Candidate gene analysis of QTL *qTGW6.1* for 1000-grains weight on the short arm of rice chromosome 6

G08 奉保华 欧承刚 张振华

粒重

产量三要素之一

品质：短圆，细长，垩白

粒长：3 mm~11 mm

粒宽：1.2 mm~3.8 mm

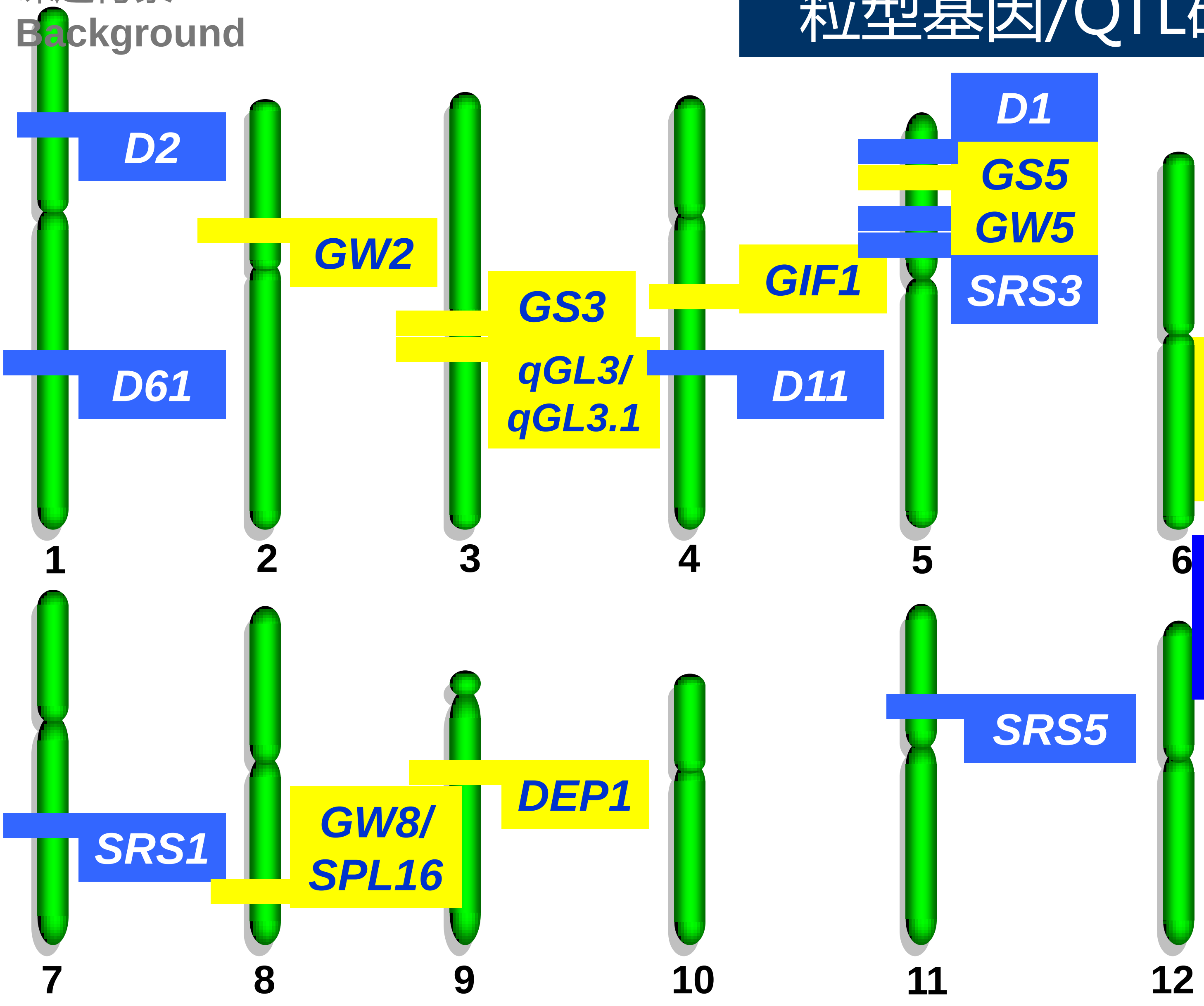
粒厚

灌浆

千粒重：10 g~70 g

Huang et al., 2012, TRPLSC

粒型基因/QTL研究进展



QTL克隆8个

突变体克隆7个

粒型基因/QTL研究进展

Gene	Trait	Encoded protein	Note
GW2	Grain width	RING-type E3 ubiquitin ligase	Ubiquitin–proteasome pathway
GS3	Grain length	Membrane protein with multiple domains	MAS
qGL3/qGL3.1	Grain length	Protein phosphatase with Kelch-like repeat domain	
GIF1	Grain filling	Cell wall invertase	
GW5/qGW5	Grain width	Arginine-rich protein of 144 amino acids	Ubiquitin–proteasome pathway
GS5	Grain shape	Serine carboxypeptidase	
GW8/SPL16	Grain width	SQUAMOSA promoter-binding protein-like 16 (OsSPL16)	SBP domain-containing transcription factors
DEP1	Grain length	Phosphatidylethanolamine-binding protein-like domain protein	
D61/OsBRI		BRI-like leucine-rich repeat (LRR) receptor kinase	BR receptor
D2/CYP90D2	Seed size	Cytochrome P450 (CYP90D) enzyme	BR biosynthesis
D11/CYP724B1	Seed length	Cytochrome P450 (CYP724B1) enzyme	BR biosynthesis
D1	Seed size	Heterotrimeric G protein alpha subunit (RGA1)	
SRS3	Seed size	Kinesin 13 protein	
SRS1/DEP2	Seed size	Protein of 1365 amino acids with unknown function	
SRS5	Seed size	Alpha-tubulin protein	

课题背景
Background

珍汕97B/密阳46-RIL

qTGW6.1

珍汕97B/密阳 BC_2F_5

高代回交自交梯系材料

精细定位

候选基因分析

Chromsome 6

Si9291~Si9396

105 kb

获取*Japonica* 序列 Gramene

不同软件的基因预测

Software/Datebase	Speices	No. of Predicted Gene
AUGUSTUS	Maize	17
Genscan	Maize	21
Softberry	Rice	11
MSU annotation		12
RGP Predicted		13

Softberry与MSU基因预测比较

Softberry			MSU		
Predicted genes1			**LOC_1**		
1 CDSf	9293370	9294632	1CDSf	9293370	9294632
2 CDSl	9294731	9296556	2CDSl	9294731	9296556
Predicted genes2			**LOC_2**		
1 CDSf	9297391	9300925	1CDSf	9297391	9300925
2 CDSl	9300997	9303352	2CDSl	9300997	9303352
Predicted genes3			**LOC_3**		
1 CDSf	9315158	9315255	1CDSf	9315158	9315255
2 CDSi	9315357	9315490	2CDSi	9315357	9315490
3 CDSi	9315591	9315663	3CDSi	9315591	9315663
4 CDSi	9315776	9315663	4CDSi	9315776	9315920
5 CDSi	9316133	9316205	5CDSi	9316133	9316205
6 CDSi	9316402	9316462	6CDSi	9316402	9316462
7 CDSi	9316567	9316821	7CDSi	9316567	9316821
8 CDSi	9318123	9318465	8CDSi	9318123	9318465
9 CDSi	9320184	9320564	9CDSi	9320184	9320564
10 CDSl	9322050	9322419	10CDSl	9322050	9322419
Predicted genes4			**LOC_4**		
1 CDSl	9323262	9323581	1CDSl	9323262	9323581
2 CDSi	9323742	9323807	2CDSi	9323742	9323807
3 CDSi	9323906	9323930	3CDSi	9323906	9323930
4 CDSf	9324279	9324420	4CDSf	9324279	9324420
Predicted genes 5			**LOC_5**		
1 CDSl	9325254	9325667	1CDSl	9325254	9325667
2 CDSi	9326079	9326245	2CDSi	9326079	9326245
3 CDSi	9326294	9326492			
4 CDSi	9326603	9326795	3CDSi	9326603	9326795
5 CDSf	9326852	9327155	4CDSf	9326921	9327155

Softberry			MSU		
Predicted genes 6			**LOC_6**		
1CDS	9332213	9332864	1CDS	9332213	9332864
Predicted genes 7			**LOC_7**		
1 CDSf	9334555	9334613	1CDSf	9335535	9336363
2 CDSi	9335497	9336363	2CDSl	9337000	9337360
3 CDSl	9337000	9337360			
			LOC_8		
			1CDSf	9349260	9349580
			2CDSi	9348668	9348723
			3CDSl	9348327	9348435
Predicted genes 8			**LOC_9**		
1CDSf	9351724	9352087	1CDSf	9353006	9353030
2CDSi	9353112	9353232	2CDSi	9353130	9353232
3CDSi	9353340	9353660	3CDSi	9353340	9353660
4CDSi	9354450	9354612	4CDSi	9354450	9354612
5CDSi	9355334	9355419	5CDSi	9355334	9355419
6CDSi	9355535	9355685	6CDSi	9355535	9355685
7CDSi	9355771	9355819	7CDSi	9355771	9355819
8CDSi	9355888	9355957	8CDSi	9355888	9355957
9CDSi	9356080	9356784	9CDSi	9356080	9356784
10CDSi	9357609	9357745	10CDSi	9357609	9357745
11CDSi	9357819	9358040	11CDSi	9357819	9358040
12CDSi	9358538	9358670	12CDSi	9358538	9358670
13CDSi	9359499	9359590	13CDSi	9359499	9359590
14CDSi	9360061	9360109	14CDSi	9360061	9360109
15CDSi	9360187	9360316	15CDSi	9360187	9360316
16CDSi	9360540	9360810	16CDSl	9360540	9360810

