谷子K3Y762基因结构分析与功能挖掘

任课教师：罗静初 教 授
小组成员：胡利芹 尹丽娟 李珊珊 杜丽媛
报告人： 胡利芹
报告时间：2013年1月20日

中国农业科学院2012级生物信息学硕士班

汇报内容

一、研究背景

二、NAC家族蛋白结构与功能简介

三、K3Y762序列、结构分析

四、核定位信号和亚细胞定位分析

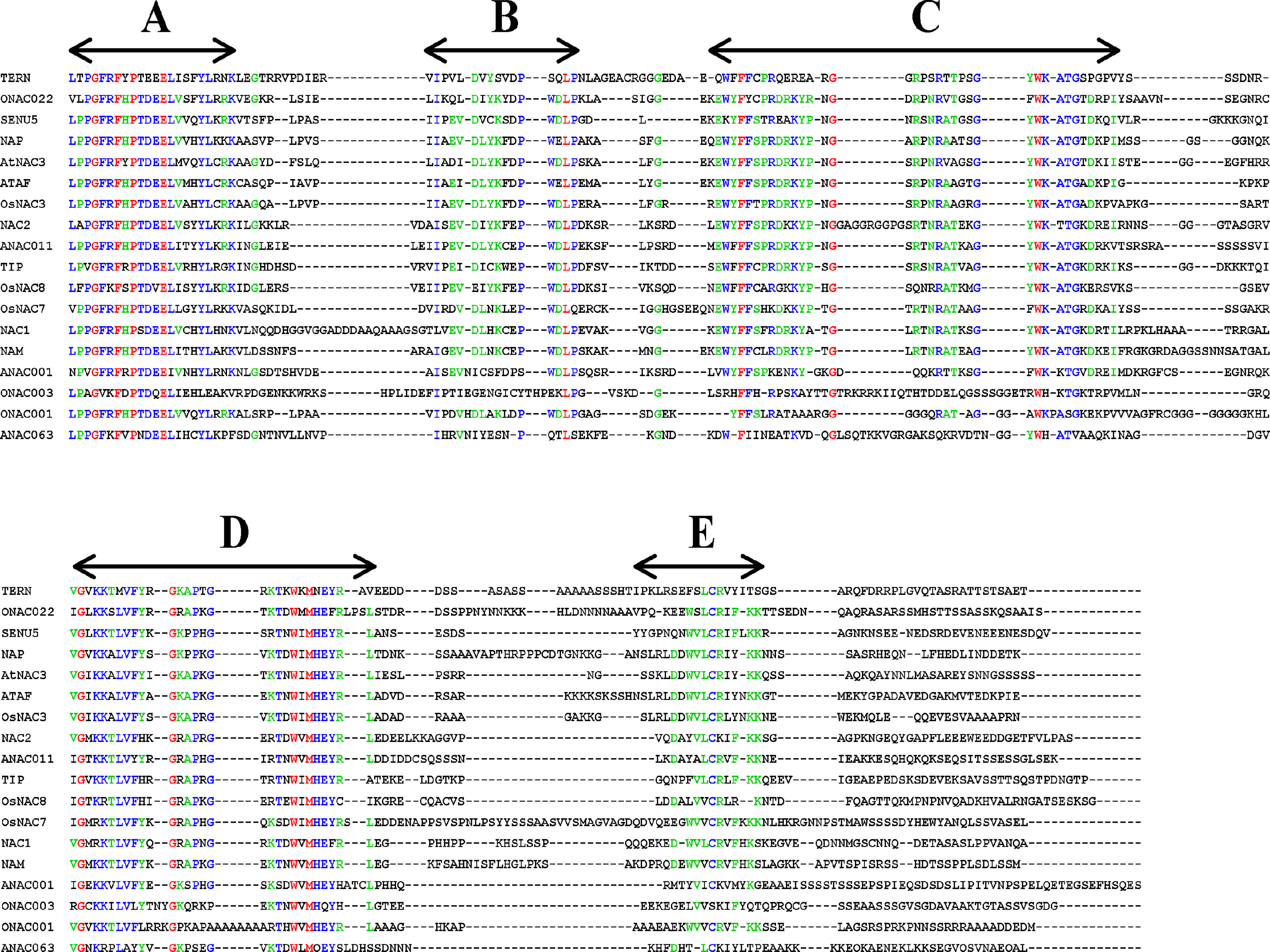
五、

研究背景

转录因子(Transcription factor)是一类调节基因表达水平上的重要调控基因，通过与靶标基因启动子中特定的DNA序列结合，激活或抑制靶标基因的转录表达。NAC转录因子是植物特有的一类转录因子,最早在矮牵牛(*Pharbifis nil*)*NAM*、拟南芥(*Ara-bidopsis thaliana*)*ATAF1/2*和*CUC2*基因编码蛋白的N端中发现一段高度保守的氨基酸序列,由约150个氨基酸组成的NAC结构域,并命名为NAC结构域,并将包含NAC结构域的蛋白称为NAC转录因子，而C端为高度变异的转录调控区。NAC结构域一般由约150个氨基酸组成，可分为A、B、C、D、E等5个亚结构域，而最初鉴定的NAM结构域则由A、B、C、D亚结构域组成，其中A、C、D 3个亚结构域在不同的物种中都是高度保守的，C、D亚结构域中含有核定位信号，而B、E亚结构域的保守性不强。

研究表明，NAC转录因子不仅参与植物生长发育的调控，而且在植物抗逆反应中具有重要的调控作用。

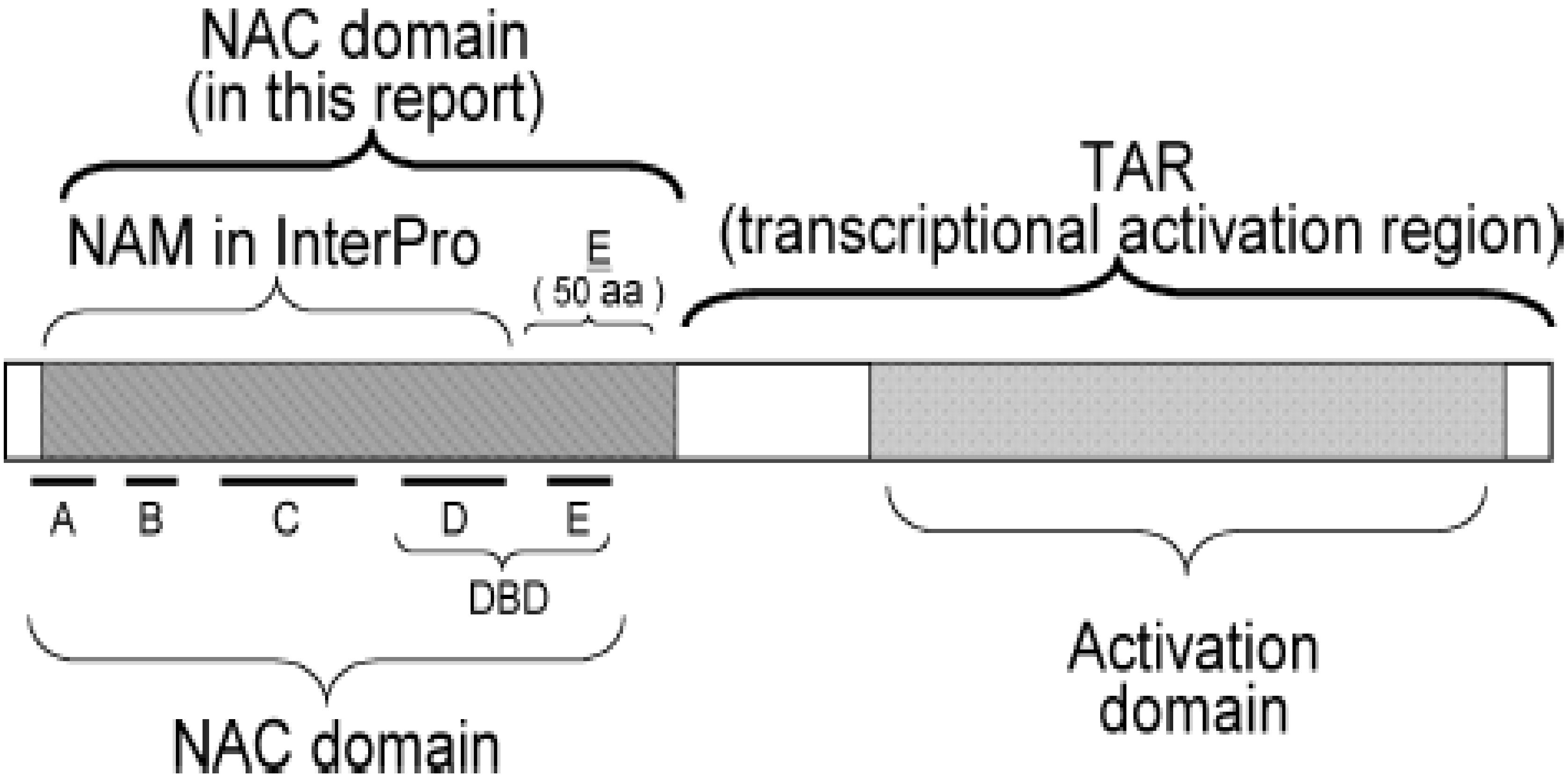
相关文献：Functions of NAC transcription factors in biotic and abiotic stress responses in plants
SUN Li-Jun, LI Da-Yong, ZHANG Hui-Juan, SONG Feng-Ming *State Key Laboratory for Rice Biology*, *Institute of Biotechnology*, *Zhejiang University*, *Hangzhou* 310029, *China*



