

FT蛋白（Flowering Locus T）与植物花发育的生物信息学分析

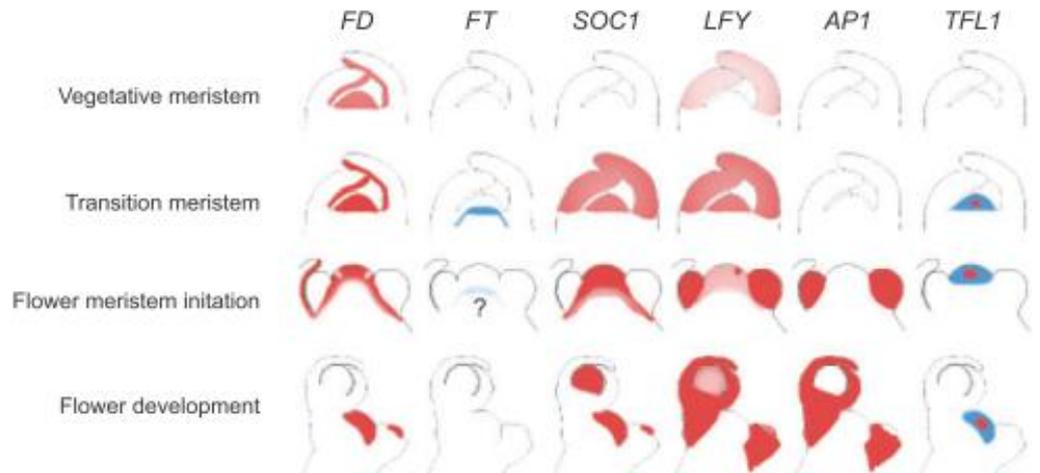
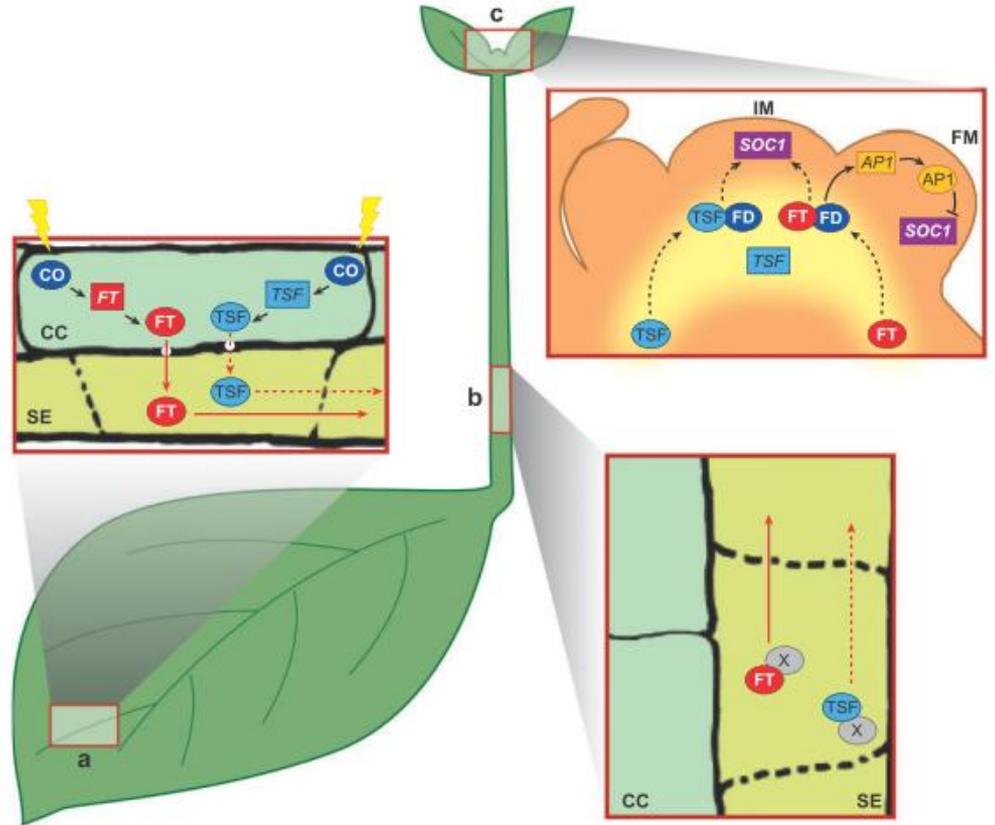
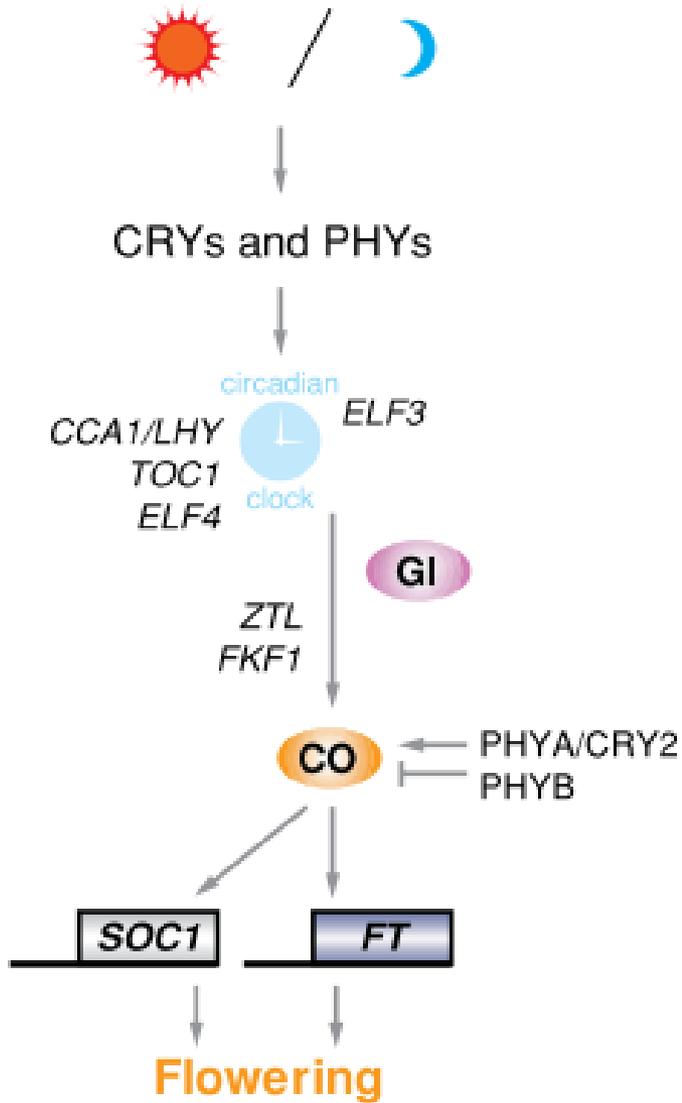
组号：15

组员：张小兵、孙占敏、徐明、
于洋

什么是FT?

- FT是开花素的可能组成物质，促使植物生殖生长向花发育阶段转变.

Arabidopsis thaliana



- 其在开花诱导过程中同FD/BZIP14存在相互作用；
- 蛋白主要定位于叶片筛管，翻译产物可通过植物韧皮部进行运输，速度在 1.2 ~ 3.5 mm/h ；
- 基因的表达受到光节律调控。基因表达受到FLC的抑制；
- FT蛋白的Tyr-85突变为His得到TFL1，是开花的抑制蛋白；
- FT蛋白属于PEBP家族。（phosphatidylethanolamine-binding protein family）

AtFT蛋白的直系同源与旁系同源关系

AtFT Ortholog

FT_ARATH Protein FLOWERING LOCUS T OS=Arabidopsis...

