Some bioinformatic tools for plant biology research in the legumes

Focusing on:
the “Legume Federation” project: legumefederation.org
and some of the constituent federation projects:
legumeinfo.org, peanutbase.org, soybase.org and others

Contact us: steven.cannon@ars.usda.gov
or legumefederation.org/contact
There are ~20,000 species in the legumes, and ~two dozen crop legumes: alfalfa, soybean, peanut, lupin, mungbean, ...

Fascinating biology: nitrogen-fixing symbiosis, special phosphorus uptake mechanisms, high-protein- and oil seeds, floral/developmental biology, secondary metabolism, domestication histories, etc.
The NSF “Legume Federation” project objectives are to facilitate development, data sharing and standards, and common tools and interfaces across many on-line resources for the legumes: 

**Legume Information System, PeanutBase, SoyBase, SoyKB, Cool Season Food Legume database, KnowPulse, LegumeIP, MedicagoGenome, PhaseolusGenes, LegumeBase, MedicagoHapmap**, etc.

Steven Cannon (USDA-ARS)  
David Fernandez-Baca (ISU)  
Andrew Farmer (NCGR)  
Jeremy DeBarry (CyVerse)  
Ethy Cannon (ISU)  
Chris Town (JCVI)

Sudhansu Dash, Vivek Krishnakumar, Nathan Weeks, Jacqueline Campbell, Akshay Yadov, Joel Berendsen, Sam Hokin,  
Others in “federation projects” ... Tripal: Dorrie Main,  
Stephen Ficklin, Lacey Sanderson, Sook Jeung;  
Phytozome, SoyBase, KnowPulse,  
MedicagoGenome, USAID chickpea project, Noble Foundation alfalfa project, ...
LegumeFederation.org – right now, mostly a jumping-off point to federated projects...

Welcome to Legume Federation

The Legume Federation is an NSF project to build a federation of legume databases through data standards, distributed development and comparative analysis, to support research across the legume family, and to support robust agriculture for a world that is significantly "legume-fed".

Participating Genomic Data Portals (GDPs) currently include, but are not limited to:

- Medicago genome/JCVI
- SoyBase
- Legume Information System (LIS)
- Cool Season Food Legume Database
- PeanutBase
- KnowPulse
- Feed the Future Climate Resilient Chickpea
- Alfalfa Genome
- Medicago truncatula HapMap

The goals of the Legume Federation include:

1. share knowledge, development efforts, and data sets across all legume crops;
2. define standards for data formats, metadata, Web service protocols and ontology use;
A collaborative, community resource to facilitate crop improvement by integrating genetic, genomic, and trait data across legume species. (Cite LIS)

**Browsers**

- Gene Families
- Gene Search
- Maps
- Protein Domains
- QTL Search
- Context Viewer
- Legume Mines

**Search sequences and features in genome sequences**

- BLAST Sequence Search
- BLAT Sequence Search
- Keyword Search

**Browse genomes**

- Adzuki bean (3.0)
- Chickpea (Kabuli)
- Chickpea (Desi)
- Common bean
- Lotus (3.0)
- Medicago (4.0)
- Mungbean
- Pigeonpea
- Red clover (2.1)
- Arachis duranensis (PeanutBase)
- Arachis ipaensis (PeanutBase)
- Soybean (SoyBase)

**NEWS**

- June 2016 update: New query tool (InterMine); BioProject search for legumes at NCBI; more genomes with gene family assignments; newsletter
- May 2016 update: Upload data to geographic map; gene family alignment viewer; Desi chickpea; improve search
- Apr 2016 update: Incorporation of user-supplied germplasm data in maps, utilities for finding recent legume publications

More News
Browsers with synteny mappings between genomes
A collaborative, community resource to facilitate crop improvement by integrating genetic, genomic, and trait data across legume species. (Cite LIS)

Updates in first week of every month
A collaborative, community resource to facilitate crop improvement by integrating genetic, genomic, and trait data across legume species. (Cite LIS)

