

用生物信息技术分析拟南芥中的 病程相关蛋白PR1

The bioinformatic analysis of PR1 in *Arabidopsis thaliana*

汇报人：张亚林

小组成员：邵冰欣 刘利彩

张天豹 张亚林

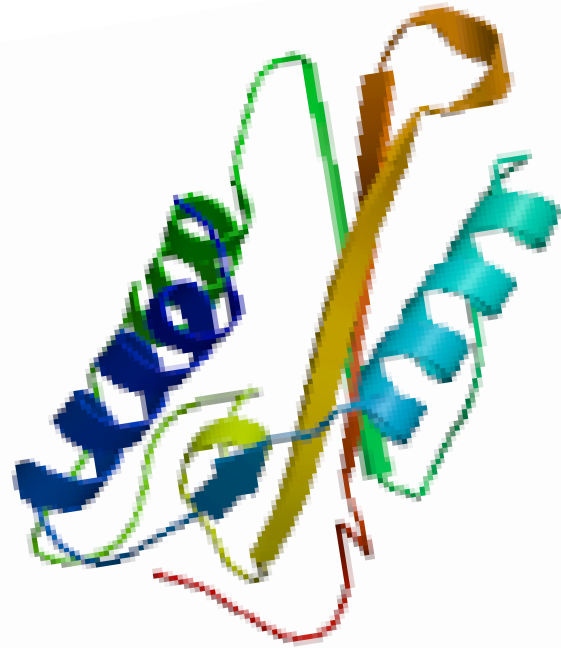
汇报提纲

- 1 研究背景
- 2 核酸序列分析
- 3 蛋白序列及结构分析
- 4 构建进化树
- 5 研究展望

1 研究背景

- PR蛋白：

病程相关蛋白PR是植物受病原物侵染或非生物因子刺激后产生的一类水溶性蛋白，参与了植物的诱导抗病性，主要功能是攻击病原物，降解细胞壁大分子，降解病原物毒素，抑制病毒外壳蛋白与植物受体分子结合等。



病程相关蛋白

根据氨基酸序列的相似程度可分为

Table 9.1 The families of pathogenesis-related proteins

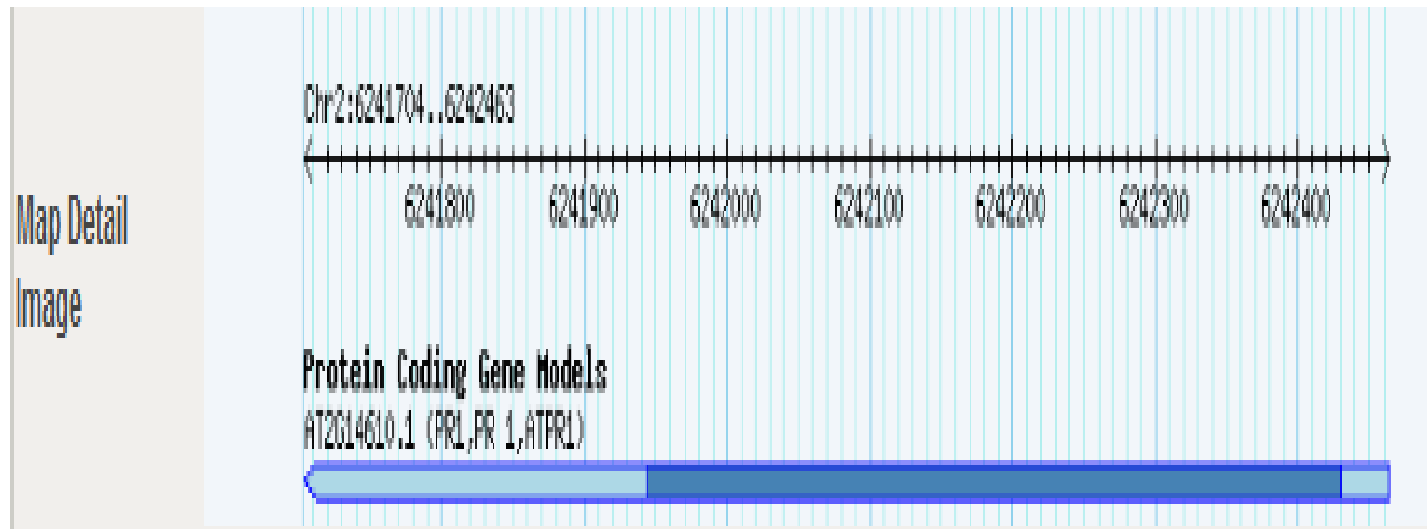
Family	Type member	Properties
PR-1	Tobacco PR-1a	Unknown
PR-2	Tobacco PR-2	Endo β -1,3-glucanase
PR-3	Tobacco P,Q	Type I, II, IV, V, VI, VII chitinases
PR-4	Tobacco 'R'	Type I, II chitinases
PR-5	Tobacco S	Thaumatin-like
PR-6	Tomato Inhibitor I	Protease inhibitor
PR-7	Tomato P ₆₉	Endoprotease
PR-8	Cucumber chitinase	Type III chitinase
PR-9	Tobacco 'lignin-forming peroxidase'	Peroxidase
PR-10	Parsley 'PR1'	Ribonuclease
PR-11	Tobacco class V chitinase	Type I chitinase
PR-12	Radish Rs-AFP3	Defensin
PR-13	<i>Arabidopsis</i> THI2.1	Thionin
PR-14	Barley LTP4	Lipid-transfer protein