Query ID Id|76223
Description FT_ARATH Protein FLOWERING LOCUS T OS=Arabidopsis thaliana GN=FT PE=1
SV=2
Molecule type amino acid
Query Length 175

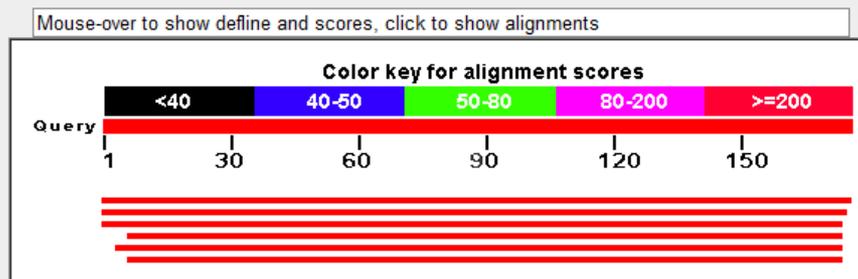
Database Name swissprot
Description Non-redundant SwissProt sequences
Program BLASTP 2.2.26+ [Citation](#)

No new sequences were found above the 0.005 threshold

Other reports: [Search Summary](#) [Taxonomy reports](#) [Distance tree of results](#) [Multiple alignment](#)

Graphic Summary

Distribution of 6 Blast Hits on the Query Sequence



Descriptions

Legend for links to other resources: [U](#) UniGene [E](#) GEO [G](#) Gene [S](#) Structure [M](#) Map Viewer [P](#) PubChem BioAssay

NEW - alignment score below the threshold on the previous iteration

- alignment was checked on the previous iteration

Run PSI-Blast iteration 3 with max

Sequences producing significant alignments with E-value BETTER than threshold

Accession	Description	Max score	Total score	Query coverage	E value	Links
<input checked="" type="checkbox"/> Q9SXZ2.2	RecName: Full=Protein FLOWERING LOCUS T	309	309	100%	1e-113	GM
<input checked="" type="checkbox"/> Q9S7R5.1	RecName: Full=Protein TWIN SISTER of FT; AltName: Full=TFL1-like p	304	304	99%	2e-111	GM
<input checked="" type="checkbox"/> P93003.1	RecName: Full=Protein TERMINAL FLOWER 1	287	287	98%	7e-105	GM
<input checked="" type="checkbox"/> Q9ZNV5.1	RecName: Full=Protein CENTRORADIALIS-like	280	280	95%	5e-102	GM
<input checked="" type="checkbox"/> Q9FIT4.1	RecName: Full=Protein BROTHER of FT and TFL 1	273	273	97%	2e-99	GM
<input checked="" type="checkbox"/> Q9XFK7.1	RecName: Full=Protein MOTHER of FT and TF 1	265	265	95%	3e-96	GM

Run PSI-Blast iteration 3 with max

- 通过对AtFT基因的直系同源分析显示Flowering locus T蛋白与protein TWIN SISTER of FT蛋白、TERMINAL flowering 1、CENTRORADIALIS-like、PROTEIN BROTHER of and TFL1、PROTEIN BROTHER of and TFL1，同源相似性较高，几个蛋白均参与到植物花发育及花器官形态建成的相关过程。

AtFT paralog use Uniprot

PSI blast Iteration 2

FT_ARATH Protein FLOWERING LOCUS T OS=Arabidopsis...

[Skip to the first new sequence](#)

Query ID Icd|65921
Description FT_ARATH Protein FLOWERING LOCUS T OS=Arabidopsis thaliana GN=FT PE=1 SV=2
Molecule type amino acid
Query Length 175

Database Name swissprot
Description Non-redundant SwissProt sequences
Program BLASTP 2.2.26+ [▶ Citation](#)

Other reports: [▶ Search Summary](#) [\[Taxonomy reports\]](#) [\[Distance tree of results\]](#) [\[Multiple alignment\]](#)

▼ Sequences producing significant alignments with E-value BETTER than threshold

Accession	Description	Max score	Total score	Query coverage	E value	Links
<input checked="" type="checkbox"/> Q9SXZ2.2	RecName: Full=Protein FLOWERING LOCUS T	271	271	100%	4e-93	GM
<input checked="" type="checkbox"/> Q93WI9.1	RecName: Full=Protein HEADING DATE 3A; AltName: Full=FT-like prot	263	263	97%	5e-90	G
<input checked="" type="checkbox"/> Q9S7R5.1	RecName: Full=Protein TWIN SISTER of FT; AltName: Full=TFL1-like p	262	262	99%	1e-89	GM
<input checked="" type="checkbox"/> Q8VWH2.1	RecName: Full=Protein HEADING DATE 3B; AltName: Full=FT-like prot	259	259	99%	1e-88	G
<input checked="" type="checkbox"/> Q9XH43.1	RecName: Full=CEN-like protein 2	253	253	98%	5e-86	
<input checked="" type="checkbox"/> Q9XH42.1	RecName: Full=CEN-like protein 4	251	251	98%	1e-85	
<input checked="" type="checkbox"/> O82088.1	RecName: Full=Protein SELF-PRUNING	249	249	98%	1e-84	G
<input checked="" type="checkbox"/> P93003.1	RecName: Full=Protein TERMINAL FLOWER 1	245	245	98%	6e-83	GM
<input checked="" type="checkbox"/> Q9XH44.1	RecName: Full=CEN-like protein 1	241	241	98%	1e-81	
<input checked="" type="checkbox"/> Q41261.1	RecName: Full=Protein CENTRORADIALIS	239	239	98%	2e-80	
<input checked="" type="checkbox"/> Q9ZNV5.1	RecName: Full=Protein CENTRORADIALIS-like	238	238	98%	2e-80	GM
<input checked="" type="checkbox"/> Q9FIT4.1	RecName: Full=Protein BROTHER of FT and TFL 1	232	232	97%	5e-78	GM
<input checked="" type="checkbox"/> Q9XFK7.1	RecName: Full=Protein MOTHER of FT and TF 1	222	222	96%	7e-74	GM
<input checked="" type="checkbox"/> P31729.2	RecName: Full=OV-16 antigen; Flags: Precursor	195	195	93%	3e-63	
<input checked="" type="checkbox"/> P54186.1	RecName: Full=Protein D1	193	193	82%	7e-63	
<input checked="" type="checkbox"/> P13696.2	RecName: Full=Phosphatidylethanolamine-binding protein 1; Short=	181	181	82%	1e-57	GM
<input checked="" type="checkbox"/> P54188.1	RecName: Full=Protein D3	178	178	74%	3e-57	
<input checked="" type="checkbox"/> Q8VIN1.1	RecName: Full=Phosphatidylethanolamine-binding protein 2; Short=	178	178	80%	2e-56	G
<input checked="" type="checkbox"/> Q3YIX4.1	RecName: Full=Phosphatidylethanolamine-binding protein 1; Short=	177	177	82%	3e-56	GM
<input checked="" type="checkbox"/> Q8MK67.1	RecName: Full=Phosphatidylethanolamine-binding protein 1; Short=	177	177	82%	5e-56	GM
<input checked="" type="checkbox"/> Q16264.1	RecName: Full=Phosphatidylethanolamine-binding protein homolog f	178	178	95%	5e-56	G
<input checked="" type="checkbox"/> P31044.3	RecName: Full=Phosphatidylethanolamine-binding protein 1; Short=	176	176	82%	1e-55	GM
<input checked="" type="checkbox"/> Q5R4R0.3	RecName: Full=Phosphatidylethanolamine-binding protein 1; Short=	175	175	82%	4e-55	GM
<input checked="" type="checkbox"/> P48737.2	RecName: Full=Phosphatidylethanolamine-binding protein 1; Short=	175	175	82%	4e-55	

AtFT paralog use refseq

PSI blast Iteration 2

FT_ARATH Protein FLOWERING LOCUS T OS=Arabidopsis...

[Skip to the first new sequence](#)

Query ID |cl|60146
Description FT_ARATH Protein FLOWERING LOCUS T OS=Arabidopsis thaliana GN=FT PE=1 SV=2
Molecule type amino acid
Query Length 175

