葫芦科中促使两性花向单性花转化的WIP转录因子基因家族的分析

2025

汇报人：孙雯

汇报时间：2025.04.27

目录

01.

课题背景

蓖麻（Ricinus communis L.）是大戟科蓖麻属植物

课题内容：课题组研究团队正在进行**野生蓖麻（Ricinus communis L.）和栽培蓖麻品种的杂交育种项目**。

创新点：**通过野外观察，试验田中偶有发生非分枝表型，以此为契机，作为未来育种目标的重点领域。**

已完成工作：**通过分子标记辅助选择，已经成功地鉴定了多个与表型相关的遗传位点，**这些遗传位点控制着关键的农艺性状，包括开花季节、花序发育和种子大小。

多年生

一年生

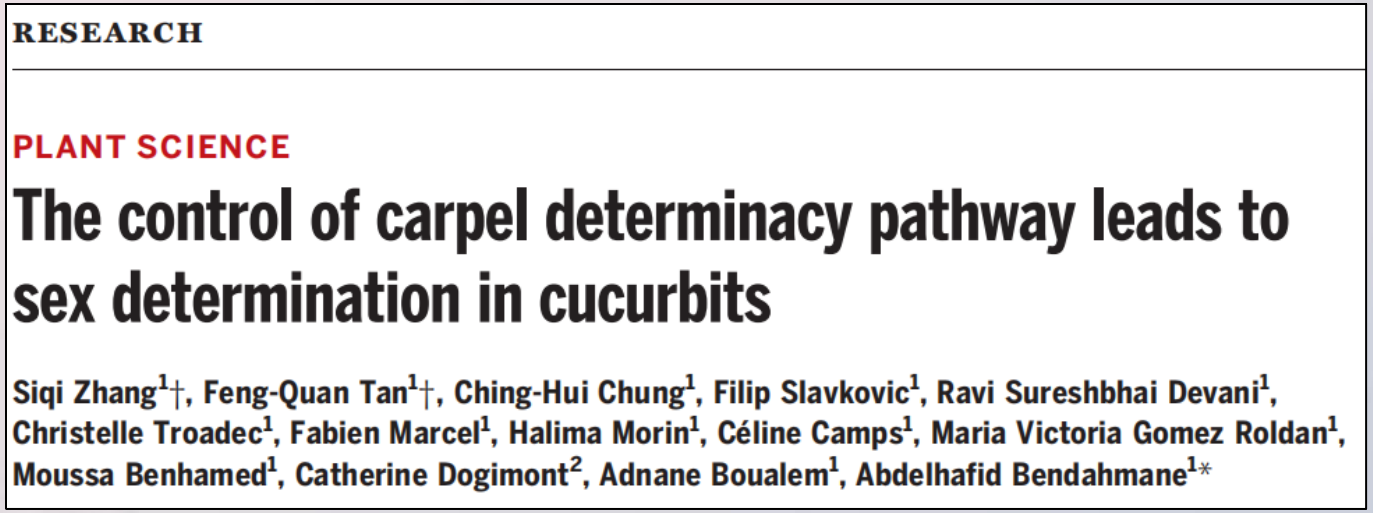
课题目的：种植一种不分枝的植物，以方便机械收获，

例如不分枝的向日葵

获得一年生蓖麻油植物将有利于其种植和谷物的机械收获

阻碍形成层活性 = 阻碍蓖麻分枝 = 阻碍分生组织通路基因表达

《Science》2022.11发表文章



RESEARCH

PLANT SCIENCE

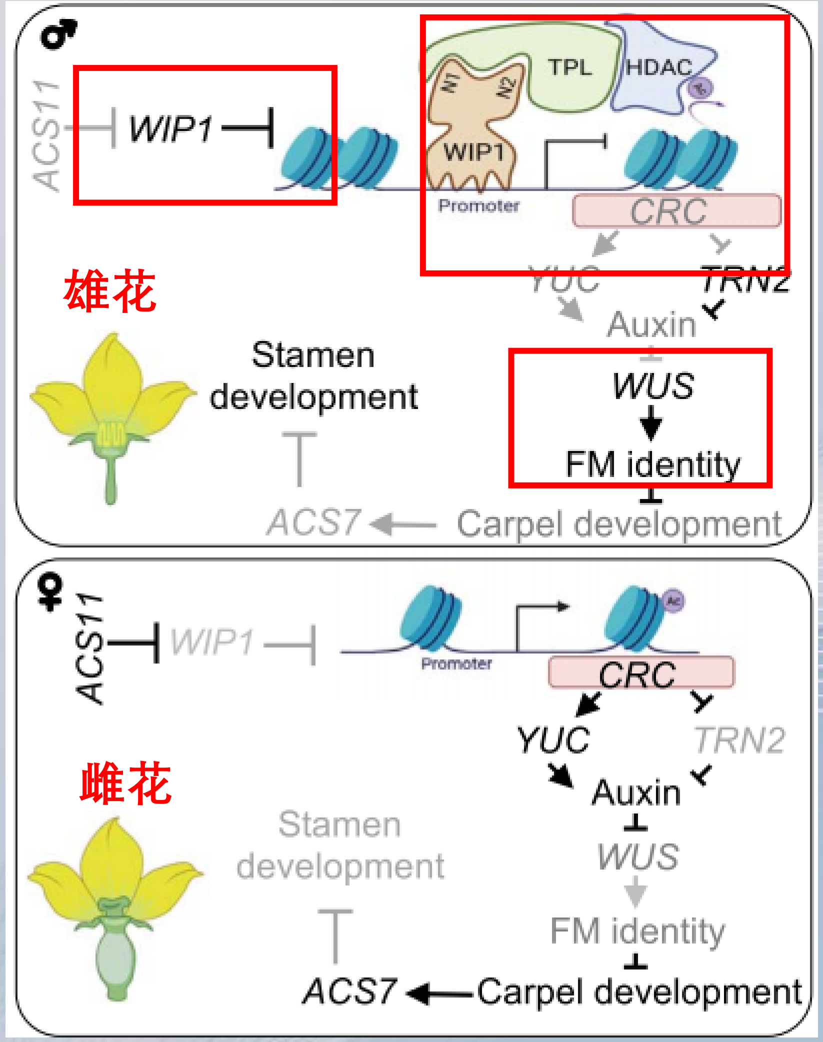
The control of carpel determinacy pathway leads to sex determination in cucurbits

Siqi Zhang[1]†, Feng-Quan Tan[1]†, Ching-Hui Chung[1], Filip Slavkovic[1], Ravi Sureshbhai Devani[1], Christelle Troadec[1], Fabien Marcel[1], Halima Morin[1], Céline Camps[1], Maria Victoria Gomez Roldan[1], Moussa Benhamed[1], Catherine Dogimont[2], Adnane Boualem[1], Abdelhafid Bendahmane[1]*

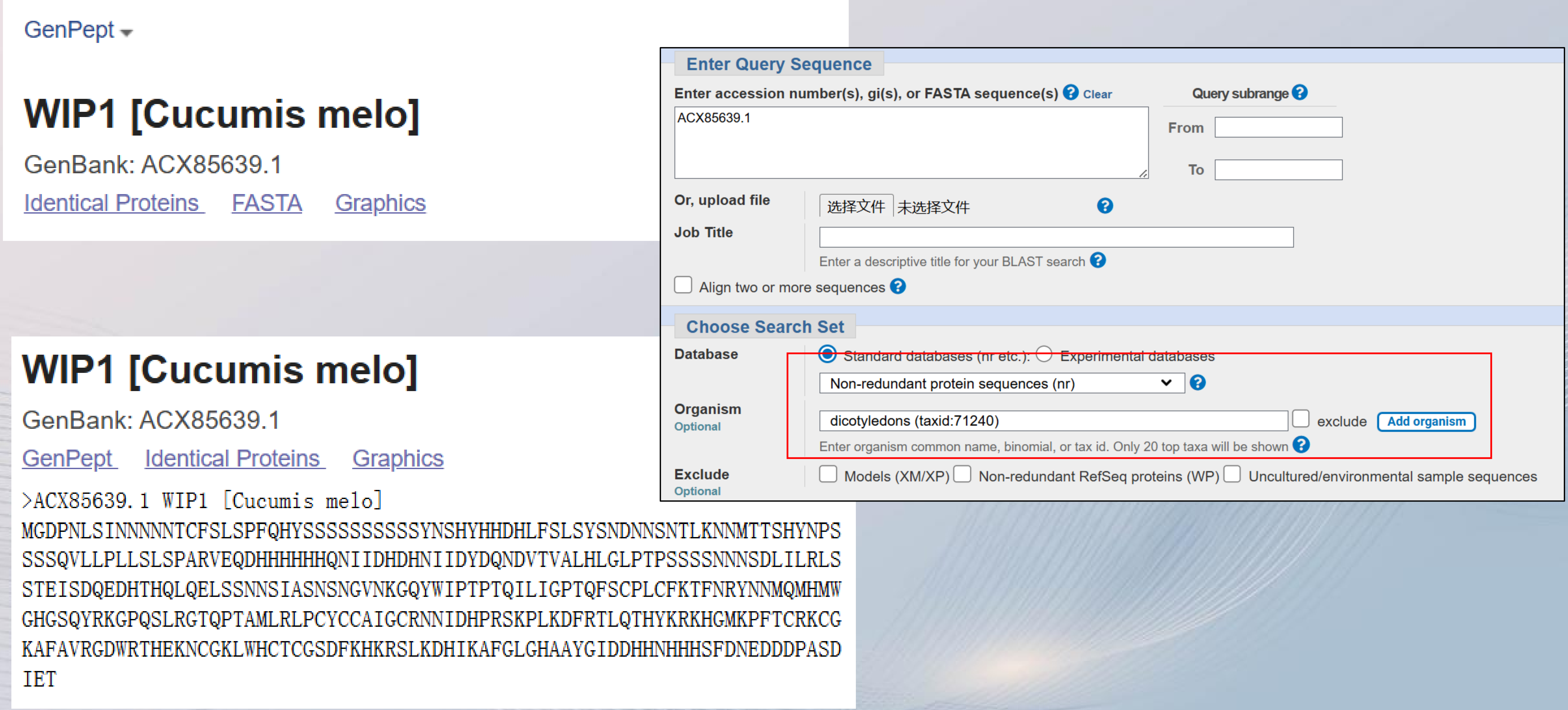
已知分生组织的通路基因表达与控制雌雄蕊转化的基因有关，

WIP1 ⊣ 雌蕊生长 → 分生组织（葫芦科植物）

现可能发现有关蓖麻中控制雌雄蕊转化的基因，基于此在进行基因功能验证之前做相关基因分析



WIP1基因家族



WIP1 [Cucumis melo]

GenBank: ACX85639.1

WIP1 [Cucumis melo]

GenBank: ACX85639.1

```
>ACX85639.1 WIP1 [Cucumis melo]
MGDPNLSINNNNNTCFSLSPFQHYSSSSSSSSSSYNSHYHHDHLFSLSYSNDNNSNTLKNNMTTSHYNPS
SSSQVLLPLLSLSPARVEQDHHHHHHQNIIDHDHNIIDYDQNDVTVALHLGLPTPSSSSNNNSDLILRLS
STEISDQEDHTHQLQELSSNNSIASNSNGVNKGQYWIPTPTQILIGPTQFSCPLCFKTFNRYNNMQMHMW
GHGSQYRKGPQSLRGTQPTAMLRLPCYCCAIGCRNNIDHPRSKPLKDFRTLQTHYKRKHGMKPFTCRKCG
KAFAVRGDWRTHEKNCGKLWHCTCGSDFKHKRSLKDHIKAFGLGHAAYGIDDHHNHHHSFDNEDDDPASD
IET
```

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s): ACX85639.1

Or, upload file: 选择文件 未选择文件

Job Title: Enter a descriptive title for your BLAST search

☐ Align two or more sequences

Choose Search Set

Database: ● Standard databases (nr etc.): ○ Experimental databases

Non-redundant protein sequences (nr)

Organism (Optional): dicotyledons (taxid:71240) ☐ exclude

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown

Exclude (Optional): ☐ Models (XM/XP) ☐ Non-redundant RefSeq proteins (WP) ☐ Uncultured/environmental sample sequences

Blast比对

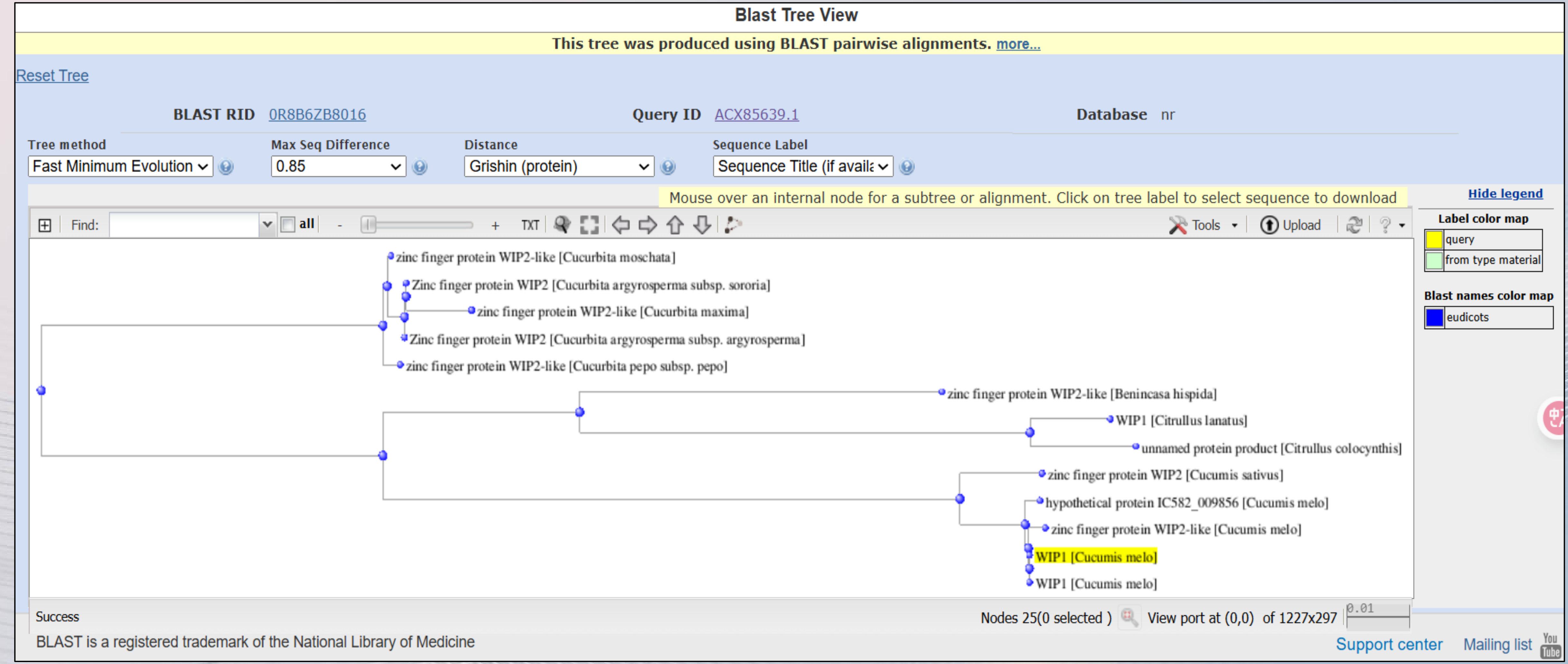
Sequences producing significant alignments

☑ select all *50 sequences selected*

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	WIP1 [Cucumis melo]	Cucumis melo	729	729	100%	0.0	100.00%	353	ACX85637.1
☑	hypothetical protein IC582_009856 [Cucumis melo]	Cucumis melo	728	728	100%	0.0	99.72%	353	KAL0550793.1
☑	zinc finger protein WIP2-like [Cucumis melo]	Cucumis melo	722	722	100%	0.0	99.43%	352	XP_050941016.1
☑	zinc finger protein WIP2 [Cucumis sativus]	Cucumis sativus	678	678	100%	0.0	94.97%	353	XP_004142226.2
☑	zinc finger protein WIP2-like [Benincasa hispida]	Benincasa hispida	450	450	100%	2e-156	74.67%	368	XP_038902545.1
☑	zinc finger protein WIP2-like [Cucurbita moschata]	Cucurbita moschata	436	436	88%	1e-151	72.81%	333	XP_022944308.1
☑	Zinc finger protein WIP2, partial [Cucurbita argyrosperma subsp. sororia]	Cucurbita argyrosperma subsp. sororia	437	437	88%	1e-151	72.81%	339	KAG6570388.1
☑	Zinc finger protein WIP2, partial [Cucurbita argyrosperma subsp. argyrosperma]	Cucurbita argyrosperma subsp. argyros…	437	437	88%	2e-151	72.81%	340	KAG7010263.1
☑	zinc finger protein WIP2-like [Cucurbita maxima]	Cucurbita maxima	435	435	92%	4e-151	71.34%	334	XP_022985746.1
☑	WIP1 [Citrullus lanatus]	Citrullus lanatus	436	436	97%	5e-151	73.10%	364	QEP54301.1
☑	zinc finger protein WIP2-like [Cucurbita pepo subsp. pepo]	Cucurbita pepo subsp. pepo	435	435	88%	8e-151	72.73%	335	XP_023513261.1
☑	unnamed protein product [Citrullus colocynthis]	Citrullus colocynthis	431	431	97%	6e-149	71.12%	370	CAK9314187.1
	hypothetical protein I3760_14G123100 [Carya illinoinensis]	Carya illinoinensis	419	419	95%	8e-145	63.25%	337	KAG2671206.1

通过限定双子叶植物比对，前12个序列均属于葫芦科植物，并且属于均具有锌指蛋白结构域，属于C2H2基因家族

12个葫芦科植物发育树分析



Blast Tree View

This tree was produced using BLAST pairwise alignments. more...

