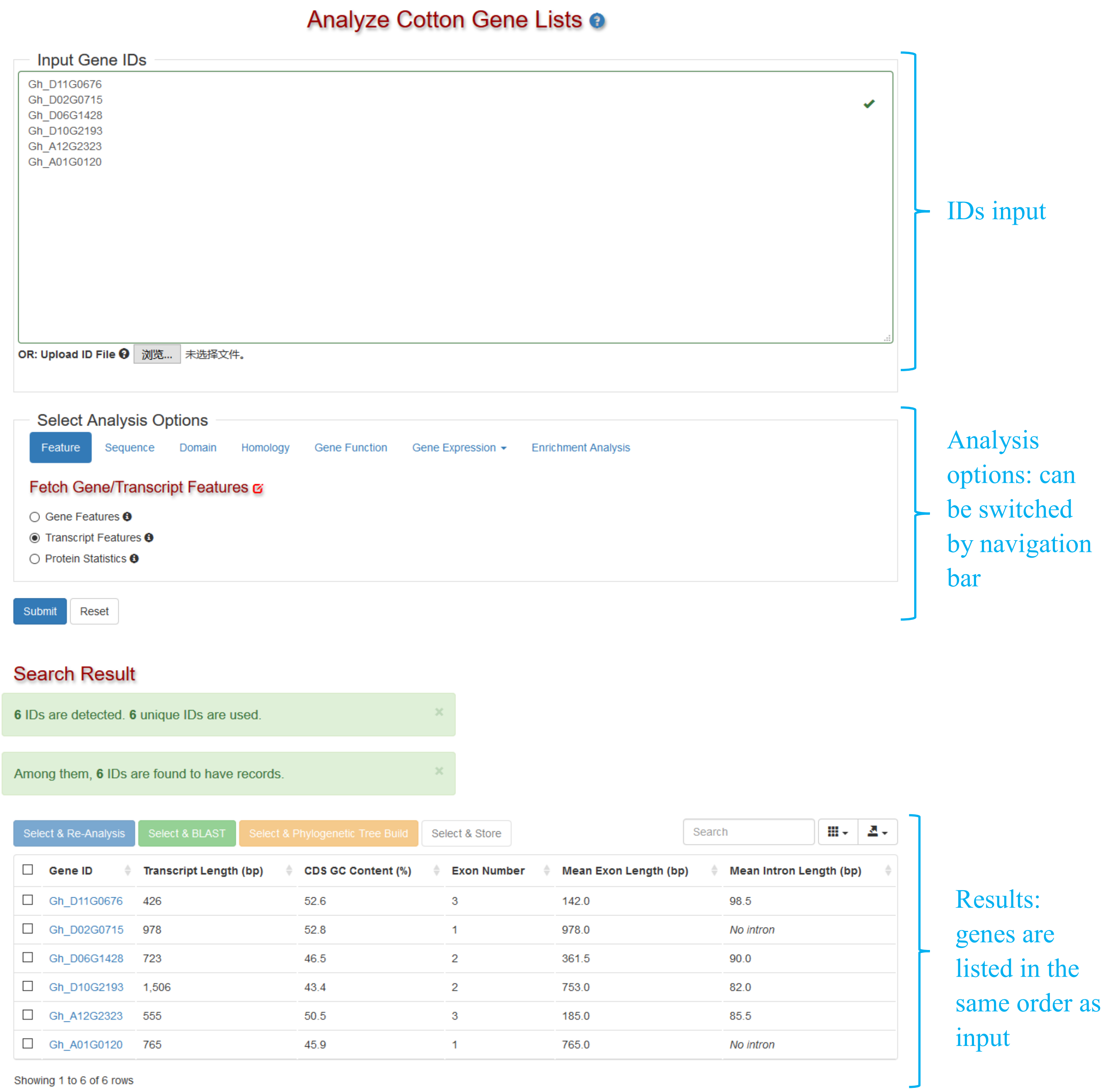
Addition File 6: Snapshots of the analysis module



