

RESEARCH ARTICLE

Open Access

# Genetic variants in root architecture-related genes in a *Glycine soja* accession, a potential resource to improve cultivated soybean

Silvas J Prince<sup>1†</sup>, Li Song<sup>1†</sup>, Dan Qiu<sup>1</sup>, Joao V Maldonado dos Santos<sup>1,2</sup>, Chenglin Chai<sup>1</sup>, Trupti Joshi<sup>2,3</sup>, Gunvant Patil<sup>1</sup>, Babu Valliyodan<sup>1</sup>, Tri D Vuong<sup>1</sup>, Mackensie Murphy<sup>1</sup>, Konstantinos Krampis<sup>4</sup>, Dominic M Tucker<sup>4</sup>, Ruslan Biyashev<sup>4</sup>, Anne E Dorrance<sup>5</sup>, MA Saghai Maroof<sup>4</sup>, Dong Xu<sup>2,3</sup>, J Grover Shannon<sup>1</sup> and Henry T Nguyen<sup>1,2\*</sup>

## Abstract

**Background:** Root system architecture is important for water acquisition and nutrient acquisition for all crops. In soybean breeding programs, wild soybean alleles have been used successfully to enhance yield and seed composition traits, but have never been investigated to improve root system architecture. Therefore, in this study, high-density single-feature polymorphic markers and simple sequence repeats were used to map quantitative trait loci (QTLs) governing root system architecture in an inter-specific soybean mapping population developed from a cross between *Glycine max* and *Glycine soja*.

**Results:** Wild and cultivated soybean both contributed alleles towards significant additive large effect QTLs on chromosome 6 and 7 for a longer total root length and root distribution, respectively. Epistatic effect QTLs were also identified for taproot length, average diameter, and root distribution. These root traits will influence the water and nutrient uptake in soybean. Two cell division-related genes (*D type cyclin* and *auxin efflux carrier protein*) with insertion/deletion variations might contribute to the shorter root phenotypes observed in *G. soja* compared with cultivated soybean. Based on the location of the QTLs and sequence information from a second *G. soja* accession, three genes (*slow anion channel associated 1 like*, *Auxin responsive NEDD8-activating complex* and *peroxidase*), each with a non-synonymous single nucleotide polymorphism mutation were identified, which may also contribute to changes in root architecture in the cultivated soybean. In addition, *Apoptosis inhibitor 5-like* on chromosome 7 and *slow anion channel associated 1-like* on chromosome 15 had epistatic interactions for taproot length QTLs in soybean.

**Conclusion:** Rare alleles from a *G. soja* accession are expected to enhance our understanding of the genetic components involved in root architecture traits, and could be combined to improve root system and drought adaptation in soybean.

**Keywords:** Root, Quantitative trait locus, Soybean, Wild soybean, Root architecture, Non-synonymous SNP, Microarray, Single feature polymorphism, DNA sequencing

## Background

*Glycine soja*, the annual wild progenitor of cultivated soybean, is widely distributed among East Asian countries. In China, the cultivated soybean was domesticated from wild soybean more than 5,000 years ago [1] and underwent two rounds of whole genome duplication [2].

\* Correspondence: [nguyenhenry@missouri.edu](mailto:nguyenhenry@missouri.edu)

†Equal contributors

<sup>1</sup>National Center for Soybean Biotechnology and Division of Plant Sciences, University of Missouri, Columbia, MO 65211, USA

<sup>2</sup>Christopher S. Bond Life Sciences Center, University of Missouri, Columbia, MO 65211, USA

Full list of author information is available at the end of the article

The first genome duplication occurred within the last 60 million years and the latter between 5 and 13 million years ago. Both *G. soja* and *Glycine max* have prominent differences for various morphological and physiological characters, known as domestication syndrome [3]. In soybean, the process of plant breeding accelerated genetic gain and narrowed the genetic base [4]. The genetic diversity among 99% of North American cultivars released between 1947 and 1988 could be traced back to only 0.02% of the landraces [1]. This loss in diversity among high-yielding adapted lines ultimately inhibits

future genetic gains in productivity, broadens susceptibility to new pests and diseases, and acts as a threat to food security [4]. In contrast to modern soybean cultivars, wild soybeans are genetically diverse, with valuable rare alleles [5]. Recent advances in sequencing technologies also highlighted the uniqueness of genomic content in both cultivated and wild soybean, and provide an opportunity to use *G. soja* to broaden the genetic base of cultivated soybean [6,7]. In addition, assessing genomic differences for key traits will provide insights into the process of speciation and domestication, and will deepen our understanding of the origin of genes involved in complex traits [8].

Earlier studies showed that the presence of unique alleles in wild/weedy species and primitive land races could be used to improve agronomic traits in crop plants [9]. Later, alleles were successfully introgressed from wild species and deployed in different crops through genetic mapping and molecular marker approaches [9,10]. A number of array-based high-throughput marker genotyping platforms have been used in plant breeding, especially marker-assisted selection, to understand crop domestication and plant evolution [11]. These microarray-based markers have been used for high-density molecular map construction, quantitative trait locus (QTL)/expression QTL mapping, and genetic diversity analysis [11]. Among these array-based markers, single-feature polymorphism (SFP) was originally used for fine mapping and positional cloning of genes in yeast [12]. Later, it was used in plant species with both small and complex genomes [11]. SFPs have been widely used for different applications, such as molecular linkage map construction and QTL mapping in *Arabidopsis* [13], as well as in major cereal crops [14] and legumes [15].

The effective use of wild relatives to improve a wide variety of traits from yield to stress tolerance in cultivated/domesticated crops was reviewed [16] and has been successfully applied in rice [17] and wheat [18]. Similarly, inter-specific variation in soybean was used to identify novel alleles in *G. soja* that influence various traits, including domestication [19], alkaline and salt tolerance [20], dehydration tolerance [21], yield [22], resistance to pathogens and pests, and seed compositional traits [23]. Among abiotic stresses, drought stress causes tremendous yield losses in soybean [24]. Drought avoidance is considered to be the most relevant process to mitigate agricultural drought and maintain crop performance [25]. Root system architecture (RSA) and root hydraulics are the key traits that affect water capture under drought-prone environments [26,27] and sustain yield in sub-optimal conditions. Thus, RSA and root distribution within the environment are important to understand nutrient and water use efficiency in plants [28]. Recent studies in rice have shown that an increase

in root depth leads to an increase in water uptake, which is translated into higher grain yield under rain-fed conditions [29]. The existence of genetic variation for root growth and architecture within various crop species makes RSA a promising target for crop improvement programs [30]. A recent study of inter-specific tomato introgression lines also emphasized the need to identify genes associated with favorable root traits and their transcription regulation [31]. To the best of our knowledge, *G. soja* alleles have never been used to improve root system architecture. This is understandable because *G. soja* roots are often very thin, with narrow hairs, as well as reduced root mass and volume. Thus, the objectives of the present study were (i) to identify novel alleles from a *G. soja* accession to explore the possibility of enhancing root architectural traits in cultivated soybean; (ii) to detect significant QTL regions and identify candidate genes governing root traits, and (iii) to understand the mechanisms regulating the transcript variation in an inter-specific mapping population.

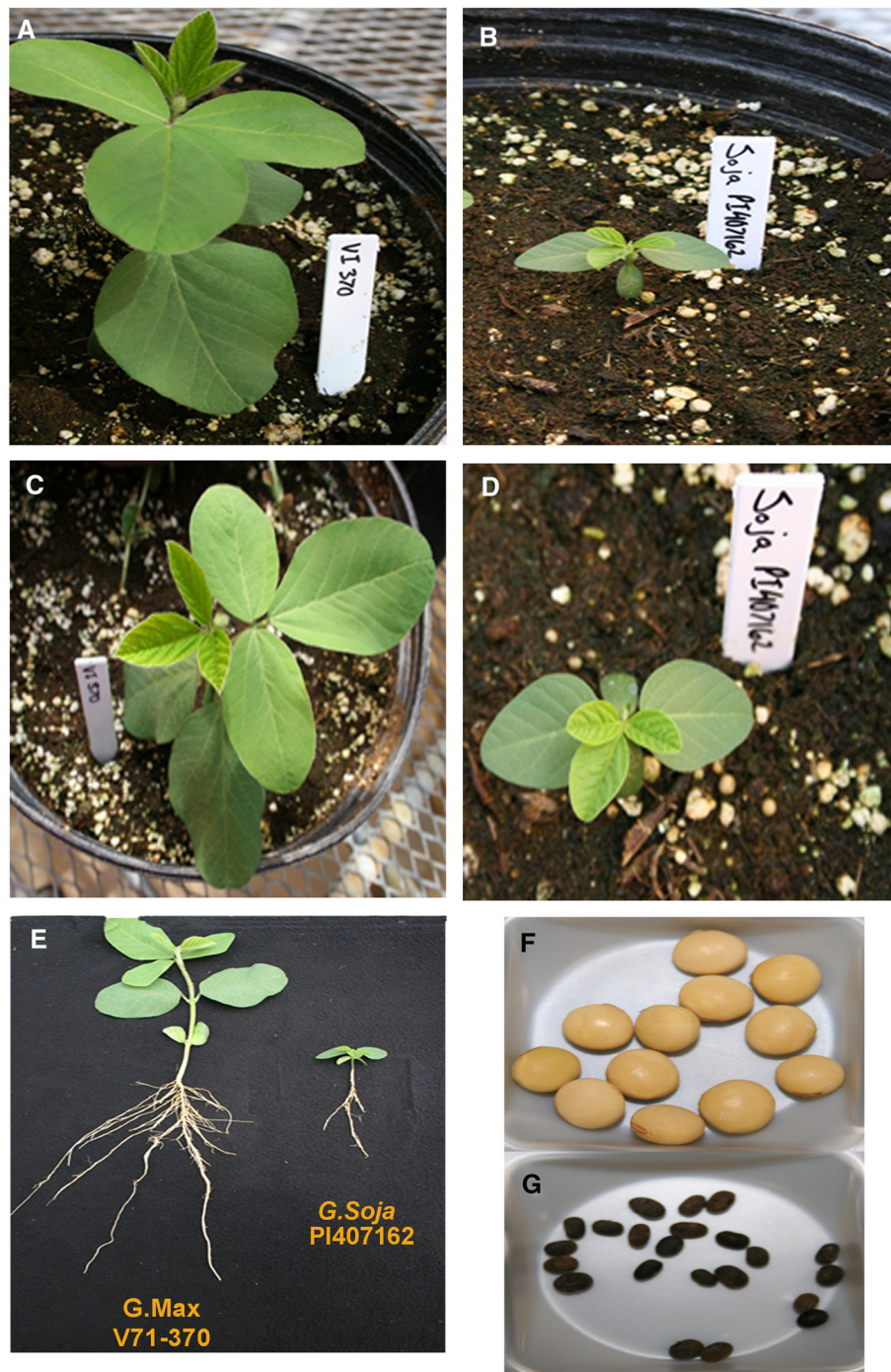
## Results

### Phenotypic variation of root traits

The parents, V71-370 and PI 407162, show significant variation for a number of common soybean traits including: plant stature, root morphology, and seed size (Figure 1). The *G. max* V71-370 parent develops a larger root system than the wild soybean parent, *G. soja* PI407162. The recombinant inbred lines (RILs) developed from these two parents showed a transgressive segregation for root morphology (Figure 2): many genotypes had longer or shorter taproots and varied total root length compared with the *G. max* and wild parent, respectively (Table 1). The phenotypic mean of all the RILs were similar to the mid-parent values for all remaining traits that were measured. A Shapiro–Wilk test showed that the frequency distributions of these traits were approximately normal (Additional file 1: Figure S1). Most of the previous mapping works in root studies focused on coarse/thicker roots (tap or lateral roots); however, understanding the finer roots and their distribution are important, because they are the ones involved in nutrient and water absorption by increasing the root surface area. Significant positive correlations ( $P < 0.01$ ) were found amongst various root traits measured in this study (Table 2). The taproot length and tertiary root length were highly correlated (0.8) with root surface area, which influences the plant nutrient and water absorption. Most of the fine roots and their distribution had similar correlations with root volume (Table 2).

### Root QTLs and their interaction

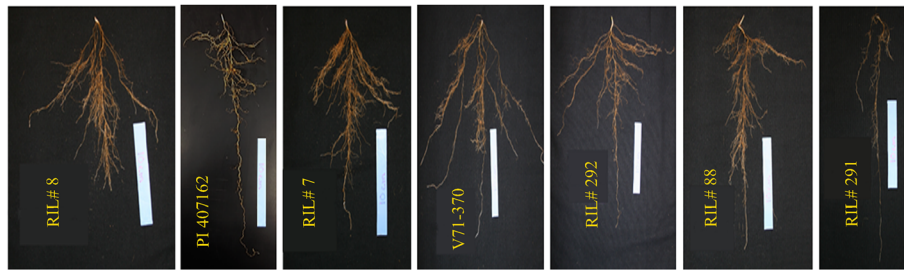
The root QTLs identified in this study was flanked by Affymetrix probe sets (Table 3). The genes within each



**Figure 1** Variation among parental lines, V71-370 and PI407162 for plant morphology (A-E), first trifoliate leaf size, root architecture, and seed traits (A, C, F: V71-370; B, D, G: PI407162).

