

Exploring the transcriptional regulation of Arabidopsis pollen germination

2018/01/14

报告人：张立光、黄勋

组员

G10: 傅偲、黄勋、张二禾、俞琴琴

G15: 杜丽萍、郑馨慧、贾名康、张东亮

G19: 张立光、周梦璇、王妍妍

Outline

1. Promoter analysis of *JINGUBANG*

- 1.1 How to find the promoter of a gene .
- 1.2 Prediction of Transcription factors.

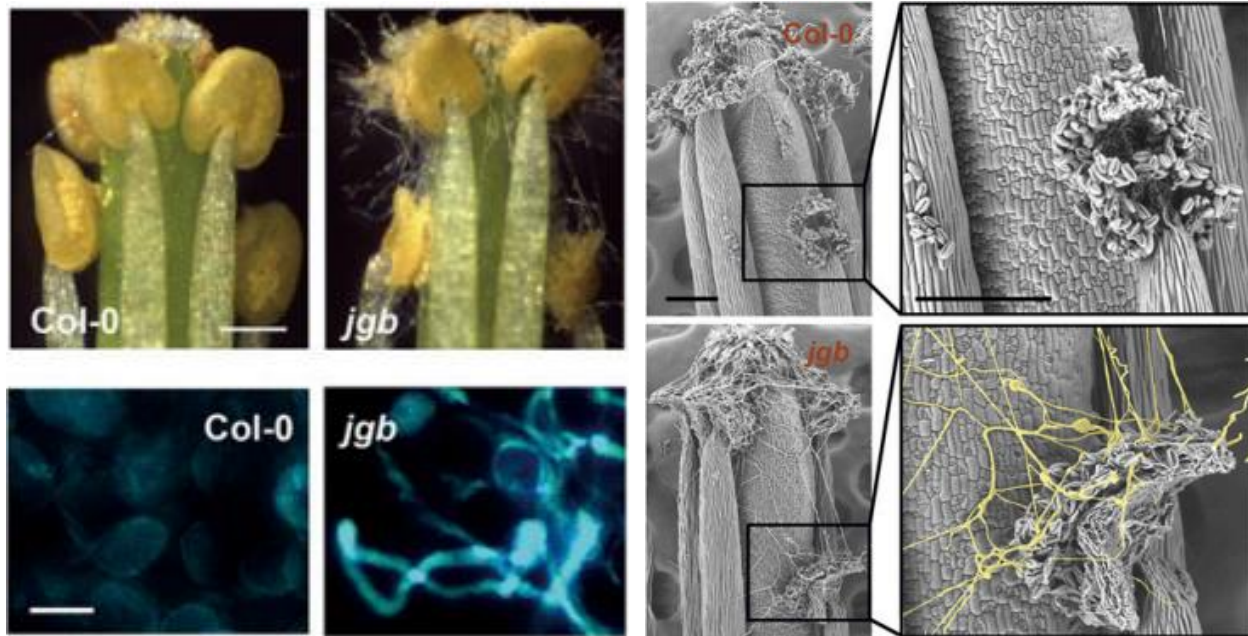
2. Introduction to MADS family

- 2.1 MADS gene evolution in land plants.

3. MADS Transcription factors

- 3.1 Overview of MADS family in *Arabidopsis thaliana*.
- 3.2 MIKC*-type genes in *A. thaliana*

JGB, a Negative Regulator of Pollen Germination



jgb mutant pollen germinated faster than wild type in moist environment.

(Ju *et al.*, 2016)

Outline

1. Promoter analysis of *JINGUBANG*

1.1 How to find the promoter of a gene .

1.2 Prediction of Transcription factors.

2. Introduction to MADS family

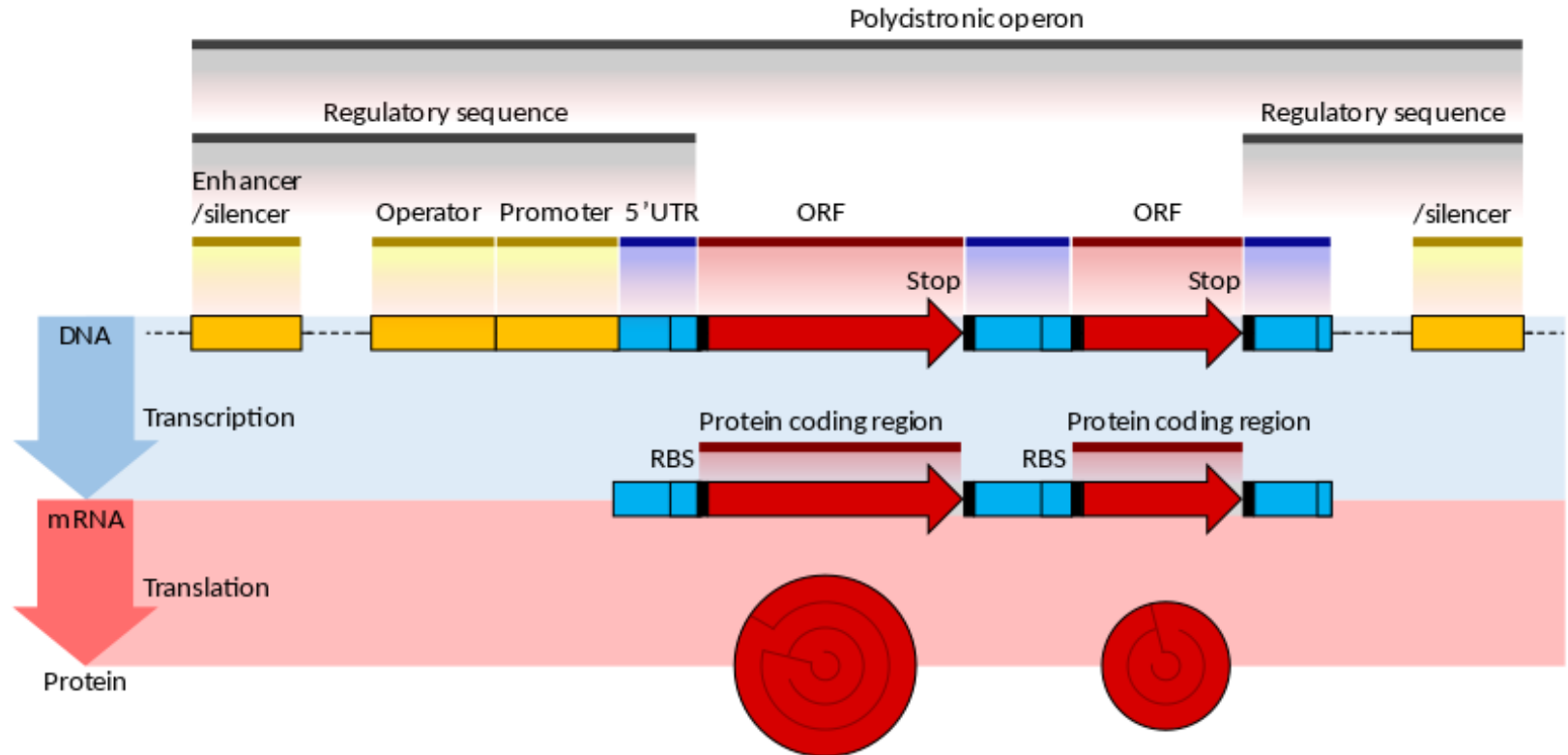
2.1 MADS gene evolution in land plants.

3. MADS Transcription factors

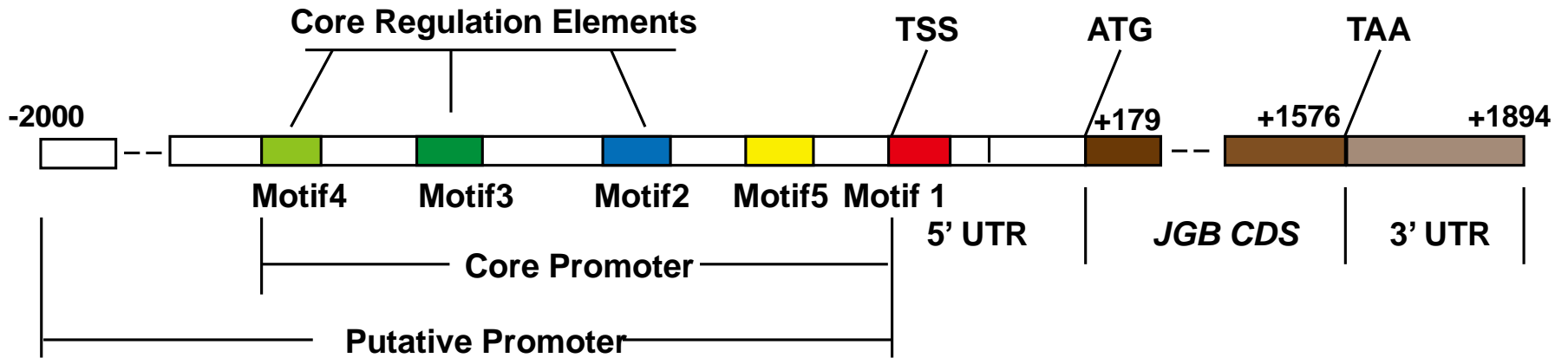
3.1 Overview of MADS family in *Arabidopsis thaliana*.

3.2 MIKC*-type genes in *A. thaliana*.

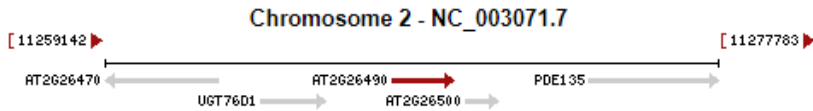
The Gene Structure of Bacteria



The Gene Structure of *JGB*



How to Find Promoter



GenBank

Change region shown

Whole sequence (abbreviated view)

Selected region

from: 11267857 to: 11269750

Update View

Customize view

Abbreviated view

Customize

Basic Features

Default features

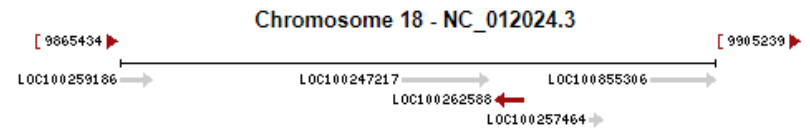
Gene, RNA, and CDS features only

Display options

Show sequence

Show reverse complement

Update View



GenBank

Change region shown

Whole sequence (abbreviated view)

Selected region

from: 9890439 to: 9892407

Update View

Customize view

Abbreviated view

Customize

Basic Features

Default features

Gene, RNA, and CDS features only

Display options

Show sequence

Show reverse complement

Update View

Outline

1. Promoter analysis of *JINGUBANG*

1.1 How to find the promoter of a gene .

1.2 Prediction of Transcription factors.

2. Introduction to MADS family

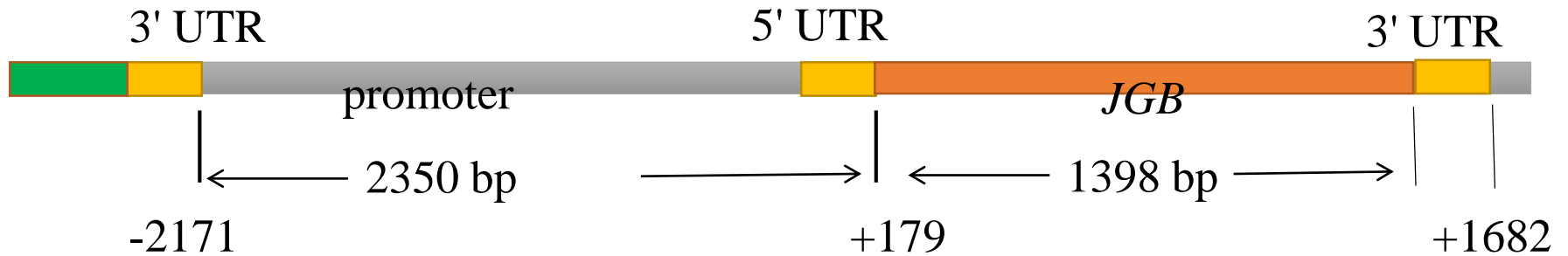
2.1 MADS gene evolution in land plants.