Sequence search (BLAST & BLAT); keyword search (of features in the browsers)
A collaborative, community resource to facilitate crop improvement by integrating genetic, genomic, and trait data across legume species. (Cite LIS)

Gene families, domains, synteny
LIS gene family

Nodes and genes are clickable

Prefix and color indicate species

Legend:
- root node
- internal node
- glyma (Glycine max, soybean)
- phavu (Phaseolus vulgaris, common bean)
- cajca (Cajanus cajan, pigeonpea)
- vigra (Vigna radiata, mungbean)
- car (Cicer arietinum, chickpea)
- mdt (Medicago truncatula, barrel medic)
- lotja (Lotus japonicus, birdsfoot trefoil)
- adhu (Arachis duranensis, Arachis duranensis / wild p
- araip (Arachis ipaensis, Arachis ipaensis / wild peanut
- prupe (Prunus persica, peach)
- arath (Arabidopsis thaliana, mouse-ear cress)
- solly (Solanum lycopersicum, tomato)
- vitvi (Vitis vinifera, wine grape)
- zeama (Zea mays, maize)
- orysa (Oryza sativa, rice)
- ambral (Amborella trichopoda, Amborella trichopoda)
- hilite: araip.Araip.30XFU.1, true
<table>
<thead>
<tr>
<th>Label</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>phytozome_10_2.59028464</td>
<td>MSVSLSDCSDDLLCGEDDSSESVEVLSGESPEVSSSDLESSEASSAVEEES</td>
</tr>
<tr>
<td>zeama.GRMZM2G476685_P01</td>
<td>YDLCAGEDDSALLCTDDDGCSAVEDDSAAAAAAAS</td>
</tr>
<tr>
<td>vitiv.GSVIT01020167001</td>
<td>MSLYSDRFSDLGLCEDRSTILSGDLPESCSSDLESPTDI-EEES</td>
</tr>
<tr>
<td>vigra.Vradi0384s00010.1</td>
<td>MSLYSRDFSDDLCGESDDSGVGLESPECSDSLDDSSPPSSEEVV</td>
</tr>
<tr>
<td>solly.Soly05g006050.2.1</td>
<td>MSVSCSNCPDLDLGCEDSSGVEGSLGESPECSDSLDSSPPSSEEVV</td>
</tr>
<tr>
<td>prupe.ppa008337m</td>
<td>MSVSCSNDLCSDDLCCEDYSGILSGESPECSDSLDSEPPACSES</td>
</tr>
<tr>
<td>phavu.Phuvi003G054900.1</td>
<td>DDDASYLLECAEDAGAAVFDDDICTCTTTEDEDECCSVGEEES</td>
</tr>
<tr>
<td>phavu.Phuvi002G074800.1</td>
<td>DDDASYLLECAEDAGAAVFDDDICTCTTTEDEDECCSVGEEES</td>
</tr>
<tr>
<td>orysa.LOC_092g11450.1</td>
<td>DDDASYLLECAEDAGAAVFDDDICTCTTTEDEDECCSVGEEES</td>
</tr>
<tr>
<td>orysa.LOC_089g32540.1</td>
<td>DDDASYLLECAEDAGAAVFDDDICTCTTTEDEDECCSVGEEES</td>
</tr>
<tr>
<td>medt.Medtr8g063120.1</td>
<td>TGELLCGEDDSSVEVLTGDLPECSDDLDSEFFAEEBBBES</td>
</tr>
<tr>
<td>medt.Medtr5g035360.1</td>
<td>TGELLCGEDDSSVEVLTGDLPECSDDLDSEFFAEEBBBES</td>
</tr>
<tr>
<td>lotja.Lj2g3v1373380.1</td>
<td>TGELLCGEDDSSVEVLTGDLPECSDDLDSEFFAEEBBBES</td>
</tr>
<tr>
<td>lotja.Lj0g3v0115929.1</td>
<td>TGELLCGEDDSSVEVLTGDLPECSDDLDSEFFAEEBBBES</td>
</tr>
</tbody>
</table>
Find similar "genome contexts" for this gene
“Genome contexts” view: find similar regions from other legume species

