

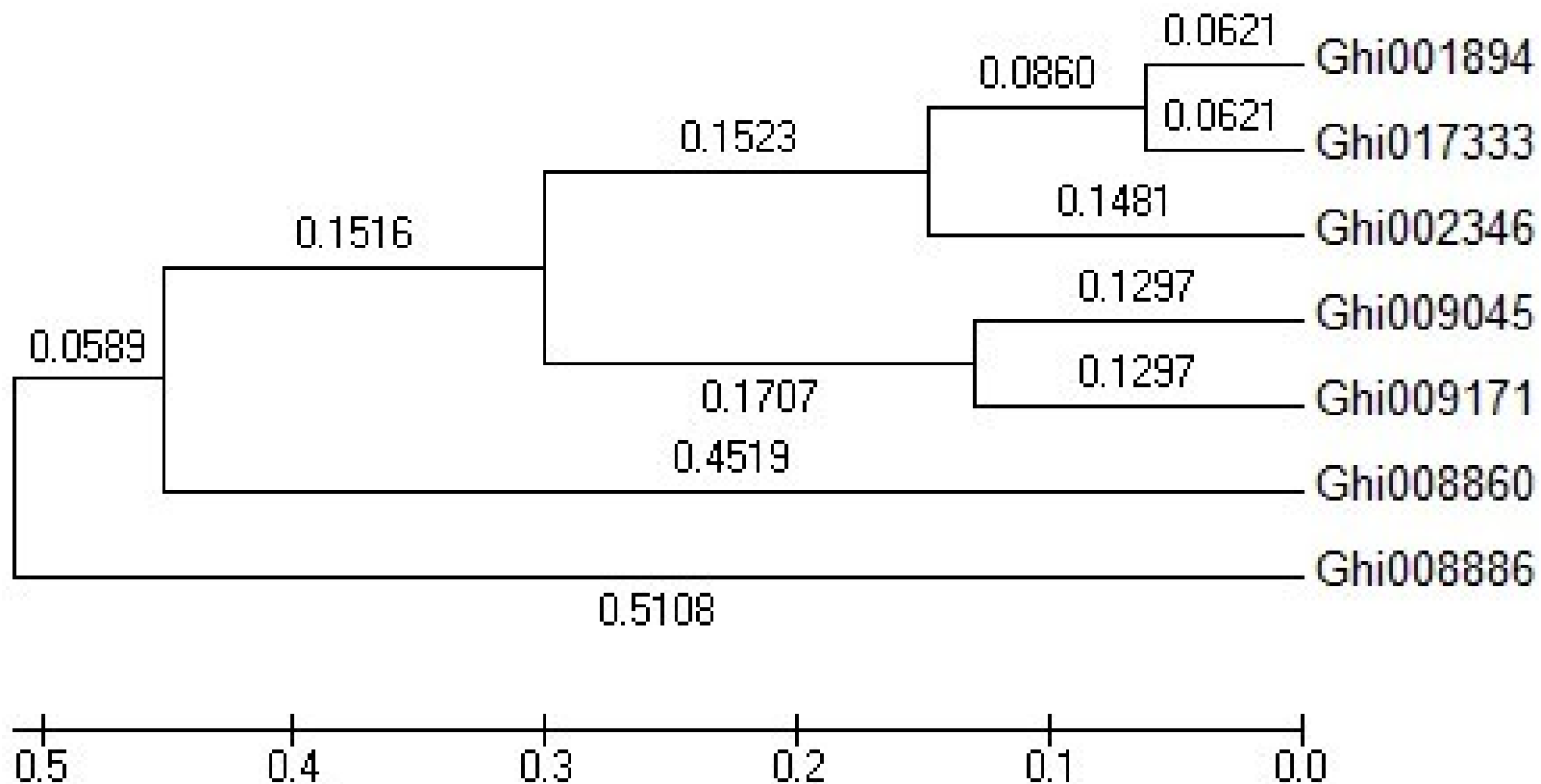
棉花ARF转录因子家族 生物信息学分析

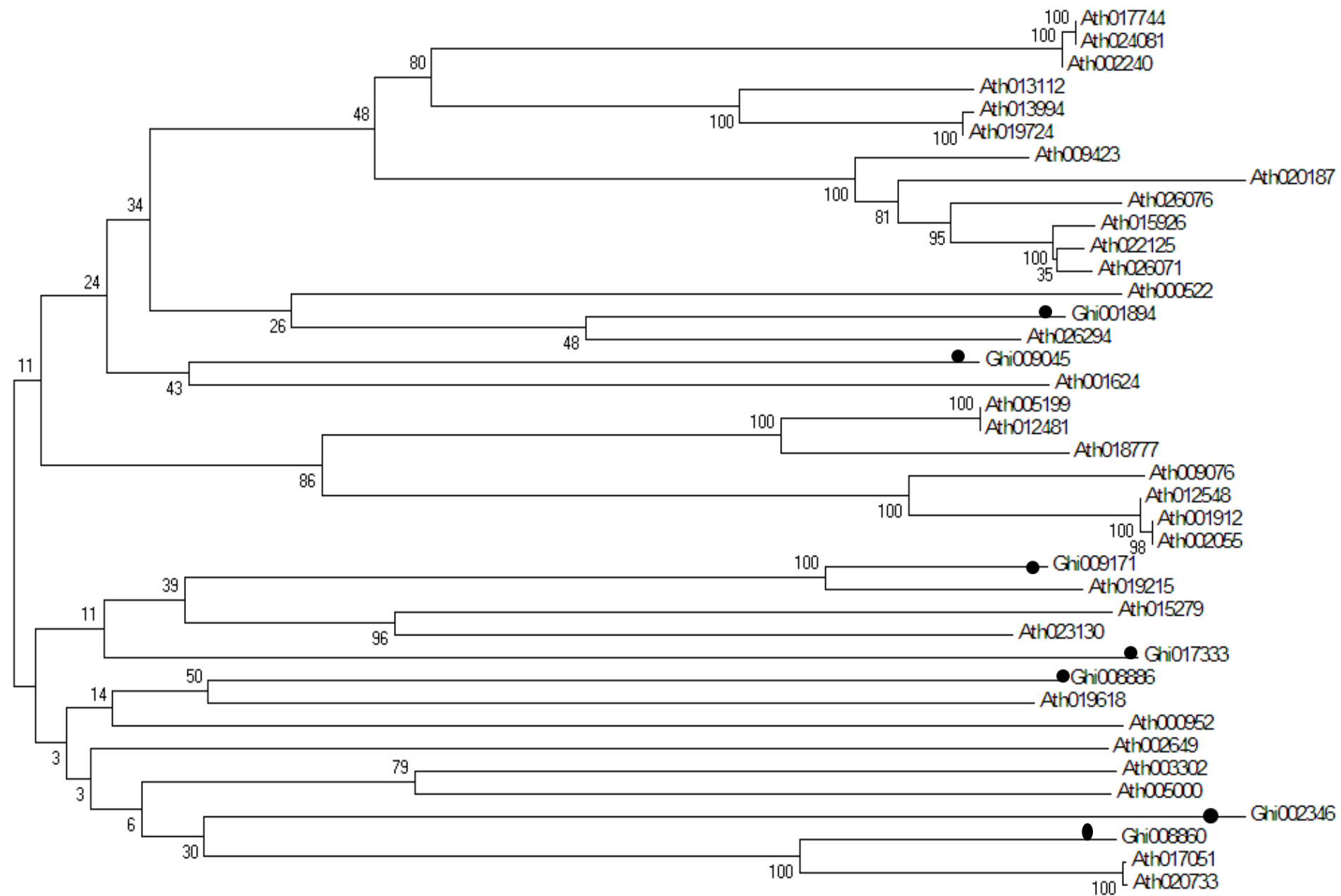
组成员 崔兴雷 刘国元
李黎贝 赵艳艳

报告人 崔兴雷

- 生长素响应因子（**ARF**）是一类调控生长素响应基因表达的转录因子，在生长素的信号传导过程中处于中心位置，它可与生长素响应元件特异结合，促进或抑制基因的表达。
- 生长素响应因子（**ARF**）基因家族在棉花中有**7**个成员,在拟南芥中有**32**个成员。

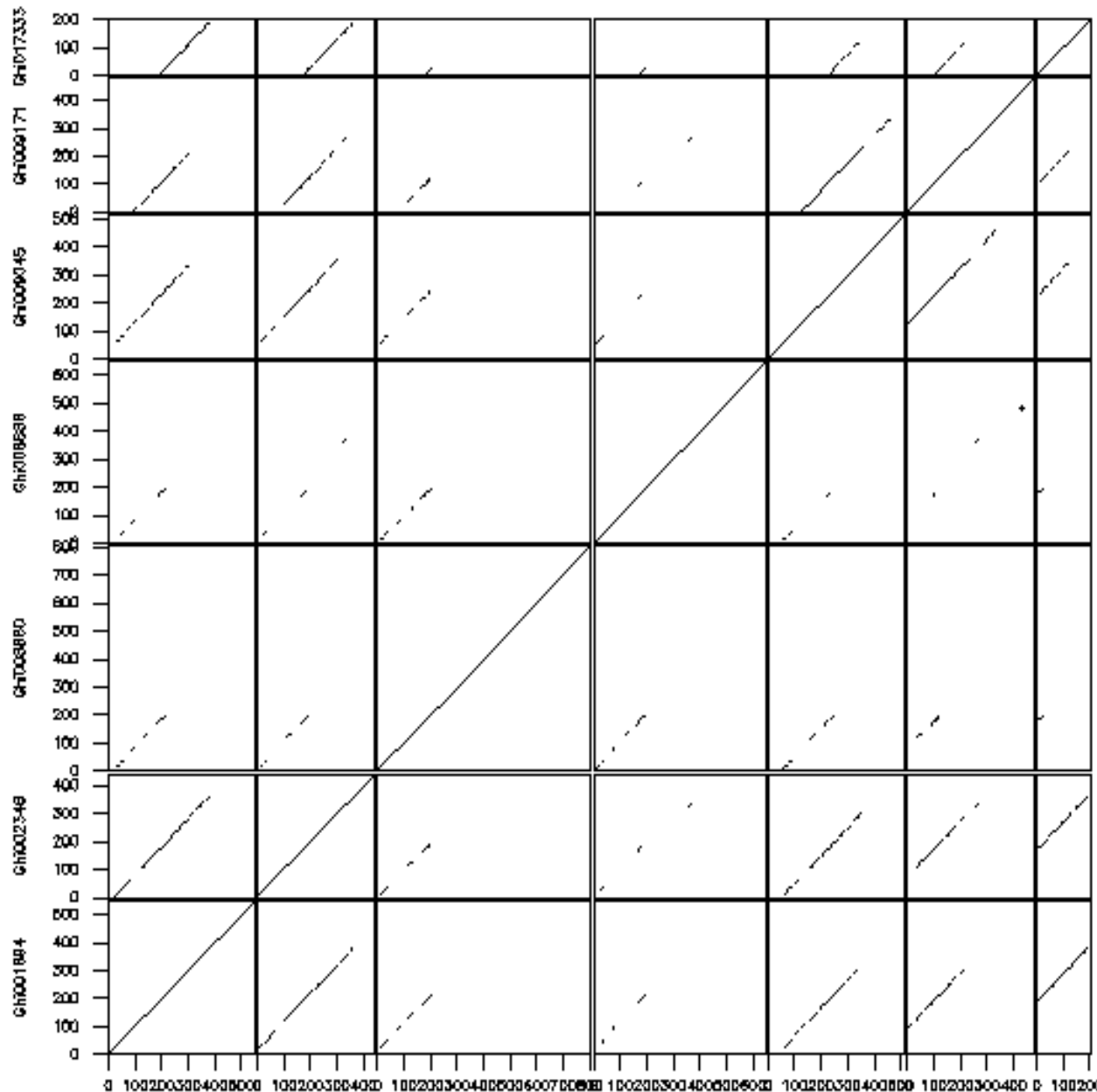
棉花ARF转录因子家族蛋白质进化树





Poly dotplot of 635508

Wed 19 Jun 2013 15:45:23



No.	Length	Lines	Points	Sequence
1	601	62	1247	GHI001894
2	442	63	1083	GHI002346
3	804	34	1062	GHI008860
4	647	22	798	GHI008886
5	515	58	1143	GHI009045
6	477	46	1036	GHI009171
7	206	33	626	GHI017333

- >Ghi001894
- >Ghi002346
- >Ghi008860
- >Ghi008886
- >Ghi009045
- >Ghi009171
- >Ghi017333

1	-----	0	Ghi017333
1	MANRVESFSQTNVVSSEGNGCDDLMEWKLKAGPLVEVPRAEQRVYYFPQGHMEQLEAS	60	Ghi001894
1	-----	0	Ghi017333
61	TNQTILNQRIPLFNLPSKILCRVVHIQLLAEKETDEVYAQVTLLEPEPSQPEPTTPDLCPE	120	Ghi001894
1	-----	0	Ghi017333
121	SQRPTVHSFCKVLTASDTSTHGGFSVLRKHATDCLPKLNMNDATPTQELVAKDLHGWEWR	180	Ghi001894