Softberry			MSU		
Predicted genes 9			**LOC_10**		
1CDSf	9362511	9362706			
2CDSi	9363409	9363510			
3CDSi	9363889	9364520	1CDSf	9363917	9364520
4CDSi	9364609	9364783	2CDSi	9364609	9364834
5CDSi	9364922	9364988	3CDSi	9364922	9364988
6CDSi	9365070	9365139	4CDSi	9365070	9365139
7CDSi	9365237	9365309	5CDSi	9365237	9365309
8CDSi	9365390	9365582	6CDSi	9365390	9365582
9CDSl	9366297	9366384	7CDSl	9366297	9366384
Predicted genes 10			**LOC_11**		
1CDSl	9367544	9369202	1CDSl	9367544	9369202
2CDSi	9369539	9369849	2CDSi	9369539	9369849
3CDSf	9369994	9370948	3CDSf	9369994	9370948
Predicted genes 11			**LOC_12**		
1CDSf	9395465	9395847	1CDSf	9395465	9395847
2CDSi	9396258	9396411	2CDSi	9396258	9396411
3CDSi	9396519	9396723	3CDSi	9396519	9396723
4CDSi	9396892	9396942	4CDSi	9396892	9397358

候选基因的初步分析

ID	Gene Product
LOC_1	Retrotransposon protein
LOC_2	Retrotransposon protein
LOC_3	BAK1 precursor
LOC_4	Expressed protein
LOC_5	Peroxidase precursor
LOC_6	Expressed protein
LOC_7	CCT/B-Box zinc finger protein
LOC_8	Expressed protein
LOC_9	*Hd1*
LOC_10	Helix-loop-helix DNA-binding domain containing protein
LOC_11	Armadillo/beta-catenin repeat family protein
LOC_12	Amino acid transporter

qTGW6.1

qTGW6.1蛋白序列（预测）BlastP Searches (UniRef 100)

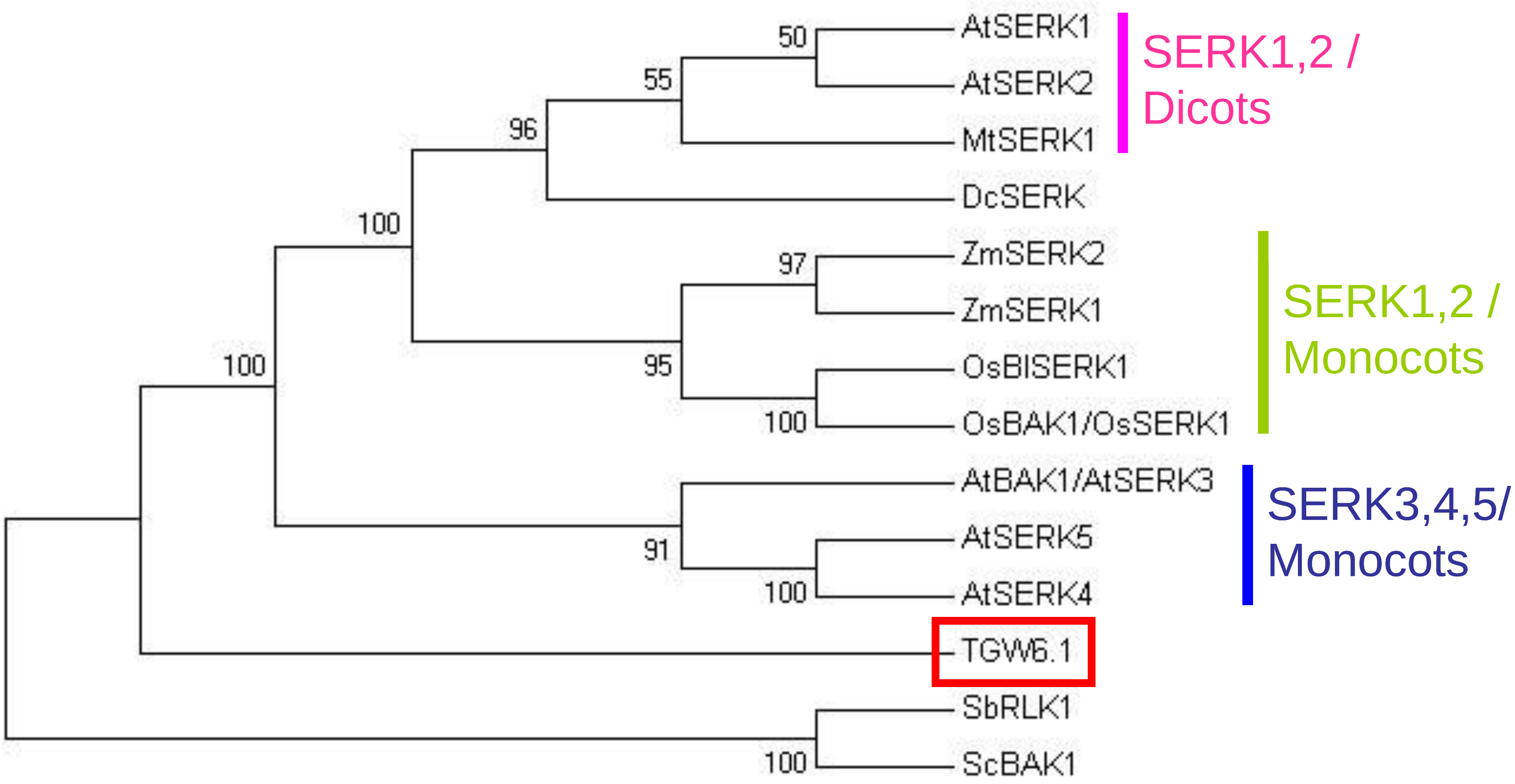
Entry name	St.	Protein names	Length	Identity	Score	E-value
I1Q1H7_ORYGL	☆	Uncharacterized protein	641	99	3281	0
A3BAL8_ORYSJ	☆	Putative uncharacterized protein	671	92	3152	0
D4N3U5_9ORYZ	☆	**Putative somatic embryogenesis protein kinase 1**	636	94	3085	0
D4N3T2_9ORYZ	☆	**Putative somatic embryogenesis protein kinase 1**	632	92	3043	0
B8B0K0_ORYSI	☆	Putative uncharacterized protein	629	91	2969	0
D4N3R9_9ORYZ	☆	**Putative somatic embryogenesis protein kinase 1**	643	89	2961	0
D4N3V6_ORYBR	☆	**Putative somatic embryogenesis protein kinase 1**	640	89	2933	0
K3XVV7_SETIT	☆	Uncharacterized protein	627	80	2653	0
I1GZ29_BRADI	☆	Uncharacterized protein	629	79	2632	0
K3YQV1_SETIT	☆	Uncharacterized protein	624	78	2590	0
C0P934_MAIZE	☆	Uncharacterized protein	632	78	2586	0
C5Z8J1_SORBI	☆	Putative uncharacterized protein Sb10g010010	629	79	2585	0
B9F2H8_ORYSJ	☆	Putative uncharacterized protein	627	77	2553	0
B6SWS1_MAIZE	☆	**BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1**	636	77	2552	0
B8AI39_ORYSI	☆	Putative uncharacterized protein	627	77	2551	0
I1P3V1_ORYGL	☆	Uncharacterized protein	627	77	2548	0
C5XZE6_SORBI	☆	Putative uncharacterized protein Sb04g029170	626	77	2545	0
J3LGP6_ORYBR	☆	Uncharacterized protein	626	77	2543	0
F2D480_HORVD	☆	Predicted protein	625	77	2533	0
B8A2N9_MAIZE	☆	Uncharacterized protein	626	77	2527	0
I1IEA0_BRADI	☆	Uncharacterized protein	628	76	2514	0
C0PH57_MAIZE	☆	Uncharacterized protein	589	78	2450	0
B9T3S1_RICCO	☆	**BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1, puta**	624	70	2277	0
D7SMJ9_VITVI	☆	Putative uncharacterized protein	622	70	2264	0
D7TAB0_VITVI	☆	Putative uncharacterized protein	625	69	2249	0

SERK

(体细胞胚发生受体类激酶）

BAK1

对qTGW6.1与具有代表性的SERK蛋白进行系统发育分析



AtSERK1, 2, 3, 4 and 5 from Arabidopsis; OsBAK1 and OsBISERK1 from rice; ZmSERK1 and 2 from maize; MtSERK1 from Medicago truncatula; DcSERK from carrot; SbRLK from sorghum(高粱); and ScBAK1 from sugarcane(甘蔗)

