PeanutBase: New genome assemblies and breeding support

PeanutBase.org

Ethy Cannon
Iowa State University
July 12th, 2018

Talk Overview

PeanutBase is the community resource for *Arachis* reasearch and breeding.

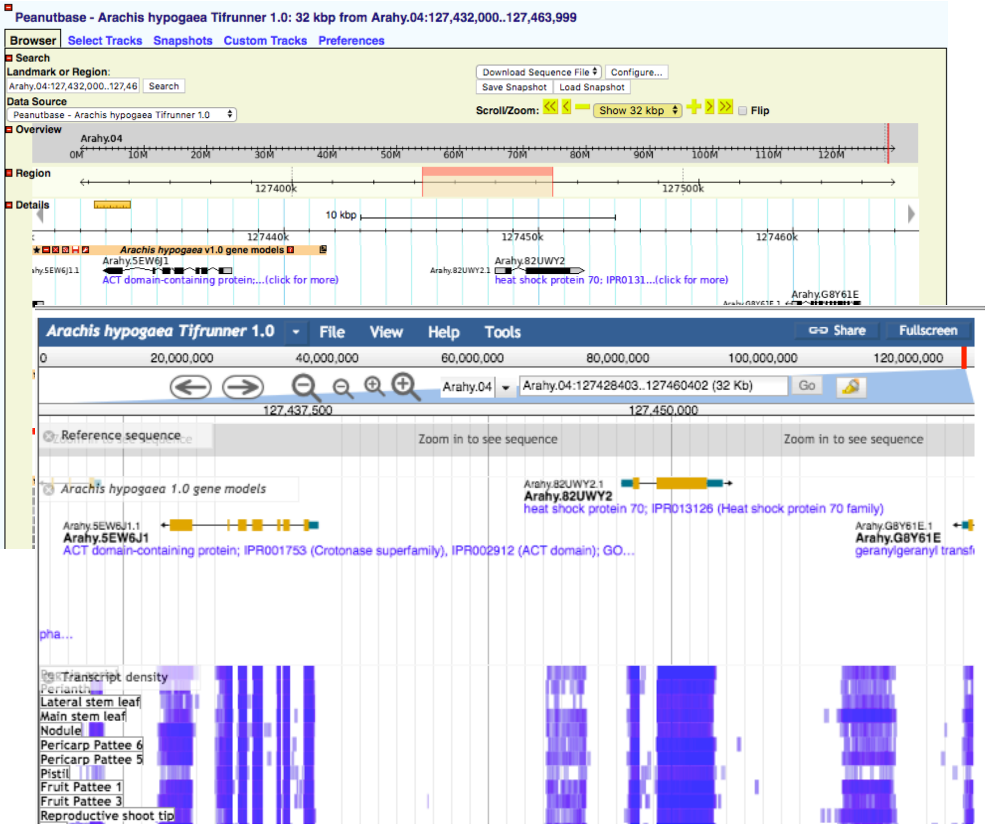
1. Brief overview of PeanutBase resources
2. Data management at PeanutBase

PeanutBase Resources

Genomic resources

- Genome assembly browsers
    - ***A. hypogaea*, Tifrunner***
    - *A. duranensis*
    - *A. ipaensis*

*Released in December, 2017 under Fort Lauderdale Agreement

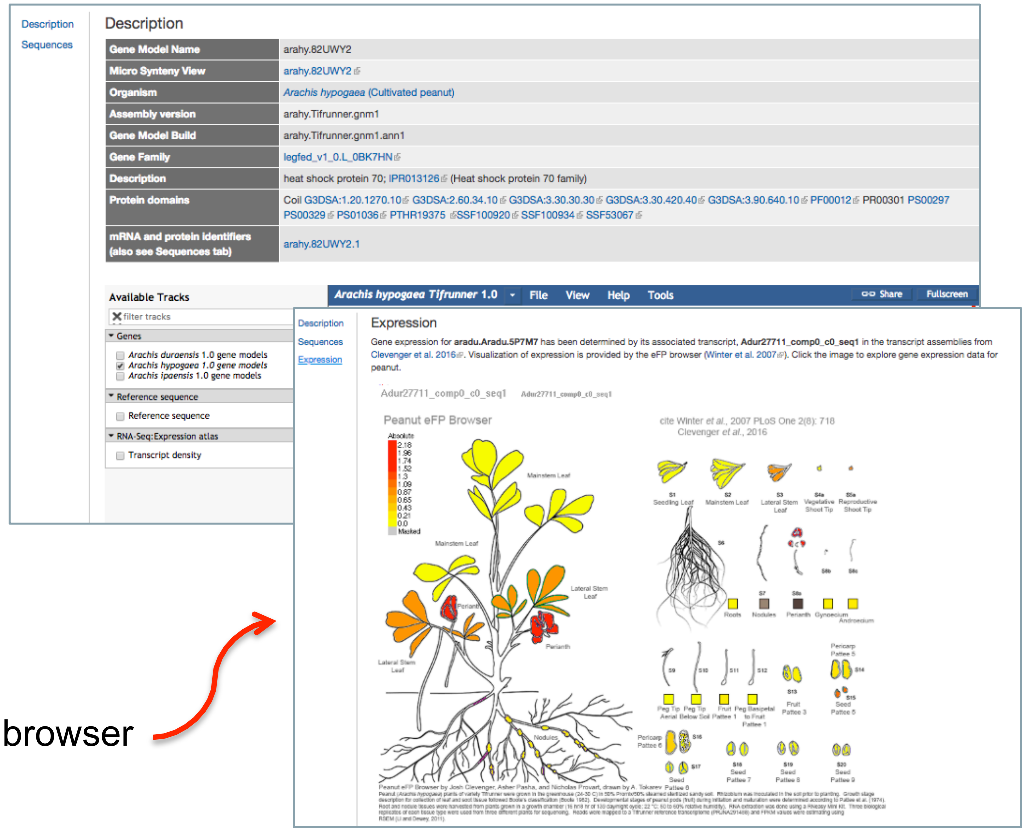


PeanutBase Resources

Genomic resources

- Genome assembly browsers
- Gene models
    - ***A. hypogaea* Tifrunner***
    - *A. duranensis*
    - *A. ipaensis*

*April, 2018; Jacqueline Campbell and Andrew Farmer



Description

Gene Model Name	arahy.82UWY2
Micro Synteny View	arahy.82UWY2
Organism	*Arachis hypogaea* (Cultivated peanut)
Assembly version	arahy.Tifrunner.gnm1
Gene Model Build	arahy.Tifrunner.gnm1.ann1
Gene Family	legfed_v1_0.L_0BK7HN
Description	heat shock protein 70; IPR013126 (Heat shock protein 70 family)
Protein domains	Coil G3DSA:1.20.1270.10 G3DSA:2.60.34.10 G3DSA:3.30.30.30 G3DSA:3.30.420.40 G3DSA:3.90.640.10 PF00012 PR00301 PS00297 PS00329 PS01036 PTHR19375 SSF100920 SSF100934 SSF53067
mRNA and protein identifiers (also see Sequences tab)	arahy.82UWY2.1

