

# Data search and visualization tools at the Comparative Evolutionary Genomics of Cotton Web resource

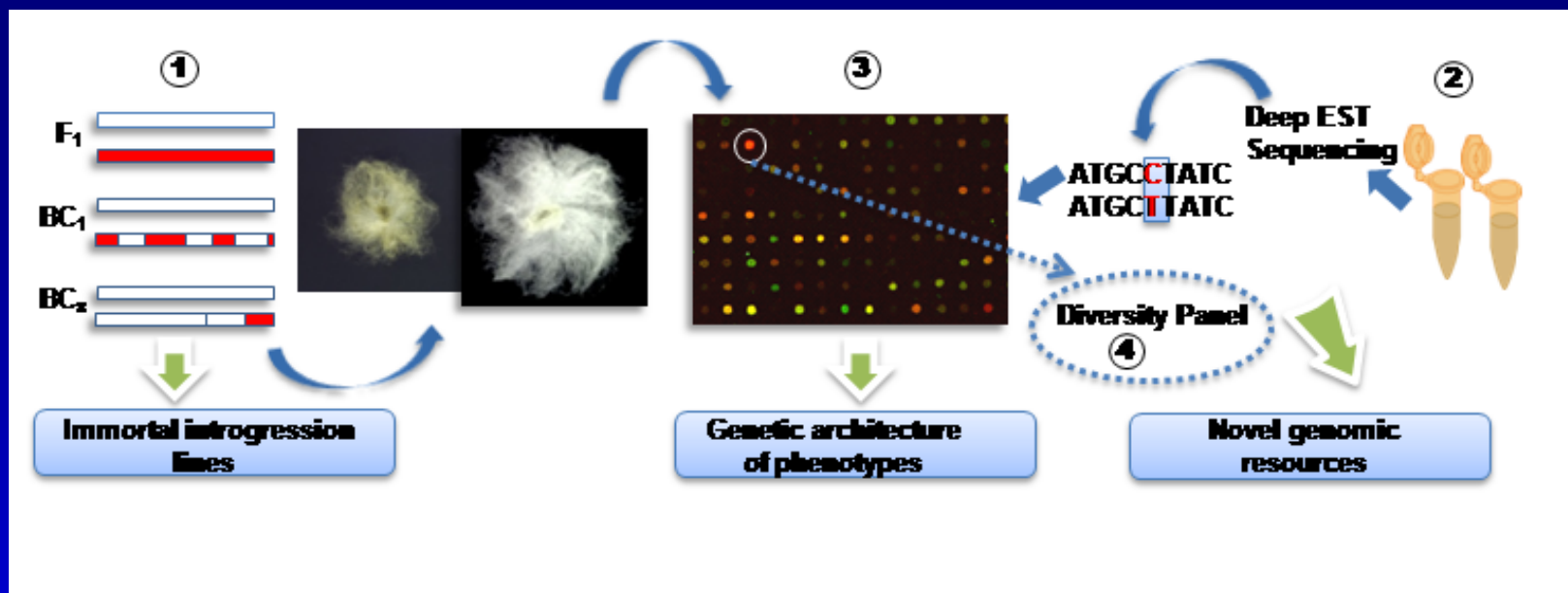
*Alan R. Gingle*  
*Andrew H. Paterson*  
*Joshua A. Udall*  
*Jonathan F. Wendel*

# CEGC project goals

## *set the context*

- **Develop and characterize immortal introgression populations, to reduce complex morphology into defined constituents amenable to functional genomic analyses.**
- **Develop a novel homoeolog-specific comparative expression profiling platform using a vastly enriched EST resource.**
- **Reveal perturbations in genetic networks and gene expression associated with naturally occurring variation in fiber phenotypes using the introgression lines.**
- **Provide a foundation for understanding the effects of selection on genetic diversity in cotton.**

