Emerging Transcriptomic Databases and Their Use in Gene Expression Profiling

Carlos M. Hernandez-Garcia
Department of Horticulture and Crop Science, OARDC/The Ohio State University, 1680 Madison Ave., Wooster, OH 44691, USA
hernandez-garcia.1@buckeyemail.osu.edu

Main purposes of this tutorial

- ☐ Provide an updated list of plant gene-expression databases and related resources
- ☐ Provide step-by-step instructions to generate gene expression profiles
- ☐ Review considerations relevant to the use of gene expression databases
- ☐ Use web-based tools for visualization of transcriptomic data

Background

- Expression databases hosting microarray -derived data have been fundamental to study gene expression in many plants; however, this technology is biased toward known RNAs used to generate the probes in the chips.

- With the advent of next-generation sequencing (RNA-Seq), global RNA analysis (transcriptome) is becoming routine for many plant species.

- RNA-Seq is a powerful tool not only to validate gene annotation but also to unravel quantitative gene expression for all sets of genes transcribed in a sample.

- The vast amount of information generated using RNA-Seq technology allows the generation of databases that capture a wider snapshoot of the transcriptome, including absolute numbers of transcripts for most of the genes in the genome.

Biological rationale for RNA-Seq

Next-generation sequencing technologies such as Solexa, Illumina, and 454 can be applied to transcriptome sequencing. These technologies detect short reads of RNA present in biological samples, including coding and non-coding RNA. These reads are short, but long enough to be aligned uniquely to genes lying on a reference genome. Thus, reads can be assigned to their respective gene.

Further information on next-generation sequencing for plant breeders can be accessed at http://www.extension.org/article/32489

Specific objective:

Demonstrate how transcript profiles can be generated for two sets of soybean candidate genes. Two different transcriptomic databases from soybean will be used for more consistency (Soybase and Transcriptome Atlas).

Rationale:

We are interested in soybean seed-specific promoters. As gene expression is largely regulated by promoters, one approach is to first identify genes playing a major role in determining seed composition (e.g. oil content, fatty acid, etc.).

Goal:

Examine how putative seed specific genes are transcribed in various tissues and seeds at different developmental stages using databases of RNA-Seq experiments.

Current plant transcriptomic databases, including microarray-derived data

Arabidopsis Transcriptome Genomic Express Database (RNA-Seq data)
http://signal.salk.edu/cgi-bin/atta

RiceGE Japonica: Rice Functional Genomic Express Database (RNA-Seq data)
http://signal.salk.edu/cgi-bin/RiceGE

RiceGE: Rice (*indica*) Functional Genomic Express Database (RNA-Seq data)
http://signal.salk.edu/cgi-bin/RiceiGE

PopGenIE: The Populus Genome Integrative Explorer (cDNA array)
http://www.popgenie.org/

Medicago truncatula Gene Expression Atlas (Affimetrix data)
http://mtgea.noble.org/v2/

Maize C3/C4 Transcriptomic Database (RNA-Seq data)
http://c3c4.tc.cornell.edu/search.aspx

Tomato Expression Database (cDNA array and Affimetrix) http://ted.bti.cornell.edu/

Generating expression profiles for two sets of soybean genes

For this tutorial, two sets of soybean genes will be used as examples of how to build expression profiles using transcript databases. The first set was identified in the soybean genome by Dr. Robert Bouchard* using the N-terminal amino acid sequences for reported proteins found in seeds (Vodkin and Raikhel 1986; Kalinski et al. 1989; Natarajan et al. 2007). We named this group *SEED* genes. The second group of genes, identified by Dr. Leah McHale*, contains candidate genes mapping to known fatty acid regions. These genes are therefore predicted to be involved in fatty acid biosynthesis and were termed for these tutorial *FAB* (*Fatty Acid Biosynthesis*) genes.

Promoters from both set of genes are being validated by Dr. John Finer* with the aim of isolating soybean seed specific promoters. Transcript profiles for these genes may predict tissue-specific expression driven by their promoters. More information about validation of promoters from these and other sets of soybean genes is available at http://www.oardc.ohio-state.edu/SURE/

*The Ohio State University/OARDC

First set of genes – *SEEDs*

Gene	Gene ID	Gene Family
SEED1	Glyma19g34780.1	Proglycinin
SEED2	Glyma03g32030.1	Proglycinin
SEED3	Glyma08g12270.1	P34
SEED5	Glyma13g18450.1	Glycinogin B
SEED6	Glyma10g04280.1	Glycinogin B
SEED7	Glyma01g10900.1	Kunitz Trypsin Inhibitor KTI1
*SEED10**	Glyma02g01590.1	Lectin
SEED11	Glyma20g28650.1	β-Conglycinin A
SEED12	Glyma10g39160.1	β-Conglycinin A

**SEED10* is the previously identified *Lectin 1* gene

Second set of genes – *FAB* (*<u>F</u>atty <u>A</u>cid <u>B</u>iosynthesis*)

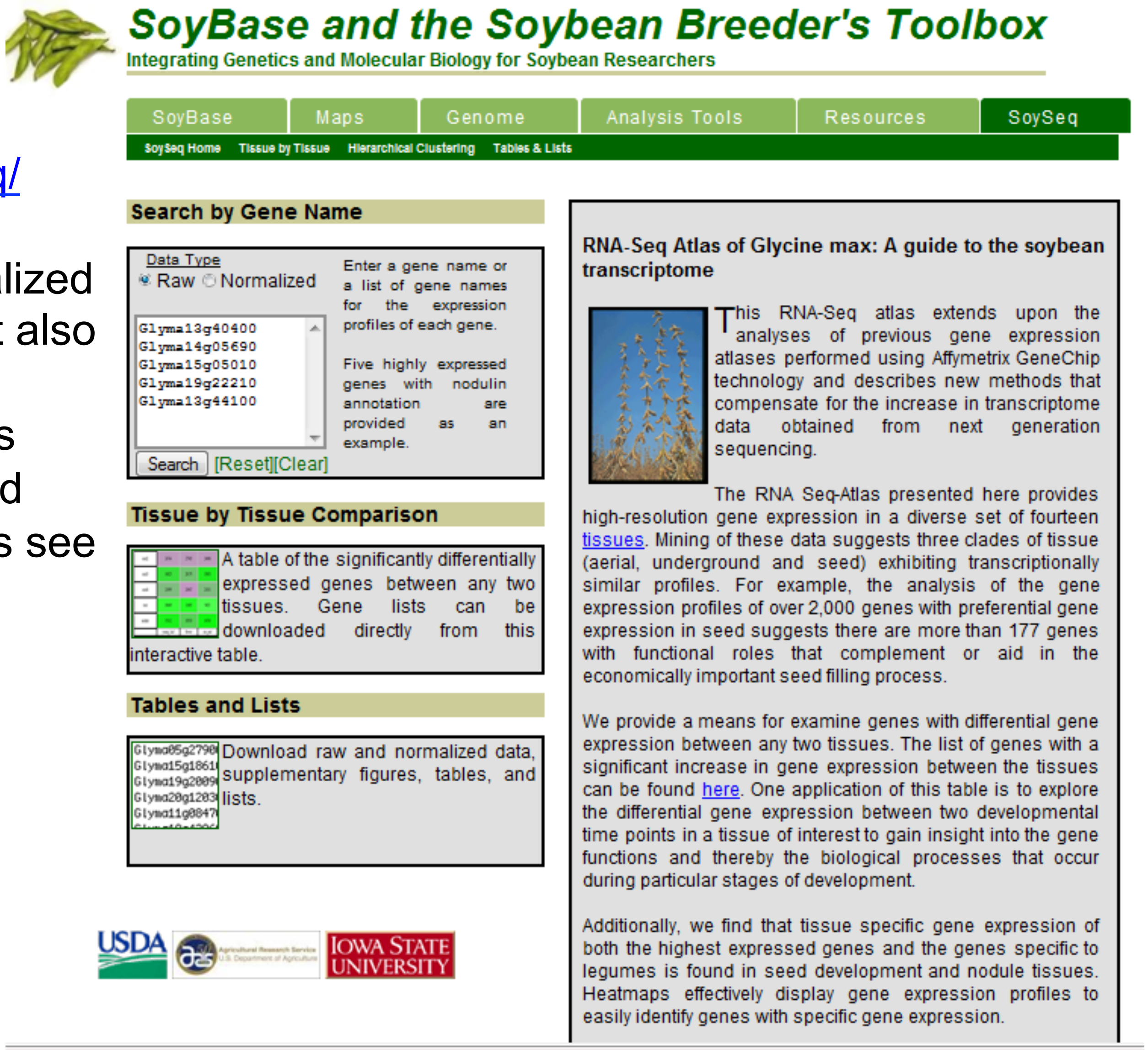
Gene*	Gene ID	QTL Name**
FAB1	Glyma14g22840	Ole1-5
FAB2	Glyma14g27990	Fas_Stearic2-2
FAB3	Glyma14g38180	Fan
FAB4	Glyma05g03100	Palm2-1
FAB5	Glyma05g36450	Ole1-1; Linole1-1; Linole1-2; Linole1-3
FAB6	Glyma09g15600	Linolen1-6
FAB7	Glyma14g09100	Palm1-2