Reset Tree

BLAST RID 0R8B6ZB8016 **Query ID** ACX85639.1 **Database** nr

Tree method: Fast Minimum Evolution

Max Seq Difference: 0.85

Distance: Grishin (protein)

Sequence Label: Sequence Title (if availa

Mouse over an internal node for a subtree or alignment. Click on tree label to select sequence to download

Label color map

🟨 query

🟩 from type material

Blast names color map

🟦 eudicots

Success

Nodes 25(0 selected) View port at (0,0) of 1227x297

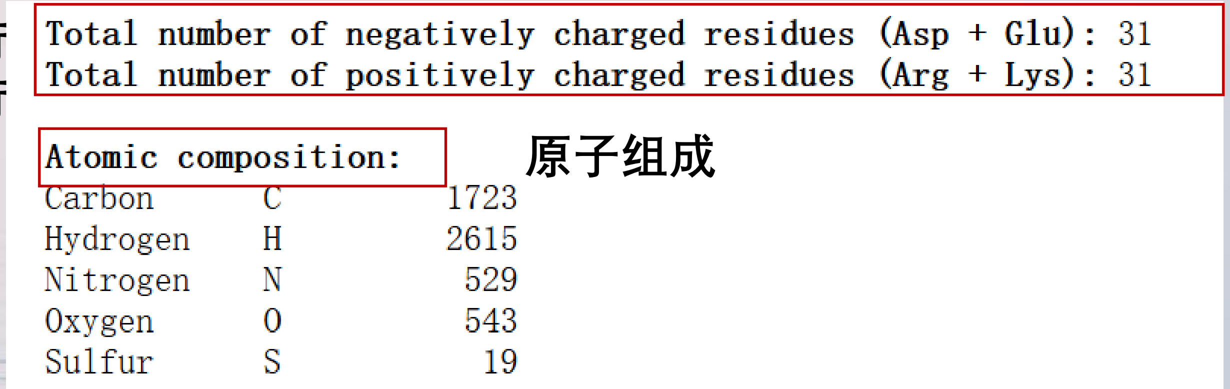
BLAST is a registered trademark of the National Library of Medicine

ProtParam

D0UIX1 · D0UIX1_CUCME

Protein	WIP1	**Amino acids**	353 (go to sequence)
Gene	WIP1	**Protein existence**	Inferred from homology
Status	UniProtKB unreviewed (TrEMBL)	**Annotation score**	1/5
Organism	Cucumis melo (Muskmelon)		

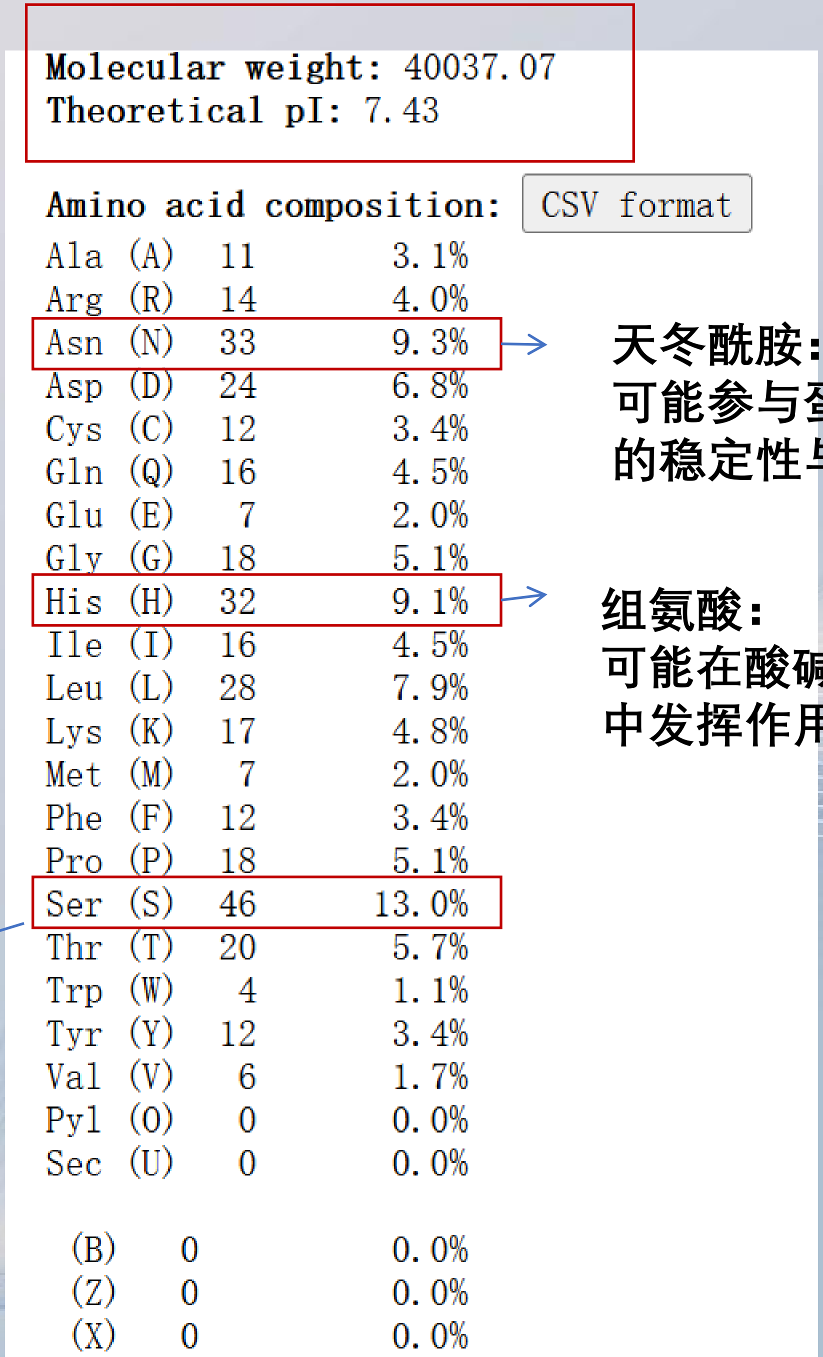
负电荷
正电荷



```
Total number of negatively charged residues (Asp + Glu): 31
Total number of positively charged residues (Arg + Lys): 31

Atomic composition:
Carbon       C         1723
Hydrogen     H         2615
Nitrogen     N          529
Oxygen       O          543
Sulfur       S           19
```

原子组成

分子量
等电点



```
Molecular weight: 40037.07
Theoretical pI: 7.43

Amino acid composition:
Ala (A)  11        3.1%
Arg (R)  14        4.0%
Asn (N)  33        9.3%
Asp (D)  24        6.8%
Cys (C)  12        3.4%
Gln (Q)  16        4.5%
Glu (E)   7        2.0%
Gly (G)  18        5.1%
His (H)  32        9.1%
Ile (I)  16        4.5%
Leu (L)  28        7.9%
Lys (K)  17        4.8%
Met (M)   7        2.0%
Phe (F)  12        3.4%
Pro (P)  18        5.1%
Ser (S)  46       13.0%
Thr (T)  20        5.7%
Trp (W)   4        1.1%
Tyr (Y)  12        3.4%
Val (V)   6        1.7%
Pyl (O)   0        0.0%
Sec (U)   0        0.0%

 (B)   0           0.0%
 (Z)   0           0.0%
 (X)   0           0.0%
```

天冬酰胺：
可能参与蛋白质
的稳定性与折叠

组氨酸：
可能在酸碱平衡
中发挥作用

丝氨酸：
可能与蛋白质亲水性
和磷酸化有关

02.

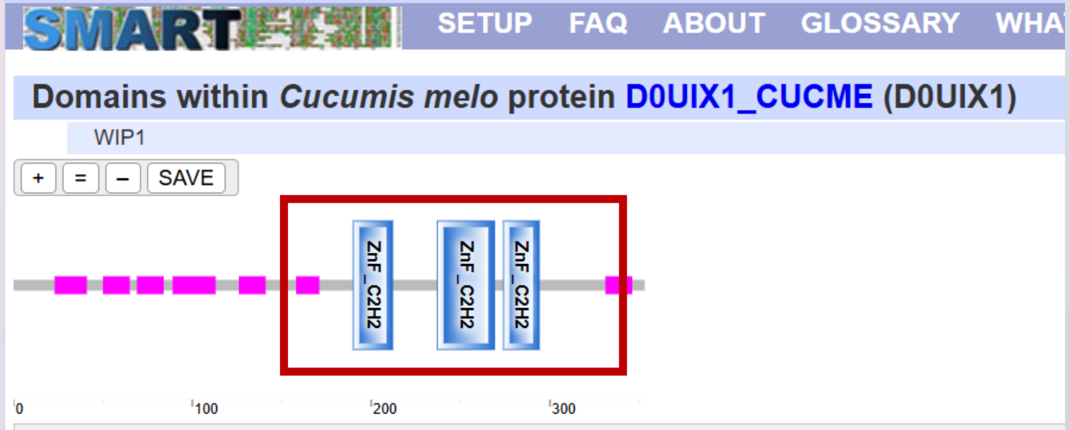
生信工具

NCBI	Needle
UniProt	Dotpath
ProtParam	STRING
SMART	Expasy
Interpro	MEGA7
TAIR	
EMBOSS	
TBtools	
plantTFDB	

03.

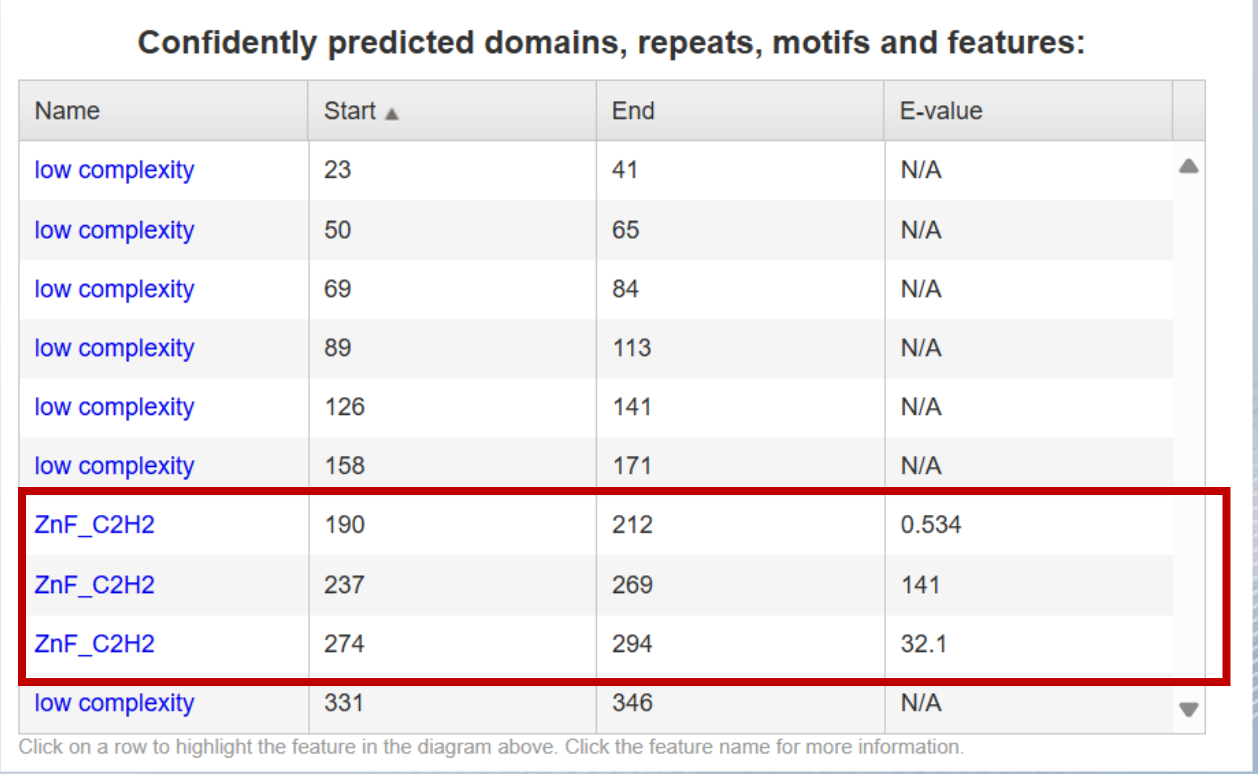
分析

结构域预测 SMART Interpro



Domains within *Cucumis melo* protein D0UIX1_CUCME (D0UIX1)

WIP1



Confidently predicted domains, repeats, motifs and features:

Name	Start ▲	End	E-value
low complexity	23	41	N/A
low complexity	50	65	N/A
low complexity	69	84	N/A
low complexity	89	113	N/A
low complexity	126	141	N/A
low complexity	158	171	N/A
ZnF_C2H2	190	212	0.534
ZnF_C2H2	237	269	141
ZnF_C2H2	274	294	32.1
low complexity	331	346	N/A

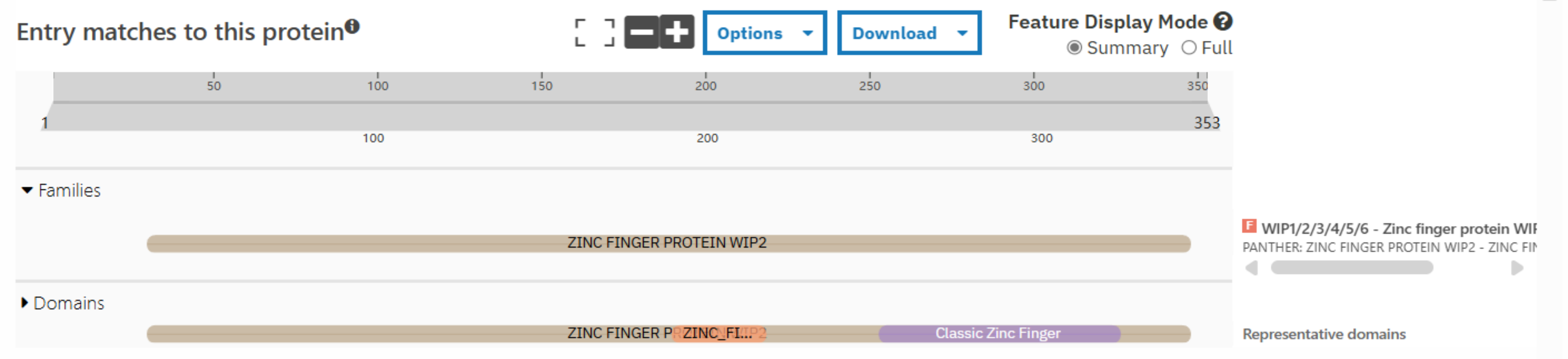
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.

190---212

237---269

274---294

锌指蛋白结构域



Entry matches to this protein

Feature Display Mode

● Summary ○ Full

▾ Families

▸ Domains

WIP1/2/3/4/5/6 - Zinc finger protein WI

PANTHER: ZINC FINGER PROTEIN WIP2 - ZINC FI

Representative domains

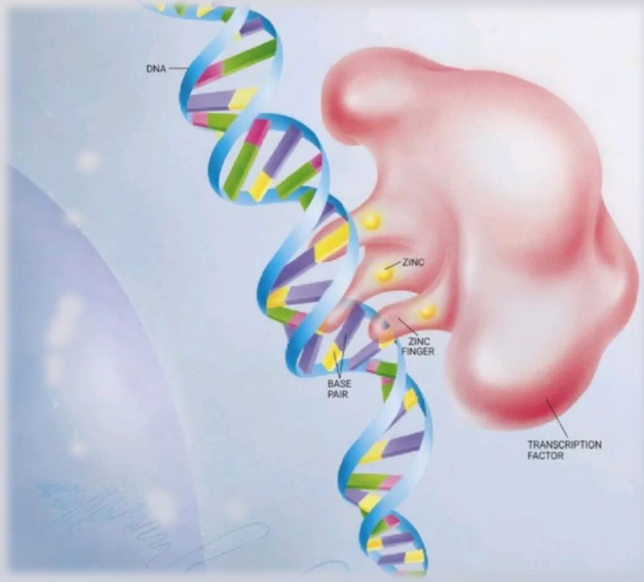
C2H2锌指蛋白基因家族

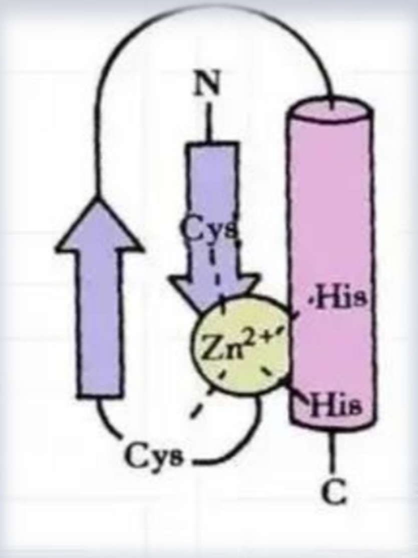
C2H2 家族简介:

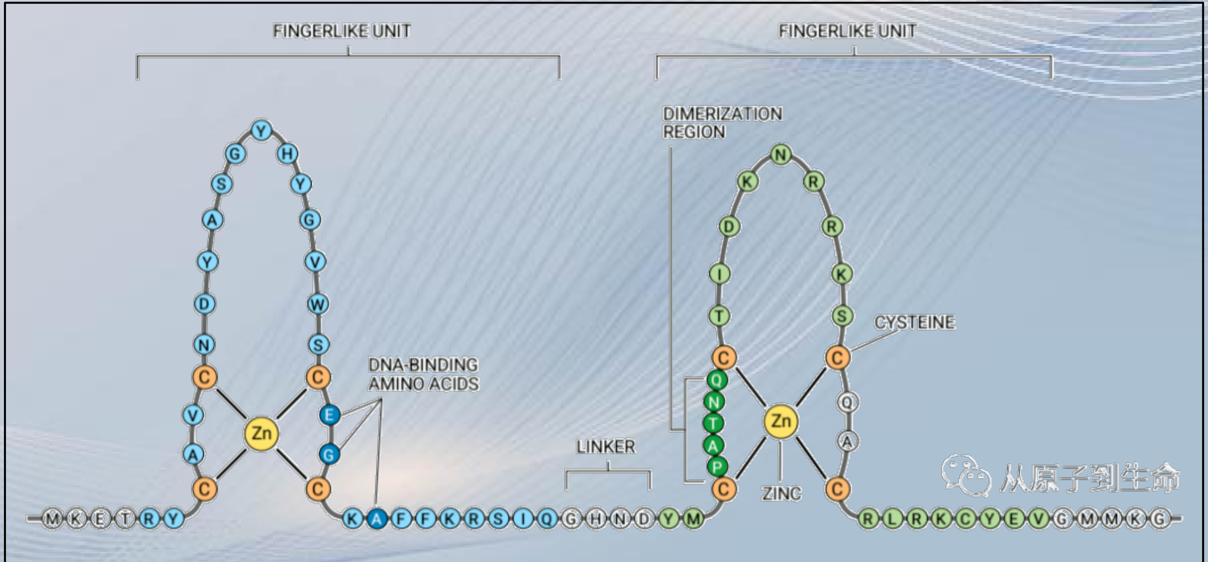
C2H2 型锌指蛋白是最明确的锌指转录因子，在人类、动物和植物中广泛存在。

已知的植物 C2H2 锌指蛋白主要参与植物的生长发育和对环境胁迫的响应。

迄今为止，已在包括天竺葵、拟南芥、小麦和水稻在内的植物中报道了 50 多种 C2H2 锌指蛋白，**其中大多数在锌指结构域中具有植物特有的 QALGGH 基序**。