QTL interval were selected based on transcript abundance from a previous study [32] and are listed in Table 4. Four significant QTLs were associated with different root architectural traits on chromosomes 6 and 7 (Table 3). Two significant large effect QTLs for root surface area (SA) and thickness (RDT2) were also identified

on chromosomes 6 and 7 (Figure 3). The QTL on chromosome 6 (flanked by 4222.1.S1\_10 and 77599.1.S1\_7) is contributed by PI 407162 and explained >10% of the phenotypic variation for SA and total root length (TRL), with a higher additive effect (Table 3). This array-based mapping approach narrowed the confidence



**Figure 2** Transgressive segregation pattern for root traits among recombinant inbred (RI) lines of the mapping population (V71-370/PI407162).

interval of this key QTL region on chromosome 6 to 1.3 cM. This genomic region was also associated with other root traits, such as total root length (Table 3), tertiary root length, and root volume (Additional file 2: Table S1). The root thickness QTL on chromosome 7 (between 59884.1.S1\_8 and 8398.1.S1\_11) was contributed by V71-370, explaining 15% of the phenotypic variation. The QTL region on chromosome 7 contributed to root distribution in different diameter classes (Table 3) and other traits (Additional file 2: Table S1). Interestingly, both loci had additive effects for their respective traits. Epistatic effects were detected for three pairs of loci (Table 5), and none of these loci were identified as QTLs with single effects. Two loci that contributed to taproot length on chromosome 7 (between 6648.1.S1\_11 and 5451.1.S1\_5) interacted with the QTL on chromosome 15 (between 6807.1.S1\_10 and 9882.1.S1\_10) (Figure 4). This epistatic interaction accounted for 8% of the phenotypic variance. A similar interaction was detected for the average diameter between chromosome 4 (between Satt164 and 4792.1.A1\_5) and chromosome 15 (between 15910.1.A1\_10 and 6807.1.S1\_10), which

accounted for 7% of the phenotypic variance. For root thickness class 2, an interaction between chromosome 8 (between 70452.1.S1\_3 and 55124.1.S1\_7) and 9 (between Sat043 and 16443.1.A1\_2) was identified that explained 6% of the phenotypic variance.

#### Identification of genes associated with the root QTLs

Twenty-three genes were selected based on their transcript abundance (Table 4) in the microarray analysis [32] within the mapped QTL intervals (Tables 3 and 5; Additional file 2: Table S1). To study the genes specific to the wild soybean within the QTL interval on chromosome 6, sequence information of 162 genes (Table 6) were extracted from another *G. soja* accession, IT182932 [8] and annotated as shown in Additional file 3: Figure S2. The sequence of each of the 162 genes was compared between Williams 82 with another *G. soja* accession; 18 of these genes (Tables 7 and 8) were selected for expression analysis using qRT-PCR. Nine of the 18 genes (Table 7) had high transcript abundance in root tissues already in the soybean transcriptome database [33], while the remaining nine genes had non-synonymous mutations

**Table 1** Phenotypic variation of root traits significant at P value < 0.0001, based on analysis of variance

Traits	Parental lines		RI population		
	V71-370	PI407162	Mean ± SD	Minimum	Maximum
Tap root length (cm)	21.5	17.2	24.3 ± 0.42	12.7	36.8
Root fresh weight (g)	2.0	0.4	0.9 ± 0.03	0.2	2.3
Total root length (cm)	425.0	238.0	430.9 ± 11.4	123.4	805.1
Surface area (cm <sup>2</sup> )	73.2	29.1	53.2 ± 1.2	16.3	92.3
Average diameter (mm)	0.6	0.4	0.4 ± 0.01	0.3	0.6
Root volume (cm <sup>3</sup> )	1.0	0.2	0.54 ± 0.01	0.2	1.3
Lateral average diameter (mm)	0.6	0.4	0.5 ± 0.01	0.3	0.7
Tertiary root number	865	439	635.5 ± 22.1	116.0	1650.0
Tertiary root length (cm)	317	191	259.6 ± 8.8	33.5	727.7
Root distribution based on length in diameter (1.0-1.5 mm)	24.0	2.9	9.8 ± 0.4	1.7	28.2
Root distribution based on surface area in diameter (1.0-1.5 mm)	9.2	1.4	3.6 ± 0.2	0.6	10.1
Root distribution based on volume in diameter (1.0-1.5 mm)	0.3	0.04	0.1 ± 0.0	0.0	0.3
Root distribution based on thickness in diameter classification 2 (0.5-1.0 mm)	11.0	0.3	10.8 ± 0.8	0.3	46.5

RI, Recombinant inbred; SD, Standard deviation.

**Table 2 Correlation coefficients among various root traits measured in the present study**

Traits	TRTL	RFW	TRL	SA	AD	RV	LAD	TERN	TERL	L3	SA3	PA3	V3	T2
TRTL	1													
RFW	-.199*	1												
TRL	.478**	.005	1											
SA	.420**	.285**	.897**	1										
AD	-.324**	.421**	-.509**	-.147	1									
RV	.239**	.531**	.518**	.832**	.327**	1								
LAD	-.346**	.479**	-.504**	-.138	.792**	.342**	1							
TERN	.132	.152	.741**	.682**	-.338**	.412**	-.314**	1						
TERL	.308**	.176*	.824**	.799**	-.317**	.532**	-.307**	.924**	1					
RDL3	.241**	.389**	.137	.482**	.466**	.739**	.477**	.029	.157	1				
RDSA3	.230**	.399**	.131	.474**	.478**	.738**	.486**	.022	.148	.999**	1			
RDV3	.130	.361**	.100	.413**	.436**	.661**	.457**	.066	.133	.855**	.857**	.853**	1	
RDT2	.421**	-.052	.165*	.220**	.010	.186*	-.083	-.141	.022	.351**	.344**	.340**	.272**	1

Correlations were performed using 160 RI lines of the population. Data used were the means of four replications of independent measurements. Pairwise correlation coefficients were significant at the \*5% or \*\*1% significance level.

TRTL, Tap root length (cm); RFW, Root fresh weight (g); TRL, Total root length (cm); SA, Surface area (cm<sup>2</sup>); AD, Average diameter (mm); RV, Root volume (cm<sup>3</sup>); LAD, Lateral average diameter (mm); TERN, Tertiary root number; TERL, Tertiary root length (cm); RDL3, Root distribution in length classification 3 (1.0–1.5 mm); RDSA3, Root distribution in surface area classification 3 (1.0–1.5 mm); RDV3, Root distribution in volume classification 3 (1.0–1.5 mm); RDT2, Root distribution in thickness classification 2 (0.5–1.0 mm).

(Table 8). Based on the normalized tissue-specific expression pattern in root tissues in the soybean Affymetrix gene chip dataset [34], key genes were identified (Figure 5) for quantitative RT-PCR analysis.

#### Presence of non-synonymous mutation in root-related genes

Sequence data for the parental lines V71-370 and PI 407162 were analyzed for the presence of non-synonymous SNPs within the 23 selected genes, based on transcript abundance from a microarray analysis, to identify candidate genes that might contribute to variation in root phenotypes. Only three of these genes (Table 4), had non-synonymous SNPs, two in the *G. soja* line, PI 407162 (*Glyma07g09860* and *Glyma15g42220*) and one gene (*Glyma07g32480*) in both parental lines that had altered the amino acid content (Table 9). The gene *Glyma07g09860* encodes triglyceride lipase, showed higher transcript abundance and was in the QTLs identified for root distribution based on length and thickness

(Table 3). The remaining two genes, *Glyma07g32480* and *Glyma15g42220*, encode apoptosis inhibitory 5 family protein and oxidoreductase/transition metal ion binding protein, respectively. These two genes also showed higher transcript abundance and mapped to the taproot length QTL intervals. Missense mutations were also identified in four of 18 genes (Table 8) with significant differences in transcript abundance selected within the root total length and surface area QTL confidence interval on chromosome 6 (Table 7) (*Glyma06g45920*, *Glyma06g44900*, *Glyma06g46170*, and *Glyma06g45910*) (Table 9). There was an insertion/deletion variation (Indel) in the coding sequence of two genes from the *G. soja* parent, *Glyma06g45510* (insertion) and *Glyma06g45261* (deletions), which were in the cell cycle-associated *D6 type cyclin* gene and the key hormone auxin-associated gene, *auxin efflux carrier protein* gene. Five out of nine genes (Table 8) selected from another wild soybean IT182932 [8] had similar conserved missense mutations (Table 10) to those in the wild

**Table 3 List of large-effect QTLs identified for root architectural traits in the V71-370/PI407162 mapping population using composite interval mapping (CIM) analysis**

S. No	Trait	Chr.	Marker interval	LOD value	R <sup>2</sup> value	Additive effect
1	SA	6	4222.1.S1_10 - 77599.1.S1_7	4.5	13.0	-5.23
2	TRL	6	4222.1.S1_10 - 77599.1.S1_7	3.5	11.0	-46.2
3	RDT2	7	59884.1.S1_8- 8398.1.S1_11	5.1	15.0	3.40
4	RDL3	7	8398.1.S1_11 - 1900.1.S1_3	3.6	10.0	1.71

SA, Surface area (cm<sup>2</sup>); TRL, Total root length (cm); RDT2, Root distribution in thickness classification 2 (0.5–1.0 mm); RDL3, Root distribution in length classification 3 (1.0–1.5 mm).

**Table 4 List of genes identified within each potential root QTL interval based on transcript abundance in microarray analyses**

S. no	Affymetrix probe ID	Annotation based on expression profiling*	Gene ID	Gene details #	Root QTLs flanked
1	Gma.4222.1.S1_10	Probable carboxylesterase 6-like	<i>Glyma06g46680</i>	Alpha/beta hydrolase	Surface area, Total length
2	GmaAffx.77599.1.S1_7	Uncharacterized	<i>Glyma06g46850</i>	Histone-like CCAAT Transcription Factor	
3	GmaAffx.59884.1.S1_8	Kinesin-like protein KIF2C-like	<i>Glyma07g09530</i>	Kinesin like protein	Root length (1.0–1.5 mm)
			<i>Glyma09g32280</i>		Root thickness (0.5–1.0 mm)
4	Gma.8398.1.S1_11	Lipase 1-like	<i>Glyma07g09860</i>	Triglyceride lipase-cholesterol esterase	
			<i>Glyma09g31950</i>		
5	Gma.1900.1.S1_3	DEAD-box ATP-dependent RNA helicase 20-like	<i>Glyma07g11880</i>	ATP-dependent RNA helicase	
			<i>Glyma08g20670</i>		
			<i>Glyma07g01260</i>		
6	Gma.6648.1.S1_11	Apoptosis inhibitor 5-like	<i>Glyma07g32480</i>	Apoptosis Inhibitor 5-related	Tap root length
			<i>Glyma13g24090</i>		
7	Gma.5451.1.S1_5	Uncharacterized	<i>Glyma02g15190</i>	Apoptosis-promoting RNA-binding protein	
			<i>Glyma07g33300</i>		
8	Gma.6807.1.S1_10	Cysteine synthase-like	<i>Glyma15g41600</i>	Cystathionine beta-synthase and related enzymes	
9	Gma.9882.1.S1_10	Uncharacterized	<i>Glyma15g42220</i>	slow anion channel associated 1-like	
10	Gma.4792.2.S1_5	Uncharacterized	<i>Glyma04g42630</i>	BTB domain transcription factor	Average diameter
			<i>Glyma06g12140</i>		
11	Gma.15910.1.S1_10	Uncharacterized	<i>Glyma11g26970</i>	Nuclear distribution protein NUDC	
			<i>Glyma18g07050</i>		
	Gma.6807.1.S1_10	Cysteine synthase-like	<i>Glyma15g41600</i>	Cystathionine beta-synthase and related enzymes	
12	GmaAffx.70452.1.S1_3	Calmodulin-binding transcription activator 2-like	<i>Glyma08g19100</i>	CAMTA Transcription factor	Root thickness (0.5–1.0 mm)
			<i>Glyma15g05900</i>		
13	GmaAffx.55124.1.S1_7	Metacaspase-1	<i>Glyma08g19050</i>	Metacaspase involved in regulation of apoptosis	
14	Gma.16443.1.A1_2	Histone-lysine N-methyltransferase ASHH3-like	<i>Glyma09g28430</i>	Uncharacterized	

\*The full expression profiling data of mock control plants can be accessed from the NCBI database. (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE11611>).

