GRADUATE SCHOOL OF CHINESE ACADEMY OF AGRICULTURAL SCIENCES
中国农业科学院研究生院

甘薯块根膨大转录组数据分析

Transcriptome data analysis of sweet potato root enlargement

G1组成员：李成阳_G1A
张若晗_G1B
任丽莹_G1C
王　鹏_G1D

2022年12月4日

目　录
CONTENTS

一、课题背景

课题背景——李成阳_G1A

甘薯块根是收获的主要器官，其发育和扩展直接影响甘薯的产量(Fan *et al*., 2021)，目前，伴随分子生物学技术的快速发展、组学测序技术的不断应用以及甘薯二倍体祖先种*I.trifida*和*I.triloba*参考基因组的破译，块根发育生理过程的解析也取得了一定进展，但分子机理方面的研究还处于相对滞后的阶段(Song *et al*., 2021)。已鉴定基因数量远远不够，对其所行使的功能也鲜有探究，并且它们相互之间的关系、如何系统的调控甘薯块根膨大以及同激素等信号通路的相互作用等都不明确，因此对甘薯块根形成和膨大相关分子机理的研究空间仍然很大。

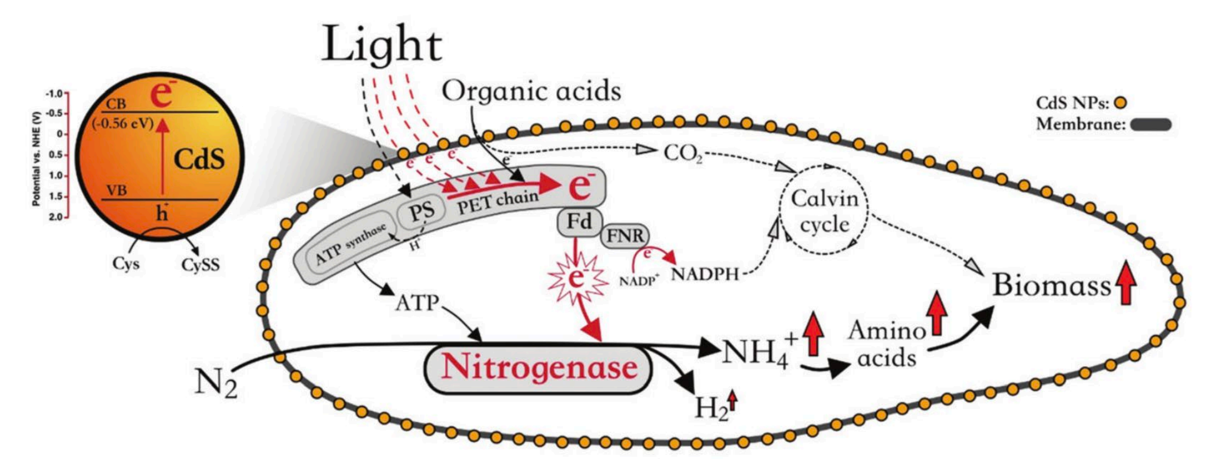
课题背景——张若晗_G1B

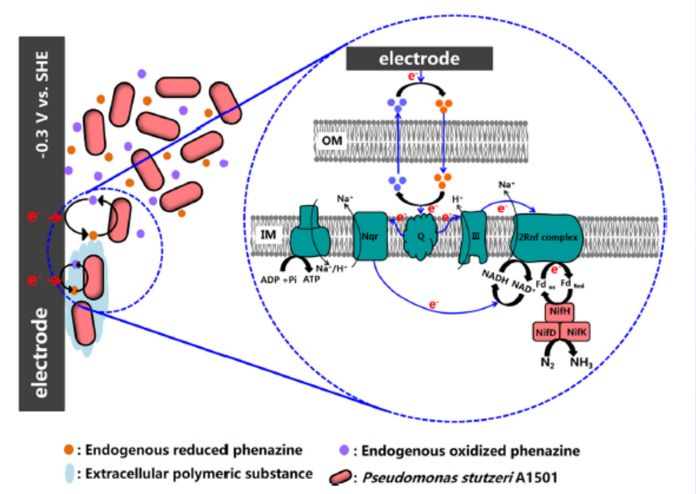
- 聚乙烯（Polyethylene, PE）广泛应用于牛奶桶、玩具及塑料薄膜等产品，2016 年PE的年产量占欧洲塑料总产量的29.8%。但是由于PE的大分子量、强疏水性以及低表面能等很难被自然降解，就导致了严重的“白色污染”。

- 目前对酶降解PE的研究较少且均不深入，针对前期工作得知的芽孢杆菌YP1能够降解PE的现象，期望找到功能基因以及关键酶，分析其分子机制。

课题背景——任丽莹_G1C

基于光/电的电子传递策略驱动微生物固氮





生物固氮是指微生物利用自身固氮酶，　在常温常压下，将N_2直接还原为NH_3 。

$$N_2 + 8H^+ + 16ATP + 8e^- \rightarrow 2\ NH_3 + H_2 + 16ADP\ /\ Pi$$

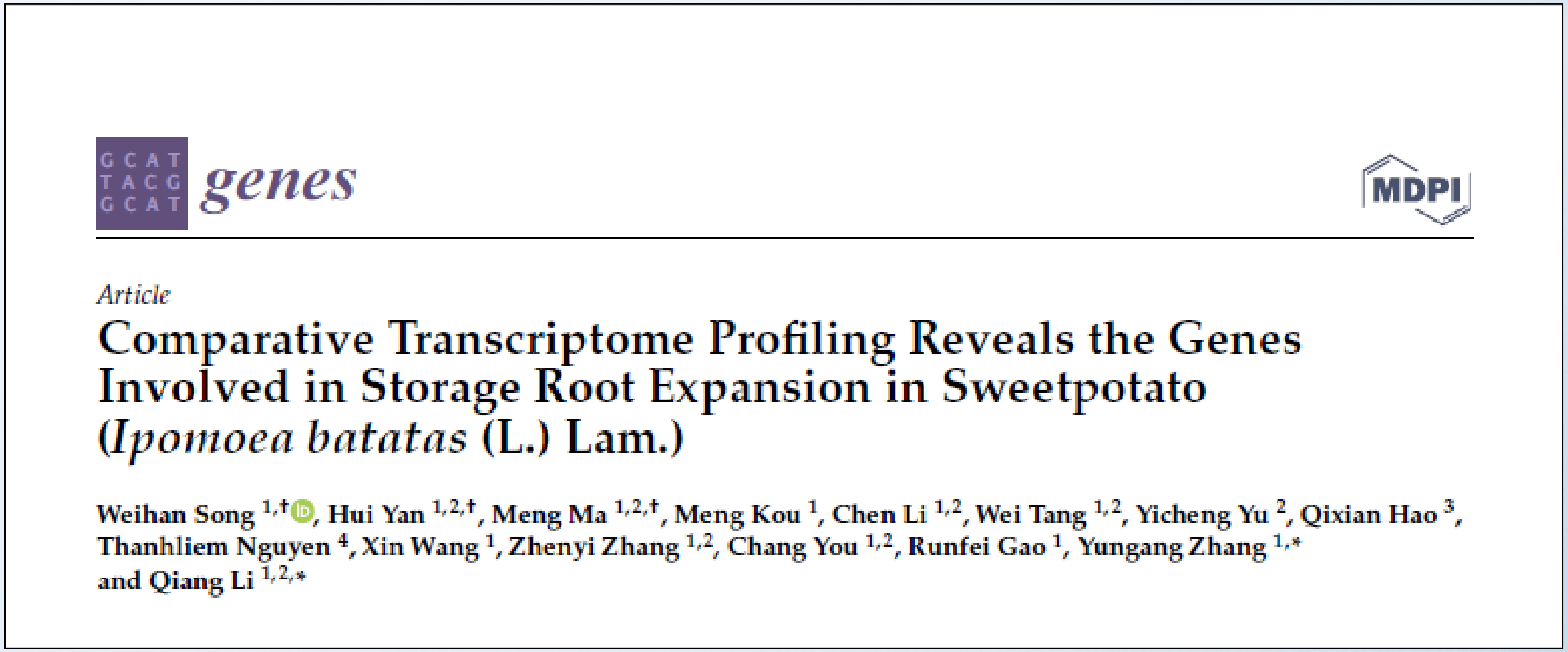
生物固氮过程中，常见的电子来源包括有机碳（如乙酸盐、丙酸盐、乳酸钠等）和还原型无机物（如Fe^{2+}，H_2，　H_2S，CO等），而电子传递速率已成为限制微生物固氮效率的决定性因素，亟需立足于天然机制对电子供给方案进行创新和优化。

课题背景——王鹏_G1D

- 牡丹（*Paeonia suffruticosa* Andrews）在分类学上属于芍药科（Paeoniaceae）芍药属（*Paeonia*）牡丹组（Sect. *Moutan*）的一种木本花卉。
- 牡丹这一名称最早出现在《神农本草经》中， 《神农本草经》将牡丹列为中品，性微寒。
- 牡丹是中国的传统名花，不仅花型花色丰富，而且集药食赏于一体，牡丹的花、叶和株型皆有较高的观赏价值，牡丹的花和其他食品加工产品具有较好的食用口感，牡丹种子可以榨油，其油富含α-亚麻酸，牡丹皮可以入药，丹皮酚是一种珍贵的药材，主治活血化瘀，消炎止痒。

二、研究相关汇报

1、转录组数据来源



genes

MDPI

Article

Comparative Transcriptome Profiling Reveals the Genes Involved in Storage Root Expansion in Sweetpotato (*Ipomoea batatas* (L.) Lam.)

Weihan Song [1,†], Hui Yan [1,2,†], Meng Ma [1,2,†], Meng Kou [1], Chen Li [1,2], Wei Tang [1,2], Yicheng Yu [2], Qixian Hao [3], Thanhliem Nguyen [4], Xin Wang [1], Zhenyi Zhang [1,2], Chang You [1,2], Runfei Gao [1], Yungang Zhang [1,*] and Qiang Li [1,2,*]

Cite: Song W, Yan H, Ma M, Kou M, Li C, et al. Comparative Transcriptome Profiling Reveals the Genes Involved in Storage Root Expansion in Sweetpotato (*Ipomoea batatas* (L.) Lam.). Genes (Basel), 2022, 13(7):1156. doi: 10.3390/genes13071156.

1、转录组数据来源

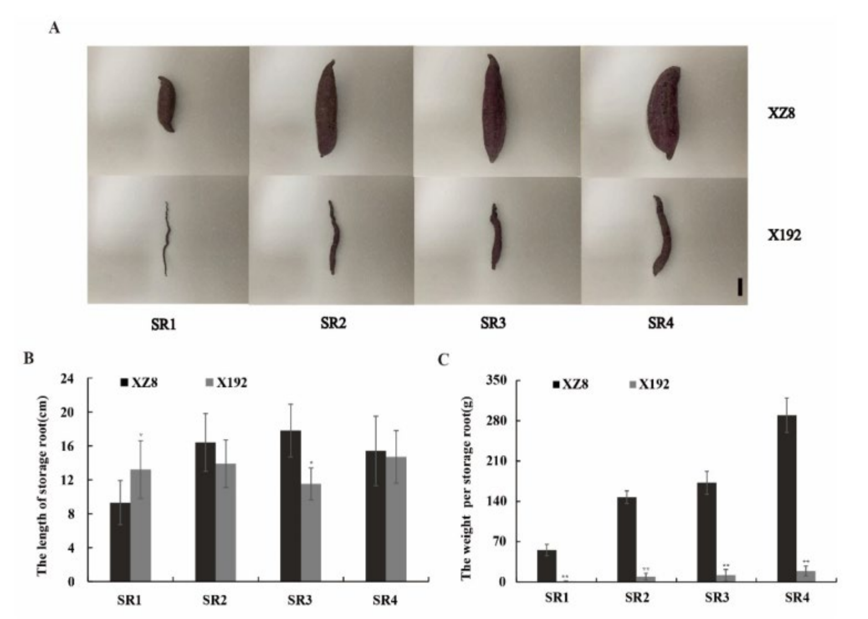


Figure 1. Phenotypic comparison of XZ8 and X192 at different SR expansion stages

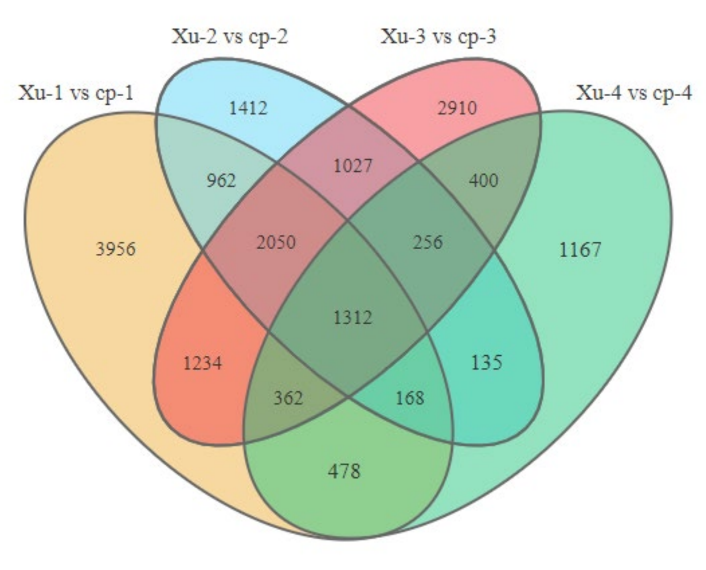


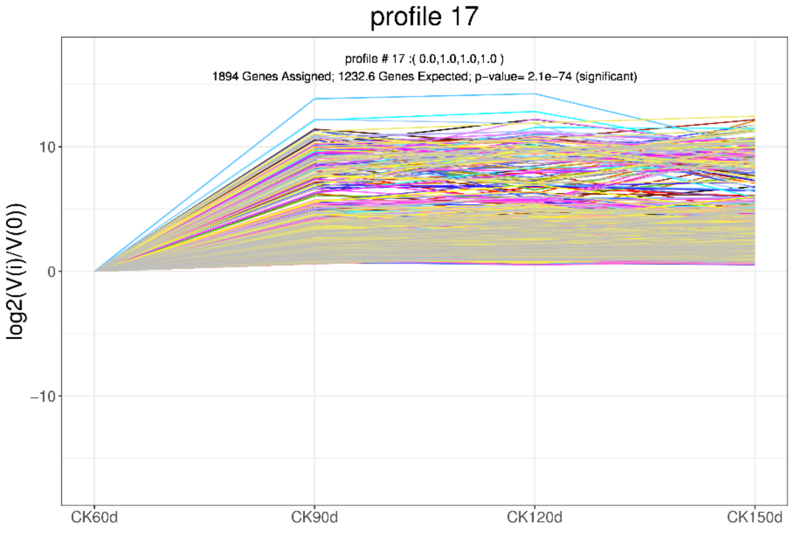
Figure 2. Venn diagram of DEGs in XZ8 and X192 in the four groups.

