



Applied Bioinformatics Course

Bioinformatics Analysis of E3 Ubiquitin Ligase in *Oryza sativa*

Group 12: 江红梅、任磊、
师翠兰、周晶

Reporter: 任磊

Content

1. Background

2. Reason

3. Gene Analysis

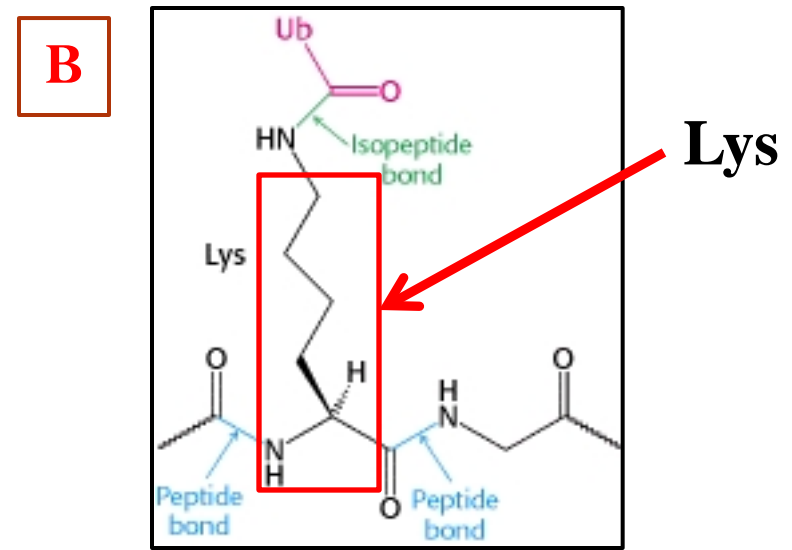
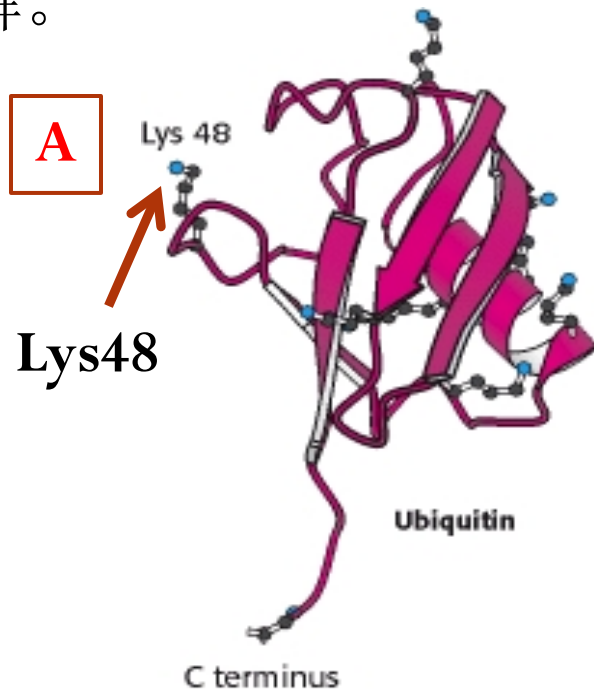
4. Protein Analysis

5. Acknowledgement

Background

1. What is ubiquitin and ubiquitin ligase

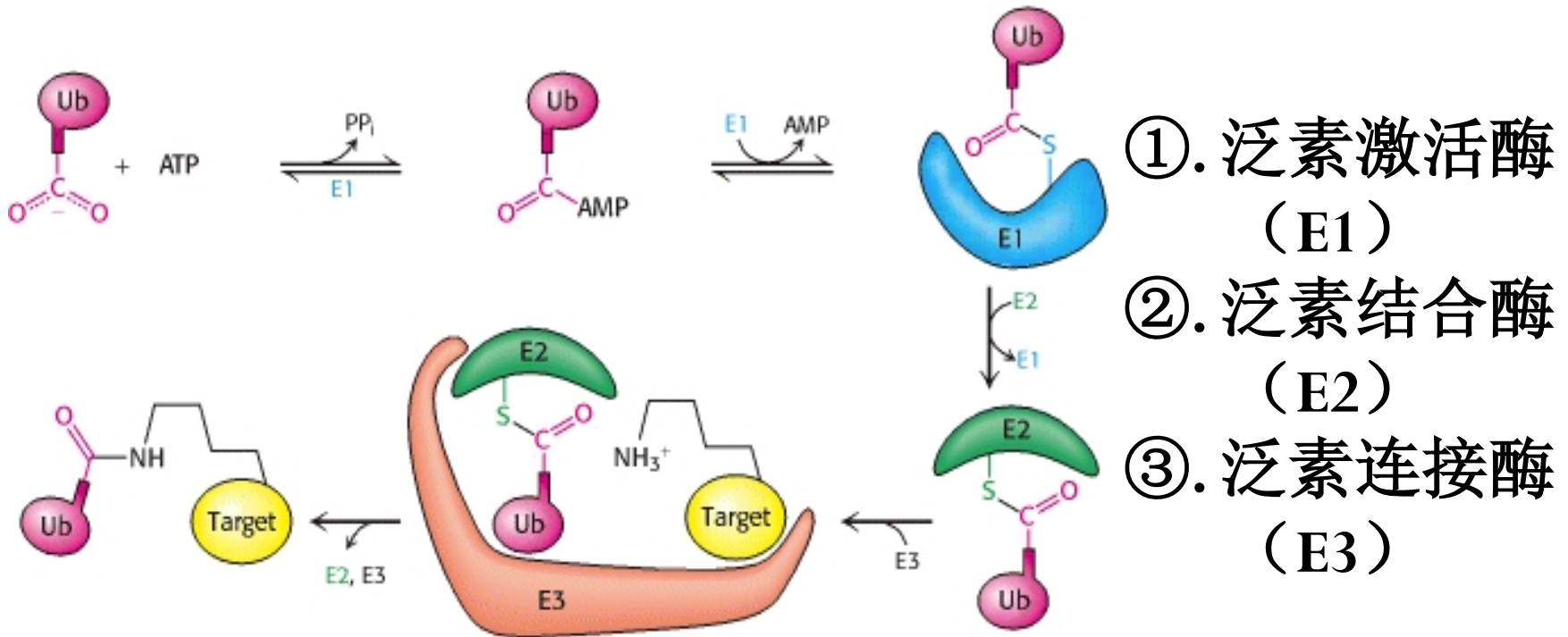
泛素 (Ubiquitin, Ub): 是一种存在于大多数真核细胞中的小蛋白, 它的主要功能是标记需要分解掉的蛋白质, 使其被水解。



A为Ub的结构图, **B**为Ub与目标蛋白的结合方式

Background

泛素化(Ubiquitination), 一共有三个酶参与



- 特点:
- ① 依赖于ATP
 - ② E1种类较少, 多数真核只有一种或很少差别的E1
 - ③ E2、E3种类较多, 但E2来自于同一家族, E3家族较多

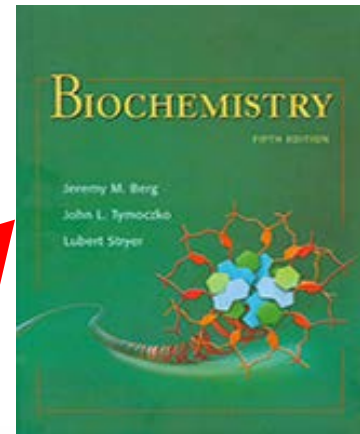
Background

泛素连接酶E3主要有三大类：

主要根据连接酶中包含的结构域进行划分。

- ①、HECT结构域家族；
- ②、RING结构域家族；
- ③、U.box家族

以上内容来自NCBI-Bookshelf中的
Biochemistry 5th edition



2004年，阿龙·切哈诺沃、阿夫拉姆·赫什科、欧文·罗斯因发现了泛素调节的蛋白质降解过程而获得了诺贝尔化学奖。

Reason

ARTICLES

nature
genetics

A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase

Xian-Jun Song^{1,2}, Wei Huang^{1,2}, Min Shi¹, Mei-Zhen Zhu¹ & Hong-Xuan Lin¹

文献背景:

中科院上海生物所一个小组，检测到一个控制水稻粒型和粒重的主效QTL-GW2。经过精细定位发现一个编码RING-E3泛素连接酶的新基因。研究发现当该基因功能缺失时水稻籽粒变宽，粒重增加。因此该基因与水稻产量密切相关。

Reason

存在问题:

该研究虽然证明了该基因（GW2）的表达与否影响了水稻籽粒的大小，但是，正如文中Discussion中描述的，该基因和所产生的RING-E3泛素连接酶的遗传机制与作用机制所知甚少，且该酶的蛋白结构并未得到解析。

我们分析的目标:

- *1. 从核酸水平对该基因进行初步分析;
- *2. 从蛋白水平分析序列、结构特征;
- *3. 尝试利用同源模建，分析该酶的高级结构、结合位点，以及核心作用区域等。

目的：为研究该酶的作用机制提供参考。

Reason

分析基础:

- ①、该基因全长，6937bp;
- ②、mRNA序列，1278bp;
- ③、蛋白序列425aa。



采取的策略:

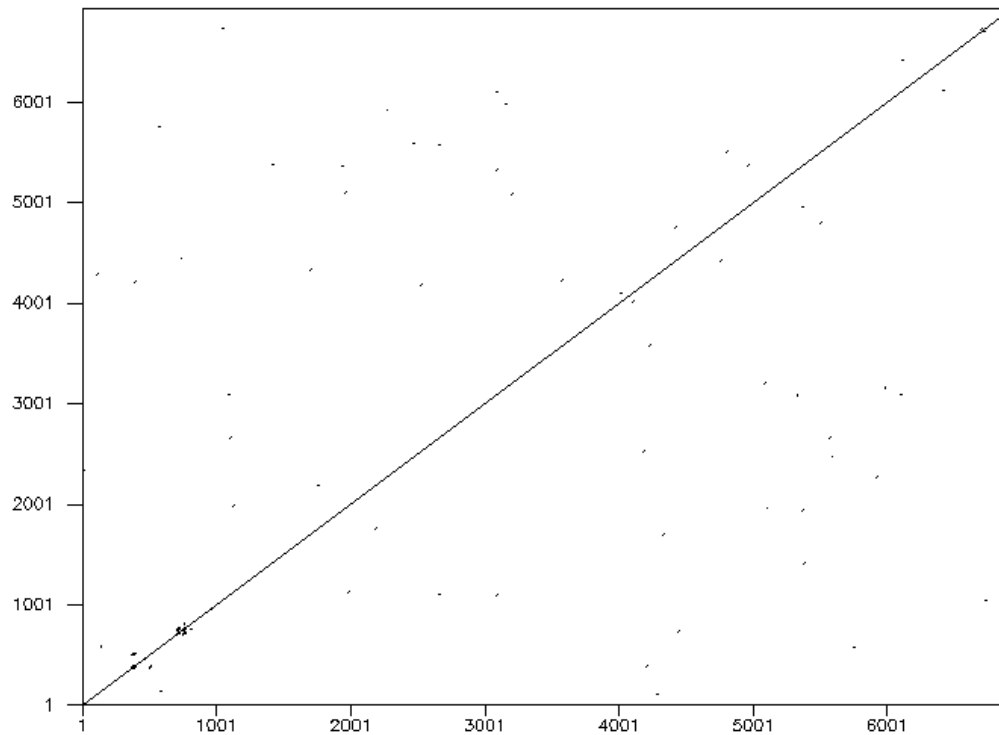
使用生物信息学软件预测分析，并与已有信息进行比较；
同时，利用生物信息学工具对未知信息进行预测。

Gene Analysis

使用Dottup查看基因结构

Dottup: raw::682897 vs raw::682898

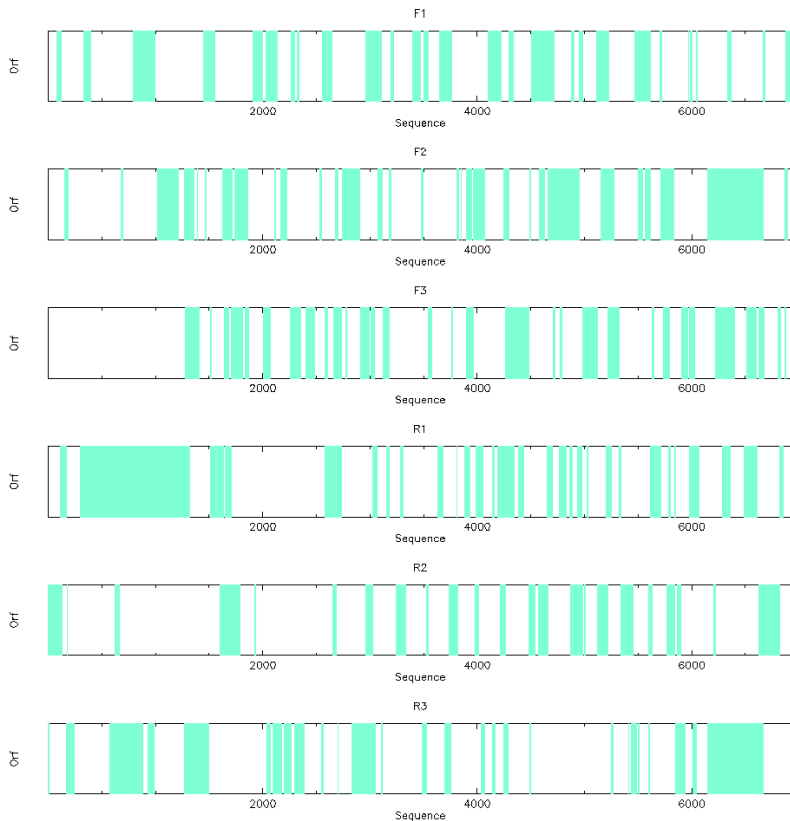
Sat 30 Nov 2013 21:15:41



无重复结构

Gene Analysis

使用多种方法预测ORF



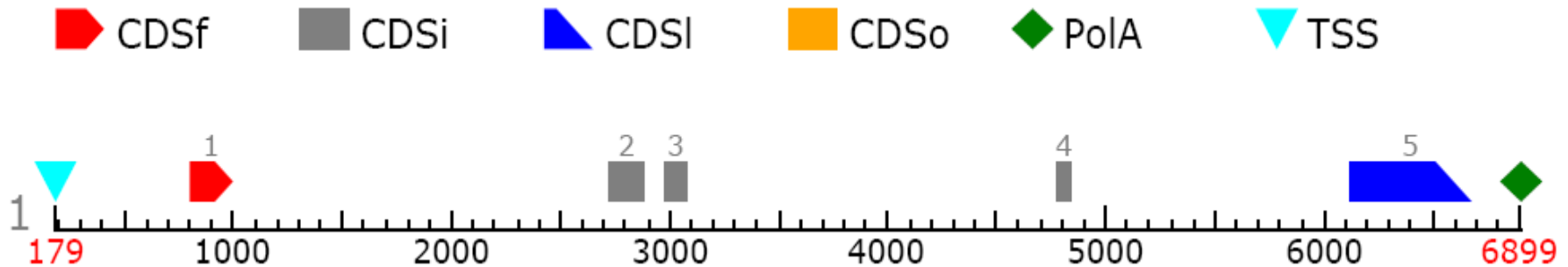
Plotorf

```
>_1 [403 - 996]
ATEVAVSLAVAAAE TRPLFSSLLPFLVHSVAASCCSSREVSTLSSSHPEKAKKRKI IFQK
KSSPRPRVVVPI TPPPPLRVRVRIPPPPSPPPYVSTVRRRRRRRRRRRSGEGFICGGGGR
RREBGRSGRGMGNRIGGRRKAGVEERYTRPQGLYHRDIDQKCLRKLILEAKLAPCYMGA
DDAAAAADLEECPICFV
>_2 [5936 - 6670]
AFFLLYSLLLERCKLYVKCCFIFEFRFLGINNKQSSCIAYRKVCIYFPNTVNLFSEQEQ
GSIGNPVCGNFMPVTEPSPRERQPFVPAASLEIPHGGGFSCAVAAMAHQPPSMDFSYMA
GSSAFPVDFMFRPCNIAGGSMCNLESSPESWSGIAPSCSREVVRREEGCSADHWSEGAE
AGTSYAGSDIVADAGTMPQLPFAENFAMAPSHFRPESIBEQMMF SMALSLADGHGRTHSQ
GLAWL
>_3 [6683 - 6150] (REVERSE SENSE)
SALPTTMTPLASVFFHDHLLKKEPWKTSVLRCSFGGSGLAPWRSSRRKAAAASRHLPR
CLSPRSLFRPLHPPSTSGQPSTLLLLLPCCSLVLSRSSPVSSPDYTCFHQQCCMVAGT
CRKLGMRCCQPCKRSPYVWAGAQPLQPHRKIHHHEVFLKQLERMAGVHAEMAQLQA
>_4 [1359 - 298] (REVERSE SENSE)
LSKNNTNPKTVPMSTEILAPKHCWSSISITATHRRHSPQSNRSKSNQPPNRNAPRHHP
ATRNPRPTLATTSTSSSITTKSHQKREKLLKWKKEAGGGSEEEGRDLFLPSTRHISPEAV
LYQEADGALLEVGGGGVVGAVHVARRELGLDQLPELLLDVDPVLVQPLRPRVPLLPRL
PPPPYPVPHPPRSTLLPPPPSSSADETLTPPPPPPPPPPHRRYIRWWWRRWWWYTHAY
SERRRGDGDDDARARRAFFLKNNSFFGFLGVGGKSGDLSAAATASRYAMHKRKRKR
EERPCLCRDCERDCDLSLAAAAAAAIVAIISVPEPHSRSLQRGGGLPV
```

Getorf

Gene Analysis

使用Softberry预测ORF

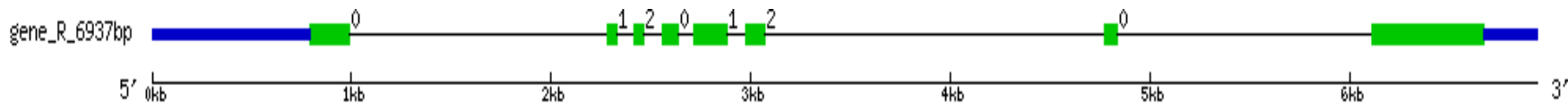


1	+	TSS	179			-1.68			
1	+	1 CDSf	793	-	993	24.77	793	-	993 201
1	+	2 CDSi	2717	-	2879	6.93	2717	-	2878 162
1	+	3 CDSi	2972	-	3074	9.17	2974	-	3072 99
1	+	4 CDSi	4771	-	4837	13.74	4772	-	4837 66
1	+	5 CDSl	6110	-	6673	29.06	6110	-	6673 564
1	+	PoIA	6899			-1.06			