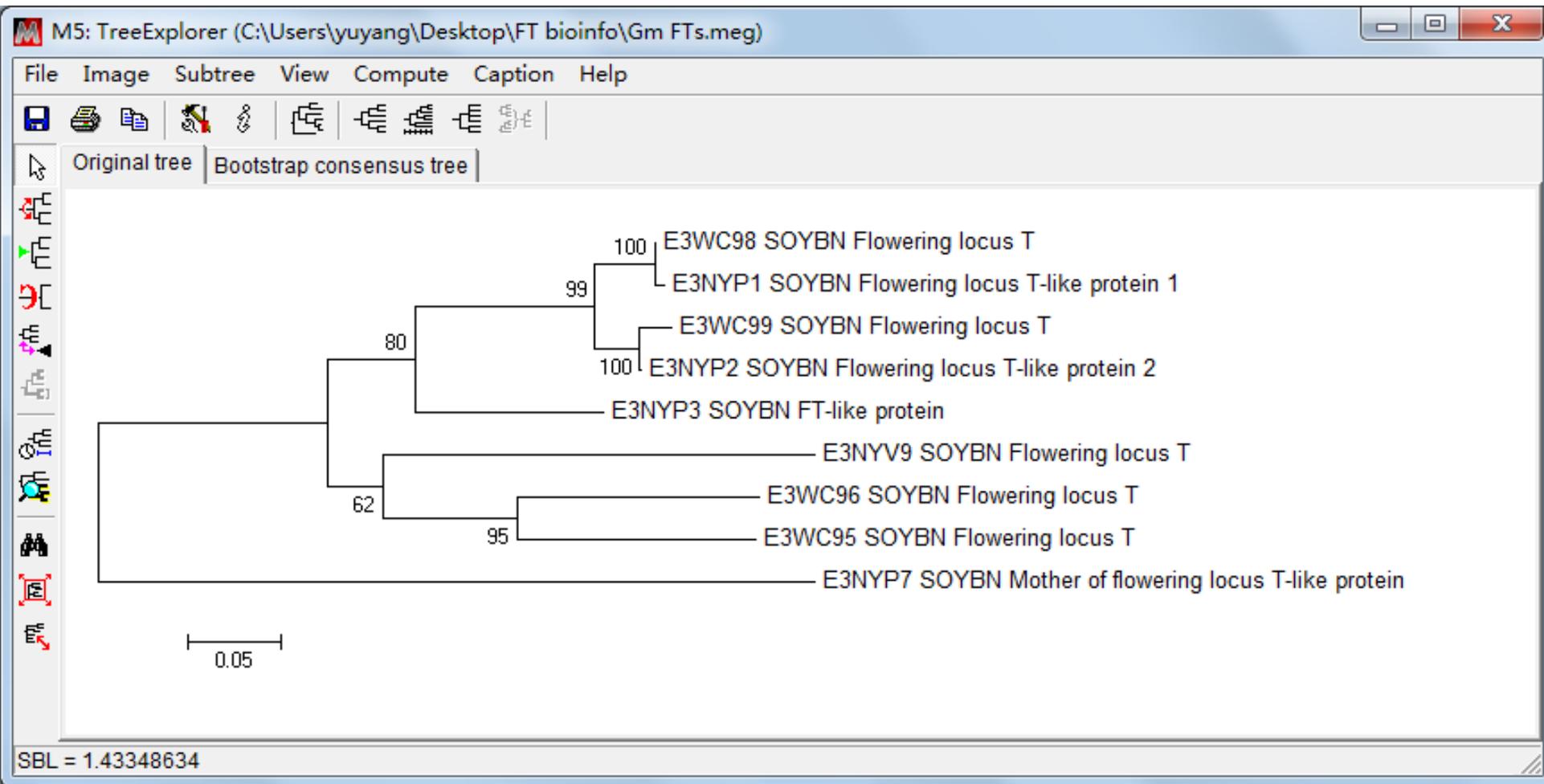
Database Name refseq_protein
Description NCBI Protein Reference Sequences
Program BLASTP 2.2.26+ [Citation](#)

Other reports: [Search Summary](#) [Taxonomy reports](#) [Distance tree of results](#) [Multiple alignment](#)

Accession	Description	Max score	Total score	Query coverage	E value	Links
<input checked="" type="checkbox"/> XP_002270408.1	PREDICTED: flowering locus T-like protein [Vitis vinifera]	352	352	96%	3e-124	UG
<input checked="" type="checkbox"/> XP_002334492.1	predicted protein [Populus trichocarpa] >ref XP_002334306.1 pr	350	350	96%	4e-123	UG
<input checked="" type="checkbox"/> XP_002311264.1	predicted protein [Populus trichocarpa]	349	349	96%	1e-122	UG
<input checked="" type="checkbox"/> NP_176726.1	protein FLOWERING LOCUS T [Arabidopsis thaliana]	348	348	100%	2e-122	UGM
<input checked="" type="checkbox"/> XP_002886920.1	hypothetical protein ARALYDRAFT_894103 [Arabidopsis lyrata sub	346	346	100%	9e-122	G
<input checked="" type="checkbox"/> NP_001106252.1	LOC100127524 [Zea mays]	346	346	100%	9e-122	UG
<input checked="" type="checkbox"/> XP_002436509.1	hypothetical protein SORBIDRAFT_10g003940 [Sorghum bicolor]	343	343	97%	2e-120	UG
<input checked="" type="checkbox"/> NP_001056860.1	Os06g0157700 [Oryza sativa Japonica Group]	343	343	97%	2e-120	UG
<input checked="" type="checkbox"/> XP_002867884.1	predicted protein [Arabidopsis lyrata subsp. lyrata]	340	340	99%	2e-119	G
<input checked="" type="checkbox"/> XP_002514784.1	phosphatidylethanolamine-binding protein, putative [Ricinus com	337	337	96%	5e-118	G
<input checked="" type="checkbox"/> NP_001106251.1	ZCN14 protein [Zea mays]	336	336	96%	1e-117	UG
<input checked="" type="checkbox"/> NP_193770.1	protein TWIN SISTER of FT [Arabidopsis thaliana]	334	334	99%	5e-117	UGM
<input checked="" type="checkbox"/> XP_002528554.1	phosphatidylethanolamine-binding protein, putative [Ricinus com	334	334	97%	9e-117	G
<input checked="" type="checkbox"/> XP_002446704.1	hypothetical protein SORBIDRAFT_06g020850 [Sorghum bicolor]	332	332	96%	3e-116	UG
<input checked="" type="checkbox"/> NP_001056859.1	Os06g0157500 [Oryza sativa Japonica Group]	332	332	96%	3e-116	UG
<input checked="" type="checkbox"/> XP_002454134.1	hypothetical protein SORBIDRAFT_04g025210 [Sorghum bicolor]	331	331	96%	7e-116	UG
<input checked="" type="checkbox"/> XP_002457125.1	hypothetical protein SORBIDRAFT_03g001700 [Sorghum bicolor]	336	336	96%	1e-115	UG
<input checked="" type="checkbox"/> NP_001106253.1	ZCN16 protein [Zea mays]	330	330	96%	2e-115	UG
<input checked="" type="checkbox"/> XP_002489297.1	hypothetical protein SORBIDRAFT_0010s003120 [Sorghum bicolor]	329	329	96%	9e-115	UG
<input checked="" type="checkbox"/> NP_001042411.1	Os01g0218500 [Oryza sativa Japonica Group]	332	332	96%	9e-115	UG
<input checked="" type="checkbox"/> NP_001053150.1	Os04g0488400 [Oryza sativa Japonica Group]	327	327	96%	5e-114	UG
<input checked="" type="checkbox"/> NP_001106254.1	ZCN17 protein [Zea mays]	325	325	96%	3e-113	UG
<input checked="" type="checkbox"/> NP_001106257.1	ZCN25 protein [Zea mays]	324	324	96%	6e-113	UG
<input checked="" type="checkbox"/> NP_001106256.1	ZCN19 protein [Zea mays]	324	324	96%	6e-113	UG
<input checked="" type="checkbox"/> NP_001236597.1	brother of FT and TFL1 protein [Glycine max]	311	311	97%	6e-108	G
<input checked="" type="checkbox"/> XP_002282828.1	PREDICTED: similar to TFL1B protein [Vitis vinifera]	311	311	97%	6e-108	UG
<input checked="" type="checkbox"/> XP_002451827.1	hypothetical protein SORBIDRAFT_04g008320 [Sorghum bicolor]	310	310	96%	2e-107	G
<input checked="" type="checkbox"/> XP_002438551.1	hypothetical protein SORBIDRAFT_10g021790 [Sorghum bicolor]	309	309	96%	5e-107	UG
<input checked="" type="checkbox"/> XP_002312811.1	predicted protein [Populus trichocarpa]	308	308	97%	1e-106	UG
<input checked="" type="checkbox"/> XP_002442808.1	hypothetical protein SORBIDRAFT_08g003210 [Sorghum bicolor]	307	307	95%	3e-106	UG

- 旁系同源BLAST结果显示，FT在多种主要作物中存在保守存在。
- 本课题组通过同源克隆获得了大豆GmFT2a，构建过表达载体转化获得了转基因大豆株系，转基因后代性状稳定，转基因株系开花提前，使得南方短日照转基因受体品种可在北方长日照条件下提前开花。
- 对大豆中存在的FT基因进行直系同源BLAST并构建系统发育树。