1	-----	0	Ghi017333
181	-----MRHLLTTGWSTFVTSKRLVAGDSFVFLRGENGELRVGVRRVAHQSSMPTS FKHIFRGQPRRHLLTTGWSTFVTSKRLVAGDSFVFLRGENGELRVGVRRVVRQHSIMPSS *****. : * : *	51 240	Ghi017333 Ghi001894
52	VISSQSMHLGVLATASHAVSTQTLFVVYYKPRTSQFIIGLNKYLEALS NKFTVGMRFKMR	111	Ghi017333
241	VISSQSMHVGVLATASHAVSTQTLFVVYYKPRTSQFIIGLNKYLEALS NKFAVGMRFKMK ***** : ***** : ***** :	300	Ghi001894
112	FEGEDSPERRFSGTIVGVEAFSPYWKDSKWRSLKVQWDEPASIPRPDRVSPWEIEPFAAP	171	Ghi017333
301	FEGEDSPERRFSGTIVGVEDFSPLWKDSKWRSLKVQWDEPASIPRPDRVSPWEIEPFAAP ***** ** *	360	Ghi001894
172	IPPTLGQPVAANKRPRRTNAEIPALNYLQRISPGI-----	206	Ghi017333
361	IPPSLTQPVAANKRPRPPAEIPATDLSSTTSAPYSGVTHCHDLTPQNIAAEAKGNENPV *** : * ***** ***** : . *	420	Ghi001894
207	-----	206	Ghi017333
421	MGHHMWTEMSGCSVVKTLNKGSWLSSPGTCVPHHLVPEAREDKTRFSAWPVPSGFSNP	480	Ghi001894
207	-----	206	Ghi017333
481	QLQKIQPLDFTEKAKRSETSANCRLSGRTDKPFCSSNPLERTPGQLSTMKMGKTEVPGPS	540	Ghi001894
207	-----	206	Ghi017333
541	NLSSLNSDQKS	551	Ghi001894