# The gene annotation information was from SoyKB.

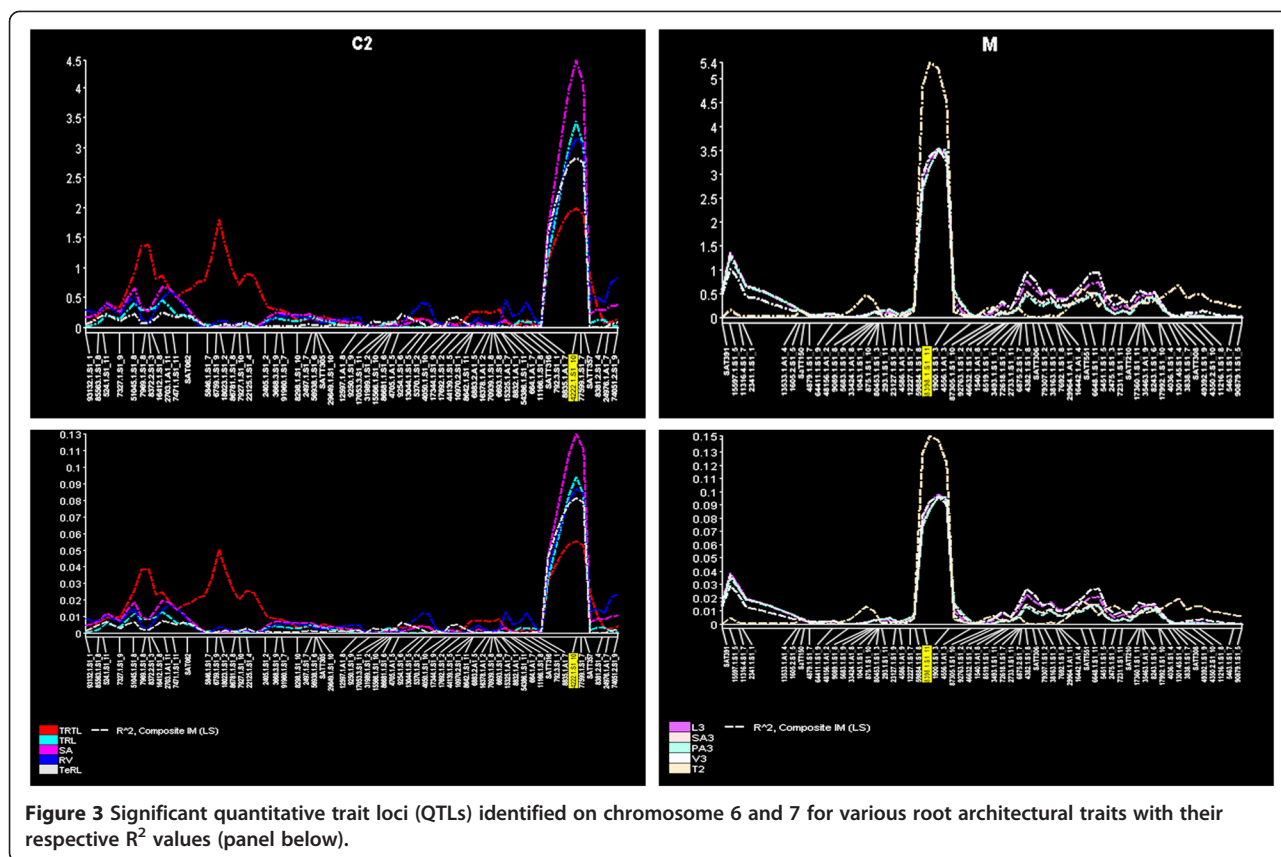
soybean, PI407162, used in this study. *Glyma06g46490* encodes a TPR transcription factor with high expression that is limited to the root pericycle cells (Figure 5).

### Expression patterns of root-related genes

#### (a) Parental genotypes

The differential expression of root-related genes from the *G. soja* accession and cultivated soybean parent enabled us to gain an understanding of the gene regulation associated with various root architecture traits. Based on transcript abundance, 10 genes were identified with significant expression fold changes among the parental lines (Figure 6). Nine genes had a

significantly higher expression in the *G. soja* parental line compared with the *G. max* parent, while gene *Glyma15g42220* showed the opposite trend. When sequence comparisons were made, three genes (*Glyma07g09860*, *Glyma07g32480*, and *Glyma15g42220*) had non-synonymous SNPs (Table 9). Only the gene *Glyma07g32480* had a non-synonymous mutation in both parental lines. The *kinesin motor family protein* (*Glyma09g32280*) had the highest expression in wild soybean, which may have contributed to the smaller and finer roots. Eighteen genes within the QTL region on chromosome 6 were selected based on sequence polymorphisms with *G. soja* accession IT182932



(Tables 7 and 8). Among nine genes listed in Table 7, three *Glyma06g45910*, *Glyma06g45980*, and *Glyma06g44880*, showed higher expression in the *G. soja* parent PI 407162 (Figure 7A); and four genes, *Glyma06g44900*, *Glyma06g45261*, *Glyma06g45810*, and *Glyma06g45920*, showed higher expression in the *G. max* parent V71-370 (Figure 7A). Interestingly, an uncharacterized protein (*Glyma06g45980*) gene had higher expression levels in the *G. soja* parent than in *G. max*. The remaining nine genes in the QTL region, each with a non-synonymous mutation, (Table 8) showed higher expression in *G. max* than in *G. soja*, except for *Glyma06g46210* (Figure 8). Based on 15× depth sequence data from PI 407162, five genes exhibited missense mutations (Table 10). These mutations were conserved when this sequence was compared with other public sequence databases of diverse *G. soja* lines from Korea [8] and China [35].

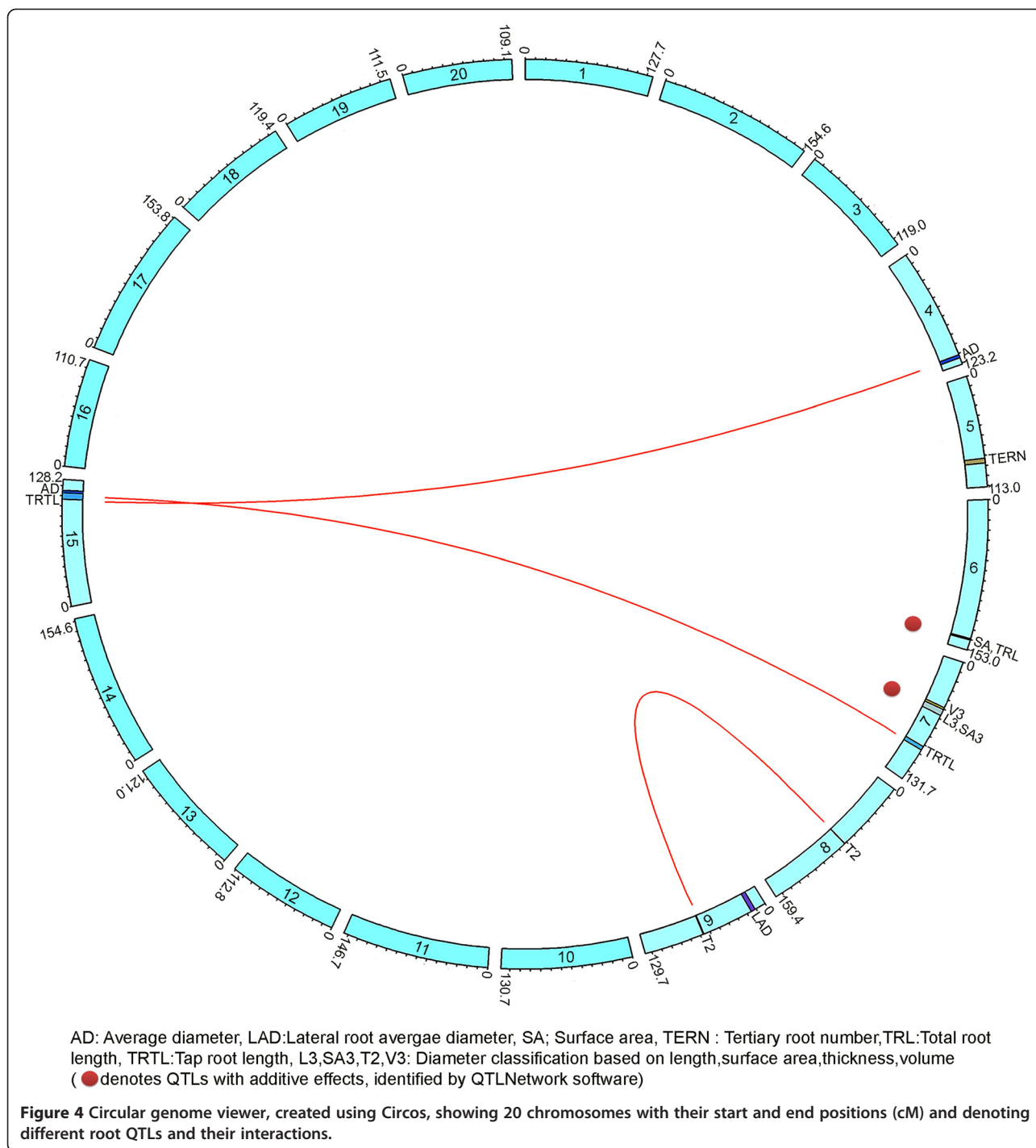
(b) *RILs with extreme root phenotypes*

Eleven genes were selected based on transcript abundance (Table 4) from within the QTL intervals on chromosome 7 for the root diameter distribution based on length and thickness (Table 3) Three of these genes, *Kinesin like proteins*, *triglyceride lipase*, and *ATP-dependent RNA helicase* showed higher expression in RILs with the smallest root phenotypes. These three genes could represent prime candidate genes that play a critical role in regulating fine root development and distribution based on length and thickness. The additive effect taproot length QTL (Table 5) involves interaction of gene(s) on chromosome 7 (*Glyma07g32480*) and chromosome 15 (*Glyma15g42220*), both of which show high levels of expression in RILs with extreme root phenotypes (Figure 6). Non-synonymous mutations were present in both of these genes in V71-370 and PI407162. The

**Table 5** Estimated additive × additive epistatic effect QTLs detected by QTLNetwork for root architectural traits

Trait	Chr.	Marker interval	Chr.	Marker interval	Epistasis	$h^2$ (aa)
TRTL	7	6648.1.S1_11 - 5451.1.S1_5	15	6807.1.S1_10 - 9882.1.S1_10	-1.52	0.08
AD	4	Satt164-4792.1.A1_5	15	15910.1.A1_10 - 6807.1.S1_10	0.02	0.07
RDT2	8	70452.1.S1_3 - 55124.1.S1_7	9	Sat043 - 16443.1.A1_2	2.41	0.06

TRTL, Tap root length (cm); AD, Average diameter (mm); RDT2, Root distribution in thickness classification 2 (0.5–1.0 mm).



genes in the QTL interval region for root thickness of 0.5–1.0 mm (*Glyma08g19050*) and taproot length (*Glyma15g42220*) showed higher expression in RILs with roots smaller than the *G. soja* parent. Among the 18 genes selected in QTLs based on the *G. soja* IT182932 sequence (Tables 7 and 8), only three genes (Table 7) encoding *peroxidase* (two

genes) and *CASP like protein* showed differential gene expression associated with the root phenotype of the parents and RILs (Figure 7B). The missense mutations in these two *peroxidase* genes (Table 9) might contribute to the higher levels of expression in *G. soja* than in *G. max*; however, this needs to be validated with further gene knockout experiments.



**Table 6 Genes selected from wild soybean variety IT182932<sup>a</sup>, based on the QTL confidence interval on chromosome 6, with their annotations**

Gene ID	Annotation
<i>Glyma06g44810</i>	Tetraspanin family protein
<i>Glyma06g44980</i>	Epoxide hydrolase 2-like
<i>Glyma06g44010</i>	8-hydroxyquercetin 8-o-methyltransferase-like isoform 1
<i>Glyma06g44660</i>	Transcription factor bhlh36-like
<i>Glyma06g44990</i>	Epoxide hydrolase 2-like
<i>Glyma06g44600</i>	O-acyltransferase wsd1
<i>Glyma06g44650</i>	Alpha-farnesene synthase
<i>Glyma06g44830</i>	Accelerated cell death 6
<i>Glyma06g44890</i>	Protein
<i>Glyma06g44800</i>	Ribosomal-protein-alanine acetyltransferase-like
<i>Glyma06g44620</i>	ATP synthase mitochondrial f1 complex assembly factor 1-like
<i>Glyma06g44770</i>	MYB Transcription factor
<i>Glyma06g44780</i>	Hypothetical protein PRUPE_ppa014299mg
<i>Glyma06g44970</i>	Gdsl esterase lipase xl3-like
<i>Glyma06g44930</i>	Expansin-b3-like precursor
<i>Glyma06g44730</i>	Probable serine threonine-protein kinase at1g54610-like
<i>Glyma06g44740</i>	PREDICTED: uncharacterized protein LOC100805467
<i>Glyma06g44630</i>	Mitochondrial dihydroorotase
<i>Glyma06g44880</i>	Ankyrin repeat-containing protein at3g12360-like
<i>Glyma06g44790</i>	Caax amino terminal protease family protein
<i>Glyma06g44720</i>	L-type lectin-domain containing receptor kinase -like
<i>Glyma06g44750</i>	Mannan endo-beta-mannosidase 2-like
<i>Glyma06g44640</i>	O-acyltransferase wsd1-like
<i>Glyma06g44900</i>	Ankyrin repeat-containing protein at3g12360-like
<i>Glyma06g45100</i>	Probable protein phosphatase 2c 52-like
<i>Glyma06g45280</i>	Wound-induced protein
<i>Glyma06g45210</i>	Micronuclear linker histone
<i>Glyma06g45590</i>	G-type lectin s-receptor-like serine threonine-protein kinase
<i>Glyma06g45560</i>	Myb-related protein myb4-like
<i>Glyma06g45910</i>	Peroxidase 3
<i>Glyma06g45820</i>	Riboflavin synthase alpha chain
<i>Glyma06g45860</i>	Xyloglucan endotransglucosylase hydrolase protein 9
<i>Glyma06g45740</i>	Probable histone-lysine n-methyltransferase atrx3-like
<i>Glyma06g45300</i>	Unnamed protein product
<i>Glyma06g45420</i>	Wound-induced protein
<i>Glyma06g45230</i>	Tpa: duf566 domain containing family protein
<i>Glyma06g45720</i>	Formin-like protein 5-like
<i>Glyma06g45430</i>	Wound-induced protein
<i>Glyma06g45290</i>	Uncharacterized loc101222779
<i>Glyma06g45920</i>	Peroxidase 3
<i>Glyma06g45800</i>	Proline-rich protein
<i>Glyma06g45550</i>	Myb-related protein myb4-like

