CottonGen

Development of an integrated Genomics, Genetics and Breeding Database for the Cotton Research Community

Jing Yu[1], Sook Jung[1], Chun-Huai Cheng[1], Stephen Ficklin[1], Taein Lee[1], Ping Zheng[1], Don Jones[2], Richard Percy[3], Dorrie Main[1]

1. Washington State University, 2. Cotton Incorporated, 3. USDA-ARS

2012 ICGI Conference Raleigh, North Carolina

Topics

- Introduction
  - What is CottonGen
  - Infrastructure
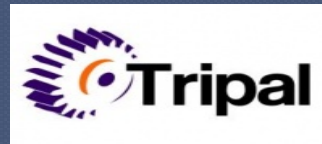
- Demo of CottonGen
  - Genome

- Future Work

What is CottonGen?

- A new cotton community database to further enable basic, translational and applied cotton research.

- Built using the new open-source Tripal database infrastructure used by several other databases

- Consolidate and expand CottonDB and CMD to include transcriptome, genome sequence and breeding data

- CottonDB still available for easy transition

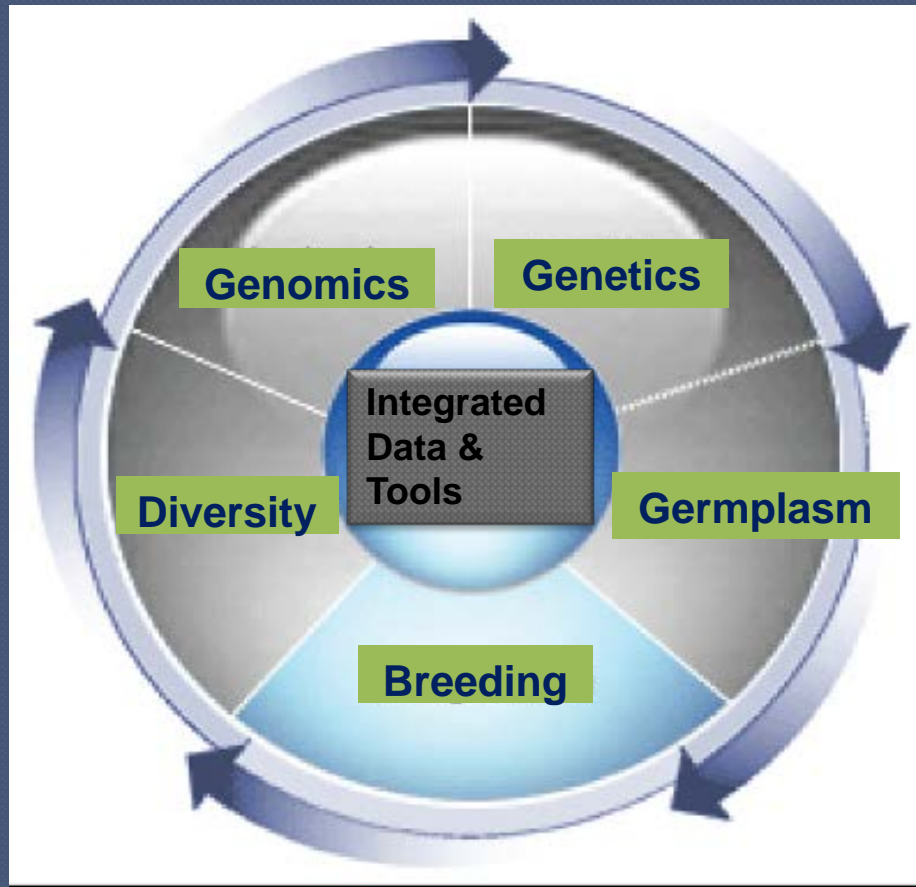# CottonGen Structure

Content Management System

Drupal modules as web front-end for Chado

Generic Database schema

Integrated Data Facilitates Discovery

Basic Science

Structure and evolution of genome, gene function, genetic variability, mechanism underlying traits

Translational Science

QTL /marker discovery, genetic mapping, Breeding values

Applied Science

Utilization of DNA information in breeding decisions

CottonGen Homepage

Gene/Sequence Search

Gene/Sequence Page

Species Page

Gossypium spp.

Gossypium spp.

Germplasm

The following stocks are available for this organism.

Gossypium spp.

Data Type Summary

The following data types are currently present for this organism

Feature Type	Count
genetic_marker	9,169
primer	7,224
marker_locus	6,684
linkage_group	908
sequence_feature	315
QTL	213
heritable_phenotypic_marker	166
chromosome_arm	3

Resources

- Species Details
- Maps

Resources

- Species Details
- Maps
- Germplasm
- Data Type Summary

Gene/Sequence Search cont.