Input Gene IDs

Gh_D11G0676
Gh_D02G0715
Gh_D06G1428
Gh_D10G2193
Gh_A12G2323
Gh_A01G0120

○ Gene Features
● Transcript Features
○ Protein Statistics

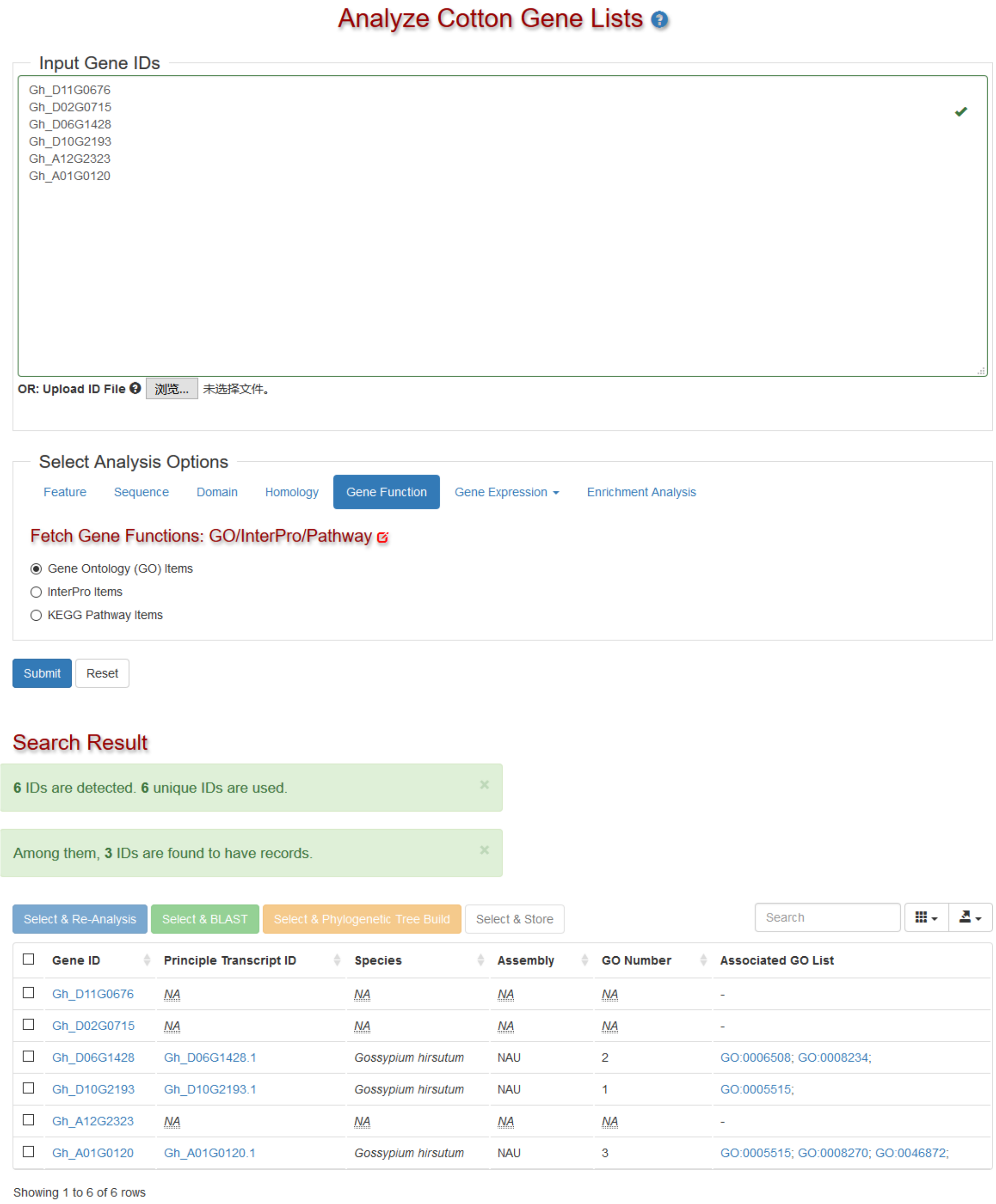
6 IDs are detected. **6** unique IDs are used.

Among them, **6** IDs are found to have records.

☐	Gene ID	Transcript Length (bp)	CDS GC Content (%)	Exon Number	Mean Exon Length (bp)	Mean Intron Length (bp)
☐	Gh_D11G0676	426	52.6	3	142.0	98.5
☐	Gh_D02G0715	978	52.8	1	978.0	*No intron*
☐	Gh_D06G1428	723	46.5	2	361.5	90.0
☐	Gh_D10G2193	1,506	43.4	2	753.0	82.0
☐	Gh_A12G2323	555	50.5	3	185.0	85.5
☐	Gh_A01G0120	765	45.9	1	765.0	*No intron*

Showing 1 to 6 of 6 rows

Figure S6-1. Snapshots of the analysis page: fetch structure information. This snapshot displayed fetching transcript structure information in six genes in *G. hirsutum* (NAU assembly). Genes are: Gh_D11G0676, Gh_D02G0715, Gh_D06G1428, Gh_D10G2193, Gh_A12G2323 and Gh_A01G0120. The fetching results are in exactly the same order as the inputted IDs. (Note: this snapshot hid several columns in the result table in order to show the main ones.)



