Evolutionary History and Molecular Mechanism of Differentiation of Cold Response in Chinese Arabidopsis thaliana Populations along the Yangtze River

> 2019/06/16 Yang Jixuan, Song Zihan Qu and Gu Lab School of Life Sciences, Peking University

Evolutionary history and cold response of Chinese Arabidopsis thaliana populations

Yang Jixuan 2019/06/16

What is Arabidopsis thaliana?



Capsella rubella Arabidopsis thaliana A. lyrata

Annual, dicot Cruciferae, *Arabidopsis*

- Small plant
- Short life cycle
- High self-fertilization
- Small genome

Gross morphological variation in *Arabidopsis* and relatives.

What is evolution?

- **Biological** (or **organic**) **evolution** is change in the properties of groups of organisms over the course of generations.
- The development, or **ontogeny**, of an individual organism is not considered evolution: individual organisms do not evolve.
- Groups of organisms, which we may call **populations**, do evolve: they undergo *descent with modification*.

What is adaptation?

- An adaptation is a characteristic that enhances the survival or reproduction of organisms that bear it relative to alternative character states (especially the ancestral condition in the population in which the adaptation evolved).
- **Natural selection** is the only mechanism known to cause the evolution of adaptations.

Phylogenetic tree



The time scale in most phylogenetic trees is a relative one, but the tree always implies the passage of time from the root of the tree toward the tips of the branches.

Apomorphy and Plesiomorphy

APOMORPHY VERSUS PLESIOMORPHY

Apomorphy refers to a novel evolutionary character unique to a particular clade and all its descendants

An apomorphic character is similar throughout the clade

The two types are autapomorphy and synapomorphy Plesiomorphy refers to an evolutionary character homologous within a particular clade but, not unique to all members of that particular clade

A plesiomorphic character is not similar throughout the clade

Not divided further

A derived or specialized character

An apomorphic character can be used as a defining character for the clade

Ex: absence of legs in snakes

A primitive or ancestral character

A plesiomorphic character cannot be used as a defining character for the clade

Ex: presence of legs in reptiles

Visit www.pediaa.com

https://pediaa.com/difference-between-apomorphy-and-plesiomorphy/

Divergent evolution

• Divergent evolution is the process whereby groups from the same common ancestor evolve and accumulate differences, resulting in the formation of new species.



The homologous bones found in the forelimb of four different types of mammal.

Darwin's finches

Convergent Evolution

• Convergent evolution is the process in which organisms that are not closely related independently evolve similar features





The skin stretched between the limbs of a Sugar Glider and a Flying Squirrel, evolved to allow gliding motion.

The different internal bone structures of wings in: 1) Reptilian pterosaurs (*Pterosauria*). 2)Mammalian bats (*Chiroptera*). 3) Birds (*Aves*).

The Great Wall of China: a physical barrier to gene flow?

• As an artificial physical barrier, the Great Wall could be an excellent model for studying its effect on the genetic differentiation of plant populations around it.





Control group, mountain path

Juyong-guan

Su et al., Heredity, 2003

The Great Wall of China: a physical barrier to gene flow?



- Prunus armeniaca
- Ziziphus jujuba
- Vitex negundo
- Heteropappus hispidus
- Cleistogenes caespitosa
- Ulmus pumila
- Ulmus macrocarpa



The Juyong-guan Great Wall has served as a physical barrier to gene flow of separated plant subpopulations.

Geographical distribution of Arabidopsis thaliana





Geographical distribution of *Arabidopsis thaliana* around the world

Arabidopsis habitats

Koornneef *et al.*, Annu. Rev. Plant Biol. ,2004 Mitchell-Olds & Schmitt., Nature, 2006

How to study the origin and evolution of Chinese Arabidopsis thaliana populations?

• Chloroplast DNA sequence variability



Distribution map of the 19 accessions of *Arabidopsis thaliana* from China and one from India (Kas-2).

Primer sequence
5'-GGATAGGACATAGGTCTTCTAA-3' 5'-CACAGTGGAAGAACAGATAATG-3'
5'-CCCTTCAAATTGTATCTGATTAAA-3' 5'-GATTTGAACTGGGGAAAAAGGATT-3'
5'-GCGGATATAGTCGAATGGTAAA-3' 5'-GTAAGATTCCGTCGCTAAGTGA-3'
5'-CGCATAGCTTCATAATAATTCTGT-3' 5'-TCTACATAACAGTTCCAATGTTAC-3'
5'-TCCTCTGCTCTACCAACTGA-3' 5'-AAATCGTGAGGGTTCAAGTC-3'
5'-CTAGCTGCTGCTTGTGAAGTATGG-3' 5'-TAAAATTGAACCACGATTTTTCCA-3'
5'-CAATTGCCGGAAAGACTAGG-3' 5'-GTTCACAAGCGGCTGAATCT-3'
5'-TATCGAATACTGGTAATAATATCA-3' 5'-ATAGTTAAAGCTGCTAGTAGAAAA-3'
5'-CTAGGGGTGTTATGCCGATT-3' 5'-GTTCGGTTCGTTAGCAGGTT-3'
5'-CATGGAACGGGATGTTTTA-3' 5'-GTTCTACGCCTCCGAGCTAT-3'
5'-TCCTCGATGTTGTTTACGAAATCT-3' 5'-TCGAACTATTTATGGGGTTTTAGG-3'

Information about eleven pairs of primers for PCR amplification.

Evolutionary history of Chinese Arabidopsis thaliana populations



- The origin types of Chinese wild *Arabidopsis thaliana* populations can be divided into two categories.
- MDHC-I: Yangtze River Basin Populations
- MDHC-II: Xinjiang Populations
- Rapid population expansion
 along the Yangtze River was
 estimated to be about 90,000
 years ago.

