COTTONGEN

Overview for Breeders

Jing Yu, Sook Jung, Chun-Huai Chung, Taein Lee, Ping Zheng, Jodi Humann, Deah McGaughey, Morgan Frank, Kirsten Scott, Heidi Hough, Todd Campbell, Josh Udall*, Don Jones, Dorrie Main

Cotton Breeders Tour 2017
Chandler, Arizona


NIH
NHGRI

Community Databases Increasingly Important

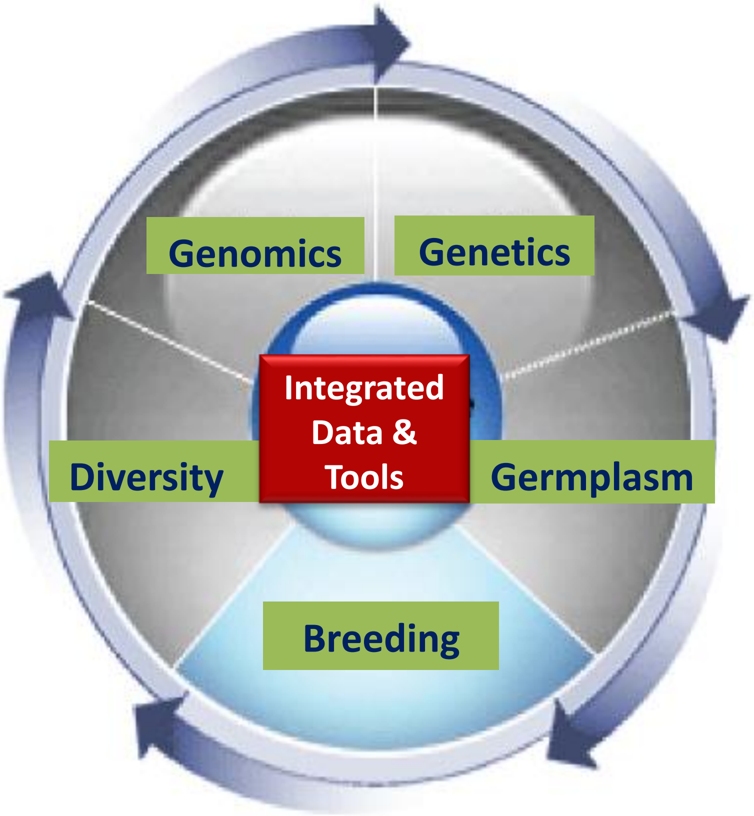
- CottonGen's primary mission is to support research by connecting datasets

- How can CottonGen be useful to you???
  - BIMS for cotton breeding
  - Identify genetic markers to track traits
  - Compare results of previous work with your own (e.g. QTLs)
  - DNA sequence analysis (BLAST, PATHWAYS, sequence retrieval)
  - Recent publications

Integrated Data Facilitates Discovery and Application

Basic Science

- Structure and evolution of genomes
- Gene function
- Genetic variability
- Mechanism underlying traits



Translational Science

- Trait discovery
- Marker development
- Genetic mapping
- Breeding values

Applied Science

Utilization of DNA information in breeding decisions

About CottonGen

- Hosts the ICGI website within it and serves as a communication portal for cotton GGB science
- It is funded by a partnership between the cotton industry and federal programs

Cotton Incorporated

USDA ARS

Bayer CropScience

Dow AgroSciences

MONSANTO

SAAESD Southern Association of Agricultural Experiment Station Directors

NRSP10 CROP DATABASE RESOURCES

Data Available on CottonGen

- Genomics Data
  - Annotated genome sequences (A2, D5, AD1, and AD2)
  - Annotated RNA Seq reference transcriptomes (A2, D5, AD1, and AD2)
  - Annotated *Gossypium* EST unigene
  - CottonCyc metabolic pathways (D5 and AD1)
  - Synteny (D5 and AD1)

- Genetics and Breeding Data
  - 103 genetic maps
  - 404K genetic markers
  - 1026 QTLs
  - 16K germplasm from GRIN, NCGC, China, Uzbekistan, etc. collections
  - 12K digital images of 2K germplasm from USDA-ARS NCGC
  - 492K trait scores from evaluations of US NCGC, GRIN, China and Uzbekistan
  - 60K trait scores for 20 fiber traits from 14 years of RBTN trials (newly added)

3 Takehome points of CottonGen

1. As reviewers and editors, please require data submission to CottonGen. Receive a CottonGen ID number.

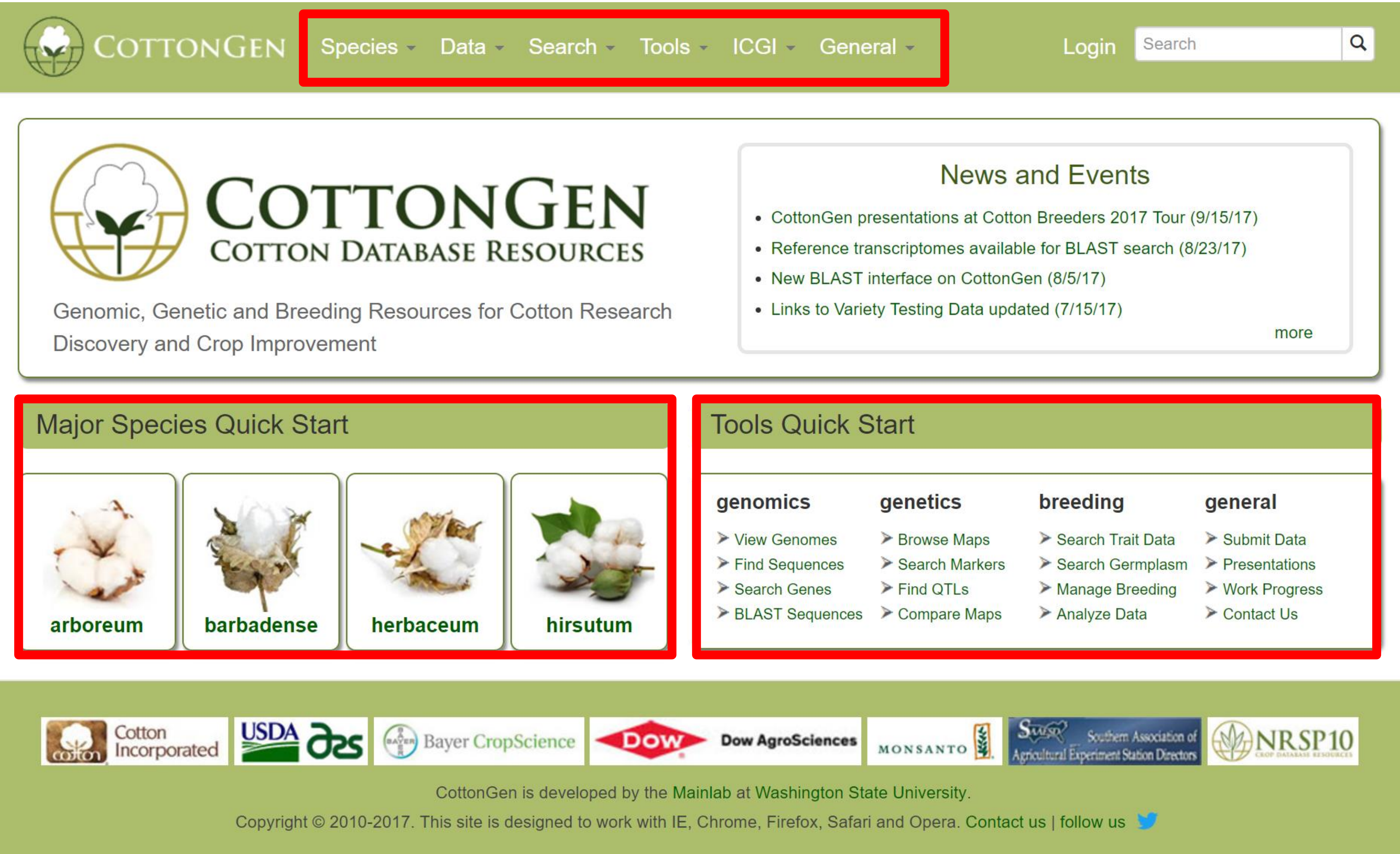
2. As you design and publish your experiments, use common descriptors and methods

3. Use CottonGen! It is your resource. It is excellent.

Tools Available on CottonGen

- Implemented Tools
  - Sequence BLAST Search
  - Genome Browsers
  - Pathway Cyc
  - Map Comparison
  - Sequence Retrieval
- Query search pages for
  - Gene and transcripts
  - Germplasm
  - Marker
  - publication
  - QTL
  - Phenotype data

http://www.cottongen.org



COTTONGEN

COTTON DATABASE RESOURCES

Genomic, Genetic and Breeding Resources for Cotton Research Discovery and Crop Improvement

