CottonGen: An Integrated Web Database for Cotton Genomics, Genetics and Breeding Data

Plant and Animal Genome Conference XXIII

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 * corresponding authors: jing.yu@wsu.edu, dorrie@wsu.edu

Abstract

CottonGen (http://www.cottongen.org) is a curated web-based relational database providing a central repository and data-mining resource for genomics, genetics, and breeding data of cotton. CottonGen contains whole genome sequences and annotation of both *G. raimondii* and *G. arboretum* - the ancestors of tetraploid commercial cottons, the unigenes for the family and the genera, gene sequences from NCBI, genetic maps, trait loci, germplasm, marker diversity, breeding, and publication data. All predicted genes of the whole genome sequence, unigenes and NCBI genes have been further annotated by homology to genes in other species, InterPro protein domains, GO terms and KEGG pathway terms. Whole genome sequences and annotation including genes, mRNAs, markers, re-sequencing data, NCBI sequences and homologous genes from other species, can be viewed through GBrowse and JBrowse. Pathway data can also be accessed through the CottonCyc Pathways. New querying functionality includes an advanced marker search, gene search, publications search and a sequence retrieval tool. Data submission templates, tutorials and a frequently asked questions section have also been added. Future development will include implementation of a breeder's toolbox, and addition of synteny, gene/genome curation tools as well as more map, marker and trait data.

Function Overview

Genomes: *G. raimondii* and *G. arboreum* genome sequence assembles and functional annotations (predicted genes, expression data, InterPro, GO terms, and KEGG assignments.

Maps: 49 map data sets with 34,559 loci consisting of 44 genetic, 2 association, 1 bin, 1 consensus, and 1 *in-silico maps*, which covers cotton genome groups AD, A, D, and G.

Markers: Over 265,000 genetic markers consisting of 3,541 RFLPs, 78,340 SSRs, 183,035 SNPs, etc.

Germplasm: Over 15,000 germplasm including populations or individual entries identified from nearly 50k names or entries obtained from 14 collections from United States, Uzbekistan, and China.

Trait: 107,889 trait scores from USDA GRIN, China and Uzbekistan germplasm evaluations.

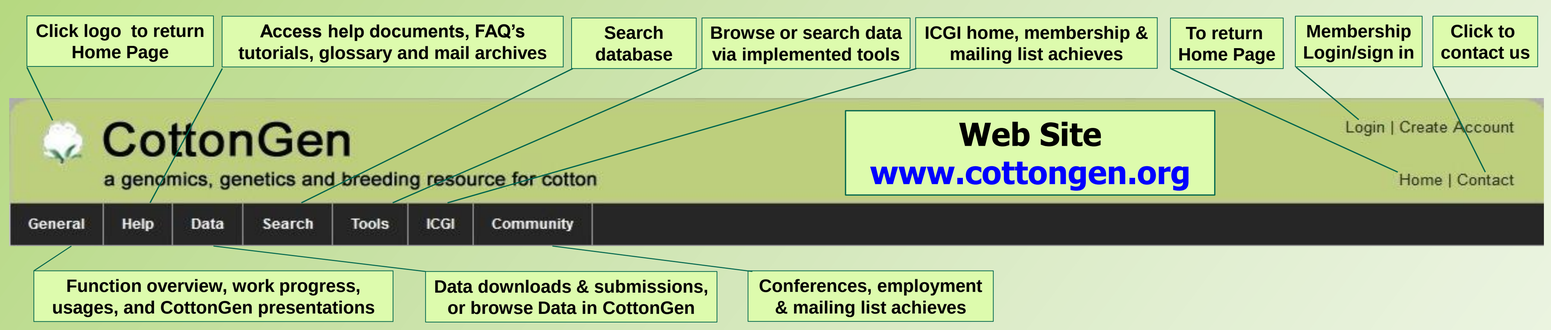
Images: 12,269 digital images from the USDA-ARS National Cotton Germplasm Characterization Project.

BLAST: Search for sequence similarity matches; select the best target sequence database and alignment parameters for your search.