不同病程相关蛋白的功能特性

家族	代表成员	特性
PR-1	Tobacco PR-1a	未知
PR-2	Tobacco PR-2	β -1, 3- 葡聚糖酶
PR-3	Tobacco P. Q	I, II, IV, V, VI, VII 型几丁质酶
PR-4	Tobacco "R"	I, II型几丁质酶
PR-5	Tobacco S	类甜蛋白 (TIP)
PR-6	Tomato Inhibitor I	蛋白酶抑制因子
PR-7	Tomato P	蛋白内切酶
PR-8	Cucumber chitinase	III型几丁质酶
PR-9	Tobacco "lignin-forming peroxidase"	过氧化物酶
PR-10	Parsley "PR1"	类核糖核酸酶
PR-11	Tobacco class V chitinase	I型几丁质酶
PR-12	Radish Rs-1/FP3	防卫素
PR-13	Arabidopsis TH I2.1	硫素
PR-14	Barley LIP4	脂质转移蛋白

2 核酸序列分析

编码PR-1基因有760个核苷酸，第35-520位核苷酸是CDS区



From TAIR

通过softBerry预测基因结构，有两个基因，有九个外显子

```
FGENESH 2.6 Prediction of potential genes in Phytophthora genomic DNA
Time      :   Thu May 30 11:11:23 2013
Seq name: gi|1228948|emb|X96600.1| A.thaliana PR1 genes in tandem repeat
Length of sequence: 2616
Number of predicted genes 2:  in +chain 1,  in -chain 1.
Number of predicted exons 9:  in +chain 5,  in -chain 4.
Positions of predicted genes and exons: Variant 1 from 1, Score:31.050195
```

G Str	Feature	Start	End	Score	ORF	Len
1 +	TSS	114		2.07		
1 +	1 CDSf	320 -	333	1.83	320 -	331 12
1 +	2 CDSi	719 -	744	0.75	720 -	743 24
1 +	3 CDSi	790 -	806	4.35	792 -	806 15
1 +	4 CDSi	1003 -	1014	4.65	1003 -	1014 12
1 +	5 CDSl	1069 -	1086	6.14	1069 -	1086 18
1 +	PolA	1151		1.88		
2 -	PolA	1234		1.88		
2 -	1 CDSl	1419 -	1426	6.74	1419 -	1424 6
2 -	2 CDSi	1561 -	1581	13.85	1562 -	1579 18
2 -	3 CDSi	1632 -	1637	7.38	1633 -	1635 3
2 -	4 CDSi	1810 -	1818	1.11	1811 -	1816 6

Predicted protein(s):

```
>FGENESH: [mRNA] 1 5 exon (s) 320 - 1086 87 bp, chain +
ATGCTCGAGACGACACGAAAGAAACTTGTCCATAGAGGACGATGTGGTGGTTCCTTTCTA
ACTATTTATTATCGGAATCTCTTTTAA
>FGENESH: 1 5 exon (s) 320 - 1086 28 aa, chain +
MLETTRKKLVHRGRCGGSFLTIYYRNLF
>FGENESH: [mRNA] 2 4 exon (s) 1419 - 1818 42 bp, chain -
AGAAAACATTACATTTTTGGAGATCCTGACATATCGATATGA
>FGENESH: 2 4 exon (s) 1419 - 1818 13 aa, chain -
RKHYIFGDPDISI
```