大豆FT蛋白进化关系

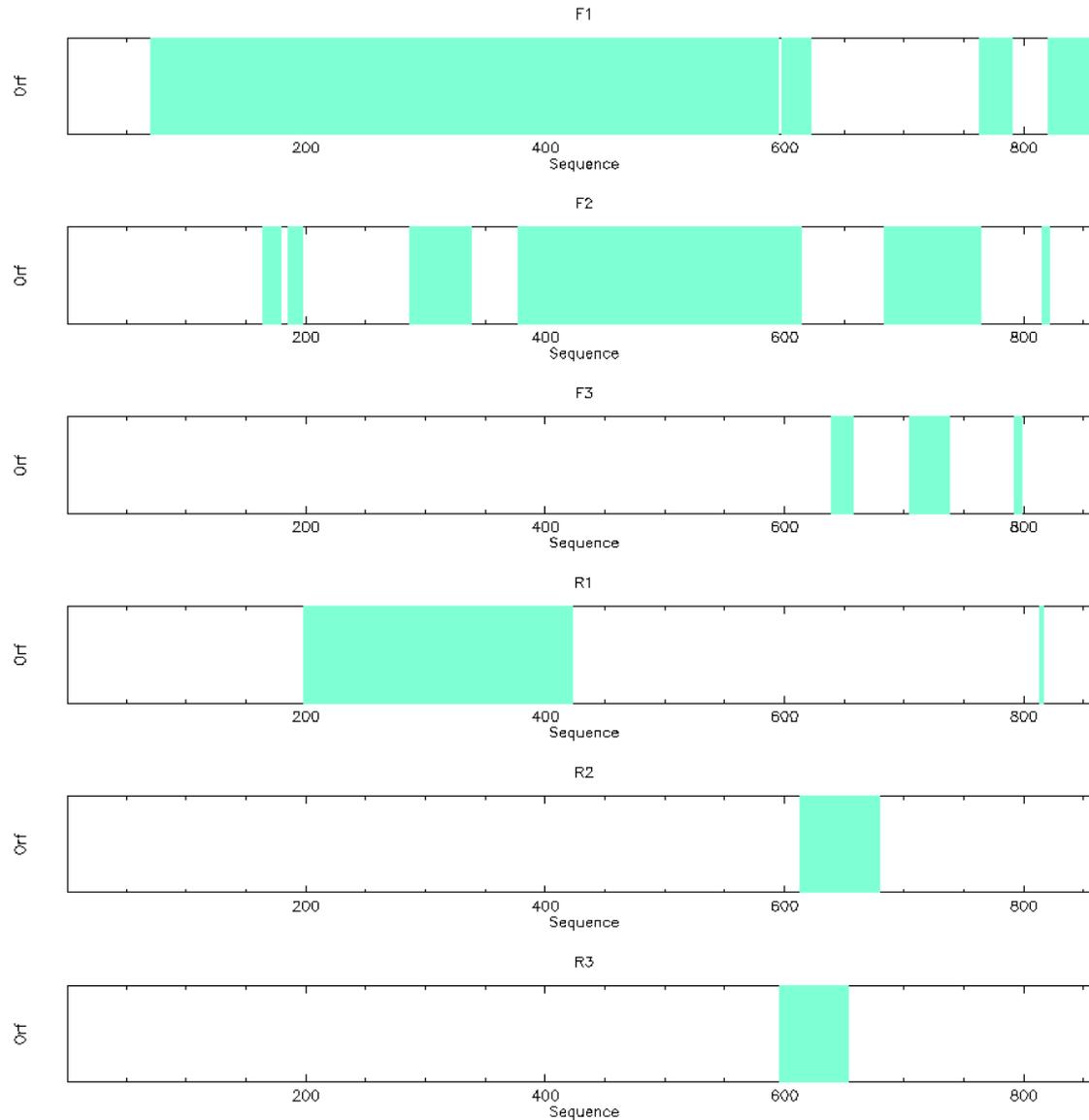


AtFT基因ORF相关分析

GETORF: OUTPUT

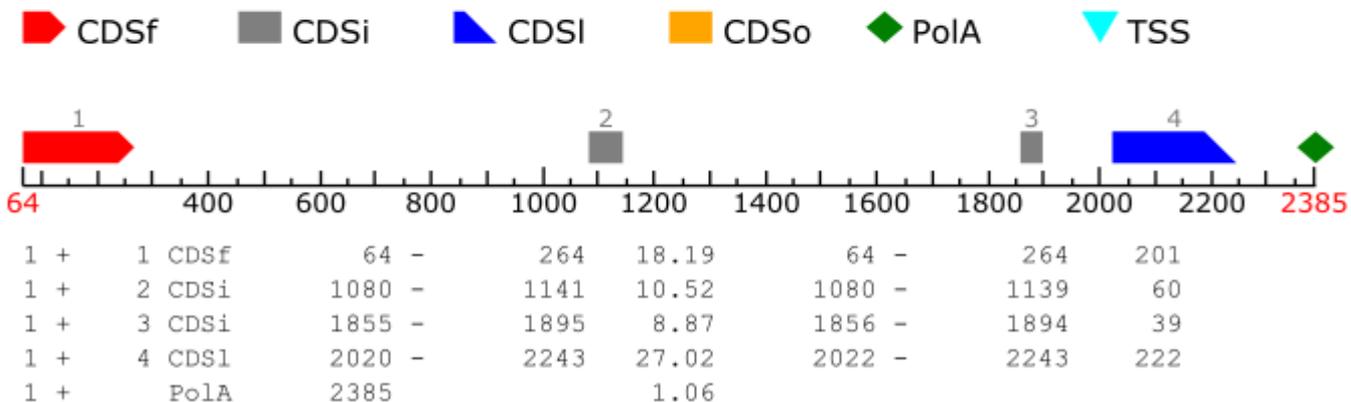
OUTPUT FILE: outseq [[RIGHT CLICK TO SAVE](#)]

```
>AB027504.1_1 [70 - 594] Arabidopsis thaliana FT (FLOWERING LOCUS T) mRNA, complete cds  
MSINIRDPLIVSRVVGDVLDPFNRSITLKVITYGQREVINGLDLRPSQVQNKPRVEIGGED  
LRNFYTLVMVDPDPVSPSPNPHLREYLHWLVTDIPATTGTTFGNEIVCYENPSPTAGIHRV  
VFILFRQLGRQTVYAPGWRQNFNTREFAEIYNLGLPVAAVFYNCQRESGCGGRRL
```



- 调出AtFT基因组片段信息，用softberry软件对上述基因组序列进行预测分析，参数设置：organism选择：**Dicot plants (Arabidopsis)**

FGENESH 2.6 Prediction of potential genes in Dicot_arab genomic DNA
 Seq name: gi|6117977|gb|AF152096.1| Arabidopsis thaliana flowering locus T (FT) gene, co
 Length of sequence: 2483
 Number of predicted genes 1: in +chain 1, in -chain 0.
 Number of predicted exons 4: in +chain 4, in -chain 0.
 Positions of predicted genes and exons: Variant 1 from 1, Score:45.537515

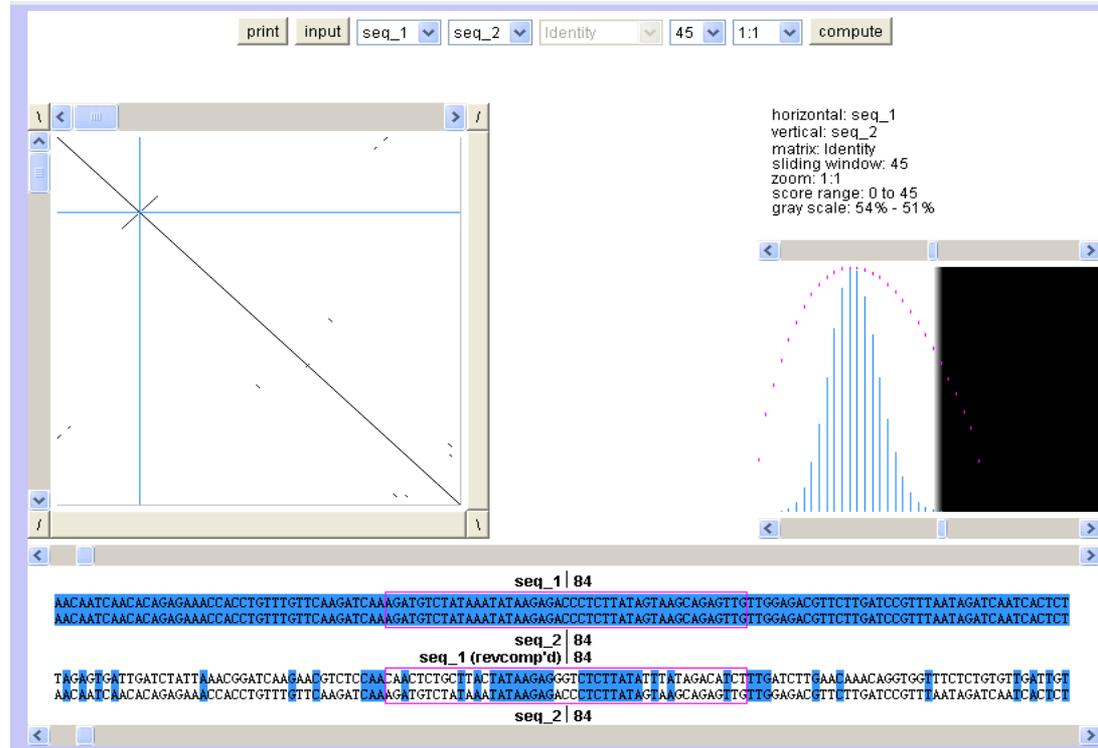
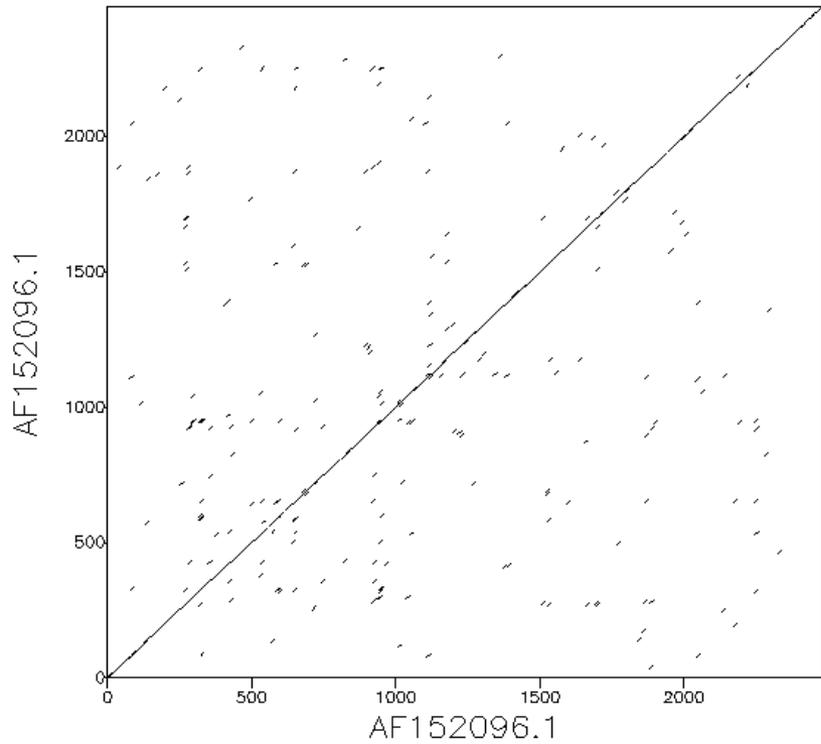


