

水稻斑点叶 *spl7* 的 生物信息学分析

第二组成员：

王彩红、黄奇娜(主讲)、
李赓觅、刘群恩



致谢！

吉祥如意

- 感谢罗静初老师！
- 感谢11级生物信息学同学！
- 感谢全体组员！



主要内容

- 一、研究背景
- 二、基因结构与分析
- 三、蛋白质结构预测与分析
- 四、系统发育树构建



一、研究背景

1、水稻

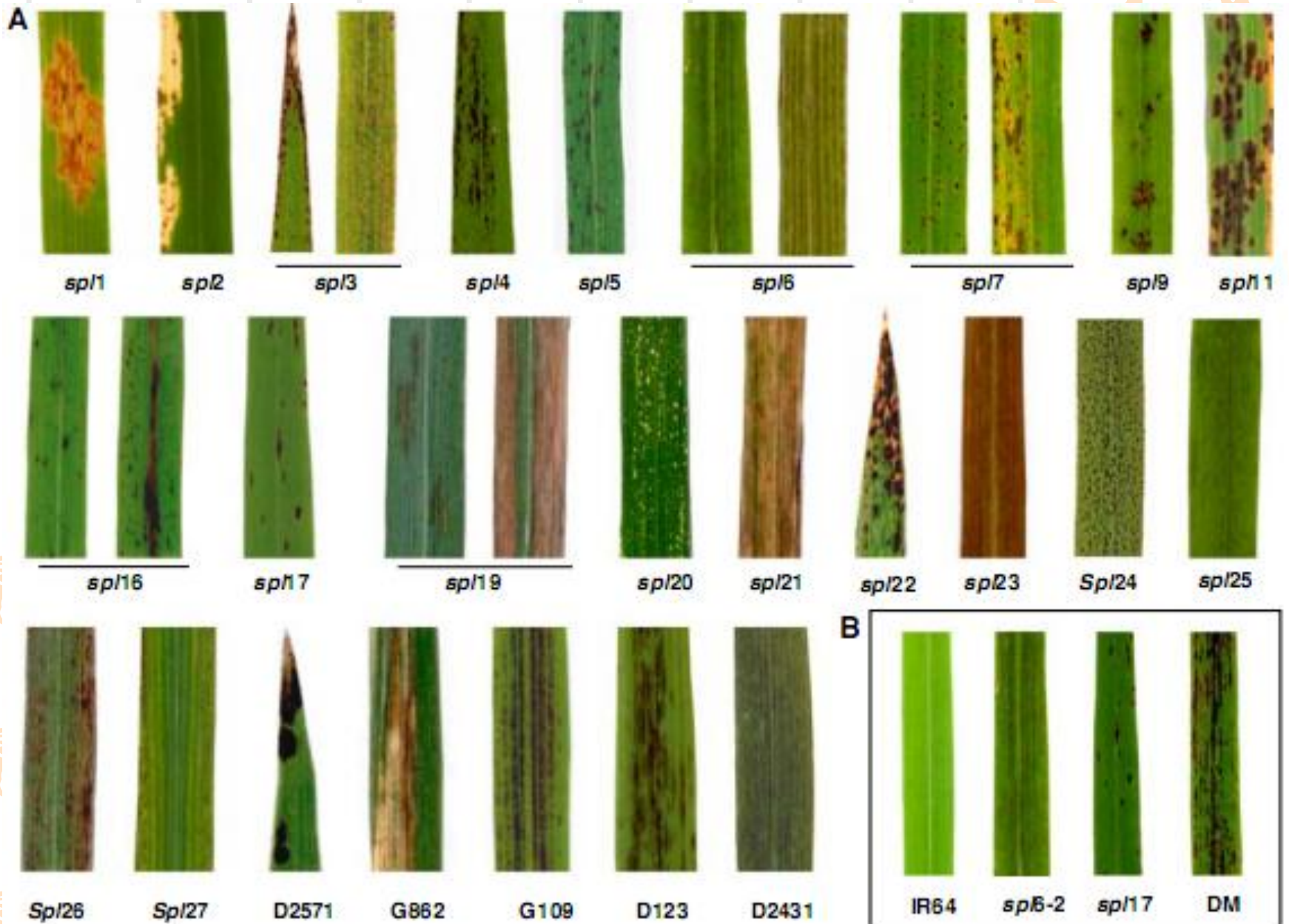
水稻是单子叶、一年生禾本科植物，是我国主要的粮食作物，自从2002年测序之后被作为模式生物。



2、水稻斑点叶

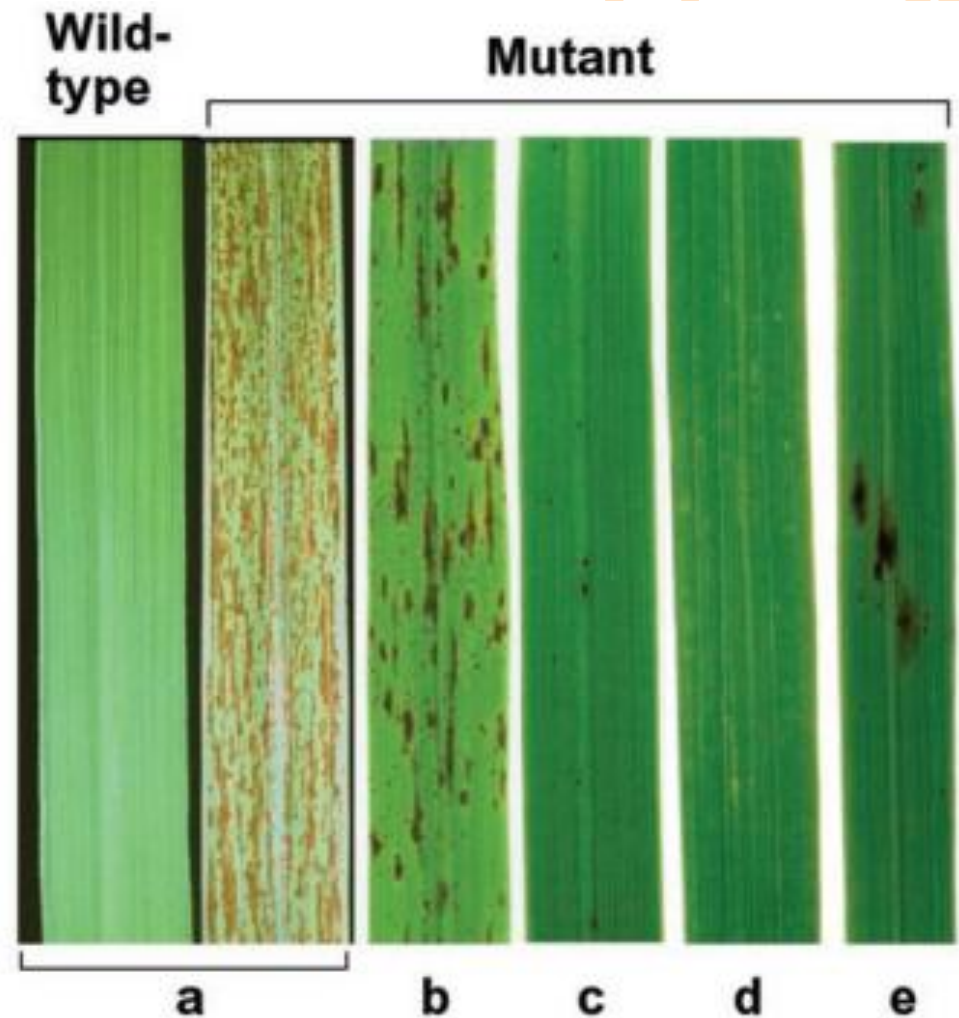
水稻斑点叶（spotted-leaf）的斑点是在**没有**明显逆境，机械和农药损伤或者病原菌侵染的条件下，由植物体**自发**形成的，大多数情况下与无毒病原菌侵染时产生的病斑相似。许多学者也称之为类病变或类病斑（lesion mimic, lesion simulating disease）。

水稻斑点叶（spotted-leaf）属于肉眼可见的叶片形态变异，发生变异的叶片上产生色泽、形状和大小不同的斑点，在有些突变体中斑点还会出现在叶鞘、枝梗和谷壳上。