43 records were returned.   Download the results

Gossypium unigene v1.0

Unigene Details

Analysis Name	Gossypium unigene v1.0
Unigene Name	Gossypium unigene v1.0
Software	CAP3
Source	Genbank Gossypium ESTs (September 16, 2012)
Date constructed	2012-09-27
Materials & Methods	This is the first version of the Gossypium unigene. This build was used many sequencing projects around the world are depositing ESTs from the genus Gossypium in the NCBI dbEST database. The Gossypium ESTs included in this assembly were downloaded on September 16, 2012.

This is the first version of the Gossypium unigene. This build was used many sequencing projects around the world are depositing ESTs from the genus Gossypium in the NCBI dbEST database. The Gossypium ESTs included in this assembly were downloaded on September 16, 2012.

Not all of the Gossypium ESTs are of high quality. To filter, we crossmatched the public sequences against NCBI's UniVec database and used the BLAST sequence similarity algorithm to remove species-specific chloroplast, mitochondrial, tRNA, and rRNA sequences. To reduce redundancy and create longer transcripts we assembled these ESTs using the CAP3[1] program. The final assembly has been annotated by BLAST sequence similarity searching agaist Swiss-Prot[2],TrEMBL[3],TAIR[4]Arabidopsis proteinsPrunus persica[5],Populus trichocarpa[6] and Vitis vinifera[7].

For more information on this project please contact the Cottongen development team.

Resources

- Search Gossypium ESTs
- Search Gossypium unigene v1.0 Contigs
- BLAST Gossypium unigene v1.0
- Unigene Details
- Library Information
- Homology
- KEGG
- Microsatellite Analysis
- Contact
- Downloads

Processing Summary	
Number of ESTs available	442,954
Number of ESTs available after filtering	437,185

Gene/Sequence Search cont.

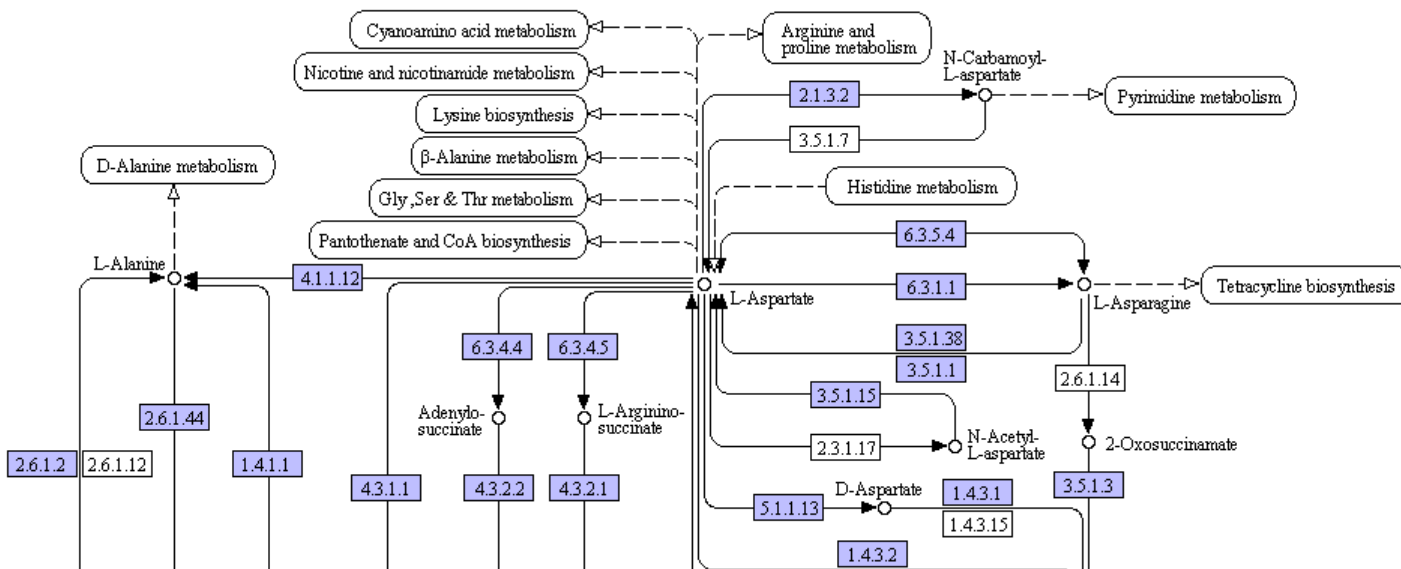**KEGG** Alanine, aspartate and glutamate metabolism - Reference pathway (KO)