Expression

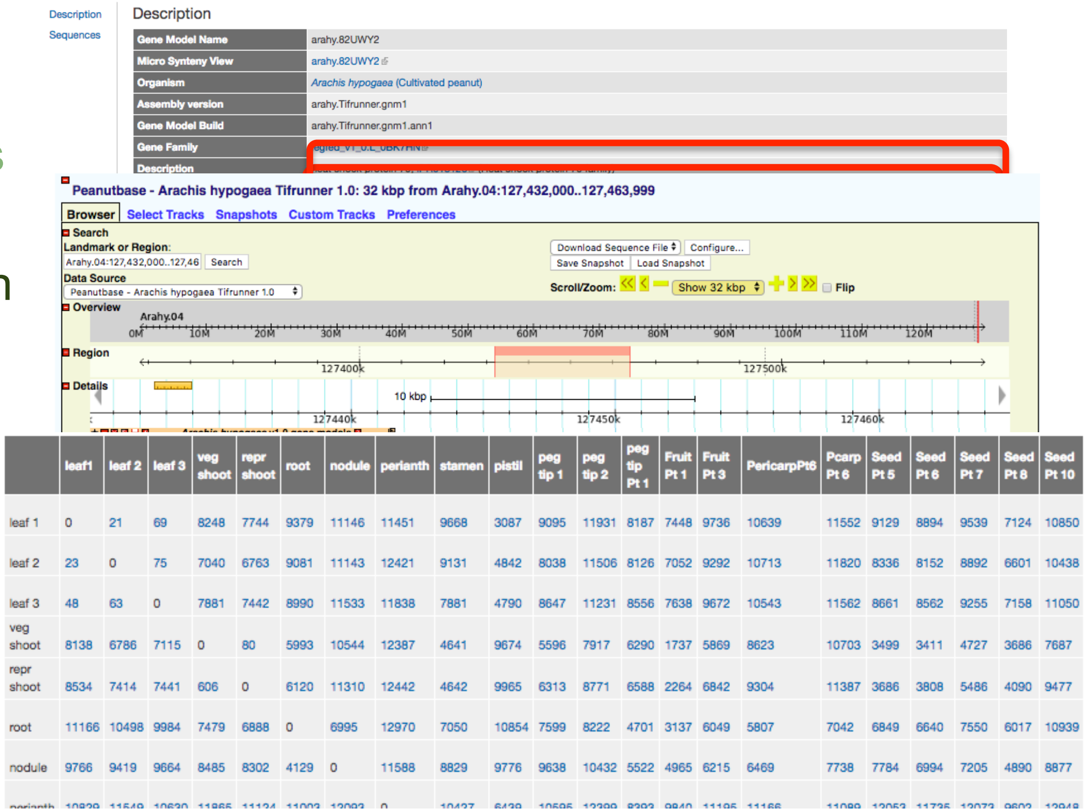
Gene expression for **aradu.Aradu.5P7M7** has been determined by its associated transcript, **Adur27711_comp0_c0_seq1** in the transcript assemblies from Clevenger et al. 2016. Visualization of expression is provided by the eFP browser (Winter et al. 2007). Click the image to explore gene expression data for peanut.

expression browser

PeanutBase Resources

Genomic resources

- Genome assembly browsers
- Gene models
    - Gene function prediction
    - Protein domains
    - Tissue expression atlas aligned on 3 genomes
    - 5 expression atlases



Description

Gene Model Name	arahy.82UWY2
Micro Synteny View	arahy.82UWY2
Organism	*Arachis hypogaea* (Cultivated peanut)
Assembly version	arahy.Tifrunner.gnm1
Gene Model Build	arahy.Tifrunner.gnm1.ann1
Gene Family	
Description	

Peanutbase - Arachis hypogaea Tifrunner 1.0: 32 kbp from Arahy.04:127,432,000..127,463,999

Landmark or Region: Arahy.04:127,432,000..127,46

Data Source: Peanutbase - Arachis hypogaea Tifrunner 1.0

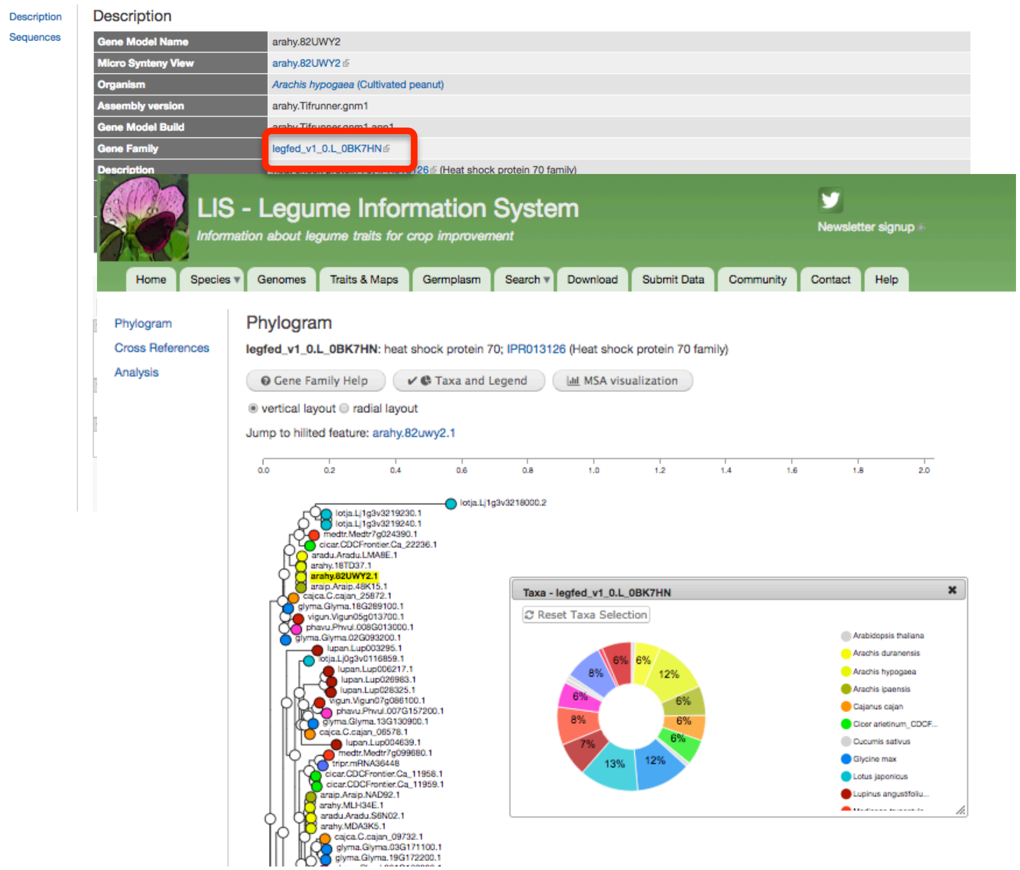
	leaf1	leaf 2	leaf 3	veg shoot	repr shoot	root	nodule	perianth	stamen	pistil	peg tip 1	peg tip 2	peg tip Pt 1	Fruit Pt 1	Fruit Pt 3	PericarpPt6	Pcarp Pt 6	Seed Pt 5	Seed Pt 6	Seed Pt 7	Seed Pt 8	Seed Pt 10
leaf 1	0	21	69	8248	7744	9379	11146	11451	9668	3087	9095	11931	8187	7448	9736	10639	11552	9129	8894	9539	7124	10850
leaf 2	23	0	75	7040	6763	9081	11143	12421	9131	4842	8038	11506	8126	7052	9292	10713	11820	8336	8152	8892	6601	10438
leaf 3	48	63	0	7881	7442	8990	11533	11838	7881	4790	8647	11231	8556	7638	9672	10543	11562	8661	8562	9255	7158	11050
veg shoot	8138	6786	7115	0	80	5993	10544	12387	4641	9674	5596	7917	6290	1737	5869	8623	10703	3499	3411	4727	3686	7687
repr shoot	8534	7414	7441	606	0	6120	11310	12442	4642	9965	6313	8771	6588	2264	6842	9304	11387	3686	3808	5486	4090	9477
root	11166	10498	9984	7479	6888	0	6995	12970	7050	10854	7599	8222	4701	3137	6049	5807	7042	6849	6640	7550	6017	10939
nodule	9766	9419	9664	8485	8302	4129	0	11588	8829	9776	9638	10432	5522	4965	6215	6469	7738	7784	6994	7205	4890	8877
perianth	10829	11549	10630	11865	11124	11003	12093	0	10427	6439	10595	12399	8393	9840	11195	11166	11089	12053	11735	12073	9602	12948

