	PdCR	PsL	PdL	Annotation
LAGI01_23220	6.3	16.6	1.3	7.5	MGDG synthase
LAGI01_41314	5.0	14.0	3.6	8.0	SPX domain
LAGI01_24353	2.6	13.6	0	0	Phosphate transporter
LAGI01_3168	1.3	9.6	0	4.5	Phospholipase
LAGI01_53773	2.0	7.0	0	5.0	MGDG synthase
LAGI01_47701	0	5.6	1.6	7.0	SPX domain
LAGI01_62170	0	4.3	0	1.0	SPX domain
LAGI01_31197	0	10.0	0	19.5	Pyrophosphatase
LAGI01_79787	2.0	11.6	0	14.0	Purple acid phosphatase
LAGI01_35670	3.3	14.3	5.6	16.5	Glycosyltransferase
LAGI01_57598	6.0	60.0	2.6	21.5	GPX-PDE8
LAGI01_81721/74210/75025/63491	4.3	22.3	27.4	80.0	Ferritin

Expression patterns (as RPKM) for LAGI 1.0 transcripts identified as potential candidates for phosphorus deficiency marker-assisted selection.

PsR, phosphate-sufficient normal roots; PdCR, phosphate-deficient cluster roots; PsL, phosphate-sufficient leaves; PdL, phosphate-deficient leaves.

The multiple transcripts (LAGI01_81721, LAGI01_74210, LAGI01_75025 and LAGI01_63491) all share the same expression patterns and all annotate as ferritin, suggesting that all the transcripts encode the same gene.

acclimation in *Brassica napus*. We found LPR1, MGDG synthase and G-6-PT all highly upregulated in P_i-deficient cluster roots and designated them as important candidates for MAS.

Overall, our white lupin transcriptome analyses have provided novel insight and detail into how plants acclimate to P_i deficiency. We have shown previously unknown genes and processes to be major effectors for P_i deficiency tolerance. The lupin RNA-seq data also provided a foundation for gene calling and identification of splice variants upon sequencing of the *L. angustifolius* genome.

CONCLUSIONS

Studies using high-throughput DNA and RNA sequencing have resulted in profound paradigm shifts in plant genetics and biochemistry. Our fundamental understanding of genome architecture (Jiao *et al.*, 2012), species evolution (Lam *et al.*, 2010), adaptation (O'Rourke *et al.*, 2013b), transcript variants (Libault *et al.*, 2010), gene networks (Tomato Genome Consortium, 2012) and biochemical pathways (Li *et al.*, 2012; O'Rourke *et al.*, 2013b) has been expanded. Divergent plant lines such as ecotype collections (Branca *et al.*, 2011; Varshney *et al.*, 2013), mutant populations (Bolon *et al.*, 2011; Fu *et al.*, 2013; O'Rourke *et al.*, 2013a) and nested association lines (Fu *et al.*, 2013) can now be readily sequenced at both RNA and DNA levels. Such studies allow the identification of variants for almost any trait at the nucleotide level (Severin *et al.*, 2010a; Peiffer *et al.*, 2012). Access to genome-wide transcripts facilitates the identification of transcript variants (Libault *et al.*, 2010; Severin *et al.*, 2010b), miRNAs and their target RNAs (Sunkar, 2010), cell- and tissue-specific genes (Libault *et al.*, 2010; Severin *et al.*, 2010b) and epigenetic regulation (Weigel and Colot, 2012). The power of both forward and reverse genetics can be more fully realized through rapid, low-cost sequencing. The broad umbrella of genomics extended by high-throughput sequencing now impacts the disciplines of agronomy, biochemistry, physiology and microbiology. However, to translate the vast array to improved plants still lies through plant breeding and genetics (Supplementary Fig. S4). Multidisciplinary collaboration among plant biologists will be essential to feeding the worlds 7.5 billion inhabitants.