# Experimental Plan: *major components and related data types*

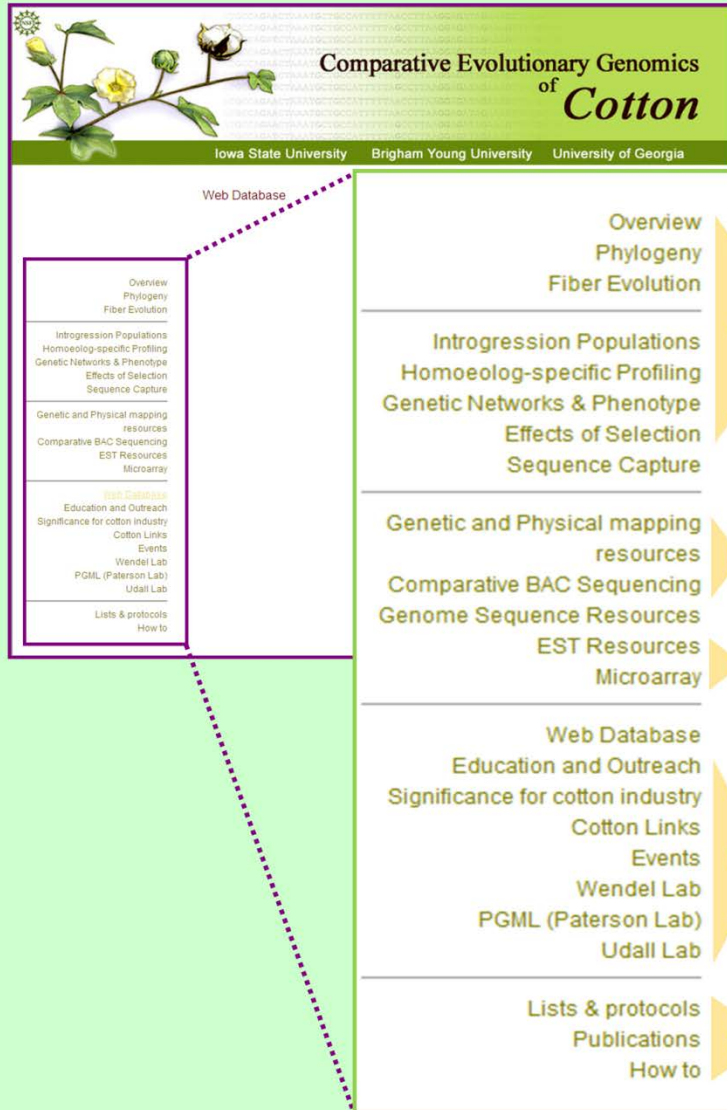


- **1. Near-isogenic introgression lines:** genetic map, marker and fiber quality phenotypic data
- **2. EST sequencing & assembly:** both Sanger and NGS (454) read/quality and SNP polymorphism data
- **3. Homoeolog-specific expression profiling:** microarray and RNAseq
- **4. Diversity & effects of selection:** microarray based sequence capture, NGS and SNP polymorphism data

# CEGC Web overview

<http://cottonrevolution.info>

## Data and Tools



The CEGC search, display and download tools for cotton data are accessible via the sidebar menu shown in both the page context and magnified views.

**Overview - Effects of Selection** provide general information on our research activities, including overviews, background and strategies by topic area.

**Sequence Capture** provides eSNP and SSR diversity data/tools. (*upcoming*)

**Genetic and Physical mapping resources** includes an implementation of CMap.

### Genome Sequence Resources

**EST Resources – Microarray** provide EST assembly comparative annotation and gene expression data/tools.

**Web Database** includes a complete list of our data search and display tools.

**Education and Outreach – Udall Lab** provide outreach as well as a range of other project related information and links.

**Lists & protocols – How to** contain protocols, publications and additional information related to accessing our project deliverables.

# Search & data display tools

- A battery of search and display tools have been integrated into the CEGC Web resource.
- The available tools are listed on the site's "Web Database" page, shown on the right.
- Community developed tools like Bioconductor, CMap, GBrowse, SMD and wwwBLAST have been integrated to provide familiar views and functionality.

## Web Database

---

[Overview](#) | [People](#) | [Publications](#) | [Further Information](#) | [Web Statistics](#)

---

### Overview

We have developed this Web-database resource (<http://cottonrevolution.info>) to provide access to project data, information and materials. It includes components and tools for each of the project focus areas. It also provides convenient access to project materials such as our spotted oligo microarrays and EST data, including new NGS read sequence and quality data. We are building upon this resource to enhance data availability, integration and presentation. Links to our developing battery of search and visualization tools are listed below.

#### EST tools:

- BLAST to our EST assembly
- Search and view our EST assemblies
- View contig-to-*Arabidopsis* alignments

#### Genetic and Physical Map tools:

- Consensus map
- Mapped objects portal
- Physical mapping database.

#### Microarray tools:

- Arrayed objects portal
- Microarray expression database

#### Other tools:

- Access to the [G. raimondii draft sequence](#) using blast or a genome browser.



# Search Tools

Objects ☒ is an EST ☐ has BAC hits [Help](#)

Name keyword:

Show objects from 1 to 15 | page 1, begins with GA\_Ea0001A01.f

Name	EST Data	Genetic Map	BAC Hits (GaD)	BAC Hits (Gcnh)	BAC Hits (Gr)	BAC Hits (maxxa)
GA_Ea0001A01.f	<a href="#">get EST Data</a>	<a href="#">get Map Data</a>				
GA_Ea0001A03.f	<a href="#">get EST Data</a>	<a href="#">get Map Data</a>				
GA_Ea0001A09.f	<a href="#">get EST Data</a>	<a href="#">get Map Data</a>				
GA_Ea0001A10.f	<a href="#">get EST Data</a>	<a href="#">get Map Data</a>				
GA_Ea0001A14.f	<a href="#">get EST Data</a>	<a href="#">get Map Data</a>				
GA_Ea0001A15.f	<a href="#">get EST Data</a>	<a href="#">get Map Data</a>				
GA_Ea0001A17.f	<a href="#">get EST Data</a>	<a href="#">get Map Data</a>				
GA_Ea0001B01.f	<a href="#">get EST Data</a>	<a href="#">get Map Data</a>				
GA_Ea0001B11.f	<a href="#">get EST Data</a>	<a href="#">get Map Data</a>				

Search for cotton contigs that have BLAST hits (E-values  $\leq 1e-10$ ) to *Arabidopsis Thaliana* genes (TAIR9) with any of the following annotation categories and terms. Select the desired criteria and then click on .