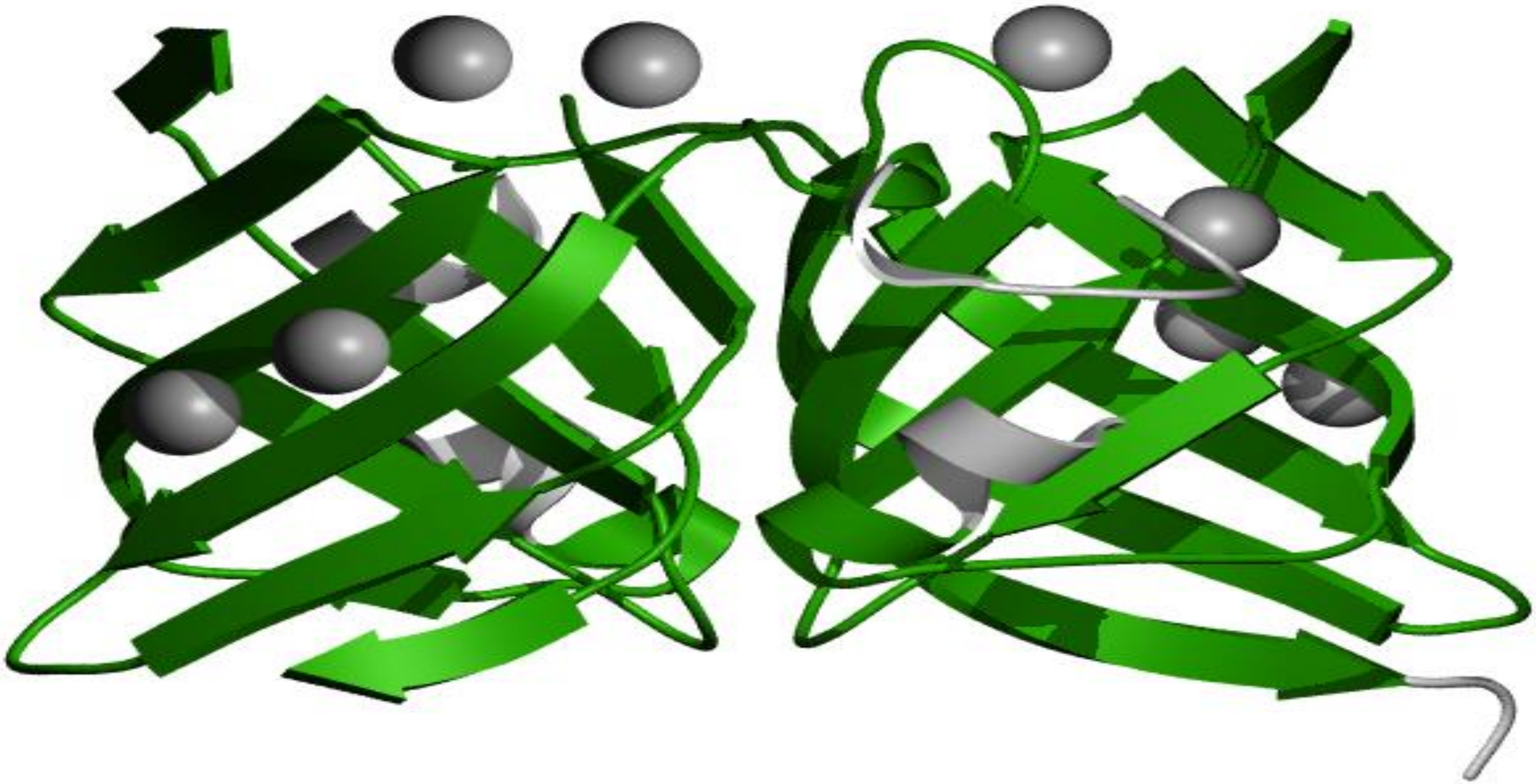
```
TERN      LTPGFRFYPTEEELISFYLRNKLEGTRRVPDIER-------------VIPVL-DVYSVDP---SQLPNLAGEACRGGGEDA--E-QWFFFCPRQEREA-RG---------GRPSRTTPSG------YWK-ATGSPGPVYS------------SSDNR-
ONAC022   VLPGFRFHPTDEELVSFYLRRKVEGKR--LSIE--------------LIKQL-DIYKYDP---WDLPKLA----SIGG-----EKEWYFYCPRDRKYR-NG---------DRPNRVTGSG------FWK-ATGTDRPIYSAAVN--------SEGNRC
SENU5     LPPGFRFHPTDEELVVQYLKRKVTSFP--LPAS--------------IIPEV-DVCKSDP---WDLPGD------L-------EKEKYFFSTREAKYP-NG---------NRSNRATGSG------YWK-ATGIDKQIVLR---------GKKKGNQI
NAP       LPPGFRFHPTDEELVVHYLKKKAASVP--LPVS--------------IIAEV-DLYKFDP---WELPAKA-----SFG-----EQEWYFFSPRDRKYP-NG---------ARPNRAATSG------YWK-ATGTDKPIMSS------GS----GGNQK
AtNAC3    LPPGFRFYPTDEELMVQYLCRKAAGYD--FSLQ--------------LIADI-DLYKFDP---WDLPSKA-----LFG-----EKEWYFFSPRDRKYP-NG---------SRPNRVAGSG------YWK-ATGTDKIISTE-----GG----EGFHRR
ATAF      LPPGFRFHPTDEELVMHYLCRKCASQP--IAVP--------------IIAEI-DLYKFDP---WELPEMA-----LYG-----EKEWYFFSPRDRKYP-NG---------SRPNRAAGTG------YWK-ATGADKPIG---------------KPKP
OsNAC3    LPPGFRFHPTDEELVAHYLCRKAAGQA--LPVP--------------IIAEV-DLYKFDP---WDLPERA-----LFGR-----REWYFFTPRDRKYP-NG---------SRPNRAAGRG------YWK-ATGADKPVAPKG------------SART
NAC2      LAPGFRFHPTDEELVSYYLKRKILGKKLR----------------VDAISEV-DIYKFEP---WDLPDKSR----LKSRD----LEWYFFSPRDKKYP-NGGAGGRGGPGSRTNRATEKG------YWK-TTGKDREIRNNS----GG---GTASGRV
ANAC011   LPPGFRFHPTDEELITYYLKRKINGLEIE----------------LEIIPEV-DLYKCEP---WDLPEKSF----LPSRD----MEWFFFSPRDRKYP-NG---------SRTNRATKAG------YWK-ATGKDRKVTSRSRA-------SSSSSVI
TIP       LPVGFRFRPTDEELVRHYLRGKINGHDHSD---------------VRVIPEI-DICKWEP---WDLPDFSV----IKTDD----SEWFFFCPRDRKYP-SG---------SRSNRATVAG------YWK-ATGKDRKIKS------GG---DKKKTQI
OsNAC8    LFPGFKFSPTDVELISYYLKRKIDGLERS----------------VEIIPEV-EIYKFEP---WDLPDKSI----VKSQD----NEWFFFCARGKKYP-HG---------SQNRRATKMG------YWK-ATGKERSVKS--------------GSEV
OsNAC7    VPPGFRFHPTDEELLGYYLRKKVASQKIDL----------------DVIRDV-DLNKLEP---WDLQERCK----IGGHGSEEQNEWYFFSHKDKKYP-TG---------TRTNRATAAG------FWK-ATGRDKAIYSS-----------SSGAKR
NAC1      LPPGFRFHPSDEELVCHYLHNKVLNQQDHGGVGGADDDAAQAAAGSGTLVEV-DLHKCEP---WDLPEVAK----VGG------KEWYFFSFRDRKYA-TG---------LRTNRATKSG------YWK-ATGKDRTILRPKLHAAA-----TRRGAL
NAM       LPPGFRFHPTDEELITHYLAKKVLDSSNFS---------------ARAIGEV-DLNKCEP---WDLPSKAK----MNG-----EKEWYFFCLRDRKYP-TG---------LRTNRATEAG------YWK-ATGKDKEIFRGKGRDAGGSSNNSATGAL
ANAC001   NPVGFRFRPTDEEIVNHYLRNKNLGSDTSHVDE--------------AISEVNICSFDPS---WDLPSQSR----IKSRD----LVWYFFSPKENKY-GKGD---------QQKRTTKSG------FWK-KTGVDREIMDKRGFCS------EGNRQK
ONAC003   LPAGVKFDPTDQELIEHLEAKVRPDGENKKWRKS-------HPLIDEFIPTIEGENGICYTHPEKLPG---VSKD--G------LSRHFFH-RPSKAYTTGTRKRRKIIQTHTDDELQGSSSGGETRWH-KTGKTRPVMLN--------------GRQ
ONAC001   LPPGFRFHPTDEELVVQYLRRKALSRP--LPAA--------------VIPDVHDLAKLDP---WDLPGAG-----SDGEK-------YFFSLRATAAARGG---------GGGQRAT-AG--GG--AWKPASGKEKPVVVAGFRCGGG--GGGGGKHL
ANAC063   LPPGFKFVPNDEELIHCYLKPFSDGNTNVLLNVP--------------IHRVNIYESN-P---QTLSEKFE-----KGND----KDW-FIINEATKVD-QGLSQTKKVGRGAKSQKRVDTN-GG--YWH-ATVAAQKINAG--------------DGV
```

```
TERN      VGVKKTMVFYR--GKAPTG------RKTKWKMNEYR--AVEEDD----DSS----ASASS----AAAAASSSHTIPKLRSEFSLCRVYITSGS--------ARQFDRRPLGVQTASRATTSTSAET---------------
ONAC022   IGLKKSLVFYR--GRAPKG------TKTDWMMHEFRLPSLSTDR----DSSPPNYNNKKK----HLDNNNNAAAVPQ-KEEWSLCRIF-KKTTSEDN----QAQRASARSSMHSTTSSASSKQSAAIS-------------
SENU5     VGLKKTLVFYK--GKPPHG------SRTNWIMHEYR---LANS-----ESDS----------------------YYGPNQNWVLCRIFLKKR---------AGNKNSEE-NEDSRDEVENEEENESDQV------------
NAP       VGVKKALVFYS--GKPPKG------VKTDWIMHEYR---LTDNK----SSAAAVAPTHRPPPCDTGNKKG---ANSLRLDDWVLCRIY-KKNNS--------SASRHEQN--LFHEDLINDDETK----------------
AtNAC3    VGIKKALVFYI--GKAPKG------TKTNWIMHEYR---LIESL----PSRR----------------NG-----SSKLDDWVLCRIY-KKQSS--------AQKQAYNNLMASAREYSNNGSSSSS--------------
ATAF      VGIKKALVFYA--GKAPKG------EKTNWIMHEYR---LADVD----RSAR-------------KKKKSKSSHNSLRLDDWVLCRIYNKKGT--------MEKYGPADAVEDGAKMVTEDKPIE----------------
OsNAC3    VGIKKALVFYS--GKAPRG------VKTDWIMHEYR---LADAD----RAAA-------------GAKKG-----SLRLDDWVLCRLYNKKNE--------WEKMQLE---QQEVESVAAAAPRN----------------
NAC2      VGMKKTLVFHK--GRAPRG------ERTDWVMHEYR---LEDEELKKAGGVP-------------------------VQDAYVLCKIF-KKSG--------AGPKNGEQYGAPFLEEEWEEDDGETFVLPAS---------
ANAC011   IGTKKTLVYYR--GRAPHG------IRTNWVMHEYR---LDDIDDCSQSSSN-------------------------LKDAYALCRVF-KKNE--------IEAKKESQHQKQKSEQSITSSESSGLSEK-----------
TIP       IGVKKTLVFHR--GRAPRG------TRTNWIMHEYR---ATEKE--LDGTKP-------------------------GQNPFVLCRLF-KKQEEV------IGEAEPEDSKSDEVEKSAVSSTTSQSTPDNGTP-------
OsNAC8    IGTKRTLVFHI--GRAPKG------ERTEWIMHEYC---IKGRE--CQACVS-------------------------LDDALVVCRLR--KNTD--------FQAGTTQKMPNPNVQADKHVALRNGATSESKSG------
OsNAC7    IGMRKTLVFYK--GRAPHG------QKSDWIMHEYRS--LEDDENAPPSVSPNLPSYYSSSAASVVSMAGVAGDQDVQEEGWVVCRVFKKKNLHKRGNNPSTMAWSSSSDYHEWYANQLSSVASEL---------------
NAC1      VGMRKTLVFYQ--GRAPNG------RKTDWVMHEFR---LEG-----PHHPP----KHSLSSP----------QQQEKED-WVLCRVFHKSKEGVE--QDNNMGSCNNQ--DETASASLPPVANQA---------------
NAM       VGMKKTLVFYK--GRAPKG------EKTNWVMHEYR---LEG-----KFSAHNISFLHGLPKS----------AKDPRQDEWVVCRVFHKSLAGKK--APVTSPISRSS--HDTSSPPLSDLSSM----------------
ANAC001   IGEKKVLVFYE--GKSPHG------SKSDWVMHEYHATCLPHHQ----------------------------------RMTYVICKVMYKGEAAEISSSSTSSSEPSPIEQSDSDSLIPITVNPSPELQETEGSEFHSQES
ONAC003   RGCKKILVLYTNYGKQRKP------EKTNWVMHQYH---LGTEE-------------------------------EEKEGELVVSKIFYQTQPRQCG----SSEAASSSGVSGDAVAAKTGTASSVSGDG-----------
ONAC001   VGVKKTLVFLRRKGPKAPAAAAAAAARTHWVMHEYR---LAAAG----HKAP----------------------AAAEAEKWVVCRVF-KKSSE-------LAGSRSPRKPNNSSRRRAAAADDEDM--------------
ANAC063   VGNKRPLAYYV--GKPSEG------VKTDWLMQEYSLDHSSDNNN-------------------------------KHFDHT-LCKIYLTPEAAKK----KKEQKAENEKLKKSEGVQSVNAEQAL---------------
```

