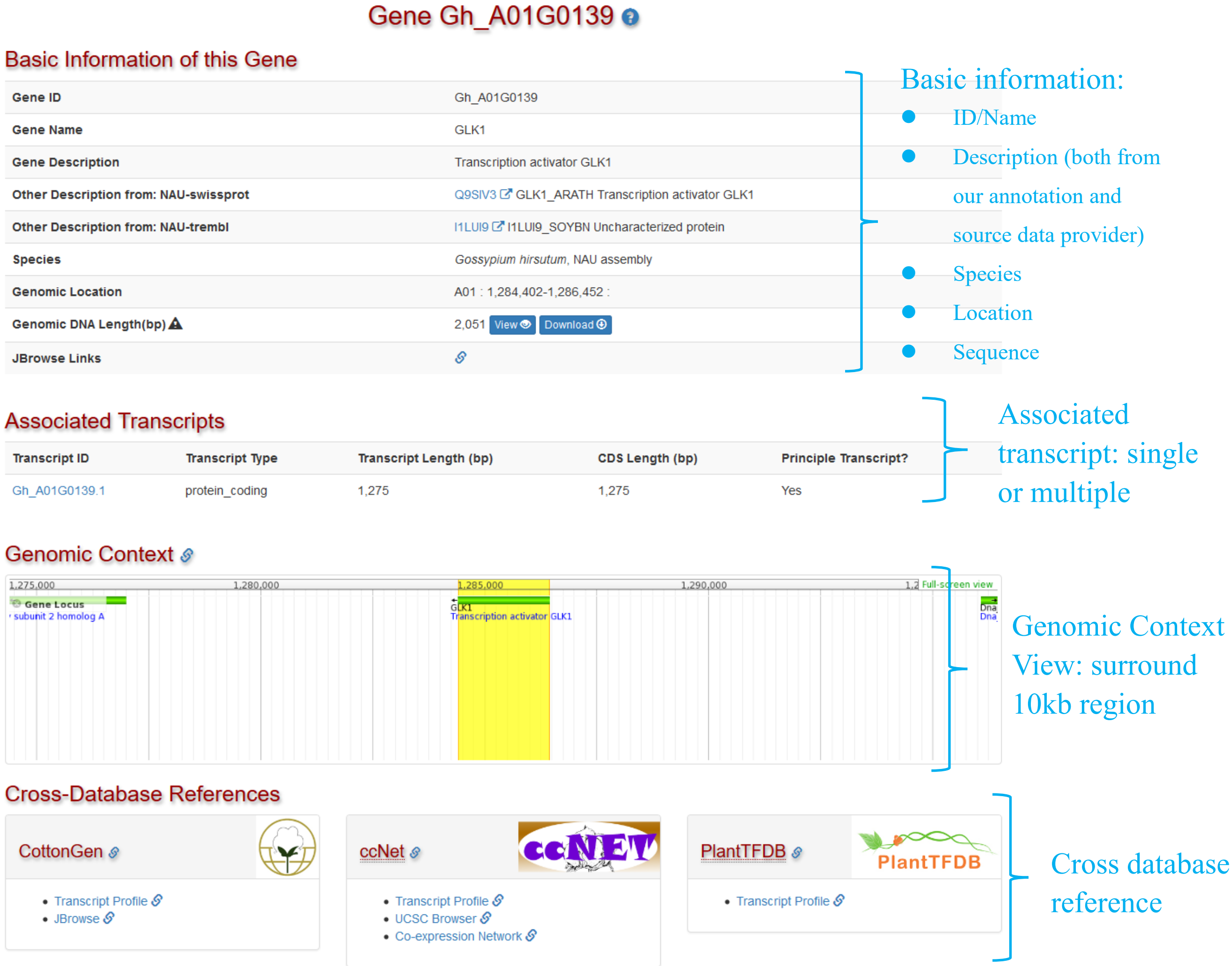
Addition File 5: Snapshots of the profile module



Gene Gh_A01G0139

Basic Information of this Gene

Gene ID	Gh_A01G0139
Gene Name	GLK1
Gene Description	Transcription activator GLK1
Other Description from: NAU-swissprot	Q9SIV3 GLK1_ARATH Transcription activator GLK1
Other Description from: NAU-trembl	I1LUI9 I1LUI9_SOYBN Uncharacterized protein
Species	*Gossypium hirsutum*, NAU assembly
Genomic Location	A01 : 1,284,402-1,286,452 :
Genomic DNA Length(bp)	2,051 View Download
JBrowse Links	

Basic information:
- ID/Name
- Description (both from our annotation and source data provider)
- Species
- Location
- Sequence

Associated Transcripts

Transcript ID	**Transcript Type**	**Transcript Length (bp)**	**CDS Length (bp)**	**Principle Transcript?**
Gh_A01G0139.1	protein_coding	1,275	1,275	Yes

