

# 开花调控基因CONSTANS的生物信息学分析

汇报人	刘	薇
小组	G04	
组员	简银巧	刘薇友哲

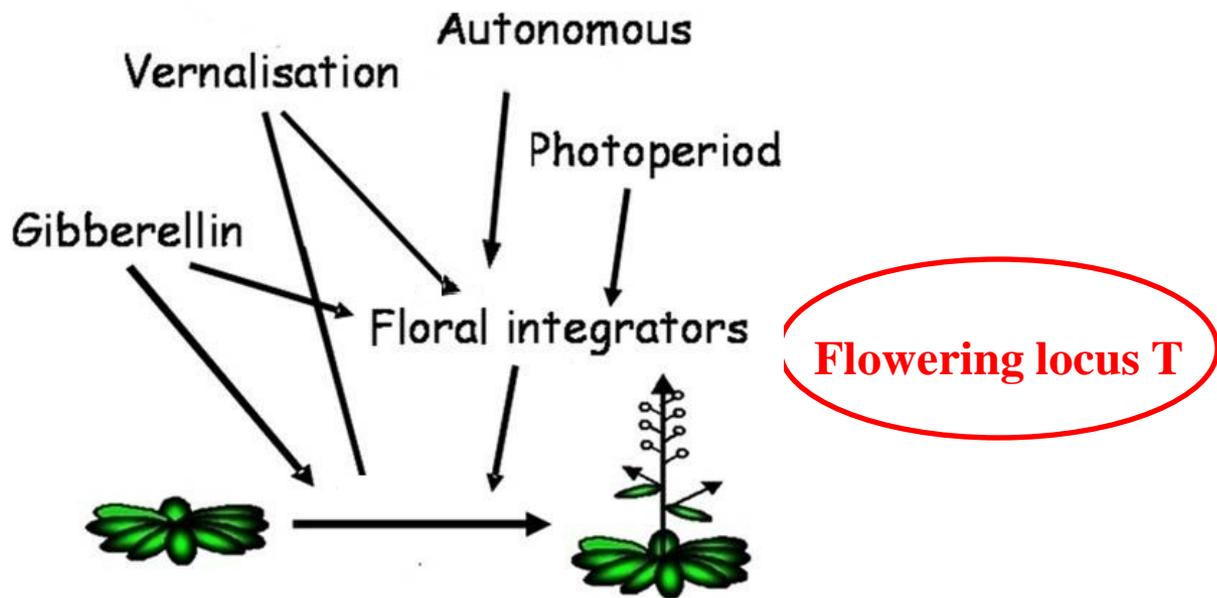
**背景介绍**

**CO序列及结构分析**

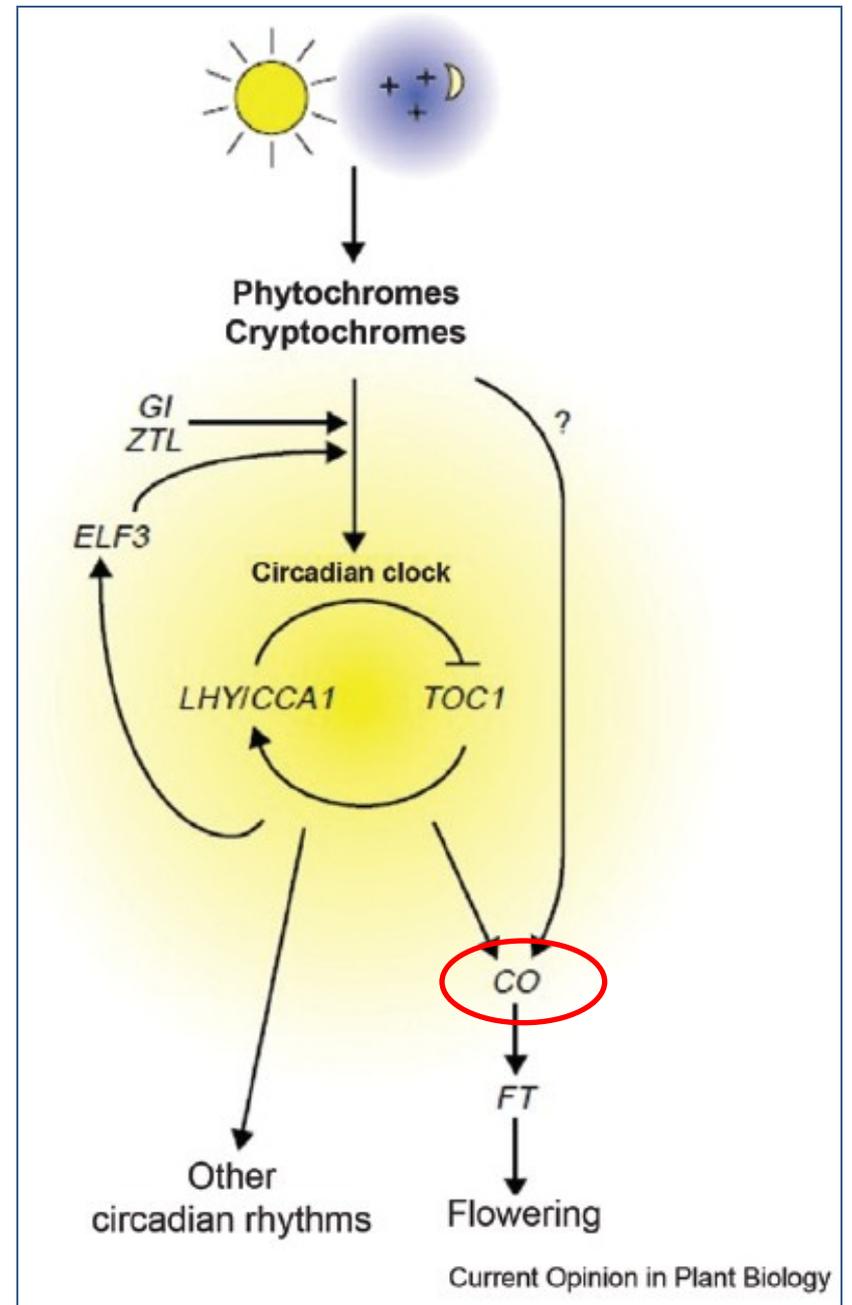
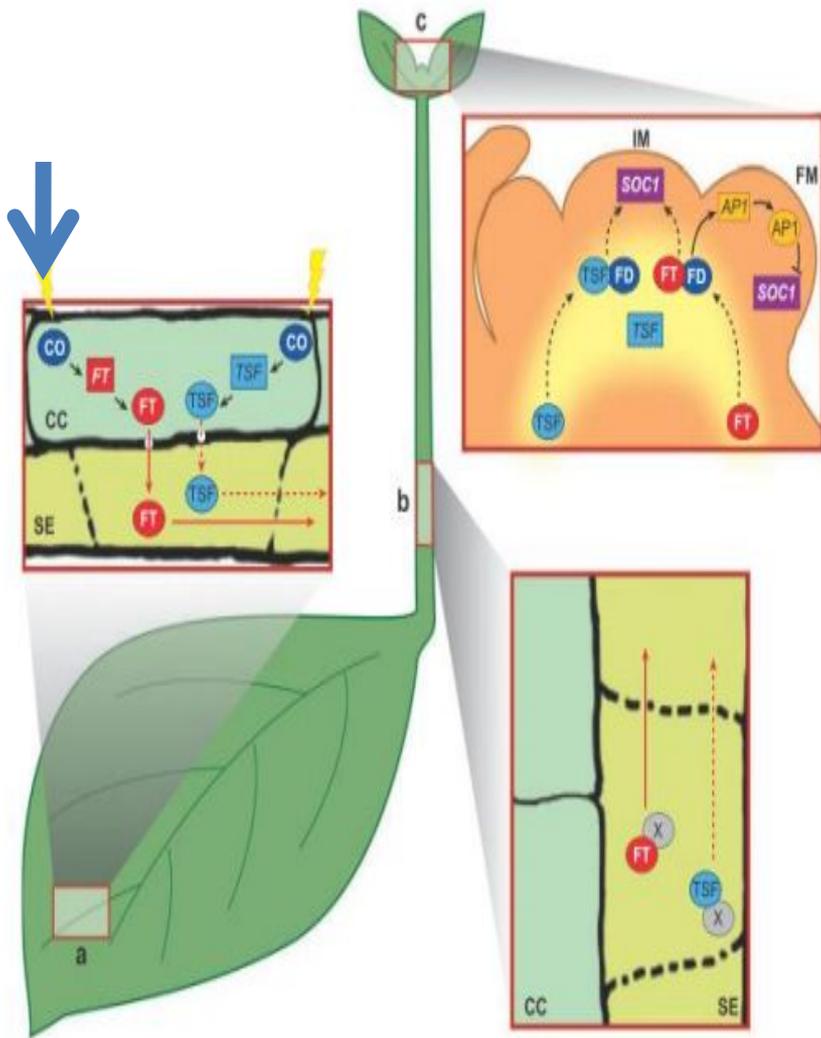
**蛋白质结构预测**