3. MADS Transcription factors

3.1 Overview of MADS family in *Arabidopsis thaliana*.

3.2 MIKC*-type genes in *A. thaliana*.

Putative promoter and putative TFs



PlantRegMap Plant Transcriptional Regulatory Map





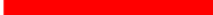






















Home PlantTFDB *cis*-Map Network Tools ATRM Download Help About

Binding Site Prediction Results

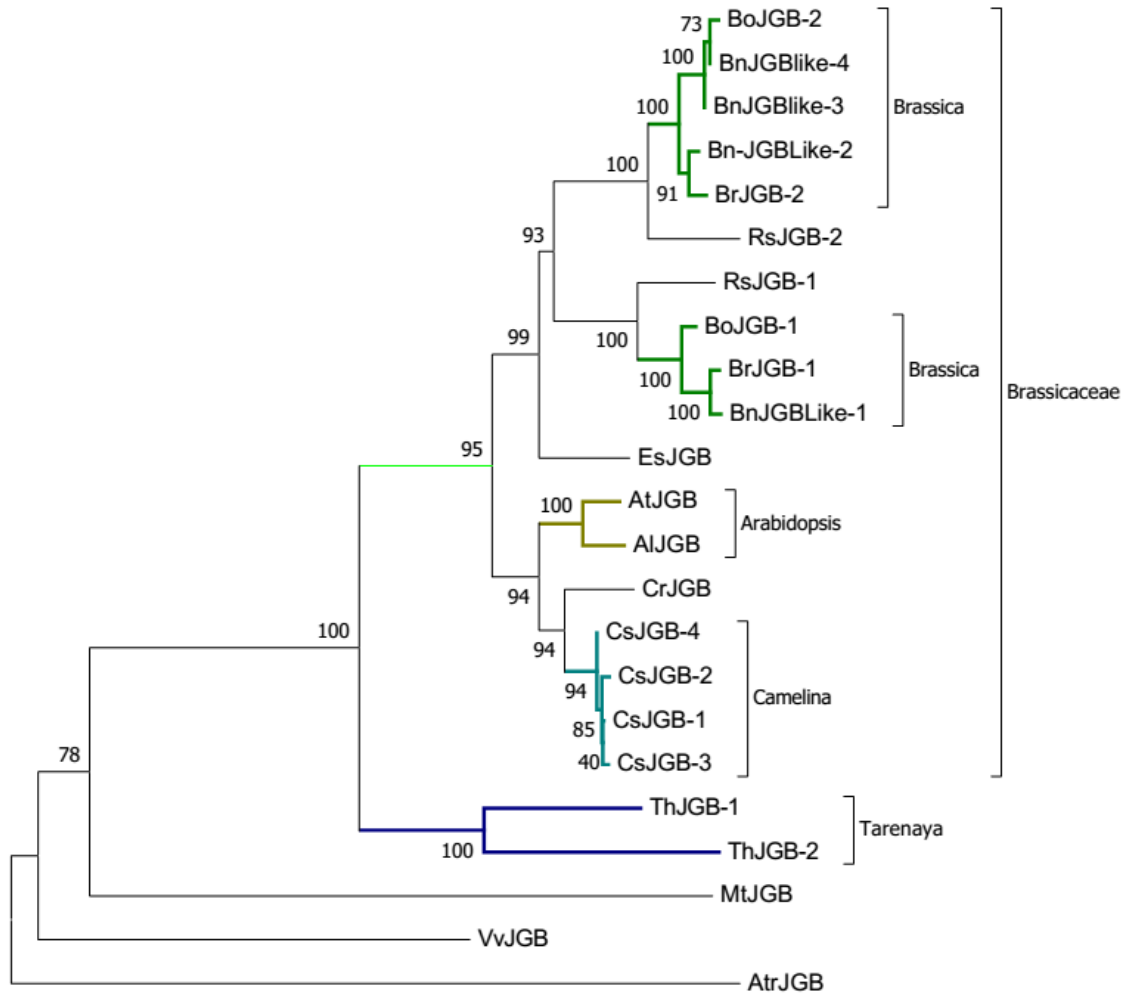
466 binding sites of 201 TFs are identified from 1 input sequences. You can download the results [here](#).

How to screen the possible TFs?

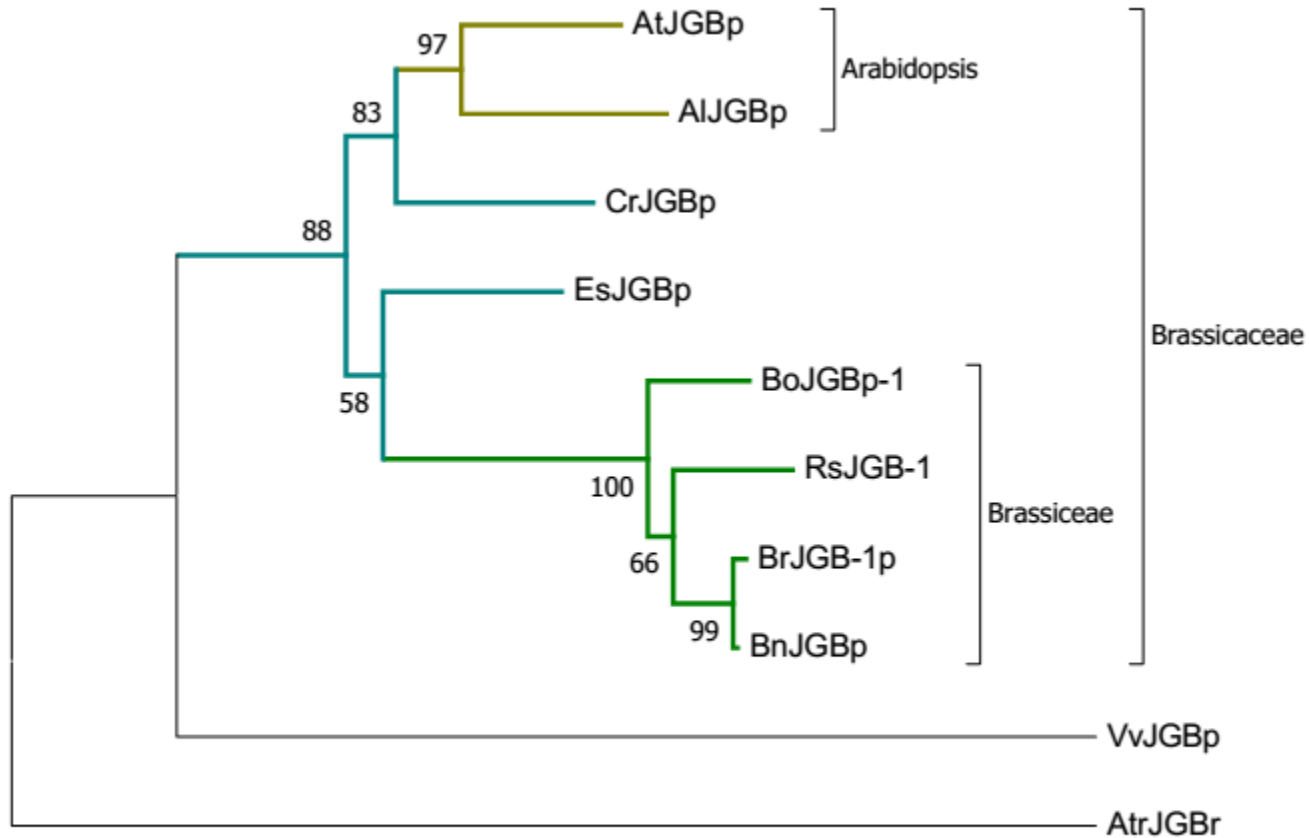
Conserved Gene may have Conserved Promoter

Protein names		Identity
 At2g26490 (Arabidopsis thaliana)		100.0%
 Uncharacterized protein (Arabidopsis thaliana)		99.8%
 Uncharacterized protein (Eutrema salsugineum)		97.4%
 Uncharacterized protein (Capsella rubella)		97.2%
 BnaC08g32900D protein (Brassica napus)		94.6%
 Uncharacterized protein (Arabis alpina)		94.8%
 BnaA09g40460D protein (Brassica napus)		94.2%
 Uncharacterized protein (Brassica rapa subsp. pekinen..)		94.4%
 BnaA04g15460D protein (Brassica napus)		91.1%
 Uncharacterized protein (Brassica rapa subsp. pekinen..)		91.1%
 BnaC04g38480D protein (Brassica napus)		91.1%
 F-box and wd40 domain protein, putative (Ricinus communis)		76.6%
 Uncharacterized protein (Jatropha curcas)		74.3%