Vicentini et al., 2009, PCR

qTGW6.1BlastP搜索直系同源基因

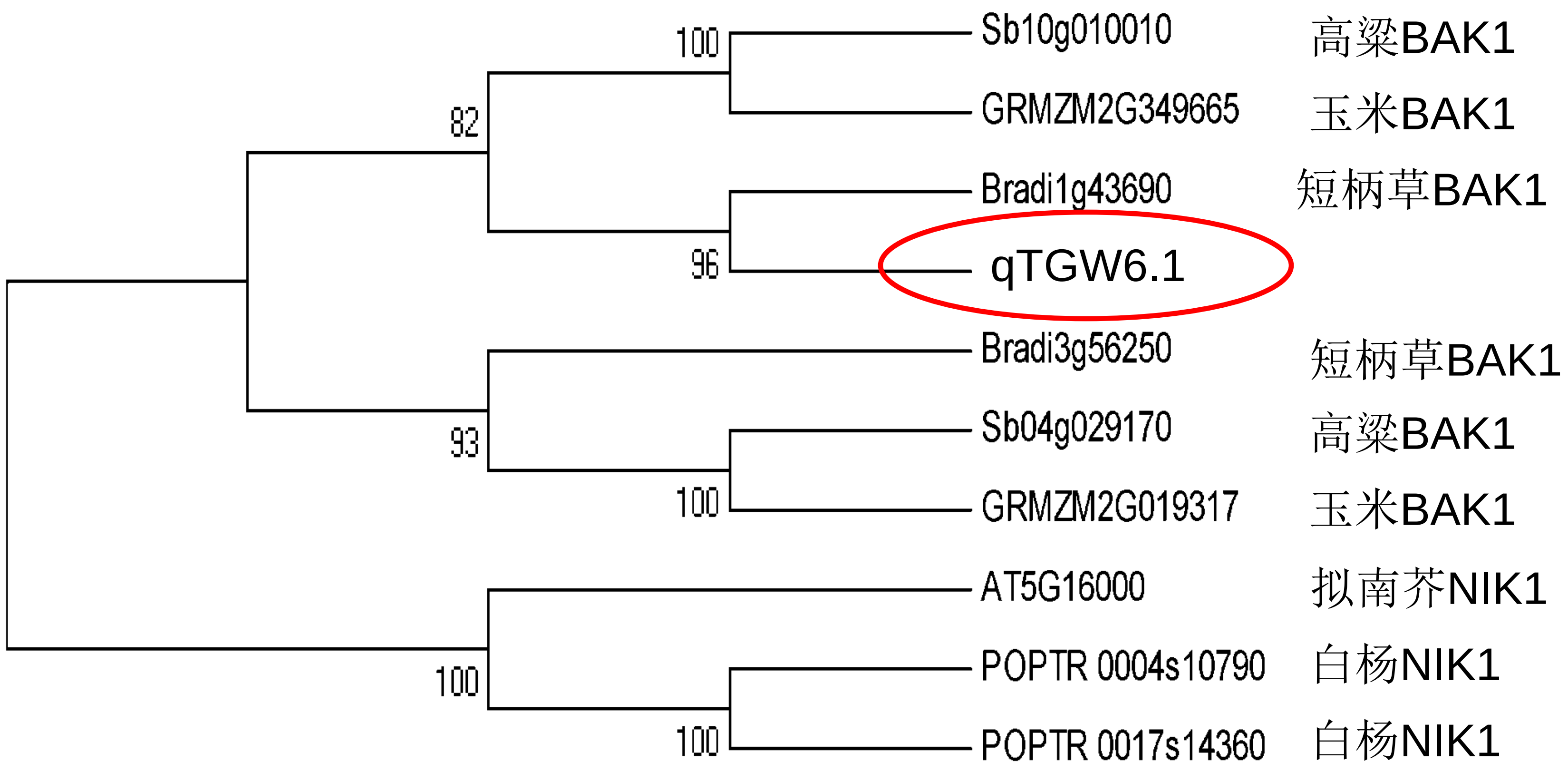
Species	Orthologous gene	Putative function
Brachypodium	Bradi1g43690	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
Brachypodium	Bradi3g56250	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
Maize	GRMZM2G019317	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
Maize	GRMZM2G349665	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
Sorghum	Sb04g029170	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
Sorghum	Sb10g010010	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
Poplar	POPTR_0004s10790	NIK1 (NSP-INTERACTING KINASE 1); kinase
Poplar	POPTR_0017s14360	NIK1 (NSP-INTERACTING KINASE 1); kinase
Arabidopsis	AT5G16000	NIK1 (NSP-INTERACTING KINASE 1); kinase

LRR-RLKII

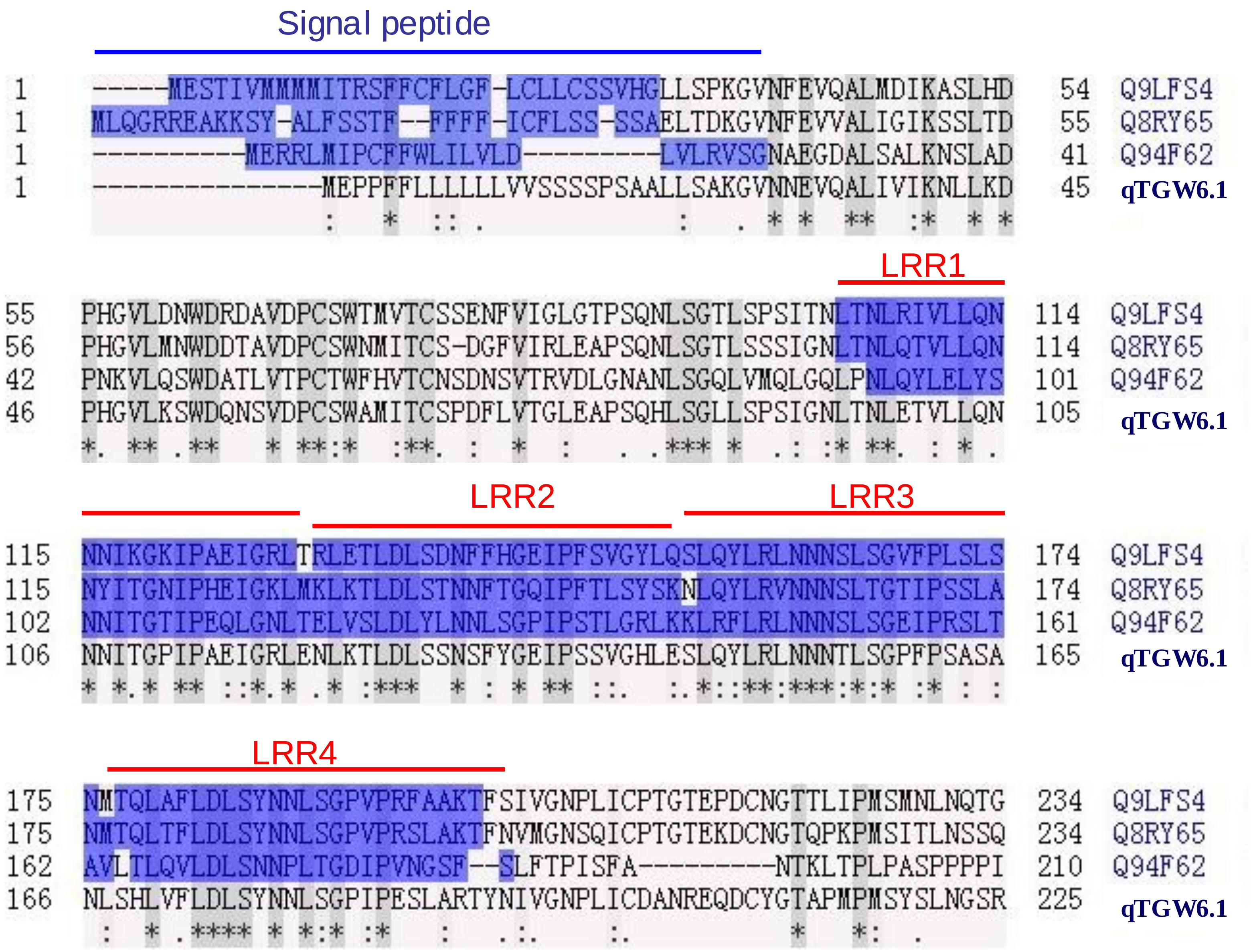
Are NIK-like and SERK-like receptors functionally redundant?