Different ecotypes of Arabidopsis thaliana



Col-0, USA



Ler-0, Germany



Cvi-0, Cape Verde



Per-1, Russia



Bla-1, Spain



Vajug-1, Serbia



Altai-5, China



Bor-1, Czech



Haz-2, USA



Ale-Stenar-56-14, IP-Mos-1, Portugal Sweden







How to study *Arabidopsis* natural variation?



Trontin *et al.*, Current Opinion in Plant Biology, 2011

Quantitative Trait Loci (QTL)-Mapping





SNP-index: Short reads generated by wholegenome sequencing are aligned to the reference sequence. If 10 short reads cover a given nucleotide position, the coverage of the site is 10. Among the 10 reads, if four contain a SNP different from the reference nucleotide, the SNP-index is defined as 0.4. On the other hand, if all the reads harbor a SNP different from the reference, the SNP-index is 1.0.

CRISPR-Cas9



Step IV: Delivery to plants via different delivery system like Geminivirus mediated delivery, agrobacterium transformation and protoplast transformation.



Step VI: Regeneration and screening of mutant plants for miRNA/IncRNA targeting events.

Series of events to generate a clustered regulatory interspaced short palindromic repeats/CRISPRassociated (CRISPR-Cas) mutagenised plant

> Basak & Nithin, Front Plant Sci., 2015 18 Sternberg & Doudna, Mol Cell., 2015

1001 Genomes

Arabidopsis 1001 genomes

1001 Genomes

A Catalog of Arabidopsis thaliana Genetic Variation.

Tools

Explore the variants. We maintain several tools for data download, visualization, and analysis.

Go

Download

Visit the Data Center and download whole sets of SNPs, indels, SVs, and genome sequences.

Go

Get Seeds

Seed sets of natural accessions are available for

Complete set 80 strains (D. Weigel lab, MPI) 195 strains (J. Ecker lab, Salk) 180 strains (M. Nordborg Lab, GMI)

1135 Accessions Final Set

180 GMI Accessions (GMINordborg2010) 80 MPI Accessions (MPICao2010) 195 Salk Accessions

Legacy Projects

Cold response in plants



Physiological manifestation of cold stress by LTsensitive, -insensitive and -tolerant plants



The involvement of plant hormone signaling in plant responses to cold stress Sharma *et al.*, BioEssays, 2005 Javier & Julio, Survival Strategies in Extreme Cold and Desiccation, 2018

What are the CBFs?