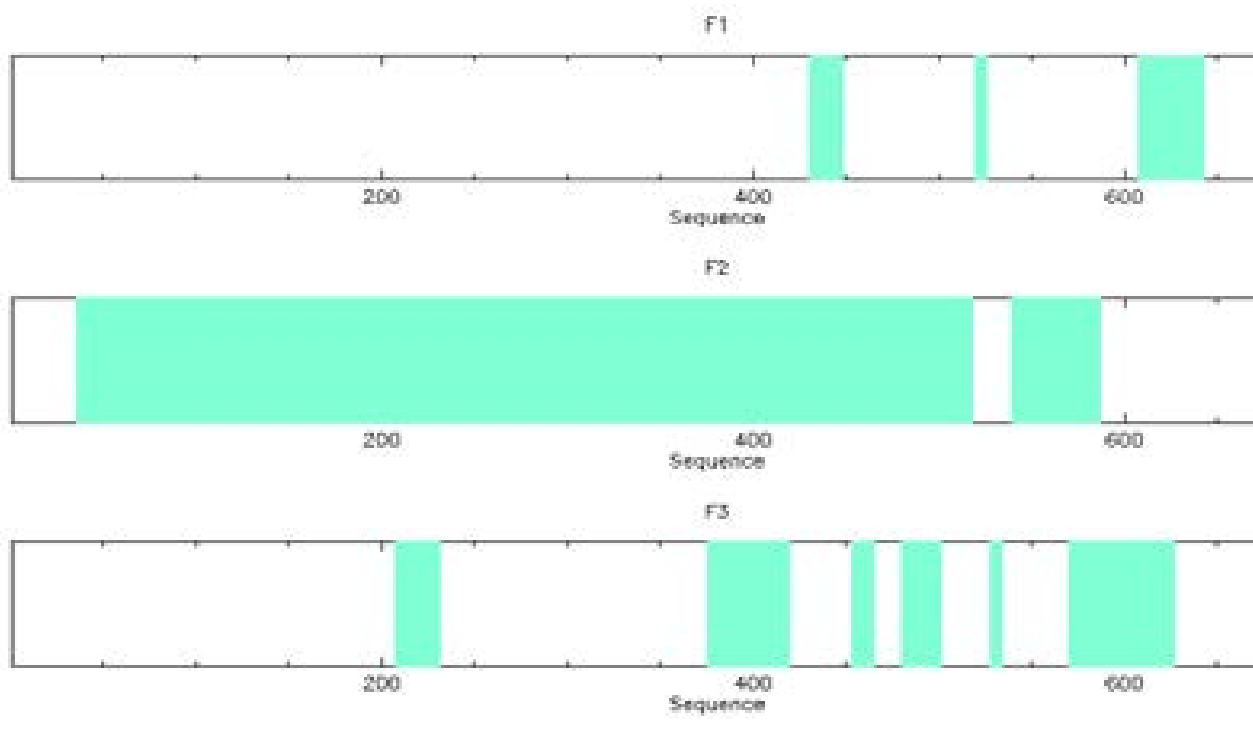
编码区序列

- >gi|30679169:35-520 Arabidopsis thaliana, complete cds

ATGAATTTTACTGGCTATTCTCGATTTTTTAATCGTCTTTGTA
GCTCTTGTAGGTGCTCTTGTTCTTCCCTCGAAAGCTCAAGATAGC
CCACAAGATTATCTAAGGGTTCACAACCAGGCACGAGGAGCGGT
AGGCGTAGGTCCCATGCAGTGGGACGAGAGGGTTGCAGCCTAT
GCTCGGAGCTACGCAGAACAACTAAGAGGGCAACTGCAGACTCAT
ACACTCTGGTGGGCCTTACGGGGAAAACCTTAGCCTGGGGTAGCG
GTGACTTGTCTGGCGTCTCCGCCGTGAACATGTGGGTTAGCGAG
AAGGCTAACTACAACACTACGCTGCGAACACGTGCAATGGAGTTTG
TGGTCACTAFACTCAAGTTGTTTGGAGAAAGTCAGTGAGACTCG
GATGTGCCAAAGTGAGGTGTAACAATGGTGGGAACCATAATCAGT
TGCAACTATGATCCTCGTGGGAATTATGTGAACGAGAAGCCATA
CTAA

