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*?

Annual, dicot
Cruciferae, *Arabidopsis*

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

Gross morphological variation in *Arabidopsis* and relatives.

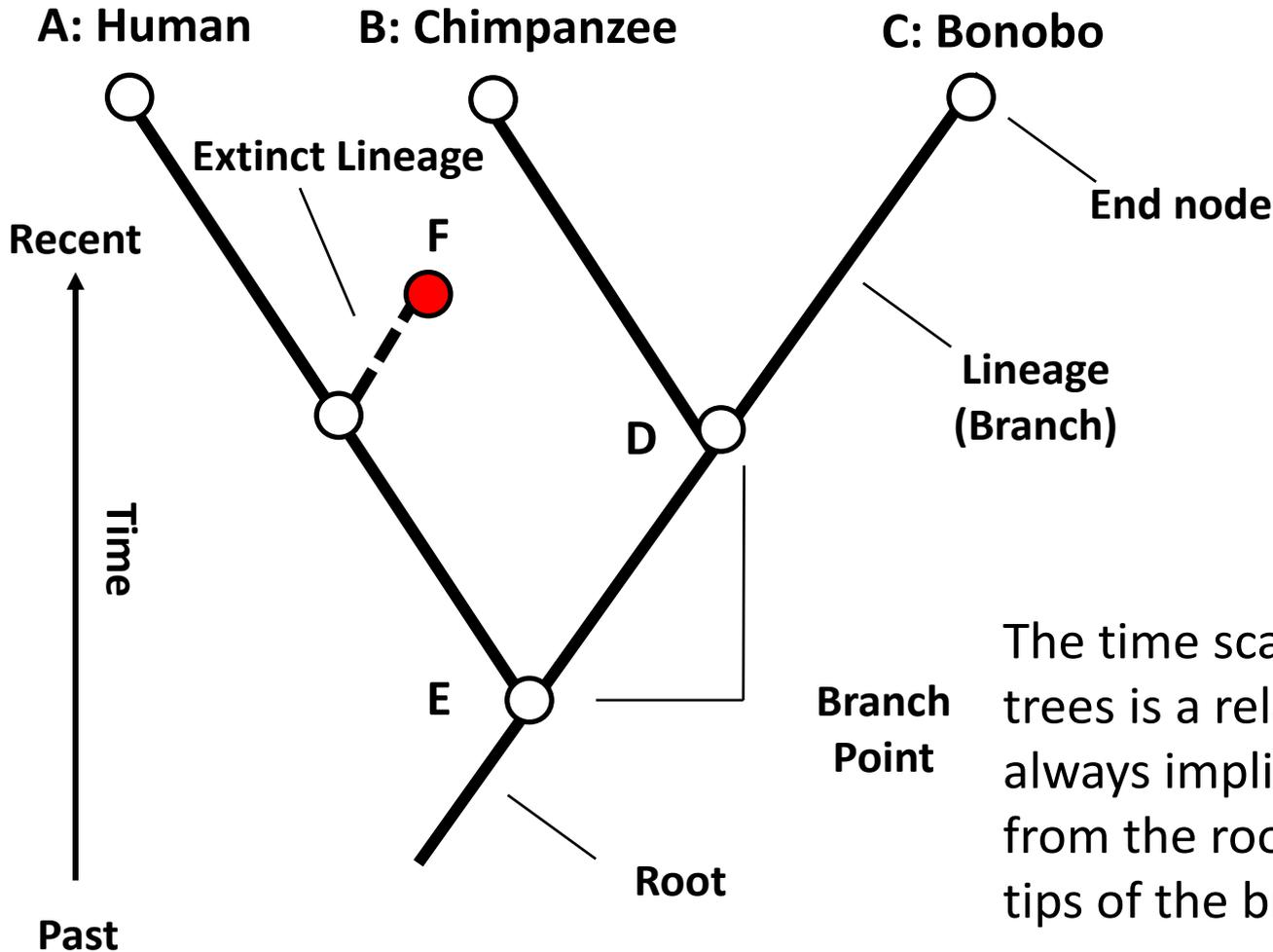*Capsella rubella*   *Arabidopsis thaliana*   *A. lyrata*

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*.

4

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.

5

Phylogenetic tree

A: Human   B: Chimpanzee   C: Bonobo

Extinct Lineage

End node

Recent

F

Lineage
(Branch)

Time

D

E

Branch
Point

Root

Past

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	Plesiomorphy refers to an evolutionary character homologous within a particular clade but, not unique to all members of that particular clade
An apomorphic character is similar throughout the clade	A plesiomorphic character is not similar throughout the clade
The two types are autapomorphy and synapomorphy	Not divided further

| A derived or specialized character	A primitive or ancestral character
An apomorphic character can be used as a defining character for the clade	A plesiomorphic character cannot be used as a defining character for the clade
Ex: absence of legs in snakes	Ex: presence of legs in reptiles

Visit www.pediaa.com

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.
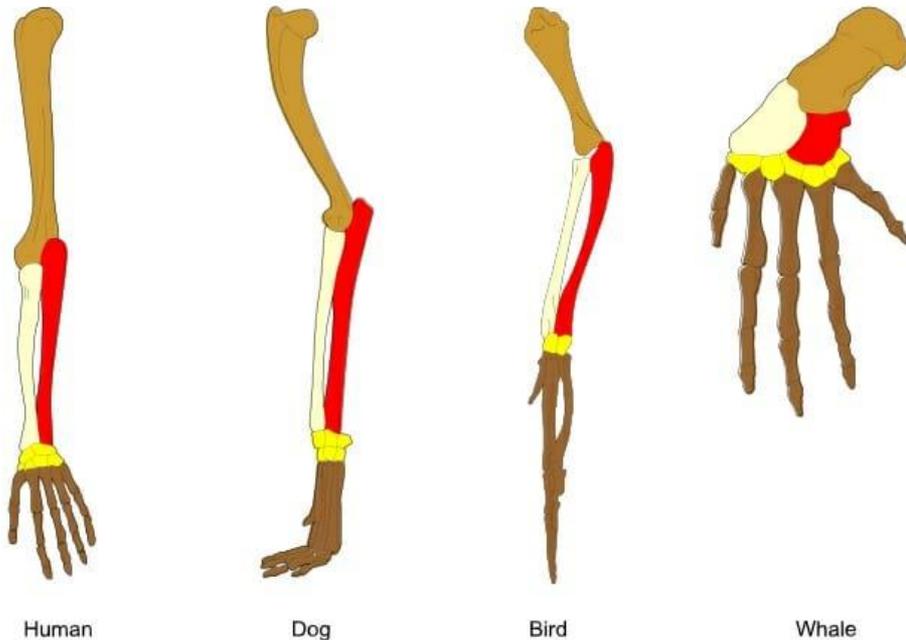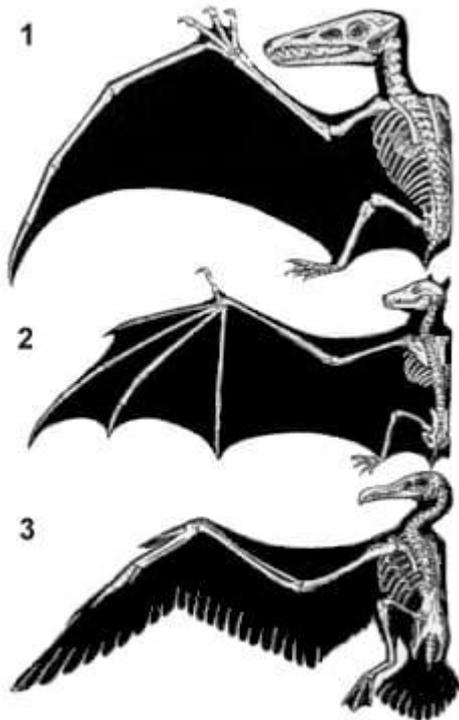