3、Spl7突变体的表型分析

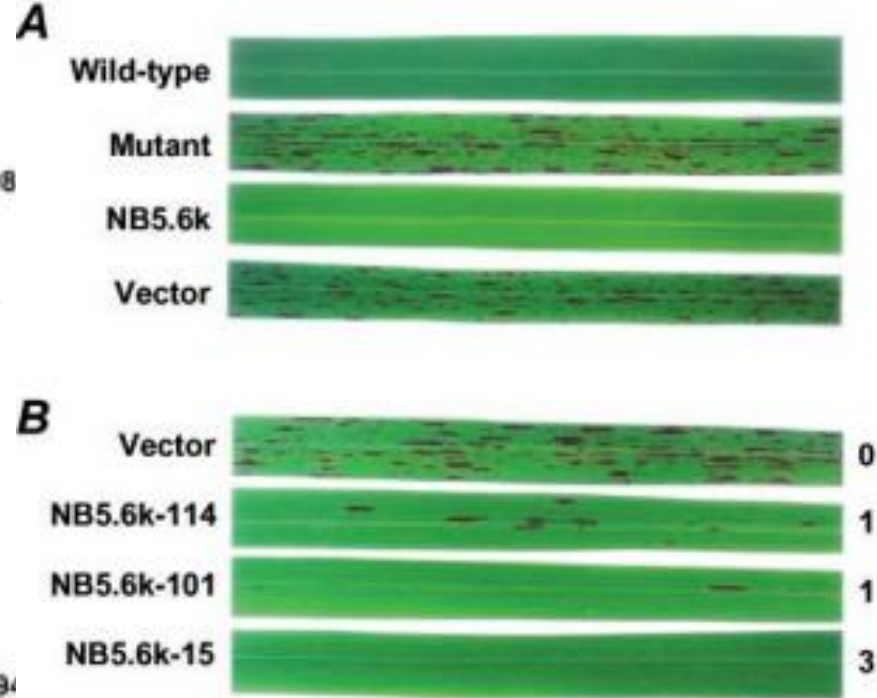
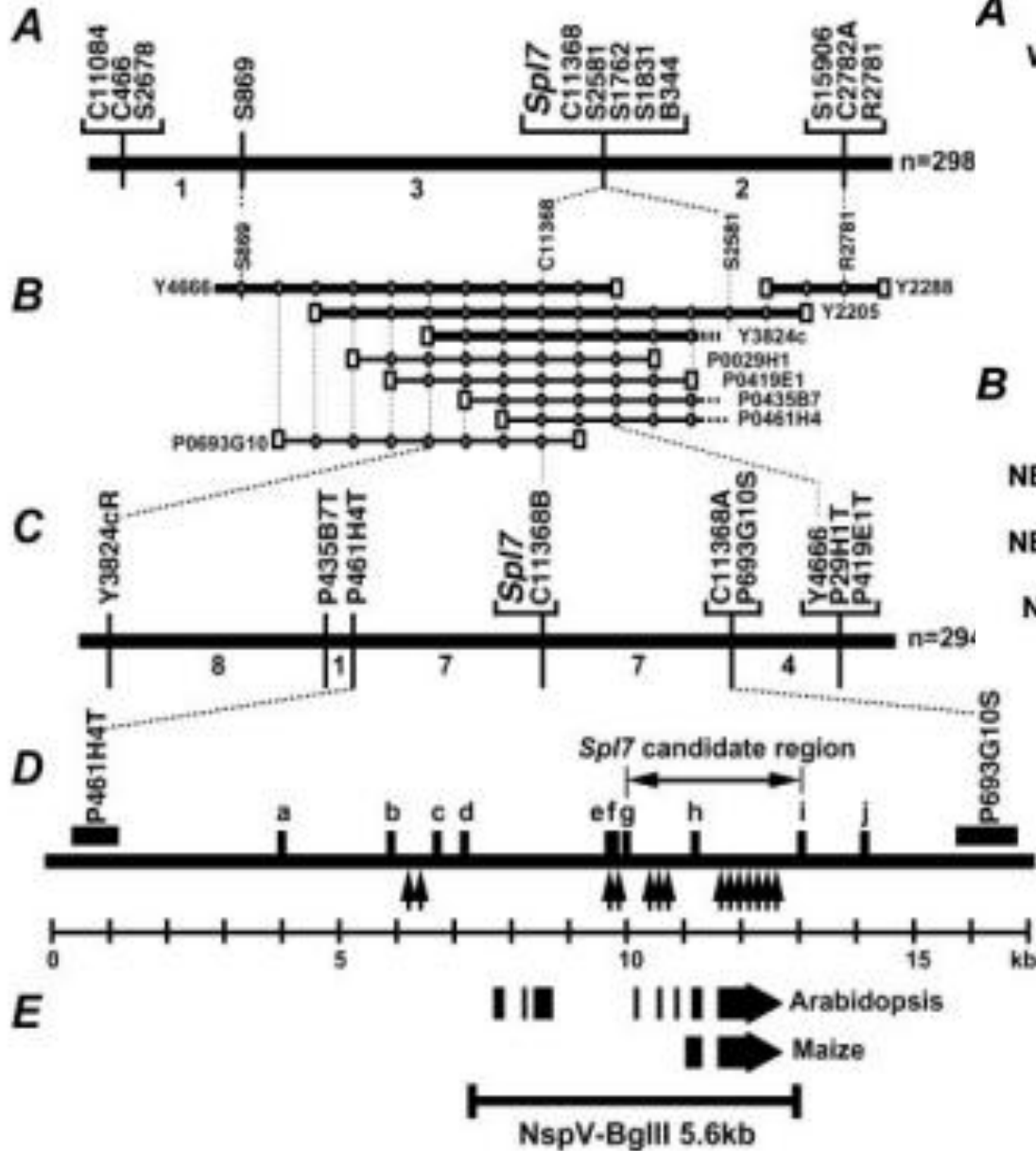
突变体*spl7* 从分蘖到抽穗，整个叶片表面分布着相对较小的红棕色小点；当降低的温度，或者遮去光照中的紫外线，叶片斑点症状会减轻。但是，无论是高温或紫外线处理，突变体在幼苗和嫩叶中没有发现任何损伤。



说明：

Spl7突变体是受
紫外线和**高温**诱导的。

4、Spl7的基因定位



通过互补性分析：
NspV-BglIII野生
型染色体片段
5.6kb包含整个
*spl7*基因序列和
完整的ORF区域。

二、基因结构与分析

1、基因注释信息

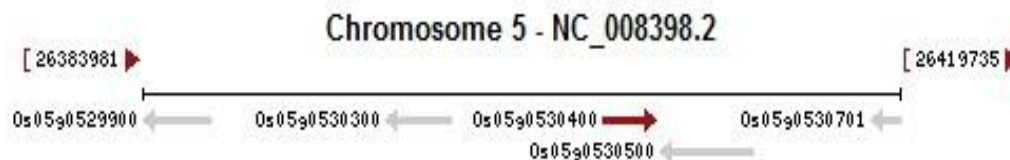
Summary

Gene symbol	Os05g0530400
Gene description	Os05g0530400
Gene type	protein coding
RefSeq status	PROVISIONAL
Organism	Oryza sativa Japonica Group (cultivar: Nipponbare)
Lineage	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP clade; Ehrhartoideae; Oryzeae; Oryza

Genomic context

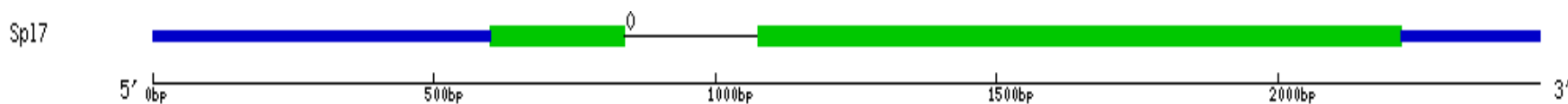
Sequence: Chromosome: 5; NC_008398.2 (26405672..26408134)

See Os05g0530400 in [MapViewer](#)



2、基因结构

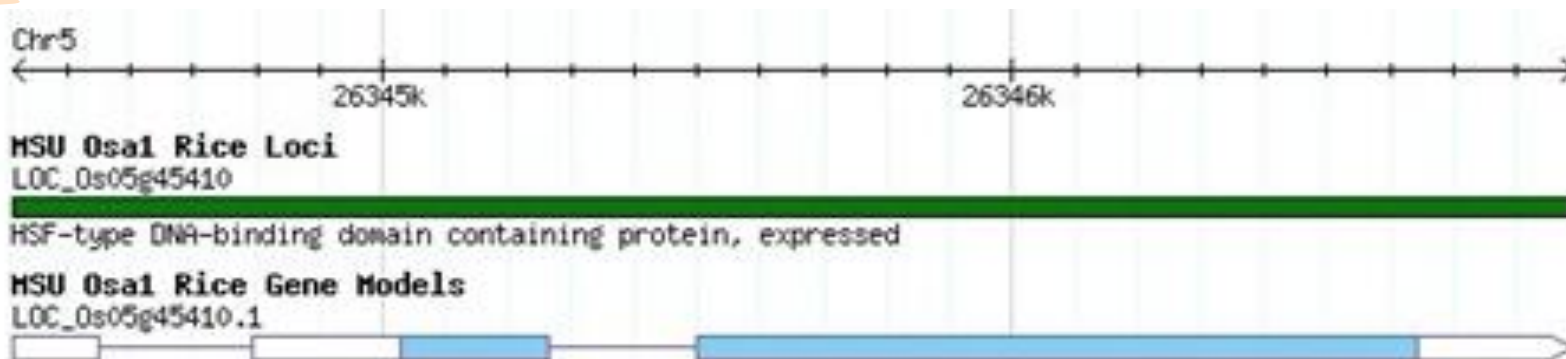
我们根据已知的gene序列和CDS序列，使用GSDS预测基因结构如下：



Legend:

■ exon ■ marked region — intron ■ UTR 0 1 2: intron phase

Rice Genome Annotation Project给出的结构图：



三、蛋白质结构预测与分析

1、蛋白质结构注释

Basic Information ? help		Back to Top			
Species	<i>Oryza sativa subsp. japonica</i>				
TF ID	Osj008119				
Family	HSF 属于: HSF家族				
Protein Properties	Length: 459aa MW: 51163.9 Da PI: 4.9482				
Description	HSF-type DNA-binding domain containing protein, expressed				
Gene Model	Gene Model ID	Type	Source	Coding Sequence	
	LOC_Os05g45410.1	genome	TIGR6.1	View CDS	
	NM_001062662	refseq	Refseq	View CDS	

Signature Domain ? help		Back to Top					
No.	Domain	Score	E-value	Start	End	HMM Start	HMM End
1	HSF_DNA-bind	110	1.7e-34	21	112	2	101
HSF_DNA-bind	2	HHHHHHHHCTGGGTTTSEES----EEEEES-HHHHHHHTHHHHSTT--HHHHHHHHHH---EE---S-B----TTSEEEES----- CS	98	FlkkllyeiledpelseliswsengnsfiildeeefakkvLpkyFkhsnfASFvRqLnmYgFkkvkdeekksksekiweFkhksFkkgdkellekikr			
Osj008119	21	FLIKTYEMVEDAATNHVVSWGPGGASFVWVWPLDFSRDLLPKYFKHNNFSSFIRQLNTYGFGRKIDPER-----WEFANEDFIRGHTLLKNIHR 109	pp	g*****999.....*****			
HSF_DNA-bind	99	--- CS	101	kks			
Osj008119	110	985 PP	112	RKP			

功能：是转录调控因子，具有HSF结构域，能够特异性地结合热激蛋白因子，而且该结构域从21个到112个氨基酸残基，共101个氨基酸残基。

2、水稻中同源基因蛋白质序列比对

HFA2B_ORYSJ	SHGDGGGLEVVV	-GEDGAAAVAAGVAPRPM	EGLHDAGPPPFLTKTYDMVDDAGTDAAVSW	72
HFA2E_ORYSJ	GPSTG	-----	VANGQPPRPM DGLADGGPPPFLTKTYDMVDDPTTDAVWSW	58
HFA2D_ORYSJ	PPSSPEE	-----	GEAPRPM EGLHEVGPPPFLTKTFDLVADPATDGVVSW	57
HFA2C_ORYSJ	LESQQQQRQ	-----	EDGGAAPRPM EGLHEVGPPPFLTKTYDLVEDPATDGVVSW	62
HFA2A_ORYSJ	DFAAAAAAAAAA	EGEGSPSSWAVGVMDLPRPMEGLGEAGPPPFLCKTYEVVDDPGTDTVISW	73	
HFA6B_ORYSJ	KPEAGEGWGGGD	-----	LGVVPPPPRPM EGLGEAGPAPFVAKTYEMVADAATDAVWSW	113
HSFA1_ORYSJ	AVTTAVAP	-----	PPGAAVSNVATAPPPFLMKTYEMVDDPATDAVWSW	57
HFA4D_ORYSJ	-----	-----	LGGGGGGGGGGGPPPFLIKTYEMVEDAATNHVWSW	40
HFA4B_ORYSJ	-----	-----	GGGGGSLPPFLSKTYEMVDDPSTDVAVGW	32
HSFA5_ORYSJ	-----	-----	ARGGGAGGGGGPAPFLKTYEMVDDPSTDVAVWSW	41

