

Addition File 4: Snapshots of the search module

Figure S4-1: Snapshots of the Browse page: browse genes by genomic position.....	2
Figure S4-2: Snapshots of the Browse page: browse genes by genetic marker.....	4
Figure S4-3: Snapshots of the BLAST page.....	6
Figure S4-4: Snapshots of the Search page: search by gene ID/name.....	7
Figure S4-5: Snapshots of the Search page: search by domain or function annotations.	9
Figure S4-6: Snapshots of the Search page: search by gene expression profiles.	11
References.....	12

Browse Genes in Cotton ?

By Genomic Position By Genetic Marker

Options

Species Gossypium hirsutum, NAU Chromosome A01 Start 1000000 - 3000000 End Update

User specify genome location

Show 150 Genes in Region A01:1,000,000-3,000,000 in *Gossypium hirsutum* (NAU assembly)

Select & Analysis Select & BLAST Select & Phylogenetic Tree Build Select & Store Search ⌵

<input type="checkbox"/>	Gene ID	Gene Name	Description	Start	End	Strand	Links
<input type="checkbox"/>	Gh_A01G0119	ERF053	Ethylene-responsive transcription factor ERF053	1,008,923	1,010,191	+	Structure; Domain; Homology; GO;
<input type="checkbox"/>	Gh_A01G0120	RMA1H1	E3 ubiquitin-protein ligase RMA1H1	1,023,424	1,024,188	-	Structure; Domain; Homology; GO;
<input type="checkbox"/>	Gh_A01G0121	UBC38	Putative ubiquitin-conjugating enzyme E2 38	1,041,997	1,044,247	-	Structure; Domain; Homology; GO;
<input type="checkbox"/>	Gh_A01G0122	NA	NA	1,048,448	1,048,951	-	Structure; Domain; Homology; GO;
<input type="checkbox"/>	Gh_A01G0123	NSE4A	Non-structural maintenance of chromosomes element 4 homolog A	1,053,235	1,055,912	+	Structure; Domain; Homology; GO;
<input type="checkbox"/>	Gh_A01G0124	RPS5	Disease resistance protein RPS5	1,064,885	1,067,521	+	Structure; Domain; Homology; GO;
<input type="checkbox"/>	Gh_A01G0125	At5g63020	Probable disease resistance protein At5g63020	1,081,116	1,083,758	+	Structure; Domain; Homology; GO;
<input type="checkbox"/>	Gh_A01G0126	NA	NA	1,123,288	1,123,602	+	Structure; Domain; Homology; GO;
<input type="checkbox"/>	Gh_A01G0127	NA	NA	1,124,825	1,125,130	-	Structure; Domain; Homology; GO;
<input type="checkbox"/>	Gh_A01G0128	NA	NA	1,178,670	1,178,861	-	Structure; Domain; Homology; GO;

Showing 1 to 10 of 150 rows 10 records per page < 1 2 3 4 5 ... 15 >

Result gene list table

JBrowse View ?



An embedded JBrowse view

Figure S4-1: Snapshots of the Browse page: browse genes by genomic position.

This snapshot displayed 150 genes in region A01:1,000,000-3,000,000 in *G. hirsutum* (NAU assembly), which was the default value when first visiting the Browse page.

Browse Genes in Cotton ?

By Genomic Position By Genetic Marker

Options

Species: Gossypium hirsutum, NAU Start Marker: JESPR0152 - NAU3040 End Marker: Update

