

## **CottonGen, a Central Platform to Facilitate Knowledge Discovery in Cotton Research**



www.cottongen.org

Jing Yu<sup>1</sup>, Sook Jung<sup>1</sup>, Chun-Huai Cheng<sup>1</sup>, Taein Lee<sup>1</sup>, Ping Zheng<sup>1</sup>, Katheryn Buble<sup>1</sup>, Jodi L. Humann<sup>1</sup>,

James Crabb<sup>1</sup>, Heidi Hough<sup>1</sup>, B. Todd Campbell<sup>2</sup>, Don C. Jones<sup>3</sup> Josh Udall<sup>4</sup>, and Dorrie Main<sup>1</sup>

1. Washington State University, Pullman, WA, 2. USDA-ARS, Florence, SC, 3. Cotton Incorporated, Cary, NC, 4. USDA-ARS, College Station, TX

## Abstract

CottonGen (www.cottongen.org) integrated significantly more data in 2019: New whole genome assembly data include those from 5 diploids (G. arboreum, G. raimondii, G. turneri, G. austral, Gossypioides kirkii) and 5 tetraploids (G. hirsutum, G. barbadense, G. mustelinum). Cotton sequences, curated from NCBI, has also been updated and now contains over 53K new genes and over 205K new proteins. Seven new (G. arboretum, G. barbadense, G. turneri, G. kirkii, G. austral, G. mustelinum) or updated (G. hirsutum) pathways are available from CottonCyc database. 282 cotton trait descriptors, developed from QTL publications and evaluation data, have been submitted to Crop Ontology (www.cropontology.org). Other new data added in 2019 include 8000 non-fiber traits and 5,000 new fiber trait data from RBTN, 321 QTLs, 594 QTL trait data, 2,230 SSR markers collected from peer-viewed publications. New and updated tools include the interactive and customizable search tool, MegaSearch and new functionalities including SNP genotype

data management in BIMS.