预测结果得到5个外显子， ± 1098 bp
的CDS，编码365个氨基酸

Gene Analysis

实际的ORF分析



Legend:

■ exon ■ marked region — intron ■ upstream/downstream ◊ 1 2: intron phase

图中详细显示了外显子在序列中的分析，一共有8个外显子。

Gene Analysis

对ORF进行初步的统计分析

#CdsCount: 1

#Coding GC 52.82%

#1st letter GC 52.82%

#2nd letter GC 48.59%

#3rd letter GC 57.04%

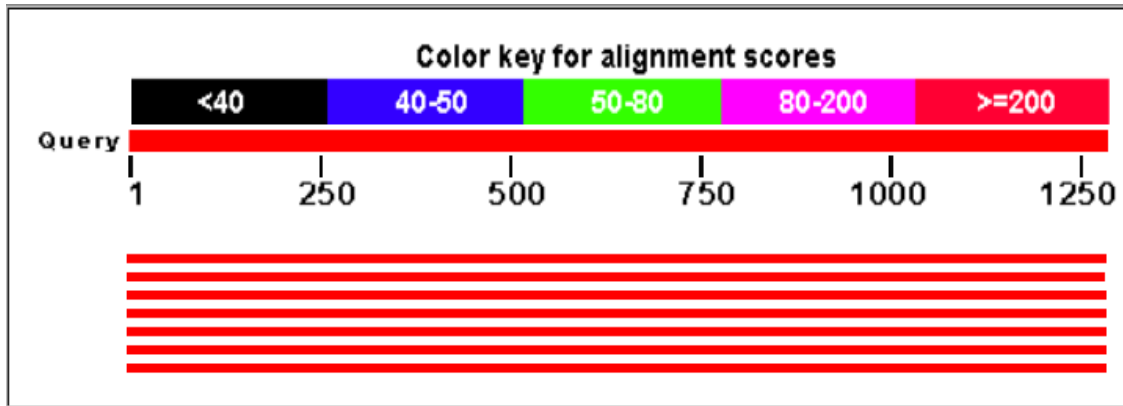
```
BsaXI
| MwoI
| BglI      MwoI
| BstMWI    BstMWI
| HpyF10VI  HpyF10VI
\ \
CAGGGGCTGTACGAGCACAGGGACATCGACCAGAAGAAGCTCCGGAAGCTGATCCTCGAG
-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|
              70           80           90           100          110          120
GTCCCCGACATGCTCGTGTCCCTGTAGCTGGTCTTCTTCGAGGCCTTCGACTAGGAGCTC
          /                               /
          HpyF10VI                       HpyF10VI
          BstMWI                         BstMWI
          MwoI                           MwoI
```