利用MEME网站分析7个蛋白质序列的保守结构域

ALL MOTIFS

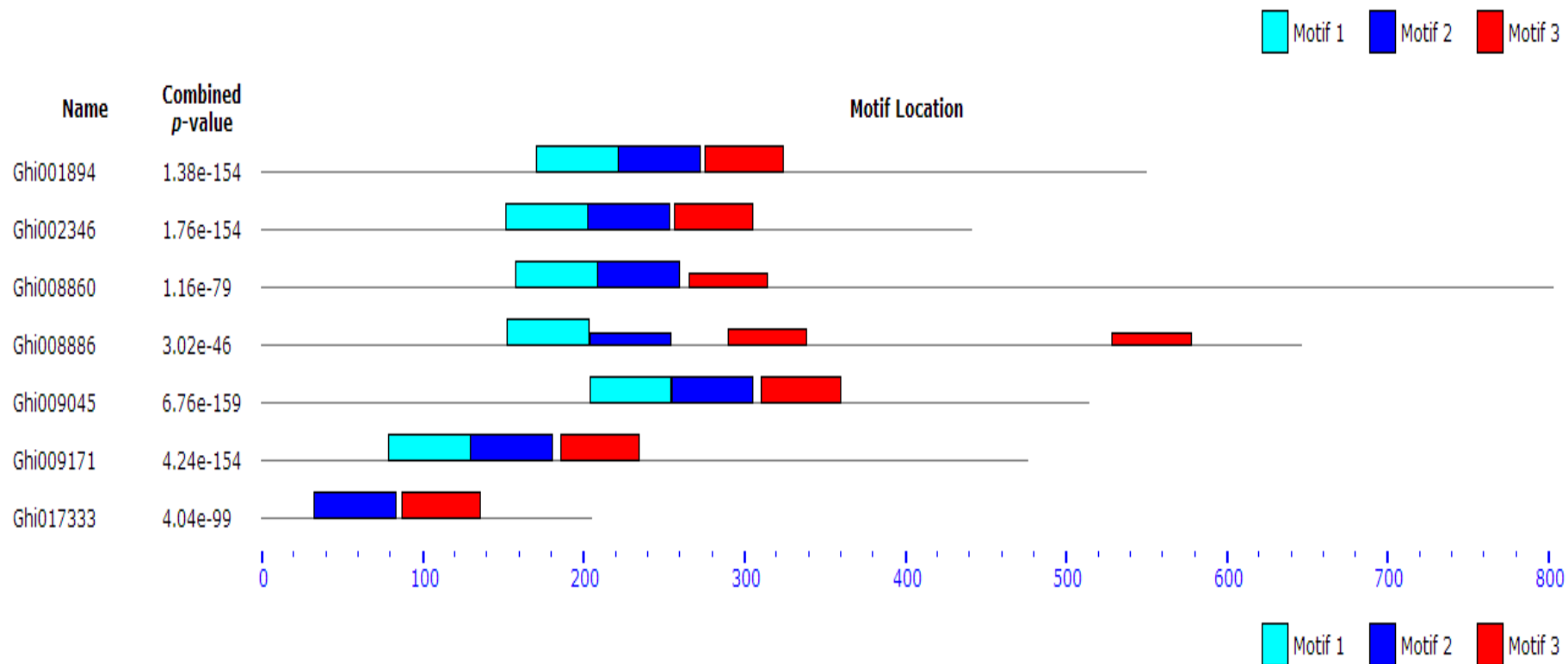
[Top](#)

Combined Block Diagrams [?](#)

Non-overlapping sites with a p -value better than 0.0001.

The height of the motif "block" is proportional to $-\log(p\text{-value})$, truncated at the height for a motif with a p -value of $1e-10$.

Click on any row to highlight sequence in all motifs. The motif blocks have tool tips with more information.



MOTIF 1

[Next Top](#)

Summary ?

E-value 1.7e-147
Width 50
Sites 6
[show more](#)

Sequence Logo ?



Standard

Download LOGO ? Orientation: SSC: Format: Width: cm Height: cm

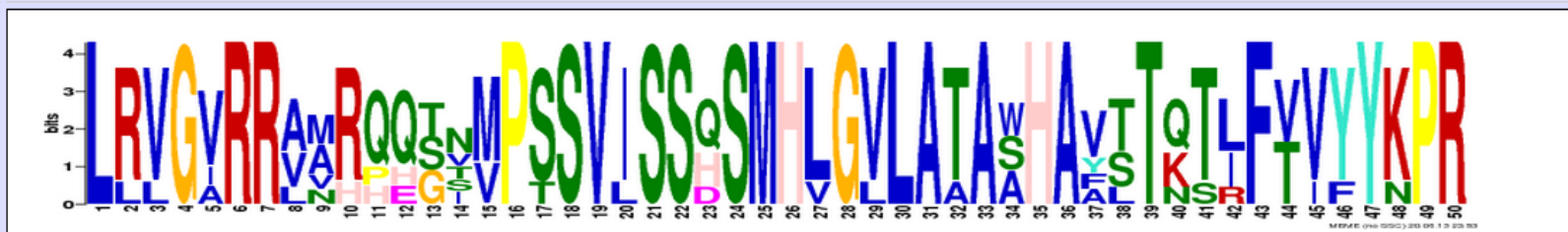
MOTIF 2

[Previous Next Top](#)

Summary ?

E-value 5.6e-120
Width 50
Sites 6
[show more](#)

Sequence Logo ?



Standard

Download LOGO ? Orientation: SSC: Format: Width: cm Height: cm

MOTIF 3

[Previous Top](#)

Summary ?

E-value 2.5e-095
Width 48
Sites 5
[show more](#)

Sequence Logo ?