Discovery of C-repeat (CRT) sequences

DREB1A

$\begin{array}{cccccccc} TGGCTTCGATTAGGAGTCTTCGGTTTCCTCAGGCGGTGATTATTTCCGAGGCTTGCGAGCAGCTGCCCCAAGAA\\ \mathsf{G & S & D & Y & E & S & V & S & S & G & G & D & Y & I & P & T & L & A & S & S & C & P & \underline{K} & \underline{K} \\ ACCGGCGGGGTGTAAGAAGTTTCGTGAGACTCGTCACCCAATATACAGAGGAGTTCGTCGGAGAAACTCCGGTAA\\ P & A & G & \underline{R} & \underline{K} & \underline{K} & F & \underline{R} & E & T & \underline{R} & H & P & \underline{I} & Y & R & G & V & R & R & N & S & G & K \\ STGGGTTTGTGAGGTTAGAGAACCAAGAAACAAGAAAACAAGGATTTGGCTCGGAACATTTCAAACCGCTGAGATGGC \\ & \mathsf{W & V & C & E & V & R & \underline{R} & N & K & T & R & I & W & L & G & T & F & O & T & A & E & M & A \\ AGCTCGAGCTCACGACGTTGCCGCTTTAGCCCTTCGTGGCCGATCAGCCTGTCCAATTTCGCTGACTCGGCTTG \\ & R & A & H & D & V & A & A & L & R & M & M \\ GA\mathsf{GACTCCGAATCCCGAATCAACTTGCGCTAAGGACATCCAAAAGGCGGCGCGCGC$															М	N	S	F	S	Α	F	S	E	Μ	F
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	TG	GCT	CCGP	TTA	CGA	GTC	TTT	GGT	TTT	CTC	CAG	CGG	STG/	TT	ATA	TTC	GAG	CGC	TGG	GAG	CAG	CTO	SCCC	CAA	GAA
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	G	S	D	Y	E	S	S	V	S	S	G	G	D	Y	I	P	т	L	A	S	S	C	Ρ	K	K
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	AC	CGGG	GGG	TCO	TAP	GAZ	GTT	TCC	GTG/	GAG	TCC	TCA	ACCO	CAAS	TAT?	CAG	GAG	GAG'	TCC	TCO	GAG	AAA	CTO	CGG	TAA
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	Ρ	A	G	R	K	K	F	R	E	т	R	H	P	I	Y	R	G	V	R	R	R	N	S	G	K
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	GT	GGGT	TTT	TGA	GGT	TAC	AGA	ACC	CAAA	CA	GA	AAG	CAAC	GA	TTT	GCT	CGC	GAAG	CATT	TCA	AA	CGG	TG/	GAT	GGC
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	W	V	C	E	V	R	E	P	N	K	K	т	R	I	W	L	G	т	F	0	т	A	Е	Μ	A
A R A H D V A A L A L R G R S A C L N F A D S A W GAGACTCCGAATCCCGGAATCAACTTGGGCTAGGACATCCAAAAGGCGGCGGCTGCAGGTGGCGTGGGGGTTTCA R L R I P E S T C A K D I Q K A A A E A A L A F Q GGATGAGATGGTGGTGGCGACGATCATGGCTTCGACATGGAGGAGACGTTGGTGGGAGGCTATTTACACGGC D E M C D A T T D H G F D M E E T L V E A I Y T A GGAACAGAGGGAAAATGCGTTTTATATGCACGATGAGGCGATGTTTGAGATGCCGAGTTTGTTGGCTAATATGGC E Q S E N A F Y M H D E A M F E M P S L L A N M A AGAAGGGATGCTTTTACGCTCCGTCAGTGGAATCATAATCATGAAGTCGACGGCGATGATGACGACGT E G M L L P L P S V Q W N H N H E V D G D D D V ATCGTTATGGAGTTATTATAACTCAGATTATTATTTCCATTTTTATTATGTACGATGATCACGATCATTTTTTTT	AG	CTCC	GAGO	TCA	CGA	ACG	TGG	CGG	CTTT	TAG	CCCI	TCO	TG	GCCC	GATO	CAGO	CT	GTC	CA.	TTT	CGG	TGA	ACTO	GGG	TTG
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	A	R	A	H	D	V	A	A	L	A	L	R	G	R	S	A	C	L	N	F	A	_ D	S	A	W
R L R I P E S T C A K D I Q K A A A E A A L A F Q GGATGAGATGTGTGATGCGACGACGACGACGATCATGGCTTCGACATGGACGACGACGTGGTGGAGGCTATTACACGGC D E M C D A T T D H G F D M E E T L V E A I Y T A GGAACAGACGAAATGCGTTTATATGCACGATGAGGCGATGATTGAGAGGCGAGTTGTTGCGAGTTGGTGGACTATATGGC E Q S E N A F Y M H D E A M F E M P S L L A N M A AGAAGGGATGCTTTGCGGCTTCGGTCCGTCCGTCGGTGGAATCATAATCATGAAGTCGACGGCGATGATGACGACGT E G M L L P L P S V Q W N H N H E V D G D D D V AATCGTTATGAGAGTTATAAACTCAGATTATTATTTCCATTTTTATTGTACGATACTTTATTTA	GA	GACT	rccc	AAT	CCC	GGI	ATC	AAC	CTTC	CGG	CTAP	GGA	ACAS	CC1	AAA	AGG	CGGG	CGGG	TG/	AGO	TGO	GTT	rGGG	GTT	TCA
$\begin{array}{llllllllllllllllllllllllllllllllllll$	R	L	R	I	P	E	S	т	C	A	K	D	Ι	Q	K	A	A	A	E	A	A	L	A	F	0
D E M C D A T T D H G F D M E E T L V E A I Y T A GGAACAGAGCGAAAATGCGTTTTATATGCACGATGAGGCGATGTTTGAGATGCCGAGTTTGTTGGCTAATATGGC E Q S E N A F Y M H D E A M F E M P S L L A N M A AGAAGGGATGCTTTGCGCCCCTCCGTCCGTCAGTGGAATCATAATCATGAAGTCGACGGCGATGATGACGACGGCGATGATGACGACGGCGATGATGACGACGCGA TG D D D V A CGTTATGGAGTTATTATATATATATTATTTTCCATTTTAGTACGATACTTTTATTTTATTATTATTATTTT S L W S Y	GG.	ATGA	GAT	GTG	TGA	TGO	GAG	GAG	CGGI	ATC/	TGC	CT	CG/	CAT	rGG/	GGZ	GAG	CGT	rGGT	GGA	GGG	TAT	TT7	CAC	GGC
$\begin{array}{llllllllllllllllllllllllllllllllllll$	D	E	М	C	D	Α	т	т	D	Н	G	F	D	Μ	E	E	т	L	v	E	A	I	Y	т	A
E Q S E N A F Y M H D E A M F E M P S L L A N M A AGAAGGAATGCTTTTGCCGCTCCGTCGTGCAATGAATCATGAAGTCGACGGCGATGATGACGACGT E G M L L P L P S V Q W N H N H E V D G D D D V ATCGTTATGGAGTATTATTAAAACTCAGATTATTATTTCCATTTTTAGTACGATACTTTTTATTATTATTATTATTATTATTATTATTATTA	GG.	AACA	GAC	CGA	AAA	TGO	GT	TT7	ATAT	GC/	ACGA	TGA	AGGO	GAT	IGT?	TGA	GAT	rgco	GAG	TTT	GTT	GGG	TAZ	TAT	GGC
AGAAGGGATGCTTTTGCCGCTTCCGTCCGTACAGTGGAATCATAATCATGAAGTCGACGACGGCGATGATGACGACGT E G M L L P L P S V Q W N H N H E V D G D D D D V ATCGTTATGGAGTTATTAAAACTCAGATTATTATTTCCATTTTTAGTACGATACTTTTTATTTTTATTATTATTTTT S L W S Y	E	Q	S	E	N	Α	F	Y	Μ	H	D	E	A	М	F	E	М	P	S	L	L	A	N	Μ	Α
E G M L L P L P S V Q W N H N H E V D G D D D V ATCGTTATGGAGTTATTAAAAACTCAGATTATTATTATTATTATTATTATTATTATTATTATTAT	AG.	AAGO	GAT	GCT	TTT	GCC	GCT	TCC	CGTC	CGT	TACA	GTO	GA	TCA	TAT	ATCA	TGA	AAG	CGA	CGG	CGA	TGA	TGA	CGA	CGT
ATCGTTATGGAGTTATTAAAAACTCAGATTATTATTTCCATTTTTAGTACGATACTTTTTATTTTTATTATTATTATTTT SLWSY	Е	G	Μ	L	L	P	L	P	S	V	Q	W	Ν	H	N	Н	E	V	D	G	D	D	D	D	V
S L W S Y	AT	CGT	TATO	GAG	TTA	TT	AAZ	CTO	CAG	TT/	TTA	TT	ICC/	TT	PTT/	GTA	ACG	ATA	TTT	TTA	TT	TAT	TAT	TAT	TTT
	S	L	W	S	Y																				
	TYT.	CACT	TATA	AAA	AAA	AAA	AAZ	AAZ	AAA	AAZ	AAZ														

Discovery of C-repeat-binding factor 3 (CBF3) sequences



Discovery of CBF1 and CBF2 sequences

Yamaguchi-Shinozaki & Shinozaki, Plant Cell, 1994 Liu *et al.*, Plant Cell, 1998 21 Shinwari *et al.*, Biochem Biophys Res Commun., 1998

CBF signaling pathway



Regulation of the CBF signaling pathway during cold acclimation.