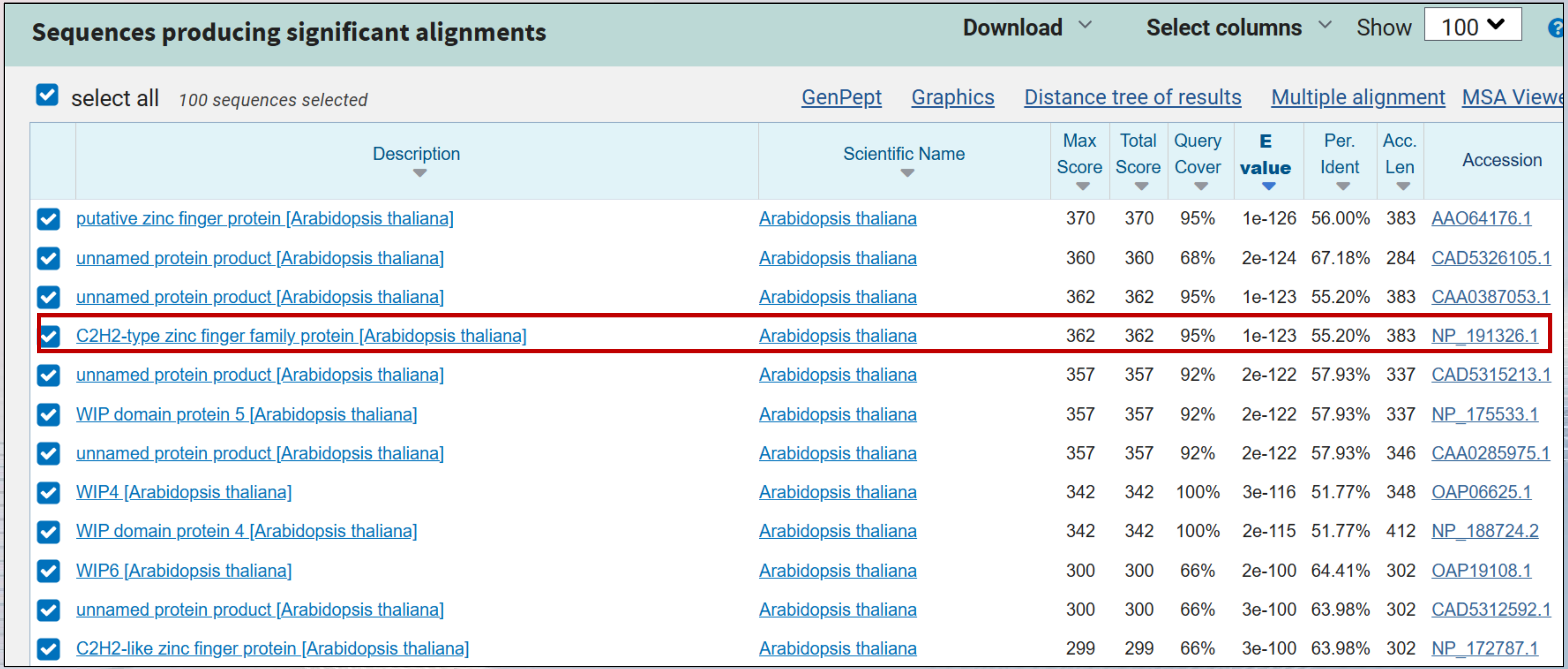






雌激素受体DNA结合域的三维视图

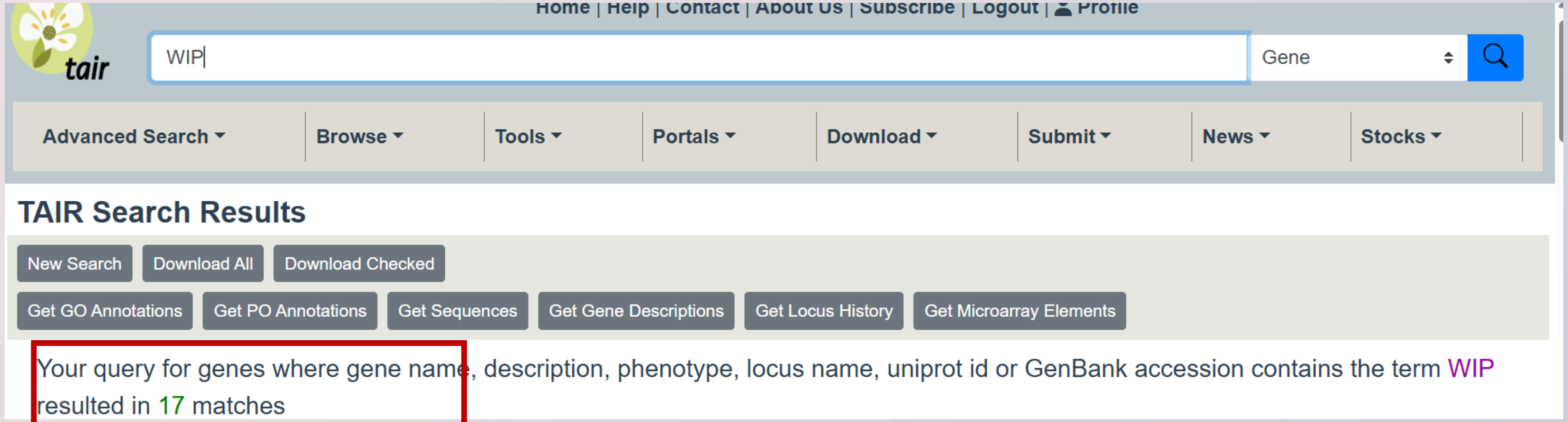
NCBI　Blast Protein序列比对



	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	putative zinc finger protein [Arabidopsis thaliana]	Arabidopsis thaliana	370	370	95%	1e-126	56.00%	383	AAO64176.1
☑	unnamed protein product [Arabidopsis thaliana]	Arabidopsis thaliana	360	360	68%	2e-124	67.18%	284	CAD5326105.1
☑	unnamed protein product [Arabidopsis thaliana]	Arabidopsis thaliana	362	362	95%	1e-123	55.20%	383	CAA0387053.1
☑	C2H2-type zinc finger family protein [Arabidopsis thaliana]	Arabidopsis thaliana	362	362	95%	1e-123	55.20%	383	NP_191326.1
☑	unnamed protein product [Arabidopsis thaliana]	Arabidopsis thaliana	357	357	92%	2e-122	57.93%	337	CAD5315213.1
☑	WIP domain protein 5 [Arabidopsis thaliana]	Arabidopsis thaliana	357	357	92%	2e-122	57.93%	337	NP_175533.1
☑	unnamed protein product [Arabidopsis thaliana]	Arabidopsis thaliana	357	357	92%	2e-122	57.93%	346	CAA0285975.1
☑	WIP4 [Arabidopsis thaliana]	Arabidopsis thaliana	342	342	100%	3e-116	51.77%	348	OAP06625.1
☑	WIP domain protein 4 [Arabidopsis thaliana]	Arabidopsis thaliana	342	342	100%	2e-115	51.77%	412	NP_188724.2
☑	WIP6 [Arabidopsis thaliana]	Arabidopsis thaliana	300	300	66%	2e-100	64.41%	302	OAP19108.1
☑	unnamed protein product [Arabidopsis thaliana]	Arabidopsis thaliana	300	300	66%	3e-100	63.98%	302	CAD5312592.1
☑	C2H2-like zinc finger protein [Arabidopsis thaliana]	Arabidopsis thaliana	299	299	66%	3e-100	63.98%	302	NP_172787.1

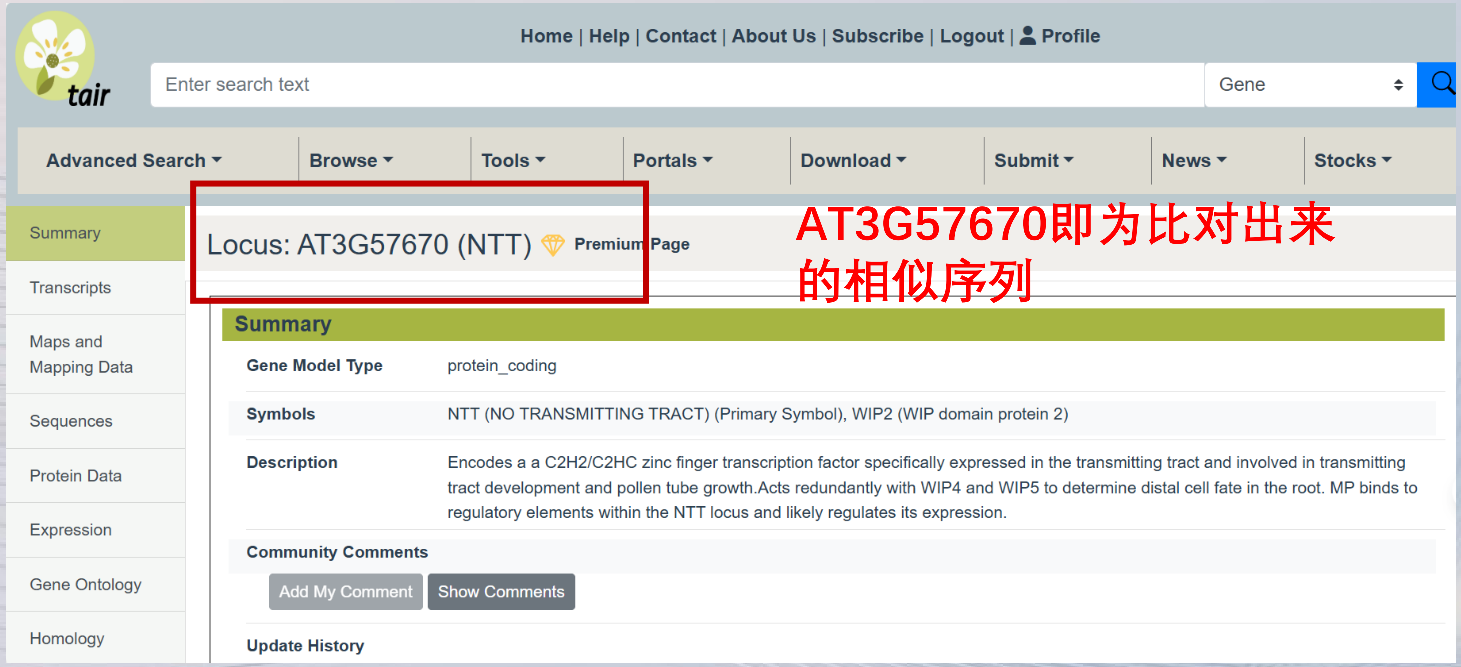
通过输入葫芦科植物甜瓜的蛋白序列进行Blast序列比对，限定拟南芥物种，发现在拟南芥C2H2锌指蛋白家族存在相似性高的序列

拟南芥数据库TAIR



TAIR Search Results

Your query for genes where gene name, description, phenotype, locus name, uniprot id or GenBank accession contains the term WIP resulted in 17 matches



Locus: AT3G57670 (NTT)

AT3G57670即为比对出来的相似序列

Summary

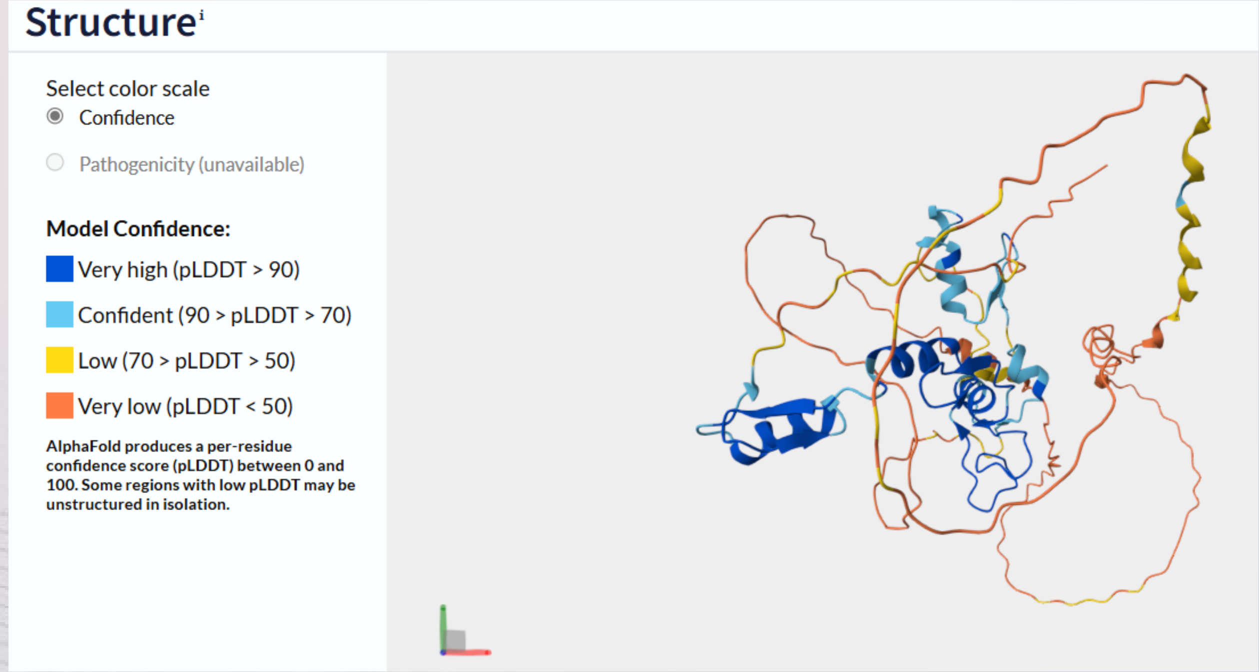
Gene Model Type	protein_coding
Symbols	NTT (NO TRANSMITTING TRACT) (Primary Symbol), WIP2 (WIP domain protein 2)
Description	Encodes a a C2H2/C2HC zinc finger transcription factor specifically expressed in the transmitting tract and involved in transmitting tract development and pollen tube growth.Acts redundantly with WIP4 and WIP5 to determine distal cell fate in the root. MP binds to regulatory elements within the NTT locus and likely regulates its expression.

Community Comments

Update History

使用TAIR数据库查找WIP，得到有关拟南芥中17个WIP转录因子具有C2H2锌指结构域，其中有的转录因子具有两个或多个剪接体，下载序列后得到19个结果

UniProt



Structure

Select color scale

● Confidence

○ Pathogenicity (unavailable)

Model Confidence:

🟦 Very high (pLDDT > 90)

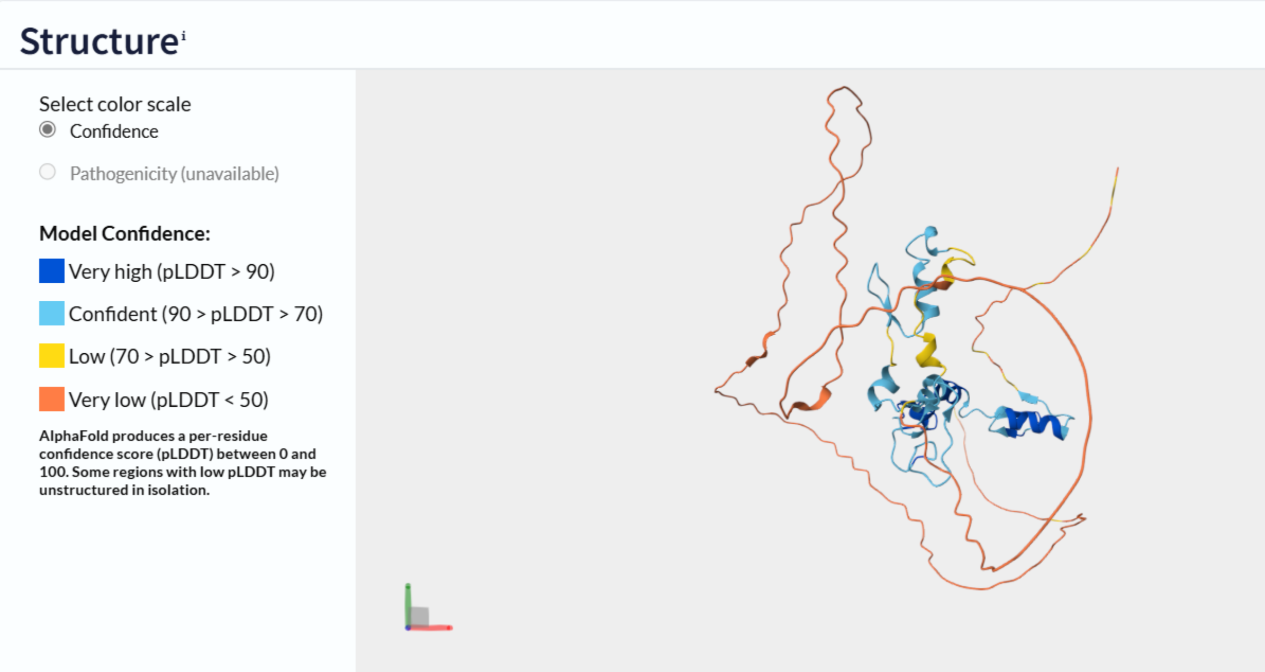
🟦 Confident (90 > pLDDT > 70)

🟨 Low (70 > pLDDT > 50)

🟧 Very low (pLDDT < 50)

AlphaFold produces a per-residue confidence score (pLDDT) between 0 and 100. Some regions with low pLDDT may be unstructured in isolation.