Standard

Download LOGO ? Orientation: SSC: Format: Width: cm Height: cm

棉花ARF转录因子家族蛋白 亚细胞定位预测



TargetP 1.1 Server - prediction results

Technical University of Denmark

```
### targetp v1.1 prediction results #####  
Number of query sequences: 7  
Cleavage site predictions not included.  
Using PLANT networks.
```

Name	Len	cIP	mIP	SP	other	Loc	RC
Ghi001894	551	0.167	0.110	0.016	0.883	-	2
Ghi002346	442	0.052	0.176	0.086	0.799	-	2
Ghi008860	804	0.063	0.333	0.114	0.496	-	5
Ghi008886	647	0.056	0.153	0.045	0.796	-	2
Ghi009045	515	0.231	0.104	0.054	0.831	-	3
Ghi009171	477	0.253	0.045	0.044	0.901	-	2
Ghi017333	206	0.107	0.521	0.010	0.141	M	4
cutoff		0.000	0.000	0.000	0.000		

>Ghi001894/129-230

FCKVLTASDTSTHGGFSVLRKHATDCIpKLNmndaTPTQELVAKDLHGYESWRFKHIfRGQ
PRRHLLTTGWSTFVTSKRLVAGDSFVFLRGENGELRVGVERRV

>Ghi009171/37-138

FCKILTASDTSTHGGFSVLRRAHADECipPLDmsrqPPTQELVAKDLHGNEWRFRRHIfRGQ
PRRHLLQSGWSVFVSSKRLVAGDAFIFLARGENGELRVGVERRA

>Ghi008886/111-212

FAKTLTQSDANNGGGFSVPRYCAETIfpRLDytadPPVQTVIARDVHGEIWKFRHIyRGT
PRRHLLTTGWSSFVNHHKLVAGDSIVFLRAENGELCVGIRRA

>Ghi009045/162-263

FCKTLTASDTSTHGGFSVLRRAHADECipPLDmsrqPPTQELVAKDLHGNEWRFRRHIfRGQ
PRRHLLQSGWSVYVSSKRLVAGDAFIFLARGENGELRVGVERRA

>Ghi002346/110-210

FCKILTASDTSTHGGFSVLRKHATECipPLDmnqaTPTQELAAKDLHGYESWRFKHIfRGQ
PRRHLLTTGWSTFVTSKRLVAGDAFVFLRGDNGELRVGARR-

>Ghi008860/116-217

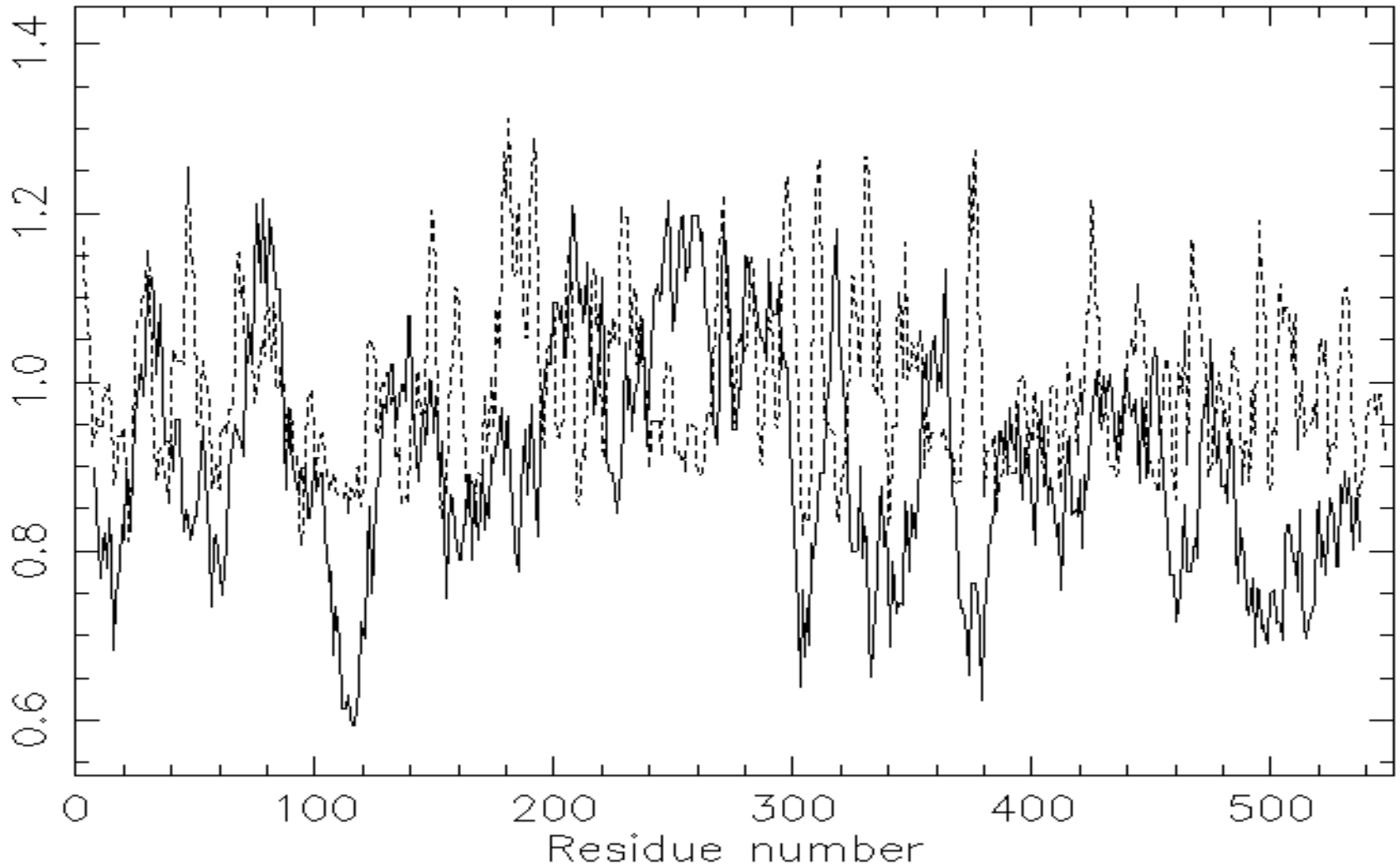
FCKTLTASDTSTHGGFSVPRRAAEKVfsPLDfsqqPPAQELIARDLHDNEWKFRRHIfRGQ
PKRHLLTTGWWSVFVSAKRLVAGDSVLFIWNEKNQLLLGIRRA