. . .** : **:::* * . * : . : *

HFA2B_ORYSJ	SATSN	--SFVWDPHAFATVLLPRFFKHNNF	SSFVRQLNTYGFRKVPDRWEFANENFLR	130
HFA2E_ORYSJ	SATNN	--SFVWDPHLFGNVLLPRYFKHNNF	SSFVRQLNTYGFRKVPDKWEFANEGLR	116
HFA2D_ORYSJ	GRAGS	--SFVWDPHVF AAVFLPRFFKHNNF	SSFVRQLNTYGFRKIDPDRWEFANDGFLR	115
HFA2C_ORYSJ	SRAGN	--SFVWDPHVFADLLPRLFKHNNF	SSFVRQLNTYGFRKVPDRWEFANEGLR	120
HFA2A_ORYSJ	GFAGN	--SFVWDANAF AAVLLPRYFKHSNF	SSFVRQLNTYGFRKVPDRWEFANEGLR	131
HFA6B_ORYSJ	GPGSGAS	SFVWDPHALAAGVLPFRFFKHANF	SSFVRQLNTYGFRKVTPDRWEFANEFLA	173
HSFA1_ORYSJ	GPGNN	--SFVWNTPEFARDLLPKYFKHSNF	SSFVRQLNTYGFRKVPDRWEFANEGLR	115
HFA4D_ORYSJ	GPG-G	-ASFVWNPLDFSRDLLPKYFKHNNF	SSFIRQLNTYGFRKIDPERWEFANEDEFIR	98
HFA4B_ORYSJ	TPA-G	-TSFVVANQPEFCRDLLPKYFKHNNF	SSFVRQLNTYGFRKVDPEQWEFANEDEFIK	90
HSFA5_ORYSJ	SDASD	-ASFVWNHPEFAARLLPAYFKHSNF	SSFIRQLNTYGFRKIDPERWEFANEYFIK	100

. **** : : .** *** *****:*****: *:*****: *:

在HSF结构域区域的序列相似性较高，尤其是HFA4D与HFA4B，高度保守。

Identity: 194/481 (40.3%) Similarity: 274/481 (57.0%)



3、与拟南芥中同源基因蛋白质序列比对

```

HSFA2_ARATH      PFLTPTYEMVEDPATDTVVSWSNGRNSFVWVDSHKFSTLLPRYFKHSNFSSFIRQLNTY 103
HFA6B_ARATH      PFLTPTYDLVEDSRTNHVVSWSKSNNSFVWDPQAFSVTLLPRPFKHNNFSSFVRQLNTY 120
HFA1D_ARATH      PFLSKTYDMVDDHNTDSIVSWSANNSFVWKPPEFARDLLPKNFKHNNFSSFVRQLNTY 96
HFA1A_ARATH      PFLSKTYDMVEDPATDAIVSWSPTNNSFVWDPPEFSRDLLPKYFKHNNFSSFVRQLNTY 111
HFA1B_ARATH      PFLSKTYDMVDDPLTNEVVSWSGNSFVWVSAPEFSKVLLPKYFKHNNFSSFVRQLNTY 86
HFA1E_ARATH      PFLSKTYDMVDDPLTDDVVSWSGNSFVWVNPPEFAKQFLPKYFKHNNFSSFVRQLNTY 82
HSFA3_ARATH      PFLSKTFDLVDDPTLDPVISWGLTGASFVWVDPLEFARIILPRNFKHNNFSSFVRQLNTY 114
HFA4A_ARATH      PFLTPTYEMVDDSSSDSIVSWSQSNKSFVWNPPEFSRDLLPRPFKHNNFSSFIRQLNTY 74
HFA4C_ARATH      PFLTPTYEMVDDSSSDSVVAWSENNSFVKNPAEFSRDLLPRPFKHNFSSFIRQLNTY 72
HFA4D_ORYSJ      PFLIKTYEMVEDAATNHVVSWSGPGGASFVWVNPPLDFSRDLPKYFKHNNFSSFIRQLNTY 79
HSFA5_ARATH      PFLVPTYEMVDDSSSTDQIVSWSANNSFVWNHAEFSRLLPTYFKHNNFSSFIRQLNTY 82
*** **:::*:* : :::* **:* . *: **: ***.*****:*****

```



```

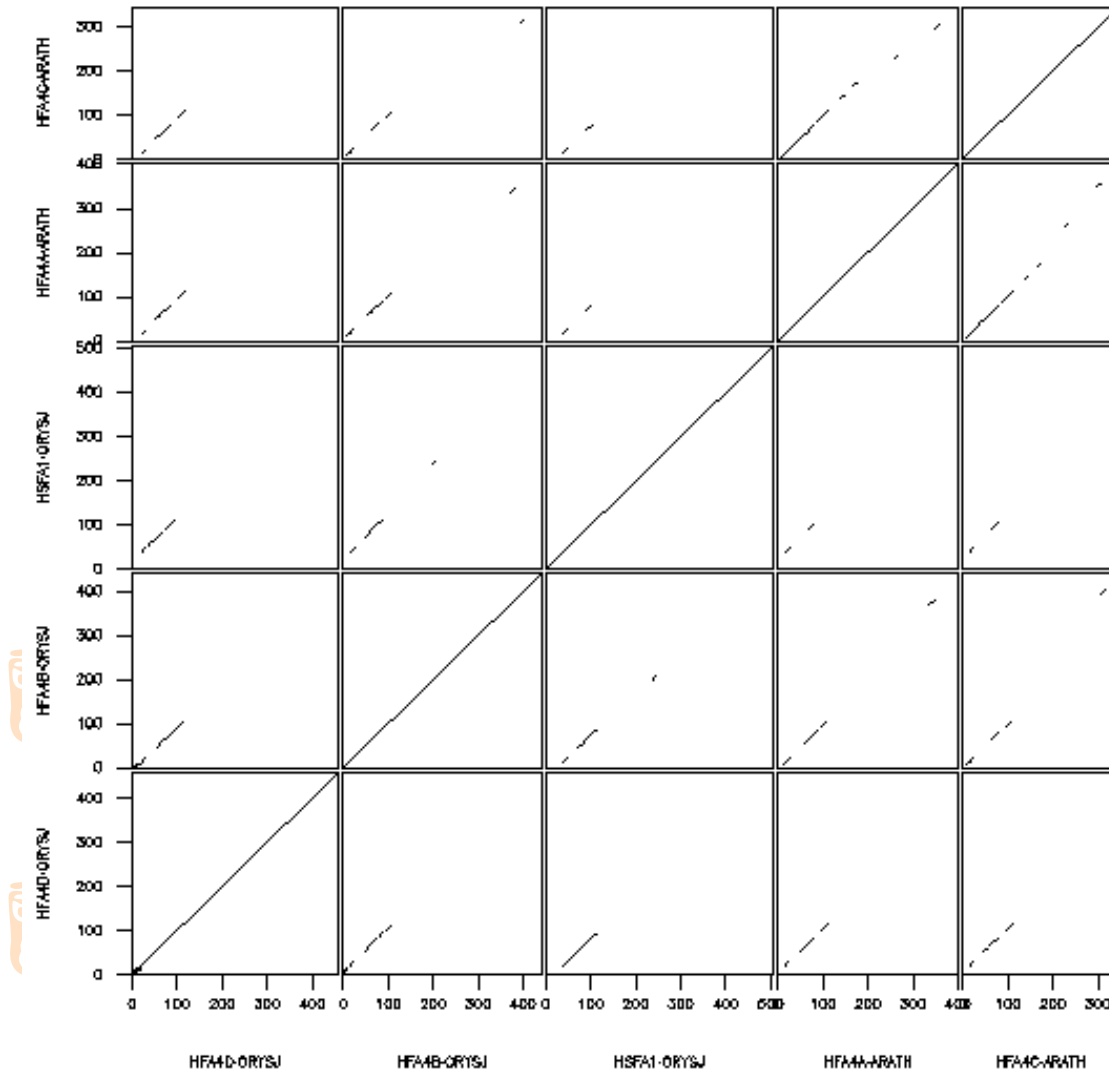
HSFA2_ARATH      GFRKIDPDRWEFANEGFLAGQKHLKNIKRRRNMG-----LQNVNQQ--GSGM 149
HFA6B_ARATH      GFRKVNPDPRWEFANEGFLRGQKHLKNIIRRRKTSNN-----SNQMQQPQSSEQQ--SLDN 173
HFA1D_ARATH      GFRKVPDPRWEFANEGFLRGQKHLQSI TRRKP AHGQGGQ---HQRSQHSNGQN--SSVS 151
HFA1A_ARATH      GFRKVPDPRWEFANEGFLRGQKHLKKISRRKSVQGHGSSSSNPQSQQLSQGQGSMAALS 171
HFA1B_ARATH      GFRKVPDPRWEFANEGFLRGRKQLLKSIVRRKPS-----HVQQNQQTQVQS--SSVG 137
HFA1E_ARATH      GFRKVPDPRWEFANEGFLRGQKQILKSIVRRKPA-----QVQP-PQQPQVQH--SSVG 132
HSFA3_ARATH      GFRKIDTDKWEFANEAFLRGKHLKNIHRRRSPQ-----SNQTCSSSTSQSQG 163
HFA4A_ARATH      GFRKADPEQWEFANDDFVRGQPHLMKNIHRRKPVHS-----HSLPNLQAQLN 121
HFA4C_ARATH      GFRKVDPEKWEFLNDDFVRGRPYLMKNIHRRKPVHS-----HSLVNLQAQ-N 118
HFA4D_ORYSJ      GFRKIDPERWEFANEDFIRGHTHLLKNIHRRKPVHS-----HSLQNQING-- 124
HSFA5_ARATH      GFRKIDPERWEFLNDDFIKDQKHLKNIHRRKPIHS-----HSHPPASST-- 127
**** ..:::** *:*: .: :::* **