拟南芥WIP2/NTT/AT3G57670.1蛋白结构



Structure

Select color scale

● Confidence

○ Pathogenicity (unavailable)

Model Confidence:

🟦 Very high (pLDDT > 90)

🟦 Confident (90 > pLDDT > 70)

🟨 Low (70 > pLDDT > 50)

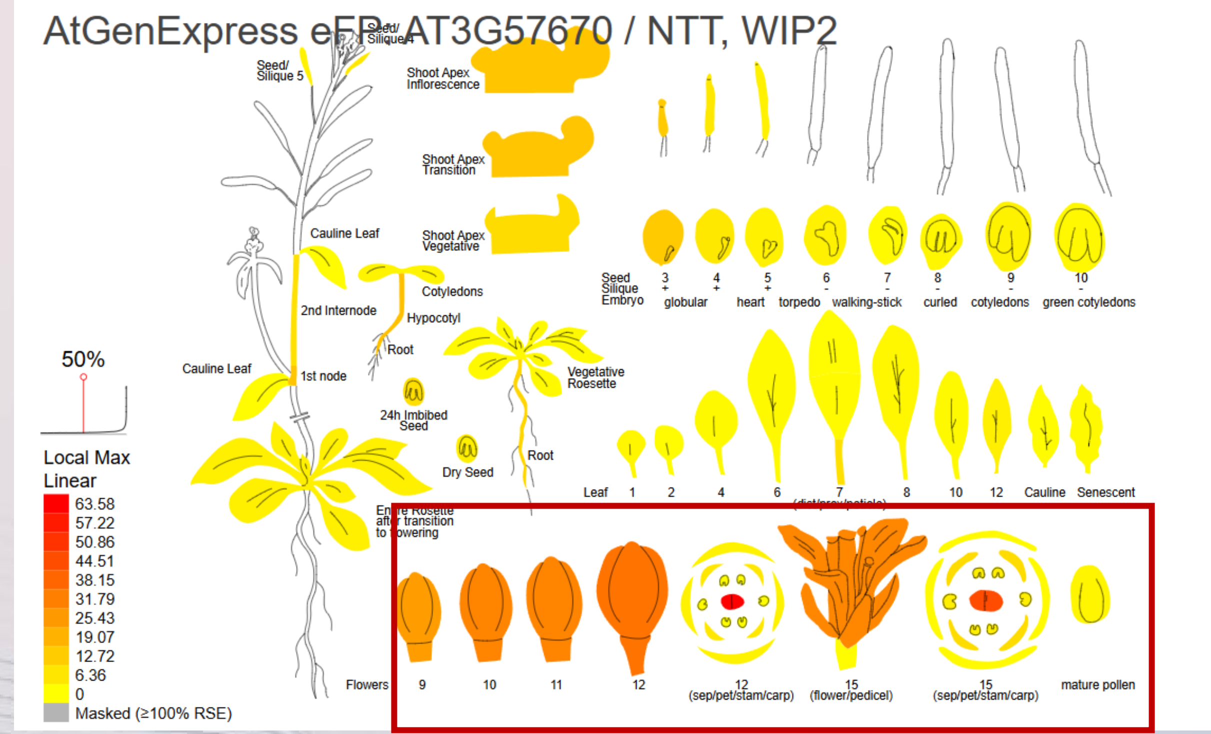
🟧 Very low (pLDDT < 50)

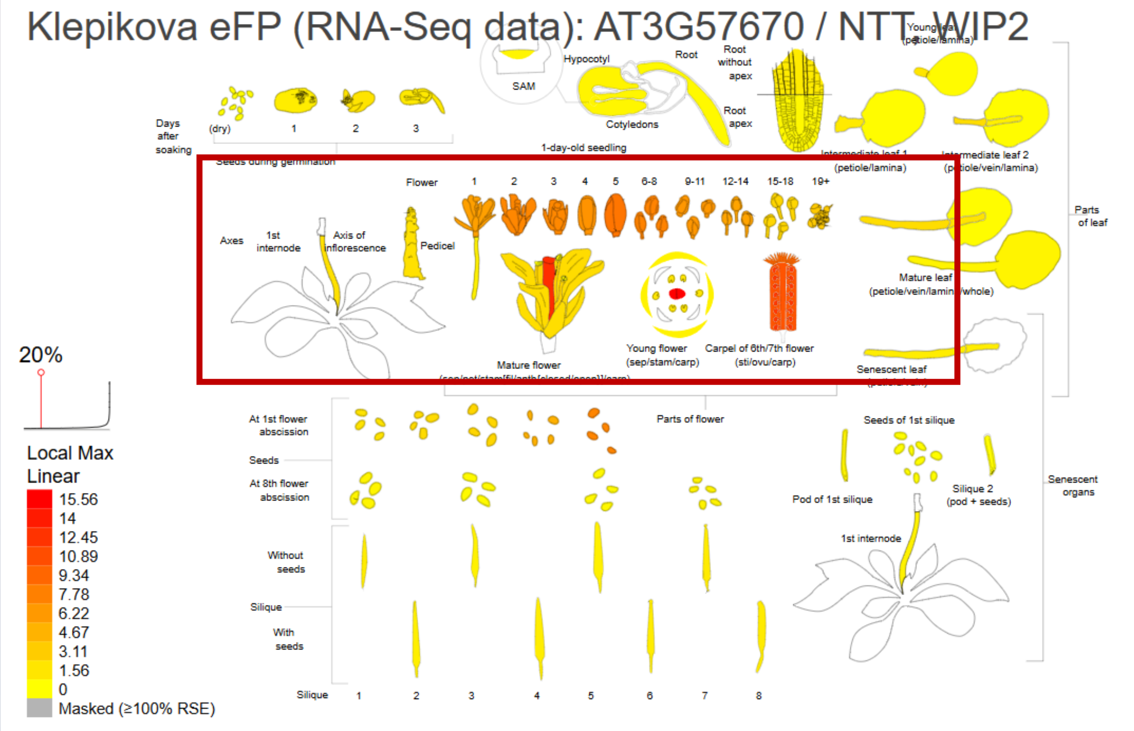
AlphaFold produces a per-residue confidence score (pLDDT) between 0 and 100. Some regions with low pLDDT may be unstructured in isolation.

甜瓜WIP1蛋白结构

通过模型对某蛋白质结构进行预测的可视化结果，提供了蛋白质不同区域的三维结构及其预测置信度

TAIR 拟南芥WIP2/NTT表达模式图





花序分生组织：表达量较高（橙色）

花器官：在不同花器官（如花萼、花瓣、雄蕊、雌蕊）中表达量不同，尤其在雄蕊和雌蕊中表达量较高

花粉：在成熟花粉中表达量较高（橙色）

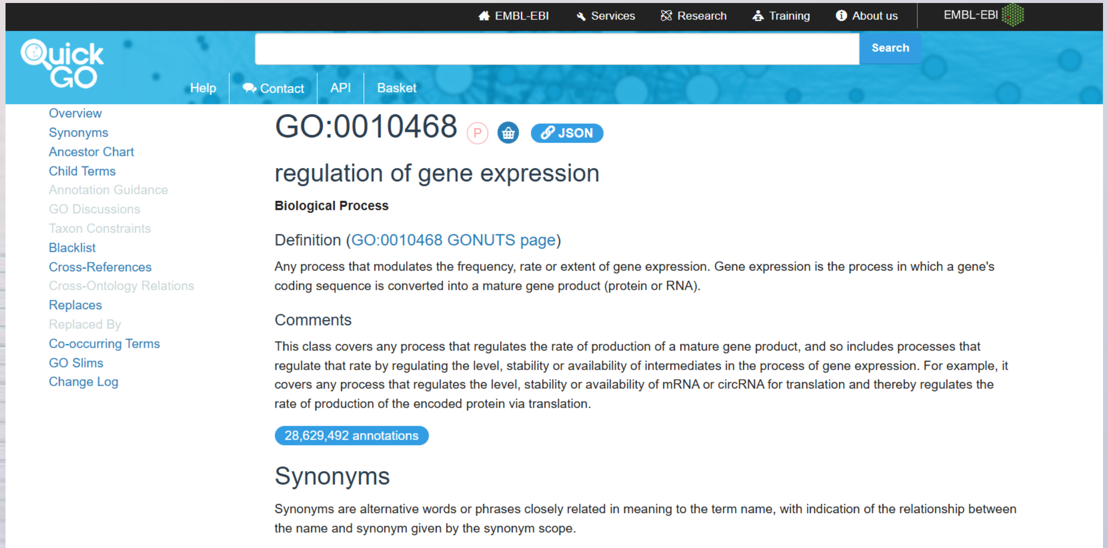
这种表达模式提示 NTT/WIP2 可能在植物生殖发育和种子成熟过程中发挥重要作用

以下简称该序列为WIP2

GO

用于描述基因产物功能和相关生物学信息的标准化术语。GO计划旨在为生物学家提供一个统一的框架，用于描述基因产物在以下三个主要方面的功能

分子功能；生物功能；细胞组分



GO:0010468

regulation of gene expression

Biological Process

Definition (GO:0010468 GONUTS page)

Any process that modulates the frequency, rate or extent of gene expression. Gene expression is the process in which a gene's coding sequence is converted into a mature gene product (protein or RNA).

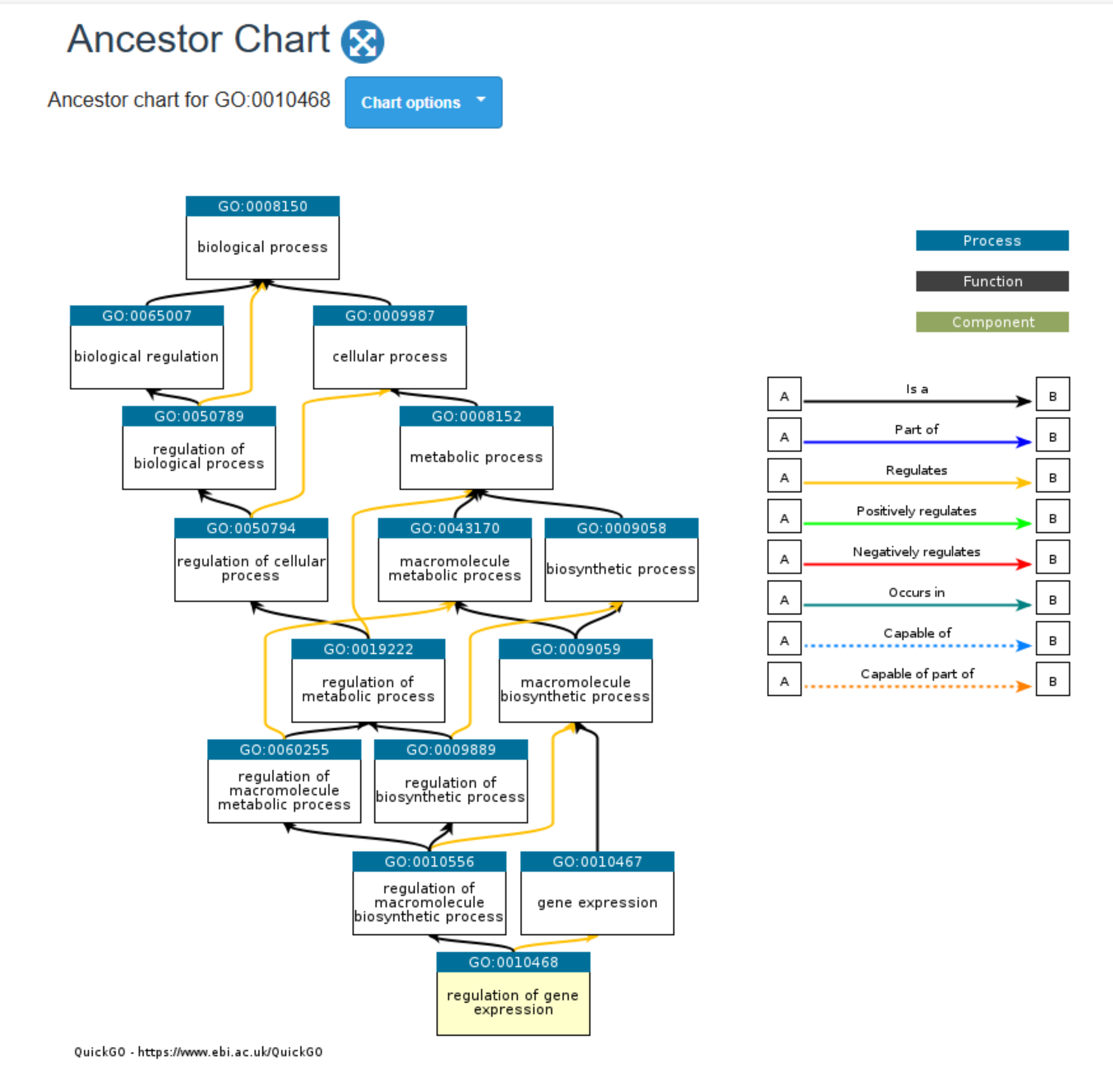
Comments

This class covers any process that regulates the rate of production of a mature gene product, and so includes processes that regulate that rate by regulating the level, stability or availability of intermediates in the process of gene expression. For example, it covers any process that regulates the level, stability or availability of mRNA or circRNA for translation and thereby regulates the rate of production of the encoded protein via translation.

28,629,492 annotations

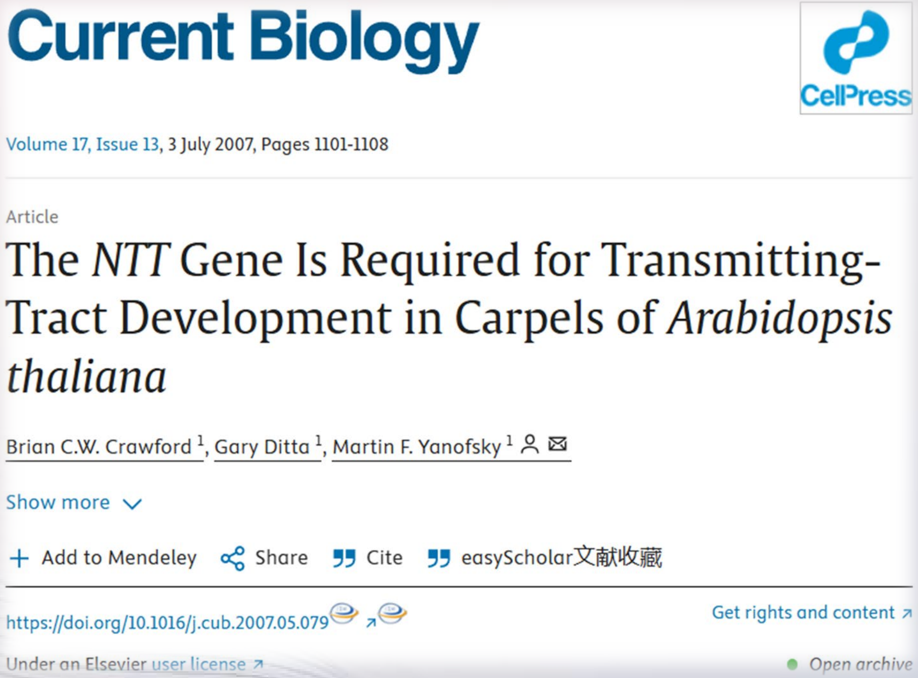
Synonyms

Synonyms are alternative words or phrases closely related in meaning to the term name, with indication of the relationship between the name and synonym given by the synonym scope.



基因本体（Gene Ontology, GO）祖先图

NTT/WIP2/AT3G57670.1



Current Biology

CellPress

Volume 17, Issue 13, 3 July 2007, Pages 1101-1108

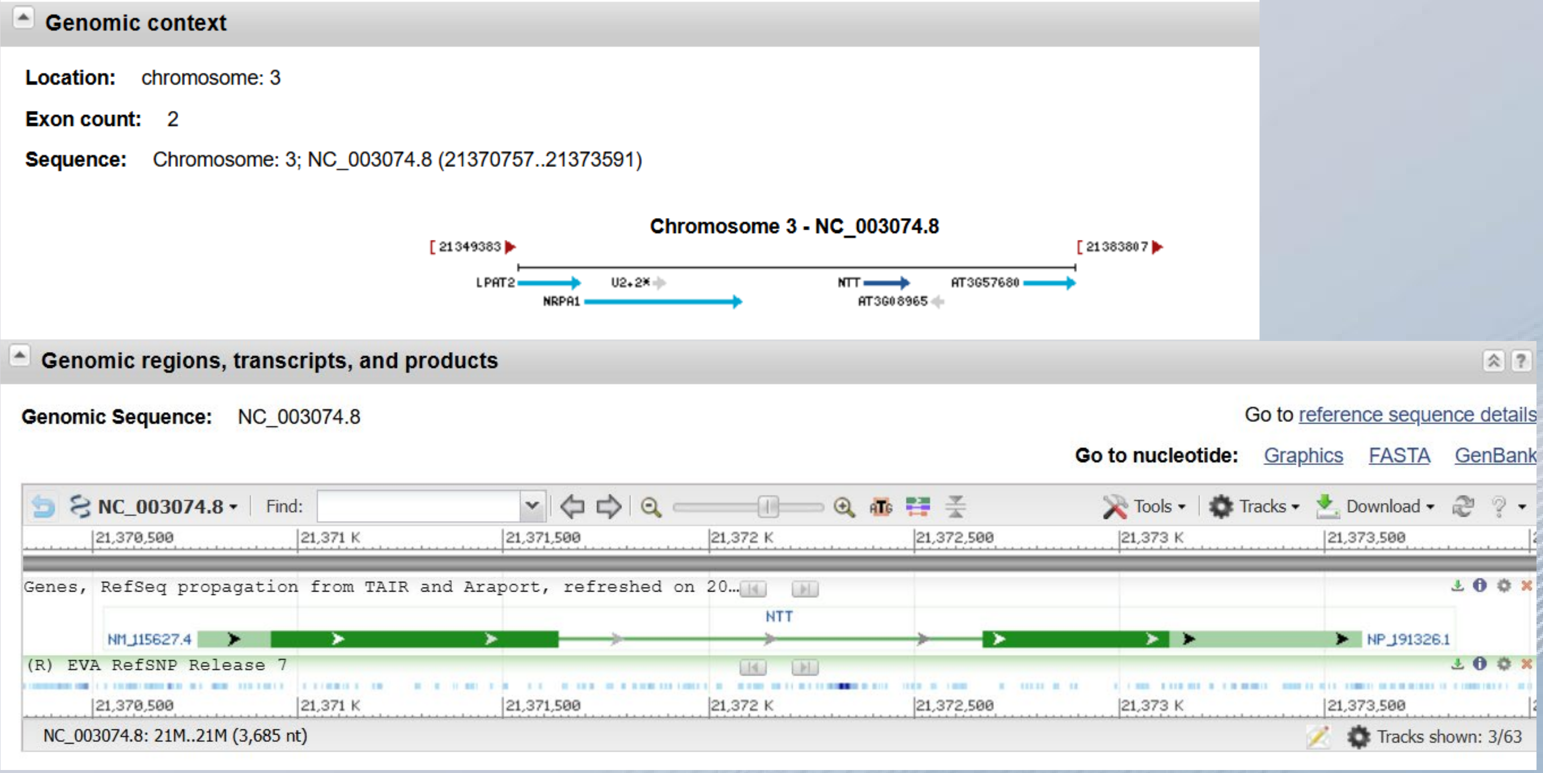
Article

The *NTT* Gene Is Required for Transmitting-Tract Development in Carpels of *Arabidopsis thaliana*

