



Welcome to the first issue of the CottonGen newsletter in 2022. This newsletter is issued to inform users about **new or updated data and tools in CottonGen**. In addition to new and updated data, each issue will provide more information on data or tools on the **featured tool/data** section.

CottonGen Manuscript Published (11/2021)

Yu J, Jung S, Cheng C-H, Lee T, Zheng P, Buble K, Crabb J, Humann J, Hough H, Jones D, Campbell JT, Udall J, Main D. CottonGen: The Community Database for Cotton Genomics, Genetics, and Breeding Research. *Plants*. 2021; 10(12):2805.

<https://doi.org/10.3390/plants10122805>

(This article belongs to the Special Issue Plant Genetic Resources and Their Use in Cotton Improvement)

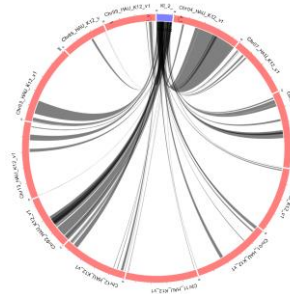
CottonGen Cotton Trait Ontology v2 on Crop Ontology (10/2021)

The structured vocabulary--[CottonGen Cotton Trait Ontology](#)--was established in 2016 by CottonGen curator (Dr. Jing Yu) with input from Drs., [Lori Hinze](#), [Richard Percy](#), and [Russell Kohel](#) (USDA-ARS, College Station, TX). This vocabulary was developed in response to the lack of common, structured names to describe cotton phenotypic traits when integrating into CottonGen germplasm evaluation data collected from four countries and QTL-trait association data obtained from over one hundred peer-reviewed publications. This structured vocabulary was validated as new data was imported to CottonGen, and this cotton ontology later become part of the agricultural data resource known as [Crop Ontology](#) (created by the CGIAR as a source of traits and variables that support the standardization of breeding databases) in 2019. The [Cotton Crop Ontology](#) (v2) has centralized categorical trait descriptors used in different countries as well as some newly added vocabularies.

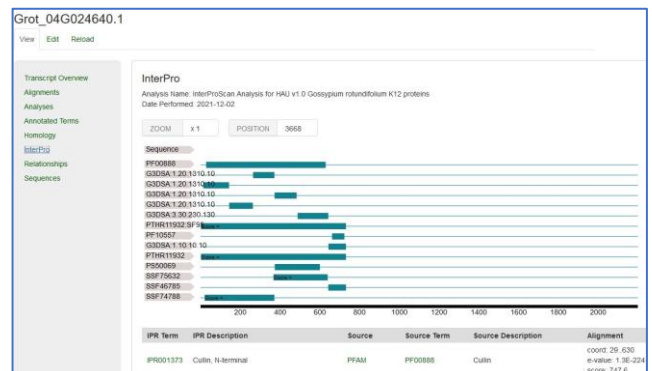
New CottonGen Annotation on Genomes

New CottonGen Functional Analysis (InterProScan and KEGG), Protein Homologies, CottonGen Marker and Reftran Alignments have been added to version of *G.arboreum* (A2), *G.raimondii* (D5), and *G.rotundifolium* (K12) Assemblies ([Wang, et al. Mol Biol Evol. 2021](#)). Access or download data from [CottonGen SyntenyViewer](#), [CottonGen Data Download](#), or from the individual genome pages linked below:

- [G. arboreum \(A2\) 'SXY1' HAU v1](#)
- [G. raimondii \(D5\) 'Grai D502' HAU v1](#)
- [G. rotundifolium \(K12\) 'Grot K201' HAU v1](#)



Synteny view between *G.rotundifolium* (K12) and *G.kirkii* chromosomes genomes and detailed information of a specific syntenic block





Featured Data and Tools

SNP Genotype data

Over 25 million SNP genotype data points currently hosted in CottonGen. The majority of the data have been produced through the efforts of [the TAMU CottonSNP63K Array project](#) (Hulse-Kemp *et al.* 2015).

[Search SNP Genotype](#) is a page for users to search for the SNP genotype dataset based on the germplasm and SNP markers used in the dataset.