CUSP对密码子使用情况分析

Remap对限制性酶切位点分析

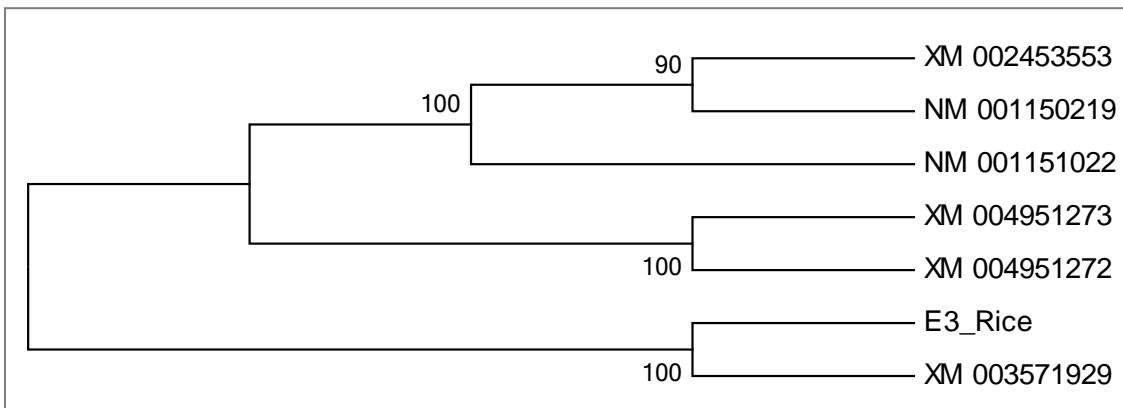
Gene Analysis

对ORF进行初步的统计分析



对mRNA序列进行比对，结果中所包含的多为未注释的基因。

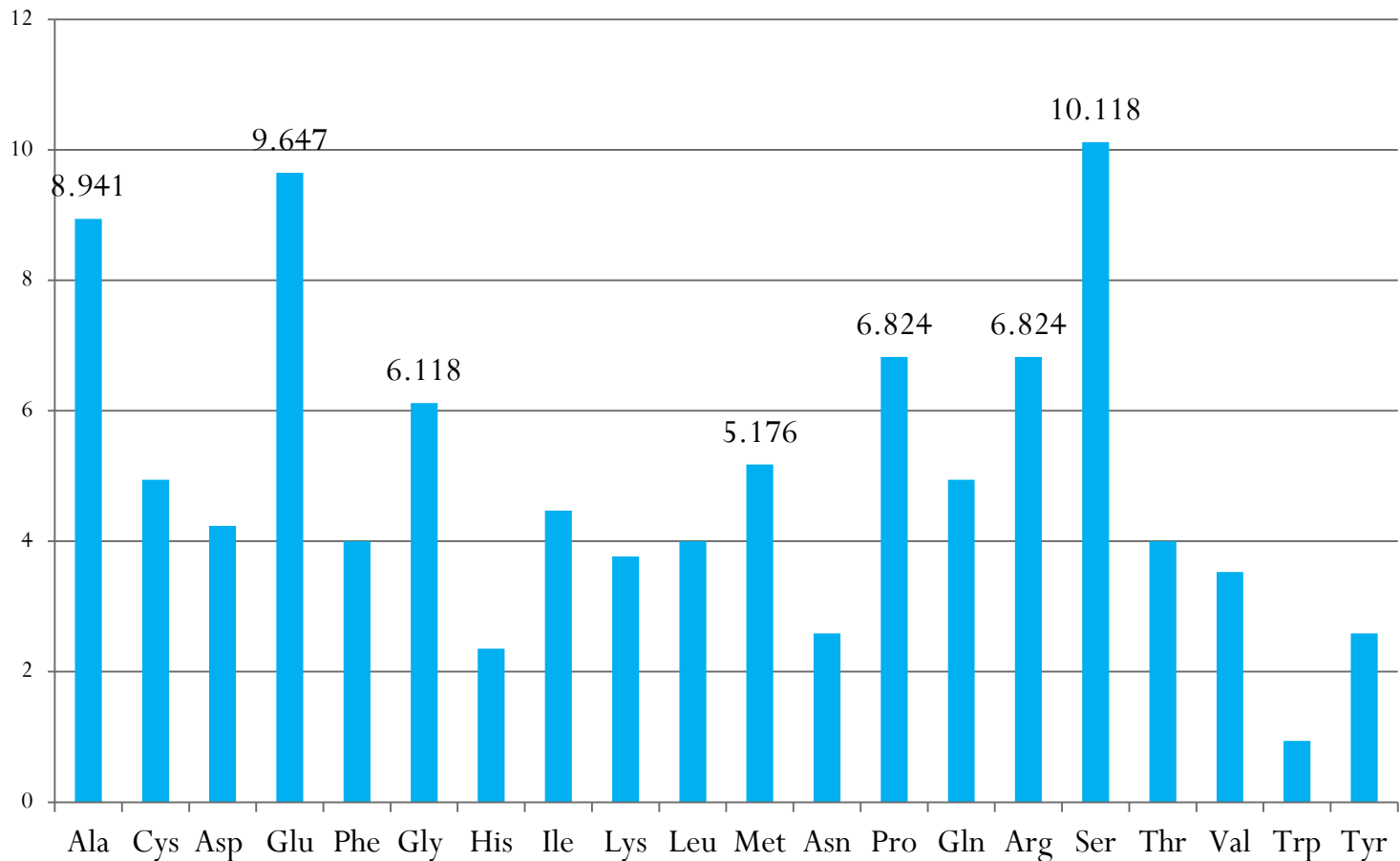
Database: NCBI_refseq, E-value=0.001



对比对结果进行建树分析，可以发现，这个基因普遍存在于水稻、杂草、玉米、高粱等植物中。

Protein Analysis

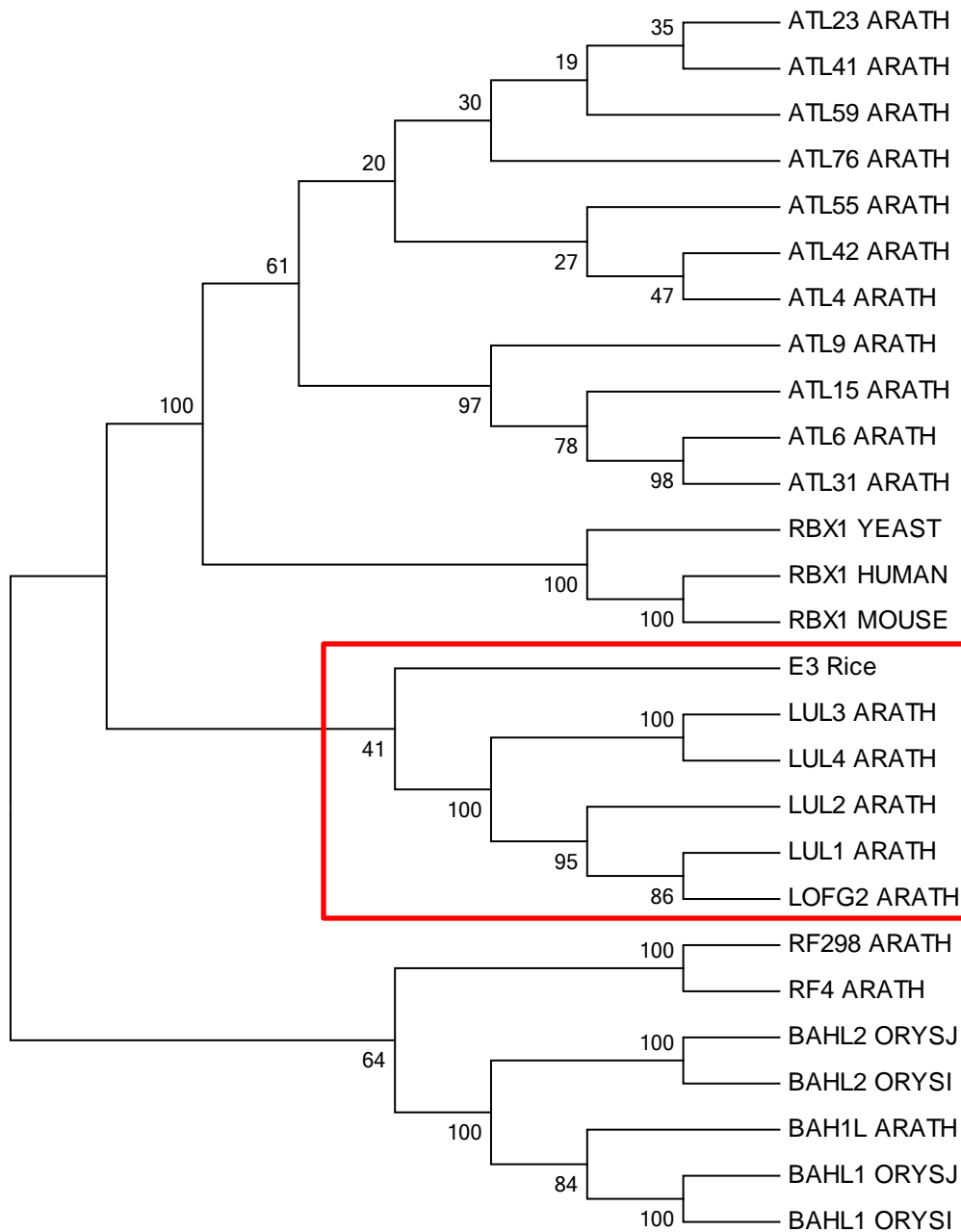
利用Pepstats分析氨基酸组成



Protein Analysis

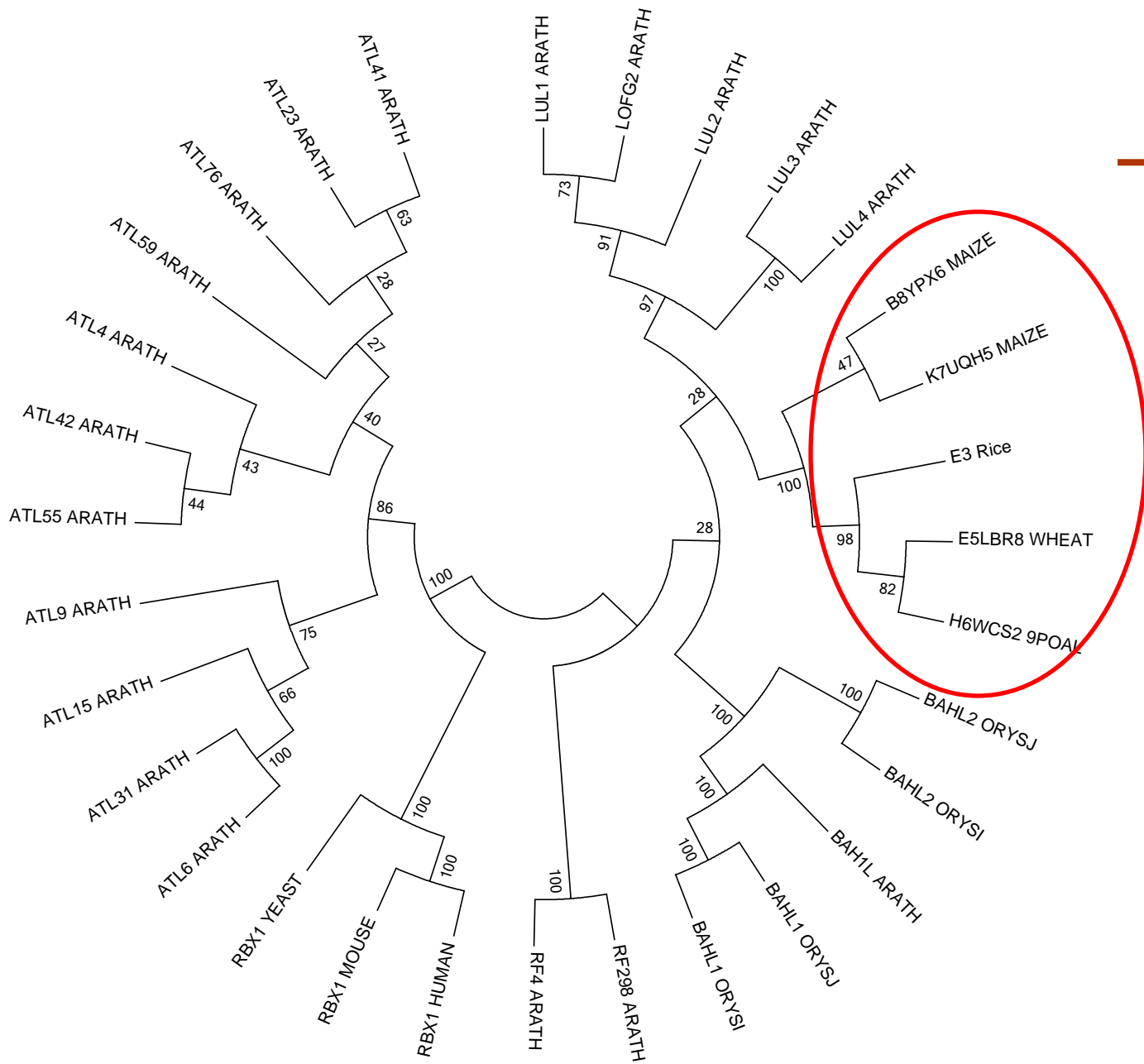
利用Pepstats分析氨基酸组成

Property	Residues	Number	Mole%
Tiny	(A+C+G+S+T)	145	34.118
Small	(A+B+C+D+G+N+P+S+T+V)	218	51.294
Aliphatic	(A+I+L+V)	89	20.941
Aromatic	(F+H+W+Y)	42	9.882
Non-polar	(A+C+F+G+I+L+M+P+V+W+Y)	219	51.529
Polar	(D+E+H+K+N+Q+R+S+T+Z)	206	48.471
Charged	(B+D+E+H+K+R+Z)	114	26.824
Basic	(H+K+R)	55	12.941
Acidic	(B+D+E+Z)	59	13.882