PlotORF程序分析

用WebLab中PlotORF程序分析其可能的读码框。由图可知，所得结果中只有F2具有连续片段，符合coding sequence的序列特征，应是正确的阅读框。



ShowORF程序分析

从图中可看出，在`1`行编码M的密码子为起始密码子，在后边有两个连续的终止密码子UAA、UGA。

```
-----|-----|-----|-----|-----|
1  ATAACACAACAATAAACCATTATCAACTTAGAAAAATGAATTTTACTGGCT 50
F2  1  * H N N N H Y Q L R K N N P T G Y 16
-----|-----|-----|-----|-----|
51  ATTCTCGAATTTTAAATGTCITTTGTAGCTCTTGTAGGTGCTCTTGTCTT 100
F2  17  S R F L I V F V A L V G A L V L 32
-----|-----|-----|-----|-----|
101 CCTCGAAAGCTCAAGATAGCCACAAGATTATCTAAGGGTTCACAACCA 150
F2  33  P S K A Q D S P Q D Y L R V H N Q 49

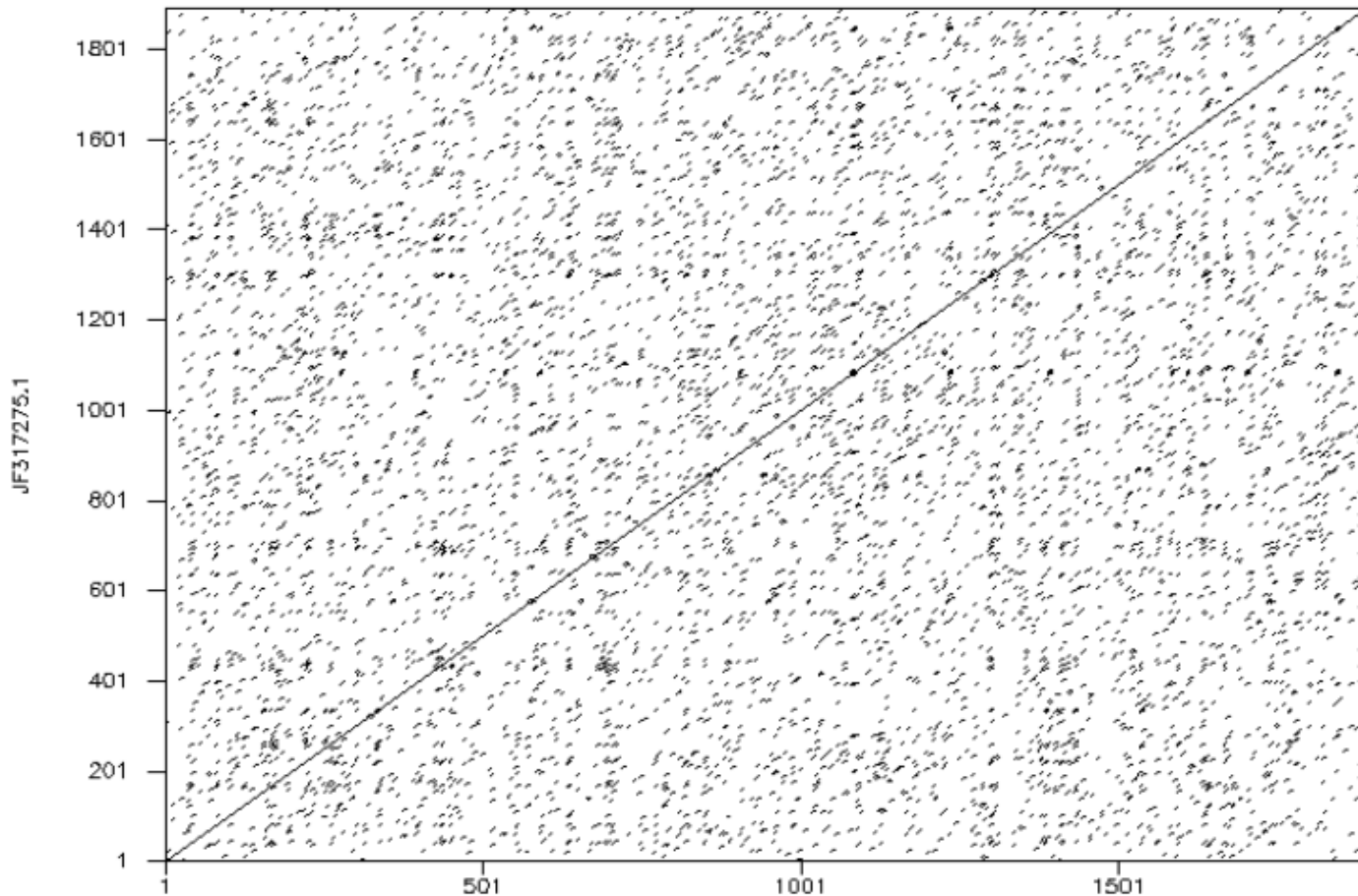
501 TGAACGAGAAGCCATAATAATGAAGTAATGATGTGATCATGCATACACAC 550
F2  167  N E K P Y * * S N D V I M H T H 9
-----|-----|-----|-----|-----|
551 GTACATAAAGGAGGTGTATAATGATCAGTATTTCAATAAGGAGCATCATA 600
F2  10  V H K G R V Y V S V F Q * G A S Y 4
-----|-----|-----|-----|-----|
601 TGCAGGATGTATCAATATTTATCAATAAATAACAAATAAGAGCTGAGATTA 650
F2  5  A G C I N I Y Q I I Q I R A E I T 21
```