Gh_D11G0676
Gh_D02G0715
Gh_D06G1428
Gh_D10G2193
Gh_A12G2323
Gh_A01G0120

Fetch Gene Functions: GO/InterPro/Pathway

● Gene Ontology (GO) Items
○ InterPro Items
○ KEGG Pathway Items

Search Result

6 IDs are detected. **6** unique IDs are used.

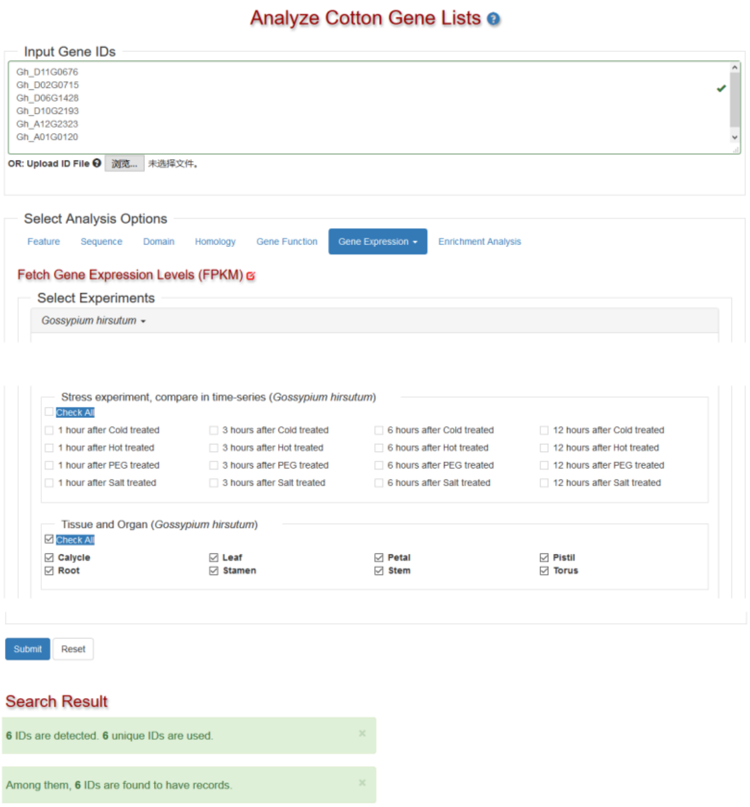
Among them, **3** IDs are found to have records.

☐	Gene ID	Principle Transcript ID	Species	Assembly	GO Number	Associated GO List
☐	Gh_D11G0676	NA	NA	NA	NA	-
☐	Gh_D02G0715	NA	NA	NA	NA	-
☐	Gh_D06G1428	Gh_D06G1428.1	*Gossypium hirsutum*	NAU	2	GO:0006508; GO:0008234;
☐	Gh_D10G2193	Gh_D10G2193.1	*Gossypium hirsutum*	NAU	1	GO:0005515;
☐	Gh_A12G2323	NA	NA	NA	NA	-
☐	Gh_A01G0120	Gh_A01G0120.1	*Gossypium hirsutum*	NAU	3	GO:0005515; GO:0008270; GO:0046872;

Showing 1 to 6 of 6 rows

Results: genes are listed in the same order as input

Figure S6-2. Snapshots of the analysis page: fetch function information. This snapshot displayed fetching associated GO terms in six genes in *G. hirsutum* (NAU assembly). Genes are: Gh_D11G0676, Gh_D02G0715, Gh_D06G1428, Gh_D10G2193, Gh_A12G2323 and Gh_A01G0120. Note that even for genes with no GO terms, they are still listed following the original order. Therefore, it is convenient to fetch different information type separately and then connect them together.



Analyze Cotton Gene Lists

Input Gene IDs

Gh_D11G0676
Gh_D02G0715
Gh_D06G1428
Gh_D10G2193
Gh_A12G2323
Gh_A01G0120

OR: Upload ID File 浏览... 未选择文件。

Select Analysis Options

Fetch Gene Expression Levels (FPKM)

Select Experiments

Gossypium hirsutum

Stress experiment, compare in time-series (*Gossypium hirsutum*)

☐ Check All

☐ 1 hour after Cold treated
☐ 1 hour after Hot treated
☐ 1 hour after PEG treated
☐ 1 hour after Salt treated
☐ 3 hours after Cold treated
☐ 3 hours after Hot treated
☐ 3 hours after PEG treated
☐ 3 hours after Salt treated
☐ 6 hours after Cold treated
☐ 6 hours after Hot treated
☐ 6 hours after PEG treated
☐ 6 hours after Salt treated
☐ 12 hours after Cold treated
☐ 12 hours after Hot treated
☐ 12 hours after PEG treated
☐ 12 hours after Salt treated