- 述预测得到的mRNA编码序列和Uniprot中的mRNA编码序列进行比对，采用NEEDLE，参数设置：matrix file 选择：EDNAFULL for nucleic sequence,gap opening penalty:10,gap extension penalty:05,比对结果如下：

Table. Pairwise Alignment Result

Length	Score	Identity	Similarity	Gaps
528	2640.0	528/528 (100.0%)	528/528 (100.0%)	0/528 (0.0%)

DotMather和Dotlet对AtFT基因gDNA进行分析显示结果



PEBP家族相关信息

PSI blast Iteration 1

E3NYP3_SOYBN FT-like protein OS=Glycine max...

Query ID Id|71198
Description E3NYP3_SOYBN FT-like protein OS=Glycine max GN=FTL3 PE=2 SV=1
Molecule type amino acid
Query Length 176

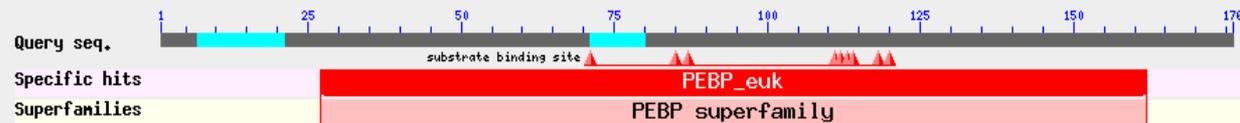
Database Name swissprot
Description Non-redundant SwissProt sequences
Program BLASTP 2.2.25+ [Citation](#)

Other reports: [Search Summary](#) [Taxonomy reports](#) [Distance tree of results](#) [Multiple alignment](#)

▼ Graphic Summary

▼ [Show Conserved Domains](#)

Putative conserved domains have been detected, click on the image below for detailed results.



磷脂酰乙醇胺结合蛋白

- 广泛存在于真核生物中，与系统发育分化紧密相关，在FT/TFL亚家族中主要同花发育相关。

[No Wikipedia article](#) [Pfam](#) [Interpro](#)

This tab holds the annotation information that is stored in the Pfam database. As we move to using Wikipedia as our main source of annotation, the contents of this tab will be gradually replaced by the Wikipedia tab.

Phosphatidylethanolamine-binding protein [Add annotation](#)

No Pfam abstract.

Literature references

- Banfield MJ, Barker JJ, Perry AC, Brady RL; , Structure 1998;6:1245-1254.: Function from structure? The crystal structure of human phosphatidylethanolamine-binding protein suggests a role in membrane signal transduction. [PUBMED:9782050](#)
- Serre L, Vallee B, Bureaud N, Schoentgen F, Zelwer C; , Structure 1998;6:1255-1265.: Crystal structure of the phosphatidylethanolamine-binding protein from bovine brain: a novel structural class of phospholipid-binding proteins. [PUBMED:9782057](#)

External database links

HOMSTRAD:	PBP
PANDIT:	PF01161
PROSITE:	PDOC00938
Pseudofam:	PF01161
SCOP:	1beh
SYSTERS:	PBP

Example structure

[PDB entry 3N08](#): Crystal Structure of a Putative Phosphatidylethanolamine-Binding Protein (PEBP) Homolog CT736 from Chlamydia trachomatis D/UW-3/CX

View a different structure:

3N08

Pfam domains

This image shows the arrangement of the Pfam domains that we found on this sequence. below gives the domain boundaries for each of the domains.



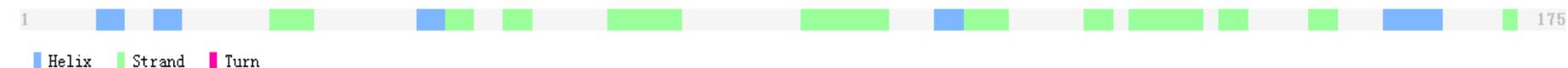
Source	Domain	Start	End
Pfam A	PBP	25	163

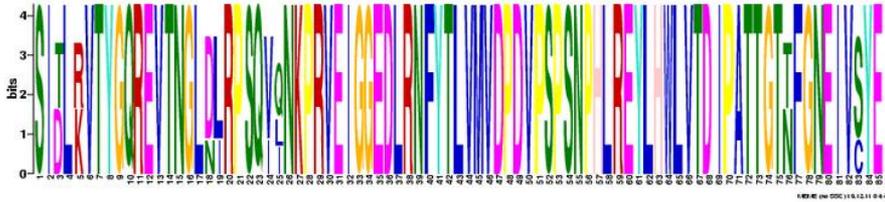
相关突变体注释信息

Sequence annotation (Features)

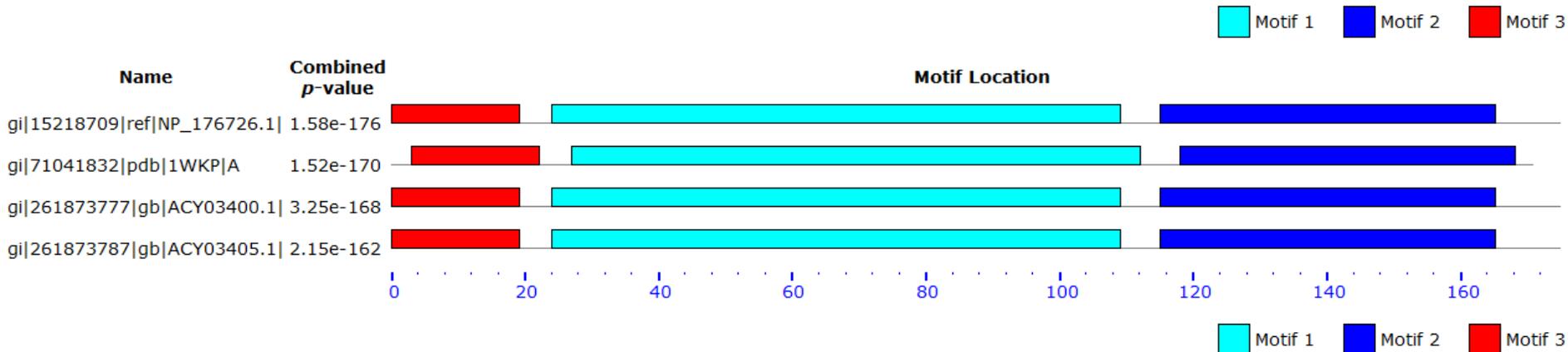
Feature key	Position(s)	Length	Description	Graphical view	Feature identifier
Molecule processing					
<input type="checkbox"/> Chain	1 - 175	175	Protein FLOWERING LOCUS T		PRO_0000204762
Natural variations					
<input type="checkbox"/> Alternative sequence	102 - 122	21	GNEIV...HRVVF → AWQANSVCTRVAPELQHSRV C in isoform Short .		VSP_004543
<input type="checkbox"/> Alternative sequence	123 - 175	53	Missing in isoform Short .		VSP_004544
Experimental info					
<input type="checkbox"/> Mutagenesis	84	1	E → K in ft-4; late-flowering.		
<input type="checkbox"/> Mutagenesis	85	1	Y → H: Inhibition of terminal flower formation, but weak effect on flowering time. Ref. 8		
<input type="checkbox"/> Mutagenesis	94	1	P → L in ft-6; late-flowering.		
<input type="checkbox"/> Mutagenesis	110	1	N → M: No effect on terminal flower formation. Ref. 8		
<input type="checkbox"/> Mutagenesis	119	1	R → H in ft-3; late-flowering.		
<input type="checkbox"/> Mutagenesis	120	1	V → F: No effect on terminal flower formation. Ref. 8		
<input type="checkbox"/> Mutagenesis	171	1	G → E in ft-1; late-flowering.		