NAM and NAC domains and TAR in a NAC family protein.



Crystal structure of the NAC domain



NAC转录因子在植物抗逆反应中的作用

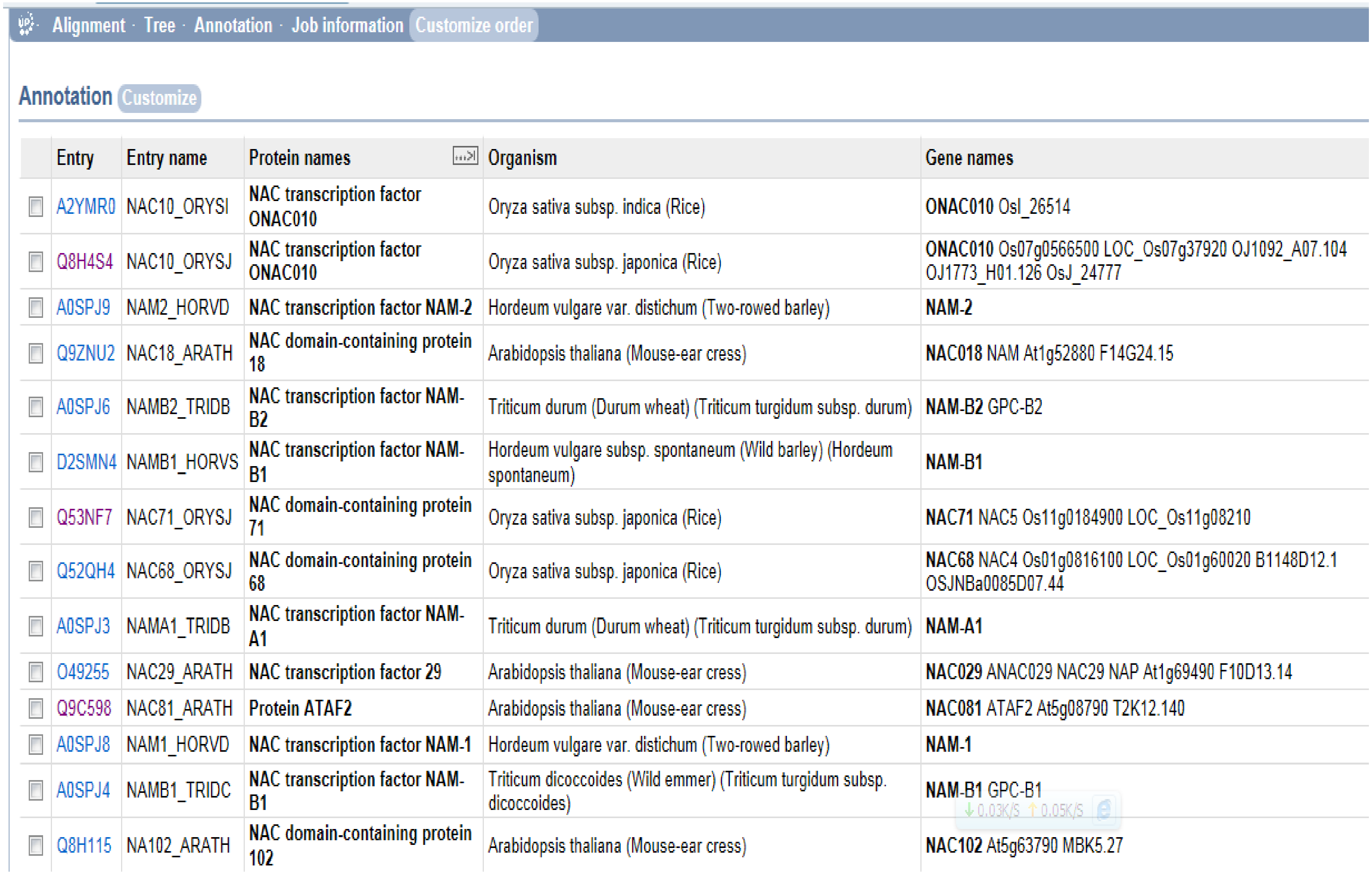
1、NAC转录因子在植物抗非生物逆境胁迫中的作用

干旱、高盐是影响植物生长发育的主要逆境因子。当植物处于干旱、高盐的胁迫下,植物细胞会感知外界的胁迫信号,通过一系列复杂的信号传导途径如ABA信号途径,把信号传递到胁迫应答的转录因子,由各类转录因子启动胁迫应答反应基因表达,从而激活植物抗逆反应,降低或消除干旱、高盐逆境所造成的伤害。研究发现,NAC转录因子直接参与或通过调控参与干旱、高盐应答基因的表达,在植物抗干旱、高盐等非生物逆境胁迫中起重要作用。