Orthologous Or Paralogous of *JGB*

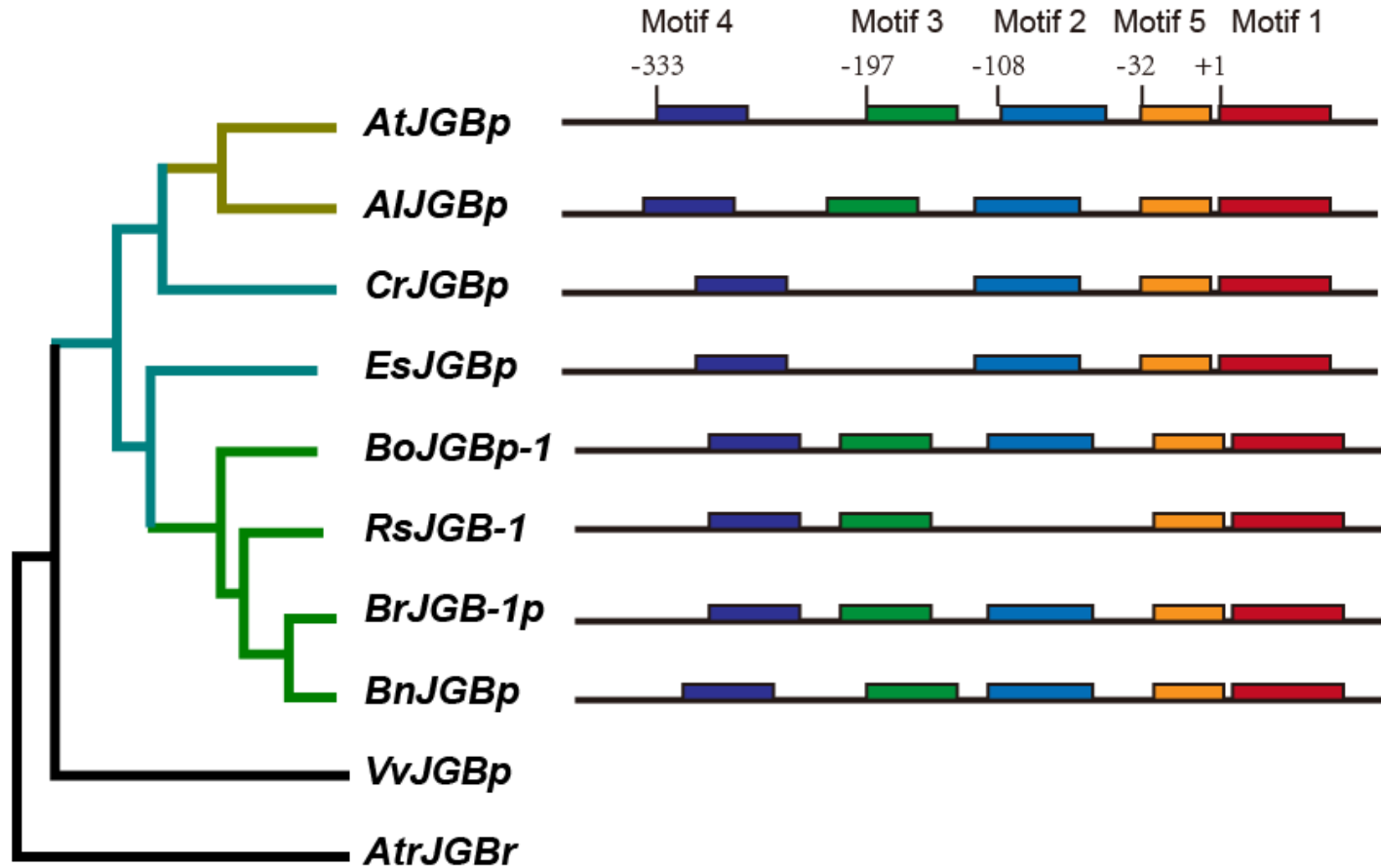


JGB Orthologous Promoters Reserved Evolution Feature as their Genes

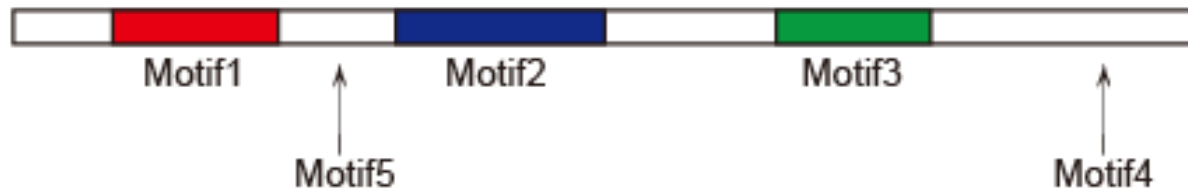
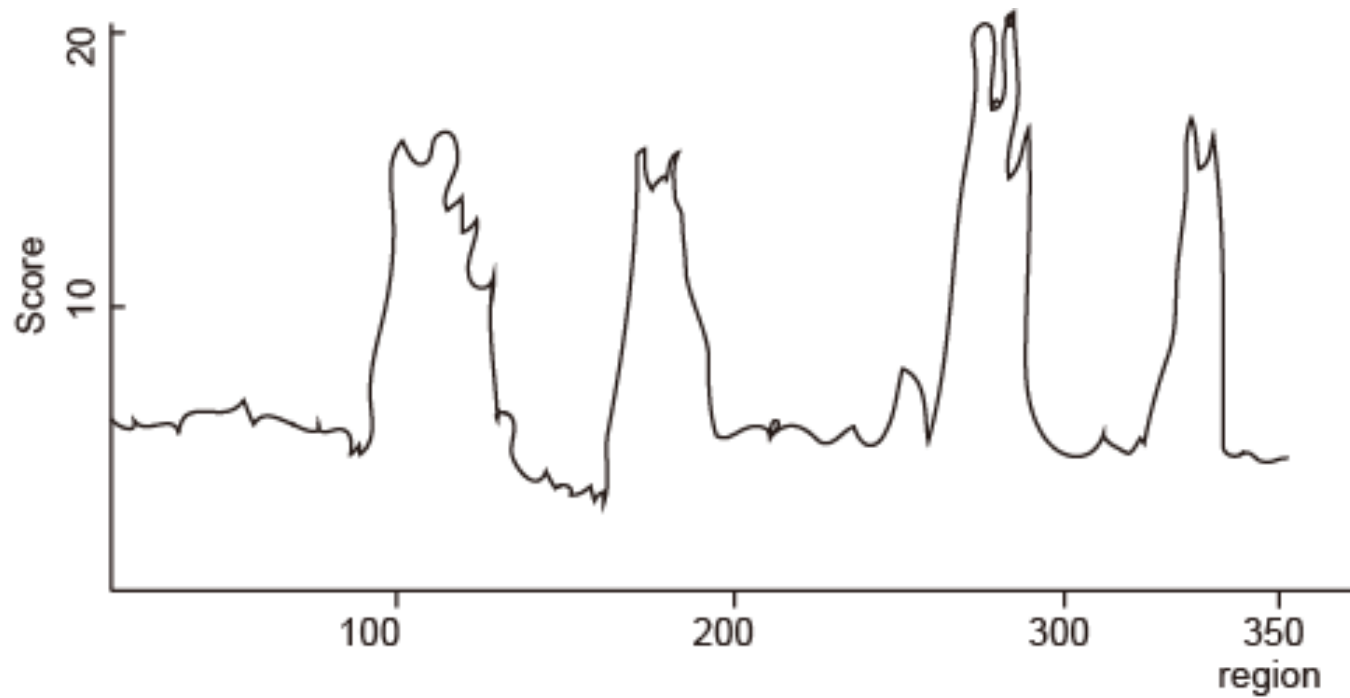


0.10

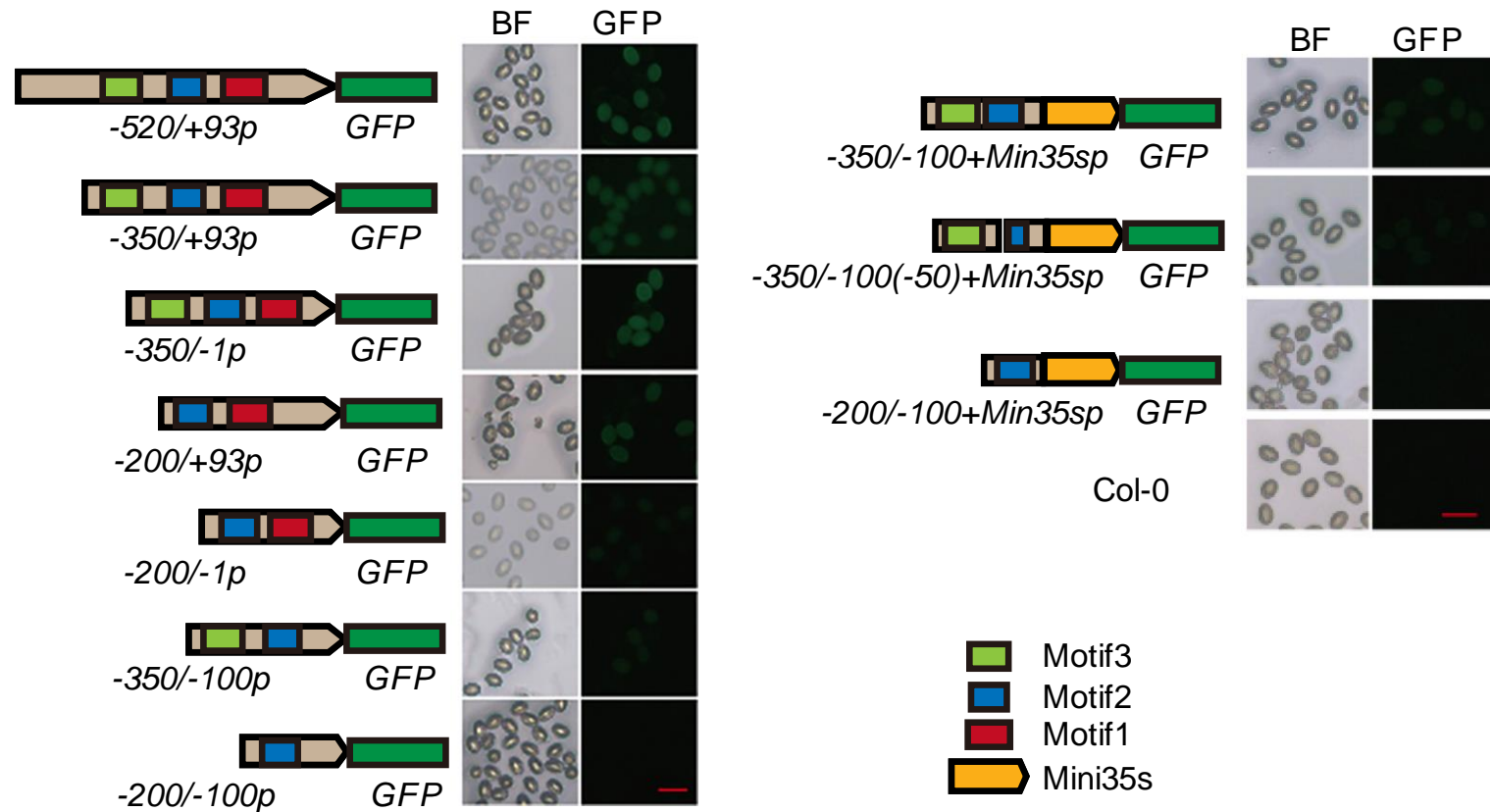
Conserved motifs can be found in *JGB* orthologous Promoter



