

CottonGen

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

Jing Yu¹, Sook Jung¹, Chun-Huai Cheng¹, Stephen Ficklin¹, Taein Lee¹, Ping Zheng¹, Don Jones², Richard Percy³, Dorrie Main¹

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

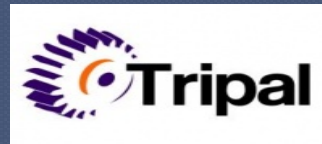
What is CottonGen?

- 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

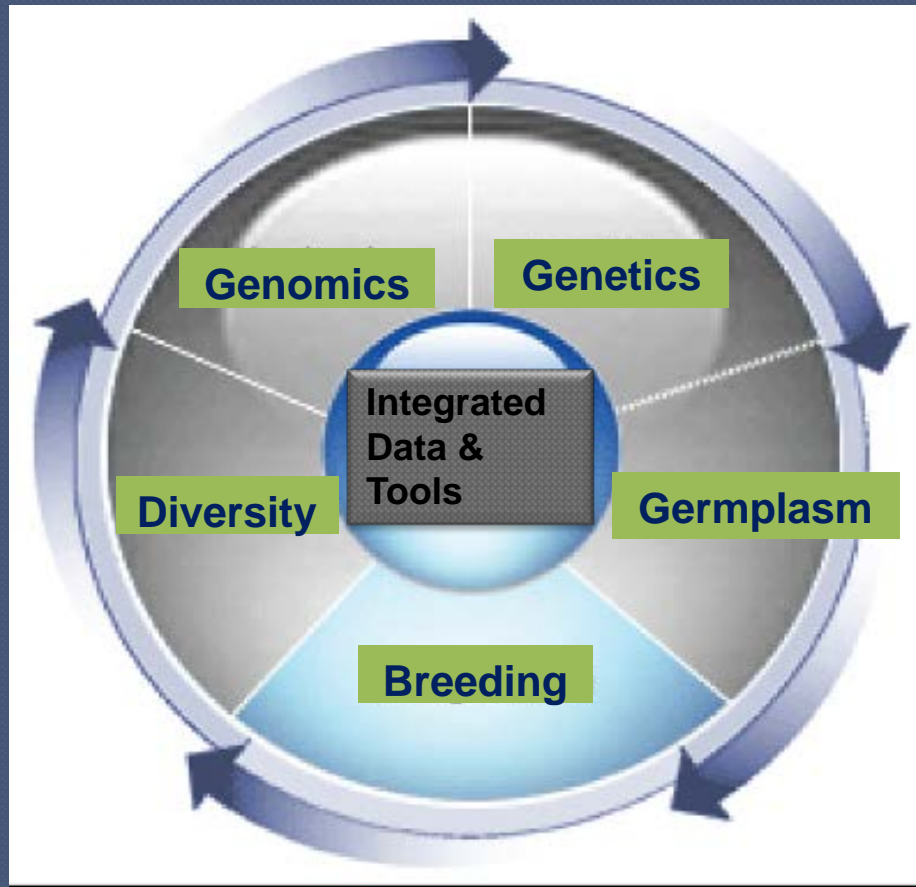


Drupal modules as web front-end for Chado



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



CottonGen

a genomics, genetics and breeding resource for cotton

[Login](#) | [Create Account](#)

[Home](#) | [Contact](#)

General	Data	Tools	Search	ICGI	Mailing Lists
About CottonGen	Germplasm	CMap	Genes	ICGI Home	Subscribe/Unsubscribe
Function Overview	Maps	FPC	Germplasm	ICGI News	The CottonGen List Archives
Steering Committee	Marker	GBrowse	Maps	Becon	The ICGI List Archives
News Archive	Traits	NCBI BLAST	Markers	Forgotten Account Details?	
Contact	ICGI	Batch Editor	Traits	Membership	
Disclaimer		SSR Server		ICGI Mailing List Archives	
				ICGI Meetings	
				ICGI Elections	
				ICGI 2012 Conference	

