



用生物信息技术分析拟南芥中 PP2C 16 (HAB1)

The bioinformatic analysis of PP2C 16 HAB1
in *Arabidopsis thaliana*

报告人：徐伟娜

G09组员：李艳妃 徐伟娜
刘慧 赵立萍



主要内容

- ▶ 一 目的
- ▶ 二 相关背景介绍
- ▶ 三 序列分析
 - ▶ (一) 核酸序列分析
 - ▶ (二) 蛋白质序列分析
- ▶ 四 功能及结构分析
- ▶ 五 讨论



目的

我们组选取已经研究的比较透彻的拟南芥基因HAB1，编码蛋白质为PP2C 16。我们希望通过对这个基因及表达蛋白的分析来巩固在这门课上所学到的知识，并与已知的研究结果进行对比，找出并讨论我们在实际的分析过程中所出现的各种问题，以便更好的为我们将要开始的实验做指导。



一 相关背景介绍

1、干旱对农业的影响

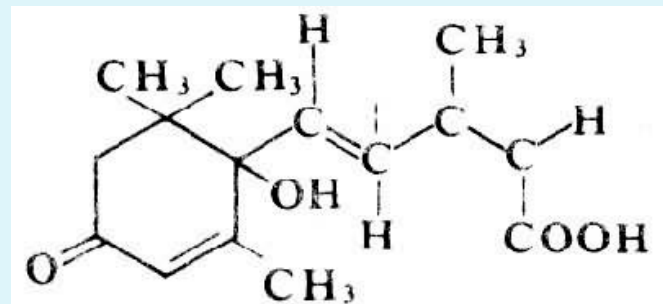
干旱一直是制约我国及全球农业发展的主要因素之一。随着人口的增加、工农业用水需求的增大，以及全球气候变暖和生态环境的恶化，干旱将进一步加剧。世界上有将近1/3的可耕地处于半干旱及干旱地区。较之其他的自然灾害，干旱发生的频率高、影响范围广，持续时间长，在农业的发展中造成的损失最大。

由此可见，深入研究**作物抗旱调控机制**，发掘并利用作物抗旱的优异基因资源，提高作物的抗旱性即水分利用效率，培育抗旱新品种，对于缓解水资源短缺，保障国家粮食安全、生态安全和社会的可持续发展具有重要意义。

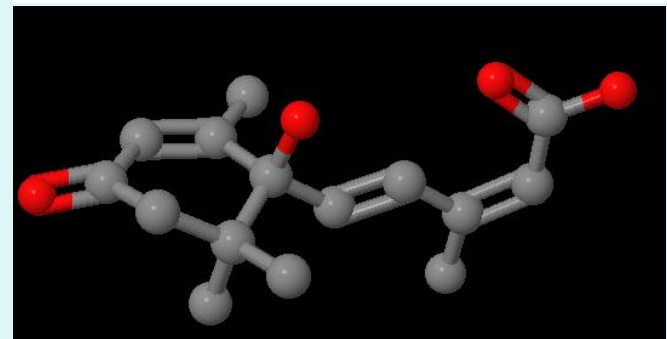


2、ABA(脱落酸)

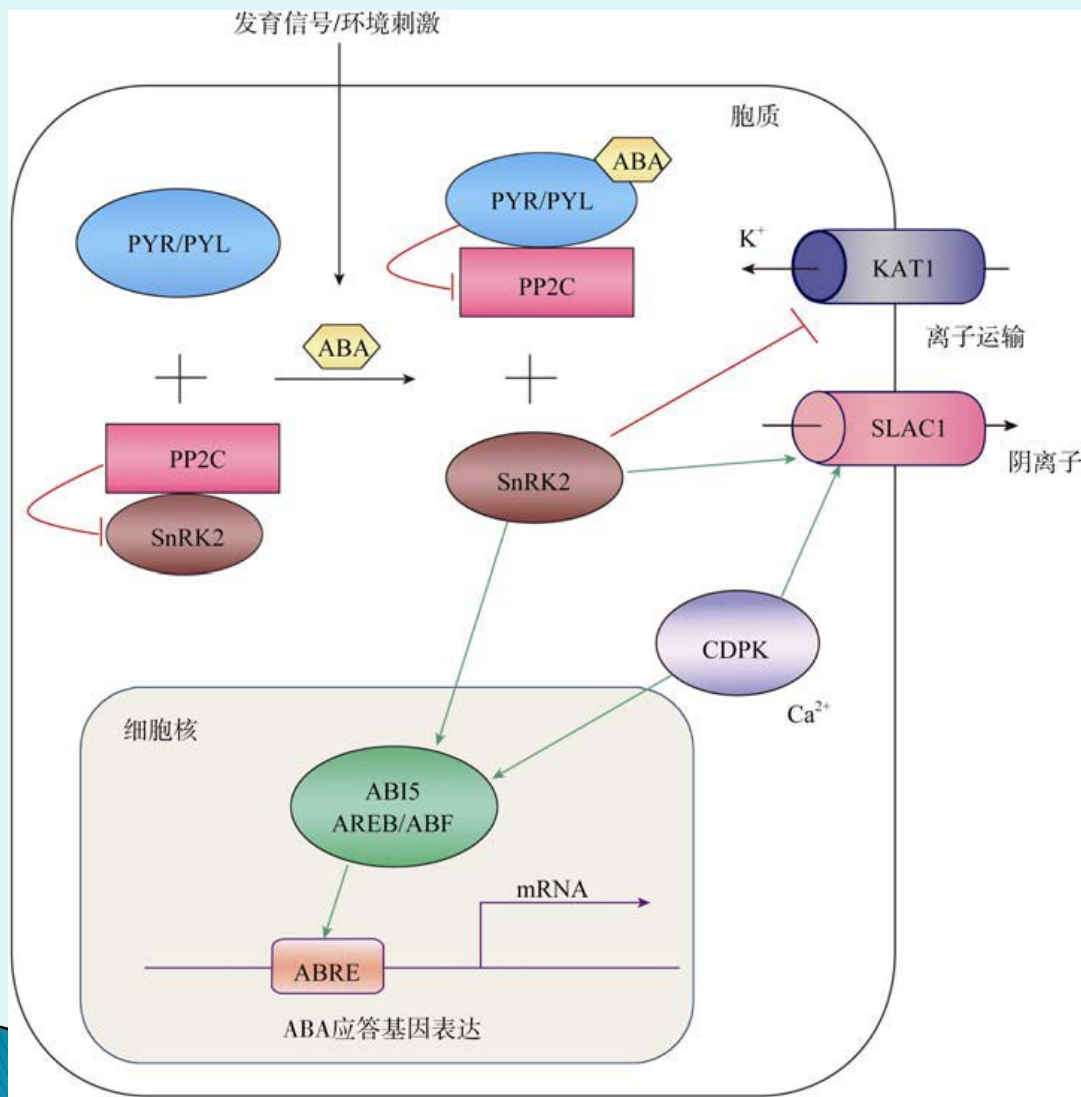
脱落酸是一种具有15碳的**倍半萜结构**植物激素。天然脱落酸与生长素、乙烯、赤霉素、**细胞分裂素**并列为**植物五大激素**，它可以提高植物的**逆能力**，ABA还是抑制种子萌发的有效抑制剂，因此可以用于种子贮藏，保证种子、果实的贮藏质量。此外，ABA还能引起叶片**气孔的迅速关闭**，可用于花的保鲜、调节花期、促进生根等，在花卉园艺上有较大的应用价值。对ABA及其应答基因的研究可揭示植物**抗逆生理**反应的分子过程，从而为定向增强作物对环境的适应力奠定基础。



脱落酸结构式



ABA信号转导的核心途径



大量的研究表明：
PYR/PYL/RCAR、
PP2C和**SnRK2**这3
 种信号组份在ABA
 信号转导途径中
 相互联系，共同
 组成了一个**双重
 负调控系统**，通
 过调节下游的作
 用元件而形成一
 条完整的ABA信号
 通路。