Query gene family

Other gene families
Prefix gives genus & species: aradu, araip, glyma phavu
plot region with respect to region of our gene
A new federated version with Arabidopsis and other species (from legumeinfo.org or legumefederation.org)
Examples of interconnections and shared development
These mines were built from various data sources for individual legumes:

- **BeanMine** - built entirely from the LIS multi-species database.
- **SoyMine** - genomic data from the LIS multi-species database.
- **PeanutMine** - built from Peanutbase (chado and legacy.

This mine combines the data contained in the following:

- **LegumeMine** - combination of the above mines.

These mines are in steady development, but the public release is available:

"InterMine" for four legume species (bean, soybean, peanut, *Medicago*), at LegumeFederation.org and LegumeGenome.org with synteny- and gene family-based linkages.
The “Mines” specialize in powerful operations on lists (genes, regions, expression, etc.)
Welcome to the LegFed Soymine!

This mine integrates many types of data for soybean. It is currently under development by LegFed/NCGR as a demo built from the LIS tripal.chado database (genomic data) and exports from Soybase (genetic data). Contact Sam Hokin shokin at ngr dot org.

Latest updates (v0.3, 1 June 2016)
- Added Plant Reactome pathways to polypeptides with external links to Plant Reactome

Stored query “templates” ...
Result of one of the example templates, relating scaffolds, gene families, genes
PeanutBase.org: similar interface and capabilities as LegumeInfo.org, but with some special focuses and tools...

- (July, 2016) New search page for peanut projects recorded at NCBI
- (June, 2016) New PeanutMine query tool, more intuitive trait search, improved marker display, new "A" genome report
- (May, 2016) Upload data to geographic map, improved QTL search
- (Apr, 2016) A utility for finding peanut/Arachis publications

Support
This resource is being developed for U.S. and international peanut researchers and breeders, with support from The Peanut Foundation and the many contributors that have made the Peanut Genomics Initiative possible.
## PeanutBase

**Genetic and genomic data to enable more rapid crop improvement in peanut.**

### Search Arachis QTL

All fields are optional and partial names are accepted. Click column headers to sort.

**Total QTL count: 232**

<table>
<thead>
<tr>
<th>Species</th>
<th>Trait Class</th>
<th>Trait</th>
<th>QTL Symbol in Pub</th>
<th>Trait Name in Pub</th>
<th>From this Pub</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Arachis hypogaea</em></td>
<td>oil</td>
<td></td>
<td>Linoic acid, PE (LG01, PM137-Ah1TC6H03)</td>
<td>Linoic acid</td>
<td></td>
</tr>
<tr>
<td><em>Arachis hypogaea</em></td>
<td>oil</td>
<td></td>
<td>Linoic acid, El (LG07, IPAHM295)</td>
<td>Linoic acid</td>
<td></td>
</tr>
<tr>
<td><em>Arachis hypogaea</em></td>
<td>oil</td>
<td></td>
<td>Linoic acid, PE (LG08, Ah1TC3A12-PM433)</td>
<td>Linoic acid</td>
<td></td>
</tr>
<tr>
<td><em>Arachis hypogaea</em></td>
<td>oil</td>
<td></td>
<td>Linoic acid, El (pPGSSeq11G07)</td>
<td>Linoic acid</td>
<td></td>
</tr>
<tr>
<td><em>Arachis hypogaea</em></td>
<td>oil</td>
<td></td>
<td>Linoic acid, EII (LG01, Ah1TC6H03-Ah2TC11A04)</td>
<td>Linoic acid</td>
<td></td>
</tr>
</tbody>
</table>