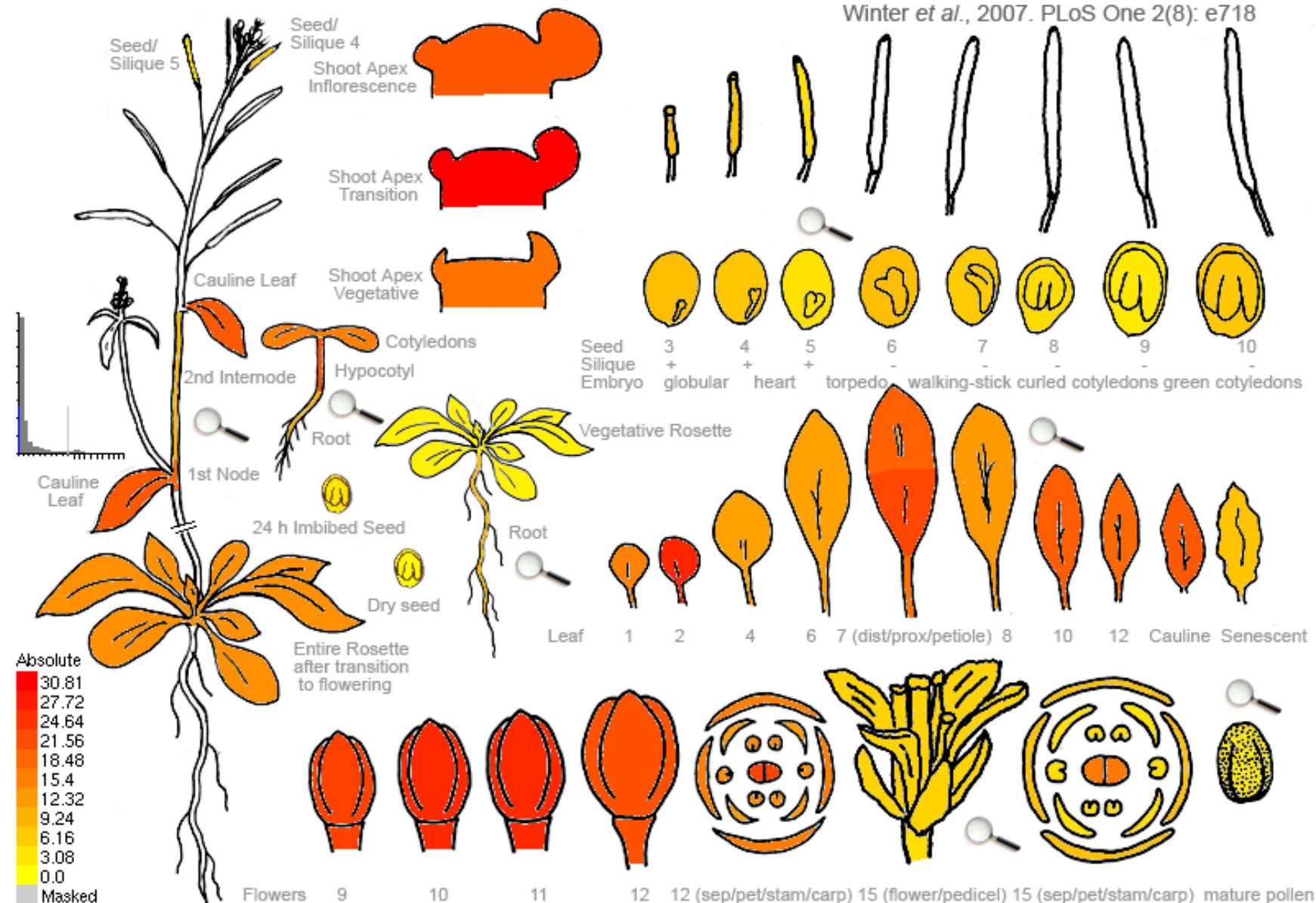
# 背景介绍

CONSTANS: 1995年从拟南芥晚花突变体中发现。  
可以整合光信号和生物钟信号，节律性地激活  
Flowering locus T 的表达，从而诱导开花。

