松树AP2家族两个基因的生物信息学分析

汇报人：李魁鹏
组长：李荣琴
组员：解婷婷、雷俊

一. 基因的信息

二. 直系同源基因的寻找方法

三. 基因的生物信息学分析

一 基因的基本信息

根据AP2/EREBP 基因中所含AP2功能域数量的不同, 将其分为3类:

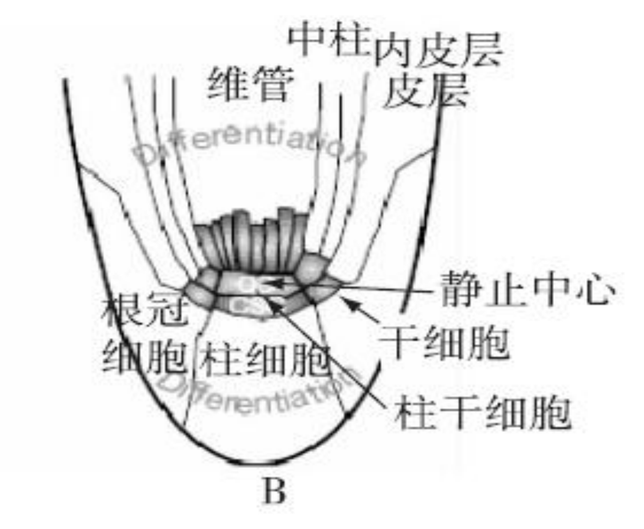
第一类是AP2亚家族, 在其蛋白结构中含有两个AP2结构域。

第二类是EREBP亚家族, 含有一个AP2/EREBP 结构域。

第三类是RAV1和RAV2, 含有两个不同的结构域，AP2/EREBP 结合域和类似B3的结构域。

静止中心位于根尖分生组织中心,干细胞围绕在静止中心细胞周围。静止中心形成和干细胞区特化是植物根尖分生组织确立的标志。

依赖于生长素调控的*PLT*基因维持静止中心细胞稳定的作用。



生长素的梯度分布引起PLT表达差异:

高PLT活性可维持干细胞特性。

低活性诱导干细胞子细胞进行有丝分裂。

最低活性可以引起细胞分化。

The PLETHORA genes mediate patterning of the Arabidopsis root stem cell niche，Cell, 2004

Arabidopsis tyrosylprotein sulfotransferase acts in the auxin /PLETHORA pathway in regulating postembryonic maintenance of the root stem cell niche，Plant Cell，2010

本课题首先关注的是：

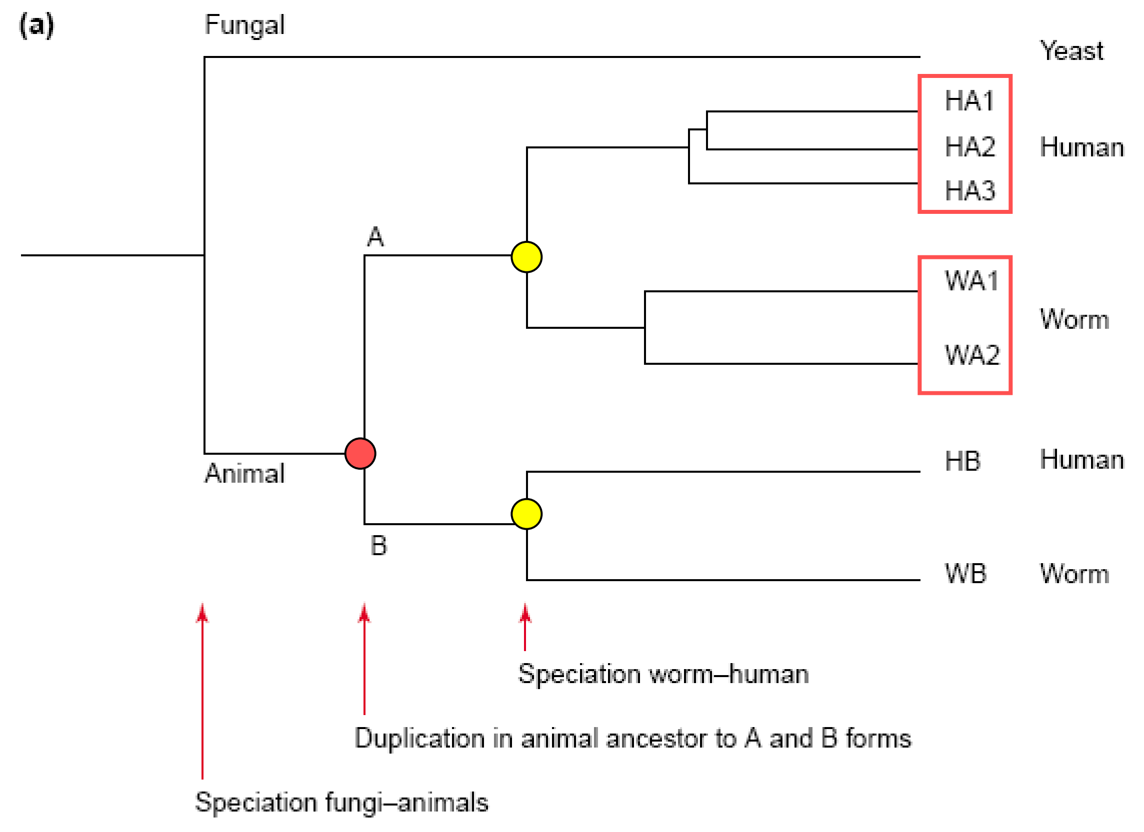
如何找到拟南芥*PLT1*在松树中的直系同源基因？

二 直系同源基因的寻找方法

直系同源基因（Orthology）最初的定义是由最近祖先物种的一个基因在两个不同物种中分别演化而来的基因。

现在很多研究人员认为直系同源基因是在不同物种中行使相同功能的基因。

双向 blast（ bidirectional best hit method in BLAST ）：用物种A的基因组的序列搜索(BLAST)物种B的基因组的序列，搜到的最高相似的序列反过来再BLAST物种A的基因组(阈值E小于$1e^{-10}$)。如果两条序列在这两个BLAST中均是最相似的称为 “双向BLAST最高相似蛋白”），则认为是直系同源关系。

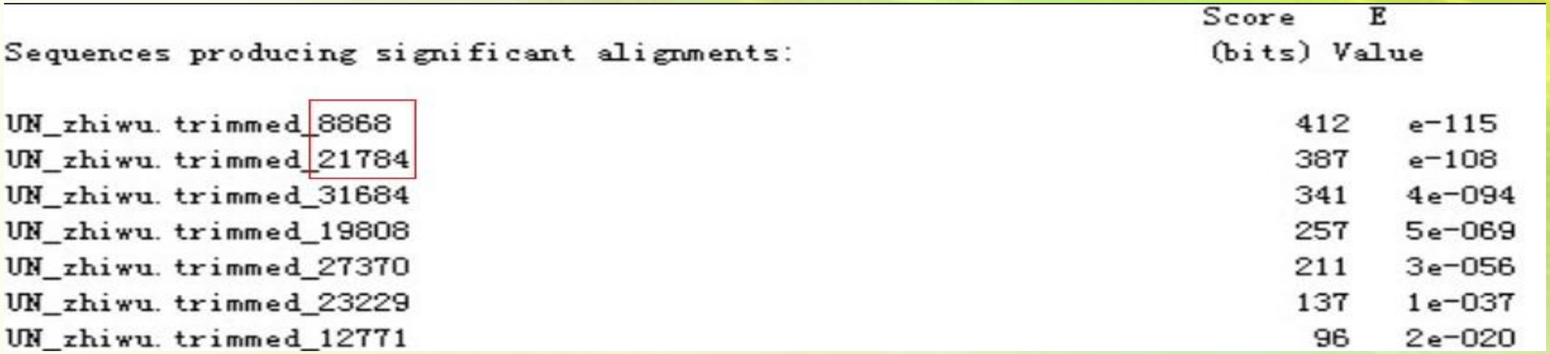


Orthology,paralogy and proposed classification for paralog subtypes

TRENDS in Genetics Vol.18 No.12 December 2002

存在一(多) 对多的直系同源关系,若在物种形成后所产生的旁系同源基因还相当保守或者还没有明显异化, 此时可看成各旁系同源体簇之间是直系同源关系。

用拟南芥***PLT1***基因搜索转录组数据库：



```
                                                                 Score    E
Sequences producing significant alignments:                      (bits) Value

UN_zhiwu.trimmed_8868                                              412   e-115
UN_zhiwu.trimmed_21784                                             387   e-108
UN_zhiwu.trimmed_31684                                             341   4e-094
UN_zhiwu.trimmed_19808                                             257   5e-069
UN_zhiwu.trimmed_27370                                             211   3e-056
UN_zhiwu.trimmed_23229                                             137   1e-037
UN_zhiwu.trimmed_12771                                              96   2e-020
```

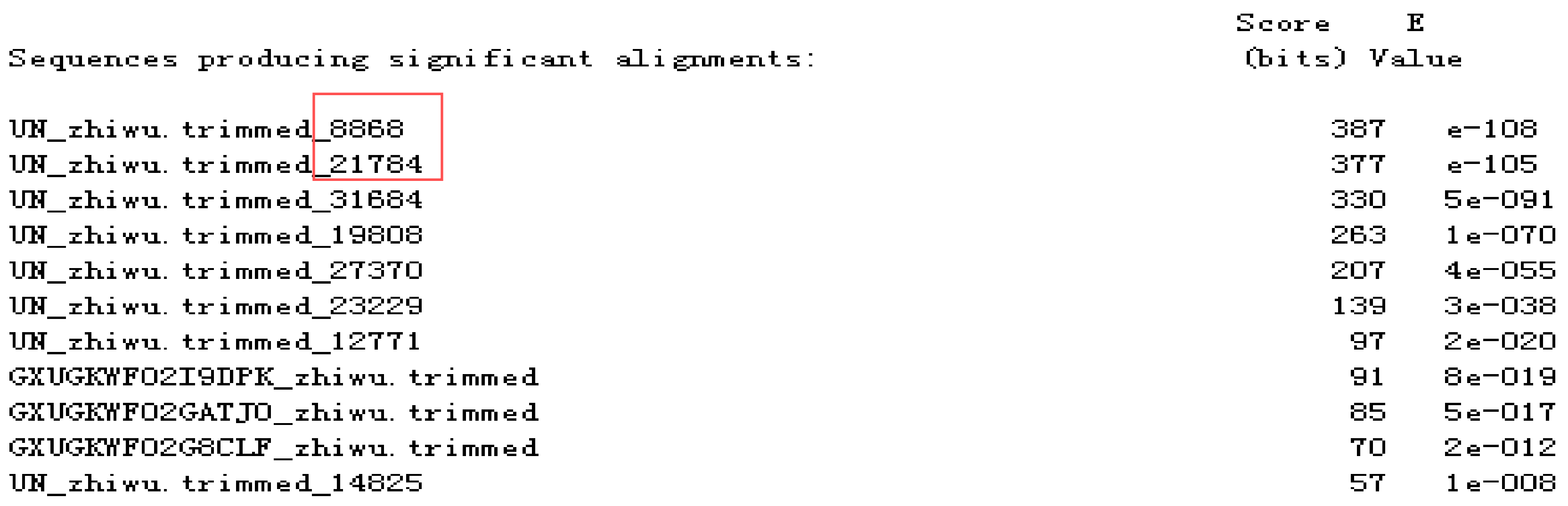
gene 1 (8868)与拟南芥基因进行序列比对结果：

ecName: Full=AP2-like ethylene-responsive transcription facto PLT2; ltName: Full=Protein AINTEGUMENTA-LIKE 4; AltName: Full=Prc	413	413	37%	3e-135
ecName: Full=AP2-like ethylene-responsive transcription facto PLT1; ltName: Full=Protein AINTEGUMENTA-LIKE 3; AltName: Full=Prc	410	410	41%	5e-134
ecName: Full=AP2-like ethylene-responsive transcription factor BBM; Short=AtBBM; AltName: Full=Protein AINTEGUMENTA-LIKE 2; AltN	400	431	66%	4e-130
ecName: Full=AP2-like ethylene-responsive transcription factor AIL5; AltName: Full=Protein AINTEGUMENTA-LIKE 5	393	393	41%	2e-127
ecName: Full=AP2-like ethylene-responsive transcription factor AIL1; AltName: Full=Protein AINTEGUMENTA-LIKE 1	363	363	41%	4e-118

gene 2 (21784)与拟南芥基因进行序列比对结果：

Full=AP2-like ethylene-responsive transcription facto AIL5; ltName: Full=Protein AINTEGUMENTA-LIKE 5	392	392	47%	2e-130
Full=AP2-like ethylene-responsive transcription facto PLT1 AltName: Full=Protein AINTEGUMENTA-LIKE 3; AltName: Full=Prc	391	522	42%	8e-130
Full=AP2-like ethylene-responsive transcription facto PLT2 AltName: Full=Protein AINTEGUMENTA-LIKE 4; AltName: Full=Prc	390	390	43%	1e-129
Full=AP2-like ethylene-responsive transcription factor BBM; Short=AtBBM; AltName: Full=Protein AINTEGUMENTA-LIKE 2; AltN	385	385	44%	1e-127