*Number of *FAB* genes are solely for the purpose of this tutorial
**QTL data are from Soybase database

RNA-Seq Expression Databases - Soybase

http://soybase.org/soyseq/

Soybase contains normalized and raw transcript data. It also allows tissue by tissue comparison and facilitates construction of figures and tables. For specific details see Severin et al. 2010



SoyBase and the Soybean Breeder's Toolbox

Integrating Genetics and Molecular Biology for Soybean Researchers

Search by Gene Name

Data Type

● Raw ○ Normalized

Glyma13g40400
Glyma14g05690
Glyma15g05010
Glyma19g22210
Glyma13g44100

Enter a gene name or a list of gene names for the expression profiles of each gene.

Five highly expressed genes with nodulin annotation are provided as an example.

Tissue by Tissue Comparison

A table of the significantly differentially expressed genes between any two tissues. Gene lists can be downloaded directly from this interactive table.

Tables and Lists

Download raw and normalized data, supplementary figures, tables, and lists.

RNA-Seq Atlas of Glycine max: A guide to the soybean transcriptome

This RNA-Seq atlas extends upon the analyses of previous gene expression atlases performed using Affymetrix GeneChip technology and describes new methods that compensate for the increase in transcriptome data obtained from next generation sequencing.

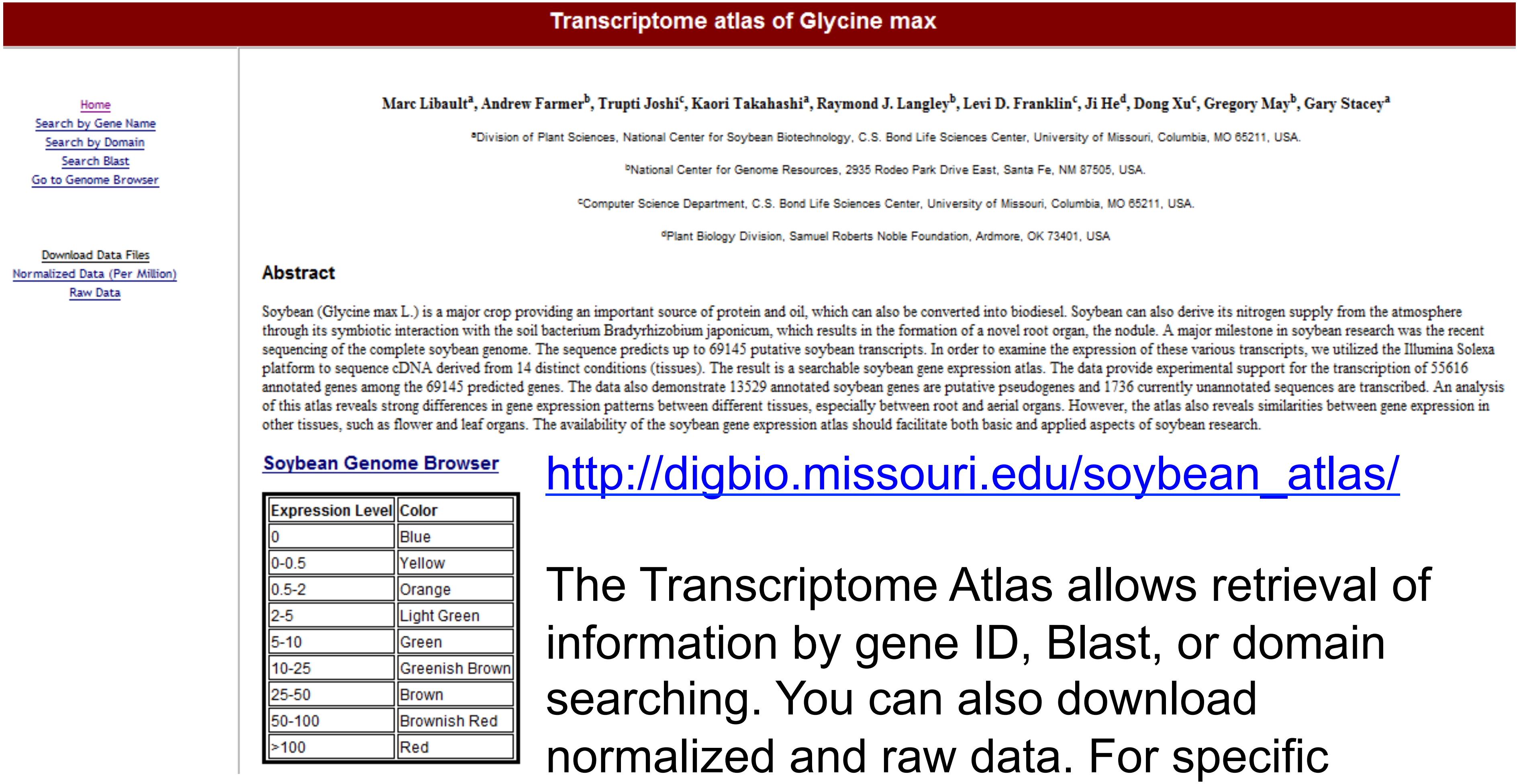
The RNA Seq-Atlas presented here provides high-resolution gene expression in a diverse set of fourteen tissues. Mining of these data suggests three clades of tissue (aerial, underground and seed) exhibiting transcriptionally similar profiles. For example, the analysis of the gene expression profiles of over 2,000 genes with preferential gene expression in seed suggests there are more than 177 genes with functional roles that complement or aid in the economically important seed filling process.

We provide a means for examine genes with differential gene expression between any two tissues. The list of genes with a significant increase in gene expression between the tissues can be found here. One application of this table is to explore the differential gene expression between two developmental time points in a tissue of interest to gain insight into the gene functions and thereby the biological processes that occur during particular stages of development.

Additionally, we find that tissue specific gene expression of both the highest expressed genes and the genes specific to legumes is found in seed development and nodule tissues. Heatmaps effectively display gene expression profiles to easily identify genes with specific gene expression.

Snapshoot from the SoyBase and the Soybean Breeder's Toolbox

RNA-Seq Expression Databases – Transcriptome Atlas



Transcriptome atlas of Glycine max

Home
Search by Gene Name
Search by Domain
Search Blast
Go to Genome Browser

Download Data Files
Normalized Data (Per Million)
Raw Data

Marc Libault[a], Andrew Farmer[b], Trupti Joshi[c], Kaori Takahashi[a], Raymond J. Langley[b], Levi D. Franklin[c], Ji He[d], Dong Xu[c], Gregory May[b], Gary Stacey[a]

[a]Division of Plant Sciences, National Center for Soybean Biotechnology, C.S. Bond Life Sciences Center, University of Missouri, Columbia, MO 65211, USA.

[b]National Center for Genome Resources, 2935 Rodeo Park Drive East, Santa Fe, NM 87505, USA.

[c]Computer Science Department, C.S. Bond Life Sciences Center, University of Missouri, Columbia, MO 65211, USA.