直接在Uniprot中检索E3泛素连接酶RING家族，得到26个结果，将得到结果与目标序列比较，建树。

Uniprot检索策略：
Family:RING and name:E3
and reviewed



以目标序列在 Uniprot 中进行 BLAST，获得相似性较高的 E3 泛素连接酶添加到之前比对序列中，重新建树。

Protein Analysis

以氨基酸序列比对分析进行建树分析

获得的E3 泛素连接酶（RING），与其他已报道RING家族E3泛素连接酶（reviewed）关系较远；

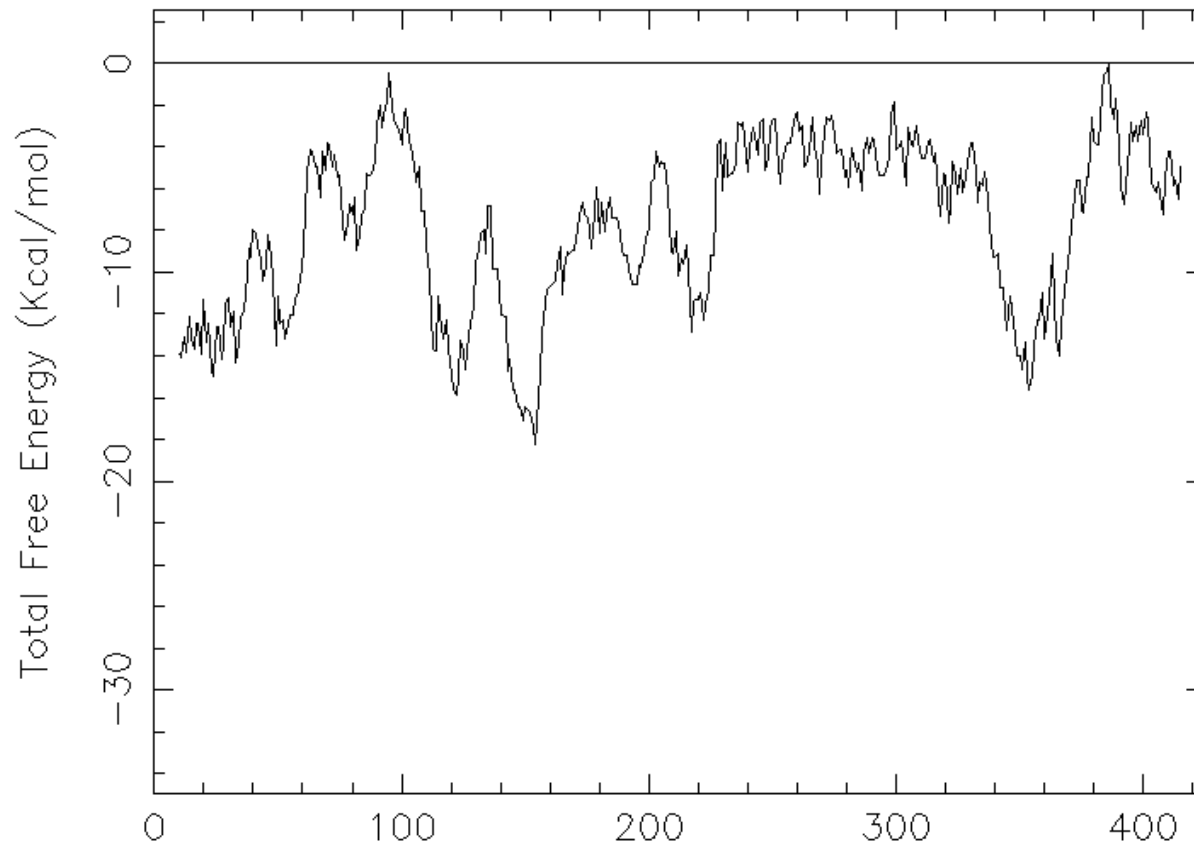
然而与其他部分未通过Swissprot审核的E3泛素连接酶（RING）关系很近。

初步的判断：所得到的的E3 泛素连接酶（RING）是一类不同于之前所报道的E3泛素链激酶（RING），且广泛存在于植物中（玉米、小麦、水稻、高粱等都有报道）。

Protein Analysis

利用Octanol分析蛋白亲疏水性

White-Wimley Plot



可以看出，该蛋白有较高的亲水性。

Protein Analysis

利用Compute pI/Mw分析蛋白等电点和分子量



Expasy

Bioinformatics Resource Portal

Compute pI/Mw

Compute pI/Mw

Theoretical pI/Mw (average) for the user-entered sequence:

<u>10</u>	<u>20</u>	<u>30</u>	<u>40</u>	<u>50</u>	<u>60</u>
MGNRIGGRRK	AGVEERYTRP	QGLYEHRDID	QKKLRKLILE	AKLAPCYMGA	DDAAAAADLE
<u>70</u>	<u>80</u>	<u>90</u>	<u>100</u>	<u>110</u>	<u>120</u>
ECPICFLYYP	SLNRSKCCSK	GICTECFLQM	KPTHTAQPTQ	CPFCKTPSYA	VEYRGVKTKE
<u>130</u>	<u>140</u>	<u>150</u>	<u>160</u>	<u>170</u>	<u>180</u>
ERSIEQFEEQ	KVIEAQMRMR	QQALQDEEDK	MKRKQNRCS	SRTITPTKEV	EYRDICSTSF
<u>190</u>	<u>200</u>	<u>210</u>	<u>220</u>	<u>230</u>	<u>240</u>
SVPSYRCAEQ	ETECCSSEPS	CSAQTSMPF	HSRHRNDDNI	DMNIEDMMVM	EAIWRSIQEQ
<u>250</u>	<u>260</u>	<u>270</u>	<u>280</u>	<u>290</u>	<u>300</u>
GSIGNPFCGN	FMPVTEPSR	ERQPFVPAAS	LEIPHGGGFS	CAVAAMAEHQ	PPSMDFSYMA
<u>310</u>	<u>320</u>	<u>330</u>	<u>340</u>	<u>350</u>	<u>360</u>
GSSAFPVFD	FRRPCNIAGG	SMCNLESSPE	SWSGIAPSCS	REVVREEGEC	SADHWSEGAE
<u>370</u>	<u>380</u>	<u>390</u>	<u>400</u>	<u>410</u>	<u>420</u>
AGTSYAGSDI	VADAGTMPQL	PFAENFAMAP	SHFRPESIEE	QMMFSMALS	ADGHGRTHSQ

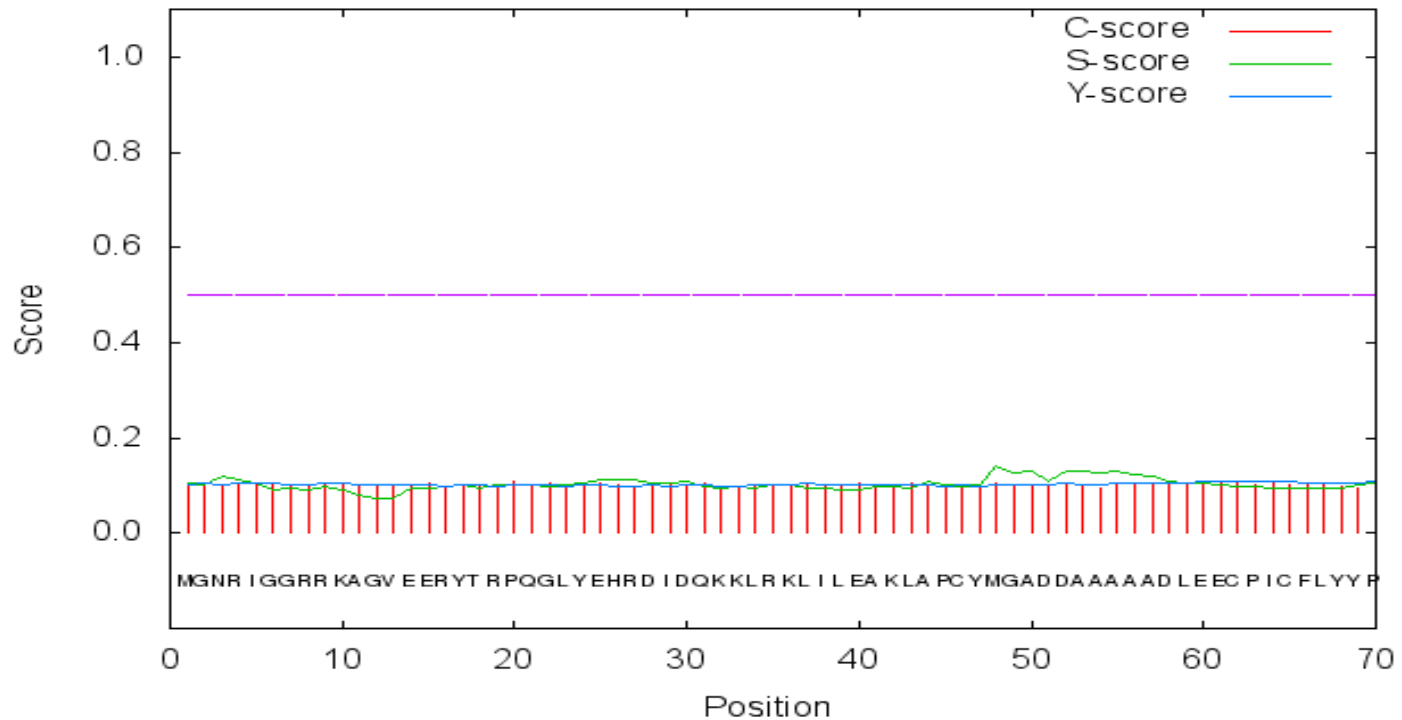
SLAWL

Theoretical pI/Mw: 5.32 / 47399.41

Protein Analysis

利用Signal P预测是否有信号肽

SignalP-4.1 prediction (euk networks): Sequence



#	Measure	Position	Value	Cutoff	signal peptide?
	max. C	20	0.108		
	max. Y	64	0.110		
	max. S	48	0.141		
	mean S	1-63	0.104		
	D	1-63	0.107	0.450	NO

Protein Analysis

亚细胞定位



SeqID: E3

Analysis Report:

CMSVM_a	Unknown	[No details]
CWSVM_a	Unknown	[No details]
CytoSVM_a	Cytoplasmic	[No details]
ECSVM_a	Unknown	[No details]
ModHMM_a	Unknown	[No internal helices found]
Motif_a	Unknown	[No motifs found]
Profile_a	Unknown	[No matches to profiles found]
SCL-BLAST_a	Unknown	[No matches against database]
SCL-BLASTe_a	Unknown	[No matches against database]
Signal a	Unknown	[No signal peptide detected]

Localization Scores:

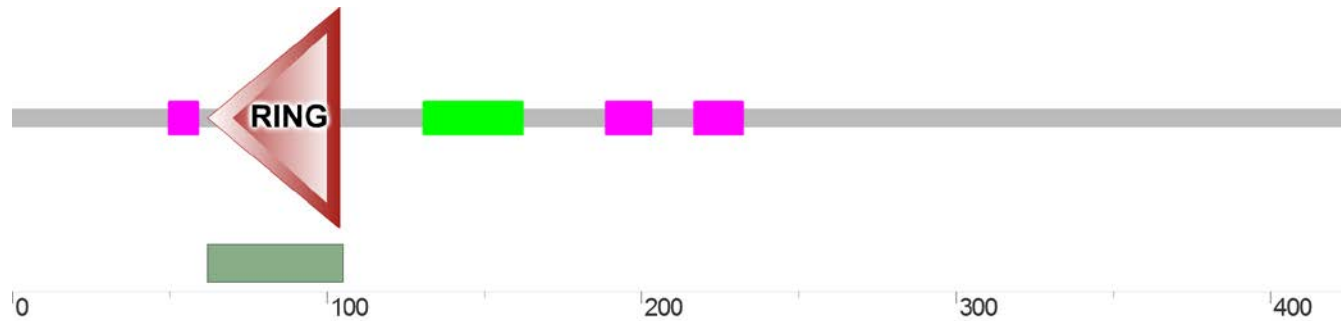
Cytoplasmic	7.50
CytoplasmicMembrane	1.00
Cellwall	0.87
Extracellular	0.63

Final Prediction:
Cytoplasmic 7.50

定位的结果：存在于细胞质中

Protein Analysis

利用SMART分析保守结构域



Confidently predicted domains, repeats, motifs and features:

Name	Start	End	E-value
low complexity	50	58	N/A
RING	62	104	0.726
coiled coil	131	161	N/A
low complexity	189	202	N/A
low complexity	217	231	N/A

该序列中存在一个RING保守结构域。

Protein Analysis

对RING结构域进行多序列比对分析

Species/Abbrv	*	*									*				*							*	*																														
1. E3	C	F	I	C	F	L	Y	Y	S	---	L	N	S	K	C	C	S	K	-	G	I	C	T	E	C	F	L	Q	M	R	S	---	T	H	I	A	D	P	T	C	C	F	C	-									
2. RING_ARATH	C	F	I	C	L	D	T	L	F	N	---	F	Y	A	L	S	-	C	G	H	-	L	F	C	R	G	C	A	C	G	A	A	S	V	Y	I	F	Q	G	V	K	S	A	F	F	E	A	K	C	F	V	C	H
3. RING_RICE	C	S	V	C	L	S	K	F	E	S	V	E	---	I	L	R	L	L	E	K	R	H	-	A	F	H	I	G	C	I	D	---	Q	W	L	E	Q	H	A	T	C	E	L	C	R								
4. RING_ARATH(2)	C	A	I	C	L	N	E	F	E	D	D	E	---	T	L	R	L	L	P	K	C	D	H	-	V	F	H	P	H	C	I	G	---	A	W	L	Q	G	H	V	T	C	F	V	C	H							
5. RING_RICE(2)	C	F	I	C	L	D	T	L	F	N	---	F	Y	A	L	S	-	C	G	H	-	L	F	C	R	G	C	A	C	G	A	A	S	V	Y	I	F	Q	G	V	K	S	A	F	F	E	A	K	C	F	V	C	H
6. RING_ARATH(3)	C	V	V	C	L	S	E	F	F	D	---	T	T	V	L	E	-	C	R	H	M	C	M	C	S	G	C	A	K	---	A	L	R	F	Q	T	-	N	L	C	F	V	C	H									
7. RING_RICE(3)	C	S	I	C	L	D	T	V	F	D	---	F	V	A	L	S	-	C	G	H	-	I	Y	C	Y	L	C	S	C	S	A	A	S	V	T	I	V	D	G	L	K	S	A	E	R	K	S	K	C	F	L	C	H
8. RING_ARATH(4)	C	V	M	C	L	S	E	E	M	S	---	V	I	F	L	E	-	C	A	H	Q	V	L	C	S	K	C	N	Q	---	L	H	E	K	E	A	M	E	D	C	F	S	C	H									
9. RING_RICE(4)	C	F	I	C	L	D	T	L	F	N	---	F	Y	A	L	S	-	C	G	H	-	L	F	C	R	G	C	A	C	G	A	A	S	V	Y	I	F	Q	G	V	K	S	A	F	F	E	A	K	C	F	V	C	H
10. RING_HUMAN	C	I	E	C	Q	A	N	Q	A	S	A	T	S	E	E	C	T	V	A	W	G	V	C	N	H	-	A	F	H	F	H	C	I	S	---	R	W	L	R	T	R	Q	V	C	F	L	D	N					
11. RING_MOUSE	C	I	E	C	Q	A	N	Q	A	S	A	T	S	E	E	C	T	V	A	W	G	V	C	N	H	-	A	F	H	F	H	C	I	S	---	R	W	L	R	T	R	Q	V	C	F	L	D	N					

Protein Analysis

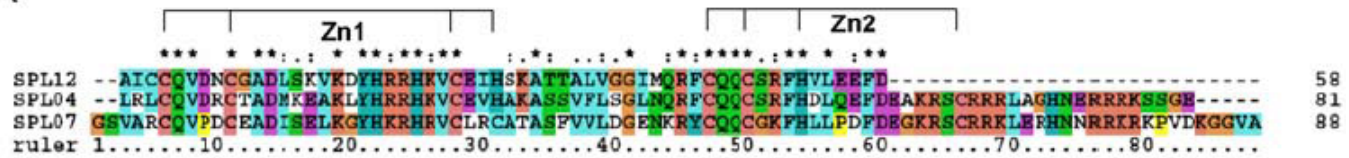
对RING结构域进行多序列比对分析

Species/Abbrv	*	*								*																								*	*	*																	
1. E3	C	P	I	C	F	L	Y	Y	F	S	---	L	N	R	S	K	C	C	S	K	-	G	I	C	T	E	C	F	L	Q	M	K	---	F	T	H	T	A	Q	P	T	Q	C	P	F	C	-						
2. RING_ARATH	C	P	I	C	L	D	T	L	F	N	---	F	Y	A	L	S	-	C	G	H	-	L	F	C	K	G	C	A	C	G	A	A	S	V	Y	I	F	Q	G	V	K	S	A	P	P	E	A	K	C	P	V	C	R
3. RING_RICE	C	S	V	C	L	S	K	F	E	S	V	E	I	L	R	L	L	P	K	C	R	H	-	A	F	H	I	G	C	I	D	---	Q	W	L	E	Q	H	A	T	C	P	L	C	R								
4. RING_ARATH(2)	C	A	I	C	L	N	E	F	E	D	E	T	L	R	L	L	P	K	C	D	H	-	V	F	H	P	H	C	I	G	---	A	W	L	Q	G	H	V	T	C	P	V	C	R									
5. RING_RICE(2)	C	P	I	C	L	D	T	L	F	N	---	F	Y	A	L	S	-	C	G	H	-	L	F	C	K	G	C	A	C	G	A	A	S	V	Y	I	F	Q	G	V	K	S	A	P	P	E	A	K	C	P	V	C	R
6. RING_ARATH(3)	C	V	V	C	L	S	E	E	F	A	D	---	T	T	V	L	L	P	K	C	R	H	M	C	M	C	S	G	C	A	K	---	A	L	R	F	Q	T	-	N	L	C	P	V	C	R							
7. RING_RICE(3)	C	S	I	C	L	D	T	V	F	D	---	F	V	A	L	S	-	C	G	H	-	I	Y	C	Y	L	C	S	C	S	A	A	S	V	T	I	V	D	G	L	K	S	A	E	R	K	S	K	C	P	L	C	R
8. RING_ARATH(4)	C	V	M	C	L	S	E	E	M	S	---	V	I	F	L	P	-	C	A	H	Q	V	L	C	S	K	C	N	Q	---	L	H	E	K	E	A	M	E	D	C	P	S	C	R									
9. RING_RICE(4)	C	P	I	C	L	D	T	L	F	N	---	F	Y	A	L	S	-	C	G	H	-	L	F	C	K	G	C	A	C	G	A	A	S	V	Y	I	F	Q	G	V	K	S	A	P	P	E	A	K	C	P	V	C	R