Welcome to CottonGen

CottonGen is a new cotton community genomics, genetics and breeding database being developed to enable basic, translational and applied research in cotton. It is being built using the open-source [TriPal database infrastructure](#). CottonGen will initially consolidate the data from [CottonDB](#) and the [Cotton Marker Database](#), which includes sequences, genetic and physical maps, genotypic and phenotypic markers and polymorphisms, QTLs, pathogens, germplasm collections and trait evaluations, pedigrees, and relevant bibliographic citations. It will be expanded to include annotated transcriptome, genome sequence, marker-trait-locus and breeding data, as well as enhanced tools for easy querying and visualizing research data. It will facilitate community communication and collaboration through housing the [ICGI website](#), [forums](#) and [mailing lists](#). This project is funded by

Conference

International Cotton Genome Initiative Research Conference
in Raleigh, North Carolina, USA, October 9 -12, 2012.
Cotton Incorporated, the meeting will bring together
to highlight new developments and opportunities to
cotton genome.



News

- The draft genome of a diploid cotton *Gossypium raimondii* published in Nature Genetics by Wang et al., 26 August 2012
- CottonGen Steering Committee formed
- ICGI Cotton Research Conference to be held in Raleigh, USA, October 9-12, 2012
- New Cotton Database being developed

Gene/Sequence Search

Gene/Sequence Search

Gene/Sequence

InterPro

Gene Ontology

KEGG

► Gene/Sequence Search Criteria

21,698 records were returned.

[Download the results](#)

Feature Name	Common Name	Type	Source
Gossypium_Contig1	Gossypium_Contig1	contig	Gossypium unigene v1.0
Gossypium_Contig10	Gossypium_Contig10	contig	Gossypium unigene v1.0
Gossypium_Contig100	Gossypium_Contig100	contig	Gossypium unigene v1.0
Gossypium_Contig1000	Gossypium_Contig1000	contig	Gossypium unigene v1.0
Gossypium_Contig10000	Gossypium_Contig10000	contig	Gossypium unigene v1.0
Gossypium_Contig10001	Gossypium_Contig10001	contig	Gossypium unigene v1.0
Gossypium_Contig10002	Gossypium_Contig10002	contig	Gossypium unigene v1.0
Gossypium_Contig10003	Gossypium_Contig10003	contig	Gossypium unigene v1.0
Gossypium_Contig10004	Gossypium_Contig10004	contig	Gossypium unigene v1.0
Gossypium_Contig10005	Gossypium_Contig10005	contig	Gossypium unigene v1.0
Gossypium_Contig10006	Gossypium_Contig10006	contig	Gossypium unigene v1.0
Gossypium_Contig10007	Gossypium_Contig10007	contig	Gossypium unigene v1.0
Gossypium_Contig10008	Gossypium_Contig10008	contig	Gossypium unigene v1.0
Gossypium_Contig10009	Gossypium_Contig10009	contig	Gossypium unigene v1.0

Gene/Sequence Page

Gossypium_Contig1 (contig) Gossypium spp.

ExPASy Swiss-Prot Homologs

Analysis Date: 2012-10-03 (Blastx of Gossypium unigene v1.0 contigs vs UniProtKB/Swiss-Prot)
Best 10 Hits Shown

Click a description for more details

Match Name	E value	Identity	Description
------------	---------	----------	-------------

Resources

- [contig Details](#)
- [ExPASy Swiss-Prot Homologs](#)
- [Populus trichicarpa v2.0 Homologs](#)
- [Prunus Persica v1.0](#)

Gossypium_Contig1 (contig) Gossypium spp.

contig Details

Name	Gossypium_Contig1
Unique Name	Gossypium_Contig1
Type	contig
Organism	Gossypium spp. (Cotton)

Resources

- [contig Details](#)
- [ExPASy Swiss-Prot Homologs](#)
- [Populus trichicarpa v2.0 Homologs](#)
- [Prunus Persica v1.0 Homologs](#)
- [Vitis vinifera Homologs](#)
- [TAIR10 Homologs](#)
- [Analyses](#)

Species Page

Gossypium spp.

Gossypium spp.