>Ghi017333/2-41

--RHLLTTGWSTFVTSKRLVAGDSFVFLRGENGELRVGVERRV

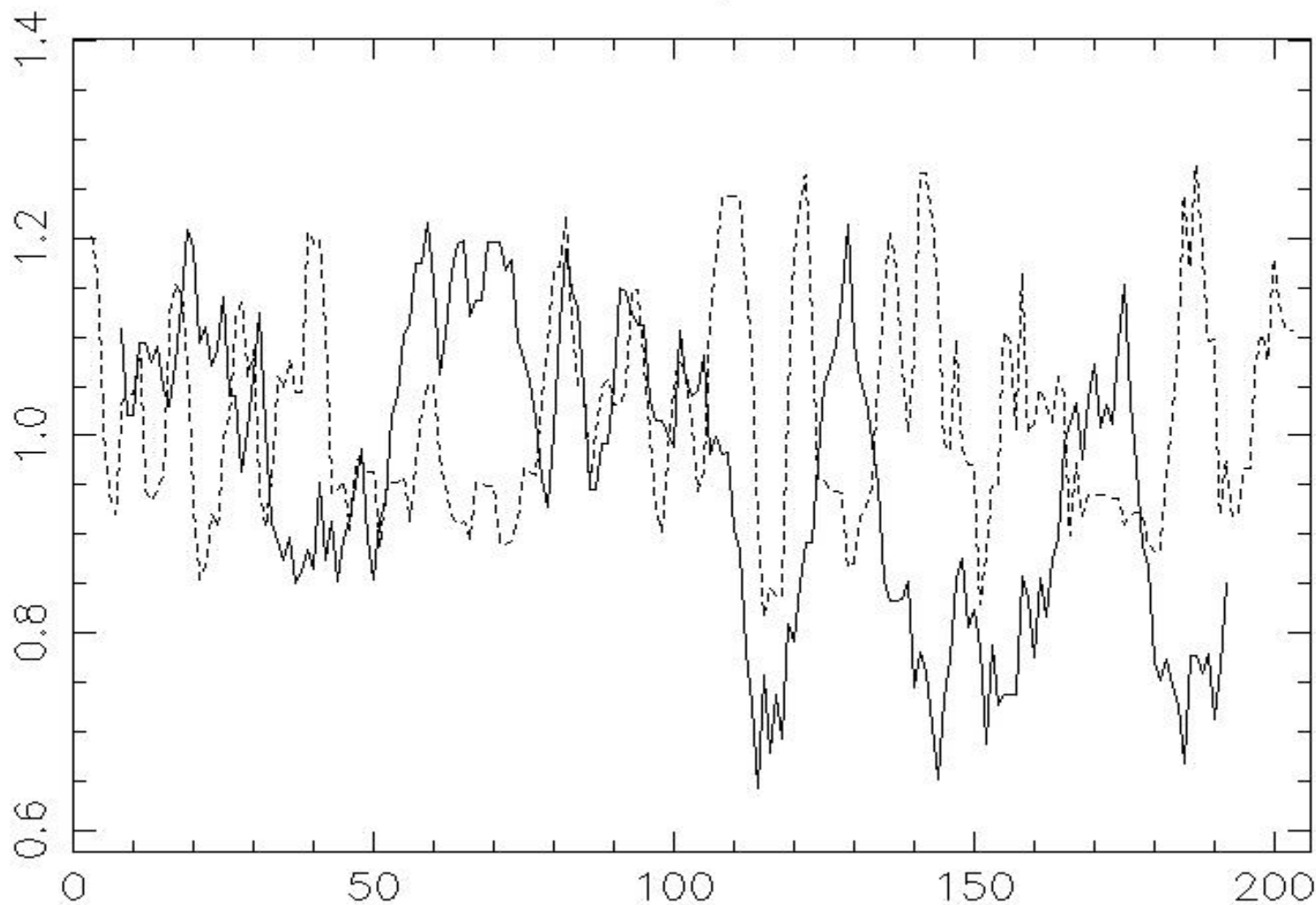
棉花ARF转录因子家族蛋白 Ghi001894跨膜结构预测

Tmap



棉花ARF转录因子家族蛋白 Ghi017333跨膜结构预测

Tmap



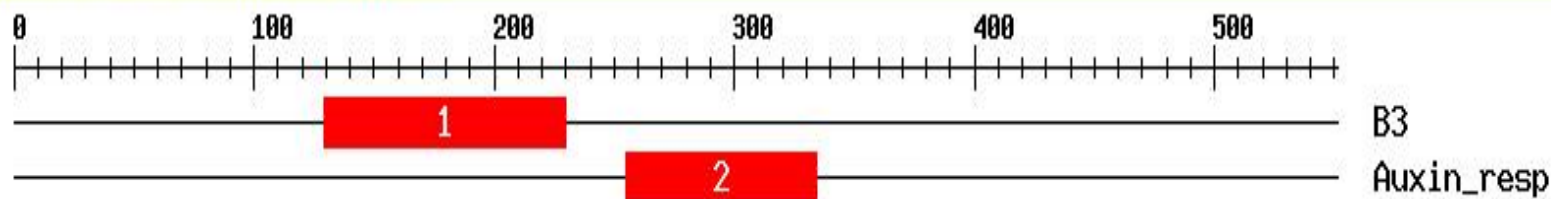
用smart网站做Ghi001894的DNA结合域预测，结果与PlantTDF上公布的一致

Confidently predicted domains, repeats, motifs and features:

Name	Start ▲	End	E-value
B3	129	231	7.29e-22
low complexity	233	249	98500

Click on a row to highlight the feature in the diagram above. Click the feature name for more information.

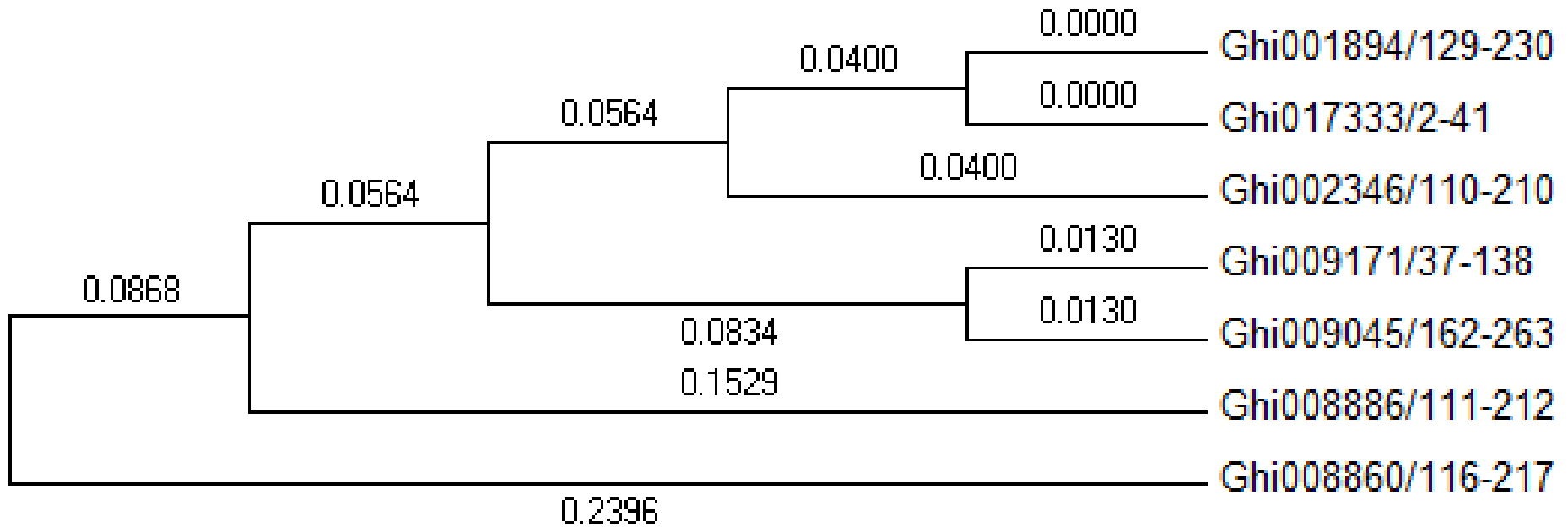
Signature Domain [? help](#) Back to Top