Associated transcript: single or multiple

Genomic Context

Genomic Context View: surround 10kb region

Cross-Database References

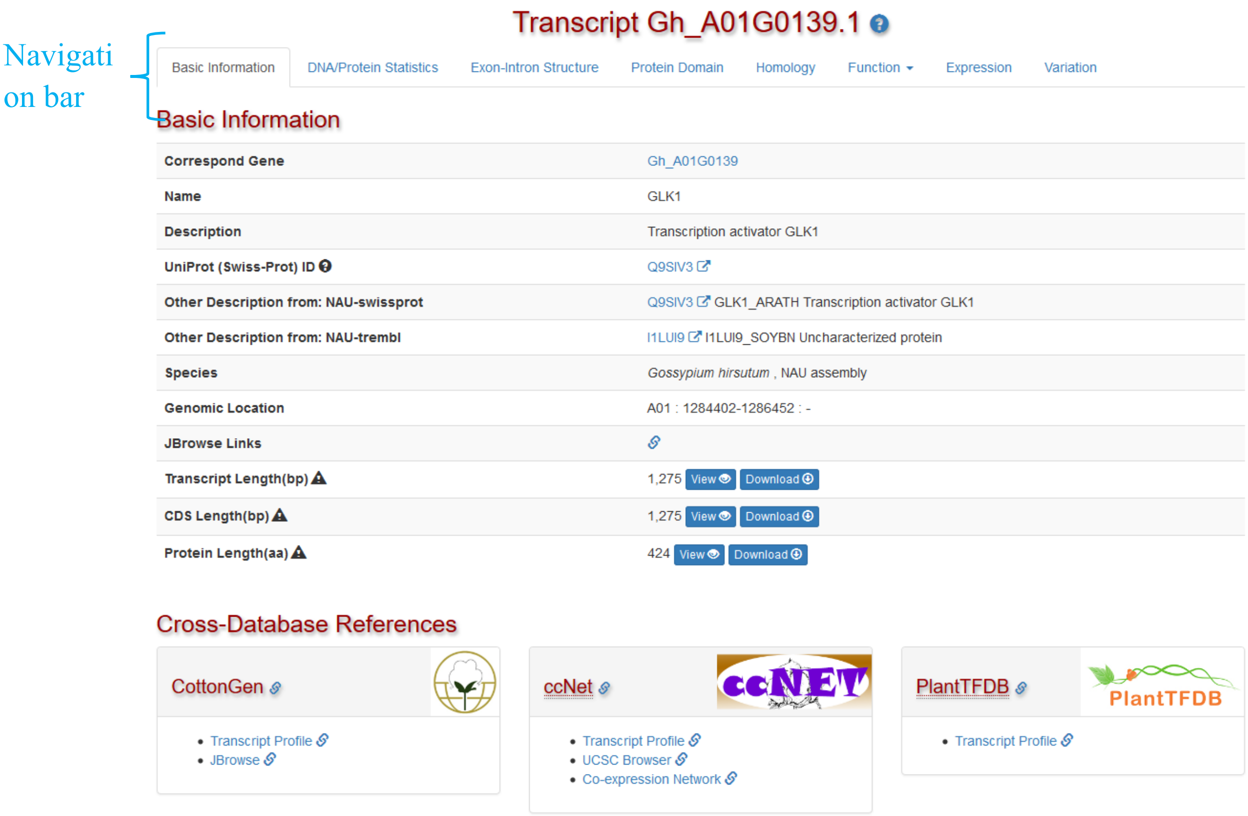
CottonGen
- Transcript Profile
- JBrowse

ccNet
- Transcript Profile
- UCSC Browser
- Co-expression Network

PlantTFDB
- Transcript Profile

Cross database reference

Figure S5-1. Snapshots of the gene profile page. This snapshot displayed profile pages of gene Gh_A01G0139 (*GLK1*) in *G. hirsutum* (NAU assembly).



Transcript Gh_A01G0139.1

Basic Information

Correspond Gene	Gh_A01G0139
Name	GLK1
Description	Transcription activator GLK1
UniProt (Swiss-Prot) ID	Q9SIV3
Other Description from: NAU-swissprot	Q9SIV3 GLK1_ARATH Transcription activator GLK1
Other Description from: NAU-trembl	I1LUI9 I1LUI9_SOYBN Uncharacterized protein
Species	*Gossypium hirsutum* , NAU assembly
Genomic Location	A01 : 1284402-1286452 : -
JBrowse Links	
Transcript Length(bp)	1,275
CDS Length(bp)	1,275
Protein Length(aa)	424