Brian C.W. Crawford [1], Gary Ditta [1], Martin F. Yanofsky [1]

https://doi.org/10.1016/j.cub.2007.05.079

Under an Elsevier user license



Genomic context

Location: chromosome: 3

Exon count: 2

Sequence: Chromosome: 3; NC_003074.8 (21370757..21373591)

Genomic regions, transcripts, and products

Genomic Sequence: NC_003074.8

Go to reference sequence details

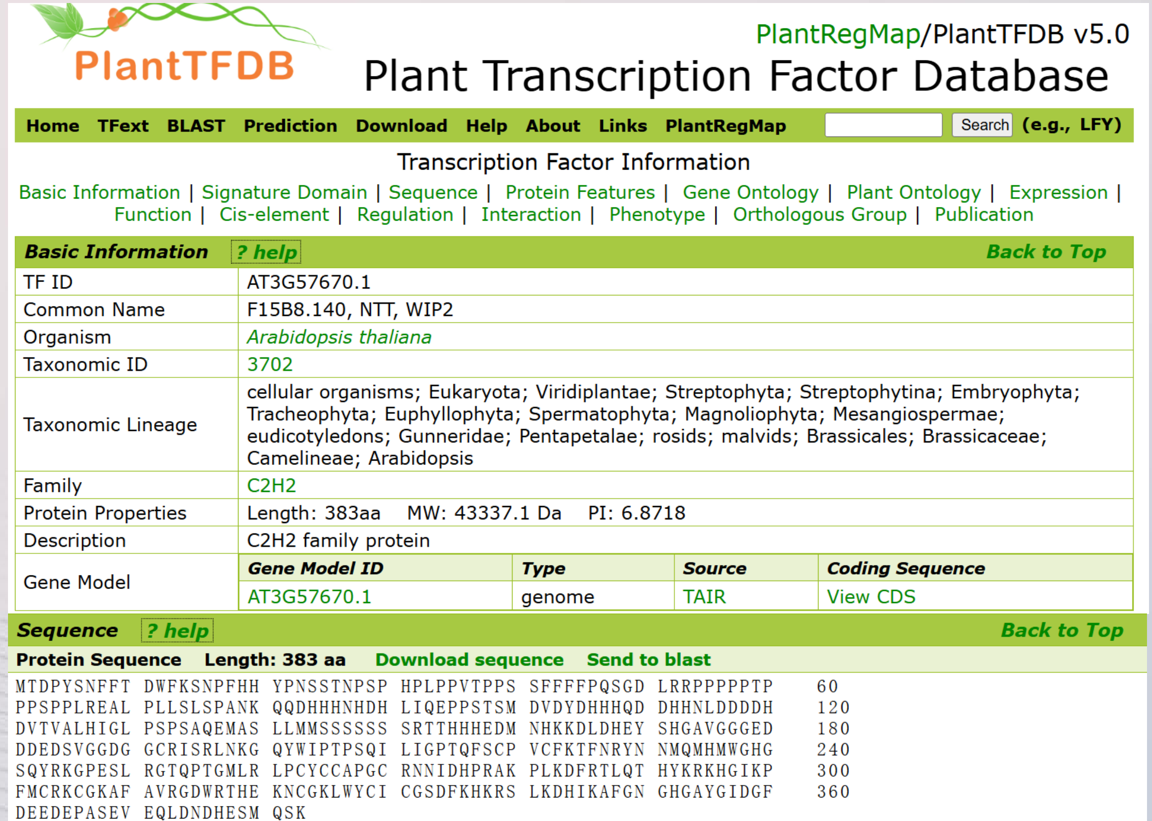
Go to nucleotide: Graphics FASTA GenBank

NC_003074.8: 21M..21M (3,685 nt)

组织特异性表达：NTT基因在拟南芥的多个组织中表达，但在花序分生组织、花器官（尤其是雄蕊和雌蕊）、果实和种子发育阶段表达量较高

生殖发育：在花器官的发育过程中，NTT基因在雄蕊和雌蕊中的高表达表明其在生殖发育中的重要作用

plantTFDB



PlantRegMap/PlantTFDB v5.0

Plant Transcription Factor Database

Transcription Factor Information

Basic Information | Signature Domain | Sequence | Protein Features | Gene Ontology | Plant Ontology | Expression | Function | Cis-element | Regulation | Interaction | Phenotype | Orthologous Group | Publication

Basic Information

TF ID	AT3G57670.1
Common Name	F15B8.140, NTT, WIP2
Organism	*Arabidopsis thaliana*
Taxonomic ID	3702
Taxonomic Lineage	cellular organisms; Eukaryota; Viridiplantae; Streptophyta; Streptophytina; Embryophyta; Tracheophyta; Euphyllophyta; Spermatophyta; Magnoliophyta; Mesangiospermae; eudicotyledons; Gunneridae; Pentapetalae; rosids; malvids; Brassicales; Brassicaceae; Camelineae; Arabidopsis
Family	C2H2
Protein Properties	Length: 383aa MW: 43337.1 Da PI: 6.8718
Description	C2H2 family protein

Gene Model

Gene Model ID	Type	Source	Coding Sequence
AT3G57670.1	genome	TAIR	View CDS

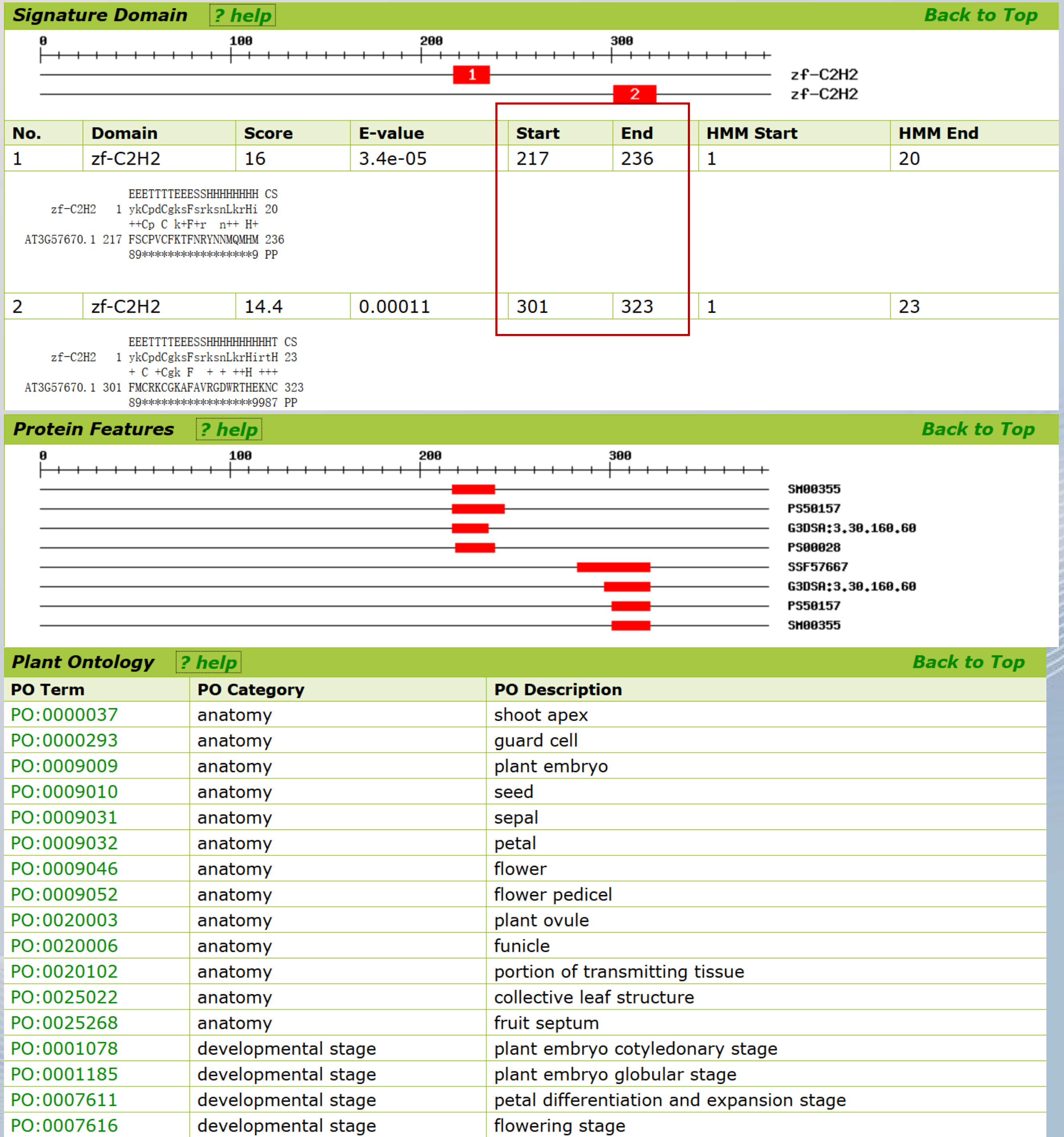
Sequence

Protein Sequence **Length: 383 aa** **Download sequence** **Send to blast**

```
MTDPYSNFFT DWFKSNPFHH YPNSSTNPSP HPLPPVTPPS SFFFFPQSGD LRRPPPPPTP    60
PPSPPLREAL PLLSLSPANK QQDHHHNHDH LIQEPPSTSM DVDYDHHHQD DHHNLDDDDH   120
DVTVALHIGL PSPSAQEMAS LLMMSSSSSS SRTTHHHEDM NHKKDLDHEY SHGAVGGGED   180
DDEDSVGGDG GCRISRLNKG QYWIPTPSQI LIGPTQFSCP VCFKTFNRYN NMQMHMWGHG   240
SQYRKGPESL RGTQPTGMLR LPCYCCAPGC RNNIDHPRAK PLKDFRTLQT HYKRKHGIKP   300
FMCRKCGKAF AVRGDWRTHE KNCGKLWYCI CGSDFKHKRS LKDHIKAFGN GHGAYGIDGF   360
DEEDEPASEV EQLDNDHESM QSK
```

这张蛋白质特征图展示了一个蛋白质序列中不同特征或结构域的位置和类型

表格提供了一组标准化的术语，用于描述植物的不同解剖结构和发育阶段。



No.	Domain	Score	E-value	Start	End	HMM Start	HMM End
1	zf-C2H2	16	3.4e-05	217	236	1	20
2	zf-C2H2	14.4	0.00011	301	323	1	23

```
                     EEETTTTEEESSHHHHHHHH CS
      zf-C2H2   1 ykCpdCgksFsrksnLkrHi 20
                    ++Cp C k+F+r  n++ H+
  AT3G57670.1 217 FSCPVCFKTFNRYNNMQMHM 236
                    89******************9 PP
```

```
                     EEETTTTEEESSHHHHHHHHHHT CS
      zf-C2H2   1 ykCpdCgksFsrksnLkrHirtH 23
                    + C +Cgk F  + + ++H +++
  AT3G57670.1 301 FMCRKCGKAFAVRGDWRTHEKNC 323
                    89*****************9987 PP
```

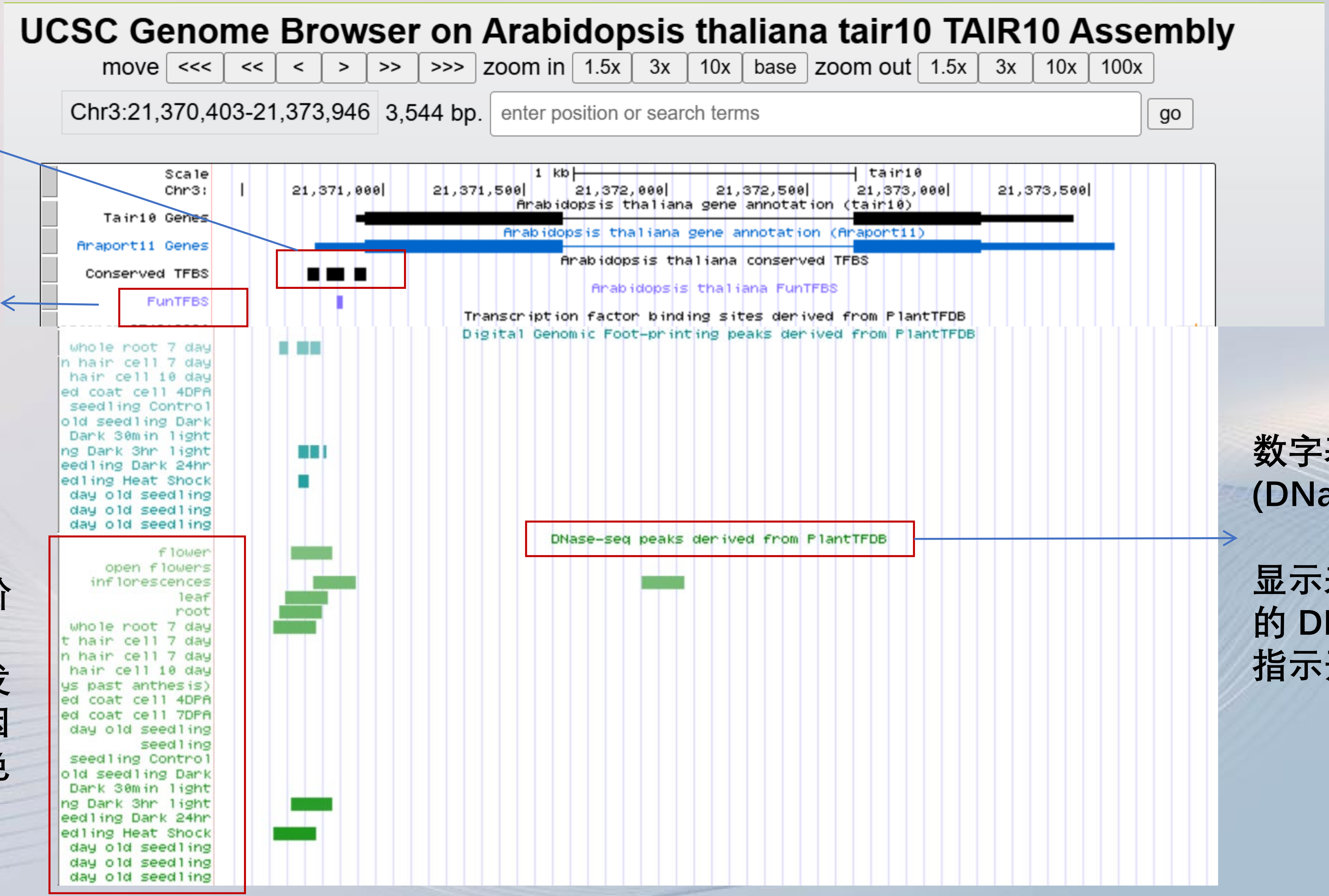
PO Term	PO Category	PO Description
PO:0000037	anatomy	shoot apex
PO:0000293	anatomy	guard cell
PO:0009009	anatomy	plant embryo
PO:0009010	anatomy	seed
PO:0009031	anatomy	sepal
PO:0009032	anatomy	petal
PO:0009046	anatomy	flower
PO:0009052	anatomy	flower pedicel
PO:0020003	anatomy	plant ovule
PO:0020006	anatomy	funicle
PO:0020102	anatomy	portion of transmitting tissue
PO:0025022	anatomy	collective leaf structure
PO:0025268	anatomy	fruit septum
PO:0001078	developmental stage	plant embryo cotyledonary stage
PO:0001185	developmental stage	plant embryo globular stage
PO:0007611	developmental stage	petal differentiation and expansion stage
PO:0007616	developmental stage	flowering stage