[d]Plant Biology Division, Samuel Roberts Noble Foundation, Ardmore, OK 73401, USA

Abstract

Soybean (Glycine max L.) is a major crop providing an important source of protein and oil, which can also be converted into biodiesel. Soybean can also derive its nitrogen supply from the atmosphere through its symbiotic interaction with the soil bacterium Bradyrhizobium japonicum, which results in the formation of a novel root organ, the nodule. A major milestone in soybean research was the recent sequencing of the complete soybean genome. The sequence predicts up to 69145 putative soybean transcripts. In order to examine the expression of these various transcripts, we utilized the Illumina Solexa platform to sequence cDNA derived from 14 distinct conditions (tissues). The result is a searchable soybean gene expression atlas. The data provide experimental support for the transcription of 55616 annotated genes among the 69145 predicted genes. The data also demonstrate 13529 annotated soybean genes are putative pseudogenes and 1736 currently unannotated sequences are transcribed. An analysis of this atlas reveals strong differences in gene expression patterns between different tissues, especially between root and aerial organs. However, the atlas also reveals similarities between gene expression in other tissues, such as flower and leaf organs. The availability of the soybean gene expression atlas should facilitate both basic and applied aspects of soybean research.

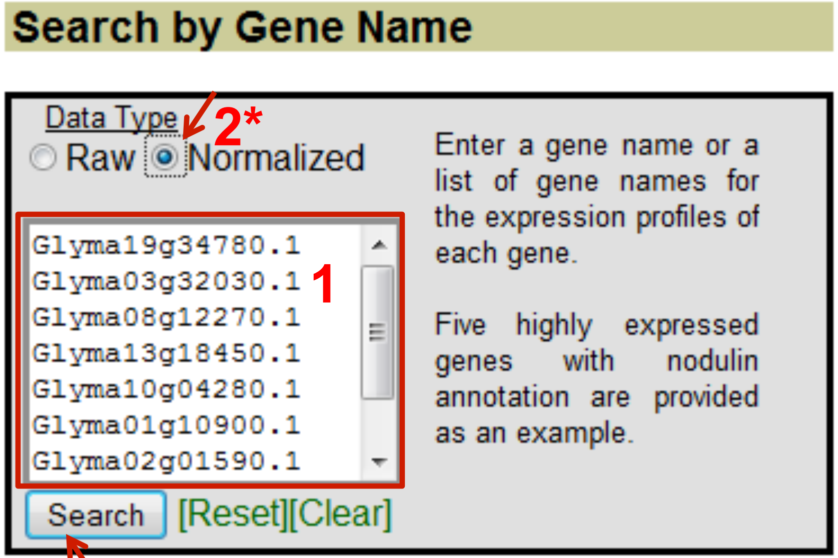
Soybean Genome Browser

Expression Level	Color
0	Blue
0-0.5	Yellow
0.5-2	Orange
2-5	Light Green
5-10	Green
10-25	Greenish Brown
25-50	Brown
50-100	Brownish Red
>100	Red

http://digbio.missouri.edu/soybean_atlas/

The Transcriptome Atlas allows retrieval of information by gene ID, Blast, or domain searching. You can also download normalized and raw data. For specific details see Libault et al. 2010

Downloading transcript data – Soybase



Search by Gene Name

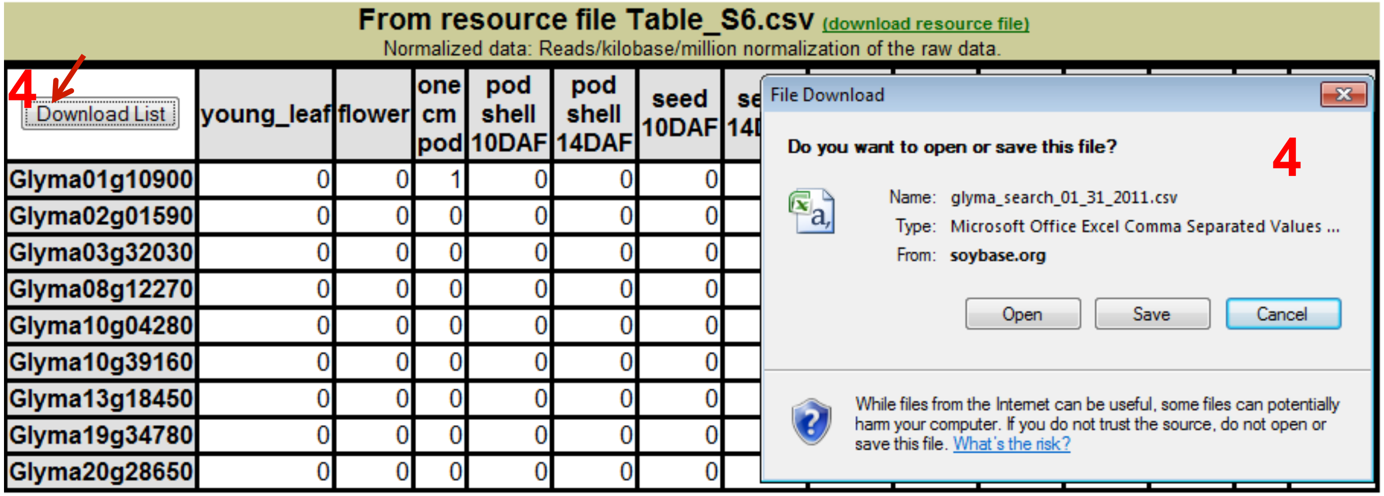
```
Glyma19g34780.1
Glyma03g32030.1
Glyma08g12270.1
Glyma13g18450.1
Glyma10g04280.1
Glyma01g10900.1
Glyma02g01590.1
```

Enter a gene name or a list of gene names for the expression profiles of each gene.

Five highly expressed genes with nodulin annotation are provided as an example.

http://soybase.org/soyseq/

1. Enter the list of gen IDs
2. Check "normalized"
3. Click "search"
4. Download list as comma delimited file in Excel



From resource file Table_S6.csv (download resource file)

Normalized data: Reads/kilobase/million normalization of the raw data.

Download List	young_leaf	flower	one cm pod	pod shell 10DAF	pod shell 14DAF	seed 10DAF	se 14D
Glyma01g10900	0	0	1	0	0	0	
Glyma02g01590	0	0	0	0	0	0	
Glyma03g32030	0	0	0	0	0	0	
Glyma08g12270	0	0	0	0	0	0	
Glyma10g04280	0	0	0	0	0	0	
Glyma10g39160	0	0	0	0	0	0	
Glyma13g18450	0	0	0	0	0	0	
Glyma19g34780	0	0	0	0	0	0	
Glyma20g28650	0	0	0	0	0	0	

File Download

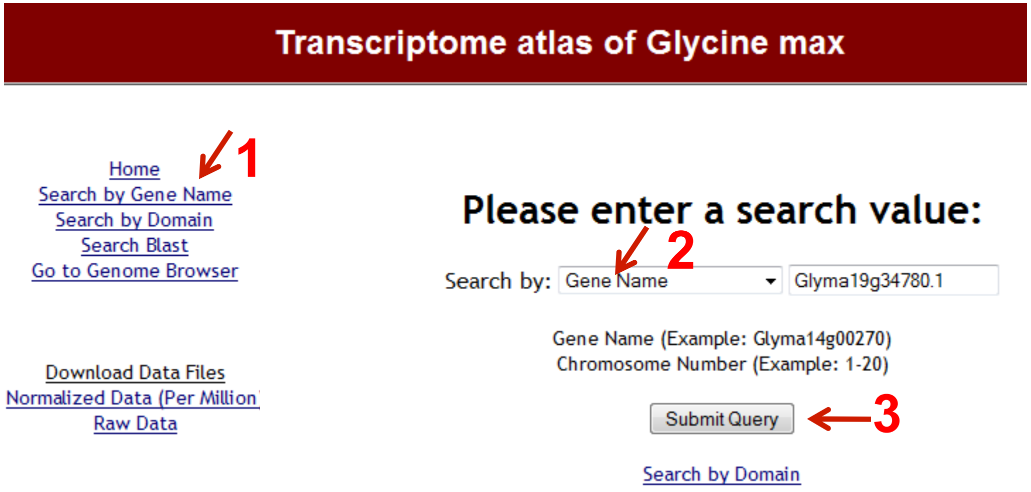
Do you want to open or save this file?

Name: glyma_search_01_31_2011.csv
Type: Microsoft Office Excel Comma Separated Values ...
From: **soybase.org**

While files from the Internet can be useful, some files can potentially harm your computer. If you do not trust the source, do not open or save this file. What's the risk?

*The RPKM (reads/Kb/Million) method for normalization corrects for biases in total gene exon size and normalizes for the total read sequences obtained in each library. Thus, normalized data is comparable between genes and samples.