Cross-Database References

CottonGen

- Transcript Profile
- JBrowse

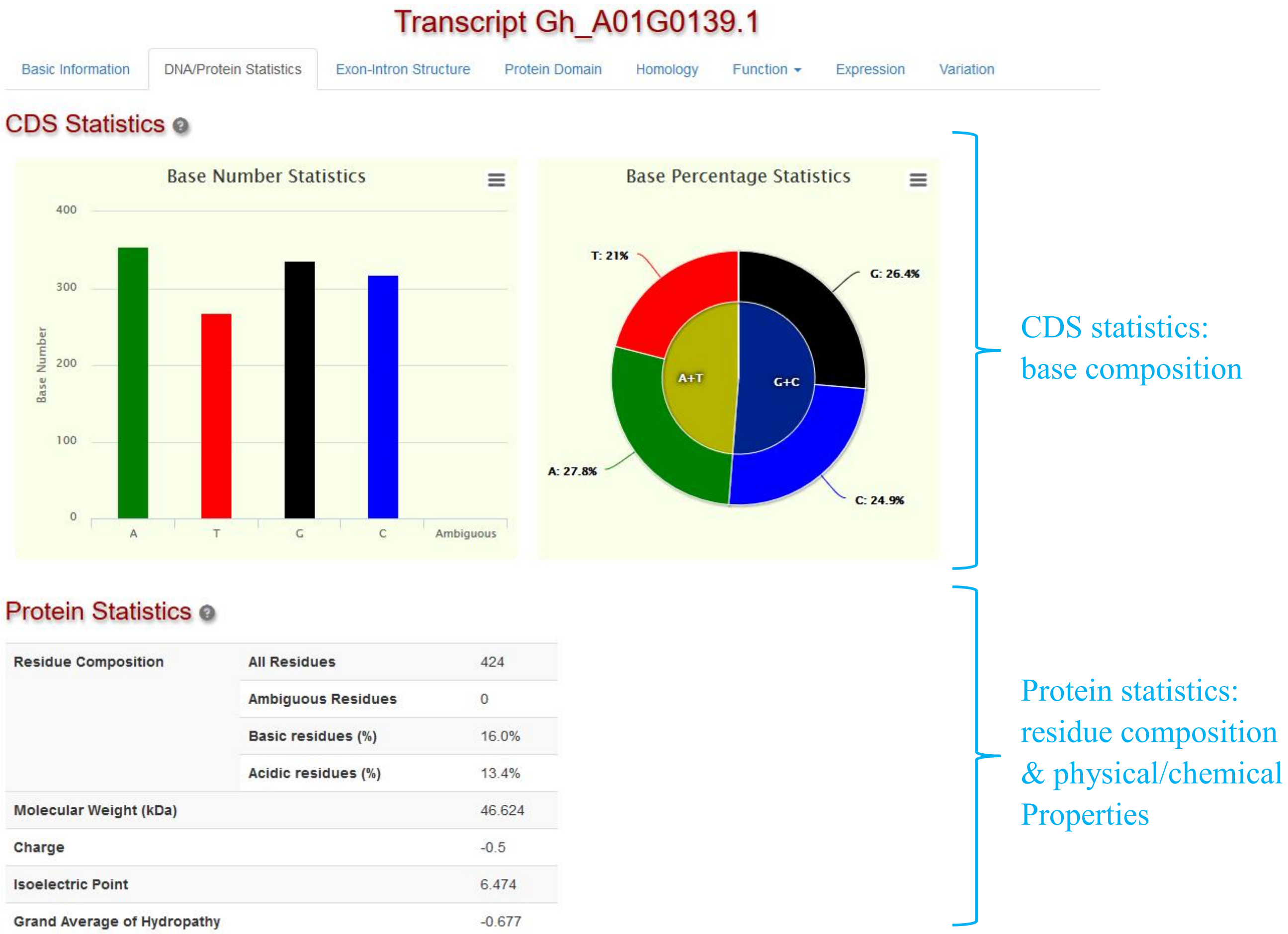
ccNet

- Transcript Profile
- UCSC Browser
- Co-expression Network

PlantTFDB

- Transcript Profile

Figure S5-2. Snapshots of the transcript profile page: Basic information. This snapshot displayed ID/name, transcript/CDS/protein sequences and cross-database references of transcript Gh_A01G0139.1 (*GLK1*) in *G. hirsutum* (NAU assembly).

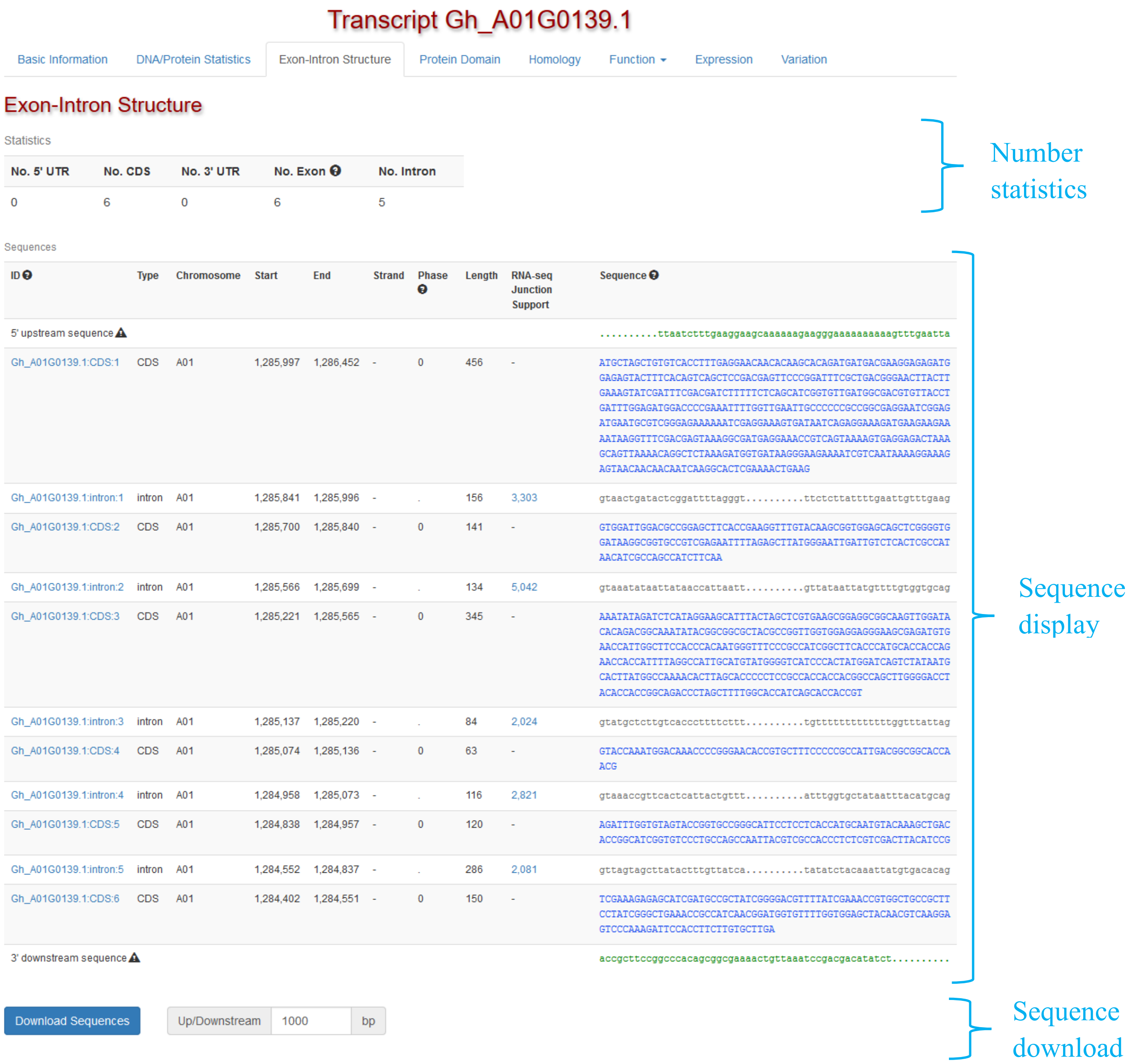


Residue Composition	**All Residues**	424
	Ambiguous Residues	0
	Basic residues (%)	16.0%
	Acidic residues (%)	13.4%
Molecular Weight (kDa)		46.624
Charge		-0.5
Isoelectric Point		6.474
Grand Average of Hydropathy		-0.677