植物调控图：基因注释、转录因子结合位点和染色质开放性数据

保守转录因子的结合位点

基于该数据库转录因子的结合位点

不同组织和发育阶段：
显示不同组织和发育阶段（的转录因子结合位点和染色质开放性数据。

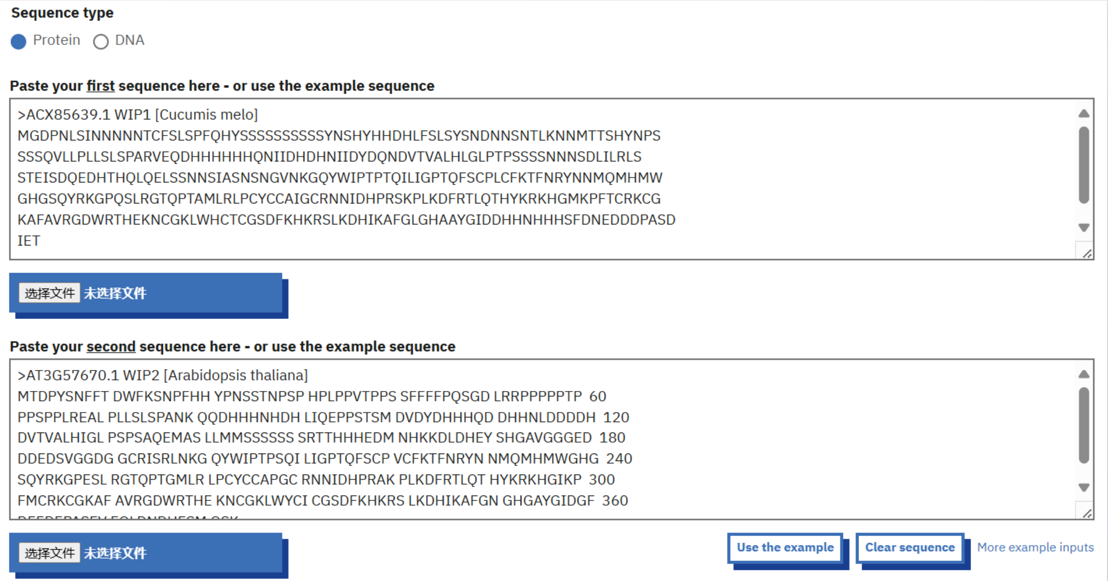


UCSC Genome Browser on Arabidopsis thaliana tair10 TAIR10 Assembly

数字基因组足迹 (DNase-seq)：

显示来自 PlantTFDB 的 DNase-seq 峰值，指示开放染色质区域

EBI Needle



Sequence type

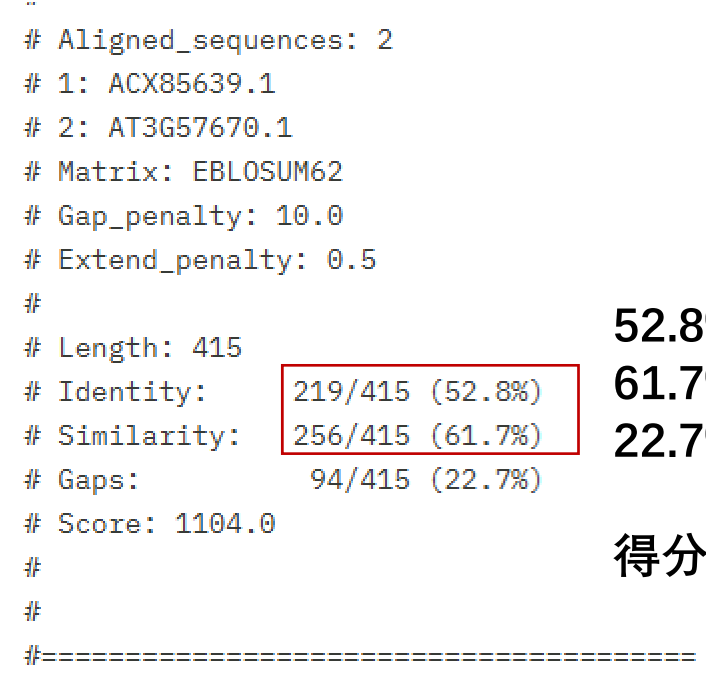
● Protein ○ DNA

Paste your first sequence here - or use the example sequence

```
>ACX85639.1 WIP1 [Cucumis melo]
MGDPNLSINNNNNTCFSLSPFQHYSSSSSSSSSSSYNSHYHHDHLFSLSYSNDNNSNTLKNNMTTSHYNPS
SSSQVLLPLLSLSPARVEQDHHHHHHQNIIDHDHNIIDYDQNDVTVALHLGLPTPSSSSNNNSDLILRLS
STEISDQEDHTHQLQELSSNNSIASNSNGVNKGQYWIPTPTQILIGPTQFSCPLCFKTFNRYNNMQMHMW
GHGSQYRKGPQSLRGTQPTAMLRLPCYCCAIGCRNNIDHPRSKPLKDFRTLQTHYKRKHGMKPFTCRKCG
KAFAVRGDWRTHEKNCGKLWHCTCGSDFKHKRSLKDHIKAFGLGHAAYGIDDHHNHHHSFDNEDDDPASD
IET
```

Paste your second sequence here - or use the example sequence

```
>AT3G57670.1 WIP2 [Arabidopsis thaliana]
MTDPYSNFFT DWFKSNPFHH YPNSSTNPSP HPLPPVTPPS SFFFFPQSGD LRRPPPPPTP  60
PPSPPLREAL PLLSLSPANK QQDHHHNHDH LIQEPPSTSM DVDYDHHHQD DHHNLDDDDH  120
DVTVALHIGL PSPSAQEMAS LLMMSSSSSS SRTTHHHEDM NHKKDLDHEY SHGAVGGGED  180
DDEDSVGGDG GCRISRLNKG QYWIPTPSQI LIGPTQFSCP VCFKTFNRYN NMQMHMWGHG  240
SQYRKGPESL RGTQPTGMLR LPCYCCAPGC RNNIDHPRAK PLKDFRTLQT HYKRKHGIKP  300
FMCRKCGKAF AVRGDWRTHE KNCGKLWYCI CGSDFKHKRS LKDHIKAFGN GHGAYGIDGF  360
```



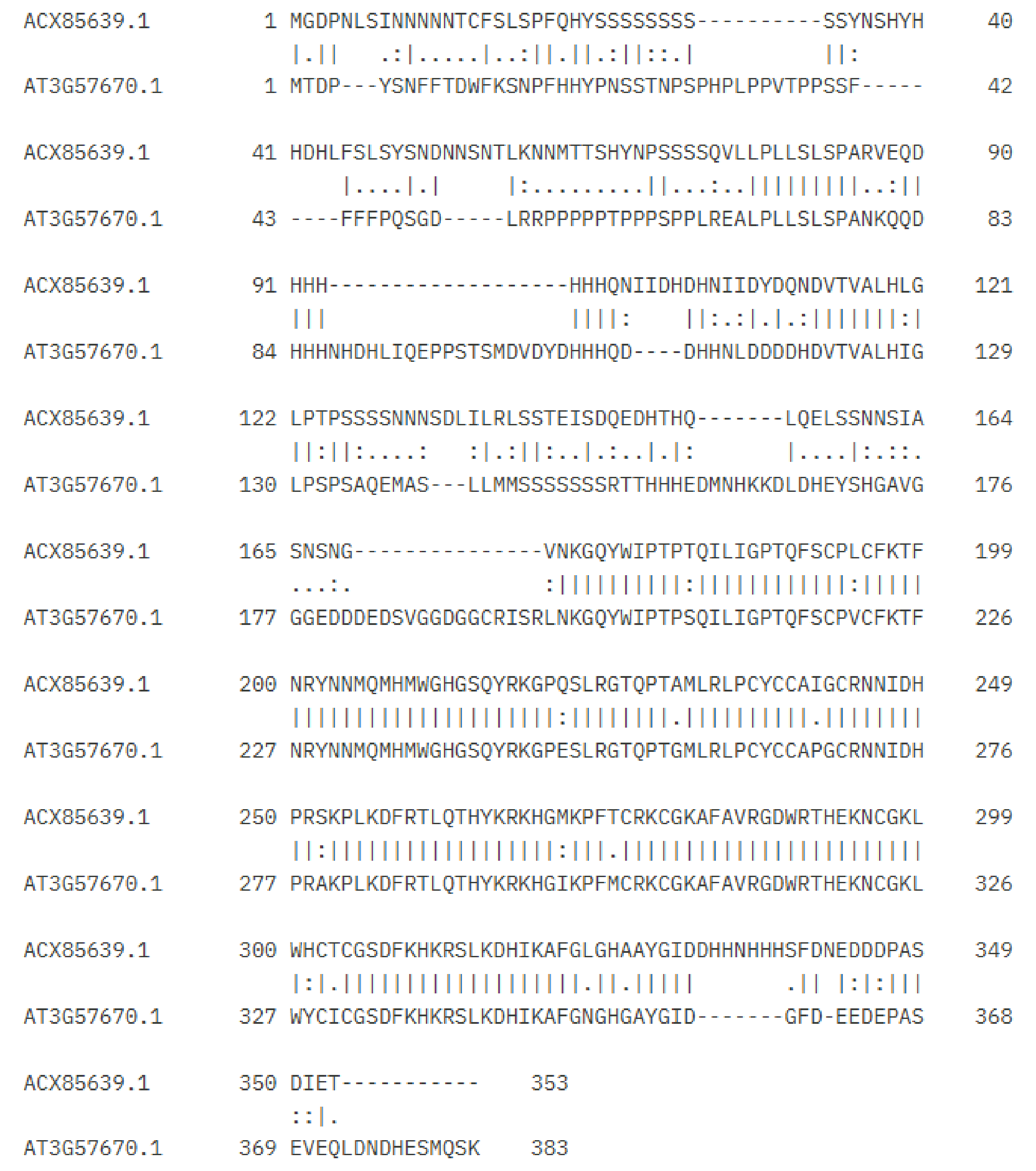
```
# Aligned_sequences: 2
# 1: ACX85639.1
# 2: AT3G57670.1
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 415
# Identity:     219/415 (52.8%)
# Similarity:   256/415 (61.7%)
# Gaps:          94/415 (22.7%)
# Score: 1104.0
#
#
#=======================================
```

52.8%的同一性

61.7%的相似性

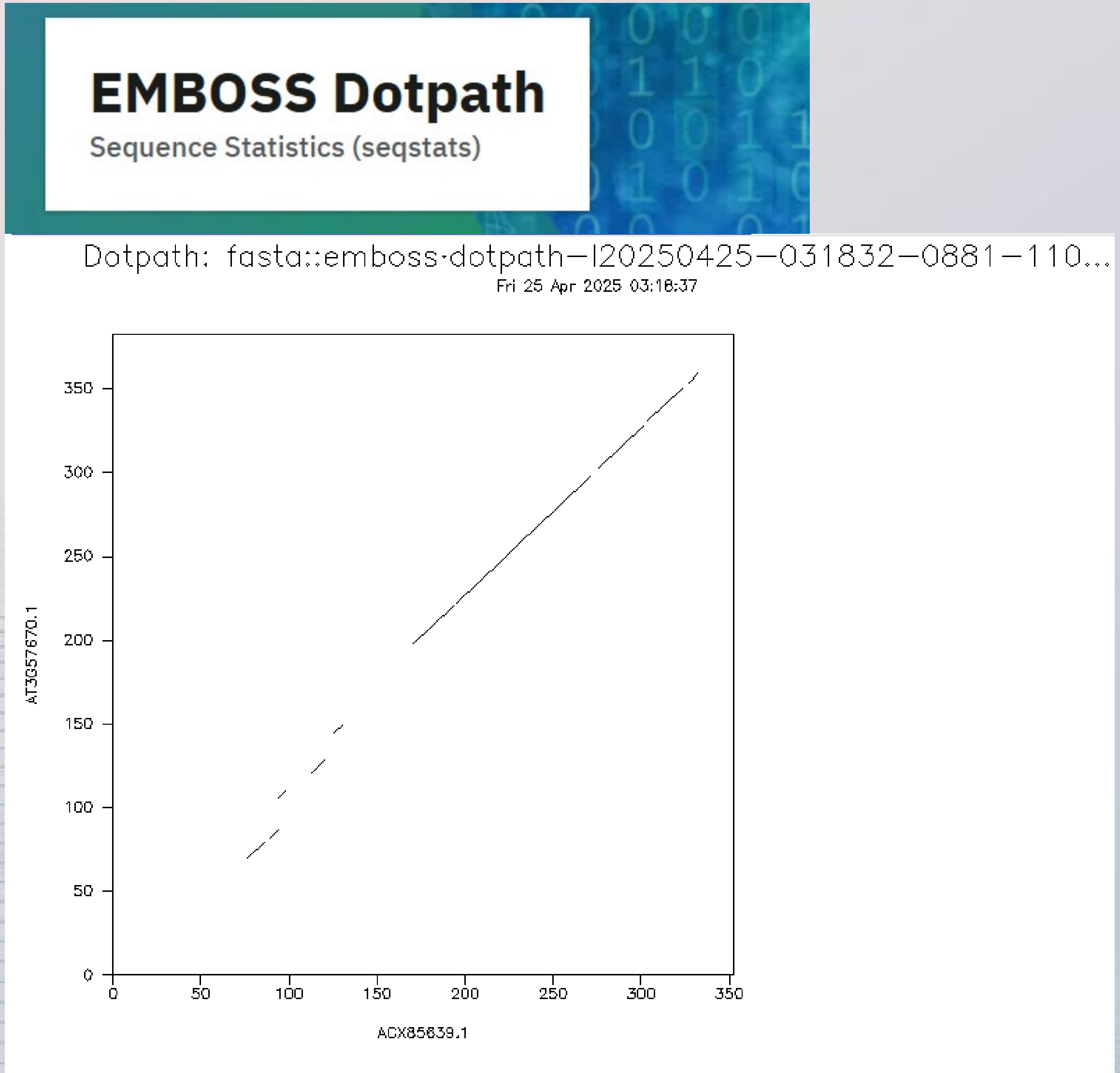
22.7%差距比例

得分1104.0，结果可靠



```
ACX85639.1         1 MGDPNLSINNNNNTCFSLSPFQHYSSSSSSSS----------SSYNSHYH     40
                     |.||   .:|.....|..:||.||.:||::.|          ||:
AT3G57670.1        1 MTDP---YSNFFTDWFKSNPFHHYPNSSTNPSPHPLPPVTPPSSF-----     42

ACX85639.1        41 HDHLFSLSYSNDNNSNTLKNNMTTSHYNPSSSSQVLLPLLSLSPARVEQD     90
                         |....|.|     |:.........||...:..|||||||||..:||
AT3G57670.1       43 ----FFFPQSGD-----LRRPPPPPTPPPSPPLREALPLLSLSPANKQQD     83

ACX85639.1        91 HHH-------------------HHHQNIIDHDHNIIDYDQNDVTVALHLG    121
                     |||                   ||||:    ||:.:|.|.:|||||||:|
AT3G57670.1       84 HHHNHDHLIQEPPSTSMDVDYDHHHQD----DHHNLDDDDHDVTVALHIG    129

ACX85639.1       122 LPTPSSSSNNNSDLILRLSSTEISDQEDHTHQ-------LQELSSNNSIA    164
                     ||:||:....:   :|.:||:..|.:..|.|:       |....|:.::.
AT3G57670.1      130 LPSPSAQEMAS---LLMMSSSSSSSRTTHHHEDMNHKKDLDHEYSHGAVG    176

ACX85639.1       165 SNSNG---------------VNKGQYWIPTPTQILIGPTQFSCPLCFKTF    199
                     ...:.               :||||||||||:||||||||||||:|||||
AT3G57670.1      177 GGEDDDEDSVGGDGGCRISRLNKGQYWIPTPSQILIGPTQFSCPVCFKTF    226

ACX85639.1       200 NRYNNMQMHMWGHGSQYRKGPQSLRGTQPTAMLRLPCYCCAIGCRNNIDH    249
                     |||||||||||||||||||||:||||||||.||||||||||.||||||||
AT3G57670.1      227 NRYNNMQMHMWGHGSQYRKGPESLRGTQPTGMLRLPCYCCAPGCRNNIDH    276

ACX85639.1       250 PRSKPLKDFRTLQTHYKRKHGMKPFTCRKCGKAFAVRGDWRTHEKNCGKL    299
                     ||:||||||||||||||||||:|||.||||||||||||||||||||||||
AT3G57670.1      277 PRAKPLKDFRTLQTHYKRKHGIKPFMCRKCGKAFAVRGDWRTHEKNCGKL    326

ACX85639.1       300 WHCTCGSDFKHKRSLKDHIKAFGLGHAAYGIDDHHNHHHSFDNEDDDPAS    349
                     |:|.|||||||||||||||||||.||.|||||       .|| |:|:|||
AT3G57670.1      327 WYCICGSDFKHKRSLKDHIKAFGNGHGAYGID-------GFD-EEDEPAS    368

ACX85639.1       350 DIET-----------    353
                     ::|.
AT3G57670.1      369 EVEQLDNDHESMQSK    383
```

Dotpath



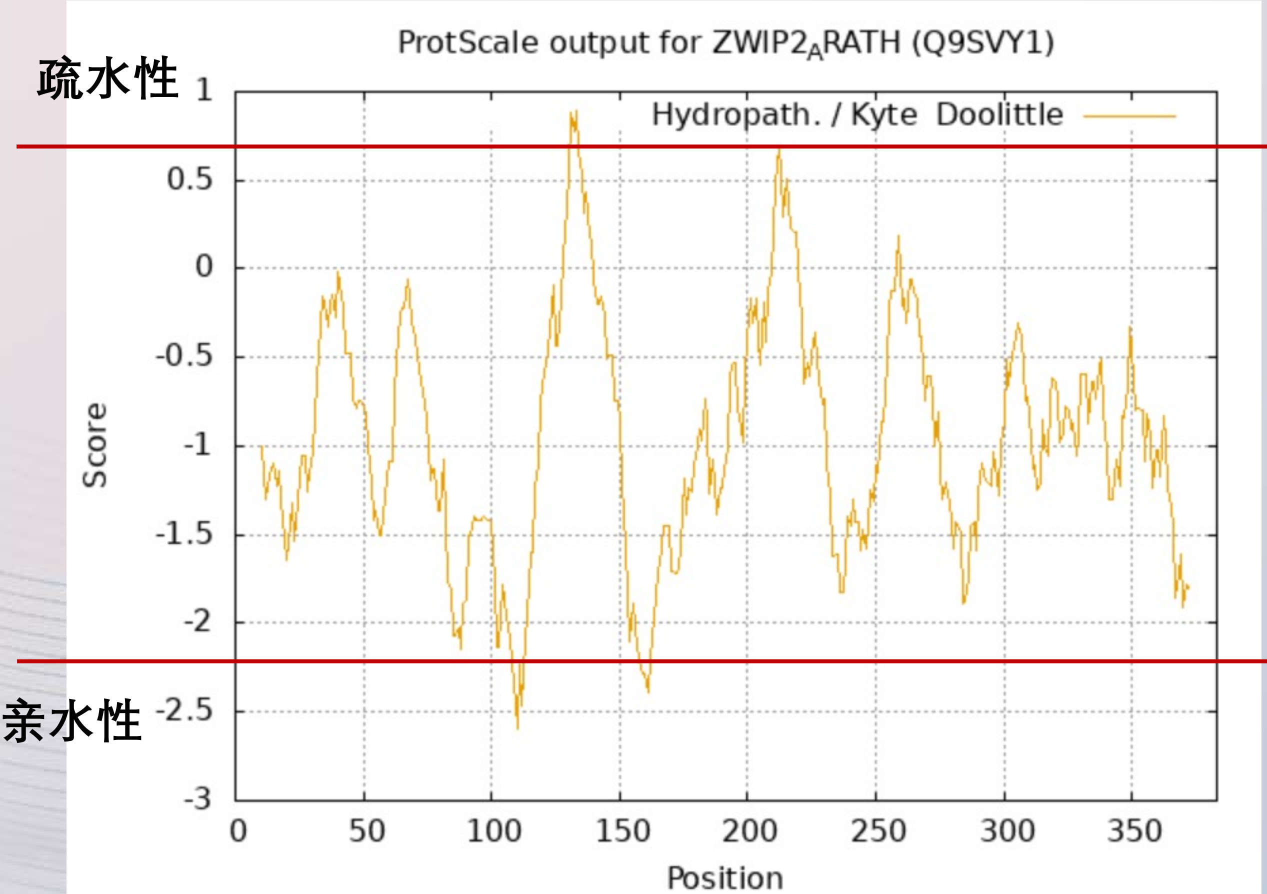
相似性：两个序列在大部分区域具有相似性，但斜线的不均匀性表明在某些区域相似性可能较低