Cold signaling involves multiple regulatory mechanisms, including transcriptional, posttranscriptional, and posttranslational regulation, which are triggered by unknown temperature sensors.

Some key components of the circadian clock also have effects on plant cold responses.

CBF signaling pathway



Hierarchical clustering of all *COR* genes in the WT and three single *cbf* mutants after treatment with cold for 0 h, 2 h and 12 h (n = 3,766)



Schematic diagram of CBF regulation pathway

Shi et al., J Integr Plant Biol., 201723Data from Dr. Shi Yihao, unpublished

CBF variation in worldwide Arabidopsis populations



Genomic map and orientation of *CBFs* (A), and gene structures and polymorphism illustrations of *CBF1* (B), *CBF2* (C) and *CBF3* (D).

CBF variation in European Arabidopsis populations



CBF variation in Yangtze River Arabidopsis populations

Natural variation of *C-repeat-binding factor* (*CBFs*) genes is a major cause of divergence in freezing tolerance among a group of *Arabidopsis thaliana* populations along the Yangtze River in China

Juqing Kang¹, Huiting Zhang¹, Tianshu Sun¹, Yihao Shi¹, Jianqiao Wang¹, Baocai Zhang², Zhiheng Wang^{3,4}, Yihua Zhou² and Hongya Gu^{1,5}

¹State Key Laboratory for Protein and Plant Gene Research, College of Life Sciences, Peking University, Beijing 100871, China; ²The Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China; ³College of Urban and Environmental Sciences, Peking University, Beijing 100871, China; ⁴Center for Macroecology, Evolution and Climate, Department of Biology, University of Copenhagen, Universitetsparken 15, DK-2100, Copenhagen Denmark; ⁵The National Plant Gene Research Center (Beijing), Beijing 100101, China

Summary

Author for correspondence: Hongya Gu Tel: +86 10 6275 1847 Email: guhy@pku.edu.cn

Received: 8 March 2013 Accepted: 18 April 2013

New Phytologist (2013) 199: 1069–1080 doi: 10.1111/nph.12335

Key words: Arabidopsis thaliana, C-repeatbinding factor (CBF), freezing tolerance, local adaptation, natural variation, Yangtze River. We used a monophyletic group of four natural populations of Arabidopsis thaliana expanded from a single ancestor along the Yangtze River c. 90 000 yr ago to study the molecular mechanism of the divergence in their freezing tolerance, in order to gain an insight into the genetic basis of their local adaption to low temperatures.

 Freezing tolerance assays, measurements of metabolites in the raffinose biosynthesis pathway and transactivation-activity assays of variation in forms of cold-responsive transcription factors were conducted on the four populations. Quantitative trait locus mapping was adopted with F₂ populations of the most- and least freezing-tolerant populations.

 The degree of freezing tolerance among the four populations was negatively correlated with the lowest monthly average temperature of January in their native habitats, and positively correlated to the expression level of some cold-regulated genes. We identified a major locus harboring three cold-responsive transcription factor genes CBF1-3, and found a nucleotide insertion in CBF2 in all populations except SXcgx, which generated a dysfunctional CBF2 protein.

 The CBF2 in SXcgx experienced a stronger natural selection in the cooler environment after CBF3 lost its response to low temperature, which possibly reflects a local adaptation of these populations during the expansion from a common ancestor.

CBF variation in Yangtze River Arabidopsis populations

QTL- Mapping $\rightarrow CBF1/2/3$

The time course of expression of the three *CBF* genes in four natural populations and the Col ecotype of *Arabidopsis thaliana* under cold treatment.



He *et al.*, Chinese Science Bulletin , 20087 Kang *et al.*, New Phytologist , 2013

CBF variation in Yangtze River Arabidopsis populations



In silico studies of a gene or gene family (take *CBFs* as examples)

Song Zihan 2019/06/16

Use TAIR10 to find sequence of CBF3

https://www.arabidopsis.org/

AT4G25480 Other names: ATCBF3, C-REPEAT BINDING FACTOR 3, CBF3, DEHYDRATION RESPONSE ELEMENT B1A, DREB1A

encodes a member of the DREB subfamily A-1 of ERF/AP2 transcription factor family (CBF3). The protein contains one AP2 domain. There are six members in this subfamily, including CBF1, CBF2, and CBF3. This gene is involved in response to low temperature and abscisic acid.

Sequence: AT4G25480.1

Update History 🖗	AT4G25480.1 replaced At4g25480 on 2003-08-13 AT4G25480.1 replaces AT4G25480.1 on 2003-01-09 AT4G25480.1 replaces AT4G25480.1 on 2003-01-09	
Date last modified Name Tair Accession	2017-07-24 AT4G25480.1 Sequence:1005020824	
GenBank Accession Sequence Length (bp)	NM_118680 651	1 MNSFSAFSEM FGSDYESSVS SGGDYIPTLA SSCPKKPAGR KKFRETRHPI 51 NRCVRRRNSC NWYCHVRRDN NYTRIWLCTR OTAFWAARAN DVAALALRCR
Sequence Send to WU-BLAST	1 ATGAACTCAT TTTCTGCTTT TTCTGAAATG TTTGGCTCCG ATTACGAGTC 51 TTCGGTTTCC TCAGGCGGTG ATTATATTCC GACGCTTGCG AGCAGCTGCC 101 CCAAGAAACC GGCGGGTCGT AAGAAGTTTC GTGAGACTCG TCACCCAATA 151 TACAGAGGAG TTCGTCGGGAG AAACTCCGGT AAGTGGGTTT GTGAGGTTAG 201 AGAACCAAAC AAGAAAACAA GGATTTGGCT CGGAACATTT CAAACCGCTG 251 AGATGGCAGC TCGAGCTCAC GACGTTGCCG CTTTAGCCCT TCGTGGCCGA 301 TCAGCCTGTC TCAATTTCGC TGACTCGGCT TGGAGACTCC GAATCCCGGA 351 ATCAACTTGC GCTAAGGACA TCCAAAAGGC GGCGGCTGAA GCTGCGTTGG 401 CGTTTCAGGA TGAGATGTG GATGCGACGA CGGATCATGG CTTCGACATG 451 GAGGAGACGT TGGTGGAGGC TATTTACACG GCGGAACAGA GCGAAAATGC 501 GTTTTATATG CACGATGAGG CGATGTTGA GATGCCGAGT TTGTTGGCTA 551 ATATGGCAGA AGGGATGCTT TTGCCGCTTC CGTCCGTACA GTGGAATCAT 601 AATCATGAAG TCGACGGCGA TGATGACGAC GTATCGTTAT GGAGTTATTA 651 A	51 TRGVRRRNSG RWVCEVREPN RKTRIWLGTF QTAEMAARAA DVAALALRGR 101 SACLNFADSA WRLRIPESTC AKDIQKAAAE AALAFQDEMC DATTDHGFDM 151 EETLVEAIYT AEQSENAFYM HDEAMFEMPS LLANMAEGML LPLPSVQWNH 201 NHEVDGDDDD VSLWSY Amino acid sequence of CBF3