CDS statistics: base composition

Protein statistics: residue composition & physical/chemical Properties

Figure S5-3. Snapshots of the transcript profile page: DNA/protein statistics. This snapshot displayed base composition, protein residue composition and protein physical/chemical properties of transcript Gh_A01G0139.1 (*GLK1*) in *G. hirsutum* (NAU assembly).



Transcript Gh_A01G0139.1

Exon-Intron Structure

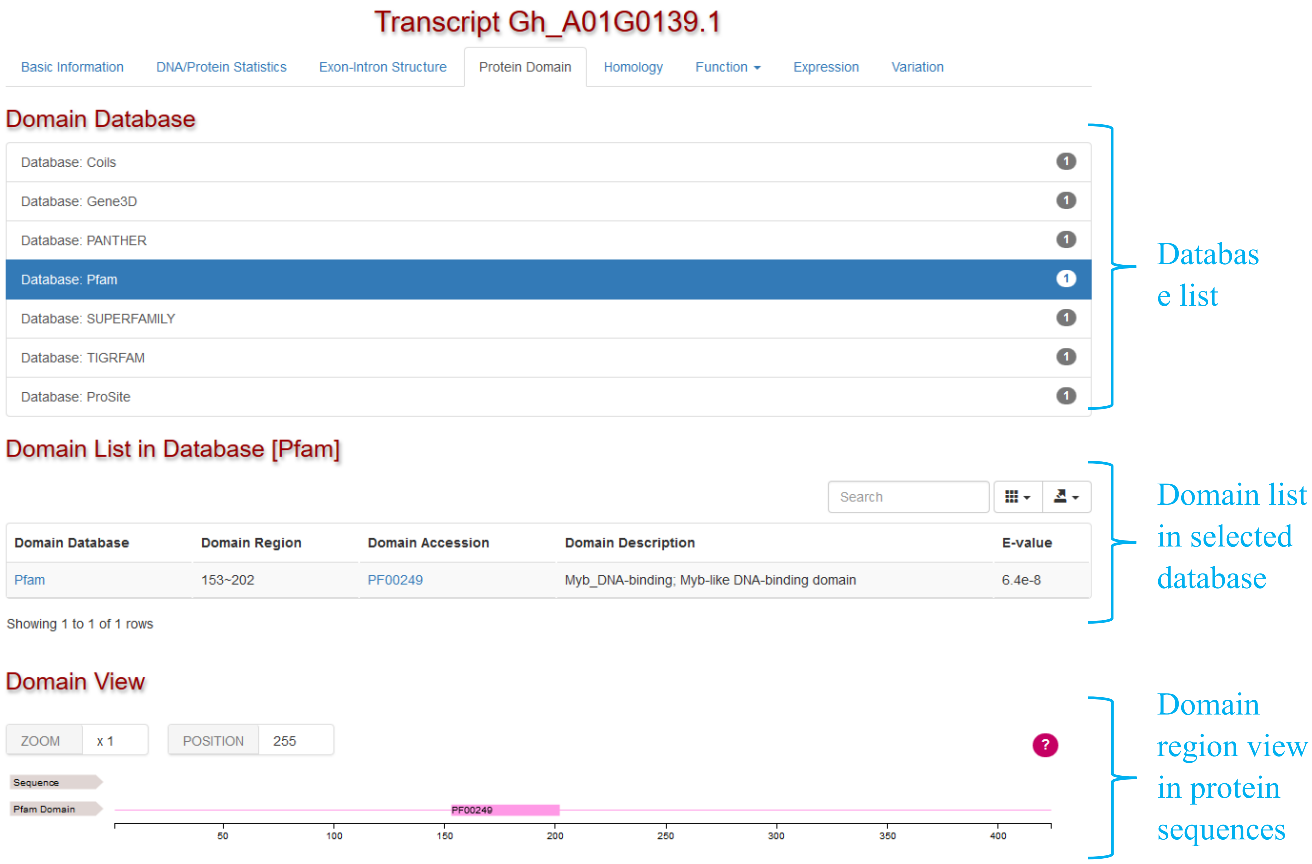
Statistics

No. 5' UTR	No. CDS	No. 3' UTR	No. Exon ❓	No. Intron
0	6	0	6	5

Sequences

ID ❓	Type	Chromosome	Start	End	Strand	Phase ❓	Length	RNA-seq Junction Support	Sequence ❓
5' upstream sequence ⚠									..........ttaatctttgaaggaagcaaaaaagaagggaaaaaaaaaagtttgaatta
Gh_A01G0139.1:CDS:1	CDS	A01	1,285,997	1,286,452	-	0	456	-	ATGCTAGCTGTGTCACCTTTGAGGAACAACACAAGCACAGATGATGACGAAGGAGAGATG GAGAGTACTTTCACAGTCAGCTCCGACGAGTTCCCGGATTTCGCTGACGGGAACTTACTT GAAAGTATCGATTTCGACGATCTTTTTCTCAGCATCGGTGTTGATGGCGACGTGTTACCT GATTTGGAGATGGACCCCGAAATTTTGGTTGAATTGCCCCCCGCCGGCGAGGAATCGGAG ATGAATGCGTCGGGAGAAAAAATCGAGGAAAGTGATAATCAGAGGAAAGATGAAGAAGAA AATAAGGTTTCGACGAGTAAAGGCGATGAGGAAACCGTCAGTAAAAGTGAGGAGACTAAA GCAGTTAAAACAGGCTCTAAAGATGGTGATAAGGGAAGAAAATCGTCAATAAAAGGAAAG AGTAACAACAACAATCAAGGCACTCGAAAACTGAAG
Gh_A01G0139.1:intron:1	intron	A01	1,285,841	1,285,996	-	.	156	3,303	gtaactgatactcggattttagggt..........ttctcttattttgaattgtttgaag
Gh_A01G0139.1:CDS:2	CDS	A01	1,285,700	1,285,840	-	0	141	-	GTGGATTGGACGCCGGAGCTTCACCGAAGGTTTGTACAAGCGGTGGAGCAGCTCGGGGTG GATAAGGCGGTGCCGTCGAGAATTTTAGAGCTTATGGGAATTGATTGTCTCACTCGCCAT AACATCGCCAGCCATCTTCAA
Gh_A01G0139.1:intron:2	intron	A01	1,285,566	1,285,699	-	.	134	5,042	gtaaatataattataaccattaatt..........gttataattatgttttgtggtgcag
Gh_A01G0139.1:CDS:3	CDS	A01	1,285,221	1,285,565	-	0	345	-	AAATATAGATCTCATAGGAAGCATTTACTAGCTCGTGAAGCGGAGGCGGCAAGTTGGATA CACAGACGGCAAATATACGGCGGCGCTACGCCGGTTGGTGGAGGAGGGAAGCGAGATGTG AACCATTGGCTTCCACCCACAATGGGTTTCCCGCCATCGGCTTCACCCATGCACCACCAG AACCACCATTTTAGGCCATTGCATGTATGGGGTCATCCCACTATGGATCAGTCTATAATG CACTTATGGCCAAAACACTTAGCACCCCCTCCGCCACCACCACGGCCAGCTTGGGGACCT ACACCACCGGCAGACCCTAGCTTTTGGCACCATCAGCACCACCGT