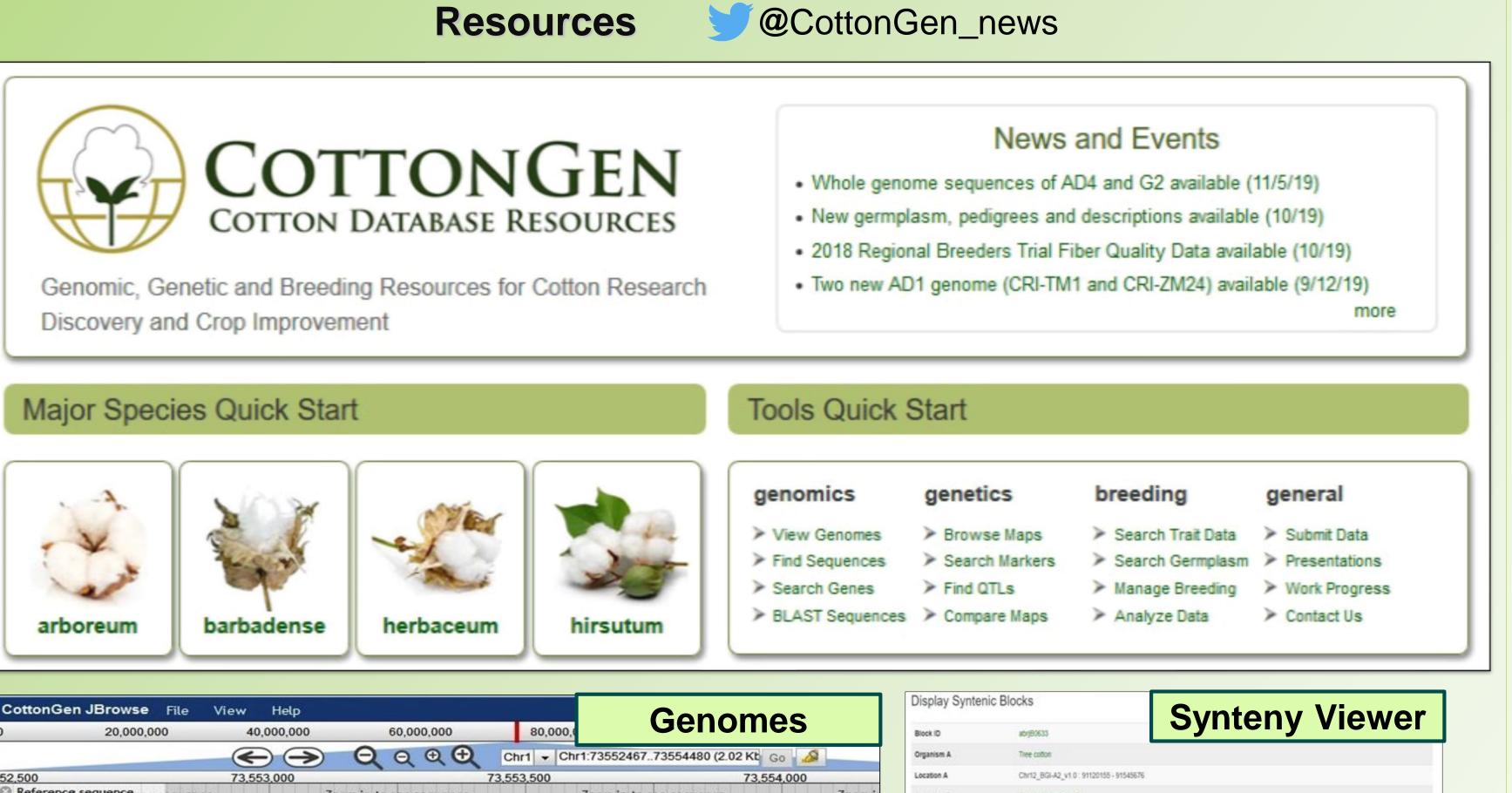
## **Key Features**

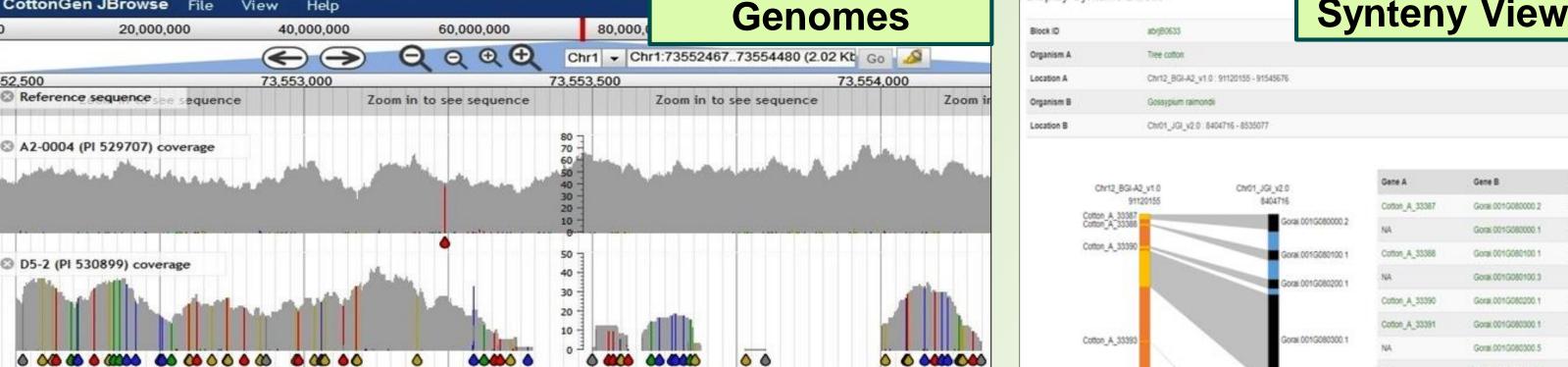
- Easy-to-use search interfaces for retrieving data on markers, genes, transcripts, sequences, QTLs, and traits
- JBrowse and Syntenic Viewer with genome comparisons with orthologs data
- MapViewer for interactive map viewing of genetic maps
- BIMS, Breeder Data Management and Analysis System
- Assembled and annotated reference transcriptome (RefTrans) which is searchable using BLAST
- PathwayCyc with eight species (5 diploids and 3 tetraploids)

