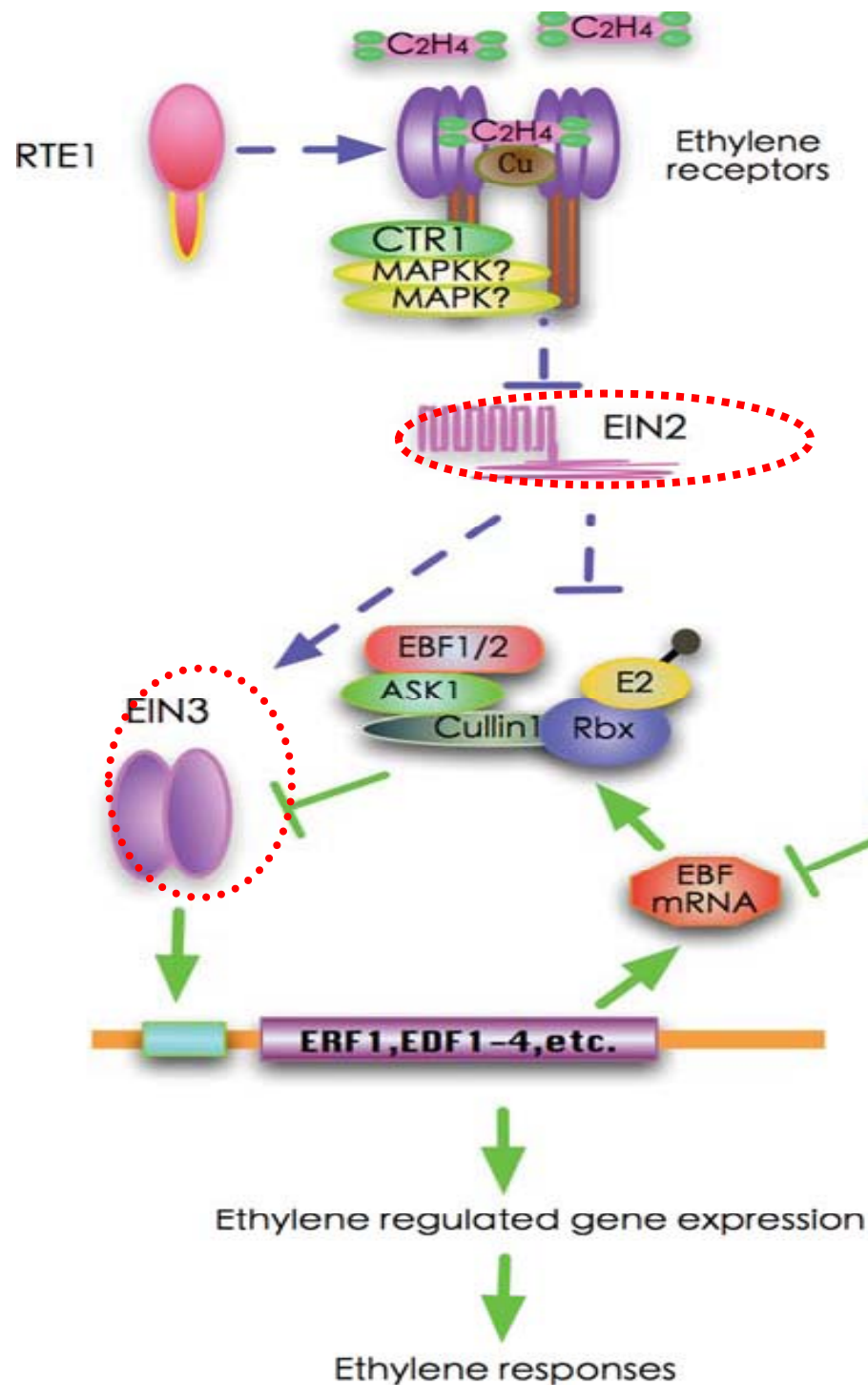




Analysis of two important genes in ethylene signaling pathway

李文阳, 张新岩, 何文容
指导老师: 罗静初, 郭红卫
Peking University



EIN2: a central component
 the first positive regulator
 the biochemical mechanism X
 the subcellular localization X

EIN3: a plant-specific primary TF
 the downstream of EIN2
 the EILs(EIN3 like) family