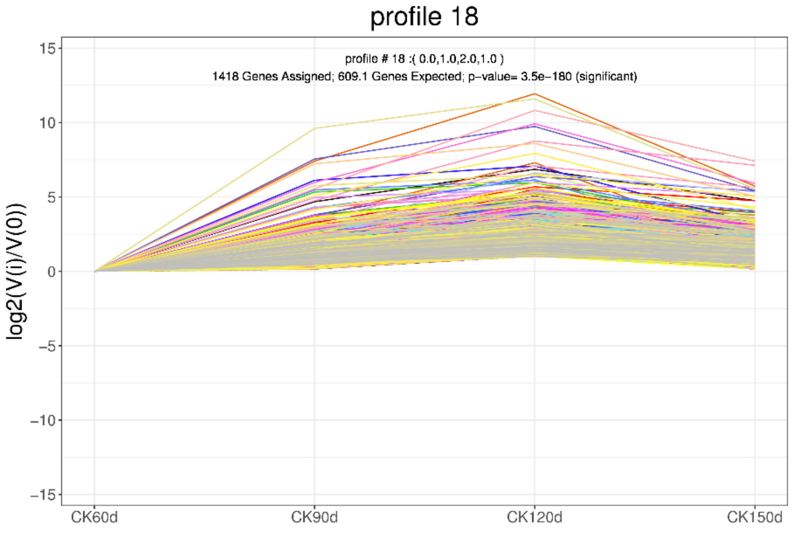
(Song *et al*., GENES, 2022)

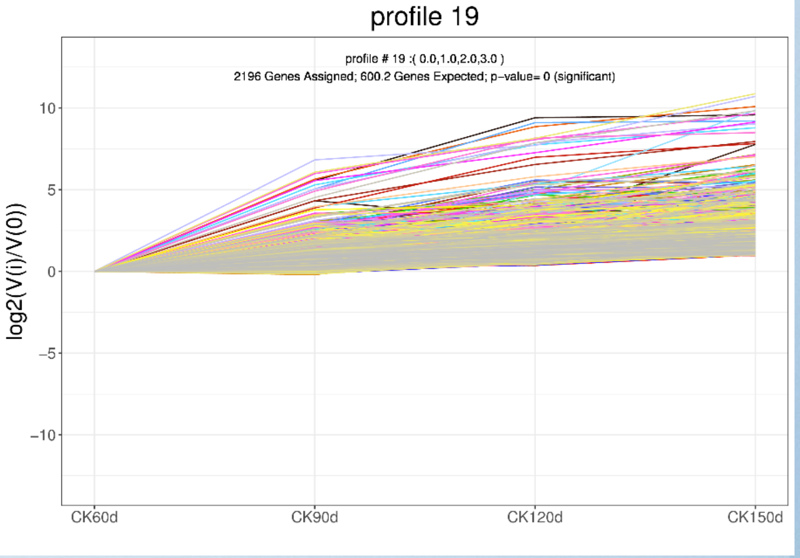
基于甘薯块根膨大转录组数据挖掘差异表达基因并进行功能分析，深入探索甘薯块根膨大机理及调控机制，利用育种及基因工程手段筛选培育高产、抗逆和优质甘薯新品种。

2、转录组数据分析

甘薯作为块根类作物，地下部膨大往往在生育前期明显受膨大基因调控作用，而生育后期的膨大发展方向基本确定，其受膨大相关主效基因调控作用不明显；基于此，筛选主要依据为生育前期显著上调表达的基因。





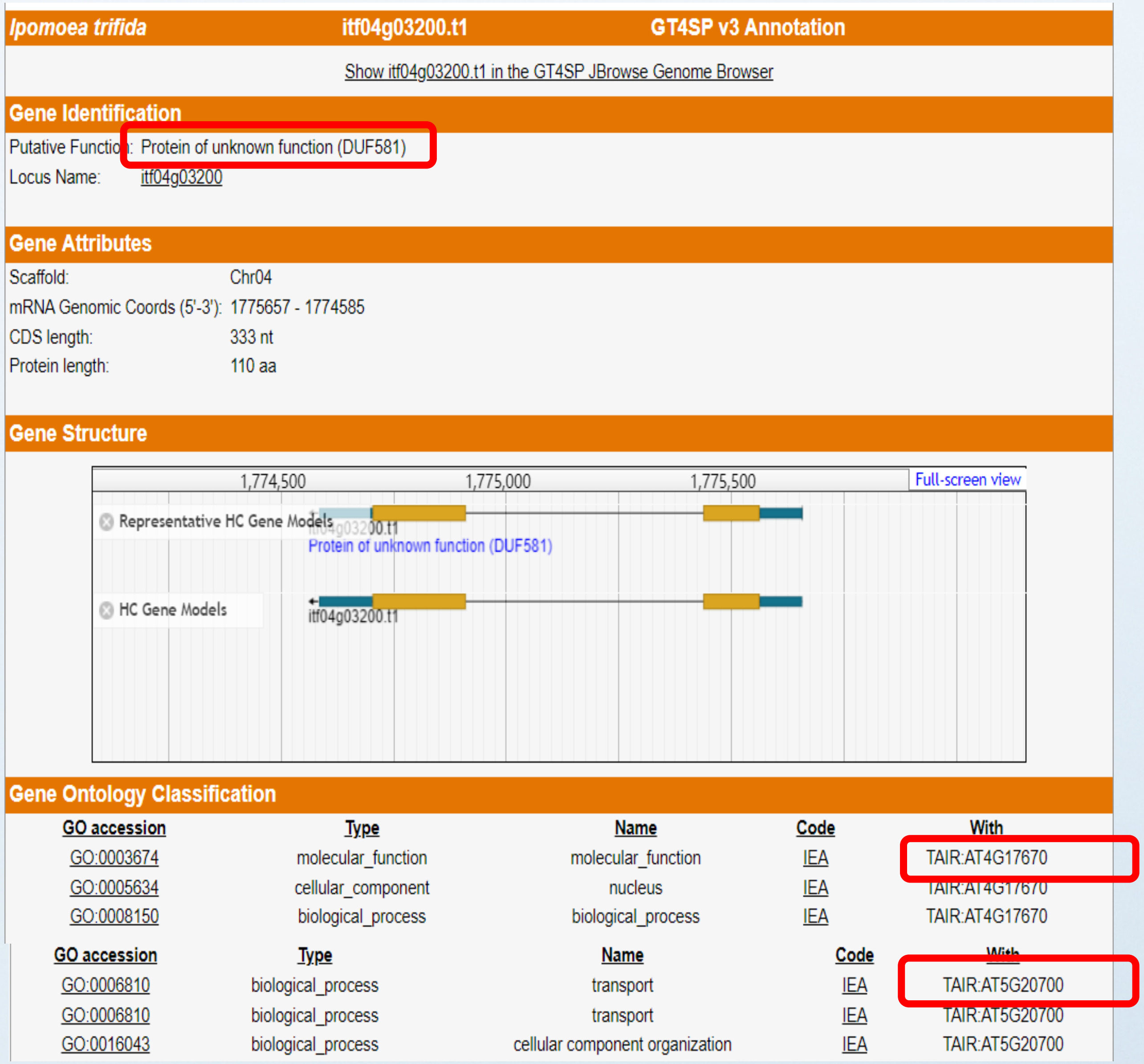


3、转录组数据分析

来源	id	60d	90d	120d	150d	90d/60d	120d/60d	150d/60d	Symbol	Description		Cloned species	Research progress
XZ8 (膨大)	G43403	6.203	208.503	376.72	105.897	33.61	60.73	17.07	--	protein MARD1-like [Cajanus cajan]	甘薯基因库中未知功能蛋白，Mediator of ABA-regulated dormancy 1(MADR1)，衰老家族蛋白(含有锌指结构DFU581)。 转录组数据库中共有8个相关基因，其中以这3个表达量相对突出。	拟南芥MARD家族基因中少部分MARD基因已克隆，其他物种暂未克隆。	拟南芥中报道该锌指结构基因调控SnRK1蛋白激酶活性以维持植物饥饿状态下生长，调控休眠。 其在块根、块茎类作物中地下部研究尚未见报道
	G8143	5.69	162.737	243.34	50.113	28.60	42.77	8.81	--	protein MARD1-like [Cajanus cajan]			
	G7856	0.667	33.907	57.27	12.24	50.84	85.86	18.35	--	protein MARD1-like [Cajanus cajan]			
	G23756	6.567	263.753	405.687	56.623	40.16	61.78	8.62	MARD1	PREDICTED: protein MARD1-like [Nicotiana attenuata]			
	G43081	4.053	126.533	184.857	28.503	31.22	45.61	7.03	MARD1	PREDICTED: protein MARD1-like [Nicotiana attenuata]			
	G24130	2.14	71.14	95.04	13.56	33.24	44.41	6.34	--	PREDICTED: protein MARD1-like [Nicotiana attenuata]			
	G39203	49.893	304.22	401.11	187.16	6.86	8.04	3.75	MARD1	PREDICTED: protein MARD1-like [Nicotiana attenuata]			
	G39152	20.61	141.82	155.59	76.17	6.88	7.55	3.70	MARD1	PREDICTED: protein MARD1-like [Nicotiana attenuata]			
X192 (不膨大)	G43403	52.53	264.52	181.67	172.63	5.04	3.46	3.29	--	protein MARD1-like [Cajanus cajan]			
	G8143	0.70	2.86	2.74	1.96	4.07	3.89	2.79	--	protein MARD1-like [Cajanus cajan]			
	G7856	1.40	6.21	3.66	1.66	4.44	2.61	1.18	--	protein MARD1-like [Cajanus cajan]			
	G23756	11.81	66.96	64.16	41.80	5.67	5.43	3.54	MARD1	PREDICTED: protein MARD1-like [Nicotiana attenuata]			
	G43081	36.36	194.02	121.76	117.33	5.34	3.35	3.23	MARD1	PREDICTED: protein MARD1-like [Nicotiana attenuata]			
	G24130	10.14	54.76	46.80	34.45	5.40	4.62	3.40	--	PREDICTED: protein MARD1-like [Nicotiana attenuata]			
	G39203	144.54	289.05	305.54	363.20	2.00	2.11	2.51	MARD1	PREDICTED: protein MARD1-like [Nicotiana attenuata]			
	G39152	26.09	55.96	53.56	50.58	2.14	2.05	1.94	MARD1	PREDICTED: protein MARD1-like [Nicotiana attenuata]			

该基因在膨大品种(XZ8)中尤其是前、中生育时期显著上调表达，但在不膨大品种(X192)中几乎不表达。

4、数据库注释信息



Ipomoea trifida **itf04g03200.t1** **GT4SP v3 Annotation**

Show itf04g03200.t1 in the GT4SP JBrowse Genome Browser

Gene Identification

Putative Function:	Protein of unknown function (DUF581)
Locus Name:	itf04g03200

Gene Attributes

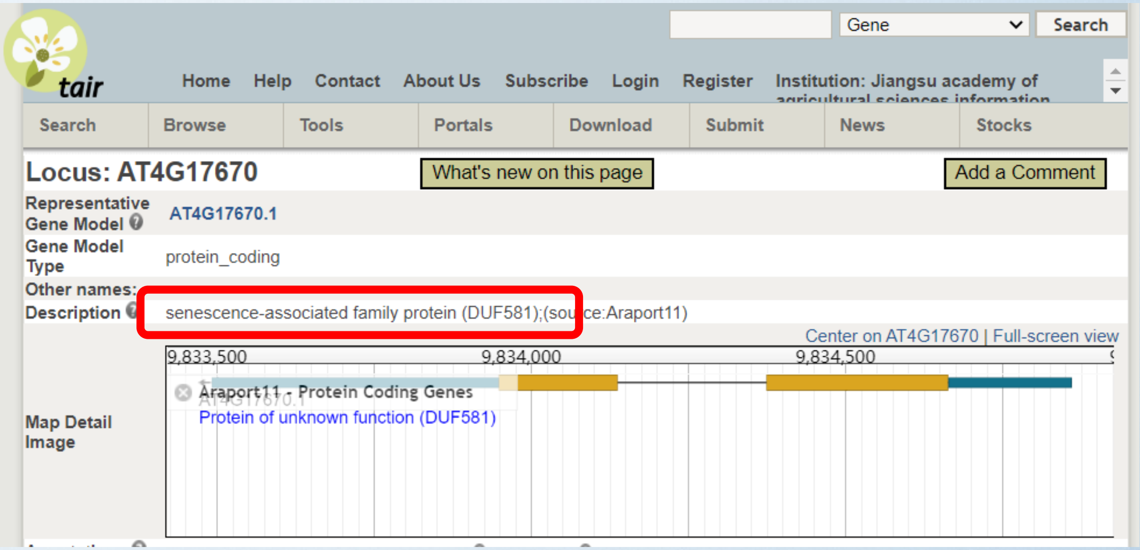
Scaffold:	Chr04
mRNA Genomic Coords (5'-3'):	1775657 - 1774585
CDS length:	333 nt
Protein length:	110 aa