No.	Domain	Score	E-value	Start	End	HMM Start	HMM End
1	B3	80.2	2.2e-25	129	230	1	95

EEEE--HHHHTT..-EE-----TTT-..-----SEEEEE-----EEEEEE..EE-----EEE-----HHHHHHT-----EEEEEE-SSSEEEEEEE CS

棉花ARF转录因子家族蛋白质 BDB区域进化树



Alignment

Annotation

No sequence annotation (1) available for this local

Amino acid properties

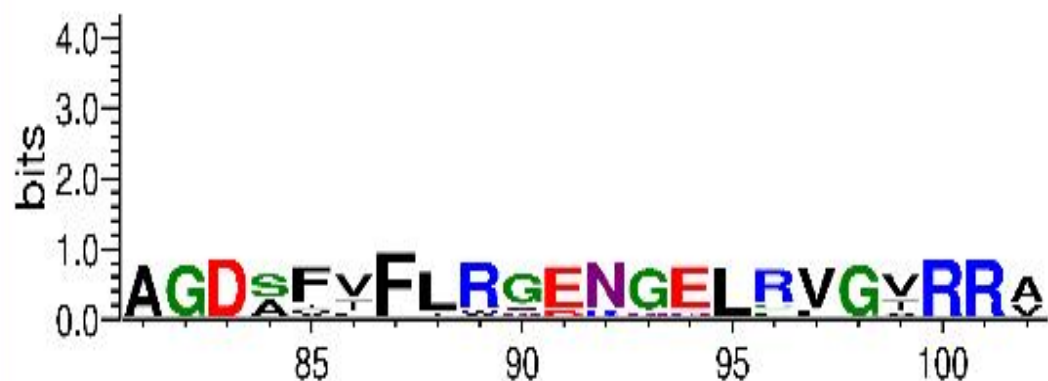
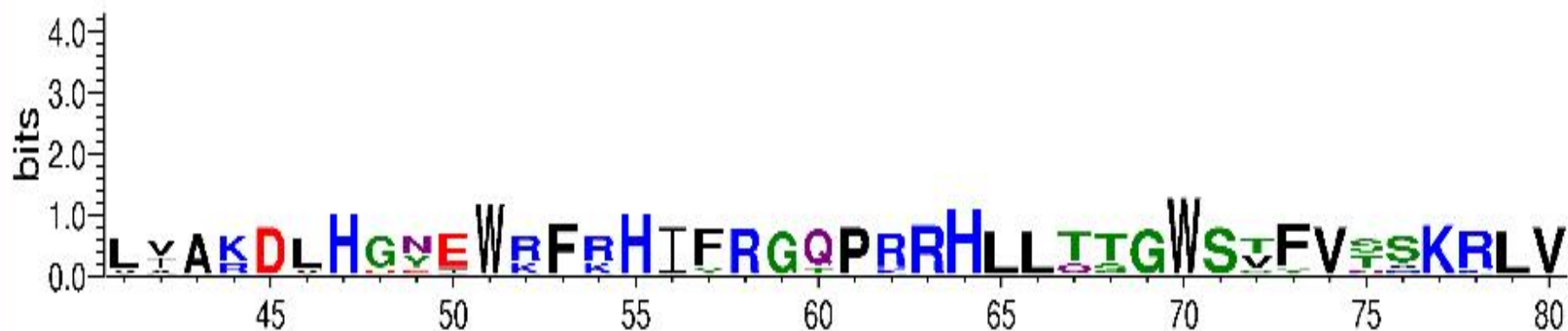
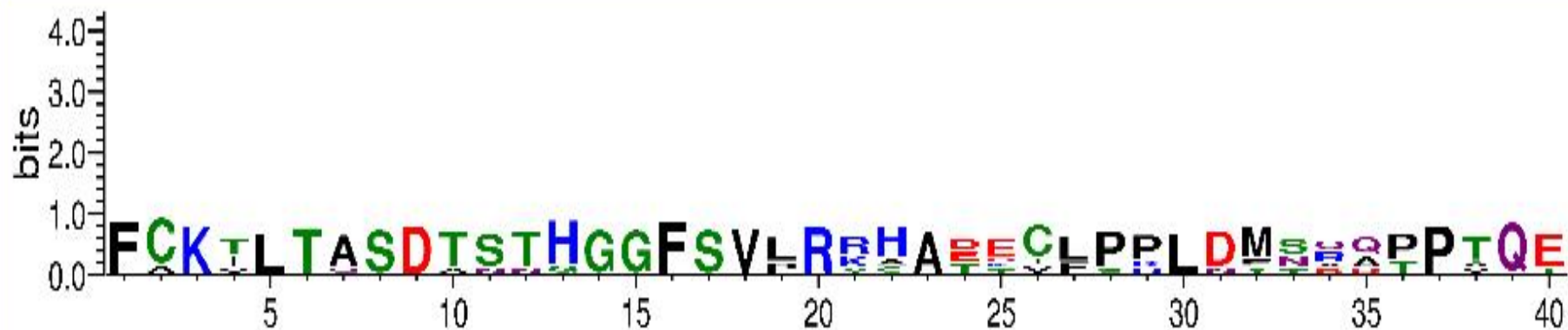
- Similarity
- Hydrophobic
- Negative
- Positive
- Aliphatic
- Tiny
- Aromatic
- Charged
- Small
- Polar
- Big
- Serine Threonine



Learn how to print this alignment in color

```
1  FCKVLTASDTSTHGGFSVLRKHATDClpKLNmndaTPTQELVAKDLHG YEWRFKHI fRGQ 60 Ghi001894/129-230
1  FCKILTASDTSTHGGFSVLRRAHADEClpPLDmsrqPPTQELVAKDLHGNEWRFRI fRGQ 60 Ghi009171/37-138
1  FAKLTIQSDANNGGGFSVPRYCAETIfpRLDytadPPVQTVIARDVHGEIWKFRHI yRGT 60 Ghi008886/111-212
1  FCKILTASDTSTHGGFSVLRRAHADEClpPLDmsrqPPTQELVAKDLHGNEWRFRI fRGQ 60 Ghi009045/162-263
1  FCKILTASDTSTHGGFSVLRKHATEClpPLDmngaTPTQELAAKDLHG YEWRFKHI fRGQ 60 Ghi002346/110-210
1  FCKILTASDTSTHGGFSVPRRAAEKVisPLDfsqgPPAQELIARDLHDNEWKFRHI fRGQ 60 Ghi008860/116-217
*. * ** ** :.. ***** * * *: *. * : *:* * *:*:* **
```