Gh_A01G0139.1:intron:3	intron	A01	1,285,137	1,285,220	-	.	84	2,024	gtatgctcttgtcacccttttcttt..........tgttttttttttttttggtttattag
Gh_A01G0139.1:CDS:4	CDS	A01	1,285,074	1,285,136	-	0	63	-	GTACCAAATGGACAAACCCCGGGAACACCGTGCTTTCCCCCGCCATTGACGGCGGCACCA ACG
Gh_A01G0139.1:intron:4	intron	A01	1,284,958	1,285,073	-	.	116	2,821	gtaaaccgttcactcattactgttt..........atttggtgctataatttacatgcag
Gh_A01G0139.1:CDS:5	CDS	A01	1,284,838	1,284,957	-	0	120	-	AGATTTGGTGTAGTACCGGTGCCGGGCATTCCTCCTCACCATGCAATGTACAAAGCTGAC ACCGGCATCGGTGTCCCTGCCAGCCAATTACGTCGCCACCCTCTCGTCGACTTACATCCG
Gh_A01G0139.1:intron:5	intron	A01	1,284,552	1,284,837	-	.	286	2,081	gttagtagcttatactttgttatca..........tatatctacaaattatgtgacacag
Gh_A01G0139.1:CDS:6	CDS	A01	1,284,402	1,284,551	-	0	150	-	TCGAAAGAGAGCATCGATGCCGCTATCGGGGACGTTTTATCGAAACCGTGGCTGCCGCTT CCTATCGGGCTGAAACCGCCATCAACGGATGGTGTTTTGGTGGAGCTACAACGTCAAGGA GTCCCAAAGATTCCACCTTCTTGTGCTTGA
3' downstream sequence ⚠									accgcttccggcccacagcggcgaaaactgttaaatccgacgacatatct..........

Figure S5-4. Snapshots of the transcript profile page: Exon-intron structure. This snapshot displayed all the sequences of each exon and intron in transcript Gh_A01G0139.1 (*GLK1*) in *G. hirsutum* (NAU assembly). Users can download these sequences together with upstream/downstream sequences in custom length.



Transcript Gh_A01G0139.1

Domain Database

Database: Coils 1
Database: Gene3D 1
Database: PANTHER 1
Database: Pfam 1
Database: SUPERFAMILY 1
Database: TIGRFAM 1
Database: ProSite 1

Domain List in Database [Pfam]

Domain Database	Domain Region	Domain Accession	Domain Description	E-value
Pfam	153~202	PF00249	Myb_DNA-binding; Myb-like DNA-binding domain	6.4e-8

Showing 1 to 1 of 1 rows

Domain View

Figure S5-5. Snapshots of the transcript profile page: Protein domain. This snapshot displayed domains in protein Gh_A01G0139.1 (*GLK1*) in *G. hirsutum* (NAU assembly). Users can switch among different domain databases.