[Pathway menu | Organism menu | Pathway entry | Download KGML | User data mapping]

Reference pathway (KO) ▼ | Go | 100% ▼ | |

Help

ALANINE, ASPARTATE AND GLUTAMATE METABOLISM

DVI: Dvir_GJ12156
AGA: AgaP_AGAP006360
AAG: AaeL_AAEL014768
CQU: CpipJ_CPIJ007345
AME: 413372

Return t

Marker Search

Search Markers on Nearby Loci

Locus [   ] Distance [   ] cM [Submit]

3501 records were returned

Locus	Map	Linkage Group	Position	Neighbor	Position
A1046	A1-97 x A2-47, F2 (1999)	A1-97 x A2-47, F2 (1999).AA-F2_A04	22	A1102	34
A1046	A1-97 x A2-47, F2 (1999)	A1-97 x A2-47, F2 (1999).AA-F2_A04	22	G1008	8.5
A1046	A1-97 x A2-47, F2 (1999)	A1-97 x A2-47, F2 (1999).AA-F2_A04	22	A1834	34
A1046	A1-97 x A2-47, F2 (1999)	A1-97 x A2-47, F2 (1999).AA-F2_A04	22	A1417	16
A1046	A1-97 x A2-47, F2 (1999)	A1-97 x A2-47, F2 (1999).AA-F2_A04	22	A1826	11
A1046	A1-97 x A2-47, F2 (1999)	A1-97 x A2-47, F2 (1999).AA-F2_A04	22	A1808	52
A1046	A1-97 x A2-47, F2 (1999)	A1-97 x A2-47, F2 (1999).AA-F2_A04	22	A1794	33
A1046	A1-97 x A2-47, F2 (1999)	A1-97 x A2-47, F2 (1999).AA-F2_A04	22	G1140	31
A1046	A1-97 x A2-47, F2 (1999)	A1-97 x A2-47, F2 (1999).AA-F2_A04	22	A1737	11
A1046	A1-97 x A2-47, F2 (1999)	A1-97 x A2-47, F2 (1999).AA-F2_A04	22	A1720b	35
A1046	A1-97 x A2-47, F2 (1999)	A1-97 x A2-47, F2 (1999).AA-F2_A04	22	A1099	17.5

BNL0137	SSR	BNL0137	genomic	G.h.fbr-sw	Deltapine 90	Gossypium hirsutum
BNL0140	SSR	BNL0140	genomic	G.h.fbr-sw	Deltapine 90	Gossypium hirsutum
BNL0148	SSR	BNL0148	genomic	G.h.fbr-sw	Deltapine 90	Gossypium hirsutum
BNL0150	SSR	BNL0150	genomic	G.h.fbr-sw	Deltapine 90	Gossypium hirsutum

A1097	RFLP	A1097	Gtr x Gra, F2 (2004)	Gtr x Gra, F2 (2004).D_D02	D_chr.02	45.8	45.8
A1097	RFLP	A1097	CAMD-E x Seaberry, F2 (1998)	CAMD-E x Seaberry, F2 (1998).CS-F2_chr01	AD_chr.01	38.5	38.5

Trait Search

Search By Collection

Collection Name

--Any--
Asiatic Cotton Collection (USA)
China Cotton Germplasm Collection
CMD Panel

Germplasm Name

[   ] [Apply]

27,977 records were returned

Search Mapped Sequences by Map Name

Map Name

[   ] [Apply]

Search By Pedigree

Germplasm

[Contains ▼]

[   ]

Enter the name or partial name of a germplasm

Search Sites for Germplasm Mapped sequences

Search Germplasm by Country

search_germplasm_by_country: Country

| --Any-- | [Apply]
|---|
| Afghanistan |
| Algeria |
| American Samoa |

Search Mapped Sequences by Genome Group

Genome group

[   ] [Apply]

Search Mapped Sequences by Chromosome Number

Chromosome number

[   ] [Apply]

Implemented Tools

Future Work

- Finish transferring CottonDB and CMD data to CottonGen

- Implement and develop new Drupal interfaces to browse, query and download data according to user requirements

- Add annotated genome sequence, transcriptome, genotype and phenotype data

- Develop genome annotation community annotation tool

Acknowledgements

THANKS FOR YOUR ATTENTION!

Questions?