插入或缺失：这些区域存在插入或缺失。

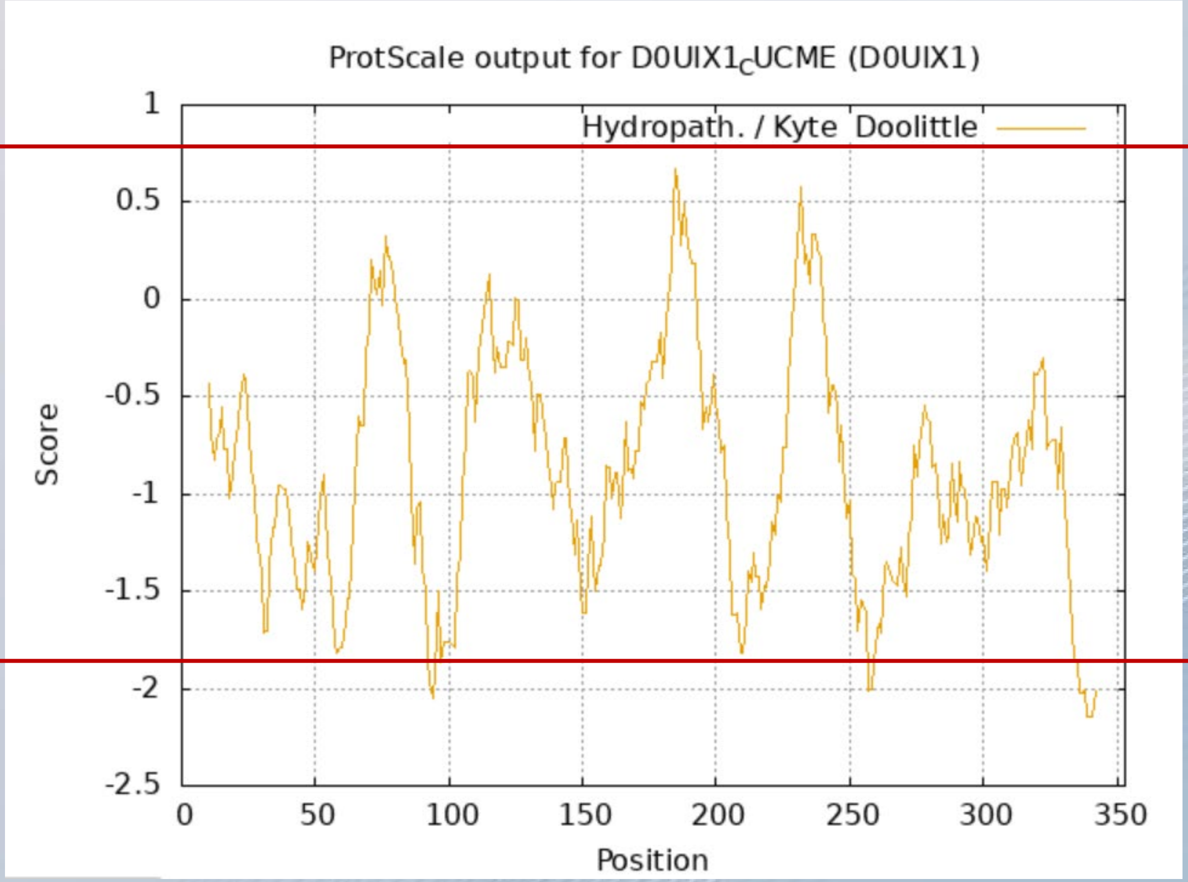
整体相似性：从图中可以看出，这两个序列在大部分区域具有较高的相似性。

Expasy

拟南芥WIP2



甜瓜WIP1

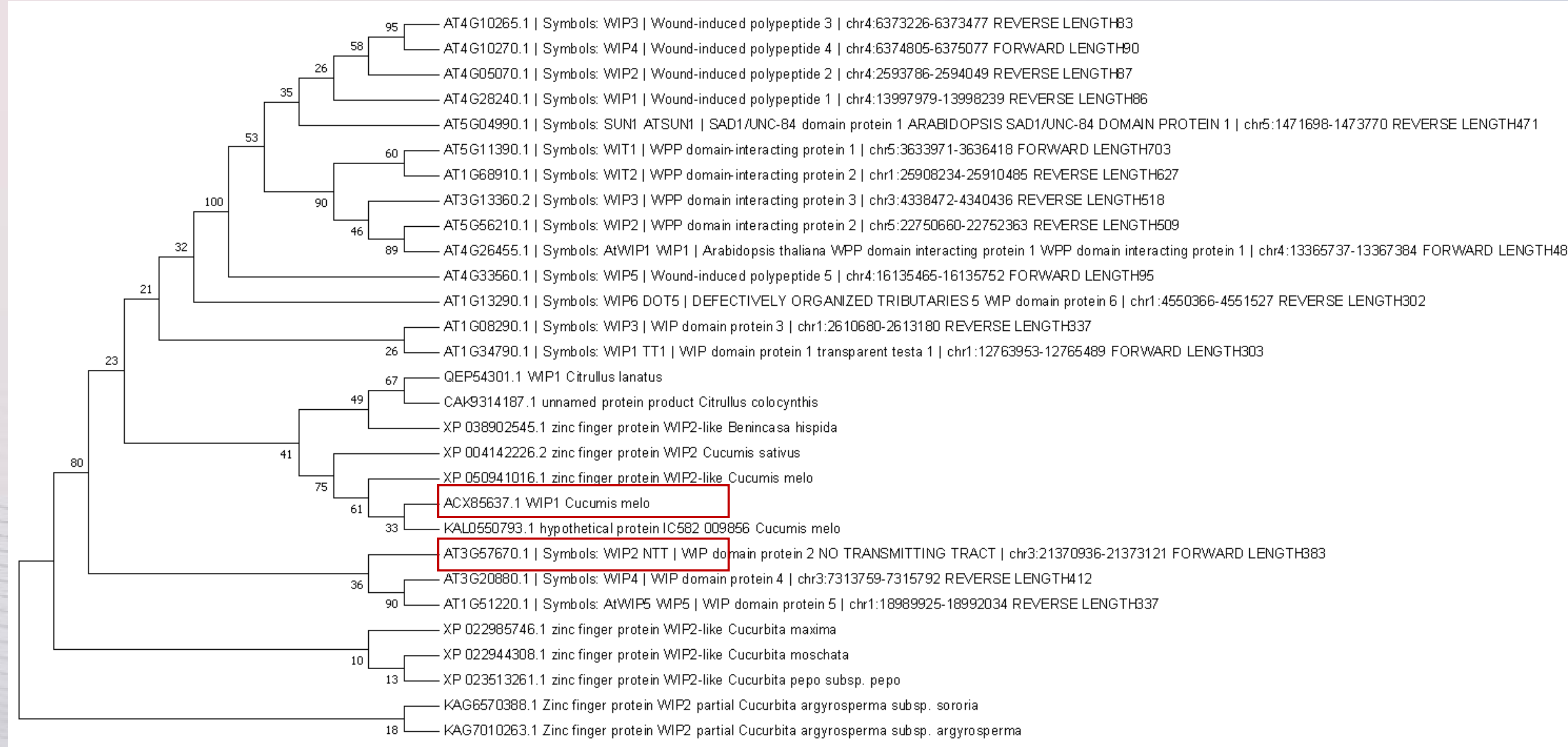


高疏水性区域可能形成跨膜结构或与其他疏水性分子相互作用，该蛋白质可能定位于细胞膜

亲水性区域可能参与与其他蛋白质、核酸或其他极性分子的相互作用

MEGA7

19个拟南芥C2H2锌指家族WIP转录因子和12个葫芦科植物相似序列比对



ACX85637.1/WIP1

来自甜瓜的蛋白质

AT3G57670.1/WIP2

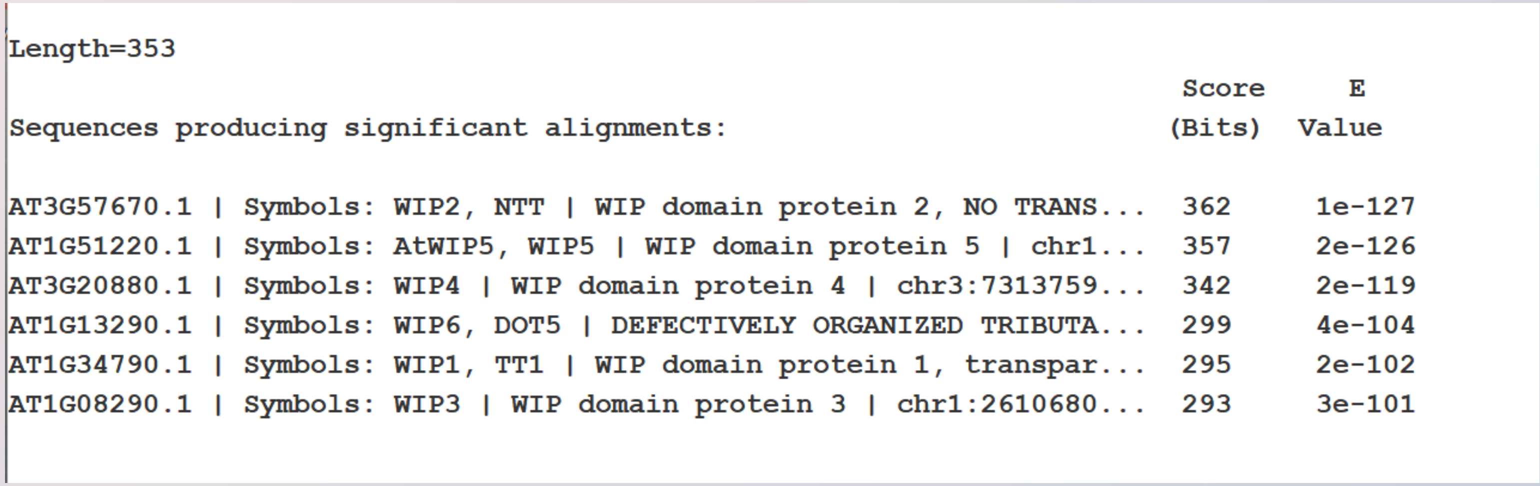
来自拟南芥的蛋白质，

标注为 WIP2 NTT

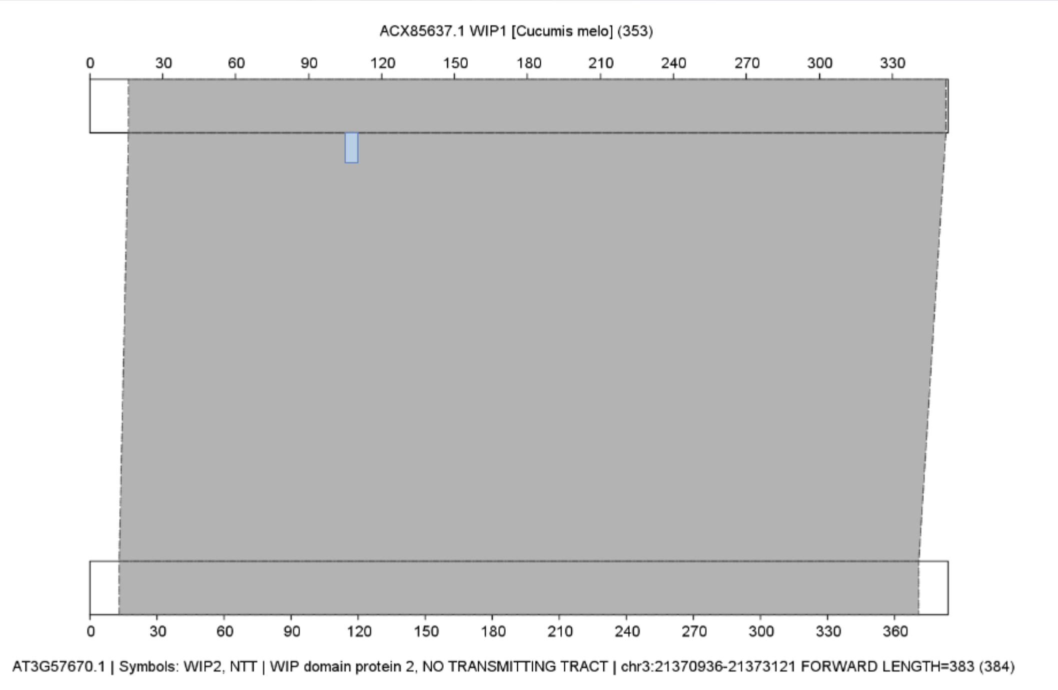
ACX85637.1 和 AT3G57670.1 之间的分支长度较长，说明序列差异较大，进化关系相对较远，可能在进化过程中从一个共同祖先分化而来，各自适应了不同的生物体和功能需求

作为 WIP 家族的成员，可能在功能上具有一定的相似性，但具体功能和作用机制可能因物种不同而有所差异

TBtools

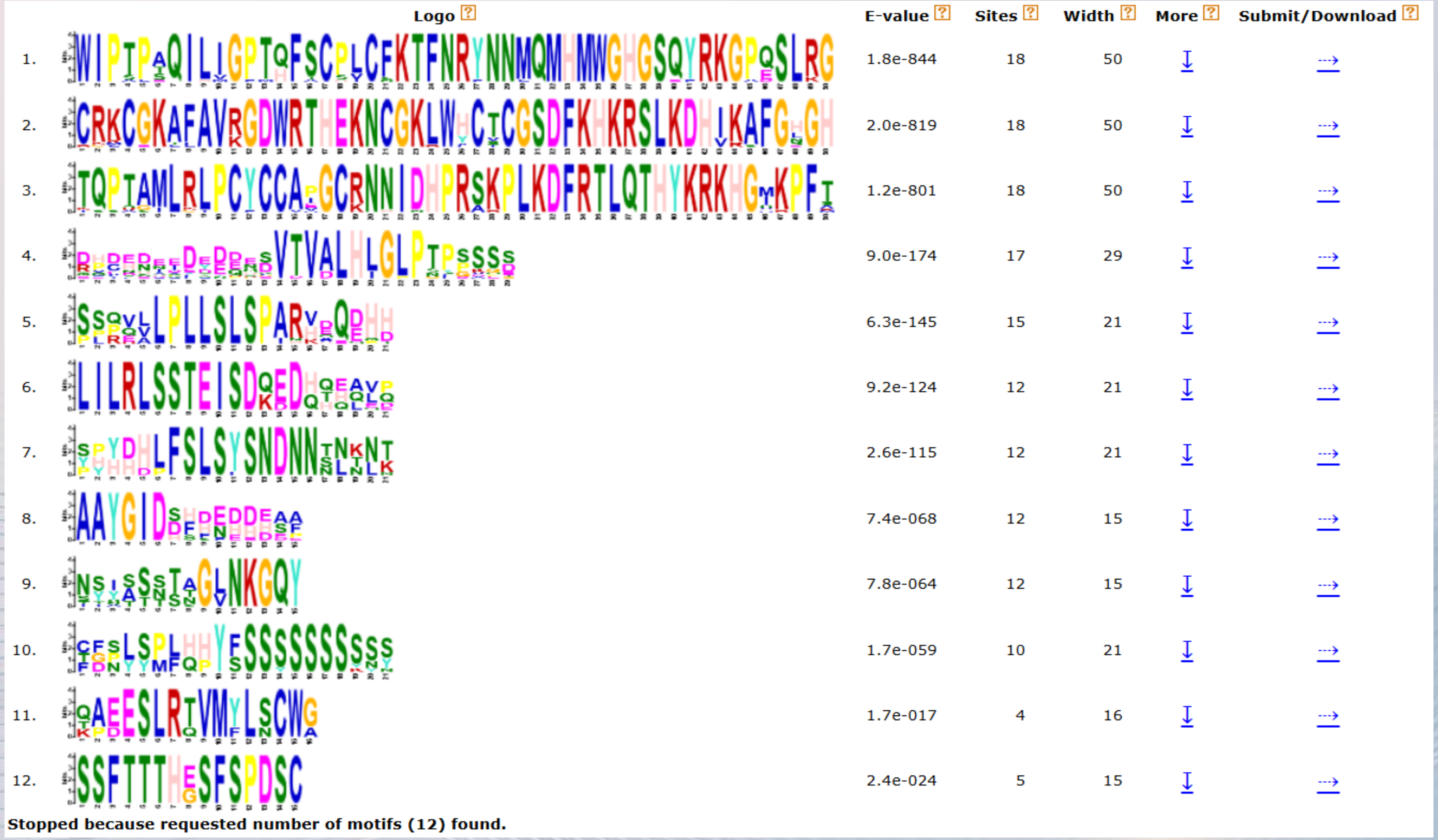


```
Length=353
                                                                          Score     E
Sequences producing significant alignments:                              (Bits)  Value

AT3G57670.1 | Symbols: WIP2, NTT | WIP domain protein 2, NO TRANS...  362      1e-127
AT1G51220.1 | Symbols: AtWIP5, WIP5 | WIP domain protein 5 | chr1...  357      2e-126
AT3G20880.1 | Symbols: WIP4 | WIP domain protein 4 | chr3:7313759...  342      2e-119
AT1G13290.1 | Symbols: WIP6, DOT5 | DEFECTIVELY ORGANIZED TRIBUTA...  299      4e-104
AT1G34790.1 | Symbols: WIP1, TT1 | WIP domain protein 1, transpar...  295      2e-102
AT1G08290.1 | Symbols: WIP3 | WIP domain protein 3 | chr1:2610680...  293      3e-101
```