Transcript Gh_A01G0139.1

Homologs List

Orthologs Among Other Plants

☐	Subject Species	Subject ID	Subject Length (aa)	Query Coverage (%)	E value	Identity (%)
☑	*Gossypium barbadense, NAU assembly*	GOBAR_AA28660.1	424	100	0	100
☑	*Gossypium raimondii, JGI assembly*	Gorai.002G020300.1	424	100	0	97
☐	*Gossypium raimondii, BGI assembly*	Cotton_D_gene_10018000.1	424	100	0	97
☑	*Gossypium arboreum, BGI assembly*	Cotton_A_00064.1	424	100	0	99
☑	*Theobroma cacao*	Thecc1EG008193t1	427	100	0	76
☐	*Arabidopsis thaliana*	AT2G20570.2	437	100	8e-105	52
☐	*Brassica rapa*	Brara.I04696.1.p	417	100	5e-97	50
☐	*Carica papaya*	evm.model.supercontig_83.93	446	99	4e-98	51
☐	*Citrus sinensis*	orange1.1g013528m	442	100	2e-171	65
☐	*Populus trichocarpa*	Potri.T098900.1	435	100	3e-165	67

Showing 1 to 10 of 26 rows 10 records per page

Users would select what they want and draw alignments

Orthologs in other *Gossypium* and other plants

Multiple Sequence Alignment for Orthologs

```
ID  Label
1   Ghir(NAU)   MLAVSPLRNNTSTDDDEGEMESTFTVSSDEFPDFADGNLLESIDFDDLFLSIGVDGDVLPDLEMDPEI
2   Garb(BGI)   MLAVSPLRNNTSTDDDEGEMESTFTISSDEFPDFADGNLLESIDFDDLFLSIGVDGDVLPDLEMDPEI
3   Gbar(NAU)   MLAVSPLRNNTSTDDDEGEMESTFTVSSDEFPDFADGNLLESIDFDDLFLSIGVDGDVLPDLEMDPEI
4   Grai(JGI)   MLAVSPLRNNTSKDDDEGEMESTFTISSDEFPDFADGNLLESIDFDDLFLSIDVDGDVLPDLEMDPEI
5   Tcac        MLAVSPLR-NTTNDENKGEMES-FTISSEEFPDFADGNLLESIDFDDLFVSIN-EGDMLPDLEMDPEI
```

An interactive MSA viewer drawing alignments

Paralog within this Species

Subject Species	Subject ID	Subject Length (aa)	Query Coverage (%)	E value	Identity (%)
Gossypium hirsutum, NAU assembly	Gh_D01G0183.1	424	100	0	96

Show paralog within the same species (assembly)

Assembly Conversion

Subject Species	Subject ID	Subject Length (aa)	Query Coverage (%)	E value	Identity (%)
Gossypium hirsutum, BGI assembly	CotAD_24836.1	424	100	0	97

Switch among different assemblies in the same species

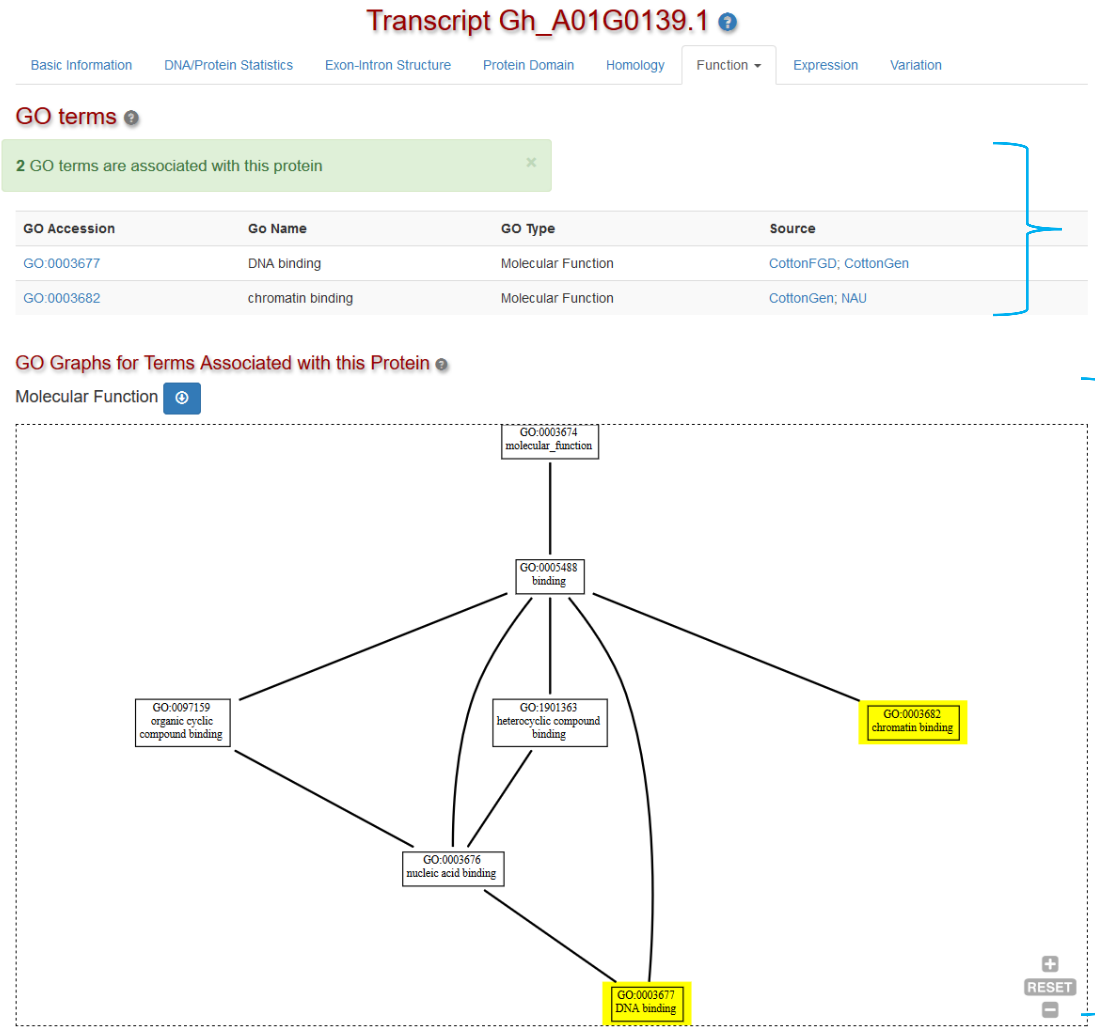
Cross Database References

PHD

- Link to PHD database

Links to PHD database based on the ortholog in *Arabidopsis thaliana*

Figure S5-6. Snapshots of the transcript profile page: Homology. This snapshot displayed orthologs and paralogs in protein Gh_A01G0139.1 (*GLK1*) in *G. hirsutum* (NAU assembly). Users can select several target species and CottonFGD would draw an interactive alignment.