Use phytozome or Gramene for other plants

https://phytozome.jgi.doe.gov/pz/portal.html

		THE	PLANT GENOMI	CS RESOURCE							
JG	Species 🗸	Tools 🕶	Info 🕶	Download 🗸	Help 🕶	Cart	Subscribe				
Species 🗸	Amaranthus	s hypochondria	cus v1.0	Cucumis sati	vus v1.0		Panicum h	allii var. hallii v2.1			
Dhuton	Amaranthus	s hypochondria	icus v2.1	Daucus carot	a v2.0		Panicum v	irgatum v1.1			
FIIytoz	Amborella ti	richopoda v1.0)	Dunaliella sal	ina v1.0		Panicum v	irgatum v4.1			
	Anacardium	occidentale v	0.9	Eucalyptus gr	andis v2.0		Phaseolus	vulgaris v2.1			
	Ananas com	nosus v3		Eutrema sals	ugineum v1.0		Physcomit	rella patens v3.3			
	Aquilegia co	perulea v3.1		Fragaria vesc	a v1.1		Populus de	eltoides WV94 v2.1			
	Arabidopsis	halleri v1.1		Glycine max \	Vm82.a2.v1		Populus tri	ichocarpa v3.0			
All re spe	Arabidopsis	lyrata v2.1		Gossypium h	irsutum v1.1		Populus tri	ichocarpa v3.1			
4	Arabidopsis	thaliana Arap	ort11	Gossypium ra	imondii v2.1		Porphyra u	ımbilicalis v1.5			
	Arabidopsis	thaliana TAIR	10	Helianthus ar	nuus r1.2		Prunus per	rsica v2.1			
	Asparagus o	officinalis V1.1		Hordeum vulg	gare r1		Ricinus co	mmunis v0.1			
t	Boechera st	ricta v1.2		Kalanchoe fe	dtschenkoi v1	.1	Salix purpu	urea v1.0			
	Botryococcu	us braunii v2.1		Kalanchoe lax	kiflora v1.1		Selaginella	a moellendorffii v1.0			
	Brachypodiu	um distachyon	Bd21-3 v1.1	Lactuca sativ	a V8		Setaria ital	lica v2.2			
	Brachypodiu	um distachyon	v3.1	Linum usitati	ssimum v1.0		Setaria viri	idis v1.1			
	Brachypodiu	ım hybridum v	1.1	Malus domes	tica v1.0		Setaria viri	idis v2.1			
	Brachypodiu	um stacei v1.1		Manihot escu	lenta v6.1		Solanum lycopersicum iTAG2.4				
	Brachvnodi	ım svlvaticum	v1.1	Marchantia n	olvmornha v.3	1	Solanum tuberosum v4 03				

Use phytozome or Gramene for other plants

http://gramene.org/



• Search a gene

hytozo	ome quick se	arch (advance	d)				
		Flagships	Clustered Genomes	and Families	Unclustered Genon	nes	
a coerulea	Arabidopsis halleri v1.1	Arabidopsis lyrata	Arabidopsis thaliana TAIR10	Asterid	Boechera stricta v1.2	Brachypodium distachyon v3.1	Brachypodium stacei v1.1
•							Þ
ſ							_
	Search in Arab	dopsis thaliana colu	mbia TAIR10	for AT4	G25480	13	GO

Gene AT4G25480

▼Gene Info						
Organism	Arabidopsis tha	aliana columbia				
Transcript Name	AT4G25480.1 (primary)				
Location:	Chr4:13018214	I13019121 reverse				
Alias	ATCBF3 CBF3	DREB1A				
Description	dehydration res	ponse element B1A				
Gene Atlas Desc	Highly expresse	ed in root				
Links	в м 🛞 ta	air UniProt UniPr	ot			
Functional Annotation	Genomic Seque	ences Protein Homolo	igs Gene Ancestry	Expression		
			Protein do	main view		
1						216
			Functional annota	tions for this lo	cus	
ID ID	Туре	Description				

• Analyze your gene: gene structure and alternative splicing

Functional Annotation	Genomic	Sequences	Protein Homologs	Gene Ancestry	Expression			
			Chr4:1	3017851130194	84 :: View in JB	rowse		* III
13,018,000			13,018,	500		13,019,000	Full-screen view	Fullscree
💿 Transcript		+ AT4G25480.1						+
4				111				