对甜瓜的WIP1转录因子比对拟南芥19个序列，依然获得相似性较高的序列WIP2

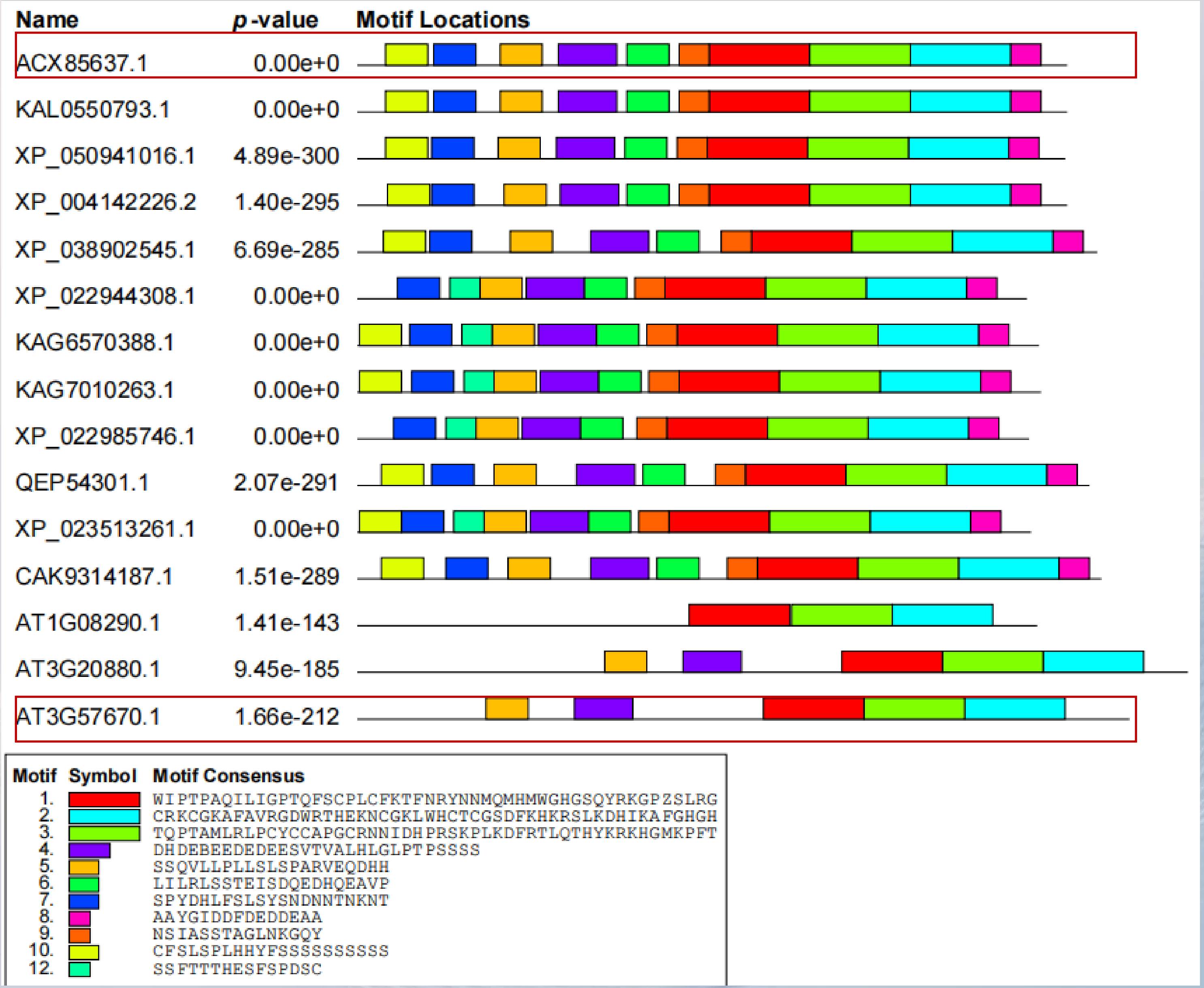
MEME



	Logo	E-value	Sites	Width	More	Submit/Download
1.		1.8e-844	18	50	⤓	--->
2.		2.0e-819	18	50	⤓	--->
3.		1.2e-801	18	50	⤓	--->
4.		9.0e-174	17	29	⤓	--->
5.		6.3e-145	15	21	⤓	--->
6.		9.2e-124	12	21	⤓	--->
7.		2.6e-115	12	21	⤓	--->
8.		7.4e-068	12	15	⤓	--->
9.		7.8e-064	12	15	⤓	--->
10.		1.7e-059	10	21	⤓	--->
11.		1.7e-017	4	16	⤓	--->
12.		2.4e-024	5	15	⤓	--->

Stopped because requested number of motifs (12) found.

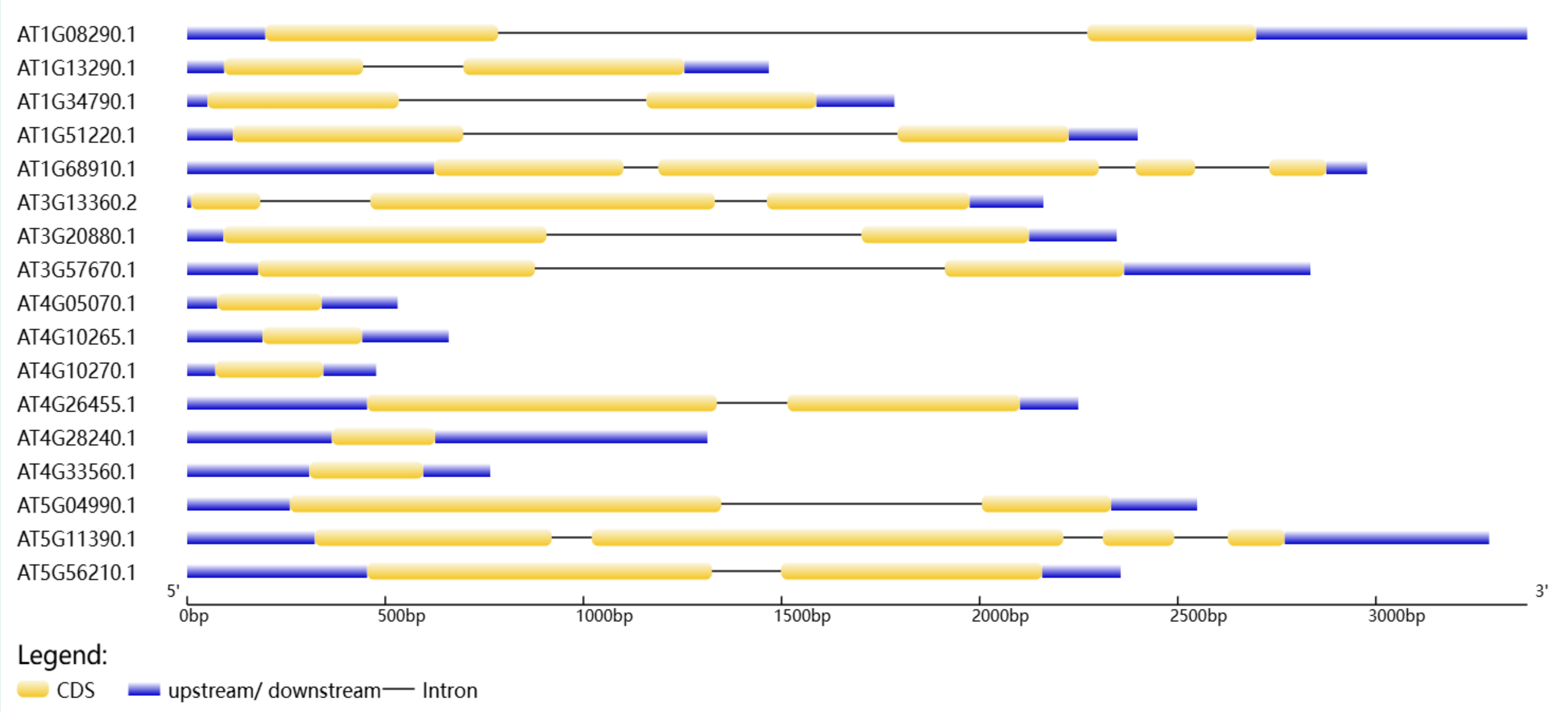
MEME



Name	*p*-value	Motif Locations
ACX85637.1	0.00e+0	
KAL0550793.1	0.00e+0	
XP_050941016.1	4.89e-300	
XP_004142226.2	1.40e-295	
XP_038902545.1	6.69e-285	
XP_022944308.1	0.00e+0	
KAG6570388.1	0.00e+0	
KAG7010263.1	0.00e+0	
XP_022985746.1	0.00e+0	
QEP54301.1	2.07e-291	
XP_023513261.1	0.00e+0	
CAK9314187.1	1.51e-289	
AT1G08290.1	1.41e-143	
AT3G20880.1	9.45e-185	
AT3G57670.1	1.66e-212	

Motif	Symbol	Motif Consensus
1.	🟥	WIPTPAQILIGPTQFSCPLCFKTFNRYNNMQMHMWGHGSQYRKGPZSLRG
2.	🟦	CRKCGKAFAVRGDWRTHEKNCGKLWHCTCGSDFKHKRSLKDHIKAFGHGH
3.	🟩	TQPTAMLRLPCYCCAPGCRNNIDHPRSKPLKDFRTLQTHYKRKHGMKPFT
4.	🟪	DHDEBEEDEDEESVTVALHLGLPTPSSSS
5.	🟧	SSQVLLPLLSLSPARVEQDHH
6.	🟩	LILRLSSTEISDQEDHQEAVP
7.	🟦	SPYDHLFSLSYSNDNNTNKNT
8.	🟪	AAYGIDDFDEDDEAA
9.	🟧	NSIASSTAGLNKGQY
10.	🟨	CFSLSPLHHYFSSSSSSSSSS
12.	🟩	SSFTTTHESFSPDSC

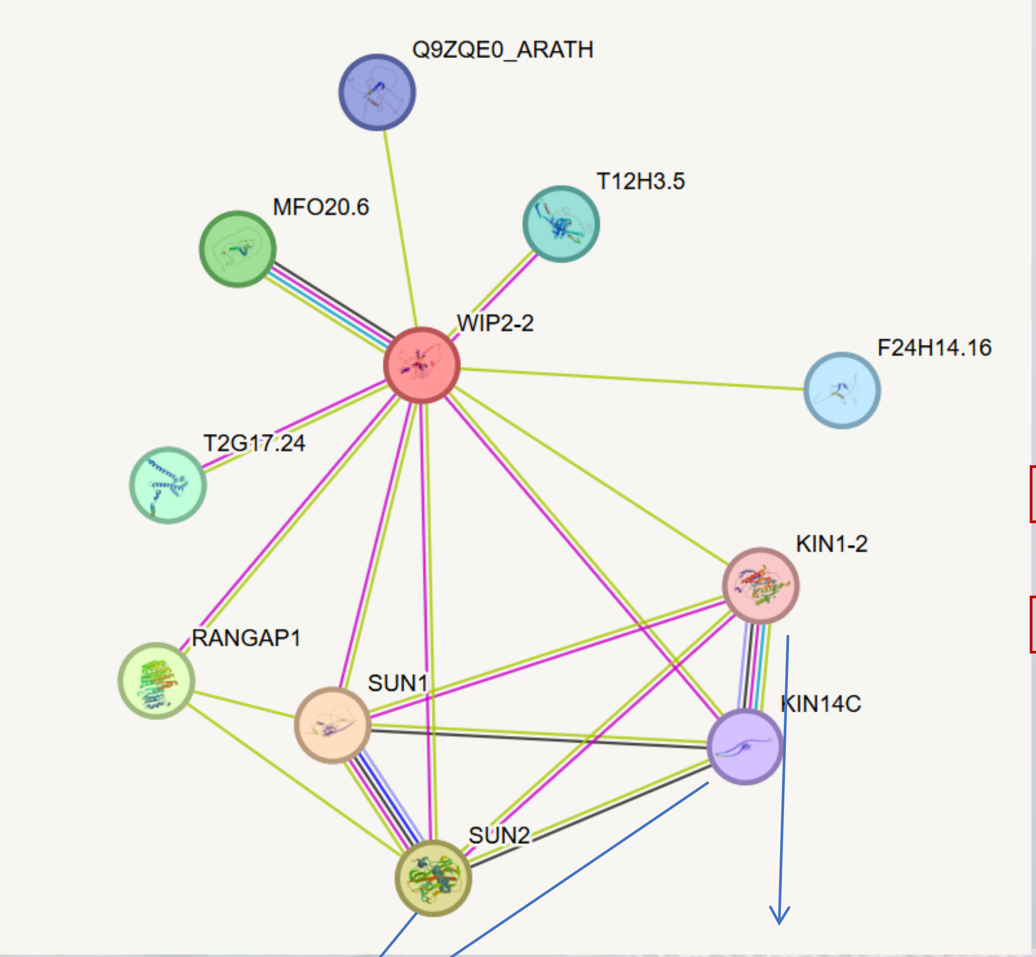
内含子与外显子结构

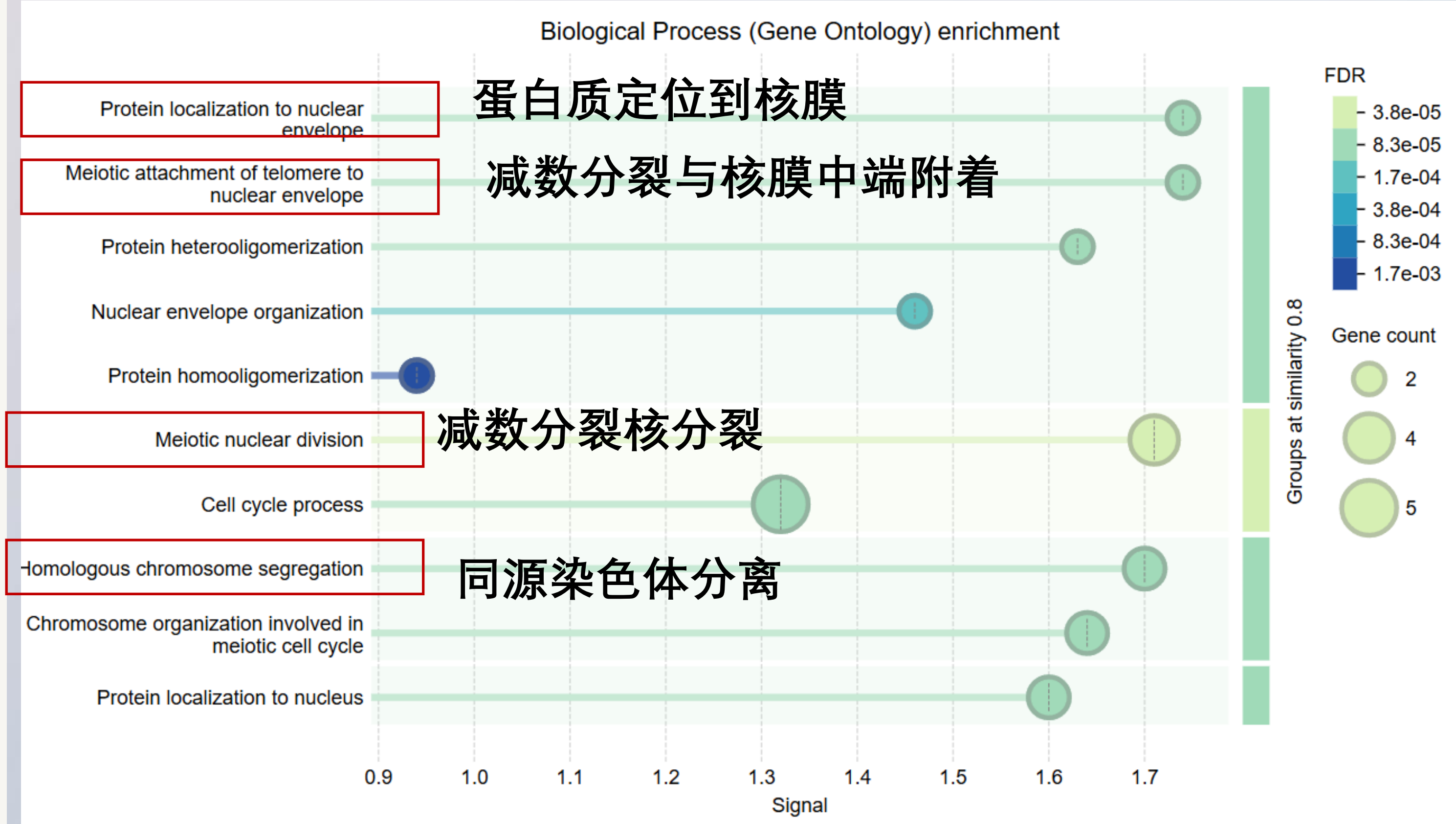


从拟南芥基因转录因子数据库 TAIR得到 17个基因19种剪接体的全长基因和编码区核苷酸序列，拟南芥WIP转录因子家族 17个成员基因结构存在差异较小，含有2-4个内含子

潜在互作蛋白

STRING





蛋白质定位到核膜

减数分裂与核膜中端附着

减数分裂核分裂

同源染色体分离

减数分裂和有丝分裂中在纺锤体发挥作用

在雌配子发生中的核分裂周期中发挥作用。

在所分析的数据集中，与核膜、减数分裂、细胞周期和蛋白质定位相关的生物学过程显著富集

04.

规划

后续生信分析：将以上分析运用于蓖麻中控制两性花到单性花的相关基因上

表观遗传与调控机制探索

甲基化数据挖掘：从公共表观数据库获取雌雄花不同发育阶段的甲基化测序数据（如WGBS），分析WIP1基因启动子或编码区的甲基化差异

染色质可及性预测

ATAC-seq或DNase-seq数据分析：评估WIP1所在区域的染色质开放状态，推测其转录活性与性别发育阶段的关联

实验验证的生物信息学衔接

CRISPR靶点设计：

使用CRISPR-P或CHOPCHOP设计针对蓖麻相关基因的sgRNA，确保靶向保守功能域（如锌指结构），避免脱靶

转录组数据分析：

差异表达基因筛选：对蓖麻基因敲除/过表达植株的RNA-seq数据进行差异分析（DESeq2/edgeR），筛选下游靶基因（如雌蕊抑制基因AGAMOUS-like、雄蕊促进基因AP3/PI同源基因）。

ChIP-seq数据分析：

结合位点鉴定：若蓖麻相关基因为转录因子，通过ChIP-seq预测其直接调控的靶基因启动子区域，结合Motif分析验证结合基序。

后续研究方向：

功能验证实验：构建蓖麻相关基因的**转基因/基因编辑植株**，观察雌雄蕊表型变化，结合qRT-PCR验证候选靶基因表达。

激素互作研究：分析蓖麻相关基因与乙烯/赤霉素信号通路的交叉调控（如通过外源激素处理后的表达变化）。

跨物种比较：在单性花与两性花中比较蓖麻相关功能差异，揭示其在性别分化中的保守性与特异性。

THANK YOU

2030