3、蛋白磷酸酶@蛋白磷酸激酶

蛋白磷酸酶和蛋白激酶分别参与蛋白质的磷酸化和去磷酸化过程。

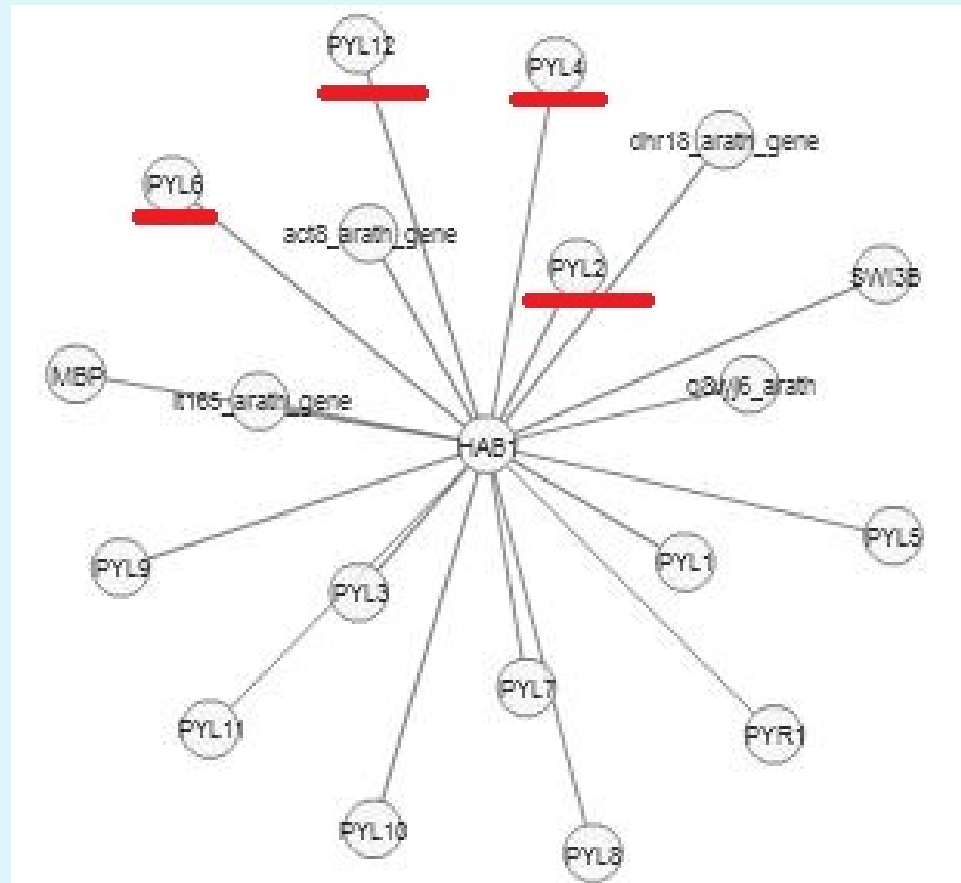
蛋白磷酸酶：

A phosphoprotein + H₂O = a protein + phosphate

蛋白激酶：

A protein + phosphate = a phosphoprotein + H₂O

蛋白质的相互作用



蛋白磷酸酶 (PP2C)

PP2C蛋白质家族能结合二价锰离子或二价镁离子，作用于丝氨酸/苏氨酸的。它是蛋白磷酸酶中的一大类，与植物的生长发育、细胞周期调节、信号转导及环境胁迫应答等各种生物学过程相关。研究显示，这个蛋白质家族作为 **ABA 信号通路** 的关键组分和**抑制物**来调节大量的 ABA 参与的信号响应。

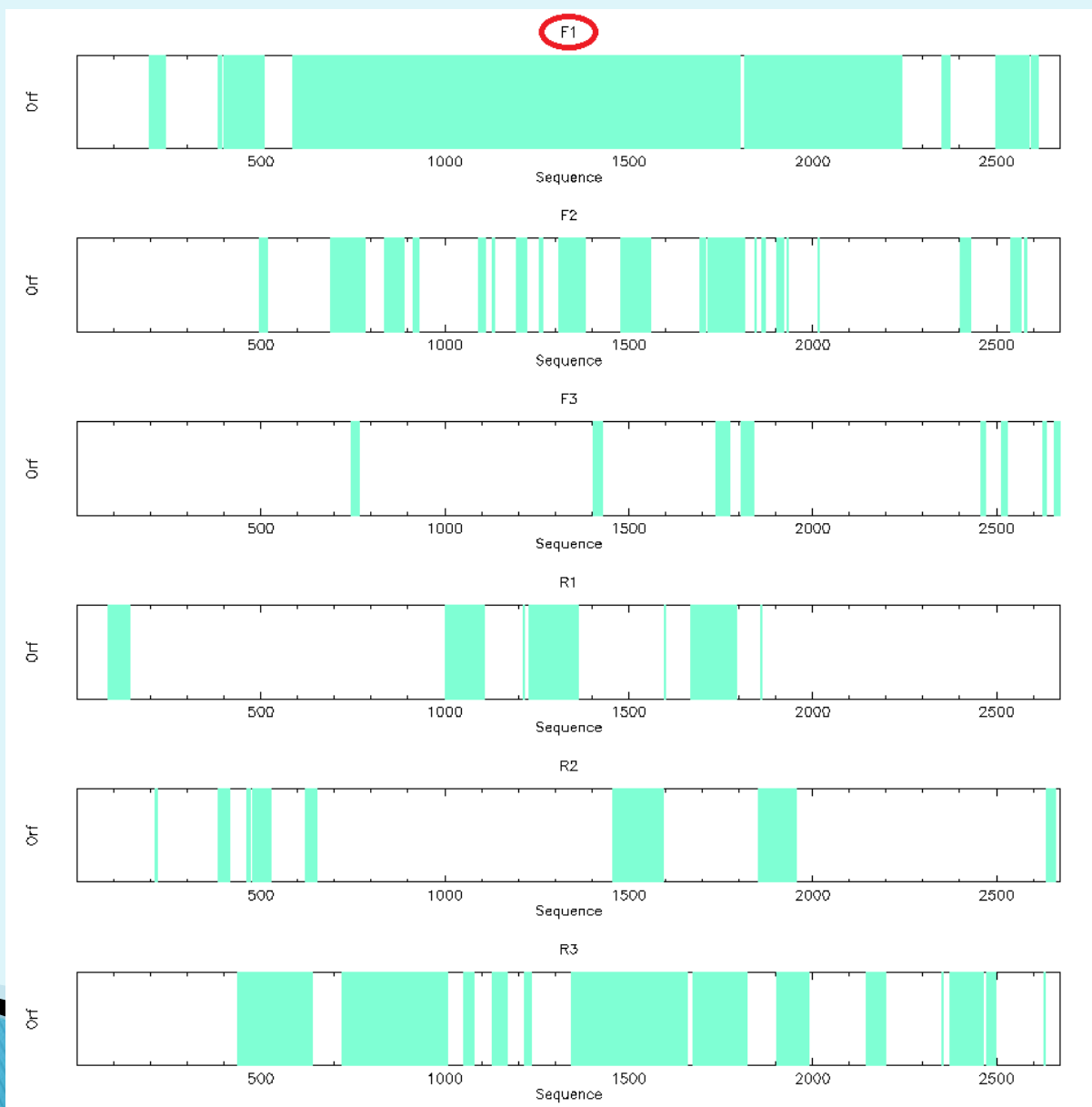


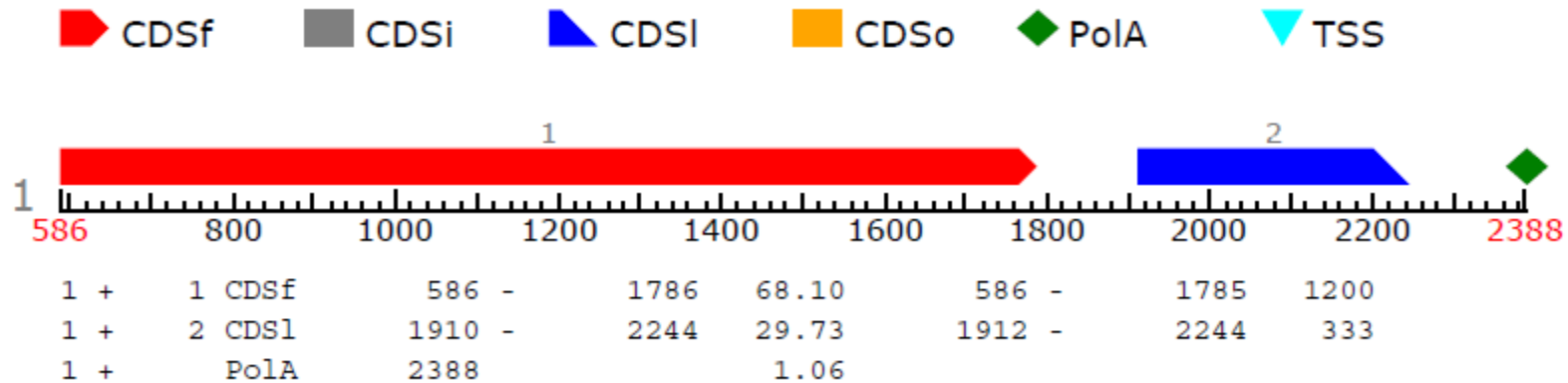


二 序列分析

(一) 核酸序列分析

Plotorf 分析结果





1 2 exon (s) 586 - 2244 1536 bp, chain +

Softberry: 两个外显子

- ▶ CDSf: 起始外显子; CDSi: 居间外显子;
- ▶ CDSl: 末端外显子; CDSo: 单外显子;
- ▶ PoIA: 末端polyA 区; TSS: 转录起始位点。