News and Events

- CottonGen presentations at Cotton Breeders 2017 Tour (9/15/17)
- Reference transcriptomes available for BLAST search (8/23/17)
- New BLAST interface on CottonGen (8/5/17)
- Links to Variety Testing Data updated (7/15/17)

more

Major Species Quick Start

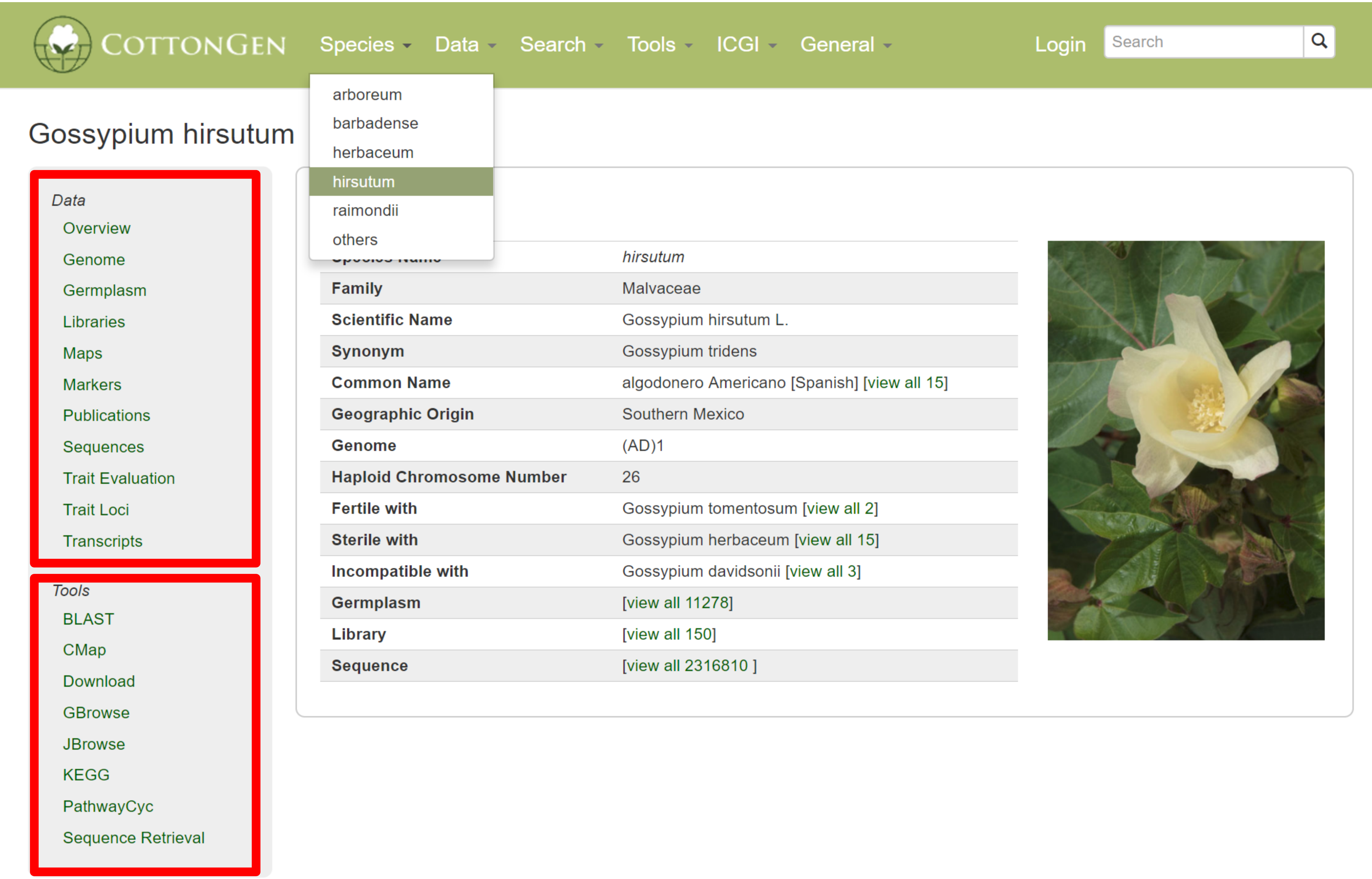
arboreum **barbadense** **herbaceum** **hirsutum**

Tools Quick Start

genomics	genetics	breeding	general
View Genomes	Browse Maps	Search Trait Data	Submit Data
Find Sequences	Search Markers	Search Germplasm	Presentations
Search Genes	Find QTLs	Manage Breeding	Work Progress
BLAST Sequences	Compare Maps	Analyze Data	Contact Us

CottonGen is developed by the Mainlab at Washington State University.

Species Dropdown: Links to Species Data and Tools



Gossypium hirsutum

Data

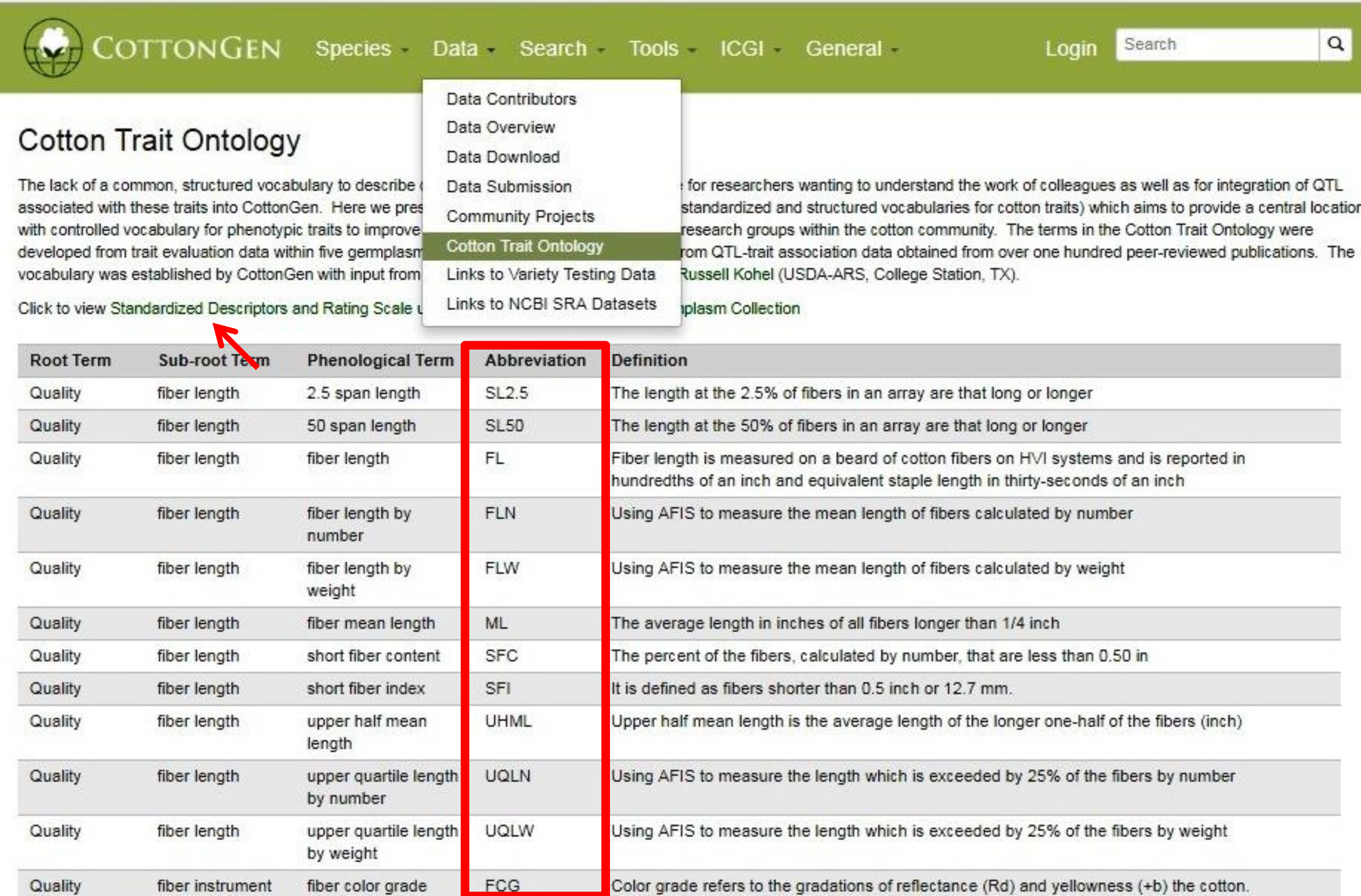
- Overview
- Genome
- Germplasm
- Libraries
- Maps
- Markers
- Publications
- Sequences
- Trait Evaluation
- Trait Loci
- Transcripts

Tools

- BLAST
- CMap
- Download
- GBrowse
- JBrowse
- KEGG
- PathwayCyc
- Sequence Retrieval

Species Name	*hirsutum*
Family	Malvaceae
Scientific Name	Gossypium hirsutum L.
Synonym	Gossypium tridens
Common Name	algodonero Americano [Spanish] [view all 15]
Geographic Origin	Southern Mexico
Genome	(AD)1
Haploid Chromosome Number	26
Fertile with	Gossypium tomentosum [view all 2]
Sterile with	Gossypium herbaceum [view all 15]
Incompatible with	Gossypium davidsonii [view all 3]
Germplasm	[view all 11278]
Library	[view all 150]
Sequence	[view all 2316810]

Data Dropdown: Cotton TO & NCGC Std. rating scales



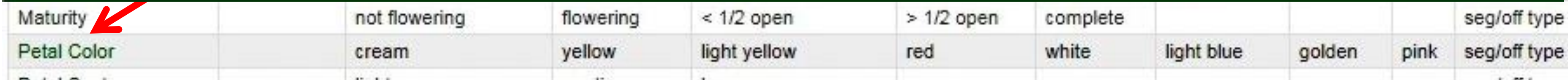
Cotton Trait Ontology

The lack of a common, structured vocabulary to describe … for researchers wanting to understand the work of colleagues as well as for integration of QTL
associated with these traits into CottonGen. Here we pres … standardized and structured vocabularies for cotton traits) which aims to provide a central locatio
with controlled vocabulary for phenotypic traits to improve … research groups within the cotton community. The terms in the Cotton Trait Ontology were
developed from trait evaluation data within five germplasm … rom QTL-trait association data obtained from over one hundred peer-reviewed publications. The
vocabulary was established by CottonGen with input from … Russell Kohel (USDA-ARS, College Station, TX).