User specify a pair of genetic marker IDs

Marker Information

	Start Marker	End Marker
ID	JESPR0152	NAU3040
Alias	JESPR152	<u>NA</u>
Type	SSR	SSR
Source Species	<i>hirsutum</i>	<i>raimondii</i>
Map	AD-genome wide Reference Map (2009)	AD-genome wide Reference Map (2009)
Linkage Group	AD-genome wide Reference Map (2009).Ref-chr10	AD-genome wide Reference Map (2009).Ref-chr15
Sequence	GATGCACCAGATCCCTTTTATTAGTTTATAAGGATTACCAACTAAATAGAAGAAGAAGAAG AAGAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAATAGAAAGAAAGAAAGAAAG AAG AAGAAGAAGAAGAAAAGAAATGAAATACCCACTGTGATCCGATGTACC	AAAAGCCGTGCTAACTGAATACAGTTCTTTTAGGCTTTTGCAACAGGGAAGCTTCTTCC TTCCTTCTACAGTATGTCTTTTACTTATGTGAAATGTTTATAGGTTTCGAAATGCAAAA ACTTGGACAATTACCGATTTCCAAGTTTTGCCATAAAGACTTTTTATCAGTCTCGATGATG ATAAACCCGTAGTTTCGGGTGTTTGTGATT
Primer 1	GATGCACCAGATCCCTTTTATTAG	AAAAGCCGTGCTAACTGAAT
Primer 2	GATGCACCAGATCCCTTTTATTAG	AATCACAACACCCGAACT
Physical Positions (Best)	<ul style="list-style-type: none"> <i>Gossypium raimondii</i> (JGI): <u>NA</u> <i>Gossypium raimondii</i> (BGI): <u>NA</u> <i>Gossypium arboreum</i> (BGI): <u>NA</u> <i>Gossypium hirsutum</i> (BGI): scaffold238.1:1,447,946-1,448,279 <i>Gossypium hirsutum</i> (NAU): D01:59,289,789-59,290,090 <i>Gossypium barbadense</i> (NAU): D01:58,919,930-58,920,135 	<ul style="list-style-type: none"> <i>Gossypium raimondii</i> (JGI): Chr02:61,729,007-61,729,216 <i>Gossypium raimondii</i> (BGI): Chr02:1,035,849-1,036,058 <i>Gossypium arboreum</i> (BGI): Chr02:69,391,280-69,391,482 <i>Gossypium hirsutum</i> (BGI): A09:41,262,967-41,263,172 <i>Gossypium hirsutum</i> (NAU): D01:60,440,750-60,440,959 <i>Gossypium barbadense</i> (NAU): scaffold_1679.D01:156,016-156,225

Show marker information imported from CottonGen database

Show markers physical positions among cotton species

Continue on the next page...

Continue the previous page ...

Physical Position in Selected Genome

Filter By

Min. BLAT Query Coverage (%)

Min. BLAT Identity (%)

Max. Return Hit Number Per Marker

Show relevant marker position in genomes. The mapping parameters could be customized.

Selected Position: D01:59,289,789-59,290,090

Selected Position: D01:60,440,750-60,440,959

Select & Confirm Position

Start Marker: JESPR0152

Chromosome	Start	End	Coverage	Identity
<input checked="" type="radio"/> D01	59,289,789	59,290,090	98.71	98.28

Showing 1 to 1 of 1 rows

End Marker: NAU3040

Chromosome	Start	End	Coverage	Identity
<input checked="" type="radio"/> D01	60,440,750	60,440,959	100	98.1
<input type="radio"/> A01	98,793,134	98,793,339	100	95.71

Showing 1 to 2 of 2 rows

Users should select positions they want.

Confirm Your Selection & Fetch Genes

Show 116 Genes in Region D01:59,289,789-60,440,959 in *Gossypium hirsutum* (NAU assembly)

Select & Analysis | Select & BLAST | Select & Phylogenetic Tree Build | Select & Store

Gene ID	Gene Name	Description	Start	End	Strand	Links
<input type="checkbox"/> Gh_D01G2029	PSK6	Putative phytolectin-like protein 6	59,308,542	59,309,042	+	Structure; Domain; Homology
<input type="checkbox"/> Gh_D01G2030	At3g23620	Ribosome production factor 2 homolog	59,310,810	59,312,600	+	Structure; Domain; Homology
<input type="checkbox"/> Gh_D01G2031	FK	Delta(14)-sterol reductase	59,313,576	59,319,041	-	Structure; Domain; Homology
<input type="checkbox"/> Gh_D01G2032	HDT1	Histone deacetylase HDT1	59,320,151	59,322,851	+	Structure; Domain; Homology
<input type="checkbox"/> Gh_D01G2033	NA	Enolase	59,326,180	59,329,485	-	Structure; Domain; Homology
<input type="checkbox"/> Gh_D01G2034	PRXIE	Peroxisomal protein, chloroplastic	59,334,370	59,335,053	+	Structure; Domain; Homology
<input type="checkbox"/> Gh_D01G2035	At5g12100	Pentatricopeptide repeat-containing protein At5g12100, mitochondrial	59,335,370	59,337,823	-	Structure; Domain; Homology
<input type="checkbox"/> Gh_D01G2036	CSLA9	Glucosyltransferase 9	59,340,699	59,345,467	-	Structure; Domain; Homology
<input type="checkbox"/> Gh_D01G2037	NA	NA	59,364,867	59,366,338	+	Structure; Domain; Homology
<input type="checkbox"/> Gh_D01G2038	PXC1	Leucine-rich repeat receptor-like protein kinase PXC1	59,380,005	59,382,115	+	Structure; Domain; Homology

Showing 1 to 10 of 116 rows records per page

...