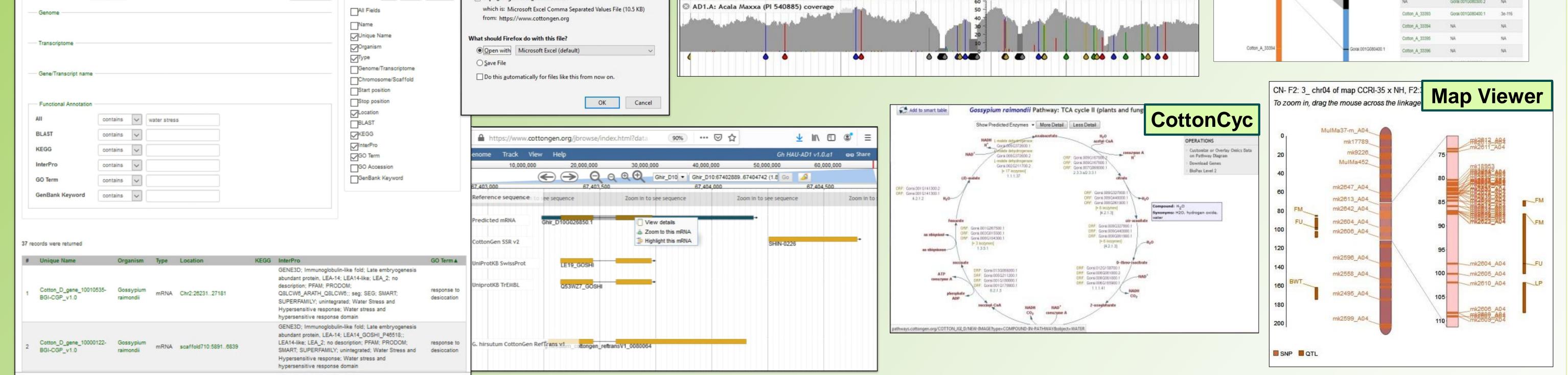
## **Available Data**

- 20 whole genome sequences for 8 diploid and 12 tetraploid cotton
- 575,844 genetic markers and 111 genetic maps
- 5,522 QTLs, 12,337 images, and 19,563 germplasm
- 25,227,605 genotype scores and 383,431 phenotype scores

Data Type Gene/Transcript Reset   0 Gene/Transcript. Note: actual rows in downloaded file depend on the selected fields.	Gene Sea	arch
Query	Downloadable Fields	Opening tripal_megasearch_download.csv
Sequence Type Any V Refresh Count	View FASTA Download	You have chosen to open:







									🚍 Manage Trait			BI
arch Trait Descri	iptors			Trait Sear	ch				AREA (%)	*	Click here to ge	o Data Analysis $\rightarrow$ Statistical Analysis to view stats from many re-
p Any ~									BWT (g)		Nama	MIC (unit)
									E1 (%)		Name	N/A
gory morphological trai	it ~								FUI (%) FUR (%)		Alias	numeric
vord contains 🗸	bract					Germplasm	bract color		GTO (%)	E	Tornat	
					tion, CS12 D1-35 US_TX_CS_2012 2413	D1-35	green		LI (g)			MIC (micronaire) is an HVI measurement, the fineness of the sample t from the ginned lint, measured by a Fibronaire and expressed in stan
rch Reset					tion, CS15 Assili US_TX_CS_2015 2101	Assili	green		LP (%)		Definition	(curvilinear scale) micronaire units.
ords were returned				Download Tab	tion, CS15 Sake US_TX_CS_2015 2102	Sake	green		LY (lb/acre)			
ius were returned				Download Tac	ition, CS14[Maarad US_1X_CS_2014]2906[[]	Maarad	green		MIC (unit) QS (1,1,3,5)			data_unit = unit
Group Descriptor	Category	Trait	Definition		tion, CS15 Maarad US_TX_CS_2015 2223	Maarad	green		QS (2,1,4,3)		Properties	
NCGC bract color	morphological trait	bract color	bract color descriptor code definitio	on: 1=green; 2=red; 3=sun red; 9=seg/off type	tion, CS14 GB-1425 US_TX_CS_2014 2907	GB-1425	green		QS (4.5,2.5,0,3)			
	morphological		bract nectaries descriptor code def	finition: 0=absent; 1=present; 2=reduced; 3=inactive;	tion, CS14 Giza 3 US_TX_CS_2014 2908	Giza 3	green		QS (5,2.5,1,1.5)	+		
NCGC bract nectaries	trait	bract nectaries	9=seg/off type		tion, CS15 Giza 3 US_TX_CS_2015 2103	Giza 3	green		Click a trait to see details			MIC (unit)
NCGC bract teeth	morphological	bract teeth	bract teeth number descriptor code	e definition: 1=medium; 2=many; 3=few; 9=seg/off type	tion, CS15 GB-1433 US_TX_CS_2015 2106	GB-1433	green				450	
number	trait morphological	number		efinition: 0=none; 1=medium; 2=small; 3=large; 9=seg/off	tion, CS15 GB-1434 US_TX_CS_2015 2107	GB-1434	green				400 350	
NCGC bract teeth size	trait	bract teeth size	type	sindon, o-none, i-medium, z-sinan, s-large, s-segrun	tion, CS14 GB-1435 US_TX_CS_2014 2913	GB-1435	green				300	
NCGC bract type	morphological	bract type	bract type descriptor code definition	n: 1=normal; 2=frego; 3=flared; 4=recurved; 9=seg/off type	tion, CS14 GB-1436 US_TX_CS_2014 2914	GB-1436	green				250	
	trait				tion, CS14 Pima 67 US_TX_CS_2014 2917	Pima 67 (GB-1441)					200	
CN_COT bract shape	morphological trait	bract shape	6=trangle, 7=three teeth	ordate, 2=narrow coil, 3=ovule, 4=sword, 5=lanceolate,	tion, CS15 GB-1442 US_TX_CS_2015 2108	GB-1442	green				150	
					aluation, CS15 GB-1443 US_TX_CS_2015 2109    aluation, CS15 Giza 31 US_TX_CS_2015 2110	GB-1443	green					
					aluation, CS14 Giza 31 (PI 407500) US_TX_CS_2014 1424	Giza 31 (GB-1444) Giza 31 (PI 407500)	Contraction of the second seco		BRACT COLOR			
Descriptor Over	view				aluation, CS14 GB-1448 US_TX_CS_2014 2921	GB-1448	green	4				
Trait Descriptor	bract color				aluation, CS15 GB-1449 US_TX_CS_2015 2111	GB-1449	sun red	EASY SQUAREJR		EASY SQUAR	EJR.	ે મેં મેં છે
Of Trait	bract color				aluation, CS15 CNH-83-69 US_TX_CS_2015 2112	CNH-83-69	green	EAST SOL CUT	EASY AREJR.	by SHABOD	X -	
Definition		intor code definition:	1=green; 2=red; 3=sun red; 9=seg/off		aluation, CS15 GB-1455 US_TX_CS_2015 2116	GB-1455	green			TAL		Max Min Mean STD
Alias	[view all 1]	iptor code demition.	T-green, 2-reu, 3-sun reu, 3-segron	.,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,	aluation, CS15 GB-1456 US_TX_CS_2015 2117	GB-1456	green	A ST	2-1-+++			6.17 2.83 4.84 0.40
Format	categorical				aluation, CS14 GB-1457 US_TX_CS_2014 2923	GB-1457	green					
Trait Category	morphological tra	uit			aluation, CS14 Giza 30 (P8-75) US_TX_CS_2014 2925	Giza 30 (P8-75)	green				17	
Datasets	[view all 18]				aluation, CS14 Pima Direx US_TX_CS_2014 2926	Pima Direx	green		WEST WARREN, MA 01092	+++++		
										DED (2)	22	
Germplasm	[view all 5154]			Page 1 2 3	4 5 6 7 8 9 10 11 12 13	14   15     Last		GREEN (1)	SUN RED (3)	RED (2)		

