

Welcome to the first issue of the CottonGen newsletter in 2016. This newsletter is issued to inform users about new or updated data and features in CottonGen and community related news between October 1, 2015 and March 31, 2016. Please feel free to provide feedback and news via <u>https://www.cottongen.org/contact</u>. The online version can be found at <u>https://www.cottongen.org/CottonGen newsletter</u>.

# What's new at CottonGen?

- 1. USDA-ARS NCGC phenotypic data
- 2. Newly designed CottonGen website
- 3. Genome sequence assembly and annotation data
- 4. CottonCyc databases
- 5. Genome synteny analysis
- 6. CottonSNP63K AD genome intra species maps
- 7. CottonGen BLAST server sequences
- 8. New data available to download
- 9. Work in Progress
- 10. Subscription
- 11. Employment opportunity
- 12. CottonGen Community News

### 1. USDA-ARS NCGC phenotypic data

This dataset is part of USDA-ARS Project: Genotypic and Phenotypic Analysis and Digital Imaging of Accessions in the US National Cotton Germplasm Collection (NCGC). Major data added includes:

- 383,935 trait scores of 50 traits evaluated on 8,976 germplasm in 38 environments between 2006 to 2015
- Trait scores can be queried and downloaded via Trait Search

### 2. Newly designed CottonGen website

CottonGen (Figure 1 - home page, Figure 2 – Expanded Navigation Bar) has been redesigned in a new easier to use theme and format and converted to Drupal v7 and Tripal v2. Major new features include major species overview pages (Figure 3) to more easily view the data and tools available for the major species from a single page as well as quick links to the most used features from the home page.





Species	Data	Search	Tools	ICGI	General
arboreum barbadense herbaceum hirsutum raimondii others	Overview Data contributors Data Download Data Submission Environment Array Gene Genome Gerome Germplasm	Overview Genes Germplasm Markers Publications QTLS Sequences Trait Evaluation	BLAST CMap CottonCyc GBrowse GBrowse_syn JBrowse Sequence Retrieval	ICGI Home ICGI FAQ ICGI News Become an ICGI Member Forgotten Account Details? Membership ICGI Conferences ICGI Elections ICGI Mailing List	About CottonGen Conferences CottonGen Mailing List Data Contributors Disclaimer Employment External Resources Function Overview Mailing List Archives
	Map Marker Publication Species QTL Trait Variety Testing				News Archive Presentations Reference Us Steering Committee Usage Work Completed Work in Progress

Figure 2: CottonGen Navigation Bar Organization

Data		
Overview	Overview	
Genetic Maps	Species Name	barbadense
Genome	Family	Malvaceae
Germplasm	Scientific Name	Gossypium barbadense L.
Libraries	Synonym	Gossypium vitifolium
Markers	Common Name	algodonero de las Barbados [Spanish] [view all 18]
Publications	Geographic Origin	NW South America
Sequences	Genome	(AD)2
Trait Evaluation	Haploid Chromosome Number	26
Trait Loci	Fertile with	Gossypium hirsutum [view all 2]
Transcripts	Sterile with	Gossypium stocksii [view all 13]
Tools	Incompatible with	Gossypium klotzschianum [view all 2]
BLAST	Germplasm	[view all 2005]
СМар	Library	[view all 17]
Download	Sequence	[view all 993923 ]
GBrowse		
GBrowse_Syn		
JBrowse		
PathwayCyc		

Figure 3: CottonGen Example of a Species Page

# 3. Genome sequence assembly and annotation data

- G. barbadense 3-79 (AD2-NBI) genome sequence assembly and annotation available on Data Download
- *G. barbadense* 3-79 (AD2-NBI) genome sequence and annotation data available on the <u>BLAST servers</u>
- New alignments with CottonGen markers, unignenes, EST sequences added and available on <u>Data Download</u>
- New protein alignments and homologs added and available on Data Download

# 4. CottonCyc pathways database

- <u>CottonCyc G. hirsutum (AD1-BGI.v2)</u> pathway database created
- <u>CottonCyc G. raimondii (D5-JGI.v2)</u> updated to 1.0.1 by running against the newest MetaCyc and updated pathway tools.