Germplasm

The following stocks are available for this organism.

Resources

- [Species Details](#)

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](#)
- [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.

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 CAP3¹ program. The final assembly has been annotated by BLAST sequence similarity searching against Swiss-Prot², TrEMBL³, TAIR⁴ Arabidopsis proteins, Prunus persica⁵, Populus trichocarpa⁶ and Vitis vinifera⁷.

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.

43



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

Help

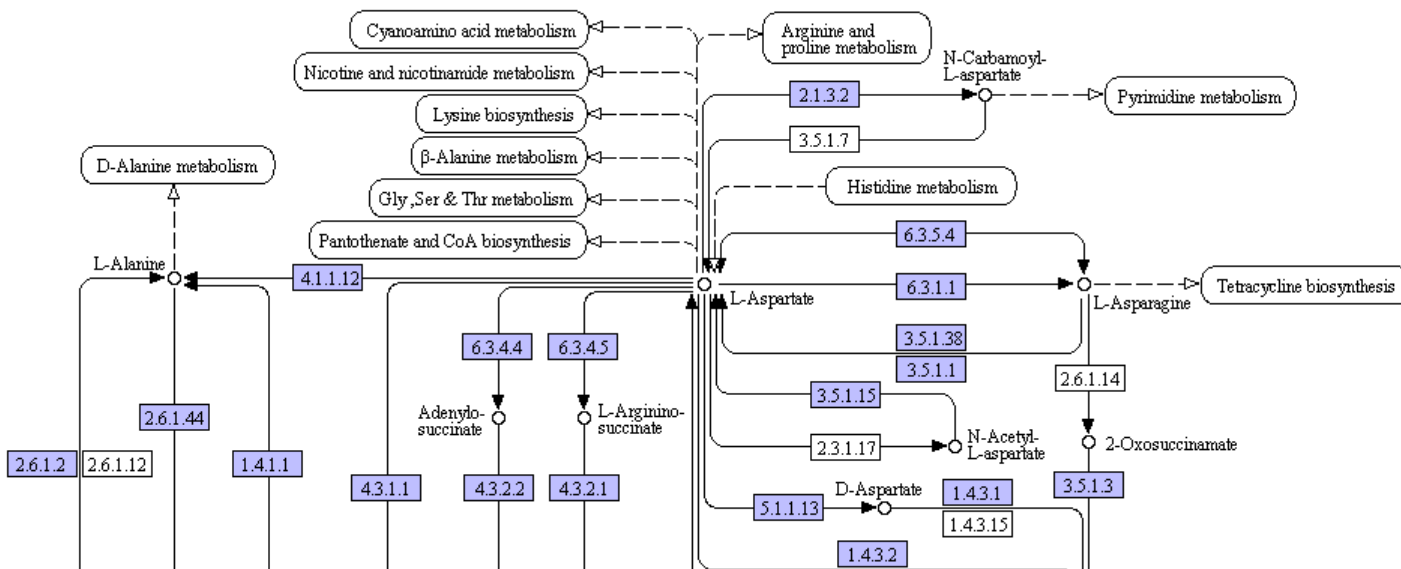
[Pathway menu | Organism menu | Pathway entry | Download KGML | User data mapping]

Reference pathway (KO)

Go

100%

ALANINE, ASPARTATE AND GLUTAMATE METABOLISM



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

Return

Marker Search

Search Markers on Nearby Loci

Locus Distance cM

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

Quantitative Trait Evaluation Search

C

Trait1 Between And:

Qualitative Trait Evaluation Search

Trait1 Value1

Trait2 Value2

Trait3 Value3

Submit



- Any--
- Any--
- Calyx hairs
- Leaf color
- Leaf hairs
- Leaf shape
- Lint color
- Maturity
- Petal color
- Petal spot
- Plant type
- Pollen color
- Productiveness
- Resistance to bacterial blight
- Seed type
- STIGMA