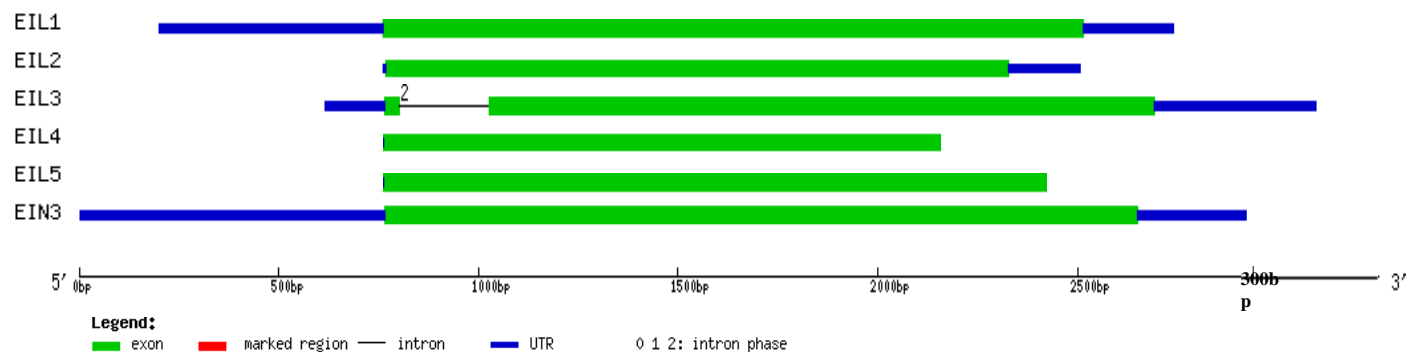
HJ Li and HW Guo, 2007

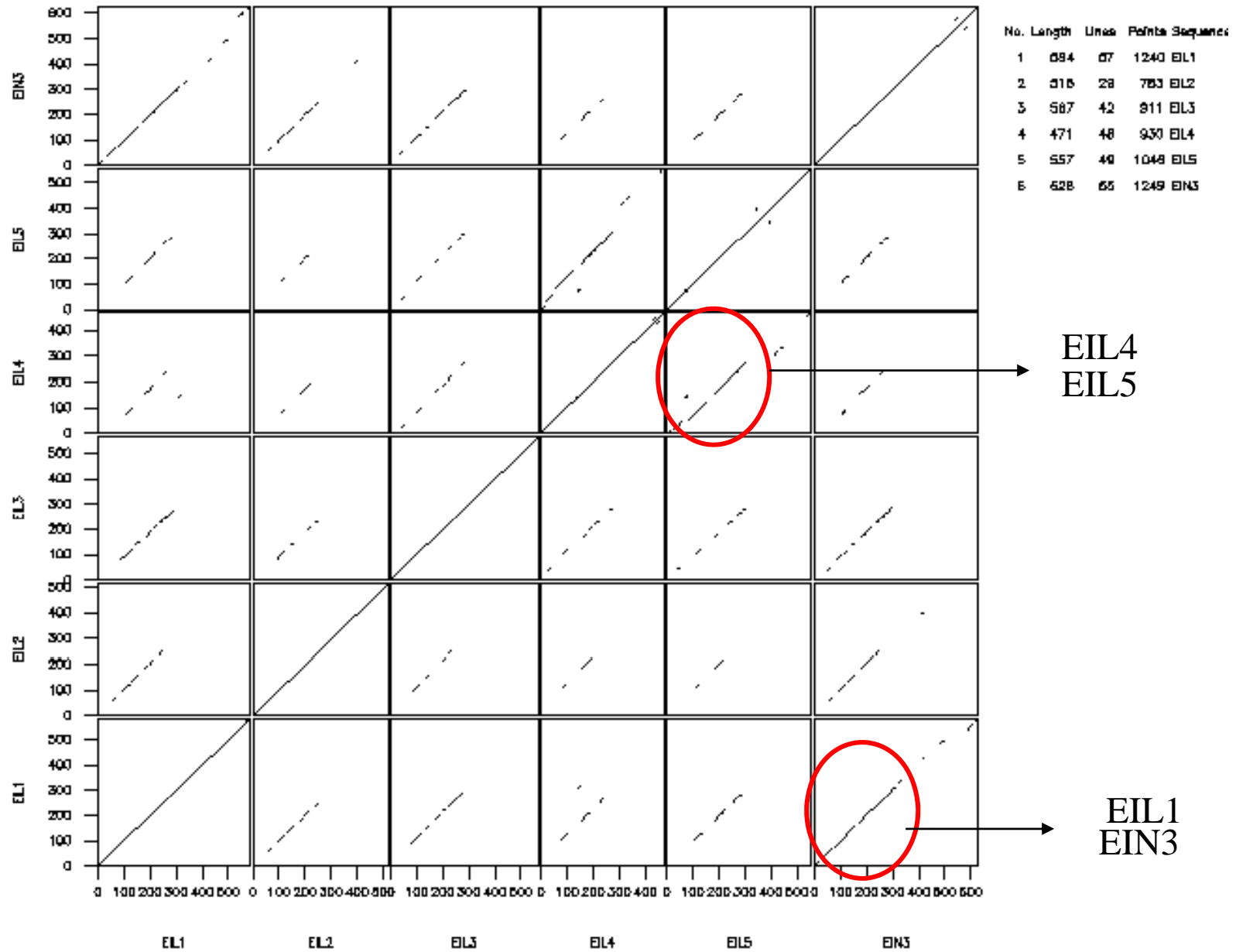
Part I---EIN3

At. EIL family TFs

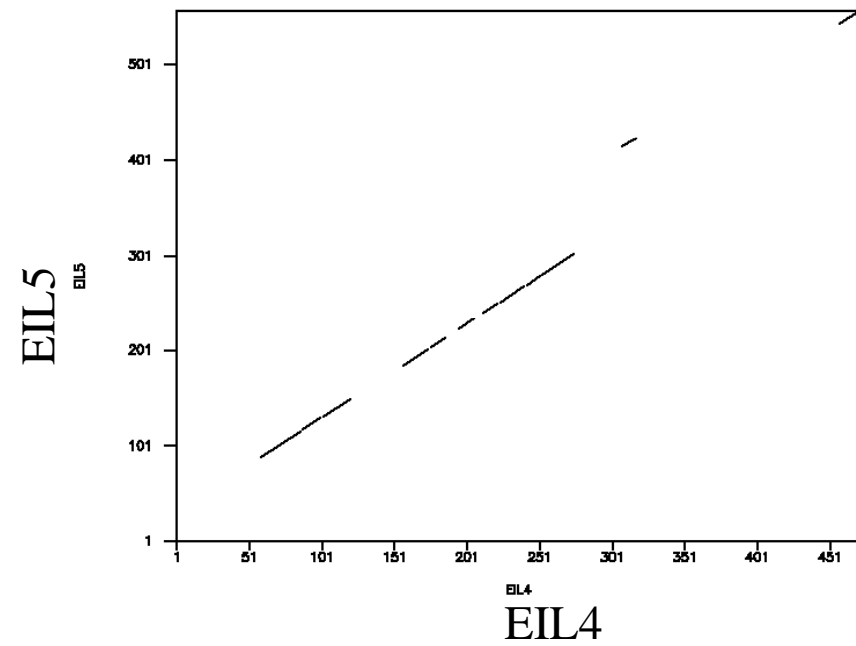
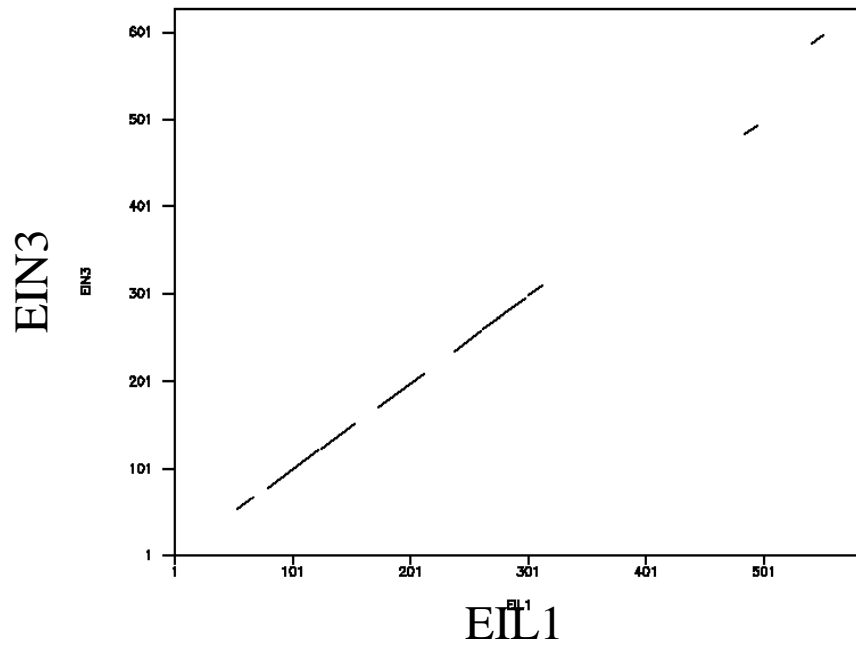
NO.	Name	TAIR ID	RefSeq ID	Ex	GenPet ID	L(aa)	Gene ID
1	AtEIL1	AT2G27050.1	NM_128263.4	2	NP_180273.1	584	S17247
2	AtEIL2	AT5G21120.1	NM_122120.2	1	NP_197611.1	518	S32237
3	AtEIL3	AT1G73730.1	NM_106032.4	3	NP_177514	567	S43708
4	AtEIL4	AT5G10120.1	NM_121050.1	1	NP_196574.1	471	S30876
5	AtEIL5	AT5G65100.1	NM_125909.1	1	NP_201315.1	557	S36634
6	AtEIN3	AT3G20770.1	NM_112968.3	2	NP_188713.1	628	S21625

Ex:外显子数目; L:蛋白序列长度

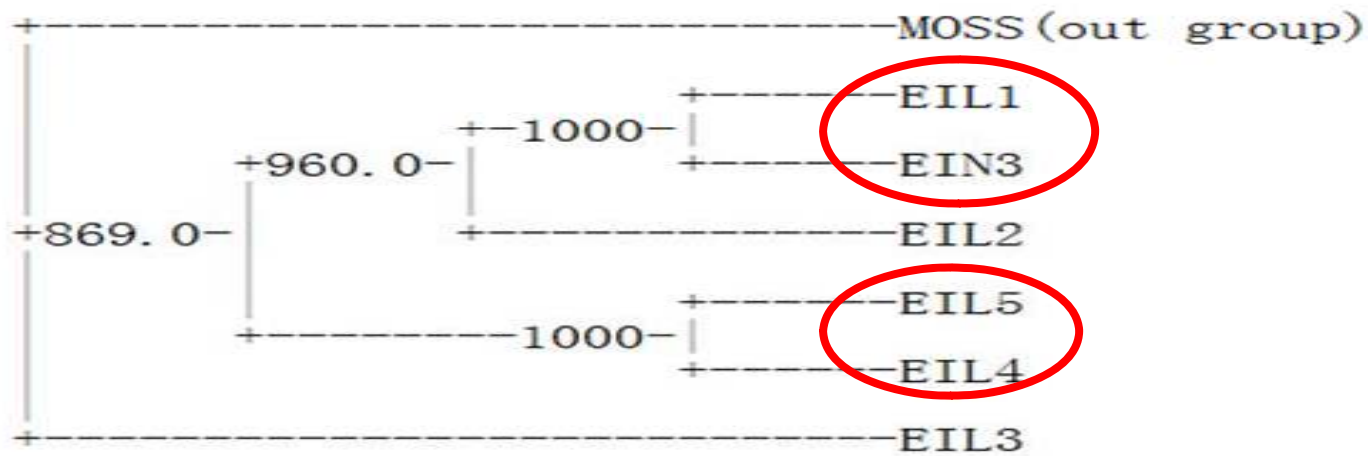




EIL family shares high identical sequences in N terminal; more, EIN3&EIL1, EIL4&EIL5 have higher identity.



S \ I	EIL1	EIL2	EIL3	EIL4	EIL5	EIN3
EIL1		38.9%	41.4%	36.5%	37.0%	66.8%
EIL2	49.2%		34.9%	33.3%	33.7%	37.8%
EIL3	52.4%	47.2%		35.7%	34.9%	40.8%
EIL4	46.7%	43.1%	47.9%		56.4%	35.1%
EIL5	47.0%	43.4%	45.2%	63.1%		35.2%
EIN3	73.7%	46.6%	50.7%	45.9%	45.8%	



Links	Name	Expect	Motifs
S A ?	EIL1	8.8e-296	10 5 8 1 7 2 3 6 4 9
S A ?	EIN3	1.6e-294	10 5 8 1 7 2 3 6 10 4 9
S A ?	EIL4	1.3e-220	5 8 1 7 2 3 6 4
S A ?	EIL5	4.2e-220	5 8 1 7 2 3 6 4
S A ?	EIL3	1.1e-191	5 8 1 7 2 3 6 8
S A ?	EIL2	3.3e-185	10 5 1 7 2 3 6 4
SCALE			1 25 50 75 100 125 150 175 200 225 250 275 300 325 350 375 400 425 450 475 500 525 550 575 600 625

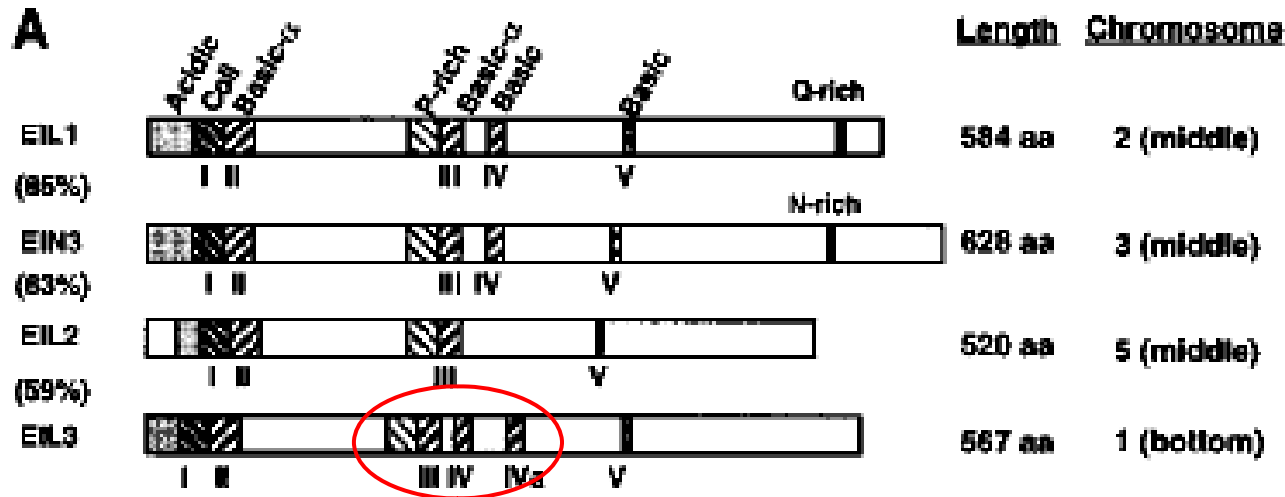
SECTION III: ANNOTATED SEQUENCES



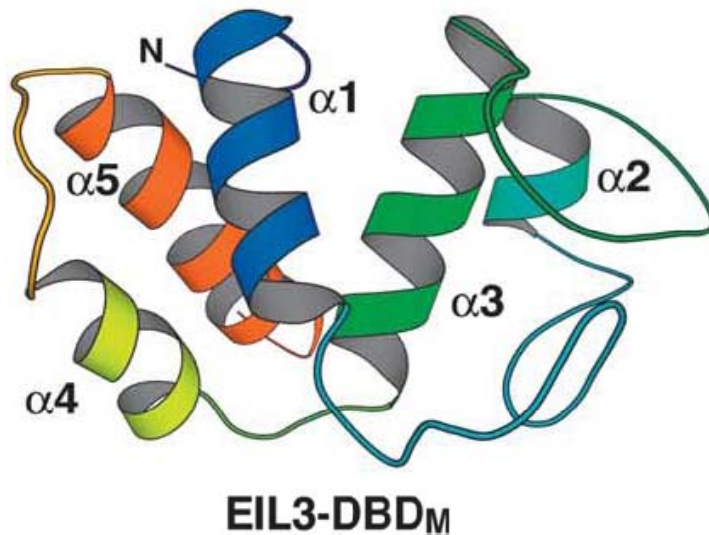
Clue on:

- Study of EIL1 function via investigation of EIN3;
- Our strategy is to study EIN3 first!