SUPPLEMENTARY DATA

Supplementary data are available online at www.aob.oxfordjournals.org and consist of the following. File S1: soybean NILs of the 1000 most highly expressed genes in each tissue. File S2: tissue-specific genes identified from the soybean NIL RNA-seq data. File S3: the RNA-seq expression profiles of soybean NILs and annotations of all soybean genes identified as transporters. File S4: the RNA-seq expression profiles of soybean NILs for genes involved in nitrogen assimilation. File S5: the RNA-seq expression profiles of soybean NIL genes involved in oil biosynthesis. File S6: RNA-seq expression profiles and annotation of transcription factors expressed in developing seed tissues. File S7: the 1000 most highly expressed genes in white lupin. File S8: transcripts differentially expressed due to phosphate deficiency in white lupin. File S9: gene expression patterns of hormone-related genes due to P_i deficiency in white lupin. File S10: transporters differentially expressed due to P_i deficiency in white lupin. Figure S1: legume phylogenetic tree.

Figure S2: a metabolism overview of transcripts differentially expressed due to P_i deficiency in leaves using MapMan. Figure S3: a metabolism overview of transcripts differentially expressed due to P_i deficiency in roots using MapMan. Figure S4: the umbrella of genomics.

ACKNOWLEDGEMENTS

The analyses were carried out in part using resources provided by the University of Minnesota Supercomputing Institute. This work was supported by the US Department of Agriculture-Agricultural Research Service [CRIS number 3640–21000–028–00D] and the United Soybean Board [projects 1240 and 2240].

LITERATURE CITED

- Agrawal GK, Hajduch M, Graham K, Thelen JJ. 2008. In-depth investigation of the soybean seed-filling proteome and comparison with a parallel study of rapeseed. *Plant Physiology* **148**: 504–518.
- Al-Ghazi Y, Muller B, Pinloche S, *et al.* 2003. Temporal responses of Arabidopsis root architecture to phosphate starvation: evidence for the involvement of auxin signalling. *Plant, Cell and Environment* **26**: 1053–1066.
- Atwell S, Huang YS, Vihjalmsson BJ, *et al.* 2010. Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. *Nature* **465**: 627–631.
- Atwood SE, O'Rourke JA, Peiffer GA, *et al.* 2014. Replication protein A subunit 3 and the iron efficiency response in soybean. *Plant, Cell and Environment*, **37**: 213–234.
- Bari R, Pant BD, Stitt M, Scheible W-R. 2006. *PHO2*, microRNA399 and *PHR1* define a phosphate signaling pathway in plants. *Plant Physiology* **141**: 988–999.
- Belfield EJ, Gan X, Mithani A, *et al.* 2012. Genome-wide analysis of mutations in mutant lineages selected following fast-neutron irradiation mutagenesis of *Arabidopsis thaliana*. *Genome Research* **22**: 1306–1315.
- Benedito VA, Torres-Jerez I, Murray JD, *et al.* 2008. A gene expression atlas of the model legume *Medicago truncatula*. *The Plant Journal* **55**: 504–513.
- Bielecki RL. 1973. Phosphate pools, phosphate transport, and phosphate availability. *Annual Review of Plant Physiology* **24**: 225–252.
- Bolon Y-T, Vance CP. 2012. Characterization of the linkage group I seed protein QTL in soybean. In: Wilson RF, ed. *Designing soybeans for the 21st century*. AOCSS Press.
- Bolon Y-T, Joseph B, Cannon SB, *et al.* 2010. Complementary genetic and genomic approaches help characterize the linkage group I seed protein QTL in soybean. *BMC Plant Biology* **3**: 10–41.
- Bolon Y-T, Haun WJ, Xu WW, *et al.* 2011. Phenotypic and genomic analyses of a fast neutron mutant population resource in soybean. *Plant Physiology* **156**: 240–253.
- Bolon Y-T, Hyten DL, Orf JH, Vance CP, Muehlbauer GJ. 2013. eQTL networks reveal complex genetic architecture in the immature soybean seed. *Plant Genome* **7**: 10.3835/plantgenome2013.08.0027.
- Boscari A, Del Giudice J, Ferrarini A, *et al.* 2013. Expression dynamics of the *Medicago truncatula* transcriptome during the symbiotic interaction with *Sinorhizobium meliloti*: which role for nitric oxide? *Plant Physiology* **161**: 425–439.
- Branca A, Paape TD, Zhou P, *et al.* 2011. Whole-genome nucleotide diversity, recombination, and linkage disequilibrium in the model legume *Medicago truncatula*. *Proceedings of the National Academy of Sciences USA* **108**: E864–E870.
- Cannon SB, Shoemaker RC. 2012. Evolutionary and comparative analysis of the soybean genome. *Breeding Science* **61**: 437–444.
- Cao J, Schneeberger K, Ossowski S, *et al.* 2011. Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nature Genetics* **43**: 956–963.
- Casimiro I, Marchant A, Bhalerao RP, *et al.* 2001. Auxin transport promotes *Arabidopsis* lateral root initiation. *The Plant Cell* **13**: 843–852.
- Cheng L, Bucciarelli B, Liu J, *et al.* 2011. White lupin cluster root acclimation to phosphorus deficiency and root hair development involve unique glycerophosphodiester phosphodiesterases. *Plant Physiology* **156**: 1131–1148.