Click to view Standardized Descriptors and Rating Scale u … plasm Collection

Root Term	Sub-root Term	Phenological Term	Abbreviation	Definition
Quality	fiber length	2.5 span length	SL2.5	The length at the 2.5% of fibers in an array are that long or longer
Quality	fiber length	50 span length	SL50	The length at the 50% of fibers in an array are that long or longer
Quality	fiber length	fiber length	FL	Fiber length is measured on a beard of cotton fibers on HVI systems and is reported in hundredths of an inch and equivalent staple length in thirty-seconds of an inch
Quality	fiber length	fiber length by number	FLN	Using AFIS to measure the mean length of fibers calculated by number
Quality	fiber length	fiber length by weight	FLW	Using AFIS to measure the mean length of fibers calculated by weight
Quality	fiber length	fiber mean length	ML	The average length in inches of all fibers longer than 1/4 inch
Quality	fiber length	short fiber content	SFC	The percent of the fibers, calculated by number, that are less than 0.50 in
Quality	fiber length	short fiber index	SFI	It is defined as fibers shorter than 0.5 inch or 12.7 mm.
Quality	fiber length	upper half mean length	UHML	Upper half mean length is the average length of the longer one-half of the fibers (inch)
Quality	fiber length	upper quartile length by number	UQLN	Using AFIS to measure the length which is exceeded by 25% of the fibers by number
Quality	fiber length	upper quartile length by weight	UQLW	Using AFIS to measure the length which is exceeded by 25% of the fibers by weight
Quality	fiber instrument	fiber color grade	FCG	Color grade refers to the gradations of reflectance (Rd) and yellowness (+b) the cotton.

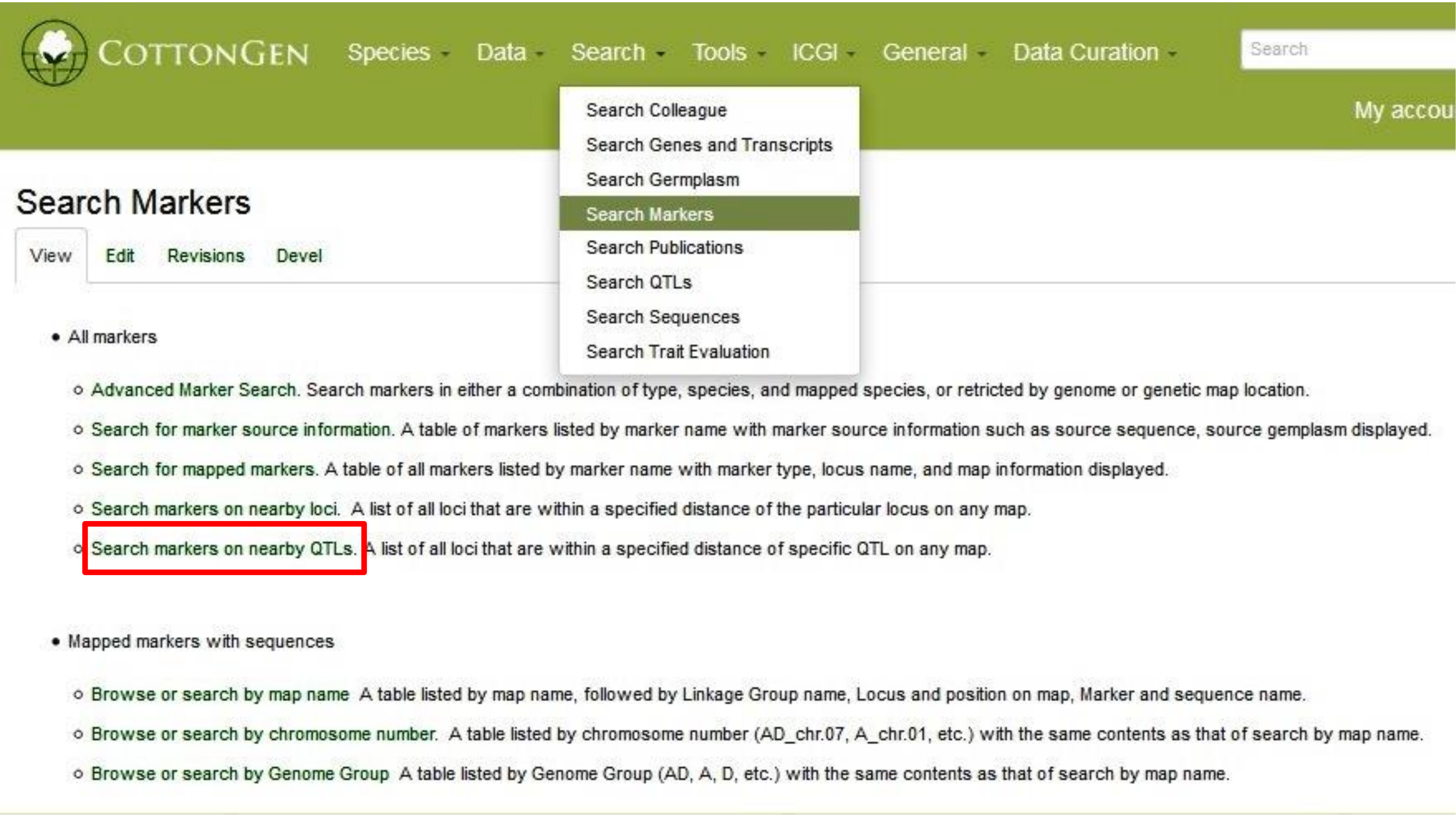
Data Dropdown: Cotton TO & NCGC Std. rating scales



Maturity		not flowering	flowering	< 1/2 open	> 1/2 open	complete				seg/off type
Petal Color		cream	yellow	light yellow	red	white	light blue	golden	pink	seg/off type

Query Examples

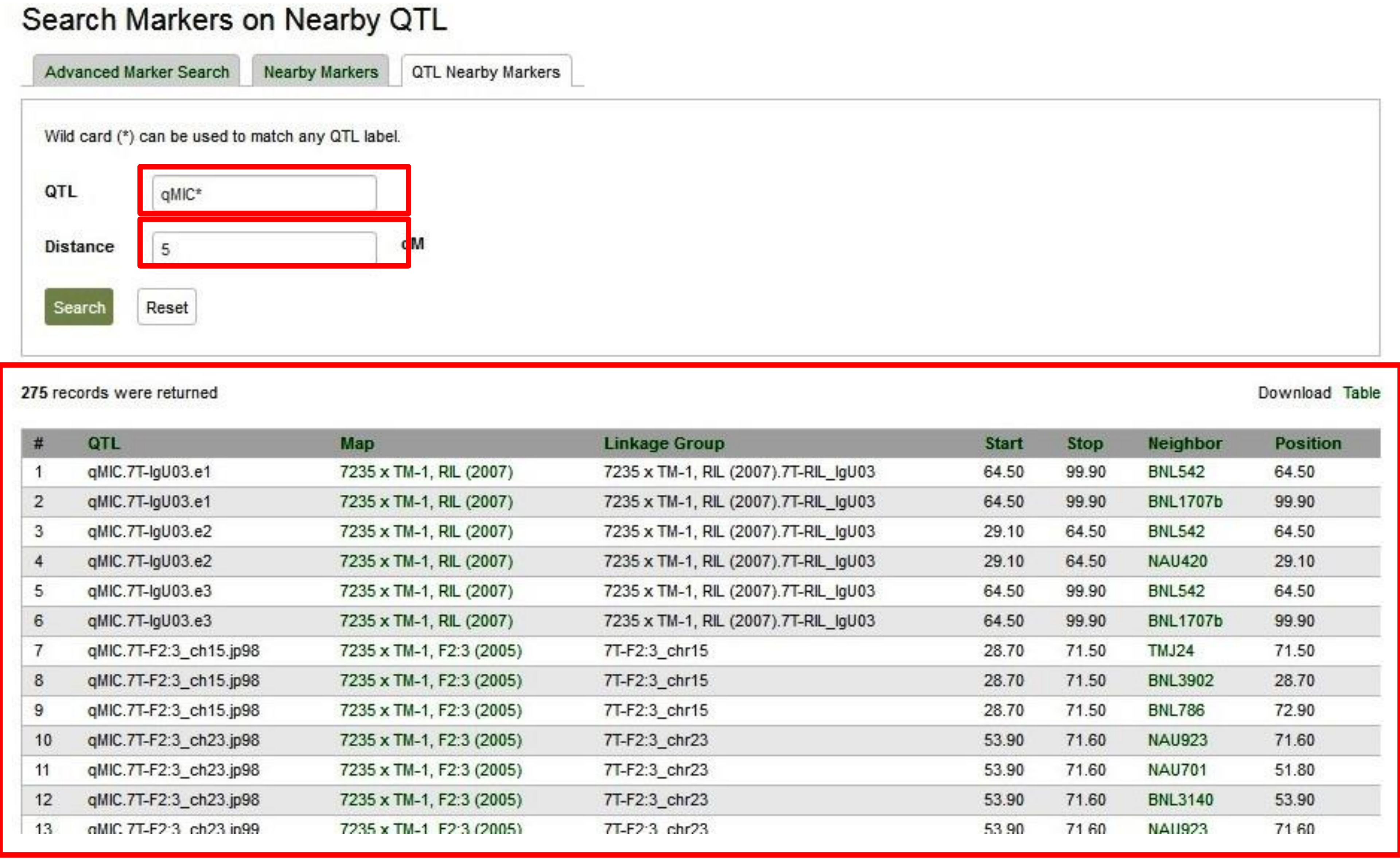
Marker Search



Search Markers

- All markers
    - Advanced Marker Search. Search markers in either a combination of type, species, and mapped species, or retricted by genome or genetic map location.
    - Search for marker source information. A table of markers listed by marker name with marker source information such as source sequence, source gemplasm displayed.
    - Search for mapped markers. A table of all markers listed by marker name with marker type, locus name, and map information displayed.
    - Search markers on nearby loci. A list of all loci that are within a specified distance of the particular locus on any map.
    - Search markers on nearby QTLs. A list of all loci that are within a specified distance of specific QTL on any map.
- Mapped markers with sequences
    - Browse or search by map name A table listed by map name, followed by Linkage Group name, Locus and position on map, Marker and sequence name.
    - Browse or search by chromosome number. A table listed by chromosome number (AD_chr.07, A_chr.01, etc.) with the same contents as that of search by map name.
    - Browse or search by Genome Group A table listed by Genome Group (AD, A, D, etc.) with the same contents as that of search by map name.