Cusp: 密码子偏好性分析

```
#CdsCount: 1
#Coding GC 44.14%
#1st letter GC 55.04%
#2nd letter GC 40.29%
#3rd letter GC 37.10%
#Codon AA Fraction Frequency Number
GCA A 0.276 19.656 8
GCC A 0.207 14.742 6
GCG A 0.172 12.285 5
GCT A 0.345 24.570 10
TGC C 0.700 17.199 7
TGT C 0.300 7.371 3
GAC D 0.179 12.285 5
GAT D 0.821 56.511 23
GAA E 0.579 54.054 22
GAG E 0.421 39.312 16
TTC F 0.308 9.828 4
TTT F 0.692 22.113 9
GGA G 0.182 14.742 6
GGC G 0.242 19.656 8
GGG G 0.182 14.742 6
GGT G 0.394 31.941 13
CAC H 0.182 4.914 2
CAT H 0.818 22.113 9
ATA I 0.421 19.656 8
ATC I 0.263 12.285 5
ATT I 0.316 14.742 6
AAA K 0.667 24.570 10
```

```
AAG K 0.333 12.285 5
CTA L 0.118 9.828 4
CTC L 0.118 9.828 4
CTG L 0.088 7.371 3
CTT L 0.176 14.742 6
TTA L 0.206 17.199 7
TTG L 0.294 24.570 10
ATG M 1.000 29.484 12
AAC N 0.312 12.285 5
AAT N 0.688 27.027 11
CCA P 0.300 7.371 3
CCC P 0.400 9.828 4
CCG P 0.100 2.457 1
CCT P 0.200 4.914 2
CAA Q 0.400 9.828 4
CAG Q 0.600 14.742 6
AGA R 0.471 19.656 8
AGG R 0.235 9.828 4
CGA R 0.118 4.914 2
CGC R 0.000 0.000 0
CGG R 0.000 0.000 0
CGT R 0.176 7.371 3
AGC S 0.154 14.742 6
AGT S 0.154 14.742 6
TCA S 0.205 19.656 8
TCC S 0.103 9.828 4
TCG S 0.051 4.914 2
```

```
TCT S 0.333 31.941 13
ACA T 0.136 7.371 3
ACC T 0.273 14.742 6
ACG T 0.136 7.371 3
ACT T 0.455 24.570 10
GTA V 0.140 14.742 6
GTC V 0.116 12.285 5
GTG V 0.093 9.828 4
GTT V 0.651 68.796 28
TGG W 1.000 7.371 3
TAC Y 0.000 0.000 0
TAT Y 1.000 9.828 4
TAA * 0.000 0.000 0
TAG * 0.000 0.000 0
TGA * 1.000 2.457 1
```



Remap分析结果

Restriction enzyme	Number	Restriction enzyme	Number	Restriction enzyme	Number
MspJI	476	Sse9I	18	SsiI	13
SgeI	204	TasI	18	TspDTI	13
FaiI	42	Tsp509I	18	AgsI	12
SetI	34	TspEI	18	BfuCI	12
LpnPI	32	HpyCH4V	15	Bsp143I	12
MnlI	29	AluBI	14	BssMI	12
CviJI	28	AluI	14	BstKTI	12
CviKI-1	28	AciI	13	BstMBI	12
MboII	26	BspACI	13	CviAII	12
MluCI	18	Hpy188I	13	DpnI	12



(二) 蛋白质序列分析

用softberry预测的编码区翻译的蛋白质 (511aa)

```
>PP2C16_ARATH  
MEEMTPAVAMTSLAANTMCESSPVEITQLKNVTDAADLLSDSENQSF CNGGTECTME  
DVSELEEVGEQDLLKTLSDTRSGSSNVFDEDDVLSVVEDNSAVISEGLLVVDAGSELSLSN  
TAMEIDNGRVLATAIIVGESSIEQVPTAEVLIAGVNQDTNTSEVVIRLPDENSNHLVKGRS  
VYELDCIPLWGTVSIQGNRSEMEDAFVSPHFLKLPKMLMGDHEGMSPSLTHLTGHFFG  
VYDGHGGHKVADYCRDRLHFALAEIERIKDELCKRNTGEGRQVQWDKVFTSCFLTVD  
GEIEGKIGRAVVGSSDKVLEAVASETVGSTAVVALVCSSHIVVSNCGDSRAVLFRGKEAM  
PLSVDHKPDREDEYARIENAGGKVIQWQGARVFGVLAMSRSIGDRYLKPYVIPEPEVTFM  
PRREDECLILASDGLWDVMNNQEVCEIARRRILMWHKKN GAPPLAERGKGIDPACQAA  
ADYLSMLALQKGSKDNISIIVIDLKAQRKFKTRT
```