Santos et al., 2009, Journal of Experimental Botany

对qTGW6.1与直系同源蛋白进行系统发育分析

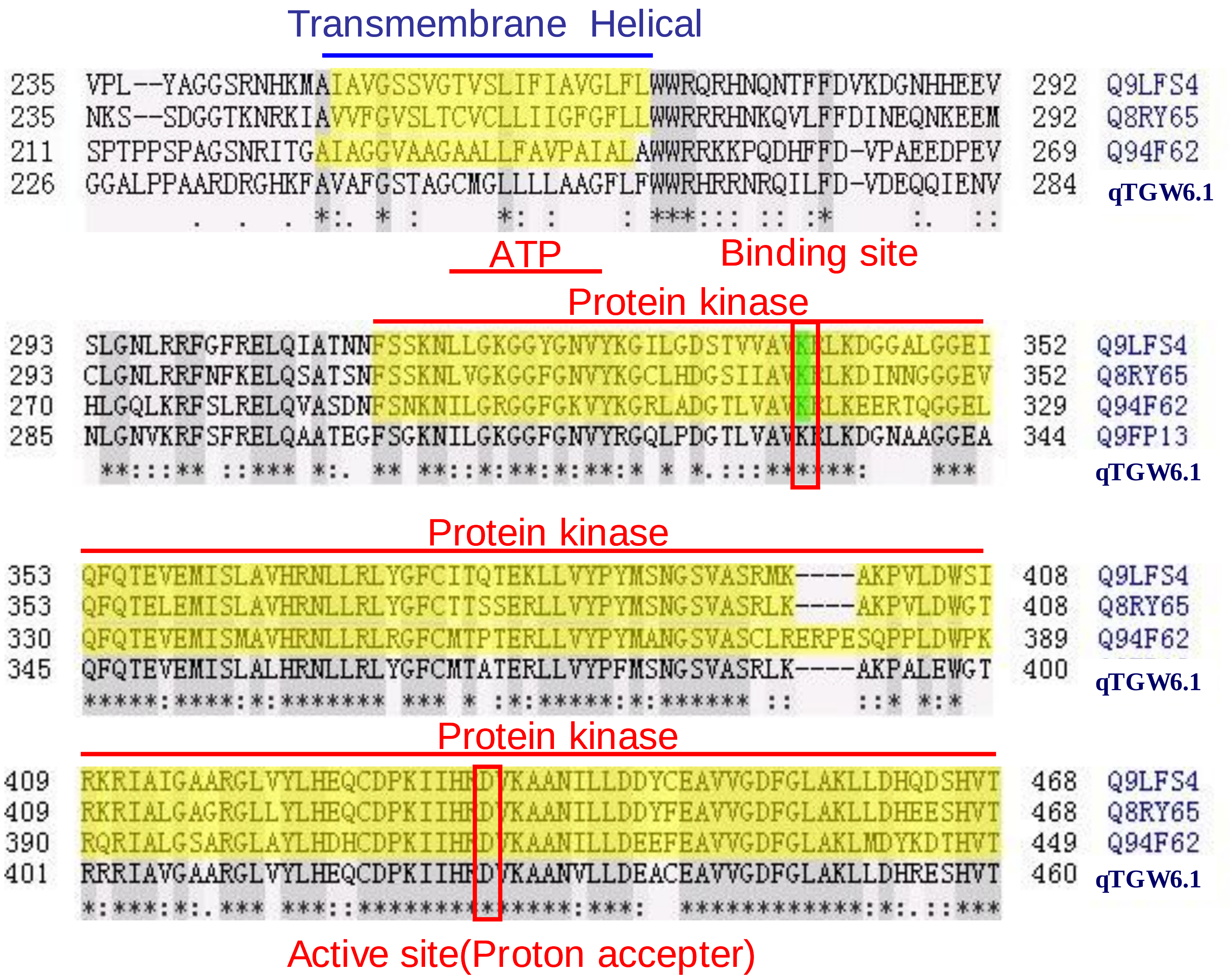


Aligment of qTGW6.1, AtNIK1, AtNIK2 and AtBAK1



```
1     -----MESTIVMMMMITRSFFCFLGF-LCLLCSSVHGLLSPKGVNFEVQALMDIKASLHD  54   Q9LFS4
1     MLQGRREAKKSY-ALFSSTF--FFFF-ICFLSS-SSAELTDKGVNFEVVALIGIKSSLTD  55   Q8RY65
1     ----------MERRLMIPCFFWLILVLD---------LVLRVSGNAEGDALSALKNSLAD  41   Q94F62
1     ---------------MEPPFFLLLLLLVVSSSSPSAALLSAKGVNNEVQALIVIKNLLKD  45   qTGW6.1
                     :   *  :: .            :   . * *  **  :*  * *

55    PHGVLDNWDRDAVDPCSWTMVTCSSENFVIGLGTPSQNLSGTLSPSITNLTNLRIVLLQN  114  Q9LFS4
56    PHGVLMNWDDTAVDPCSWNMITCS-DGFVIRLEAPSQNLSGTLSSSIGNLTNLQTVLLQN  114  Q8RY65
42    PNKVLQSWDATLVTPCTWFHVTCNSDNSVTRVDLGNANLSGQLVMQLGQLPNLQYLELYS  101  Q94F62
46    PHGVLKSWDQNSVDPCSWAMITCSPDFLVTGLEAPSQHLSGLLSPSIGNLTNLETVLLQN  105  qTGW6.1
      *. ** .**   * **:*  :**. :  *  :   . .*** *  .: :* **. : * .

115   NNIKGKIPAEIGRLTRLETLDLSDNFFHGEIPFSVGYLQSLQYLRLNNNSLSGVFPLSLS  174  Q9LFS4
115   NYITGNIPHEIGKLMKLKTLDLSTNNFTGQIPFTLSYSKNLQYLRVNNNSLTGTIPSSLA  174  Q8RY65
102   NNITGTIPEQLGNLTELVSLDLYLNNLSGPIPSTLGRLKKLRFLRLNNNSLSGEIPRSLT  161  Q94F62
106   NNITGPIPAEIGRLENLKTLDLSSNSFYGEIPSSVGHLESLQYLRLNNNTLSGPFPSASA  165  qTGW6.1
      * *.* ** ::*.* .* :***  * : * ** ::.  :.*::**:***:*:* :* : :

175   NMTQLAFLDLSYNNLSGPVPRFAAKTFSIVGNPLICPTGTEPDCNGTTLIPMSMNLNQTG  234  Q9LFS4
175   NMTQLTFLDLSYNNLSGPVPRSLAKTFNVMGNSQICPTGTEKDCNGTQPKPMSITLNSSQ  234  Q8RY65
162   AVLTLQVLDLSNNPLTGDIPVNGSF--SLFTPISFA---------NTKLTPLPASPPPPI  210  Q94F62
166   NLSHLVFLDLSYNNLSGPIPESLARTYNIVGNPLICDANREQDCYGTAPMPMSYSLNGSR  225  qTGW6.1
       :  * .**** * *:* :*   :   .:.    :.          *   *:  .
```

Aligment of qTGW6.1, AtNIK1, AtNIK2 and AtBAK1



```
235  VPL--YAGGSRNHKMAIAVGSSVGTVSLIFIAVGLFLWWRQRHNQNTFFDVKDGNHHEEV  292  Q9LFS4
235  NKS--SDGGTKNRKIAVVFGVSLTCVCLLIIGFGFLLWWRRRHNKQVLFFDINEQNKEEM  292  Q8RY65
211  SPTPPSPAGSNRITGAIAGGVAAGAALLFAVPAIALAWWRRKKPQDHFFD-VPAEEDPEV  269  Q94F62
226  GGALPPAARDRGHKFAVAFGSTAGCMGLLLLAAGFLFWWRHRRNRQILFD-VDEQQIENV  284  qTGW6.1
            .  .  . *:. * :     *: :    : ***::: :: :*     :.  ::

293  SLGNLRRFGFRELQIATNNFSSKNLLGKGGYGNVYKGILGDSTVVAVKRLKDGGALGGEI  352  Q9LFS4
293  CLGNLRRFNFKELQSATSNFSSKNLVGKGGFGNVYKGCLHDGSIIAVKRLKDINNGGEV  352  Q8RY65
270  HLGQLKRFSLRELQVASDNFSNKNILGRGGFGKVYKGRLADGTLVAVKRLKEERTQGGEL  329  Q94F62
285  NLGNVKRFSFRELQAATEGFSGKNILGKGGFGNVYRGQLPDGTLVAVKRLKDGNAAGGEA  344  Q9FP13
      **:::** ::*** *:. ** **::*:**:*:**:* * *.::::******:   ***        qTGW6.1

353  QFQTEVEMISLAVHRNLLRLYGFCITQTEKLLVYPYMSNGSVASRMK----AKPVLDWSI  408  Q9LFS4
353  QFQTELEMISLAVHRNLLRLYGFCTTSSERLLVYPYMSNGSVASRLK----AKPVLDWGT  408  Q8RY65
330  QFQTEVEMISMAVHRNLLRLRGFCMTPTERLLVYPYMANGSVASCLRERPESQPPLDWPK  389  Q94F62
345  QFQTEVEMISLALHRNLLRLYGFCMTATERLLVYPFMSNGSVASRLK----AKPALEWGT  400  qTGW6.1
     *****:****:*:******* *** * :*:*****:*:****** ::    ::* *:*

409  RKRIAIGAARGLVYLHEQCDPKIIHRDVKAANILLDDYCEAVVGDFGLAKLLDHQDSHVT  468  Q9LFS4
409  RKRIALGAGRGLLYLHEQCDPKIIHRDVKAANILLDDYFEAVVGDFGLAKLLDHEESHVT  468  Q8RY65
390  RQRIALGSARGLAYLHDHCDPKIIHRDVKAANILLDEEFEAVVGDFGLAKLMDYKDTHVT  449  Q94F62
401  RRRIAVGAARGLVYLHEQCDPKIIHRDVKAANVLLDEACEAVVGDFGLAKLLDHRESHVT  460  qTGW6.1
     *:***:*:.*** ***::**************:***:  ************:*:.::***
```