```



与拟南芥同源基因比对发现，**HSF**结构域序列相似度较高，该区域高度保守。

4、5个同源基因比对点阵图



HFA4C_ARATH

HFA4A_ARATH

HSFA1_ORYSJ

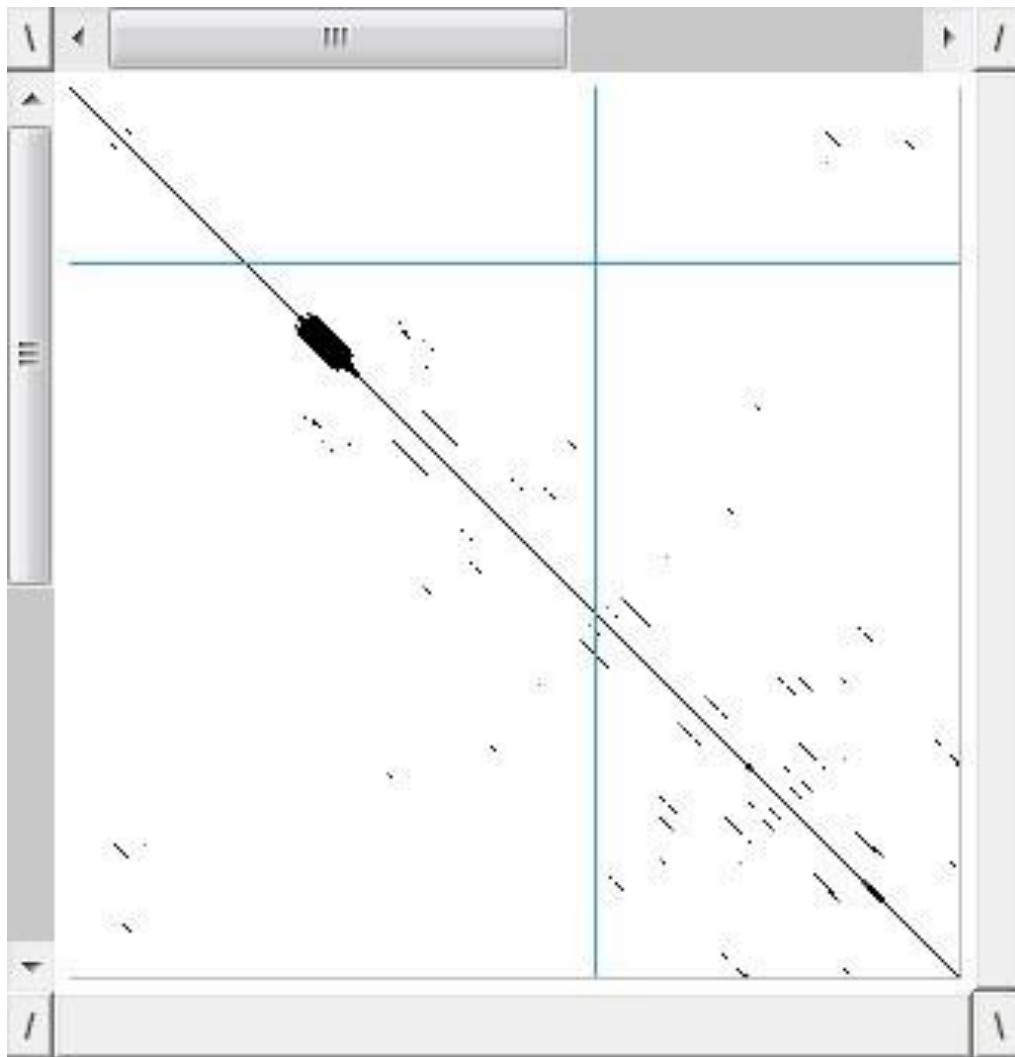
HFA4B_ORYSJ

HFA4D_ORYSJ

HSF结构域序列保守性较高

5、点阵预测重复序列

吉祥



吉祥

吉祥

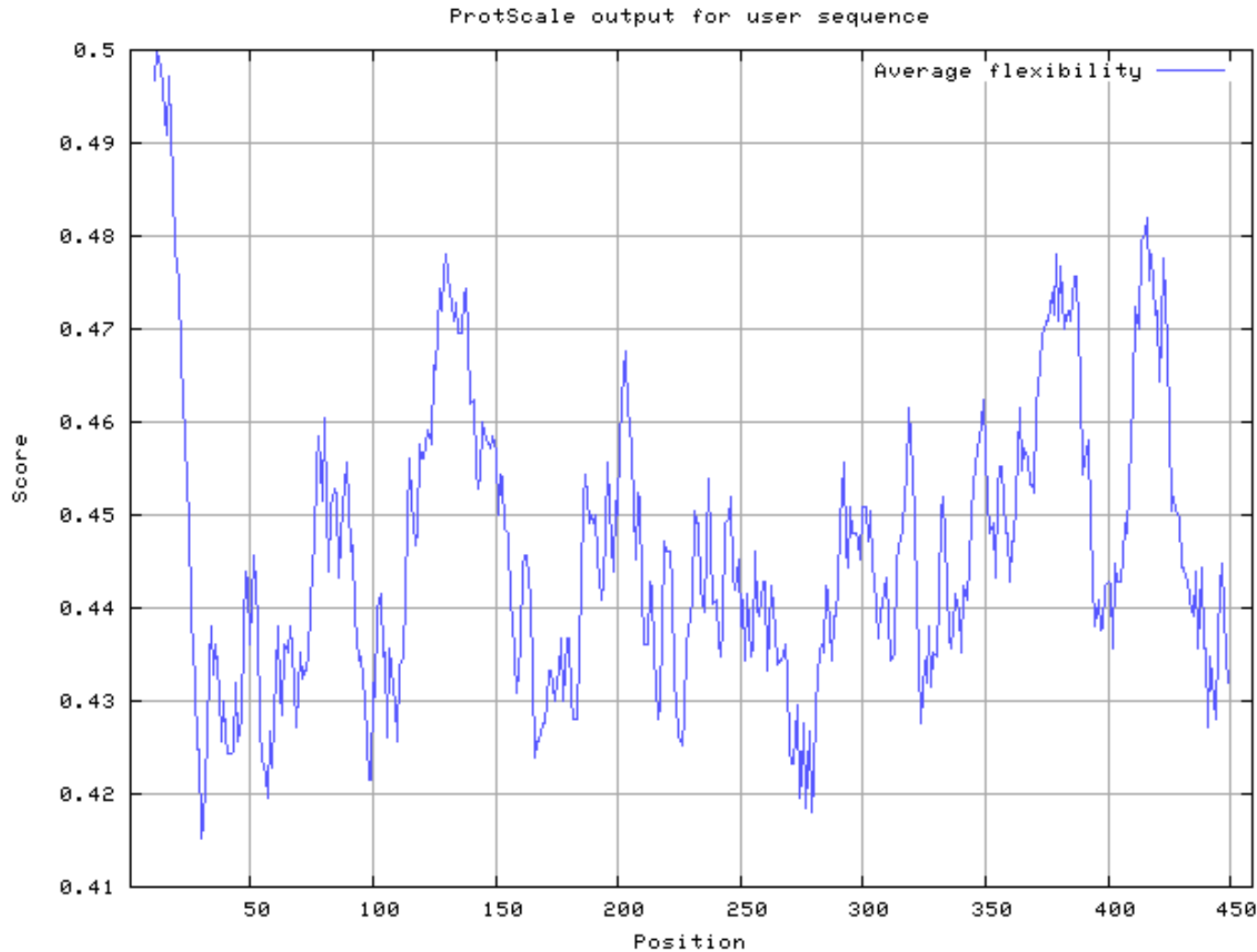
吉祥

吉祥

吉祥

吉祥

6、柔性分析



Amino acid scale values:

Ala: 0.360

Arg: 0.530

Asn: 0.460

Asp: 0.510

Cys: 0.350

Gln: 0.490

Glu: 0.500

Gly: 0.540

His: 0.320

Ile: 0.460

Leu: 0.370

Lys: 0.470

Met: 0.300

Phe: 0.310

Pro: 0.510

Ser: 0.510

Thr: 0.440

Trp: 0.310

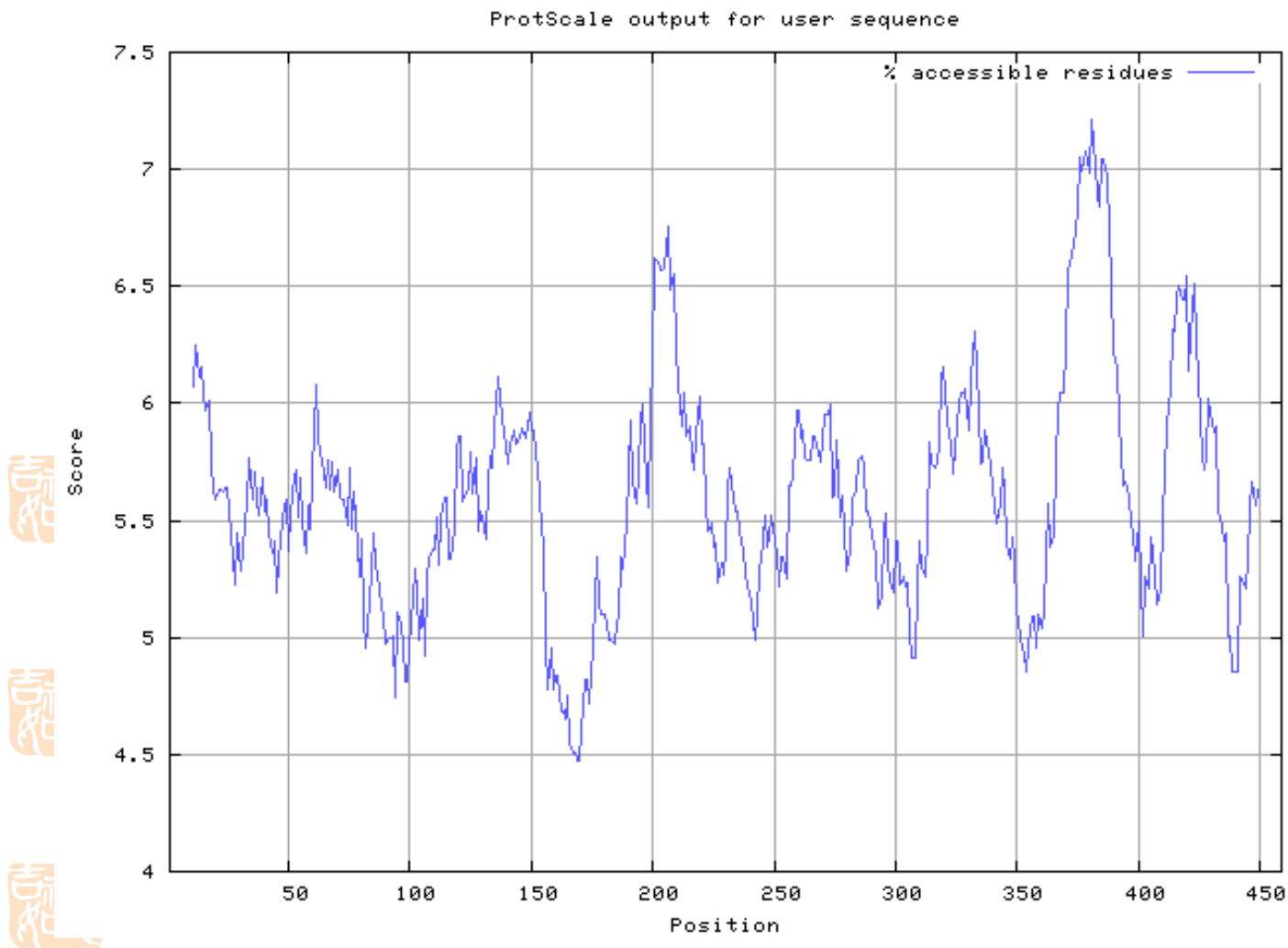
Tyr: 0.420

Val: 0.390

说明该蛋白质序列的柔性较高，蛋白质结构稳定性越好。

7、溶剂可及性分析

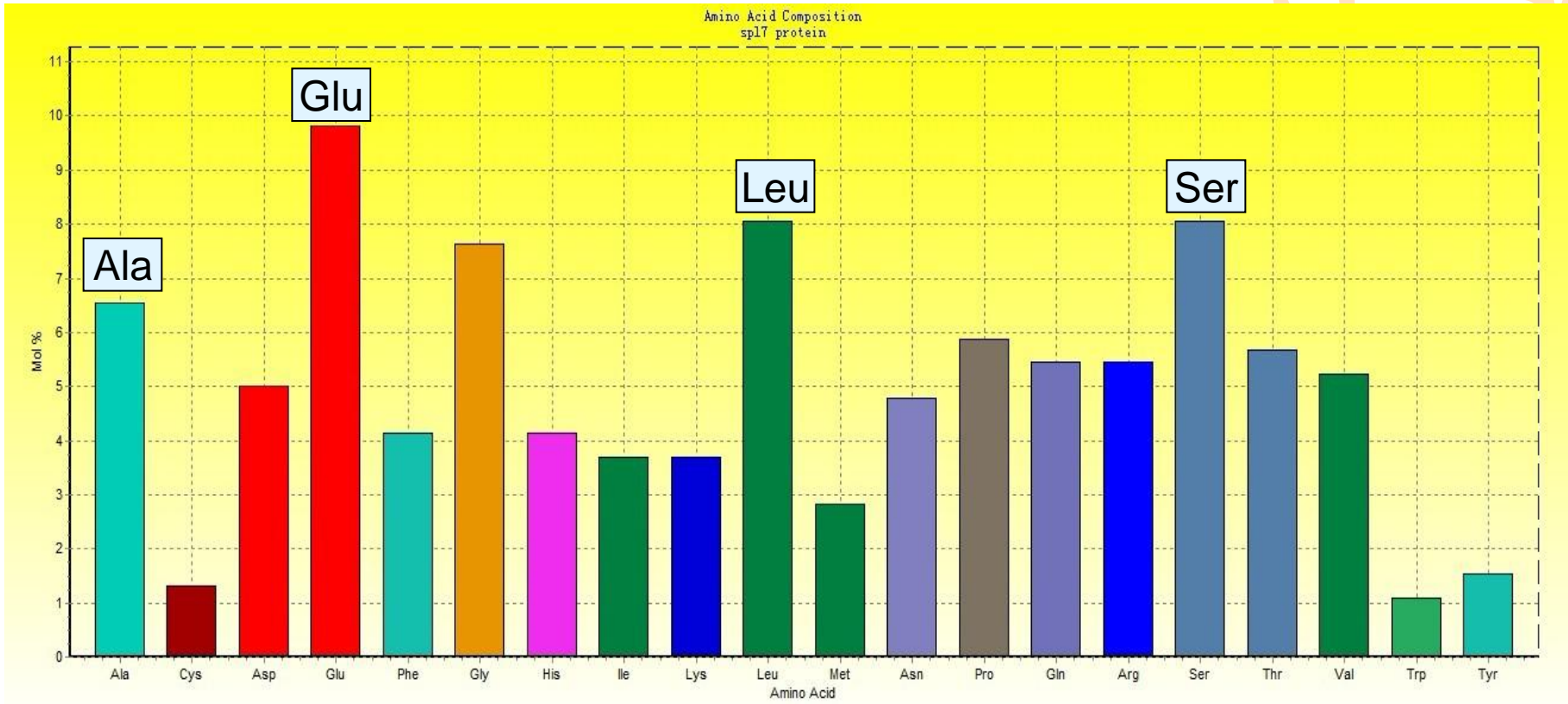
吉祥



说明溶剂可及性较高，可能有利于与DNA结合。

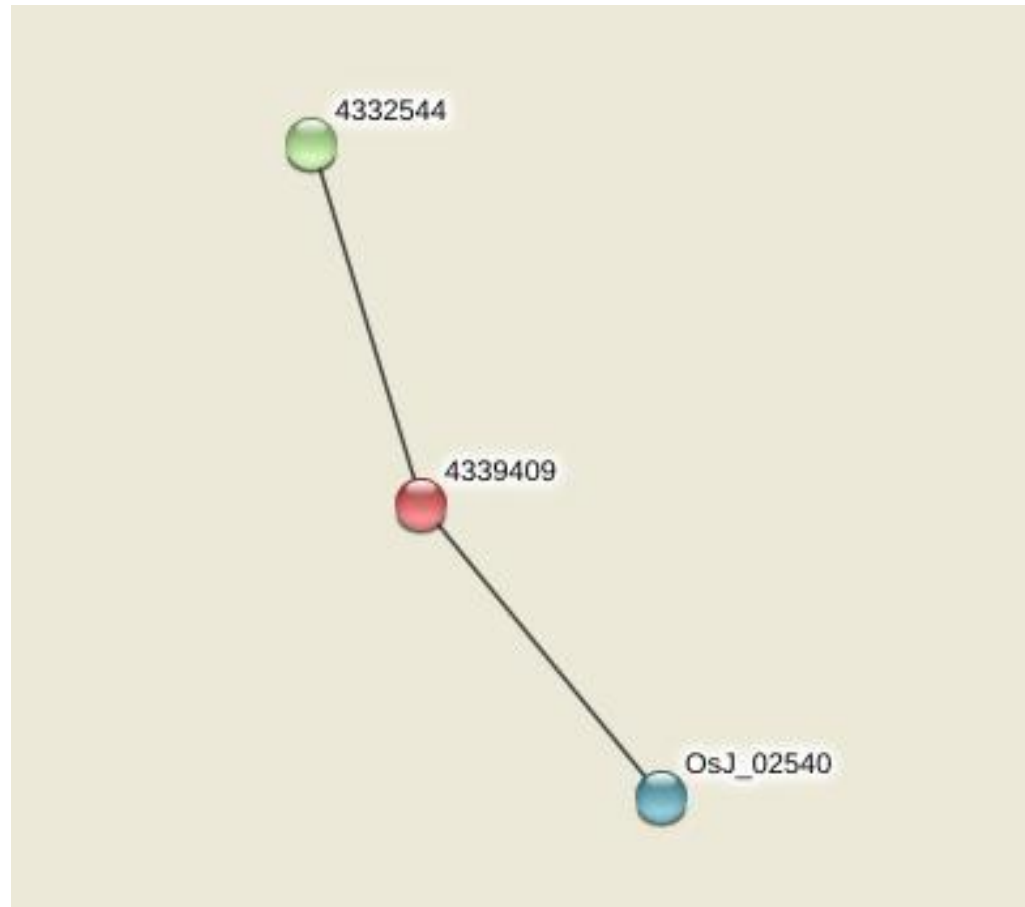
吉祥

8、蛋白质中各氨基酸的组成比例



Glu含量最高达到9.8%。

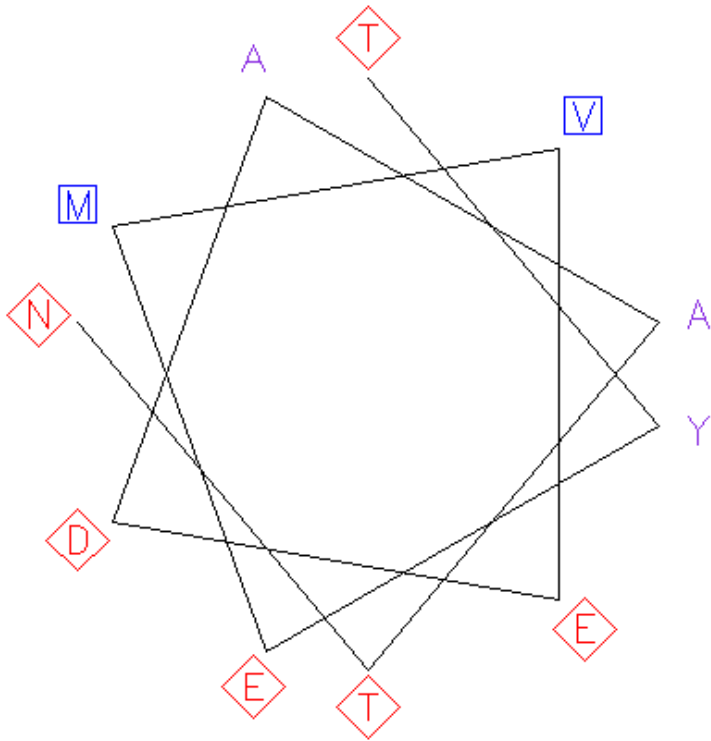
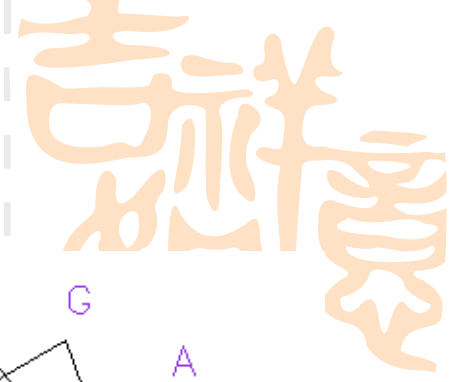
9、与HSF相互作用的蛋白质