Pepstats 分析结果

PEPSTATS of P2C16_ARATH from 1 to 511

Molecular weight = 55743.87

Residues = 511

Average Residue Weight = 109.088

Charge = -26.5

Isoelectric Point = 4.4730

A280 Molar Extinction Coefficient = 37410

A280 Extinction Coefficient mg/ml = 0.67

Probability of expression in inclusion bodies = 0.701

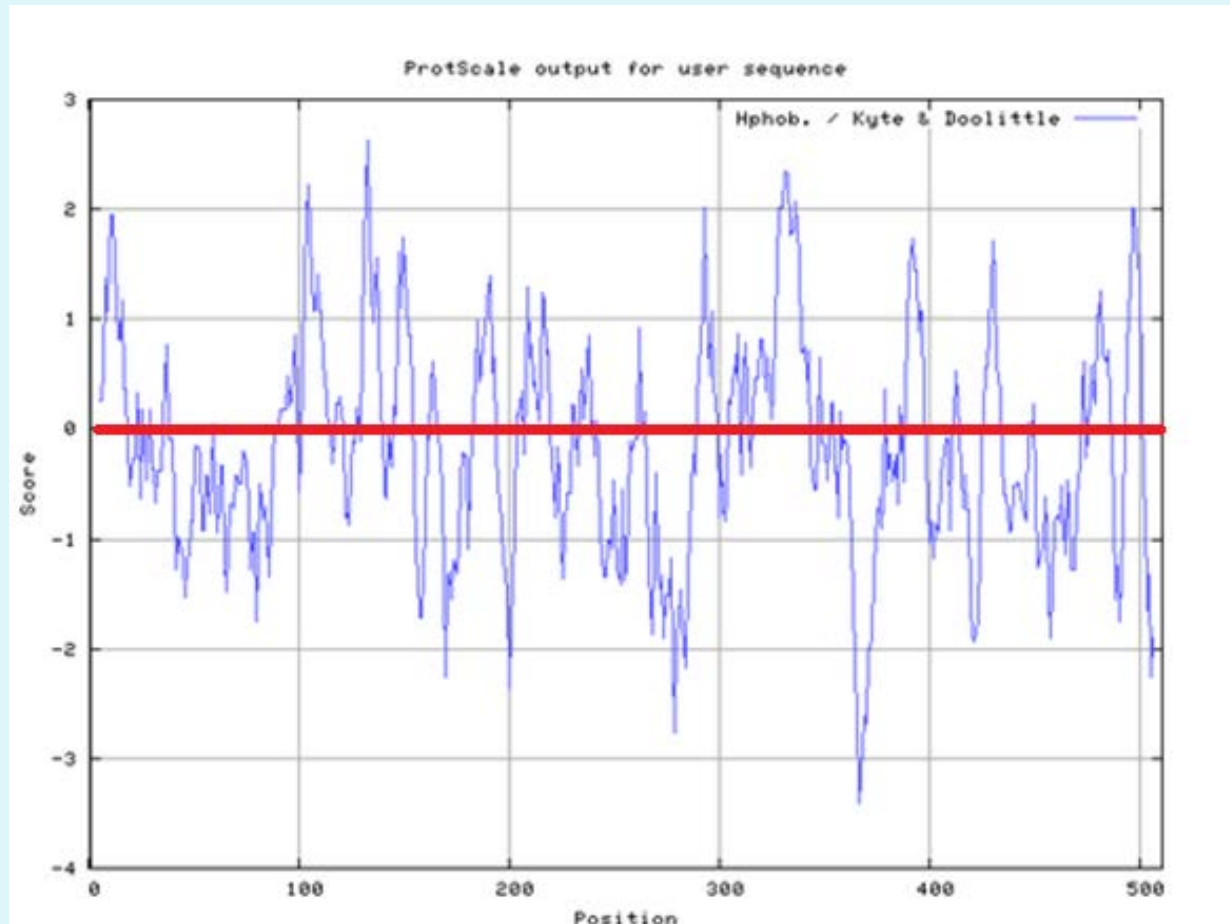
Residue	Number	Mole%	DayhoffStat
A = Ala	39	7.632	0.887
B = Asx	0	0.000	0.000
C = Cys	12	2.348	0.810
D = Asp	36	7.045	1.281
E = Glu	45	8.806	1.468
F = Phe	13	2.544	0.707
G = Gly	38	7.436	0.885
H = His	11	2.153	1.076
I = Ile	28	5.479	1.218
J = ---	0	0.000	0.000
K = Lys	23	4.501	0.682
L = Leu	44	8.611	1.164
M = Met	16	3.131	1.842
N = Asn	20	3.914	0.910
O = ---	0	0.000	0.000
P = Pro	17	3.327	0.640
Q = Gln	14	2.740	0.702
R = Arg	26	5.088	1.038
S = Ser	44	8.611	1.230
T = Thr	25	4.892	0.802
U = ---	0	0.000	0.000
V = Val	48	9.393	1.423
W = Trp	5	0.978	0.753
X = Xaa	0	0.000	0.000
Y = Tyr	7	1.370	0.403
Z = Glx	0	0.000	0.000



极性氨基酸和非极性氨基酸大约各占一半

Property	Residues	Number	Mole%	
Tiny	(A+C+G+S+T)		158	30.920
Small	(A+B+C+D+G+N+P+S+T+V)	279		54.599
Aliphatic	(A+I+L+V)	159	31.115	
Aromatic	(F+H+W+Y)	36	7.045	
Non-polar	(A+C+F+G+I+L+M+P+V+W+Y)	267		52.250
Polar	(D+E+H+K+N+Q+R+S+T+Z)	244		47.750
Charged	(B+D+E+H+K+R+Z)		141	27.593
Basic	(H+K+R)		60	11.742
Acidic	(B+D+E+Z)	81		15.851

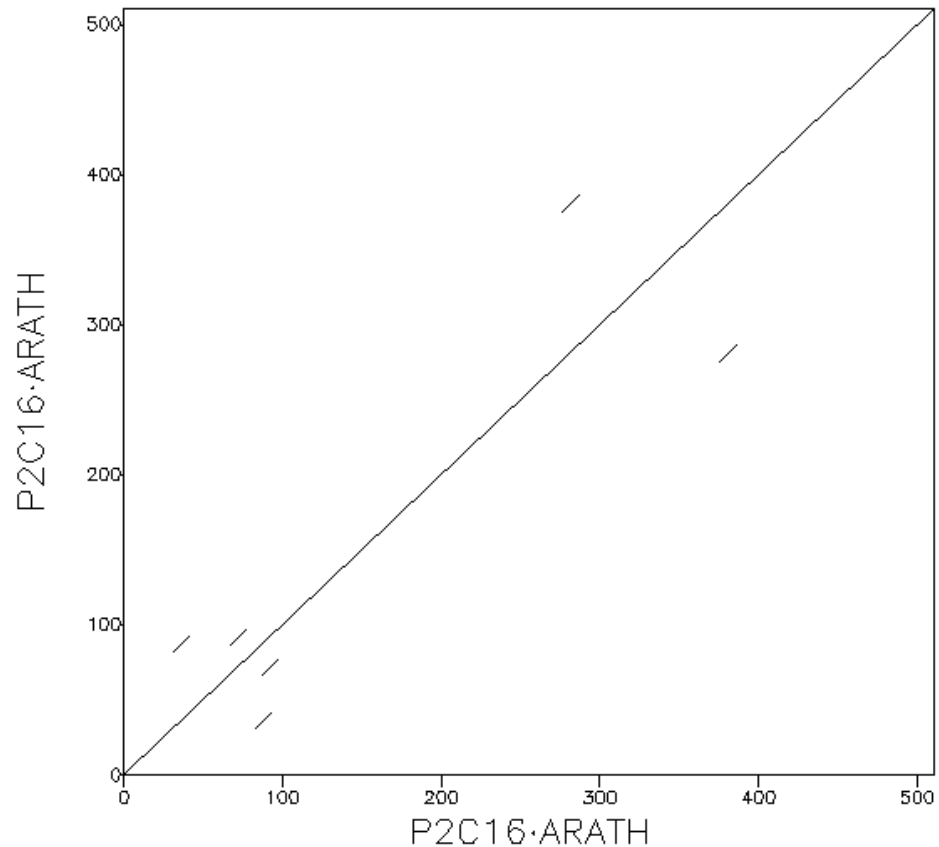
ProtScale 分析结果





dotmatcher 分析结果

Dotmatcher: fasta::698300:P2C16-ARATH vs fasta::698300:P...
(windowsize = 10, threshold = 23.00 11/01/14)



573 Sequence

SPQCAXPCARATHPRXTEINPHXSPHATASECX SARAXIDXP S ISTHALIANAGNHAXPESVMEEMTPAVAMTLSLAANT	80
MCESSPVEITQLKNVTD AADLLSDSENQSF CNGGTECTMEDVSELEEVGEQDLLKTLSDTRSGSSNVFDEDDVLSVVEDN	160
SAVISEGLLVVDAGSELSLSNTAMEIDNGRVLATAIIVGESSIEQVPTAEVLIAGV NQDINTSEVVIRLPDENS NHLVKG	240
RSVYELDCIPLWGT VSIQGNRSEMEDAF AVSPHFLKLP IKMLMGDHEGMSPSLTHLTGHFFGVYDGHGGHKVADYCRDRL	320
HFALAE EIERIKDELCKRNTGEGRQVQWDKVFTSCFLTVDGEIEGKIGRAVVGSSDKVLEAVASETVGSTAVVALVCSSH	400
IVVSNCGDSRAVLF RGKEAMPLSVDHKPDREDEYARIENAGGKVIQWQGARVFGV LAMSRSIGDRYLKPYVIPEPEVTFM	480
PRSRDECLILASDGLWDMNMQEVCEIARRRILMWHKKN GAPPLAERGKGIDPACQAAAADYLSMLALQKGSKDNIS IIV	560
IDLKAQRKFKTRT	640
.....T.....S...T.....S.....T.....T	80
...S.....S.S...S.....T...S.....S.SS.....S.....	160
.....S.S.....SS.....	240
.S.....S.....S.....S.....	320
.....T.....S.....	400
.....Y.....S.S...Y.....	480
.S.....Y.....S...	560
.....	640

Phosphorylation sites predicted: Ser: 24 Thr: 6 Tyr: 3

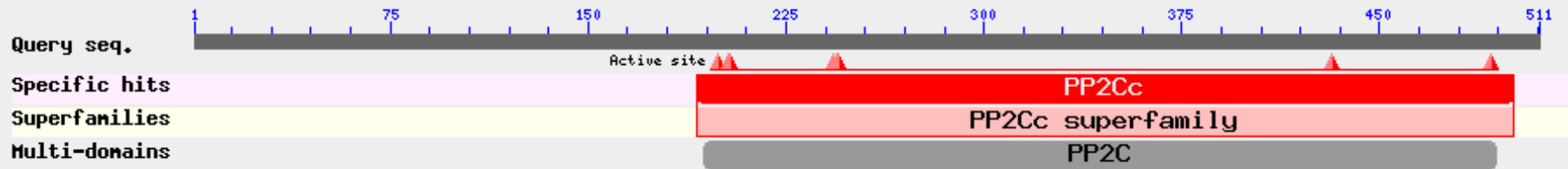
用NetPhos 2.0对PP2C磷酸化位点进行预测发现共**24个Ser**、**6个Thr**和**3个Tyr**潜在磷酸化位点