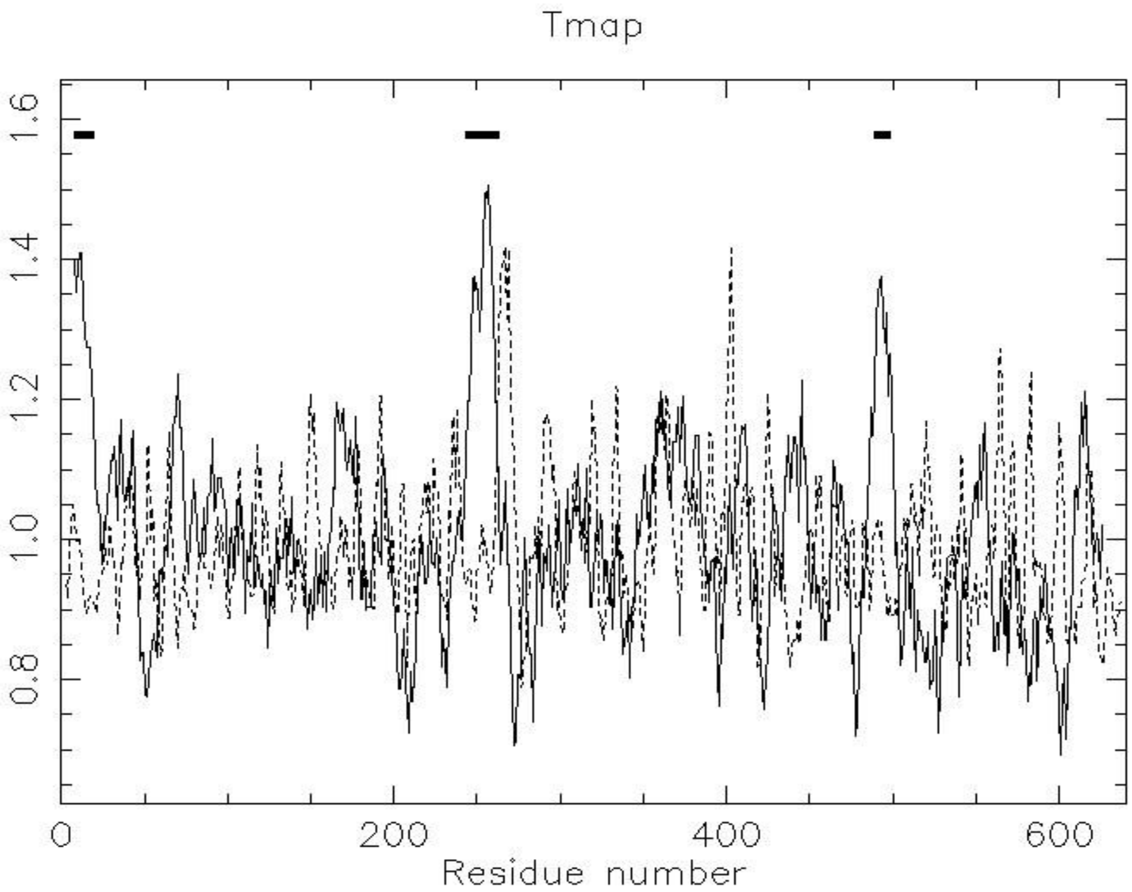
Aligment of qTGW6.1, AtNIK1, AtNIK2 and AtBAK1

```
469  TAVRGTVGHIAPEYLSTGQSSEKTDVFGFGILLLELVTGQRAFEFGKAANQKGV-MLDWV  527  Q9LFS4
469  TAVRGTVGHIAPEYLSTGQSSEKTDVFGFGILLLELITGLRALEFGKAANQRGA-ILDWV  527  Q8RY65
450  TAVRGTIGHIAPEYLSTGKSSEKTDVFGYGVMLLELITGQRAFDLARLANDDDVMLLDWV  509  Q94F62
461  TAVRGTVGHIAPEYLSTGQSSDRTDVFGFGILLLELVTGQTALEFGKSSNHKGA-MLDWV  519  qTGW6.1
     ******:***********:**::*****:*::****:**  *:::.: :*.  . :****

528  KKIHQEKKLELLVDKELLKKKSYDEIELDEMVRVALLCTQYLPGHRPKMSEVVRMLEGDG  587  Q9LFS4
528  KKLQQEKKLEQIVDKDL--KSNYDRIEVEEMVQVALLCTQYLPIHRPKMSEVVRMLEGDG  585  Q8RY65
510  KGLLKEKKLEALVDVDL--QGNYKDEEVEQLIQVALLCTQSSPMERPKMSEVVRMLEGDG  567  Q94F62
520  KKMQSEKKVEVLVDKGL--GGGYDRVEVEEMVQVALLCTQYLPAHRPRMSDVVRMLEGDG  577  qTGW6.1
     * : .***:* :**  *     *.  *::::::*******  * .**:**:*********

588  LAEKWEASQRSDSVSKCSN----------RI--NELMSSSDRYSDLTDDSSLLVQAMELS  635  Q9LFS4
586  LVEKWEASSQRAETNRSYS----------KP--N-EFSSSERYSDLTDDSSVLVQAMELS  632  Q8RY65
568  LAERWEEWQKEEMFRQDFNYPTHHPAVSGW---------------IIGDSTSQIENEYPS  612  Q94F62
578  LADRWEKASGHSTAAADSLSHSHRTSDPAPPAADFAAAFGRCFSDLTDDSSLLVQAVELS  637  qTGW6.1
     *.::**  .                                    :  **:  ::    *

636  GPR  638  Q9LFS4
633  GPR  635  Q8RY65
613  GPR  615  Q94F62
638  GPR  640  qTGW6.1
     ***
```

跨膜螺旋预测

TMAP

Predict and plot transmembrane segments in protein sequences



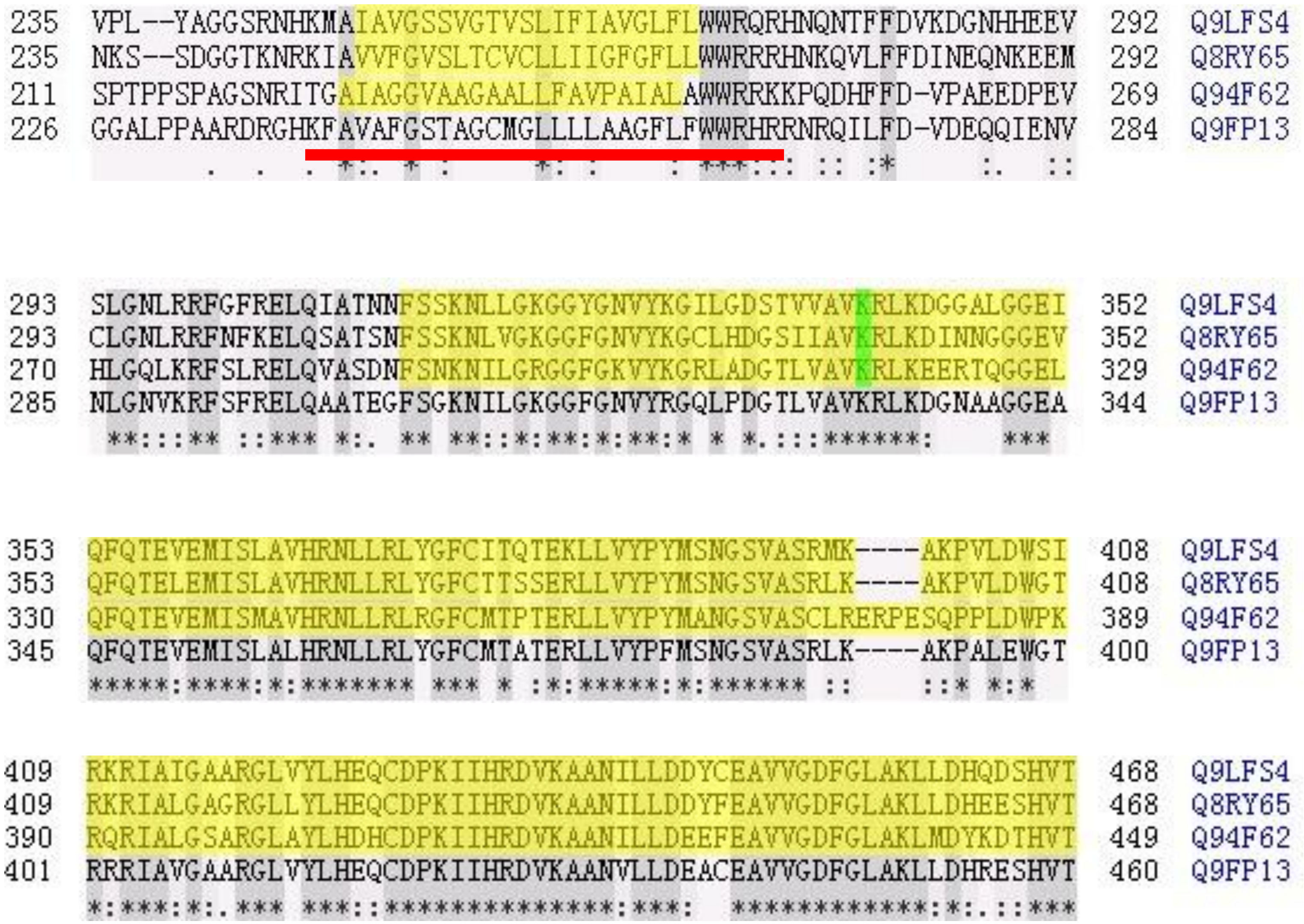
Start	End	TransMem Sequence
4	23	1 PFFLLLLLLVVSSSSPSAAL
239	267	2 KFAVAFGSTAGCMGLLLLAAGFLFWWRHR
485	502	3 DVFGFGILLLELVTGQTA