The homologous bones found in the forelimb of four different types of mammal.
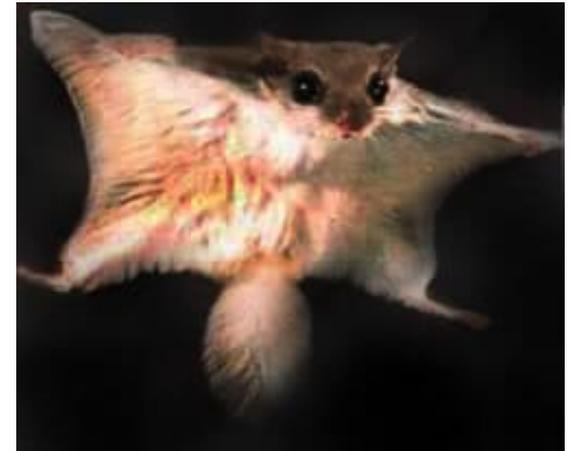
Darwin's finches

https://biologydictionary.net/divergent-evolution/
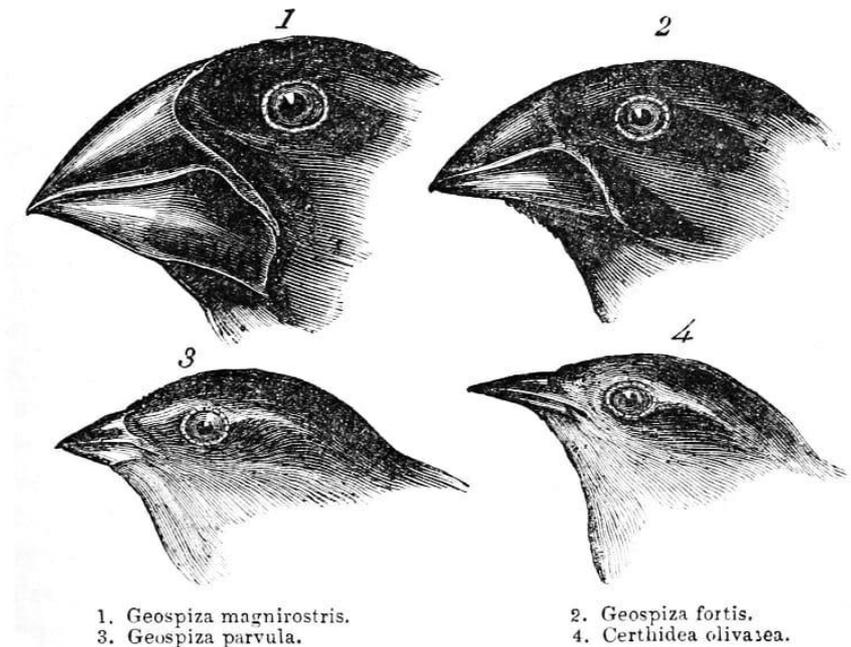
Convergent Evolution

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

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

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

9

https://biologydictionary.net/Convergent-evolution/

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.
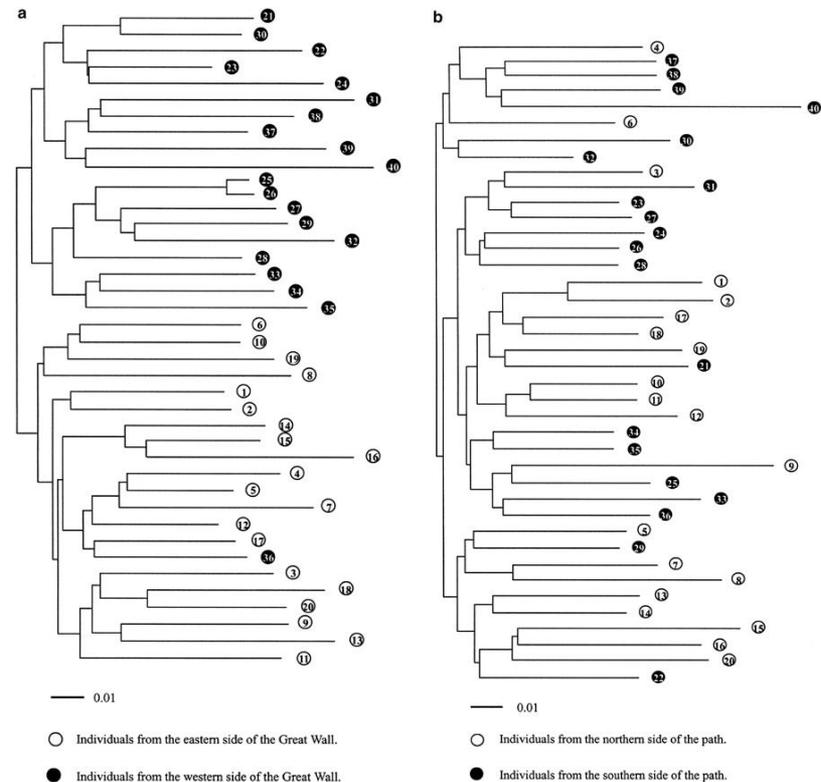
Juyong-guan

Control group, mountain path

Su *et al.*, Heredity, 2003

The Great Wall of China: a physical barrier to gene flow?
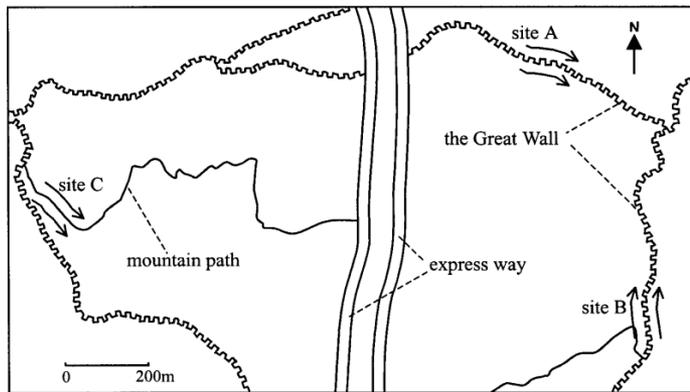
- *Prunus armeniaca*

- *Ziziphus jujuba*

- *Vitex negundo*
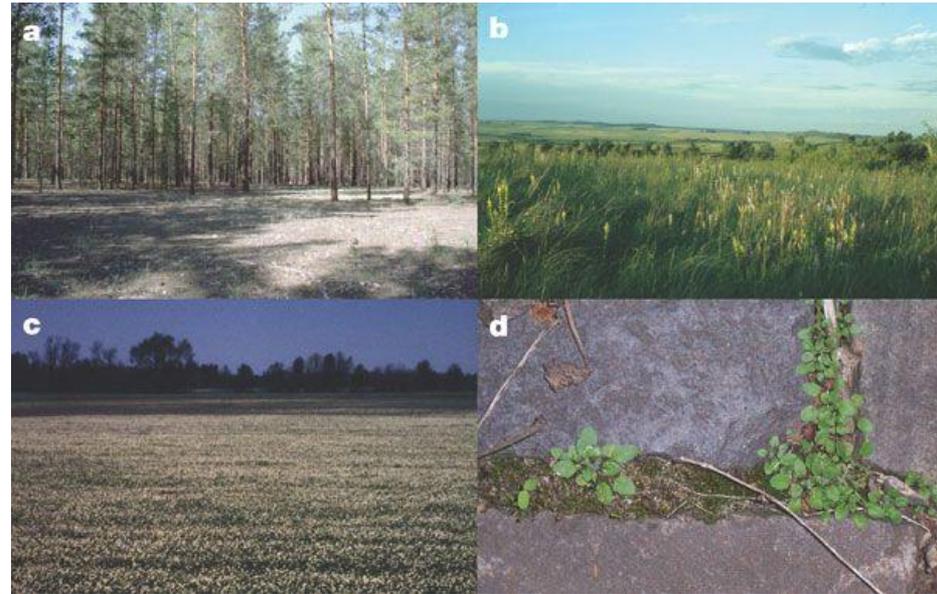
- *Heteropappus hispidus*

- *Cleistogenes caespitosa*

- *Ulmus pumila*

- *Ulmus macrocarpa*

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

11

Su *et al.*, Heredity, 2003

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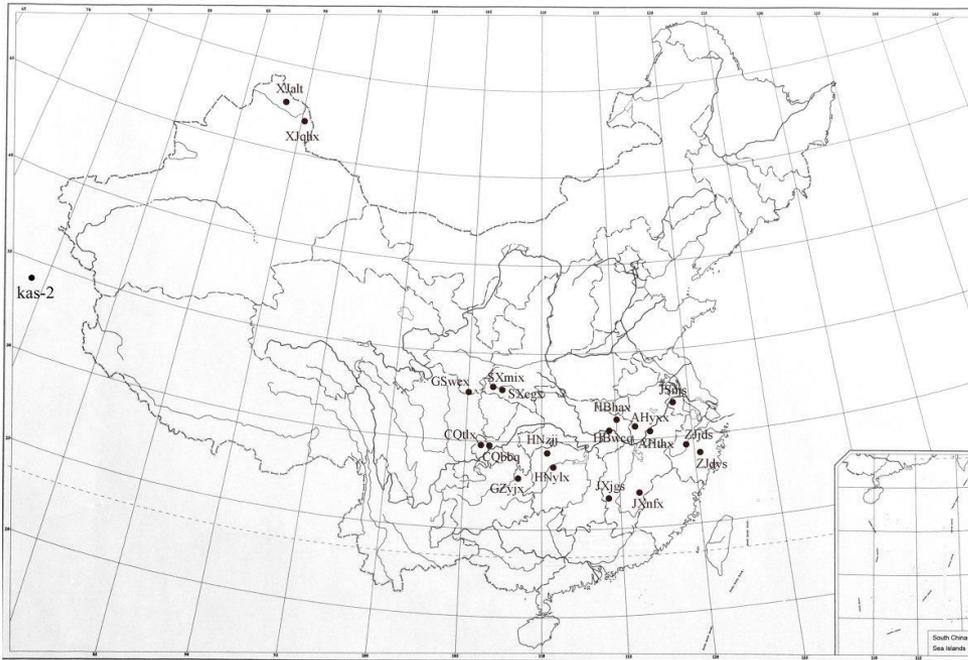
Distribution map of the 19 accessions of *Arabidopsis thaliana* from China and one from India (Kas-2).