Blastp 结果：属于PP2C蛋白超家族

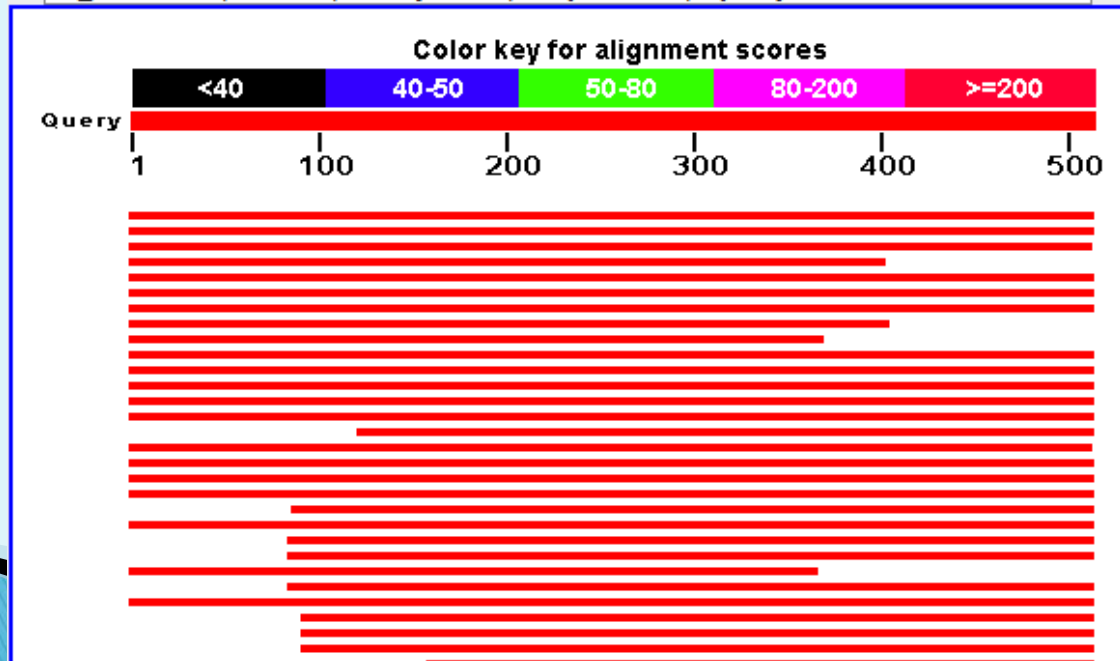
（没有找到小麦的序列，但是有玉米和水稻序列。）

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

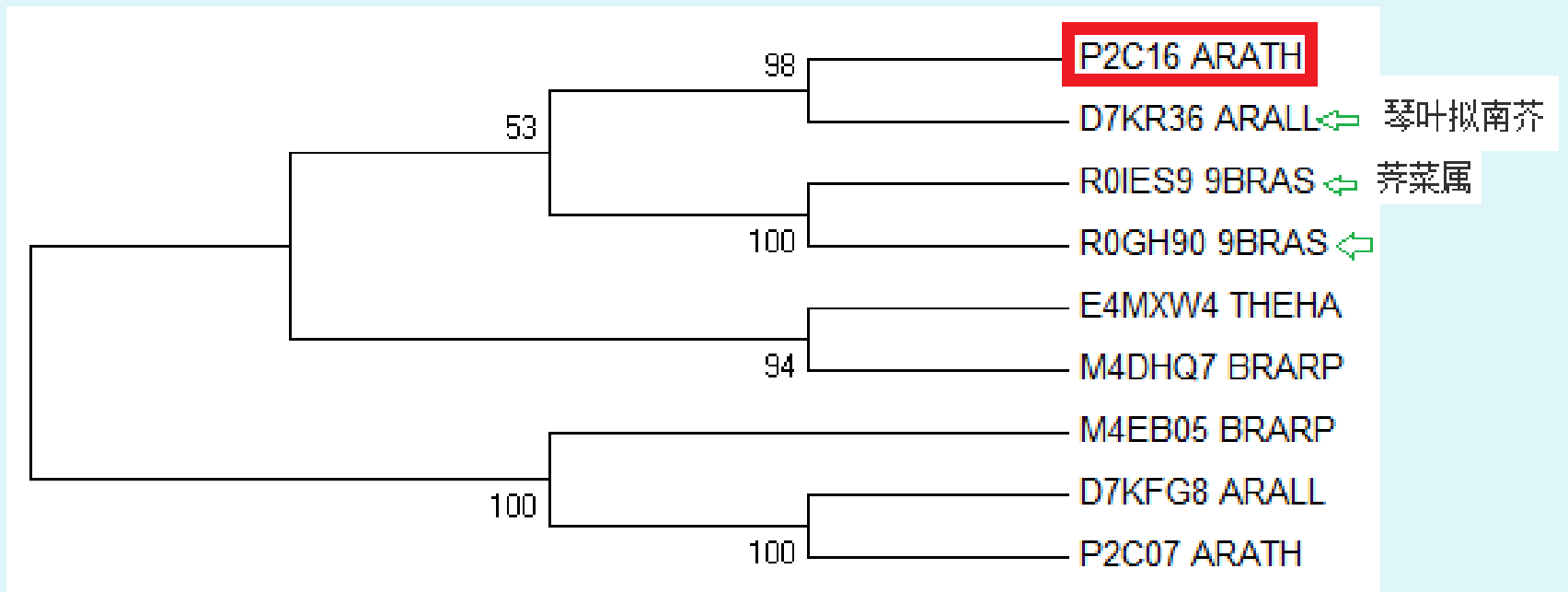


Distribution of 102 Blast Hits on the Query Sequence

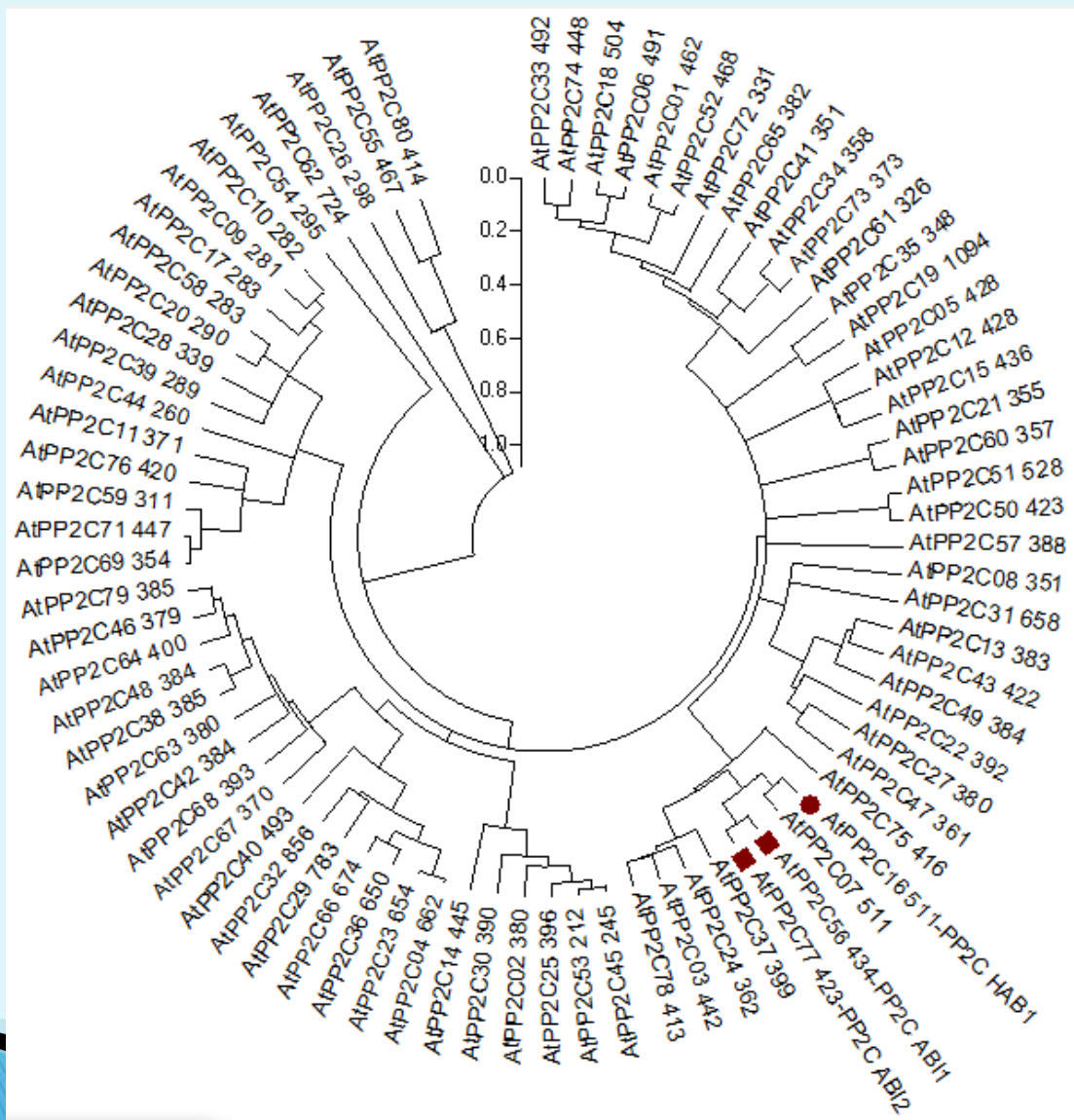
XP_002890233 predicted protein [Arabidopsis lyrata subsp. lyrata] S=622 E=0