TAIR9 annotation category ☐ functions in ☐ involved in ☐ located in

Term

INVOLVED\_IN: cell wall modification during multidimensional cell growth

Show objects from 1 to 15 | page 1, begins with contig34932

Contig Name	Assembly	E-value	Best BLAST hit to <i>Arabidopsis</i>
contig34932	Cotton32	1e-77	<a href="#">contig34932 190-504 Chr2 12432826-12433140 -76 86.66 147 36</a>
contig31723	Cotton32	3e-61	<a href="#">contig31723 238-521 Chr2 12432833-12433116 -60 85.56 120 23</a>
Contig_40924	Cotton46	1e-59	<a href="#">Contig_40924 332-615 Chr2 12432833-12433116 -58 85.21 117 35</a>
Contig_44628	Cotton46	1e-56	<a href="#">Contig_44628 336-619 Chr2 12432833-12433116 -56 84.85 112 34</a>
Contig_35234	Cotton46	5e-49	<a href="#">Contig_35234 367-689 Chr2 12432823-12433145 -48 82.66 99 46</a>
contig63031	Cotton32	8e-48	<a href="#">contig63031 259-579 Chr2 12432823-12433143 -47 82.55 97 41</a>
Contig_28833	Cotton46	1e-47	<a href="#">Contig_28833 252-608 Chr3 743487-743843 -46 81.79 97 38</a>
Contig_52888	Cotton46	1e-47	<a href="#">Contig_52888 645-965 Chr2 12432823-12433143 -46 82.55 97 25</a>
contig63032	Cotton32	1e-47	<a href="#">contig63032 259-579 Chr2 12432823-12433143 -46 81.08 96 41</a>

Selection criteria:

☒ top 100 lint yield ☐ top 100 lint % ☐ top 100 micronaire ☐ top 100 length uhm

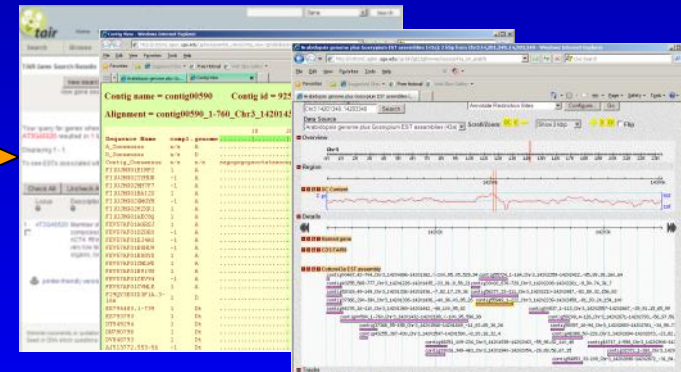
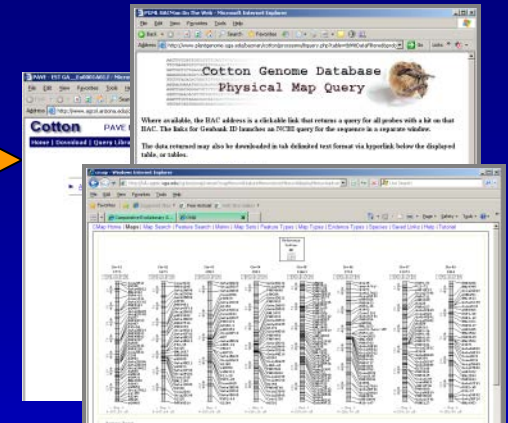
☐ top 100 unif index ☐ top 100 strength hvi ☐ top 100 elongation ☐ top 100 seed index

any state any year

Name keyword:

Show objects from 1 to 15 | page 1, begins with 8839-3-10-2

genotype	datasets	max lint yield	max lint %	max micronaire	max length uhm	max strength hvi
8839-3-10-2	GA-01	1935	45.2	5.1	1.15	29.4
ACSI Fibermex FM 958	MS-00	1877	40.3	5.3	1.18	33.4
BCG 28 R	MS-04	1906	43	5.37	1.16	31.2
BCG 28R	MS-02 NC-02 NC-04	1823	43	32.9	74	84.6
DP 393	MS-04 MS-06	1934	43.8	5.13	1.19	33.53
DP 432 RR	MS-04 MS-06	2212	42.07	5.3	1.17	31.5
DP 434 RR	MS-04 MS-06	2027	44.39	5	1.22	29.38
DP 444 BG/RR	MS-04 MS-06	2239	43.16	4.7	1.18	30.83
DP 445 BG/RR	MS-04 MS-06	2054	43.48	5.13	1.2	32.67
DP 449 BG/RR	MS-04 MS-06	2023	41.35	5.4	1.17	34.57
DP 451 BG/RR	MS-04	1911	38.68	5.2	1.17	29.43
DP 455 BG/RR	MS-04 MS-06 NC-04	2029	43.92	4.93	1.21	34.27
DP 488 BG/RR	MS-04 MS-06	2284	43.25	5.4	1.21	33.3
DP 491	MS-01 MS-04	2036	44.19	5.3	1.27	34.77
DP 494 RR	MS-04 MS-06	1905	43.36	5.33	1.22	34.4



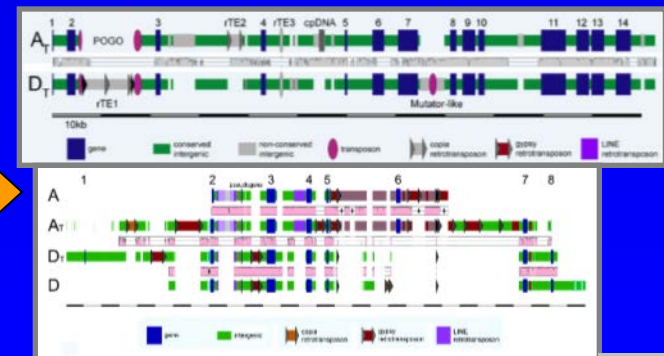
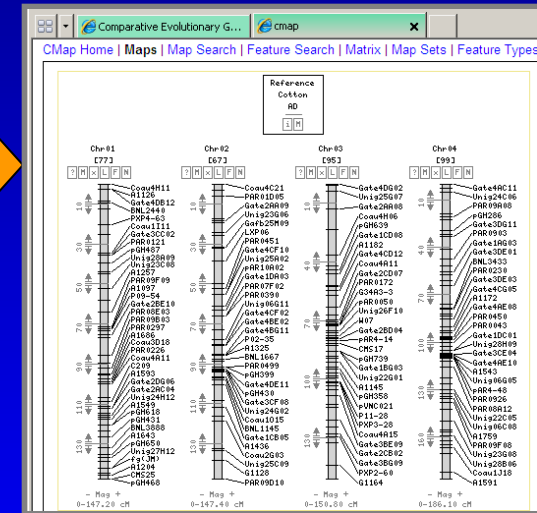
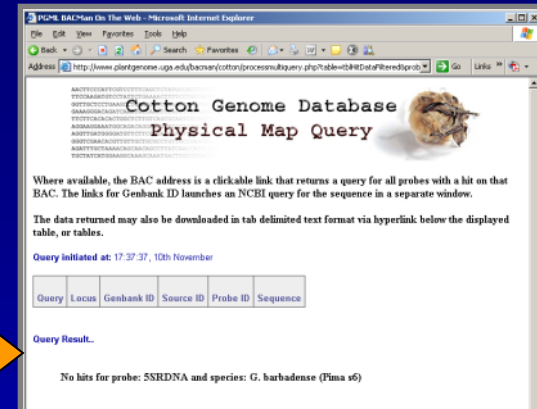
Search tools link to external and collaborator Web resources for a wide range of data types.

# Data displays

- The CEGC search tools link to collaborator implemented display tools like BACman for BAC, CMap for genetic map and GBrowse plus ContigView for EST related data.
- In addition, CEGC search interfaces link to external databases like GO, Pfam, InterPro and TAIR.
- SMD has been implemented and its data/information displays are available for microarray related expression data.

# Genetic map and BAC related resources

- Our mapped objects search tool links to BACman for hybridization data and to a CMap display for genetic map data (Paterson lab).
- BAC sequence alignment views are displayed in the resource's "Comparative BAC Sequencing" component.





# EST assembly

Select assembly  and

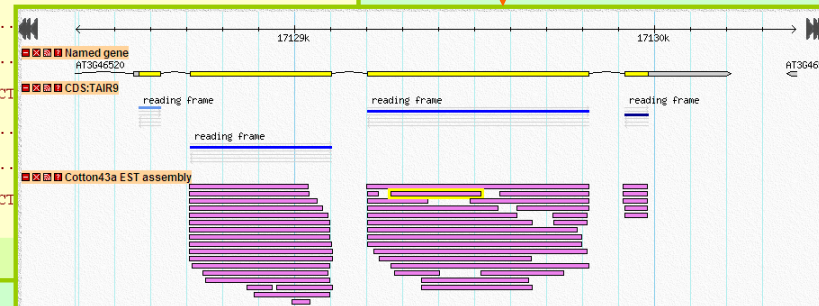
Showing objects on page 1501 | page 1501 begins with contig 22029