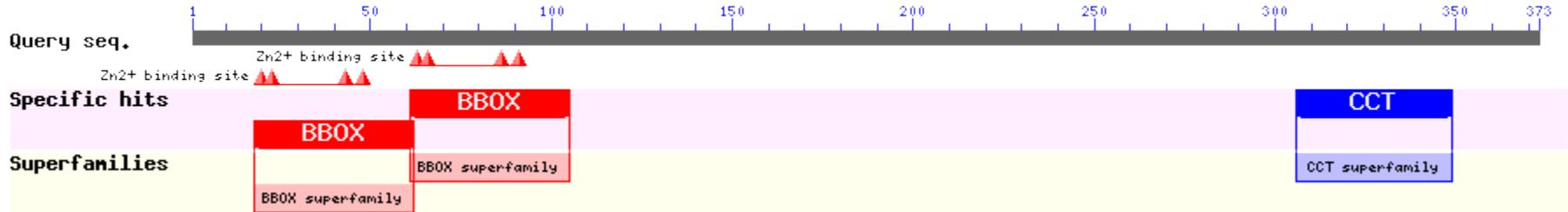


调控开花的四种途径：





# CDD

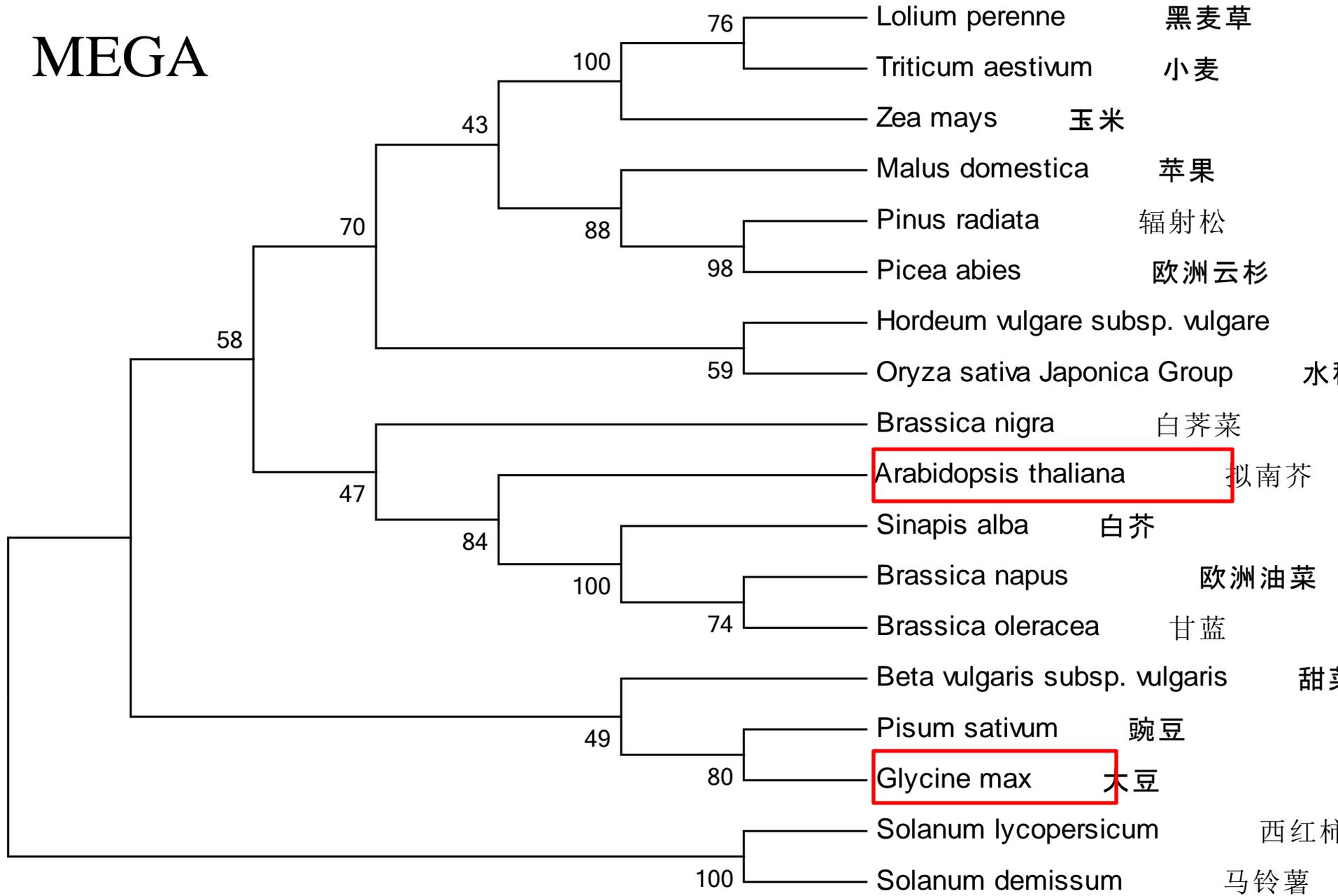


B-BOX： 锌指蛋白。

结合锌离子后能够稳定  $\alpha$ -螺旋结构，使  $\alpha$ -螺旋镶嵌于DNA的大沟中，因此含锌指结构的蛋白质都能与DNA或RNA结合。

CCT： 具有核定位信号

# MEGA



# 拟南芥CO家族成员序列比对

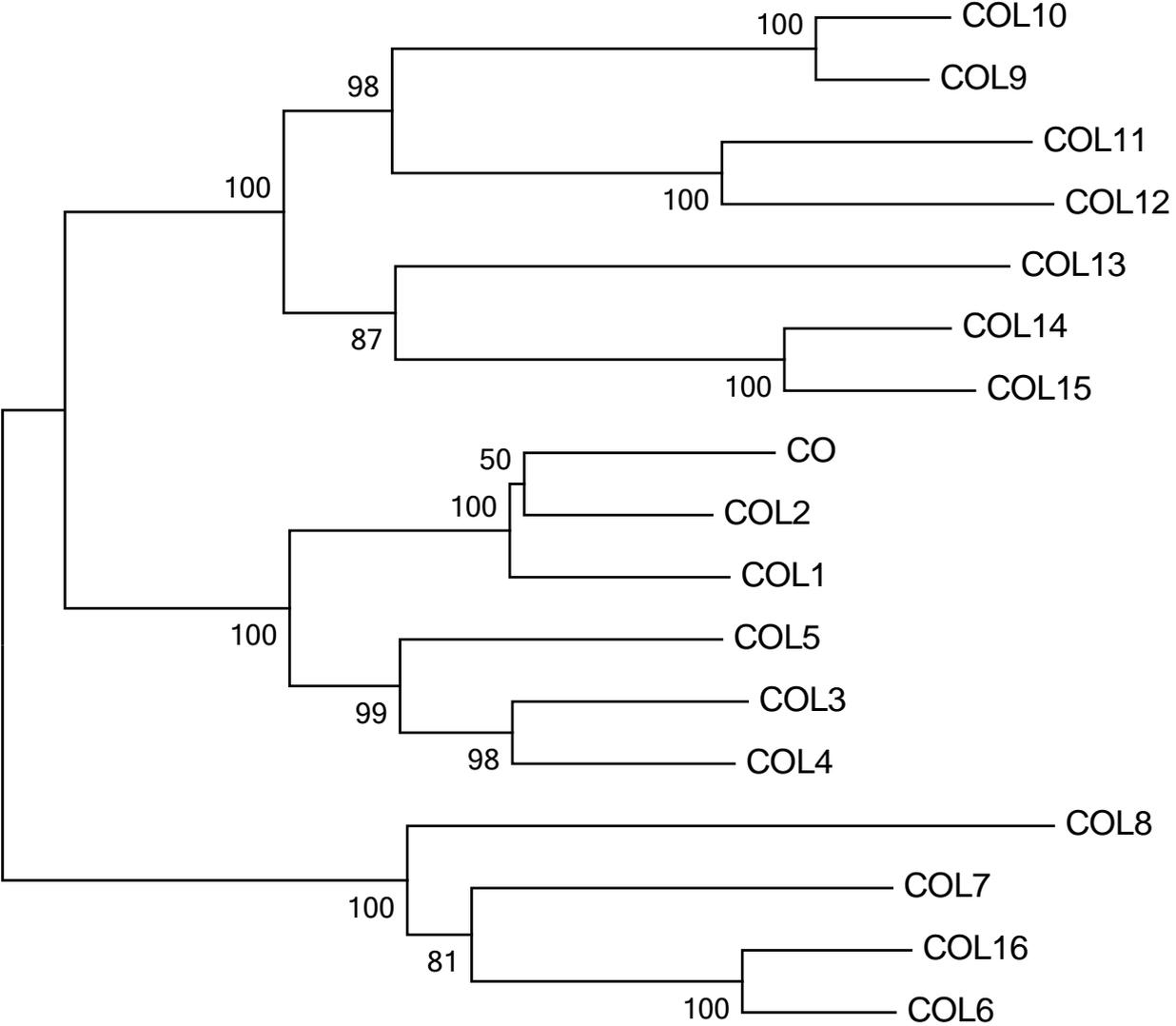
## N 端

MLKQESNDIGSGENNRARPCDTCRSN	ACTVYCHADSAYLCMSQDAQVHSANRVA SRHKRVRCESCEERAPAAFLCEADDA	LCTACDSEVHSAN	PLARRHQVRPIL	106	Q39057	CONS_ARATH
MEPKCDHCATS	QAL IYCKSDLAKLCLMCDVHVHSANPLSHRH IRLSLICEKCFSPQAAI RCLDERVS	YCQGCHWHE SMC	SELGHRVQSLN	89	Q9LJ44	COL12_ARATH
ML-----KVESNWAQA	CDTCRSA	ACTVYCRADSAYLCSSCDAQVHAANRLA SRHERVRCQSCERAPAAFFCKADAAS	LCTTCDSEIHSAN	98	050055	COL1_ARATH
MLKEESNESG-----TWARA	CDTCRSA	ACTVYCEAD SAYLCTTCDARVHAANRVA SRHERVRCQSCESAPAAFLCKADAAS	LCTACDAEIHSAN	102	Q96502	COL2_ARATH
-----MGYMCDFCGEQ	RSVVYCRSDAACLCLCDNRVHSANALSKRH SRTLVCERCNAQPA SVRCSDERSV	LCQNCDSWGHGKNS	TTTSHKRRQTI N	93	Q9LUA9	COL10_ARATH
-----MEAEEGHQDRDL	CDYCDSS	VALVYCKADSAKLCLACDKQVHVANQLPAKHFRSLLCDSNE SPSSLFCETERSV	LCQNCDSWGHHTAS	97	082256	COL13_ARATH
-----MGTSTTESVVA	CEFCGER	TAVLFCRADTAKLCLPCDQVHSANLLSRKH VRSQICDNC SKEPV SVRCFTDNLV	LCQECDWDVHG SC	98	022800	COL14_ARATH
-----MASSSRL	CDSCKST	AATLFCRADAAPFCGDCDGKIHTANKLA SRHERVWLCVEVCEQAPAHVTCADAAA	LCVTCDRDIHSAN	94	Q9SK53	COL3_ARATH
MGFGLSEIKSISGGWGAAARS	CDACKSV	TAAVFCRVD SAFLCIACTDRIHSF-----TRHERVWVCEVCEQAPAAVTCADAAA	LCVSCDADIHSAN	104	Q9FHH8	COL5_ARATH
-MFCAEIMI SKYQEDVQAPRA	CELCLNK	HAVVYCA SDDAFLCHVCDSEVHSANHVATKHERVCLRTNEI S-----NDVR-----	GGTTLT SVWH-----	92	Q9M9B3	COL8_ARATH
-----MEAR	CDFCGTE	KALIYCKSDSAKLCLMCDVHVHSANPLSQRHTRSLLECKCSLQPTAVHGMNENVS	LCQGCQWTA SMC	89	023379	COL11_ARATH
-----MSSSERVP	CDFCGER	TAVLFCRADTAKLCLPCDQVHTANLLSRKH VRSQICDNCGNEPV SVRCFTDNLV	LCQECDWDVHG SC	95	Q9C7E8	COL15_ARATH
MM-----KSLANAVG	AKTARA	CDSCVKR	RARVYCAADDAFLCQSCDSL VHSANPLARRHERVRLKTA SPA-----VVKHSMHSSA	94	Q8RWD0	COL16_ARATH
-----MA SKL	CDSCKSA	TAALYCRPDA AFLCLCD SKVHAANKLA SRHARVWVCEVCEQAPAHVTCADAAA	LCVTCDRDIHSAN	92	Q940T9	COL4_ARATH
MM-----KSLA SAVG	AKTARA	CDSCVKR	RARVYCAADDAFLCHACDC SVHSANPLARRHERVRLK SASAG-----KYRHA-----	88	Q8LG76	COL6_ARATH
MVVDVESRTASVTGEMKMAARG	CDACKRSRA	SWYCPADDAFLCQSCDA SIHSANHLAKRHERVRLQSSSPT-----ETA-----	DKTTSVWY-----	91	Q9C9A9	COL7_ARATH
-----MGYMCDFCGEQ	RSVVYCRSDAACLCLCDNRVHSANALSKRH SRTLVCERCNAQPA TVRCVEERVS	LCQNCDSWGHNSNNNNSSSSSTSPQKHRRQTI S		100	Q9SSE5	COL9_ARATH