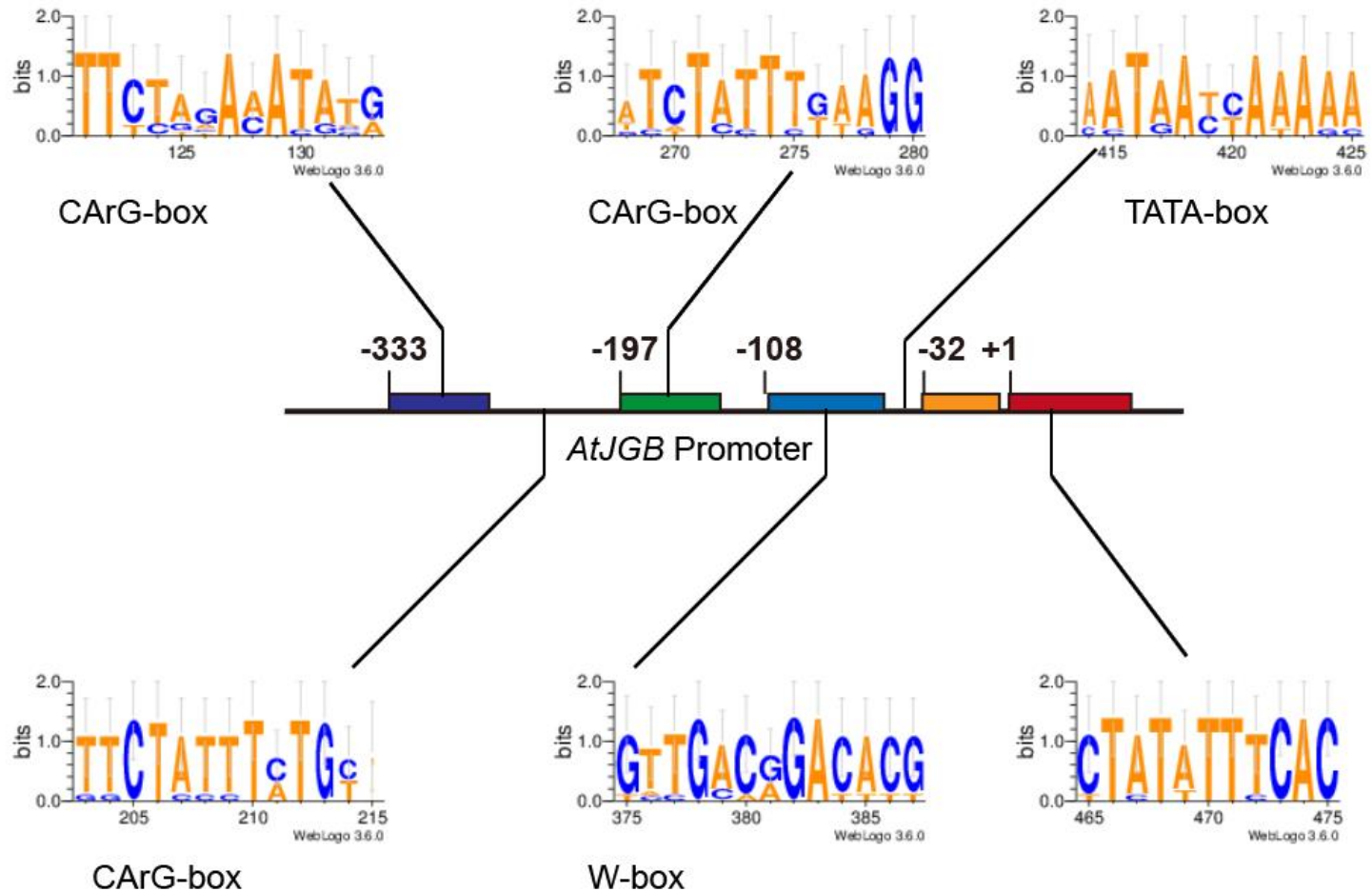
Conserved motifs have high score



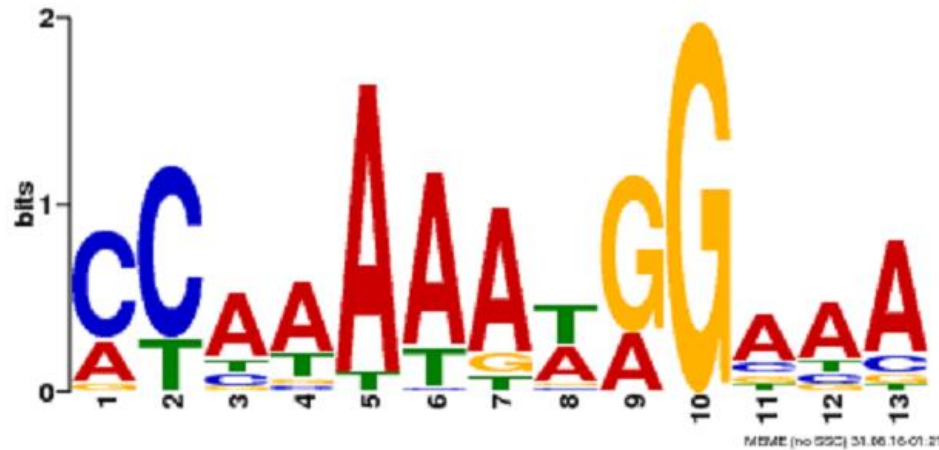
The Conserved Region Contained Core CAREs



Possible *cis*-acting regulation elements in *JGB* Promoter



CArG-Box—The Binding site of MADS TFs



AtSEP3 binding sites sequence Logo

Outline

1. Promoter analysis of JINGUBANG

1.1 How to find the promoter of a gene .

1.2 Prediction of Transcription factors.

2. Introduction to MADS family

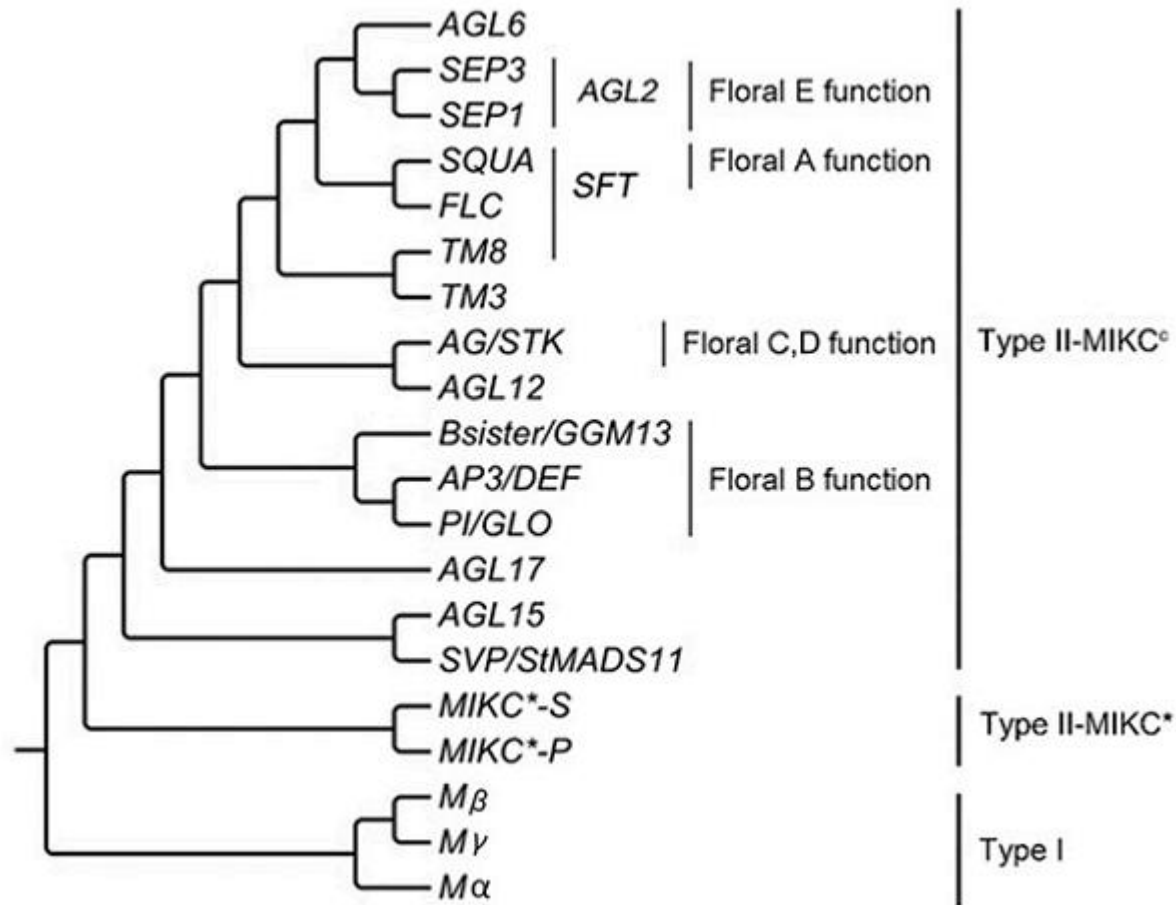
2.1 MADS gene evolution in land plants.

3. MADS Transcription factors

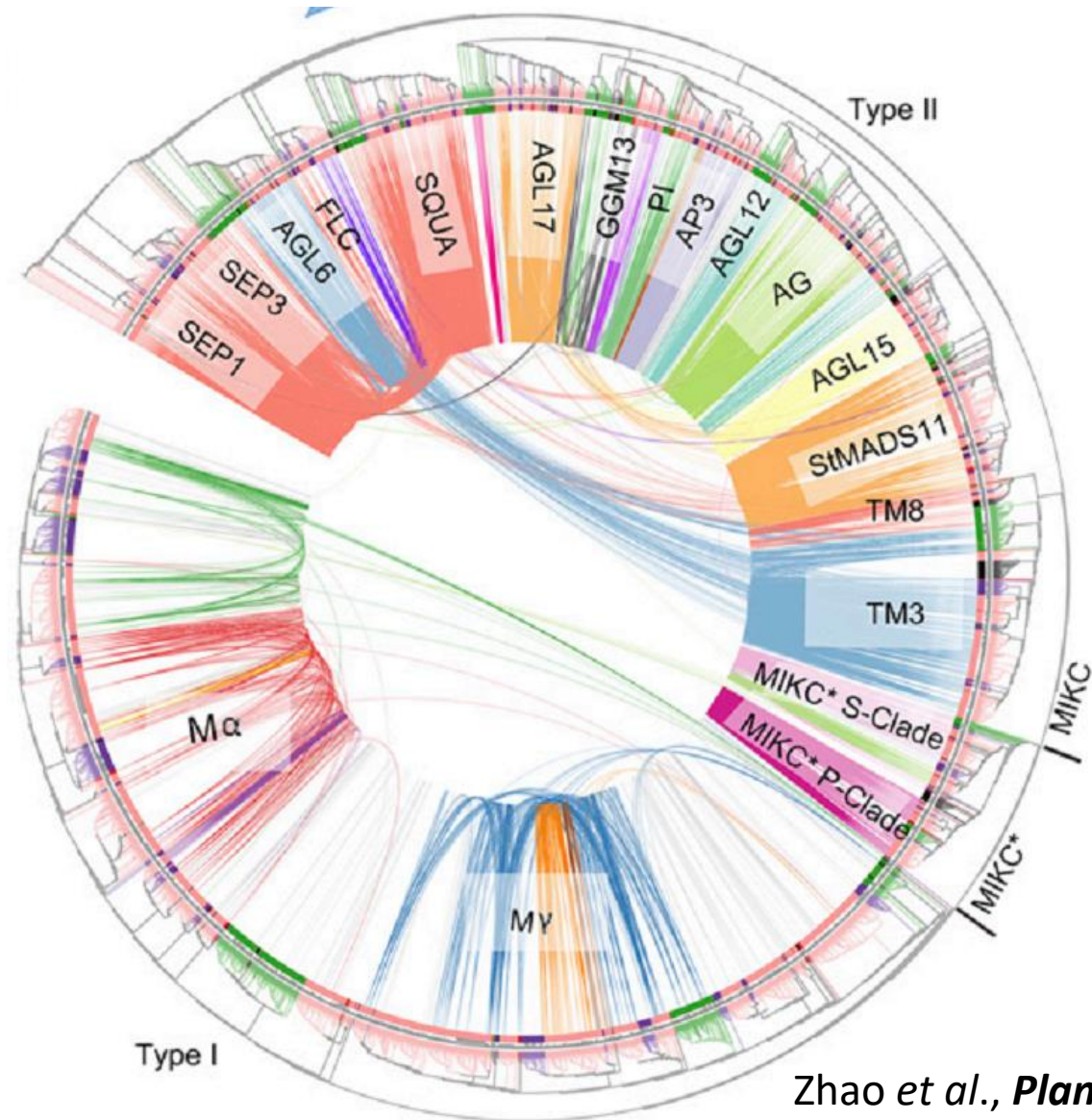
3.1 Overview of MADS family in *Arabidopsis thaliana*.