Contig Name	Best BLAST hit to <i>Arabidopsis</i>
<a href="#">contig22029 (Cotton43a)</a>	<a href="#">contig22029</a> 320-431 Chr4 11992938-11993049 -8 82.14 32 9
<a href="#">contig22030 (Cotton43a)</a>	<a href="#">contig22030</a> 163-311 Chr2 7988711-7988859 -30 86.57 69 22
<a href="#">contig22031 (Cotton43a)</a>	
<a href="#">contig22032 (Cotton43a)</a>	
<a href="#">contig22033 (Cotton43a)</a>	<a href="#">contig22033</a> 8-57 Chr2 829454-829503 -5 88 26 37
<a href="#">contig22034 (Cotton43a)</a>	
<a href="#">contig22035 (Cotton43a)</a>	<a href="#">contig22035</a> 87-229 Chr1 1290844-1290986 -15 82.51 43 51
<a href="#">contig22036 (Cotton43a)</a>	
<a href="#">contig22037 (Cotton43a)</a>	<a href="#">contig22037</a> 181-250 Chr2 12618622-12618691 -12 88.57 38 21
<a href="#">contig22038 (Cotton43a)</a>	
<a href="#">contig22039 (Cotton43a)</a>	<a href="#">contig22039</a> 536-741 Chr3 2533381-2533586 -30 83.49 70 16
<a href="#">contig22040 (Cotton43a)</a>	
<a href="#">contig22041 (Cotton43a)</a>	
<a href="#">contig22042 (Cotton43a)</a>	<a href="#">contig22042</a> 62-297 Chr4 14940344-14940579 -15 79.66 44 22
<a href="#">contig22043 (Cotton43a)</a>	

BLAST alignment: Contig\_begin-end\_At chr\_begin-end\_Evalue\_%ident\_score\_%CigLength

Contig name = contig02310    Contig id = 94262    Assembly\_id = 220  
 Alignment = contig02310\_1-252\_Ch3\_17129266-17129517\_-20\_80.15\_52\_100

Sequence Name	compl.	genome	10	20	30	40	50	60	70
A_Consensus	n/a	A	no_a_consensus_sequence						
D_Consensus	n/a	D	TCCACATGCCATCCTCCGCTAGACCTCGCAGGTCGTGACCTTACTGATGCCCTTATGAAAATCCTG-ACCGAGCGTGTTATTC						
Contig_Consensus	n/a	n/a	TCCACATGCCATCCTCCGCTAGACCTCGCAGGTCGTGACCTTACTGATGCCCTTATGAAAATCCTG-ACCGAGCGTGTTATTC						
GR_Eb023019.r.520-768.fm2311	1	D	TCCACATGCCATCCTCCGCTAGACCTCGCAGGTCGTGACCTTACTGATGCCCTTATGAAAATCCTG-ACCGAGCGTGTTATTC						
F29QY3E01BWZ9	-1	D	.....TGCCATCCTCCGCTAGACCTCGCAGGTCGTGACCTTACTGATGCCCTTATGAAAATCCTG-ACCGAGCGTGTTATTC						
F29QY3E01C8KU4.30-264.fm2311	1	D	TCCACATGCCATCCTCCGCTAGACCTCGCAGGTCGTGACCTTACTGATGCCCTTATGAAAATCCTG-ACCGAGCGTGTTATTC						
F29QY3E01BDWOW.6-200	1	D	.....						
F29QY3E01CF8FA.6-182	1	D	.....						
F29QY3E01D9ASQ.162-1.fm2311	-1	D	TCCACATGCCATCCTCCGCTAGACCTCGCAGGTCGTGAC-ITACT						
F29QY3E01BC8E6.6-116	1	D	.....						
F29QY3E01D0YZ3.6-98	1	D	.....						
DT554115.472-699.fm2311	1	H	TCCACATGCCATCCTCCGCTAGACCTCGCAGGTCGTGACCTTACT						



Our contig search tool facilitates searches for contig structure and significant alignments to the *Arabidopsis* genome (TAIR10).

The search tool facilitates searches by assembly and contig.

Contig structure is displayed by the contig alignment viewer developed in the Udall lab.

The GMOD tool, GBrowse has been implemented to display significant *Gossypium* EST contig BLAST alignments to *Arabidopsis* genomic sequence (TAIR10).

# Functional annotation and expression data

Search for cotton contigs that have BLAST hits (E-values <= 1e-50) to *Arabidopsis Thaliana* genes (TAIR9) with any of the following annotation categories and terms. Select the desired criteria and then click on Search.