Search Marker on Nearby QTLs



Search Markers on Nearby QTL

Wild card (*) can be used to match any QTL label.

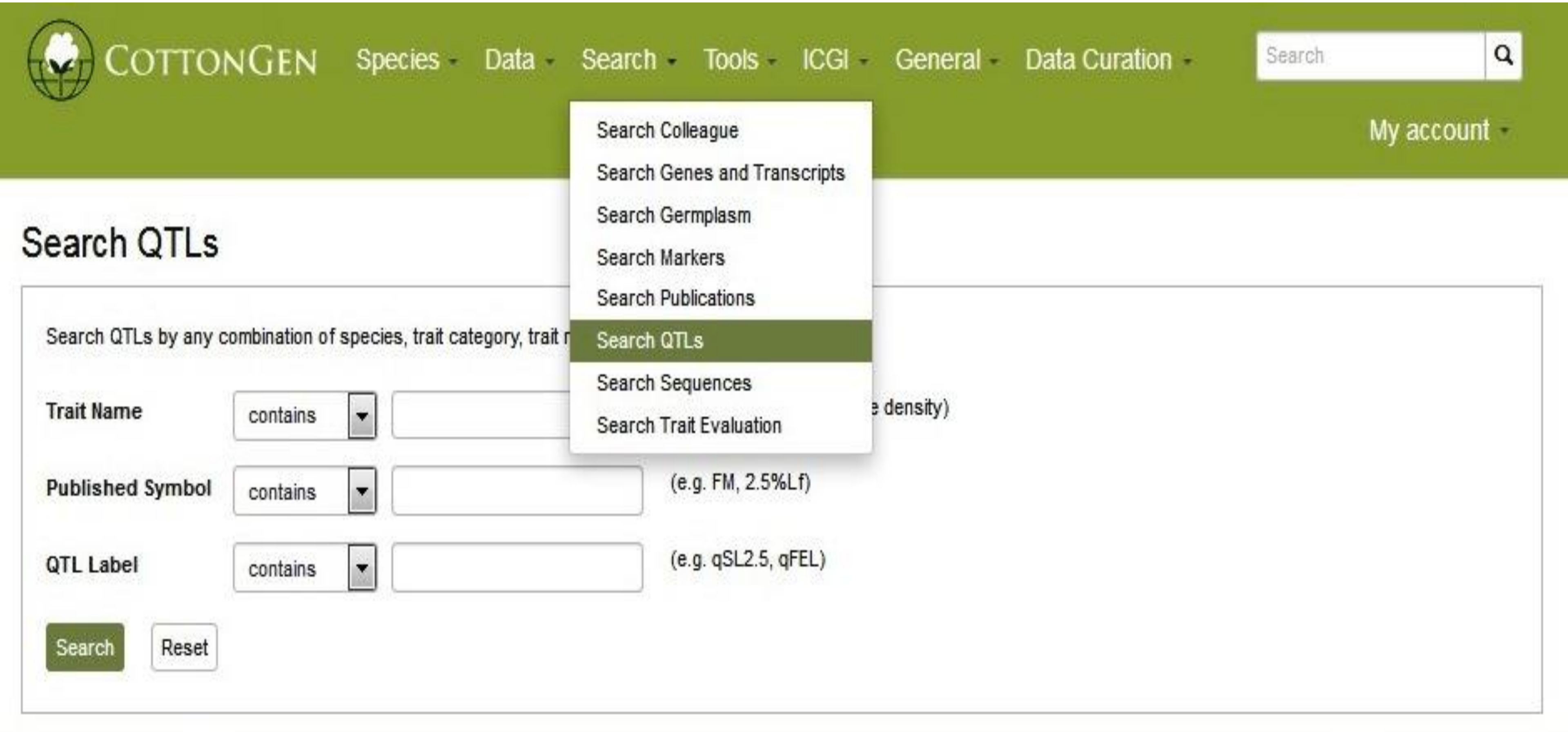
QTL qMIC*

Distance 5 cM

275 records were returned

#	QTL	Map	Linkage Group	Start	Stop	Neighbor	Position
1	qMIC.7T-IgU03.e1	7235 x TM-1, RIL (2007)	7235 x TM-1, RIL (2007).7T-RIL_IgU03	64.50	99.90	BNL542	64.50
2	qMIC.7T-IgU03.e1	7235 x TM-1, RIL (2007)	7235 x TM-1, RIL (2007).7T-RIL_IgU03	64.50	99.90	BNL1707b	99.90
3	qMIC.7T-IgU03.e2	7235 x TM-1, RIL (2007)	7235 x TM-1, RIL (2007).7T-RIL_IgU03	29.10	64.50	BNL542	64.50
4	qMIC.7T-IgU03.e2	7235 x TM-1, RIL (2007)	7235 x TM-1, RIL (2007).7T-RIL_IgU03	29.10	64.50	NAU420	29.10
5	qMIC.7T-IgU03.e3	7235 x TM-1, RIL (2007)	7235 x TM-1, RIL (2007).7T-RIL_IgU03	64.50	99.90	BNL542	64.50
6	qMIC.7T-IgU03.e3	7235 x TM-1, RIL (2007)	7235 x TM-1, RIL (2007).7T-RIL_IgU03	64.50	99.90	BNL1707b	99.90
7	qMIC.7T-F2:3_ch15.jp98	7235 x TM-1, F2:3 (2005)	7T-F2:3_chr15	28.70	71.50	TMJ24	71.50
8	qMIC.7T-F2:3_ch15.jp98	7235 x TM-1, F2:3 (2005)	7T-F2:3_chr15	28.70	71.50	BNL3902	28.70
9	qMIC.7T-F2:3_ch15.jp98	7235 x TM-1, F2:3 (2005)	7T-F2:3_chr15	28.70	71.50	BNL786	72.90
10	qMIC.7T-F2:3_ch23.jp98	7235 x TM-1, F2:3 (2005)	7T-F2:3_chr23	53.90	71.60	NAU923	71.60
11	qMIC.7T-F2:3_ch23.jp98	7235 x TM-1, F2:3 (2005)	7T-F2:3_chr23	53.90	71.60	NAU701	51.80
12	qMIC.7T-F2:3_ch23.jp98	7235 x TM-1, F2:3 (2005)	7T-F2:3_chr23	53.90	71.60	BNL3140	53.90
13	qMIC.7T-F2:3_ch23.jp99	7235 x TM-1, F2:3 (2005)	7T-F2:3_chr23	53.90	71.60	NAU923	71.60

Find QTL(s) of SI that mapped on Chrom 23



Search QTLs

Search QTLs by any combination of species, trait category, trait r

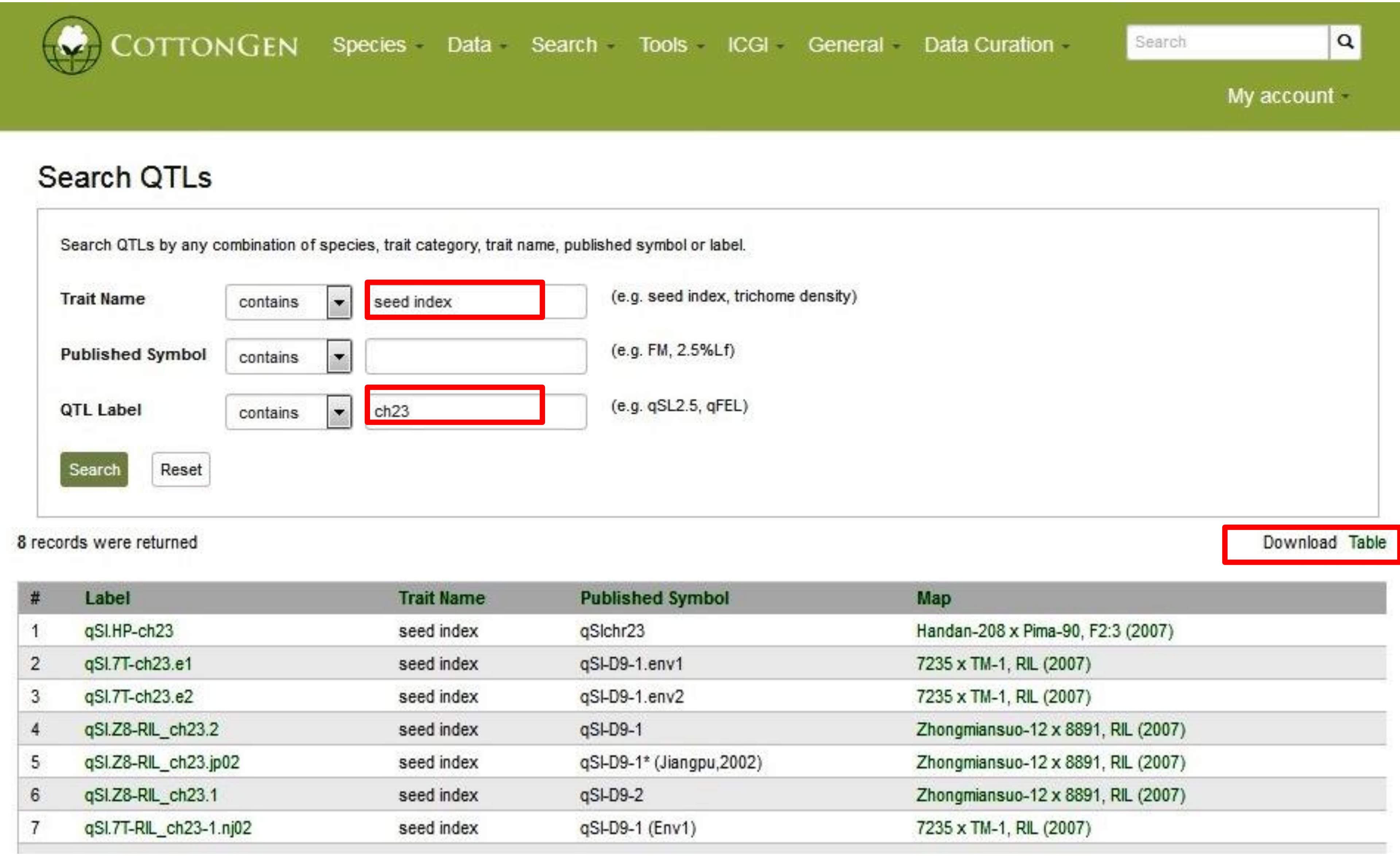
Trait Name contains e density)