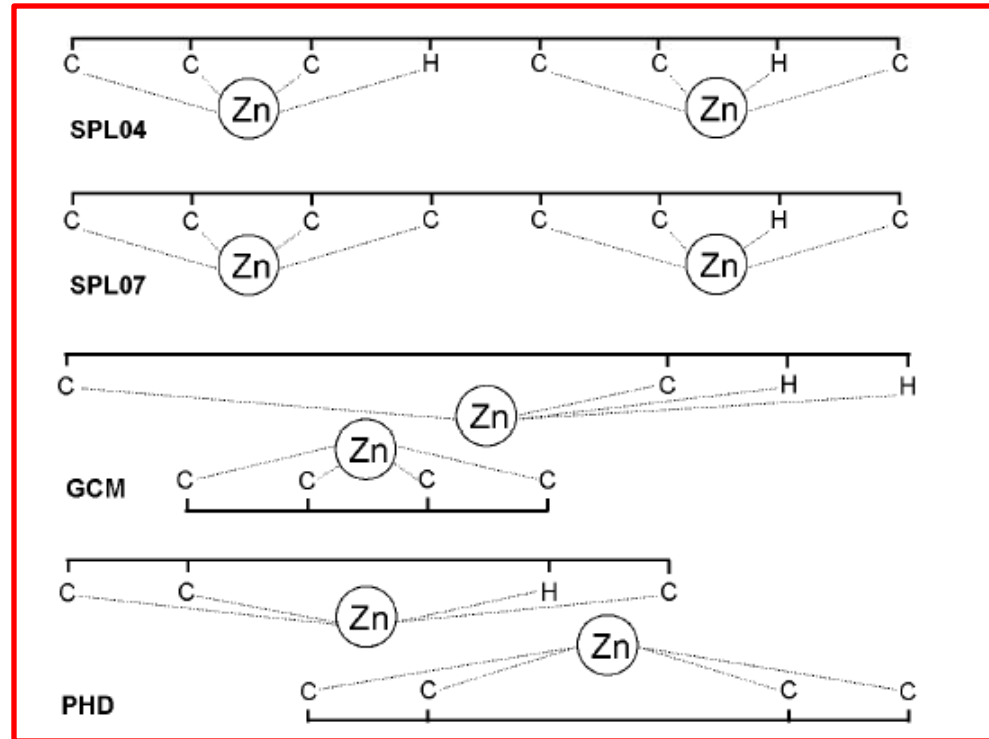
Protein Analysis

根据序列比对预测分析锌指结构

A



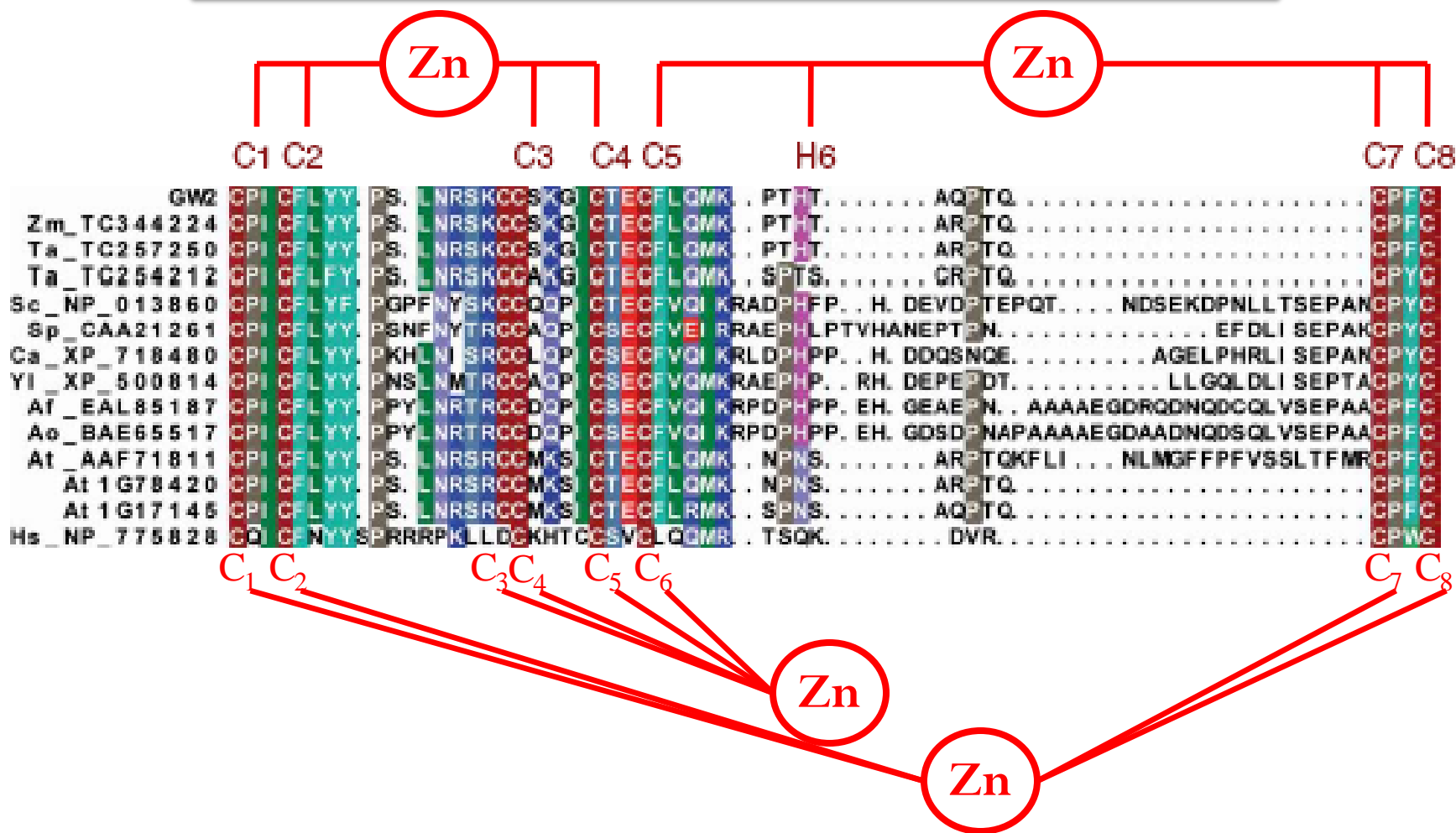
B



不同的结构里锌指结构会有所差异

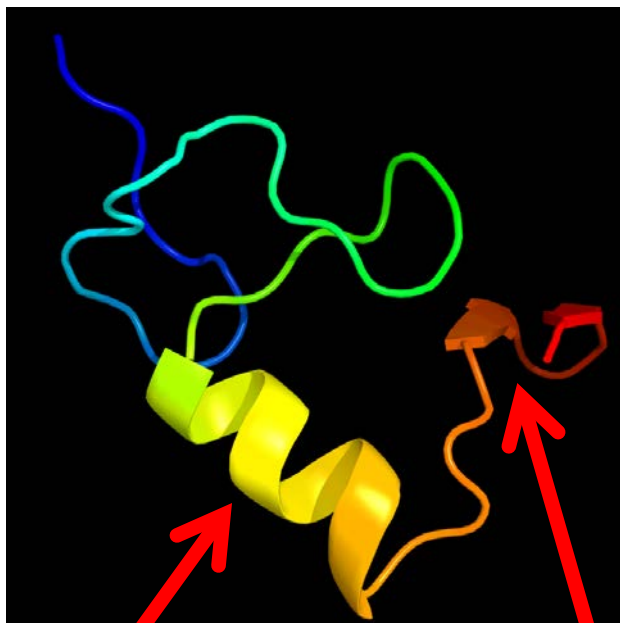
Protein Analysis

根据序列比对预测分析锌指结构

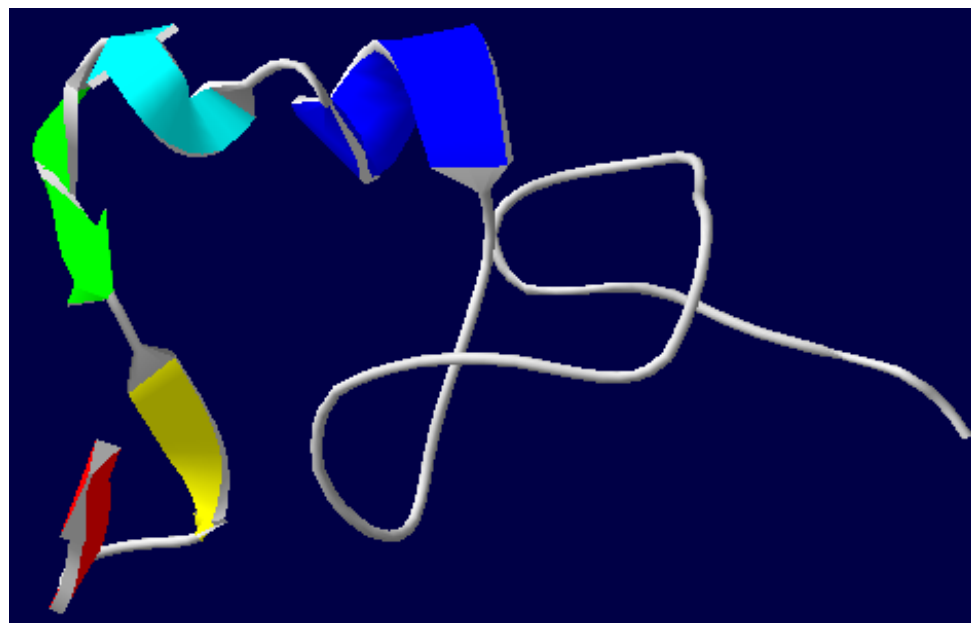


Protein Analysis

使用Phyre2预测对蛋白结构进行预测



螺旋结构



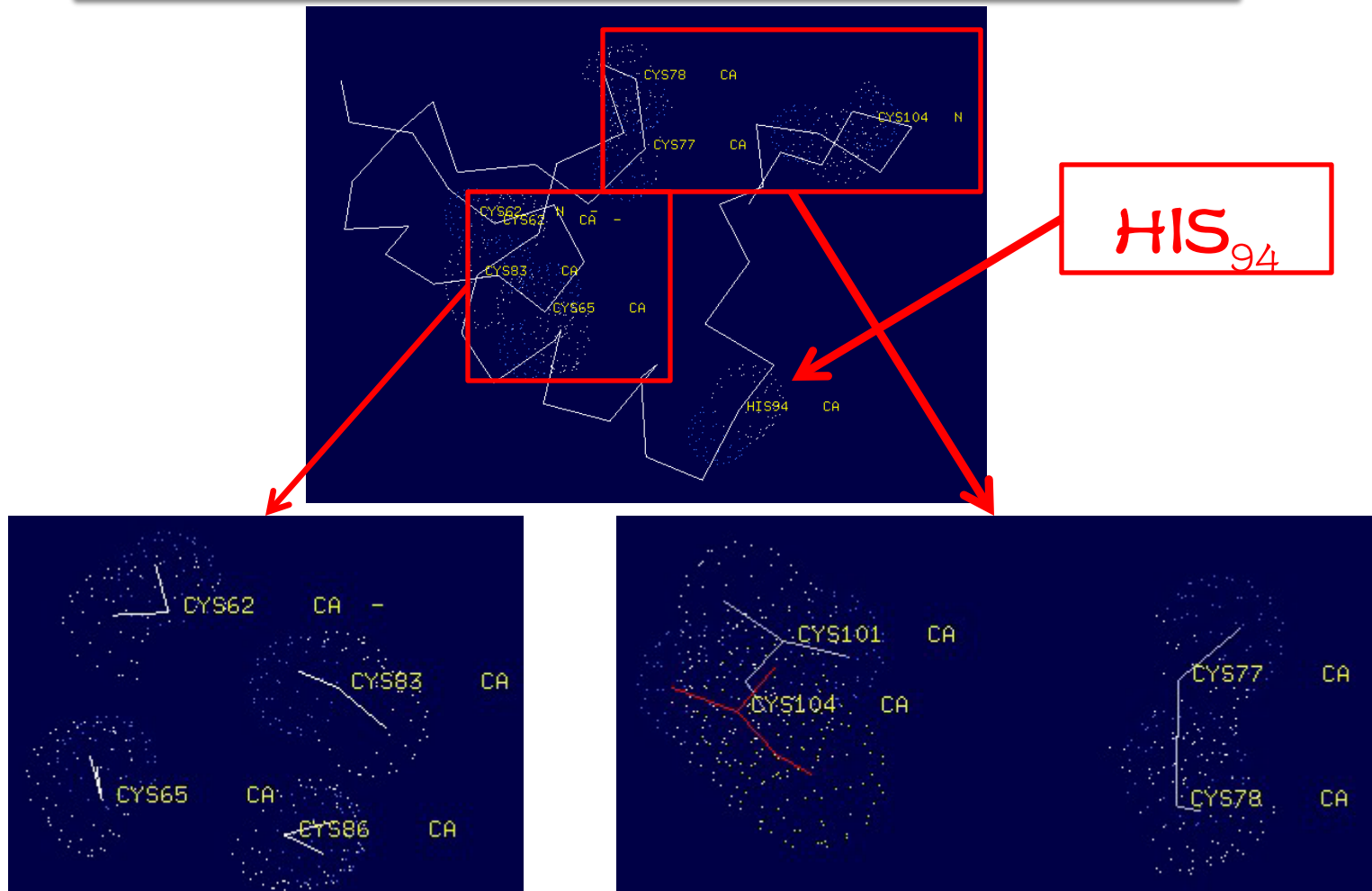
折叠结构

实际只对58-107位置进行了预测

Phyre²

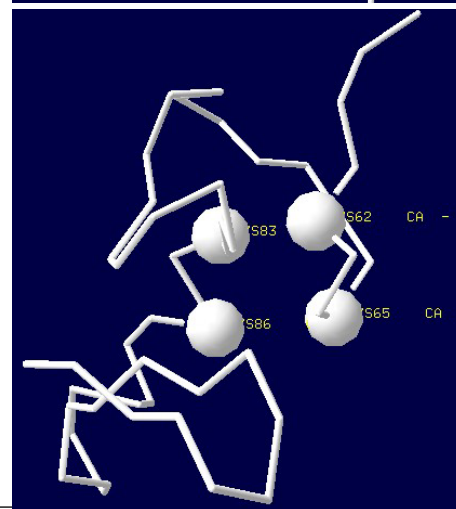
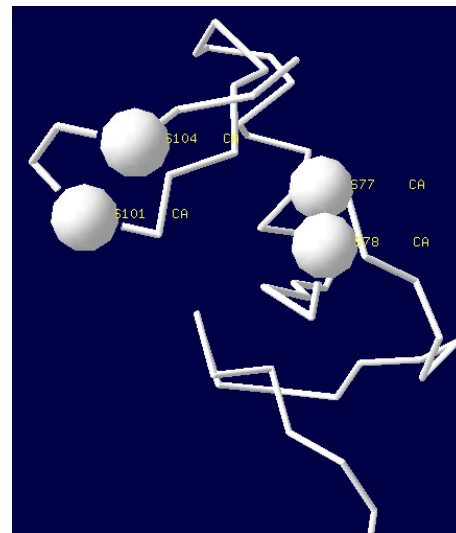
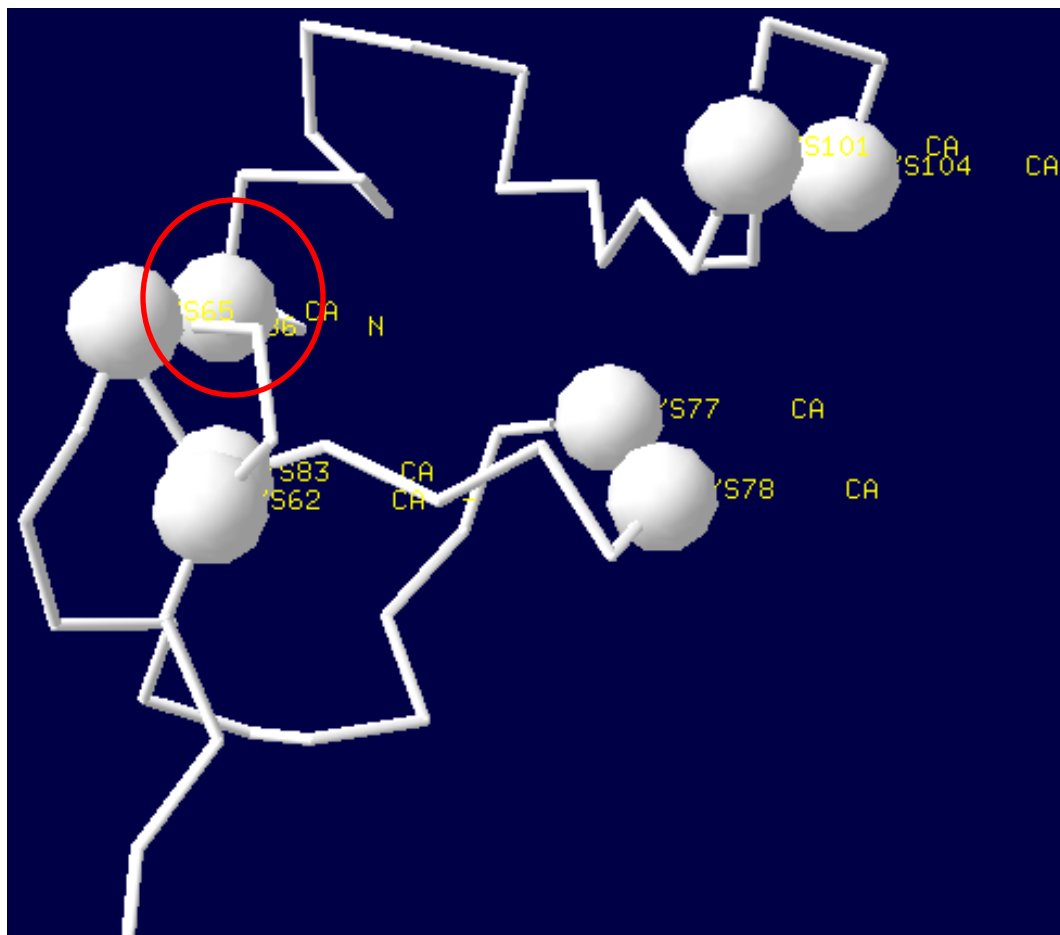
Protein Analysis

使用SPDB-Viewer寻找锌指结构



Protein Analysis

使用SPDB-Viewer寻找锌指结构



Conclusion

- 一、分别从Gene、mRNA和氨基酸水平对序列信息进行了初步的统计分析；
- 二、从蛋白序列水平分析了结构特征信息；
- 三、通过蛋白结构预测软件对RING结构域进行了预测分析，并分析了其中的核心结构区域。

Reference

1. Xian-Jun Song, Wei Huang, Min Shi etc., QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. *Nature genetics*, 39(5), 2007
2. NCBI_bookshelf, Biochemistry 5th edition
3. Parsimony Methods. Molecular Biology and Evolution Tamura K, Peterson D, Peterson N, Stecher G etc., MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution*. 2011, 28: 2731-2739.

Acknowledgement

感谢师翠兰提供的序列信息；

感谢周晶、江红梅、师翠兰为课题分析完成所做的努力；

感谢罗老师一直以来认真、严谨的教导；

同时感谢大家对我的支持！