Transcript Gh_A01G0139.1

GO terms

2 GO terms are associated with this protein

GO Accession	Go Name	GO Type	Source
GO:0003677	DNA binding	Molecular Function	CottonFGD; CottonGen
GO:0003682	chromatin binding	Molecular Function	CottonGen; NAU

GO Graphs for Terms Associated with this Protein

Molecular Function

List associated GO terms from both our annotation and other provider

Graphs showing the relationship between associated GO terms and their parent terms

Figure S5-7. Snapshots of the transcript profile page: Function. This snapshot displayed associated GO terms in protein Gh_A01G0139.1 (*GLK1*) in *G. hirsutum* (NAU assembly). Interactive graphs showing the relationship between associated GO terms and their parent terms were also presented in three categories: Molecular Function, Biology Process and Cell Component. (This gene only has GO terms in “Molecular Function”).

Transcript Gh_A01G0139.1

Expression Groups

Cytoplasmic male sterility (CMS) study of strain JA and JB, sporogenous cell stage	1
Fiber development, time-series	2
Ovule development, time-series	3
Seed germination, cotyledon time-series	5
Seed germination, organ comparison	2
Seed germination, root time-series	2
Seed germination, seed time-series	3
Stress experiment, Cold treat in time-series	4
Stress experiment, compare with control	5
Stress experiment, Hot treat in time-series	4
Stress experiment, PEG treat in time-series	4
Stress experiment, Salt treat in time-series	4
Tissue and Organ	5

[Stress experiment, Salt treat in time-series] Expression Levels

Project	Run	Sample	FPKM
SRP044705	SRR1768512;	1 hour after Salt treated	84.27
SRP044705	SRR1768513;	3 hours after Salt treated	96.93
SRP044705	SRR1768514;	6 hours after Salt treated	55.27
SRP044705	SRR1768515;	12 hours after Salt treated	2.45

Showing 1 to 4 of 4 rows

[Stress experiment, Salt treat in time-series] Expression Difference

WARNING: **No** Significant Expression Difference in this group.