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'-CATGGAACGGGATGTTTTTA-3'
5'-GTTCTACGCCTCCGAGCTAT-3'

5'-TCCTCGATGTTGTTTACGAAATCT-3'
5'-TCGAACTATTTATGGGGTTTTAGG-3'

Information about eleven pairs of primers for PCR amplification.

13

Yin *et al.*, BMC Plant Biology, 2010

Evolutionary history of Chinese *Arabidopsis thaliana* populations
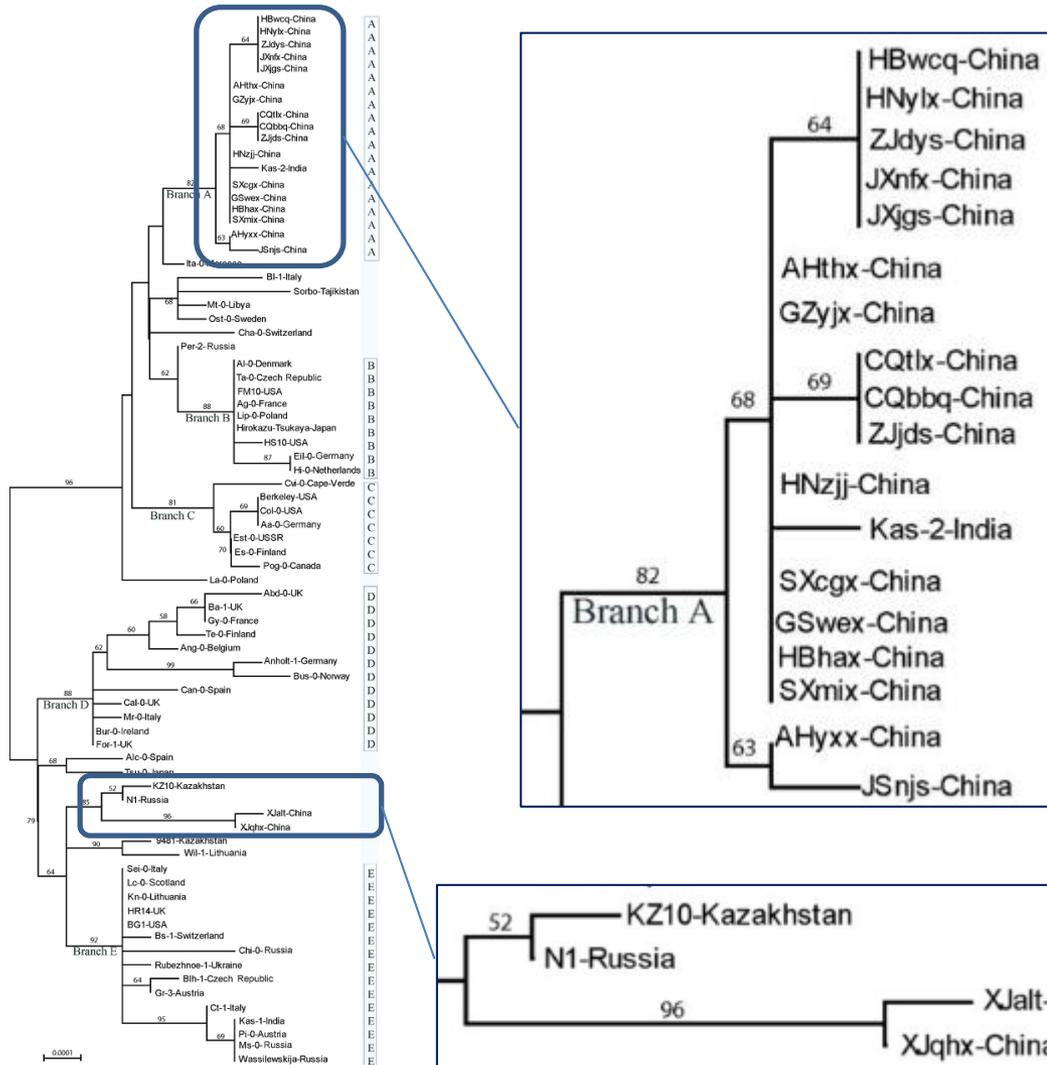
- The origin types of Chinese wild *Arabidopsis thaliana* populations can be divided into two categories.

- **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.

14

Yin *et al.*, BMC Plant Biology, 2010

Different ecotypes of *Arabidopsis thaliana*

Col-0, USA

Ler-0, Germany

Cvi-0, Cape Verde
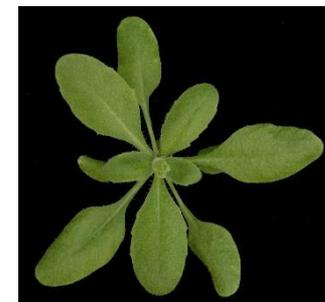
Per-1, Russia

Bla-1, Spain

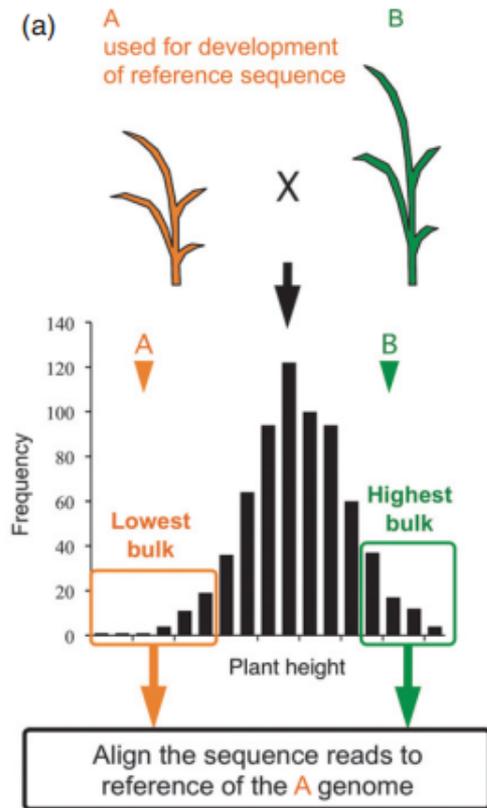Vajug-1, Serbia

Altai-5, China

Bor-1, Czech

Haz-2, USA

Ale-Stenar-56-14, Sweden

IP-Mos-1, Portugal

Neo-6, Tajikistan

How to study *Arabidopsis* natural variation?

