PeanutBase
The peanut community database

Ethy Cannon
Iowa State University
APRES 2017
Contributors to PeanutBase

Iowa State University/PeanutBase
- Ethy Cannon – lead
- Wei Huang – computational biologist
- Paul Otyama – PhD student

Iowa State University/Legume Federation
- Jacqueline Campbell – data curator and outreach

USDA-ARS at Ames, IA
- Steven Cannon – lead scientist, ARS legume database project
- Nathan Weeks – IT specialist and computational biologist
- Scott Kalberer – data curator
- Andrew Wilkey – programmer

National Center for Genomic Resources
- Andrew Farmer – lead, Legume Information System
- Sudhansu Dash - geneticist and computational biologist
- Alex Rice - programmer
PeanutBase is a community resource

Let us know what you need

**Cite PeanutBase and peanut genome**

- Click for more on PeanutBase and the Peanut Genomics Initiative ...

**Support**

This resource is being developed for U.S. and international peanut research programs.
PeanutBase is a community resource

Report errors

**Cite PeanutBase and peanut genome

* Click for more on PeanutBase and the Peanut Genomics Initiative ...

Support

This resource is being developed by US and international research groups and

NEWS

- (June, 2017) PeanutBase now hosts Peanut Bioscience content
- (May, 2017) Peanut downloads in the data store, more marker data
- (Mar 2017) Interactive tour for QTL data and new marker track
- (Feb 2017) SNPs from the Arachis SNP chip now available

More News
PeanutBase is a community resource

Work with us to host your public data
PeanutBase is a community resource

Sign up for the ~quarterly newsletter
Hot off the presses

PeanutBase now hosts PeanutBioscience.com

- AAGB archives and information
- IPGC and PGP meeting minutes and information
Hot off the presses

Data store for downloading datasets for peanut and related species
What is in PeanutBase?

1. Genomes
2. Genes
3. Maps & Markers
4. Germplasm & Traits
5. Germplasm geographic viewer
6. Sequence search tools
7. Gene families & synteny
8. Website tours (and other help)
1. **Genome sequences for A. duranensis and A. ipaensis**

Genome browsers maintained by Wei Huang
2. Predicted genes, with sequences, structure, functional descriptions, and expression data.
2. Predicted **genes**, with sequences, structure, functional descriptions, and **expression** data.

(thanks Josh, Peggy et al., and Nick Provart lab for these views)
2. Predicted genes, with sequences, structure, functional descriptions, and expression data.
3. Various genetic map and marker sets, including the Affymetrix SNP array. (Pandey et al., 2017, Clevenger et al., 2017)
4. germplasm + traits
4. germplasm + traits
### Overview

<table>
<thead>
<tr>
<th>Stock Name</th>
<th>PI 196635</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other Name(s)</td>
<td>47-16, CC270</td>
</tr>
<tr>
<td>Stock type</td>
<td>Accession</td>
</tr>
<tr>
<td>Market type</td>
<td>Spanish</td>
</tr>
<tr>
<td>Organism</td>
<td><em>Arachis hypogaea</em> (Cultivated peanut)</td>
</tr>
<tr>
<td>GRIN Global accession</td>
<td>PI 196635</td>
</tr>
<tr>
<td>Origin</td>
<td>Madagascar</td>
</tr>
<tr>
<td>Geographic location</td>
<td>GIS</td>
</tr>
</tbody>
</table>

### Description

#### Traits

- **1995 peanut increase in Griffin.**
  - Color of the leaf at 60 - 90 days: Dark green
  - Flowers on main axis at 60 - 90 days: No
  - Growth habit recorded at 60 to 70 days: Spreading and bunch

- **Accession which comprise the U.S. cultivated peanut core collection**
  - A flag to indicate the accession is part of the core subset: Yes, accession is part of the core

- **Contains data taken at the Southern RPIS in Griffin, GA.**
  - The information presented under this environment represents data which has been recorded over several years at the Southern Regional Plant Introduction Station. No one specific environment exists for the data.
  - Resistance to leafspot infection: Intermediate

- **Fatty Acid composition of the peanut core collection analyzed using a gas chromatograph (GC)**
  - Arachidate (20:0) methyl ester percentage: 1.21
  - Behenate (22:0) methyl ester percentage: 3.64

---

**Image:** Mini core regeneration Chen
Arachis species, S. America
Brazilian *A. hypogaea* accessions colored by pod type.
6. Sequence search tools

PeanutBase is a community resource. Please help us meet your needs by taking our survey.