2、NAC转录因子在抗病反应中的作用

植物在其生长发育的环境中，经常会受到细菌、真菌和病毒等病原物的侵害，促使植物形成了一系列复杂的抗病机制。在受到病原菌的侵染时,植物通常通过水杨酸(Salicylic acid，SA)、茉莉酸(Jasmonic acid，JA)、乙烯(Ethylene，ET)等抗病信号传导途径来活化一大批防卫反应基因的协同表达，从而激活植物的抗病防卫反应。研究显示，AP2/ERF、bZIP、WRKY等不同类型的转录因子参与植物抗病反应中防卫基因的表达调控。近年来的研究表明，NAC转录因子家族基因在植物对不同类型病原菌的抗病反应中起重要作用。

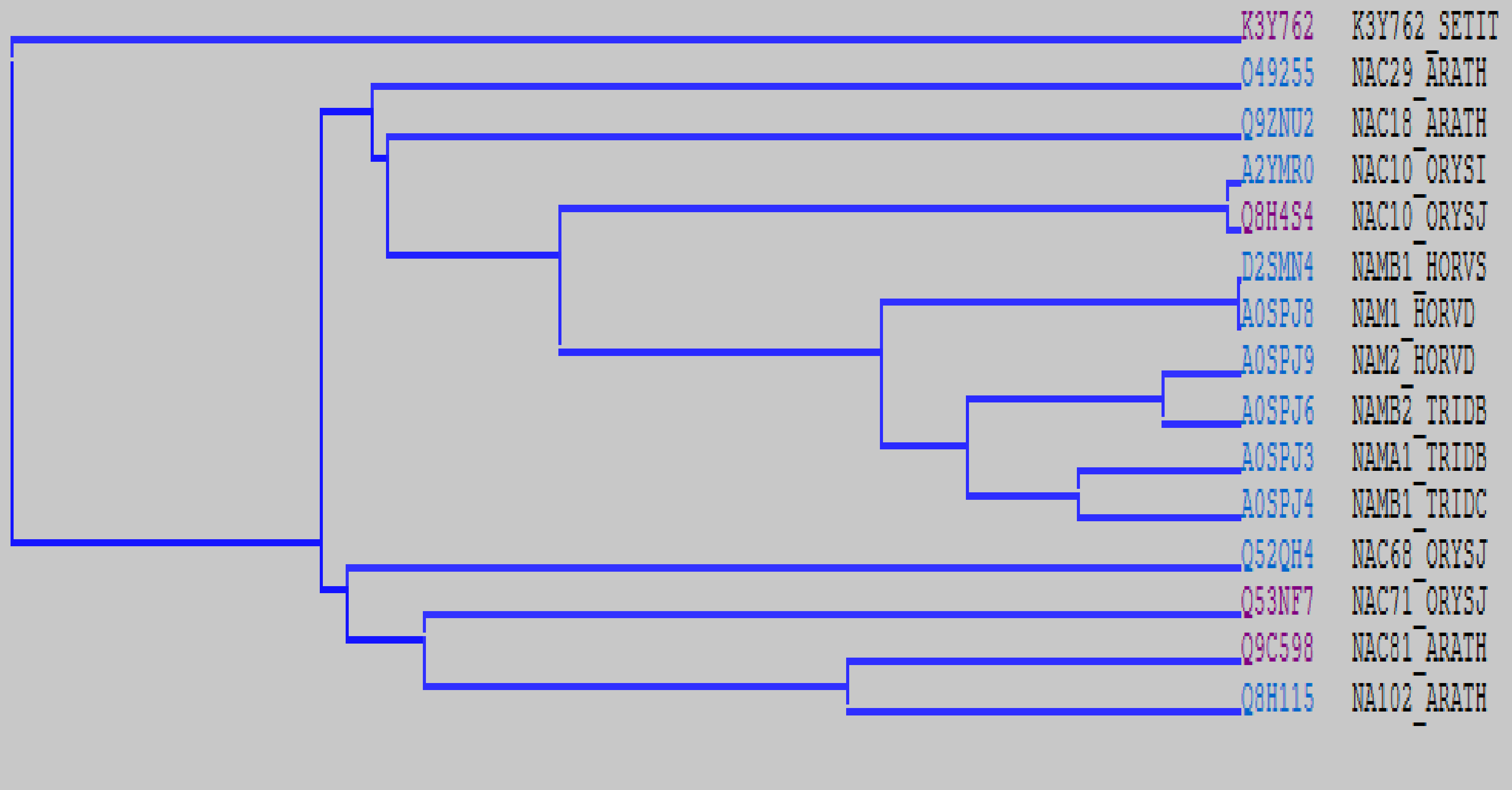
K3Y762序列

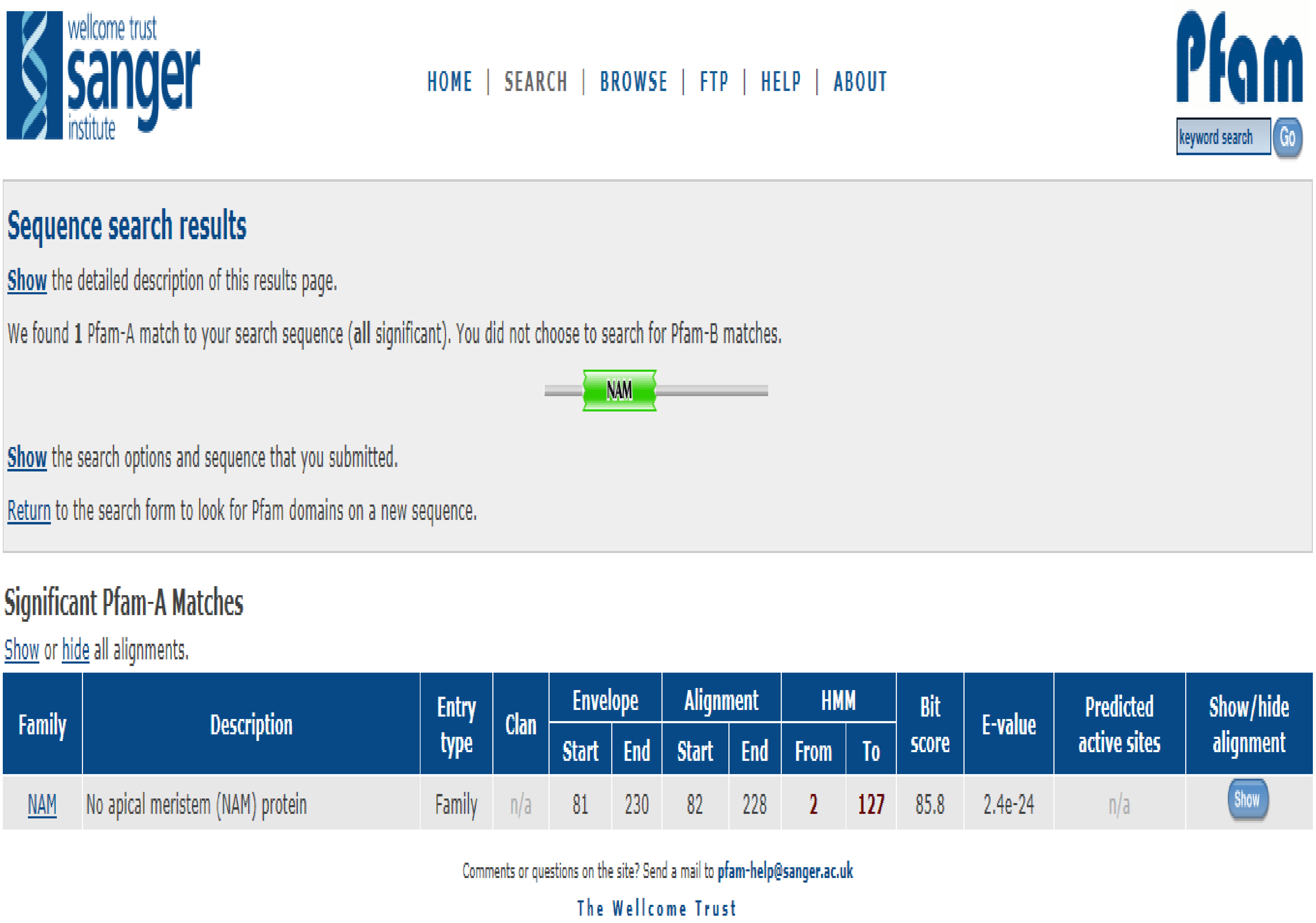


Annotation

	Entry	Entry name	Protein names	Organism	Gene names
☐	A2YMR0	NAC10_ORYSI	NAC transcription factor ONAC010	Oryza sativa subsp. indica (Rice)	**ONAC010** OsI_26514
☐	Q8H4S4	NAC10_ORYSJ	NAC transcription factor ONAC010	Oryza sativa subsp. japonica (Rice)	**ONAC010** Os07g0566500 LOC_Os07g37920 OJ1092_A07.104 OJ1773_H01.126 OsJ_24777
☐	A0SPJ9	NAM2_HORVD	NAC transcription factor NAM-2	Hordeum vulgare var. distichum (Two-rowed barley)	**NAM-2**
☐	Q9ZNU2	NAC18_ARATH	NAC domain-containing protein 18	Arabidopsis thaliana (Mouse-ear cress)	**NAC018** NAM At1g52880 F14G24.15
☐	A0SPJ6	NAMB2_TRIDB	NAC transcription factor NAM-B2	Triticum durum (Durum wheat) (Triticum turgidum subsp. durum)	**NAM-B2** GPC-B2
☐	D2SMN4	NAMB1_HORVS	NAC transcription factor NAM-B1	Hordeum vulgare subsp. spontaneum (Wild barley) (Hordeum spontaneum)	**NAM-B1**
☐	Q53NF7	NAC71_ORYSJ	NAC domain-containing protein 71	Oryza sativa subsp. japonica (Rice)	**NAC71** NAC5 Os11g0184900 LOC_Os11g08210
☐	Q52QH4	NAC68_ORYSJ	NAC domain-containing protein 68	Oryza sativa subsp. japonica (Rice)	**NAC68** NAC4 Os01g0816100 LOC_Os01g60020 B1148D12.1 OSJNBa0085D07.44