VEQQP----	DPASQMITVTIQLSPND	FEARVRLYRREKRETRREFEET IRYASREAYAE I RPRVNGRFAKREIEA--	FEQGF-NTMLMWTGYGIVPSF	373	Q39057	CONS_ARATH	
ESEA--P-WEIN----	FEVSC-PQARWFAKLRYEPEKELERS	FGDQ IRYASREKARADTRERVEGRFVVEAGISYDVPSSPT-----	TNN-----	364	Q9LJ44	COL12_ARAI	
TIQFP----	YPPAQE----	LSPRD	FEARVRLYRREKRETRREFEET IRYASREAYAEERPRI EGRFAKRE	EDVDEFAWQAF-STMIIFDTGYGIVPSF	355	050055	COL1_ARATH
IDQLS----	GPPTQV--	VQQLTP	FEARVRLYRREKRETRREFDET IRYASREAYAE I RPRI EGRFAKRE	IETFAAEFIIF-STSLMS EITGYGIVPSF	347	Q96502	COL2_ARATH
LSREPQP-WCHPTAQDI	IASSHATT	FWAVRFRYRREKRETRREFDERVRYVS REERADVRRRVEGRFVVEAG	GEAYDYDVPSSPT-----	RSY-----	373	Q9LUA9	COL10_ARAI
TVMVP----	VITSTR	WHEINSLERWS ALSRYEPEKESRRYERH IRYESREVRAPESRIRI EGRFAKRE	ADP-----	332	082256	COL13_ARAI	
IAITSWRATRLVAVINADLEQ	WAGN	WVAVRFRYRREKRETRRYDET IRYETRKAETRLRVEGRFVVEAG	TDF-----	402	022800	COL14_ARAI	
-----TN-----	T-----	EIPAVQLSPA	FEARVRLYRREKRETRREFEET IRYASREAYAEERPRI EGRFAKRE	IDSREND--GGDV--G VYGGFGVVPSP	294	Q9SE53	COL5_ARATH
T-----	-----	GGIEHQASS	FEARVRLYRREKRETRREFEET IRYASREAYAESRPRI EGRFAKRE	IETENDDIFLSHVYASAAHAQYGVVPTF	355	Q9FHH8	COL5_ARATH
PG-----	-----	IPEERVRS	FEARVRYRDERERLFEKI IRYEVREVAADERPRVEGRFVVEAG	SLAIDS-----	319	Q9M9B3	COL8_ARATH
IGIS--P-WESN----	VEVSNPEL	RDEAKERYEQEESERWFGDQ IRYASREKARADTRERVEGRFVVEAG	GETIFEYDPSLVM-----	330	023379	COL11_ARAI	
IAGTSCETITRLVA	TKADLERLAGN	FGDAWFRYRREKRETRRYDET IRYESREKARADTRRVEGRFVVEAG	SEAPYP-----	433	Q9C7E8	COL15_ARAI	
MGGESIEHQEQYVGGCLPSSGFGDGG	FEARVRSRYRREKRETRLFS	EKI IRYEVRELVAEERPRVEGRFVVEAG	ASLAAAASPLGVWY-----	417	Q8RWD0	COL16_ARAI	
SGADP----	G-----	TQRAVPLTSA	FEARVRYRREKRETRREFEET IRYASREAYAEERPRI EGRFAKRE	ITDINESDWDVHG--GIFSGFLVPTF	362	Q940T9	COL4_ARATH
SGAAHHHHHFRGLGHLGDAGDGG	FEARVRSRYRREKRETRLFS	EKI IRYEVRELVAEERPRVEGRFVVEAG	SSIGVAH-----	406	Q8LG76	COL6_ARATH	
YEKLSSSDGSVTRQGRDGGGSDGE	FEARVRLYRREKRETRLFS	EKI IRYEVRELVAEERPRI EGRFVVEAG	ISLIT-----	392	Q9C9A9	COL7_ARATH	
LSGEP-P-WYPPTLQINWACSB	SVT	FWAVRFRYRREKRETRREFDERVRYASREKARADVRRRVEGRFVVEAG	GEAYDYDPLIPT-----	RSY-----	372	Q9SSE5	COL9_ARATH

## C 端

# 拟南芥CO家族成员进化树分析

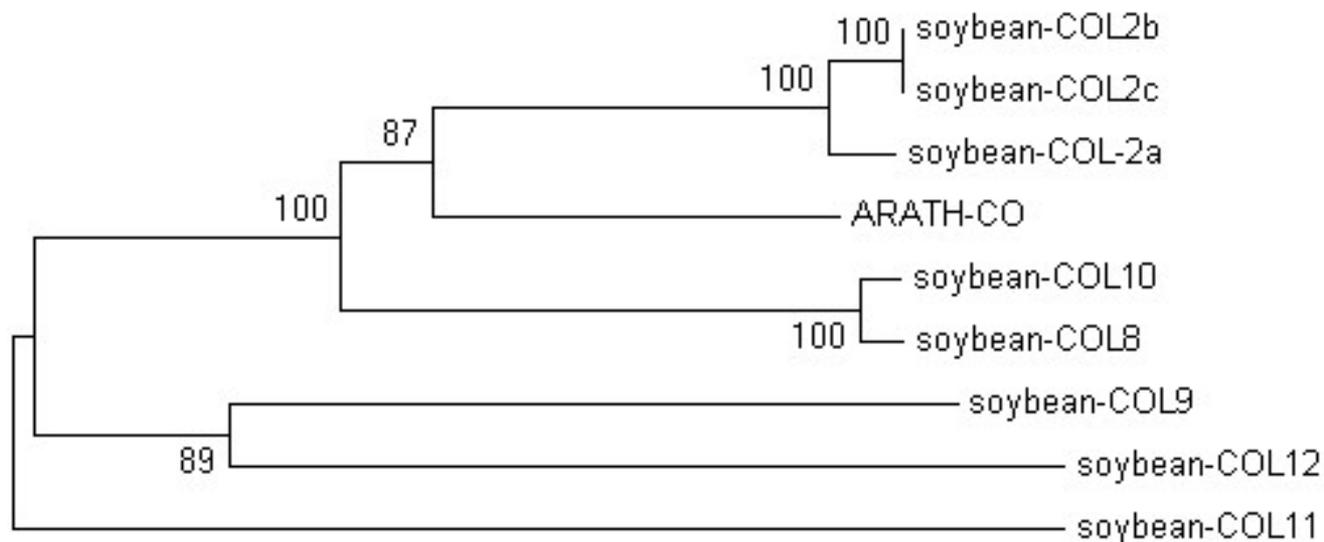


0.1

- 大豆是典型的短日作物，不同品种早晚有较大的差异。
- 影响大豆开花候选基因的克隆，表达分析，转基因研究，鉴定其在成花转变中的作用，对探究基因与开花整合因子FT的相互关系以及调节大豆开花的分子网络的描绘有重要意义。

# 大豆CO-Like基因分析

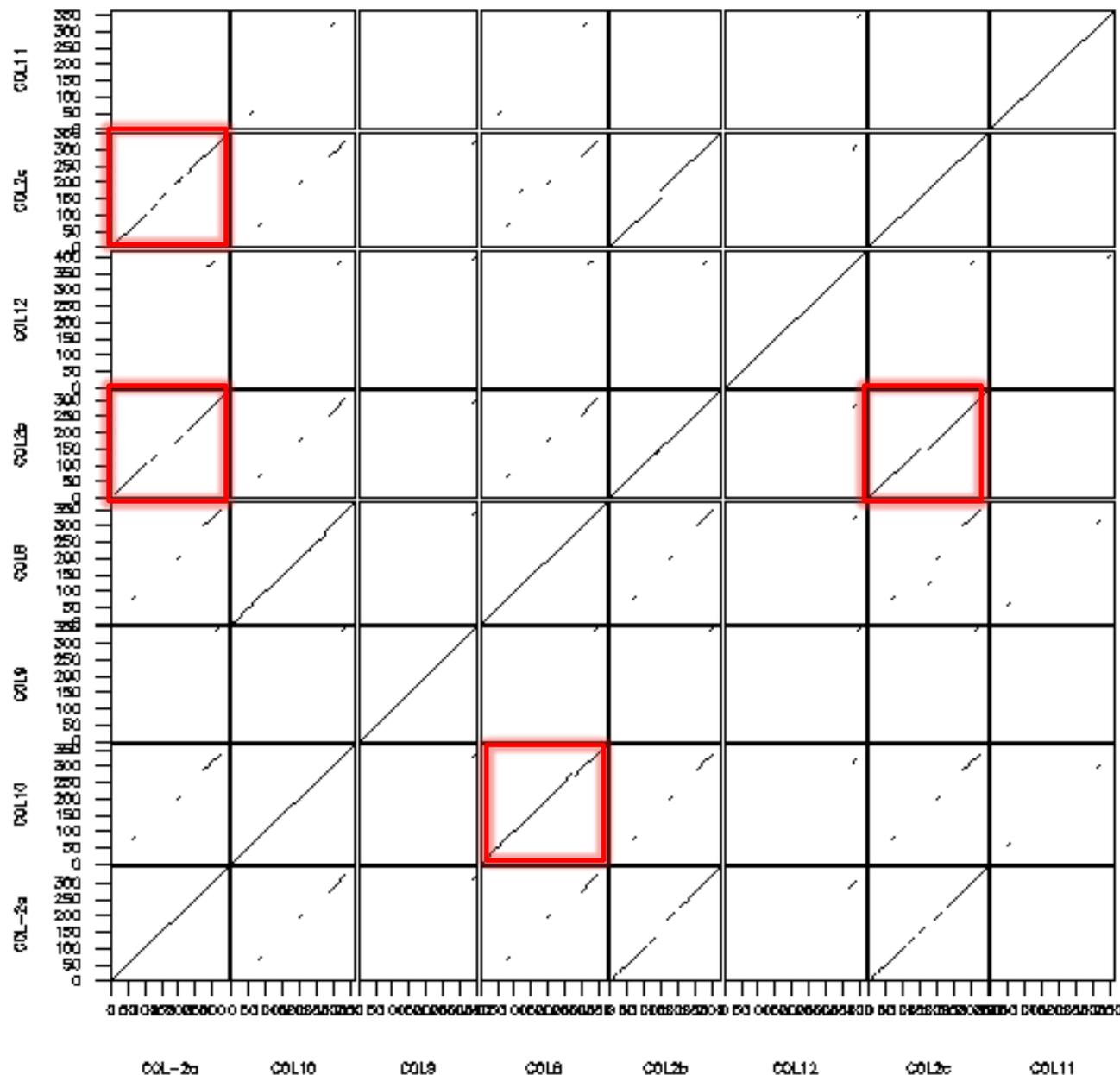
Entry	Entry name	Status	Protein names	Gene names	Organism	Length
<input checked="" type="checkbox"/> B9MSS5	B9MSS5_SOYBN	★	CONSTANS-like 2a	COL2a	Glycine max (Soybean) (Glycine hispida)	348
<input checked="" type="checkbox"/> D0EP05	D0EP05_SOYBN	★	CONSTANS-like zinc finger protein	COL10	Glycine max (Soybean) (Glycine hispida)	365
<input checked="" type="checkbox"/> D0EP04	D0EP04_SOYBN	★	CONSTANS-like zinc finger protein	COL9	Glycine max (Soybean) (Glycine hispida)	351
<input checked="" type="checkbox"/> D0EP03	D0EP03_SOYBN	★	CONSTANS-like zinc finger protein	COL8	Glycine max (Soybean) (Glycine hispida)	374
<input checked="" type="checkbox"/> B9MSS6	B9MSS6_SOYBN	★	CONSTANS-like 2b	COL2b	Glycine max (Soybean) (Glycine hispida)	328
<input checked="" type="checkbox"/> D0EP07	D0EP07_SOYBN	★	CONSTANS-like zinc finger protein	COL12	Glycine max (Soybean) (Glycine hispida)	418
<input checked="" type="checkbox"/> B9MSS7	B9MSS7_SOYBN	★	Zinc finger protein CONSTANS-LIKE 2	COL2c	Glycine max (Soybean) (Glycine hispida)	352
<input checked="" type="checkbox"/> D0EP06	D0EP06_SOYBN	★	CONSTANS-like zinc finger protein	COL11	Glycine max (Soybean) (Glycine hispida)	366