**Cite PeanutBase and peanut genome**

› Click for more on PeanutBase and the Peanut Genomics Initiative ...
... against either genome
... gives an overview picture ...
The following table summarizes the results of your BLAST. Click on a triangle on the left to see the alignment and a visualization of the hit, and click the target name to get more information about the target hit.

<table>
<thead>
<tr>
<th>#</th>
<th>Query Name (Click for alignment &amp; visualization)</th>
<th>Target Name</th>
<th>E-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>partial lipoxygenase Glyma15g03040</td>
<td>Aradu.A08</td>
<td>3.03876E-69</td>
</tr>
</tbody>
</table>

**Hit Visualization**

The image above shows the relationship between query and target for this particular BLAST hit.

... and details about the matches to each chromosome.
... which link to GBrowse at that location (which may be large)
7. Gene families and synteny with other legumes

BLAST Search

Search for one or more of your sequences (using BLAST) against the genome sequences for two wild pear species, *P. duranensis* and *A. ipaensis*. First pick a query type (nucleotide or protein). You will be able to set search parameters later in the page.

Choose the appropriate program based on the Query type and Target database type. Please click on the programs in the table below to view the search form.

<table>
<thead>
<tr>
<th>Query Type</th>
<th>Database Type</th>
<th>BLAST Program</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nucleotide</td>
<td>Nucleotide</td>
<td>blastn: Search a nucleotide database using a nucleotide query.</td>
</tr>
<tr>
<td>Protein</td>
<td>Nucleotide</td>
<td>blastx: Search protein database using a translated nucleotide query.</td>
</tr>
<tr>
<td>Protein</td>
<td>Protein</td>
<td>blastp: Search protein database using a protein query.</td>
</tr>
<tr>
<td>Protein</td>
<td>Nucleotide</td>
<td>tblastn: Search translated nucleotide database using a protein query.</td>
</tr>
</tbody>
</table>

A blastp or blastx search ...
Enter Protein Query Sequence

Enter one or more queries in the top text box or use the browse button to upload a file from your local disk. The file may contain a single sequence or a list of sequences. In both cases, the data must be in FASTA format.

Enter FASTA sequence(s)

> Glyma.02g227200.1 fatty acid desaturase 8
MVKDTKPLAYAANNYYQQGSSFDPSAPPFKIAEIRASIPKHCVKNNPWSLSYVLRDVLVIAAVAAAIHDNWLWLLWLINKIYCQGT
MFWALFVLGHDCGHGSFSDPLNLSVLGHSLVGLPHYGRISHRTHQNHGHIKDESWVPILEKTYKELSDMTRLIRMFTVFFLVFY
PIYLFRSRPGKEGSHFNPYSNLFPPSERKGIATLCTWATMFSLLIYLSFITSPPLLKVLYGIPYWIFVMWLDFTYLLHHHHHGHQKLPWYRGK
EWSYLRGLCTTTDREDYGWNIINHHDIGTHVIHLFPQIPHYHLVEATQAAKPVLDGYYREPERSAPLPFHLKYLIQSMRQDFHVSĐTGD

Enter query sequence(s) in the text area.

Or upload your own query FASTA:

Choose File: no file selected

The file should be a plain-text FASTA (.fasta, .fna, .fa, .fas) file. In other words, it cannot have formatting as is the case with MS Word (.doc, .docx) or Rich Text Format (.rtf). It cannot be greater than in size. Don't forget to press the Upload button before attempting to submit your BLAST.

Choose Search Target

Choose from one of the protein BLAST databases listed below.

Protein BLAST Databases:

- Select a Dataset
  - All proteins
  - Arachis duranensis – proteins
  - Arachis ipaensis – proteins
  - Gene Family Consensus

... against target “Gene Family Consensus”
7. Gene families and synteny with other legumes

... leads to a matching gene family or families ...
... leads to a gene family tree view ...

note the tour
... and you can learn more about each gene or node ...

prefix and color indicate species
7. Gene families and synteny with other legumes

... for example, about related regions in other legume genomes ...