- Chia JM, Ware D. 2011. Sequencing for the cream of the crop. *Nature Biotechnology* **29**: 138–139.
- Chiou T-J, Aung K, Lin S-I, Wu C-C, Chiang S-F, Su C-L. 2006. Regulation of phosphate homeostasis by microRNA in Arabidopsis. *The Plant Cell* **18**: 412–421.
- Chiou T-J, Lin S-I. 2011. Signaling network in sensing phosphate availability in plants. *Annual Review of Plant Biology* **62**: 185–206.
- Cordell D, Drangert J, White S. 2009. The story of phosphorus: global food security and food for thought. *Global Environmental Change* **19**: 292–305.
- Cui H, Hao Y, Kovtun M, et al. 2011. Genome-wide direct target analysis reveals a role for SHORT-ROOT in root vascular patterning through cytokinin homeostasis. *Plant Physiology* **157**: 1221–1231.
- Edwards D, Batley J, Snowdon RJ. 2013. Accessing complex crop genomes with next-generation sequencing. *Theoretical and Applied Genetics* **126**: 1–11.
- Fedorova M, van de Mortel J, Matsumoto PA, Cho J, Town CD, VandenBosch KA. 2002. Genome-wide identification of nodule specific transcripts in the model legume *Medicago truncatula*. *Plant Physiology* **130**: 519–537.
- Fu J, Cheng Y, Linghu J, et al. 2013. RNA sequencing reveals the complex regulatory network in the maize kernel. *Nature Communications* **4**: 2832.
- Fujii H, Chiou T-J, Lin S-I, Aung K, Zhu J-K. 2005. A miRNA involved in phosphate-starvation response in Arabidopsis. *Current Biology* **15**: 2038–2043.
- Gan X, Stegle O, Behr J, et al. 2011. Multiple reference genomes and transcriptomes for Arabidopsis thaliana. *Nature* **477**: 419–423.
- Garg R, Jain M. 2013. RNA-Seq for transcriptome analysis in non-model plants. *Methods in Molecular Biology* **1069**: 43–58.
- Garg R, Patel RK, Tyagi AK, Jain M. 2011. *De novo* assembly of chickpea transcriptome using short reads for gene discovery and marker identification. *DNA Research* **18**: 53–63.
- Gilbert GA, Knight JD, Vance CP, Allan DL. 2000. Proteoid root development of phosphorus deficient lupin is mimicked by auxin and phosphonate. *Annals of Botany* **85**: 921–928.
- Gilbert N. 2009. Environment: the disappearing nutrient. *Nature* **461**: 716–718.
- Gill N, Findley S, Walling JG, et al. 2009. Molecular and chromosomal evidence for allopolyploidy in soybean. *Plant Physiology* **151**: 1167–1174.
- Graham PH, Vance CP. 2003. Legumes: importance and constraints to greater use. *Plant Physiology* **131**: 872–877.
- Hajduch M, Ganapathy A, Stein JW, Thelen JJ. 2005. A systematic proteomic study of seed filling in soybean. *Establishment of high-resolution two-dimensional reference maps, expression profiles, and an interactive proteome database*. *Plant Physiology* **137**: 1397–1419.
- Hajduch M, Hearne LB, Miernyk JA, et al. 2010. Systems analysis of seed filling in Arabidopsis: using general linear modeling to assess concordance of transcript and protein expression. *Plant Physiology* **152**: 2078–2087.