**Found 42 QTL**

<table>
<thead>
<tr>
<th>Organism</th>
<th>Species</th>
<th>Trait Class</th>
<th>Trait</th>
<th>QTL name</th>
<th>QTL Symbol in Publication</th>
<th>Trait Name in Publication</th>
<th>Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Arachis hypogaea</em></td>
<td>oil</td>
<td>Seed linoleic 1-1</td>
<td></td>
<td>Seed linoleic 1-1</td>
<td>Linoic acid, PE (LG01, PM137-Ah1TC6H03)</td>
<td>Linoic acid</td>
<td>Sarvamangala, Gowda et al., 2011a</td>
</tr>
<tr>
<td><em>Arachis hypogaea</em></td>
<td>oil</td>
<td>Seed linoleic 1-12</td>
<td></td>
<td>Seed linoleic 1-12</td>
<td>Linoic acid, El (LG07, IPAHM295)</td>
<td>Linoic acid</td>
<td>Sarvamangala, Gowda et al., 2011a</td>
</tr>
<tr>
<td><em>Arachis hypogaea</em></td>
<td>oil</td>
<td>Seed linoleic 1-13</td>
<td></td>
<td>Seed linoleic 1-13</td>
<td>Linoic acid, PE (LG08, Ah1TC3A12-PM433)</td>
<td>Linoic acid</td>
<td>Sarvamangala, Gowda et al., 2011a</td>
</tr>
<tr>
<td><em>Arachis hypogaea</em></td>
<td>oil</td>
<td>Seed linoleic 1-15</td>
<td></td>
<td>Seed linoleic 1-15</td>
<td>Linoic acid, El (pPGSSeq11G07)</td>
<td>Linoic acid</td>
<td>Sarvamangala, Gowda et al., 2011a</td>
</tr>
<tr>
<td><em>Arachis hypogaea</em></td>
<td>oil</td>
<td>Seed linoleic 1-3</td>
<td></td>
<td>Seed linoleic 1-3</td>
<td>Linoic acid, EII (LG01, Ah1TC6H03-Ah2TC11A04)</td>
<td>Linoic acid</td>
<td>Sarvamangala, Gowda et al., 2011a</td>
</tr>
</tbody>
</table>
QTLs, traits, ... on maps
### Species distributions in a GIS context

#### PeanutBase | Germplasm Map

- **Refine map to accessions**
- **Add my data**
- **Base map**
- **Go to geographic coordinates**

---

**Results: 2000** Your max # of results are listed below, but may appear to be clustered at the center of the map. Try zooming the map in, change search criteria, or increase the max # of results.

<table>
<thead>
<tr>
<th>accession</th>
<th>taxon</th>
<th>crop name</th>
<th>orig country</th>
<th>collection site</th>
</tr>
</thead>
<tbody>
<tr>
<td>PI 536259</td>
<td><em>Arachis hypogaea var. fastigiata</em></td>
<td>Valencia peanut</td>
<td>BRA</td>
<td>Village market, Vilhena, Ro...</td>
</tr>
<tr>
<td>PI 536258</td>
<td><em>Arachis hypogaea var. fastigiata</em></td>
<td>Valencia peanut</td>
<td>BRA</td>
<td>Village market, Vilhena, Ro...</td>
</tr>
</tbody>
</table>
SoyBase.org: with special focuses and tools for soybean researchers...
QTLs, traits, ... on maps
Gene expression in browsers and other contexts

Gene models - Glyma.Wm82.a2.v1

Soybean gene expression atlas - Libault et al. ...

Soybean gene expression atlas - Severin et al. ...

Soybean gene expression atlas - Schmutz et al. ...

Glyma.15g026400

LIP0XYGENASE...(click for more)

Glyma.15g026300

LIP0XYGENASE...(click for more)
Some bioinformatic tools for plant biology research in the legumes

Focusing on:
the “Legume Federation” project: legumefederation.org
and some of the constituent federation projects:
legumeinfo.org, peanutbase.org, soybase.org and others

Please let us know if you have questions, requests, or data to contribute!

steven.cannon@ars.usda.gov
legumefederation.org/contact