3.2 MIKC*-type genes in *A. thaliana*.

Phylogenetic tree of MADS Gene family



Phylogenetic tree with Syntenic connects



Outline

1. Promoter analysis of JINGUBANG

- 1.1 How to find the promoter of a gene .
- 1.2 Prediction of Transcription factors.

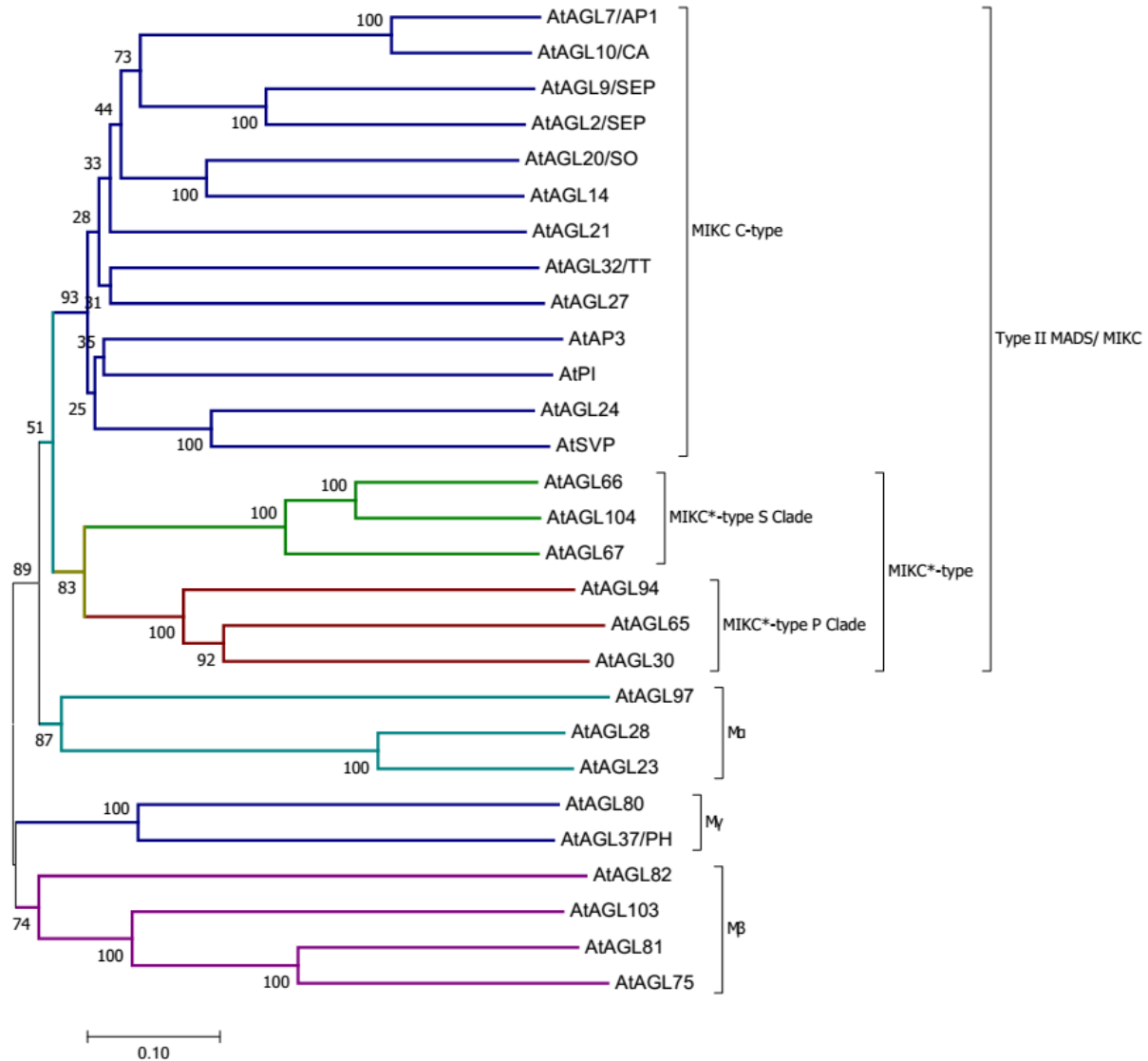
2. Introduction to MADS family

- 2.1 MADS gene evolution in land plants.

3. MADS Transcription factors

- 3.1 Overview of MADS family in *Arabidopsis thaliana*.
- 3.2 MIKC*-type genes in *A. thaliana*.

Phylogeny tree with Syntenic connects



Outline

1. Promoter analysis of JINGUBANG

- 1.1 How to find the promoter of a gene .
- 1.2 Prediction of Transcription factors.

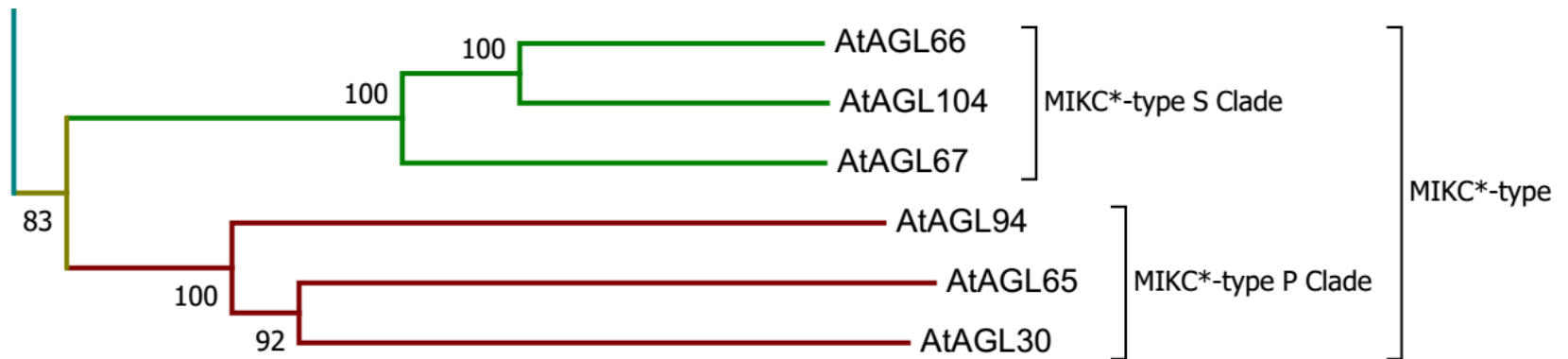
2. Introduction to MADS family

- 2.1 MADS gene evolution in land plants.

3. MADS Transcription factors

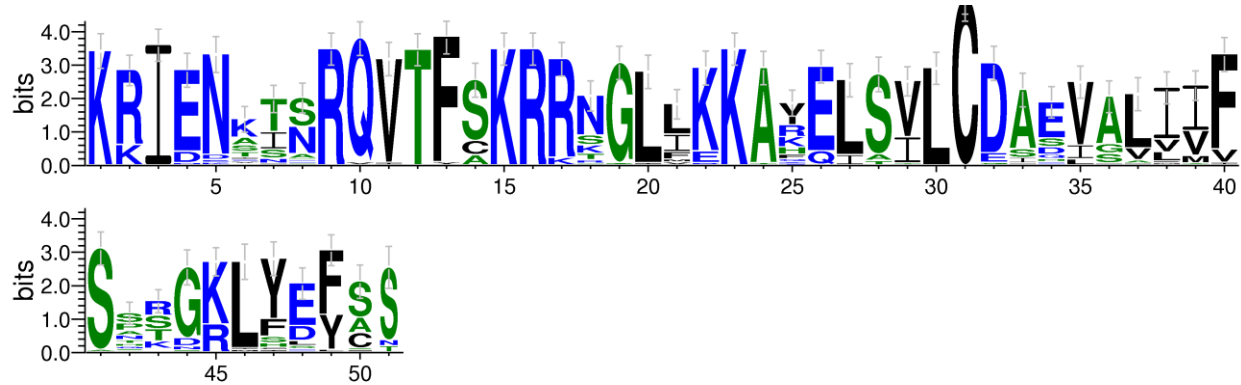
- 3.1 Overview of MADS family in *Arabidopsis thaliana*.
- 3.2 MIKC*-type genes in *A. thaliana*.

