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
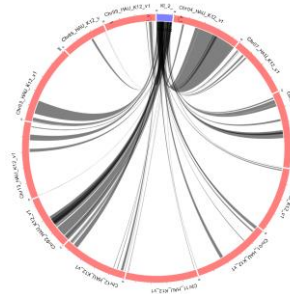
Synteny view between G.rotundifolium (K12) and G.kirkii chromosomes genomes and detailed information of a specific synteny block
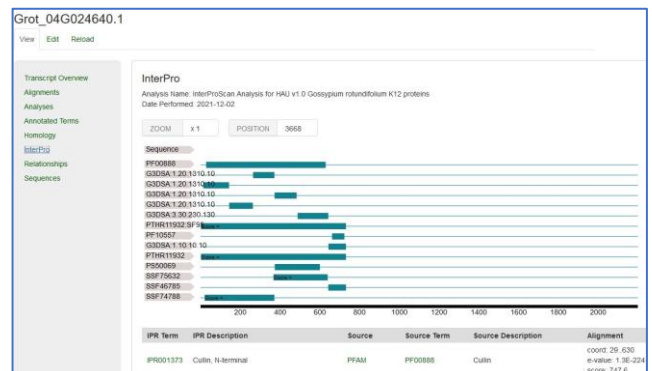
Featured Data and Tools

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 genotyope dataset based on the germplasm and SNP markers used in the dataset.

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

Join the CottonGen Mailing List and follow us on Twitter
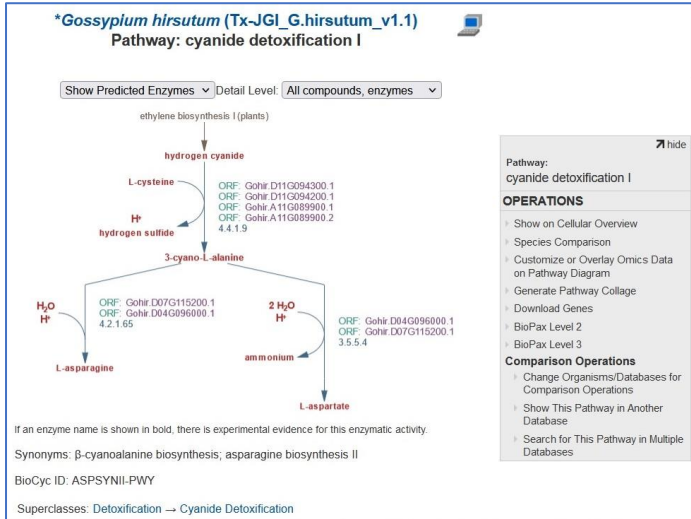
PathwayCyc

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.