PeanutBase Resources

Genomic resources

- Genome assembly browsers
- Gene models
- Legume gene families*

*May, 2018; Steven Cannon and the Legume Federation



Description

Gene Model Name	arahy.82UWY2
Micro Synteny View	arahy.82UWY2
Organism	*Arachis hypogaea* (Cultivated peanut)
Assembly version	arahy.Tifrunner.gnm1
Gene Model Build	arahy.Tifrunner.gnm1.ann1
Gene Family	legfed_v1_0.L_0BK7HN
Description	26 (Heat shock protein 70 family)

LIS - Legume Information System

Information about legume traits for crop improvement

Phylogram

legfed_v1_0.L_0BK7HN: heat shock protein 70; IPR013126 (Heat shock protein 70 family)

● vertical layout ○ radial layout

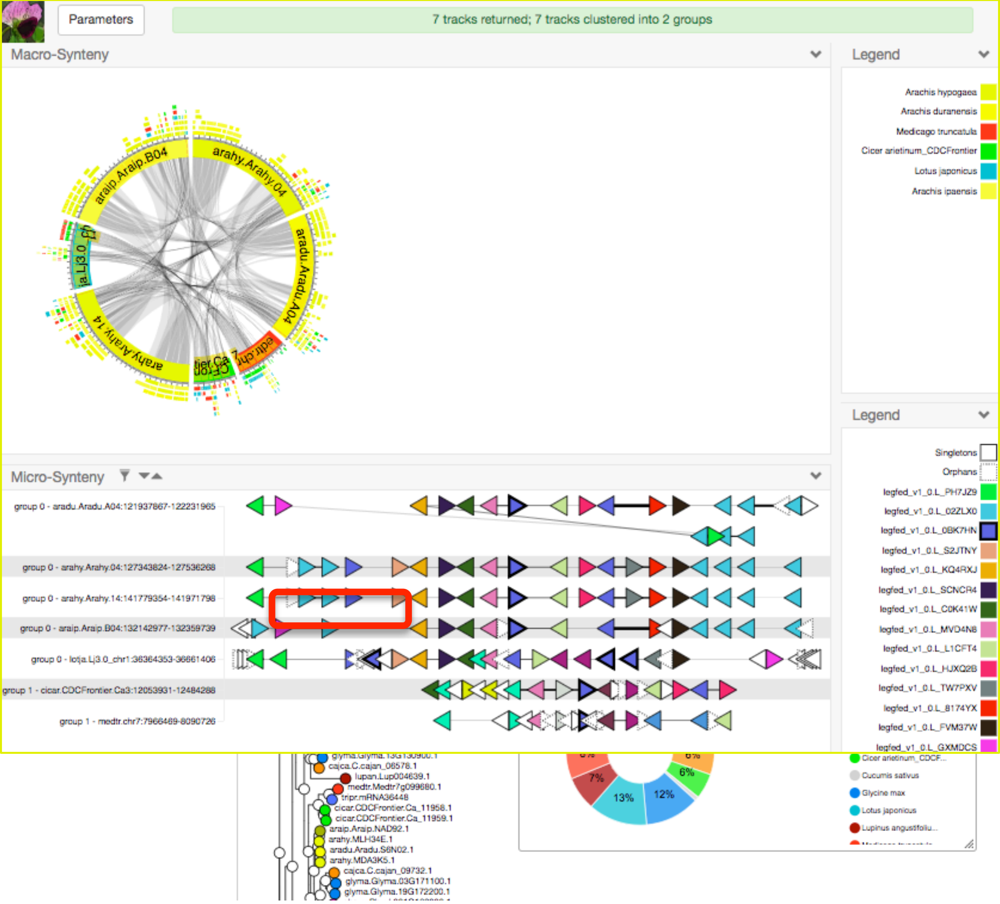
Jump to hilited feature: arahy.82uwy2.1

PeanutBase Resources

Genomic resources

- Genome assembly browsers
- Gene models
- Legume gene families
- Synteny*

* Alan Cleary



7 tracks returned; 7 tracks clustered into 2 groups

PeanutBase Resources

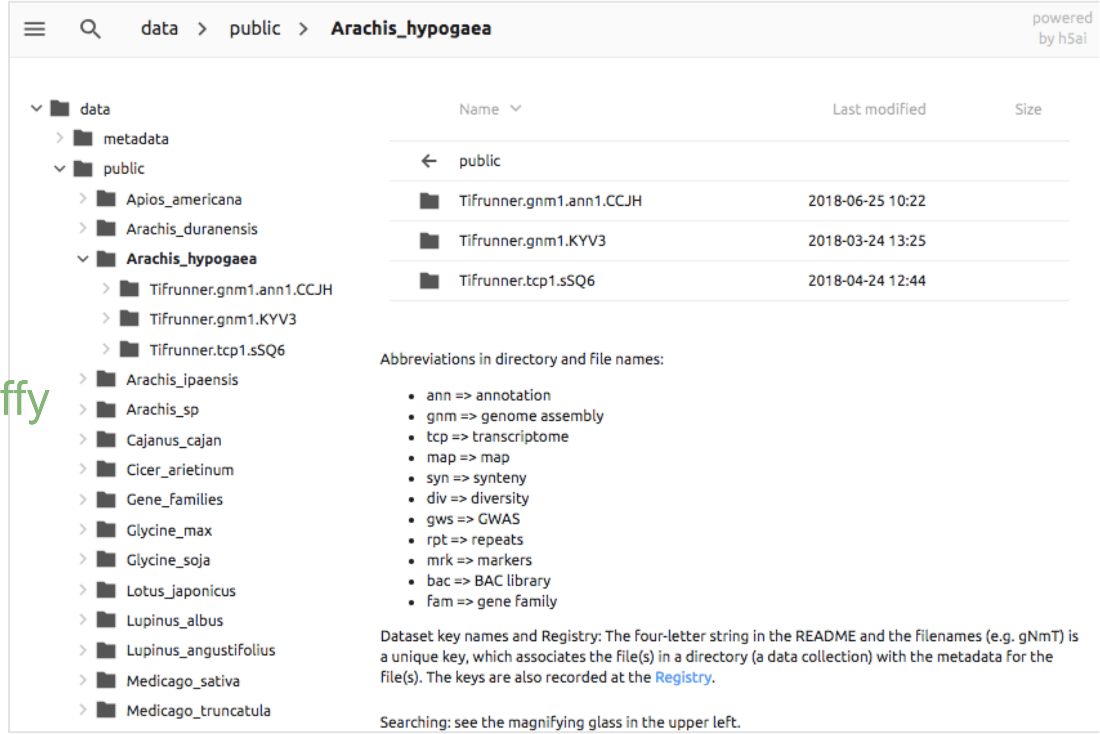
Genomic resources

- Genome assembly browsers
- Gene models
- Legume gene families
- Synteny
- SNP markers, including the Affy SNP chip markers.
  - Aligned on assemblies
  - Searchable in database
  - Download files available.