Genomic tab of CBF3



Fig. 9 Genomic tab in Gene View

• Analyze your gene: get sequence

Functional Annotation	Genomic	Sequences	Protein Homologs	Gene Ancestry	Expression				
Gen	omic sequence	Transcript	sequence CDS set	quence	e sequence S	how all key: 5' UT	R CDS	3' UTR	
∧ Genomic Sequence [908]					BLAST this sequence	e at Phytozome	NCBI	
 Transcript Sequence 	[908]					BLAST this sequence	e at Phytozome	NCBI	1
▲ CDS Sequence [651]						BLAST this sequence	e at Phytozome	NCBI	
Functional Annotation	Genomic	Sequences	Protein Homologs	Gene Ancestry	Expression				
Geno	omic sequence	Transcript s	equence CDS see	quence	le sequence	how all key: 5' U	TR CDS	3' UTR	
>4T4G25480 Chr	- ·	13019121 r	everse						
CCTGAACTAGAACAGA	AAGAGAGAGAGAA	ACTATTATTTC	AGCAAACCATACCAAC	CAAAAAAGACAGAG	ATCTTTTAGTTAG	CCTTATCCAGTTTCTTGAAAC			
AGAGTACICITCIGAT	CAAIGAACICA	ITTICIGCITI	TICIGAAAIGIIIGGC	TCCGATTACGAGT	CITCGGITTCCTC	AGGCGGIGAIIATATTCCGA			

This tab contains genomic locus, transcript, CDS, and translated protein sequences for this gene. Each sequence is color-coded to indicate boundaries between UTR, coding and intronic sequences.

• Analyze your gene: protein homologs and evolution history

Func	tional	I Annotatior	n Ge	enomic Se	Sequences	Protein Homologs	Gene Ancest	ry Express	ion				
				All protoine	Drimer								
			snow:	All proteins	Primar	ry proteins							
		Views	Org	Defline			MRSF	Relationship	Score	Similarity	1	this gene	216
_			-										
		GB	Cru	Carubv10005	5709m - (1 o	of 4) PTHR31839:SF8 -	D BRA	M-1	1147	92.6%			
_													v

Protein Homologs tab of Gene View

Fun	ctional A	nnotation	Genomic	Sequences	Protein Homologs	Gene	Ancest	ry	Expres	sion												
	Views	Size	Node	Desc	ription		Мро	Ppa	Sfa	Smo	Ac	Atr	Mac	Spo	Ζ	Bdi	Bsta	Osa	Oth	Pha	Pvi	Sit
	F	1612	Viridiplantae	Auto:	(E=2e-144) PF00847 -	AP2	3	27	13	5	13	15	54	13	18	36	31	29	-	34	48	39
	F	1610	Embryophyte	Auto:	(E=2e-144) PF00847 -	AP2	3	27	13	5	13	15	54	13	18	36	31	29	-	34	48	39
	F	1183	Angiosperm	hypot	hetical protein		4		111													>

Gene Ancestry view

CENECC

VEC

111

DVI

Bst: Bostr.7867s0360.1.p Cru: Carubv10006965m Cgr: Cagra.2248s0026.1.p Car: Cagra.2248s0025.1.p

consensus/100%

• Analyze your gene: protein homologs and evolution history

Memb	ership At	th Bst (Cgr Cru 2 1											
Genes in	n Family F	unction	al Annotation	MSA	Family Histor	У								
M	Views	Org	ID	Alias/S	Symbol	Defline	Domains	S	ynteny	Exor	าร			
F	GB	Ath	AT4G25480.1	ATCBF	-3	dehydration response element B1A -{	\sim	Ю	-			Э		
Pi	GB	Bst	Bostr.7867s0360	D.1		(1 of 4) PTHR31839:SF8 - DEHYDRATION-R		-	Ю			-DH		
Pi	GB	Cgr	Cagra.2248s002	5.1		(1 of 4) PTHR31839:SF8 - DEHYDRATION-R			÷					×
F	G B	Car	Caara.2248s002	6.1		(1 of 4) PTHR31839:SF8 - DEHYDRATION-R				- 4		-CH		Ж
Genes ir	Family F	unctiona	al Annotation	MSA Descrip	Family Histor	у		Aha	Aly	Ath E	Bst Bo	I Bra	Cgr	Cru
Genes ir	n Family F	unctiona	Al Annotation Type PFAM	MSA Descrip AP2 do	Family Histor	у		Aha	Aly	Ath E	Bst Bo	I Bra	Cgr 2	Cru 1
Genes ir ID PF(GO	n Family F 00847 0003677	unctiona	Type PFAM GO	MSA Descrip AP2 do Any mo	Family Histor otion omain olecular functi	y on by which a gene product interacts selectively	and n	Aha -	Aly -	Ath E 1 1	Bst Bo 1 -	I Bra	Cgr 2	Cru 1

COVUDACDVVEDETDU

• Analyze your gene: expression information

Gene AT4G25480

▼Gene Info							
Organism	Arabidopsis th	naliana co	olumbia				
Transcript Name	AT4G25480.1	(primary))				
Location:	Chr4:1301821	413019	121 reverse				
Alias	ATCBF3 CBF	3 DREB1	IA				
Description	dehydration re	esponse e	element B1A				
Gene Atlas Desc	Highly express	sed in roo	ot				
Links	в м 🍪	tair Uni	Prot UniProt				
Functional Annotation	Genomic Seq	uences	Protein Homologs	Gene Ancestry	Expression		
Show:	Expression	Coexpr	ession				
Experiment Group		Experimen	it Name		FPKM	Locus DE	Library DE
GeneAtlas Tissue Sample		Leaf.amm	nonia		0.986		

Expression and co-expression of a gene

For more information, please look into HELP for help. https://phytozome.jgi.doe.gov/pz/QuickStart.html

Use phytozome to do BLAST



Search for genes, families and sequences

1. Select a Target 0 species select	ted 💌	2. Build your query GO
Target set: Phytozome 12.1 Pre-release species Target type: Ancestor nodes Species		Search type: Keyword BLAST
Select or enter a species or node (e.g. Arabidopsis, Rosid) Panicum virgatum v1.1	▲	NFADSAWRLRIPESTCAKDIQKAAAEAALAFQDETCDTTTTNHGLDMEETM VEAIYTPEQSEGAFYMDEETMFGMPTLLDNMAEGMLLPPPSVQWNHNYD GEGDGDVSLWSY
Setaria italica v2.2 Setaria viridis v1.1 Sorghum bicolor v3.1.1 Zea mays Ensembl-18	÷	Algorithm parameters
Eudicot Aquilegia coerulea v3.1		Query name: (optional)
Pentapetalae Amaranthus hypochondriacus v1.0		 View results in browser Notify by email (long or multifasta jobs)
Daucus carota v2.0	•	Target type: Target type: Genome ▼ Program: Target type: Genome totide db ▼ Expect (E) threshold: Target type: Proteome Target type: Proteome

Use phytozome to do BLAST

• As *CBF3* is conserved in *Viridiplantae*, we can choose some plants as representatives.