Downloading transcript data – Transcriptome Atlas



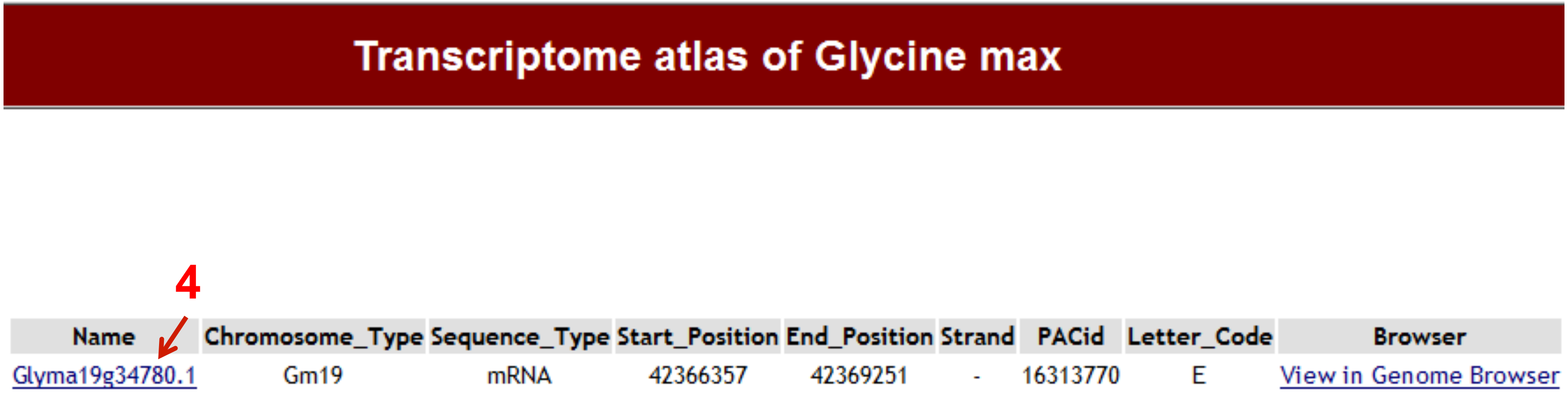
Please enter a search value:

Search by: Gene Name Glyma19g34780.1

Gene Name (Example: Glyma14g00270)
Chromosome Number (Example: 1-20)

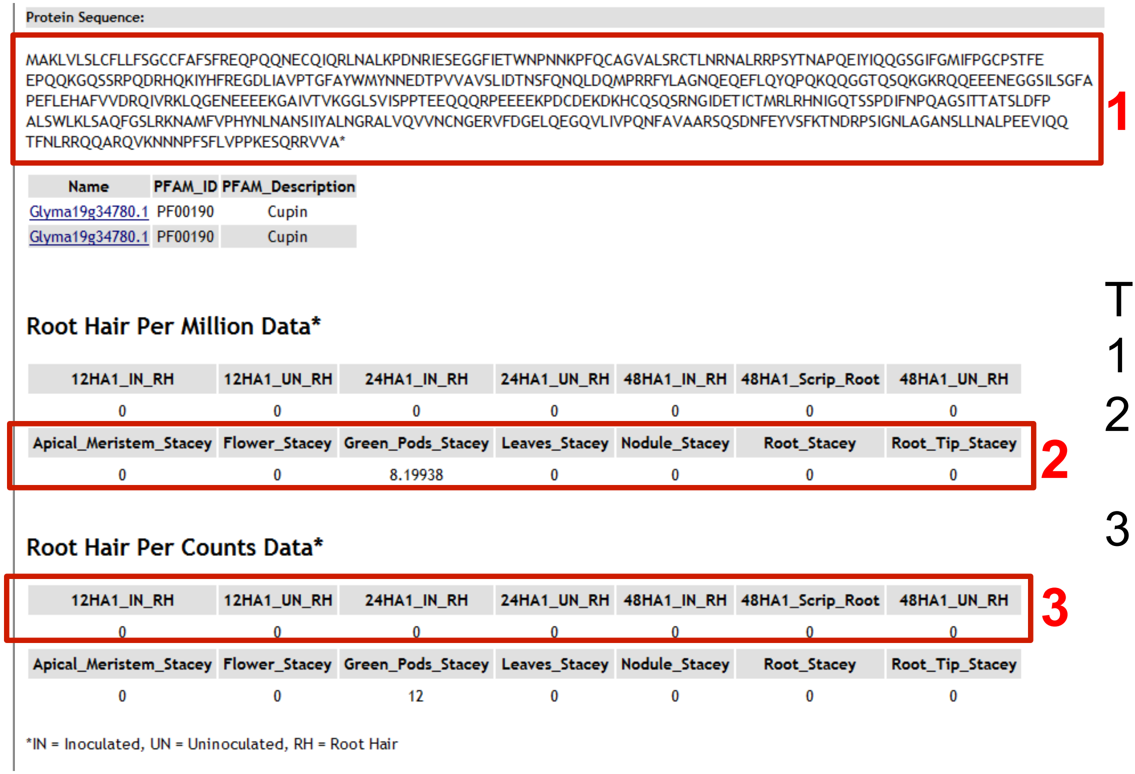
http://digbio.missouri.edu/soybean_atlas/

1. Select “Search by Gene Name”
2. Enter each gen ID individually
3. Click on “Submit Query”
4. Then, click on the Glyma link



Name	Chromosome_Type	Sequence_Type	Start_Position	End_Position	Strand	PACid	Letter_Code	Browser
Glyma19g34780.1	Gm19	mRNA	42366357	42369251	-	16313770	E	View in Genome Browser

Downloading transcript data – Transcriptome Atlas



Protein Sequence:

```
MAKLVLSLCFLLFSGCCFAFSFREQPQQNECQIQRLNALKPDNRIESEGGFIETWNPNNKPFQCAGVALSRCTLNRNALRRPSYTNAPQEIYIQQGSGIFGMIFPGCPSTFE
EPQQKGQSSRPQDRHQKIYHFREGDLIAVPTGFAYWMYNNEDTPVVAVSLIDTNSFQNQLDQMPRRFYLAGNQEQEFLQYQPQKQQGGTQSQKGKRQQEEENEGGSILSGFA
PEFLEHAFVVDRQIVRKLQGENEEEEKGAIVTVKGGLSVISPPTEEQQQRPEEEEKPDCDEKDKHCQSQSRNGIDETICTMRLRHNIGQTSSPDIFNPQAGSITTATSLDFP
ALSWLKLSAQFGSLRKNAMFVPHYNLNANSIIYALNGRALVQVVNCNGERVFDGELQEGQVLIVPQNFAVAARSQSDNFEYVSFKTNDRPSIGNLAGANSLLNALPEEVIQQ
TFNLRRQQARQVKNNNPFSFLVPPKESQRRVVA*
```

Name	PFAM_ID	PFAM_Description
Glyma19g34780.1	PF00190	Cupin
Glyma19g34780.1	PF00190	Cupin

Root Hair Per Million Data*

12HA1_IN_RH	12HA1_UN_RH	24HA1_IN_RH	24HA1_UN_RH	48HA1_IN_RH	48HA1_Scrip_Root	48HA1_UN_RH
0	0	0	0	0	0	0

Apical_Meristem_Stacey	Flower_Stacey	Green_Pods_Stacey	Leaves_Stacey	Nodule_Stacey	Root_Stacey	Root_Tip_Stacey
0	0	8.19938	0	0	0	0

Root Hair Per Counts Data*

12HA1_IN_RH	12HA1_UN_RH	24HA1_IN_RH	24HA1_UN_RH	48HA1_IN_RH	48HA1_Scrip_Root	48HA1_UN_RH
0	0	0	0	0	0	0

Apical_Meristem_Stacey	Flower_Stacey	Green_Pods_Stacey	Leaves_Stacey	Nodule_Stacey	Root_Stacey	Root_Tip_Stacey
0	0	12	0	0	0	0

*IN = Inoculated, UN = Uninoculated, RH = Root Hair

The output file also contains:

1. Predicted protein sequence
2. Transcript number for different organ/tissues
3. Transcript number for other conditions

*Data are normalized as transcripts per million

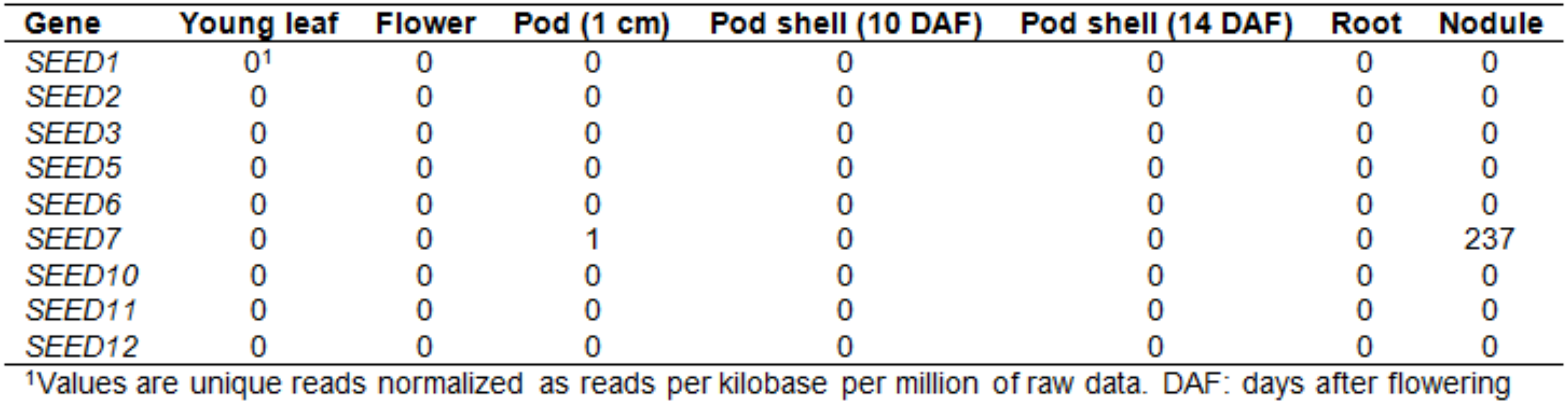
If you experience problems when retrieving your data using Internet Explorer you may try Mozilla Firefox!

Data Analysis

To determine seed-specificity for each set of genes, transcript data from the two transcriptomic databases were grouped into two categories: developing seeds and other tissues.

Results -Tissue-specific expression of *SEED* genes

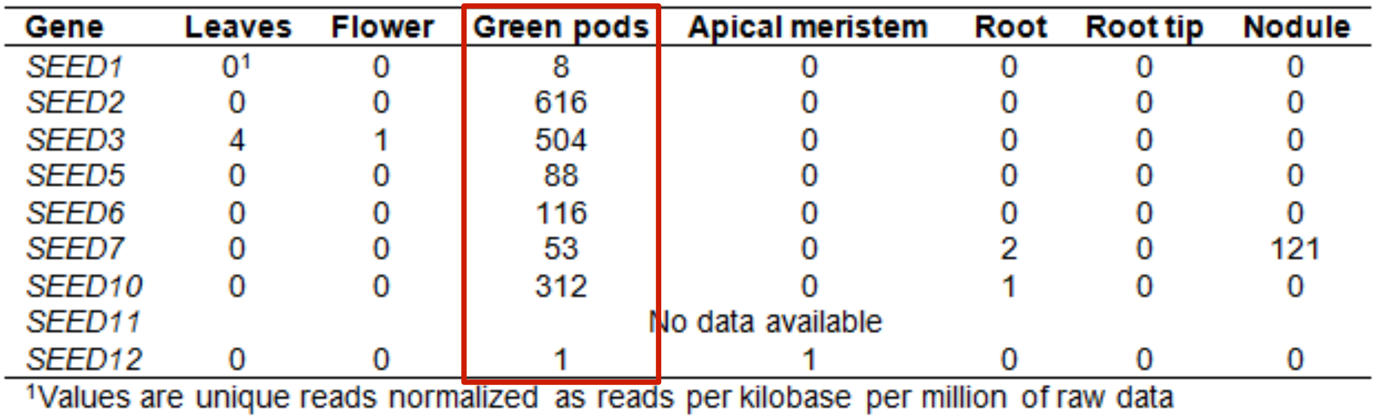
A Soybase



Gene	Young leaf	Flower	Pod (1 cm)	Pod shell (10 DAF)	Pod shell (14 DAF)	Root	Nodule
SEED1	0[1]	0	0	0	0	0	0
SEED2	0	0	0	0	0	0	0
SEED3	0	0	0	0	0	0	0
SEED5	0	0	0	0	0	0	0
SEED6	0	0	0	0	0	0	0
SEED7	0	0	1	0	0	0	237
SEED10	0	0	0	0	0	0	0
SEED11	0	0	0	0	0	0	0
SEED12	0	0	0	0	0	0	0

[1] Values are unique reads normalized as reads per kilobase per million of raw data. DAF: days after flowering

B Transcriptome Atlas



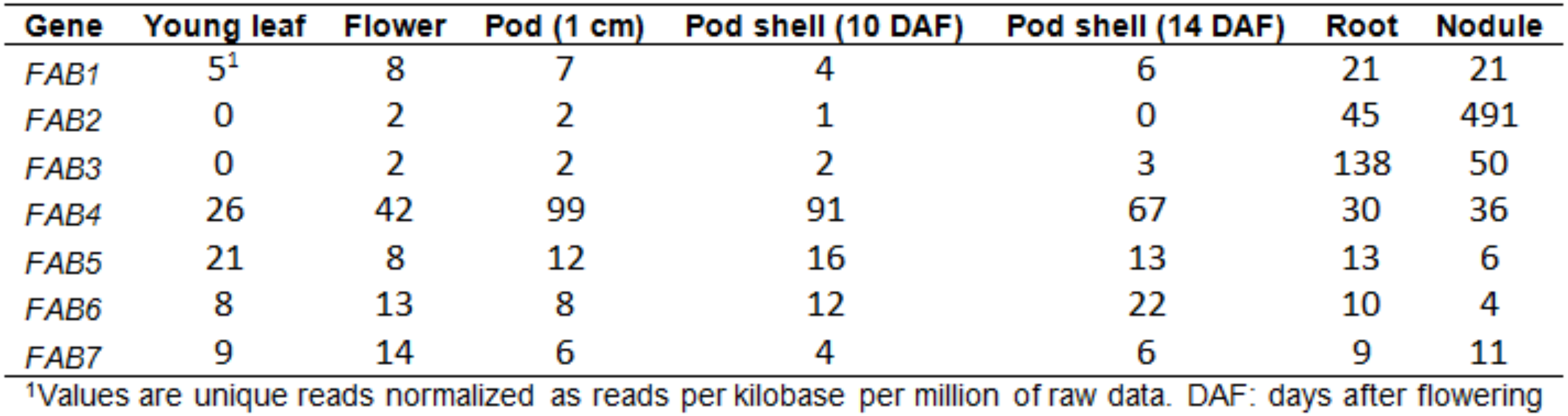
Gene	Leaves	Flower	Green pods	Apical meristem	Root	Root tip	Nodule
SEED1	0[1]	0	8	0	0	0	0
SEED2	0	0	616	0	0	0	0
SEED3	4	1	504	0	0	0	0
SEED5	0	0	88	0	0	0	0
SEED6	0	0	116	0	0	0	0
SEED7	0	0	53	0	2	0	121
SEED10	0	0	312	0	1	0	0
SEED11	No data available						
SEED12	0	0	1	1	0	0	0

[1] Values are unique reads normalized as reads per kilobase per million of raw data

- ***SEED* genes were expressed mainly in developing seeds and in green pods containing seeds at full stage (R6). Intriguingly, *SEED7* also showed expression in nodules consistently in both transcriptomic databases.**