PeanutBase Resources

Genomic resources

- Genome assembly browsers
- Gene models
- Legume gene families
- Synteny
- SNP markers, including the Affy SNP chip markers
- Full dataset downloads



- data
  - metadata
  - public
    - Apios_americana
    - Arachis_duranensis
    - **Arachis_hypogaea**
      - Tifrunner.gnm1.ann1.CCJH
      - Tifrunner.gnm1.KYV3
      - Tifrunner.tcp1.sSQ6
    - Arachis_ipaensis
    - Arachis_sp
    - Cajanus_cajan
    - Cicer_arietinum
    - Gene_families
    - Glycine_max
    - Glycine_soja
    - Lotus_japonicus
    - Lupinus_albus
    - Lupinus_angustifolius
    - Medicago_sativa
    - Medicago_truncatula

Name	Last modified	Size
public		
Tifrunner.gnm1.ann1.CCJH	2018-06-25 10:22	
Tifrunner.gnm1.KYV3	2018-03-24 13:25	
Tifrunner.tcp1.sSQ6	2018-04-24 12:44	

Abbreviations in directory and file names:

- ann => annotation
- gnm => genome assembly
- tcp => transcriptome
- map => map
- syn => synteny
- div => diversity
- gws => GWAS
- rpt => repeats
- mrk => markers
- bac => BAC library
- fam => gene family

Dataset key names and Registry: The four-letter string in the README and the filenames (e.g. gNmT) is a unique key, which associates the file(s) in a directory (a data collection) with the metadata for the file(s). The keys are also recorded at the Registry.

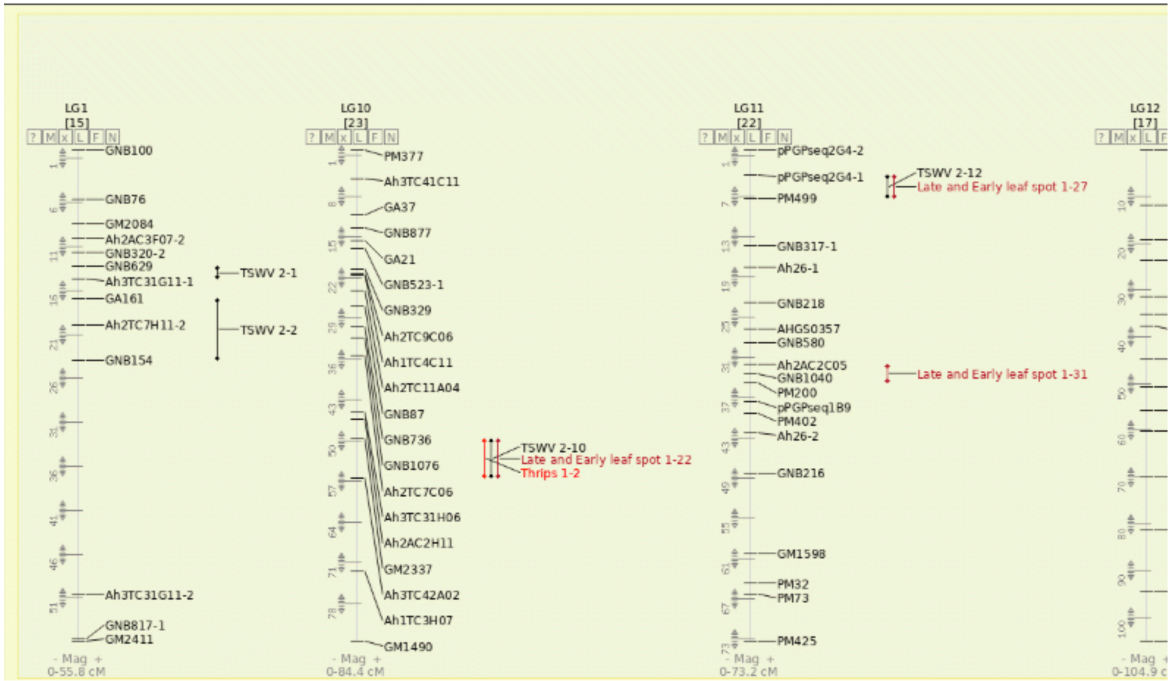
Searching: see the magnifying glass in the upper left.

PeanutBase Resources

Genetic resources

- Genetic maps collected from literature

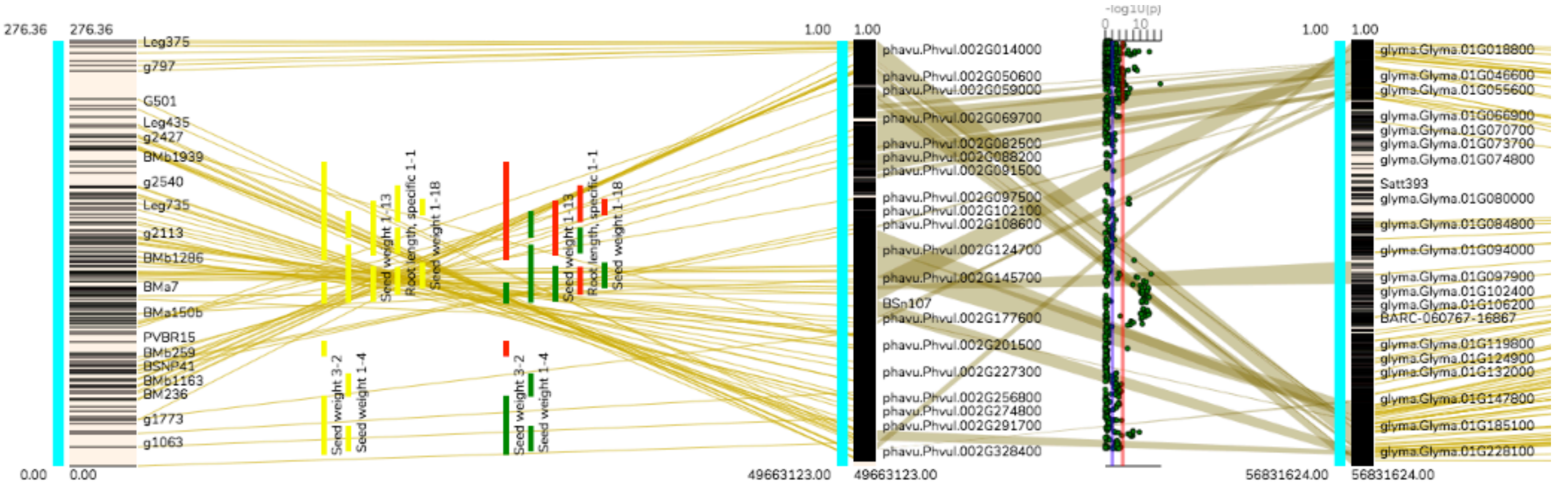
cMap
(last updated in 2008)