## 5. Genome synteny analysis

• CottonGen analysis of synteny between *G. arboreum* (A2-BGI.v2) and *G. raimondii* (D5-JGI.v2) added. View on <u>GBrowse Syn</u> (Figure 4).



Figure 4: GBrowse-Syn browse showing synteny between *G. arboreum* (A2-BGI.v2) and *G. raimondii* (D5-JGI.v2)

### 6. CottonSNP63K AD genome intra species maps in CMap

- TAMU CottonSNP63K AD genome inter species maps added
  - o <u>3T-F2</u> 19,191 loci
  - o <u>PS-F2</u>, 7,171 loci

# 7. CottonGen <u>BLAST server</u> sequences

- Added G. barbadense 3-79 genome, genes, CDS and protein sequences
- Updated UniprotKB Gossypium protein sequences
- Updated GenBank Gossypium protein sequences
- Updated GenBank Gossypium CoreNucleotide sequence
- Updated Gossypium GenBank EST sequences
- Updated *Gossypium* GenBank GSS sequences

### 8. New data available from <u>Data Download</u>

• G. barbadense 3-79 (AD2-NBI) genome sequence assembly and annotations

- Joshua Udall's SNP data from 15 *G. hirsutum* and 16 *G. barbadense* accessions. SNPs were called according to the D5 reference position. For the At- and Dtgenomes, SNPs information exists between species, and within species
- <u>Tables</u> created providing summary and links to NCBI SRA RNASeq and GBS datasets by major species

### 9. Work in Progress

- Adding the TAMU cottonSNP63k map dataset to chado
- Creating CottonGen reference transcriptomes using public RNASeq and EST data
- Adding more map, marker and QTL data to get CottonGen current on these published data sets
- Developing further terms for Trait Ontology to describe cotton traits.
- Developing a Breeding Information Management System (BIMS, Figure 5) to provide a comprehensive resource for uploading, storing, managing and analyzing private breeding data within a secure portal that all enables secure and integrated access to all public relevant data. For more information on this see the description at https://www.cottongen.org/BIMS.
- Developing short audio-visual tutorials on how to use CottonGen for various activities.

Archive	E Program Selection					
Manage Breeding	Instructions					
Breeding Program Manage List Location Trial Cross	before you can create your own program. Within your program yo	u can create multiple trial ding program with which y	have not created your own program. You need to ask for a breeder account is and make multiple crosses in multiple locations. You do not need to creat our do not want to share your own private breeding data. Once you create			
	Program Selection	Breeding Program Details				
	CN_COT GR8N	Owner	Itaein			
	NCOC UZ_COT	Program Name NCGC				
		Description	NCGC (the U.S. National Cotton Germplasm Collection)			
Data Import	Select a Program Create a New Breeding Program					
earch			This project is to conserve, describe, evaluate, and distribute accessions of the National Cotton			
lownload			Marine Marine Marine			
ield Book App Management	look App Management		Show Descriptors			
Data Analysis		Descriptors				

Figure 5: CottonGen BIMS (will replace this with a more readable image before sending out)

### 10. Subscription

Visit our mailing lists be added/removed to the available <u>CottonGen mailing lists</u>. More information is available on each page on how to use them. If you have any difficulties please <u>contact us</u>.

## **11. Employment Opportunity**

We are currently seeking to add a postdoctoral bioinformaticist to work on sequence analysis and other aspects of CottonGen. Must have appropriate skillset, good communication skills and able to work well with the <u>Mainlab team</u>. Opportunity exists to work remotely within the US for the right person and a good compensation package is offered with good opportunities for career development. Interested applicants should contact <u>Dorrie Main by email</u>.

### **12.** CottonGen Related Community News

- Efforts toward development of a comprehensive CottonGen Breeding Information Management System (BIMS) was presented at the Cotton Breeder's Workshop, Beltwide, New Orleans, Louisiana, USA (1/5/16)
- CottonGen was presented at the Beltwide Cotton Conference, New Orleans, Louisiana, USA (1/7/16)
- CottonGen was presented at the International Plant and Animal Genome Conference XXIV Computer Demo, San Diego, California, USA (1/12/16)
- A CottonGen Steering Committee Meeting was held (3/23/2016)
- CottonGen was presented at the 2016 ICGI Conference, Goiânia, Brazil (05/03/16)