用拟南芥***AIL5***基因搜索转录组数据库：



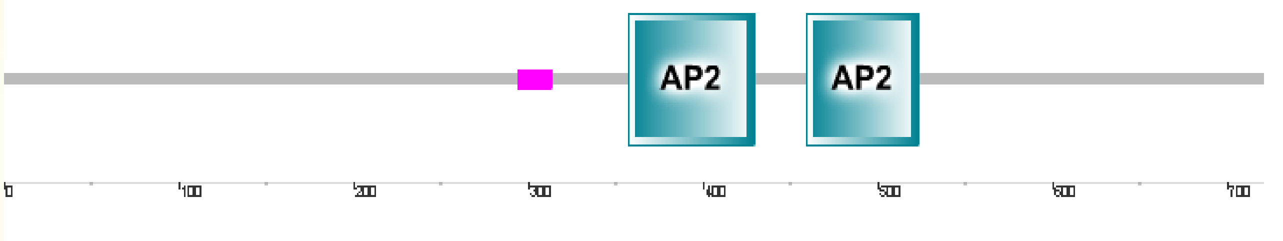
```
                                                                Score     E
Sequences producing significant alignments:                    (bits)  Value

UN_zhiwu.trimmed_8868                                             387   e-108
UN_zhiwu.trimmed_21784                                            377   e-105
UN_zhiwu.trimmed_31684                                            330   5e-091
UN_zhiwu.trimmed_19808                                            263   1e-070
UN_zhiwu.trimmed_27370                                            207   4e-055
UN_zhiwu.trimmed_23229                                            139   3e-038
UN_zhiwu.trimmed_12771                                             97   2e-020
GXUGKWF02I9DPK_zhiwu.trimmed                                       91   8e-019
GXUGKWF02GATJO_zhiwu.trimmed                                       85   5e-017
GXUGKWF02G8CLF_zhiwu.trimmed                                       70   2e-012
UN_zhiwu.trimmed_14825                                             57   1e-008
```

我们推测基因1和2与***PLT1*** 和 ***AIL5***都是同源直系的关系。

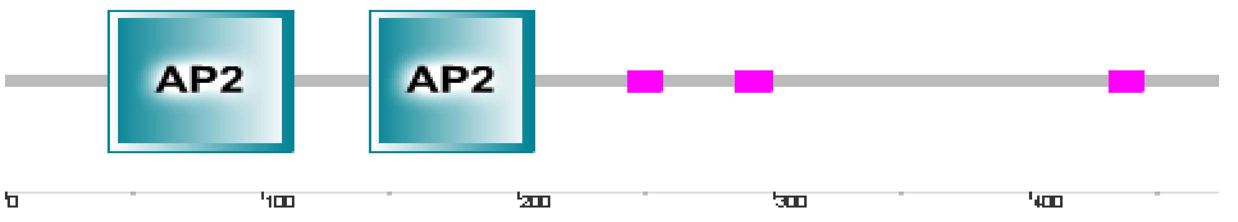
三 基因的生物信息学分析

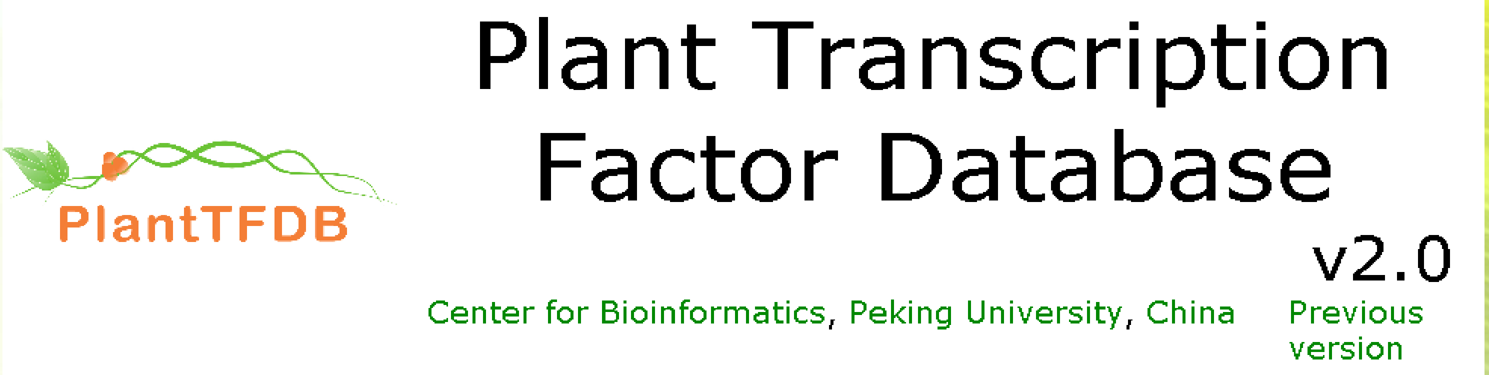
SMART程序分析：

gene 1：



gene 2:





PlantTFDB

Plant Transcription Factor Database

v2.0

Center for Bioinformatics, Peking University, China

Previous version

被子植物门

双子叶植物：拟南芥25个AP2基因(Ath)
单子叶植物：水稻的27个AP2基因(Osi)

裸子植物门

白云衫的7个AP2基因(Pg)
火炬松的4个AP2基因(Pta)
本课题研究的Gene A、B

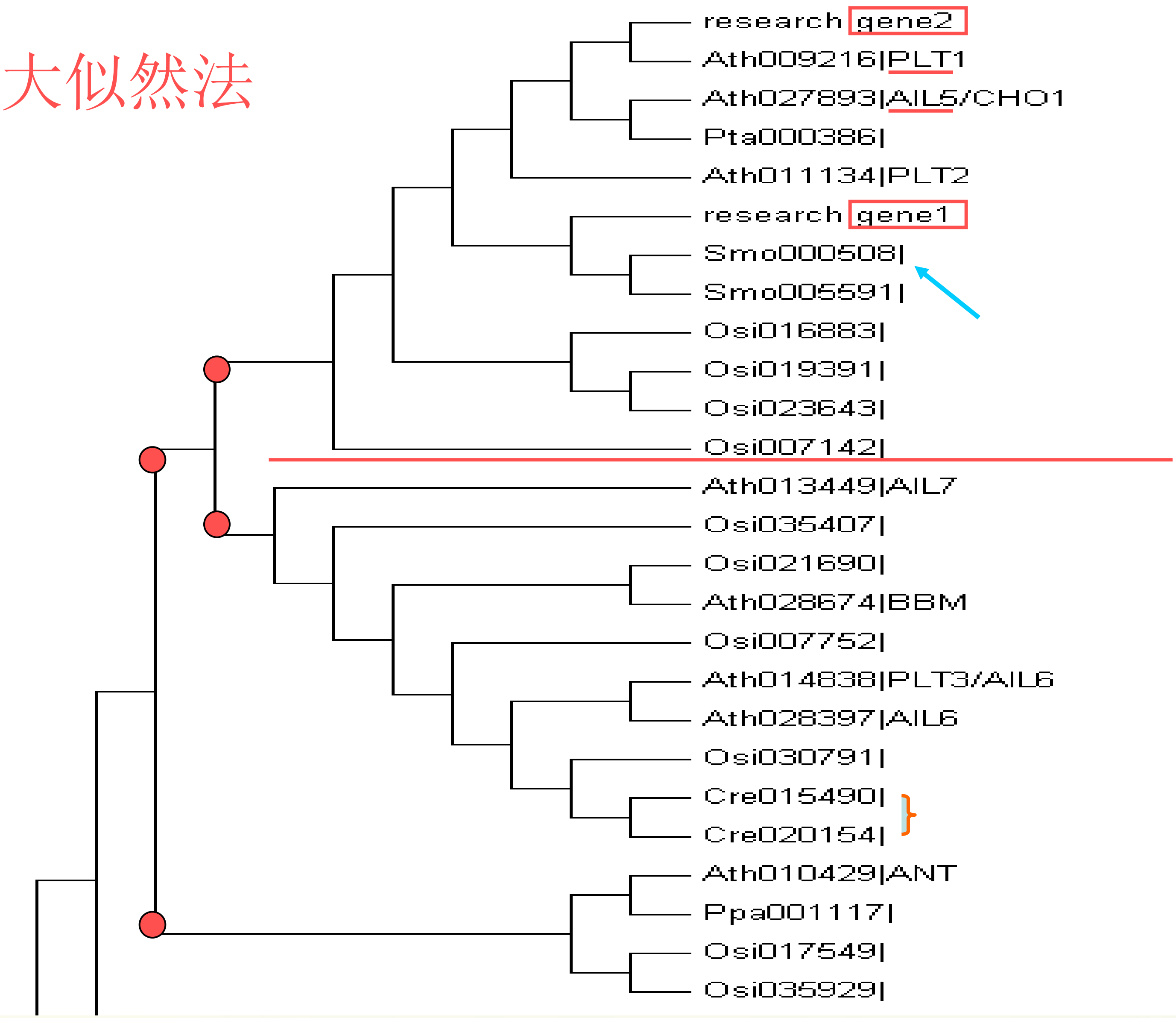
蕨类植物门

江南卷柏的16个AP2基因(Smo)
苔藓门植物：小立碗藓的14个基因(Ppa)

藻类

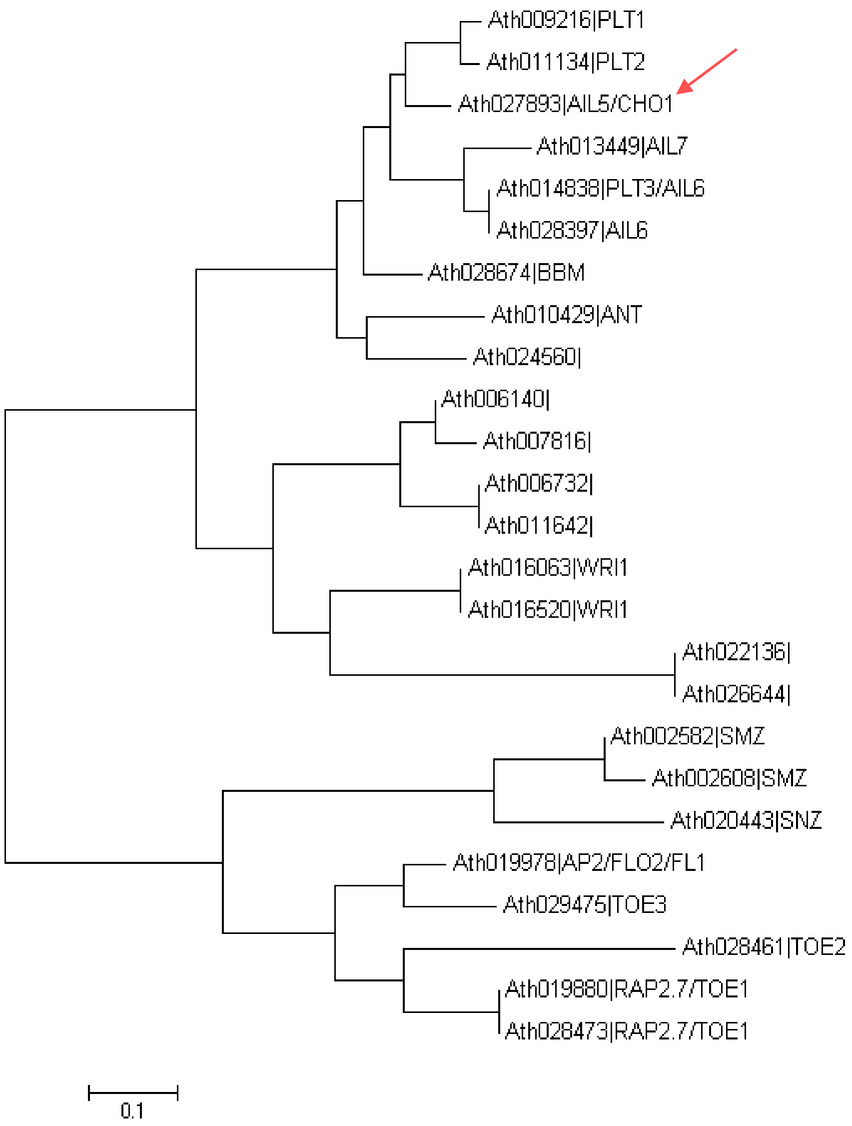
莱茵衣藻的9个AP2基因(Cre)