Published Symbol contains (e.g. FM, 2.5%Lf)

QTL Label contains (e.g. qSL2.5, qFEL)

Search Dropdown
→ Search QTLs

Find QTL(s) of SI that mapped on Chrom 23



Search QTLs

Search QTLs by any combination of species, trait category, trait name, published symbol or label.

Trait Name contains seed index (e.g. seed index, trichome density)

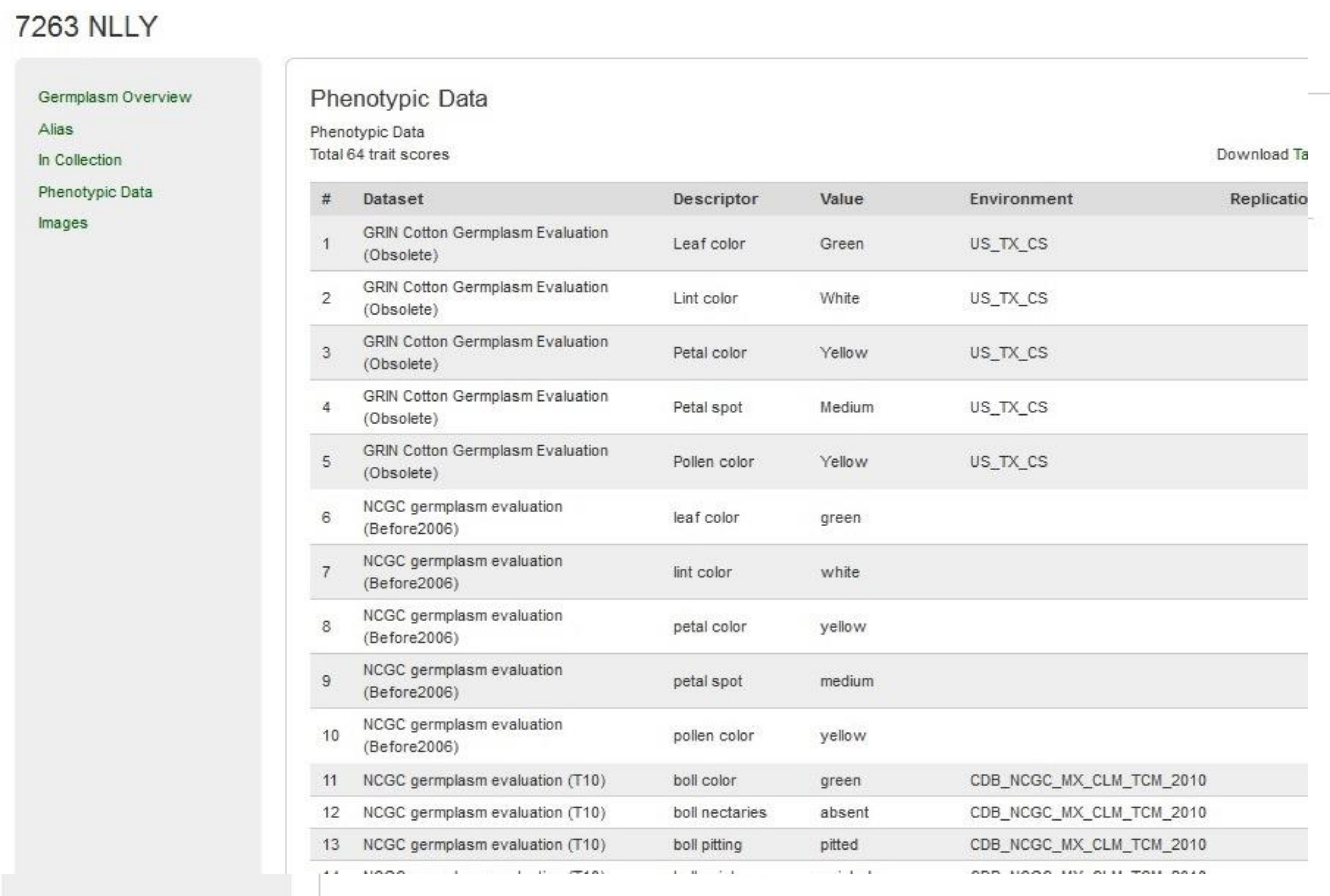
Published Symbol contains (e.g. FM, 2.5%Lf)

QTL Label contains ch23 (e.g. qSL2.5, qFEL)

8 records were returned

#	Label	Trait Name	Published Symbol	Map
1	qSI.HP-ch23	seed index	qSIchr23	Handan-208 x Pima-90, F2:3 (2007)
2	qSI.7T-ch23.e1	seed index	qSI-D9-1.env1	7235 x TM-1, RIL (2007)
3	qSI.7T-ch23.e2	seed index	qSI-D9-1.env2	7235 x TM-1, RIL (2007)
4	qSI.Z8-RIL_ch23.2	seed index	qSI-D9-1	Zhongmiansuo-12 x 8891, RIL (2007)
5	qSI.Z8-RIL_ch23.jp02	seed index	qSI-D9-1* (Jiangpu,2002)	Zhongmiansuo-12 x 8891, RIL (2007)
6	qSI.Z8-RIL_ch23.1	seed index	qSI-D9-2	Zhongmiansuo-12 x 8891, RIL (2007)
7	qSI.7T-RIL_ch23-1.nj02	seed index	qSI-D9-1 (Env1)	7235 x TM-1, RIL (2007)

Further exploration on germplasm ‘7263 NLLY’