PeanutBase Resources

Genetic resources

- Genetic maps collected from literature
- Cmap-js, a new comparative map viewer ... is coming soon.*

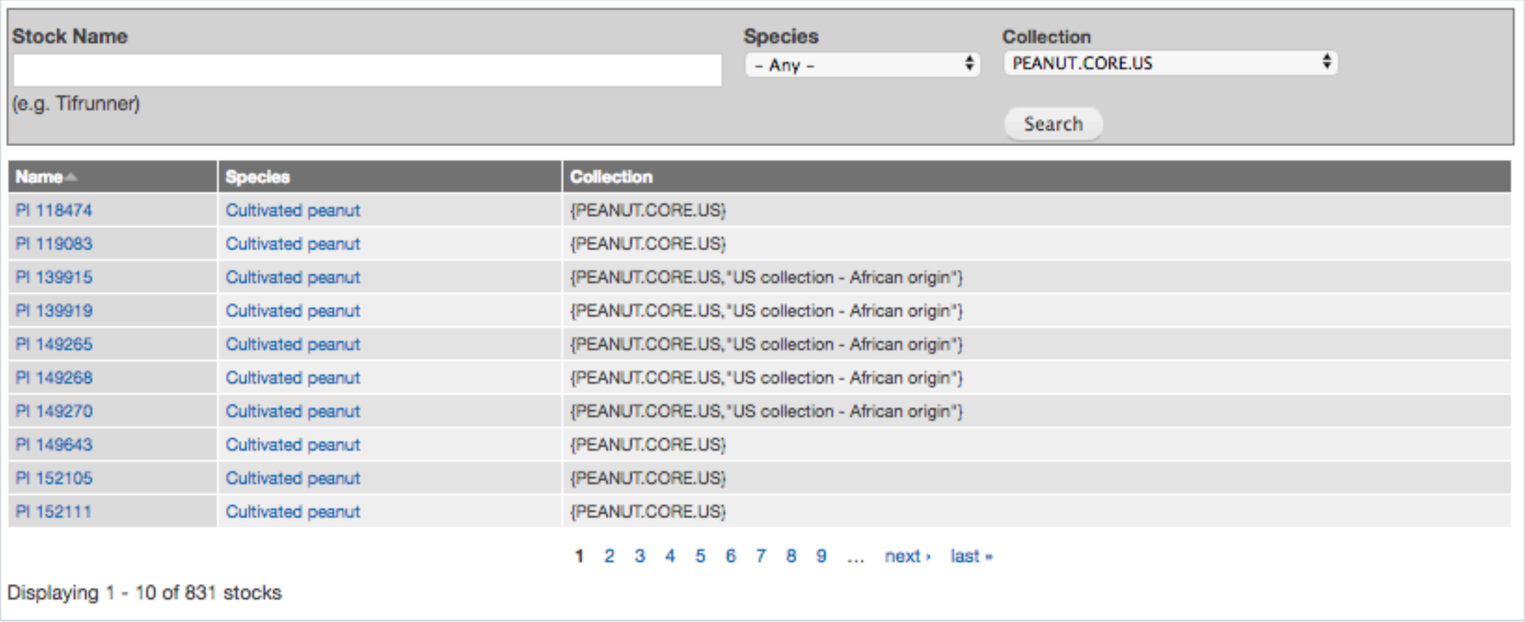


*Andrew Wilkey

PeanutBase Resources

Germplasm data

- All GRIN germplasm and trait data mirrored at PeanutBase



Stock Name

(e.g. Tifrunner)

Species: – Any –

Collection: PEANUT.CORE.US

Name	Species	Collection
PI 118474	Cultivated peanut	{PEANUT.CORE.US}
PI 119083	Cultivated peanut	{PEANUT.CORE.US}
PI 139915	Cultivated peanut	{PEANUT.CORE.US,"US collection - African origin"}
PI 139919	Cultivated peanut	{PEANUT.CORE.US,"US collection - African origin"}
PI 149265	Cultivated peanut	{PEANUT.CORE.US,"US collection - African origin"}
PI 149268	Cultivated peanut	{PEANUT.CORE.US,"US collection - African origin"}
PI 149270	Cultivated peanut	{PEANUT.CORE.US,"US collection - African origin"}
PI 149643	Cultivated peanut	{PEANUT.CORE.US}
PI 152105	Cultivated peanut	{PEANUT.CORE.US}
PI 152111	Cultivated peanut	{PEANUT.CORE.US}

Displaying 1 - 10 of 831 stocks

PeanutBase Resources

Germplasm data

- All GRIN germplasm and trait data mirrored at PeanutBase
- GIS view of germplasm origins

Leaflet | Data: USDA GRIN/NPGS, Tiles © Esri — National Geographic, Esri, DeLorme, NAVTEQ, UNEP-WCMC, USGS, NASA, ESA, METI, NRCAN, GEBCO, NOAA, iPC

Results: 200 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 parameters, or increase the max # of results.

accession	taxon	crop name	orig country		collection site	acq date	elev (m.)
PI 475871	● Arachis hypogaea var. hypogaea		BRA		Farm, Guajara-Mirim	1981-12-01	140
PI 590516	● Arachis hypogaea var. fastigiata	Valencia peanut	BRA		Cruizeiro do Sul. Cultivated upriver in ...	1993-12-09	300
[illegible]	● Arachis hypogaea var. fastigiata	Valencia peanut	BRA		Cruzeiro do Sul. Cultivated upriver in t	1993-12-09	300

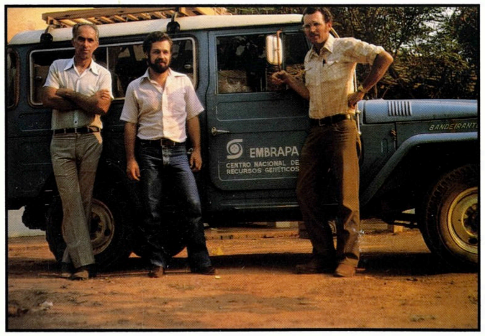
PeanutBase Resources

Germplasm data

- All GRIN germplasm and trait data mirrored at PeanutBase
- GIS view of germplasm origins
- Traits from multiple studies attached to germplasm