ch Trait Descrip	otors			Trait Sac	roh					Click have to a	o Data Analysis → Statistical Analysis to view stats from n
				Trait Sea					AREA (%) BWT (g)	Click here logo	· Data Pata your - Dialistical Patalysis to blew stats fi Olit In
Any ~								1	E1 (%)	Name	MIC (unit)
morphological trait	~								FUI (%)	Alias	N/A
						Germplasm	bract color		FUR (%)	Format	numeric
contains v br	ract				tion, CS12 D1-35 US_TX_CS_2012 2413	D1-35	green		GTO (%)		MIC (micronaire) is an HVI measurement, the fineness of the s
Reset					tion, CS15 Assili US_TX_CS_2015 2101	Assili	green		LI(g) LP(%)	Definition	from the ginned lint, measured by a Fibronaire and expressed (curvilinear scale) micronaire units.
					tion, CS15 Sake US_TX_CS_2015 2102	Sake	green		LY (lb/acre)		
were returned				Download	Table tion, CS14 Maarad US_TX_CS_2014 2906	Maarad	green		MIC (unit)		
oup Descriptor	Category	Trait	Definition		tion, CS15 Maarad US_TX_CS_2015 2223	Maarad	green		QS (1,1,3,5)		data_unit = unit
GC bract color	morphological			n: 1=green; 2=red; 3=sun red; 9=seg/off type	tion, CS14 GB-1425 US_TX_CS_2014 2907	GB-1425	green		QS (2,1,4,3)	Properties	
	trait	bract color			tion, CS14 Giza 3 US_TX_CS_2014 2908	Giza 3	green		QS (4.5,2.5,0,3) QS (5,2.5,1,1.5)		
GC bract nectaries	morphological trait	bract nectaries	bract nectaries descriptor code defin 9=seg/off type	nition: 0=absent; 1=present; 2=reduced; 3=inactive;	tion, CS15 Giza 3 US_TX_CS_2015 2103	Giza 3	green		Click a trait to see details		MIC (unit)
GC bract teeth	morphological	bract teeth		definition: 1=medium; 2=many; 3=few; 9=seg/off type	tion, CS15 GB-1433 US_TX_CS_2015 2106	GB-1433	green			450	
number	trait	number			tion, CS15 GB-1434 US_TX_CS_2015 2107	GB-1434	green			400	
GC bract teeth size	morphological trait	bract teeth size	bract teeth size descriptor code defi type	inition: 0=none; 1=medium; 2=small; 3=large; 9=seg/off	tion, CS14 GB-1435 US_TX_CS_2014 2913	GB-1435	green			350	
CO hastbar	morphological	here at the sec		Amazanak Onferna Onferna konstruction ( 7)	tion, CS14 GB-1436 US_TX_CS_2014 2914	GB-1436	green			300	
GC bract type	trait	bract type		: 1=normal; 2=frego; 3=flared; 4=recurved; 9=seg/off typ	uon, C314 Pima 07 03_1X_C3_2014 2917	Pima 67 (GB-1441)	green			250 200	
_COT bract shape	morphological trait	bract shape	Bract shape. Code Definition: 1=core 6=trangle, 7=three teeth	date, 2=narrow coil, 3=ovule, 4=sword, 5=lanceolate,	tion, CS15 GB-1442 US_TX_CS_2015 2108	GB-1442	green			150	
	udit		o-uangie, r-unee teeth		evanuation, CS15 GB-1443 US_TX_CS_2015 2109	GB-1443	green	I		150	
					evaluation, CS15 Giza 31 US_TX_CS_2015 2110	Giza 31 (GB-1444)	green		BRACT COLOR		
Descriptor Overvi	view				evaluation, CS14 Giza 31 (PI 407500) US_TX_CS_2014 1424	Giza 31 (PI 407500)		4 6 4			
					evaluation, CS14 GB-1448 US_TX_CS_2014 2921	GB-1448	green		EAS	Y SQUAREJR.	5. ×. ×. 5.
Trait Descriptor	bract color				evaluation, CS15 GB-1449 US_TX_CS_2015 2111	GB-1449	sun red	EASY SOULA EJR	R. EASY AREJR.	SHABOUT STATISTICS	
Of Trait	bract color				evaluation, CS15 CNH-83-69 US_TX_CS_2015 2112    evaluation, CS15 GB-1455 US_TX_CS_2015 2116	CNH-83-69 GB-1455	green	by a start	by SF		Max Min Mean
Definition		ptor code definition:	1=green; 2=red; 3=sun red; 9=seg/off ty		evaluation, CS15 GB-1455 US_TX_CS_2015 2117	GB-1455	green				6.17 2.83 4.84
Alias	[view all 1]				evaluation, CS14 GB-1457 US_TX_CS_2014 2923	GB-1450 GB-1457	green green				
Format	categorical	14			evaluation, CS14 Giza 30 (P8-75) US_TX_CS_2014 2925	Giza 30 (P8-75)	green				
Trait Category	morphological tra	it			evaluation, CS14 Pima Direx US_TX_CS_2014 2926	Pima Direx	green		WEST WARREN, MA 01092		
Datasets	[view all 18]								www.azquit.com	WEST WARDEN WARDER	
Germplasm	[view all 5154]			Page 1 2	3 4 5 6 7 8 9 10 11 12 13	[ 14 ] [ 15 ] [ Last ]		GREEN (1)	SUN RED (3)	RED (2)	