Tissue and Organ (*Gossypium hirsutum*)

☑ Check All

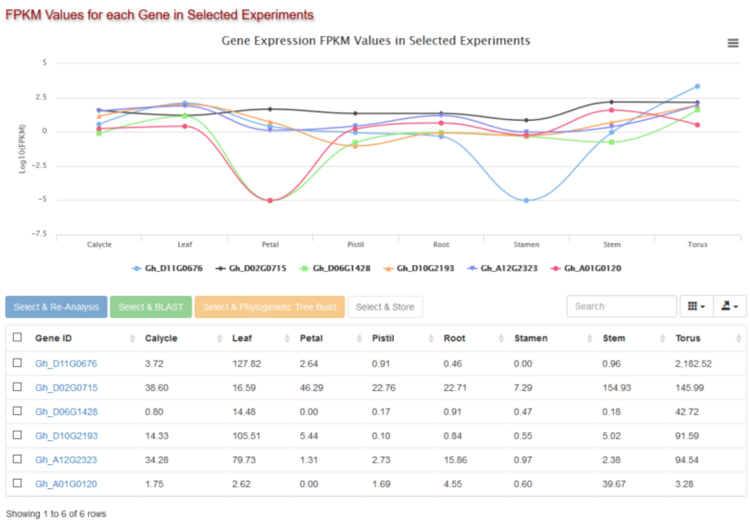
☑ **Calycle**
☑ **Root**
☑ **Leaf**
☑ **Stamen**
☑ **Petal**
☑ **Stem**
☑ **Pistil**
☑ **Torus**

Search Result

6 IDs are detected. **6** unique IDs are used.

Among them, **6** IDs are found to have records.

Select experiments (showing partial option only)



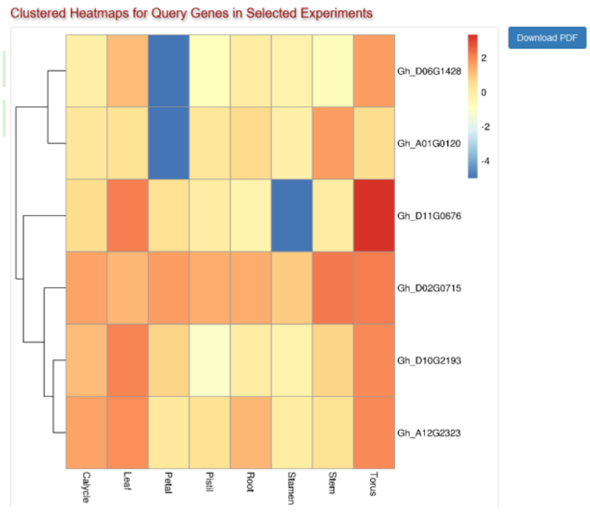
FPKM Values for each Gene in Selected Experiments

Gene Expression FPKM Values in Selected Experiments

☐	Gene ID	Calycle	Leaf	Petal	Pistil	Root	Stamen	Stem	Torus
☐	Gh_D11G0676	3.72	127.82	2.64	0.91	0.46	0.00	0.96	2,182.52
☐	Gh_D02G0715	38.60	16.59	46.29	22.76	22.71	7.29	154.93	145.99
☐	Gh_D06G1428	0.80	14.48	0.00	0.17	0.91	0.47	0.18	42.72
☐	Gh_D10G2193	14.33	105.51	5.44	0.10	0.84	0.55	5.02	91.59
☐	Gh_A12G2323	34.28	79.73	1.31	2.73	15.86	0.97	2.38	94.54
☐	Gh_A01G0120	1.75	2.62	0.00	1.69	4.55	0.60	39.67	3.28