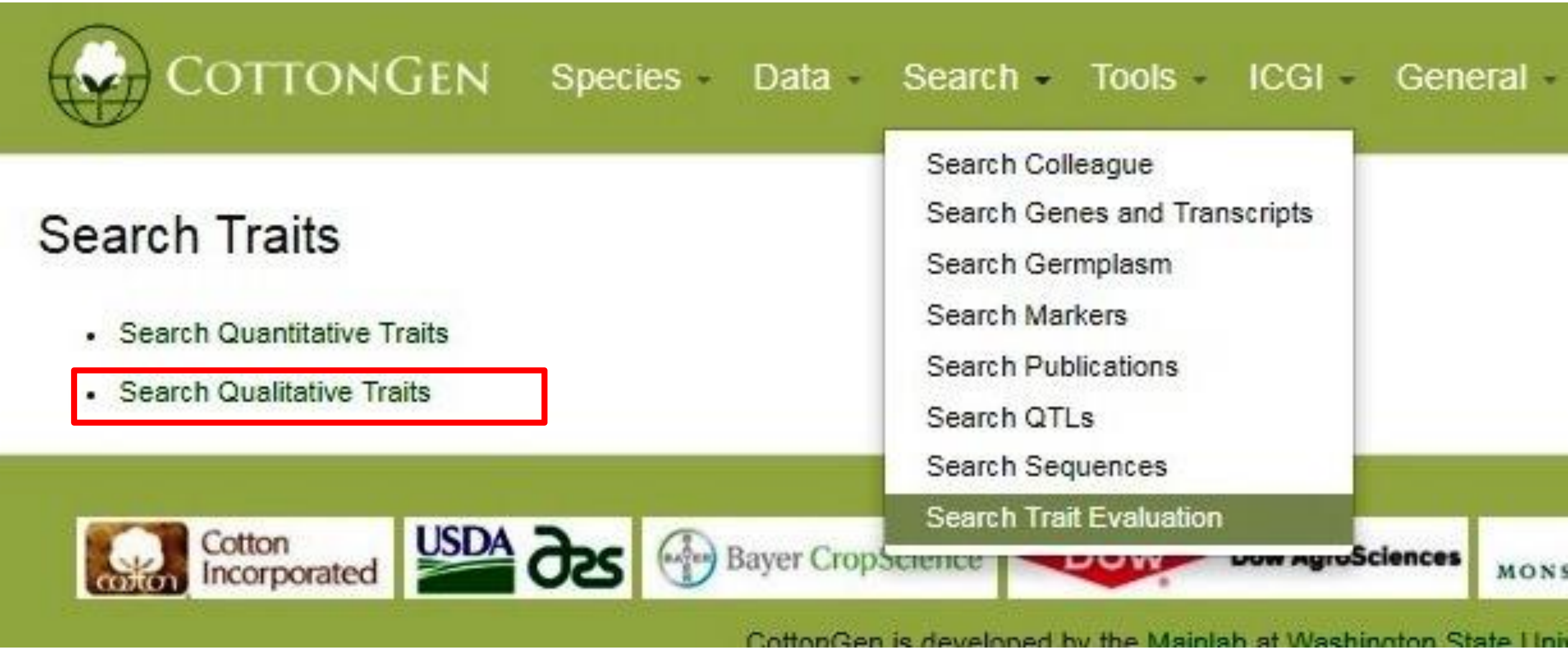
7263 NLLY

Phenotypic Data

Phenotypic Data
Total 64 trait scores

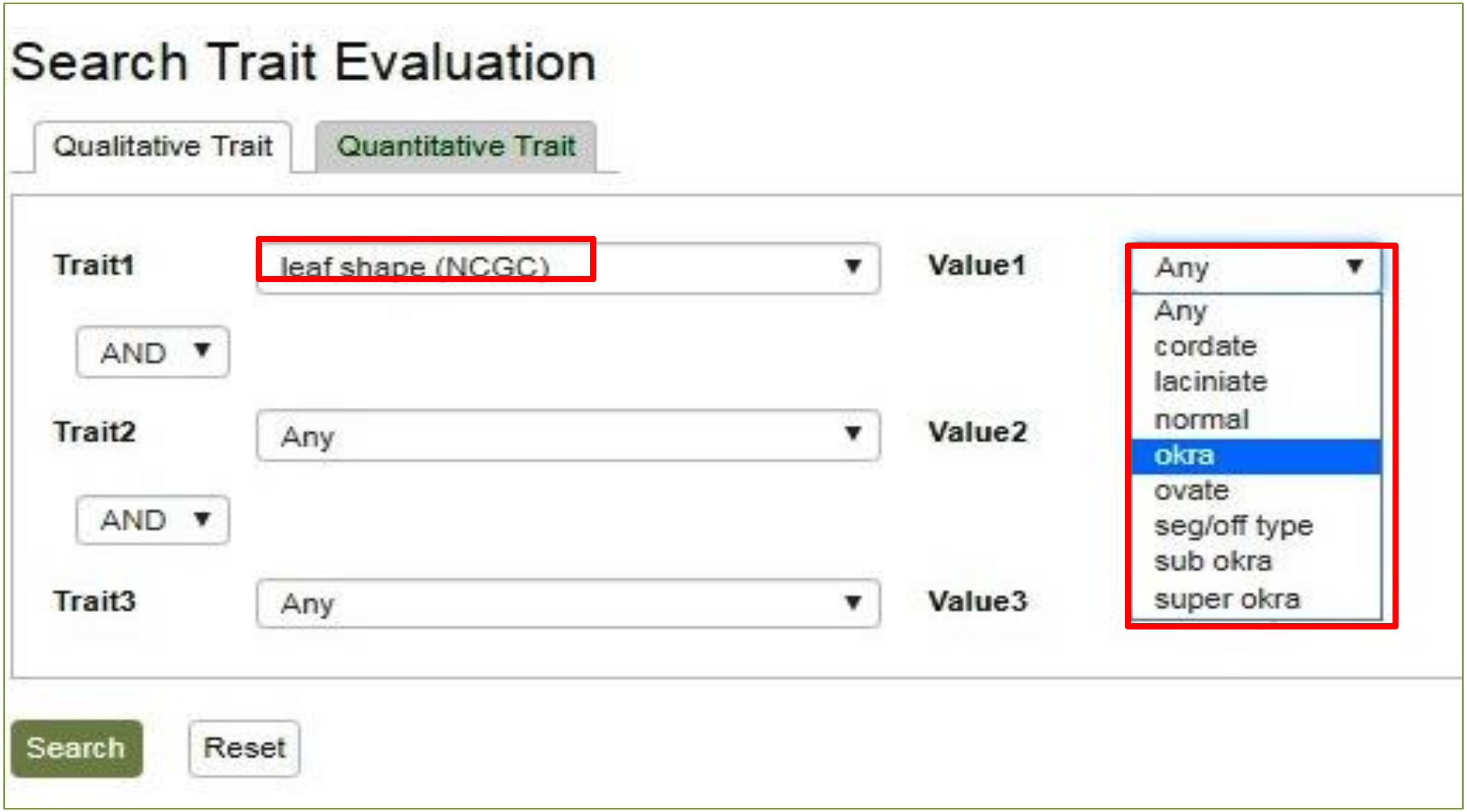
#	Dataset	Descriptor	Value	Environment	Replicatio
1	GRIN Cotton Germplasm Evaluation (Obsolete)	Leaf color	Green	US_TX_CS	
2	GRIN Cotton Germplasm Evaluation (Obsolete)	Lint color	White	US_TX_CS	
3	GRIN Cotton Germplasm Evaluation (Obsolete)	Petal color	Yellow	US_TX_CS	
4	GRIN Cotton Germplasm Evaluation (Obsolete)	Petal spot	Medium	US_TX_CS	
5	GRIN Cotton Germplasm Evaluation (Obsolete)	Pollen color	Yellow	US_TX_CS	
6	NCGC germplasm evaluation (Before2006)	leaf color	green		
7	NCGC germplasm evaluation (Before2006)	lint color	white		
8	NCGC germplasm evaluation (Before2006)	petal color	yellow		
9	NCGC germplasm evaluation (Before2006)	petal spot	medium		
10	NCGC germplasm evaluation (Before2006)	pollen color	yellow		
11	NCGC germplasm evaluation (T10)	boll color	green	CDB_NCGC_MX_CLM_TCM_2010	
12	NCGC germplasm evaluation (T10)	boll nectaries	absent	CDB_NCGC_MX_CLM_TCM_2010	
13	NCGC germplasm evaluation (T10)	boll pitting	pitted	CDB_NCGC_MX_CLM_TCM_2010	

Find all germplasm with okra leave shape



Search Dropdown
→ Trait Evaluation → Search Qualitative Traits

Find all germplasm with okra leave shape



Search Trait Evaluation

Trait1: leaf shape (NCGC) Value1: Any

AND

Trait2: Any Value2:

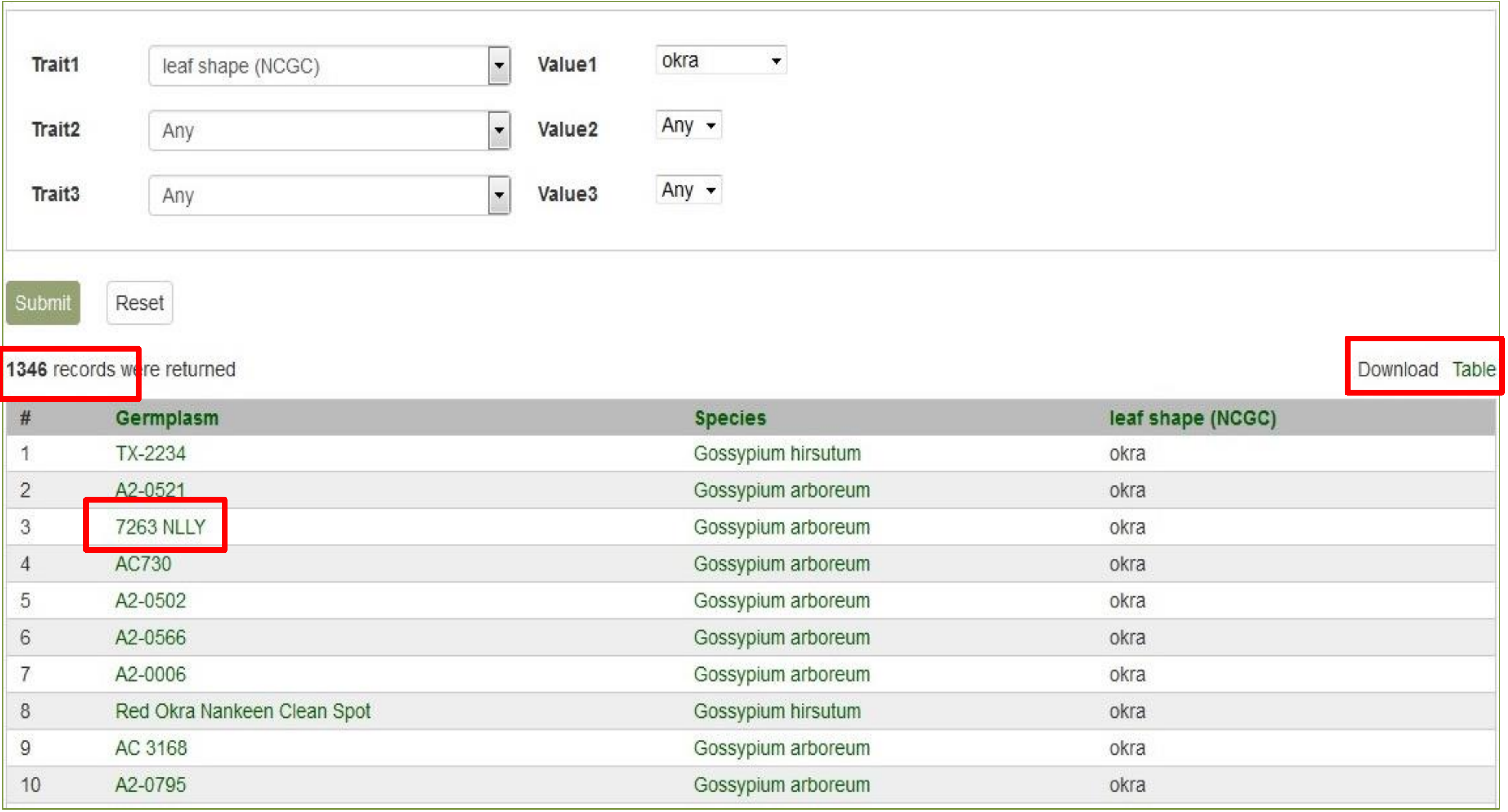
AND

Trait3: Any Value3:

Trait Evaluation Search

→ select trait=leaf shape (NCGC) & value=okra → Search

Find all germplasm with okra leave shape



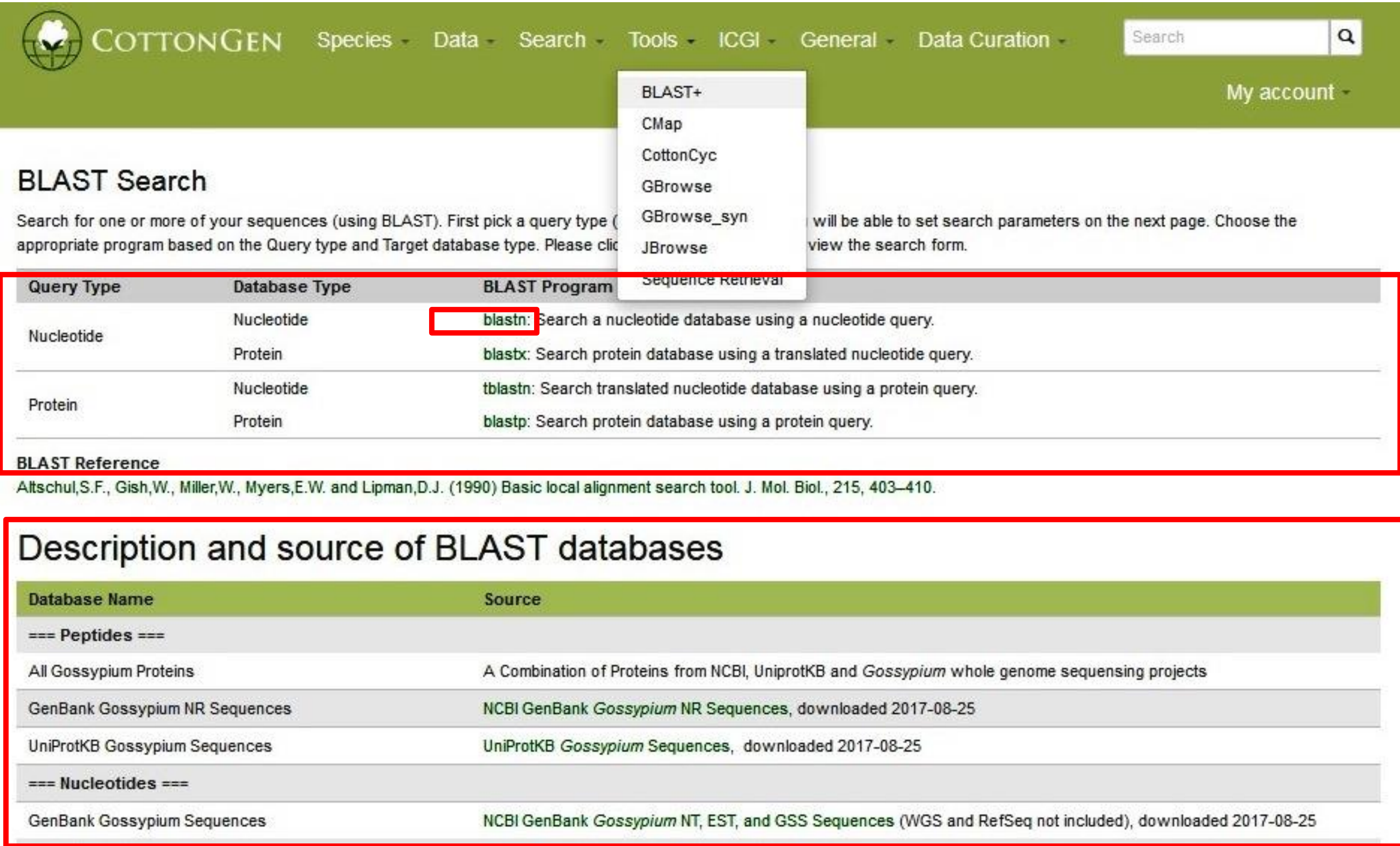
Trait1: leaf shape (NCGC) — Value1: okra
Trait2: Any — Value2: Any
Trait3: Any — Value3: Any

1346 records were returned

#	Germplasm	Species	leaf shape (NCGC)
1	TX-2234	Gossypium hirsutum	okra
2	A2-0521	Gossypium arboreum	okra
3	7263 NLLY	Gossypium arboreum	okra
4	AC730	Gossypium arboreum	okra
5	A2-0502	Gossypium arboreum	okra
6	A2-0566	Gossypium arboreum	okra
7	A2-0006	Gossypium arboreum	okra
8	Red Okra Nankeen Clean Spot	Gossypium hirsutum	okra
9	AC 3168	Gossypium arboreum	okra
10	A2-0795	Gossypium arboreum	okra

Further exploration on germplasm ‘7263 NLLY’

CottonGen BLAST Home Page



BLAST Search

Search for one or more of your sequences (using BLAST). First pick a query type (will be able to set search parameters on the next page. Choose the appropriate program based on the Query type and Target database type. Please clic view the search form.

Query Type	Database Type	BLAST Program
Nucleotide	Nucleotide	blastn: Search a nucleotide database using a nucleotide query.
	Protein	blastx: Search protein database using a translated nucleotide query.
Protein	Nucleotide	tblastn: Search translated nucleotide database using a protein query.
	Protein	blastp: Search protein database using a protein query.

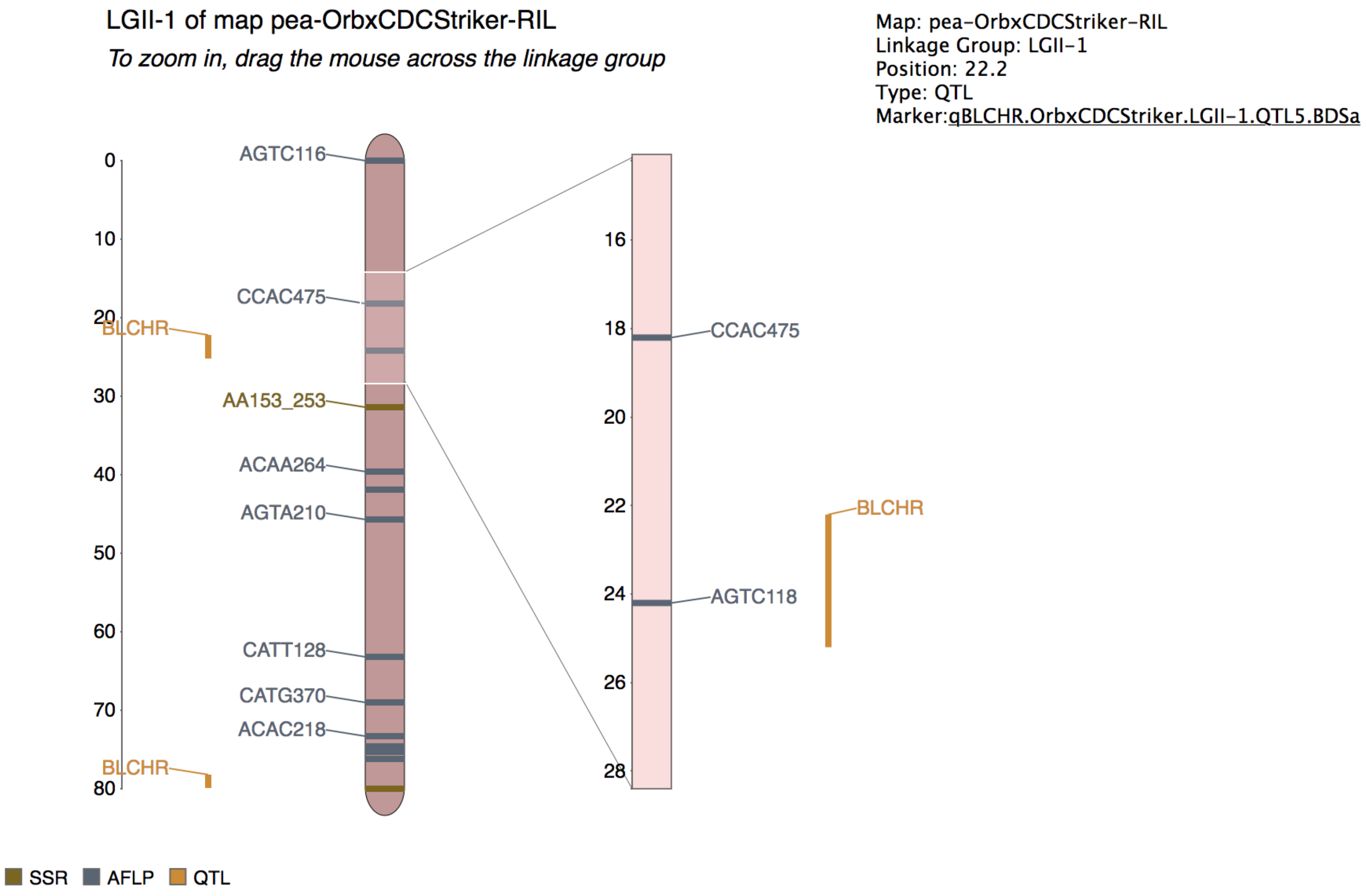
BLAST Reference

Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. J. Mol. Biol., 215, 403–410.