Result gene list table

JBrowse View

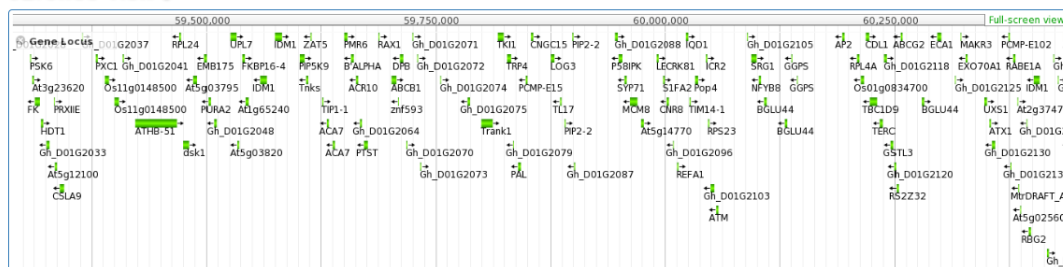


Figure S4-2: Snapshots of the Browse page: browse genes by genetic marker.

Since a single marker might map to multiple genomic locations, CottonFGD first displays all the possible mapped genomic locations to let users select the positions they want. By default, CottonFGD displays genomic locations that have $\geq 95\%$ query coverage and $\geq 90\%$ identity, which could be customized by users. Then CottonFGD

displays the result gene list table based on users' selections. This snapshot displayed 116 *G. hirsutum* genes (NAU assembly) between SSR marker JESPR0152 and NAU3040.

BLAST: Search Homologous Genes/Proteins By Sequence Similarity

SequenceServer App.

Input query

Select database

Other parameters

BLAST results

Result ID List

Query ID	Hit ID	Score	Gene Name	Description	Species	Assembly	Chromosome
CYP78A7	Gh_D06G1909.1	1155.97	CYP78A7	Cytochrome P450 78A7	Gossypium hirsutum	NAU	D06
CYP78A7	Gh_A06G1538.1	1116.68	CYP78A7	Cytochrome P450 78A7	Gossypium hirsutum	NAU	A06
CYP78A7	Gh_D12G1591.1	94.36	CYP78A5	Cytochrome P450 78A5	Gossypium hirsutum	NAU	D12
CYP78A7	Gh_A12G1471.1	93.59	CYP78A5	Cytochrome P450 78A5	Gossypium hirsutum	NAU	A12
CYP78A7	Gh_A02G1295.1	91.28	CYP78A7	Cytochrome P450 78A7	Gossypium hirsutum	NAU	A02
CYP78A7	Gh_D03G1507.1	89.74	CYP75A2	Flavonoid 3',5'-hydroxylase	Gossypium hirsutum	NAU	D03
CYP78A7	Gh_D03G0436.1	89.35	CYP78A7	Cytochrome P450 78A7	Gossypium hirsutum	NAU	D03
CYP78A7	Gh_A06G0999.1	86.66	CYP78A7	Cytochrome P450 78A7	Gossypium hirsutum	NAU	A06
CYP78A7	Gh_D03G1504.1	86.27	CYP75A2	Flavonoid 3',5'-hydroxylase	Gossypium hirsutum	NAU	D03
CYP78A7	Gh_A08G0963.1	83.96	CYP78A3	Cytochrome P450 78A3	Gossypium hirsutum	NAU	A08

Showing 1 to 10 of 10 rows

Figure S4-3: Snapshots of the BLAST page. This page mainly contains an embed SequenceServer App[1], a powerful BLAST front-end. After finishing the BLAST search, CottonFGD would automatically extract target gene IDs for each query sequence and rendered them in a table. This snapshot displayed the BLAST result using protein sequences of *CYP78A7* as query to search against all the *G. hirsutum* proteins with an E-value of 1e-05 and returning top 10 hits.

Navigation Bar

Search for Cotton Genes by Key Words ?

By Gene ID/Name | By Protein Domain | By Function | By Expression

Gene Information

Search Content:

e.g.