We have known



Qimin Chao.
CELL (89), 27 June 1997



Kazuhiko Yamasaki et al.
J. Mol. Biol. (2005) 348, 253–264

Homology-modeling –swissmodel



Peptide---swiss-prot annotation

38-68:

- Number of amino acids: 31
- Arg (R) **5** 16.1%
- Lys (K) **3** 9.7%

533-583:

- Number of amino acids: 51
- Asn (N) **22** 43.1%

Modification site prediction

Phosphorylation site prediction (pKaPS):

Position	Score	Sequence	Profile
93	1.07	VDAAKQRQSSEQARRKMSRAQDGILKYMLKMMEVCKAQQGFV	1.23
273	1.11	VIKHMFPDIAKIRKLVLRQSKCLQDKMTAKESATWLAIINQEE	1.45
416	0.13	ENLGCAHSEISRGFLDRNSRDNHQLACPHRDSRLPYGAAPSR	0.82
429	1.19	FLDRNSRDNHQLACPHRDSRLPYGAAPSRFHVNEVKPVVGFPP	1.42

Phoscan:predicted phosphorylation sites by consensus sequences:

name	kinase	site	sequence
>EIN3	PDHK	S325	C[S]QYDVE
>EIN3	PDHK	S483	M[S]MYDRN
>EIN3	CaM2	T564	VFKFD[T]AD

Sumolyzation site prediction using SUMOsp2.0:

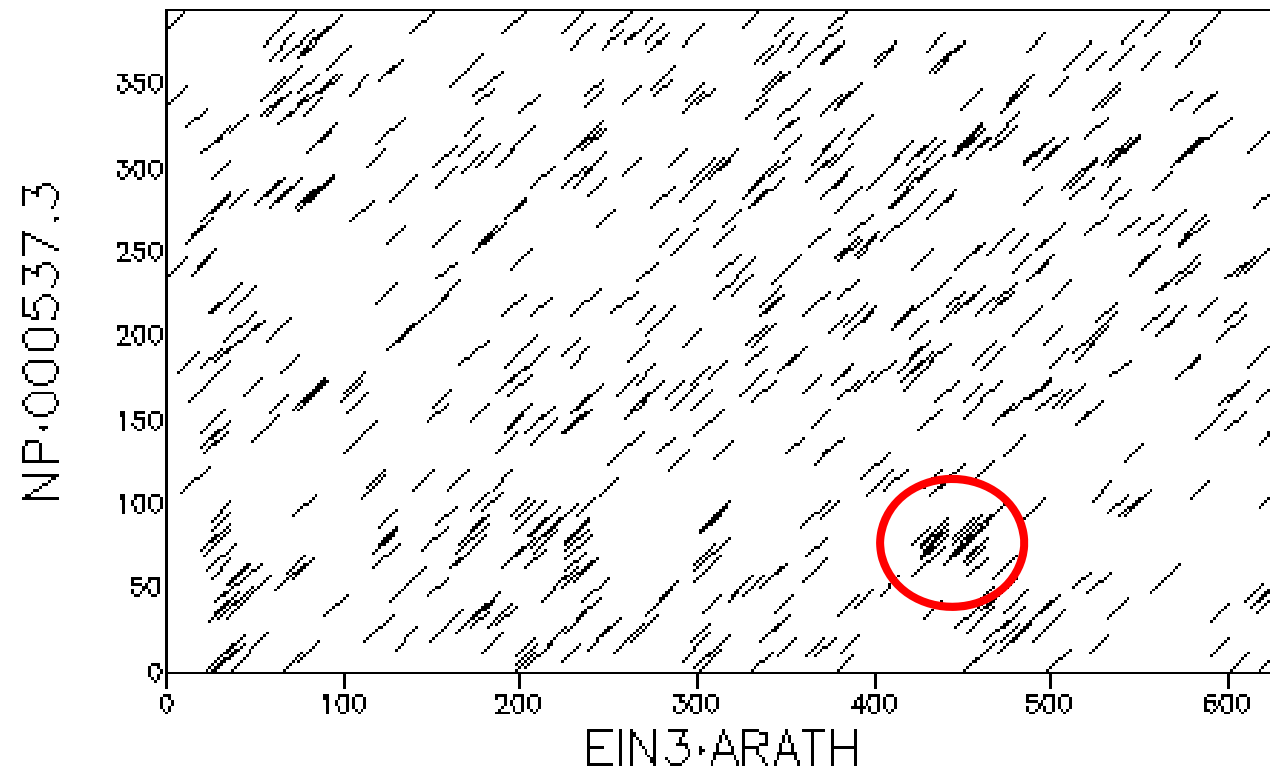
EEL K PEK	344	1.256	0.130	TypeI: ? ? - K -X-E
FPV K EEV	366	1.322	0.130	TypeI: ? ? - K -X-E

EIN3-p53 in plants?

Dotmatcher: EIN3 vs. p53

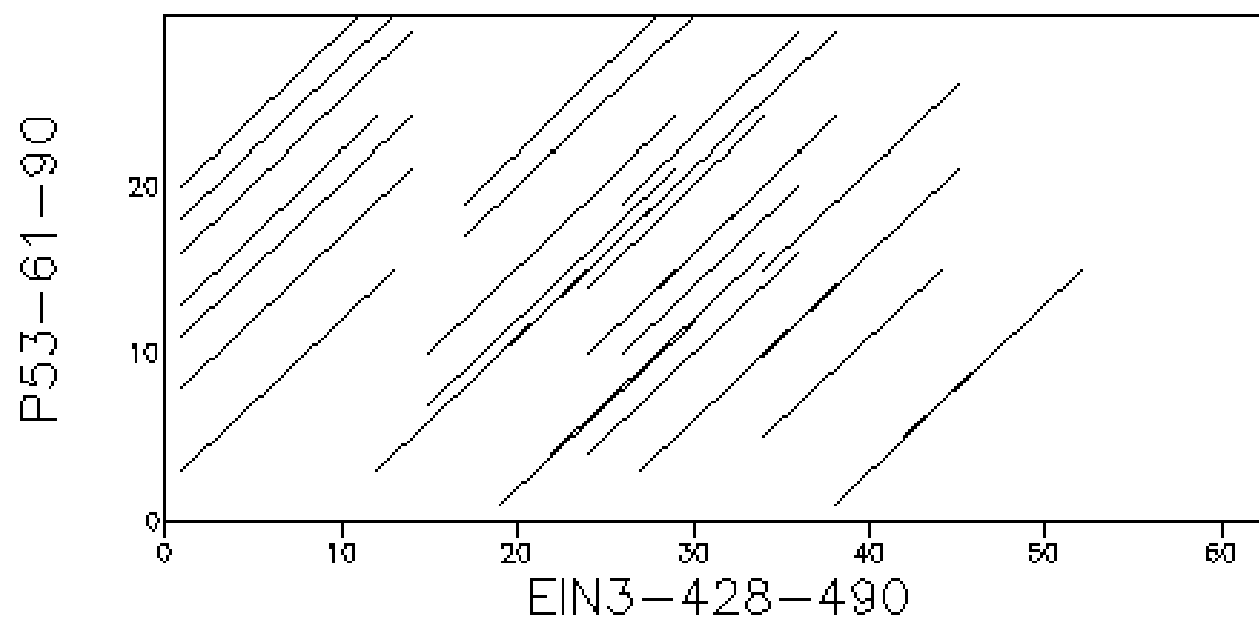
Dotmatcher: EIN3·ARATH vs NP·000537.3

(windowsize = 10, threshold = 10.00 14/06/08)



Dotmatcher: EIN3-428-490 vs P53-61-90

(windowsize = 10, threshold = 8.00 14/06/08)



p53 feature table

REGION 66 110 45 Interaction with **WWOX**

What is **WWOX**?

"WWOX, the FRA16D gene, behaves as a suppressor of tumor growth.";

[Cancer Res. 61:8068-8073\(2001\).](#)

Analysis of microarray data

Kobas for microarray data

- Microarray

 - Air/Ethylene treatment

- Microarray data normalization

- Set threshold and select the differently expressed genes

KOBAS

- Get the sequence from Tair database (www.arabidopsis.org)
- Upload sequences of gene list to KOBAS
 - ✓ KO annotation
 - ✓ Analysis with KOBAS

KOBAS: KO-Based Annotation System - Windows Internet Explorer

http://kobas.cbi.pku.edu.cn:8080/program.run.do

文件(F) 编辑(E) 查看(V) 收藏夹(A) 工具(T) 帮助(H)

KOBAS: KO-Based Annotation System

Pathway identification

Advanced KOBAS

- Login (free registration)
- User Space
- Analysis history

Help

- Tutorial
- Feedback

Summary

Red bar is for query gene set, green bar is for background gene set.

PATHWAY	COUNT AND RATIO	P-VALUE	Q-VALUE
Transcription factors	 16 / 18.39% 216 / 7.20%	0.000445879322175	0.00945087104376
Phenylalanine metabolism	 8 / 9.20% 76 / 2.53%	0.00167363134956	0.0128252766034
Methane metabolism	 8 / 9.20% 77 / 2.57%	0.00181523764704	0.0128252766034
Other enzymes	 12 / 13.79% 162 / 5.40%	0.00247881767129	0.0131352728122
Axon guidance	 2 / 2.30% 5 / 0.17%	0.00945269586641	0.0400719231444
Metabolism of xenobiotics by cytochrome P450	 3 / 3.45% 19 / 0.63%	0.0181315018772	0.0640526399367
Valine, leucine and isoleucine degradation	 3 / 3.45% 21 / 0.70%	0.023503536687	0.0711688064634
Tetracycline biosynthesis	 1 / 1.15% 1 / 0.03%	0.0285600510065	0.0756699586754
Phenylpropanoid biosynthesis	 9 / 10.34% 155 / 5.16%	0.0357375124592	0.0841659278342