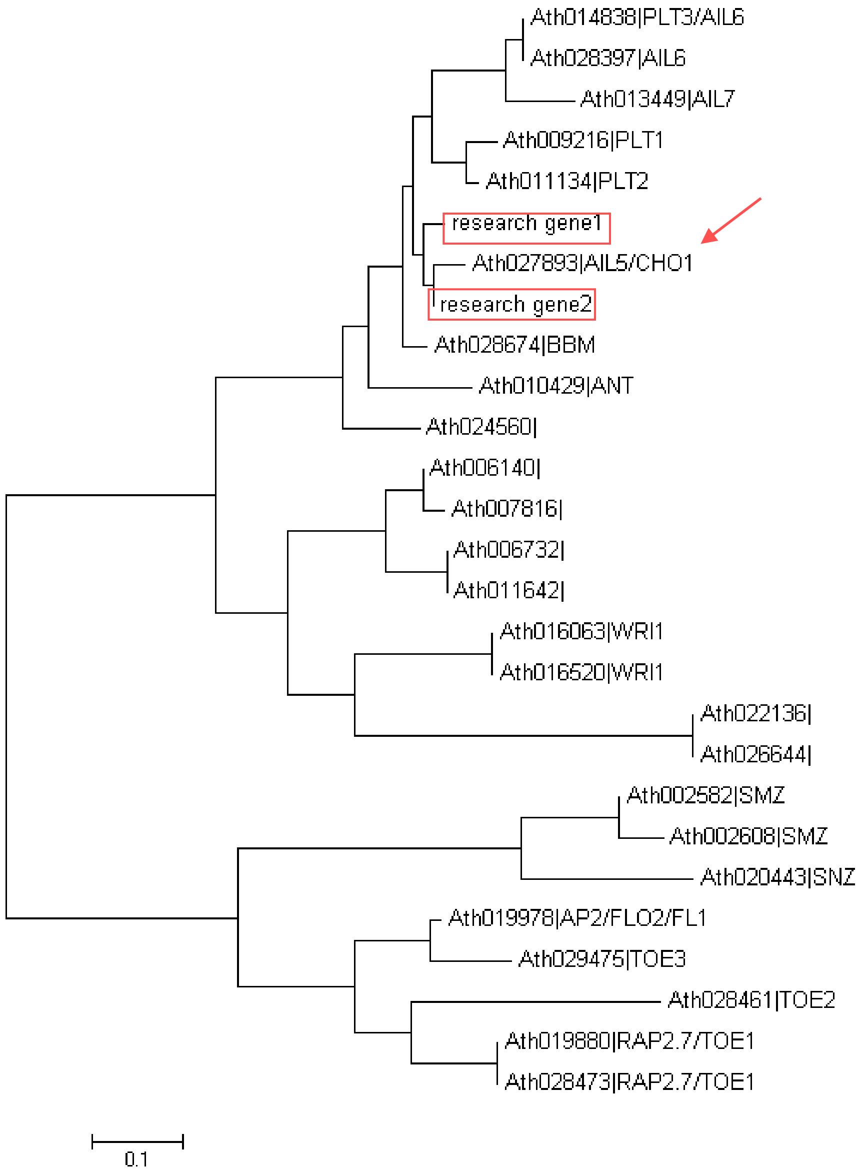
最大似然法



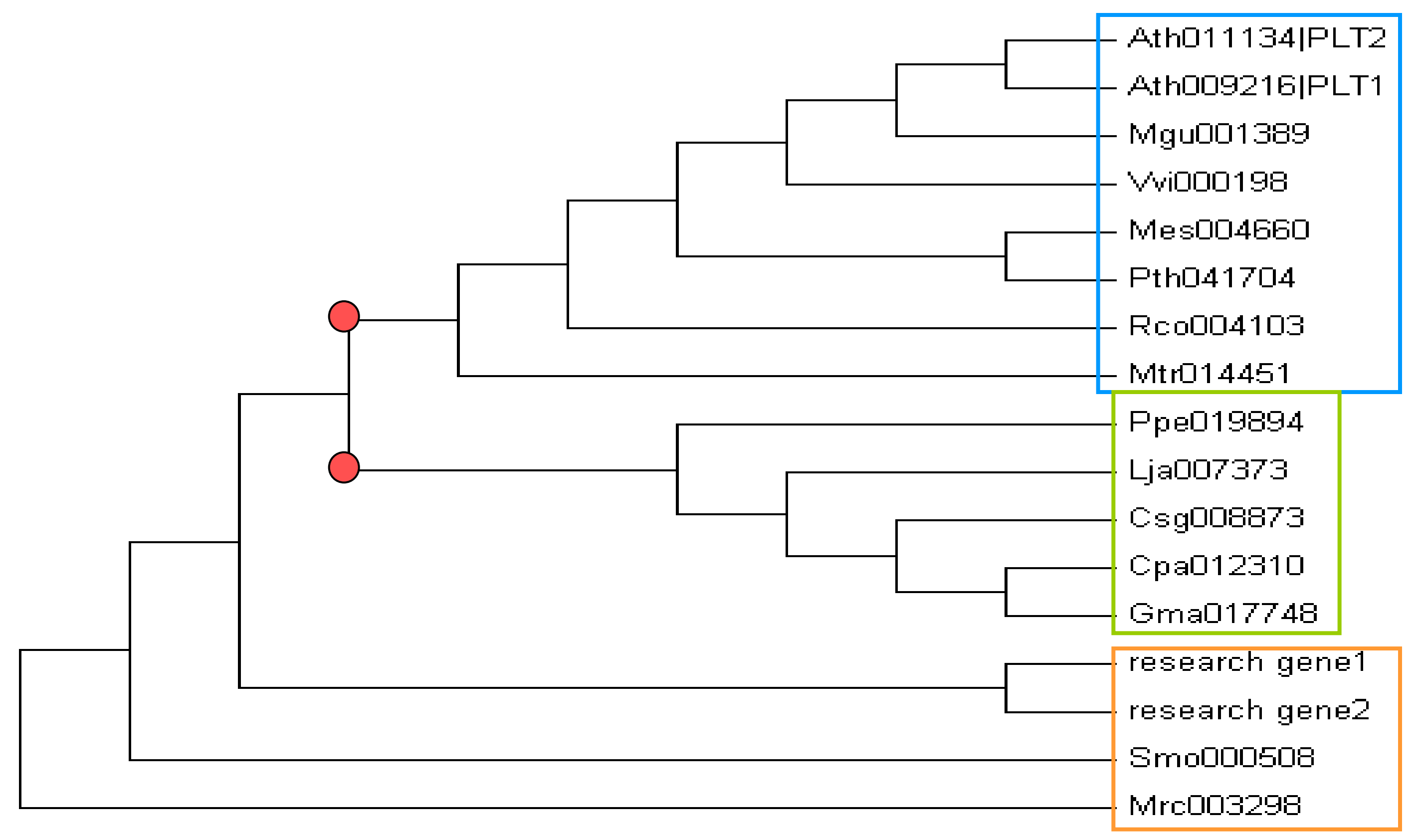
莱茵衣藻(Cre）、江南卷柏(smo) 、火炬松(Pta)、水稻(Osi)、拟南芥(Ath)

用拟南芥25个AP2基因与GeneA、B建树：　最大似然法





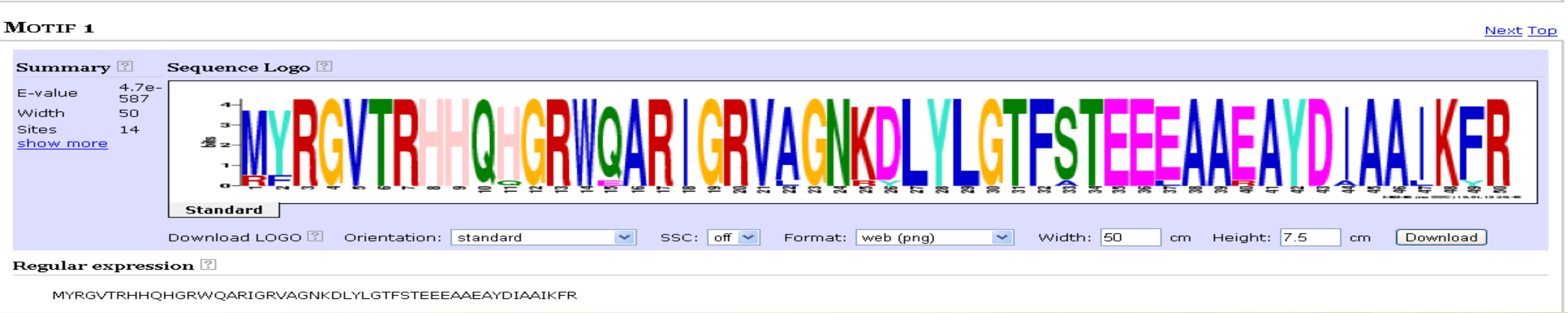
用拟南芥PLT1、其推测的直系同源基因与GeneA、B建树： 最大似然法



微单胞菌属（Mrc)、江南卷柏(smo)、大豆（Gma)、木瓜（Cpa）、黄瓜（Csg）、百脉根（Lja）、桃(Ppe)

苜蓿（Mtr)、蓖麻（Rco）、杨树（Pth）、木薯（Mes）、葡萄（Vvi）、猴面花（Mgu）、拟南芥(Ath)

拟南芥*PLT1*直系同源基因保守结构域分析（MEME）



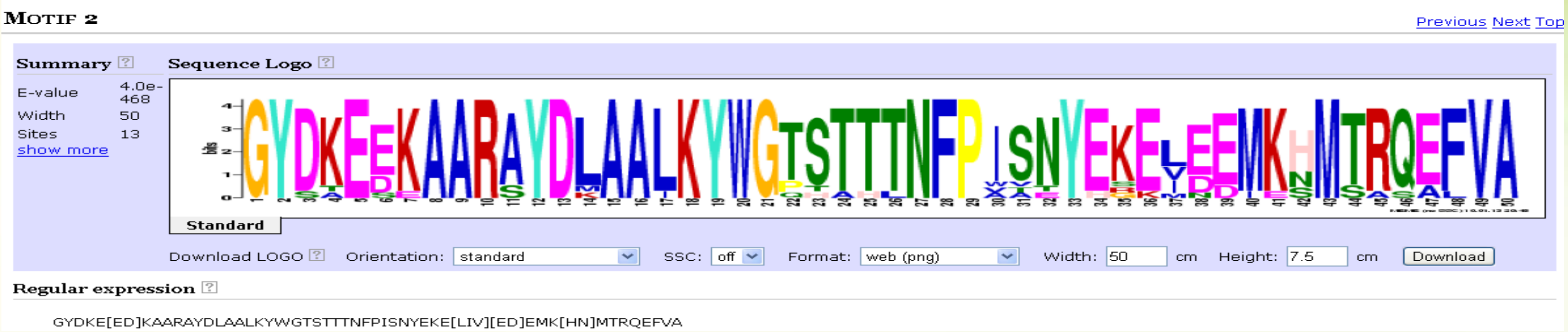
Motif 1

E-value 4.7e-587
Width 50
Sites 14
show more

Download LOGO Orientation: standard SSC: off Format: web (png) Width: 50 cm Height: 7.5 cm Download

Regular expression

MYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFSTEEEAAEAYDIAAIKFR



Motif 2

E-value 4.0e-468
Width 50
Sites 13
show more

Download LOGO Orientation: standard SSC: off Format: web (png) Width: 50 cm Height: 7.5 cm Download