Gene Structure

Gene Ontology Classification

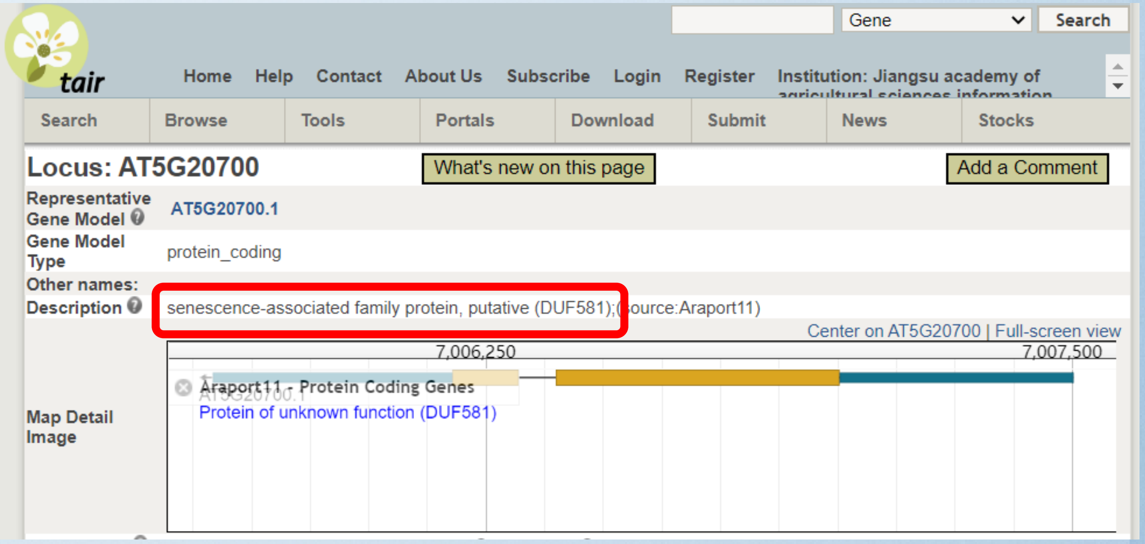
GO accession	Type	Name	Code	With
GO:0003674	molecular_function	molecular_function	IEA	TAIR:AT4G17670
GO:0005634	cellular_component	nucleus	IEA	TAIR:AT4G17670
GO:0008150	biological_process	biological_process	IEA	TAIR:AT4G17670

GO accession	Type	Name	Code	With
GO:0006810	biological_process	transport	IEA	TAIR:AT5G20700
GO:0006810	biological_process	transport	IEA	TAIR:AT5G20700
GO:0016043	biological_process	cellular component organization	IEA	TAIR:AT5G20700



Locus: AT4G17670

Representative Gene Model	AT4G17670.1
Gene Model Type	protein_coding
Other names:	
Description	senescence-associated family protein (DUF581);(sou ce:Araport11)



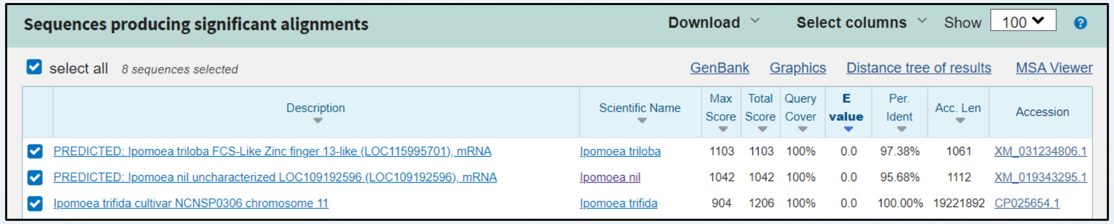
Locus: AT5G20700

Representative Gene Model	AT5G20700.1
Gene Model Type	protein_coding
Other names:	
Description	senescence-associated family protein, putative (DUF581);(ource:Araport11)

http://sweetpotato.uga.edu/index.shtml

https://www.arabidopsis.org/index.jsp

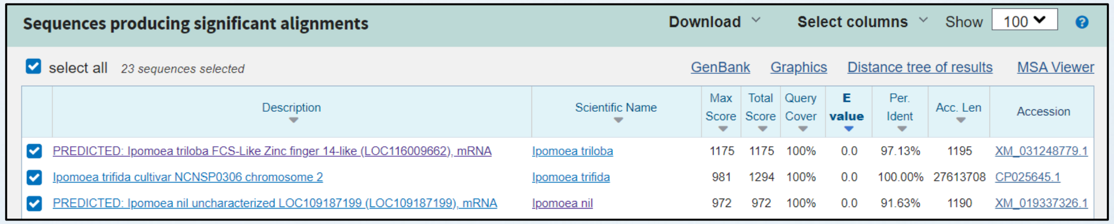
5、NCBI-Blastn、P



Sequences producing significant alignments Download Select columns Show 100

☑ select all *8 sequences selected* GenBank Graphics Distance tree of results MSA Viewer

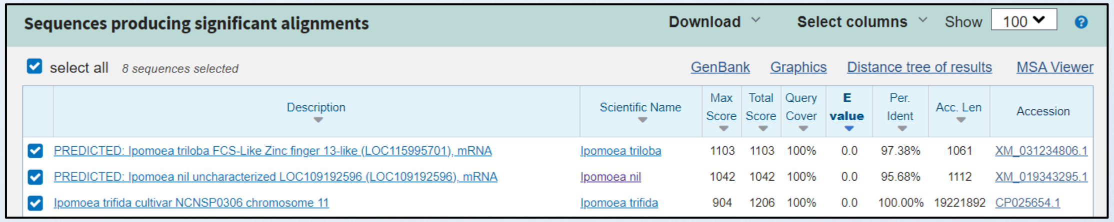
	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	PREDICTED: Ipomoea triloba FCS-Like Zinc finger 13-like (LOC115995701), mRNA	Ipomoea triloba	1103	1103	100%	0.0	97.38%	1061	XM_031234806.1
☑	PREDICTED: Ipomoea nil uncharacterized LOC109192596 (LOC109192596), mRNA	Ipomoea nil	1042	1042	100%	0.0	95.68%	1112	XM_019343295.1
☑	Ipomoea trifida cultivar NCNSP0306 chromosome 11	Ipomoea trifida	904	1206	100%	0.0	100.00%	19221892	CP025654.1



Sequences producing significant alignments Download Select columns Show 100

☑ select all *23 sequences selected* GenBank Graphics Distance tree of results MSA Viewer

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	PREDICTED: Ipomoea triloba FCS-Like Zinc finger 14-like (LOC116009662), mRNA	Ipomoea triloba	1175	1175	100%	0.0	97.13%	1195	XM_031248779.1
☑	Ipomoea trifida cultivar NCNSP0306 chromosome 2	Ipomoea trifida	981	1294	100%	0.0	100.00%	27613708	CP025645.1
☑	PREDICTED: Ipomoea nil uncharacterized LOC109187199 (LOC109187199), mRNA	Ipomoea nil	972	972	100%	0.0	91.63%	1190	XM_019337326.1

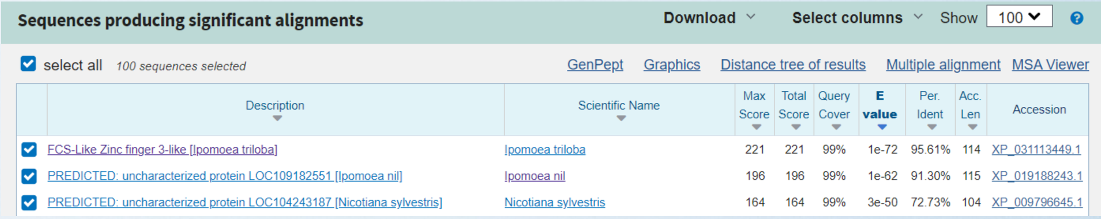


Sequences producing significant alignments Download Select columns Show 100

☑ select all *8 sequences selected* GenBank Graphics Distance tree of results MSA Viewer

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	PREDICTED: Ipomoea triloba FCS-Like Zinc finger 13-like (LOC115995701), mRNA	Ipomoea triloba	1103	1103	100%	0.0	97.38%	1061	XM_031234806.1
☑	PREDICTED: Ipomoea nil uncharacterized LOC109192596 (LOC109192596), mRNA	Ipomoea nil	1042	1042	100%	0.0	95.68%	1112	XM_019343295.1
☑	Ipomoea trifida cultivar NCNSP0306 chromosome 11	Ipomoea trifida	904	1206	100%	0.0	100.00%	19221892	CP025654.1

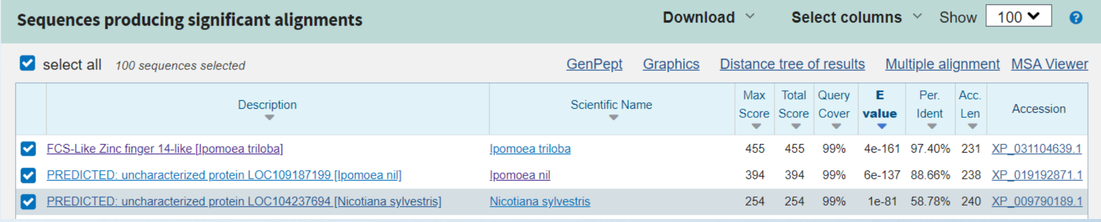
Blastn



Sequences producing significant alignments Download Select columns Show 100

☑ select all *100 sequences selected* GenPept Graphics Distance tree of results Multiple alignment MSA Viewer

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	FCS-Like Zinc finger 3-like [Ipomoea triloba]	Ipomoea triloba	221	221	99%	1e-72	95.61%	114	XP_031113449.1
☑	PREDICTED: uncharacterized protein LOC109182551 [Ipomoea nil]	Ipomoea nil	196	196	99%	1e-62	91.30%	115	XP_019188243.1
☑	PREDICTED: uncharacterized protein LOC104243187 [Nicotiana sylvestris]	Nicotiana sylvestris	164	164	99%	3e-50	72.73%	104	XP_009796645.1



Sequences producing significant alignments Download Select columns Show 100

☑ select all *100 sequences selected* GenPept Graphics Distance tree of results Multiple alignment MSA Viewer

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	FCS-Like Zinc finger 14-like [Ipomoea triloba]	Ipomoea triloba	455	455	99%	4e-161	97.40%	231	XP_031104639.1
☑	PREDICTED: uncharacterized protein LOC109187199 [Ipomoea nil]	Ipomoea nil	394	394	99%	6e-137	88.66%	238	XP_019192871.1
☑	PREDICTED: uncharacterized protein LOC104237694 [Nicotiana sylvestris]	Nicotiana sylvestris	254	254	99%	1e-81	58.78%	240	XP_009790189.1

Sequences producing significant alignments Download Select columns Show 100

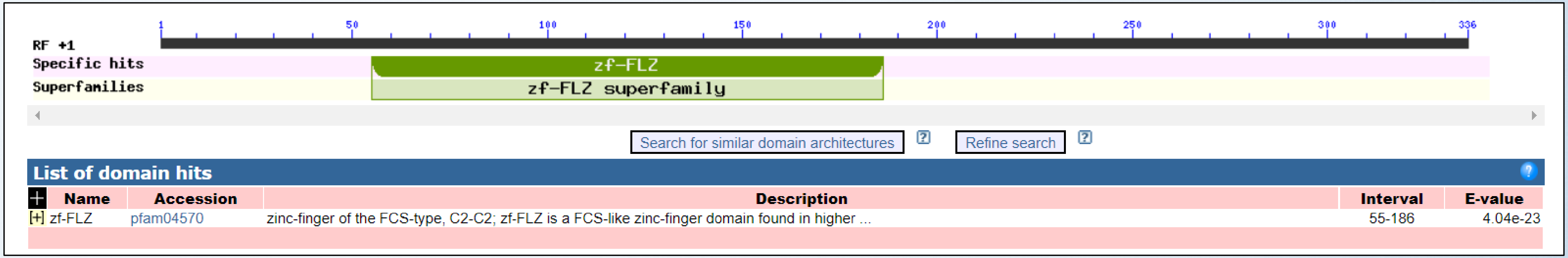
☑ select all *100 sequences selected* GenPept Graphics Distance tree of results Multiple alignment MSA Viewer

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	FCS-Like Zinc finger 13-like [Ipomoea triloba]	Ipomoea triloba	425	425	99%	1e-149	98.14%	215	XP_031090666.1
☑	PREDICTED: uncharacterized protein LOC109192596 [Ipomoea nil]	Ipomoea nil	416	416	99%	3e-146	96.28%	215	XP_019198840.1
☑	PREDICTED: protein MARD1-like [Nicotiana attenuata]	Nicotiana attenuata	272	272	99%	8e-89	61.98%	240	XP_019232387.1

Blastp

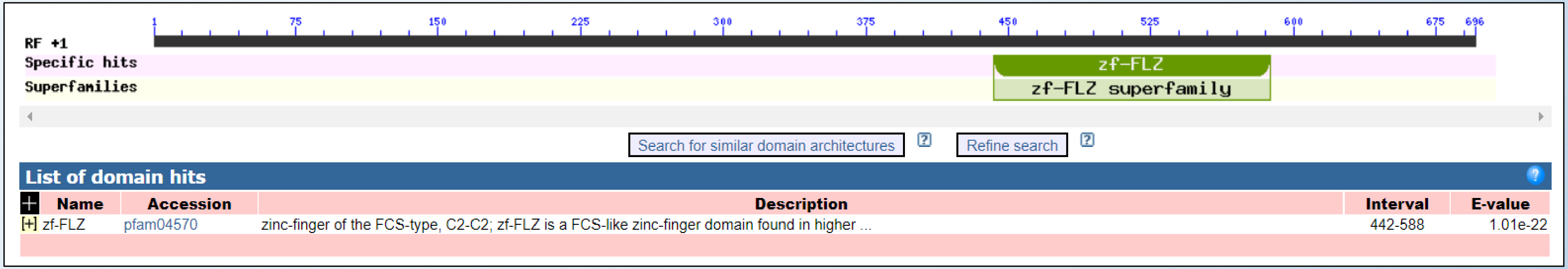
https://blast.ncbi.nlm.nih.gov/Blast.cgi

6、 NCBI-Conserved Domains



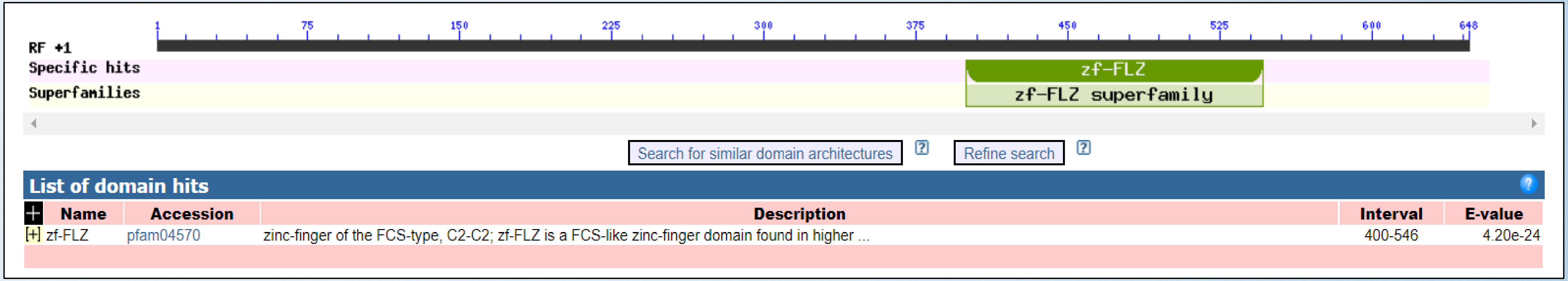
List of domain hits

Name	Accession	Description	Interval	E-value
[+] zf-FLZ	pfam04570	zinc-finger of the FCS-type, C2-C2; zf-FLZ is a FCS-like zinc-finger domain found in higher ...	55-186	4.04e-23