Results -Tissue-specific expression of *FAB* genes

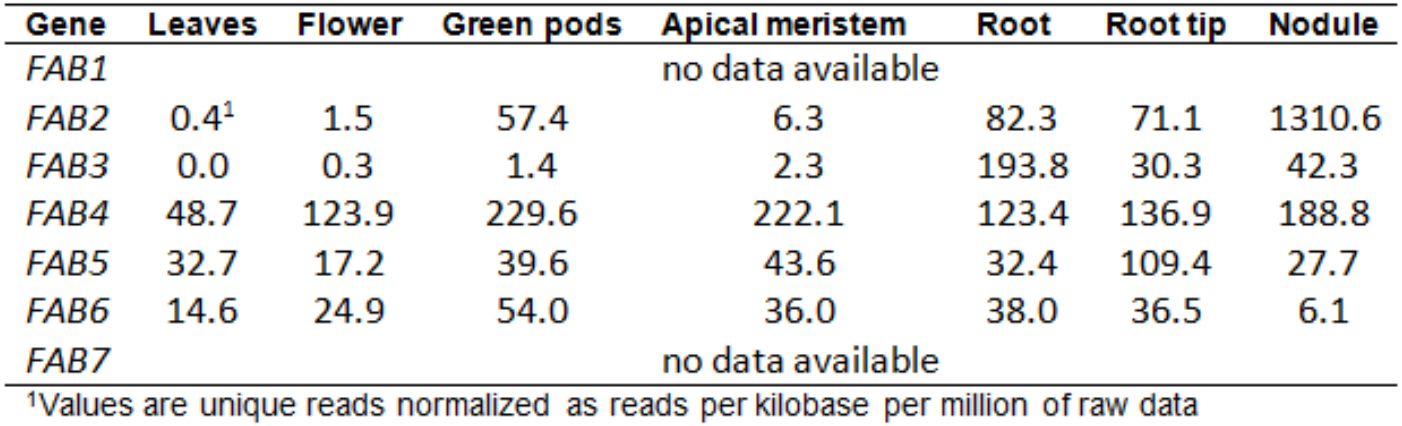
A Soybase



Gene	Young leaf	Flower	Pod (1 cm)	Pod shell (10 DAF)	Pod shell (14 DAF)	Root	Nodule
FAB1	5[1]	8	7	4	6	21	21
FAB2	0	2	2	1	0	45	491
FAB3	0	2	2	2	3	138	50
FAB4	26	42	99	91	67	30	36
FAB5	21	8	12	16	13	13	6
FAB6	8	13	8	12	22	10	4
FAB7	9	14	6	4	6	9	11

[1] Values are unique reads normalized as reads per kilobase per million of raw data. DAF: days after flowering

B Transcriptome Atlas



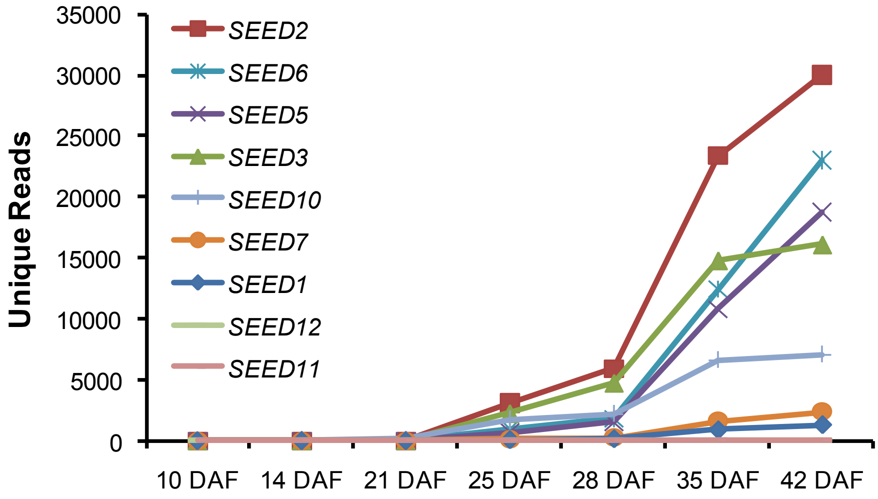
Gene	Leaves	Flower	Green pods	Apical meristem	Root	Root tip	Nodule
FAB1				no data available			
FAB2	0.4[1]	1.5	57.4	6.3	82.3	71.1	1310.6
FAB3	0.0	0.3	1.4	2.3	193.8	30.3	42.3
FAB4	48.7	123.9	229.6	222.1	123.4	136.9	188.8
FAB5	32.7	17.2	39.6	43.6	32.4	109.4	27.7
FAB6	14.6	24.9	54.0	36.0	38.0	36.5	6.1
FAB7				no data available			

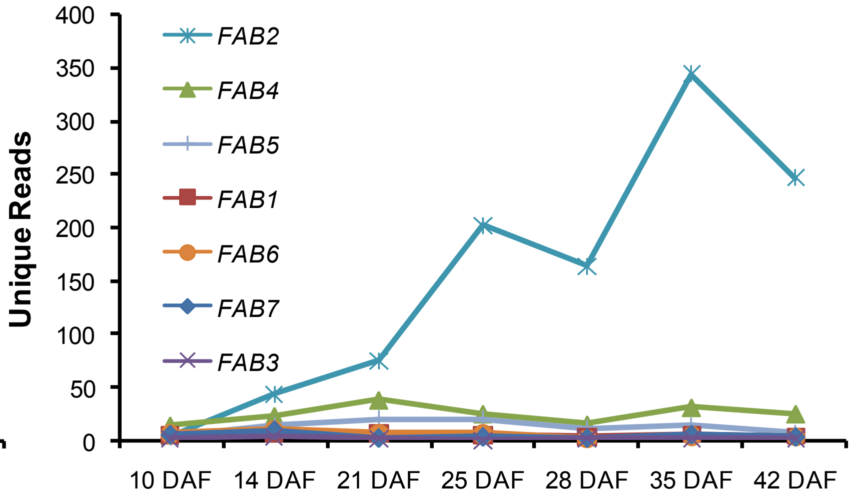
[1] Values are unique reads normalized as reads per kilobase per million of raw data

- ***FAB*** **genes showed relatively high levels of transcripts in various organs and tissues, especially in roots, root tips, and nodules. This suggests less tissue specificity than the *SEED* genes.**

Results –Expression in developing seeds

Expression profiles for *SEED* and *FAB* genes in soybean seeds at different stages. Data were obtained from Soybase http://soybase.org/soyseq/. DAF: days after flowering. Displayed data are unique reads normalized as reads per kilobase per million reads of raw data.





- Most of the *SEED* genes are actively expressed 21 days after flowering, and reach their observed maximum transcript accumulation 42 days after flowering (physiologically mature seeds).
- Only the *FAB2* candidate, which maps to Fas_Stearic2-2 QTL, clearly showed high expression in developing seeds.

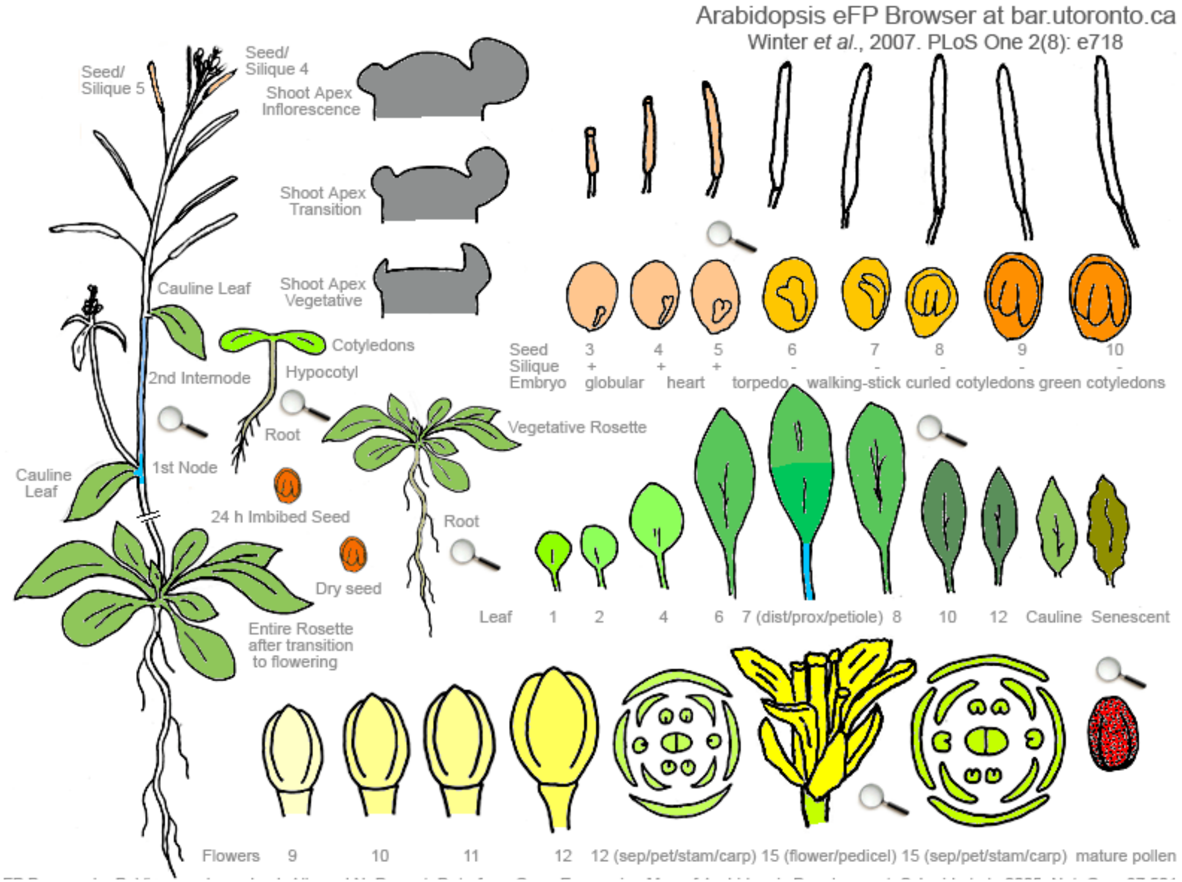
Use of web-based tools for visualization of transcriptomic data