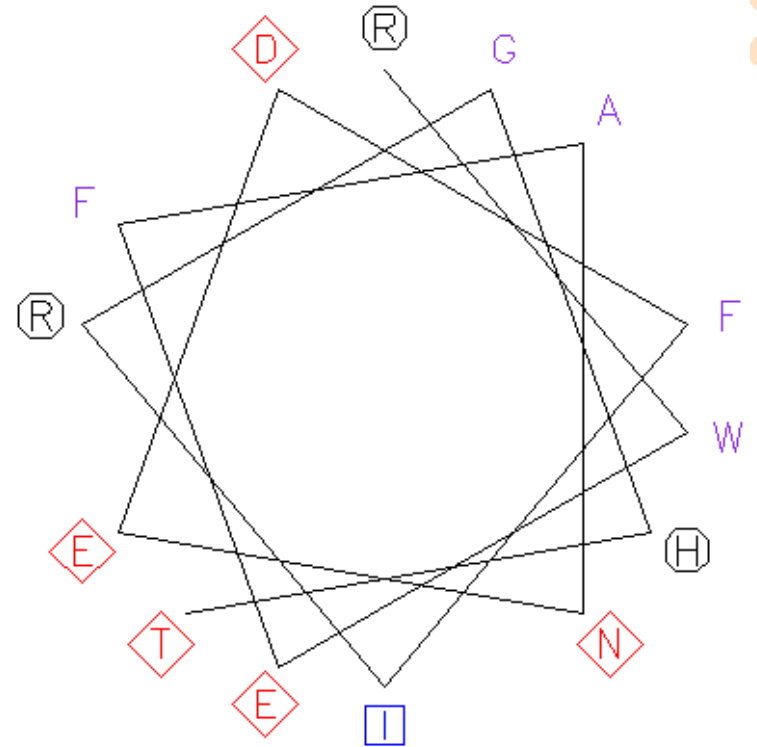
4332544: 是一个假定蛋白，含有DUF538（**Domain of unknown function**）结构域，具有177个氨基酸残基。

OsJ_02540: 是一个有332个氨基酸残基的表达蛋白。

10、Spl7二级结构预测



>alpha-1
TYEMVEDAATN



>alpha-2
RWEFANEDFIRGHT

Residue totals: H:190; E: 58; T: 83; C:128。

Percent: H: 42.9; E: 13.1; T: 18.7 ; C: 28.9。



11、同源模建



15-110氨基酸残基3-D结构图

吉祥

吉祥

吉祥

吉祥

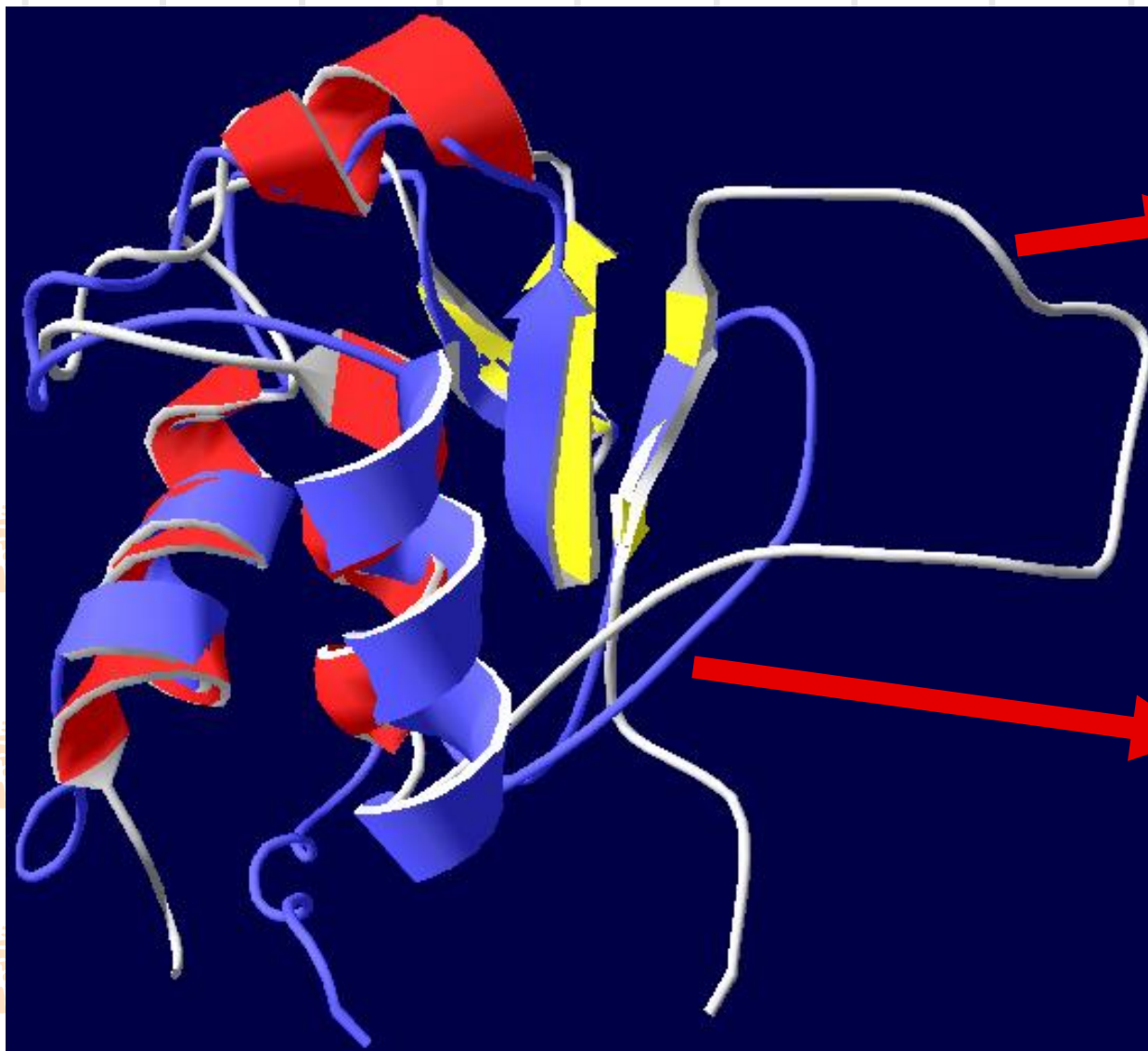
吉祥

吉祥

吉祥

水稻spl7的HSF结构域与克鲁维酵母菌属3HSF进行Magic Fit

王如洋



克鲁维酵母菌属
3HSF

水稻spl7的
HSF结构域

四、系统发育树构建

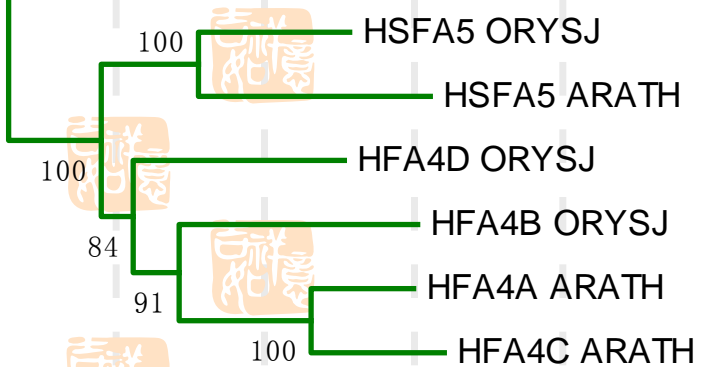
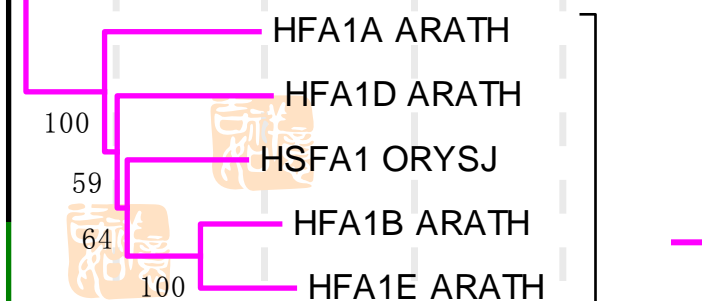
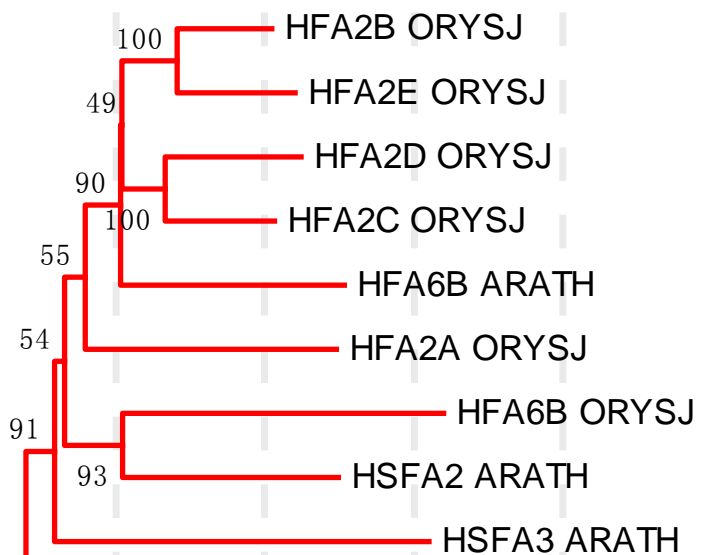
ARATH为拟南芥;