List of domain hits

Name	Accession	Description	Interval	E-value
[+] zf-FLZ	pfam04570	zinc-finger of the FCS-type, C2-C2; zf-FLZ is a FCS-like zinc-finger domain found in higher ...	442-588	1.01e-22

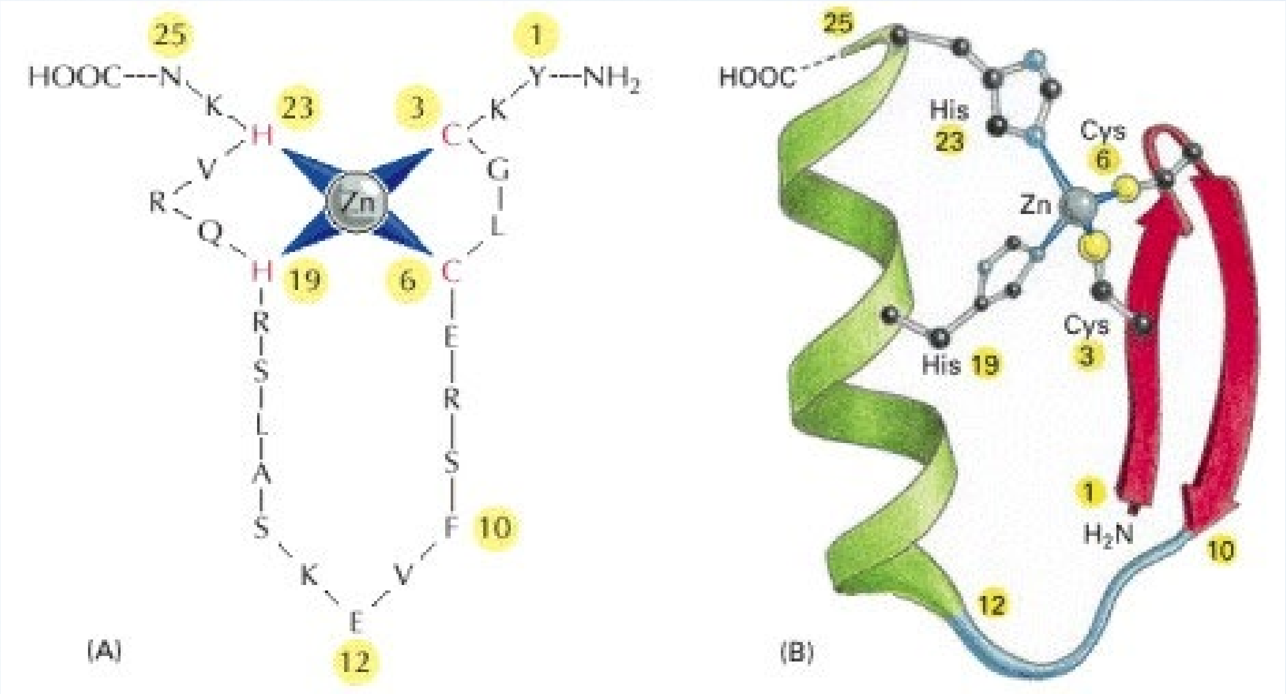


List of domain hits

Name	Accession	Description	Interval	E-value
[+] zf-FLZ	pfam04570	zinc-finger of the FCS-type, C2-C2; zf-FLZ is a FCS-like zinc-finger domain found in higher ...	400-546	4.20e-24

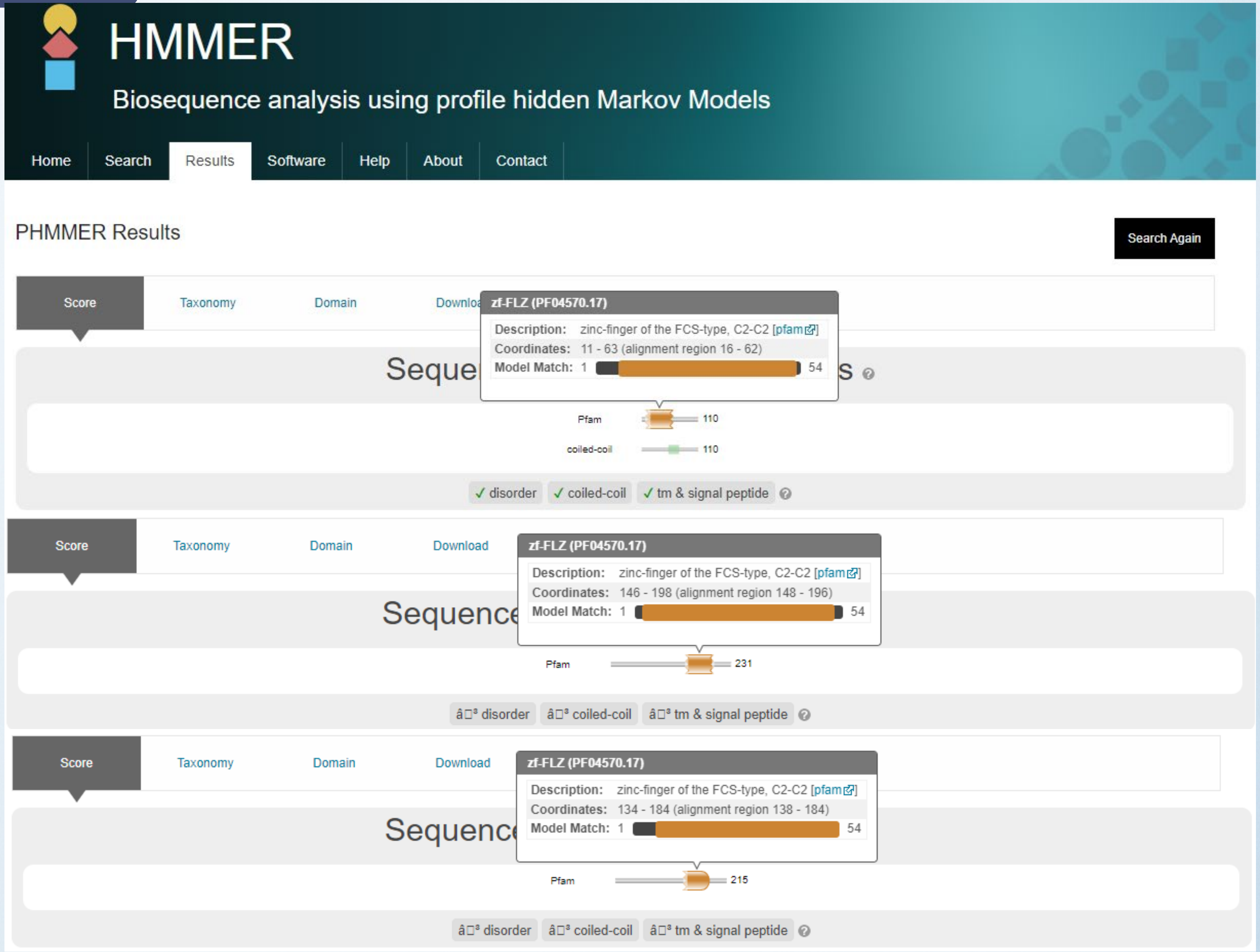
7、锌指结构

何为锌指结构?



锌指结构是由一个含有大约30个氨基酸的环和一个与环上的4个Cys(半胱氨酸) 或2个Cys和2个His(组氨酸)配位的Zn2+构成，由一个α-螺旋和两个反平行的β-折叠三个肽段组成，形似手指，具有结合锌离子的功能，具有锌指结构的蛋白大多都是与基因表达的调控相关的功能蛋白且都能与DNA或RNA结合.

8、 Markov Models



PHMMER Results

zf-FLZ (PF04570.17)

Description: zinc-finger of the FCS-type, C2-C2 [pfam]

Coordinates: 11 - 63 (alignment region 16 - 62)

Model Match: 1 54

zf-FLZ (PF04570.17)

Description: zinc-finger of the FCS-type, C2-C2 [pfam]

Coordinates: 146 - 198 (alignment region 148 - 196)

Model Match: 1 54

zf-FLZ (PF04570.17)

Description: zinc-finger of the FCS-type, C2-C2 [pfam]

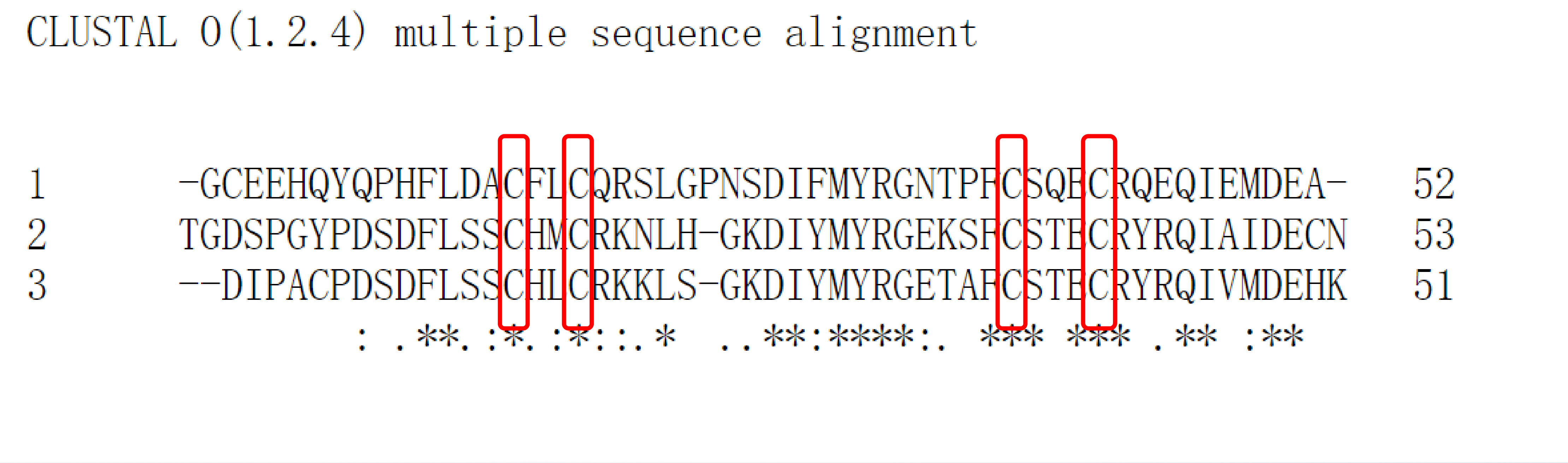
Coordinates: 134 - 184 (alignment region 138 - 184)

Model Match: 1 54

(https://www.ebi.ac.uk/Tools/hmmer/)

9、 Clustal Omega-多序列比对

基于EMBL-EBI的Clustal Omega进行多序列比对，结果如下：

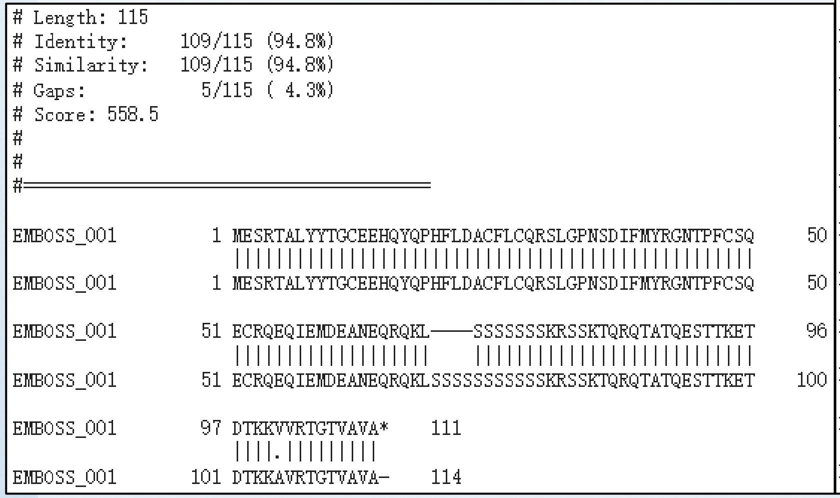


```
CLUSTAL O(1.2.4) multiple sequence alignment


1       -GCEEHQYQPHFLDACFLCQRSLGPNSDIFMYRGNTPFCSQECRQEQIEMDEA-   52
2       TGDSPGYPDSDFLSSCHMCRKNLH-GKDIYMYRGEKSFCSTECRYRQIAIDECN   53
3       --DIPACPDSDFLSSCHLCRKKLS-GKDIYMYRGETAFCSTECRYRQIVMDEHK   51
                : .**.:*.:*::.*  ..**:****:. *** *** .** :**
```

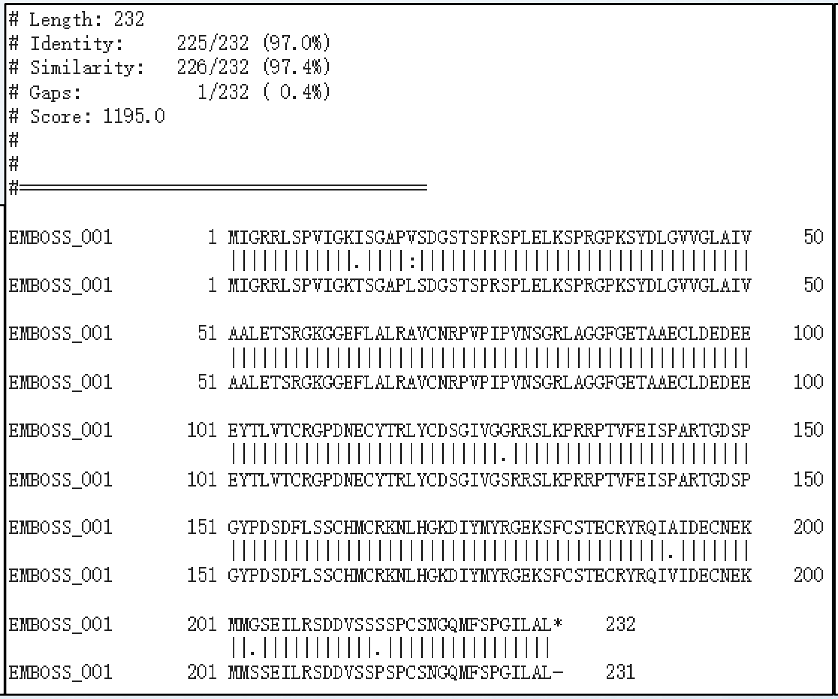
https://www.ebi.ac.uk/Tools/msa/clustalo/