利用WebLab中密码子统计CUSP程序分析如下：

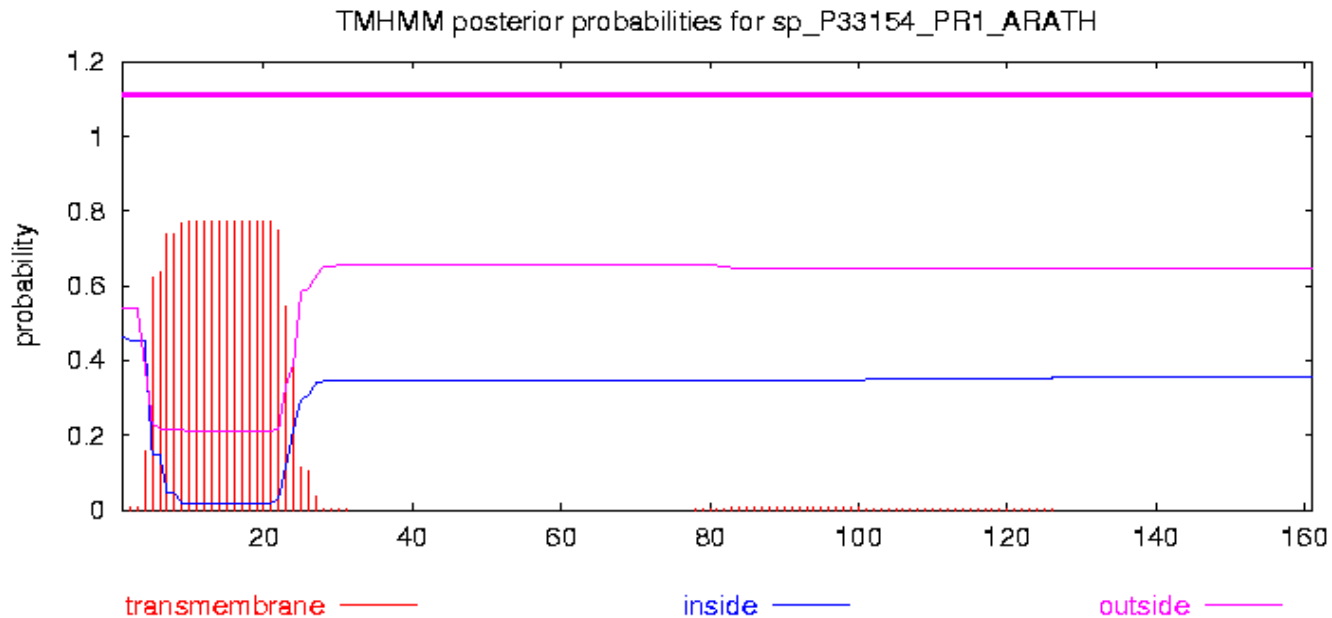
```
#CdsCount: 14
#Coding GC 40.18%
#1st letter GC 37.94%
#2nd letter GC 44.66%
#3rd letter GC 37.94%
#Codon AA Fraction Frequency Number
GCA A 0.400 7.905 2
GCC A 0.600 11.858 3
GCG A 0.000 0.000 0
GCT A 0.000 0.000 0
TGC C 0.429 23.715 6
TGT C 0.571 31.621 8
GAC D 0.000 0.000 0
GAT D 1.000 7.905 2
GAA E 0.571 31.621 8
GAG E 0.429 23.715 6
TTC F 0.167 3.953 1
TTT F 0.833 19.763 5
GGA G 0.222 7.905 2
GGC G 0.222 7.905 2
GGG G 0.111 3.953 1
GGT G 0.444 15.810 4
CAC H 0.286 7.905 2
CAT H 0.714 19.763 5
ATA I 0.500 31.621 8
ATC I 0.125 7.905 2
ATT I 0.375 23.715 6
AAA K 0.636 27.668 7
AAG K 0.364 15.810 4
CTA L 0.310 35.573 9
CTC L 0.103 11.858 3
CTG L 0.069 7.905 2
CTT L 0.103 11.858 3
TTA L 0.310 35.573 9
TTG L 0.103 11.858 3
ATG M 1.000 11.858 3
AAC N 0.111 3.953 1
AAT N 0.889 31.621 8
CCA P 0.667 7.905 2
CCC P 0.333 3.953 1
CCG P 0.000 0.000 0
```