- Hammond JP, Broadley MR, Bowen HC, Spracklen WP, Hayden RM, White PJ. 2011. Gene expression changes in phosphorus deficient potato (*Solanum tuberosum* L.) leaves and the potential for diagnostic gene expression markers. *PLoS One* **6**: e24606.
- Han Y, Kang Y., I. Torres-Jerez FC, et al. 2011. Genome-wide SNP discovery in tetraploid alfalfa using 454 sequencing and high resolution melting analysis. *BMC Genomics* **12**: 1–11.
- Hernandez G, Ramirez M, Valdes-Lopez O, et al. 2007. Phosphorus stress in common bean: root transcript and metabolic responses. *Plant Physiology* **144**: 752–767.
- Hsieh L-C, Lin S-I, Shih ACC, et al. 2009. Uncovering small RNA-mediated responses to phosphate deficiency in Arabidopsis by deep sequencing. *Plant Physiology* **151**: 2120–2132.
- Huang X, Feng Q, Qian Q, et al. 2009. High-throughput genotyping by whole-genome resequencing. *Genome Research* **19**: 1068–1076.
- Huang X, Wei X, Sang T, et al. 2010. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nature Genetics* **42**: 961–967.
- Huang X, Lu T, Han B. 2013. Resequencing rice genomes: an emerging new era of rice genomics. *Trends in Genetics* **29**: 225–232.
- Hyten DL, Song Q, Fickus EW, et al. 2010. High-throughput SNP discovery and assay development in common bean. *BMC Genomics* **11**: 475–485.
- Jiang C, Gao X, Liao L, Harberd NP, Fu X. 2007. Phosphate starvation, root architecture and anthocyanin accumulation responses are modulated by the gibberellin–DELLA signaling pathway in Arabidopsis. *Plant Physiology* **145**: 1460–1470.
- Jiao Y, Zhao H, Ren R, et al. 2012. Genome wide genetic changes during modern breeding of maize. *Nature Genetics* **44**: 812–815.
- Johnson JF, Allan DL, Vance CP. 1994. Phosphorus stress-induced proteoid roots show altered metabolism in *Lupinus albus*. *Plant Physiology* **104**: 657–665.
- Johnson J, Allan D, Vance CP, Weiblen G. 1996. Root carbon dioxide fixation by phosphorus-deficient *Lupinus albus* (contribution to organic acid exudation by proteoid roots). *Plant Physiology* **112**: 19–30.
- Jones-Rhoades MW, Bartel DP. 2004. Computational identification of plant microRNAs and their targets, including stress-induced miRNA. *Molecular Cell* **14**: 787–799.
- Joshi T, Valliyodan B, Wu J-H, Lee S-H, Xu D, Nguyen HT. 2013. Genomic differences between cultivated soybean, *G. max* and its wild relative *G. soja*. *BMC Genomics* **14** (Suppl): 55.
- Kalavacharia V, Liu Z, Meyers BC, Thimmapuram J, Melmaiee K. 2011. Identification and analysis of common bean (*Phaseolus vulgaris* L.) transcriptomes by massively parallel pyrosequencing. *BMC Plant Biology* **11**: 135–107.
- Kaur S, Cogan NO, Pembleton LW, et al. 2011. Transcriptome sequencing of lentil based on second-generation technology permits large-scale unigene assembly and SSR marker discovery. *BMC Genomics* **12**: 265.
- Kim MY, Lee S, Van K, et al. 2010. Whole-genome sequencing and intensive analysis of the undomesticated soybean (*Glycine soja* Sieb. and Zucc.) genome. *Proceedings of the National Academy of Sciences, USA* **107**: 22032–22037.