Internet | 保护模式: 启用 100%

Ethylene-induced transcription factors

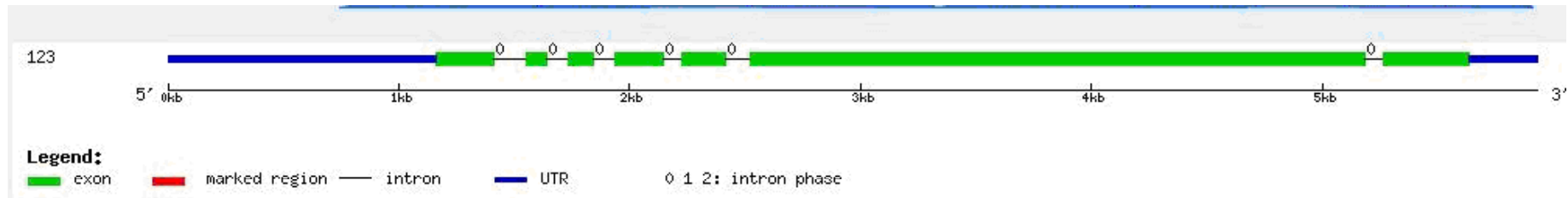
AT3G25730.1	Symbols:	AT1G18570.1	Symbols: AtMYB51, BW51A, BW51B, MYB51
AT3G53200.1	Symbols: AtMYB27	AT2G36890.1	Symbols: RAX2, MYB38, ATMYB38
AT2G47460.1	Symbols: MYB12, ATMYB12	AT4G17500.1	Symbols: AtERF-1
AT1G43160.1	Symbols: RAP2.6	AT1G74650.1	Symbols: AtY13, AtMYB31
AT1G28370.1	Symbols: ERF11, ATERF11	AT1G06180.1	Symbols: AtMYB1fgn, AtMYB13, ATMYB13
AT4G17785.1	Symbols: MYB39	AT5G06960.1	Symbols: TGA5, OBF5
AT3G14230.1	Symbols: RAP2.2	AT2G31180.1	Symbols: AtMYB14, Myb14at
AT2G36890.1	Symbols: RAX2, MYB38, ATMYB38	AT3G50060.1	Symbols: MYB77
AT5G47220.1	Symbols: ATERF2, ATERF-2, ERF2	AT4G40060.1	Symbols: ATHB16, ATHB-16
AT1G36060.1	Symbols:	AT4G32730.1	Symbols: PC-MYB1, ATMYB3R1, MYB3R-1, ATMYB3R-1
AT1G25560.1	Symbols:	AT3G14230.1	Symbols: RAP2.2
AT5G44210.1	Symbols: ERF9, ATERF9, ATERF-9	AT3G16770.1	Symbols: RAP2.3, ATEBP, ERF72
AT1G68840.1	Symbols: RAP2.8, RAV2		
AT5G53980.1	Symbols: ATHB52		
AT3G16770.1	Symbols: RAP2.3, ATEBP, ERF72		
AT3G23240.1	Symbols: ATERF1, ERF1		

Ethylene-repressed transcription factors

AT5G65790.1	Symbols: ATMYB68, MYB68
AT5G65310.1	Symbols: ATHB-5, ATHB5
AT4G37750.1	Symbols: DRG, CKC, CKC1, ANT
AT5G15310.1	Symbols: AtMYB16, AtMIXTA
AT5G15150.1	Symbols: HAT7, ATHB3, ATHB-3
AT5G60890.1	Symbols: ATMYB34, ATR1, MYB34
AT2G22430.1	Symbols: ATHB6
AT5G16600.1	Symbols: AtMYB43, MYB43
AT3G61890.1	Symbols: ATHB12, ATHB-12
AT2G46270.1	Symbols: GBF3

Part II---EIN2

EIN2 gene structure display



EIN2 gene has 7 exons ,1294 aa.

N-terminal 461aa hydrophobicity,

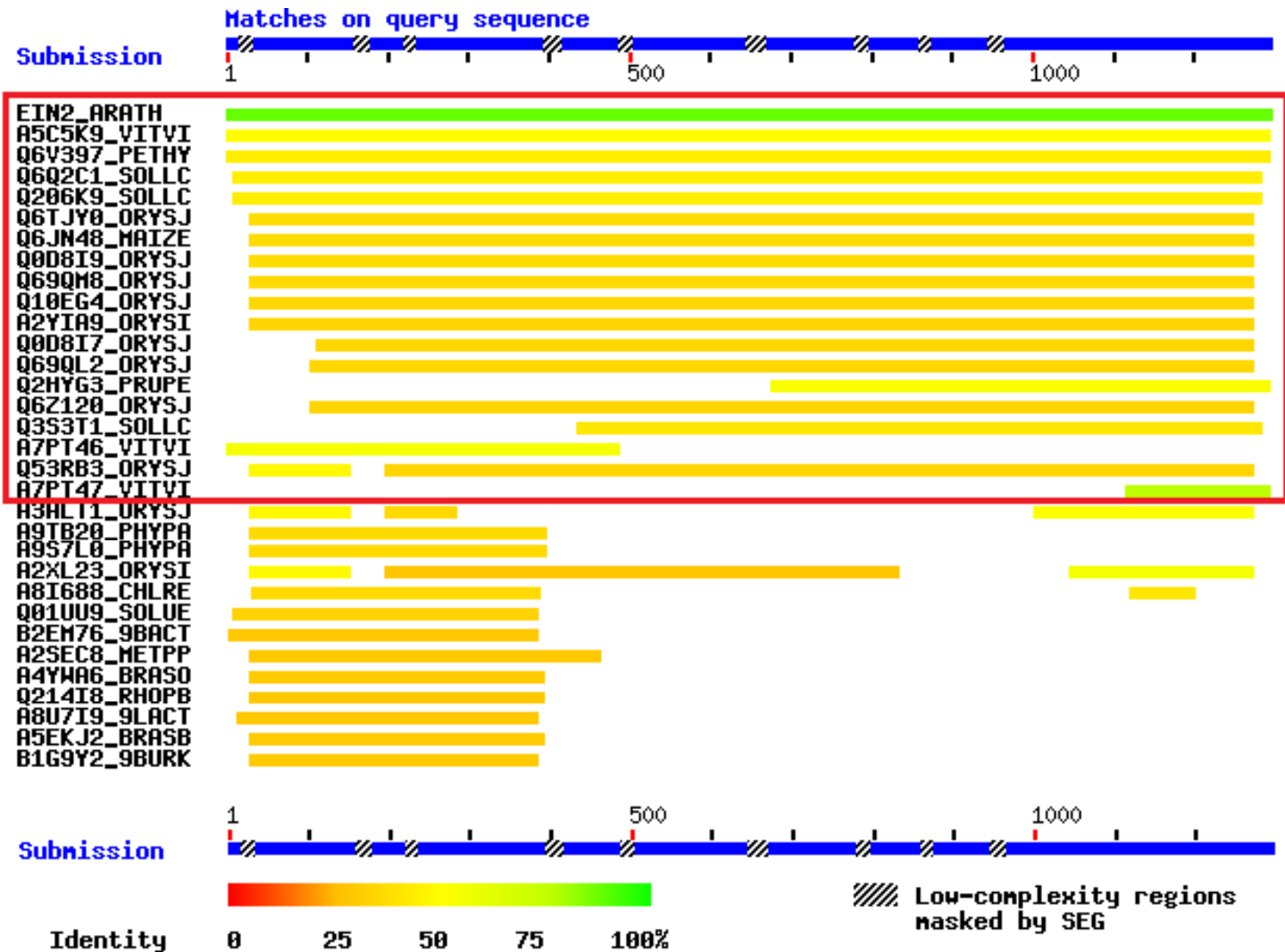
12 transmembrane helices

C-terminal 833 aa hydrophilic.