Showing 1 to 6 of 6 rows

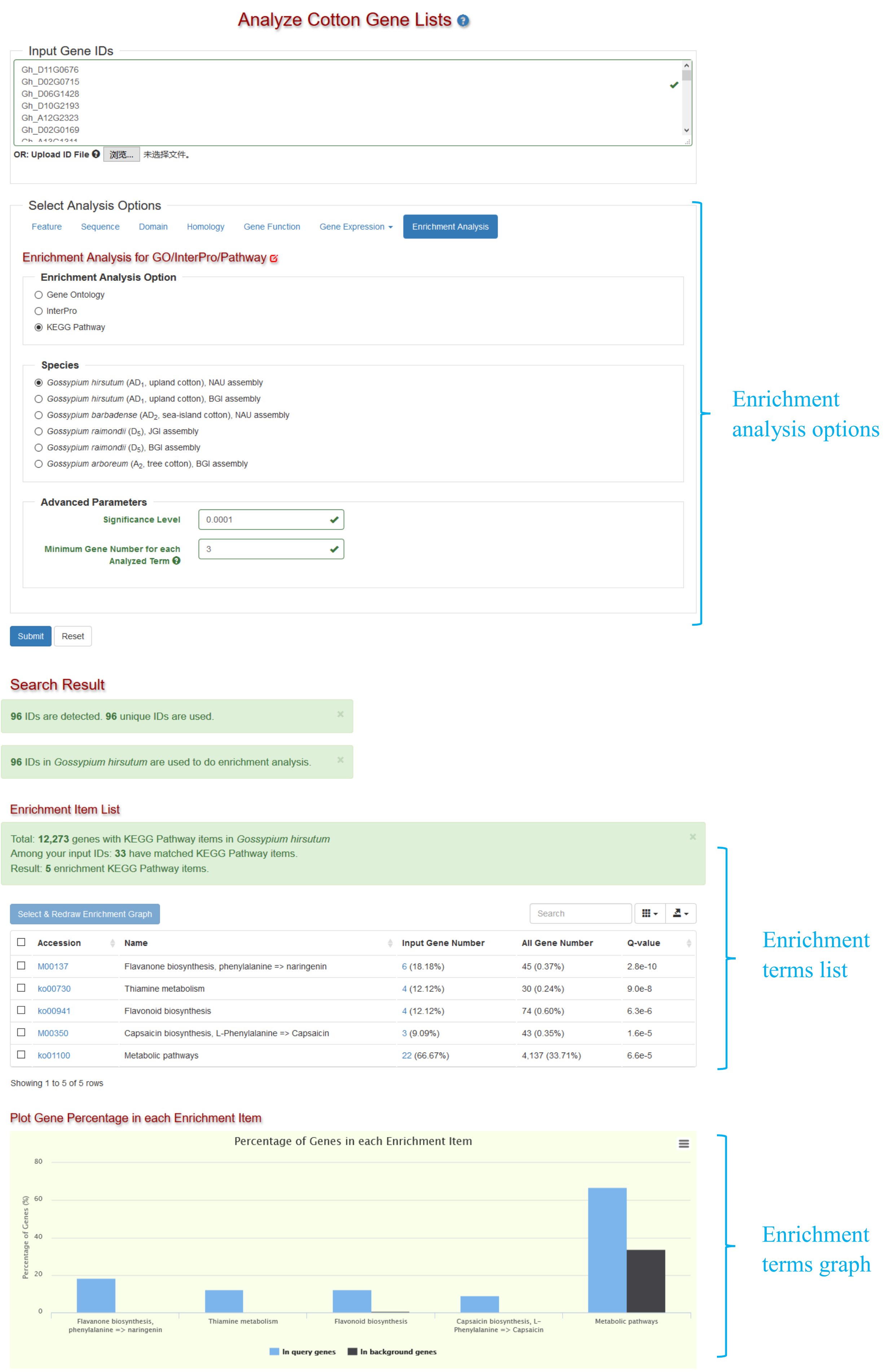
List expression level (FPKM) for each gene in selected experiments



Clustered Heatmaps for Query Genes in Selected Experiments

Cluster input genes based on their expression atlas in selected experiments

Figure S6-3. Snapshots of the analysis page: fetch expression level information. This snapshot displayed fetching expression levels (FPKM) in six genes of *G. hirsutum* (NAU assembly) in eight organs. Genes are: Gh_D11G0676, Gh_D02G0715, Gh_D06G1428, Gh_D10G2193, Gh_A12G2323 and Gh_A01G0120. In addition to the result table, CottonFGD will generate a heatmap clustered by the expression patterns for these query genes. The cluster is done by the pheatmap library (https://cran.r-project.org/web/packages/pheatmap/index.html).



Analyze Cotton Gene Lists

Input Gene IDs

Gh_D11G0676
Gh_D02G0715
Gh_D06G1428
Gh_D10G2193
Gh_A12G2323
Gh_D02G0169

OR: Upload ID File 未选择文件。

Select Analysis Options

Enrichment Analysis for GO/InterPro/Pathway

Enrichment Analysis Option

○ Gene Ontology
○ InterPro
● KEGG Pathway

Species

● *Gossypium hirsutum* (AD_1, upland cotton), NAU assembly
○ *Gossypium hirsutum* (AD_1, upland cotton), BGI assembly
○ *Gossypium barbadense* (AD_2, sea-island cotton), NAU assembly
○ *Gossypium raimondii* (D_5), JGI assembly
○ *Gossypium raimondii* (D_5), BGI assembly
○ *Gossypium arboreum* (A_2, tree cotton), BGI assembly