With the delivery of transcriptomic databases, different web-based resources are emerging for data visualization. These tools include electronic Northerns (e-Northerns) and eFP (electronic Fluorescent Pictographs) browsers to cluster genes based on expression intensity, and to draw temporal and spatial expression, respectively.

In this tutorial, we use an eFP browser to analyze the temporal and spatial expression for both set of genes.

The use of eFP browsers

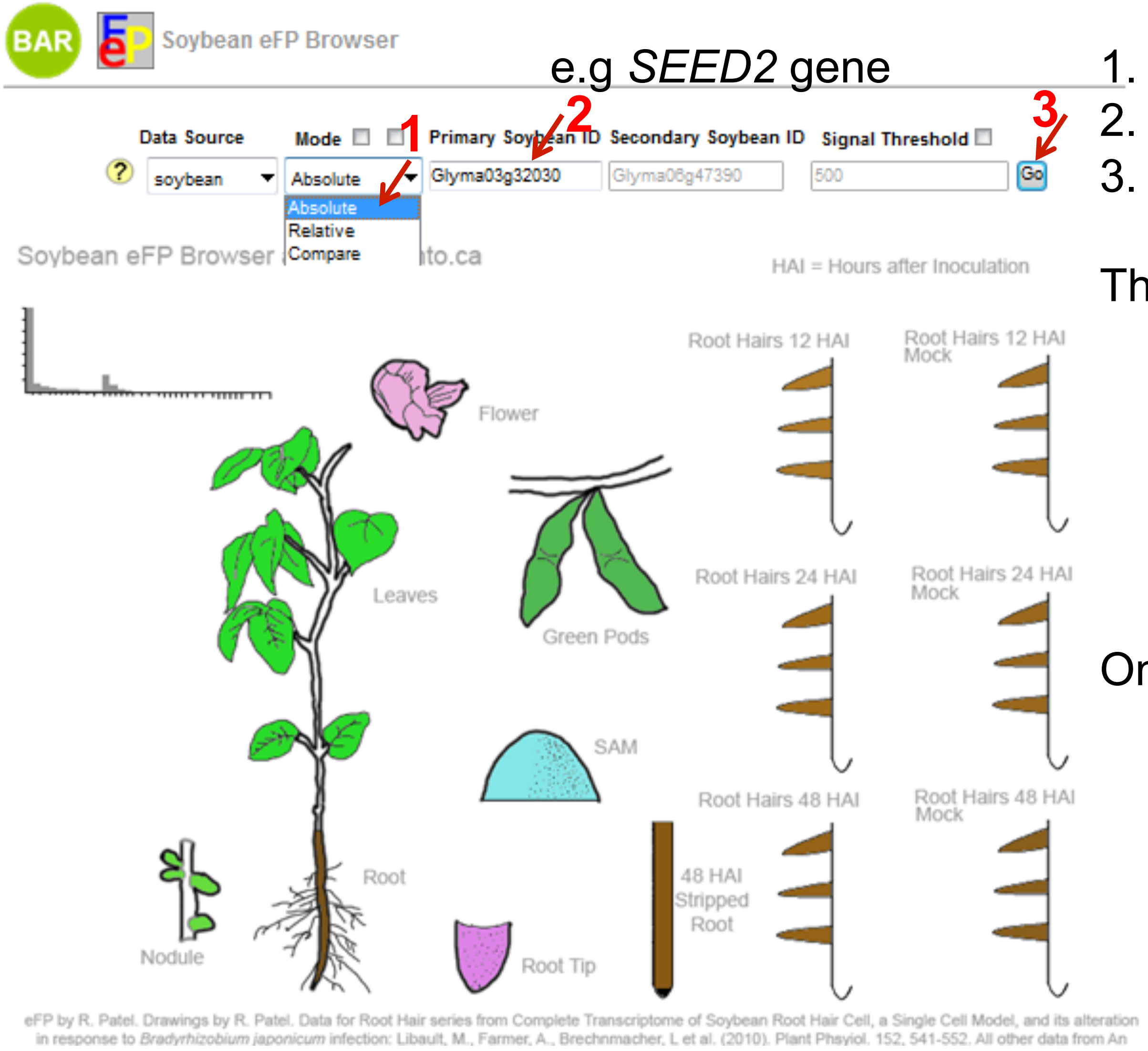
Electronic Fluorescent Pictograph Browsers (eFP browsers) are online applications to build expression maps of your gene of interest based on transcript expression data. eFP browsers for *Arabidopsis*, poplar, *Medicago truncatula*, rice, barley, maize and soybean can be freely accessed at The Bio-Array Resource for Plant Biology http://www.bar.utoronto.ca.



Snapshoot from the The Bio-Array Resource for Plant Biology

The soybean eFP browser

http://www.bar.utoronto.ca/efpsoybean/cgi-bin/efpWeb.cgi



e.g *SEED2* gene

Data Source: soybean
Mode: Absolute
Primary Soybean ID: Glyma03g32030
Secondary Soybean ID: Glyma06g47390
Signal Threshold: 500

eFP by R. Patel. Drawings by R. Patel. Data for Root Hair series from Complete Transcriptome of Soybean Root Hair Cell, a Single Cell Model, and its alteration in response to *Bradyrhizobium japonicum* infection: Libault, M., Farmer, A., Brechnmacher, L et al. (2010). Plant Phsyiol. 152, 541-552. All other data from An