biochemical mechanism X

subcellular localization X

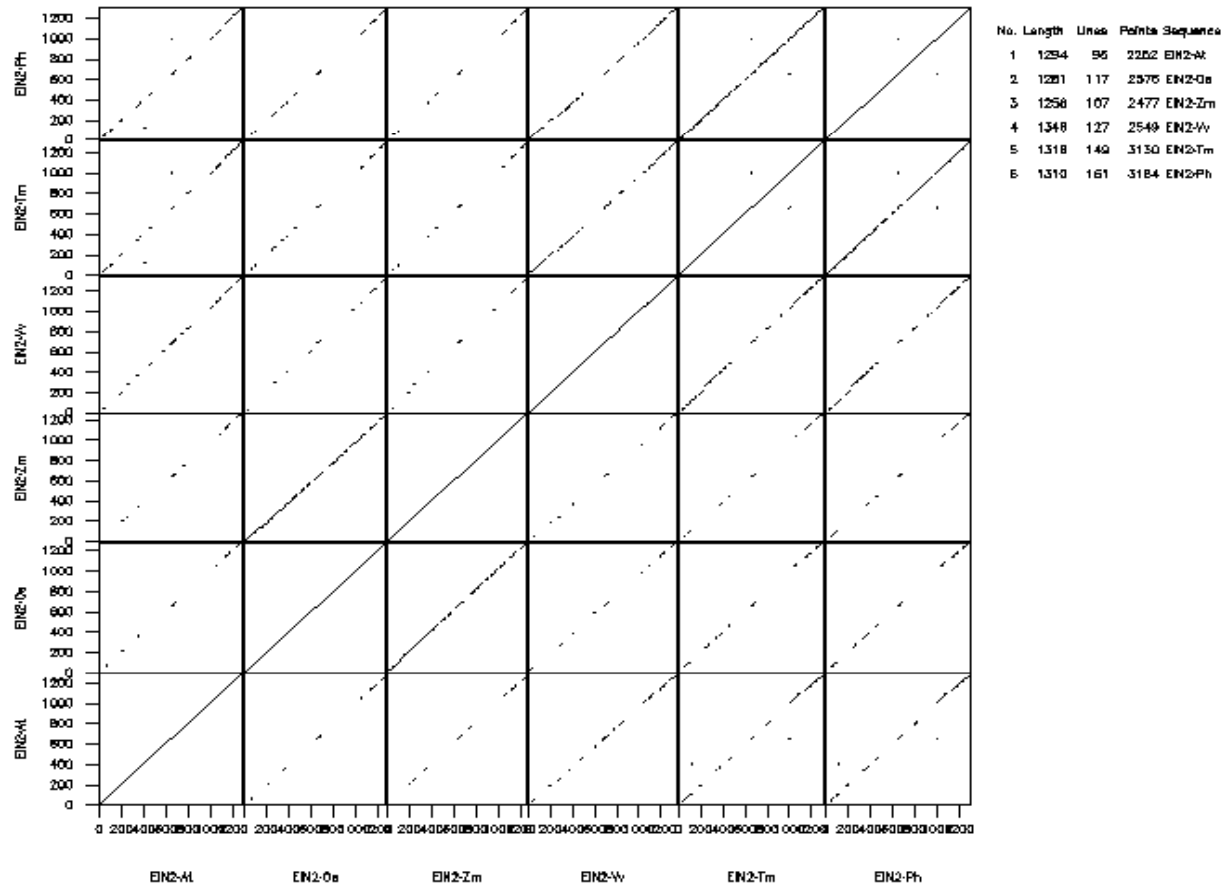
ExPASy----BLAST

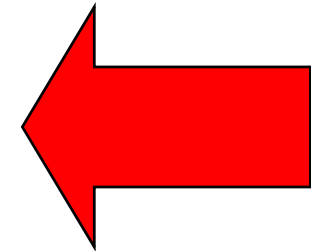
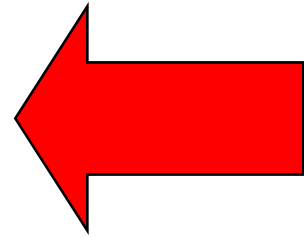
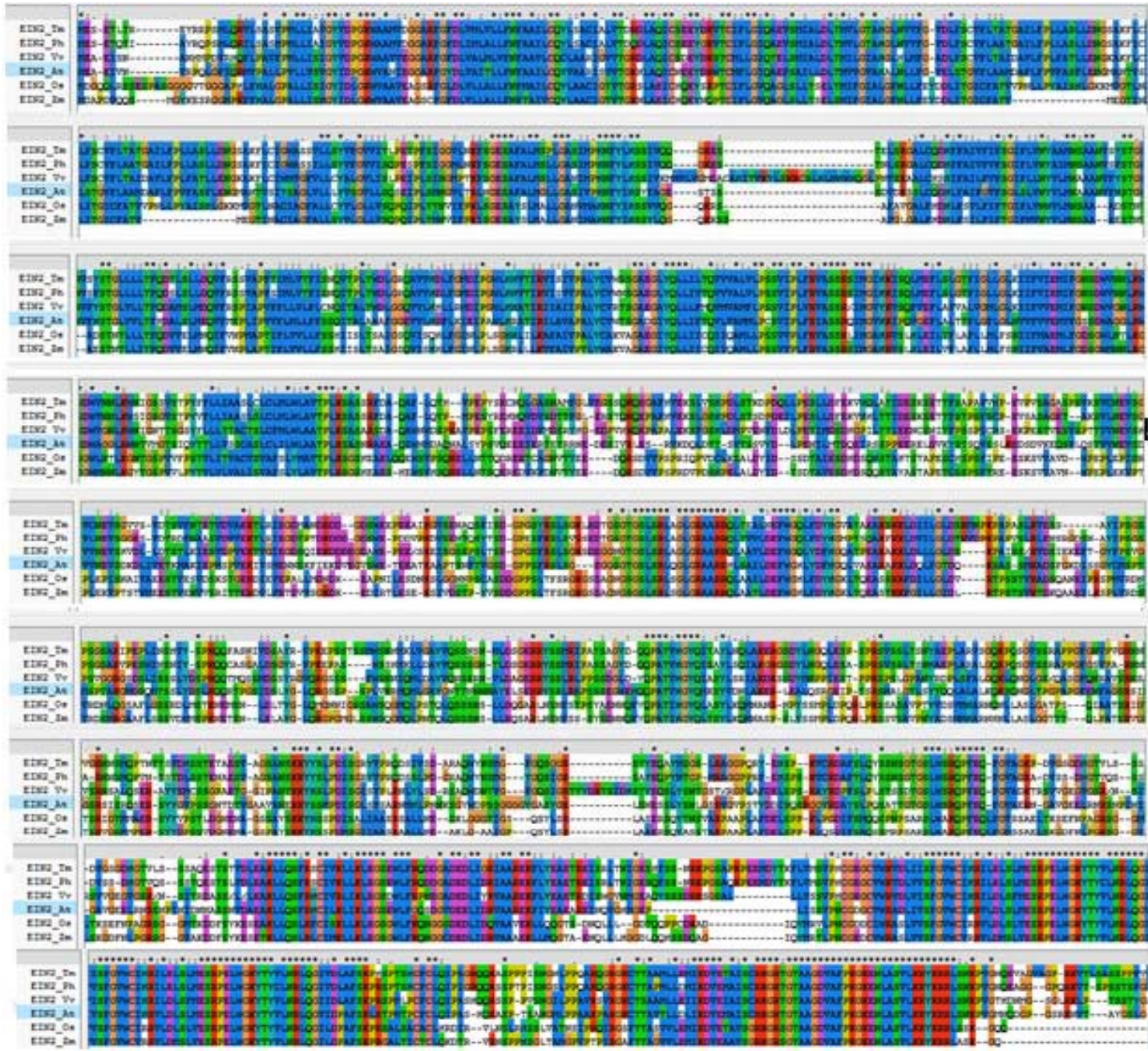


Six representative genes

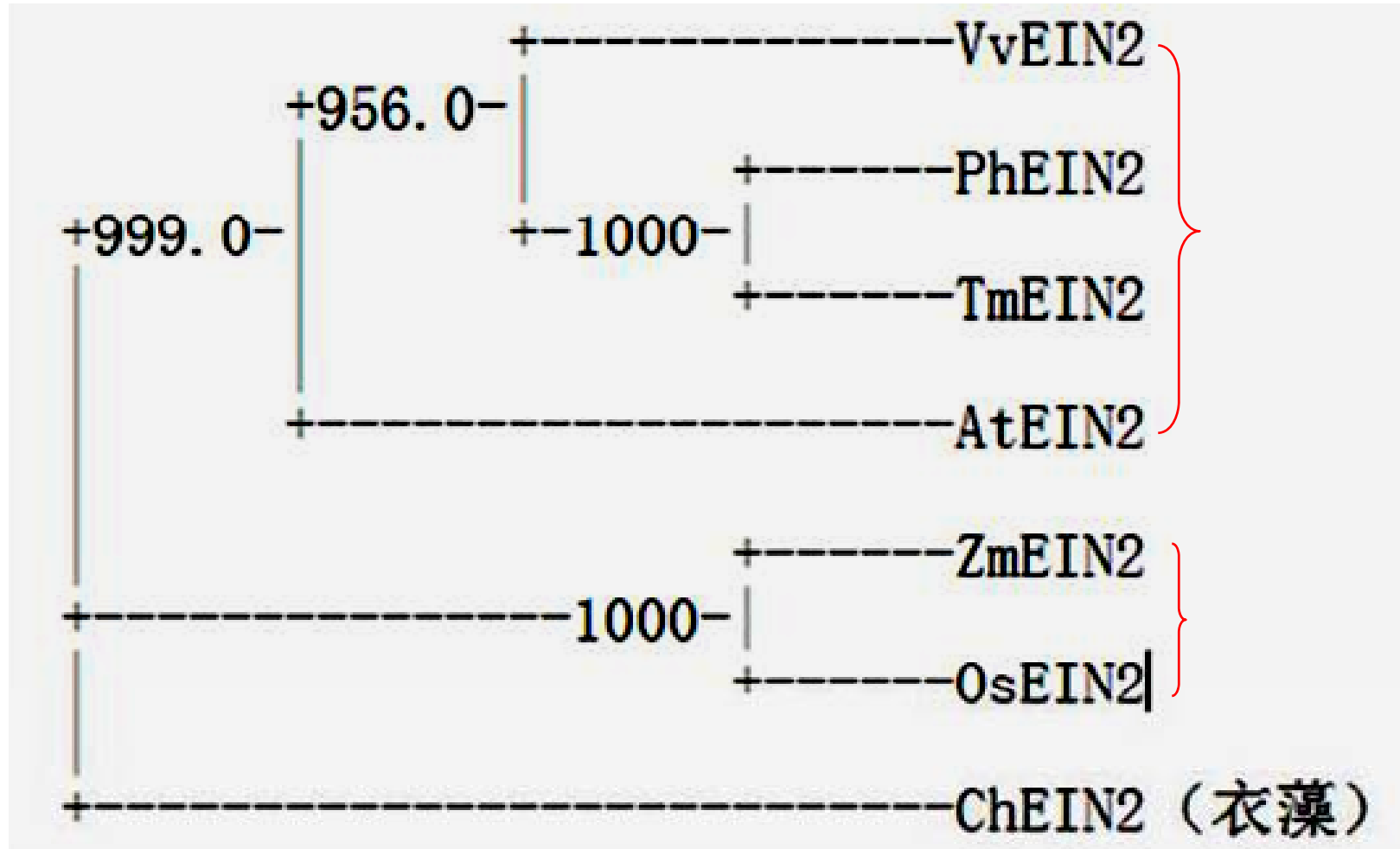
Poly dotplot of C:\Users\He Wenrong\Desktop\EIN2-6.fasta...