Regular expression

GYDKE[ED]KAARAYDLAALKYWGTSTTTNFPISNYEKE[LIV][ED]EMK[HN]MTRQEFVA

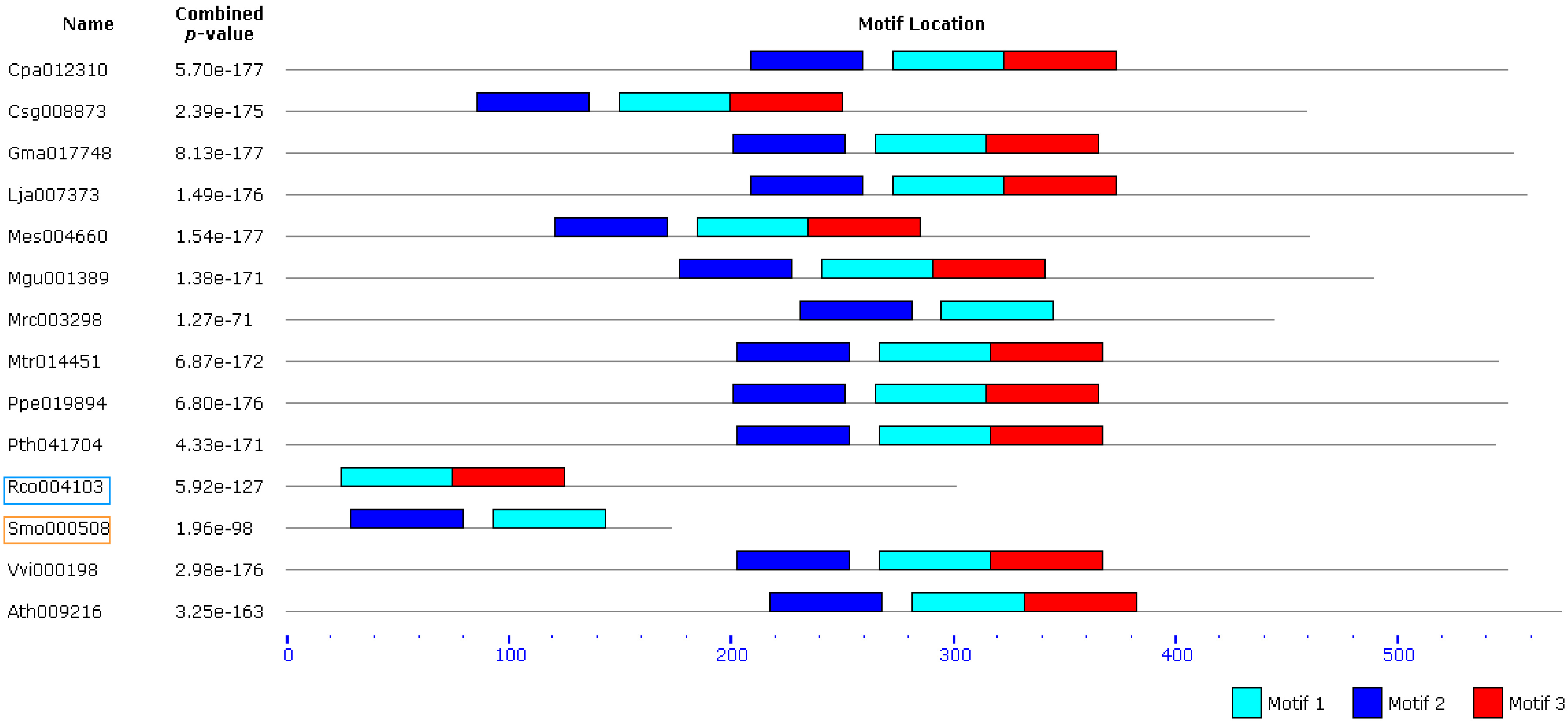
Motif 3

E-value 5.6e-442
Width 50
Sites 12
show more

Download LOGO Orientation: standard SSC: off Format: web (png) Width: 50 cm Height: 7.5 cm Download

Regular expression

GLNAVTNFDM[SN]RYDVK[SA]ILESNTLPIGGGAAKRLKEAQA[LI]ESSRKREEMI

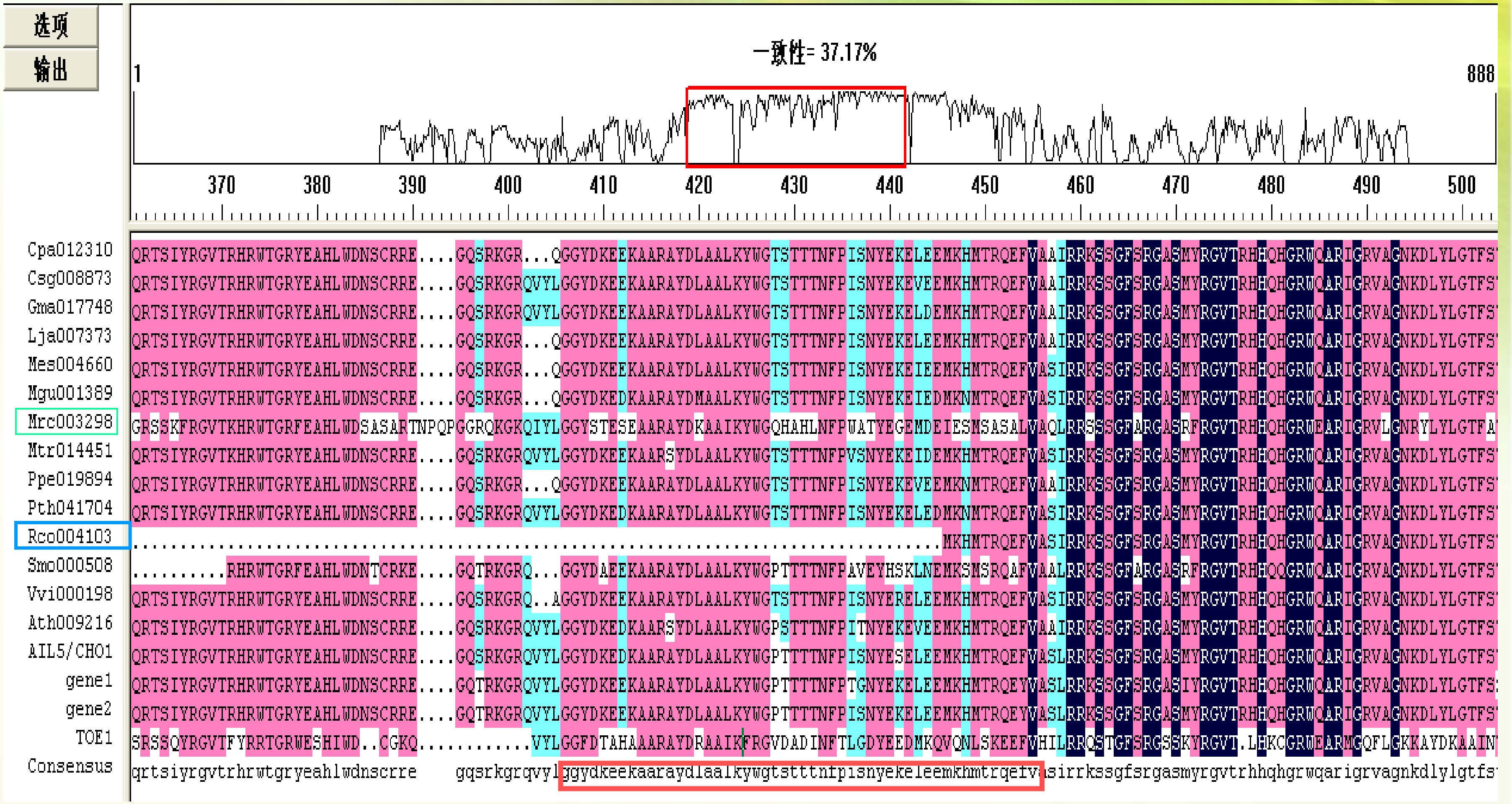


Name	Combined *p*-value
Cpa012310	5.70e-177
Csg008873	2.39e-175
Gma017748	8.13e-177
Lja007373	1.49e-176
Mes004660	1.54e-177
Mgu001389	1.38e-171
Mrc003298	1.27e-71
Mtr014451	6.87e-172
Ppe019894	6.80e-176
Pth041704	4.33e-171
Rco004103	5.92e-127
Smo000508	1.96e-98
Vvi000198	2.98e-176
Ath009216	3.25e-163

江南卷柏(smo)