☐	A0SPJ3	NAMA1_TRIDB	NAC transcription factor NAM-A1	Triticum durum (Durum wheat) (Triticum turgidum subsp. durum)	**NAM-A1**
☐	O49255	NAC29_ARATH	NAC transcription factor 29	Arabidopsis thaliana (Mouse-ear cress)	**NAC029** ANAC029 NAC29 NAP At1g69490 F10D13.14
☐	Q9C598	NAC81_ARATH	Protein ATAF2	Arabidopsis thaliana (Mouse-ear cress)	**NAC081** ATAF2 At5g08790 T2K12.140
☐	A0SPJ8	NAM1_HORVD	NAC transcription factor NAM-1	Hordeum vulgare var. distichum (Two-rowed barley)	**NAM-1**
☐	A0SPJ4	NAMB1_TRIDC	NAC transcription factor NAM-B1	Triticum dicoccoides (Wild emmer) (Triticum turgidum subsp. dicoccoides)	**NAM-B1** GPC-B1
☐	Q8H115	NA102_ARATH	NAC domain-containing protein 102	Arabidopsis thaliana (Mouse-ear cress)	**NAC102** At5g63790 MBK5.27

进化树分析


K3Y762 K3Y762_SETIT
O49255 NAC29_ARATH
Q9ZNU2 NAC18_ARATH
A2YMR0 NAC10_ORYSI
Q8H4S4 NAC10_ORYSJ
D2SMN4 NAMB1_HORVS
A0SPJ8 NAM1_HORVD
A0SPJ9 NAM2_HORVD
A0SPJ6 NAMB2_TRIDB
A0SPJ3 NAMA1_TRIDB
A0SPJ4 NAMB1_TRIDC
Q52QH4 NAC68_ORYSJ
Q53NF7 NAC71_ORYSJ
Q9C598 NAC81_ARATH
Q8H115 NA102_ARATH



Sequence search results

Show the detailed description of this results page.

We found **1** Pfam-A match to your search sequence (**all** significant). You did not choose to search for Pfam-B matches.

Show the search options and sequence that you submitted.

Return to the search form to look for Pfam domains on a new sequence.

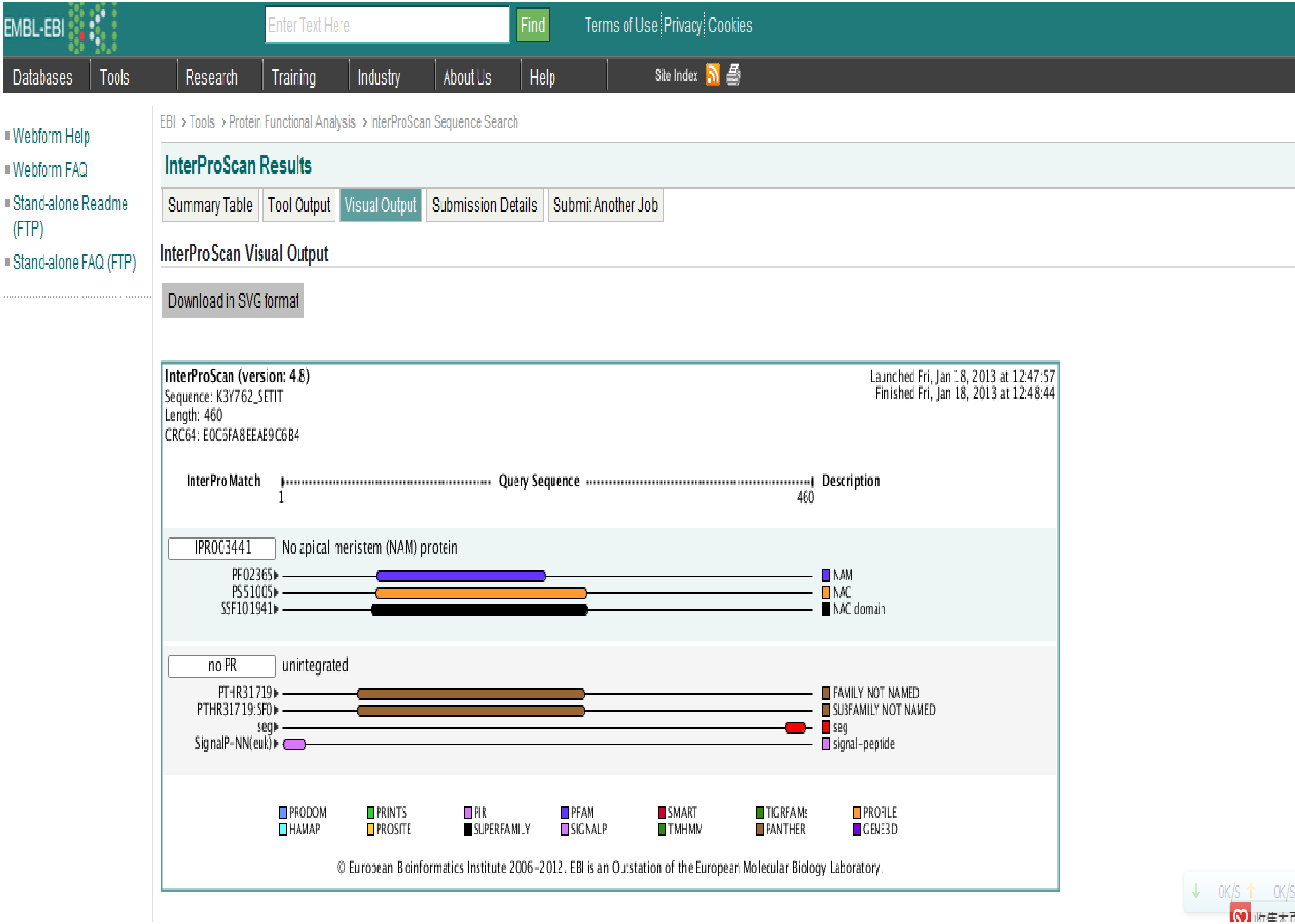
Significant Pfam-A Matches

Show or hide all alignments.