Sun 14 Jun 2008 20:50:46



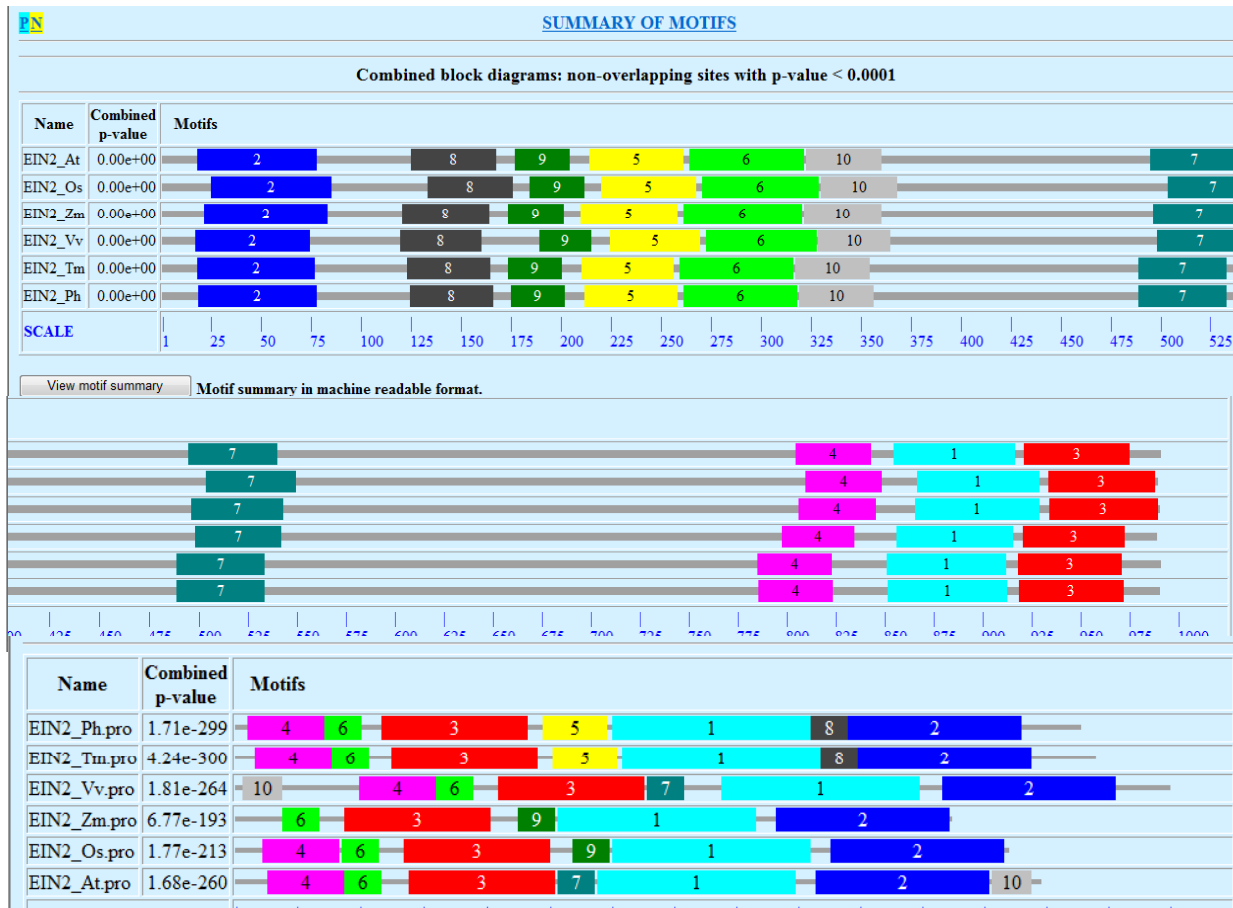


Clustering



NJ tree

MEME-Conserved domain



Minimum width: 10

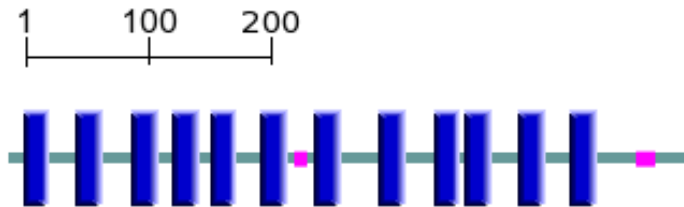
Maximum width: 80

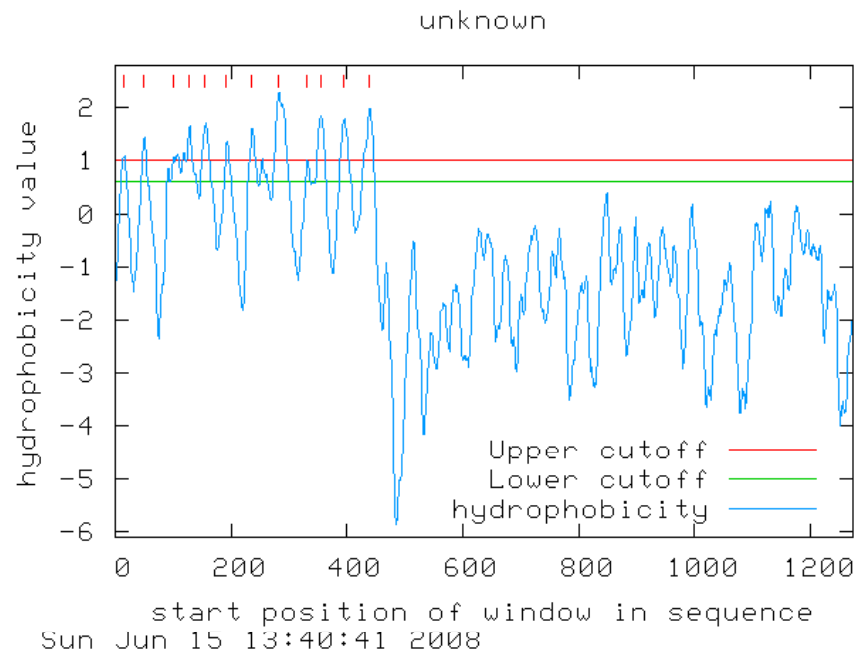
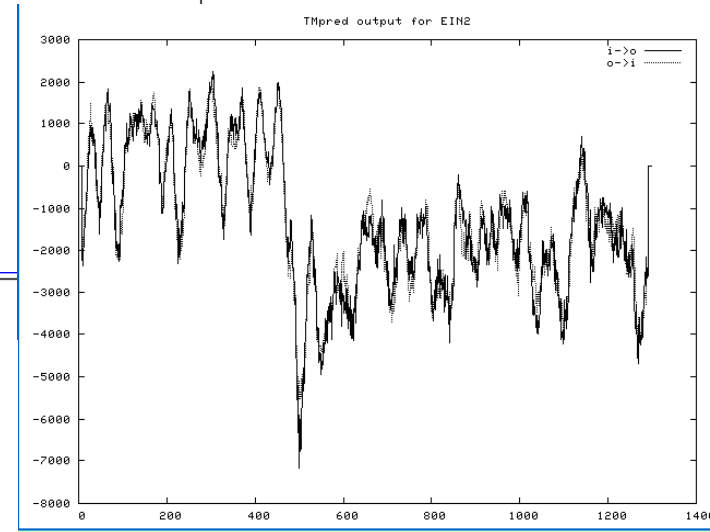
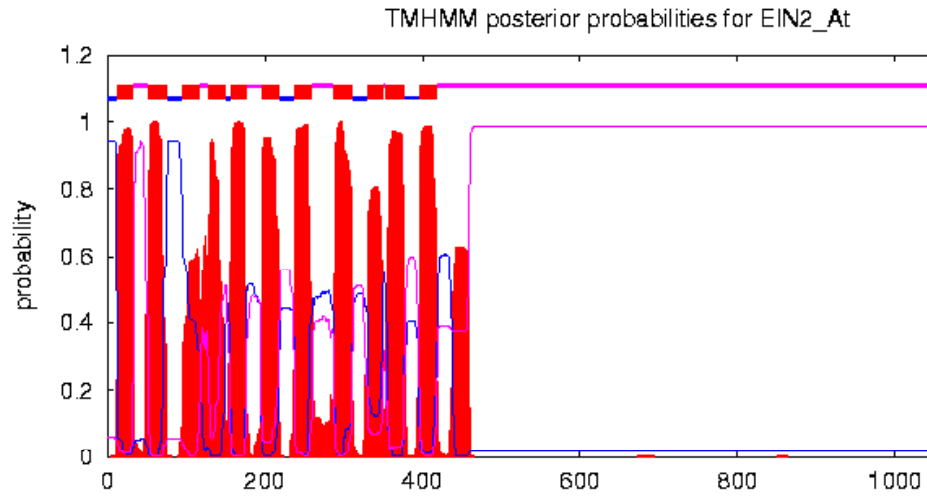
Maximum number of motifs: 10

Seven conserved domains--SMART

- 1, 15-80
- 2, 115-360
- 3, 480-535
- 4, 790-840
- 5, 850-980
- 6, 985-1105
- 7, 1115-1280

Domains within the query sequence of 535 residues

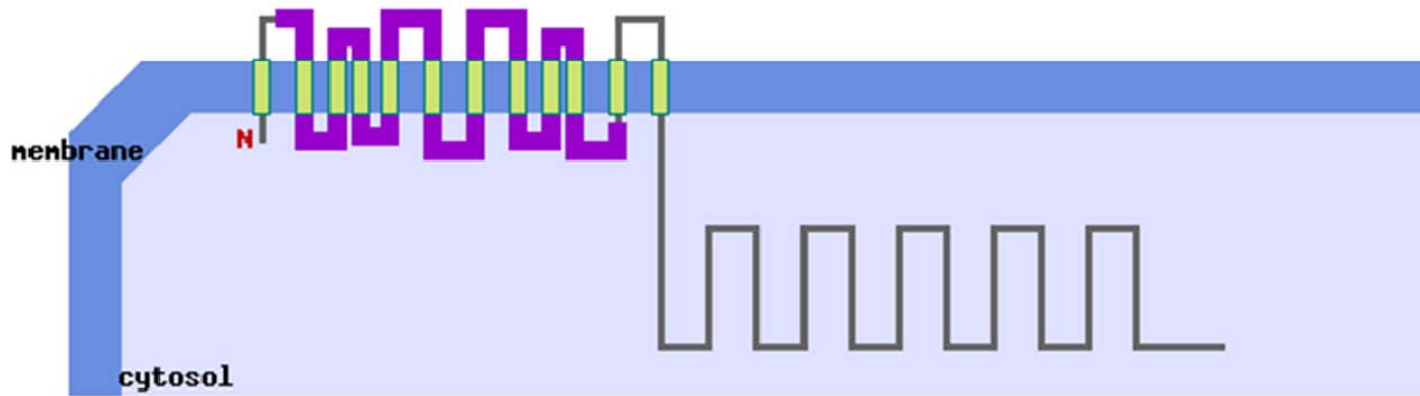




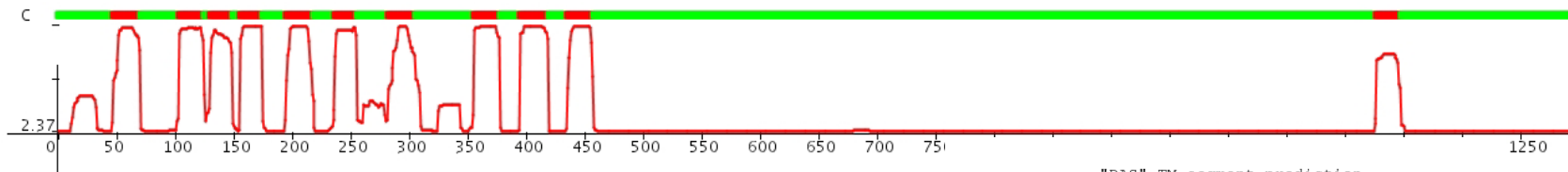
This amino acid sequence is of a MEMBRANE PROTEIN which have 10 transmembrane helices.