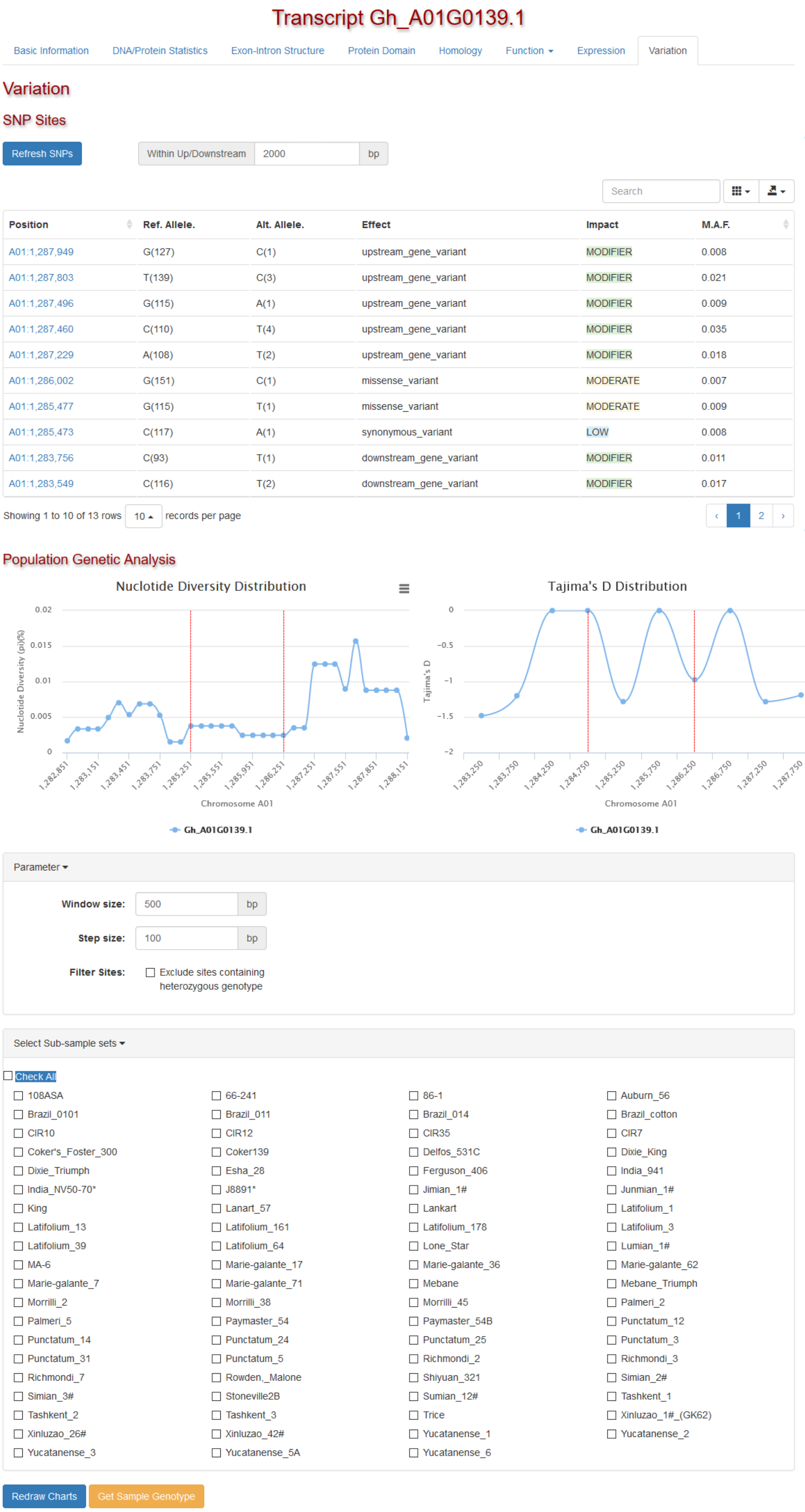
JBrowse View

Experiment List Users would select what they want

Expression levels (FPKM) in selected experiment

Visualization of RNA-seq coverages in selected experiment

Figure S5-8. Snapshots of the transcript profile page: Expression. This snapshot displayed the expression pattern of transcript Gh_A01G0139.1 (*GLK1*) in *G. hirsutum* (NAU assembly) in salt treated experiment compared in time-series (1h, 3h, 6h, 12h). CottonFGD would display a JBrowse snapshot of relevant RNA-seq coverages.



Transcript Gh_A01G0139.1

Variation

SNP Sites

Within Up/Downstream 2000 bp

Position	Ref. Allele.	Alt. Allele.	Effect	Impact	M.A.F.
A01:1,287,949	G(127)	C(1)	upstream_gene_variant	MODIFIER	0.008
A01:1,287,803	T(139)	C(3)	upstream_gene_variant	MODIFIER	0.021
A01:1,287,496	G(115)	A(1)	upstream_gene_variant	MODIFIER	0.009
A01:1,287,460	C(110)	T(4)	upstream_gene_variant	MODIFIER	0.035
A01:1,287,229	A(108)	T(2)	upstream_gene_variant	MODIFIER	0.018
A01:1,286,002	G(151)	C(1)	missense_variant	MODERATE	0.007
A01:1,285,477	G(115)	T(1)	missense_variant	MODERATE	0.009
A01:1,285,473	C(117)	A(1)	synonymous_variant	LOW	0.008
A01:1,283,756	C(93)	T(1)	downstream_gene_variant	MODIFIER	0.011
A01:1,283,549	C(116)	T(2)	downstream_gene_variant	MODIFIER	0.017

Showing 1 to 10 of 13 rows 10 records per page

‹ 1 2 ›

Population Genetic Analysis

Parameter

Window size: 500 bp

Step size: 100 bp

Filter Sites: ☐ Exclude sites containing heterozygous genotype

Select Sub-sample sets

☐ Check All
☐ 108ASA
☐ 66-241
☐ 86-1
☐ Auburn_56
☐ Brazil_0101
☐ Brazil_011
☐ Brazil_014
☐ Brazil_cotton
☐ CIR10
☐ CIR12
☐ CIR35
☐ CIR7
☐ Coker's_Foster_300
☐ Coker139
☐ Delfos_531C
☐ Dixie_King
☐ Dixie_Triumph
☐ Esha_28
☐ Ferguson_406
☐ India_941
☐ India_NV50-70*
☐ J8891*
☐ Jimian_1#
☐ Junmian_1#
☐ King
☐ Lanart_57
☐ Lankart
☐ Latifolium_1
☐ Latifolium_13
☐ Latifolium_161
☐ Latifolium_178
☐ Latifolium_3
☐ Latifolium_39
☐ Latifolium_64
☐ Lone_Star
☐ Lumian_1#
☐ MA-6
☐ Marie-galante_17
☐ Marie-galante_36
☐ Marie-galante_62
☐ Marie-galante_7
☐ Marie-galante_71
☐ Mebane
☐ Mebane_Triumph
☐ Morrilli_2
☐ Morrilli_38
☐ Morrilli_45
☐ Palmeri_2
☐ Palmeri_5
☐ Paymaster_54
☐ Paymaster_54B
☐ Punctatum_12
☐ Punctatum_14
☐ Punctatum_24
☐ Punctatum_25
☐ Punctatum_3
☐ Punctatum_31
☐ Punctatum_5
☐ Richmondi_2
☐ Richmondi_3
☐ Richmondi_7
☐ Rowden_Malone
☐ Shiyuan_321
☐ Simian_2#
☐ Simian_3#
☐ Stoneville2B
☐ Sumian_12#
☐ Tashkent_1
☐ Tashkent_2
☐ Tashkent_3
☐ Trice
☐ Xinluzao_1#_(GK62)
☐ Xinluzao_26#
☐ Xinluzao_42#
☐ Yucatanense_1
☐ Yucatanense_2
☐ Yucatanense_3
☐ Yucatanense_5A
☐ Yucatanense_6

Show SNPs within this transcript and in surround region. Users would custom the region length

Show several population genetic value calculated from the SNP data.
Users would custom the window/step size and the samples.

Figure S5-9. Snapshots of the transcript profile page: Variation (currently we only implemented SNPs in *G. hirsutum* NAU assembly and *G. barbadense* NAU assembly). This snapshot displayed the SNPs of transcript Gh_A01G0139.1 (*GLK1*) in *G. hirsutum* (NAU assembly). Length of upstream/downstream regions, parameters of population genetics analysis (currently only nucleotide diversity and Tajima's D are implemented) and sample selection could be customized.