Family	Description	Entry type	Clan	Envelope		Alignment		HMM		Bit score	E-value	Predicted active sites	Show/hide alignment
				Start	End	Start	End	From	To				
NAM	No apical meristem (NAM) protein	Family	n/a	81	230	82	228	2	127	85.8	2.4e-24	n/a	Show

Comments or questions on the site? Send a mail to **pfam-help@sanger.ac.uk**

The Wellcome Trust



InterProScan Results

InterProScan Visual Output

InterProScan (version: 4.8)
Sequence: K3Y762_SETIT
Length: 460
CRC64: E0C6FA8EEAB9C6B4

Launched Fri, Jan 18, 2013 at 12:47:57
Finished Fri, Jan 18, 2013 at 12:48:44

IPR003441 No apical meristem (NAM) protein

- PF02365 — NAM
- PS51005 — NAC
- SSF101941 — NAC domain

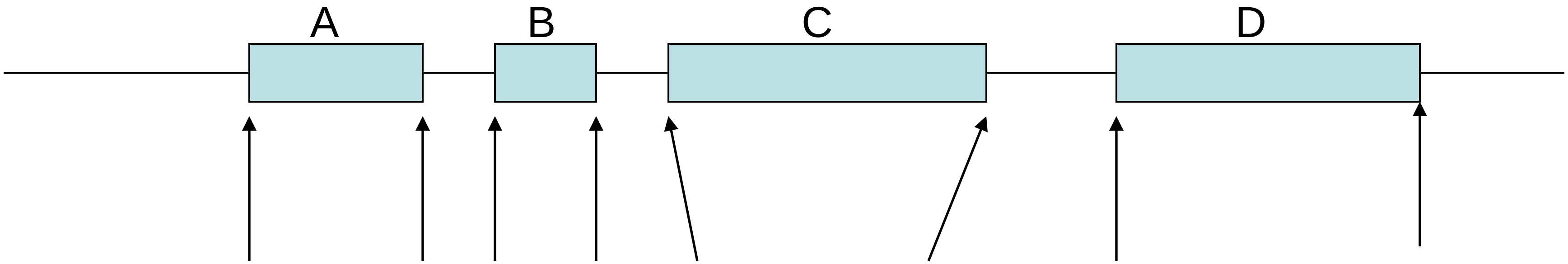
noIPR unintegrated

- PTHR31719 — FAMILY NOT NAMED
- PTHR31719:SF0 — SUBFAMILY NOT NAMED
- seg — seg
- SignalP-NN(euk) — signal-peptide

PRODOM	PRINTS	PIR	PFAM	SMART	TIGRFAMs	PROFILE
HAMAP	PROSITE	SUPERFAMILY	SIGNALP	TMHMM	PANTHER	GENE3D

© European Bioinformatics Institute 2006-2012. EBI is an Outstation of the European Molecular Biology Laboratory.

K3Y762结构预测分析



AA序列位点 81 101 115 126 139 175 196 225

其中A结构域中最保守位点：G(83)、F(86)、P(88)、E(92)

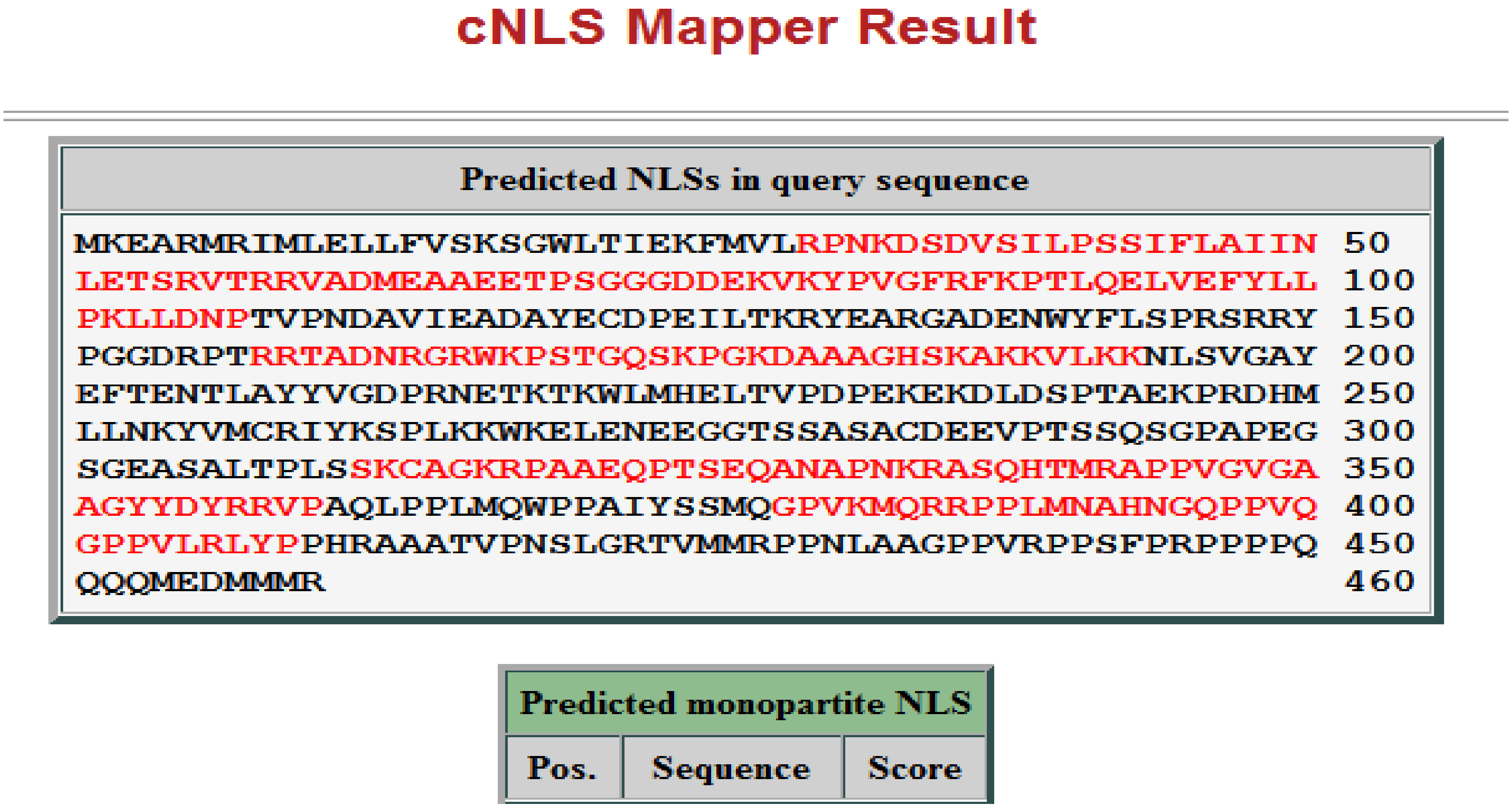
B结构域中最保守位点: L(126)

C结构域中最保守位点: F(141)、G(152)、W(166)

D结构域中最保守位点: G(197)、G(211)、W(166)、M(223)