No.	N terminal	transmembrane region	C terminal	type	length
1	12	LGFIQRMVPALLPVLVSVGYID	34	PRIMARY	23
2	50	YDLVAITLLFNFAAILCQYVAAR	72	PRIMARY	23
3	94	CMFLGIQAEFSAILLDLTMVVG	116	SECONDARY	23
4	124	FGVELSTGVFLAAMDALFPVFA	146	PRIMARY	23
5	156	TVSIYSAGLVLLLVSGVLLSQ	177	PRIMARY	22
6	236	DHLFAIFGVFSGLSLVNYVLMNA	258	SECONDARY	23
7	283	VFMSPLIPVVFLMLLFFSSQITA	305	PRIMARY	23
8	355	QLLIFTQVLVAMMLPCSVIPLFR	377	PRIMARY	23
9	395	EFLALITFLGFLGNVVFVEMV	417	PRIMARY	23
10	440	TLLVSSCASLCLILWLAATPLK	462	PRIMARY	23

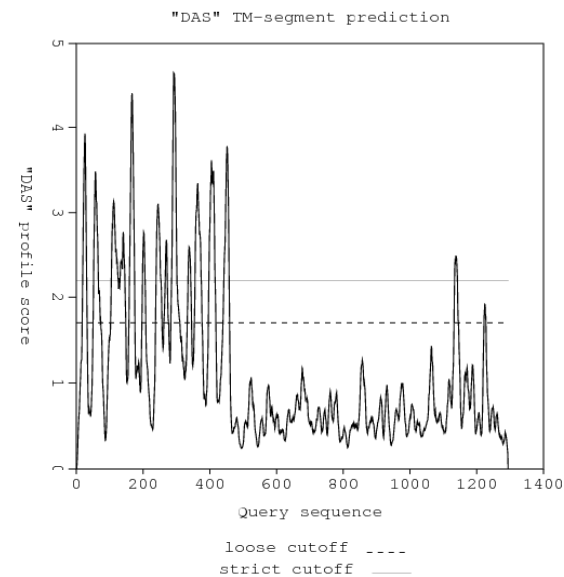
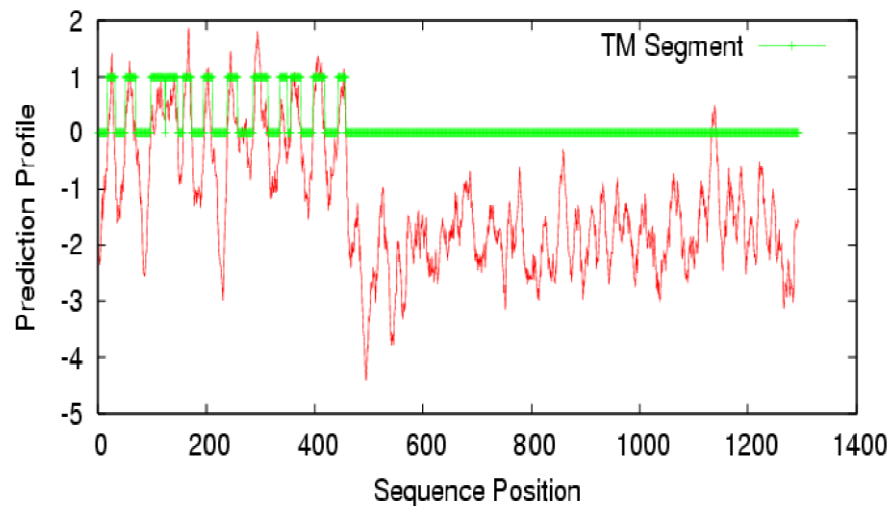
extracellular
(lumen)



|| : Predicted TM helix by Phobius. || : Logically inserted TM helix.



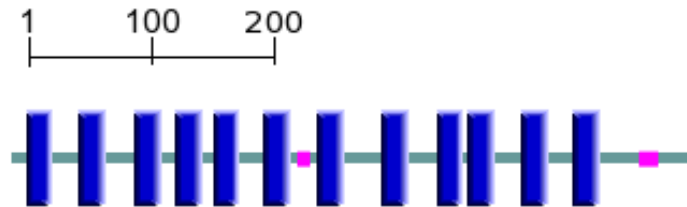
Sequence ID = AtEIN2, LENGTH = 1294



Seven conserved domains--SMART

- 1, 15-80
- 2, 115-360
- 3, 480-535
- 4, 790-840
- 5, 850-980
- 6, 985-1105
- 7, 1115-1280

Domains within the query sequence of 535 residues





Schultz et al. (1998) *Proc. Natl. Acad. Sci. USA* 95, 5857-5864
Letunic et al. (2006) *Nucleic Acids Res* 34, D257-D260

SMART MODE:

NORMAL
GENOMIC

Simple
Modular
Architecture
Research
Tool

HOME SETUP FAQ ABOUT GLOSSARY WHAT'S NEW FEEDBACK

Domains within *Arabidopsis thaliana* protein EIN2_ARATH (Q9S814)

Ethylene-insensitive protein 2

1 100 200



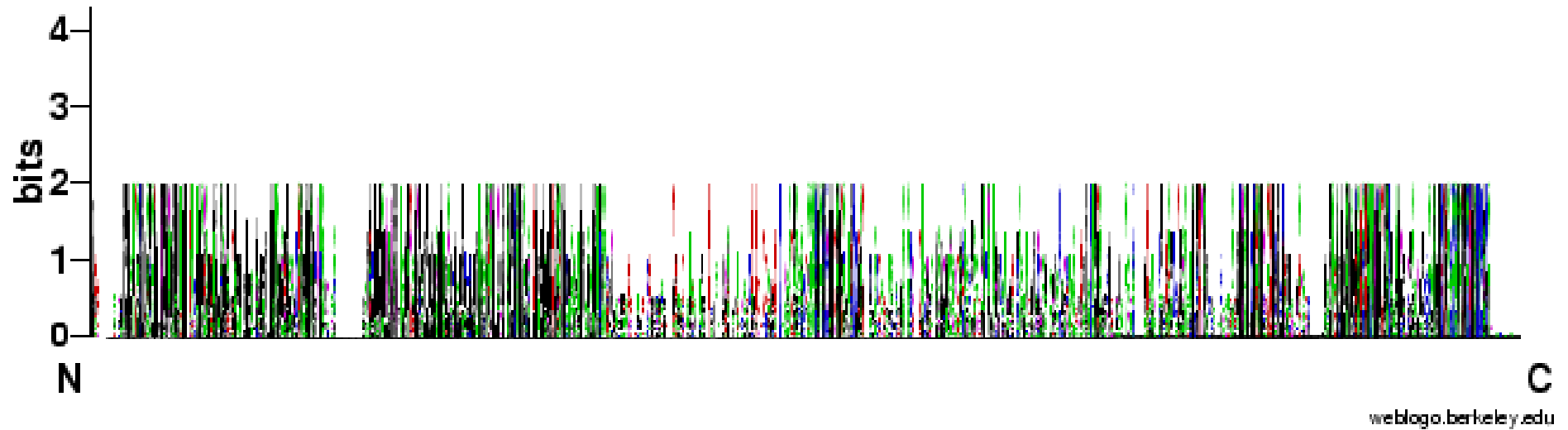
Localizome

0 250 500 750 1000 1250 1500

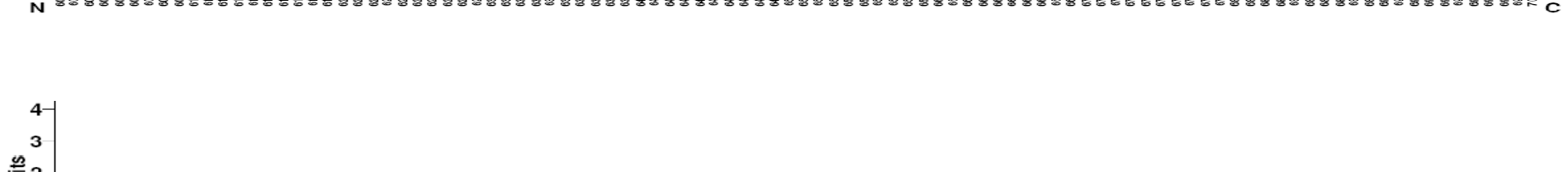
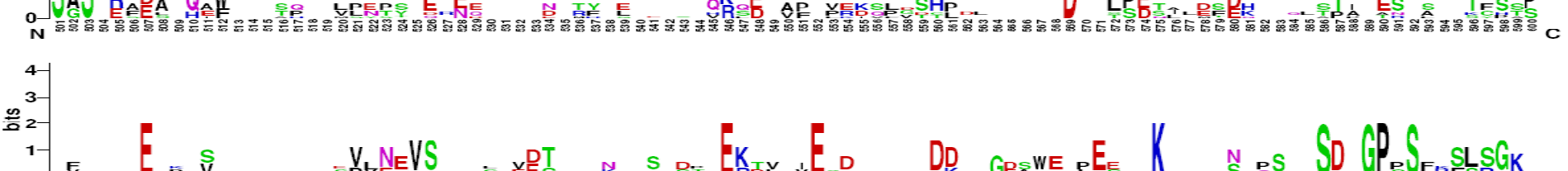
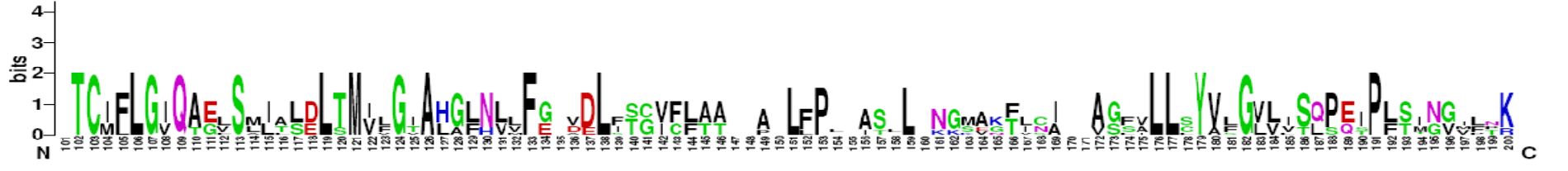
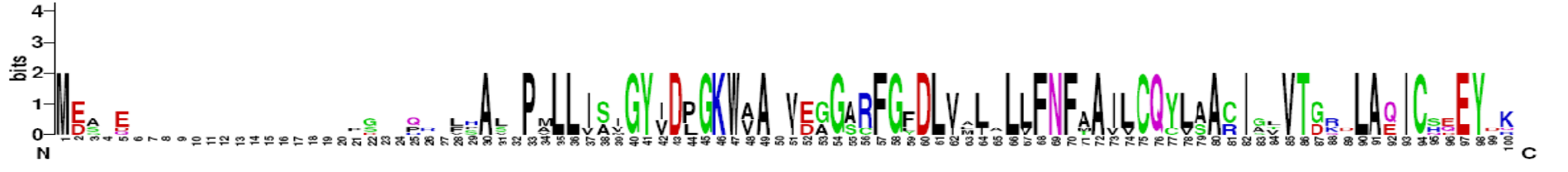