NJ法进化树的构建



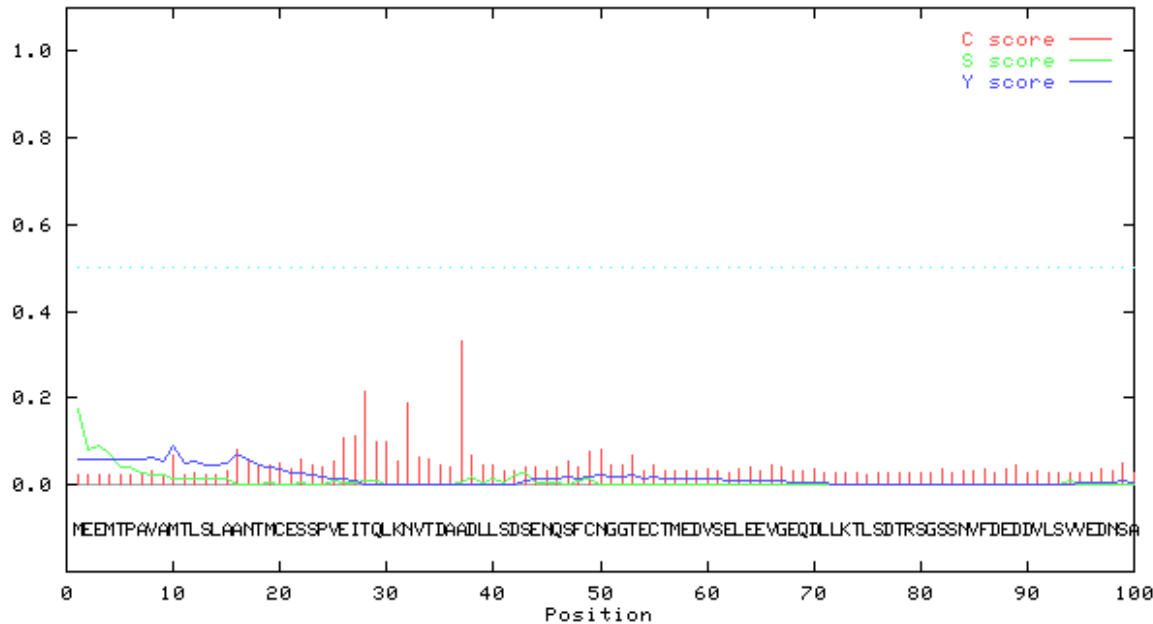
PP2C家族 (79个审阅的蛋白质序列)





信号肽预测

TatP 1.0 prediction (tat networks): sp Q9CAJ0 P2C16 ARATH



TatP 1.0 Server

<input type="checkbox"/>	Signal peptide	1 - 22	22
<input type="checkbox"/>	Chain	23 - 511	489

```
>sp_Q9CAJ0_P2C16_ARAT length = 100
# Measure Position Value Cutoff Tat signal
max. C      37      0.331  0.51  NO
max. Y      10      0.090  0.35  NO
max. S       1      0.177  0.75  NO
mean S      1-9      0.064  0.30  NO
max. D      1-9      0.077  0.36  NO
# Used regex: RR.[FGAVML][LITMVF]
```



亚细胞定位

Using PLANT networks.

Name	Len	cTP	mTP	SP	other	Loc	RC
sp_Q9CAJ0_P2C16_ARAT	511	0.284	0.034	0.366	0.532	_	5
cutoff		0.000	0.000	0.000	0.000		

TargetP 1.1 Server

Subcellular location

Cytoplasm. Nucleus. Note: Mainly cytoplasmic.





四 蛋白质功能及结构分析

MEME分析结果：3个保守结构域

MOTIF 1

[Next Top](#)

Summary

E-value 9.9e-410
Width 50
Sites 10

[show more](#)

Sequence Logo



MOTIF 2

[Previous](#) [Next Top](#)

Summary

E-value 1.9e-338
Width 50
Sites 10

[show more](#)

Sequence Logo



MOTIF 3

[Previous](#) [Next Top](#)

Summary

E-value 1.9e-328
Width 50
Sites 10

[show more](#)