BLAST Results

₹E	BLA	ST Inp	uts							
с т	Quer	y your.se t Capsel	eq (216 letters) la grandiflora v1.1 proteome (26561 sequences, 1103	38161 t	otal lette	ers)				
Pro	gran	n BLAST	P 2.2.26+			,				
Hits	Foun	id 100				Download results	elect BLAST format	▼		
		Views	Defline	Score	E	Query View	query sequence	2	.16	
		GB	Cagra.28753s0001.1 (1 of 4) PTHR31839:SF8 - DEHYDRATI	326.3	3.9E-113			1-	216	*
		GB	Cagra.2248s0026.1 (1 of 4) PTHR31839:SF8 - DEHYDRATIO	323.9	4.3E-112			2-	216	=
		GB	Cagra.0094s0011.1 (1 of 4) PTHR31839:SF8 - DEHYDRATIO	229.2	6.7E-75			1-	216	
		GB	Cagra.2248s0025.1 (1 of 4) PTHR31839:SF8 - DEHYDRATIO	158.7	4.9E-49			2-	85	
		GB	Cagra.0568s0171.1 (1 of 2) PTHR31839:SF0 - DEHYDRATIO	125.9	4.4E-35			29	9-136	+
Cor Acc	ntact cessib 997-20	Disclaimer ility / Section 017 The Reg	n 508 Statement ents of the University of California						ENT OF	Office of Science

Use protein function, length, and sequence alignment as criteria to screen your protein from the results .



A more convenient tool? GCORN Plant

http://www.plant.osakafu-u.ac.jp/~kagiana/gcorn/p/index.html





Gcorn plant: a database for retrieving functional and evolutionary traits of plant genes

Yoshiyuki Ogata, Naohiro Kimura, Ryosuke Sano

Prediction of domains Prosite

• https://prosite.expasy.org/cgi-bin/prosite/ScanView.cgi?scanfile=234757860404.scan.gz

CBF3 (216 aa)					
MNSFSAFSEMFGSDYESSV: REPNKKTRIWLGTFQTAEM LAFQDEMCDATTDHGFDME NHNHEVDGDDDDVSLWSY	SSGGDYIPTLASSCPKI AARAHDVAALALRGRSJ STLVEAIYTAEQSENAI	KPAGRKKFRETRHP <mark>IYRG</mark> <mark>ACLNFA</mark> DSAWRLRIPEST FYMHDEAMFEMPSLLANM	<mark>VRRRNSGKWVCEV</mark> CAKDIQKAAAEAA AEGMLLPLPSVQW		
Legend:					
disulfide bridge	 active site 	other 'ranges'	+ other sites		

Please note that the graphical representations of domains displayed hereafter are for illustrative purposes only, and that their colors and shapes For more information about how these graphical representations are constructed, go to https://prosite.expasy.org/mydomains/.

hits by profiles: [1 hit (by 1 profile) on 1 sequence]

Upper case represents match positions, lower case insert positions, and the '-' symbol represents deletions relative to the matching profile.



PS51032 AP2_ERF AP2/ERF domain profile :

Prediction of DNA binding site

http://biomine.cs.vcu.edu/servers/DRNApred/?tdsourcetag=s_pctim_aiomsg

DRNApred - DNA- and RNA-binding residues predictor

References | Materials | Help | Acknowledgments | Disclaimer | Biomine

DRNApred webserver

DRNApred is a server providing sequence based prediction of DNA- and RNA-binding residues.

Please follow the three steps below to make predictions:

1. Upload a file with protein sequences, or paste them into text area

Server accepts up to 100 (FASTA formated) protein sequences. Either upload a file or enter each protein in a new line in the following text field (see Help for details):

Choose a file No file chosen

><u>CBF3</u>

MNSFSAFSEMFGSDYESSVSSGGDYIPTLASSCPKKPAGRKKFRETRHPIYRGVRRRNSGKWVCEVREPNKKTRIWLGTFQ TAEMAARAHDVAALALRGRSACLNFADSAWRLRIPESTCAKDIQKAAAEAALAFQDEMCDATTDHGFDMEETLVEAIYTAE QSENAFYMHDEAMFEMPSLLANMAEGMLLPLPSVQWNHNHEVDGDDDDVSLWSY

Prediction of DNA binding site

http://biomine.cs.vcu.edu/servers/DRNApred/?tdsourcetag=s_pctim_aiomsg

>CBF3				
Amino Acid	probability_DNA	binary_DNA	probability_RNA	binary_RNA
m	0.3248	0	0.0660	0
n	0.3890	0	0.0560	0
S	0.3690	0	0.0564	0
f	0.1564	0	0.0560	0
S	0.2944	0	0.0534	0
а	0.1026	0	0. 0439	0
f	0.1633	0	0.0685	0
S	0.1187	0	0. 0497	0
е	0.1077	0	0. 0446	0
m	0.1145	0	0. 0499	0
f	0.1058	0	0.0570	0
g	0.1204	0	0.0561	0
S	0.1723	0	0.0624	0
d	0.1814	0	0.0717	0
у	0.2686	0	0.0659	0
е	0.1474	0	0.0622	0
S	0.1853	0	0.0679	0
S	0.1405	0	0.0649	0
v	0.1272	0	0.0547	0
S	0.2521	0	0.0643	0
S	0.1690	0	0.0667	0
g	0.1619	0	0.0655	0
g	0.1776	0	0.0629	0
d	0.1596	0	0.0641	0
у	0.1924	0	0.0627	0
i	0.1332	0	0.0582	0
р	0.1247	0	0.0659	0
t	0.1137	0	0. 0568	0
1	0.0881	0	0. 0531	0
a	0.1004	0	0.0604	0
S	0.1609	0	0.0777	0
S	0.1406	0	0.0724	0
с	0.1336	0	0. 0783	0
р	0.1139	0	0. 0775	0
k	0.3080	0	0. 0890	0