TAIR9 annotation category ☐ functions in ☒ involved in ☐ located in

Term cell wall modification during multidimensional cell growth

INVOLVED\_IN: cell wall modification during multidimensional cell growth

Show objects from 1 to 15 | page 1, begins with contig03773

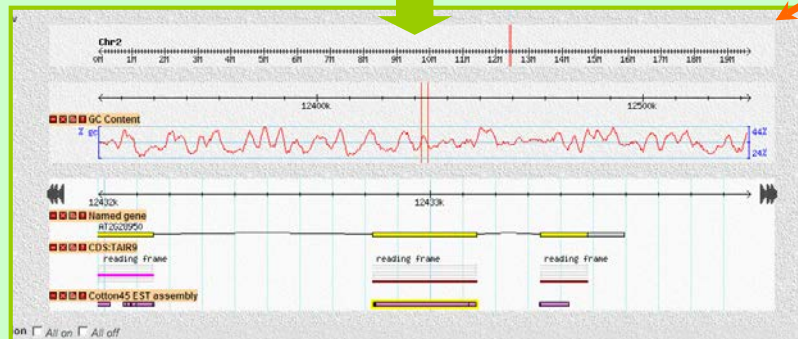
Contig Name	Assembly	E-value	Best BLAST hit to <i>Arabidopsis</i>
contig03773	Cotton45	6e-80	contig03773 419-733 Chr2 12432826-12433140 -79 86.98 151 34
contig26753	Cotton43	1e-77	contig26753 379-693 Chr2 12432826-12433140 -76 86.66 147 36
contig34932	Cotton32	1e-77	contig34932 190-504 Chr2 12432826-12433140 -76 86.66 147 36
Contig_9641	Cotton42	2e-61	Contig_9641 206-489 Chr2 12432833-12433116 -60 85.56 120 35
contig31723	Cotton32	3e-61	contig31723 238-521 Chr2 12432833-12433116 -60 85.56 120 23
Contig_40924	Cotton46	1e-59	Contig_40924 332-615 Chr2 12432833-12433116 -58 85.21 117 35
contig18013	Cotton43	6e-59	contig18013 238-521 Chr2 12432833-12433116 -58 85.21 116 23
Contig_44628	Cotton46	1e-56	Contig_44628 336-619 Chr2 12432833-12433116 -56 84.85 112 34
contig02957	Cotton45	1e-56	contig02957 250-533 Chr2 12432833-12433116 -56 84.85 112 29
contig37678	Cotton45	1e-56	contig37678 250-533 Chr2 12432833-12433116 -56 84.85 112 25
contig35445	Cotton45	2e-56	contig35445 250-533 Chr2 12432833-12433116 -55 84.85 112 22
contig37677	Cotton45	2e-56	contig37677 250-533 Chr2 12432833-12433116 -55 84.85 112 23
contig37330	Cotton45	2e-53	contig37330 177-486 Chr2 12432831-12433140 -52 83.54 106 63
contig56240	Cotton45	4e-53	contig56240 177-486 Chr2 12432831-12433140 -52 83.54 106 32
contig99231	Cotton45	4e-53	contig99231 177-486 Chr2 12432831-12433140 -52 83.54 106 33

BLAST alignment: Contig\_begin-end\_At chr\_begin-end\_Evalue\_%ident\_score\_%CtgLength

Our annotation search tool facilitates searches for *Gossypium* EST contigs based on *Arabidopsis* annotations (TAIR10).

The annotation search tool employs our database, containing the TAIR10 annotations for *Arabidopsis* genes and *Arabidopsis-Gossypium* associations based on significant BLAST alignments between EST contigs and TAIR10 genes. The tool facilitates searches based on gene function, product location and associated process.

Alignment views are displayed via GBrowse.



New gene expression datasets, associated with a study of transcriptional responses to abiotic stresses, have been added to our microarray database and submitted to GEO.



# SMD implementation

Comparative Evolutionary Genomics of Cotton - Windows Internet Explorer

http://cotton.evolution.info/

Comparative Evolutionary Genomics of Cotton

Iowa State University Brigham Young University University of Georgia

Microarray

Overview | Ordering | Probe Info | Quality Control | Analysis Aids | Expression Data (oligo)

Our microarray database, an implementation of SMD, is MIMIC compliant and includes Bioconductor generated diagnostic plots.

Overview  
Phylogeny  
Fiber Evolution

Introgression Populations  
Homeolog-specific Profiling  
Genetic Networks & Phenotype  
Effects of Selection

Genetic and Physical mapping  
resources

Comparative BAC Sequencing  
EST Resources  
Microarray

Web Database  
Education and Outreach  
Significance for cotton industry  
Cotton Links  
Events  
Wendel Lab  
PGML (Paterson Lab)  
Udall Lab

How to

turn explanations on/off

View and Sort Array Data  
Download Raw Data  
Download Original Data Files  
View Array Details

View Array Image and Grids  
Clickable Image  
Plot Array Data

SMD tools info

Your query returned 129 result sets.