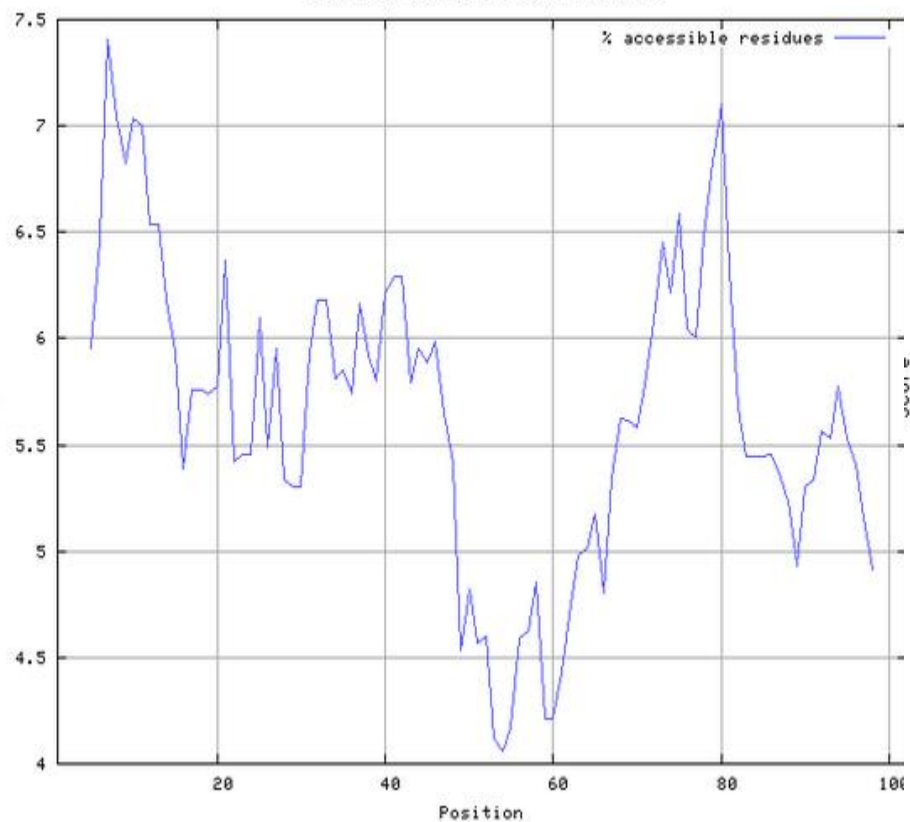
```
61 PRRHLITGWSTFVTSKRLVAGDSFVFLRGENGELRVGVRRV 102 Ghi001894/129-230
61 PRRHLIQSGWSVVFVSSKRLVAGDAFIFLARGENGELRVGVRRR 102 Ghi009171/37-138
61 PRRHLITGWSSFVNHHKLVAGDSIVFLRAENGLCVGIRRA 102 Ghi008886/111-212
61 PRRHLIQSGWSVYVSSKRLVAGDAFIFLARGENGELRVGVRRR 102 Ghi009045/162-263
61 PRRHLITGWSTFVTSKRLVAGDAFVFLRGDNGELRVGARR- 101 Ghi002346/110-210
61 PKRHLITGWSVFVSAKRLVAGDSVLFVNEKNQLLGIIRRA 102 Ghi008860/116-217
*:**** :*** :*.*:*****:..*: :: :* :* **
```