**Table 6 Genes selected from wild soybean variety IT182932<sup>a</sup>, based on the QTL confidence interval on chromosome 6, with their annotations (Continued)**

<i>Glyma06g45130</i>	Protein
<i>Glyma06g45390</i>	Wound-induced protein
<i>Glyma06g45350</i>	Wound-responsive protein
<i>Glyma06g45650</i>	F-box family protein
<i>Glyma06g45090</i>	Low quality protein: uncharacterized loc101222318
<i>Glyma06g45770</i>	Btb poz domain-containing protein at3g22104-like
<i>Glyma06g45410</i>	Wound-induced protein
<i>Glyma06g45980</i>	Uncharacterized protein LOC100786184 (Predicted)
<i>Glyma06g45020</i>	Uncharacterized protein LOC100305963
<i>Glyma06g45310</i>	Embryo defective 1923 protein
<i>Glyma06g45000</i>	Probable polyol transporter 4-like
<i>Glyma06g45260</i>	Uncharacterized loc101222779
<i>Glyma06g45150</i>	Protein strubbelig-receptor family 3-like
<i>Glyma06g45670</i>	Hypothetical protein MTR_052s0005
<i>Glyma06g45450</i>	Diacylglycerol kinase 5
<i>Glyma06g45830</i>	Uncharacterized loc101209217
<i>Glyma06g45850</i>	E3 ubiquitin-protein ligase rma1h1-like
<i>Glyma06g45370</i>	Wound-induced protein
<i>Glyma06g45520</i>	Myb-related protein myb4-like
<i>Glyma06g45840</i>	Gpi-anchored protein
<i>Glyma06g45380</i>	Wound-induced protein
<i>Glyma06g45960</i>	Plant cell wall protein 88
<i>Glyma06g45680</i>	Dehydration responsive element binding protein
<i>Glyma06g45620</i>	Zinc finger protein constans-like protein
<i>Glyma06g45640</i>	Indole-3-acetic acid-amido synthetase -like
<i>Glyma06g45810</i>	Casp-like protein rcom_1174750-like
<i>Glyma06g45730</i>	Uncharacterized loc101212188
<i>Glyma06g45220</i>	Uncharacterized protein LOC100527304
<i>Glyma06g45400</i>	Uncharacterized loc101222779
<i>Glyma06g45050</i>	Caffeic acid 3-o-methyltransferase
<i>Glyma06g45010</i>	Drought responsive element binding protein 5
<i>Glyma06g45120</i>	Probable indole-3-acetic acid-amido synthetase -like
<i>Glyma06g45160</i>	Secretory carrier-associated membrane protein 1-like
<i>Glyma06g45700</i>	Beta-amylase
<i>Glyma06g45440</i>	Protein thylakoid chloroplastic-like
<i>Glyma06g45110</i>	Cell wall
<i>Glyma06g45990</i>	Ring-box protein 1a-like
<i>Glyma06g45490</i>	Equilibrative nucleoside transporter
<i>Glyma06g45890</i>	Trab domain-containing
<i>Glyma06g45140</i>	Unknown function
<i>Glyma06g45880</i>	Transferring glycosyl
<i>Glyma06g45930</i>	Translation initiation factor
<i>Glyma06g45170</i>	H aca ribonucleoprotein complex non-core subunit naf1

**Table 6 Genes selected from wild soybean variety IT182932<sup>a</sup>, based on the QTL confidence interval on chromosome 6, with their annotations (Continued)**

<i>Glyma06g45940</i>	Triptychon and
<i>Glyma06g45510</i>	D6-type cyclin
<i>Glyma06g45950</i>	Isocitrate lyase
<i>Glyma06g45480</i>	Uncharacterized gpi-anchored protein at4g28100-like
<i>Glyma06g45780</i>	Isoprene synthase
<i>Glyma06g45360</i>	Vesicle-associated protein 4-2-like
<i>Glyma06g45200</i>	Xylosyltransferase 1-like
<i>Glyma06g45570</i>	Myb-related protein myb4-like
<i>Glyma06g45240</i>	Wound-induced protein
<i>Glyma06g45460</i>	Myb-related transcription factor
<i>Glyma06g45710</i>	Pentatricopeptide repeat-containing protein at1g08070-like
<i>Glyma06g45080</i>	Elmo domain-containing protein a-like
<i>Glyma06g45610</i>	Outer arm dynein light chain 1 protein
<i>Glyma06g45540</i>	Myb-related protein myb4-like
<i>Glyma06g45580</i>	Uncharacterized protein LOC100781575 (Predicted)
<i>Glyma06g45340</i>	Nad h dehydrogenase mitochondrial-like
<i>Glyma06g45180</i>	Protein ultrapetala 1-like
<i>Glyma06g46540</i>	Replication factor c subunit 1-like
<i>Glyma06g46500</i>	Uncharacterized protein LOC100796231 (Predicted)
<i>Glyma06g46020</i>	Duf246 domain-containing protein
<i>Glyma06g46550</i>	Adipocyte plasma membrane-associated
<i>Glyma06g46530</i>	Fasciclin-like arabinogalactan protein 17-like
<i>Glyma06g46350</i>	L-ascorbate oxidase homolog
<i>Glyma06g46370</i>	B-cell receptor-associated 31-like protein
<i>Glyma06g46880</i>	Pentatricopeptide repeat-containing protein at1g11290-like
<i>Glyma06g46360</i>	Unknown function
<i>Glyma06g46450</i>	Cellulose synthase-like protein h1
<i>Glyma06g46960</i>	Uncharacterized protein LOC100814328 (Predicted)
<i>Glyma06g46520</i>	Probable carboxylesterase 15-like
<i>Glyma06g46660</i>	Rj2 protein
<i>Glyma06g46490</i>	Outer envelope protein chloroplastic-like
<i>Glyma06g46650</i>	Protein
<i>Glyma06g46600</i>	Choline ethanolamine kinase
<i>Glyma06g46150</i>	Protein transparent testa 12-like
<i>Glyma06g46590</i>	Myb transcription factor myb142
<i>Glyma06g46110</i>	Upf0481 protein at3g47200-like
<i>Glyma06g46340</i>	Mip sip subfamily
<i>Glyma06g46320</i>	Zinc finger ccch domain-containing protein 13-like
<i>Glyma06g46290</i>	Protein
<i>Glyma06g46740</i>	Auxin-induced protein 5 ng4-like
<i>Glyma06g46710</i>	Sister chromatid cohesion protein dcc1-like
<i>Glyma06g46640</i>	Transcription initiation factor tfiid subunit 7-like

**Table 6 Genes selected from wild soybean variety IT182932<sup>a</sup>, based on the QTL confidence interval on chromosome 6, with their annotations (Continued)**

<i>Glyma06g46750</i>	Cytochrome p450
<i>Glyma06g46190</i>	Aconitate cytoplasmic-like
<i>Glyma06g46610</i>	Ring-h2 finger protein atl69-like
<i>Glyma06g46270</i>	Autophagy-related protein 8c-like
<i>Glyma06g46120</i>	Septum-promoting gtp-binding protein 1-like
<i>Glyma06g46400</i>	S-type anion channel slah1-like
<i>Glyma06g46160</i>	Uncharacterized gpi-anchored protein at1g61900-like
<i>Glyma06g46210</i>	Nedd8-activating enzyme e1 regulatory subunit-like
<i>Glyma06g46260</i>	Upf0481 protein at3g47200-like
<i>Glyma06g46620</i>	Ribosomal 15e family protein
<i>Glyma06g46680</i>	Probable carboxylesterase 6-like
<i>Glyma06g46220</i>	Rrp6-like protein 3
<i>Glyma06g46380</i>	Disease resistance response protein 206-like
<i>Glyma06g46580</i>	Uncharacterized protein LOC100527051
<i>Glyma06g46180</i>	Succinate dehydrogenase subunit 3
<i>Glyma06g46430</i>	Protein usf-like
<i>Glyma06g46630</i>	Protein
<i>Glyma06g46390</i>	Disease resistance response protein 206-like
<i>Glyma06g46760</i>	Cytochrome p450
<i>Glyma06g46300</i>	Zinc metalloprotease slr1821-like
<i>Glyma06g46240</i>	Upf0481 protein at3g47200-like
<i>Glyma06g46410</i>	Mitogen-activated protein kinase kinase kinase a-like
<i>Glyma06g46470</i>	At1g05070 t7a14_6
<i>Glyma06g46090</i>	Upf0481 protein at3g47200-like
<i>Glyma06g46250</i>	Septum-promoting gtp-binding protein 1-like
<i>Glyma06g46170</i>	Uncharacterized protein LOC100779566 (Predicted)
<i>Glyma06g46480</i>	Low quality protein: condensin complex subunit 2-like
<i>Glyma06g46130</i>	Glutamate receptor -like
<i>Glyma06g46570</i>	Proline-rich family protein
<i>Glyma06g46730</i>	Ring-h2 finger protein atl51-like
<i>Glyma06g46420</i>	WRKY transcription factor
<i>Glyma06g46000</i>	Homeobox-leucine zipper protein hdg2-like
<i>Glyma06g46230</i>	Probable beta- galactosyltransferase 2-like
<i>Glyma06g46560</i>	Yabby2-like transcription factor yab2
<i>Glyma06g46690</i>	Signal peptidase complex catalytic subunit sec11c-like

<sup>a</sup>The genes specific to *G. soja* with a deleterious mutation were obtained from Kim et al. [6,8].

One of the genes with a non-synonymous SNP on chromosome 6, *Glyma06g46210*, encodes a NEDD8-activating complex and showed an interesting pattern of expression in the RILs, with higher expression in one of the high (HR2) and low extreme RILs (LR3) (Figure 8). Higher gene expression in both the extreme root phenotypes (low and high) could possibly be

**Table 7 List of genes with transcript abundance in root tissues alone selected on chromosome 6 QTL intervals, based on re-sequence data on *G. soja* variety IT182932<sup>a</sup>**

Gene ID	Gene annotation <sup>a</sup>	Genome status*
<i>Glyma06g44981</i>	Epoxide hydrolase 2-like	N/A
<i>Glyma06g44880</i>	Ankyrin repeat-containing protein	Duplicated
<i>Glyma06g44900</i>	Ankyrin repeat-containing protein	Duplicated
<i>Glyma06g45910</i>	Peroxidase 3	Duplicated
<i>Glyma06g45920</i>	Peroxidase 3	Single copy
<i>Glyma06g45980</i>	Uncharacterized protein	N/A
<i>Glyma06g45261</i>	Uncharacterized protein	N/A
<i>Glyma06g45810</i>	CASP like protein	Single copy
<i>Glyma06g46170</i>	Uncharacterized protein	Duplicated

\*Information on genome duplication and copy number variation was obtained from Du et al. [56].

<sup>a</sup>The genes specific to *G. soja* were obtained from Kim et al. [6,8].

explained by an additive nature of the loci and their interaction (Table 5) to produce a better parental root phenotype. However, other genes with deleterious mutations (Table 8) showed significantly different expression patterns between parental lines. The LR3 line showed higher expression levels for most of the mutated genes identified in this study. Even though these RILs were selected based on allelic composition in this particular QTL confidence interval, they differed substantially for allelic composition within other regions of chromosome 6 (Additional file 4: Figure S3) and at the whole genome level. Thus, the gene expression in LR3, with the smallest root size, may reflect the enrichment of the *G. max*-derived alleles at all these loci.

**Table 8 List of genes with a deleterious mutation selected on chromosome 6 QTL intervals based on re-sequence data on *G. soja* variety IT182932<sup>a</sup>**

Gene ID	Gene annotation <sup>a</sup>	Genome status*
<i>Glyma06g45510</i>	G1/S-specific cyclin D	Duplicated
<i>Glyma06g45610</i>	Leucine rich repeat	Duplicated
<i>Glyma06g45740</i>	Histone H3 (Lys4) methyltransferase complex	Duplicated
<i>Glyma06g45850</i>	RING finger	Single copy
<i>Glyma06g45890</i>	TraB family proteins	Duplicated
<i>Glyma06g46210</i>	NEDD8-activating complex	Duplicated
<i>Glyma06g46490</i>	TPR Transcription factor	Duplicated
<i>Glyma06g46710</i>	Unknown function	Single copy
<i>Glyma06g46730</i>	Zinc finger, C3HC4 type (RING finger)	Duplicated

\*Information on genome duplication and copy number variation was obtained from Du et al. [56].