跨膜螺旋预测

TMAP

Predict and plot transmembrane segments in protein sequences

Transmembrane Helical

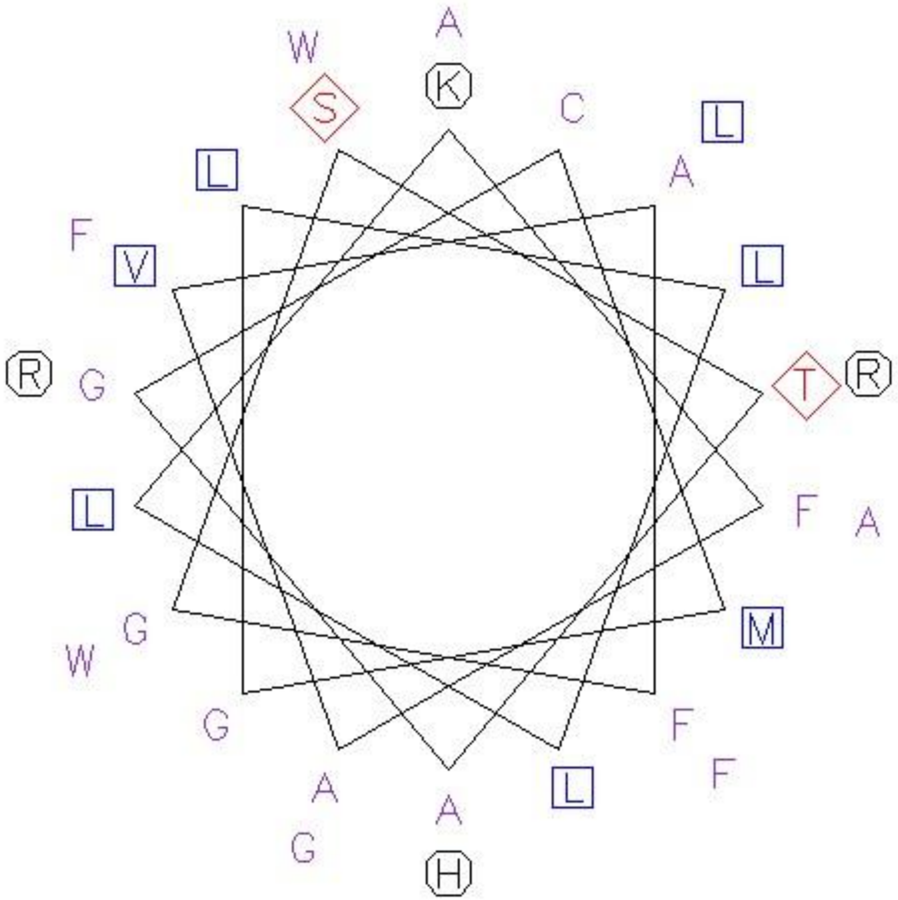


```
235  VPL--YAGGSRNHKMAIAVGSSVGTVSLIFIAVGLFLWWRQRHNQNTFFDVKDGNHHEEV  292  Q9LFS4
235  NKS--SDGGTKNRKIAVVFGVSLTCVCLLIIGFGFLLWWRRRHNKQVLFFDINEQNKEEM  292  Q8RY65
211  SPTPPSPAGSNRITGAIAGGVAAGAALLFAVPAIALAWWRRKKPQDHFFD-VPAEEDPEV  269  Q94F62
226  GGALPPAARDRGHKFAVAFGSTAGCMGLLLLAAGFLFWWRHRRNRQILFD-VDEQQIENV  284  Q9FP13
            .  .  . *:. * :     *: :    : ***::: :: :*     :.  ::


293  SLGNLRRFGFRELQIATNNFSSKNLLGKGGYGNVYKGILGDSTVVAVKRLKDGGALGGEI  352  Q9LFS4
293  CLGNLRRFNFKELQSATSNFSSKNLVGKGGFGNVYKGCLHDGSIIAVKRLKDINNGGGEV  352  Q8RY65
270  HLGQLKRFSLRELQVASDNFSNKNILGRGGFGKVYKGRLADGTLVAVKRLKEERTQGGEL  329  Q94F62
285  NLGNVKRFSFRELQAATEGFSGKNILGKGGFGNVYRGQLPDGTLVAVKRLKDGNAAGGEA  344  Q9FP13
      **:::** ::*** *:. ** **::*:**:*:**:* * *.:::******:    ***


353  QFQTEVEMISLAVHRNLLRLYGFCITQTEKLLVYPYMSNGSVASRMK----AKPVLDWSI  408  Q9LFS4
353  QFQTELEMISLAVHRNLLRLYGFCTTSSERLLVYPYMSNGSVASRLK----AKPVLDWGT  408  Q8RY65
330  QFQTEVEMISMAVHRNLLRLRGFCMTPTERLLVYPYMANGSVASCLRERPESQPPLDWPK  389  Q94F62
345  QFQTEVEMISLALHRNLLRLYGFCMTATERLLVYPFMSNGSVASRLK----AKPALEWGT  400  Q9FP13
     *****:****:*:******* *** * :*:*****:*:****** ::    ::* *:*


409  RKRIAIGAARGLVYLHEQCDPKIIHRDVKAANILLDDYCEAVVGDFGLAKLLDHQDSHVT  468  Q9LFS4
409  RKRIALGAGRGLLYLHEQCDPKIIHRDVKAANILLDDYFEAVVGDFGLAKLLDHEESHVT  468  Q8RY65
390  RQRIALGSARGLAYLHDHCDPKIIHRDVKAANILLDEEFEAVVGDFGLAKLMDYKDTHVT  449  Q94F62
401  RRRIAVGAARGLVYLHEQCDPKIIHRDVKAANVLLDEACEAVVGDFGLAKLLDHRESHVT  460  Q9FP13
     *:***:*:.*** ***::**************:***:  ************:*:.::***
```