蓖麻（Rco）

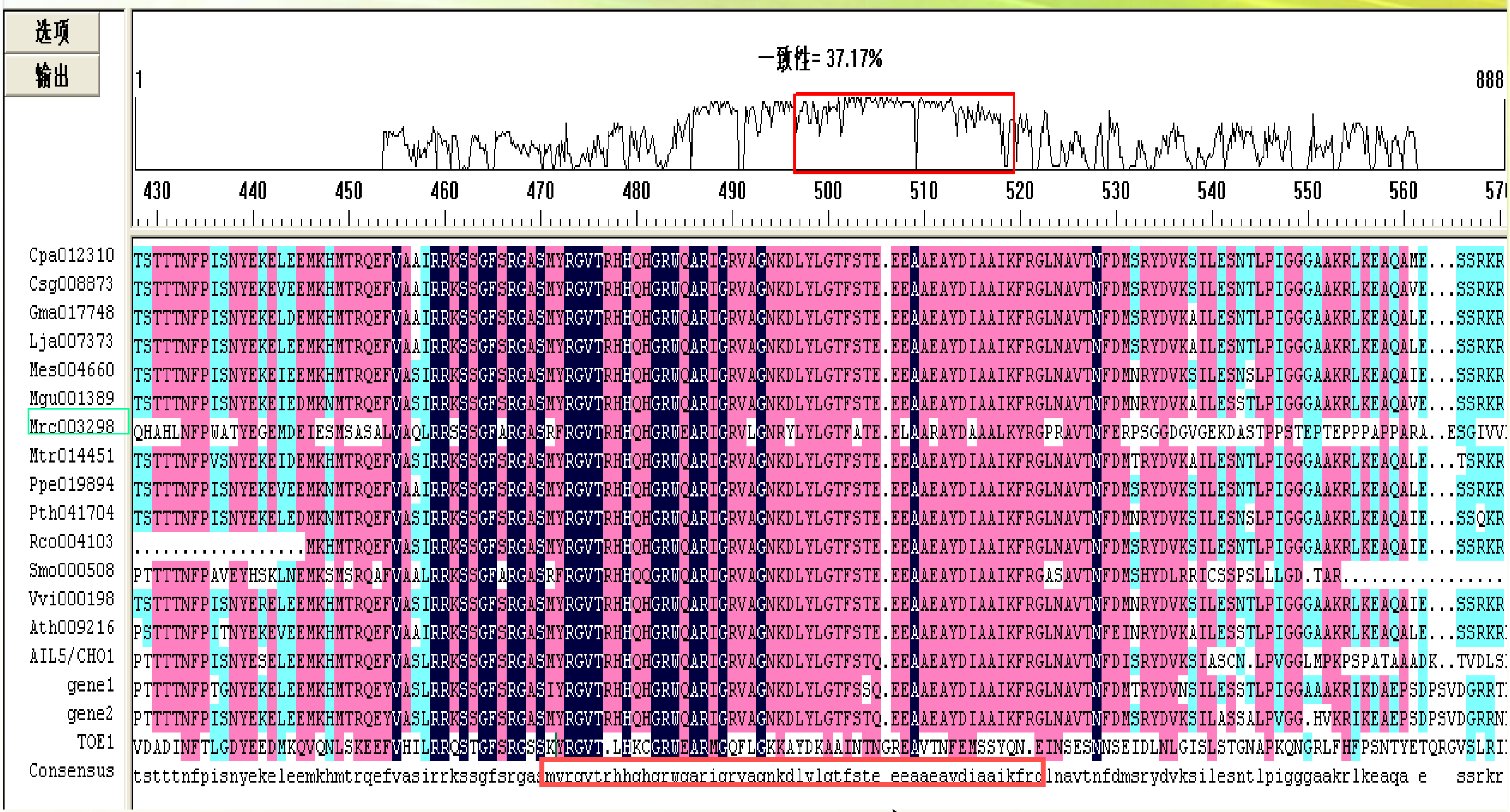
用DNAMAN进行全序列比对：



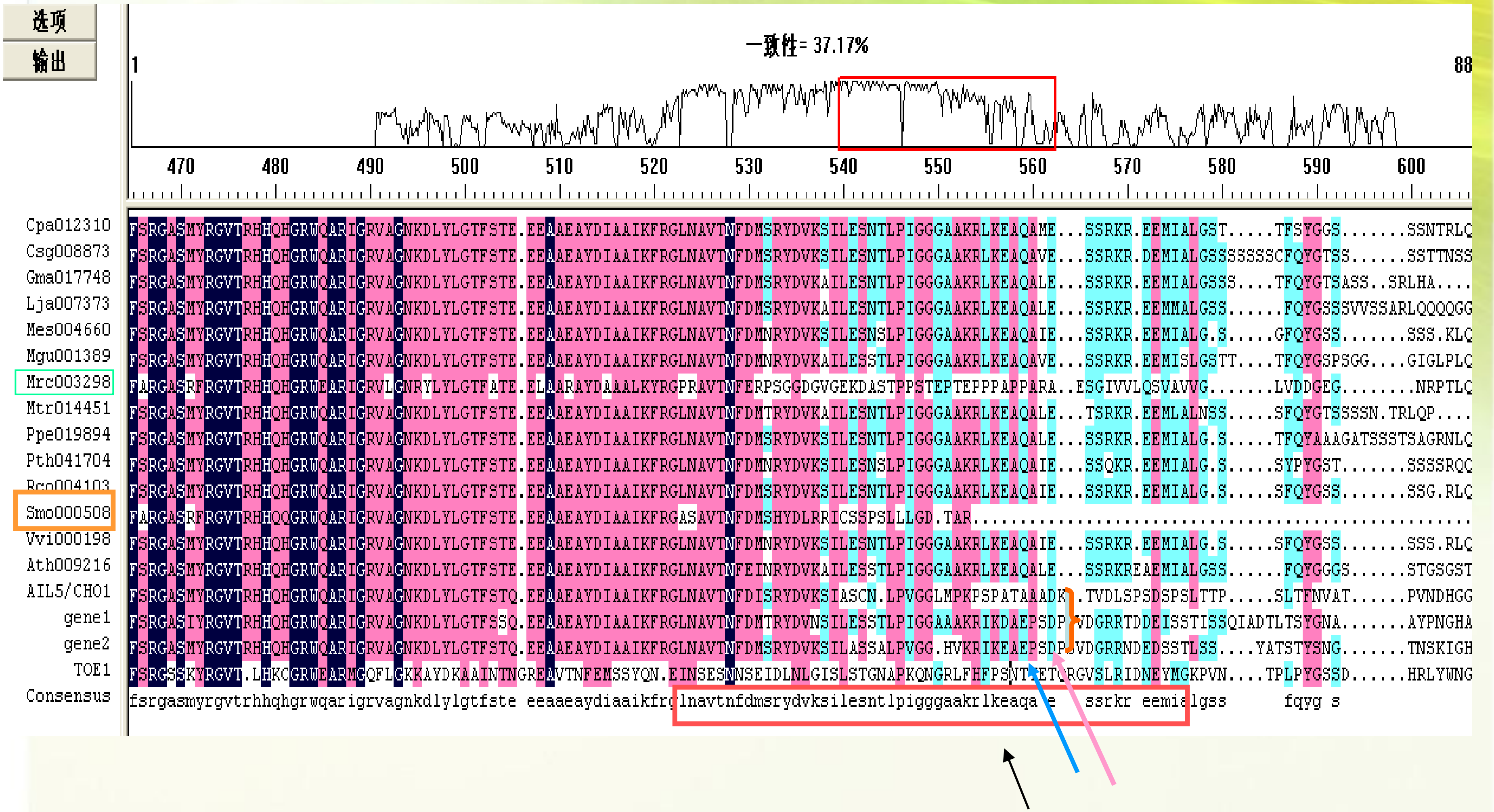
```
Cpa012310  QRTSIYRGVTRHRWTGRYEAHLWDNSCRRE....GQSRKGR...QGGYDKEEKAARAYDLAALKYWGTSTTTNFPISNYEKELEEMKHMTRQEFVAAIRRKSSGFSRGASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFS
Csg008873  QRTSIYRGVTRHRWTGRYEAHLWDNSCRRE....GQSRKGRQVYLGGYDKEEKAARAYDLAALKYWGTSTTTNFPISNYEKEVEEMKHMTRQEFVAAIRRKSSGFSRGASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFS
Gma017748  QRTSIYRGVTRHRWTGRYEAHLWDNSCRRE....GQSRKGRQVYLGGYDKEEKAARAYDLAALKYWGTSTTTNFPISNYEKELDEMKHMTRQEFVAAIRRKSSGFSRGASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFS
Lja007373  QRTSIYRGVTRHRWTGRYEAHLWDNSCRRE....GQSRKGR...QGGYDKEEKAARAYDLAALKYWGTSTTTNFPISNYEKELEEMKHMTRQEFVAAIRRKSSGFSRGASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFS
Mes004660  QRTSIYRGVTRHRWTGRYEAHLWDNSCRRE....GQSRKGR...QGGYDKEEKAARAYDLAALKYWGTSTTTNFPISNYEKEIEEMKHMTRQEFVASIRRKSSGFSRGASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFS
Mgu001389  QRTSIYRGVTRHRWTGRYEAHLWDNSCRRE....GQSRKGR...QGGYDKEDKAARAYDMAALKYWGTSTTTNFPISNYEKEIEDMKNMTRQEFVASIRRKSSGFSRGASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFS
Mrc003298  GRSSKFRGVTKHRWTGRFEAHLWDSASARTNPQPGGRQKGKQIYLGGYSTESEAARAYDKAAIKYWGQHAHLNFPWATYEGEMDEIESMSASALVAQLRRSSSGFARGASRFRGVTRHHQHGRWEARIGRVLGNRYLYLGTFA
Mtr014451  QRTSIYRGVTKHRWTGRYEAHLWDNSCRRE....GQSRKGRQVYLGGYDKEEKAARSYDLAALKYWGTSTTTNFPVSNYEKEIDEMKHMTRQEFVASIRRKSSGFSRGASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFS
Ppe019894  QRTSIYRGVTRHRWTGRYEAHLWDNSCRRE....GQSRKGR...QGGYDKEEKAARAYDLAALKYWGTSTTTNFPISNYEKEVEEMKNMTRQEFVAAIRRKSSGFSRGASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFS
Pth041704  QRTSIYRGVTRHRWTGRYEAHLWDNSCRRE....GQSRKGRQVYLGGYDKEDKAARAYDLAALKYWGTSTTTNFPISNYEKELEDMKNMTRQEFVASIRRKSSGFSRGASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFS
Rco004103  .....................................................................................MKHMTRQEFVASIRRKSSGFSRGASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFS
Smo000508  ..........RHRWTGRFEAHLWDNTCRKE....GQTRKGRQ...GGYDAEEKAARAYDLAALKYWGPTTTTNFPAVEYHSKLNEMKSMSRQAFVAALRRKSSGFARGASRFRGVTRHHQQGRWQARIGRVAGNKDLYLGTFS
Vvi000198  QRTSIYRGVTRHRWTGRYEAHLWDNSCRRE....GQSRKGRQ..AGGYDKEEKAARAYDLAALKYWGTSTTTNFPISNYERELEEMKHMTRQEFVASIRRKSSGFSRGASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFS
Ath009216  QRTSIYRGVTRHRWTGRYEAHLWDNSCRRE....GQSRKGRQVYLGGYDKEDKAARSYDLAALKYWGPSTTTNFPITNYEKEVEEMKHMTRQEFVAAIRRKSSGFSRGASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFS
AIL5/CHO1  QRTSIYRGVTRHRWTGRYEAHLWDNSCRRE....GQSRKGRQVYLGGYDKEDKAARAYDLAALKYWGPTTTTNFPISNYESELEEMKHMTRQEFVASLRRKSSGFSRGASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFS
    gene1  QRTSIYRGVTRHRWTGRYEAHLWDNSCRRE....GQTRKGRQVYLGGYDKEEKAARAYDLAALKYWGPTTTTNFPTGNYEKELEEMKHMTRQEYVASLRRKSSGFSRGASIYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFS
    gene2  QRTSIYRGVTRHRWTGRYEAHLWDNSCRRE....GQTRKGRQVYLGGYDKEEKAARAYDLAALKYWGPTTTTNFPISNYEKELEEMKHMTRQEYVASLRRKSSGFSRGASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFS
     TOE1  SRSSQYRGVTFYRRTGRWESHIWD..CGKQ............VYLGGFDTAHAAARAYDRAAIKFRGVDADINFTLGDYEEDMKQVQNLSKEEFVHILRRQSTGFSRGSSKYRGVT.LHKCGRWEARMGQFLGKKAYDKAAIN
Consensus  qrtsiyrgvtrhrwtgryeahlwdnscrre    gqsrkgrqvylggydkeekaaraydlaalkywgtstttnfpisnyekeleemkhmtrqefvasirrkssgfsrgasmyrgvtrhhqhgrwqarigrvagnkdlylgtfs
```