ExptID	Experiment	Category	SlideName	Options	ExptDate
58045	A2_A_103155	CEGC_csl_2	A2_A_103155 (GSM432873)		2009-07-01
58046	A2_B_109481	CEGC_csl_2	A2_B_109481 (GSM432874)		2009-07-01
58044	A2_C_103152	CEGC_csl_2	A2_C_103152 (GSM432872)		2009-07-01
58026	AD2_K-10-A (2126702)	CEGC_csl_2	K-10-A_2126702 (GSM432823)		2008-06-04
58027	AD2_K-10-B (2126902)	CEGC_csl_2	K-10-B_2126902 (GSM432824)		2008-06-05
58028	AD2_K-10-C (2126903)	CEGC_csl_2	K-10-C_2126903		2008

A

Click spot to see in array context

Biological Information

Sequence Name YGR199W

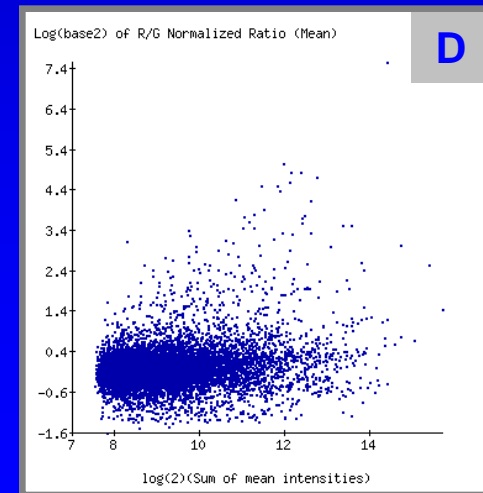
View expression history of this entity

Spot	536
SUID	1107217
Ch1 Intensity (Mean)	9447
Ch1 Background (Median)	77
Ch1 Net (Mean)	9370
Ch2 Intensity (Mean)	5402
Ch2 Net (Mean)	5342
Ch2 Background (Median)	60
Ch2 Normalized Background (Median)	59
Ch2 Normalized Net (Mean)	5252
Ch2 Normalized Intensity (Mean)	5311
Regression Correlation	.99
Spot Flag	0
% CH1 PIXELS > BG + 1SD	100
% CH2 PIXELS > BG + 1SD	100
Regression Ratio	582
Number of Background Pixels	364
Number of Spot Pixels	120
Box Top	147
Box Left	582
Box Right	595
Box Bottom	160

B

C

Dual SMD servers have been implemented to provide reliable access to our microarray expression data. The SMD data selection (A), display (B,C) and plotting (D) tools are available.



Comparative Evolutionary Genomics of Cotton - Mozilla Firefox

File Edit View History Bookmarks Tools Help

genbank - Google Search

128.192.141.98/CottonFiber/pages/sequence\_capture/screns2.aspx

Comparative Evolutionary Genomics of Cotton

Iowa State University Brigham Young University University of Georgia

Sequence Capture

Plant materials | Gene lists | Segregating Population | Sequencing Results | Results | People

Sequencing Results:

Overview  
Phylogeny  
Fiber Evolution

Introgession Populations  
Homoeolog-specific Profiling  
Genetic Networks & Phenotype  
Effects of Selection  
Sequence Capture

Genetic and Physical mapping  
resources  
Comparative BAC Sequencing  
Genome Sequence Resources  
EST Resources  
Microarray

Web Database  
Education and Outreach

An Evaluation of Sequence Capture Technologies with Polyploid Cotton (SRP009879)

Sequence capture is a revolutionary method to resequence targeted portions of the genome but it has not been tested with a polyploid plant. We constructed two capture platforms 1) a novel Nimblegen sequence capture microarray and 2) Myroarray beads containing RNA probes. Both platforms targeted the same 534 genes. On each platform, we hybridized two different accessions to this microarray to assess capture efficiency and potential of combined sequence using DNA multiplex identifiers (MIDs).

Sequence capture of 532 selected genes from *G. hirsutum* (SRP009870)

Cotton (*G. hirsutum*) is a polyploid species native to Central America. It was domesticated by the ancient inhabitants of Central America and it was quickly adopted by Europeans with colonial agricultural technology. Initially, it was grown around the Caribbean Sea. Feral cultivars from these historical growing regions and native populations have been investigated for genetic diversity with the aim to improve modern cultivated cotton. Here, we use sequence capture to investigate nucleotide diversity within and flanking 532 selected genes from the cotton genome.

# Diversity and polymorphism data

Download links for capture (e.g., fasta and ace) and sample specific (e.g., fasta, fastq and sff) data files are also included in the table.

MID	Forward Sequence	Capture 1	Capture 2	Capture 3	Capture 4
1	ACGAGTGCCT	TX_704	TX_1182	TX_2002	Coker_315
2	ACGCTCGACA	TX_786	TX_44	TX_2090	Cascot_L7
3	AGACGCACTC	TX_953	TX_959	TX_2091	Lkt_511
4	AGCACTGTAG	TX_1009	TX_1110	TX_2092	PM_145
5	ATCAGACACG	TX_1055	TX_1120	TX_2093	FM_958
6	ATATCGCGAG	TX_665	TX_1226	TX_2094	BR_110
7	CGTGTCTCTA	TX_2089	TX_1228	TX_2095	Maxxa
8	CTCGCGTGTC	TX_1037	TX_674	TX_2096	D5
9	TAGTATCAGC	TX_672	TX_1046	ARK_2402	AD3
10	TCTCTATGCG	TX_1107	TX_1982	ST825	GK
11	TGATACGTCT	TX_480	TX_1236	TX_1988	TAM
12	TACTGAGCTA	TX_1748	TX_1996	DP90	A2

We welcome your comments and suggestions.