ORYSJ为*Oryza sativa* L. japonica

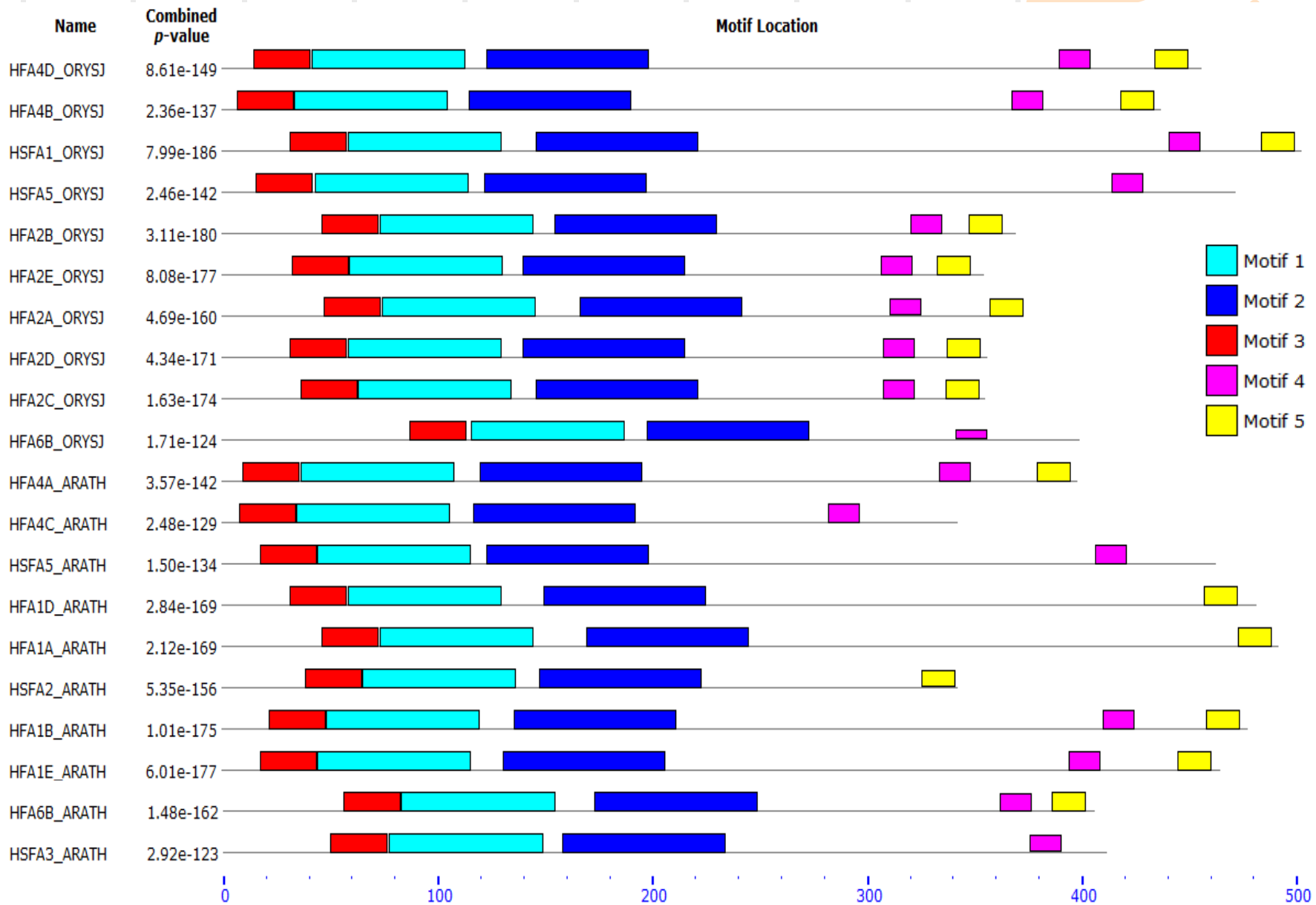
— 包括ORYSJ中的
HFA2A,2B,2C,2D,2E和HFA6B,
以及ARATH中的HSFA2和3

— 包括ORYSJ中的HSFA1和
ARATH中的HFA1A,1B,1D,1E

— 包括ORYSJ中的HFA4B,4D和
ARATH中的HFA4A,4C, 以及
ORYSY,ARATH中的HSFA5



0.1

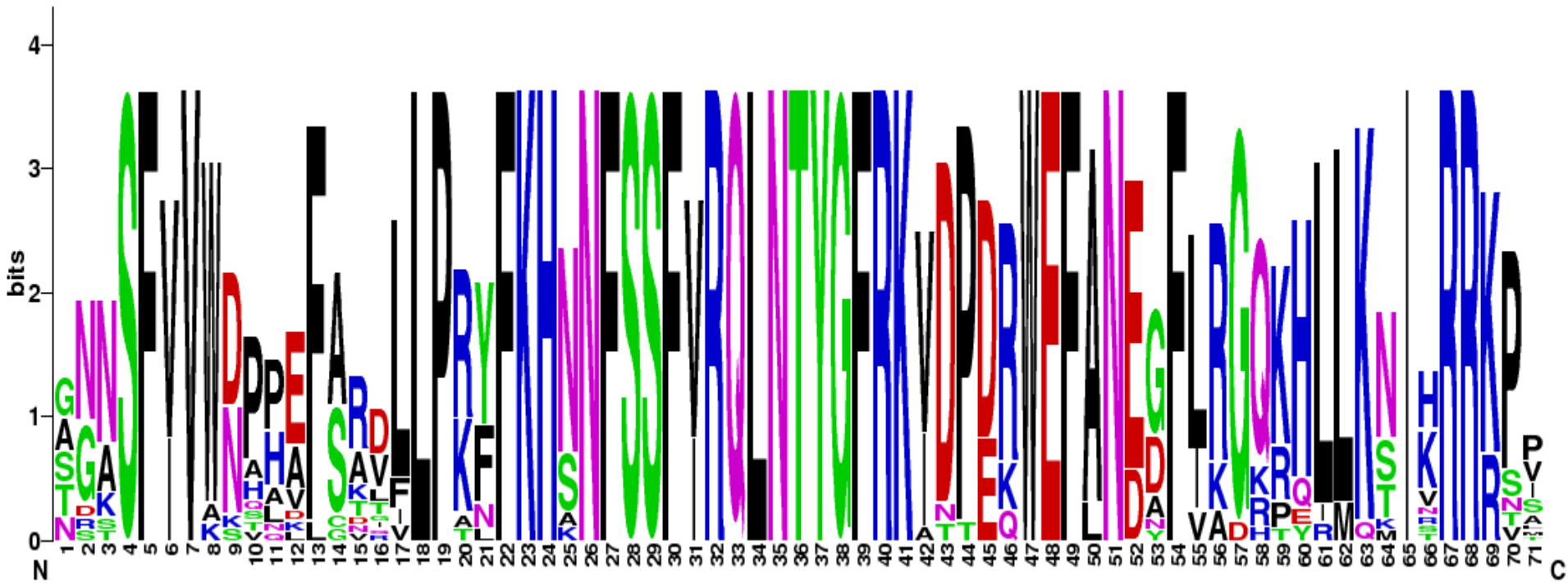


多重期望序列模体识别系统分析水稻*spl7*转录因子蛋白质序列保守结构域图形输出

水稻*spl7*基因同源序列DNA结合结构域序列和比对结果

>HFA1A_ARATH (start= 74)
TNNSFIVWDPPEFSRDLLPKYFKHNNFSSFVRQLNTYGFRKVDPDRWEFANEGFLRGQKHLKISRRKSV
>HSFA1_ORYSJ (start= 59)
GNNSFVWNTPEFARDLLPKYFKHSNFSSFVRQLNTYGFRKVDPDRWEFANEGFLRGQKHLKTIARRKPT
>HFA2B_ORYSJ (start= 74)
TSNSFVWDPHAFATVLLPRFFKHNNFSSFVRQLNTYGFRKVDPDRWEFANENFLRGQRHLLKNIKRRKPP
>HFA2E_ORYSJ (start= 60)
TNNSFVWDPHPLFGNVLLPRYFKHNNFSSFVRQLNTYGFRKVDPKWEFANEGFLRGQKHLKSIKRRKPP
>HFA2C_ORYSJ (start= 64)
AGNSFVWDPHVFAADLLPRLFKHNNFSSFVRQLNTYGFRKVDPDRWEFANEGFLRGQRHLLKTIKRRKPP
>HFA4D_ORYSJ (start= 42)
GGASFVWVWVPLDFSRDLLPKYFKHNNFSSFIRQLNTYGFRKIDPERWEFANEDFIRGHHTHLLKNIHRRKPV
>HFA1D_ARATH (start= 59)
NNNSFIVWKPPEFARDLLPKNFKHNNFSSFVRQLNTYGFRKVDPDRWEFANEGFLRGQKHLKQSIARRKPA
>HSFA5_ORYSJ (start= 44)
SDASFVWVWNHPEFAARLLPAYFKHSNFSSFIRQLNTYGFRKIDPERWEFANEYFIKGQKHLKNIHRRKPI
>HFA1B_ARATH (start= 49)
GNNSFVWVSAPEFSKVLLPKYFKHNNFSSFVRQLNTYGFRKVDPDRWEFANEGFLRGRKQLLKSIVRRKPS
>HFA2D_ORYSJ (start= 59)
AGSSFVWVWDPHVFAAVFLPRFFKHNNFSSFVRQLNTYGFRKIDPDRWEFANDGFLRGQRHLLKMIKRRRPL
>HFA1E_ARATH (start= 45)
GNNSFVWVNPPEFAKQFLPKYFKHNNFSSFVRQLNTYGFRKVDPDRWEFANEGFLRGQKQILKSIVRRKPA
>HFA6B_ARATH (start= 83)
SNNSFIVWDPQAFSVTLLPRFFKHNNFSSFVRQLNTYGFRKVNPDREWEFANEGFLRGQKHLKNIARRKTS
>HFA4A_ARATH (start= 37)
SNKSFIVWVNPPEFSRDLLPRFFKHNNFSSFIRQLNTYGFRKADPEQWEFANDDFVRGQPHLMKNIHRRKPV
>HFA2A_ORYSJ (start= 75)
AGNSFVWVDANAFAAVLLPRYFKHSNFSSFVRQLNTYGFRKVDPDRWEFANEGFLRGKKEKLLTIKRRRPP
>HSFA5_ARATH (start= 45)
NNNSFIVWVNHAEFSRLLPTFYFKHNNFSSFIRQLNTYGFRKIDPERWEFLNDDFIKDQKHLKNIHRRKPI
>HSFA2_ARATH (start= 66)
GRNSFVWVWDSHKFSTLLPRYFKHSNFSSFIRQLNTYGFRKIDPDRWEFANEGFLAGQKHLKNIKRRRNM
>HSFA3_ARATH (start= 77)
TGASFVWVWDPLEFARIILPRNFKHNNFSSFVRQLNTYGFRKIDTDKWEFANEAFLRGKKHLKNIHRRRSP
>HFA4B_ORYSJ (start= 34)
AGTSFVVANQPEFCRDLLPKYFKHNNFSSFVRQLNTYGFRKVDPEQWEFANEDFIKGQRHRLKNIHRRKPI
>HFA6B_ORYSJ (start= 117)
SGASFVWVWDPHALAAGVLPFRFFKHANFSSFVRQLNTYGFRKVTDPDRWEFANEAFLAGQKHLKNIKRRRVS
>HFA4C_ARATH (start= 35)
NNKSFIVKNPAEFSRDLLPRFFKHKNFSSFIRQLNTYGFRKVDPEKWEFLNDDFVRGRPYLMKNIHRRKPV