Advanced Parameters

Significance Level: 0.0001
Minimum Gene Number for each Analyzed Term: 3

Search Result

96 IDs are detected. **96** unique IDs are used.

96 IDs in *Gossypium hirsutum* are used to do enrichment analysis.

Enrichment Item List

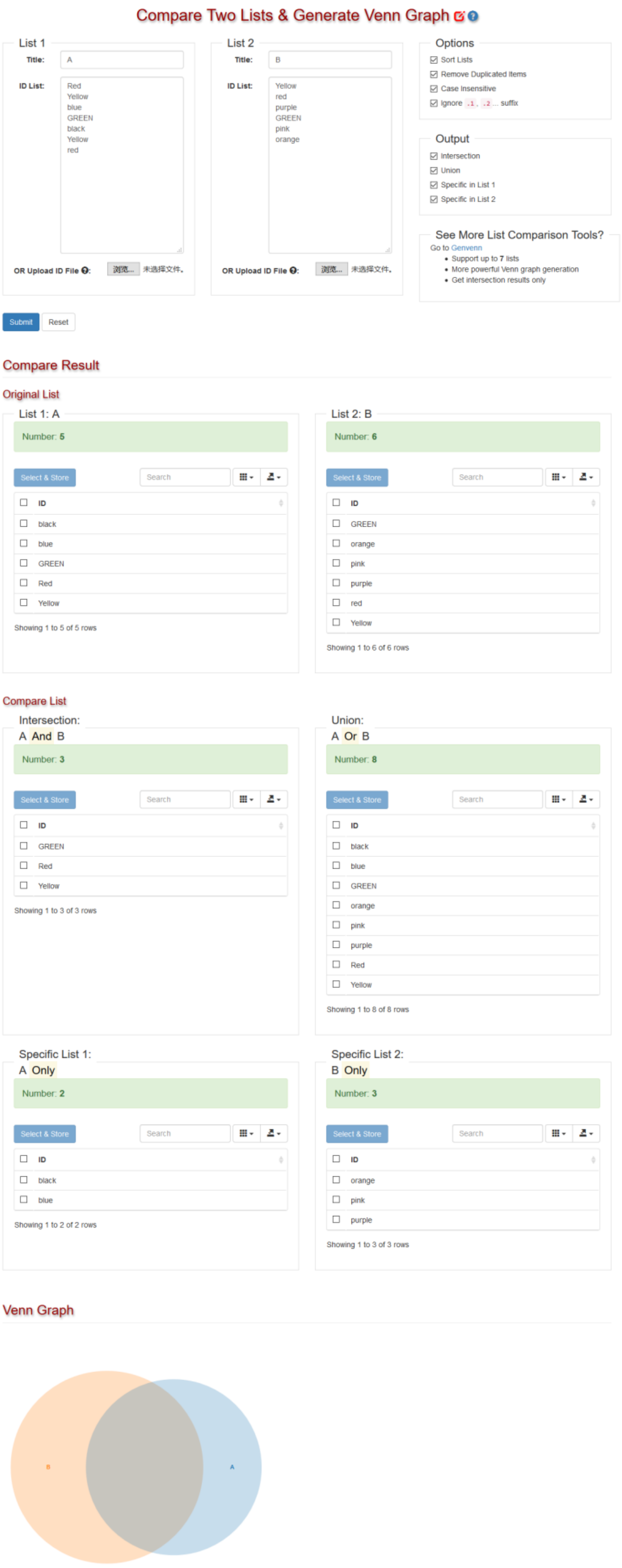
Total: **12,273** genes with KEGG Pathway items in *Gossypium hirsutum*
Among your input IDs: **33** have matched KEGG Pathway items.
Result: **5** enrichment KEGG Pathway items.

☐	Accession	Name	Input Gene Number	All Gene Number	Q-value
☐	M00137	Flavanone biosynthesis, phenylalanine => naringenin	6 (18.18%)	45 (0.37%)	2.8e-10
☐	ko00730	Thiamine metabolism	4 (12.12%)	30 (0.24%)	9.0e-8
☐	ko00941	Flavonoid biosynthesis	4 (12.12%)	74 (0.60%)	6.3e-6
☐	M00350	Capsaicin biosynthesis, L-Phenylalanine => Capsaicin	3 (9.09%)	43 (0.35%)	1.6e-5
☐	ko01100	Metabolic pathways	22 (66.67%)	4,137 (33.71%)	6.6e-5

Showing 1 to 5 of 5 rows

Plot Gene Percentage in each Enrichment Item

Figure S6-4. Snapshots of the analysis page: enrichment analysis. This snapshot displayed doing KEGG pathway enrichment analysis in 96 genes of *G. hirsutum* (NAU assembly) using $P < 0.0001$ as thresholds.