9、EMBOSS Needle双序列比对

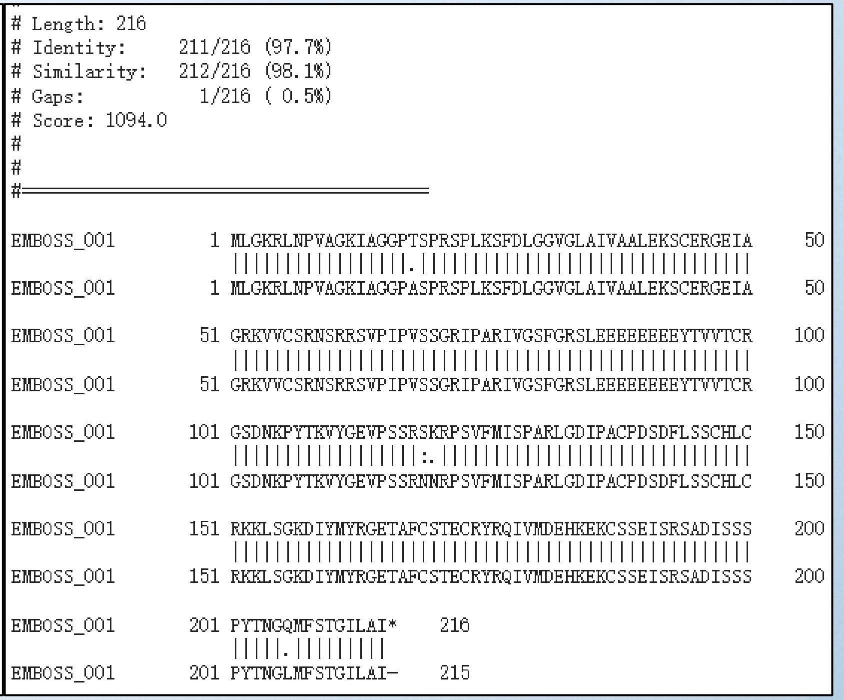
MATRIX: BLOSUM62
GAP OPEN: 10
GAP EXTEND: 0.5



```
# Length: 115
# Identity:     109/115 (94.8%)
# Similarity:   109/115 (94.8%)
# Gaps:           5/115 ( 4.3%)
# Score: 558.5
#
#
#=======================================

EMBOSS_001         1 MESRTALYYTGCEEHQYQPHFLDACFLCQRSLGPNSDIFMYRGNTPFCSQ     50
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001         1 MESRTALYYTGCEEHQYQPHFLDACFLCQRSLGPNSDIFMYRGNTPFCSQ     50

EMBOSS_001        51 ECRQEQIEMDEANEQRQKL----SSSSSSSKRSSKTQRQTATQESTTKET     96
                     |||||||||||||||||||    |||||||||||||||||||||||||||
EMBOSS_001        51 ECRQEQIEMDEANEQRQKLSSSSSSSSSSSKRSSKTQRQTATQESTTKET    100

EMBOSS_001        97 DTKKVVRTGTVAVA*    111
                     ||||.|||||||||
EMBOSS_001       101 DTKKAVRTGTVAVA-    114
```



```
# Length: 232
# Identity:     225/232 (97.0%)
# Similarity:   226/232 (97.4%)
# Gaps:           1/232 ( 0.4%)
# Score: 1195.0
#
#
#=======================================

EMBOSS_001         1 MIGRRLSPVIGKISGAPVSDGSTSPRSPLELKSPRGPKSYDLGVVGLAIV     50
                     ||||||||||||.||||:||||||||||||||||||||||||||||||||
EMBOSS_001         1 MIGRRLSPVIGKTSGAPLSDGSTSPRSPLELKSPRGPKSYDLGVVGLAIV     50

EMBOSS_001        51 AALETSRGKGGEFLALRAVCNRPVPIPVNSGRLAGGFGETAAECLDEDEE    100
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001        51 AALETSRGKGGEFLALRAVCNRPVPIPVNSGRLAGGFGETAAECLDEDEE    100

EMBOSS_001       101 EYTLVTCRGPDNECYTRLYCDSGIVGGRRSLKPRRPTVFEISPARTGDSP    150
                     ||||||||||||||||||||||||||.|||||||||||||||||||||||
EMBOSS_001       101 EYTLVTCRGPDNECYTRLYCDSGIVGSRRSLKPRRPTVFEISPARTGDSP    150

EMBOSS_001       151 GYPDSDFLSSCHMCRKNLHGKDIYMYRGEKSFCSTECRYRQIAIDECNEK    200
                     ||||||||||||||||||||||||||||||||||||||||||.|||||||
EMBOSS_001       151 GYPDSDFLSSCHMCRKNLHGKDIYMYRGEKSFCSTECRYRQIVIDECNEK    200

EMBOSS_001       201 MMGSEILRSDDVSSSSPCSNGQMFSPGILAL*    232
                     ||.|||||||||||.||||||||||||||||
EMBOSS_001       201 MMSSEILRSDDVSSPSPCSNGQMFSPGILAL-    231
```

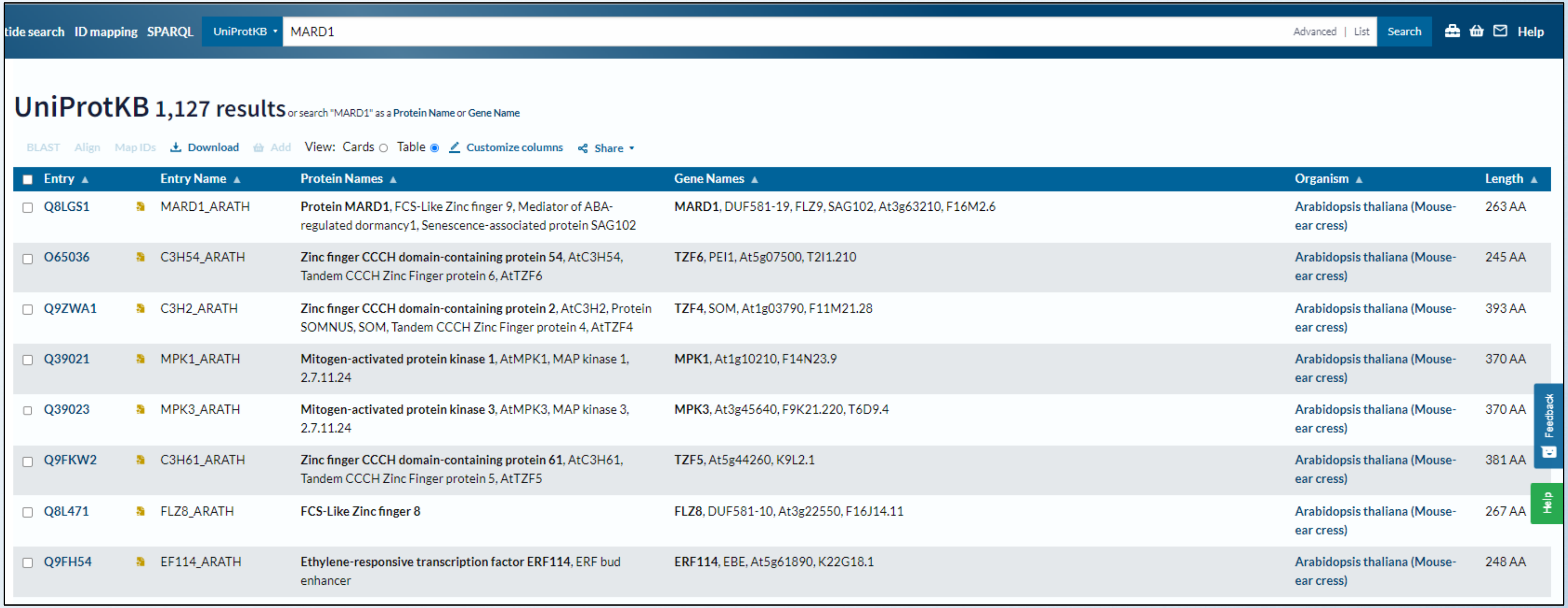


```
# Length: 216
# Identity:     211/216 (97.7%)
# Similarity:   212/216 (98.1%)
# Gaps:           1/216 ( 0.5%)
# Score: 1094.0
#
#
#=======================================

EMBOSS_001         1 MLGKRLNPVAGKIAGGPTSPRSPLKSFDLGGVGLAIVAALEKSCERGEIA     50
                     |||||||||||||||||.||||||||||||||||||||||||||||||||
EMBOSS_001         1 MLGKRLNPVAGKIAGGPASPRSPLKSFDLGGVGLAIVAALEKSCERGEIA     50

EMBOSS_001        51 GRKVVCSRNSRRSVPIPVSSGRIPARIVGSFGRSLEEEEEEEEYTVVTCR    100
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001        51 GRKVVCSRNSRRSVPIPVSSGRIPARIVGSFGRSLEEEEEEEEYTVVTCR    100

EMBOSS_001       101 GSDNKPYTKVYGEVPSSRSKRPSVFMISPARLGDIPACPDSDFLSSCHLC    150
                     ||||||||||||||||||:.||||||||||||||||||||||||||||||
EMBOSS_001       101 GSDNKPYTKVYGEVPSSRNNRPSVFMISPARLGDIPACPDSDFLSSCHLC    150

EMBOSS_001       151 RKKLSGKDIYMYRGETAFCSTECRYRQIVMDEHKEKCSSEISRSADISSS    200
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       151 RKKLSGKDIYMYRGETAFCSTECRYRQIVMDEHKEKCSSEISRSADISSS    200

EMBOSS_001       201 PYTNGQMFSTGILAI*    216
                     |||||.|||||||||
EMBOSS_001       201 PYTNGLMFSTGILAI-    215
```

比对结果显示，三个目的蛋白和甘薯野生种中注释的锌指结构蛋白绝大部分位点形同，相同位点用竖杠“|”表示，不同位点用句号“.”表示，相似位点用“：”表示；其中第一个蛋白在70-73位点有四个“S”丝氨酸的缺失，但由于该位点为无序区域，所以该位点的缺失无关紧要。

（https://www.ebi.ac.uk/Tools/psa/emboss_needle/）

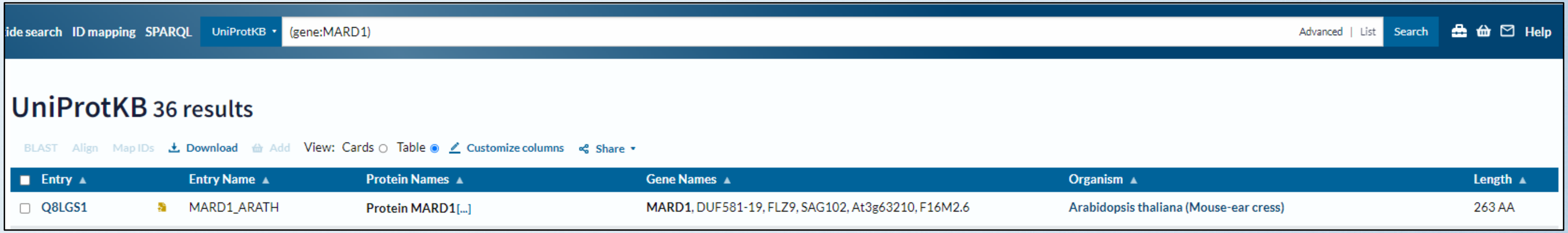
10、UniProt注释



MARD1

UniProtKB 1,127 results or search "MARD1" as a Protein Name or Gene Name

Entry	Entry Name	Protein Names	Gene Names	Organism	Length
Q8LGS1	MARD1_ARATH	**Protein MARD1**, FCS-Like Zinc finger 9, Mediator of ABA-regulated dormancy1, Senescence-associated protein SAG102	**MARD1**, DUF581-19, FLZ9, SAG102, At3g63210, F16M2.6	Arabidopsis thaliana (Mouse-ear cress)	263 AA
O65036	C3H54_ARATH	**Zinc finger CCCH domain-containing protein 54**, AtC3H54, Tandem CCCH Zinc Finger protein 6, AtTZF6	**TZF6**, PEI1, At5g07500, T2I1.210	Arabidopsis thaliana (Mouse-ear cress)	245 AA
Q9ZWA1	C3H2_ARATH	**Zinc finger CCCH domain-containing protein 2**, AtC3H2, Protein SOMNUS, SOM, Tandem CCCH Zinc Finger protein 4, AtTZF4	**TZF4**, SOM, At1g03790, F11M21.28	Arabidopsis thaliana (Mouse-ear cress)	393 AA
Q39021	MPK1_ARATH	**Mitogen-activated protein kinase 1**, AtMPK1, MAP kinase 1, 2.7.11.24	**MPK1**, At1g10210, F14N23.9	Arabidopsis thaliana (Mouse-ear cress)	370 AA
Q39023	MPK3_ARATH	**Mitogen-activated protein kinase 3**, AtMPK3, MAP kinase 3, 2.7.11.24	**MPK3**, At3g45640, F9K21.220, T6D9.4	Arabidopsis thaliana (Mouse-ear cress)	370 AA
Q9FKW2	C3H61_ARATH	**Zinc finger CCCH domain-containing protein 61**, AtC3H61, Tandem CCCH Zinc Finger protein 5, AtTZF5	**TZF5**, At5g44260, K9L2.1	Arabidopsis thaliana (Mouse-ear cress)	381 AA
Q8L471	FLZ8_ARATH	**FCS-Like Zinc finger 8**	**FLZ8**, DUF581-10, At3g22550, F16J14.11	Arabidopsis thaliana (Mouse-ear cress)	267 AA
Q9FH54	EF114_ARATH	**Ethylene-responsive transcription factor ERF114**, ERF bud enhancer	**ERF114**, EBE, At5g61890, K22G18.1	Arabidopsis thaliana (Mouse-ear cress)	248 AA