MP-1737

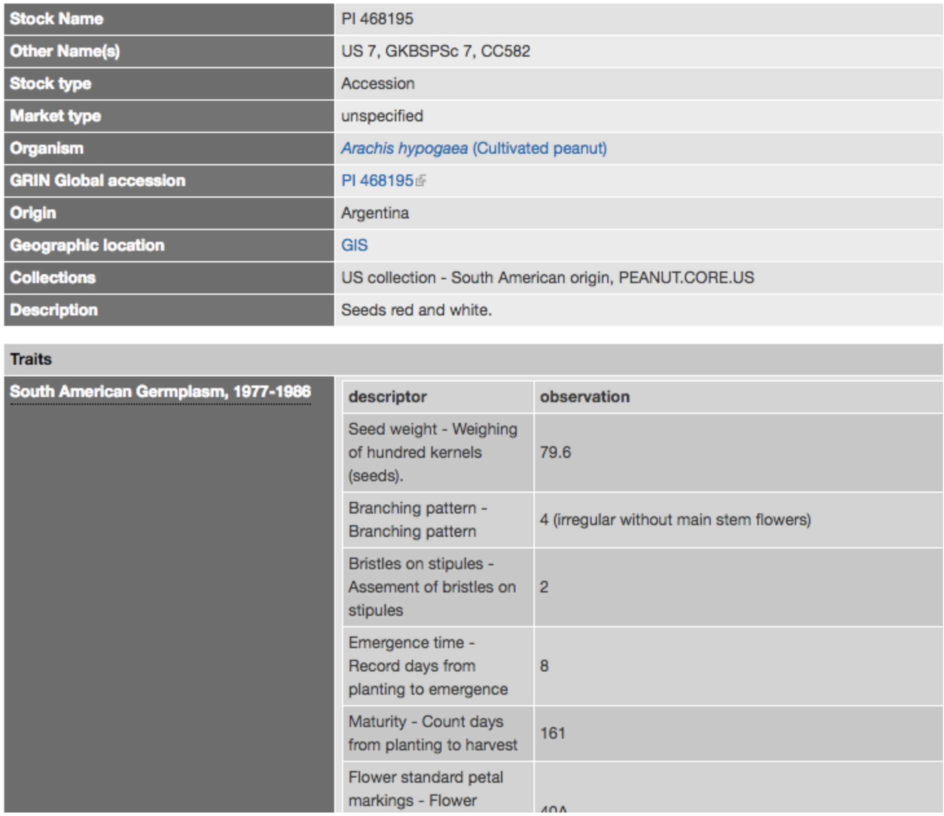
Catalog of Passport Data and Minimum Descriptors of *Arachis hypogaea* L. Germplasm Collected in South America, 1977-1986



THE TEXAS AGRICULTURAL EXPERIMENT STATION • J. Charles Lee, Interim Director • The Texas A&M University System • College Station, Texas

PeanutBase Resources

Traits collected by Charles Simpson and colleagues in 1980s



Stock Name	PI 468195
Other Name(s)	US 7, GKBSPSc 7, CC582
Stock type	Accession
Market type	unspecified
Organism	*Arachis hypogaea* (Cultivated peanut)
GRIN Global accession	PI 468195
Origin	Argentina
Geographic location	GIS
Collections	US collection - South American origin, PEANUT.CORE.US
Description	Seeds red and white.

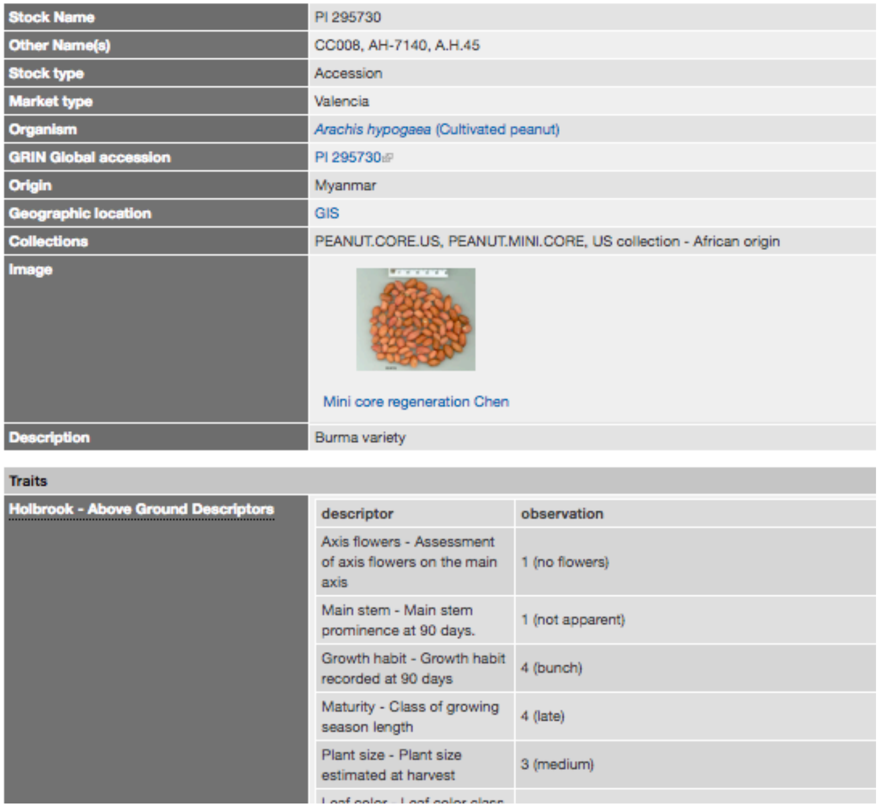
Traits

South American Germplasm, 1977-1986

descriptor	observation
Seed weight - Weighing of hundred kernels (seeds).	79.6
Branching pattern - Branching pattern	4 (irregular without main stem flowers)
Bristles on stipules - Assement of bristles on stipules	2
Emergence time - Record days from planting to emergence	8
Maturity - Count days from planting to harvest	161
Flower standard petal markings - Flower	40A

PeanutBase Resources

Mini core traits collected by Corley Holbrook in the 1990s



Stock Name	PI 295730
Other Name(s)	CC008, AH-7140, A.H.45
Stock type	Accession
Market type	Valencia
Organism	*Arachis hypogaea* (Cultivated peanut)
GRIN Global accession	PI 295730
Origin	Myanmar
Geographic location	GIS
Collections	PEANUT.CORE.US, PEANUT.MINI.CORE, US collection - African origin
Image	Mini core regeneration Chen
Description	Burma variety

Traits

Holbrook - Above Ground Descriptors

descriptor	observation
Axis flowers - Assessment of axis flowers on the main axis	1 (no flowers)
Main stem - Main stem prominence at 90 days.	1 (not apparent)
Growth habit - Growth habit recorded at 90 days	4 (bunch)
Maturity - Class of growing season length	4 (late)
Plant size - Plant size estimated at harvest	3 (medium)
Leaf color - Leaf color class	

PeanutBase Resources

Traits and breeding resources

- Collaborating with the Integrated Breeding Platform group.
- Working to improve our collection of QTL and marker-trait data from literature.