核定位信号预测

http://nlsmapper.iab.keio.ac.jp/cgibin/NLS_Mapper_form.cgi

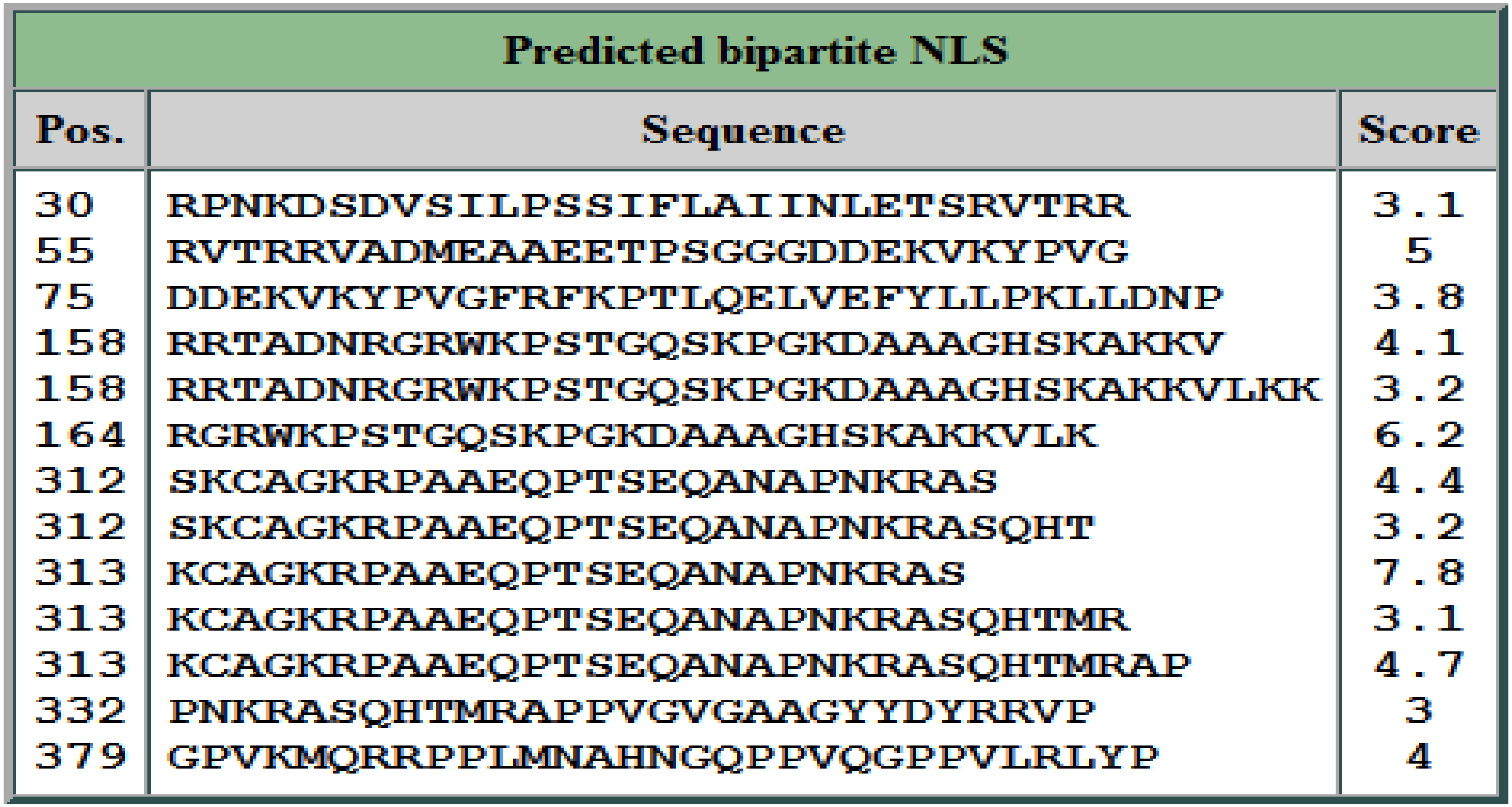


cNLS Mapper Result

Predicted NLSs in query sequence

```
MKEARMRIMLELLFVSKSGWLTIEKFMVLRPNKDSDVSILPSSIFLAIIN  50
LETSRVTRRVADMEAAEETPSGGGDDEKVKYPVGFRFKPTLQELVEFYLL 100
PKLLDNPTVPNDAVIEADAYECDPEILTKRYEARGADENWYFLSPRSRRY 150
PGGDRPTRRTADNRGRWKPSTGQSKPGKDAAAGHSKAKKVLKKNLSVGAY 200
EFTENTLAYYVGDPRNETKTKWLMHELTVPDPEKEKDLDSPTAEKPRDHM 250
LLNKYVMCRIYKSPLKKWKELENEEGGTSSASACDEEVPTSSQSGPAPEG 300
SGEASALTPLSSKCAGKRPAAEQPTSEQANAPNKRASQHTMRAPPVGVGA 350
AGYYDYRRVPAQLPPLMQWPPAIYSSMQGPVKMQRRPPLMNAHNGQPPVQ 400
GPPVLRLYPPHRAAATVPNSLGRTVMMRPPNLAAGPPVRPPSFPRPPPPQ 450
QQQMEDMMMR                                         460
```

Predicted monopartite NLS		
Pos.	**Sequence**	**Score**

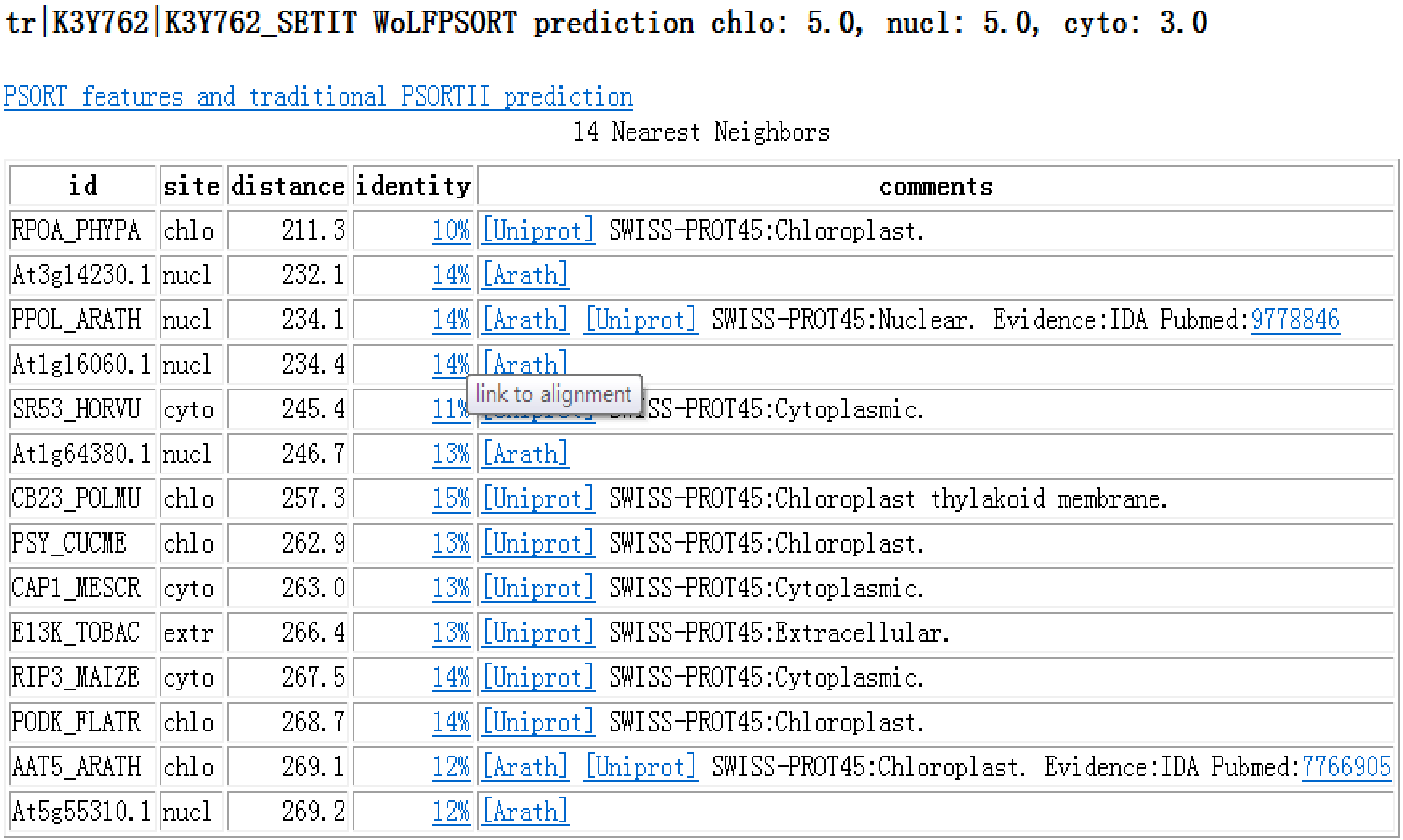


Predicted bipartite NLS

Pos.	Sequence	Score
30	RPNKDSDVSILPSSIFLAIINLETSRVTRR	3.1
55	RVTRRVADMEAAEETPSGGGDDEKVKYPVG	5
75	DDEKVKYPVGFRFKPTLQELVEFYLLPKLLDNP	3.8
158	RRTADNRGRWKPSTGQSKPGKDAAAGHSKAKKV	4.1
158	RRTADNRGRWKPSTGQSKPGKDAAAGHSKAKKVLKK	3.2
164	RGRWKPSTGQSKPGKDAAAGHSKAKKVLK	6.2
312	SKCAGKRPAAEQPTSEQANAPNKRAS	4.4
312	SKCAGKRPAAEQPTSEQANAPNKRASQHT	3.2
313	KCAGKRPAAEQPTSEQANAPNKRAS	7.8
313	KCAGKRPAAEQPTSEQANAPNKRASQHTMR	3.1
313	KCAGKRPAAEQPTSEQANAPNKRASQHTMRAP	4.7
332	PNKRASQHTMRAPPVGVGAAGYYDYRRVP	3
379	GPVKMQRRPPLMNAHNGQPPVQGPPVLRLYP	4