由结果可知密码子TGT、GAA、ATA、CTA、TTA、AAT等的使用频率是较高的。

提取PR1蛋白基因，通过点阵图分析，由分析结果看，无重复序列。



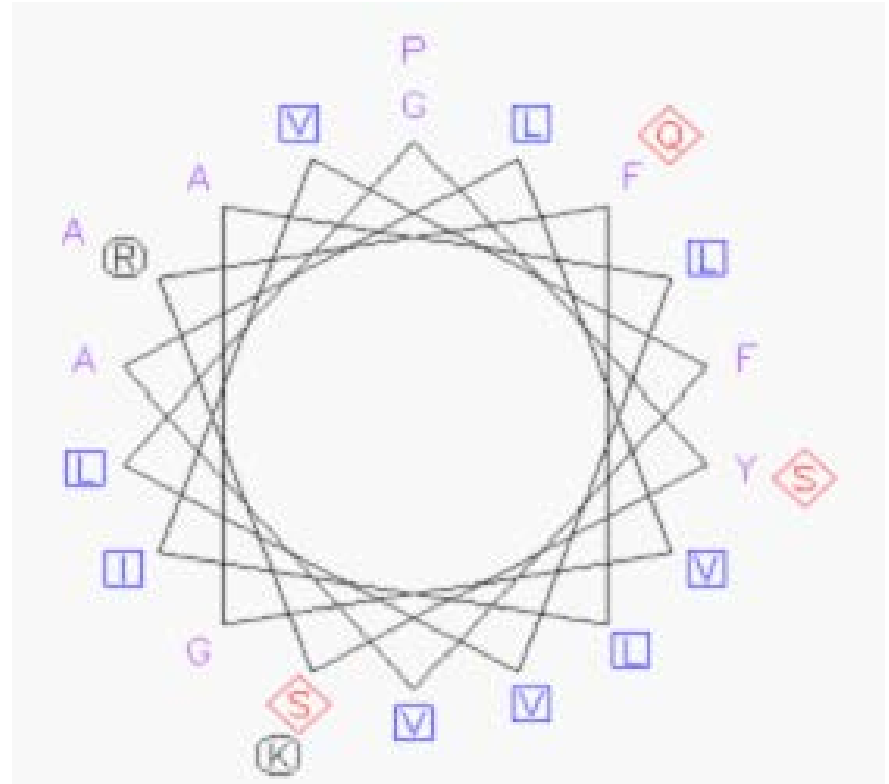
3 蛋白序列分析



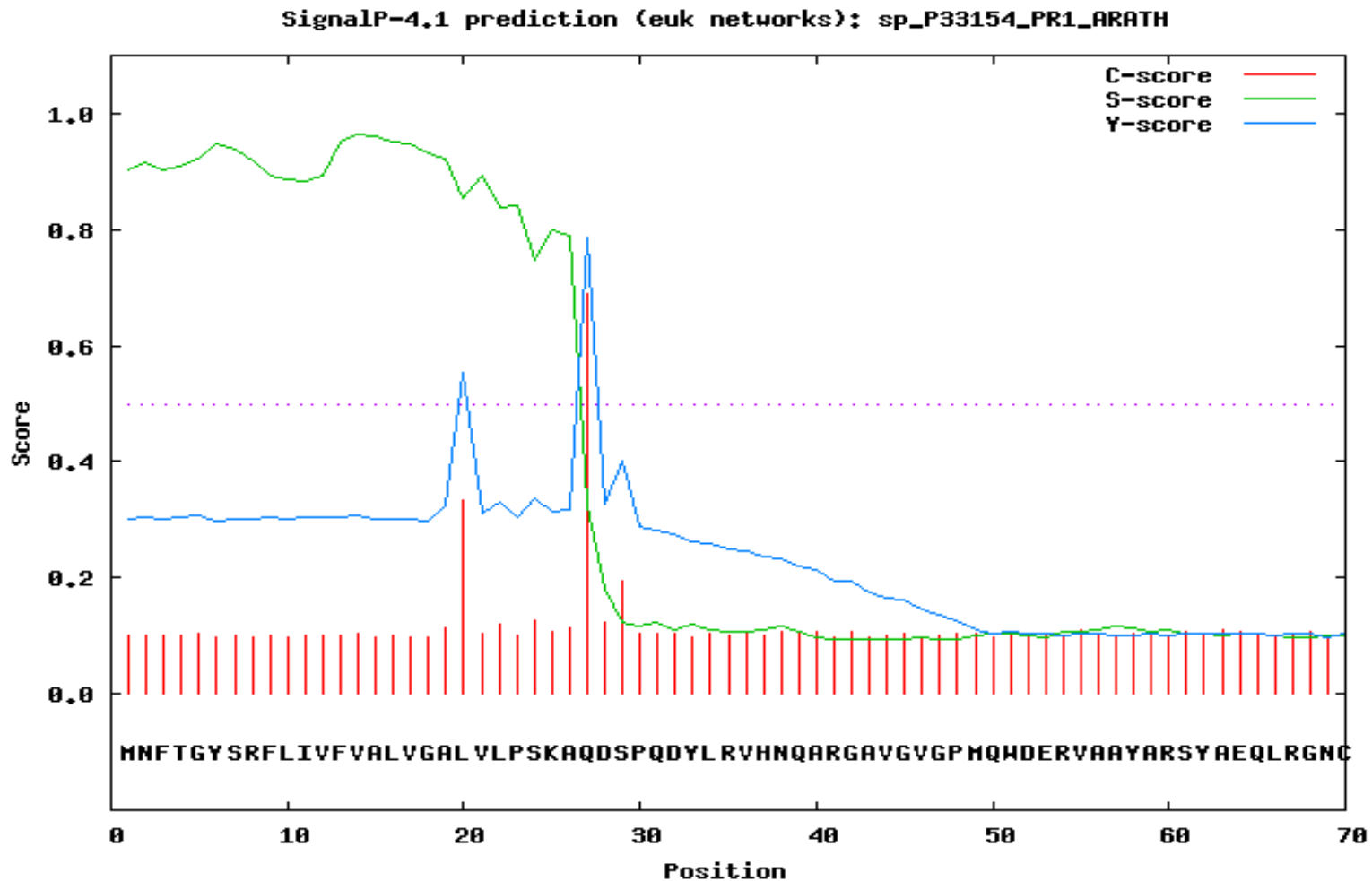
根据跨膜螺旋预测结果可知，PR1蛋白质是跨膜蛋白，有一个跨膜螺旋，长度为23个氨基酸。

利用WebLab中alpha-螺旋论显示程序，绘制PR1跨膜螺旋的螺旋论

其中黑色带圆圈的是极性带电荷的氨基酸，亲水性比较强；红色带方框的是极性不带电荷的氨基酸，也是亲水性的；紫色字母不带框的是非极性氨基酸，具有疏水性；蓝色带方框的是非极性氨基酸，疏水性比紫色的强。