蓖麻（Rco）

微单胞菌属（Mrc）



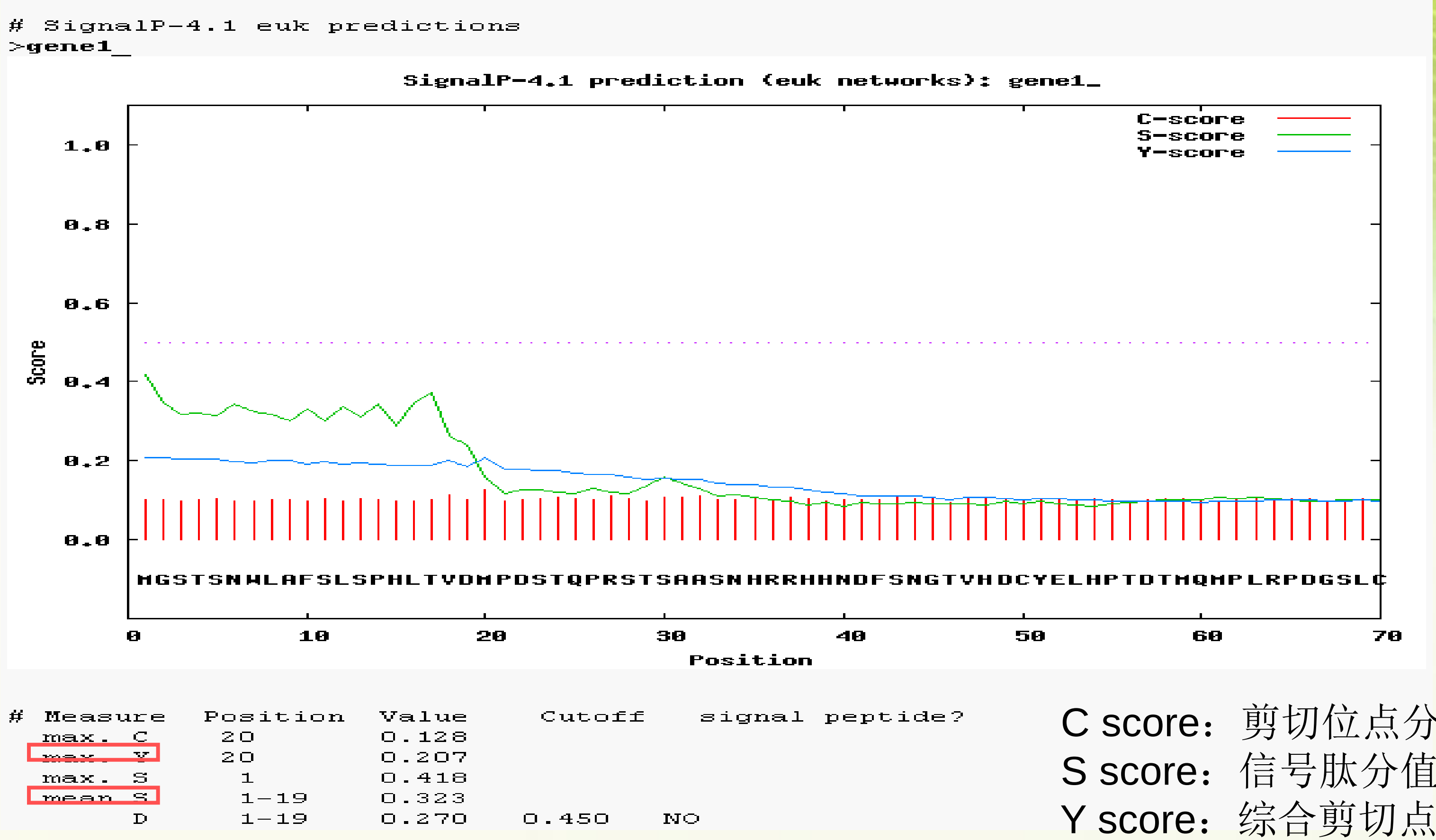
```
Cpa012310  TSTTTNFPISNYEKELEEMKHMTRQEFVAAIRRKSSGFSRGASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFSTE.EEAAEAYDIAAIKFRGLNAVTNFDMSRYDVKSILESNTLPIGGGAAKRLKEAQAME...SSRKR
Csg008873  TSTTTNFPISNYEKEVEEMKHMTRQEFVAAIRRKSSGFSRGASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFSTE.EEAAEAYDIAAIKFRGLNAVTNFDMSRYDVKSILESNTLPIGGGAAKRLKEAQAVE...SSRKR
Gma017748  TSTTTNFPISNYEKELDEMKHMTRQEFVAAIRRKSSGFSRGASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFSTE.EEAAEAYDIAAIKFRGLNAVTNFDMSRYDVKAILESNTLPIGGGAAKRLKEAQALE...SSRKR
Lja007373  TSTTTNFPISNYEKELEEMKHMTRQEFVAAIRRKSSGFSRGASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFSTE.EEAAEAYDIAAIKFRGLNAVTNFDMSRYDVKAILESNTLPIGGGAAKRLKEAQALE...SSRKR
Mes004660  TSTTTNFPISNYEKEIEEMKHMTRQEFVASIRRKSSGFSRGASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFSTE.EEAAEAYDIAAIKFRGLNAVTNFDMNRYDVKSILESNSLPIGGGAAKRLKEAQAIE...SSRKR
Mgu001389  TSTTTNFPISNYEKEIEDMKNMTRQEFVASIRRKSSGFSRGASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFSTE.EEAAEAYDIAAIKFRGLNAVTNFDMNRYDVKAILESSTLPIGGGAAKRLKEAQAVE...SSRKR
Mrc003298  QHAHLNFPWATYEGEMDEIESMSASALVAQLRRSSSGFARGASRFRGVTRHHQHGRWEARIGRVLGNRYLYLGTFATE.ELAARAYDAAALKYRGPRAVTNFERPSGGDGVGEKDASTPPSTEPTEPPPAPPARA..ESGIVV
Mtr014451  TSTTTNFPVSNYEKEIDEMKHMTRQEFVASIRRKSSGFSRGASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFSTE.EEAAEAYDIAAIKFRGLNAVTNFDMTRYDVKAILESNTLPIGGGAAKRLKEAQALE...TSRKR
Ppe019894  TSTTTNFPISNYEKEVEEMKNMTRQEFVAAIRRKSSGFSRGASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFSTE.EEAAEAYDIAAIKFRGLNAVTNFDMSRYDVKSILESNTLPIGGGAAKRLKEAQALE...SSRKR
Pth041704  TSTTTNFPISNYEKELEDMKNMTRQEFVASIRRKSSGFSRGASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFSTE.EEAAEAYDIAAIKFRGLNAVTNFDMNRYDVKSILESNSLPIGGGAAKRLKEAQAIE...SSQKR
Rco004103  ..................MKHMTRQEFVASIRRKSSGFSRGASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFSTE.EEAAEAYDIAAIKFRGLNAVTNFDMSRYDVKSILESNTLPIGGGAAKRLKEAQAIE...SSRKR
Smo000508  PTTTTNFPAVEYHSKLNEMKSMSRQAFVAALRRKSSGFARGASRFRGVTRHHQQGRWQARIGRVAGNKDLYLGTFSTE.EEAAEAYDIAAIKFRGASAVTNFDMSHYDLRRICSSPSLLLGD.TAR...............
Vvi000198  TSTTTNFPISNYERELEEMKHMTRQEFVASIRRKSSGFSRGASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFSTE.EEAAEAYDIAAIKFRGLNAVTNFDMNRYDVKSILESNTLPIGGGAAKRLKEAQAIE...SSRKR
Ath009216  PSTTTNFPITNYEKEVEEMKHMTRQEFVAAIRRKSSGFSRGASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFSTE.EEAAEAYDIAAIKFRGLNAVTNFEINRYDVKAILESSTLPIGGGAAKRLKEAQALE...SSRKR
AIL5/CHO1  PTTTTNFPISNYESELEEMKHMTRQEFVASLRRKSSGFSRGASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFSTQ.EEAAEAYDIAAIKFRGLNAVTNFDISRYDVKSIASCN.LPVGGLMPKPSPATAAADK..TVDLS
    gene1  PTTTTNFPTGNYEKELEEMKHMTRQEYVASLRRKSSGFSRGASIYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFSSQ.EEAAEAYDIAAIKFRGLNAVTNFDMTRYDVNSILESSTLPIGGAAAKRIKDAEPSDPSVDGRRT
    gene2  PTTTTNFPISNYEKELEEMKHMTRQEYVASLRRKSSGFSRGASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFSTQ.EEAAEAYDIAAIKFRGLNAVTNFDMSRYDVKSILASSALPVGG.HVKRIKEAEPSDPSVDGRRN
     TOE1  VDADINFTLGDYEEDMKQVQNLSKEEFVHILRRQSTGFSRGSSKYRGVT.LHKCGRWEARMGQFLGKKAYDKAAINTNGREAVTNFEMSSYQN.EINSESNNSEIDLNLGISLSTGNAPKQNGRLFHFPSNTYETQRGVSLRI
Consensus  tstttnfpisnyekeleemkhmtrqefvasirrkssgfsrgasmyrgvtrhhqhgrwqarigrvagnkdlylgtfste eeaaeaydiaaikfrglnavtnfdmsrydvksilesntlpigggaakrlkeaqa e   ssrkr
```



```
Cpa012310  FSRGASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFSTE.EEAAEAYDIAAIKFRGLNAVTNFDMSRYDVKSILESNTLPIGGGAAKRLKEAQAME...SSRKR.EEMIALGST.....TFSYGGS.......SSNTRLQ
Csg008873  FSRGASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFSTE.EEAAEAYDIAAIKFRGLNAVTNFDMSRYDVKSILESNTLPIGGGAAKRLKEAQAVE...SSRKR.DEMIALGSSSSSSSCFQYGTSS......SSTTNSS
Gma017748  FSRGASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFSTE.EEAAEAYDIAAIKFRGLNAVTNFDMSRYDVKAILESNTLPIGGGAAKRLKEAQALE...SSRKR.EEMIALGSSS....TFQYGTSASS..SRLHA....
Lja007373  FSRGASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFSTE.EEAAEAYDIAAIKFRGLNAVTNFDMSRYDVKAILESNTLPIGGGAAKRLKEAQALE...SSRKR.EEMMALGSS......FQYGSSSVVSSARLQQQGG
Mes004660  FSRGASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFSTE.EEAAEAYDIAAIKFRGLNAVTNFDMNRYDVKSILESNSLPIGGGAAKRLKEAQAIE...SSRKR.EEMIALG.S.....GFQYGSS.......SSS.KLQ
Mgu001389  FSRGASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFSTE.EEAAEAYDIAAIKFRGLNAVTNFDMNRYDVKAILESSTLPIGGGAAKRLKEAQAVE...SSRKR.EEMISLGSTT....TFQYGSPSGG....GIGLPLQ
Mrc003298  FARGASRFRGVTRHHQHGRWEARIGRVLGNRYLYLGTFATE.ELAARAYDAAALKYRGPRAVTNFERPSGGDGVGEKDASTPPSTEPTEPPPAPPARA..ESGIVVLQSVAVVG.......LVDDGEG........NRPTLQ
Mtr014451  FSRGASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFSTE.EEAAEAYDIAAIKFRGLNAVTNFDMTRYDVKAILESNTLPIGGGAAKRLKEAQALE...TSRKR.EEMLALNSS.....SFQYGTSSSSN.TRLQP....
Ppe019894  FSRGASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFSTE.EEAAEAYDIAAIKFRGLNAVTNFDMSRYDVKSILESNTLPIGGGAAKRLKEAQALE...SSRKR.EEMIALG.S.....TFQYAAAGATSSSTSAGRNLQ
Pth041704  FSRGASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFSTE.EEAAEAYDIAAIKFRGLNAVTNFDMNRYDVKSILESNSLPIGGGAAKRLKEAQAIE...SSQKR.EEMIALG.S.....SYPYGST.......SSSSRQQ
Rco004103  FSRGASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFSTE.EEAAEAYDIAAIKFRGLNAVTNFDMSRYDVKSILESNTLPIGGGAAKRLKEAQAIE...SSRKR.EEMIALG.S.....SFQYGSS.......SSG.RLQ
Smo000508  FARGASRFRGVTRHHQQGRWQARIGRVAGNKDLYLGTFSTE.EEAAEAYDIAAIKFRGASAVTNFDMSHYDLRRICSSPSLLLGD.TAR.....................................................
Vvi000198  FSRGASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFSTE.EEAAEAYDIAAIKFRGLNAVTNFDMNRYDVKSILESNTLPIGGGAAKRLKEAQAIE...SSRKR.EEMIALG.S.....SFQYGSS.......SSS.RLQ
Ath009216  FSRGASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFSTE.EEAAEAYDIAAIKFRGLNAVTNFEINRYDVKAILESSTLPIGGGAAKRLKEAQALE...SSRKREAEMIALGSS......FQYGGGS......STGSGST
AIL5/CHO1  FSRGASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFSTQ.EEAAEAYDIAAIKFRGLNAVTNFDISRYDVKSIASCN.LPVGGLMPKPSPATAAADK.TVDLSPSDSPSLTTP.....SLTFNVAT......PVNDHGG
gene1      FSRGASIYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFSSQ.EEAAEAYDIAAIKFRGLNAVTNFDMTRYDVNSILESSTLPIGGAAAKRIKDAEPSDP[illegible]VDGRRTDDEISSTISSQIADTLTSYGNA.......AYPNGHA
gene2      FSRGASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFSTQ.EEAAEAYDIAAIKFRGLNAVTNFDMSRYDVKSILASSALPVGG.HVKRIKEAEPSDP[illegible]VDGRRNDEDSSTLSS....YATSTYSNG.......TNSKIGH
TOE1       FSRGSSKYRGVT.LHKCGRWEARMGQFLGKKAYDKAAINTNGREAVTNFEMSSYQN.EINSESNNSEIDLNLGISLSTGNAPKQNGRLFHFPSNT[illegible]ETQRGVSLRIDNEYMGKPVN....TPLPYGSSD......HRLYWNG
Consensus  fsrgasmyrgvtrhhqhgrwqarigrvagnkdlylgtfste eeaaeaydiaaikfrglnavtnfdmsrydvksilesntlpigggaakrlkeaqa e   ssrkr eemialgss      fqyg s
```

江南卷柏(smo)

信号肽分析 SignalP 4.1 Server



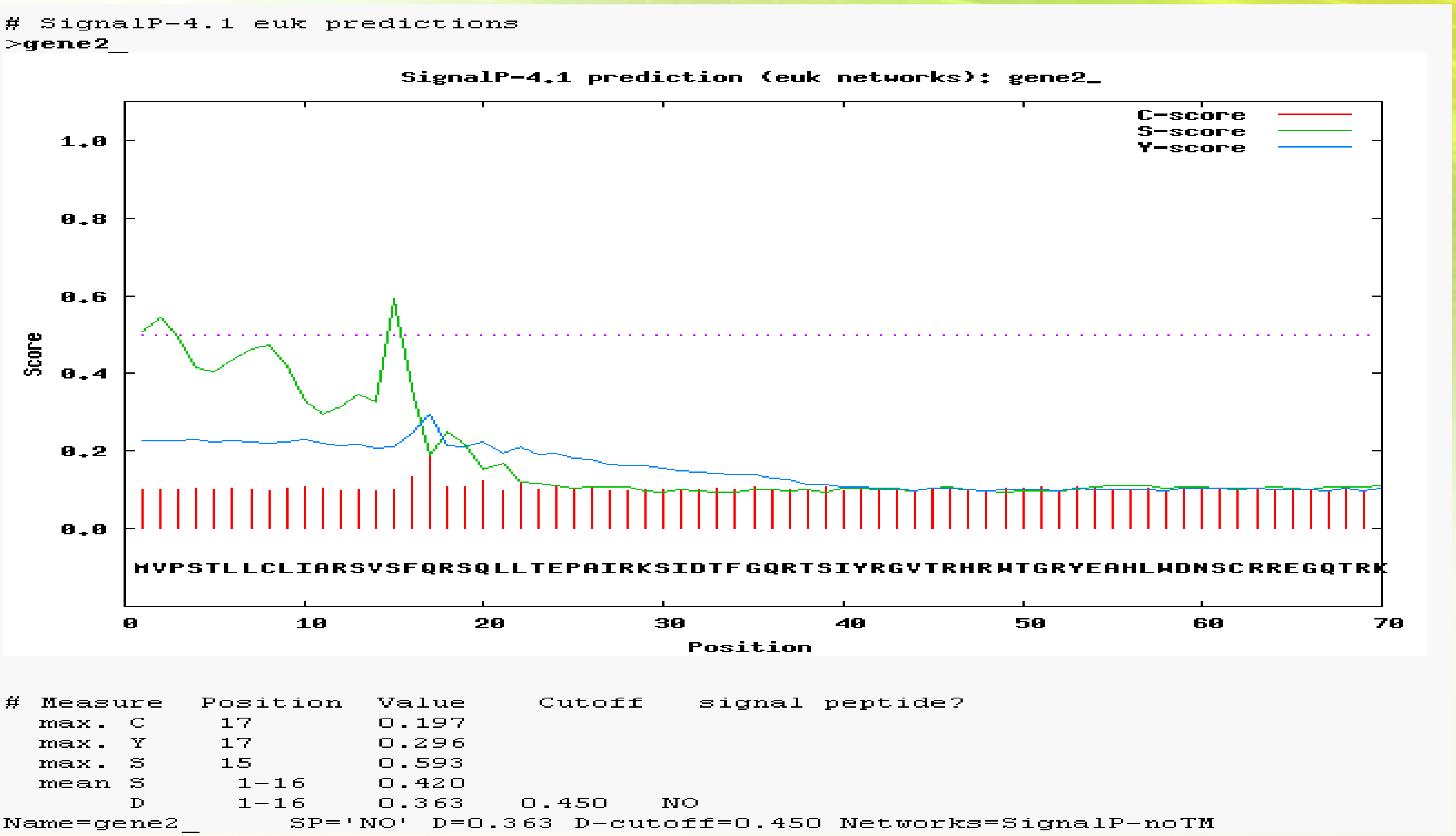
```
# SignalP-4.1 euk predictions
>gene1_
```

```
# Measure  Position  Value    Cutoff   signal peptide?
  max. C    20       0.128
  max. Y    20       0.207
  max. S     1       0.418
  mean S     1-19    0.323
       D     1-19    0.270   0.450    NO
```