Trontin *et al.*, Current Opinion in Plant Biology, 2011
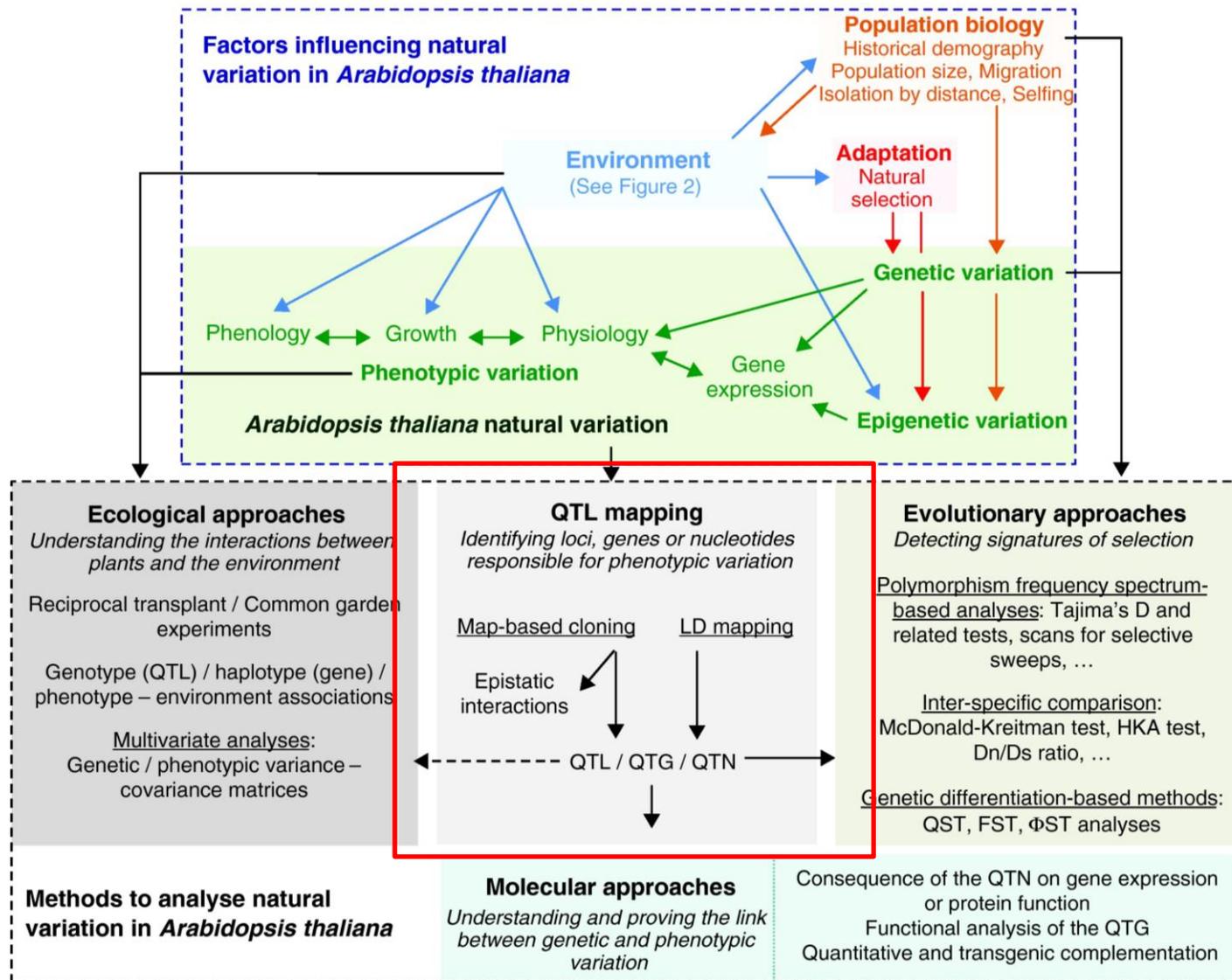
Quantitative Trait Loci (QTL)-Mapping
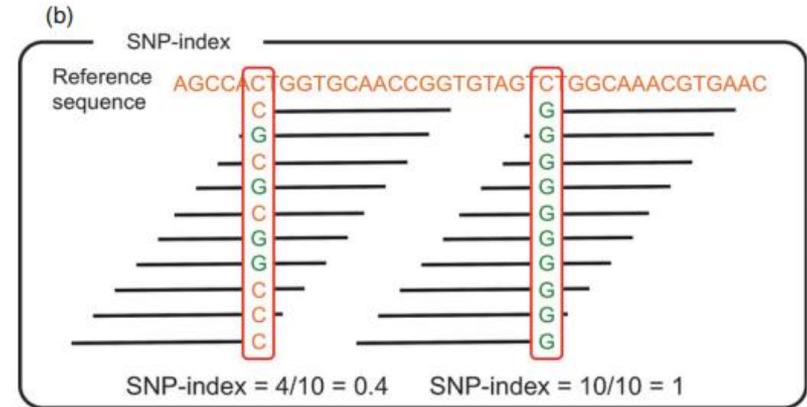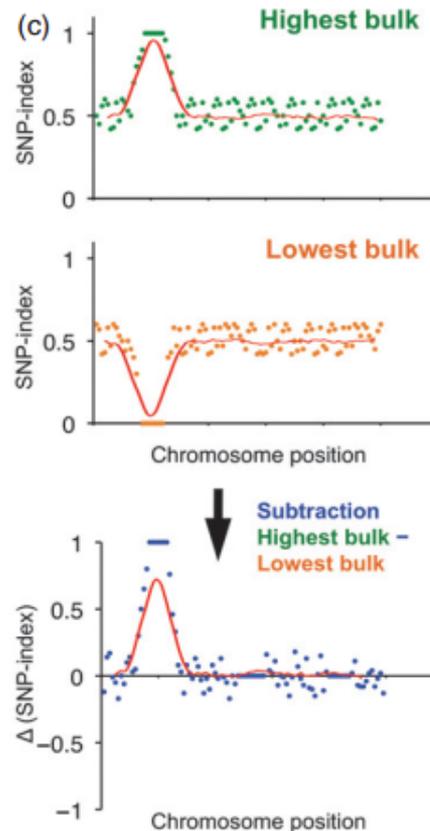
SNP-index: Short reads generated by whole-genome sequencing are aligned to the reference sequence. If 10 short reads cover a given nucleotide position, the coverage of the site is 10. Among the 10 reads, if four contain a SNP different from the reference nucleotide, the SNP-index is defined as 0.4. On the other hand, if all the reads harbor a SNP different from the reference, the SNP-index is 1.0.

CRISPR-Cas9

Step I:
Selection of the genomic target

Gene

Genomic target sequence

20bp sequence followed by the PAM sequence (NGG)

....NNNN**NNNNNNNNNNNNNNNNNNNN**NGGNNNNNN....

Step II:
sgRNA designing and its expression

guide sequence
NNNNNNNNNNNNNNNNNNNN   sgRNA backbone

expressed under a small RNA promoter

sgRNA

Step III:
Assembly of sgRNA and Cas9

Cas9   sgRNA   sgRNA:Cas9 complex

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

Step V:
Target DNA site cleavage by sgRNA:Cas9 complex

Genomic DNA   PAM

Cas9   sgRNA:Cas9 complex

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

Cas9

DNA

sgRNA

DSB

NHEJ   HDR

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

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   80 MPI Accessions   195 Salk Accessions   Legacy Projects
                             (GMINordborg2010)   (MPICao2010)