Compare Two Lists & Generate Venn Graph

List 1
Title: A
ID List:
Red
Yellow
blue
GREEN
black
Yellow
red
OR Upload ID File:

List 2
Title: B
ID List:
Yellow
red
purple
GREEN
pink
orange
OR Upload ID File:

Options
☑ Sort Lists
☑ Remove Duplicated Items
☑ Case Insensitive
☑ Ignore `.1`, `.2` ... suffix

Output
☑ Intersection
☑ Union
☑ Specific in List 1
☑ Specific in List 2

See More List Comparison Tools?
Go to Genvenn
- Support up to **7** lists
- More powerful Venn graph generation
- Get intersection results only

Compare Result

Original List

List 1: A
Number: **5**

☐	**ID**
☐	black
☐	blue
☐	GREEN
☐	Red
☐	Yellow

Showing 1 to 5 of 5 rows

List 2: B
Number: **6**

☐	**ID**
☐	GREEN
☐	orange
☐	pink
☐	purple
☐	red
☐	Yellow

Showing 1 to 6 of 6 rows

Compare List

Intersection:
A And B
Number: **3**

☐	**ID**
☐	GREEN
☐	Red
☐	Yellow

Showing 1 to 3 of 3 rows

Union:
A Or B
Number: **8**

☐	**ID**
☐	black
☐	blue
☐	GREEN
☐	orange
☐	pink
☐	purple
☐	Red
☐	Yellow

Showing 1 to 8 of 8 rows

Specific List 1:
A Only
Number: **2**

☐	**ID**
☐	black
☐	blue

Showing 1 to 2 of 2 rows

Specific List 2:
B Only
Number: **3**

☐	**ID**
☐	orange
☐	pink
☐	purple

Showing 1 to 3 of 3 rows

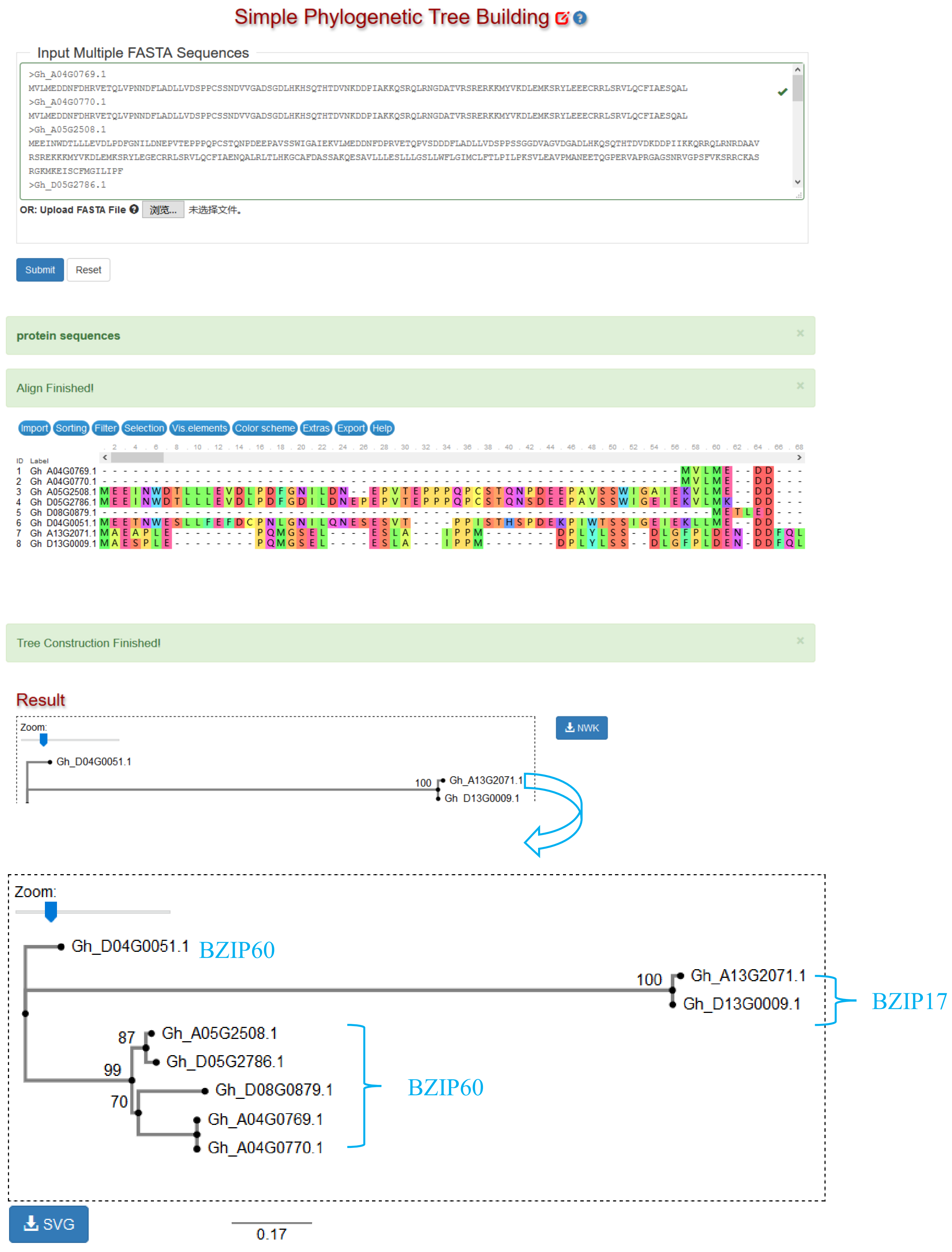
Venn Graph

Input two lists

Compare result:
Intersection
Union
Specific

Venn Graph
showing count
number

Figure S6-5. Snapshots of the list compare page. Note this page only conduct comparisons based on inputted IDs. Therefore, it could be applied to any ID lists instead of restricting on cotton gene IDs.



Simple Phylogenetic Tree Building

Input Multiple FASTA Sequences

```
>Gh_A04G0769.1
MVLMEDDNFDHRVETQLVPNNDFLADLLVDSPPCSSNDVVGADSGDLHKHSQTHTDVNKDDPIAKKQSRQLRNGDATVRSRERKKMYVKDLEMKSRYLEEECRRLSRVLQCFIAESQAL