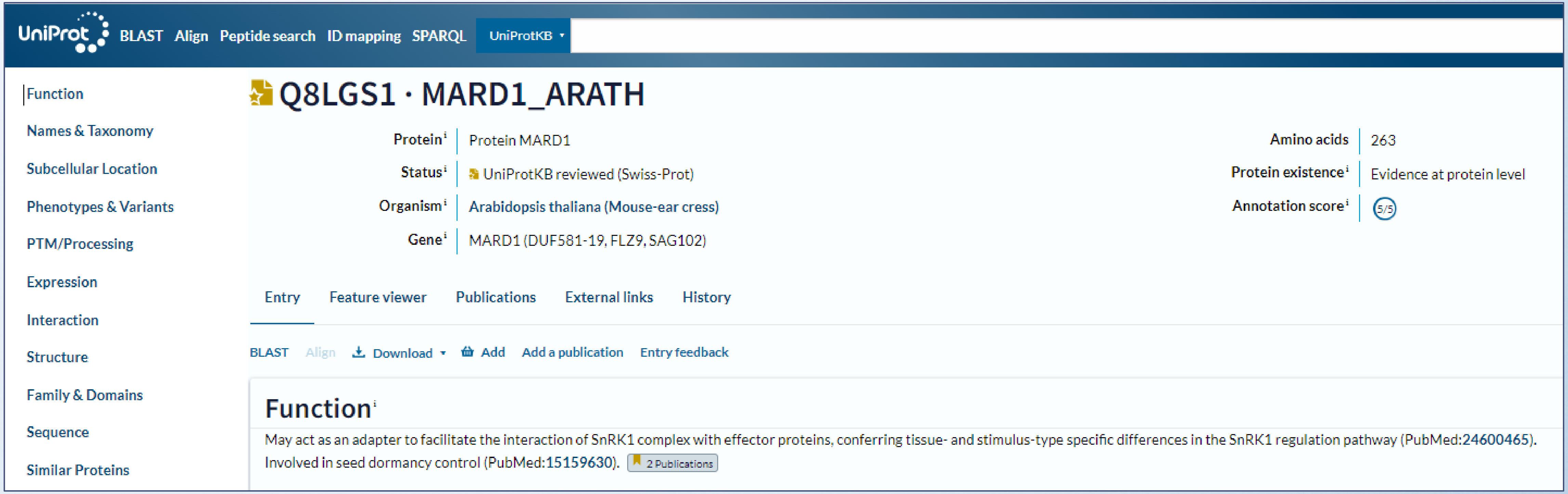


(gene:MARD1)

UniProtKB 36 results

Entry	Entry Name	Protein Names	Gene Names	Organism	Length
Q8LGS1	MARD1_ARATH	**Protein MARD1** [...]	**MARD1**, DUF581-19, FLZ9, SAG102, At3g63210, F16M2.6	Arabidopsis thaliana (Mouse-ear cress)	263 AA

10、UniProt注释



Q8LGS1 · MARD1_ARATH

Protein Protein MARD1

Status UniProtKB reviewed (Swiss-Prot)

Organism Arabidopsis thaliana (Mouse-ear cress)

Gene MARD1 (DUF581-19, FLZ9, SAG102)

Amino acids 263

Protein existence Evidence at protein level

Annotation score 5/5

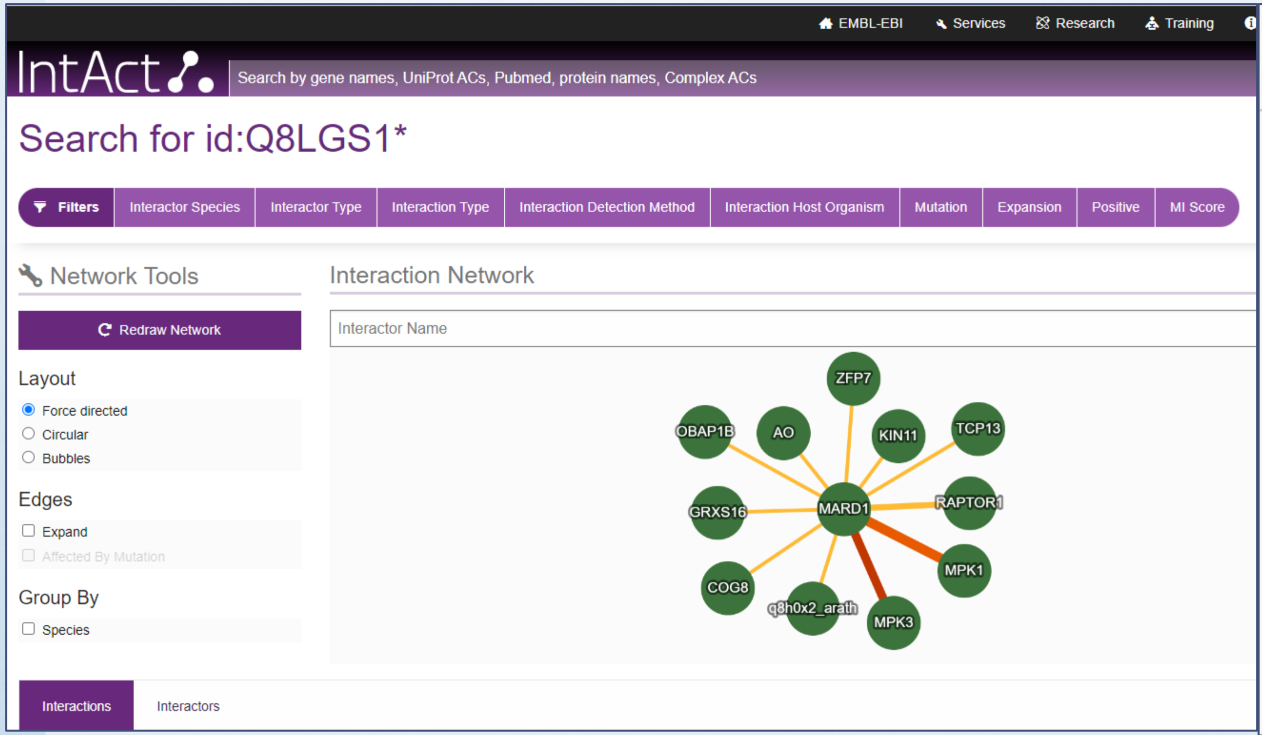
Function

May act as an adapter to facilitate the interaction of SnRK1 complex with effector proteins, conferring tissue- and stimulus-type specific differences in the SnRK1 regulation pathway (PubMed:24600465). Involved in seed dormancy control (PubMed:15159630). 2 Publications

（https://www.uniprot.org/uniprotkb/Q8LGS1/entry）

11、互作蛋白

Protein-protein interaction databases

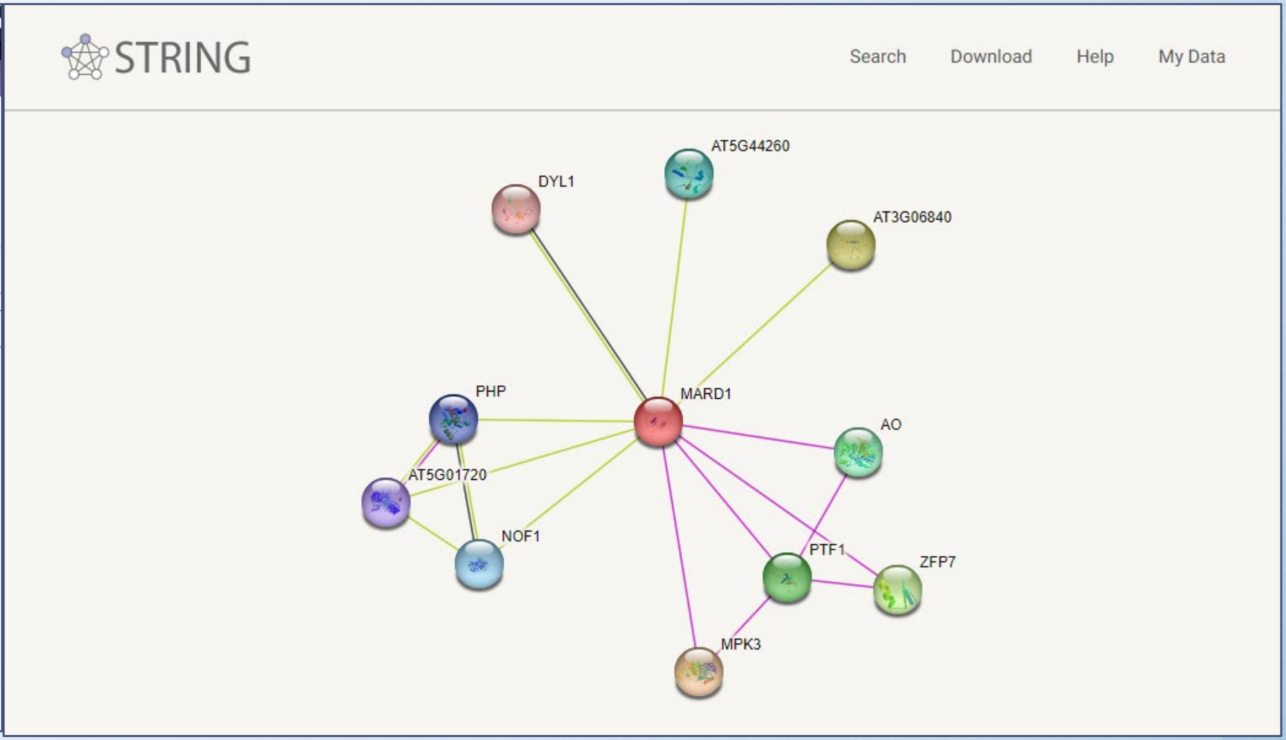


Search for id:Q8LGS1*

Layout
● Force directed
○ Circular
○ Bubbles

Edges
☐ Expand
☐ Affected By Mutation

Group By
☐ Species



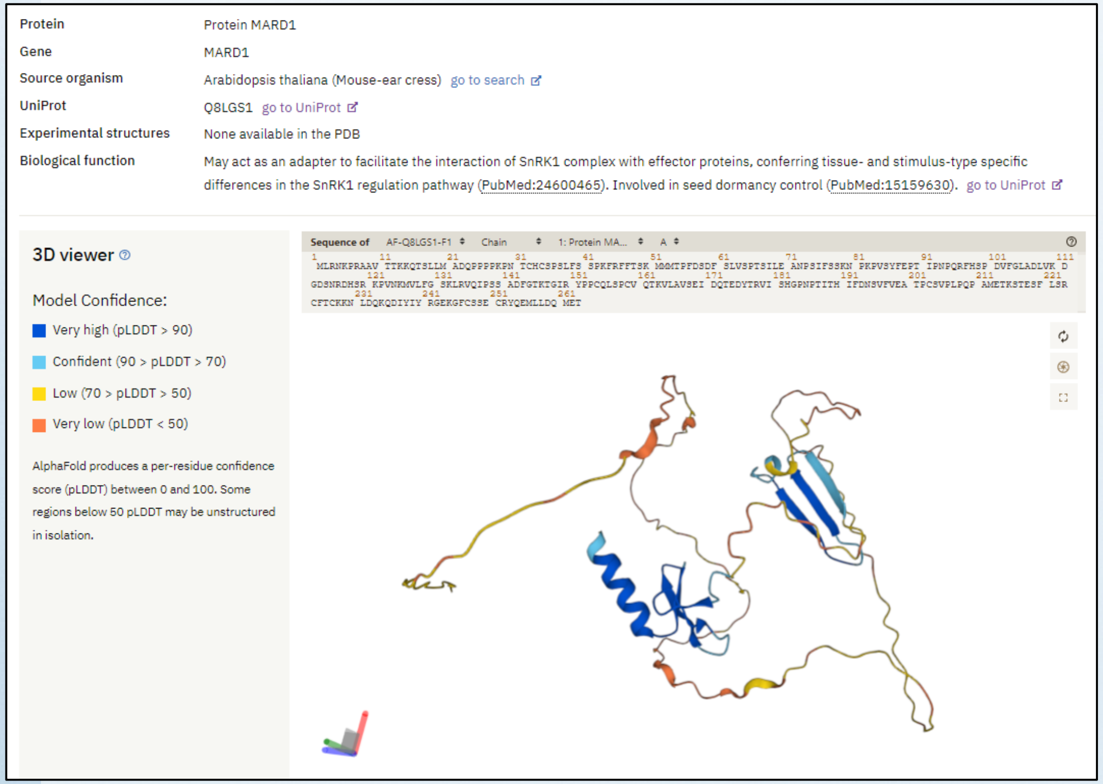
（https://www.ebi.ac.uk/intact/home）

（https://cn.string-db.org/cgi/input?sessionId=boLZZckqAtMH&input_page_show_search=on）

12、蛋白结构

3D structure databases

AlphaFoldDB



Protein	Protein MARD1
Gene	MARD1
Source organism	Arabidopsis thaliana (Mouse-ear cress) go to search
UniProt	Q8LGS1 go to UniProt
Experimental structures	None available in the PDB
Biological function	May act as an adapter to facilitate the interaction of SnRK1 complex with effector proteins, conferring tissue- and stimulus-type specific differences in the SnRK1 regulation pathway (PubMed:24600465). Involved in seed dormancy control (PubMed:15159630). go to UniProt

3D viewer

Model Confidence:

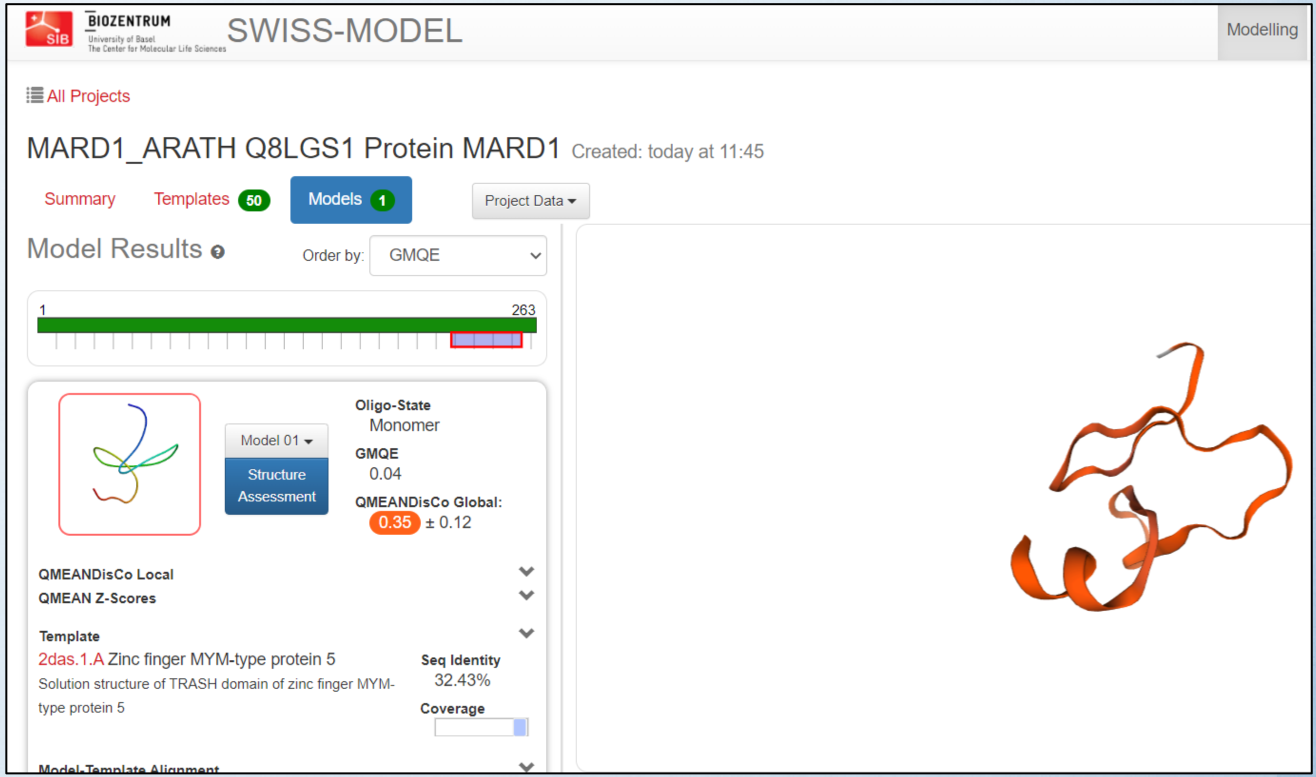
- 🟦 Very high (pLDDT > 90)
- 🟦 Confident (90 > pLDDT > 70)
- 🟨 Low (70 > pLDDT > 50)
- 🟧 Very low (pLDDT < 50)

AlphaFold produces a per-residue confidence score (pLDDT) between 0 and 100. Some regions below 50 pLDDT may be unstructured in isolation.

```
1          11         21         31         41         51         61         71         81         91         101        111
 MLRNKPRAAV TTKKQTSLLM ADQPPPPKPN TCHCSPSLFS SPKFRFFTSK MMMTPFDSDF SLVSPTSILE ANPSIFSSKN PKPVSYFEPT IPNPQRFHSP DVFGLADLVK D
        121        131        141        151        161        171        181        191        201        211        221
GDSNRDHSR KPVNKMVLFG SKLRVQIPSS ADFGTKTGIR YPPCQLSPCV QTKVLAVSEI DQTEDYTRVI SHGPNPTITH IFDNSVFVEA TPCSVPLPQP AMETKSTESF LSR
      231        241        251        261
CFTCKKN LDQKQDIYIY RGEKGFCSSE CRYQEMLLDQ MET
```

（ https://alphafold.ebi.ac.uk/ ）

SWISS-MODEL



MARD1_ARATH Q8LGS1 Protein MARD1 Created: today at 11:45

Oligo-State
Monomer

GMQE
0.04

QMEANDisCo Global:
0.35 ± 0.12

QMEANDisCo Local

QMEAN Z-Scores

Template

2das.1.A Zinc finger MYM-type protein 5

Solution structure of TRASH domain of zinc finger MYM-type protein 5

Seq Identity
32.43%

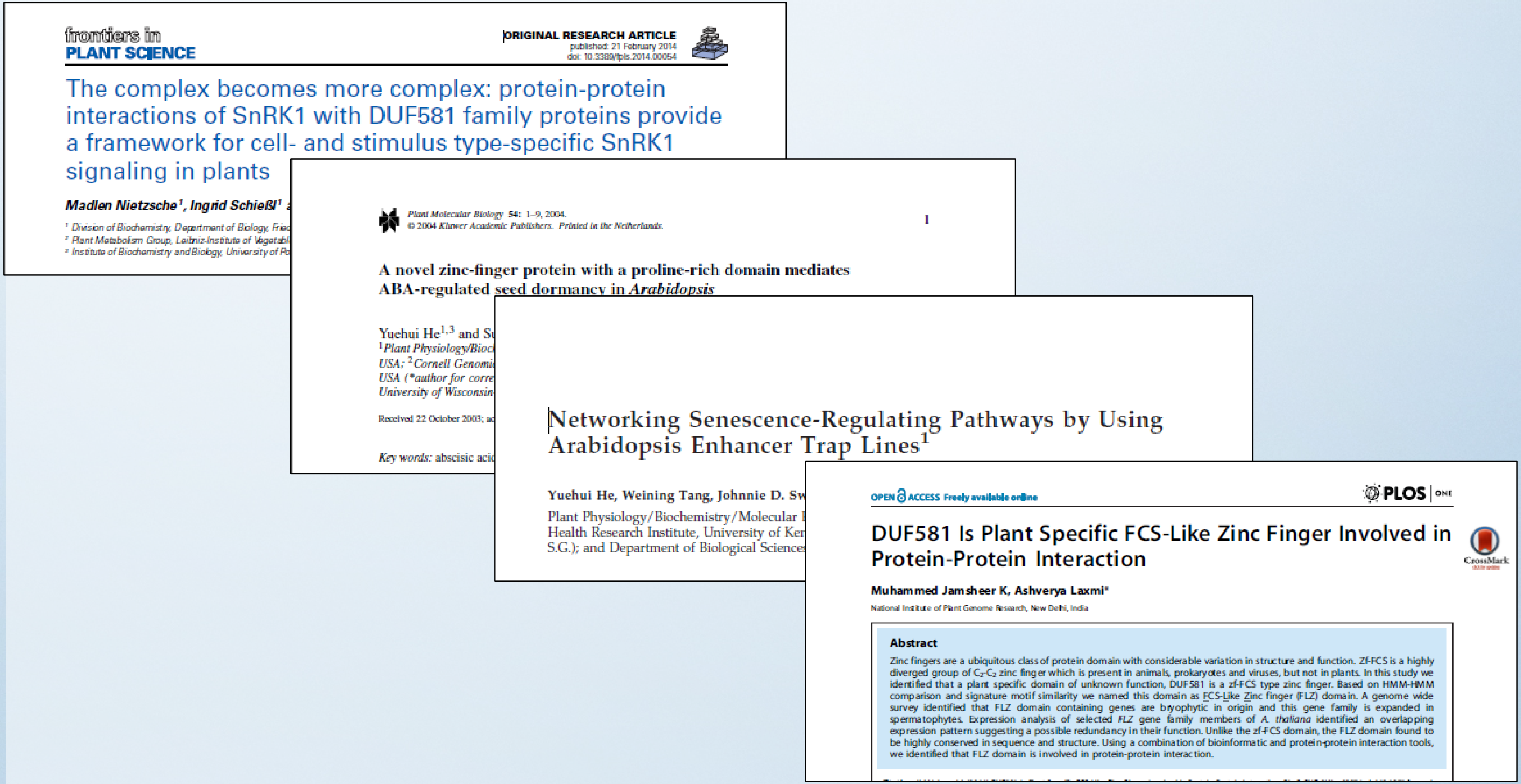
Coverage

Model-Template Alignment

（ https://swissmodel.expasy.org/ ）

13、相关文献

Pulications



frontiers in
PLANT SCIENCE

ORIGINAL RESEARCH ARTICLE
published: 21 February 2014
doi: 10.3389/fpls.2014.00054

The complex becomes more complex: protein-protein interactions of SnRK1 with DUF581 family proteins provide a framework for cell- and stimulus type-specific SnRK1 signaling in plants

Madlen Nietzsche[1], **Ingrid Schießl**[1] a
[1] Division of Biochemistry, Department of Biology, Fried
[2] Plant Metabolism Group, Leibniz-Institute of Vegetabl
[3] Institute of Biochemistry and Biology, University of Po

Plant Molecular Biology **54:** 1–9, 2004.
© 2004 *Kluwer Academic Publishers. Printed in the Netherlands.*

A novel zinc-finger protein with a proline-rich domain mediates ABA-regulated seed dormancy in *Arabidopsis*

Yuehui He[1,3] and Su
[1] *Plant Physiology/Bioc*
USA; [2] *Cornell Genomic*
*USA (*author for corre*
University of Wisconsin

Received 22 October 2003; ac

Key words: abscisic acid

Networking Senescence-Regulating Pathways by Using Arabidopsis Enhancer Trap Lines[1]

Yuehui He, Weining Tang, Johnnie D. Sw

Plant Physiology/Biochemistry/Molecular B
Health Research Institute, University of Ken
S.G.); and Department of Biological Science

OPEN ACCESS Freely available online

PLOS | ONE

DUF581 Is Plant Specific FCS-Like Zinc Finger Involved in Protein-Protein Interaction

Muhammed Jamsheer K, Ashverya Laxmi*

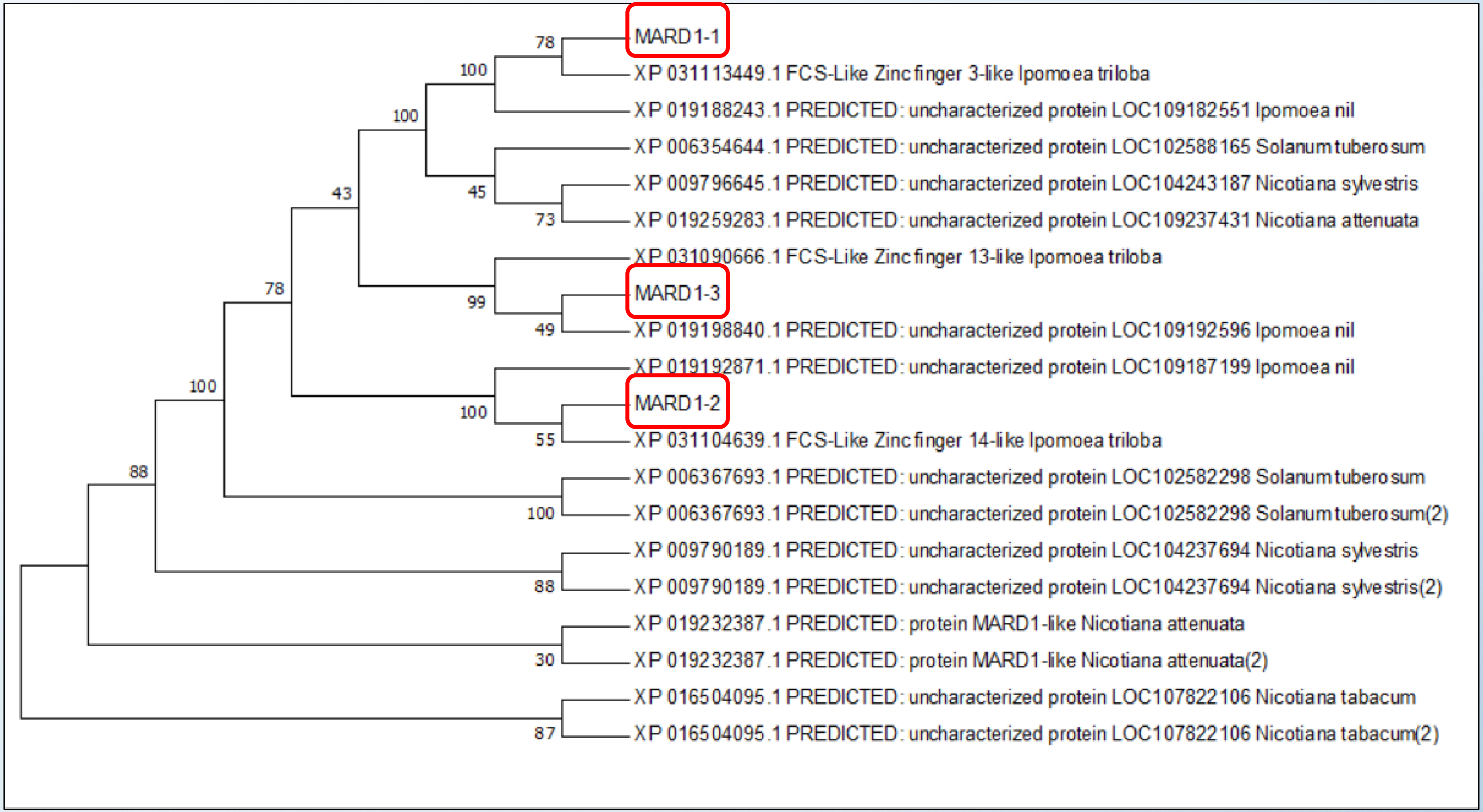
National Institute of Plant Genome Research, New Delhi, India