- Krusell L, Krause K, Ott T, et al. 2005. The sulfate transporter SST1 is crucial for symbiotic nitrogen fixation in *Lotus japonicus* root nodules. *The Plant Cell*, **17**: 1625–1636.
- Lai J, Li R, Xu X, et al. 2010. Genome-wide patterns of genetic variation among elite maize inbred lines. *Nature Genetics* **42**: 1027–1031.
- Lam H-M, Xu X, Lui X, et al. 2010. Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nature Genetics* **42**: 1053–1059.
- Lambers H, Raven JA, Shaver GR, Smith SE. 2008. Plant nutrient-acquisition strategies change with soil age. *Trends in Ecology and Evolution* **23**: 95–103.
- Le BH, Wagmaister JA, Bui AQ, Harada JJ, Goldberg RB. 2007. Using genomics to study legume seed development. *Plant Physiology* **144**: 562–574.
- Le BH, Cheng C, Bui AQ, et al. 2010. Global analysis of gene activity during Arabidopsis seed development and identification of seed-specific transcription factors. *Proceedings of the National Academy of Sciences, USA* **107**: 8063–8070.
- Lelandais-Briere C, Naya L, Sallet E, et al. 2009. Genome-wide *Medicago truncatula* small RNA analysis revealed novel microRNAs and isoforms differentially regulated in roots and nodules. *The Plant Cell* **21**: 2780–2796.
- Leshchiner I, Alexa K, Kelsey P, et al. 2012. Mutation mapping and identification by whole-genome sequencing. *Genome Research* **22**: 1541–1548.
- Li H, Peng Z, Yang X, et al. 2012. Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. *Nature Genetics* **45**: 43–50.
- Li HC, Chuang K, Henderson JT, et al. 2005. PICKLE acts during germination to repress expression of embryonic traits. *The Plant Journal* **44**: 1010–1022.
- Li L, Liu C, Lian X. 2010. Gene expression profiles in rice roots under low phosphorus stress. *Plant Molecular Biology* **72**: 423–432.
- Li X, Mo X, Shou H, Wu P. 2006. Cytokinin-mediated cell cycling arrest of pericycle founder cells in lateral root initiation of Arabidopsis. *Plant and Cell Physiology* **47**: 1112–1123.
- Libault M, Stacey G. 2010. Evolution of FW2.2-like (FWL) and PLAC8 genes in eukaryotes. *Plant Signaling and Behavior* **5**: 1226–1228.
- Libault M, Farmer A, Joshi T, et al. 2010. An integrated transcriptome atlas of the crop model *Glycine max*, and its use in comparative analyses in plants. *The Plant Journal* **63**: 86–99.
- Lister R, O'Malley RC, Tonti-Filippini J, et al. 2008. Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell* **133**: 523–536.
- Liu CM, Muchhal US, Uthappa M, Kononowicz AK, Raghothama KG. 1998. Tomato phosphate transporter genes are differentially regulated in plant tissues by phosphorus. *Plant Physiology* **116**: 91–99.
- Lo S-F, Yang S-Y, Chen K-T, et al. 2008. A novel class of gibberellin 2-oxidases control semidwarfism, tillering, and root development in rice. *The Plant Cell* **20**: 2603–2618.
- Lopez-Bucio J, Hernandez-Abreu E, Sanchez-Calderon L, Nieto-Jacobo MF, Simpson J, Herrera-Estrella L. 2002. Phosphate availability alters architecture and causes changes in hormone sensitivity in the Arabidopsis root system. *Plant Physiology* **129**: 244–256.