C score：剪切位点分值

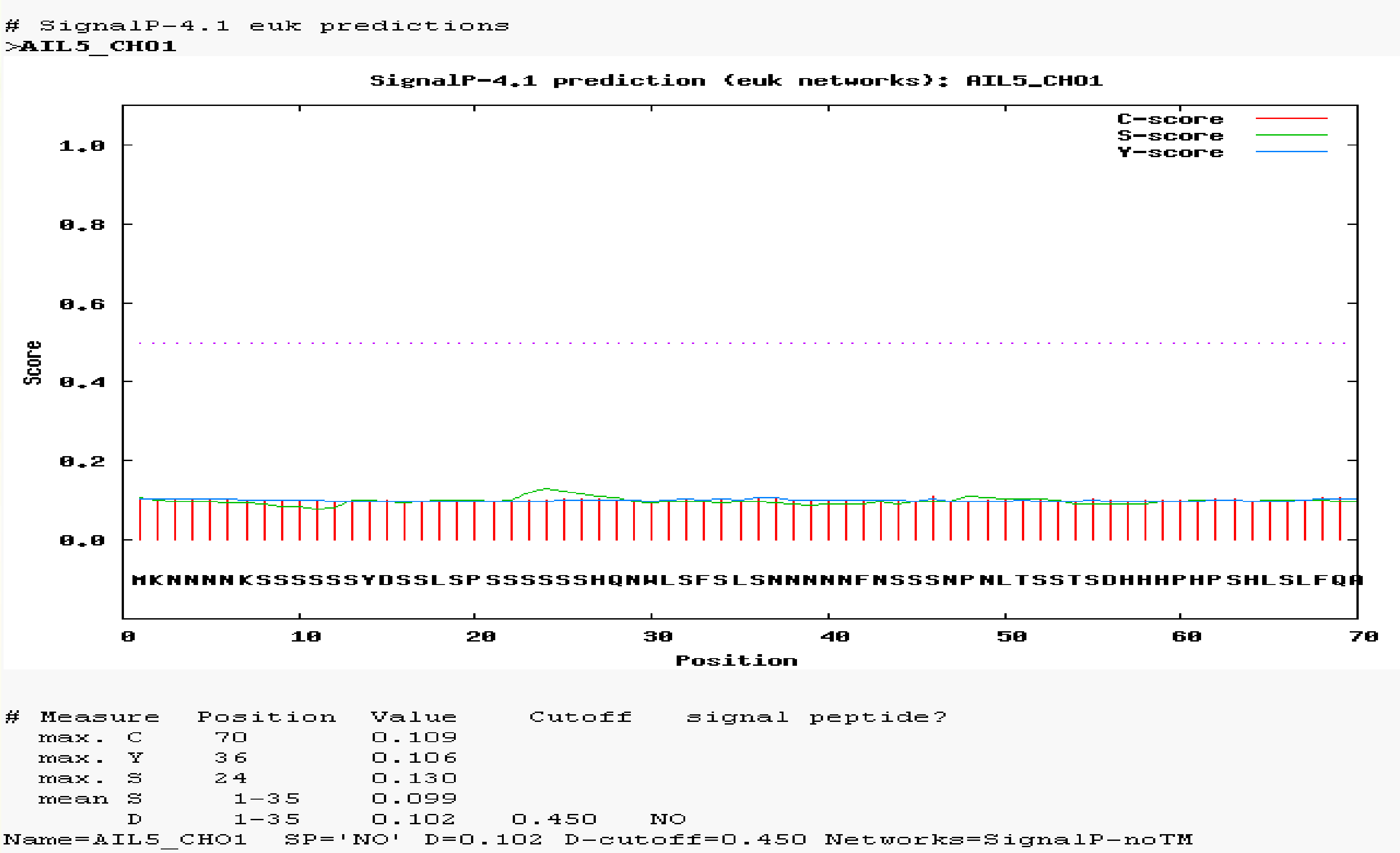
S score：信号肽分值

Y score：综合剪切点分值



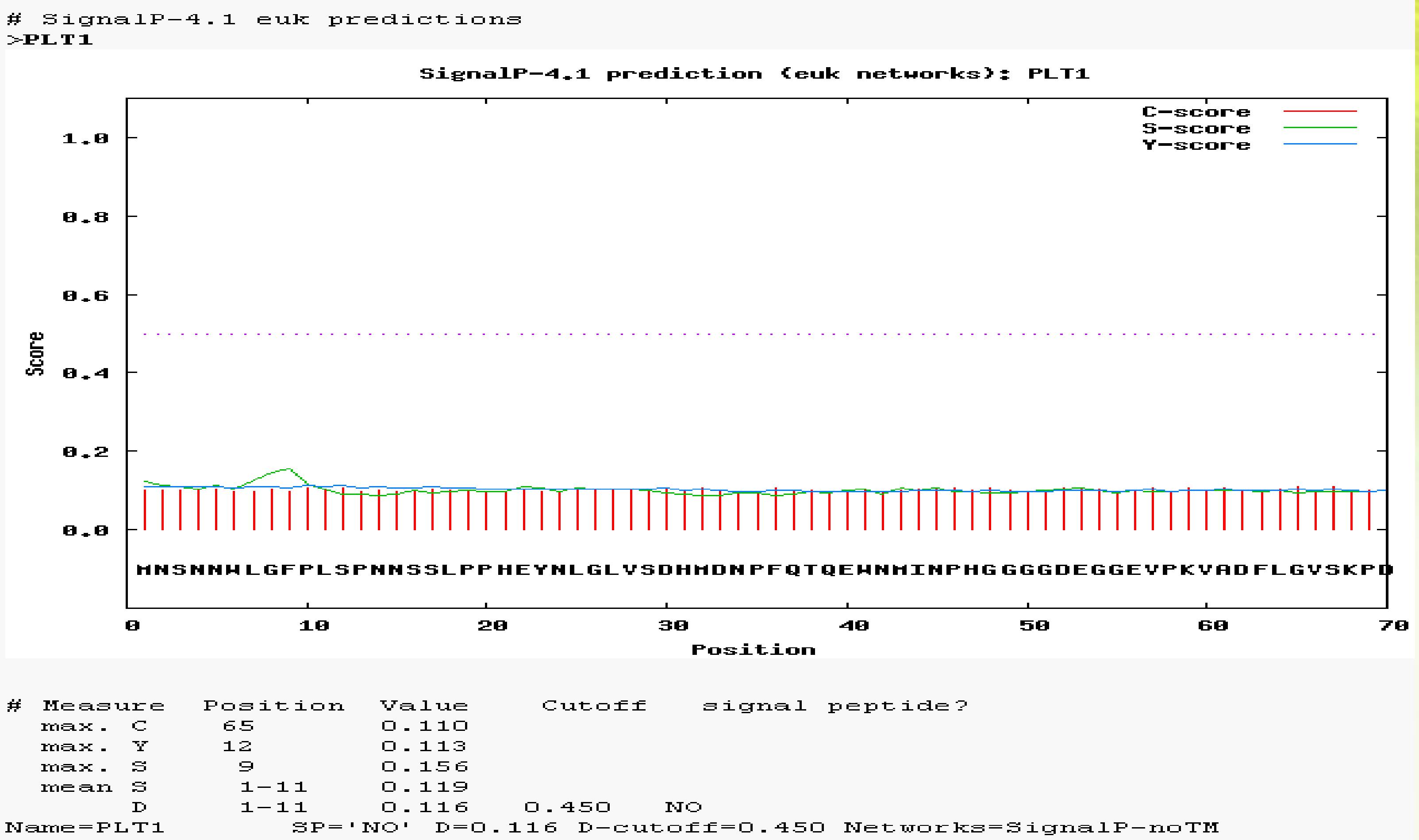
```
# SignalP-4.1 euk predictions
>gene2_
```

```
# Measure  Position  Value    Cutoff   signal peptide?
  max. C    17       0.197
  max. Y    17       0.296
  max. S    15       0.593
  mean S     1-16    0.420
       D     1-16    0.363   0.450   NO
Name=gene2_     SP='NO' D=0.363 D-cutoff=0.450 Networks=SignalP-noTM
```



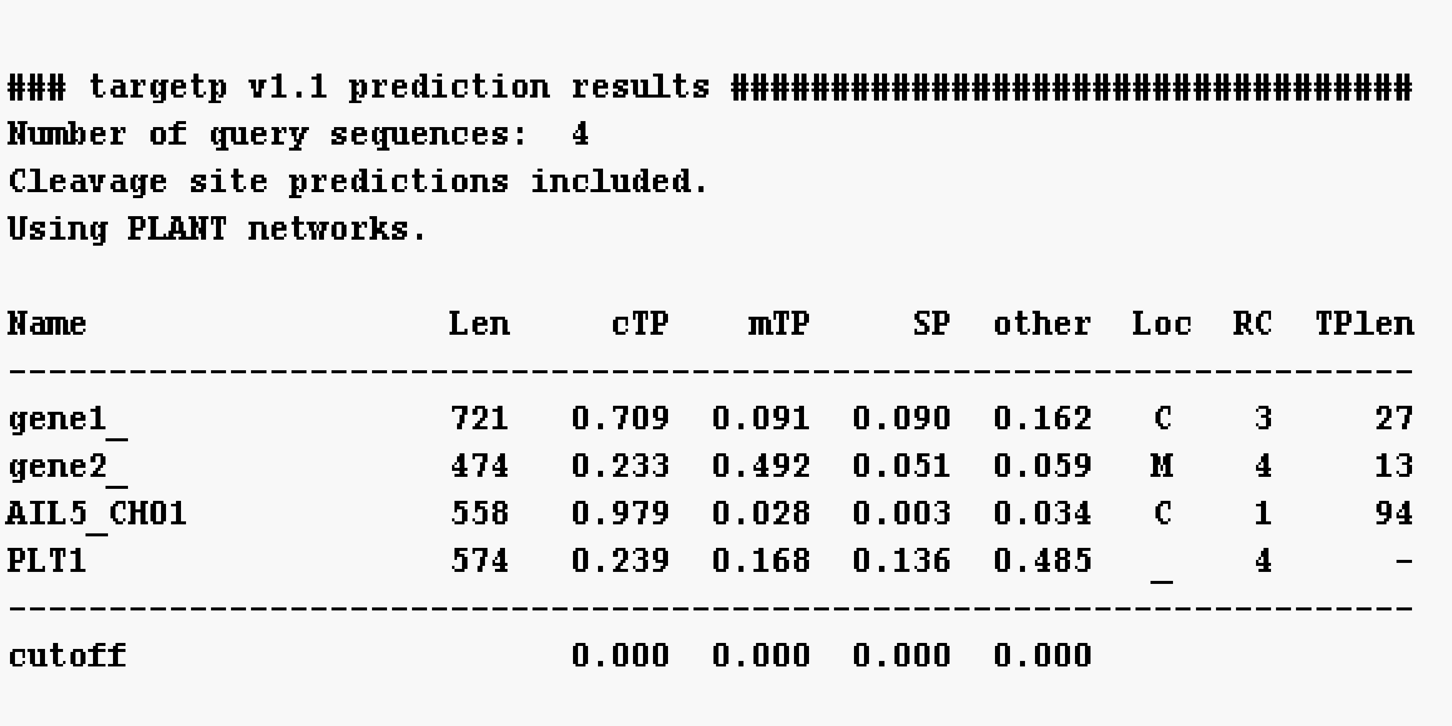
```
# SignalP-4.1 euk predictions
>AIL5_CH01
```

```
# Measure  Position  Value    Cutoff   signal peptide?
  max. C    70        0.109
  max. Y    36        0.106
  max. S    24        0.130
  mean S     1-35     0.099
       D     1-35     0.102   0.450    NO
Name=AIL5_CH01  SP='NO' D=0.102 D-cutoff=0.450 Networks=SignalP-noTM
```



```
# SignalP-4.1 euk predictions
>PLT1
```

```
# Measure  Position  Value    Cutoff   signal peptide?
  max. C    65       0.110
  max. Y    12       0.113
  max. S     9       0.156
  mean S     1-11    0.119
       D     1-11    0.116   0.450    NO
Name=PLT1       SP='NO' D=0.116 D-cutoff=0.450 Networks=SignalP-noTM
```