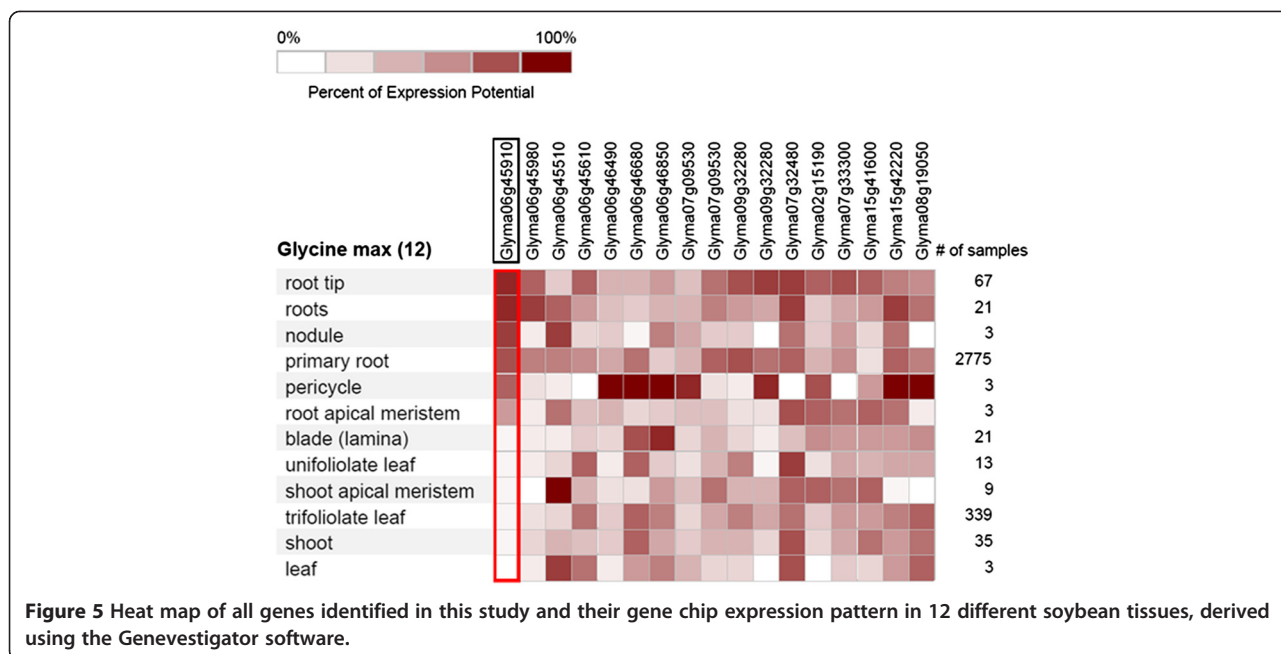
<sup>a</sup>The genes specific to *G. soja* with a deleterious mutation were obtained from Kim et al. [6,8].

## Discussion

### Novel wild soybean alleles for the improvement of root system architecture

Despite the narrow genetic base of cultivated soybeans [4], previous root mapping studies [36,37] successfully used intra-specific mapping populations to map root QTLs in soybean. A recent study [38] used a mapping population developed between the cultivar Jingdou23 and a semi-wild cultivar, ZDD2315, to identify genes and their regulation that control seedling coarse root traits. In the present study, we dissected the genetic variation for both coarse and fine roots using an inter-specific soybean mapping population. Interestingly, the *G. soja* parent (with smaller roots) alleles influenced the root traits of total root length and root volume, explaining a phenotypic variation of more than 10 per cent. Similar contributions of positive alleles for root length and surface area QTLs by a phosphorus inefficient genotype parental line were reported in an earlier study of soybean [39]. Similar useful alleles for a number of agronomic traits were identified from *G. soja* accessions in previous studies [19-23]. The seedling root trait, total root length, and root volume (which is highly correlated with surface area) are highly correlated with drought and aluminum tolerance indices in soybean [40-42]. These traits determine the overall root growth rate and the plasticity of root architecture of plants, and are important for effective soil exploration to intercept nutrients, and for communication of stress cues [43]. A deeper root (influenced by total root length) is vital to improve drought tolerance and is positively correlated with yield under drought stress in both soybean [44] and rice [29,45]. The total root length/surface area plays a crucial role in foraging and accumulation of phosphorus [39], and also improves the limiting effect of phosphorus on shoot growth [46]. The QTL region on chromosome 6 identified in this study also harbors a domestication-related trait for 100-seed weight [19] and seed yield [47]. This QTL region co-locates with another root QTL identified in other studies near Satt357-Satt202. These markers flank the QTL for mean root length [37] and yield [47] that explained phenotypic variances of 26% and 8%, respectively. This region also possesses different aquaporin genes (involved in water transport) that interact with other aquaporin family members on chromosomes 8 and 12 [37].

Alleles from the *G. max* parent, V71-370, also contributed to the development of fine root structure and distribution. Studies focused on fine root structure in legume crops have used pulses [48], but none have been completed in soybean. The fine root distribution, based on length, surface area, and volume, showed high positive correlation with total root volume. The distribution based on thickness was highly correlated with surface



area only. Small diameter roots reflect the proportion of fine lateral roots, which enhance the whole root surface area, acting as an exchange site between the plant and the soil [49]. In this study, several genes associated with different root architectural traits were identified, based

on microarray analyses (Table 4) and whole genome sequencing analysis of two *G soja* lines IT182932 and PI 407162 (Tables 7 and 8). Most of the root traits reported in this study were associated with the expression of one or polygenes. Similar genetic regulation of root traits by

**Table 9 List of genes associated with root traits and with the non-synonymous SNPs in V71-370 and PI 407162 parental lines**

S. No	Gene ID and annotation*	SNP position	W82 (Ref. genome)	V71-370	PI407162	AA change
1	<i>Glyma07g09860</i> <sup>+</sup> Triglyceride lipase-cholesterol esterase	8318354	A	A	C	Phenylalanine to Cysteine
2	<i>Glyma07g32480</i> <sup>+</sup> Apoptosis Inhibitor 5-related	37386211	G	T	G	Proline to Threonine
		37386231	C	T	C	Glycine to Glutamic Acid
		37388170	C	C	G	Glutamic Acid to Glutamine
3	<i>Glyma15g42220</i> <sup>+</sup> Uncharacterized	49621619	C	C	T	Glutamic Acid to Lysine
		49621846	C	C	T	Arginine to Glutamine
4	<i>Glyma06g45920</i> Peroxidase 3	48646118	C	C	A	Arginine to Leucine
5	<i>Glyma06g44900</i> Ankyrin repeat-containing protein	47715303	A	G	A	Glutamic Acid to Glycine
6	<i>Glyma06g46170</i> Uncharacterized	48839708	G	G	C	Glycine to Alanine
		48839816	G	G	A	Serine to Asparagine
		48839842	G	G	A	Glutamic Acid to Lysine
		48840059	G	G	A	Tryptophan (stop gained in wild soybean)
7	<i>Glyma06g45910</i> Peroxidase 3	48635068	A	A	C	Serine to Alanine
8	<i>Glyma06g46210</i> NEDD8-activating complex	48869473	A	A	C	Lysine to Asparagine
		48873032	C	C	G	Arginine to Cysteine
		48873041	C	C	A	Alanine to Glycine
		48874136	A	A	G	Threonine to Asparagine
9	<i>Glyma06g45261</i> Uncharacterized	48049483	A	G	A	Histidine to Arginine

\*Genes selected based on Affymetrix probe hybridization data are denoted with a + sign; the other genes were selected based on QTL confidence intervals.

**Table 10 Conserved non-synonymous mutations in root-related genes among wild soybean varieties**

S. No	Gene ID/Annotation*	SNP position	W82 (Ref. genome)	V71-370	PI407162	AA change	Similar non-Synonymous SNPs in other wild soybeans <sup>†</sup>
1	<i>Glyma06g46210</i> NEDD8-activating complex	48869473	A	A	C	Lysine to Asparagine	W09
		48872310	C	C	T	Arginine to Cysteine	W05, W06, W08, W12, W13, W14, W16
		48873032	C	C	G	Alanine to Glycine	W05, W08, W09, W12, W13, W14, W16, W17
		48873041	C	C	A	Threonine to Asparagine	W05, W08, W09, W12, W13, W14, W16
2	<i>Glyma06g45510</i> G1/S-specific cyclin D	48234941	G	G	A	Aspartic Acid to Asparagine	All lines except W12
		48235772	A	G	A	Methionine to Valine	W01, W10, W12, W17
		48236876	C	C	A	Histidine to Asparagine	W01, W02, W03, W04, W05, W07, W08, W09, W11, W13, W15, W16, W17
3	<i>Glyma06g45610</i> Leucine rich repeat	48346924	C	T	C	Glycine to Aspartic acid	W01, W02, W03, W04, W05, W06, W07, W08, W12, W13, W15, W16, W17
		48353134	G	G	T	Threonine to Asparagine	W05, W09, W12, W11, W16
4	<i>Glyma06g46490</i> TPR Transcription factor	49095540	G	A	G	Aspartic Acid to Asparagine	All except W07, W09 and W10
		49095620	T	A	T	Serine to Arginine	W10
		49095627	A	G	A	Asparagine to Aspartic Acid	W01, W03, W06, W08, W09
		49095824	A	G	A	Threonine to Alanine	All except W10, W13, W17
		49096075	A	G	A	Isoleucine to Methionine	All except W10
5	<i>Glyma06g46730</i> C3HC4 type (RING finger)	49310698	T	T	C	Phenylalanine to Serine	W10, W13, W14, W16
		49310719	C	C	T	Alanine to Valine	W09, W10, W12, W13, W14, W16, W17
		49310842	G	G	A	Arginine to Glutamine	All except W07, W10, W09, W13, W14, W17
		49311027	A	A	G	Isoleucine to Valine	All except W07, W09, W10, W12, W13, W14, W17
		49311204	C	C	G	Glutamine to Glutamic Acid	W12, W13, W14, W16, W17

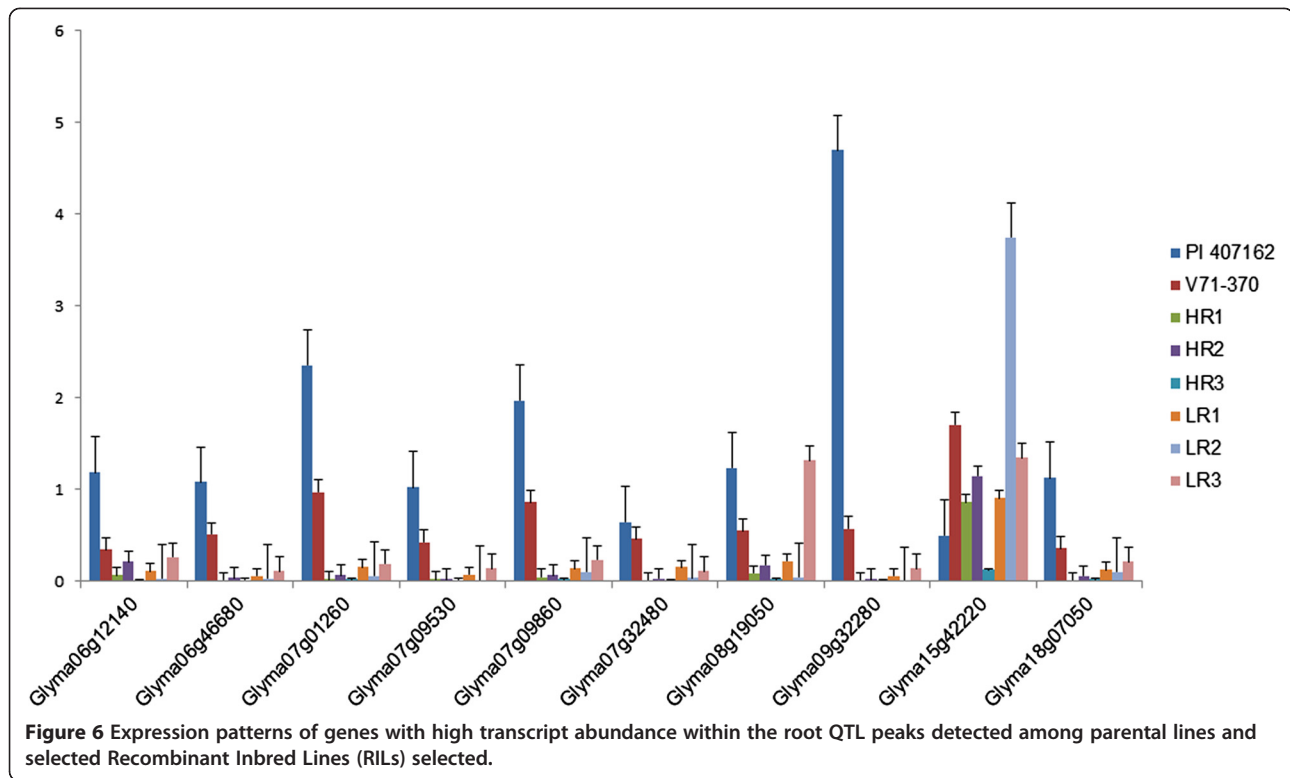
<sup>†</sup>The wild soybean designations were derived from Lam et al. [35].

major gene and polygenes was reported in soybean recently during seedling-stage development [38]. In the present study, most of the root traits showed a transgressive segregation pattern. A similar segregation pattern was reported for maximum root length and lateral root number in soybean [38]. However, tomato introgression lines developed between cultivated and wild tomatoes showed both transgressive and continuous patterns [31]. As the QTL regions identified governs both coarse and fine roots, these might be candidate regions to develop a better root ideotype in soybean.