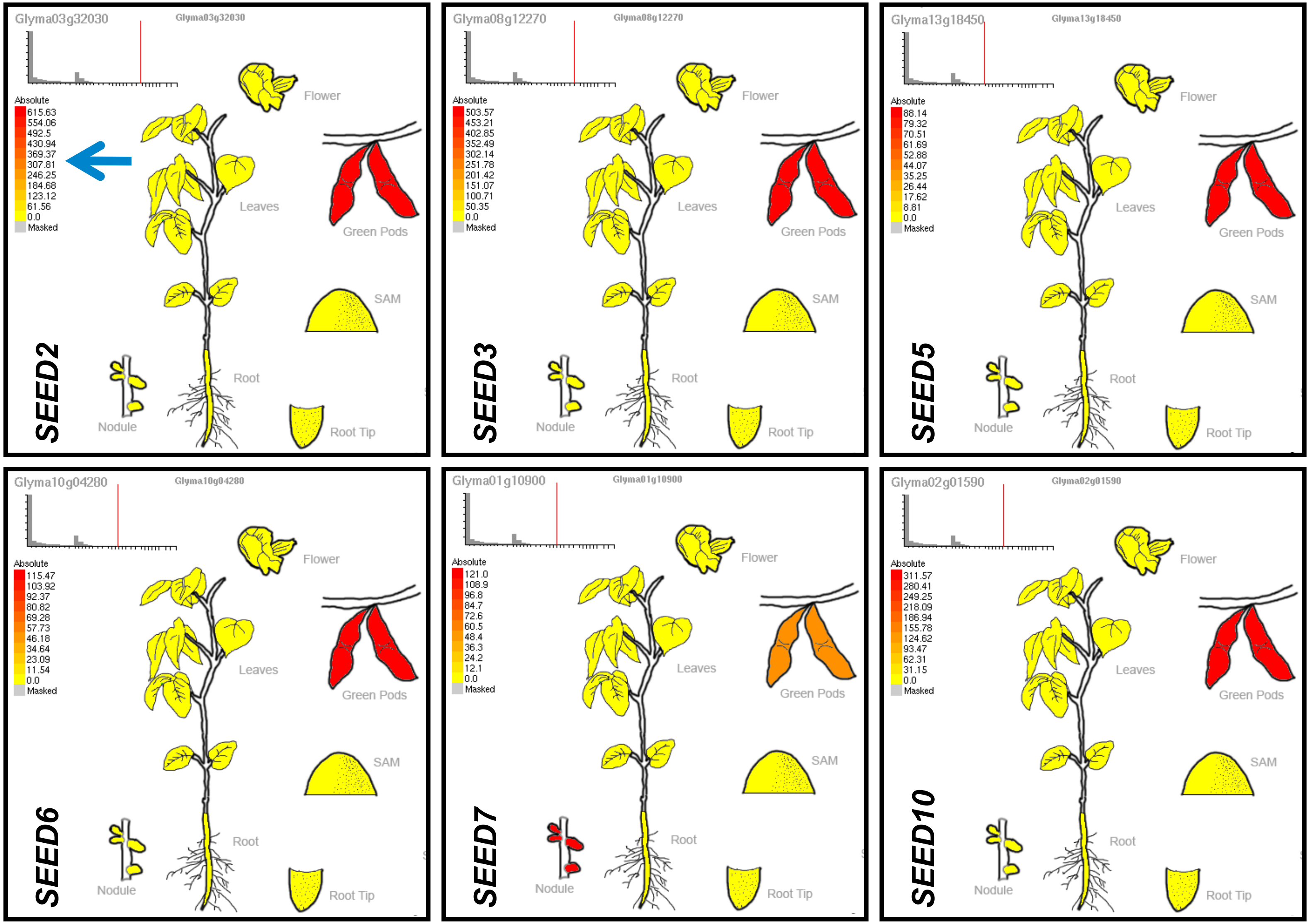
1. Select "absolute"
2. Enter gene ID as indicated
3. Click on "Go"

The soybean eFP browser generates a pictographic representation of transcriptome data from the Transcriptome Atlas database.

One can also directly compare expression of two different genes of interest.

Results

Expression profiles confirming expression of *SEED* genes exclusively to green pods with seeds at full stage (R6). Profiles were built using the soybean eFP browser. The blue arrow points the expression scale (the more intense red color, the more gene expression).



Conclusions

- Consistent results for *SEED* and *FAB* genes were obtained from both databases (Soybase and Transcriptome Atlas).
- *SEED* genes were expressed almost exclusively in developing seeds from medium to late developmental stages. Conversely, *FAB* genes showed less seed-specificity and higher levels in other organs and tissues including roots, root tips and nodules.
- The candidate *FAB2* gene showed high levels of transcripts in developing seeds. This suggests that this gene may have a major effect on the Fas-Stearic2-2 QTL.
- The *SEEDS* may be potential sources of seed-specific promoters in soybean.
- Databases based on RNA-Seq technology are a powerful source of gene expression data.

References Cited

Kalinski, A., J. M. Weisman, B. F. Mathews, and M. Herman. 1989. Molecular cloning of a protein associated with soybean seed oil bodies that is similar to thiol proteases of the papain family. Journal of Biological Chemistry 265: 13843-13848.

Libault, M, A. Farmer, T. Joshi, K. Takahashi, R. J. Langley, L. D. Franklin, J. He, D. Xu, G. May, and G. Stacey. 2010. An integrated transcriptome atlas of the crop model (*Glycine max*) and its use in comparative analyses in plants. Plant Journal 63: 86-99. (Available online at: http://dx.doi.org/10.1111/j.1365-313X.2010.04222.x) (verified 21 July 2011).

Natarajan, S. S., C. Xu, H. Bae, T. J. Caperna, and W. Garrett. 2007. Determination of optimal protein quantity required to identify abundant and less abundant soybean seed proteins by 2D-PAGE and MS. Plant Molecular Biology Report 25: 55-62.

Severin, A., J. Woody, Y. –T. Bolon, B. Joseph, B. Diers, A. Farmer, G. Muehlbauer, R. Nelson, D. Grant, J. Specht, M. Graham, S. Cannon, G. May, C. Vance, and R. Shoemaker. 2010. RNA-Seq Atlas of *Glycine max*: A guide to the soybean transcriptome. BMC Plant Biology 10: 160. (Available online at: http://dx.doi.org/doi:10.1186/1471-2229-10-160) (verified 21 July 2011).

Vodkin, L.O., and N. V. Raikhel. 1986. Soybean lectin and related proteins in seeds and roots of Le+ and Le − soybean varieties. Plant Physiology 81: 558–565.

External Links

Arabidopsis Transcriptome Genomic Express Database [Online]. Salk Institute Genomic Analysis Laboratory. Available at: http://signal.salk.edu/cgi-bin/atta (verified 21 July 2011).

The Bio-Array Resource for Plant Biology [Online]. University of Toronto. Available at: http://www.bar.utoronto.ca (verified 26 July 2011).

Maize C3/C4 Transcriptomic Database [Online]. Cornell University. Available at: http://c3c4.tc.cornell.edu/search.aspx (verified 26 July 2011).

Medicago truncatula Gene Expression Atlas [Online]. The Samuel Roberts Noble Foundation. Available at: http://mtgea.noble.org/v2/ (verified 21 July 2011).

Next Generation Sequencing for Plant Breeders [Online]. The Plant Breeding and Genomics Community, eXtension. Available at: http://www.extension.org/article/32489 (verified 26 July 2011).

PopGenIE: The Populus Genome Integrative Explorer [Online]. Popgenie.org. Available at: http://www.popgenie.org/ (26 July 2011).

RiceGE Japonica: Rice Functional Genomic Express Database [Online]. Salk Institute Genomic Analysis Laboratory. Available at: http://signal.salk.edu/cgi-bin/RiceGE (verified 21 July 2011).

RiceGE: Rice (*indica*) Functional Genomic Express Database [Online]. Salk Institute Genomic Analysis Laboratory. Available at: http://signal.salk.edu/cgi-bin/RiceiGE (verified 21 July 2011).

RNA-Seq Atlas of *Glycine max* [Online]. Soybase and the Soybean Breeder’s Toolbox. Available at: http://soybase.org/soyseq/ (verified 26 July 2011).

Soybean Upstream Regulatory Element (SURE) Database [Online]. The Ohio State University, OARDC. Available at: http://www.oardc.ohio-state.edu/SURE/ (verified 26 July 2011).

Soybean eFP Browser [Online]. The Bio-Array Resource for Plant Biology, University of Toronto. Available at: http://www.bar.utoronto.ca/efpsoybean/cgi-bin/efpWeb.cgi (verified 26 July 2011).

Tomato Expression Database (TED) [Online]. Boyce Thompson Institute for Plant Research, Cornell University. Available at: http://ted.bti.cornell.edu/ (verified 26 July 2011).

Transcriptome Atlas of *Glycine max* [Online]. Digital Biology Laboratory, University of Missouri. Available at: http://digbio.missouri.edu/soybean_atlas/ (verified 26 July 2011).

Acknowledgements

The author would like to thank Dr. David Francis, Heather Merk and Jose Zambrano for reviewing this tutorial and providing helpful suggestions. The author also thanks Drs. Robert Bouchard and Leah McHale for identification of *SEED* and *FAB* genes, respectively. Promoter analysis of *SEED* and *FAB* genes is being conducted in the laboratory of Dr. John J. finer (Department of Horticulture and Crop Science, The Ohio State University/OARDC). The author is funded by a Graduate Associateship from the Department of Horticulture and Crop Science, The Ohio State University, and partial support from CONACYT, Mexico.

THE OHIO STATE UNIVERSITY
OARDC

CONACYT
Consejo Nacional de
Ciencia y Tecnología
www.conacyt.gob.mx