http://1001genomes.org/

Cold response in plants
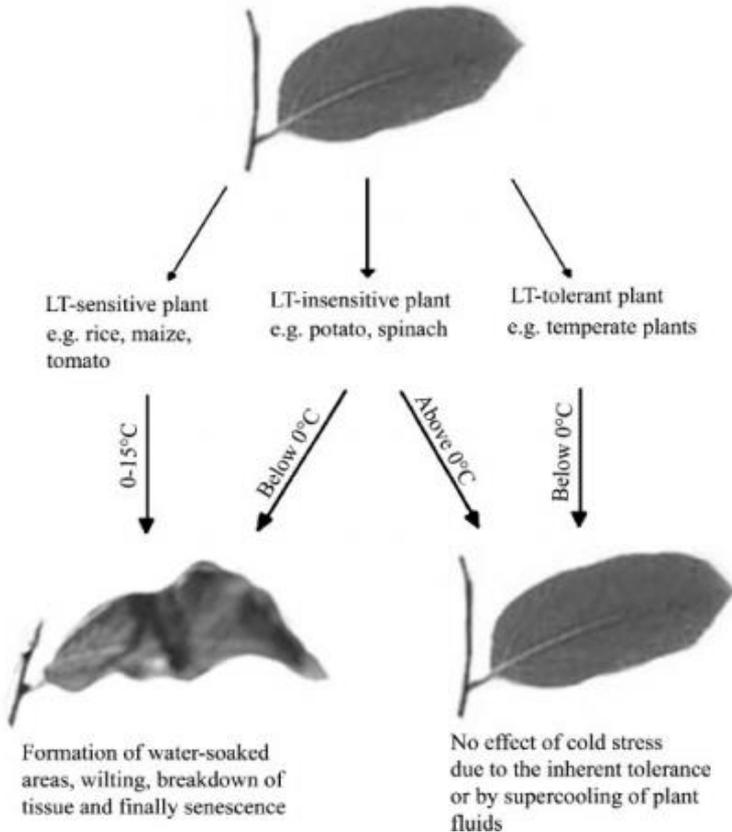
Physiological manifestation of cold stress by LT-sensitive, -insensitive and –tolerant plants
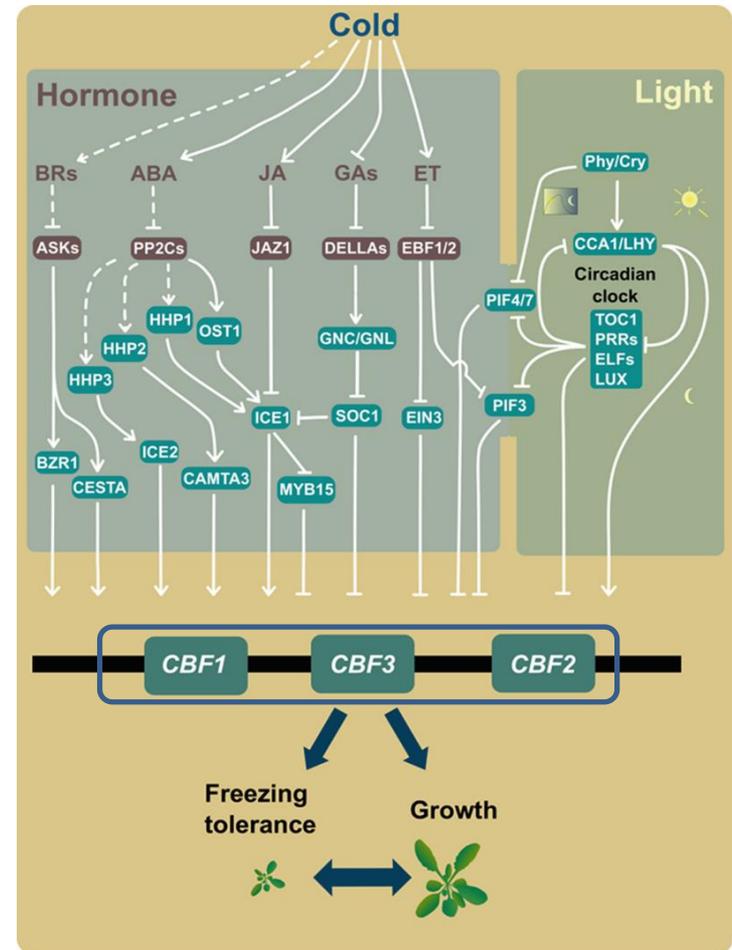
The involvement of plant hormone signaling in plant responses to cold stress

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

Discovery of *CBF1* and *CBF2* sequences

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

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

21

CBF signaling pathway
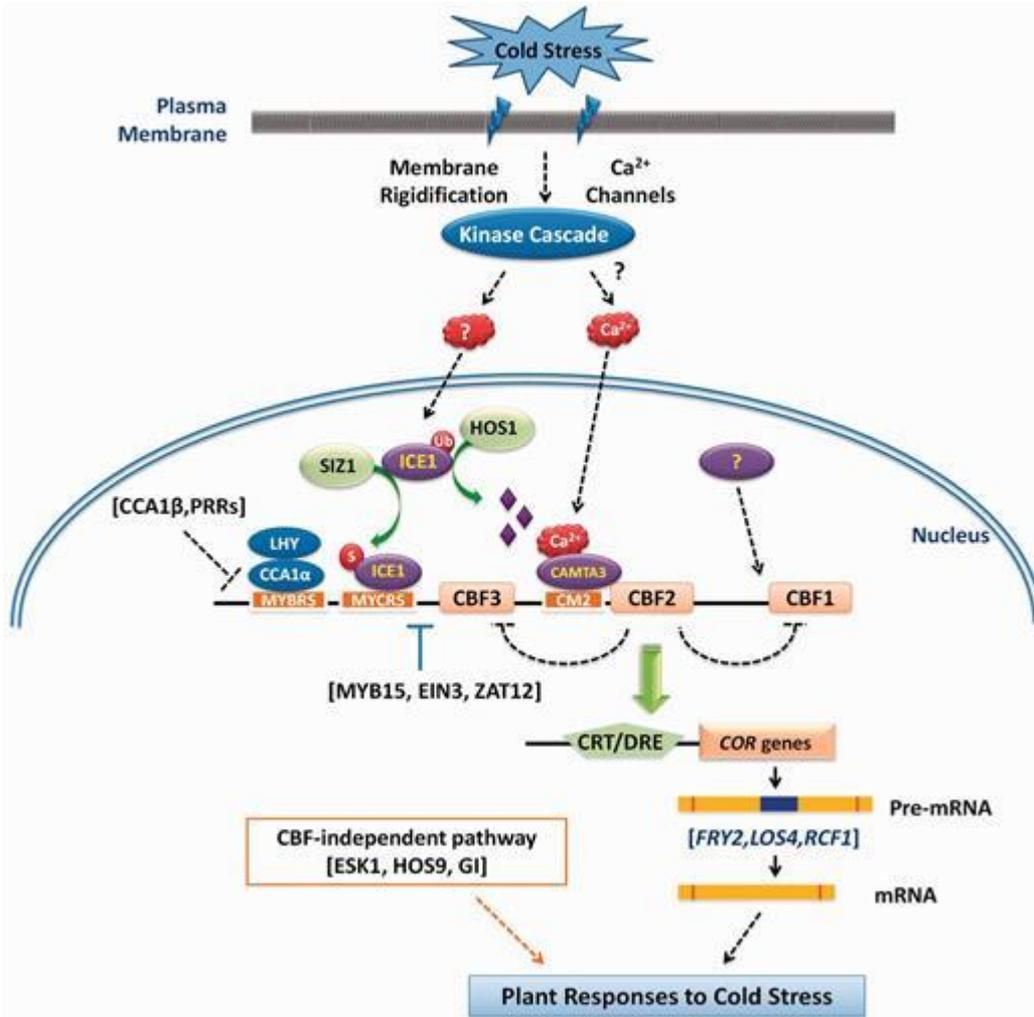
Regulation of the CBF signaling pathway during cold acclimation.

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.

Shi *et al.*, Plant Cell Physiol., 2015   22

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)
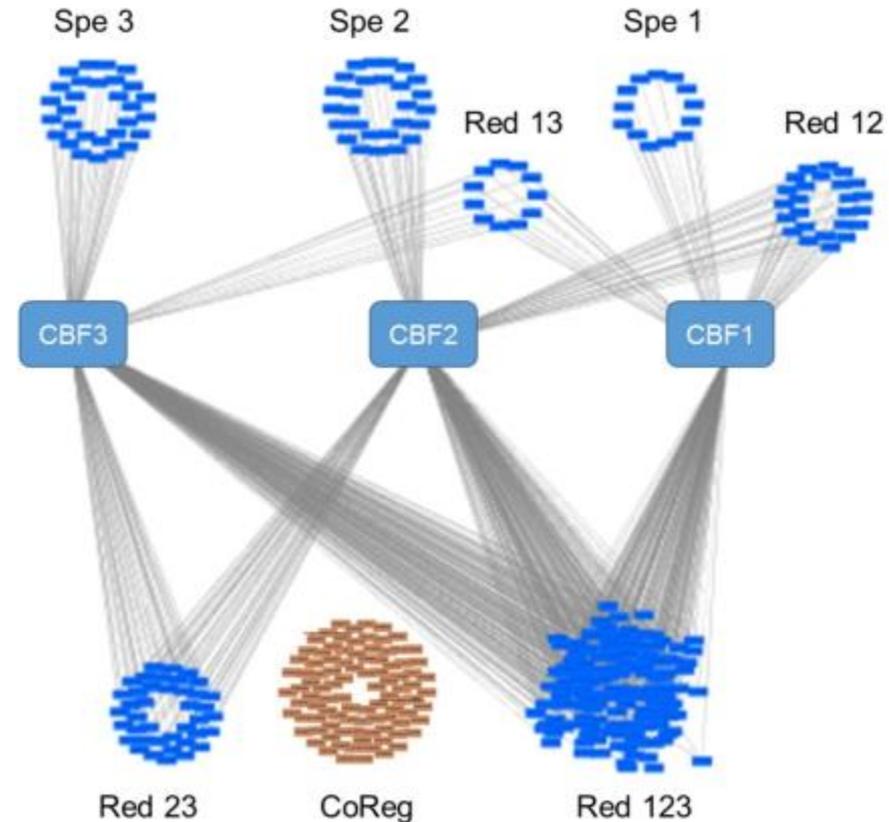
Schematic diagram of CBF regulation pathway