#### Genes associated with root QTLs

Among the 10 candidate genes (Table 4) identified based on transcript abundance from the microarray analysis, *kinesin* was also found to be highly expressed in maize root tissues [50]. Other genes, such as *carboxylesterase 6 like* and *histone-like CCAAT transcription factor*, were

also found to be highly expressed (Figure 5) in root pericycle cells, which regulate lateral root formation. These genes were also found to be upregulated in water-deficit conditions based on an Affymetrix gene chip study (Babu et al. unpublished) of different root regions. The *carboxylesterase 6 like* gene was highly upregulated under drought stress conditions in root region 2 (4- to 8-mm tip). It was also upregulated in the root region 1 (0–4 mm) following drought stress, but at lower levels compared with region 2. The role of this gene in the maintenance of root growth during drought stress conditions has yet to be studied. Candidate genes identified in the *G. soja* line, such as *Apoptosis inhibitor 5 related* (*Glyma07g32480*), *slow anion channel associated 1-like* (*Glyma15g42220*), and *Metacaspase* (*Glyma08g19050*) (Table 4), showed high expression in root-related tissues (Figure 5) and could be key candidates to improve root length and diameter in cultivated soybean, which might



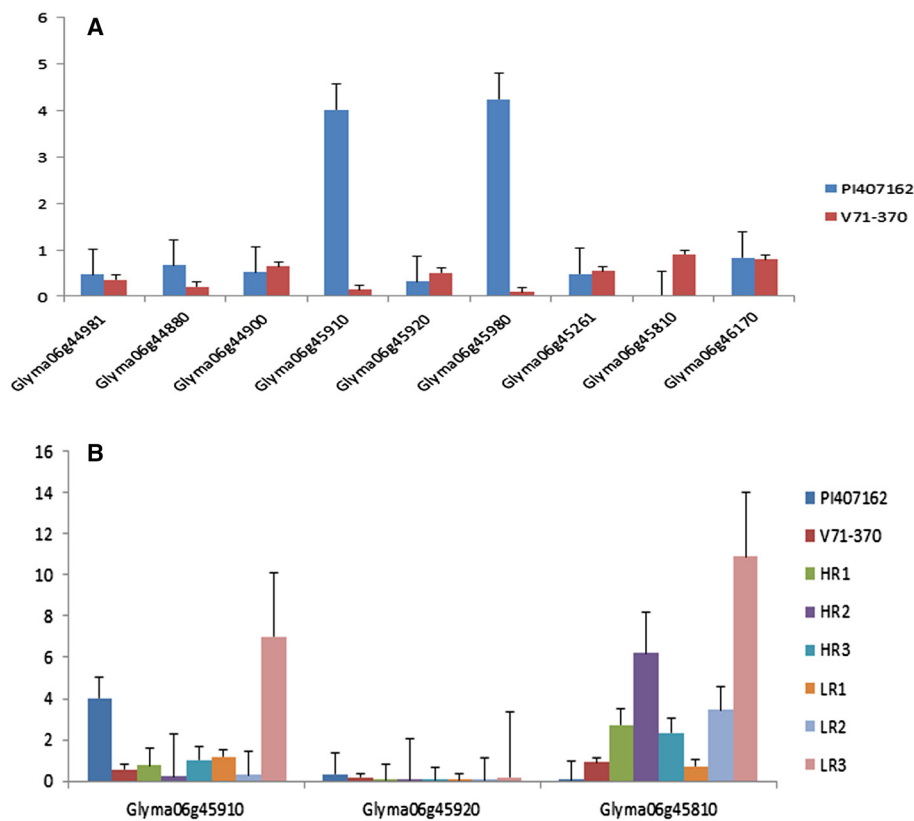
increase nutrient and water use efficiency. Emphasis should be placed on characterizing the function of an uncharacterized gene, *Glyma15g42220*, which is associated with soybean root system architecture, because it showed higher expression in RILs with the shortest root length and surface area compared with RILs with the longest root length. The sequence of this gene is similar to that of *slow anion channel associated 1* in other crops, which is involved in osmoregulation, phosphorus uptake [51], and aluminum tolerance [52]. It is also reported to be involved in regulating the exchange of water vapor and gas during drought stress [53].

The genes with high levels of transcript abundance expressed only in root tissues within the QTL region of chromosomes 6 (Table 7) had distinct expression patterns in the cultivated and *G. soja* parents. Similar trends of gene expression were reported for genes related to stress, defense response, and redox pathways observed in wild soybean [54] and wild tomato [55] compared with their respective cultivated types. Among the 18 genes identified based on sequence information from a wild soybean, two *peroxidase* genes (*Glyma06g45910* and *Glyma06g45920*) in different clades of the heat map showed high expression in the *G. soja* line (Figure 7B). One of these *peroxidase* genes maps to a duplicated region [56] and showed higher expression in the RILs with small or larger root phenotypes. A similar trend of gene expression was observed in tomato introgression lines

developed between cultivated and wild tomatoes [57]. This peroxidase gene in PI407162 also had a non-synonymous SNP variation that changed an amino acid from serine to alanine. Similar effects on root size of amino acid changes were reported in the phosphorylation of the ETHYLENE *INSENSITIVE2* gene in *Arabidopsis* [58]. Zhu et al. [59] reported that the peroxidase activity was spatiotemporal in root tissues and was involved in both cell wall loosening and tightening under water-deficit conditions. Voothuluru and Sharp [60] also reported the effect of *peroxidase* genes on apoplastic reactive oxygen species in the root apical region and its effect on cell wall modification in maize roots as a part of a drought adaptive mechanism. The single copy genes identified in the present study could also be used as informative markers for phylogenetic and comparative analyses across various taxonomic levels [61].

#### Effect of non-synonymous SNP mutations on root architecture

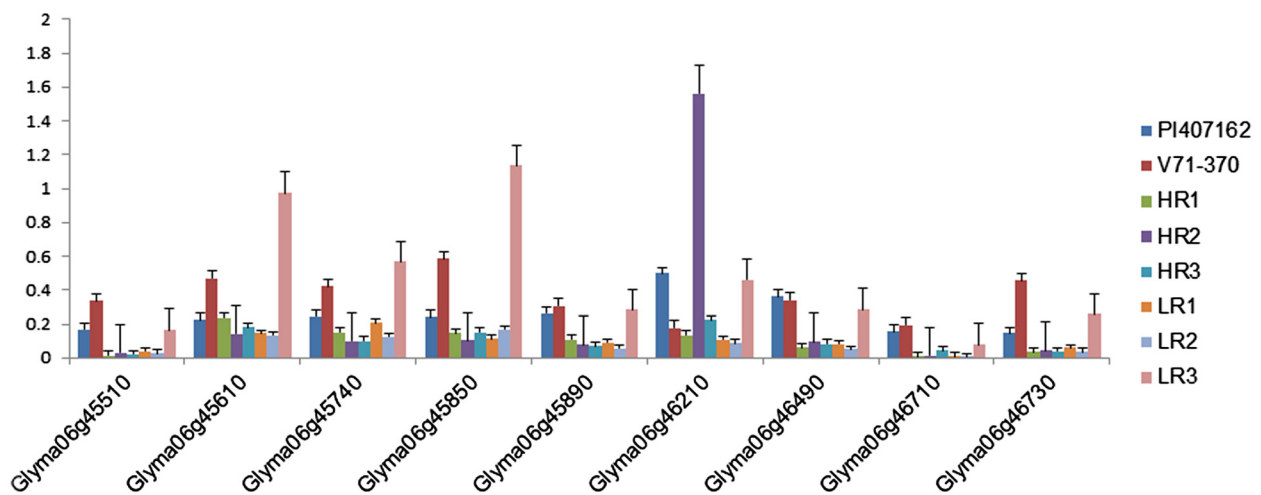
Our results indicated that non-synonymous SNP mutations in genes underlying the QTL region on chromosome 6 are critical candidate genes to study root growth and development. In particular, we identified the gene encoding NEDD8-activating complex, a ThiF protein family (Pfam ID: 00899) that is associated with production of ubiquitin-activating enzyme E1 (Panther ID: 10953), which controls phosphate starvation responses through



**Figure 7** Expression pattern of root specific candidate genes on chromosome 6. (A) Parental lines expression (B) Expression of two Peroxidase and Casp like protein genes.

shoot and root modifications. This gene showed higher expression in wild soybean PI 407162 and the HR2 RIL with higher total root length and surface area. This expression pattern indicated that the transcriptional landscape of *G. soja* lines was highly diverged relative to cultivated

soybean at this particular locus. The homologous gene to *NEDD8-activating complex* also showed higher expression in wild tomatoes than in cultivated tomatoes, as a result of accelerated evolution [55]. The HR2 RIL's higher expression for this gene might have occurred as a result of the



**Figure 8** Expression patterns of genes with a deleterious mutation on chromosome 6 among parental lines and selected RILs.

interaction between the *G. soja*-derived mutated allele (containing five non-synonymous SNPs within the coding region) and the *G. max*-derived allele and para-mutating its effect. Similar mechanisms underlying the gene expression levels were observed in a maize inter-mated mapping population [62]. Thus, this transcript with high expression from wild soybeans is a candidate to increase root length through coordinating cell division and elongation dynamics, as has been reported in wild tomatoes [31]. The TPR transcription factor, with a non-synonymous mutation, was identified among a number of *G. soja* accessions (Table 10) and showed similar expression patterns in both parental lines and RILs. A similar class of transcription factor was reported to be expressed in roots as an early response to iron availability in soybeans [63]. Mutations of key genes with insertions or deletions associated with rate of cell divisions (*D6 type cyclin* and *auxin efflux carrier protein*) might result in the shorter root phenotypes in *G. soja*, PI407162, compared with V71-370. Sequences derived from other *G. soja* accessions [8,35] also showed the same non-synonymous SNP in these same genes. As a result, the altered protein products of these two genes might have resulted in shorter root phenotypes in all wild soybeans. Similar molecules were reported in Arabidopsis and rice to influence cortex-endodermis division regulated by the *SHORT-ROOT* transcription factor [64,65].

#### Comparative genomic analysis among *G. soja* accessions

Based on the whole genome re-sequencing, information from 17 accessions, which represent diverse *G. soja* accessions from Korea [8] and China [35], also showed conserved non-synonymous SNPs in most of the root related genes identified in this study (Table 10). Among different plant species, soybeans are reported to have a higher average ratio of Nonsyn/Syn SNPs than Arabidopsis and rice, with larger effect SNPs in 4,648 genes that have greater effects on their functions [35]. In addition, 21% of potential SNP loci were fixed during the domestication process from wild soybeans. Therefore, the genes identified in the present study might represent candidates with high biological significance for root system growth and development. These genes may play a key role in maintaining short root phenotypes of wild soybeans. A strong candidate for further study is the auxin responsive gene, *NEDD8-activating complex*, because it influences the total root length and involves a key hormone, auxin, which regulates the root growth and development. In Arabidopsis, this gene interacts with a *CULLIN* gene, which then alters the growth of lateral roots and root hairs [66]. Similar results were observed in elite lines of rice in which SNPs altered the expression of *9-cis-epoxycarotenoid dioxygenase* gene associated with the ABA content. A rice line with this missense mutation had produced more lateral roots [67].

For the uncharacterized gene (*Glyma06g46170*), a gain of a stop codon in the *G. soja* PI 407162 makes this a potential candidate gene to further study the functional effects associated with root development in soybeans. Therefore, the root-related genes with non-synonymous mutations identified in this study are a valuable genetic resource to study the evolution of root system development in soybeans. The candidate genes identified from the *G. soja* accession, PI 407162, are promising targets to improve root system architecture in cultivated soybeans. However, RILs with *soja* alleles show undesirable phenotypes, such as lodging and poor yield, which could be eliminated by a series of backcrosses with the cultivated soybean, V71-370.

#### Conclusion

The QTL regions on chromosome 6 for total root length and root volume, and the QTLs on chromosome 7 for finer roots will enable us to integrate improvements in root architecture in soybean. The candidate genes associated with root traits, and with non-synonymous mutations are valuable genetic resources for understanding the evolution of the longer root phenotype in cultivated soybeans compared with the short roots found in most *G. soja* accessions. The genes identified in this study will be important for understanding the molecular mechanisms and gene regulation associated with root development. These findings also suggest that combining novel rare alleles from wild soybean with those of cultivated soybeans could be used to modify/alter the RSA of cultivated soybeans and also develop a soybean that is suited to soils with varying nutrient and water availabilities.

#### Methods

##### Plant materials

A subset of 160 F<sub>12</sub> recombinant inbred lines of a mapping population derived from a soybean inter-specific cross between *G. max* (long and robust root system) × *G. soja* (smaller roots) was selected to map QTLs for root system architectural traits. The population was first developed to map resistance genes for *Phytophthora sojae* in soybean [68,69]. However, in the seedling stage the parental lines also differ for various components of the root system architecture, which enabled us to map QTLs for various root traits in the present study.

##### Plant growth conditions and root tissue sampling

Parental lines and the RILs were grown in a cone system, replicated four times in a completely randomized block design, using DL60L cones and D20 supporting racks (Stuwe and Sons, Oregon, USA). Each replication was conducted separately in the Sears Greenhouse Facility, University of Missouri, USA, from September 2011 to December 2012. Turface (Turface Athletics, Illinois, USA)



and sand was mixed in a 1:1 ratio as a growing medium in cones to offer mechanical impedance similar to field conditions and to facilitate removal of the root system without damage. The day and night temperature in the greenhouse were maintained at 29°C and 21°C, respectively. The photoperiod was set at 12 h using overhead 400 W metal halide lamps that generated a photosynthetic photon flux density of approximately 1620  $\mu\text{mol m}^{-2} \text{s}^{-1}$ . The seedlings were grown up to V1 growth stage (approximately 14 days after sowing) and the intact seedlings from the cones were collected and analyzed [70].