Sequence Logo

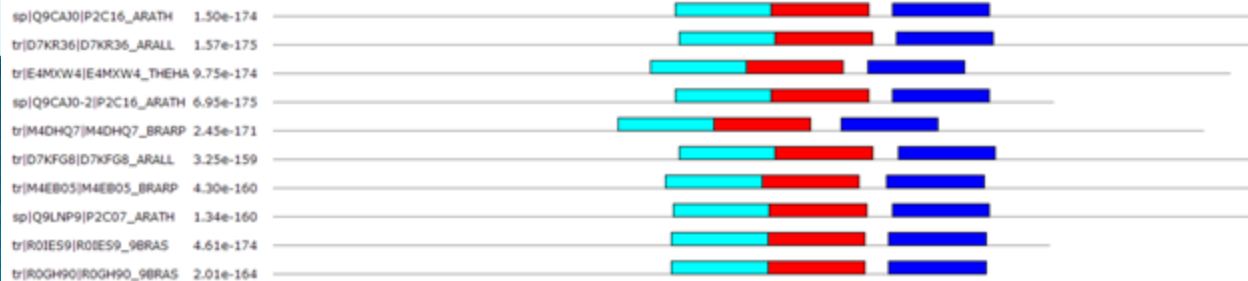


Download LOGO [?](#) Orientation: SSC: Format: Width: cm Height: cm

■ Motif 1 ■ Motif 2 ■ Motif 3

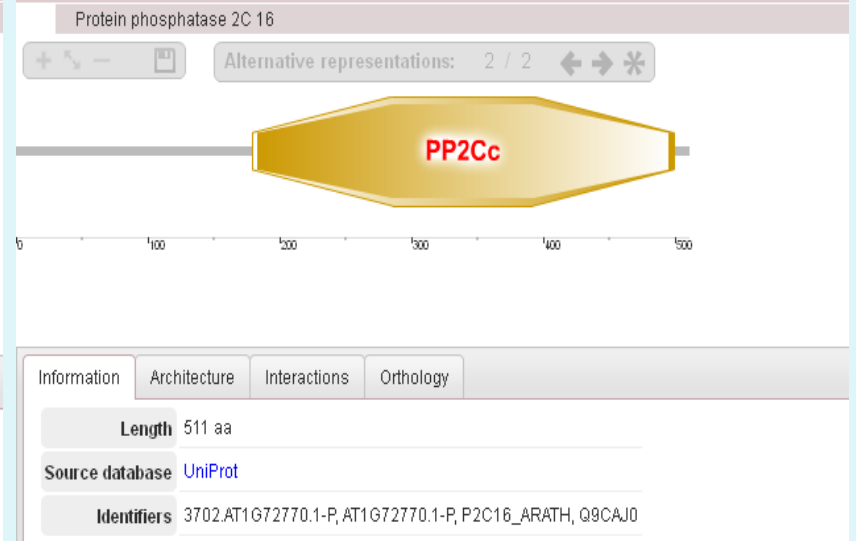
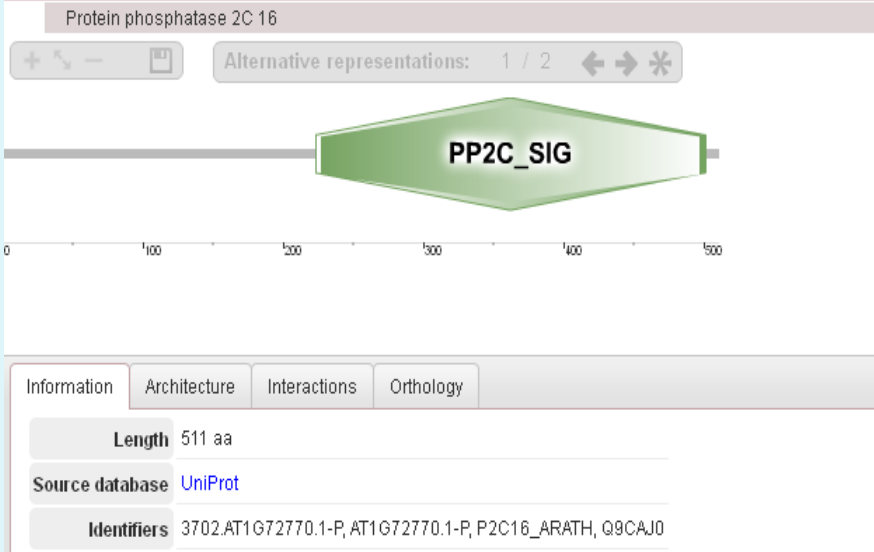
Name Combined p-value

Motif Location



Domains within *Arabidopsis thaliana* protein P2C16_ARATH (Q9CAJ0)

Domains within *Arabidopsis thaliana* protein P2C16_ARATH (Q9CAJ0)



Confidently predicted domains, repeats, motifs and features:

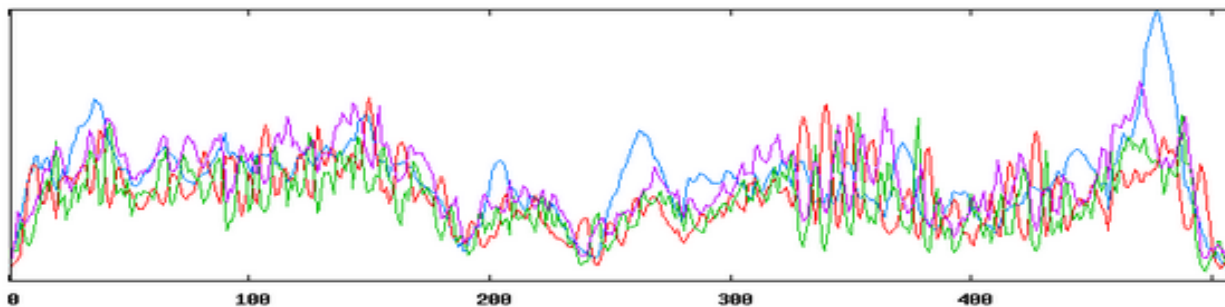
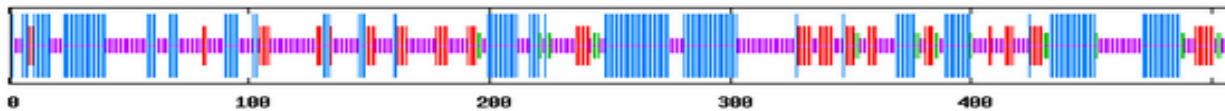
Name	Start ▲	End	E-value
PP2Cc	179	499	1e-104 更为可靠
PP2C_SIG	223	501	0.136

SMART运行预测到两种可能的结构域但PP2Cc型结构域更为可靠，位于179-499位点。推测PP2C磷酸酶催化机理与PP1，PP2A，PP2B家族丝氨酸/苏氨酸磷酸酶相似，但与PP2C与它们没有序列相似性。

SOPMA: 二级结构分析

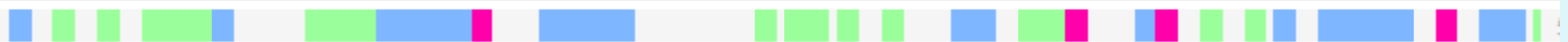
SOPMA :

Alpha helix	(Hh)	:	179	is	35.03%
310 helix	(Gg)	:	0	is	0.00%
Pi helix	(Ii)	:	0	is	0.00%
Beta bridge	(Bb)	:	0	is	0.00%
Extended strand	(Ee)	:	78	is	15.26%
Beta turn	(Tt)	:	23	is	4.50%
Bend region	(Ss)	:	0	is	0.00%
Random coil	(Cc)	:	231	is	45.21%
Ambiguous states	(?)	:	0	is	0.00%
Other states	:	:	0	is	0.00%



Parameters :

Window width	:	17
Similarity threshold	:	8
Number of states	:	4



Helix Strand Turn



SWISS-MODEL

Workunit: P000004 pp2c - Overview



Model Summary

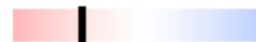


Model information:

Modelled residue range: 180 to 510
Based on template: [3rt0A] (2.11 Å)
Sequence Identity [%]: 98.79
Evalue: 0.00e-1

Quality information:
QMEAN Z-Score: -1.85

[details] ▾



Quaternary structure informat

Template (3rt0): HETERO DIMER
Model built :SINGLE CHAIN

Ligand information:

Ligands in the template: MG: 1.
Ligands in the model: MG: 1

logs: [Templates] ▾ [Alignment] ▾ [Modeling] ▾

display model: as [pdb] ⌘ - as [DeepView project] ⌘ - in [AstexViewer] ▾

download model: as [pdb] ⌘ - as [Deepview project] ⌘ - as [text] ⌘

PYL10-HAB1 complex

A: 3rt0_A (180-512)

B: 模建结果 (180-510)

C: fit (0.06埃)