23

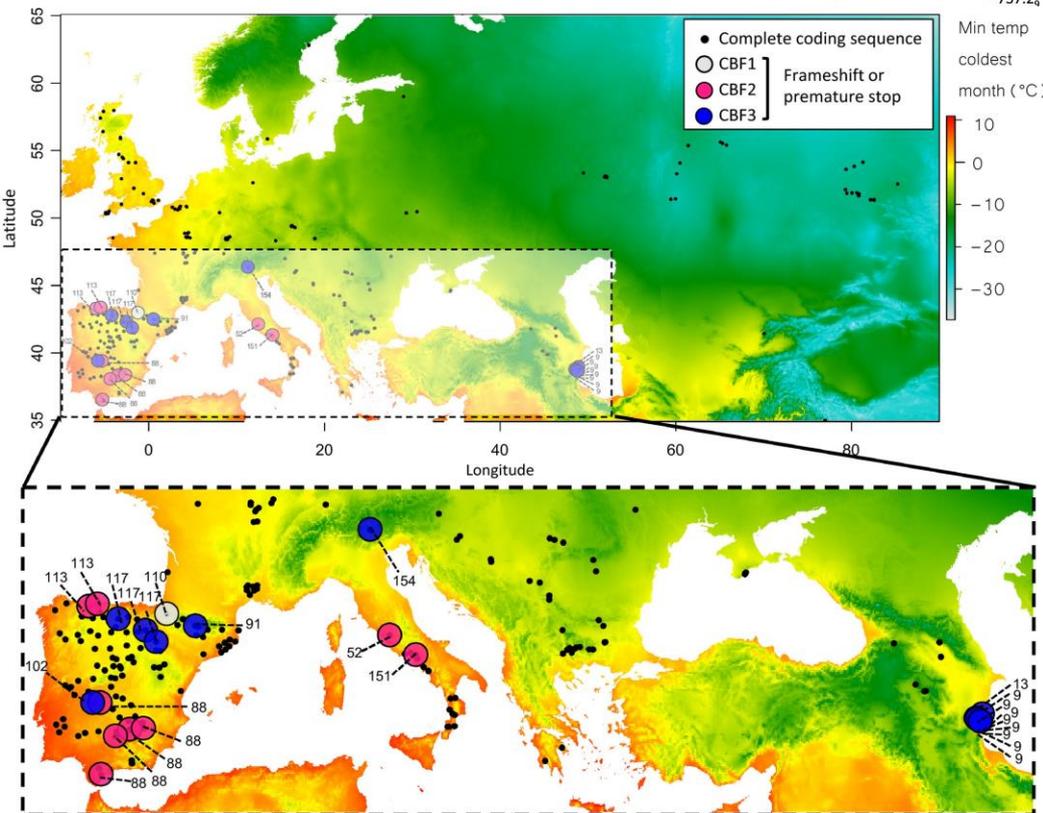# *CBF* variation in worldwide *Arabidopsis* populations
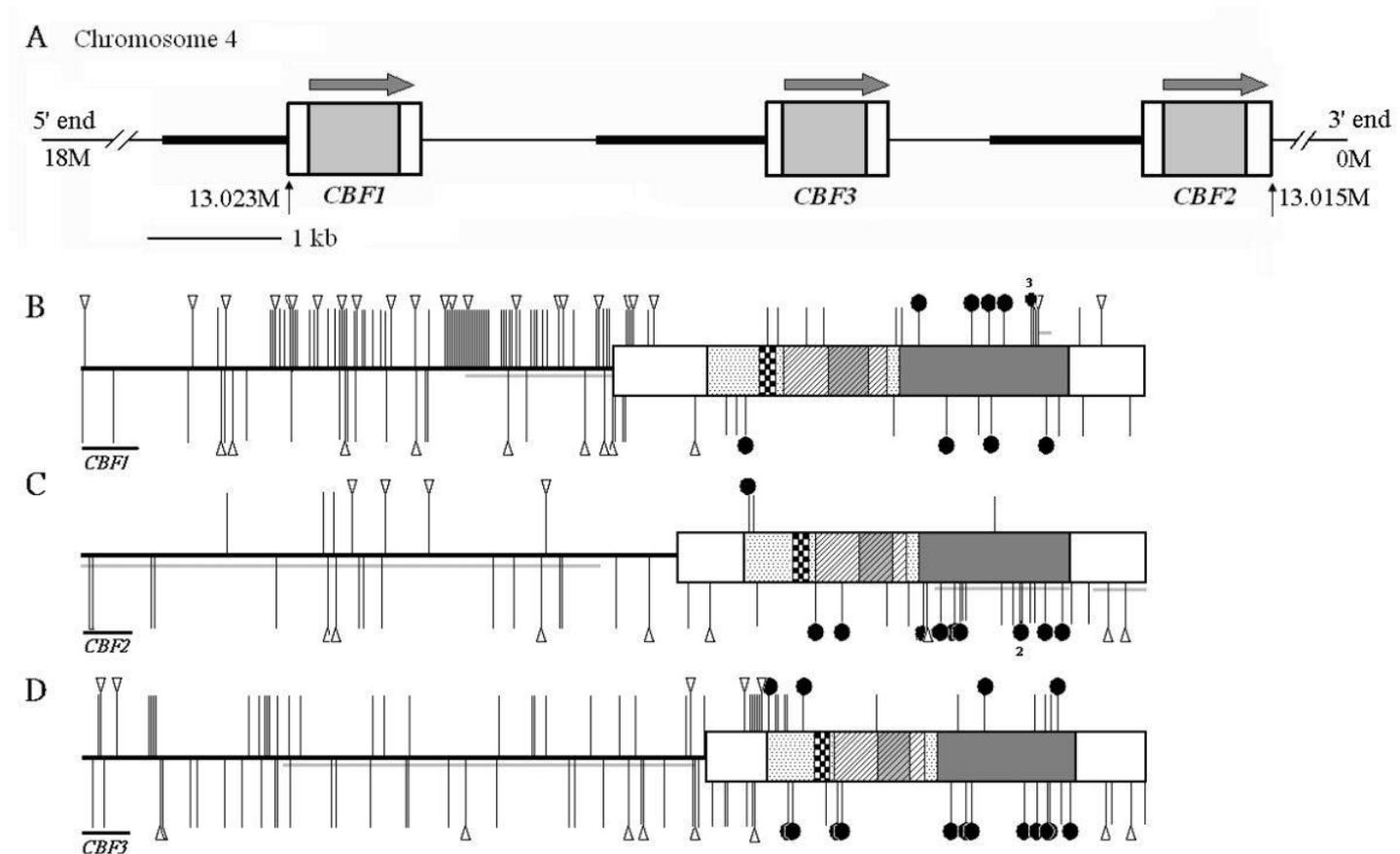
Genomic map and orientation of *CBFs* (A), and gene structures and polymorphism illustrations of *CBF1* (B), *CBF2* (C) and *CBF3* (D).

24

CBF variation in European *Arabidopsis* populations
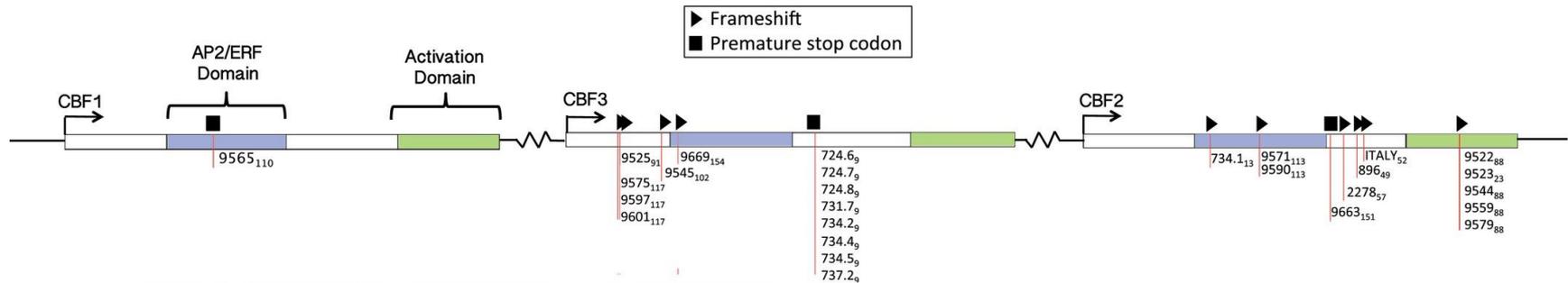
Locations of frameshift and premature stop codon mutations in *CBF1, CBF2* and *CBF3*.