Phylogenetic tree of MIKC*-type MADS



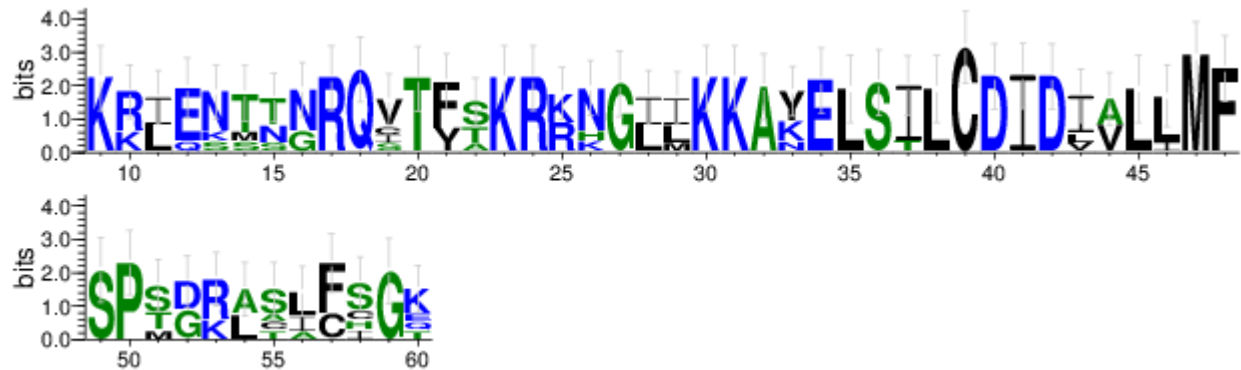
DNA Binding domain of MIKC*-type MADS

MIKC



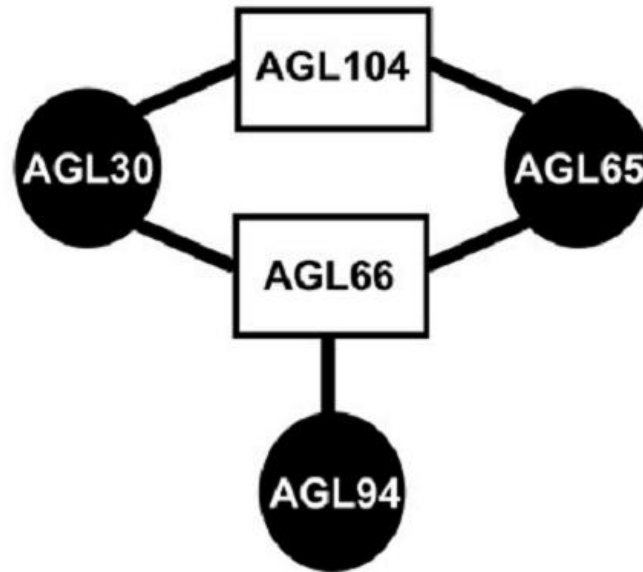
WebLogo 3.4

MIKC*
type

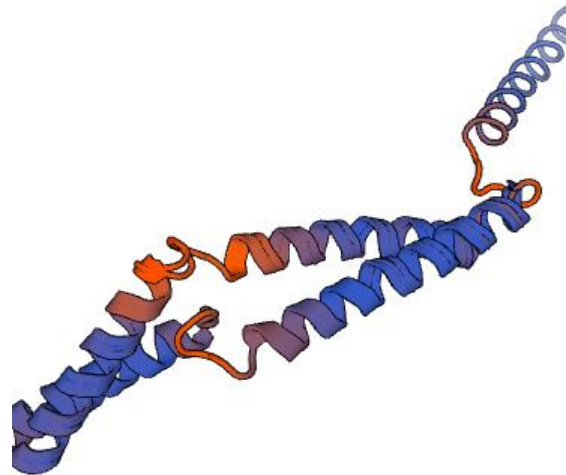


WebLogo 3.6.0

Interactions between MIKC*-S&P Clade

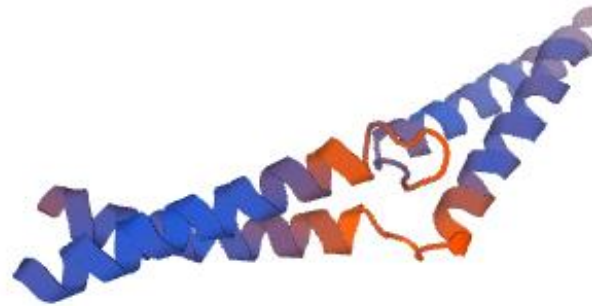


Interaction between *AtAGL30* and *AtAGL66*



Model-Template Alignment		
Model_02	MGRVKLKIKKLENTNGRQSTFAKRKNGILKKANELSILCDIDIVLLMFSPGKAAICCGTRSSMEEVIAKFSQVTPQERTKRRKFESLENL	90
4y66.1.A	-----	
Model_02	KKTFQKLDHVDVNIREFIASSNSTV EDLSTQARILQARISEIHGRLSYWTEPDKINNVEHLGQLEISIRQSLDQL RAHKEHFGQQQQAMQI	180
4y66.1.A	-----SSLLKEKALQVKLEEQRGTFRDLLK---NDPDVAQHLRNYTDIAKQEA-----	164
Model_02	ENANFVKDWSSTCSMQDGIQIPLQQLQSMWILNSNTINIVTEEHNSIQREVECSASSSFGSYPGYFGTGKSPEMTIPGQETSFLDELN	270
4y66.1.A	-----	
Model_02	TGQLKQDTSSQQQFTNNNNITAYNPNLHNDMNHHTLPPPPPLPLTLPHAQVYIPMNQREYHMNGFFEAPPPDSSAYNDNTNQTRFGSSSS	360
4y66.1.A	-----	
Model_02	SLPCSISMFDEYLFVQVTKTKLSQRF	386
4y66.1.A	-----	
Model_02	MGRVKLEIKRIENTTNRQVTFSKRRNGLIKKAYELSILCDIDIALLMFSPDRLSLFSGKTRIEDVFSRYINLSDQERENALVFPDQSRR	90
4y66.1.B	-----	
Model_02	PDFQSKEYLLRTLQQLKAENDIALQLTNPTAINSD VEELEHEVYKLOQQLLMAEEELRKYEPDPIRFTTMEEYETCEKQLMDTLT RVNQ	180
4y66.1.B	-----INDLKAQVEKLTQQGETLRITQRNLE----AAPITEVLKQEVDELQQQVS-----	133
Model_02	REHILSQDQLSSYEASALQQQSMGGPFNGDVGWLTENGPNEAHLFDASAHSAMYETLLQGSSSSSNQNNIMGESNVSNHNGDMFQEW	270
4y66.1.B	-----	
Model_02	AQAYNSTTAHNPSTLFPQMHHGLVVDPNIEEIEIPVMKKDAQADHEVSDYDIRMPQLSSQ	332
4y66.1.B	-----	

Interaction between *AtAGL30* and *AtAGL104*



```
Model_02 MGRVKLEIKRIENTTNRQVTFSKRRNGLIKKAYELSIILCDIDIALIMFSPDRLSLFSGKTRIEDVFSRFINLPKQERESALYFPDQNRDPDIQNKECLLRILQQ 105
4y66.1.B -----
Model_02 LKTENDIALQVTNPAAINSDEEELEHEVCRLQQQLQMAEEELRRYEPDPIRFTTMEEYEVSEKQLLDLTLTHVQRRDHLMSNHLSSYEASTMQPNIGGPFVNDV 210
4y66.1.B -----TEVLKQEVDELRRQVSANDEKLRLLVRE-SNAIVSDADMLTLQKNYHDAMTAWATRRAKC----- 175
Model_02 EGWLPENGTNQTHLFDASAHSNQLRELSSAMYPELLQSSSSSNQNMSECHVTNHNMGEMFPEWAQAYSSSALFASMQQQHEGVGPSIEEMMPAQSSDIPGVTA 315
4y66.1.B -----
Model_02 TQVDHEVSDYETKVPQLSSQ 335
4y66.1.B -----
Model_02 MGRVKLKIKKLENTNGRQSTFAKRRNGILKKANELSIILCDIDIVLLMFSPGKAAICCGTRSSMEEVIAKFSQVTPQERTKRRKFESLENLKKTFQKLDHVDNIRE 105
4y66.1.A -----
Model_02 FIASSNSTVEDLSTQARILQARISEIHGRLSYWTEPKINNVEHLGQLEISIRQSLDQLRAHKEHFGQQQQAMQIENANFVKDWSTCSMQDGIQIPLEQLQSMS 210
4y66.1.A -----SSLLEKALQVKLEEQRGTFRDLLK----NDPDVAQKLRNYTDIAKQEA----- 164
Model_02 WILNSNTINIVTEEHNIPQREVECSASSSFGSPYGFYGTGKSPEMTIPGQETSFLDELNTGQLKQDTSSQQQFTNNNNITAYNPNLHNDMNHHTLPPPPPLT 315
4y66.1.A -----
Model_02 LPHAQVYIPMNQREYHMNGFFEAPPPDSSAYNDNTNQTRFGSSSSSLPCSISMFDEYLFQSVIKTKLSQRF 386
4y66.1.A -----
```


The Analysis about Domain and Structure of MIKC* Family

G10 Group

傅偲 黄勋 俞琴琴 张二禾



Sequence search results

[Show](#) the detailed description of this results page.

We found **1** Pfam-A match to your search sequence (**all** significant)



[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

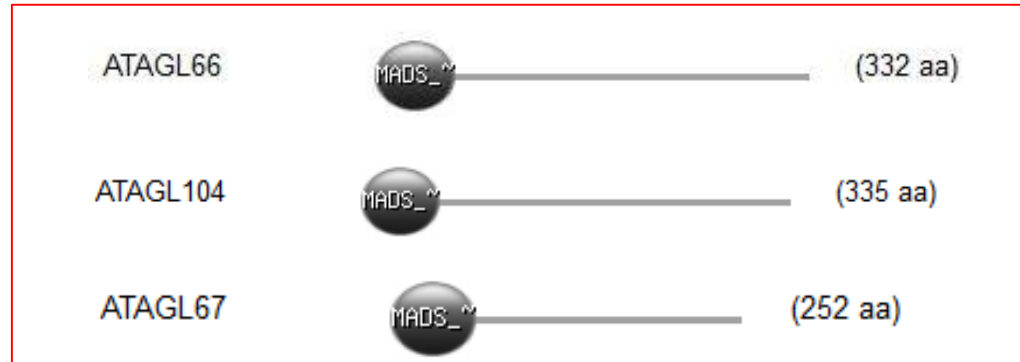
[Show](#) or [hide](#) all alignments.

Family	Description	Entry type	Clan	Envelope		Alignment		HMM		HMM length	Bit score	E-value	Predicted active sites	Show/hide alignment
				Start	End	Start	End	From	To					
SRF-TF	SRF-type transcription factor (DNA-binding and dimerisation domain) (shorten)	Domain	n/a	10	57	10	55	1	46	48	83.7	4.2e-24	n/a	Hide
#HMM	rIeneFaRqvvtFsKRrnGLfKKAsELsvLcgaevavvfvspngkly													
#MATCH	rIen ++RqvtFsKRrnGL+KKA+ELs+Lc+++a+++fsp+++1													
#PP	7*****9985													
#SEQ	rIENTNRRQVTFskRRNGLIKKAYELsILCDIDIALLMFSPsDRLS													

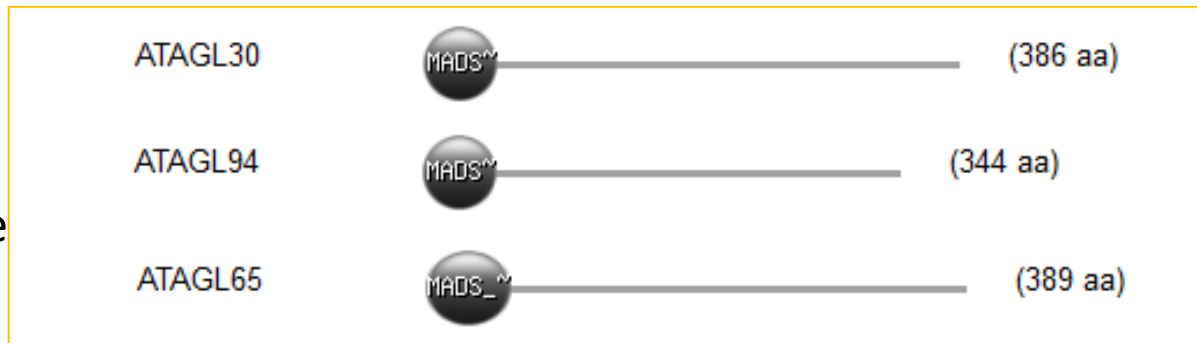
**SRF-type transcription factor
(DNA-binding and dimerisation domain)**



MIKC*s clade



MIKC*p clade



Predicted feature:

DOMAIN	1	61	MADS-box
DOMAIN	1	53	MADS-box



MEME

Multiple Em for Motif Elicitation

DISCOVERED MOTIFS

	Logo	E-value ?	Sites ?	Width ?	More ?	Submit/Download ?
1.		1.0e-154	6	59	I	---
2.		8.3e-043	6	40	I	---
3.		7.5e-035	3	60	I	---

Stopped because requested number of motifs (3) found.