- Ma Z, Baskin TI, Brown KM, Lynch JP. 2003. Regulation of root elongation under phosphorus stress involves changes in ethylene responsiveness. *Plant Physiology* **131**: 1381–1390.
- Matas AJ, Yeats TH, Buda GJ, et al. 2011. Tissue and cell-type specific transcriptome profiling of expanding tomato fruit provides insights into metabolic and regulatory specialization and cuticle formation *The Plant Cell* **23**: 3893–3910.
- McMullen MD, Kreshovich S, Villeda HS, et al. 2009. Genetic properties of the maize nested association mapping population. *Science* **7**: 737–740.
- Meyer K, Stecca KL, Ewell-Hicks K, Allen SM, Everard JD. 2012. Oil and protein accumulation in developing seeds is influenced by the expression of a cytosolic pyrophosphatase in *Arabidopsis*. *Plant Physiology* **159**: 1221–1234.
- Misson J, Raghothama K, Jain A, et al. 2005. A genome-wide transcriptional analysis using *Arabidopsis thaliana* Affymetrix gene chips determined plant responses to phosphate deprivation. *Proceedings of the National Academy of Sciences, USA* **102**: 11934–11939.
- Mochida K, Shinozaki K. 2011. Advances in omics and bioinformatics tools for systems analyses of plant functions. *Plant and Cell Physiology* **52**: 2017–2038.
- Morcuede R, Bari R, Gibon Y, et al. 2007. Genome-wide reprogramming of metabolism and regulatory networks of *Arabidopsis* in response to phosphorus. *Plant, Cell and Environment* **30**: 85–112.
- Mosquna A, Katz A, Shochat S, Grafi G, Ohad N. 2004. Interaction of FIE, a polycomb protein with pRB: a possible mechanism regulating endosperm development. *Molecular Genetics and Genomics* **271**: 651–657.
- Mudge J, Cannon SB, Kalo P, et al. 2005. Highly syntenic regions in the genomes of soybean, *Medicago truncatula*, and *Arabidopsis thaliana*. *BMC Plant Biology* **5**: 15.
- Nagarajan VK, Smith AP. 2012. Ethylene's role in phosphate starvation signaling: more than just a root growth regulator. *Plant and Cell Physiology* **52**: 277–286.
- Neumann G, Romheld V. 1999. Root excretion of carboxylic acids and protons in phosphorus-deficient plants. *Plant and Soil* **211**: 121–130.
- Nichols DM, Glover KD, Carlson SR, Specht E, Diers BW. 2006. Fine mapping of a seed protein QTL on soybean linkage group I and its correlated effects on agronomic traits. *Crop Science* **46**: 834–839.
- Niu YF, Chai RS, Jin GL, Wang H, Tang CX, Zhang YS. 2013. Responses of root architecture development to low phosphorus availability: a review. *Annals of Botany* **112**: 391–408.
- Ohlrogge JB, Kuo T-M. 1984. Control of lipid synthesis during soybean seed development: enzymic and immunochemical assay of acyl carrier protein. *Plant Physiology* **74**: 622–625.
- O'Rourke JA, Iniguez LP, Bucciarelli B, et al. 2013a. A re-sequencing based assessment of genomic heterogeneity and fast neutron-induced deletions in a common bean cultivar. *Frontiers in Plant Science* **4**: 210.
- O'Rourke JA, Yang SS, Miller SS, et al. 2013b. An RNA-Seq transcriptome analysis of orthophosphate-deficient white lupin reveals novel insights into phosphorus acclimation in plants. *Plant Physiology* **161**: 705–724.
- Ossowski S, Schneeberger K, Clark RM, Lanz C, Warthmann N, Weigel D. 2008. Sequencing of natural strains of *Arabidopsis thaliana* with short reads. *Genome Research* **18**: 2024–2033.
- Ozsolak F, Milos PM. 2011. RNA sequencing: advances, challenges and opportunities. *Nature Reviews Genetics* **12**: 87–98.
- Peaz P, Trejo MS, Iniguez LP, et al. 2012. Identification and characterization of microRNAs in *Phaseolus vulgaris* by high-throughput sequencing. *BMC Genomics* **13**: 83.
- Peiffer GA, King KE, Severin AJ, et al. 2012. Identification of candidate genes underlying an iron efficiency quantitative trait locus in soybean. *Plant Physiology* **158**: 1745–1754.
- Plaxon WC, Tran HT. 2011. Metabolic adaptations of phosphate-starved plants. *Plant Physiology* **156**: 1006–1015.
- Poland JA, Rife TW. 2012. Genotyping-by-sequencing for plant breeding and genetics. *Plant Genome* **5**: 92–102.
- Robinson JT, Thorvaldsdottir H, Winckler W, et al. 2011. Integrative genomics viewer. *Nature Biotechnology* **29**: 24–26.
- Rubio V, Bustos R, Irigoyen ML, Cardona-Lopez X, Rojas-Triana M, Paz-Ares J. 2009. Plant hormones and nutrient signaling. *Plant Molecular Biology* **69**: 361–373.
- Sato S, Nakamura Y, Kaneko T, et al. 2008. Genome structure of the legume, *Lotus japonicus*. *DNA Research* **15**: 227–239.