#### Phenotypic data

Root samples were transferred into water-filled clear trays to carefully remove surface particles firmly attached to the root. The roots were then transferred into another water-filled tray, scanned using an Epson Scanner 10000XL (Epson America Inc., CA, USA) and analyzed using WinRhizo software (Regent Instruments Inc., Canada). In addition to manual measurements of taproot length and root fresh weight, data on total root length, surface area, average diameter, root volume, lateral average diameter, tertiary root number, tertiary root length, and root distribution classification based on length, surface area, volume, and thickness were derived from the imaging analysis. The Proc General Linear Model (GLM) and analysis of variance analysis was performed using SAS (v. 9.3).

#### Genotypic data

The total RNA isolation and Affymetrix microarray data analysis and data processing were performed at the Core Laboratory Facility, Virginia Bioinformatics Institute, Virginia Tech, as described in Zhou et al. [32]. The algorithms used to identify informative SFP markers are explained in Additional file 5: File S1, Figures S4–S7. The SFP marker data generated were combined with 109 publicly available simple sequence repeat (SSR) markers that span all linkage groups, and used to construct a genetic linkage map with the Kosambi mapping function in JoinMap 3.0 [71] at the University of Missouri, Columbia. The initial linkage grouping of markers were performed with a likelihood of odds (LOD) threshold score of 3.0 and a maximum genetic distance of 50 cM. The chromosomes were numbered [72] corresponding to the designated soybean genetic linkage groups [73].

#### QTL mapping analysis

A high-density genetic map was created with 1,046 molecular markers that included 937 SFPs and 109 SSRs, with an average marker separation of less than 2 cM. The interval mapping (IM) method was conducted to predict QTLs, followed by composite interval mapping (CIM) using Qgene v4.3.6 [74], with permutations of

1000 iterations to declare the QTL as significant. For CIM, stepwise cofactor selection was used with markers as cofactors and the maximum number of cofactors was selected automatically (F to add = 0.01 and F to drop = 0.01). The permutation LOD value at  $p \leq 0.05$  was used as the threshold to declare the significance of the QTLs. The interactions among QTLs were identified using QTLNetwork-v2.1 software, with a mixed linear model based on CIM with a 10 cM window size and 1 cM walking speed. Permutation tests with 1,000 runs were used to determine the threshold of the F-value for the significance of QTLs to control genome-wide type I errors. Digenic interactions were also analyzed using a mixed-model approach. Significance levels for the genome scans for candidate intervals, QTL detection, and effects were set at 0.05, 0.001, and 0.001, respectively. The epistatic interaction between QTLs was illustrated using the Circos software [75].

#### Candidate gene selection for real-time quantitative reverse-transcription PCR (qRT-PCR) analysis

##### Based on transcript abundance in microarray analysis

For each root QTL identified, all the genes (Table 4) that were located between the flanking markers were examined for their presence on the Affymetrix chip, as well as for changes in transcript abundance. The genes were annotated as described in [32] and are listed in Table 4. The microarray data generated in this mapping population, along with the RILs [32], are available at <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE11611> in the NCBI database. The primers sequences for the genes selected based on transcript abundance are shown in Additional file 6: Table S3.

##### Based on wild soybean sequencing information

To identify the genes specific to the wild soybean at the QTL interval on chromosome 6 flanked by the Affymetrix probe sets 4222.1.S1\_10 and 77599.1.S1\_7, DNA sequencing information of a *G. soja* variety, IT182932 [8] was used. Based on the analysis, 162 genes were identified within the QTL interval region based on the SSR marker position in linkage map (Satt316; 126 cM, 47.5 Mb to Satt357; 143.6 cM, 49.8 Mb). These SSR markers flanked the Affymetrix probe sets. For these genes, BLAST analysis was performed using the BLASTX algorithm ( $E \leq 1e-6$ ) against the non-redundant protein NCBI database and annotated using Blast2GO software [76]. The number of genes were narrowed down to 18 genes (includes nine genes with non-synonymous SNPs) with high transcript abundance in root tissues (Tables 7 and 8), based on public soybean RNA sequencing data available in SoyKB [33]. The primers designed for the genes selected based on transcript abundance in microarrays [32] are shown in

Additional file 7: Table S4, and genes with missense mutations are shown in Additional file 8: Table S5.

#### RNA isolation and qRT-PCR

RNA was extracted from root tissues (100 mg tissues) collected from parental lines (V71-370 and PI407162) and selected RILs using an RNeasy Plant mini kit (Qiagen, CA, USA), according to the manufacturer's protocol. On-Column DNA digestion was performed using RNase-Free DNase Set (Qiagen), according to the manufacturer's protocol. Each sample (2 µg of total RNA) was reverse transcribed to cDNA in a 20-µL reaction volume using RNA to cDNA Ecopry™ Premix (Double primed) cDNA Synthesis Kit (Clontech, CA, USA). The qRT-PCR was performed using the cDNA product corresponding to 25 ng of total RNA in a 10-µL reaction volume and Maxima SYBR Green/ROX qPCR Master Mix (2×) (Thermo, USA) on a ABI7900HT detection system (Applied BioSystems, Foster City, CA, USA). The expression data for each sample were generated from three biological and two technical replicates. The relative expression of the selected genes were expressed as the mean standard deviation, in comparison with the transcript abundance of actin, a housekeeping gene, and analyzed using the Delta Ct method [77]. The PCR conditions were as follow: 50°C for 2 min, 95°C for 10 min, then 40 cycles of 95°C for 15 s, 60°C for 1 min. To normalize the gene expression, Actin (*Glyma18g52780*) was used as an internal control. All primers were designed using the Primer3 web-interface (<http://frodo.wi.mit.edu/primer3/input.htm>) [78]. Gene expression was evaluated among parental lines and selected RILs of extreme phenotypes (Additional file 9: Table S2). The chromosome graphical representation of the selected RILs was made with graphical genotype software GGT (v 2.0) [79].

#### DNA isolation, genome sequencing, and functional SNP identification

The DNA of the parental lines (V71-370 and PI407162) was isolated and sequenced at a depth of 15× using Illumina 90 bp paired-end sequencing technology with insert sizes of around 500 bp. The data were processed after filtering out low-quality reads and duplicate reads. The processed data were aligned to the William 82 *G. max* v1.1 from Phytozome as the reference genome [2]. SNPs and Indels were identified using an in-house built pipeline using GATK v3.0 [80] and were analyzed for possible synonymous/non-synonymous SNP variation annotations using SnpEFF [81] and v9.0 gene models from Phytozome. To detect small insertions and deletions, Indels (1-5 bp) were called by SOAP (Short Oligonucleotide Analysis Package) indel 1.09 (<http://soap.genomics.org.cn/soapindel.html>). The non-synonymous SNP variations were only considered for comparison

among 17 Chinese and 1 Korean wild soybeans, and can be viewed using SNPviz tool [82] available in SoyKB.

#### Availability of supporting data

All the supporting data are included as additional files.

#### Additional files

**Additional file 1: Figure S1.** Histogram of the frequency distribution of root traits among RI lines of the mapping population (V71-370/PI407162).

**Additional file 2: Table S1.** List of other root trait QTLs co-located on chromosome 6 and 7.

**Additional file 3: Figure S2.** Pie chart representation of the ontological classification of genes underlying the candidate QTL region on chromosome 6.

**Additional file 4: Figure S3.** Genomic compositions of selected extreme RILs on chromosome 6 (three High Roots (HR) lines denoted as 1–3, and three Low Roots (LR) lines as 4–6) for gene expression study using qRT-PCR.

**Additional file 5: File S1.** The SFPdev Min-Max Ratio algorithm. Figures S4-S7: Algorithms used for polymorphic single feature polymorphism (SFP) detection in Additional file 5: File S1.

**Additional file 6: Table S3.** The primer sequences for genes with high transcript abundance within the QTLs identified.

**Additional file 7: Table S4.** The primer sequence for genes selected based on wild soybean variety IT 182932 within the QTL interval on chromosome 6.

**Additional file 8: Table S5.** The primer sequences for genes with a deleterious mutation based on wild soybean variety IT 182932 within the QTL region on chromosome 6.

**Additional file 9: Table S2.** Extreme root phenotypic RI lines selected for qRT-PCR gene expression analysis.

#### Competing interests

The authors declare that they have no competing interests.

#### Authors' contributions

SJP and SL designed the experiment; DQ, JVMS, TJ, DX, and BV performed the DNA sequence analysis and performed the bioinformatics analysis; CC and GP performed the qRT-PCR, gene mining and annotation; TDV performed the QTL analysis; SJP and MM conducted the experiment, data collection, and interpretation; KK and DMT performed the SFP marker analysis and data interpretation; RB, AED, and MASM developed the mapping population and edited the intellectual content of the manuscript; GS and HTN conceived the study and edited the manuscript. All authors read and approved the final manuscript.

#### Acknowledgments

Funding support from the Missouri Soybean Merchandising Council (Project#220) and NSF Plant Genome Research Program (Award DBI-0211863) is greatly appreciated.

#### Author details

<sup>1</sup>National Center for Soybean Biotechnology and Division of Plant Sciences, University of Missouri, Columbia, MO 65211, USA. <sup>2</sup>Christopher S. Bond Life Sciences Center, University of Missouri, Columbia, MO 65211, USA. <sup>3</sup>Department of Computer Science, University of Missouri, Columbia, MO 65211, USA. <sup>4</sup>Department of Crop and Soil Environmental Sciences, Virginia Tech, Blacksburg, VA 24061, USA. <sup>5</sup>Department of Plant Pathology, The Ohio State University, OARDC, Wooster, OH 44691, USA.