Abstract

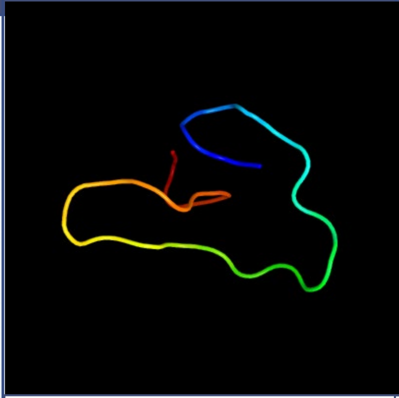
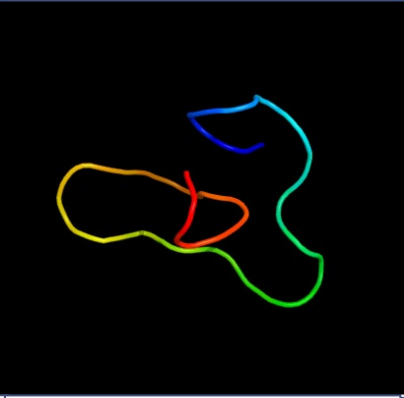
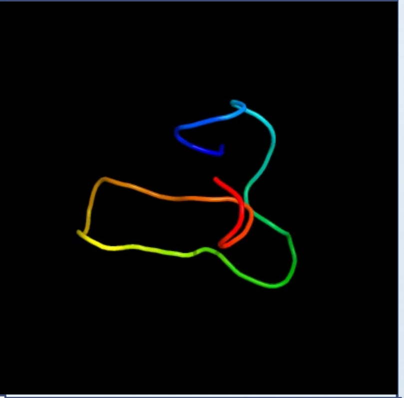
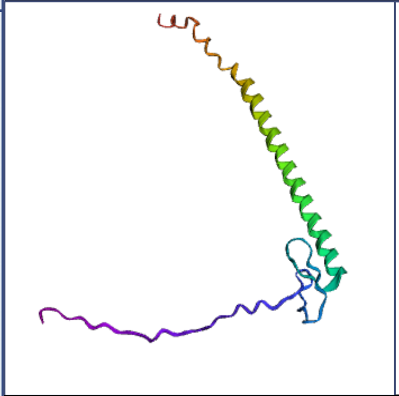
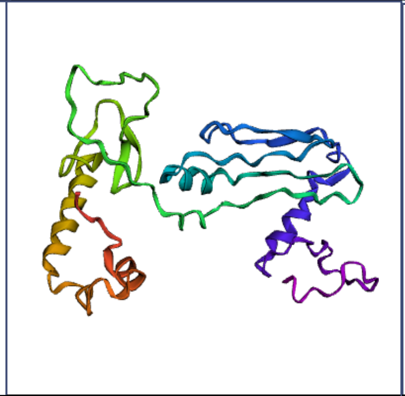
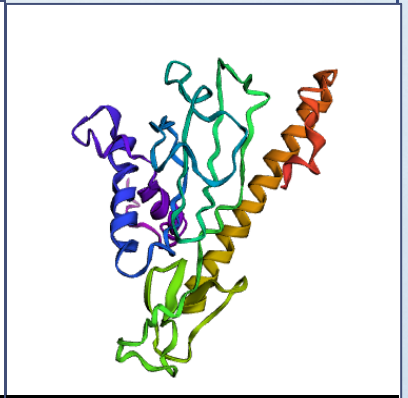
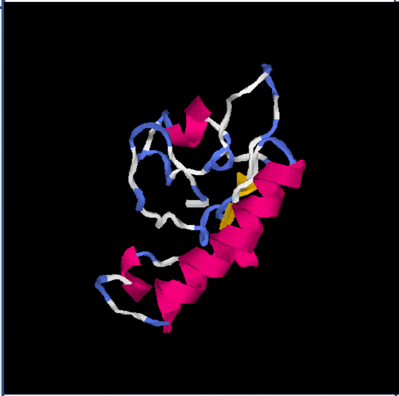
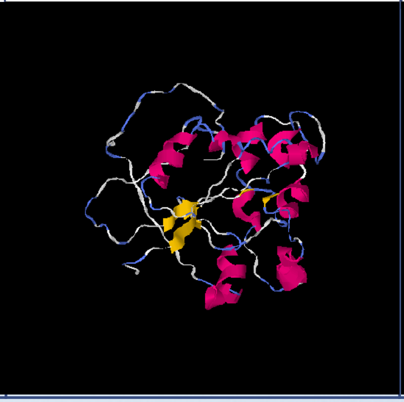
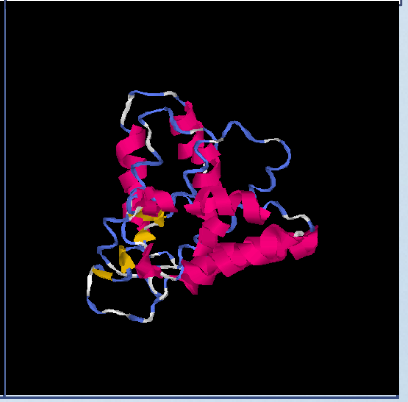
Zinc fingers are a ubiquitous class of protein domain with considerable variation in structure and function. Zf-FCS is a highly diverged group of C_2-C_2 zinc finger which is present in animals, prokaryotes and viruses, but not in plants. In this study we identified that a plant specific domain of unknown function, DUF581 is a zf-FCS type zinc finger. Based on HMM-HMM comparison and signature motif similarity we named this domain as FCS-Like Zinc finger (FLZ) domain. A genome wide survey identified that FLZ domain containing genes are bryophytic in origin and this gene family is expanded in spermatophytes. Expression analysis of selected *FLZ* gene family members of *A. thaliana* identified an overlapping expression pattern suggesting a possible redundancy in their function. Unlike the zf-FCS domain, the FLZ domain found to be highly conserved in sequence and structure. Using a combination of bioinformatic and protein-protein interaction tools, we identified that FLZ domain is involved in protein-protein interaction.

14、系统发生树构建

MEGA 7



15、蛋白质结构预测

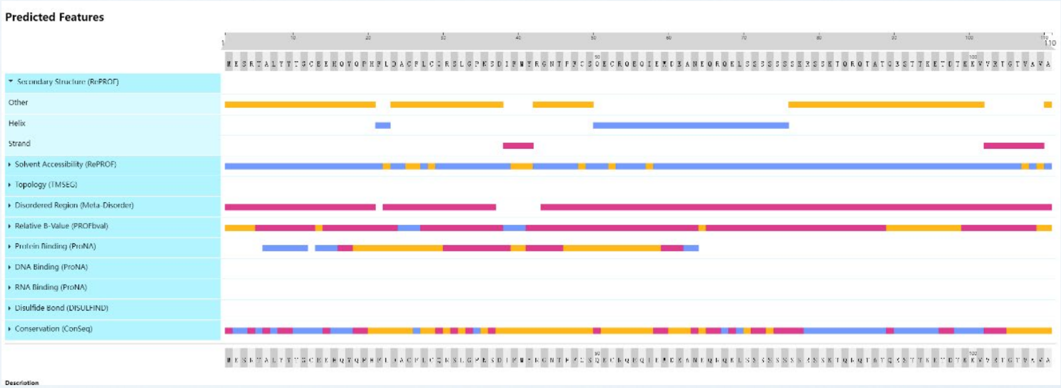
	MARD1-1	MARD1-2	MARD1-3
Phyre2			
trRosetta			
I-TASSER			

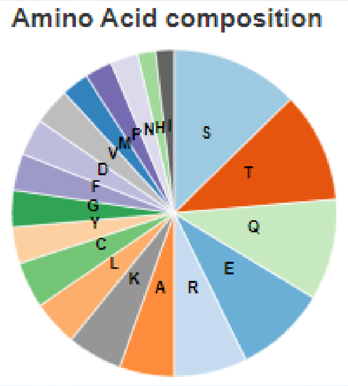
（http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index）

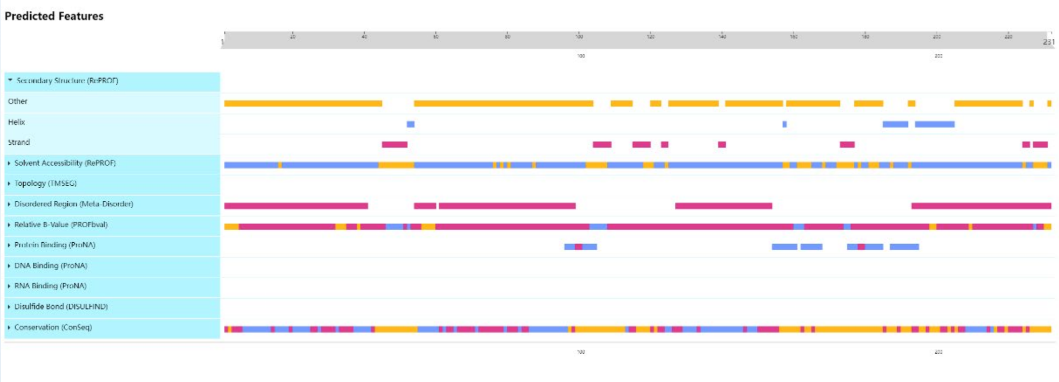
（https://yanglab.nankai.edu.cn/trRosetta/）

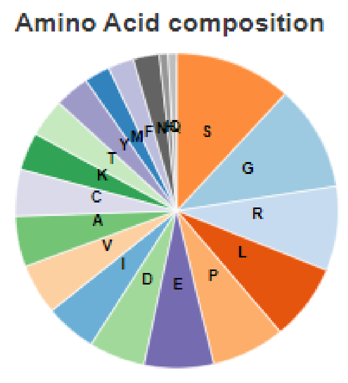
（https://zhanggroup.org/I-TASSER/）

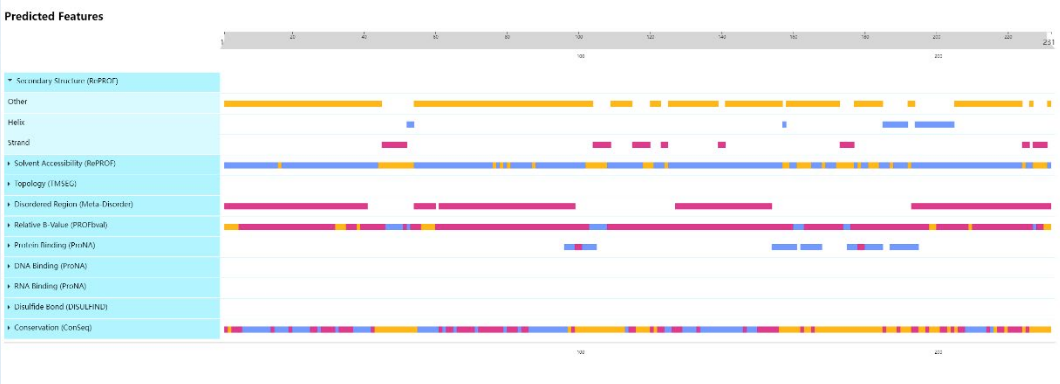
16、蛋白质二级结构预测

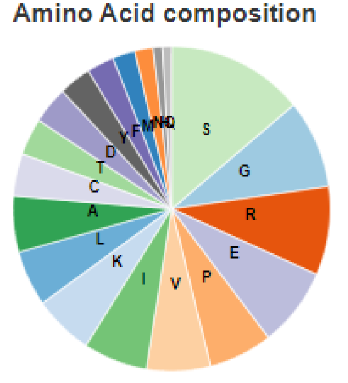






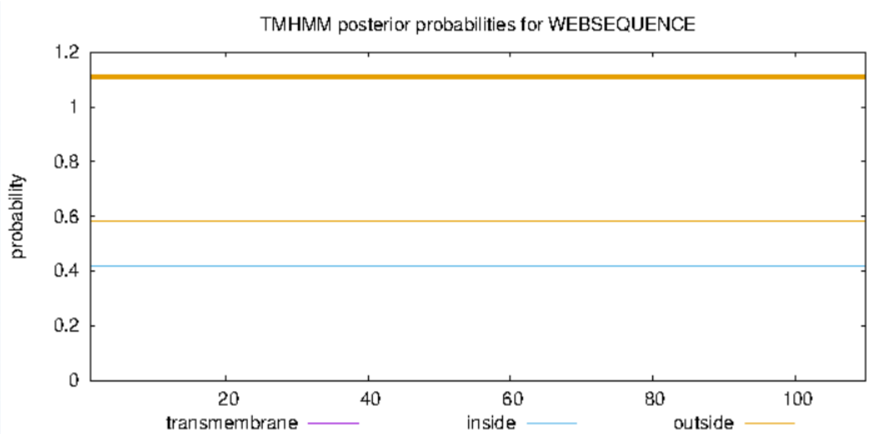


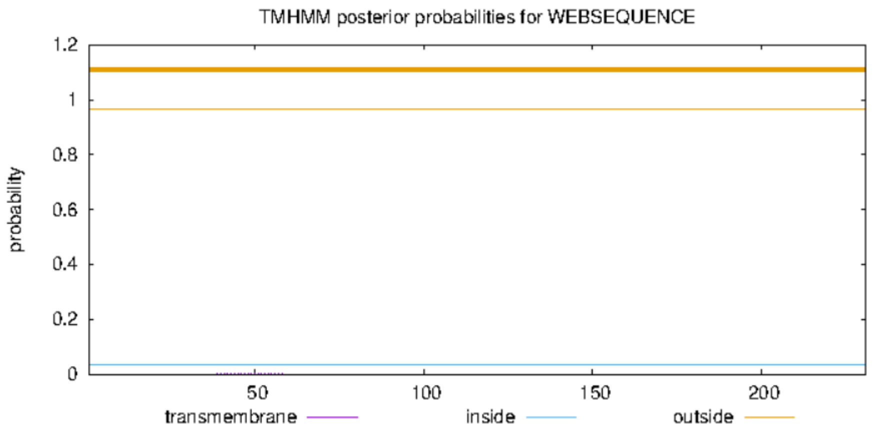


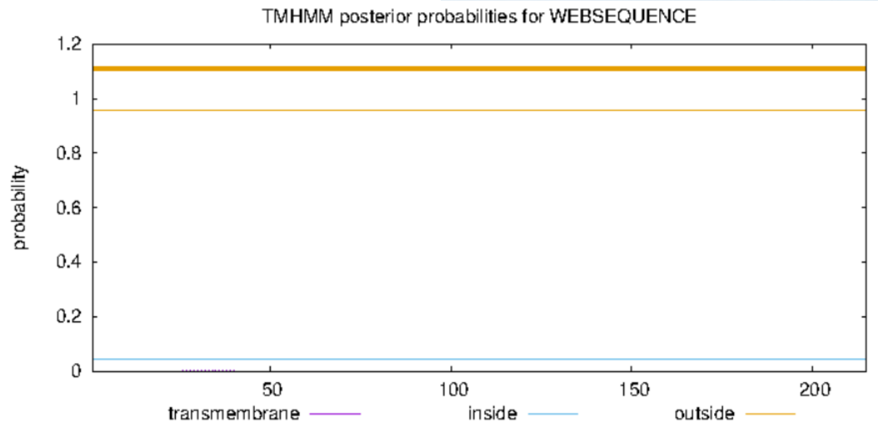


（ https://predictprotein.org ）

17、跨膜螺旋预测







（https://services.healthtech.dtu.dk/service.php?TMHMM-2.0）

三、总结

在上述研究中，运用了《实用生物信息技术》课程上学习到的部分方法和软件，并进行了拓展，这对我接下来开展研究尤其是未知功能蛋白的解析提供了很大的帮助，比如通过其他作物中该蛋白功能对研究蛋白功能进行预测，通过预测蛋白结构了解其功能等；目前，ABC网站（http://abc.gao-lab.org/）中我所熟练掌握并应用的工具并清楚其原理的只有一小部分，还有许多课程中没有提到的文章、软件等值得我们继续探索和学习。

谢谢大家！

请各位老师、同学批评指正！