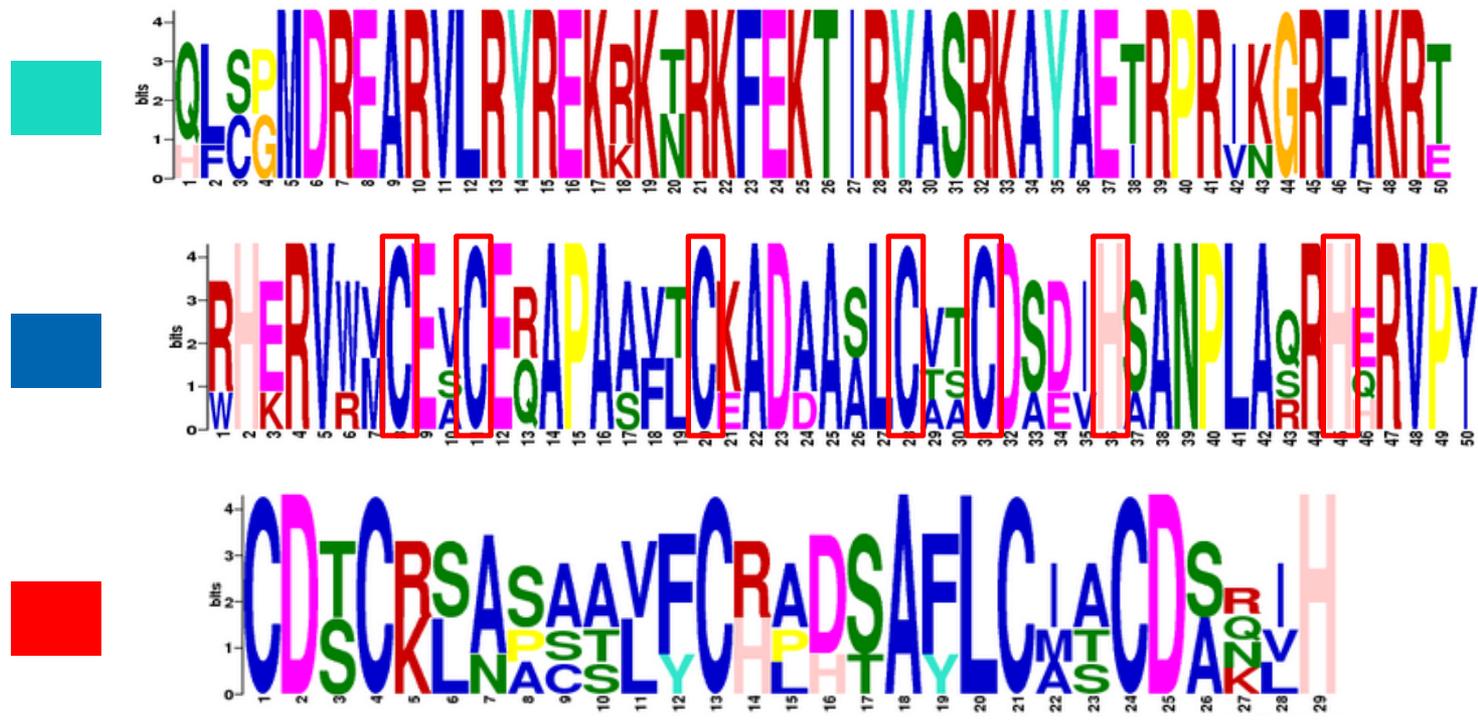
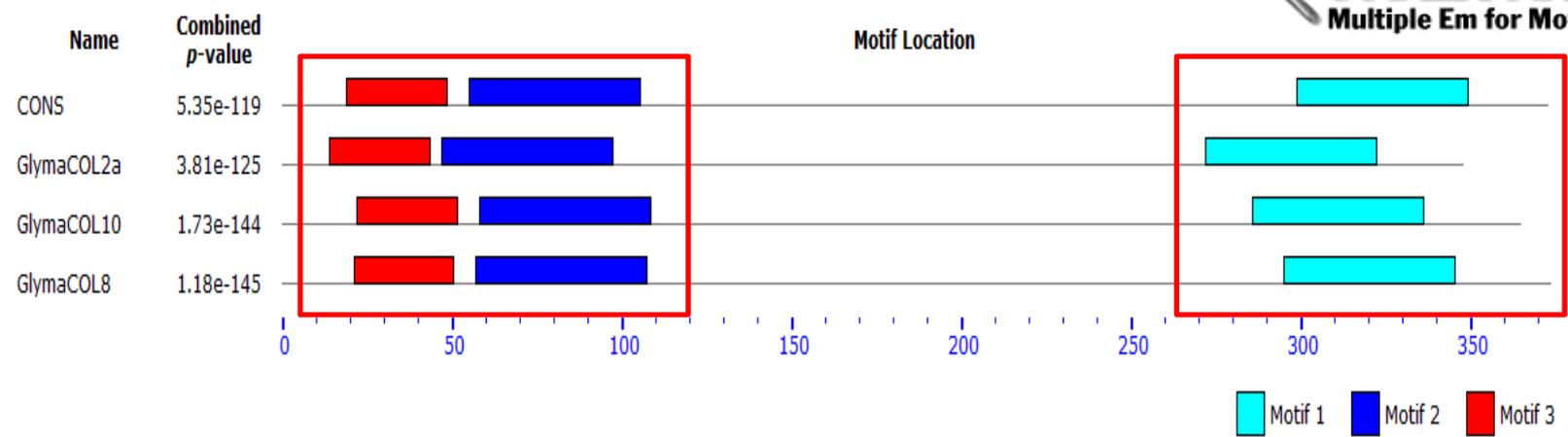