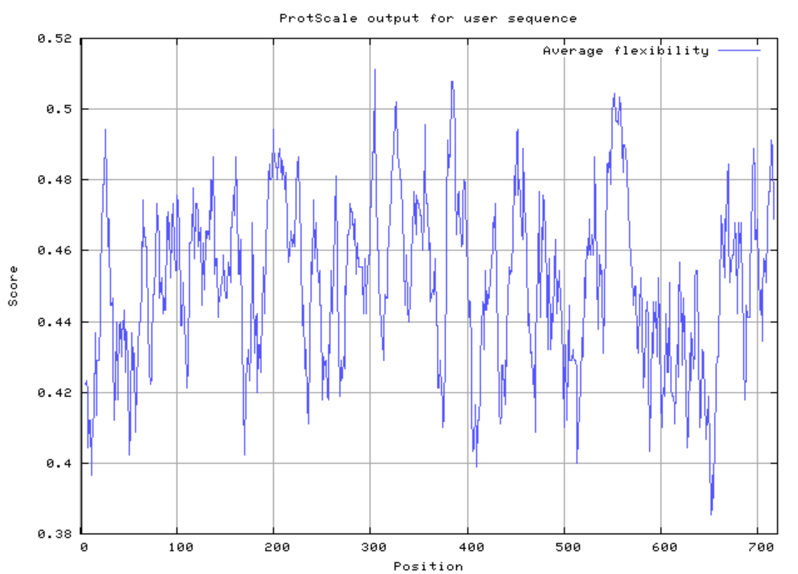


```
### targetp v1.1 prediction results ##################################
Number of query sequences:  4
Cleavage site predictions included.
Using PLANT networks.

Name                  Len     cTP    mTP     SP  other  Loc  RC  TPlen
----------------------------------------------------------------------
gene1_                721   0.709  0.091  0.090  0.162   C    3     27
gene2_                474   0.233  0.492  0.051  0.059   M    4     13
AIL5_CH01             558   0.979  0.028  0.003  0.034   C    1     94
PLT1                  574   0.239  0.168  0.136  0.485   _    4      -
----------------------------------------------------------------------
cutoff                      0.000  0.000  0.000  0.000
```

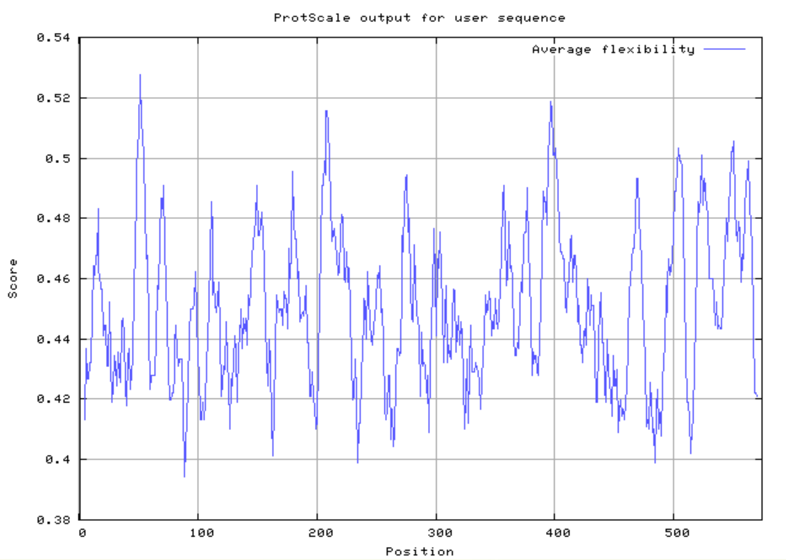
理化性质分析

柔性分析：**(EXPASSY Protscale)**

基因A:

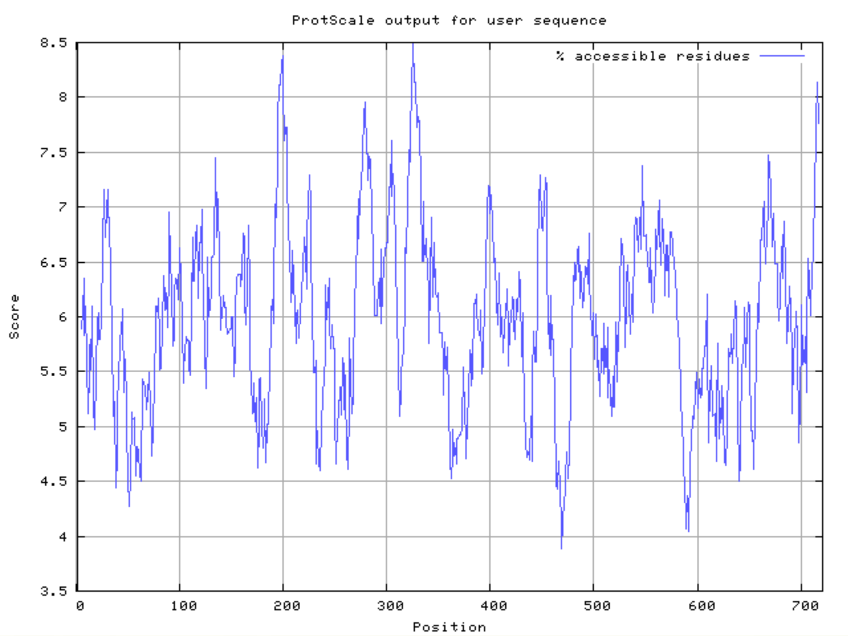


PLT1:

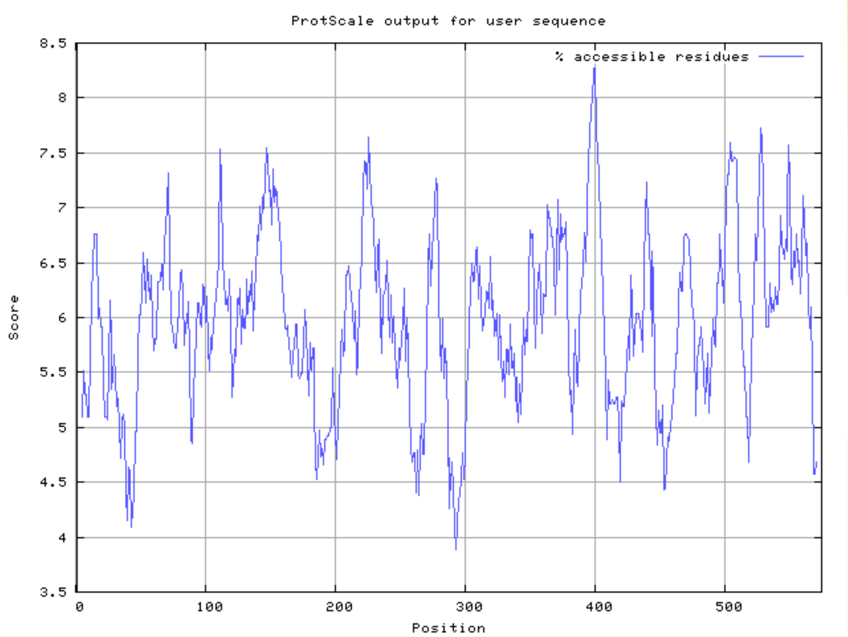


溶剂可及性分析：

基因A:

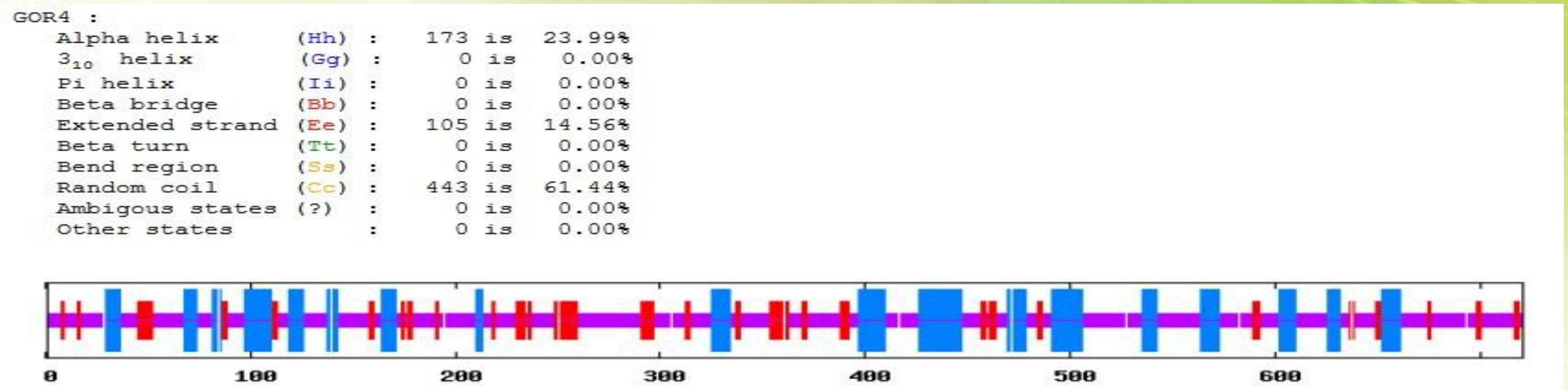


PLT1:



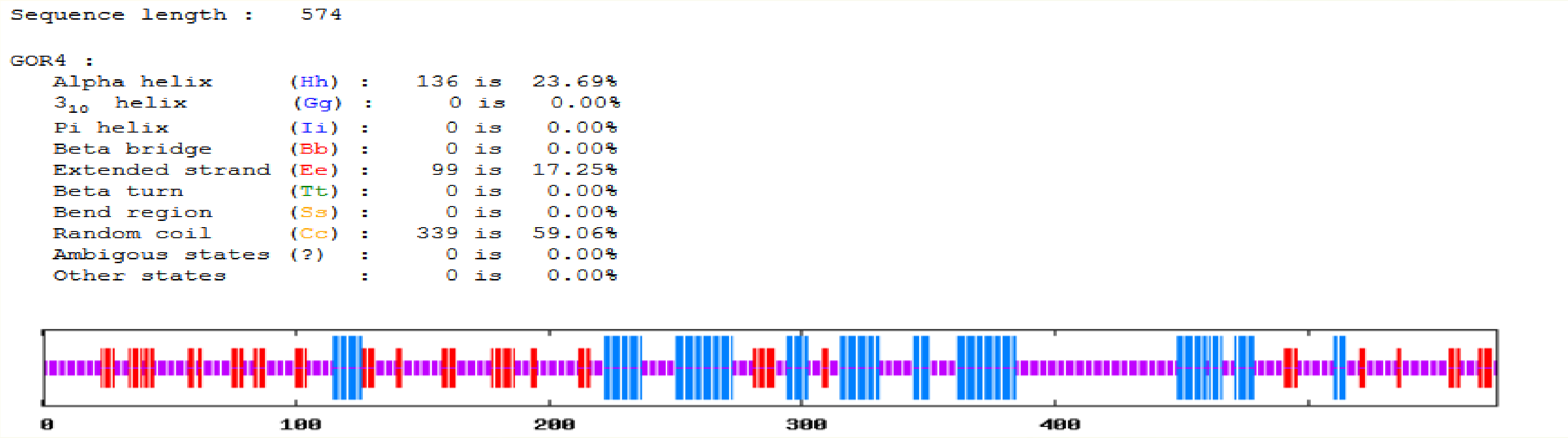
蛋白二级结构预测（GOR）

基因A:



```
GOR4 :
   Alpha helix        (Hh) :    173 is   23.99%
   3₁₀  helix         (Gg) :      0 is    0.00%
   Pi helix           (Ii) :      0 is    0.00%
   Beta bridge        (Bb) :      0 is    0.00%
   Extended strand    (Ee) :    105 is   14.56%
   Beta turn          (Tt) :      0 is    0.00%
   Bend region        (Ss) :      0 is    0.00%
   Random coil        (Cc) :    443 is   61.44%
   Ambigous states    (?)  :      0 is    0.00%
   Other states            :      0 is    0.00%
```

PLT1:



```
Sequence length :    574

GOR4 :
   Alpha helix        (Hh) :    136 is   23.69%
   3₁₀  helix         (Gg) :      0 is    0.00%
   Pi helix           (Ii) :      0 is    0.00%
   Beta bridge        (Bb) :      0 is    0.00%
   Extended strand    (Ee) :     99 is   17.25%
   Beta turn          (Tt) :      0 is    0.00%
   Bend region        (Ss) :      0 is    0.00%
   Random coil        (Cc) :    339 is   59.06%
   Ambigous states    (?)  :      0 is    0.00%
   Other states            :      0 is    0.00%
```

谢谢大家！