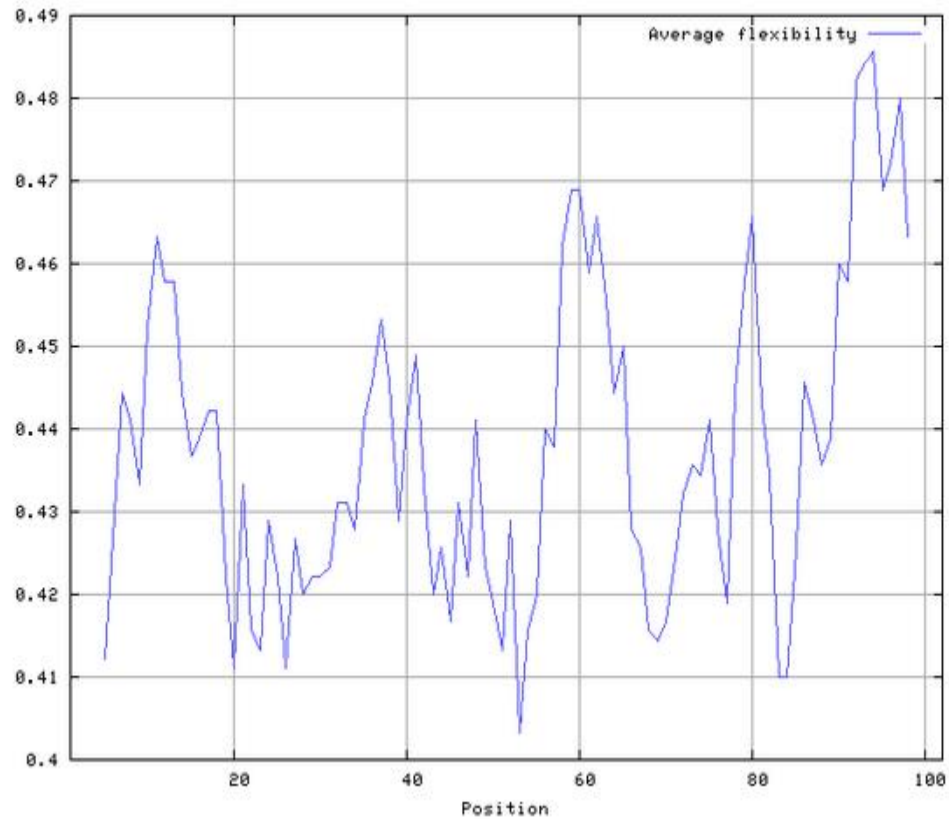
以Ghi001894蛋白为例

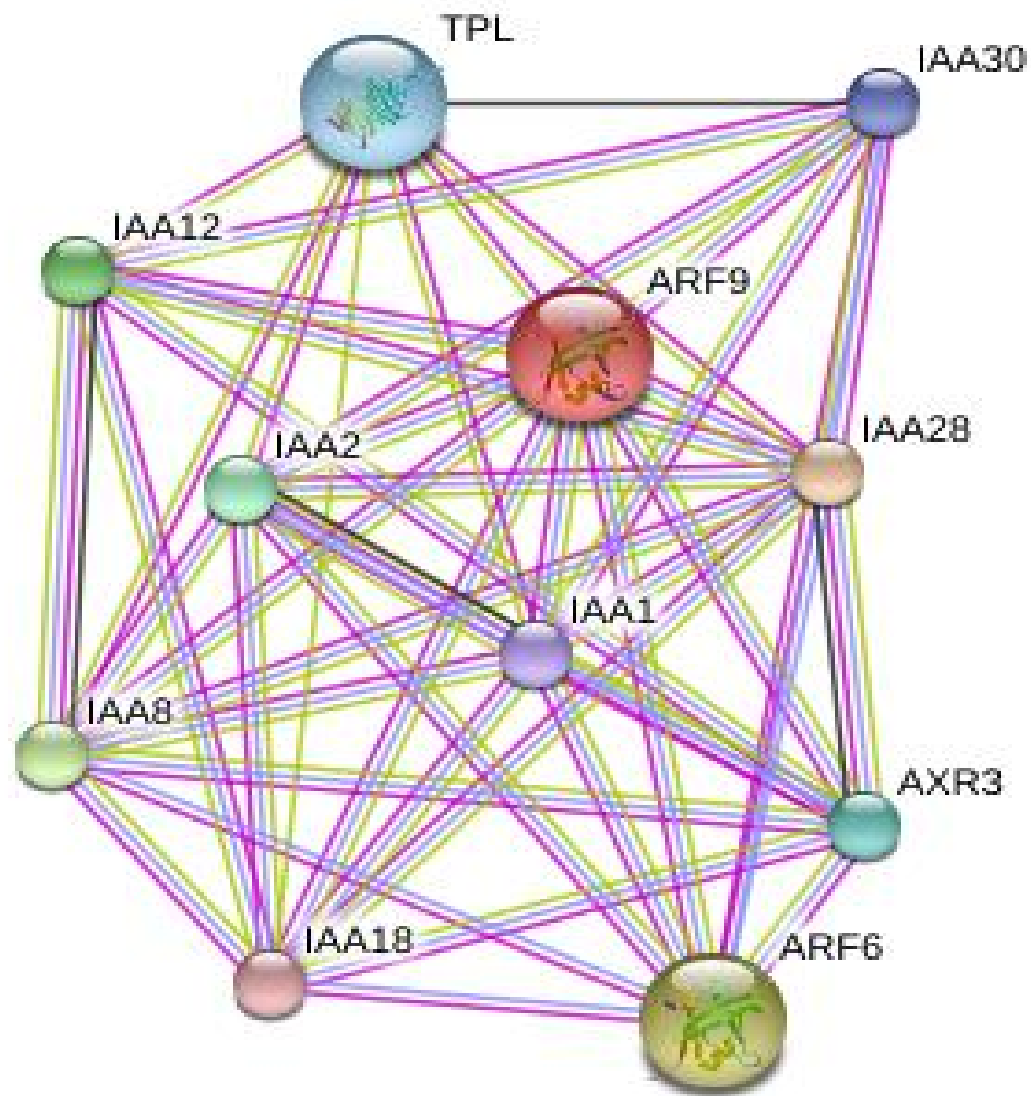
- 极性氨基酸62个，占BDB区域总氨基酸的61%。
- 疏水性氨基酸50个，占BDB区域总氨基酸的50%。

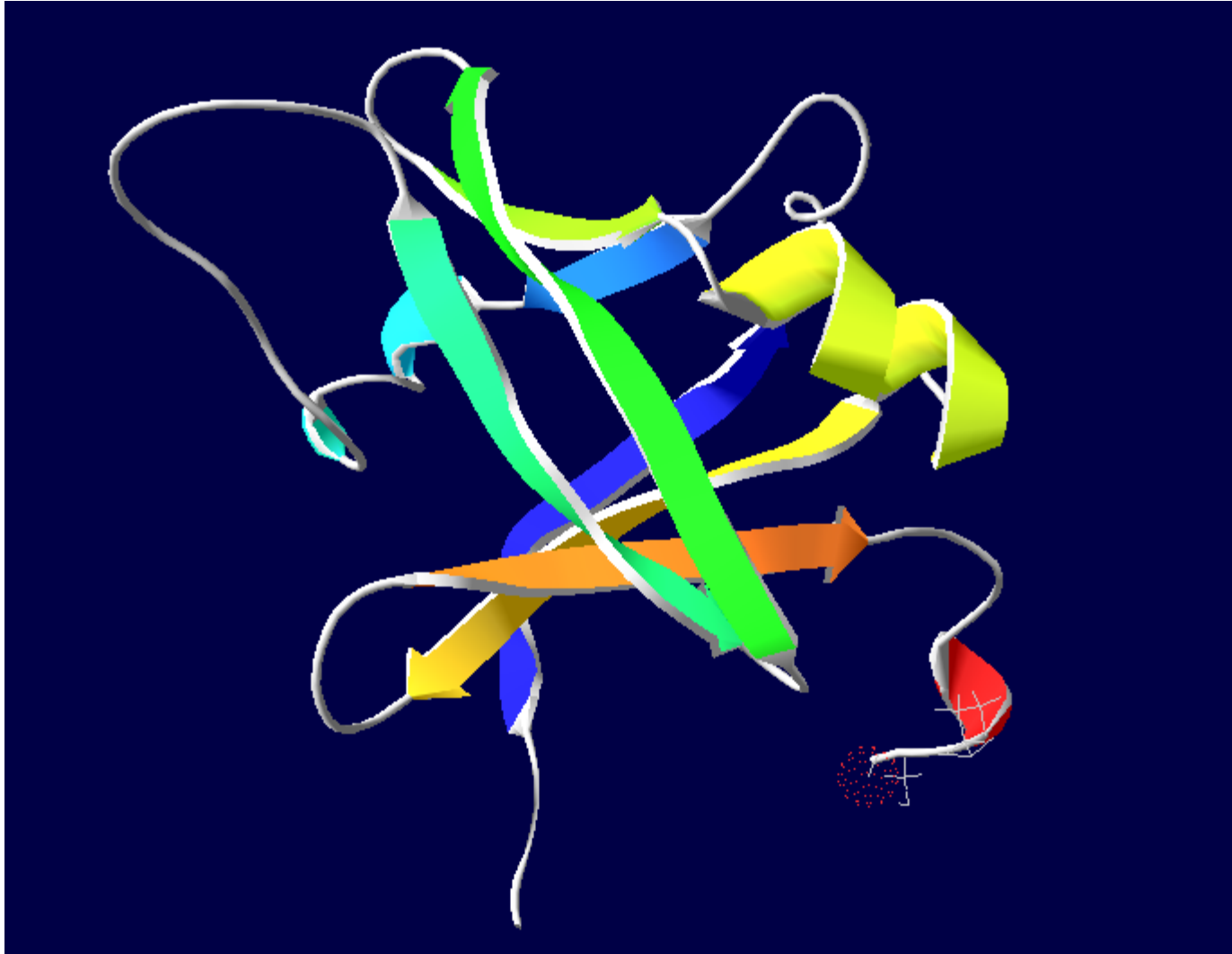
ProtScale output for user sequence



ProtScale output for user sequence







B3 DNA结合域是在高等植物的转录因子中发现的，能与其他结合域共同起作用的一种结构域，它有7个beta折叠，2个alpha罗旋，它能与大多数DNA罗旋的凹槽区相互作用。

Thank you!