Query gene family
8. Website tours (and other help)

Site Overview and Guide

PeanutBase is organized around these features and tools:

- **Genome browsers:** currently including *A. duranensis* and *A. ipaensis*. From the browsers, you can do key-word searches (enter text under "Landmark or Region" in the browser), or go to regions that you know are of interest (for example, around genetic markers that you know about; edit the chromosome and coordinates under "Landmark or Region," or just scroll and zoom).

- **Traits and Maps:** PeanutBase maintains a relatively large (and growing) collection of published genetic maps. We are working to add QTL features to these when possible - and also to link from sequence-based genetic markers to the genome browsers. This allows a breeder to identify markers close to a trait of interest (the QTL) - and ultimately, will allow researchers to identify causative genes and perfect markers for some traits.

  - **Traits overview:** A major effort at PeanutBase is to collect published peanut QTL information and integrate it into a common genetic framework. The QTL data is loaded into PeanutBase and integrated with other genetic and genomic data. This data can be searched and viewed on genetic maps.
  
  - **Maps overview:** *Arachis* maps can be viewed as interactive maps, can be searched, and downloaded in CMap format below.

- **Marker Assisted Selection (MAS):** These are custom-compiled pages that summarize what is known about markers and breeding lines relevant to important traits such as rust, nematode resistance, and oil quality. These pages require expert knowledge. If you have knowledge to contribute about existing MAS pages or new ones (for traits not yet covered), please contact us. We are looking for new information to help flesh out these pages, and we would love to credit anyone who can provide important information.
8. Website tours (and other help)

- Website Tours:
  - Germplasm Map
  - QTL search to GBrowse

  More are in the works!

FAQ

GBrowse FAQ | Traits & Maps FAQ

- Where can I download *Arachis* gene sequences?
- How can I view my data *Gbrowse*?
- Can I download track data from *Gbrowse*?
- Is there citation information about tracks in *Gbrowse*?
- Can I find the coordinates of syntenic features in *Gbrowse*, or navigate to the corresponding regions?
QTL Tour: Welcome to PeanutBase!

Let's go the Peanutbase homepage, where the QTL Tour will begin. Use the Next button or → (right arrow key) to advance the tour.

Let's go the Peanutbase homepage, where the QTL Tour will begin. Use the Next button or → (right arrow key) to advance the tour.

**Beginning the tour**

- **Traits and Maps:** PeanutBase maintains a relatively large (and growing) collection of traits and genetic maps, working to add QTL features to these when possible - and also to link from sequence data and genomic browsers. This allows a breeder to identify markers close to a trait of interest, and helps researchers to identify causative genes and perfect markers for some traits.

(enter text under "Landmark or Region" in the browser), or go to regions that you know have genetic markers that you know about; edit the chromosome and coordinates under....
QTL Tour: Getting started

This tour will provide an example of navigating Peanutbase from the genetic location of a published QTL to the associated region of the annotated genome for the species in which it was found. Now press the QTL Search button, or use the Next button or press →.
Search Arachis QTL

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

Total QTL in PeanutBase: **232**

<table>
<thead>
<tr>
<th>Species</th>
<th>Trait Class</th>
<th>Trait</th>
<th>QTL Symbol in Publication</th>
<th>Trait Name in Publication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arachis spp.</td>
<td>whole-plant</td>
<td>Aboveground weight 1-1</td>
<td>TB_WW_a05</td>
<td>Total biomass</td>
</tr>
<tr>
<td>Arachis spp.</td>
<td>whole-plant</td>
<td>Aboveground weight 1-2</td>
<td>TB_WL_a05</td>
<td>Total biomass</td>
</tr>
</tbody>
</table>
Related resource - PeanutMine

- Provided by the Legume Federation project
Search

Search this mine. Enter names, identifiers or keywords. Examples: Aradu.000JC, Aradu.000JC.1, GM1069

e.g. X, Y, Z

Analyze

Enter a list of identifiers.

Gene

e.g. Aradu.000JC

Welcome back to the LIS PeanutMine!

This mine integrates many types of data for peanuts. It is currently under development by LIS/NCGR as a demo built from Peanutbase, via chado and GFF files, database. Contact Sam Hokin for more information.