Bayer CropScience



Dow AgroSciences



MONSANTO



Sveto Southern Association of
Agricultural Experiment Station Directors



WASHINGTON STATE
UNIVERSITY

Its

Search By Collection

Collection Name

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

Germplasm Name

27,977 records were returned

Search Mapped Sequences by Map Name

Map Name

Search By Pedigree

Germplasm

Contains

Enter the name or partial name of a germplasm

Search Sites for Germplasm Mapped sequences

Search Mapped Sequences by Genome Group

Genome group

Search Germplasm by Country

search_germplasm_by_country: Country

--Any--
Afghanistan
Algeria
American Samoa

Search Mapped Sequences by Chromosome Number

Chromosome number

Implemented Tools

CMap

FPC

GBrowse

NCBI BLAST

Batch BLAST

SSR Server

Map Search | Feature Search | Matrix | Map Sets | Feature Types | Map Types | Evidence Types | Species

Map Search

Ref. Species: AD GhxGb (Gossypium hirsutum x Gossypium barbadense) Change Species

Ref. Set: --Select--

Name (optional): --Select--

Related Maps:

- Bin : AD GhxGb - Monsanto SSR Bin Map. (2009)
- Genetic : AD GhxGb - Acala-44 x Pima S-7, F2 (2004)
- Genetic : AD GhxGb - CCRI-36 x Hai-7124, F2 (2007)
- Genetic : AD GhxGb - CAMD-E x Seaberry, F2 (1998)
- Genetic : AD GhxGb - Deltapine x Giza-83, F2 (2008)
- Genetic : AD GhxGb - Deltapine-61 x Seaberry, F2 (2000)
- Genetic : AD GhxGb - Emian-22 x 3-79, BC1 (2011)
- Genetic : AD GhxGb - Guazuncho-2 x VH8-4602, BC1 (2005)
- Genetic : AD GhxGb - Guazuncho-2 x VH8-4602, BC2 (2005)
- Genetic : AD GhxGb - Guazuncho-2 x VH8-4602, RIL (2009)
- Genetic : AD GhxGb - Guazuncho-2 x VH8-4602, consensus (2009)
- Genetic : AD GhxGb - Handan-208 x Pima-90, F2:3 (2007)
- Genetic : AD GhxGb - Palmeri x K-101, F2 (2007)
- Genetic : AD GhxGb - Sic'on x F-177, F2 (2004)
- Genetic : AD GhxGb - TM-1 x 3-79, RIL (2006)
- Genetic : AD GhxGb - TM-1 x 3-79, RIL (2012)
- Genetic : AD GhxGb - TM-1 x Hai-7124, BC1 (2007)
- Genetic : AD GhxGb - TM-1 x Hai-7124, BC1 (2008)
- Genetic : AD GhxGb - Vsg x (TM-1 x Hai-7124), DH (2002)

Gossypium hirsutum: 16 kbp from hirsutumC:78,206..94,205

Browser | Select Tracks | Snapshots | Custom Tracks | Preferences

Search

Landmark or Region:

hirsutumC:78,206..94,205 Search

Examples: hirsutumC, atp*

Data Source

Gossypium hirsutum

Scroll/Zoom

Overview

hirsutumC

0k 10k 20k 30k 40k 50k 60k 70k 80k 90k

Region

0k 10k 20k 30k 40k 50k 60k 70k 80k 90k

GC Content



Details

5 kbp

78k 79k 80k 81k 82k 83k 84k 85k 86k 87k 88k 89k 90k 91k 92k 93k 94k

CDS

bB reading frame petB reading frame rpoA reading frame rps8 reading frame rps3 reading frame rpl23 reading frame

psbT reading frame petD reading frame rps11 reading frame rps19 reading frame trnI-CAU reading frame

psbN reading frame rpl36 reading frame rpl2 reading frame

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

- ◉ Cotton Incorporated
- ◉ USDA ARS
- ◉ Bayer CropScience, Dow/Phytogen, Monsanto, Association of Agricultural Experiment Station Directors
- ◉ USDA Specialty Crop Research Initiative (funding Tripal Development)
- ◉ Community of international cotton researchers
- ◉ Washington State University, Texas A&M, Clemson University

THANKS FOR YOUR ATTENTION!

Questions?