0.1

# Polydot

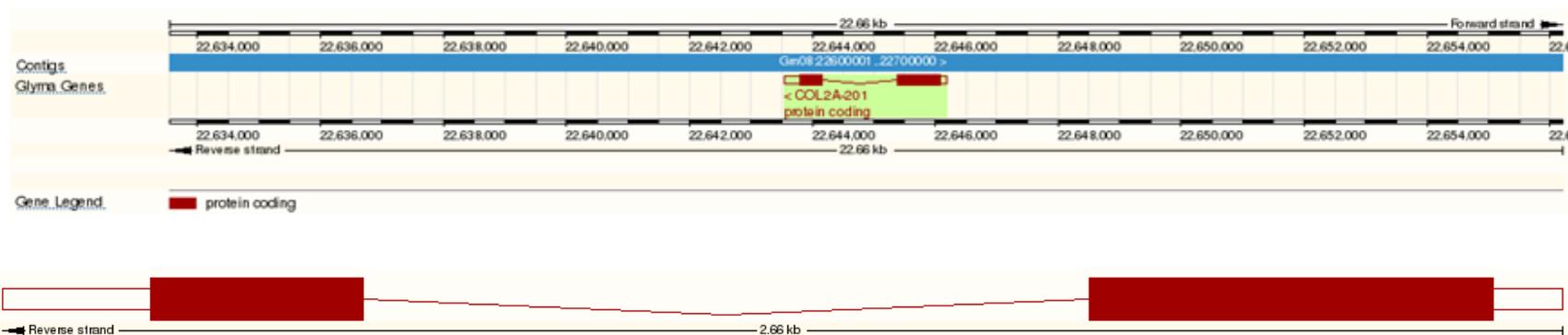
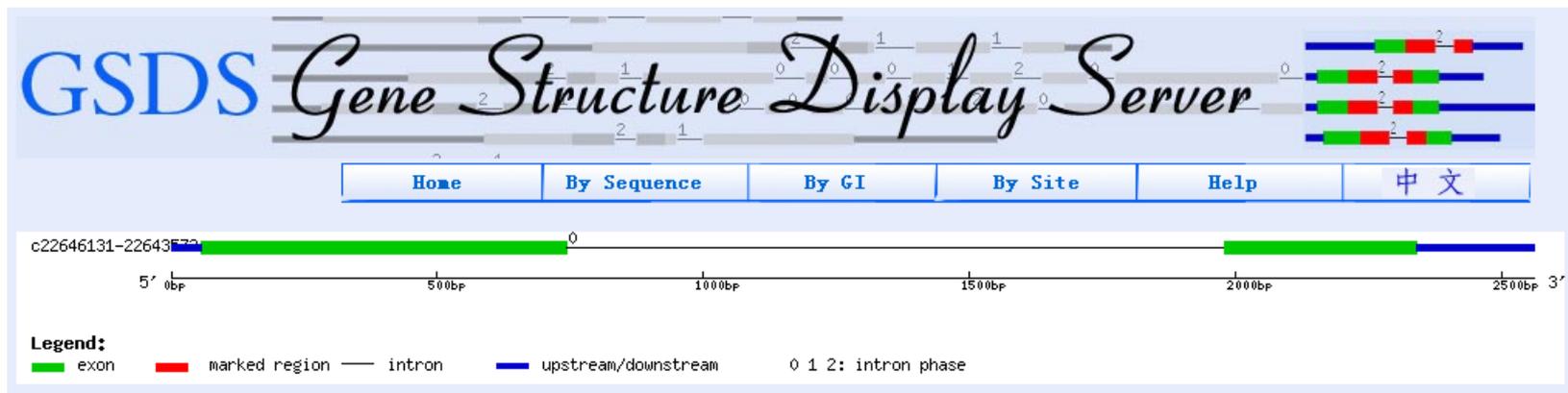
COL8和COL10  
之间，COL2a  
和COL2b和  
COL2c之间相  
似度很高





**B-box :  $CX_2CX_8CX_7CX_2CX_4HX_8H$**

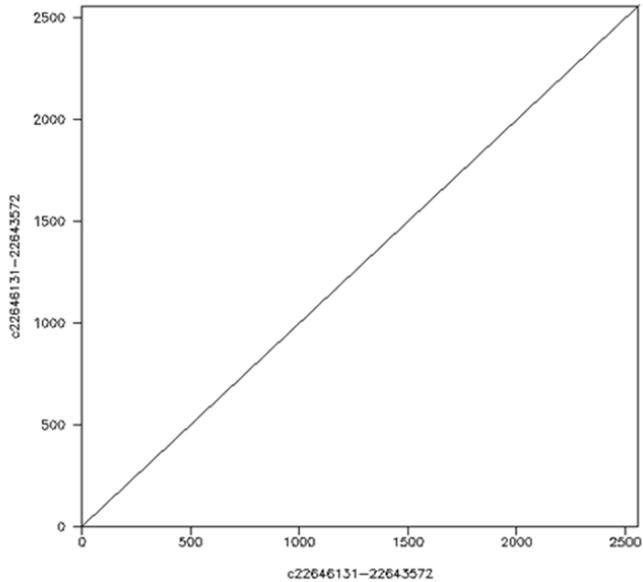
# GmCOL2a的基因结构



# GmCOL2a的重复序列分析

Dotpath: fasta::6039:c22646131-22643572 vs fasta::6040:c

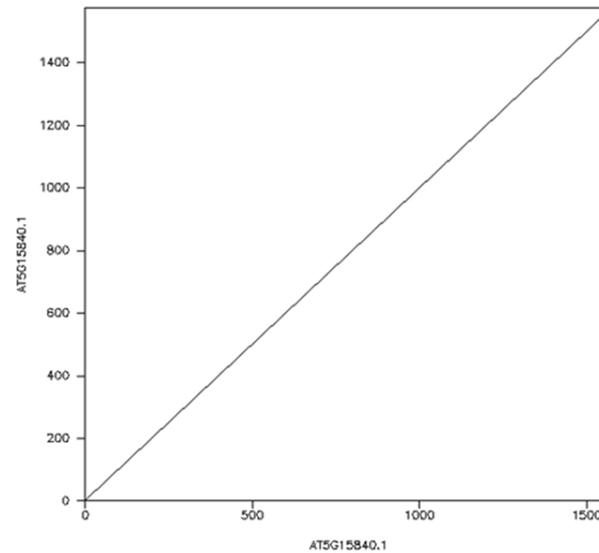
Sat 30 Nov 2013 13:51:05



大豆

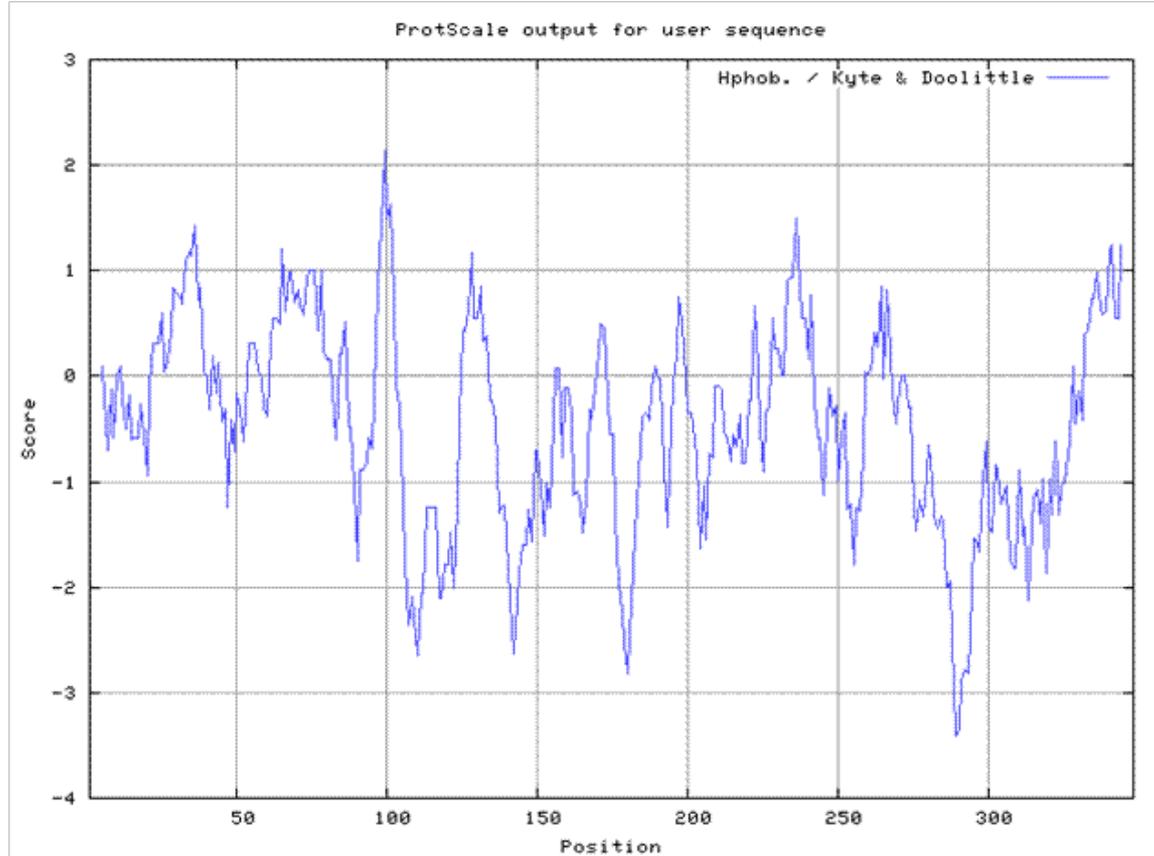
Dotpath: fasta::6019:AT5G15840.1 vs fasta::6020:AT5G1584.