Mutations of DNA binding residues may change the affinity and function of a transcription factor.

正负电荷残基数

Total number of negatively charged residues (Asp + Glu): 33 Total number of positively charged residues (Arg + Lys): 23

Prot-	Para	m		Atomic composition	☞ 原子组成	
http://www.expasy.org/tools/protparam.html			Carbon C Hydrogen H Nitrogen N	1055 1615 297		
氨基酸数目	Number of	amino acids : 2	16	Oxygen O Sulfur S	331 15	
相对分子质量	Molecular weight: 24236.12 Theoretical pI: 5.08		Formula: C ₁₀₅₅ H ₁₆₁₅ N ₂₉₇ O ₃₃₁ S ₁₅ 分子式 Total number of atoms: 3313 总原子数			
理论pI值						
	Amino acid	l campositian:	CSV format	Extinction coeffic	cients: 消光系数	
	Ala (A) 2 Arg (R) 1	7 12.5% 4 6.5%		Extinction coeffic	ients are in units of M ⁻¹ cm ⁻¹ , at 280 ;	nm measured in water.
与甘欧加大	Asn (N) Asp (D) 1 Cys (C) Cla (O)	8 3.7% 5 6.9% 5 2.3%		Ext. coefficient Abs 0.1% (=1 g/l)	36690 1.514, assuming all pairs of Cys resid	lues form cystines
虽态酸组成	Glu (E) 1 Glu (E) 1 Gly (G) 1	0 2.3% 8 8.3% 1 5.1%		Ext. coefficient Abs 0.1% (=1 g/l)	36440 1.504, assuming all Cys residues are r	reduced
	Ile (I)	6 2.8%		Estimated half-lif	fe: 半衰期	
	Leu (L) 1 Lys (K) -	4 6.5% 9 4.2%		The N-terminal of ·	the sequence considered is M (Met).	
	Met (M) 1 Phe (F) 1 Pro (P) 5 Ser (S) 1	0 4.6% 0 4.6% 9 4.2% 9 8.8%		The estimated half	-life is: 30 hours (mammalian reticulocy >20 hours (yeast, in vivo). >10 hours (Escherichia coli, i	rtes, in vitro). n vivo).
	Thr (T) 1 Trp (W)	0 4.6% 5 2.3%		Instability index	不稳定系数	
	Tyr (Y) Val (V) Pyl (O)	6 2.8% 9 4.2% 0 0.0%		The instability in This classifies th	ndex (II) is computed to be 53.99 ne protein as unstable.	
	(B) 0	0 0.0% 0.0%		Aliphatic index:	60.69 脂肪族系数	
	(Z) 0 (X) 0	0.0%		Grand average of 1	hydropathicity (GRAVY): -0.488	46

总平均亲水性

ProtScale

https://web.expasy.org/protscale/



Analysis of hydrophilicity/hydrophobicity of CBF3

SOPMA https://npsa-prabi.ibcp.fr/

α-helix, H
β-turn <i>,</i> T
random coil, C
extended strand, E

SOF	PMA :				
	Alpha helix	(Hh)	:	64 is	29.63%
	3 ₁₀ helix	(Gg)	:	0 is	0.00%
	Pi helix	(Ii)	:	0 is	0.00%
	Beta bridge	(Bb)	:	0 is	0.00%
	Extended strand	(Ee)	:	27 is	12.50%
	Beta turn	(Tt)	:	4 is	1.85%
	Bend region	(Ss)	:	0 is	0.00%
	Random coil	(Cc)	:	121 is	56.02%
	Ambiguous states	s (?)	:	0 is	0.00%
	Other states		:	0 is	0.00%



Prediction of secondary structure of CBF3 protein

Swiss-Model

https://www.swissmodel.expasy.org/





Prediction of tertiary structure of CBF3 protein

Post-translational modification Plant PTM Viewer

http://www.psb.ugent.be/PlantPTMViewer



Arabidopsis thaliana: 100,623 PTMs in 41,409 proteins

Chlamydomonas reinhardtii: 17,064 PTMs in 5,324 proteins

Oryza sativa ssp. japonica : 56,606 PTMs in 19,500 proteins

Triticum aestivum : 53,580 PTMs in 25,150 proteins

Zea mays: 143,869 PTMs in 37,099 proteins

PSORT II Prediction

https://psort.hgc.jp/

Results of the *k***-NN Prediction**

k = 9/23

47.8 %: cytoplasmic 30.4 %: nuclear 13.0 %: mitochondrial 4.3 %: cytoskeletal 4.3 %: vesicles of secretory system

>> prediction for QUERY is cyt $(k{=}23)$

Prediction of subcellular localization of CBF3

STITCH & STRING

http://stitch.embl.de & https://string-db.org/



Network of interactions between CBF3 and other proteins

Nothing in Biology Makes Sense Except in the Light of Evolution.

——Theodosius Dobzhansky, 1973

Acknowledgement

- Prof. Luo Jingchu, Prof. Gu Hongya, Prof. Qu Li-jia
- Dr. Kang Juqing
- Dr. Sun Tianshu, Dr. Shi Yihao

THANKS