绘制螺旋轮

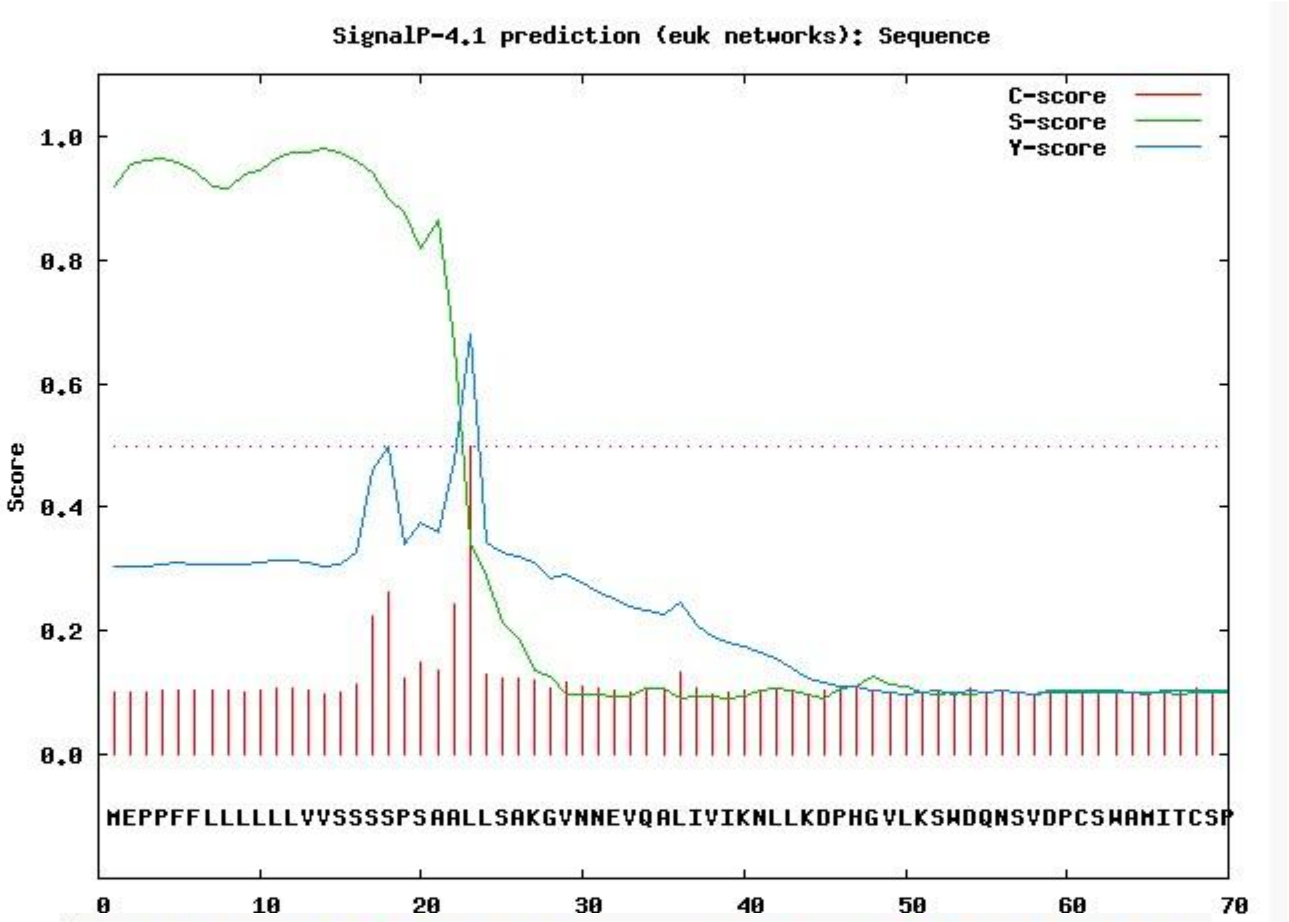
PEPWHEEL

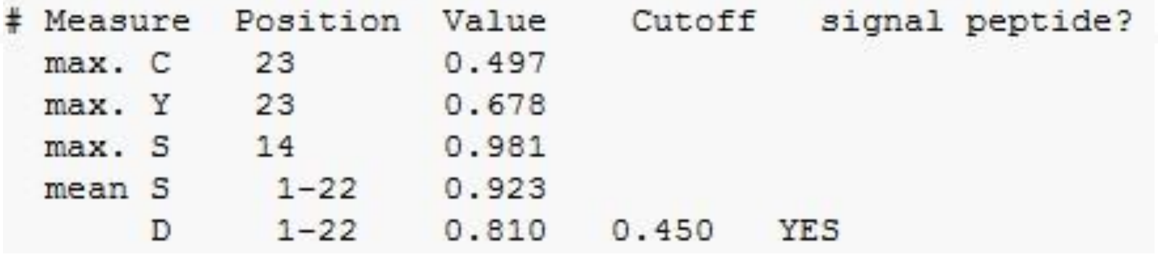
Draw a helical wheel diagram for a protein sequence



信号肽预测

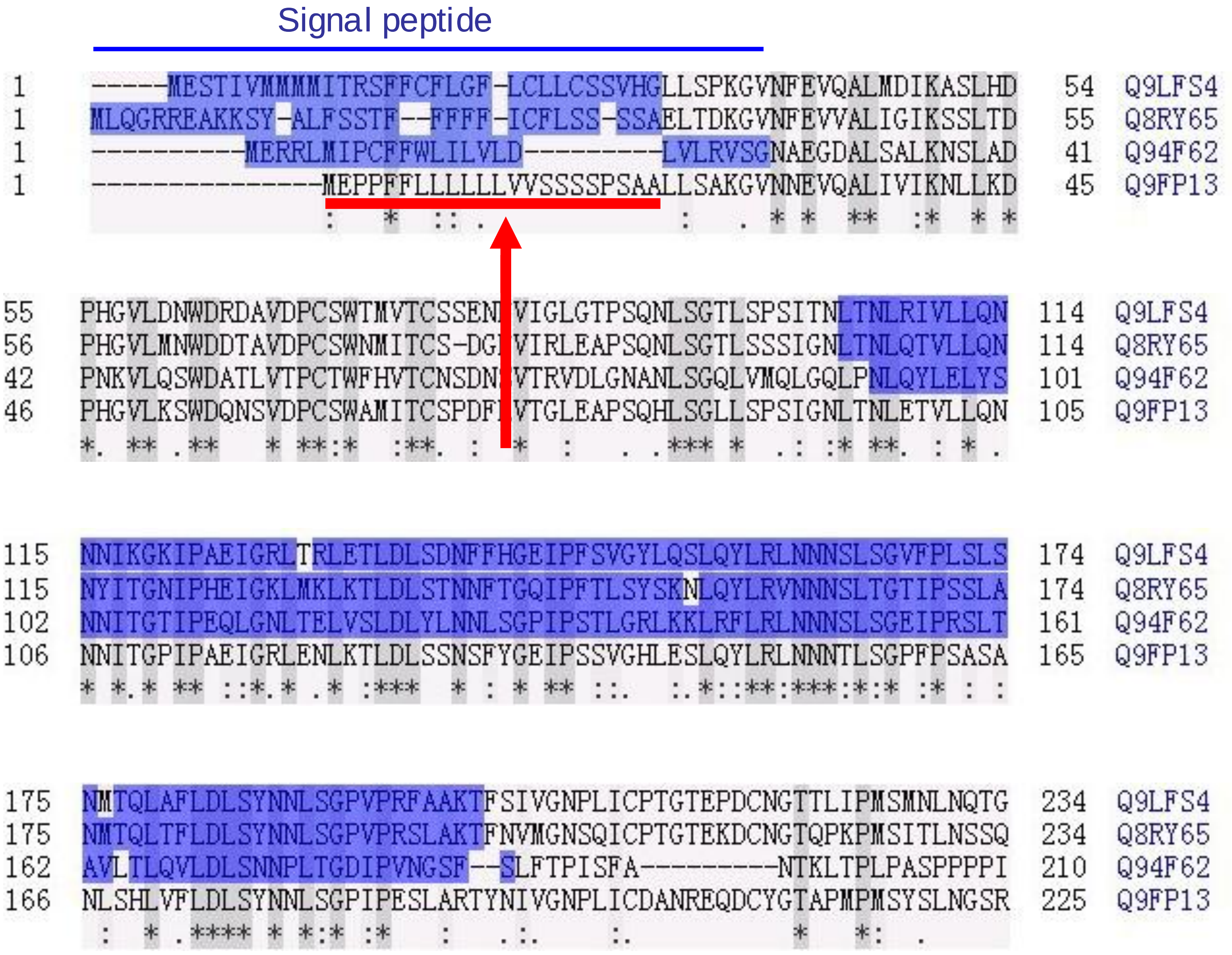
CBS >> CBS Prediction Servers >> SignalP >> Output format





```
# Measure  Position  Value    Cutoff   signal peptide?
  max. C    23       0.497
  max. Y    23       0.678
  max. S    14       0.981
  mean S     1-22    0.810
       D     1-22    0.810   0.450    YES
```