Received: 15 August 2014 Accepted: 9 February 2015

Published online: 25 February 2015

## References

- Carter Jr TE, Nelson RL, Sneller C, Cui Z. Genetic diversity in soybean. In: Boerma HR, Specht JE, editors. Soybean: improvement, production, and uses, 3rd edition. Madison: American Society for Agronomy; 2004. p. 303–416.
- Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, et al. Genome sequence of the paleopolyploid soybean. *Nature*. 2010;463:178–83.
- Broich SL, Palmer RG. A cluster analysis of wild and domesticated soybean phenotypes. *Euphytica*. 1980;29:23–32.
- Hyten DL, Song QJ, Zhu Y, Choi IY, Nelson RL, Costa JM, et al. Impacts of genetic bottlenecks on soybean genome diversity. *Proc Natl Acad Sci*. 2006;103:16666–71.
- Lee J, Yu J, Hwang Y, Blake S, So Y, Lee G, et al. Genetic diversity of wild soybean accessions from South Korea and other countries. *Crop Sci*. 2008;48:606–16.
- Kim KS, Diers BW, Hyten DL, Roufman MA, Shannon JG, Nelson RL. Identification of positive yield QTL alleles from exotic soybean germplasm in two backcross populations. *Theor Appl Genet*. 2012;125:1353–69.
- Joshi T, Valliyodan B, Wu J, Lee S, Xu D, Nguyen HT. Genomic differences between cultivated soybean, *G. max* and its wild relative *G. soja*. *BMC Genomics*. 2013;14:55.
- Kim MY, Lee S, Van K, Kim TH, Jeong SC, Choi I, et al. Whole-genome sequencing and intensive analysis of the undomesticated soybean genome. *Proc Natl Acad Sci*. 2010;107:22032–7.
- Tanksley SD, McCouch SR. Seed banks and molecular maps: unlocking genetic potential from the wild. *Science*. 1997;277:1063–6.
- Xiao J, Li J, Grandillo S, Ahn SN, Yuan L, Tanksley SD, et al. Identification of trait-improving quantitative trait loci alleles from a wild rice relative, *Oryza rufipogon*. *Genetics*. 1998;150:899–909.
- Gupta PK, Rustgi S, Mir RR. Array-based high-throughput DNA markers and genotyping platforms for cereal genetics and genomics. In: Gupta PK, Varshney RK, editors. *Cereal genomics II*. Netherlands: Springer; 2013. p. 11–55.
- Steinmetz LM, Sinha H, Richards DR, Spiegelman JJ, Oefner PJ, McCusker JH, et al. Dissecting the architecture of a quantitative trait locus in yeast. *Nature*. 2002;416:326–30.
- Childs LH, Witucka-Wall H, Günther T, Sulpice R, Korff MV, Stitt M, et al. Single feature polymorphism (SFP)-based selective sweep identification and association mapping of growth-related metabolic traits in *Arabidopsis thaliana*. *BMC Genomics*. 2010;11:188.
- Bernardo AN, Bradbury PJ, Ma H, Hu S, Bowden RL, Buckler ES, et al. Discovery and mapping of single feature polymorphisms in wheat using Affymetrix arrays. *BMC Genomics*. 2009;10:251.
- Saxena RK, Cui X, Thakur V, Walter B, Close TJ, Varshney RK. Single feature polymorphisms (SFPs) for drought tolerance in pigeon pea. *Funct Integr Genomics*. 2011;11:651–7.
- Hajjar R, Hodgkin T. The use of wild relatives in crop improvement: a survey of developments over the last 20 years. *Euphytica*. 2007;156:1–13.
- Si-bin G, Yu W, Xiao-qiong L, Kai-qiang L, Feng-kuan H, Cai-hong C. Development and Identification of Introgression Lines from Cross of *Oryza sativa* and *Oryza minuta*. *Rice Sci*. 2013;20:95–102.
- Placido DF, Campbell MT, Folsom JJ, Cui X, Kruger GR, Baenziger PS, et al. Introgression of novel traits from a wild wheat relative improves drought adaptation in wheat. *Plant Physiol*. 2013;161:1806–19.
- Liu B, Fujita T, Yan Z, Sakamoto S, Xu D, Abe J. QTL mapping of domestication-related traits in soybean. *Ann Bot*. 2007;100:1027–38.
- Tuyen DD, Lal SK, Xu DH. Identification of a major QTL allele from wild soybean for increasing alkaline salt tolerance in soybean. *Theor Appl Genet*. 2010;121:229–36.
- Chen Y, Chen P, Reyes BGD. Differential responses of the cultivated and wild species of soybean to dehydration stress. *Crop Sci*. 2006;46:2041–6.
- Li D, Pfeiffer TW, Cornelius PL. Soybean QTL for yield and yield components associated with *Glycine soja* alleles. *Crop Sci*. 2008;48:571–81.
- Kanamaru K, Wang SD, Abe J, Yamada T, Kitamura K. Identification and characterization of wild soybean (*G. soja*) strains with high lutein content. *Breed Sci*. 2006;56:231–4.
- Heatherly LG. U.S. soybean production. St. Louis, Missouri: United Soybean Board; 2009.
- Serraj R, McNally KL, Slamet-Loedin I, Kholi A, Haefele SM, Atlin G, et al. Drought resistance improvement in rice: an integrated genetic and resource management strategy. *Plant Prod Sci*. 2011;14:1–14.
- Zhao CX, Deng XP, Shan L, Steudle E, Zhang SQ, Ye Q. Changes in root hydraulic conductivity during wheat evolution. *J Integr Plant Biol*. 2005;47:302–10.
- Dorlodot S, Foster B, Pages L, Price A, Tuberosa R, Draye X. Root system architecture: opportunities and constraints for genetic improvement of crops. *Trends Plant Sci*. 2007;12:474–81.
- Prince SJ, Mutava RN, Pegoraro C, Oliveira ACD, Nguyen HT. Root characters. In: Kole C, editor. *Genomics and breeding for climate resilient crops*. Berlin: Springer; 2013. p. 67–131.
- Suji KK, Prince SJ, Mankhar SP, Kanagaraj K, Poornima R, Amutha K, et al. Evaluation of rice near isogenic lines with root QTLs for plant production and root traits in rainfed target populations of environment. *Field Crop Res*. 2012;137:89–96.
- Bouteille M, Rolland G, Balsera C, Loudet O, Muller B. Disentangling the intertwined genetic bases of root and shoot growth in *Arabidopsis*. *Plos One*. 2012;7:e32319.
- Ron M, Dorrity MW, de Lucas M, Toal T, Hernandez RI, Little SA, et al. Identification of novel loci regulating interspecific variation in root morphology and cellular development in tomato. *Plant Physiol*. 2013;162:755–68.
- Zhou L, Mideros SX, Bao L, Hanlon R, Arredondo FD, Tripathy S, et al. Infection and genotype remodel the entire soybean transcriptome. *BMC Genomics*. 2009;10:49.
- Joshi T, Fitzpatrick MR, Chen S, Liu Y, Zhang H, Endacott RZ, et al. Soybean Knowledge Base (SoyKB): a web resource for integration of soybean translational genomics and molecular breeding. *Nucl Acids Res*. 2014;42:D1245–52.
- Hruz T, Laule O, Szabo G, Wessendorf F, Bleuler S, Oertle L, et al. Genevestigator V3: a reference expression database for the meta-analysis of transcriptomes. *Adv Bioinformatics*. 2008;420747.
- Lam HM, Xu X, Liu X, Chen W, Yang G, Wong F, et al. Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nat Genet*. 2010;42:1053–9.
- Abdel-Haleem H, Lee G, Boerma RH. Identification of QTL for increased fibrous roots in soybean. *Theor Appl Genet*. 2011;122:935–46.
- Brensha W, Kantartzi SK, Meksem K, Grier RL, Bara-kat A, Lightfoot DA, et al. Genetic analysis of root and shoot traits in the ‘essex’ by ‘forrest’ recombinant inbred line population of soybean. *J Plant Genome Sci*. 2012;1:1–9.
- Liang H, Yu Y, Yang H, Xu L, Dong W, Du H, et al. Inheritance and QTL mapping of related root traits in soybean at the seedling stage. *Theor Appl Genet*. 2014;127:2127–37.
- Liang Q, Cheng X, Mei M, Yan X, Liao H. QTL analysis of root traits as related to phosphorus efficiency in soybean. *Ann Bot*. 2010;106:223–34.
- Liu Y, Gai JY, Lv HN. Identification of rhizosphere abiotic stress tolerance and related root traits in soybean. *Acta Agron Sin*. 2005;31(9):1132–7.
- Yang SP, Chen JM, He XH, Yu DY, Gai JY. Inheritance of drought tolerance and root traits of seedling in soybean. *Soybean Sci*. 2005;24(4):275–80.
- Liu Y, Gai JY, Lv HN. Genetic variation of root traits at seedling stage and their relationship with stress tolerance in soybean. *Soybean Sci*. 2007;26(2):127–33.
- Falik O, Mordoch Y, Ben-Natan D, Vanunu M, Goldstein O, Novoplansky A. Plant responsiveness to root-root communication of stress cues. *Ann Bot*. 2012;271:271–80.
- Hudak CM, Patterson RP. Vegetative growth analysis of a drought-resistant soybean plant introduction. *Crop Sci*. 1995;35:464–71.
- Bengough AG, McKenzie BM, Hallett PD, Valentine TA. Root elongation, water stress, and mechanical impedance: a review of limiting stresses and beneficial root tip traits. *J Exp Bot*. 2011;62:59–68.
- Bates TR, Lynch JP. Plant growth and phosphorus accumulation of wild type and two root hair mutants of *Arabidopsis thaliana*. *Am J Bot*. 2000;87:958–63.
- Reinprecht Y, Poysa V, Yu K, Rajcan I, Ablett G, Pauls K. Seed and agronomic QTL in low linolenic acid, lipoxigenase-free soybean germplasm. *Genome*. 2006;49:1510–27.
- Liu L, Gan Y, Bueckert R, Rees KV, Warkentin T. Fine root distributions in oilseed and pulse crops. *Crop Sci*. 2010;50:222–6.
- Eissenstat DM, Yanai RD. Root life span, efficiency, and turnover. In: Waisel Y, Eshel A, Kafkafi U, editors. *Plant roots, the hidden half*. New York: Marcel Dekker; 2002. p. 221–38.
- Sekhon RS, Lin H, Childs KL, Hansey CN, Buell CR, Leon ND, et al. Genome-wide atlas of transcription during maize development. *Plant J*. 2011;66:553–63.
- Roberts SK. Plasma membrane anion channels in higher plants and their putative functions in roots. *New Phytologist*. 2006;169:647–66.
- Kochian LV, Hoekenga OA, Piñeros MA. How do crop plants tolerate acid soils? Mechanisms of aluminum tolerance and phosphorus efficiency. *Annu Rev Plant Biol*. 2004;55:459–93.

53. Chen YH, Hu L, Punta M, Bruni R, Hillerich B, Kloss B, et al. Homologue structure of the *SLAC1* anion channel for closing stomata in leaves. *Nature*. 2010;467:1074–80.
54. Ge Y, Li Y, Zhu Y, Bai X, Lv D, Guo D, et al. Global transcriptome profiling of wild soybean roots under NaHCO<sub>3</sub> treatment. *BMC Plant Biol*. 2010;10:153.
55. Koenig D, Jiménez-Gómez JM, Kimura S, Fulop D, Chitwood DH, Headland LR, et al. Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. *Proc Natl Acad Sci*. 2013;110:E2655–62.
56. Du J, Tian Z, Sui Y, Zhao M, Song Q, Cannon SB, et al. Pericentromeric effects shape the patterns of divergence, retention, and expression of duplicated genes in the paleopolyploid soybean. *Plant Cell*. 2012;24:21–32.
57. Sun W, Xu W, Zhu H, Liu A, Liu L, Sun W. Comparative transcriptomic profiling of a salt tolerant wild tomato species and a salt sensitive tomato cultivar. *Plant Cell Physiol*. 2010;51:997–1006.
58. Ju C, Yoon GM, Shemansky JM, Lin DY, Ying ZI, Chang J, et al. CTR1 phosphorylates the central regulator EIN2 to control ethylene hormone signaling from the ER membrane to the nucleus in *Arabidopsis*. *Proc Natl Acad Sci*. 2012;109:19486–91.
59. Zhu J, Alvarez S, Marsh EL, LeNoble ME, Cho JJ, Sivaguru M, et al. Cell wall proteome in the maize primary root elongation zone. II. Region-specific changes in water soluble and lightly ionically bound proteins under water deficit. *Plant Physiol*. 2007;145(4):1533–48.
60. Voothuluru P, Sharp RE. Apoplastic hydrogen peroxide in the growth zone of the maize primary root under water stress. I. Increased levels are specific to the apical region of growth maintenance. *J Exp Bot*. 2013;64:1223–33.
61. Duarte JM, Wall PK, Edger PP, Landherr LL, Ma H, Pires JC, et al. Identification of shared single copy nuclear genes in *Arabidopsis*, *Populus*, *Vitis* and *Oryza* and their phylogenetic utility across various taxonomic levels. *BMC Evol Biol*. 2010;10:61.
62. Li L, Petsch K, Shimizu R, Liu S, Xu WW, Ying K, et al. Mendelian and non-mendelian regulation of gene expression in maize. *PLoS Genet*. 2013;9:e1003202.
63. Lauter ANM, Peiffer GA, Yin T, Whitham SA, Cook D, Shoemaker RC, et al. Identification of candidate genes involved in early iron deficiency chlorosis signaling in soybean roots and leaves. *BMC Genomics*. 2014;15(1):702.
64. Cui H, Levesque MP, Vernoux T, Jung JW, Paquette AJ, Gallagher KL, et al. An evolutionarily conserved mechanism delimiting SHR movement defines a single layer of endodermis in plants. *Science*. 2007;316:421–5.
65. Sozzani R, Cui H, Moreno-Risueno MA, Busch W, Van Norman JM, Vernoux T, et al. Spatiotemporal regulation of cell-cycle genes by *SHORTROOT* links patterning and growth. *Nature*. 2010;466:128–32.
66. Pozo JCD, Dharmasiri S, Hellmann H, Walker L, Gray WM, Estelle M. *AXR1-ECR1*-dependent conjugation of *RUB1* to the *Arabidopsis* *cullin AtCUL1* is required for auxin response. *Plant Cell*. 2002;14:421–33.
67. Lyu J, Zhang S, Dong Y, Weiming H, Jing Z, Deng X, et al. Analysis of elite variety tag SNPs reveals an important allele in upland rice. *Nat Commun*. 2013;4:2138–46.
68. Cicek MS, Chen P, Maroof MAS, Buss GR. Interrelationships among agronomic and seed quality traits in an interspecific soybean recombinant inbred population. *Crop Sci*. 2006;46:1253–9.
69. Tucker DM, Saghai Maroof MA, Mideros S, Skoneczka JA, Nabati DA, Buss GR, et al. Mapping quantitative trait loci for partial resistance to *Phytophthora sojae* in a soybean interspecific cross. *Crop Sci*. 2010;50:628–35.
70. Manavalan LP, Guttikonda SK, Nguyen VT, Shannon JG, Nguyen HT. Evaluation of diverse soybean germplasm for root growth and architecture. *Plant and Soil*. 2010;330:503–14.
71. Van Ooijen JH, Voorrips RE. JoinMap 3.0 software for the calculation of genetic linkage maps. Wageningen: Plant Research International; 2001.
72. Grant D, Nelson RT, Cannon SB, Shoemaker RC. SoyBase, the USDA-ARS soybean genetics and genomics database. *Nucleic Acids Res*. 2010;38:D843–846.
73. Song QJ, Marek LF, Shoemaker RC, Lark KG, Concibido VC, Delannay X, et al. A new integrated genetic linkage map of the soybean. *Theor Appl Genet*. 2004;109:122–8.
74. Joehanes R, Nelson JC. Q Gene 4.0 an extensible java QTL-analysis platform. *Bioinformatics*. 2008;24:2788–9.
75. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. *Genome Res*. 2009;19:1639–45.
76. Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, et al. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res*. 2008;36:3420–35.
77. Livak K, Schmittgen T. Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta C_t}$  method. *Methods*. 2001;25:402–8.
78. Rozen S, Skaletsky HJ. Primer3 on the WWW for general users and for biologist programmers. In: Misener S, Krawetz SA, editors. *Bioinformatics methods and protocols: methods in molecular biology*. Totowa (NJ): Humana Press Inc; 2000. p. 365–86.
79. Berloo V. GGT 2.0 versatile software for visualization and analysis of genetic data. *J Heredity*. 2008;99:232–6.
80. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20:1297–303.
81. Pablo C, Adrian P, Le LW, Melissa C, Tung N, Luan W, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Landes Biosci*. 2012;6:1–13.
82. Langewisch T, Zhang H, Vincent R, Joshi T, Xu D, Bilyeu K. Major soybean maturity gene haplotypes revealed by SNPviz analysis of 72 sequenced soybean genomes. *Plos One*. 2014;9(4):e94150.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