A



B



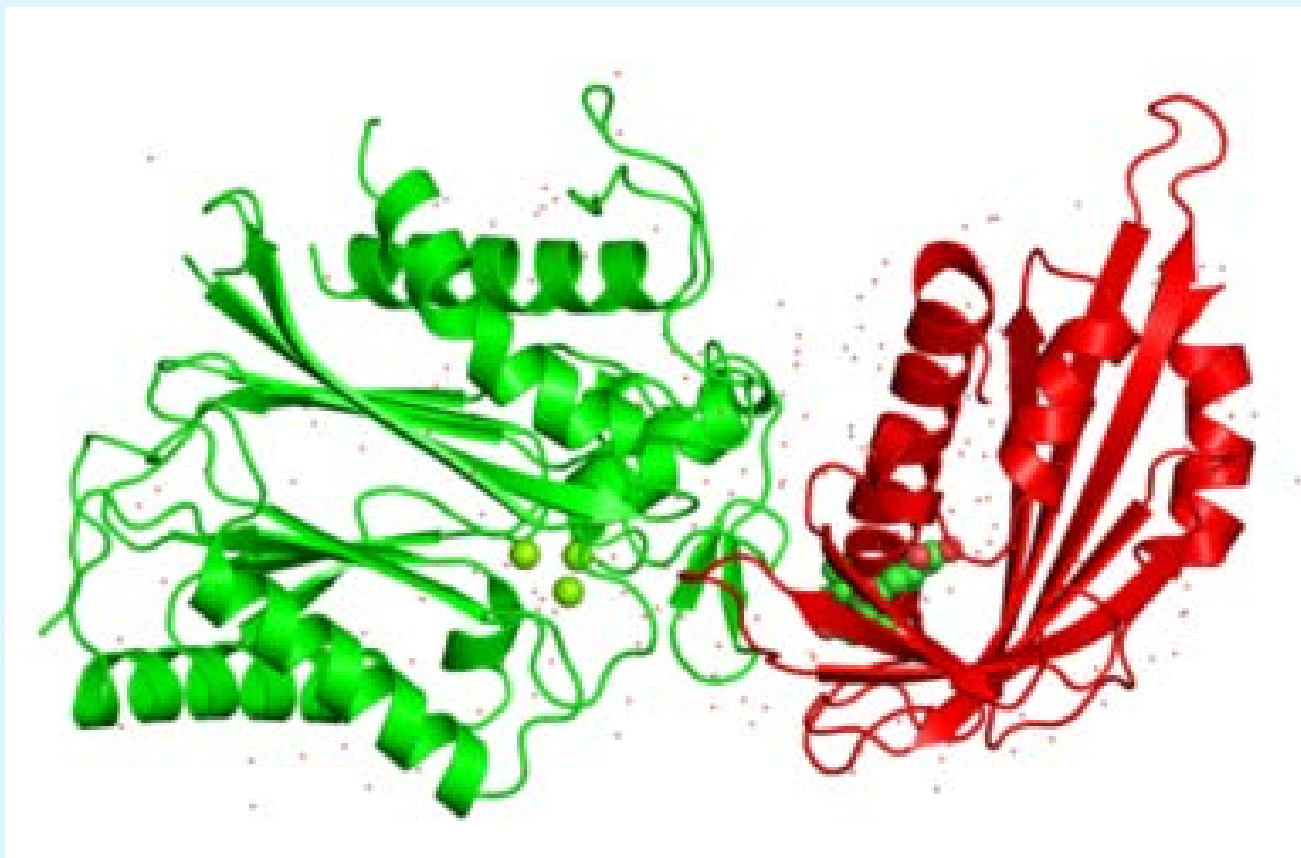
C

3kb3_B (186-506)

fit (3.7埃)

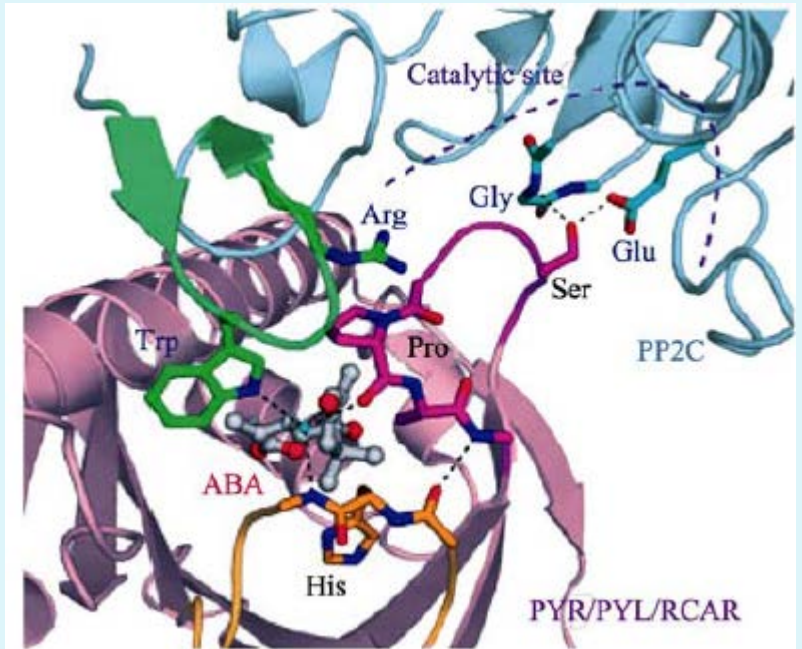
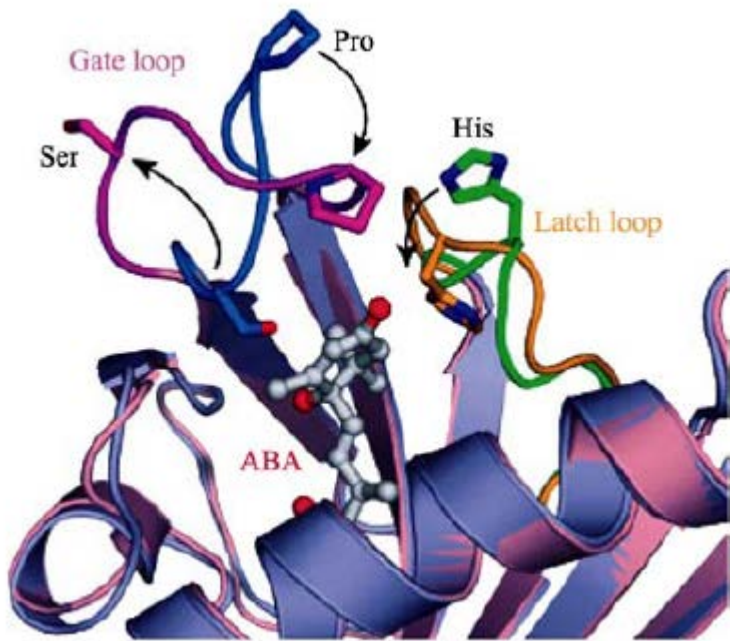
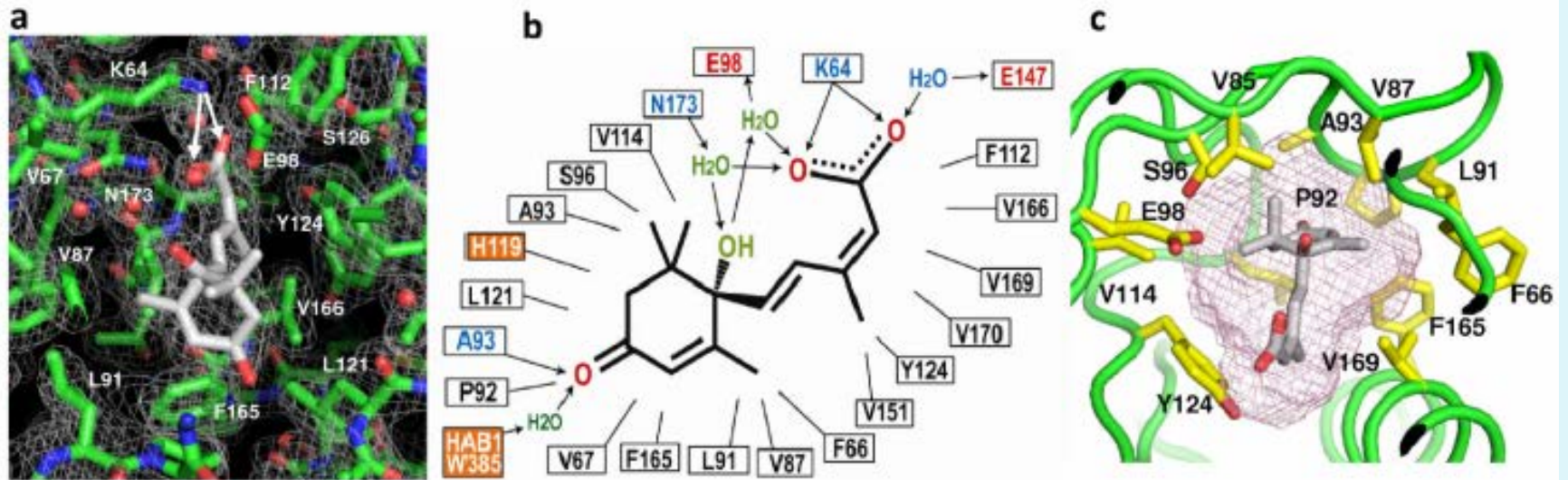


PYL2-ABA-PP2C 三元复合体



A Gate-Latch-Lock Mechanism

Nature. 2009 December 3; 462(7273): 602–608. doi:10.1038/nature08613.





总结

- ▶ 1、许多原因导致很多分析还没进行（如核酸的限制酶内切位点分析、引物设计、蛋白质保守位点的分析等），在后续会继续进行。
- ▶ 2、我们的预测结果显示：预测结果与实际结果是有差距的；所以要辩证的使用软件分析，要综合的分析所得到的结果，并通过试验进行验证。
- ▶ 3、一定要与别人多交流!



感谢罗老师的悉心指导！

感谢G09组的所有成员共同努力！

感谢同学们的帮助！



谢谢大家！

欢迎大家批评指正！