Map showing locations of Eurasian *Arabidopsis* with complete *CBF* coding sequences and those with *CBF* alleles containing frameshift or premature stop codon mutations.

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 Kang[1], Huiting Zhang[1], Tianshu Sun[1], Yihao Shi[1], Jianqiao Wang[1], Baocai Zhang[2], Zhiheng Wang[3,4], Yihua Zhou[2] and Hongya Gu[1,5]

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

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

Summary

- 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 F_2 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.
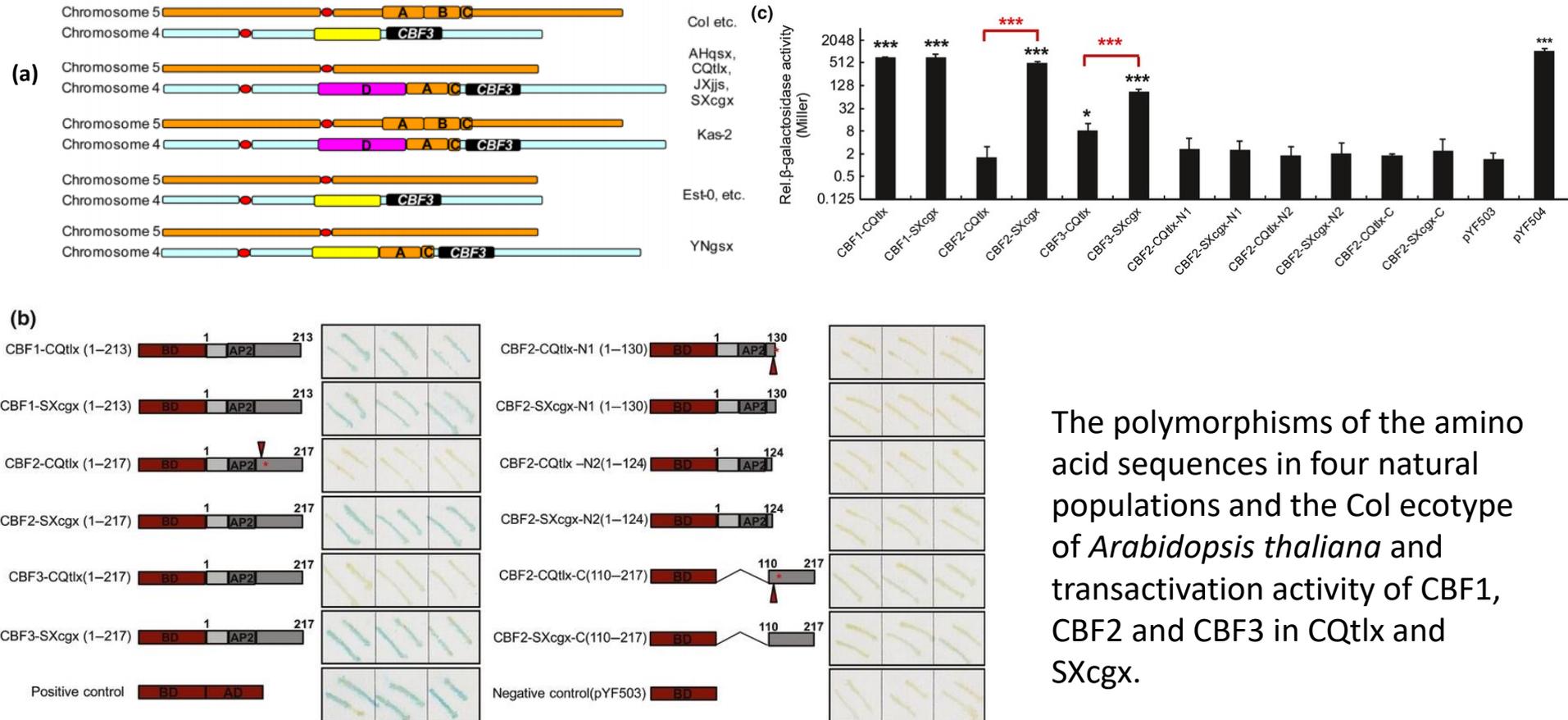
26

Kang *et al.*, New Phytologist, 2013

CBF variation in Yangtze River *Arabidopsis* populations

QTL- Mapping → *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.

The polymorphism of the amino acid sequences of CBF1, CBF2 and CBF3 among four natural populations and Col.

He *et al*., Chinese Science Bulletin , 2008
Kang *et al*., New Phytologist , 2013

CBF variation in Yangtze River *Arabidopsis* populations

The polymorphisms of the amino acid sequences in four natural populations and the Col ecotype of *Arabidopsis thaliana* and transactivation activity of CBF1, CBF2 and CBF3 in CQtlx and SXcgx.

Kang *et al.*, New Phytologist , 2013

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	2017-07-24
Name	AT4G25480.1
Tair Accession	Sequence:1005030824
GenBank Accession	NM_118680
Sequence Length (bp)	651

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 TTCGTCGGAG 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 TGAGATGTGT GATGCGACGA CGGATCATGG CTTCGACATG
451 GAGGAGACGT TGGTGGAGGC TATTTACACG GCGGAACAGA GCGAAAATGC
501 GTTTTATATG CACGATGAGG CGATGTTTGA GATGCCGAGT TTGTTGGCTA
551 ATATGGCAGA AGGGATGCTT TTGCCGCTTC CGTCCGTACA GTGGAATCAT
601 AATCATGAAG TCGACGGCGA TGATGACGAC GTATCGTTAT GGAGTTATTA
651 A
```

```
  1 MNSFSAFSEM FGSDYESSVS SGGDYIPTLA SSCPKKPAGR KKFRETRHPI
 51 YRGVRRRNSG KWVCEVREPN KKTRIWLGTF QTAEMAARAH 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

JGI — THE PLANT GENOMICS RESOURCE

| Species ▾ | Tools ▾ | Info ▾ | Download ▾ | Help ▾ | Cart | Subscribe |

Amaranthus hypochondriacus v1.0	Cucumis sativus v1.0	Panicum hallii var. hallii v2.1
Amaranthus hypochondriacus v2.1	Daucus carota v2.0	Panicum virgatum v1.1
Amborella trichopoda v1.0	Dunaliella salina v1.0	Panicum virgatum v4.1
Anacardium occidentale v0.9	Eucalyptus grandis v2.0	Phaseolus vulgaris v2.1
Ananas comosus v3	Eutrema salsugineum v1.0	Physcomitrella patens v3.3
Aquilegia coerulea v3.1	Fragaria vesca v1.1	Populus deltoides WV94 v2.1
Arabidopsis halleri v1.1	Glycine max Wm82.a2.v1	Populus trichocarpa v3.0
Arabidopsis lyrata v2.1	Gossypium hirsutum v1.1	Populus trichocarpa v3.1
Arabidopsis thaliana Araport11	Gossypium raimondii v2.1	Porphyra umbilicalis v1.5
Arabidopsis thaliana TAIR10	Helianthus annuus r1.2	Prunus persica v2.1
Asparagus officinalis V1.1	Hordeum vulgare r1	Ricinus communis v0.1
Boechera stricta v1.2	Kalanchoe fedtschenkoi v1.1	Salix purpurea v1.0
Botryococcus braunii v2.1	Kalanchoe laxiflora v1.1	Selaginella moellendorffii v1.0
Brachypodium distachyon Bd21-3 v1.1	Lactuca sativa V8	Setaria italica v2.2
Brachypodium distachyon v3.1	Linum usitatissimum v1.0	Setaria viridis v1.1
Brachypodium hybridum v1.1	Malus domestica v1.0	Setaria viridis v2.1
Brachypodium stacei v1.1	Manihot esculenta v6.1	Solanum lycopersicum iTAG2.4
Brachypodium sylvaticum v1.1	Marchantia polymorpha v3.1	Solanum tuberosum v4.03

Use phytozome or Gramene for other plants

- http://gramene.org/

How to use phytozome?

- Search a gene

How to use phytozome?