Secondary structure





搜索获得拟南芥和油菜4个FT蛋白并对其进行结构域分析预测。



GmFT2a和AtFT基因启动子预测

- 为了更好的研究FT基因在大豆以及拟南芥中的表达规律和调控模式，实验室计划分别克隆FT基因上游启动子、研究启动子特性、为阐明开花调控提供理论和实验基础。
- 在开展此项工作的同时，对上游序列进行启动子预测，提供生物信息学方面的信息支持。

BLAST results

Query: Glycine_max_FT_like_protein_(FT)_mRNA_complete_cds (888 letters)
 Target: Glycine max genome (1168 sequences, 973344380 total letters)
 Program: BLASTN 2.2.22+

6 regions identified
[revise this query](#)

[clear Gbrowse](#) *[this target organism only]*

[switch views](#) *[currently in Target View]*

HSP joining parameters: Max intron size [update](#)
 Allow query overlap

Define

Score E

Target View

[click feature to view in browser]

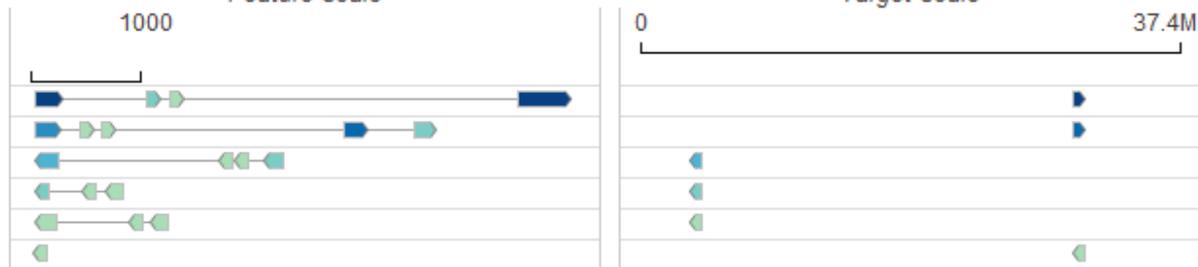


Gm16

894.0 0

Feature scale

Target scale

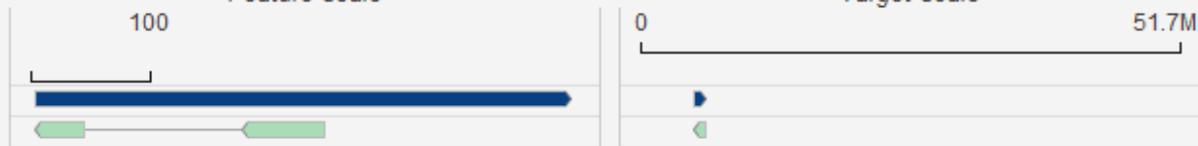


Gm02

533.3 2.4e-149

Feature scale

Target scale

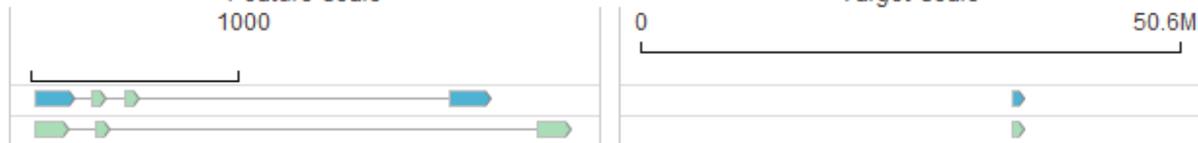


Gm19

237.5 2.6e-60

Feature scale

Target scale



Define

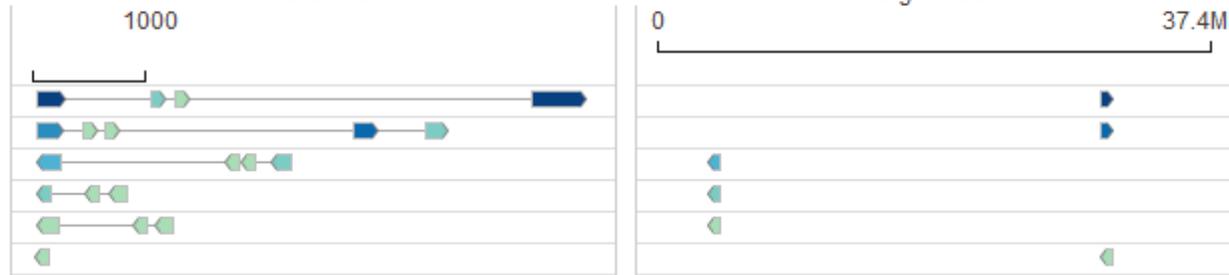
Score E

Target View

[click feature to view in browser]

Feature scale

Target scale



open all : close all

[-] Feature #1

HSP#1: Score: 894.0 bits (990.0) E-value: 0 Identity: 99.4% (502/505) Positive: 99.4% (502/505) Frame: +1/+1

```

Query      384  GGCCATGAGGTTGTAACATATGAAAGTCCAAGACCAATGATGGGGATTCAICGTTTGGTGTGGTGTATTTCGICAACTGGGTAGGGAGACCGTGTATG 483
          |||
Subject    30746100 GGCCATGAGGTTGTAACATATGAAAGTCCAAGACCAATGATGGGGATTCAICGTTTGGTGTGGTGTATTTCGICAACTGGGTAGGGAGACCGTGTATG 30746199

Query      484  CACCAGGATGGCGCCAGAATTTCAACACTAAAGAATTTGCTGAACTTTACAACCTTGGATTGCCAGTTGCTGCTGTCTATTTCAACATTAGAGGGAATC 583
          |||
Subject    30746200 CACCAGGATGGCGCCAGAATTTCAACACTAAAGAATTTGCTGAACTTTACAACCTTGGATTGCCAGTTGCTGCTGTCTATTTCAACATTAGAGGGAATC 30746299

Query      584  TGGTTCGGTGGGAAGGAGGTTATACTAAAAAAGTACTTTATATTATTGAAAAAATAAAGTAGTATAAGCTTCGTTGAGGGTTTCAGAAATATTAATT 683
          |||
Subject    30746300 TGGTTCGGTGGGAAGGAGGTTATACTAAGAAAAAGTACTTTATATTATTG-AAAAAATAAAGTAGTATAAGCTTCGTTGAGGGTTTCAGAAATATTAATT 30746398

Query      684  GGCAATCTCCACACTCTTTAGTAGTAAATGAGTGTTTTTCAACTTAATTAAGTAACTGAGCATAACAGTAAATAAATGCTAGCTCAGTTGGTAGCAGCAAG 783
          |||
Subject    30746399 GGCAATCTCCACACTCTTTAGTAGTAAATGAGTGTTTTTCAACTTAATTAAGTAACTGAGCATAACAGTAAATAAATGCTAGCTCAGTTGGTAGCAGCAAG 30746498

Query      784  TACTCTGCATATACATAAAATGAAACTGAAGCATCTAGGTTCAATTTTCTATTTGTATTATCAGTTGAAGAATGTTAAAGATATCTGATATACGTAAA 883
          |||
Subject    30746499 TACTCTGCATATACATAAAATGAAACTGAAGCATCTAGGTTCAATTTTCTATTTGTATTATCAGTTGAAGAATGTTAAAGATATCTGATATACGTAAA 30746598

Query      884  TTGGA
          |||
Subject    30746599 GTGGA
  
```




GmFT

Berkeley Drosophila Genome Project

Home

About BDGP

Searches

Neural Network Promoter Prediction

Promoter predictions for 1 eukaryotic sequence with score cutoff 0.80

Promoter predictions for Gm16:30743600..30746100 :

Start	End	Score	Promoter Sequence
87	137	0.89	TTTATCAAAATATAAAATGGGTCCAAAATATTTTAAAATAATTTTTATT
811	861	0.83	TATATATATATATATATATATATATATACAGTTAGTTATACATGTGTATA
820	870	0.99	ATATATATATATATATATACAGTTAGTTATACATGTGTATAAATACACAC
847	897	0.89	TATACATGTGTATAAATACACACAAACATACATTTATTTTATATATACA
1582	1632	0.95	GAAGCAAATTTAAAATAAGGGGAGTTAAAAGAATAAGAATTAGATTTGTTT
1933	1983	0.92	CTATCTGCTTTTTAAAACCAGAGCTTACCAGTATGTATGAGTGTTCTGG
2331	2381	0.89	ATTGCATGTATATAAATCATAAGAGTTTAAATGAGCGTCAAAGTAGATCAA