Shared Motif Search: CArG1-containing upstream gene flanking regions

Latest updates (v0.6 25 October 2016)

- Flanking regions 5kb upstream and downstream of genes have been added.
- Introns have been added.
- Linkage group and genetic map report pages now show a graphical map with the locations of markers and QTLs on the linkage group(s).
- A new Shared Motif Search is now conducted in list analysis of gene flanking regions.
- Genomic region search has a new checkbox to enforce strand-specific searches. Minus strand regions are indicated by start>end a la BLAST.
- Gene.note (from chado.featureprop) has been deprecated in favor of Gene.description (from chado.gene).
- The expression heat map now shows for Gene lists in addition to mRNA/transcript lists.

The Legume Federation is funded by the

Participants include

NSF

National Center For Genome Resources

Participants include
9. **PeanutMine** (example: motifs common among a list of genes)

### List Analysis for CArG1 upstream gene flanking regions (8 Gene Flanking Regions)

<table>
<thead>
<tr>
<th>Gene Flanking Region</th>
<th>DB identifier</th>
<th>Gene Flanking Region Length</th>
<th>Gene Flanking Region Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aradu.GF3NG 5.0kb upstream</td>
<td>Aradu.GF3NG</td>
<td>5000</td>
<td>Aradu.GF3NG</td>
</tr>
<tr>
<td>Aradu.NA9YP 5.0kb upstream</td>
<td>Aradu.NA9YP</td>
<td>5000</td>
<td>Aradu.NA9YP</td>
</tr>
<tr>
<td>Aradu.XE5HD 5.0kb upstream</td>
<td>Aradu.XE5HD</td>
<td>5000</td>
<td>Aradu.XE5HD</td>
</tr>
<tr>
<td>Araip.KK6EC 5.0kb upstream</td>
<td>Araip.KK6EC</td>
<td>5000</td>
<td>Araip.KK6EC</td>
</tr>
<tr>
<td>Araip.KLZ01 5.0kb upstream</td>
<td>Araip.KLZ01</td>
<td>5000</td>
<td>Araip.KLZ01</td>
</tr>
<tr>
<td>Araip.KW08V 5.0kb upstream</td>
<td>Araip.KW08V</td>
<td>5000</td>
<td>Araip.KW08V</td>
</tr>
<tr>
<td>Araip.TUQ7N 5.0kb upstream</td>
<td>Araip.TUQ7N</td>
<td>5000</td>
<td>Araip.TUQ7N</td>
</tr>
<tr>
<td>Araip.W6805 5.0kb upstream</td>
<td>Araip.W6805</td>
<td>5000</td>
<td>Araip.W6805</td>
</tr>
</tbody>
</table>
9. **PeanutMine** (example: motifs common among a list of genes)

Description: These upstream flanking regions all contain much of the Arabidopsis CArG1 binding site motif. The Shared Motif Search shown below will find most of that motif, GTTTACATAATGGAAAA, shared to varying degrees by these regions, with a high score indicative of the length of the motif, the number of hits, and the CG content. It will also find many other sequences common to the regions that are less interesting and have lower scores. Had these regions been chosen for other reasons, we might look into whether this high-scoring shared motif is a transcription factor binding site.

Click to toggle Shared Motif Search

**Shared Motif Search**

7 motifs close to top scorer:

```
TTACATAAAATGGAAAA
```

<table>
<thead>
<tr>
<th>Motif</th>
<th>Length</th>
<th>Num</th>
<th>Score</th>
<th>Regions</th>
</tr>
</thead>
</table>

More information
In progress

- Continued curation of marker data
- Tetraploid genome assembly
- Continued curation of marker-trait information from literature
Future plans

• Bring in all publicly-available trait data for germplasm and link to stock records
• Align publicly-available genotype-phenotype data and provide results of population structure, LD, and GWAS analyses
• Provide more integrated access to data from other legumes
PeanutBase is a community resource

Let us know what you need
Contributors to PeanutBase

**Iowa State University/PeanutBase**
- **Ethy Cannon** – lead
- Wei Huang – computational biologist
- Paul Otyama – PhD student

**Iowa State University/Legume Federation**
- Jacqueline Campbell – data curator and outreach

**USDA-ARS at Ames, IA**
- **Steven Cannon** – lead scientist, ARS legume database project
- Nathan Weeks – IT specialist and computational biologist
- Scott Kalberer – data curator
- Andrew Wilkey – programmer

**National Center for Genomic Resources**
- **Andrew Farmer** – lead, Legume Information System
- Sudhansu Dash - geneticist and computational biologist
- Alex Rice - programmer