- Gh_D07G0123
- CotAD_04554
- GOBAR_AA16059
- Cotton_A_05687
- Gorai.013G264100
- Cotton_D_gene_10040019
- ACT
- actin

Species

Select Species

- Gossypium hirsutum* (AD₁, upland cotton), NAU assembly
- Gossypium hirsutum* (AD₁, upland cotton), BGI assembly
- Gossypium barbadense* (AD₂, sea-island cotton), NAU assembly
- Gossypium raimondii* (D₅), JGI assembly
- Gossypium raimondii* (D₅), BGI assembly
- Gossypium arboreum* (A₂, tree cotton), BGI assembly

Submit Reset

Input ID/name

Select species

Search Result

159 results are found.

Select & Analysis | Select & BLAST | Select & Phylogenetic Tree Build | Select & Store | Search

Gene ID	Gene Name	Description	Species	Assembly	Chromosome
Gh_D12G0311	ADF5	Actin-depolymerizing factor 5	<i>Gossypium hirsutum</i>	NAU	D12
Gh_D06G0509	ARP6	Actin-related protein 6	<i>Gossypium hirsutum</i>	NAU	D06
Gh_D09G1029	SCAB1	Stomatal closure-related actin-binding protein 1	<i>Gossypium hirsutum</i>	NAU	D09
Gh_A03G1149	ARP3	Actin-related protein 3	<i>Gossypium hirsutum</i>	NAU	A03
Gh_A05G2093	At1g71790	Probable F-actin-capping protein subunit beta	<i>Gossypium hirsutum</i>	NAU	A05
Gh_D04G0598	AC58	Actin-58	<i>Gossypium hirsutum</i>	NAU	D04
Gh_D08G0978	ARPC3	Actin-related protein 2/3 complex subunit 3	<i>Gossypium hirsutum</i>	NAU	D08
Gh_D03G1211	MA	Actin	<i>Gossypium hirsutum</i>	NAU	D03
Gh_D03G1240	ADF1	Actin-depolymerizing factor 1	<i>Gossypium hirsutum</i>	NAU	D03
Gh_A05G0607	SCAB1	Stomatal closure-related actin-binding protein 1	<i>Gossypium hirsutum</i>	NAU	A05

Showing 1 to 10 of 159 rows | 10 records per page | 1 2 3 4 5 ... 16

Result gene list table

Figure S4-4: Snapshots of the Search page: search by gene ID/name. This snapshot displayed 159 *G. hirsutum* genes using keyword “actin” as gene name search.

Search for Cotton Genes by Key Words ?

By Gene ID/Name | By Protein Domain | By Function | By Expression

Domain Information

Domain Database: Pfam

Search Content: Thioesterase

e.g.

- PF03061
- PS00036
- 4HBT
- Thioesterase

Species

Select Species

- Gossypium hirsutum* (AD₁, upland cotton), NAU assembly
- Gossypium hirsutum* (AD₁, upland cotton), BGI assembly
- Gossypium barbadense* (AD₂, sea-island cotton), NAU assembly
- Gossypium raimondii* (D₅), JGI assembly
- Gossypium raimondii* (D₅), BGI assembly
- Gossypium arboreum* (A₂, tree cotton), BGI assembly

Submit Reset

Input domain ID/name (might be fuzzy)

Select species

Search Result

Domain List

Select & Fetch Gene Search

<input type="checkbox"/> Domain Accession	Domain Description	Exist in Cotton?
<input type="checkbox"/> PF00975	Thioesterase; Thioesterase domain	No
<input type="checkbox"/> PF01643	Acyl-ACP_TE; Acyl-ACP thioesterase	Yes
<input type="checkbox"/> PF02089	Palm_thioest; Palmitoyl protein thioesterase	Yes
<input type="checkbox"/> PF02551	Acyl_CoA_thio; Acyl-CoA thioesterase	No
<input checked="" type="checkbox"/> PF03061	4HBT; Thioesterase superfamily	Yes
<input type="checkbox"/> PF09500	YliD_C; Putative thioesterase (yliD_C-term)	No
<input type="checkbox"/> PF10862	FcoT; FcoT-like thioesterase domain	No
<input type="checkbox"/> PF12590	Acyl-thio_N; Acyl-ATP thioesterase	Yes
<input checked="" type="checkbox"/> PF13279	4HBT_2; Thioesterase-like superfamily	Yes
<input checked="" type="checkbox"/> PF13622	4HBT_3; Thioesterase-like superfamily	Yes

Showing 1 to 10 of 10 rows

Step 1: Display all possible domains in selected domain database (e.g. Pfam) Users would select their wished items and perform a secondary search.

Search Result

Associated Proteins List

30 results are found.