- Schlueter JA, Dixon P, Granger C, et al. 2004. Mining EST databases to resolve evolutionary events in major crop species. *Genome* **47**: 868–876.
- Schmitz RJ, Zhang X. 2011. High-throughput approaches for studying plant epigenomics. *Current Opinion in Plant Biology* **14**: 130–136.
- Schmutz J, Cannon SB, Schlueter J, et al. 2010. Genome sequence of the palaeopolyploid soybean. *Nature* **467**: 178–183.
- Schneeberger K, Weigel D. 2011. Fast-forward genetics enabled by new sequencing technologies. *Trends in Plant Science* **16**: 282–288.
- Seboldt AM, Shoemaker RC, Diers BW. 2000. Analysis of a quantitative trait locus allele from wild soybean that increases seed protein concentration in soybean. *Crop Science* **40**: 1438–1444.
- Severin AJ, Peiffer GA, Xu WW, et al. 2010a. An integrative approach to genomic introgression mapping. *Plant Physiology* **154**: 3–12.
- Severin AJ, Woody JL, Bolon Y-T, et al. 2010b. RNA-Seq atlas of *Glycine max*: a guide to the soybean transcriptome. *BMC Plant Biology* **10**: 160.
- Shi T, Li R, Zhao Z, et al. 2013. QTL for yield traits and their association with functional genes in response to phosphorus deficiency in *Brassica napus*. *PLoS One* **8**: e54559.
- Simon SA, Meyers BC, Sherrier DJ. 2009. MicroRNAs in the rhizobia legume symbiosis. *Plant Physiology* **151**: 1002–1008.
- Sonah H, Bastien M, Iqura E, et al. 2013. An improved genotyping by sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. *PLoS One* **8**: e54603.
- Song QX, Liu YF, Hu XY, et al. 2011. Identification of miRNAs and their target genes in developing soybean seeds by deep sequencing. *BMC Plant Biology* **10**: 5.
- Sunkar R. 2010. MicroRNAs with macro-effects on plant stress responses. *Seminars in Cell and Developmental Biology* **21**: 805–811.
- Sunkar R, Zhu J-K. 2004. Novel and stress-regulated microRNAs and other small RNAs from *Arabidopsis*. *The Plant Cell* **16**: 2001–2019.
- Svistonoff S, Creff A, Reymond M, et al. 2007. Root tip contact with low-phosphate media reprograms plant root architecture. *Nature Genetics* **39**: 792–796.
- Thibaud M-C, Arrighi J-F, Bayle V, et al. 2010. Dissection of local and systemic transcriptional responses to phosphate starvation in *Arabidopsis*. *The Plant Journal* **64**: 775–789.
- Thimm O, Blasing O, Gibon Y, et al. 2004. MAPMAN: a user driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *The Plant Journal* **37**: 914–939.
- Tomato Genome Consortium. 2012. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* **485**: 635–641.
- Torrey JG. 1976. Root hormones and plant growth. *Annual Review of Plant Physiology* **27**: 435–459.
- Toth K, Stratil TF, Madsen EB, et al. 2012. Functional domain analysis of the remorin protein LjSYMREM1 in *Lotus japonicus*. *PLoS One* **7**: e30817.
- Turner M, Yu O, Subramanian S. 2012. Genome organization and characteristics of soybean microRNAs. *BMC Genomics*, **13**: 169.
- Uhde-Stone C, Zinn KE, Ramirez-Yanez M, Li A, Vance CP, Allan DL. 2003. Nylon filter arrays reveal differential gene expression in proteoid roots of white lupin in response to phosphorus deficiency. *Plant Physiology* **131**: 1064–1079.
- Valdes-Lopez O, Yang SS, Aparicio-Fabre R, et al. 2010. MicroRNA expression profile in common bean (*Phaseolus vulgaris*) under nutrient deficiency stresses and manganese toxicity. *New Phytologist* **187**: 805–818.
- Vance CP. 2001. Symbiotic nitrogen fixation and phosphorus acquisition. *Plant nutrition in a world of declining renewable resources*. *Plant Physiology* **127**: 390–397.
- Vance CP, Heichel GH. 1981. Nitrate assimilation during vegetative regrowth of alfalfa. *Plant Physiology* **68**: 1052–1057.
- Vance CP, Uhde-Stone C, Allan DLP. 2003. Phosphorus acquisition and use: critical adaptations by plants for securing a nonrenewable resource. *New Phytologist*, **157**: 423–447.
- Varala K, Swaminathan K, Li Y, Hudson ME. 2011. Rapid genotyping of soybean cultivars using high throughput sequencing. *PLoS One* **6**: e24811.
- Varshney RK, Song C, Saxena RK, et al. 2013. Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nature Biotechnology*, **31**: 240–246.
- Varshney RK, Chen W, Li Y, et al. 2012. Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. *Nat Biotechnol*, **30**: 83–9.
- Verdier J, Kaker K., Gallardo K., et al. 2008. Gene expression profiling of *M. truncatula* transcription factors identifies putative regulators of grain legume seed filling. *Plant Molecular Biology*, **67**: 567–580.