MOTIF LOCATIONS

Only Motif Sites [?](#) Motif Sites+Scanned Sites [?](#) All Sequences [?](#)

Name ?	p-value ?	Motif Location ?		
1. AtAGL66	4.45e-162			
2. AtAGL104	1.13e-160			
3. AtAGL67	1.16e-140			
4. AtAGL65	1.51e-91			
5. AtAGL30	7.33e-88			
6. AtAGL94	6.74e-84			

MIKC*S

MIKC*P

Clustal Omega

"The last alignment program you'll ever need"

Multi-alignment result:

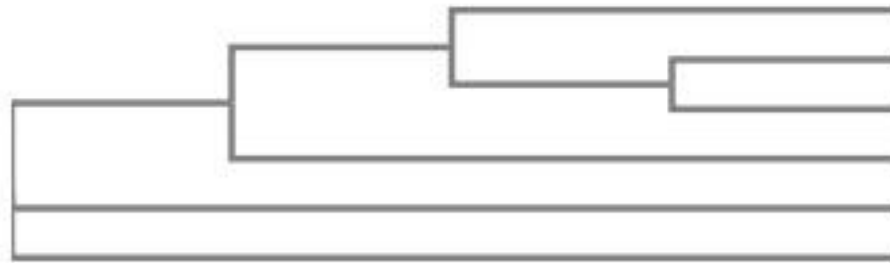
```
AtAGL67      RIEKSTNRQITFSKRKKGLIKKAYELSTLCDIDLALLMFSPDRLC--
AtAGL66      RIENTTNRQVTFSKRRNGLIKKAYELSILCDIDIALLMFSPDRLS--
AtAGL104     RIENTTNRQVTFSKRRNGLIKKAYELSILCDIDIALIMFSPDRLS--
AtAGL94      --QNMNGRQCTYTKRRHGIMKKAKELSILCDIDVLLMFSPMGK----
AtAGL65      -LESTSNRQVTYTKRKNGLKAKELSILCDIDIVLLMFSPTGRATAF
AtAGL30      -LENTNGRQSTFAKRKNGLKANELSILCDIDIVLLMFSPTGK----
              :. . .* * : * : * : * : * * * * * : . * : * * * * . :
```

Identity result :

Percent Identity Matrix - created by Clustal2.1

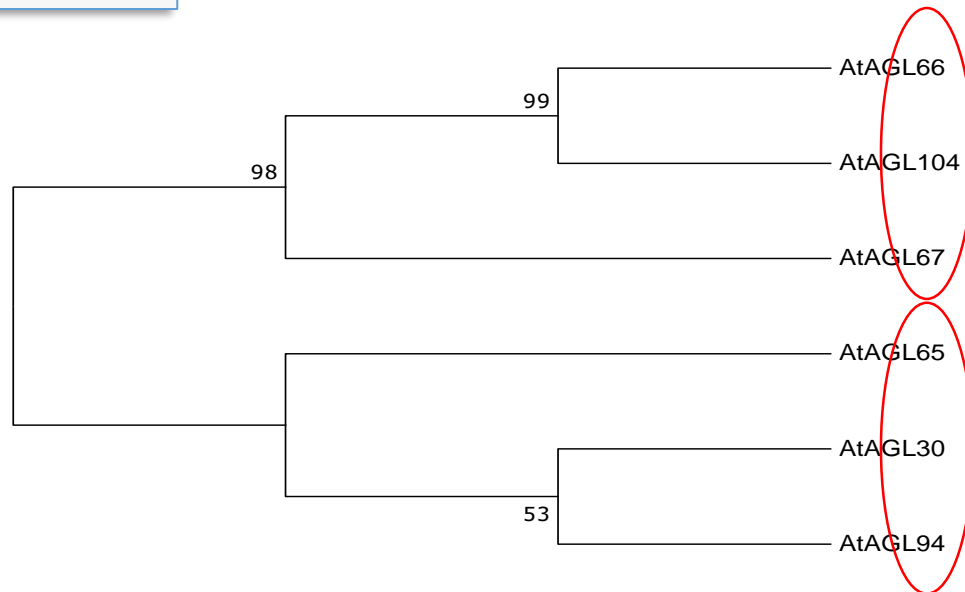
1: AtAGL67	100.00	82.61	80.43	54.76	60.00	60.47
2: AtAGL66	82.61	100.00	97.83	61.90	68.89	69.77
3: AtAGL104	80.43	97.83	100.00	59.52	66.67	67.44
4: AtAGL94	54.76	61.90	59.52	100.00	71.43	73.81
5: AtAGL65	60.00	68.89	66.67	71.43	100.00	81.40
6: AtAGL30	60.47	69.77	67.44	73.81	81.40	100.00

Clustal Omega



AtAGL67 0.12884
AtAGL66 -0.00051
AtAGL104 0.02225
AtAGL94 0.16927
AtAGL65 0.10059
AtAGL30 0.08546


MEGA7.1 Beta




Structure prediction of MIKC* family


Phyre²
Protein Homology/analogY Recognition Engine V 2.0


Subscribe to Phyre at Google Groups
Email:
[Visit Phyre at Google Groups](#)
[Follow @Phyre2server](#)




Expert Mode


One-to-one threading
Submit both a sequence and a structure. Phyre will attempt to align the sequence and structure and construct a model of your sequence


Batch processing
Submit a file containing up to 100 sequences for automated modelling by Phyre (to increase limit contact [Lawrence Kelley](#))


BackPhyre
Use Phyre in reverse. Submit a PDB structure and search that structure against a wide range of genomes


GenSearch
Search your sequence against a wide range of genomes

#	Description	JobId	Status	Hit Summary
1	sp Q1PFA4 AGL30_ARATH_Agamous-like_MADS-box_protein_AGL30_OS...	78d395697c55811f	Finished	
2	sp Q7X9I0 AGL65_ARATH_Agamous-like_MADS-box_protein_AGL65_OS...	ae2f6e9bc911608e	Finished	
3	tr F4I8L6 F4I8L6_ARATH_AGAMOUS-like_67_OS=Arabidopsis_thalia...	cbb2937d7fbbedb6	Finished	
4	tr Q766C0 Q766C0_ARATH_AGAMOUS-like_94_OS=Arabidopsis_thalia...	7c79ed5c9d89cc49	Finished	
5	sp Q9LM46 AG104_ARATH_Agamous-like_MADS-box_protein_AGL104_O...	dc6cc78947d7dc35	Finished	

Best
AGL67

Model (left) based on template [d1n6ja_](#)

Top template information

Fold:SRF-like
Superfamily:SRF-like
Family:SRF-like

Confidence and coverage

Confidence: **100.0%** Coverage: **36%**

90 residues (36% of your sequence) have been modelled with 100.0% confidence by the single highest scoring template.
Additional confident templates have been detected (see [Domain analysis](#)) which cover other regions of your sequence.
 173 residues (69%) could be modelled at >90% confidence using multiple-templates.
 You may wish to try resubmitting your sequence in "intensive" mode to model more of your sequence.

3D viewing

[Interactive 3D view in JSmol](#)

For other options to view your downloaded structure offline see the [FAQ](#)

Worst
AGL65

Model (left) based on template [d1n6ja_](#)


Top template information

Fold:SRF-like
Superfamily:SRF-like
Family:SRF-like

Confidence and coverage

Confidence: **100.0%** Coverage: **23%**

90 residues (23% of your sequence) have been modelled with 100.0% confidence by the single highest scoring template.



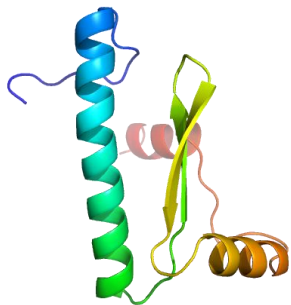
You may wish to submit your sequence to [Phyrealarm](#). This will automatically scan your sequence every week for new potential templates as they appear in the Phyre2 library.
Warning: 56% of your sequence is predicted disordered. Disordered regions cannot be meaningfully predicted.

3D viewing

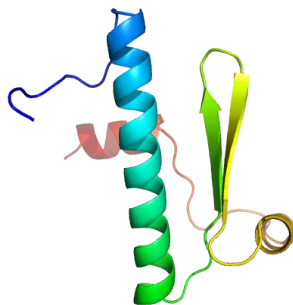
[Interactive 3D view in JSmol](#)

For other options to view your downloaded structure offline see the [FAQ](#)

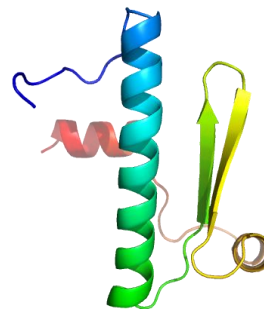
The rank1 hits of 6 members are all **1n6j**