利用 CBS网站中的SignalP程序和ExPASy中的TargetP，分别预测PR1蛋白质的信号肽、亚细胞定位。



# Measure	Position	Value	Cutoff	signal peptide?
max. C	27	0.689		
max. Y	27	0.785		
max. S	14	0.964		
mean S	1-26	0.896		
D	1-26	0.845	0.450	YES

信号肽预测：1—26位氨基酸为信号肽

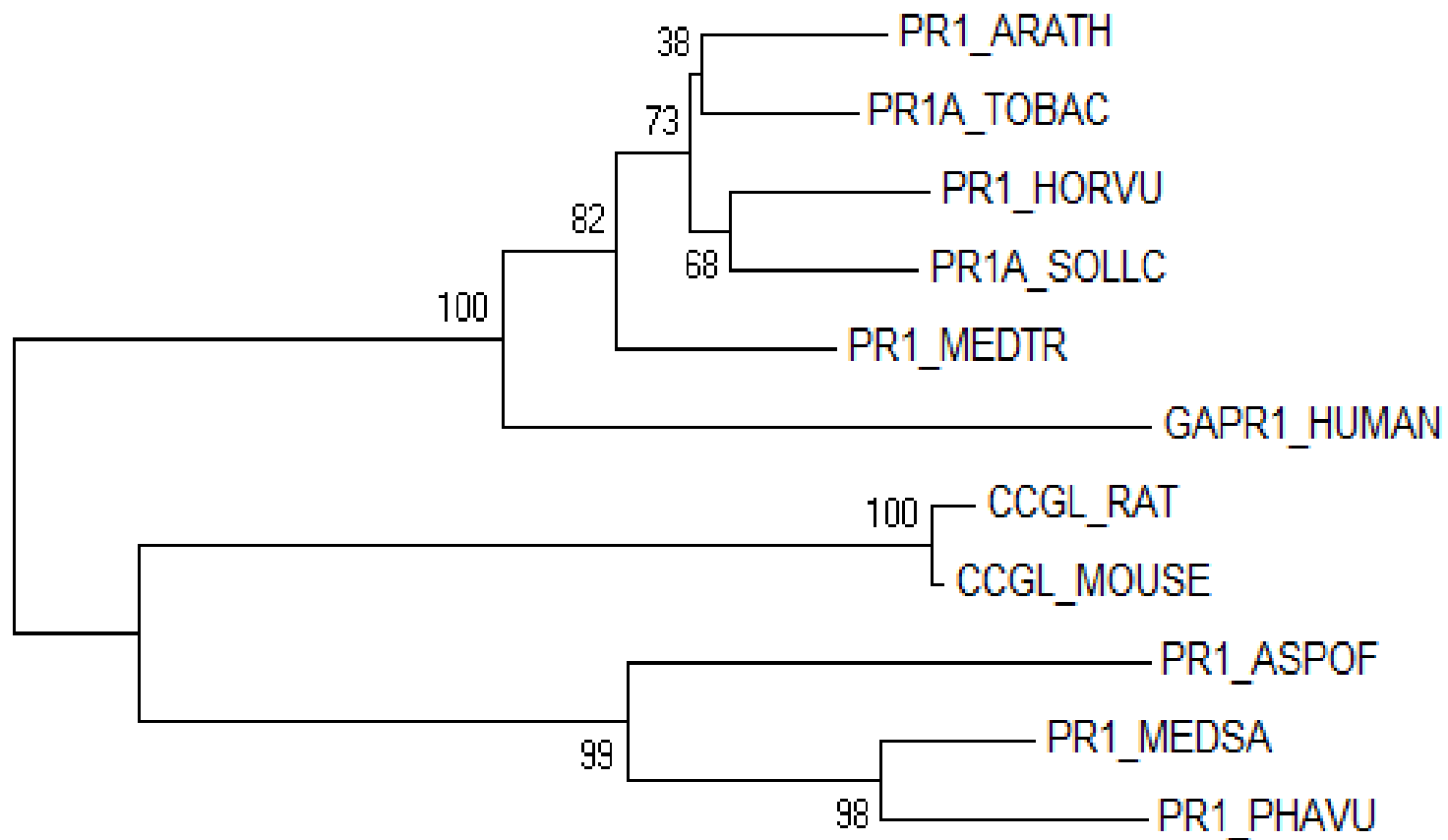
Name	Len	mTP	SP	other	Loc	RC
sp_P33154_PR1_ARATH	161	0.082	0.927	0.018	S	1
cutoff		0.000	0.000	0.000		