>Gh_A04G0770.1
MVLMEDDNFDHRVETQLVPNNDFLADLLVDSPPCSSNDVVGADSGDLHKHSQTHTDVNKDDPIAKKQSRQLRNGDATVRSRERKKMYVKDLEMKSRYLEEECRRLSRVLQCFIAESQAL
>Gh_A05G2508.1
MEEINWDTLLLEVDLPDFGNILDNEPVTEPPPQPCSTQNPDEEPAVSSWIGAIEKVLMEDDNFDPRVETQPVSDDDFLADLLVDSPPSSGGDVAGVDGADLHKQSQTHTDVDKDDPIIKKQRRQLRNRDAAV
RSREKKKMYVKDLEMKSRYLEGECRRLSRVLQCFIAENQALRLTLHKGCAFDASSAKQESAVLLLESLLLGSLLWFLGIMCLFTLPILPKSVLEAVPMANEETQGPERVAPRGAGSNRVGPSFVKSRRCKAS
RGKMKEISCFMGILIPF
>Gh_D05G2786.1
```

OR: Upload FASTA File 浏览... 未选择文件。

protein sequences

Align Finished!

```
                  . 2 . 4 . 6 . 8 . 10. 12. 14. 16. 18. 20. 22. 24. 26. 28. 30. 32. 34. 36. 38. 40. 42. 44. 46. 48. 50. 52. 54. 56. 58. 60. 62. 64. 66. 68
ID Label
1  Gh A04G0769.1  --------------------------------------------------------MVLME--DD---
2  Gh A04G0770.1  --------------------------------------------------------MVLME--DD---
3  Gh A05G2508.1  MEEINWDTLLLEVDLPDFGNILDN--EPVTEPPPQPCSTQNPDEEPAVSSWIGAIEKVLME--DD---
4  Gh D05G2786.1  MEEINWDTLLLEVDLPDFGDILDNEPEPVTEPPPQPCSTQNSDEEPAVSSWIGEIEKVLMK--DD---
5  Gh D08G0879.1  -----------------------------------------------------------METLED---
6  Gh D04G0051.1  MEETNWESLLFEFDCPNLGNILQNESESVT----PPISTHSPDEKPIWTSSIGEIEKLLME--DD---
7  Gh A13G2071.1  MAEAPLE--------PQMGSEL----ESLA---IPPM-------DPLYLSS--DLGFPLDEN-DDFQL
8  Gh D13G0009.1  MAESPLE--------PQMGSEL----ESLA---IPPM-------DPLYLSS--DLGFPLDEN-DDFQL
```

Tree Construction Finished!

Result

Figure S6-6. Snapshots of the phylogenetic tree build page. This snapshot displayed a phylogenetic tree built by six *BZIP60* (Gh_A04G0770, Gh_A04G0769, Gh_D04G0051, Gh_A05G2508, Gh_D05G2786, Gh_D08G0879) and two *BZIP17* (Gh_A13G2071, Gh_D13G0009) genes in *G. hirsutum*.