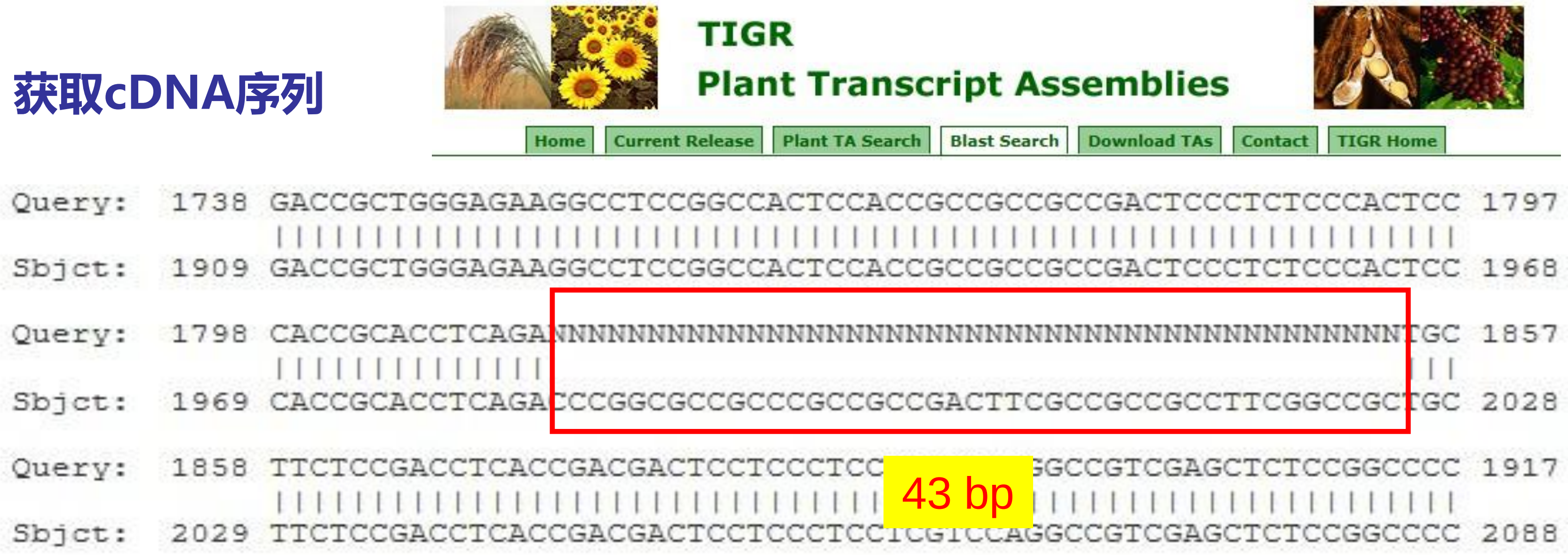
信号肽预测



Signal peptide

```
1    -----MESTIVMMMMITRSFFCFLGF-LCLLCSSVHGLLSPKGVNFEVQALMDIKASLHD  54  Q9LFS4
1    MLQGRREAKKSY-ALFSSTF--FFFF-ICFLSS-SSAELTDKGVNFEVVALIGIKSSLTD  55  Q8RY65
1    ----------MERRLMIPCFFWLILVLD---------LVLRVSGNAEGDALSALKNSLAD  41  Q94F62
1    ---------------MEPPFFLLLLLLVVSSSSPSAALLSAKGVNNEVQALIVIKNLLKD  45  Q9FP13
                    :   *  :: .            :   . * *  **  :*  * *

55   PHGVLDNWDRDAVDPCSWTMVTCSSENLVIGLGTPSQNLSGTLSPSITNLTNLRIVLLQN  114 Q9LFS4
56   PHGVLMNWDDTAVDPCSWNMITCS-DGLVIRLEAPSQNLSGTLSSSIGNLTNLQTVLLQN  114 Q8RY65
42   PNKVLQSWDATLVTPCTWFHVTCNSDNSVTRVDLGNANLSGQLVMQLGQLPNLQYLELYS  101 Q94F62
46   PHGVLKSWDQNSVDPCSWAMITCSPDFLVTGLEAPSQHLSGLLSPSIGNLTNLETVLLQN  105 Q9FP13
     *. ** .**   * **:*  :**. :  *  :   . .*** *  .: :* **. : * .

115  NNIKGKIPAEIGRLTRLETLDLSDNFFHGEIPFSVGYLQSLQYLRLNNNSLSGVFPLSLS  174 Q9LFS4
115  NYITGNIPHEIGKLMKLKTLDLSTNNFTGQIPFTLSYSKNLQYLRVNNNSLTGTIPSSLA  174 Q8RY65
102  NNITGTIPEQLGNLTELVSLDLYLNNLSGPIPSTLGRLKKLRFLRLNNNSLSGEIPRSLT  161 Q94F62
106  NNITGPIPAEIGRLENLKTLDLSSNSFYGEIPSSVGHLESLQYLRLNNNTLSGPFPSASA  165 Q9FP13
     * *.* ** ::*.* .* :***  * : * ** ::.  :.*::**:***:*:* :* : :

175  NMTQLAFLDLSYNNLSGPVPRFAAKTFSIVGNPLICPTGTEPDCNGTTLIPMSMNLNQTG  234 Q9LFS4
175  NMTQLTFLDLSYNNLSGPVPRSLAKTFNVMGNSQICPTGTEKDCNGTQPKPMSITLNSSQ  234 Q8RY65
162  AVLTLQVLDLSNNPLTGDIPVNGSF--SLFTPISFA---------NTKLTPLPASPPPPI  210 Q94F62
166  NLSHLVFLDLSYNNLSGPIPESLARTYNIVGNPLICDANREQDCYGTAPMPMSYSLNGSR  225 Q9FP13
      :  * .**** * *:* :*   :   .:.    :.          *   *:  .
```

获取cDNA序列

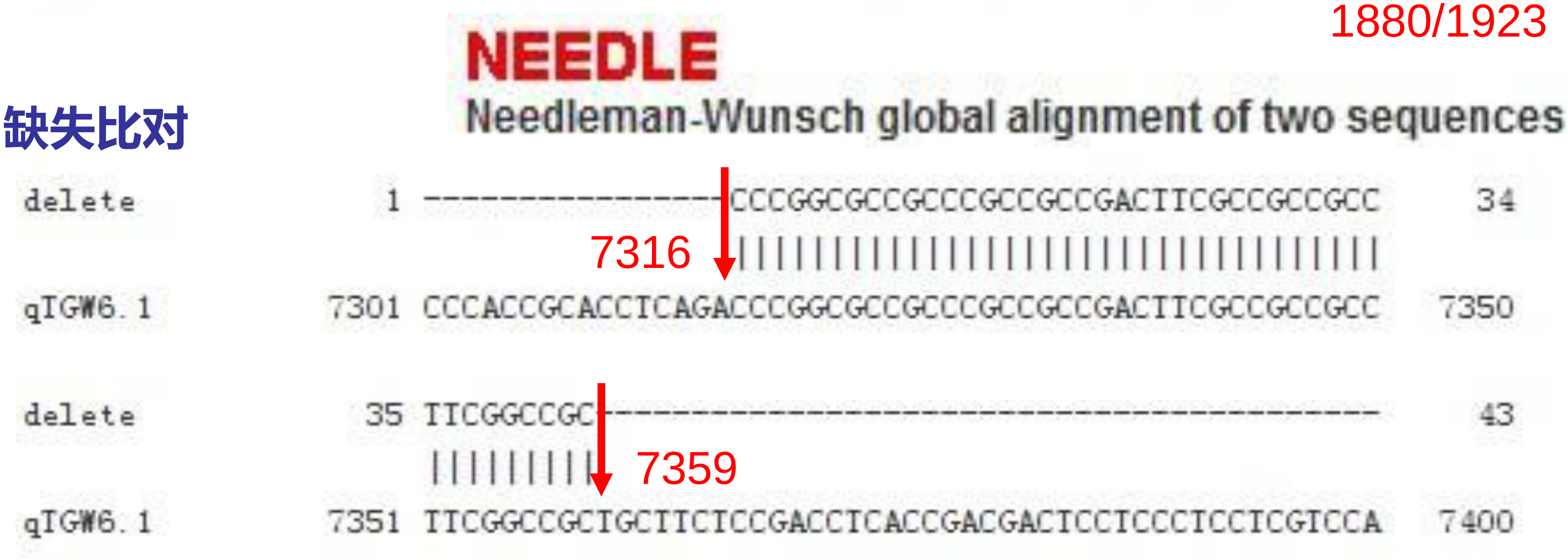


```
Query:  1738 GACCGCTGGGAGAAGGCCTCCGGCCACTCCACCGCCGCCGCCGACTCCCTCTCCCACTCC 1797
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1909 GACCGCTGGGAGAAGGCCTCCGGCCACTCCACCGCCGCCGCCGACTCCCTCTCCCACTCC 1968

Query:  1798 CACCGCACCTCAGANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTGC 1857
             ||||||||||||||                                           |||
Sbjct:  1969 CACCGCACCTCAGACCCGGCGCCGCCCGCCGCCGACTTCGCCGCCGCCTTCGGCCGCTGC 2028

Query:  1858 TTCTCCGACCTCACCGACGACTCCTCCCTCC        GGCCGTCGAGCTCTCCGGCCCC 1917
             |||||||||||||||||||||||||||||||        ||||||||||||||||||||||
Sbjct:  2029 TTCTCCGACCTCACCGACGACTCCTCCCTCCTCGTCCAGGCCGTCGAGCTCTCCGGCCCC 2088
```

43 bp

1880/1923

缺失比对



NEEDLE

Needleman-Wunsch global alignment of two sequences

```
delete       1 ----------------CCCGGCGCCGCCCGCCGCCGACTTCGCCGCCGCC     34
                               ||||||||||||||||||||||||||||||||||
qIGW6.1   7301 CCCACCGCACCTCAGACCCGGCGCCGCCCGCCGCCGACTTCGCCGCCGCC   7350

delete      35 TTCGGCCGC-----------------------------------------     43
               |||||||||
qIGW6.1   7351 TTCGGCCGCTGCTTCTCCGACCTCACCGACGACTCCTCCCTCCTCGTCCA   7400
```

7316

7359

小结
Summary

- ➢Softberry 等基因预测 12个
- ➢Blast 可能性最高的候选基因 BR
- ➢Blast, 序列比对, 系统发育树构建, 蛋白结构预测等进一步确认候选基因
- ➢TIGR 获取cDNA为下一步实验提供参照

感谢罗老师孜孜教诲

感谢奉保华和欧承刚

感谢生物信息班所有同学们

谢谢