- Verdier J, Thompson RD. 2008.** Transcriptional regulation of storage protein synthesis during dicotyledon seed filling. *Plant and Cell Physiology* **49**: 1263–1271.
- Wang F, Perry SE. 2013.** Identification of direct targets of FUSCA3, a key regulator of Arabidopsis seed development. *Plant Physiology* **161**: 1251–1264.
- Wang J, Mullighan CG, Easton J, et al. 2011.** CREST maps somatic structural variation in cancer genomes with base-pair resolution. *Nature Methods* **8**: 652–654.
- Wang L, Li P, Brutnell TP. 2010.** Exploring plant transcriptomes using ultra high-throughput sequencing. *Briefings in Functional Genomics* **9**: 118–128.
- Wang Y-H, Garin DF, Kochian LV. 2002.** Rapid induction of regulatory and transporter genes in response to phosphorus, potassium, and iron deficiencies in tomato roots. Evidence for cross talk and root/rhizosphere-mediated signals. *Plant Physiology*, **130**: 1261–1370.
- Wang Z, Gerstein M, Snyder M. 2009.** RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* **10**: 57–63.
- Wang Z, Fang B, Chen J, et al. 2010a.** De novo assembly and characterization of root transcriptome using Illumina paired-end sequencing and development of cSSR markers in sweet potato (*Ipomoea batatas*). *BMC Genomics* **11**: 726.
- Wang Z, Libault M, Joshi T, et al. 2010b.** SoyDB: a knowledge database of soybean transcription factors. *BMC Plant Biology* **10**: 14.
- Weber H, Borisjuk L, Wobus U. 2005.** Molecular physiology of legume seed development. *Annual Review of Plant Biology* **56**: 253–279.
- Weigel D, Colot V. 2012.** Epialleles in plant evolution. *Genome Biology* **13**: 249.
- Werner T, Nehnevajova E, Kollmer I, et al. 2010.** Root-specific reduction of cytokinin causes enhanced root growth, drought tolerance, and leaf mineral enrichment in Arabidopsis and tobacco. *The Plant Cell* **22**: 3905–3920.
- Wilcox JR. 2004.** World distribution and trade of soybean. In: Boerma HR, Specht JE, eds. *Soybean: improvement, production, and uses*. Madison, WI: American Society of Agronomy, 1–14.
- Xia Z, Xu H, Zhai J, et al. 2011.** RNA-Seq analysis and de novo transcriptome assembly of *Hevea brasiliensis*. *Plant Molecular Biology* **77**: 299–308.
- Yang SS, Tu JZ, Foo C, et al. 2011.** Using RNA-Seq for gene identification, polymorphism detection and transcript profiling in two alfalfa genotypes with divergent cell wall composition in stems. *BMC Genomics* **12**: 199.
- Young ND, Debelle F, Oldroyd GE, et al. 2011.** The Medicago genome provides insight into the evolution of rhizobial symbioses. *Nature* **480**: 520–524.
- Zenoni S, Ferraini A, Giacomelli E, et al. 2010.** Characterization of transcriptional complexity during berry development in *Vitis vinifera* using RNA-Seq. *Plant Physiology* **152**: 1787–1795.
- Zhang G, Guo G, Hu X, et al. 2010.** Deep RNA sequencing at single base-pair resolution reveals high complexity of the rice transcriptome. *Genome Research* **20**: 646–654.
- Zhang J, Liang S, Duan J, et al. 2012.** De novo assembly and characterization of the transcriptome during seed development, and generation of genic-SSR markers in peanut (*Arachis hypogaea* L.). *BMC Genomics* **12**: 13–90.
- Zhang YL, Lynch JP, Brown KM. 2003.** Ethylene and phosphorus availability have interacting yet distinct effects on root hair development. *Journal of Experimental Botany* **54**: 2351–2361.
- Zhu Y, Zeng H, Dong C, Yin X, Shen Q, Yang Z. 2009.** MicroRNA expression profiles associated with phosphorus deficiency in white lupin (*Lupinus albus* L.). *Plant Science* **178**: 23–29.
- Zuryn S, Le Gras S, Jamet K, Jarriault S. 2010.** A strategy for direct mapping and identification of mutations by whole-genome sequencing. *Genetics* **186**: 427–430.