Diversity panel information and sequence capture results available from CEGC as content, downloadable spreadsheets, links to SRA and downloadable flowgram, read, trace and assembly files.



# Bulk downloads and submitted data

- Links for bulk downloads of our EST related data appear on our “Analysis Aids” page. These include EST read and quality score data. These data have also been submitted to GenBank or SRA.
- Links to BACman (PGML) for BAC related data also appear on the “How to” page.
- Microarray file downloads are handled via the SMD interfaces and datasets are also submitted to GEO.
- SNPs submitted to dbSNP.

# Phenotype DB search tool (in development)

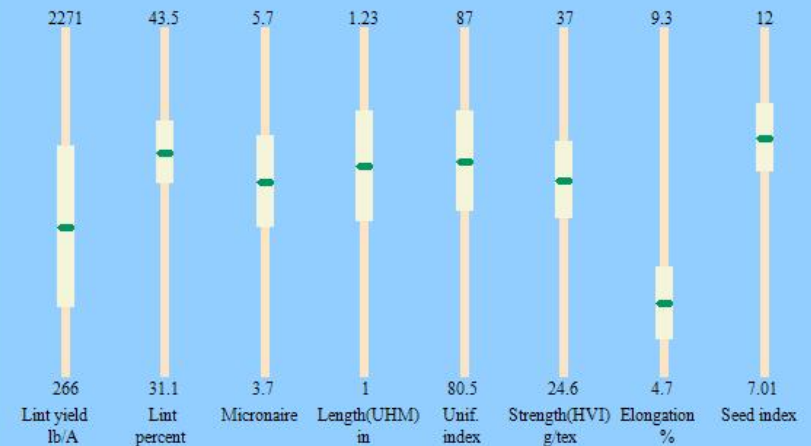
ACSI Fibermx FM 958 (MS, 1999)

Trial Information

Overall Performance

Location Performance

ACSI Fibermx FM 958



## Selection criteria:

- ☒ top 100 lint yield    ☐ top 100 lint %    ☐ top 100 micronaire    ☐ top 100 length uhm  
☐ top 100 unif index    ☐ top 100 strength hvi %    ☐ top 100 elongation    ☐ top 100 seed index
- any state    any year

Help

Name keyword:

all genotypes

Search

Show objects from 1 to 15 page 1, begins with 8839-3-10-2

genotype	datasets	max lint yield	max lint %	max micronaire	max length uhm	max strength hvi
8839-3-10-2	GA-01	1935	45.2	5.1	1.15	29.4
ACSI Fibermx FM 958	MS-00	1877	40.3	5.3	1.18	33.4
BCG 28 R	MS-04	1906	43	5.37	1.16	31.2
BCG 28R	MS-03 NC-03 NC-04	1823	43	32.9	74	84.6
DP 393	MS-04 MS-05	1934	43.8	5.13	1.19	33.53
DP 432 RR	MS-04 MS-05	2212	42.07	5.3	1.17	31.5
DP 434 RR	MS-04 MS-05	2027	44.39	5	1.22	29.38
DP 444 BG/RR	MS-04 MS-05	2239	43.16	4.7	1.18	30.83
DP 445 BG/RR	MS-04 MS-05	2054	43.48	5.13	1.2	32.67
DP 449 BG/RR	MS-04 MS-05	2023	41.35	5.4	1.17	34.57
DP 451 BG/RR	MS-04	1911	38.68	5.2	1.17	29.43
DP 455 BG/RR	MS-04 MS-05 NC-04	2029	43.92	4.93	1.21	34.27
DP 488 BG/RR	MS-04 MS-05	2284	43.25	5.4	1.21	33.3
DP 491	MS-01 MS-04	2036	44.19	5.3	1.27	34.77
DP 494 RR	MS-04 MS-05	1905	43.36	5.33	1.22	34.4

Fiber quality and other related phenotypic data will become available for NILs, developed as aids to the functional analysis of fiber development. Database tools are being adapted to make this data easily searchable.



# Read categorization pipeline (in development)

**polyCat**, a Web accessible pipeline for accurately assigning duplicated reads to their respective reference genome, is being implemented at CEGC. Interfaces will allow users to submit read files to the pipeline for A/D genome assignment. Result files will be available for download.

## Welcome to polyCat

Upload Read Tar

Job Name:

Email:

Type:

Read Tar:

## Welcome to polyCat

Upload Progress

Results ready email sent to submitter.

Uploaded  
read files  
submitted  
to pipeline

Read  
Assignment  
Pipeline

Result files  
available for  
download

# ***Acknowledgements***

- The National Science Foundation
- University of Georgia Office of the VP for Research