Description and source of BLAST databases

Database Name	Source
=== Peptides ===	
All Gossypium Proteins	A Combination of Proteins from NCBI, UniprotKB and *Gossypium* whole genome sequensing projects
GenBank Gossypium NR Sequences	NCBI GenBank *Gossypium* NR Sequences, downloaded 2017-08-25
UniProtKB Gossypium Sequences	UniProtKB *Gossypium* Sequences, downloaded 2017-08-25
=== Nucleotides ===	
GenBank Gossypium Sequences	NCBI GenBank *Gossypium* NT, EST, and GSS Sequences (WGS and RefSeq not included), downloaded 2017-08-25

Future Work - MapViewer

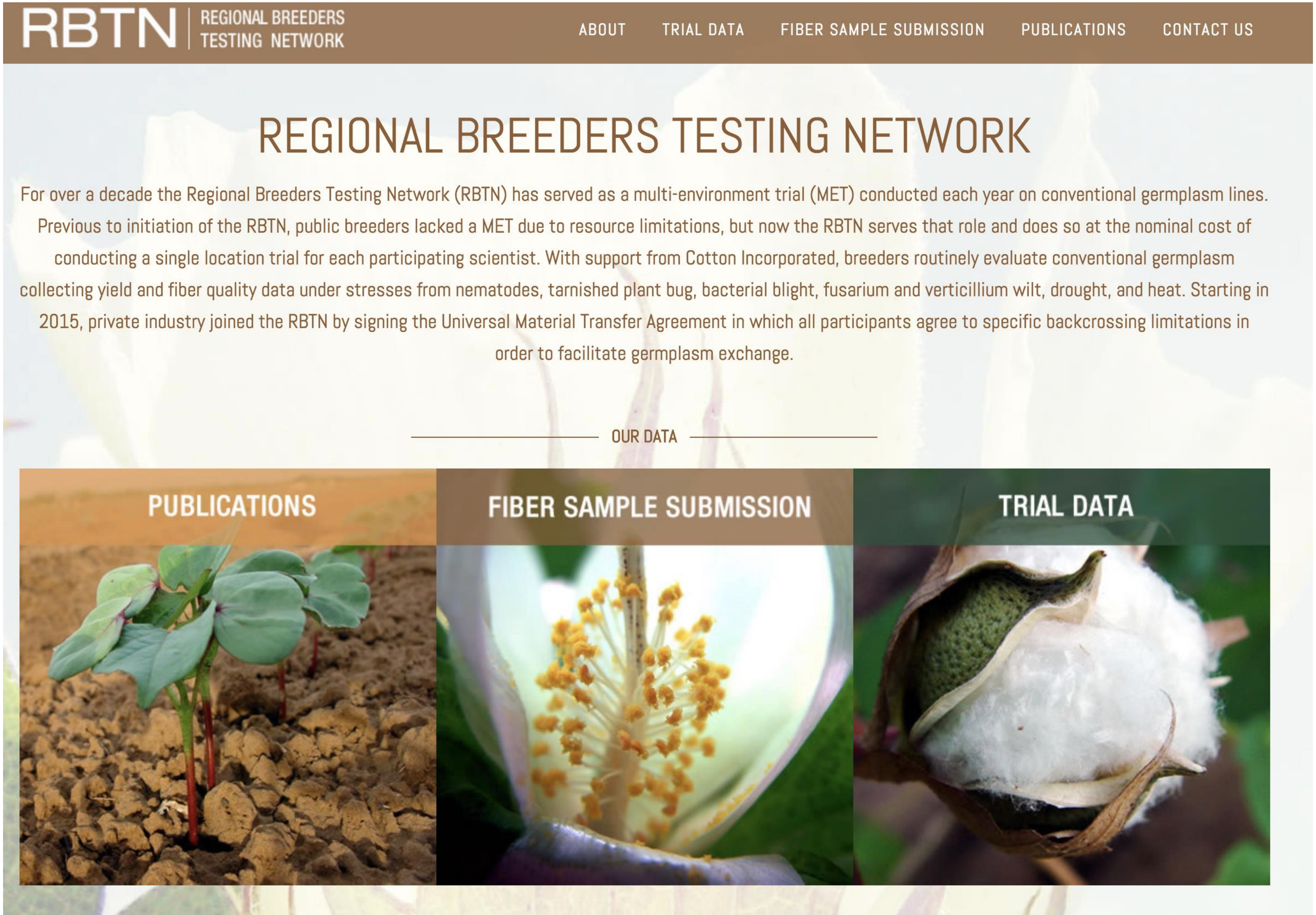


LGII-1 of map pea-OrbxCDCStriker-RIL

To zoom in, drag the mouse across the linkage group

Map: pea-OrbxCDCStriker-RIL
Linkage Group: LGII-1
Position: 22.2
Type: QTL
Marker:qBLCHR.OrbxCDCStriker.LGII-1.QTL5.BDSa

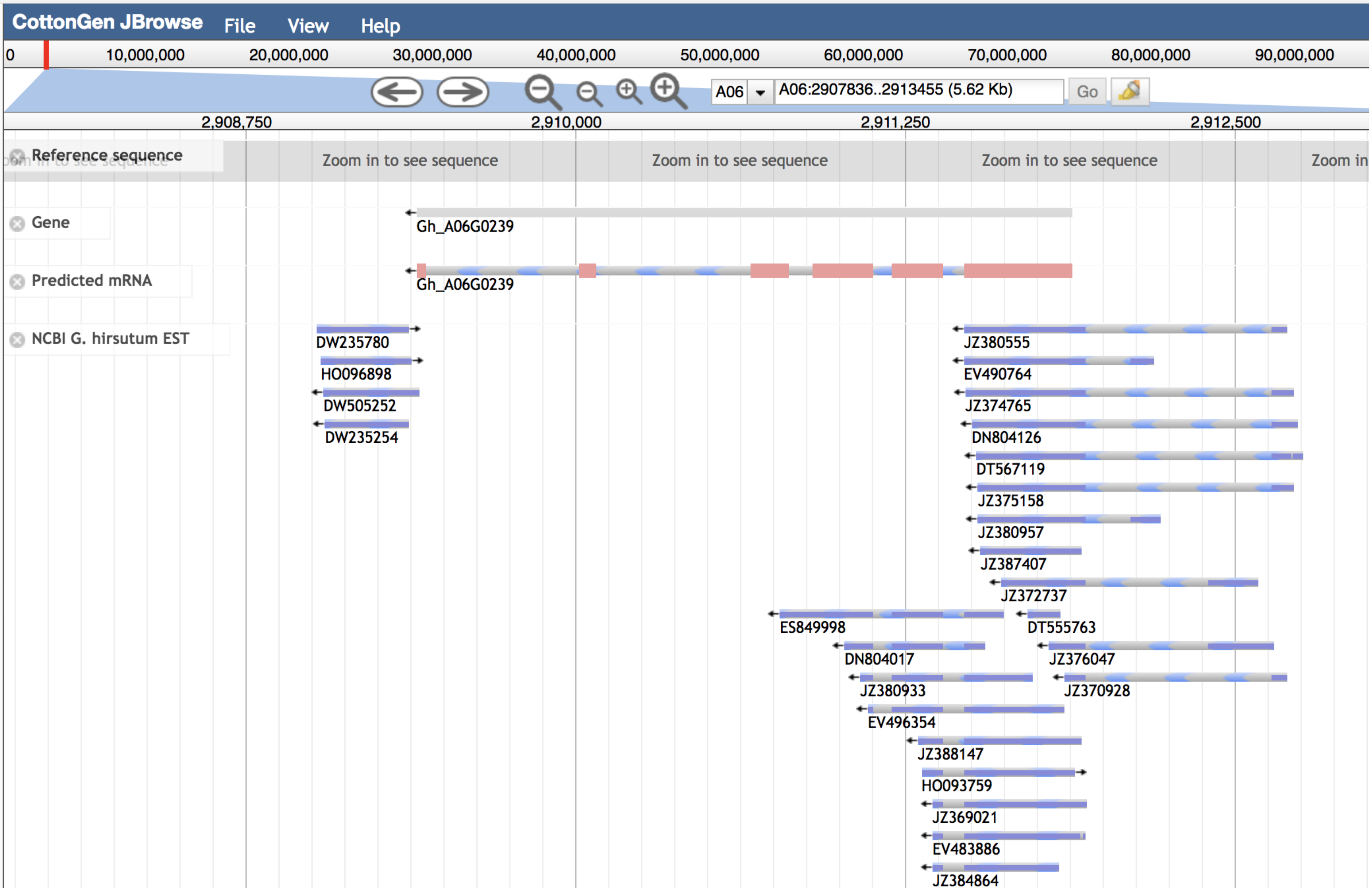
Future Work - 2002-2016 non-fiber trait data



REGIONAL BREEDERS TESTING NETWORK

For over a decade the Regional Breeders Testing Network (RBTN) has served as a multi-environment trial (MET) conducted each year on conventional germplasm lines. Previous to initiation of the RBTN, public breeders lacked a MET due to resource limitations, but now the RBTN serves that role and does so at the nominal cost of conducting a single location trial for each participating scientist. With support from Cotton Incorporated, breeders routinely evaluate conventional germplasm collecting yield and fiber quality data under stresses from nematodes, tarnished plant bug, bacterial blight, fusarium and verticillium wilt, drought, and heat. Starting in 2015, private industry joined the RBTN by signing the Universal Material Transfer Agreement in which all participants agree to specific backcrossing limitations in order to facilitate germplasm exchange.

Future Work – Integration of genomic information



Future Work

- Genomic information
- Adding RBTN 2002-2016 non-fiber trait data
- Further development of cotton breeding tools (BIMS)
- Collection and curation of more genetic maps, markers and QTLs
- Implementing new Map Viewer