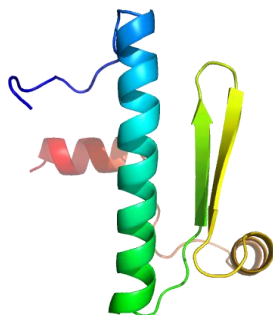
AGL66



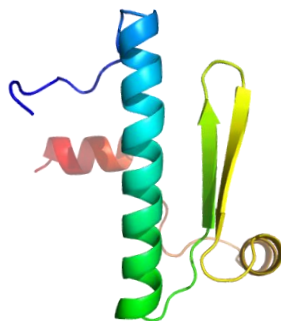
AGL104



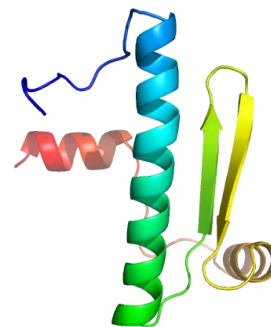
AGL67



AGL65



AGL30



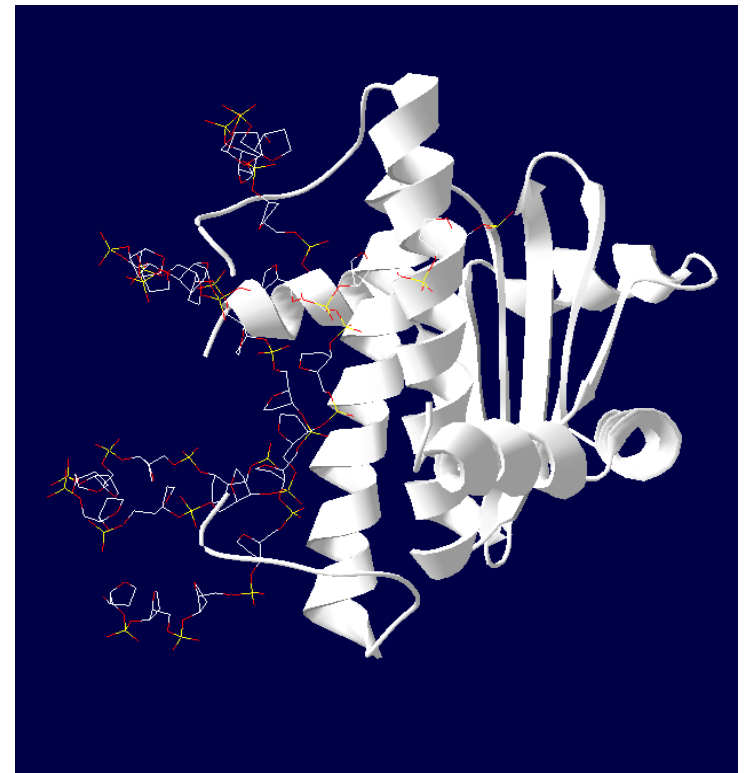
AGL104

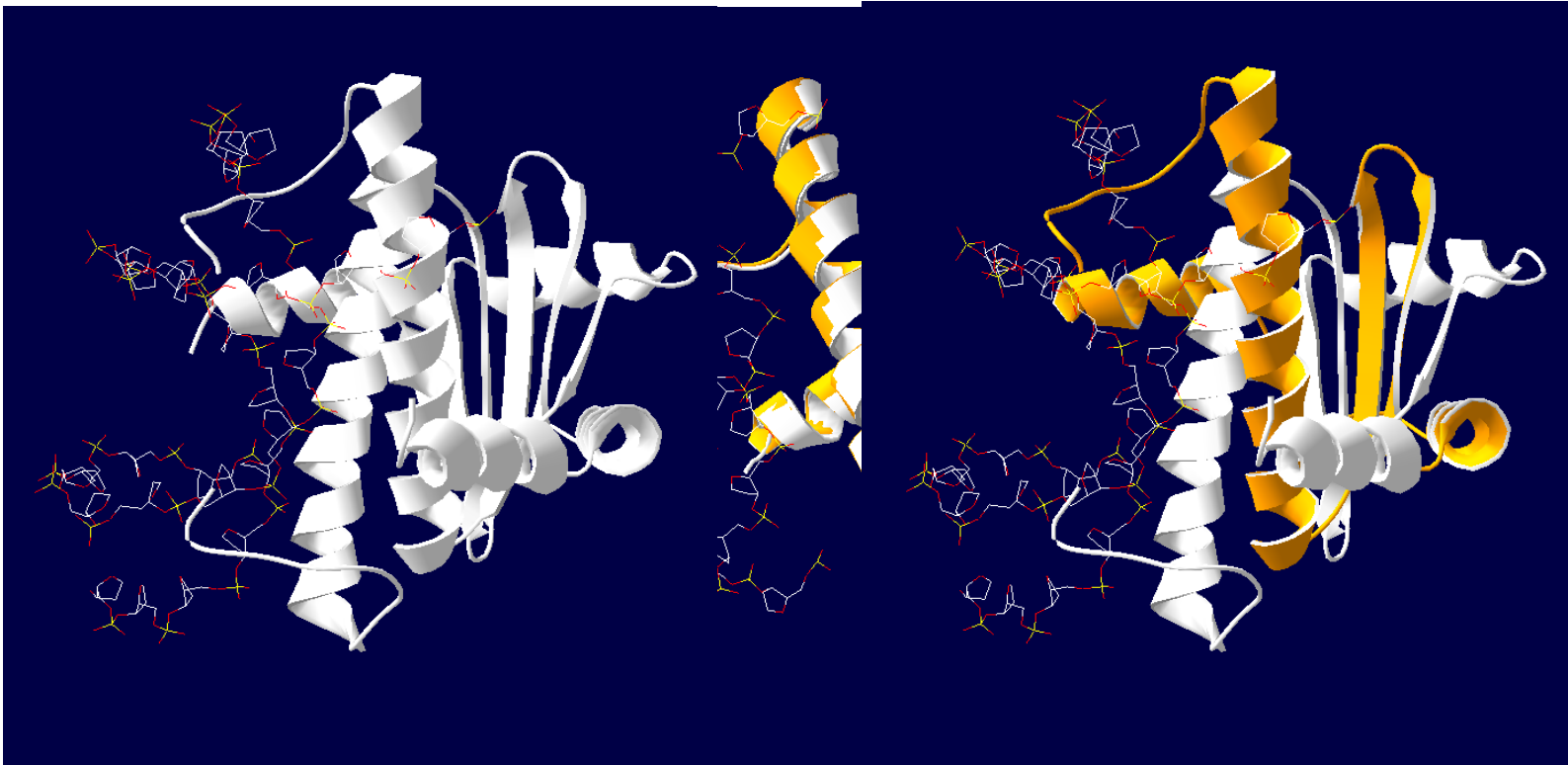
D Domain

MADS SRF-like (IPR033897)

Short name: MADS_SRF-like

In SRF, the MADS box has been shown to be involved in DNA-binding and dimerization. Proteins belonging to the MADS family function as dimers, the primary DNA-binding element of which is an anti-parallel coiled coil of two amphipathic alpha-helices, one from each subunit. The DNA wraps around the coiled coil allowing the basic N-termini of the helices to fit into the DNA major groove. The chain extending from the helix N-termini reaches over the DNA backbone and penetrates into the minor groove. A 4-stranded, anti-parallel beta-sheet packs against the coiled-coil face opposite the DNA and is the central element of the dimerization interface.





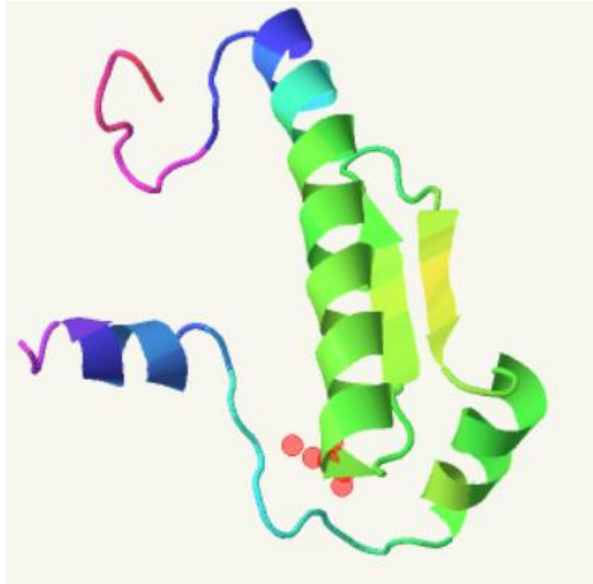
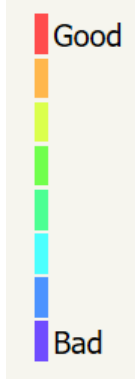
1n6j-chainA&B AGL66 VS 1n6j-chainB AGL66 VS 1n6j-chainA

■ AGL66
■ 1n6j

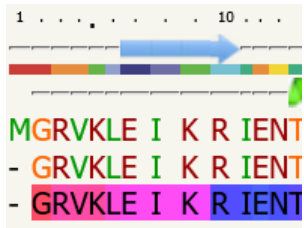
1C7U

Complex of the DNA binding core domain of the transcription factor MEF2A with a 20mer oligonucleotide

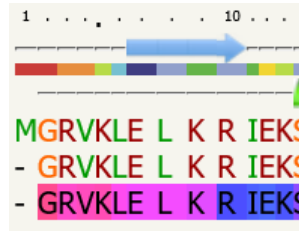




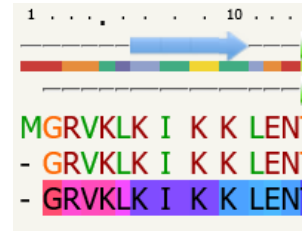
Predicted Secondary structure
 Model Secondary structure
 Query Sequence
 Modelled Residues
 ProQ2 quality assessment



AGL66



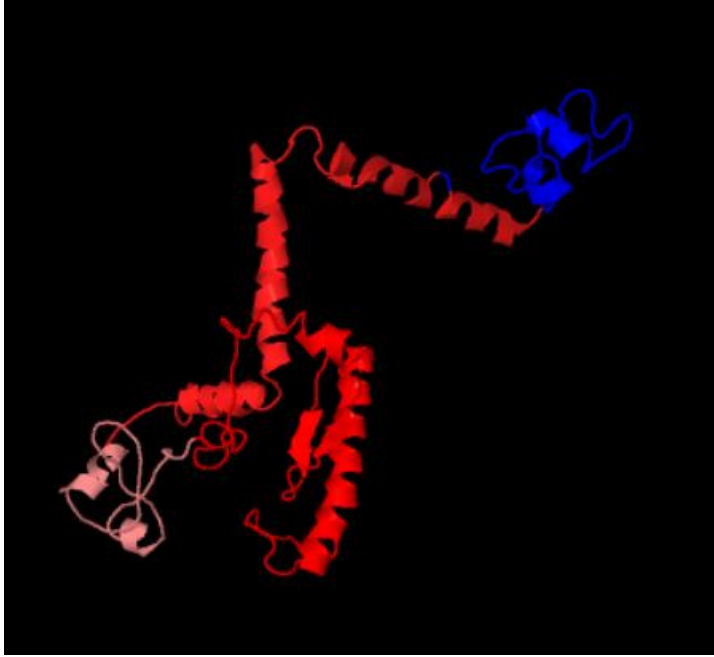
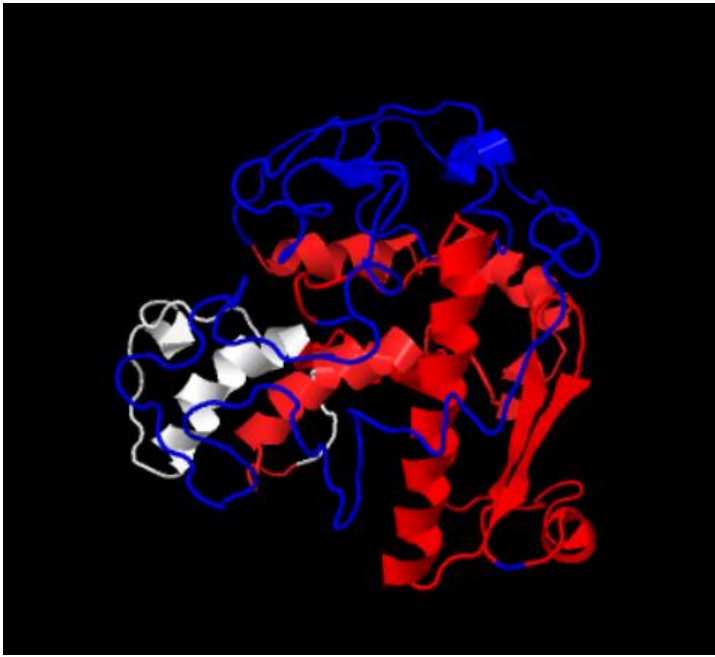
AGL67



AGL30

149 residues (45%) could be modelled at >90% confidence using multiple-templates.

173 residues (69%) could be modelled at >90% confidence using multiple-templates.



Confidence Key
High(9) | | | | | | | | | Low (0)

AGL66-Intensive

AGL67-Intensive

Reference

- 1) **Verelst, W., et al.** (2007). MIKC* MADS-Protein complexes bind motifs enriched in the proximal region of late Pollen specific Arabidopsis promoters. *Plant Physiol.*
- 2) **Ju, Y. et al.** (2016). Arabidopsis JINGUBANG is a negative regulator of pollen germination that prevents pollination in moist environments. *Plant cell.*
- 3) **Zhao, T. et al.** (2017). Phylogenomic Synteny Network Analysis of MADS-Box Transcription Factor Genes Reveals Lineage-Specific Transpositions, Ancient Tandem Duplications, and Deep Positional Conservation. *Plant cell.*

Thank you !