核定位信号序列：**RGRWKPSTGQSKPGKDAAAGHSKAKKVLK** （6.2,160-190AA） **KCAGKRPAAEQPTSEQANAPNKRAS** （7.8，>310AA）

亚细胞定位
http://wolfpsort.org/



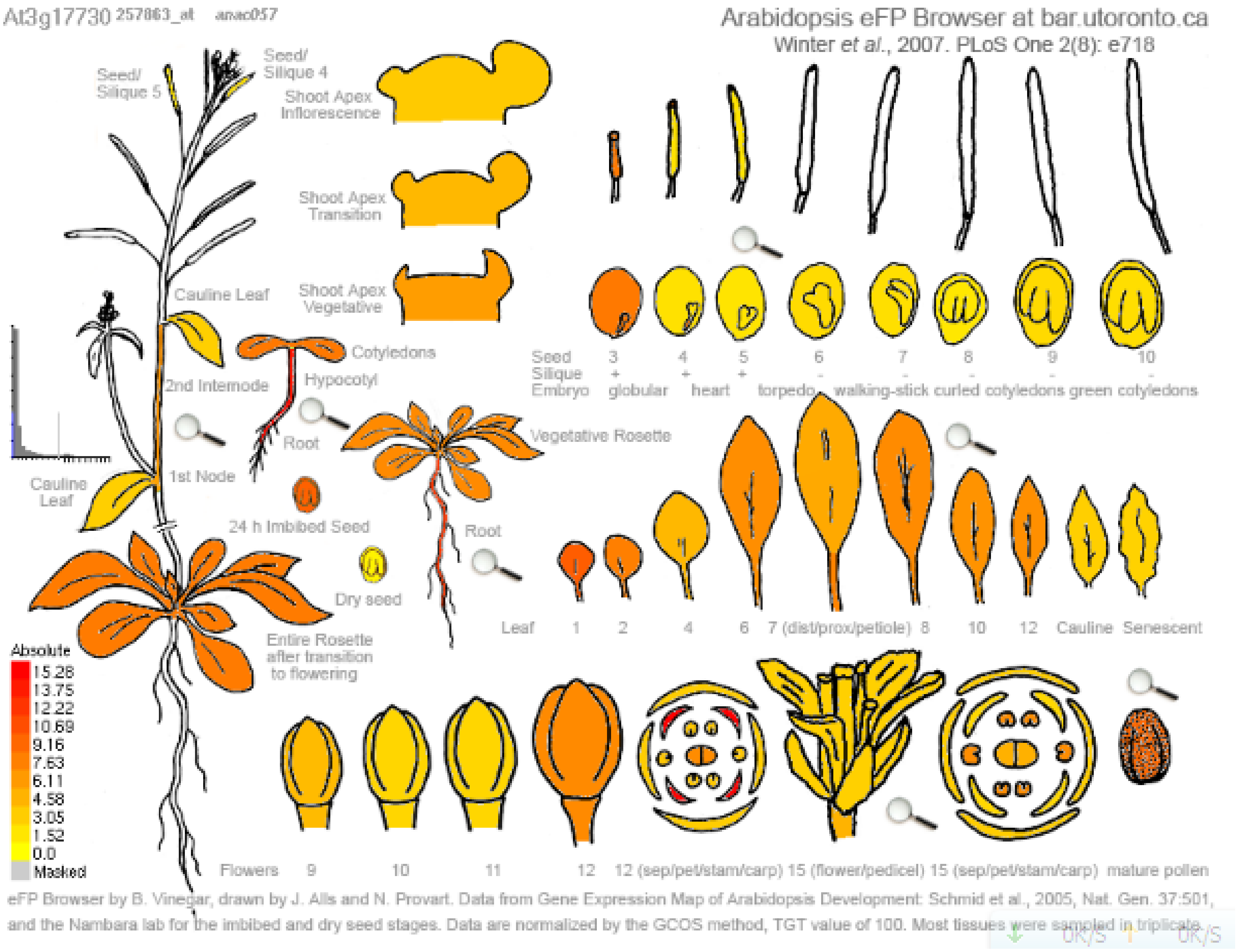
tr|K3Y762|K3Y762_SETIT WoLFPSORT prediction chlo: 5.0, nucl: 5.0, cyto: 3.0

[PSORT features and traditional PSORTII prediction]

14 Nearest Neighbors

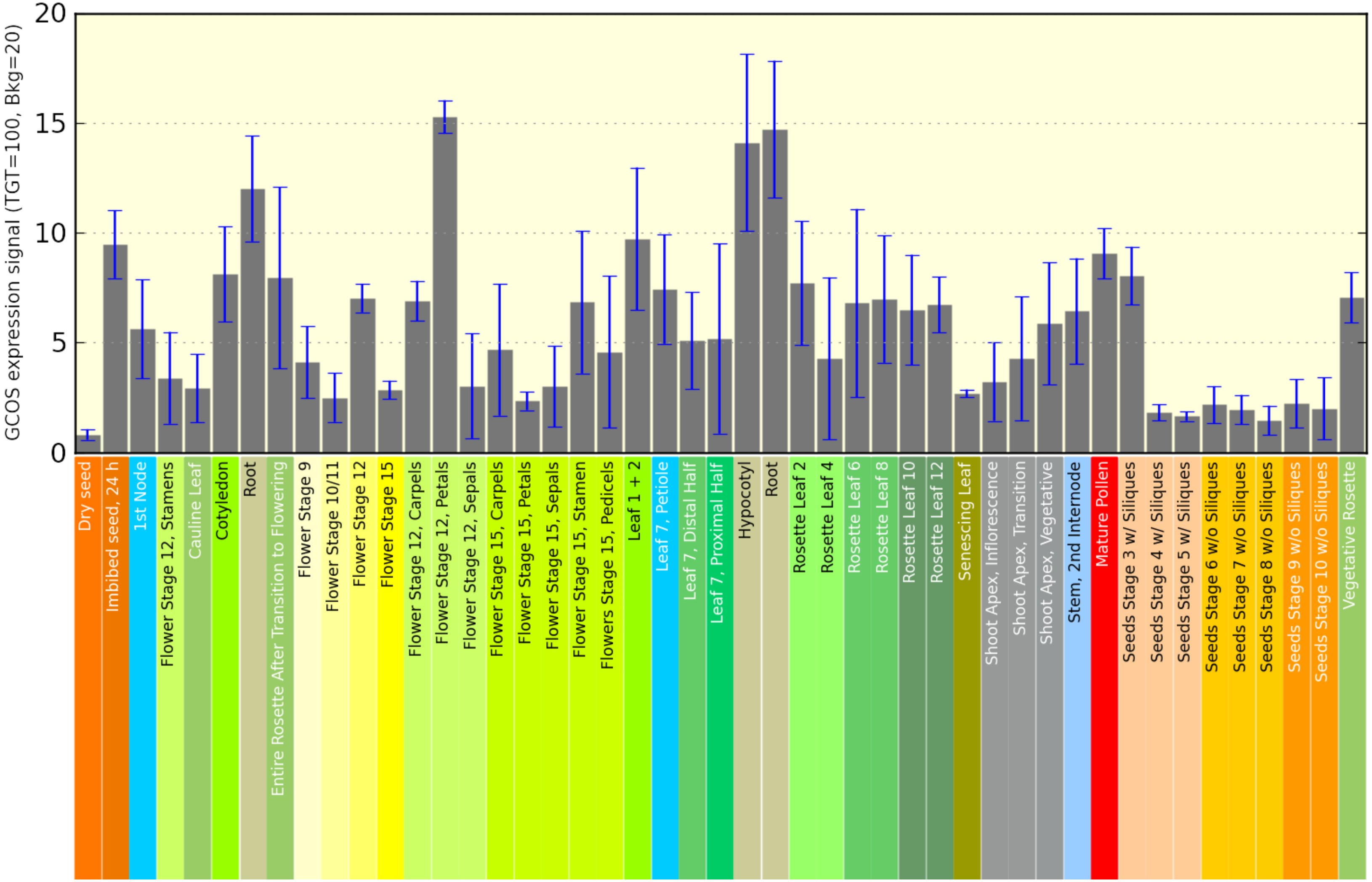
id	site	distance	identity	comments
RPOA_PHYPA	chlo	211.3	10%	[Uniprot] SWISS-PROT45:Chloroplast.
At3g14230.1	nucl	232.1	14%	[Arath]
PPOL_ARATH	nucl	234.1	14%	[Arath] [Uniprot] SWISS-PROT45:Nuclear. Evidence:IDA Pubmed:9778846
At1g16060.1	nucl	234.4	14%	[Arath]
SR53_HORVU	cyto	245.4	11%	[illegible] SS-PROT45:Cytoplasmic.
At1g64380.1	nucl	246.7	13%	[Arath]
CB23_POLMU	chlo	257.3	15%	[Uniprot] SWISS-PROT45:Chloroplast thylakoid membrane.
PSY_CUCME	chlo	262.9	13%	[Uniprot] SWISS-PROT45:Chloroplast.
CAP1_MESCR	cyto	263.0	13%	[Uniprot] SWISS-PROT45:Cytoplasmic.
E13K_TOBAC	extr	266.4	13%	[Uniprot] SWISS-PROT45:Extracellular.
RIP3_MAIZE	cyto	267.5	14%	[Uniprot] SWISS-PROT45:Cytoplasmic.
PODK_FLATR	chlo	268.7	14%	[Uniprot] SWISS-PROT45:Chloroplast.
AAT5_ARATH	chlo	269.1	12%	[Arath] [Uniprot] SWISS-PROT45:Chloroplast. Evidence:IDA Pubmed:7766905
At5g55310.1	nucl	269.2	12%	[Arath]

拟南芥AT3G17730组织特异性表达



eFP Browser by B. Vinegar, drawn by J. Alls and N. Provart. Data from Gene Expression Map of Arabidopsis Development: Schmid et al., 2005, Nat. Gen. 37:501, and the Nambara lab for the imbibed and dry seed stages. Data are normalized by the GCOS method, TGT value of 100. Most tissues were sampled in triplicate.

拟南芥AT3G17730组织特异性表达



由以上分析可得：

1：与水稻、拟南芥等序列比对虽然有结构域上的保守位点，但是进化关系较远，可以推测该基因可能有新的功能，有待研究；

2：从与该基因同源性较高的AT3G17730的Arabidopsis eFP Browser来看，该基因可能与营养的活化有关，以调控植株的生长发育，因此，可以设计相应的实验加以验证；

3：该基因含有NAC结构域，因此推测该基因可能参与抗旱、耐盐等逆境信号应答，可以设计相应的实验加以验证。

谢谢罗老师孜孜不倦的教导！

感谢小组成员的共同努力！