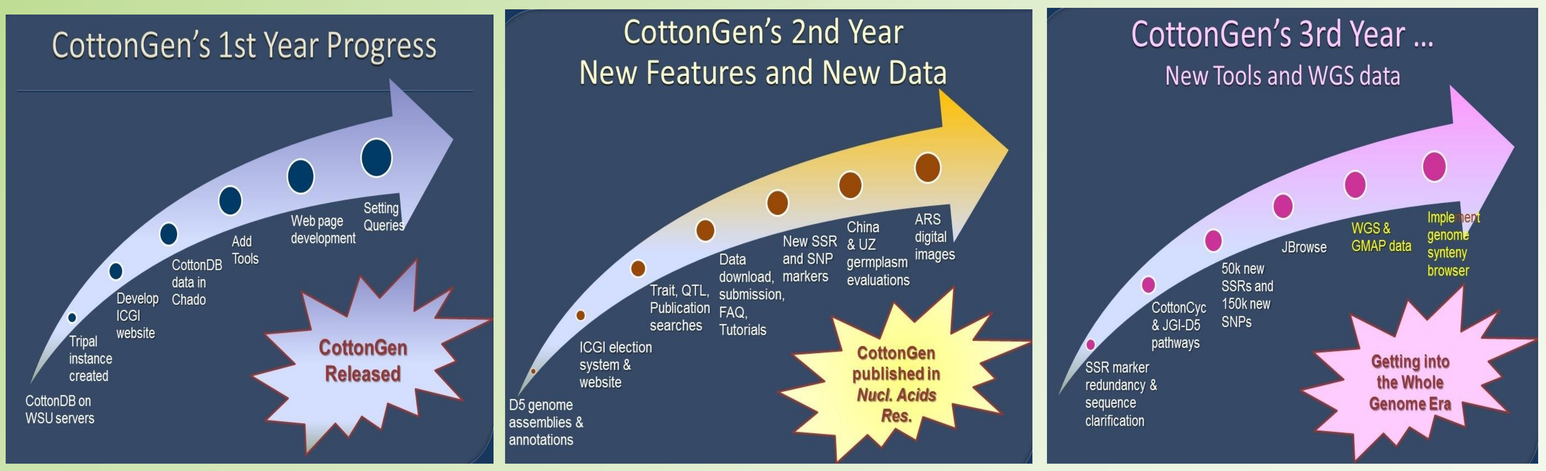
CMap: Search and view mapped genes, markers, and QTLs using the comparative mapping viewer CMap.

CottonCyc: Browse the cotton predicted pathways generated by PathwayTools using the JGI v2.0 *G. raimondii* genome gene models.

GBrowse and JBrowse: Browse and search genes, markers, expressed sequences, and other structural genomics features on the *G. raimondii* (D5) and *G. arboreum* (A2) genomes using the genome browsers GBrowse and JBrowse.



Achievements



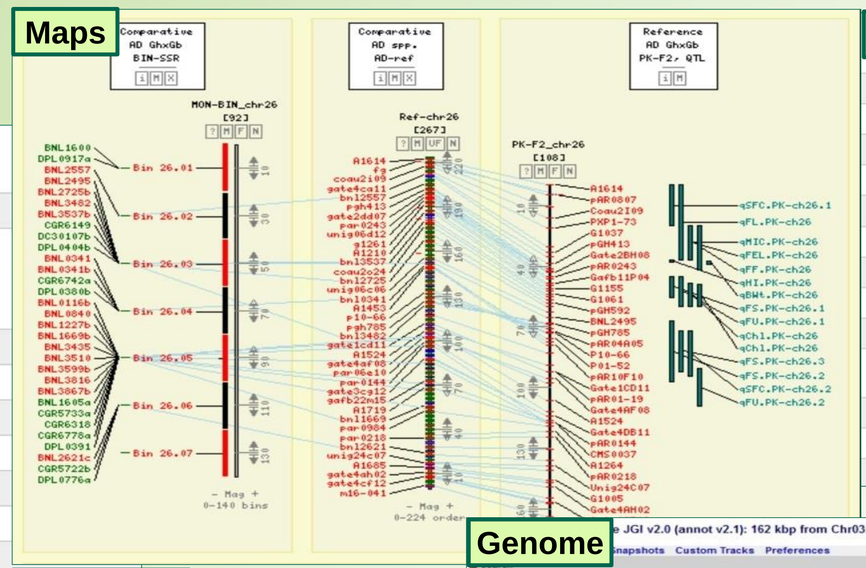
Search Markers

- All markers
  - Search for marker source information. A table of n
  - Search for mapped markers. A table of all markers
  - Search markers on nearby loci. A list of all loci th
  - Advanced Marker Search. Search markers in eith
- Mapped markers with sequences
  - Browse or search by map name A table listed by
  - Browse or search by chromosome number. A tab
  - Browse or search by Genome Group A table liste

Marker Details

Name	TMB0364
Alias	TMF14 TMB0334 TMB0364=TMB0334 TML20
Genbank ID	BV679351
Type	SSR
Probe	N/A
Species	Gossypium hirsutum
Germplasm	TM-1
Source Sequence	N/A
Source Type	genomic
Repeat Motif	(CA)9
PCR Condition	Annealing temperature: 55
Primer 1	TMB0364_F: GGCACACCGGTA
Primer 2	TMB0364_R: TTGCATGACCAA
Product Length	227