Signal sequence
Transmembrane Domain

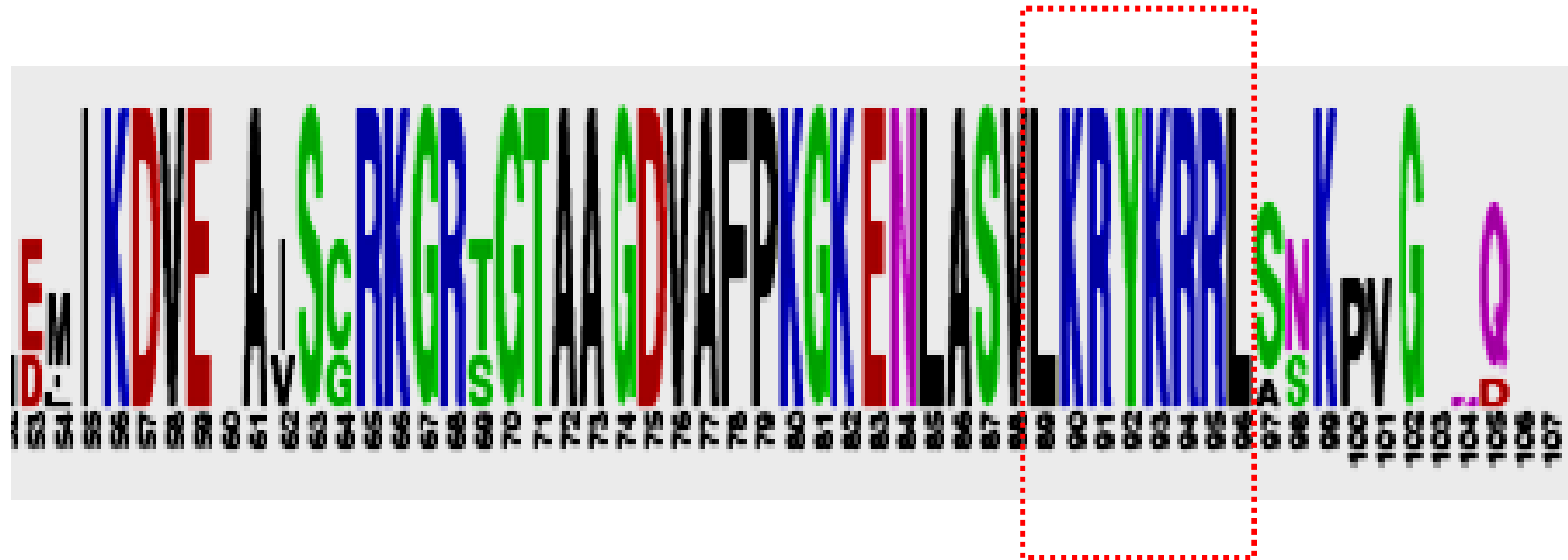
Sequence character---WEBLOGO



Full length



C-terminal 100 aa



Results Page:

- Please be patient while query is being processed.
- Waiting times could be as much as a 3-4 minutes in the
- [Help](#) on interpretation of results.

The results will be emailed to address

Input Sequence (NLS's in Red)

```
MEAEIVNVRPQLGFIQRMVPL  
FAAILCQYVAARISVVTGKH  
NLLFGVELSTGVFLAAMDAFLFPVFA  
LENGMANTVSIYSAGLVLLLYVSGVLLSQSEI  
PLSMNGVLTRLNGESAFALMGLLGA  
IVPHNFYIHSYFAGESTSSSDVDKSSLCQDHLFA  
IFGVFSGLSLVNYVLMNAAANVFH  
TGLVVLTFFHDALSIMEQVFMSP  
LIPVVFLMLLFFS  
SQITALAWAFGGEVVLHDFLKI  
EPAWLHRATIRILAVAPALYCVWT  
SGADGIYQLLI FT  
QVLVAMMLPCSVIPLFRIASS  
RIMGVHKI PQVGEFLALTTFLGFLGLNVV  
FVEMVFGS  
SDWAGGLRWNTVMGT  
SIQYTTLVSSCASLCLILWLAAT  
PLKSASNRAEAQIWNMDAQN  
ALSYPVQEEEIERTETRRNE  
LESIVRLESRVKDQLDTSVTSSVYDL  
PENILMTDQEIRS  
SPPEERELDVKYSTSQV  
SKKEDSDVKEQSVLQSTVVNEVSDKDL  
LIVETKMAKIEPMSPV  
EKIVSMENNSK  
FIEKDVGVSWETEEATKAAPT  
SNFTVGS  
DGPSPFRSLSGEGGSGTGSL  
SRLQGLGRAARRHLSA  
LDEFWGHLYDFHGQLVAE  
ARAKKLDQLFGTDQKSASSMKADSF  
GKDISSGYCMSPTAK  
MDSQMTSSLYDSLKQRT  
PGSIDSLYGLQRGSSP  
PLVNRMQML  
GAYGNTITNNNAYE  
LSERRYSSLRAPSSSE  
GWEHQQPATVHGYQMK  
SYVDNLAKERLEAL  
QSRGEIPTSRSM  
AGTLSYTQQLALALKQKS  
QNGLTPGPAPGFENFAG  
SRISIRQSERSY  
YGVPSGNTDT  
GAAVANEKKYSSMPDI  
SGLSMSARNMHL  
PNNKSGYWDPS  
SGGGGYGAS  
YGRLSNESSL  
BNLGSRVGVFPSTYDDI  
SQSRGGYRDAYS  
LPQSATTGTGSLWSR  
QPFEQF  
GVAERNGAVGEELRN  
RNPINIDNASSNVD  
AEAKLLQSF  
RHCILKLIKLEGSE  
WLFQGS  
DGVDEELIDRVAARE  
KFIYEARE  
INQVGHMGEPLI  
SSVPNCGDGC  
VWRADLIVS  
FGVW  
CIHRVLDLSLME  
SRPELWGKYTYV  
LNRLQGV  
IDPAFSKLRTP  
MTPCFCLQIP  
ASHQRASP  
TSANGMLPPAAK  
PAKPKGKCTTAV  
TLLDLIKDV  
EMAI  
SCRKGRGT  
GTAAGDVAF  
PKGKENLAS  
VLKRYKRRLSNK  
PVGMINQDGP  
GSRKNVTAYGSLG
```

```
GVAERNGAVGEELRNRSNP  
IDGVDEELIDRVAAREKFI  
YECIHRVLDLSLME  
SRPELWGKTSANGML  
PPAAKPAKPKGKCTT  
VLKRYKRRLSNK  
PVGMINQDGP
```

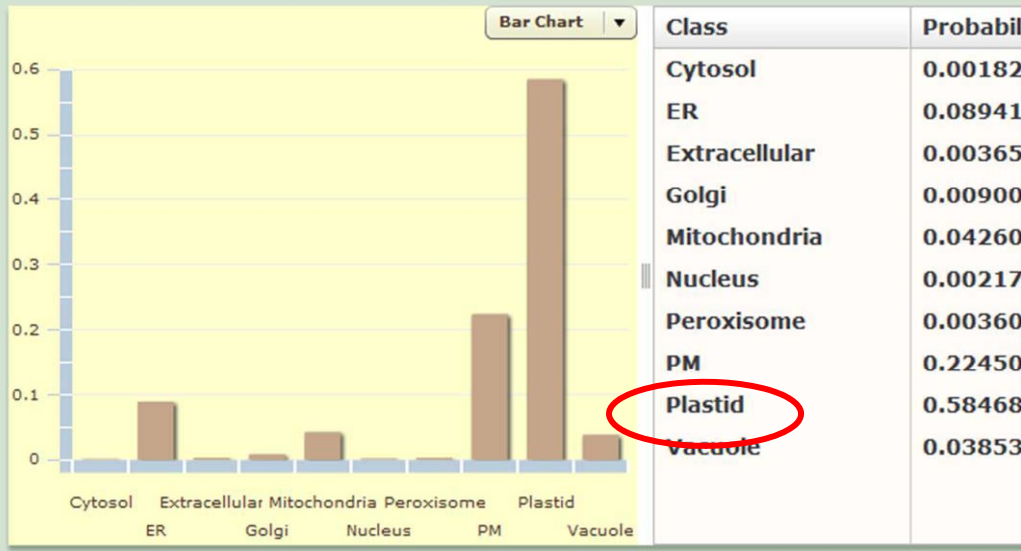




Prediction of Subcellular Localization

No.	Sequence label	First candidate		Second candidate		RI (1-10)
		Predicted location	SVM score	Predicted location	SVM score	
1	EIN2_At	secretory_pathway	0.15255575	nucleus_or_cytosol	0.08611442	1

Prediction Result of Protein AtEIN2



Protein Name = AtEIN2	
Predicted Localization Site = Plastid	
Prediction Confidence = 0.360178	
Localization Sites	Predict Probability
Cytosol	0.00182593
ER	0.0894128
Extracellular	0.00365266
Golgi	0.00900678
Mitochondria	0.0426088
Nucleus	0.00217062
Peroxisome	0.00360388
PM	0.224504
Plastid	0.584682
Vacuole	0.0385335

SLPFA

No.	Sequence label	First candidate	Second candidate
1	_EIN2_At	Nuclear_or_cytosolic(other) ---> SVM score=0.319209	Mitochondrial(mTP) ---> SVM score=0.087238

Results of your SherLoc query

ID	Predicted Location	Score
Seq 1	vacuolar	0.42

RESULTS:

#	Sequence	Subcellular localization(human protein)	Subcellular localization(eukaryotic protein)
1	>EIN2_AT	Cytoplasm	Extracellular

Conclusion

- EIN3
- Clustering
- Modification site prediction
- EIN3 vs. p53
- Microarray data analysis using KOBAS
- EIN2
- Conserved domains---- Biological Function
- Sequence motif----- Subcellular localization

Future work

- EIN3
 - Creation of point mutations;
 - Downstream responses (KOBAS);
 - Protein-protein interaction (conservative domains);
- EIN2
 - C-terminal conserved domains;
 - Localization;
 - Ion channel;

Acknowledge

- Prof. Luo
- Prof. Guo
- All the members in the class
- Our labmates