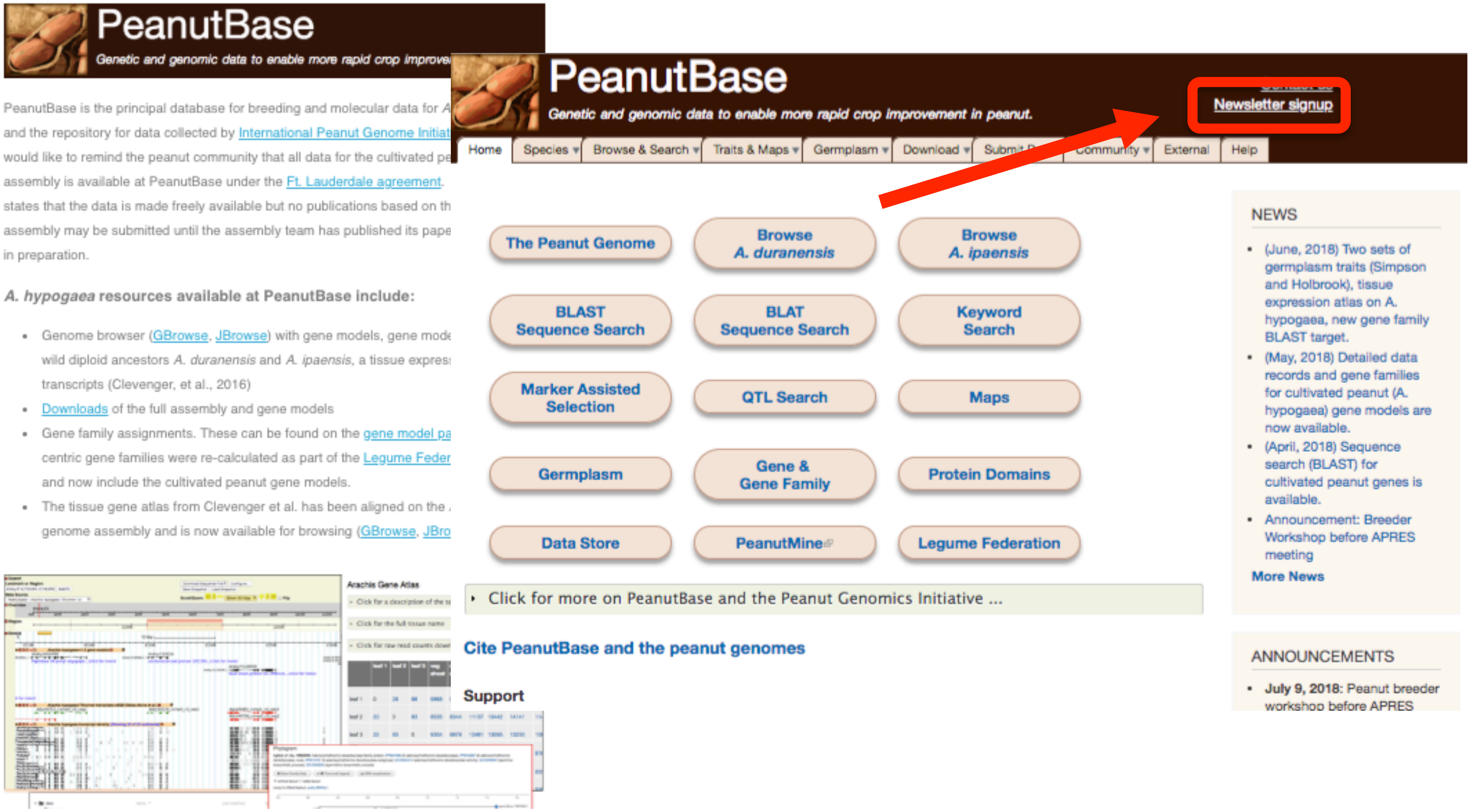
PeanutBase Resources

Community resources

- Information and archives for the Peanut Genome Initiative.
- Information and archives for Advances in Arachis Through Genomics and Biotechnology.
- List of upcoming conferences, workshops and meetings.
- Quarterly (or so) newsletter.

PeanutBase Resources

July Newsletter



PeanutBase

Genetic and genomic data to enable more rapid crop improve

PeanutBase is the principal database for breeding and molecular data for A
and the repository for data collected by International Peanut Genome Initiat
would like to remind the peanut community that all data for the cultivated pe
assembly is available at PeanutBase under the Ft. Lauderdale agreement.
states that the data is made freely available but no publications based on th
assembly may be submitted until the assembly team has published its pape
in preparation.

***A. hypogaea* resources available at PeanutBase include:**

- Genome browser (GBrowse, JBrowse) with gene models, gene mode
  wild diploid ancestors *A. duranensis* and *A. ipaensis*, a tissue expres
  transcripts (Clevenger, et al., 2016)
- Downloads of the full assembly and gene models
- Gene family assignments. These can be found on the gene model pa
  centric gene families were re-calculated as part of the Legume Feder
  and now include the cultivated peanut gene models.
- The tissue gene atlas from Clevenger et al. has been aligned on the
  genome assembly and is now available for browsing (GBrowse, JBro

Arachis Gene Atlas

PeanutBase

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

Newsletter signup

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

Cite PeanutBase and the peanut genomes

Support

NEWS

- (June, 2018) Two sets of germplasm traits (Simpson and Holbrook), tissue expression atlas on A. hypogaea, new gene family BLAST target.
- (May, 2018) Detailed data records and gene families for cultivated peanut (A. hypogaea) gene models are now available.
- (April, 2018) Sequence search (BLAST) for cultivated peanut genes is available.
- Announcement: Breeder Workshop before APRES meeting

More News

ANNOUNCEMENTS

- **July 9, 2018:** Peanut breeder workshop before APRES

Data Management at PeanutBase

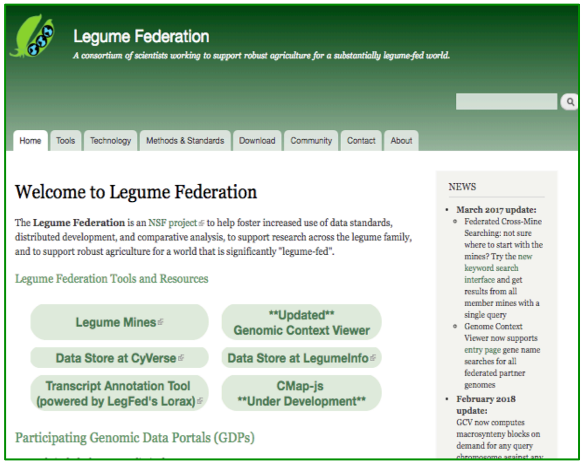
Why talk about data?

- Data is the life blood of research.
- Data is growing in size and complexity in all scientific disciplines.
- Sharing public data is important and growing more so.
- Funding agencies are alarmed at the growing costs of managing all this data.

Data Management at PeanutBase

PeanutBase is active in the data management community

Legume Federation



Legume Federation

A consortium of scientists working to support robust agriculture for a substantially legume-fed world.

Welcome to Legume Federation

The **Legume Federation** is an NSF project to help foster increased use of 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".

Legume Federation Tools and Resources

Participating Genomic Data Portals (GDPs)

NEWS

- **March 2017 update:**
  - Federated Cross-Mine Searching: not sure where to start with the mines? Try the new keyword search interface and get results from all member mines with a single query
  - Genome Context Viewer now supports entry page gene name searches for all federated partner genomes
- **February 2018 update:** GCV now computes macrosynteny blocks on demand for any query

AgBioData consortium

AgBioData

Toward enhanced genomics, genetics, and breeding research outcomes through standardization of practices and protocols across agricultural databases

Welcome to AgBioData

AgBioData is a consortium of agricultural biological databases and associated resources working together to ensure standards and best practices for acquisition, display and retrieval of genomic, genetic and breeding data.

Participating Databases and Resources

News & Events

July AgBioData Conference Call
Tue, 06/19/2018 - 16:15
No conference call due to 4th of July holiday.