TargetP亚细胞定位预测：SP值最高，定位于分泌系统，PR1蛋白是分泌蛋白。

4不同物种PR1蛋白序列比对及构建进化树

1. PR1 MEDTR	M S F R C F S F A L F L L L L F I E V S A S Y I F K K S F S R S F K F L I S C I A A A A V L R L L V D D L L H Y A A Y Y A C R A
2. PR1 ARATH	M F F G V S S F L I V F V A L V G A L V L S - Y A D S P F L V L R V H S A R G A V V G P M S D E R V A A Y A S Y A S L A
3. PR1 ASPOF	M S S G S S H E V A V V A A G M F A A M L D F L G E I V D F I A G S V V S G D C N V G I I S I V I V P A I F
4. CCGL RAT	M A I C A A Y L L G S Y S H S F E S F I L I I V C A L A V V L S V I C D G H L L V E D L F G L Y F C I S V H
5. GAPR1 HUMAN	M G A A Y F H E V L A A F E Y P A H G V P L L L E V L L S A A G Y S A L A S V I
6. CCGL MOUSE	M A I G A A H K L L G L R S H R F F S F I L I I V C A L A V V L S V I C D G H L L V E D L F G L Y F C I G H A
7. PR1A TOBAC	M G F V L F S L E F L L V S L L L F L V I H S C A A S G D Y L D A H T A A D V E V E L I D D V A A Y A S Y A S L A
8. PR1 HORVU	M T E L A I L L A L A M A A A M V L S A A S P F L V V S H A A S A V V G A V S S K L A F A S Y A S I
9. PR1A SOLLC	M S S S I F V A S F I F I I F S S A P S R F L V A H A A R V C V G P M I D G L A A Y A S Y A C A
10. PR1 MEDSA	M G V F E F E D T T I V A A A L V A L V D A S L I S V I D A I S I F I V E S G G G A G I V L I F V E G G E
11. PR1 PHAVU	M G V F E F E D T T I V A A A L V A V A A A I F S A L S A F S V E I V E S G G G A G I V L I F V E G G E

M	D	C	A	L	S	H	S	G	S	V	G	E	N	I	F	G	S	G	V	G	-	-	S	N	P	A	A	V	S	A	V	V	D	E	K	D	E	F	N	Y	H	S	G	V	D	G	E	-	-	M	G	G	
G	N	C	R	L	I	H	S	G	S	V	G	E	N	L	A	G	S	G	D	-	-	L	S	G	V	S	A	V	M	V	S	E	K	A	N	Y	A	A	N	I	C	N	G	-	-	-	-	V	C	G			
F	S	V	K	E	R	L	D	F	V	D	R	K	F	E	V	K	Q	L	V	E	G	G	L	G	M	F	E	C	A	I	-	-	-	-	I	H	F	K	F	E	S	S	M	G	-	-	-	-	G				
E	H	C	L	D	L	S	A	A	V	V	G	V	A	V	G	M	G	L	A	S	V	A	A	M	A	V	V	A	I	F	G	L	E	L	L	I	V	S	V	C	H	D	V	S	R	K	A	I	G				
L	K	H	S	E	S	S	G	S	G	E	N	L	A	A	S	Y	D	-	-	I	G	S	V	A	D	R	N	Y	S	E	I	K	N	Y	N	F	D	D	G	F	L	S	G	I	-	-	-	-	G				
E	H	C	L	D	L	S	A	A	V	V	G	V	A	V	G	M	G	L	A	S	V	A	A	M	A	V	V	A	I	F	G	L	E	L	L	I	V	S	V	C	H	D	V	S	R	K	A	I	G				
A	D	C	N	L	V	H	S	H	G	S	V	G	E	N	L	A	S	G	S	G	D	E	-	M	I	A	A	K	A	V	E	M	V	D	E	R	K	Y	Y	D	H	S	N	I	C	A	G	G	-	-	V	C	G
N	D	C	K	L	H	S	G	S	V	G	E	N	I	F	G	S	A	G	A	D	-	K	A	S	D	A	V	S	V	S	E	K	K	D	Y	D	Y	G	S	N	I	C	A	A	G	K	-	-	V	C	G		
D	D	C	G	M	I	H	S	G	S	V	G	E	N	L	-	A	A	A	F	D	-	L	N	A	A	G	A	V	K	M	D	D	E	R	K	Y	Y	D	Y	S	N	I	C	A	S	G	K	-	-	V	C	G	
I	K	Y	D	L	H	S	V	D	L	V	D	V	N	F	A	Y	N	Y	S	I	V	G	G	G	L	S	D	I	V	E	K	I	S	-	-	-	-	F	E	S	K	L	S	A	G	S	G	-	-	-	-	G	
I	K	F	V	L	H	S	I	S	I	S	E	A	N	L	G	S	Y	S	I	V	G	G	V	A	L	S	E	I	A	S	I	F	-	-	-	-	F	E	S	K	L	S	A	G	S	G	-	-	-	-	G		



0.2

5 研究展望

1. PR参与了植物的诱导抗病性,人们可通过构建PR蛋白基因的方法获得抗菌谱广、作用持久的抗病新品种。
2. 病程相关蛋白与动物体内免疫蛋白的作用方式是否类似,还有哪些重要因子参与; 基因表达涉及到哪些具体的信号传导途径.....还有待进一步研究。

小组成员



谢谢!