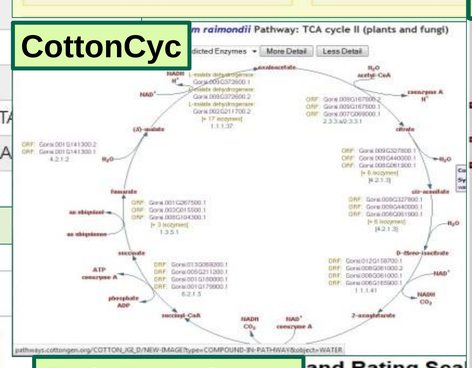
Maps



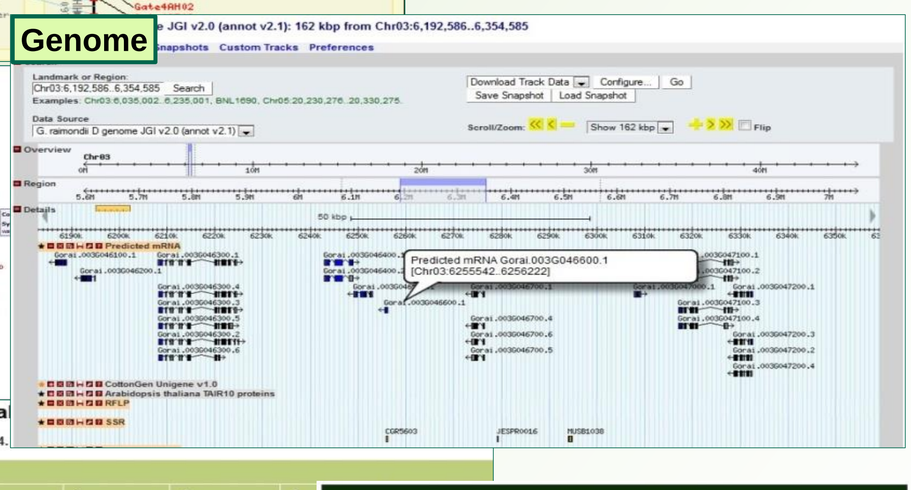
QTL Details

QTL Label	qFL.HP-ch14
Published Symbol	qFLchr14
Trait Name	Fiber_length
Trait Alias	length. Fiber.
Trait Study	HP-F2:3-2007
Population	Handan-208 x Pima-90, F2:3
Female Parent	Handan-208
Male Parent	Pima 90
Colocalizing Marker	N/A
Neighboring Marker	m7e11-360 BNL1059
LOD	3.82
Additivity Dominance Ratio	0.17
R2	0.134

CottonCyc



Genome

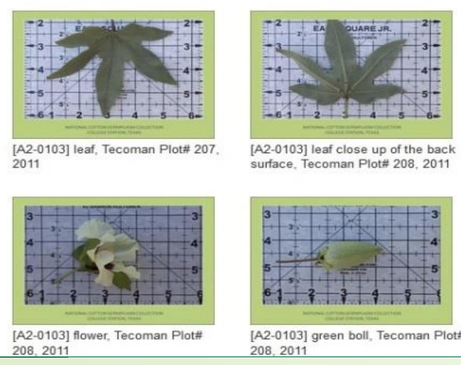


Germplasm Details

Name	TM-1
Alias	Texas-marker 1 (TM1) [view all 17]
Type	undefined
Species	*Gossypium hirsutum*
In Collection	[view all 3]
Description	genetic standard line
Origin	USDA-ARS-SPARC-CGRU
Pedigree	derived from the cultivar Deltap
Maternal Parent of	TM-1 x 3-79, F2 [view all 6]
Paternal Parent of	7235 x TM-1, RIL
Phenotypic Data	[view all 163]
Genotypic Data	[view all 2516]
Map	[view all 6]
DNA Library	[view all 13]
Sequence	[view all 48417]

Germplasm Images

Garo Hill (PI 529730)



[A2-0103] leaf, Tecoman Plot# 207, 2011

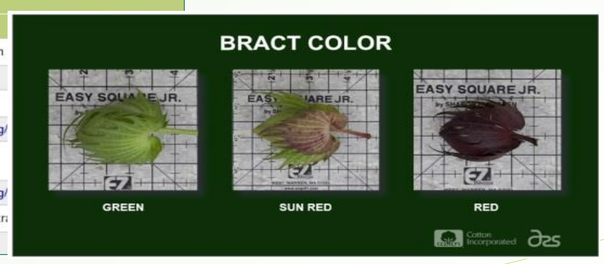
[A2-0103] leaf close up of the back surface, Tecoman Plot# 208, 2011

[A2-0103] flower, Tecoman Plot# 208, 2011

[A2-0103] green boll, Tecoman Plot# 208, 2011

Trait Descriptors

Descriptor	Description Code or Rating Scale					
	0	1	2	3	4	5
Boll Color*		green	dark green	light green	red	sun
Boll Glanding	glandless*	medium	light	heavy	seg/off type	
Boll Nectaries	absent	present	reduced	inactive	seg/off type	
Boll Pitting		smooth	light pitted	pitted	very pitted	seg/
Boll Point		Moderately pointed	pointed	blunt	seg/off type	
Boll Shape		Oval	round	cone	cone-oval	seg/
Boll Size		medium	large	small	seg/off type	extr
Bract Color		green	red	sun red	seg/off type	



Acknowledged with Thanks

The Cotton Research Community