SoftBerry



```

>Gm16:30743600..30746100
Length of sequence-      2501
Thresholds for TATA+ promoters - 0.02, for TATA-/enhancers - 0.04
  3 promoter/enhancer(s) are predicted
Promoter Pos:    889 LDF- 0.15 TATA box at    853    21.89
Promoter Pos:    300 LDF- 0.04 TATA box at    287    17.83
Promoter Pos:   2370 LDF- 0.02 TATA box at   2335    20.90
Transcription factor binding sites/RegSite DB:

```



AtFT

Berkeley Drosophila Genome Project

Home

About BDGP

Searches

Neural Network Promoter Prediction

Promoter predictions for 1 eukaryotic sequence with score cutoff 0.80

Promoter predictions for seq0 :

Start	End	Score	Promoter Sequence
429	479	0.97	TAGTTAAATCTTAAAAAGTTGCTGAAAATAGAAAACAAAC T ACTTATCTT
578	628	0.85	AAAAATGATATACACAAGTGGCGGACAATCCATCTATCTC A TTTTAGCGA
1221	1271	0.90	TAGTTACATCTATATATGTTGATGCAATGTCAAAAAAGAAA A TCTCTCAAA
1509	1559	0.97	GCCGCTTGTTTATAAAAAAGAAGAGAAATAAAACAATTG A TTTGGTTTA
1684	1734	0.89	TTGTCGACCATATAACACAAGCGGCTAGAAAAATAGGTGA C TATTCTCAA
1958	2008	0.92	TTTTTCTCTATAAACTTGGCGGTACCTACTTTTTTCT T ATTTTTATT
2377	2427	0.90	CCGAAACAGTATAAAATATGTGTAGAGGGTTCATGCCTATG A TACAAATTA



SoftBerry



```

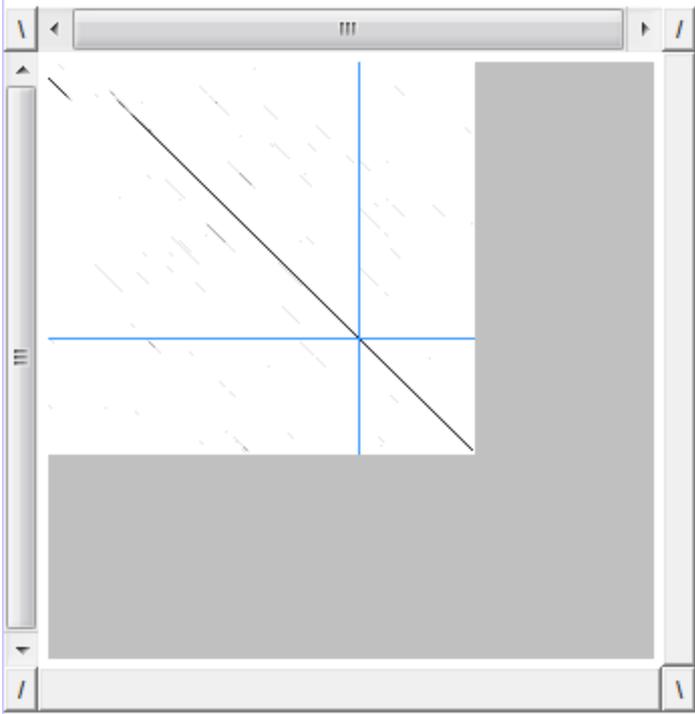
> test sequence
Length of sequence-      2501
Thresholds for TATA+ promoters - 0.02, for TATA-/enhancers - 0.04
  4 promoter/enhancer(s) are predicted
Promoter Pos:   2419 LDF- 0.06 TATA box at   2382   20.87
Promoter Pos:   1997 LDF- 0.05 TATA box at   1964   20.04
Promoter Pos:   1334 LDF- 0.03 TATA box at   1301   21.69
Promoter Pos:    609 LDF- 0.03 TATA box at    582   17.57
Transcription factor binding sites/RegSite DB:
for promoter at position -   2419

```

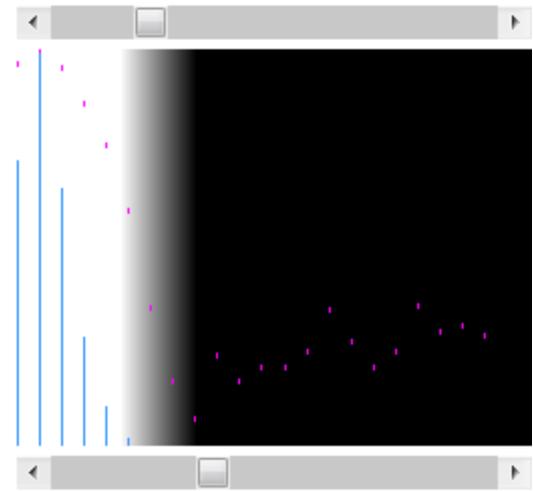


- 综合BDGP和Soft berry两个软件对*GmFT*及*AtFT* 基因的上游启动子预测结果认为上游2.5Kb中2.3~2.5Kb位置的保守结构为启动子核心序列。
- 实验室根据以上信息已开展*GmFT2a*基因的启动子克隆工作。涉及克隆片段大小为2.0Kb。

print input atff gmft Identity 23 1:1 compute



horizontal: atff
 vertical: gmft
 matrix: Identity
 sliding window: 23
 zoom: 1:1
 score range: 0 to 23
 gray scale: 35% - 20%



atff | 166

PSQVQNKPRVEIGGEDLRNFYTLVMVDPDVPSPSNPHLREYLHWLVTDIPATIGTTPGNEIVCYENPSFTAGIHRVVFILFRQLGRQTVYAPGWRQNFNTREFAEIYNLGLPVAAVFYNCQRESGC
 PSQVWQPRVNIIGGDDLRFNYTLIAVDPDAPSPSDENLREYLHWLVTDIPATIGASFGHEVVTYESPRFMMGIHRLVFLFRQLGRETVYAPGWRQNFNTKEFAELYNLGLPVAAVFYNIQRESGS

gmft | 151

- AtFT与GmFT蛋白Dotlet结果显示AtFT蛋白N端含有一段GmFT蛋白不包含的氨基酸酸序列，为下一步研究FT在不同植物作用提供了研究线索。
- 实验室计划通过同义突变的方法研究FT基因mRNA和蛋白运输机制。因此，需要在对其核苷酸序列进行突变时进行密码子偏好性分析。

AtFT cds

#1st letter GC 40.97%

#2nd letter GC 39.58%

#3rd letter GC 35.42%

#Codon AA Fraction Frequency Number

GCA	A	0.500	13.889	4
GCC	A	0.125	3.472	1
GCG	A	0.000	0.000	0
GCT	A	0.375	10.417	3
TGC	C	0.500	17.361	5
TGT	C	0.500	17.361	5
GAC	D	0.273	10.417	3
GAT	D	0.727	27.778	8
GAA	E	0.467	24.306	7
GAG	E	0.533	27.778	8
TTC	F	0.333	13.889	4
TTT	F	0.667	27.778	8

GGA	G	0.400	20.833	6	ATG	M	1.000	17.361	5	TCG	S	0.000	0.000	0
GGC	G	0.400	20.833	6	AAC	N	0.350	24.306	7	TCT	S	0.188	10.417	3
GGG	G	0.133	6.944	2	AAT	N	0.650	45.139	13	ACA	T	0.400	27.778	8
GGT	G	0.067	3.472	1	CCA	P	0.438	24.306	7	ACC	T	0.100	6.944	2
CAC	H	0.250	3.472	1	CCC	P	0.125	6.944	2	ACG	T	0.050	3.472	1
CAT	H	0.750	10.417	3	CCG	P	0.062	3.472	1	ACT	T	0.450	31.250	9
ATA	I	0.368	24.306	7	CCT	P	0.375	20.833	6	GTA	V	0.045	3.472	1
ATC	I	0.263	17.361	5	CAA	Q	0.429	10.417	3	GTC	V	0.045	3.472	1
ATT	I	0.368	24.306	7	CAG	Q	0.571	13.889	4	GTG	V	0.455	34.722	10
AAA	K	0.750	41.667	12	AGA	R	0.429	31.250	9	GTT	V	0.455	34.722	10
AAG	K	0.250	13.889	4	AGG	R	0.286	20.833	6	TGG	W	1.000	6.944	2
CTA	L	0.182	13.889	4	CGA	R	0.095	6.944	2	TAC	Y	0.267	13.889	4
CTC	L	0.227	17.361	5	CGC	R	0.095	6.944	2	TAT	Y	0.733	38.194	11
CTG	L	0.000	0.000	0	CGG	R	0.048	3.472	1	TAA	*	0.250	10.417	3
CTT	L	0.227	17.361	5	CGT	R	0.048	3.472	1	TAG	*	0.250	10.417	3
TTA	L	0.182	13.889	4	AGC	S	0.125	6.944	2	TGA	*	0.500	20.833	6
TTG	L	0.182	13.889	4	AGT	S	0.438	24.306	7					
					TCA	S	0.188	10.417	3					
					TCC	S	0.062	3.472	1					