Search Genotype

SNP Genotype | **SSR Genotype**

Search SNP Genotype is a page where users can search for the SNP genotype dataset based on the germplasm and SNP markers used in the dataset. Click the next tab to search for SSR Genotype. | [Text tutorial](#) | [Email us with problems and suggestions](#)

Dataset: TAMU_SNP63K_genotyping

Species: Any
Gossypium amourianum
Gossypium arboreum
Gossypium arboreum x thurberi [polyploid]

Germplasm Name: TM-1 (AH-213, Fang)
TM-1 (AH-712, Steily)
tomentosum 201206092.01_SteilyLabLine
Toole

SNP: contains

Genome: Gossypium hirsutum (AD1) 'TM-1' genome ZJU-improved_v2_1_a1

Chr/Scaffold: A12 between 1 and 1000000 bp

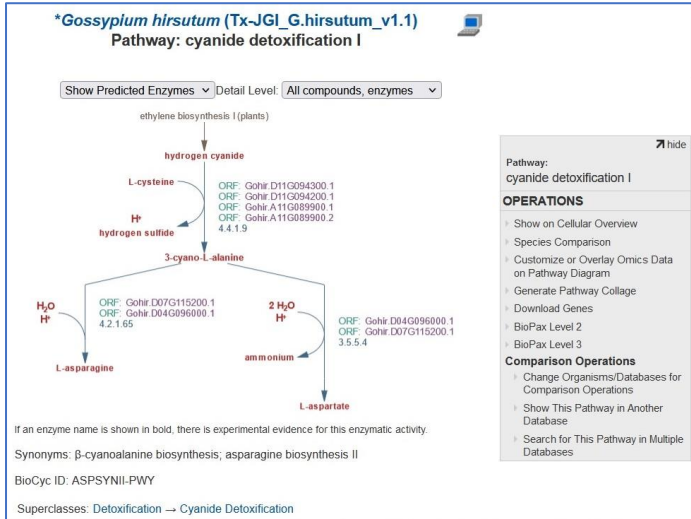
Gene Model: +/- bp

User can search SNP data using **gene model of interest** as well as SNP name, location, germplasm, and dataset name

169 records were returned

#	Array ID	Marker	Location	Allele	3-79	TM-1 (AH-213, Fang)	TM-1 (AH-712, Steily)	tomentosum 201206092.01_SteilyLabLine
1	026099Gr	TAMU_GH_TB0603G24373	Chr01:6544011..6544011	A/G	G	G	G	G
2	02843Gr	TAMU_GH_TB0609G014r23	Chr02:1851282..1851282	T/C	Y	Y	Y	Y
3	024114Gr	TAMU_GH_TB001113r106	Chr02:41443969..41443969	A/G	R	R	R	R
4	026811Gr	TAMU_GH_TB0038r021265	Chr02:50228709..50228709	A/G	G	G	G	G
5	021469Gr	USDA_CFB3145	Chr02:62752943..62752943	G/A	-	-	-	-
6	024909Gr	TAMU_GH_TB0016P11r116	Chr04:58129756..58129756	A/G	A	A	A	A
7	025400Gr	TAMU_Mus_0164r171	Chr05:41048342..41048342	A/G	A	A	A	A
8	02429Gr	TAMU_GH_TB0013M16r21	Chr06:18991481..18991481	A/G	R	R	R	R
9	027907Gr	TAMU_Gl_009156	Chr07:2473825..2473825	A/G	G	G	G	G
10	043587Gr	TAMU_GH_TB0085E07r792	Chr07:16225857..16225857	A/G	G	G	G	G
11	028755Gr	TAMU_GH_TB0036C09190	Chr07:45359081..45359081	A/G	A	A	A	A
12	06793Gr	TAMU_GH_TB0037013R46	Chr07:56983495..56983495	T/C	C	C	C	C
13	069614Gr	TAMU_Gl_097585	Chr08:111695..111695	A/G	A	A	A	A
14	07747Gr	CSIRO_D5chr08_113793	Chr08:113793..113793	G/A	A	A	A	A
15	05922Gr	TAMU_Gb379_001059	Chr08:234229..234229	T/C	C	T	T	C
16	028665Gr	TAMU_GH_TB0057A07r60	Chr08:269756..269756	A/G	A	A	A	A
17	02763Gr	TAMU_GH_TB0008E09r512	Chr08:273180..273180	A/G	A	A	A	A
18	02749Gr	CSIRO_D5chr08_285512	Chr08:285512..285512	T/G	T	T	T	T
19	021074Gr	TAMU_Gb379_002670	Chr08:285907..285907	A/C	-	-	-	-
20	040991Gr	TAMU_GH_TB0057G03r210	Chr08:301100..301100	T/C	C	C	C	C

Join the [CottonGen Mailing List](#) and follow us on [Twitter](#)



PathwayCyc

Genomes added to CottonGen are analyzed with [Pathway Tools](#) to identify metabolic pathways. The data is then displayed with an embedded instance of the software on CottonGen under the [CottonCyc tool](#).

Users can also overlay data from transcriptomics or metabolomics experiments onto the cellular overview graphic. The pathways that have corresponding data are color coded to show over- or under-expression.

PathwayCyc accounts can also be requested to allow users to upload, sort, and save data using the SmartTables feature. For more details on how to use the PathwayCyc tool, watch the video on the MainLab Bioinformatics [YouTube channel](#).

Funded by:
Cotton Inc; USDA-ARS; Bayer CS; CORTEVA;
USDA National Research Project (NRSP10)