Sat 30 Nov 2013 12:47:13

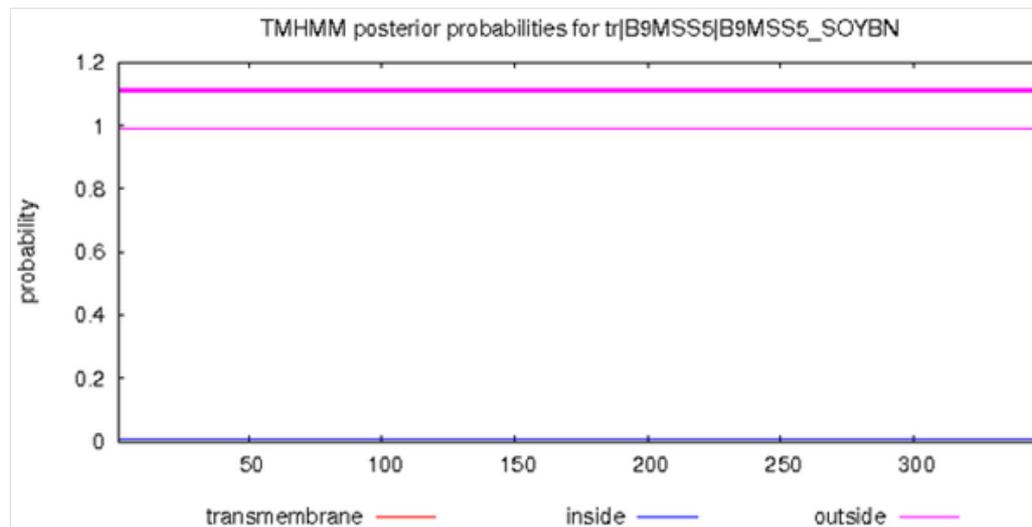


拟南芥

# ProScale



# TMHMM



# GmCOL2a的亚细胞定位预测



## TargetP 1.1 Server - prediction results

Technical University of Denmark

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

Name	Len	cTP	mTP	SP	other	Loc	RC
Soybean_COL_2a	348	0.144	0.262	0.018	0.680	_	3
cutoff		0.000	0.000	0.000	0.000		

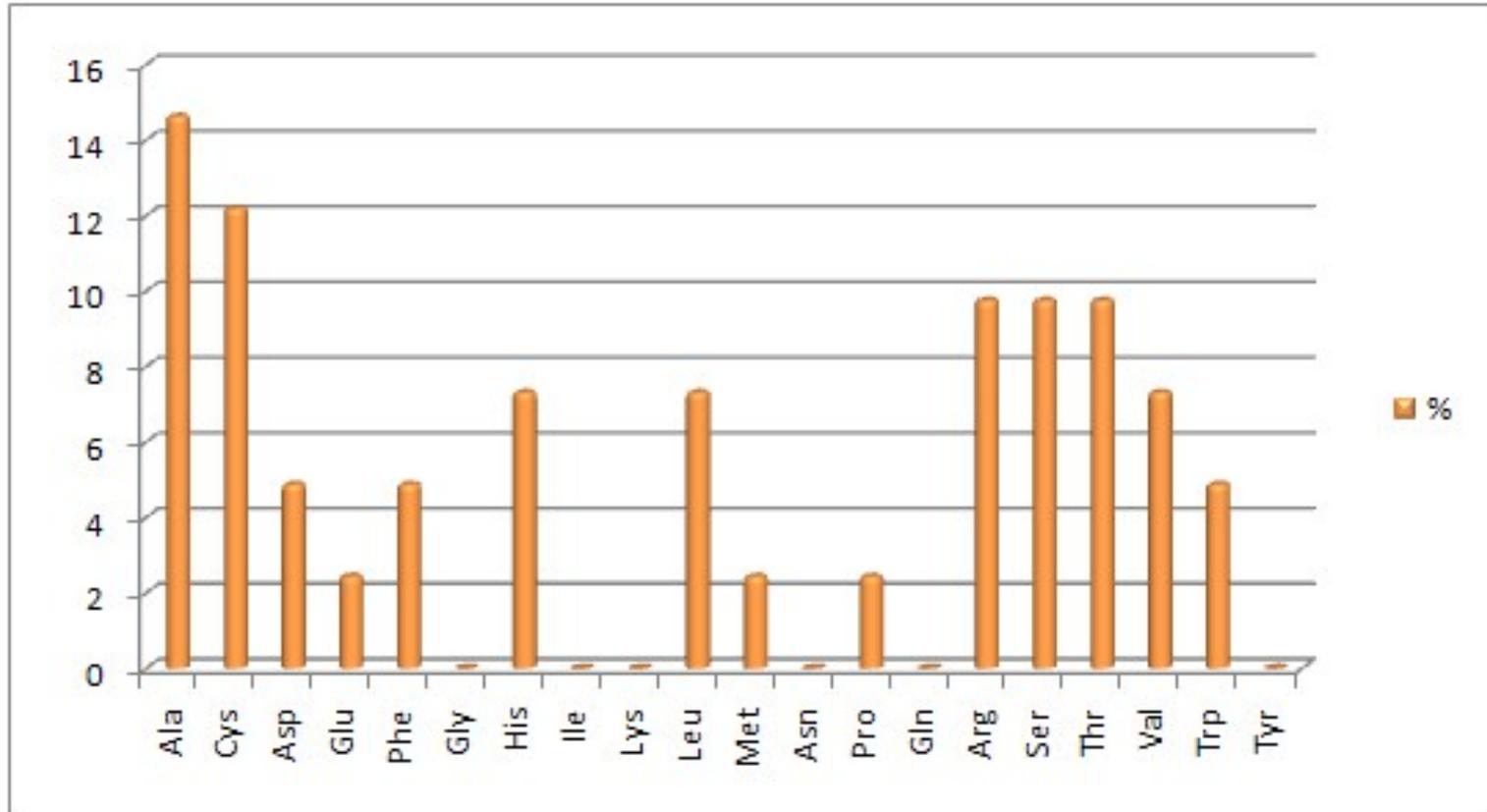
# GmCOL2a的氨基酸组分分析

Pepstats

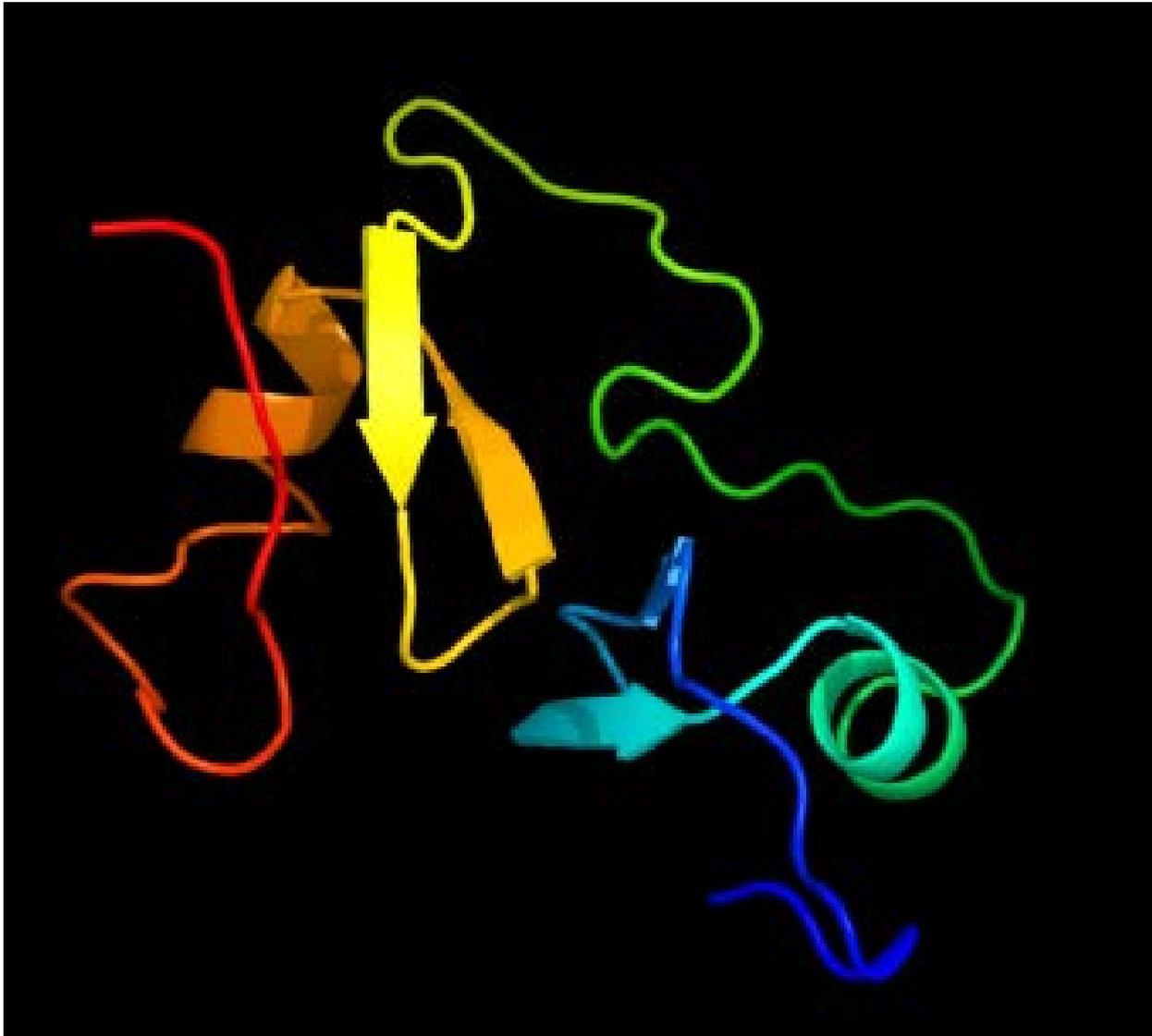
AA	SwissProt*	Glymax-COL2a	
	%	number	%
Ala	8.25	38	10.92
Cys	1.37	11	3.161
Asp	5.45	29	8.333
Met	2.42	7	2.011
Phe	3.86	15	4.31
Gln	3.93	8	2.299
His	2.27	18	5.172
Tyr	2.92	9	2.586
Ile	5.95	11	3.161
Lys	5.84	13	3.736
Gly	7.07	15	4.31
Asn	4.06	17	4.885
Pro	4.7	18	5.172
Glu	6.75	16	4.598
Arg	5.53	21	6.034
Ser	6.57	31	8.908
Thr	5.34	21	6.034
Val	6.87	27	7.759
Trp	1.08	4	1.149
Leu	9.66	19	5.46