GmFT2a cds

#Coding GC 38.51%

#1st letter GC 36.49%

#2nd letter GC 33.78%

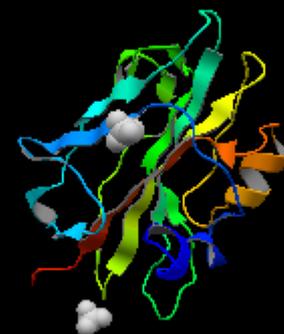
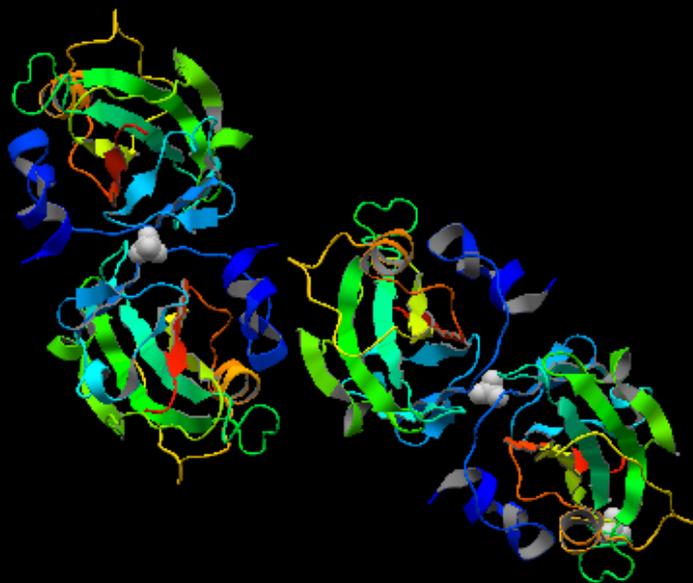
#3rd letter GC 45.27%

#Codon AA Fraction Frequency Number

GCA	A	0.500	13.514	4
GCC	A	0.375	10.135	3
GCG	A	0.000	0.000	0
GCT	A	0.125	3.378	1
TGC	C	0.400	6.757	2
TGT	C	0.600	10.135	3
GAC	D	0.200	3.378	1
GAT	D	0.800	13.514	4
GAA	E	0.250	6.757	2
GAG	E	0.750	20.270	6
TTC	F	0.400	13.514	4
TTT	F	0.600	20.270	6
GGA	G	0.235	13.514	4
GGC	G	0.059	3.378	1

GGG	G	0.529	30.405	9	CCA	P	0.429	10.135	3	ACC	T	0.263	16.892	5
GGT	G	0.176	10.135	3	CCC	P	0.000	0.000	0	ACG	T	0.053	3.378	1
CAC	H	0.455	16.892	5	CCG	P	0.286	6.757	2	ACT	T	0.316	20.270	6
CAT	H	0.545	20.270	6	CCT	P	0.286	6.757	2	GTA	V	0.105	6.757	2
ATA	I	0.333	27.027	8	CAA	Q	0.538	23.649	7	GTC	V	0.105	6.757	2
ATC	I	0.333	27.027	8	CAG	Q	0.462	20.270	6	GTG	V	0.421	27.027	8
ATT	I	0.333	27.027	8	AGA	R	0.571	13.514	4	GTT	V	0.368	23.649	7
AAA	K	0.550	37.162	11	AGG	R	0.143	3.378	1	TGG	W	1.000	16.892	5
AAG	K	0.450	30.405	9	CGA	R	0.000	0.000	0	TAC	Y	0.154	6.757	2
CTA	L	0.167	20.270	6	CGC	R	0.000	0.000	0	TAT	Y	0.846	37.162	11
CTC	L	0.028	3.378	1	CGG	R	0.143	3.378	1	TAA	*	0.409	30.405	9
CTG	L	0.167	20.270	6	CGT	R	0.143	3.378	1	TAG	*	0.182	13.514	4
CTT	L	0.139	16.892	5	AGC	S	0.174	13.514	4	TGA	*	0.409	30.405	9
TTA	L	0.056	6.757	2	AGT	S	0.087	6.757	2					
TTG	L	0.444	54.054	16	TCA	S	0.304	23.649	7					
ATG	M	1.000	30.405	9	TCC	S	0.130	10.135	3					
AAC	N	0.467	23.649	7	TCG	S	0.130	10.135	3					
AAT	N	0.533	27.027	8	TCT	S	0.174	13.514	4					
					ACA	T	0.368	23.649	7					

AtFT蛋白三维结构分析



[jV version 4](#)

Applet Launcher (new Java Plug-in) version

Style: Default Cartoon Wireframe CPK_without_water

Color: Default group chain atom

```
jV>  
5412 Atoms Selected.
```

```
jV>  
1305 Atoms Selected.
```

```
jV>
```

Display unit^{*1} (Number of chains^{*2})

Atom Type for Display and Color (Biological - Others)^{*3}

Setting for Applet Launcher Using^{*4}

Command area

Display language



ScanProsite Results Viewer

Q9SXZ2 FT_ARATH (175 aa)

RecName: Full=Protein FLOWERING LOCUS T; *Arabidopsis thaliana* (Mouse-ear cress)

```
MSINIRDFLIIVSRVVGDLDPFNRSITLKVTYGQREVTNGLDLRPSQVQNKPRVEIGGEDLRNFYT
LVMVDPDVPSPSNPHLREYLHWLVTDIPATTGTFPGNEIVCYENPSPTAGIHRVVVFILFRQLGRQT
VYAPGWRQNFNTREFAEIYNLGLFVAAVFYNCQRESGCGRRLL
```



hits by patterns: [1 hit (by 1 pattern) on 1 sequence]

Hits by PS01220 PBP Phosphatidylethanolamine-binding protein family signature :

Q9SXZ2
(FT_ARATH) (175 aa)

RecName: Full=Protein FLOWERING LOCUS T; *Arabidopsis thaliana* (Mouse-ear cress)

Hits on PDB 3D structures: [1WKP-A, 1WKP-B, 1WKP-C, 1WKP-D]

65 - 87: YtLVmVDPDvPSpsnphlreylH

Legend:

disulfide bridge active site other 'ranges' other sites

BBP, PS01220; Phosphatidylethanolamine-binding protein family signature (PATTERN)

Consensus pattern: [FYL]-x-[LVM]-[LIVF]-x-[TIVM]-[DC]-P-D-x-P-[SNG]-x(10)-H

Sequences known to belong to this class detected by the pattern: ALL

Other sequence(s) detected in Swiss-Prot: NONE.

Swiss-PdbViewer 4.0.4

File Edit Select Build Tools Fit Display Color Prefs SwissModel Wind Help

15Å 60.1° 60.8° LEU41 MUTATE FORSEION ?

Move All

Layers Info

Layer	taxon	sel	grp	vis	mov	axis	CA	O	H	Hbnd	Hdst	side	HOH	cyc	AlnWmdl	SelGrp	
a	Arabidopsis thaliana	v	1	v	v		v		v	v			v	v	v	*	162
b	Arabidopsis thaliana	v	1	v	v		v		v	v			v	v	v		162
c	Arabidopsis thaliana	v	1	v	v		v		v	v			v	v	v		162
d	Arabidopsis thaliana	v	1	v	v		v		v	v			v	v	v		162

b (940 x 506)

Control Panel

visible can move

group show side labl rbn col R

- B s LEU124 v
- B s PHE125 v
- B s ARG126 v
- B s GLN127 v
- B LEU128 v
- B GLY129 v
- B ARG130 v
- B GLN131 v
- B THR132 v
- B VAL133 v
- B TYR134 v
- B ALA135 v
- B PRO136 v
- B GLY137 v
- B TRP138 v
- B s ARG139 v
- B s GLN140 v
- B s ASN141 v
- B PHE142 v
- B ASN143 v
- B h THR144 v
- B h ARG145 v
- B h GLU146 v
- B h PHE147 v
- B h ALA148 v
- B h GLU149 v
- B h ILE150 v
- B TYR151 v
- B ASN152 v
- B LEU153 v
- B GLY154 v
- B LEU155 v
- B PRO156 v
- B VAL157 v
- B s ALA158 v
- B s ALA159 v
- B s VAL160 v
- B s PHE161 v
- B s TYR162 v

RMSD:

AB:0.22, CD:0.14; AC:0.49, BD:0.51

THANKS