Standards working groups

- Nomenclature
- Genomic metadata
- Trait terms

The FAIR data principles

Findable**:** able to locate public data of use to you
Accessible: able to easily access data in a form that works
Interoperable: able to use data across multiple resources
Reusable: able to reuse existing public data for new research

How the FAIR data principles help researchers

Findable**:**

- You can find data relevant to your research

Accessible:

Interoperable:

Reusable:

Data Management at PeanutBase

How researchers can help meet the FAIR data principles

Findable**:**

- Put your public data in appropriate public repositories. PeanutBase can help.
- Provide good metadata for your datasets, especially, describe each data field
- Use controlled vocabularies for terms (ontologies)

Accessible:

Interoperable:

Reusable:

Data Management at PeanutBase

PeanutBase and the FAIR data principles

Findable**:**

- Search pages for data types
- Consolidate data from multiple papers into one location

Accessible:

Interoperable:

Reusable:

How the FAIR data principles help researchers

Findable**:**

Accessible:

- You can acquire the data you need in a useful manner

Interoperable:

Reusable:

Data Management at PeanutBase

How researchers can help meet the FAIR data principles

Findable**:**

Accessible:

- Share your public data through long term repostories. Supplemental files with publications are not always sufficient.

Interoperable:

Reusable:

Data Management at PeanutBase

PeanutBase and the FAIR data principles

Findable**:**

Accessible:

- Browsers (genome, geographic location, maps, synteny, expression)
- Full dataset downloads: the Legume Federation Data Store

Interoperable:

Reusable:

Data Management at PeanutBase

How the FAIR data principles help researchers

Findable**:**

Accessible:

Interoperable:
- Data can be used in multiple tools and analyses
- Multiple data types are integrated for a more complete picture

Reusable:

Data Management at PeanutBase

How researchers can help meet the FAIR data principles

Findable**:**

Accessible:

Interoperable:

- Share your data in standard formats
- Use accepted controlled vocabularies

Reusable:

Data Management at PeanutBase

PeanutBase and the FAIR data principles

Findable**:**

Accessible:

Interoperable:

- Common vocabulary enables integration and sharing across resources
- Standard metadata allows datasets to be compared
- Data is integrated
    - gene models integrated with genome assemblies, gene function, gene expression, gene families
    - germplasm is integrated with trait data, assemblies

Reusable:

How the FAIR data principles help researchers

Findable**:**

Accessible:

Interoperable:

Reusable:

- Data from earlier studies can be reused with confidence.
- Your data is credited with furthering research

How researchers can help meet the FAIR data principles

Findable**:**
Accessible:
Interoperable:
Reusable:

- Provide enough information with your data to enable correct re-use...and proper citation.

Data Management at PeanutBase

PeanutBase and the FAIR data principles

Findable**:**
Accessible:
Interoperable:
Reusable:

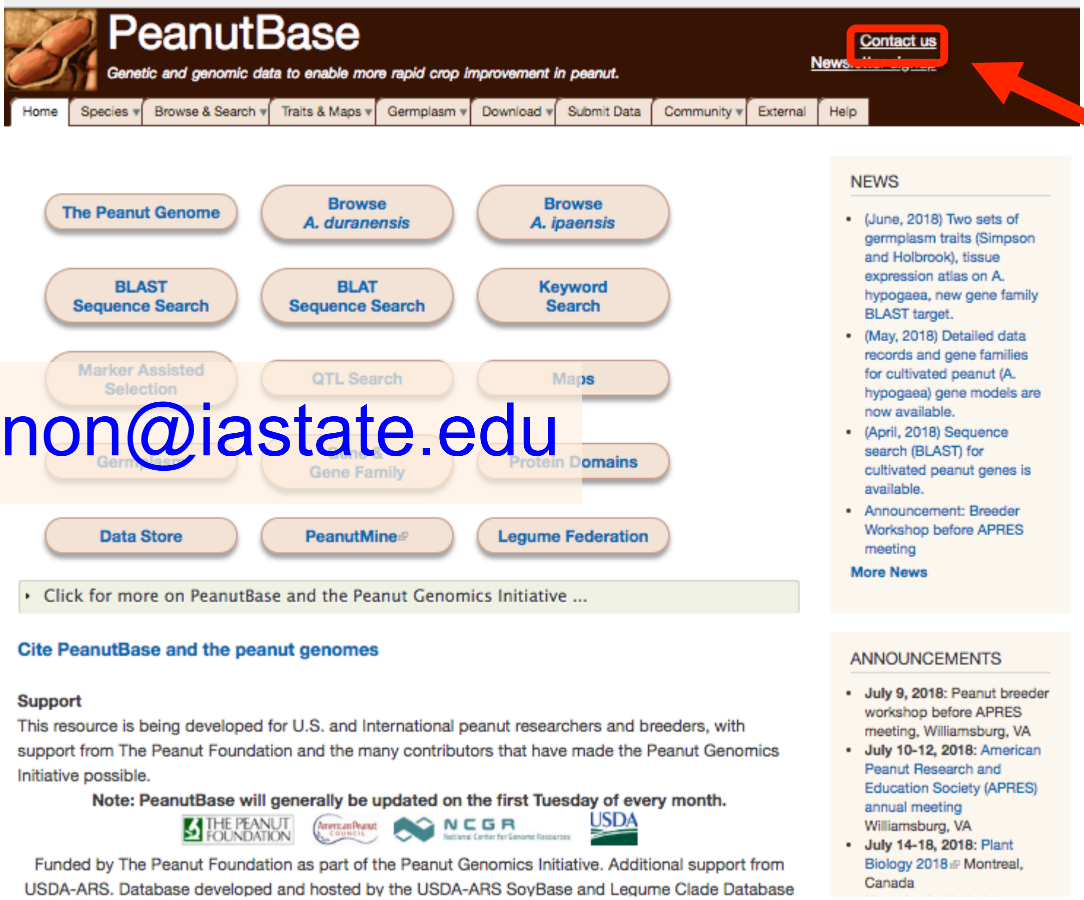
- All public data at PeanutBase can be re-used freely, except where noted otherwise
- Metadata describes data and how it was generated
- (Please cite original source and PeanutBase!)

Data Management at PeanutBase

Talk with us about your data early in your research projects, and even while writing proposals. We can help with your data management plan.

ekcannon@iastate.edu

Data Management at PeanutBase



NEWS

- (June, 2018) Two sets of germplasm traits (Simpson and Holbrook), tissue expression atlas on A. hypogaea, new gene family BLAST target.
- (May, 2018) Detailed data records and gene families for cultivated peanut (A. hypogaea) gene models are now available.
- (April, 2018) Sequence search (BLAST) for cultivated peanut genes is available.
- Announcement: Breeder Workshop before APRES meeting

More News

ANNOUNCEMENTS

- **July 9, 2018:** Peanut breeder workshop before APRES meeting, Williamsburg, VA
- **July 10-12, 2018:** American Peanut Research and Education Society (APRES) annual meeting Williamsburg, VA
- **July 14-18, 2018:** Plant Biology 2018 Montreal, Canada

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

Cite PeanutBase and the peanut genomes

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.

Note: PeanutBase will generally be updated on the first Tuesday of every month.

Funded by The Peanut Foundation as part of the Peanut Genomics Initiative. Additional support from USDA-ARS. Database developed and hosted by the USDA-ARS SoyBase and Legume Clade Database

ekcannon@iastate.edu

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

Malachy O'Connell – programmer

ekcannon@iastate.edu

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