# Pepstats

## GmCOL2a B-box区氨基酸组分分析







$CX_2CX_8CX_7CX_2CX_4HX_8H$



ool2

visible  can move

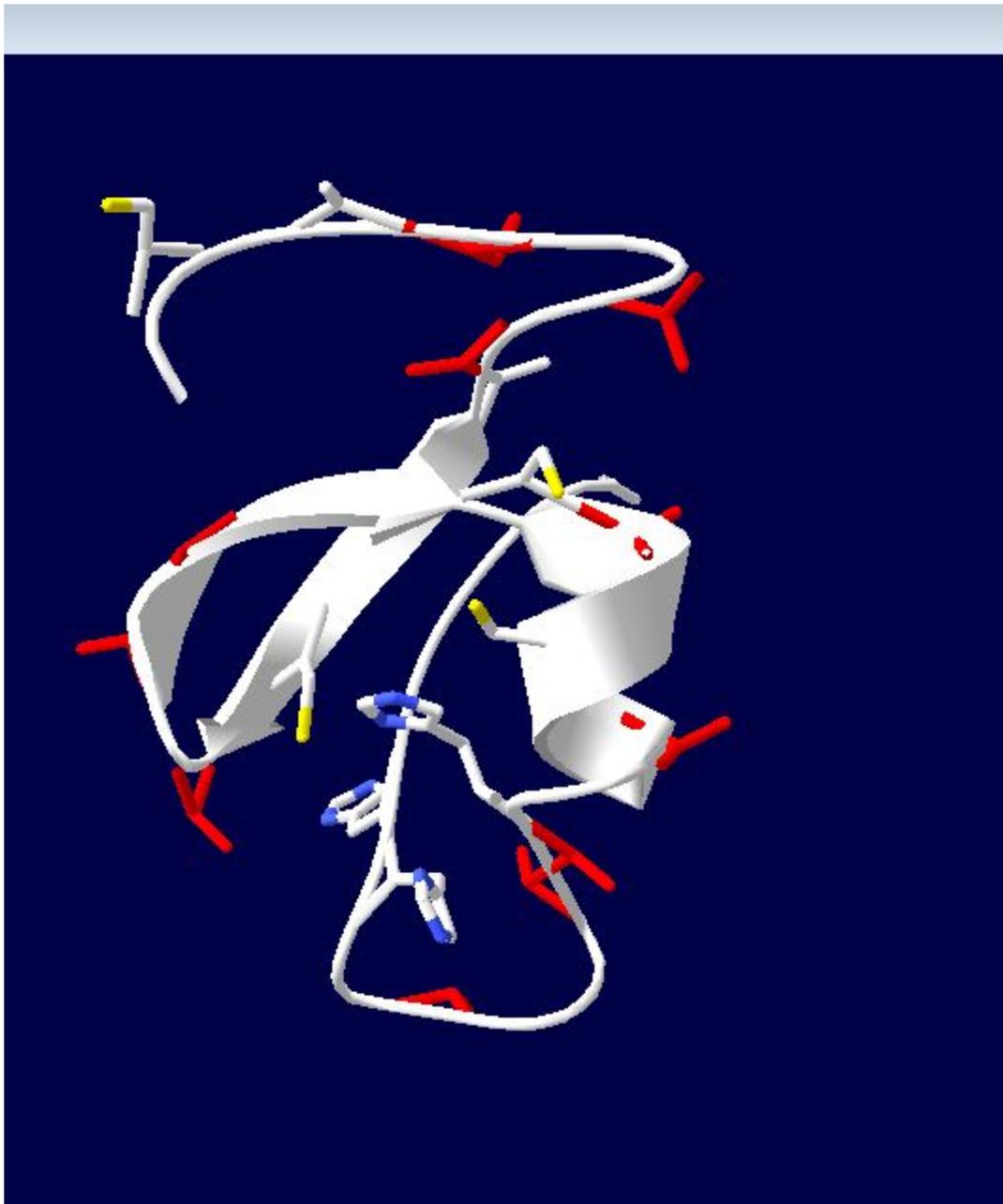
group show side labl ribn colgc

MET14	v								
CYS15	v	v							
ASP16	v								
THR17	v								
CYS18	v	v							
ARG19	v								
SER20	v								
ALA21	v								
PRO22	v								
SER23	v								
SER24	v								
s VAL25	v								
s PHE26	v								
s CYS27	v	v							
s ARG28	v								
ALA29	v								
HIS30	v	v							
s THR31	v								
s ALA32	v								
s PHE33	v								
s LEU34	v								
CYS35	v	v							
h ALA36	v								
h THR37	v								
h CYS38	v	v							
h ASP39	v								
h ALA40	v								
h ARG41	v								
h LEU42	v								
HIS43	v	v							
ALA44	v								
SER45	v								
LEU46	v								
THR47	v								
TRP48	v								
HIS49	v	v							
GLU50	v								
ARG51	v								

**CX<sub>2</sub>CX<sub>8</sub>CX<sub>7</sub>CX<sub>2</sub>CX<sub>4</sub>HX<sub>8</sub>H**



group	+	+	+	+	+
	show	side	labl	ry	ribn
					oolg
VAL52	v				<input type="checkbox"/>
TRP53	v				<input type="checkbox"/>
VAL54	v				<input type="checkbox"/>
CYS55	v	v			<input type="checkbox"/>
GLU56	v				<input type="checkbox"/>
ALA57	v				<input type="checkbox"/>
CYS58	v	v			<input type="checkbox"/>
GLU59	v				<input type="checkbox"/>
ARG60	v				<input type="checkbox"/>
ALA61	v				<input type="checkbox"/>
PRO62	v				<input type="checkbox"/>
ALA63	v				<input type="checkbox"/>
ALA64	v				<input type="checkbox"/>
s PHE65	v				<input type="checkbox"/>
s LEU66	v				<input type="checkbox"/>
s CYS67	v	v			<input type="checkbox"/>
s LYS68	v				<input type="checkbox"/>
ALA69	v				<input type="checkbox"/>
ASP70	v				<input type="checkbox"/>
s ALA71	v				<input type="checkbox"/>
s ALA72	v				<input type="checkbox"/>
s SER73	v				<input type="checkbox"/>
s LEU74	v				<input type="checkbox"/>
CYS75	v	v			<input type="checkbox"/>
h ALA76	v				<input type="checkbox"/>
h SER77	v				<input type="checkbox"/>
h CYS78	v	v			<input type="checkbox"/>
h ASP79	v				<input type="checkbox"/>
h ALA80	v				<input type="checkbox"/>
ASP81	v				<input type="checkbox"/>
ILE82	v				<input type="checkbox"/>
HIS83	v	v			<input type="checkbox"/>
ALA84	v				<input type="checkbox"/>
ALA85	v				<input type="checkbox"/>
ASN86	v				<input type="checkbox"/>
PRO87	v				<input type="checkbox"/>
LEU88	v				<input type="checkbox"/>
ALA89	v				<input type="checkbox"/>
SER90	v				<input type="checkbox"/>
ARG91	v				<input type="checkbox"/>
HIS92	v	v			<input type="checkbox"/>
HIS93	v	v			<input type="checkbox"/>
ARG94	v				<input type="checkbox"/>



visible	?		can move	
group	+	+	+	+
	show	side	labl	ribn
			oolg	
CYS55	v	v		v
GLU56				v
ALA57	v	v		v
CYS58	v	v		v
GLU59				v
ARG60				v
ALA61	v	v		v
PRO62				v
ALA63	v	v		v
ALA64	v	v		v
s PHE65				v
s LEU66				v
s CYS67	v	v		v
s LYS68				v
ALA69	v	v		v
ASP70				v
s ALA71	v	v		v
s ALA72	v	v		v
s SER73				v
s LEU74				v
CYS75	v	v		v
h ALA76	v	v		v
h SER77				v
h CYS78	v	v		v
h ASP79				v
h ALA80	v	v		v
ASP81				v
ILE82				v
HIS83	v	v		v
ALA84	v	v		v
ALA85	v	v		v
ASN86				v
PRO87				v
LEU88				v
ALA89	v	v		v
SER90				v
ARG91				v
HIS92	v	v		v
HIS93	v	v		v
ARG94				v
VAL95				v
PRO96				v
ILE97				v

**Thank you !**