Select & Analysis | Select & BLAST | Select & Phylogenetic Tree Build | Select & Store Search

<input type="checkbox"/> Protein ID	Gene Name	Species	Assembly	Domain Accession
<input type="checkbox"/> Gh_A02G0603.1	DHNAT1	<i>Gossypium hirsutum</i>	NAU	PF03061 (1);
<input type="checkbox"/> Gh_A03G0409.1	ACOT13	<i>Gossypium hirsutum</i>	NAU	PF03061 (1);
<input type="checkbox"/> Gh_A04G0840.1	slI0410	<i>Gossypium hirsutum</i>	NAU	PF03061 (1);
<input type="checkbox"/> Gh_A04G0841.1	NA	<i>Gossypium hirsutum</i>	NAU	PF03061 (1);
<input type="checkbox"/> Gh_A05G0760.1	ACOT13	<i>Gossypium hirsutum</i>	NAU	PF03061 (1);
<input type="checkbox"/> Gh_A05G3848.1	ACOT13	<i>Gossypium hirsutum</i>	NAU	PF03061 (1);
<input type="checkbox"/> Gh_A08G1159.1	ACOT13	<i>Gossypium hirsutum</i>	NAU	PF03061 (1);
<input type="checkbox"/> Gh_A09G0781.1	Ascc1	<i>Gossypium hirsutum</i>	NAU	PF03061 (2);
<input type="checkbox"/> Gh_A09G0782.1	NA	<i>Gossypium hirsutum</i>	NAU	PF03061 (1);
<input type="checkbox"/> Gh_A09G0783.1	NA	<i>Gossypium hirsutum</i>	NAU	PF03061 (2);

Showing 1 to 10 of 30 rows 10 records per page

Step 2: Result gene list table based on users' selection

Figure S4-5: Snapshots of the Search page: search by domain or function annotations. As name match might be fuzzy, CottonFGD uses a two-step search method. This snapshot displayed searching *G. hirsutum* genes containing “Thioesterase” domain. The same method is also used for function (GO/InterPro/pathway) search.

Search for Cotton Genes by Key Words ?

By Gene ID/Name By Protein Domain By Function By Expression

Expression Parameter

FPKM Vaule Cutoff ✓

Expressed In

Gossypium hirsutum ▾

Gossypium raimondii ▾

Gossypium arboreum ▾

Gossypium barbadense ▾

Fiber development, HAU, time-series (*Gossypium barbadense*)

Check All

Ovule in anthesis
 Fiber in 10 days post-anthesis
 Fiber in 20 days post-anthesis

Tissue and Organ (*Gossypium barbadense*)

Check All

Anther Fiber Leaf Ovule
 Petal Root Stem Stigma

Species

Select Species

Gossypium hirsutum (AD₁, upland cotton), NAU assembly
 Gossypium hirsutum (AD₁, upland cotton), BGI assembly
 Gossypium barbadense (AD₂, sea-island cotton), NAU assembly
 Gossypium raimondii (D₅), JGI assembly
 Gossypium raimondii (D₅), BGI assembly
 Gossypium arboreum (A₂, tree cotton), BGI assembly

Define expression thresholds

Select searched experiments (showing partial option only)

Select searched experiments

Search Result

Genes Expressed in Selected Experiments

32809 results are found. ✕

Select & Analysis Select & BLAST Select & Phylogenetic Tree Build Select & Store

<input type="checkbox"/> Protein ID	Petal
<input type="checkbox"/> GOBAR_AA00003.1	1.41
<input type="checkbox"/> GOBAR_AA00020.1	3.86
<input type="checkbox"/> GOBAR_AA00027.1	1.40
<input type="checkbox"/> GOBAR_AA00028.1	8.87
<input type="checkbox"/> GOBAR_AA00030.1	25.13
<input type="checkbox"/> GOBAR_AA00031.1	97.31

Result gene list table expressed in selected experiments

Figure S4-6: Snapshots of the Search page: search by gene expression profiles.

The expression FPKM threshold could be customized (default value is 1). CottonFGD would display a list of genes expressed in all the selected experiments. This snapshot displayed searching *G. barbadense* genes expressed in petal.

References

1. Priyam A, Woodcroft BJ, Rai V, Munagala A, Moghul I, Ter F, Gibbins MA, Moon H, Leonard G, Rumpf W: **Sequenceserver: a modern graphical user interface for custom BLAST databases.** *Biorxiv* 2015:033142.