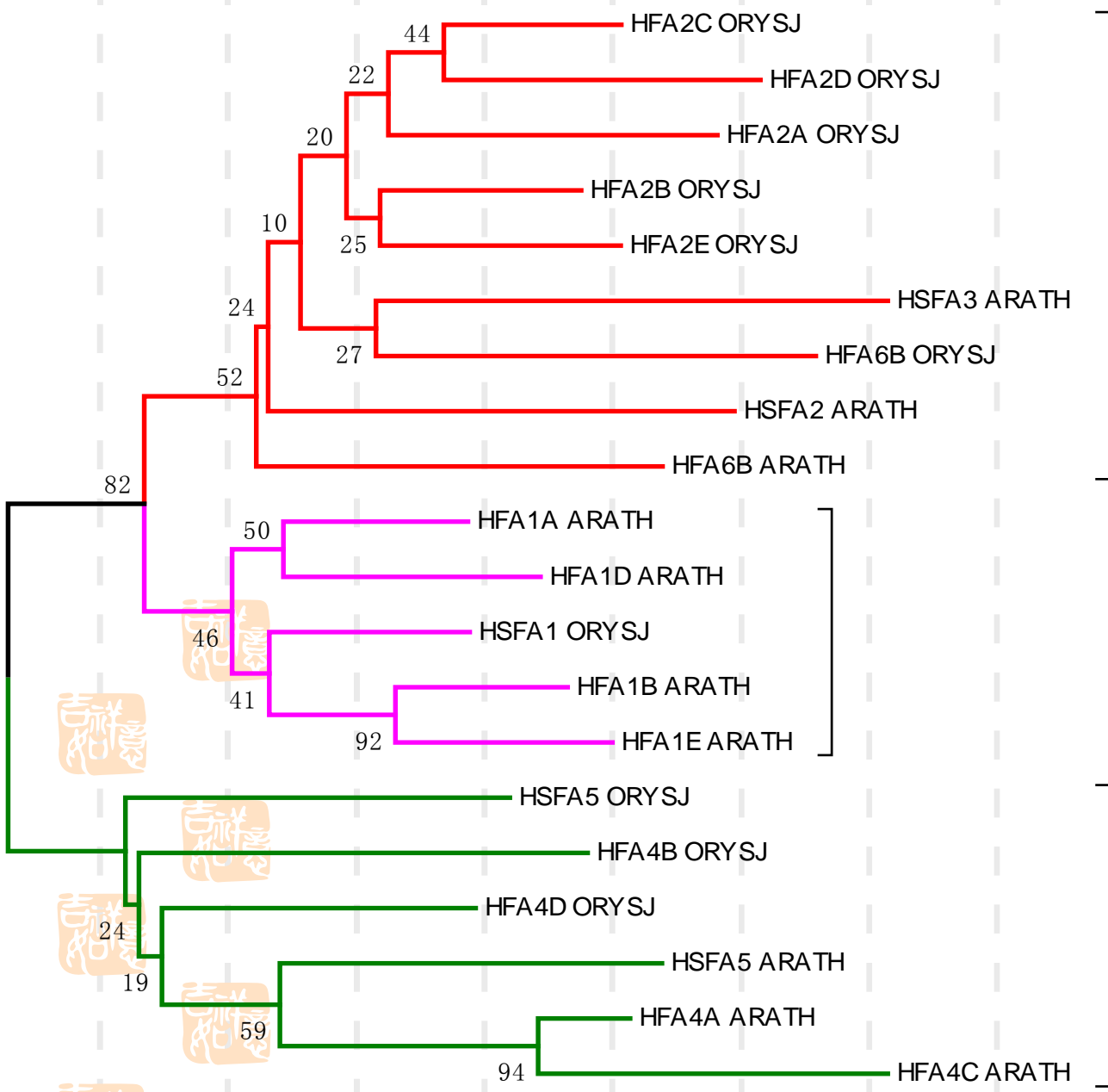


与水稻*spl7*基因同源的20个DNA结合结构域序列图标



吉祥如意

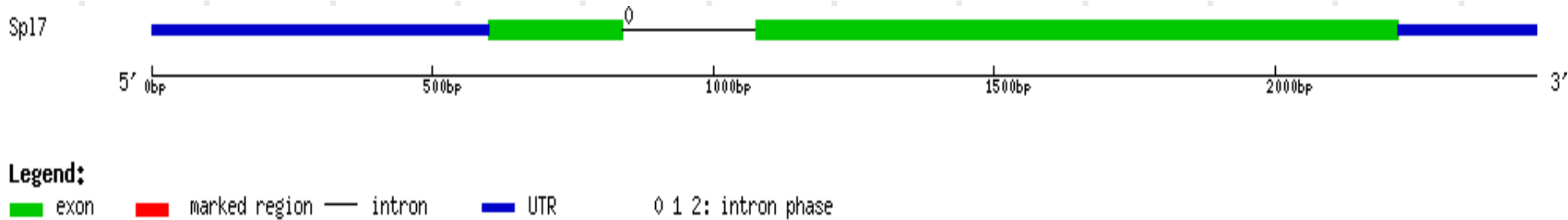
分类的趋势
与基于氨基酸
序列的相同，
相比之下，
只是每枝的
支持率降低



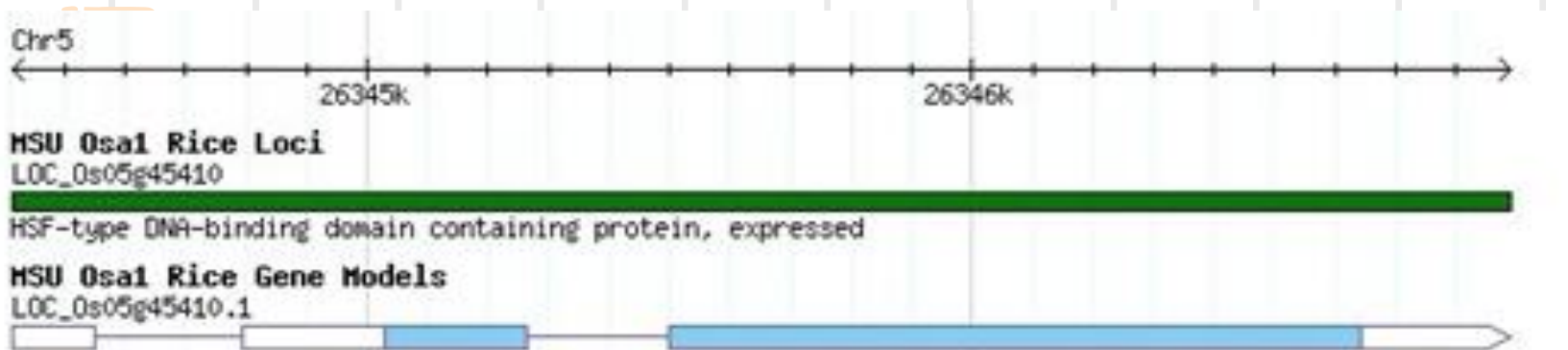
20个水稻spl7基因HSF结构域系统发育树

问题???

我们根据已知的gene序列和CDS序列，使用GSDS预测基因结构如下：



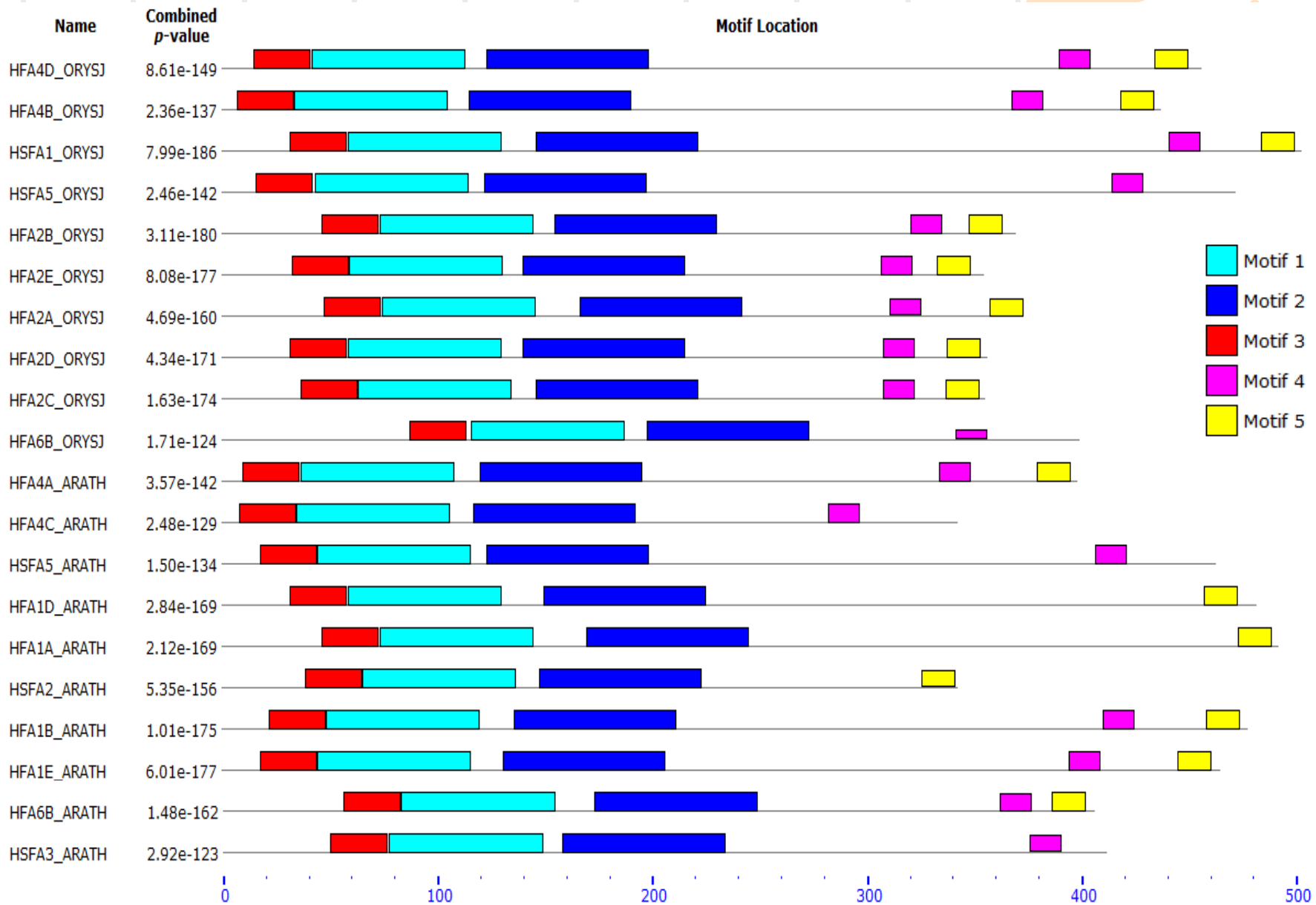
Rice Genome Annotation Project给出的结构图：




问题???

- Sequence file: **pasted_sequences**
- Distribution of motif occurrences: **Zero or one per sequence**
- Number of different motifs: **5** 问题???
- Minimum motif width: **10**
- Maximum motif width: **80**
- Statistics on your dataset:

type of sequence	protein
number of sequences	20
shortest sequence (residues)	345
longest sequence (residues)	506
average sequence length (residues)	420.4
total dataset size (residues)	8408




 多重期望序列模体识别系统分析水稻*spl7*转录因子蛋白质序列保守结构域图形输出

谢谢大家!

吉祥如意

圣诞节快乐!