- Analyze your gene: gene structure and alternative splicing
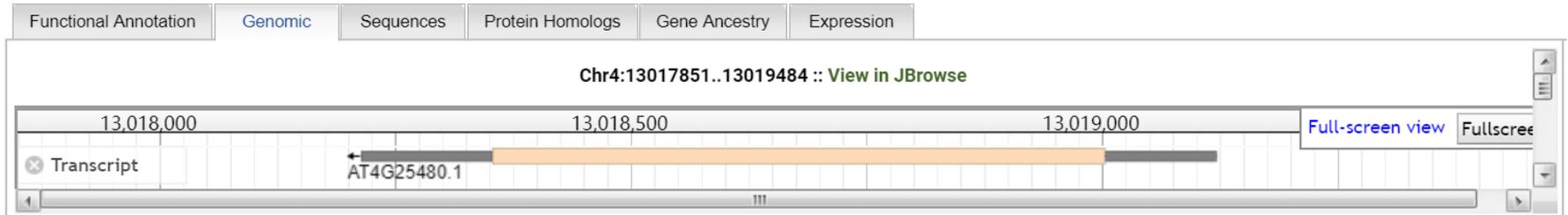
Genomic tab of *CBF3*
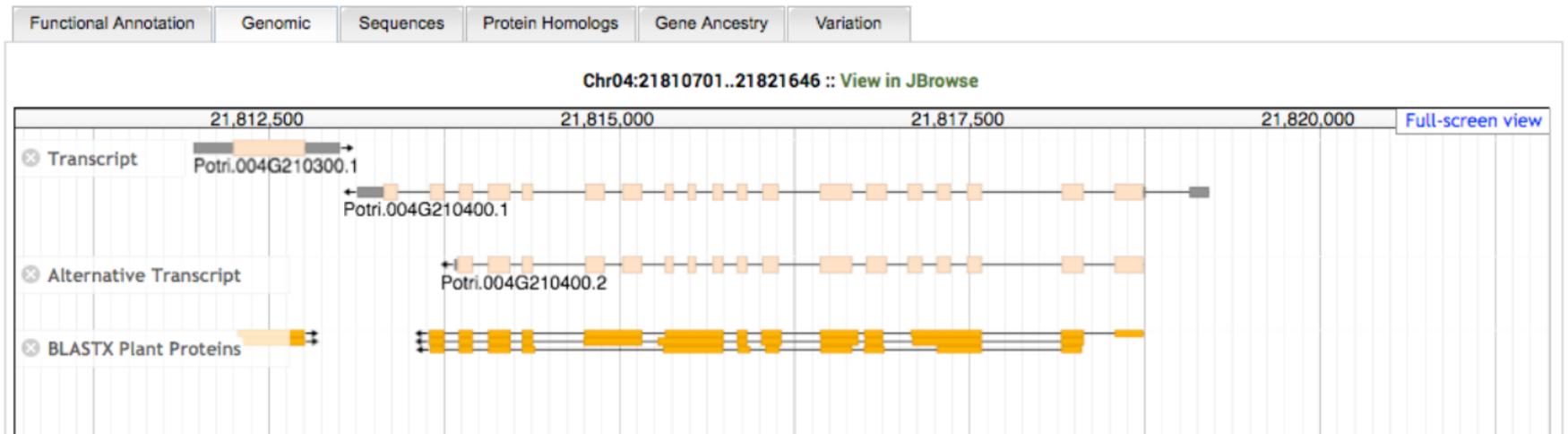
Fig. 9 Genomic tab in Gene View

How to use phytozome?

- Analyze your gene: get sequence

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.

How to use phytozome?

- Analyze your gene: protein homologs and evolution history

Protein Homologs tab of Gene View

Gene Ancestry view

How to use phytozome?

- Analyze your gene: protein homologs and evolution history

How to use phytozome?

- Analyze your gene: expression information

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

Use phytozome to do BLAST

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

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

Phylogenetic tree of *CBF3* transcription factors
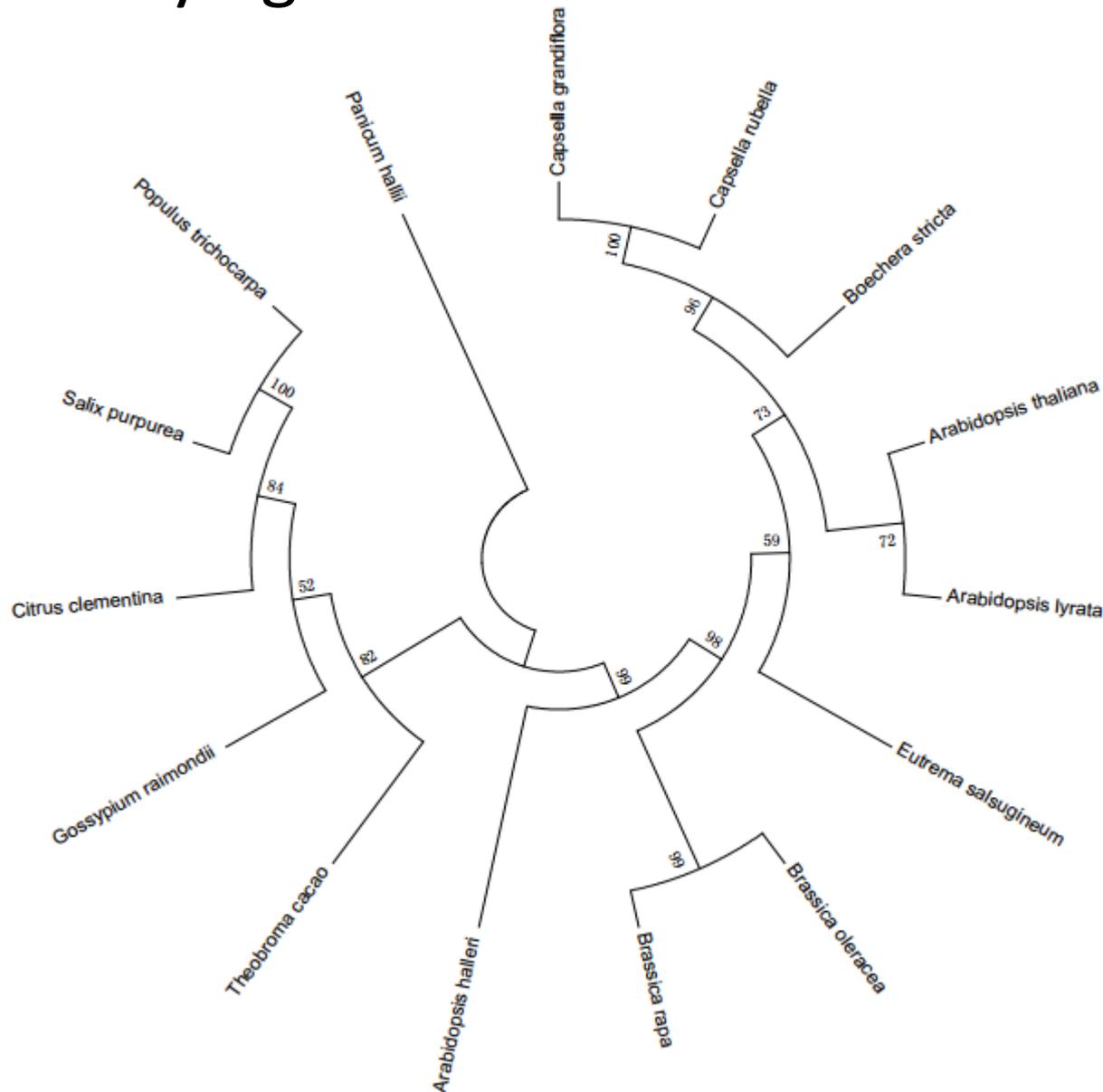
Use MEGA to build a phylogenetic tree of *CBF3* transcription factors (Neighbour-Joining)

A more convenient tool?
GCORN Plant

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

Gcorn plant
Correlation network of plant gene homology

★ Gene search

1. Keywords   AT4G25480
2. Species   Arabidopsis thaliana
3. Submit

Query: AT4G25480

Species: *Arabidopsis thaliana*

Hits: 1

★ Gene

Database	Accession	Species	Locus ID	CDD	Gene name
G	NP_567720	*Arabidopsis thaliana*	AT4G25480	237985	DREB1A; ATCBF3; C-REPEAT BINDING FACTOR 3

Plant Physiology®

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)

MNSFSAFSEMFGSDYESSVSSGGDYIPTLASSCPKKPAGRKKFRETRHP<mark>IYRGVRRRNSGKWVCEV</mark>
<mark>REPNKKTRIWLGTFQTAEMAARAHDVAALALRGRSACLNFA</mark>DSAWRLRIPESTCAKDIQKAAAEAA
LAFQDEMCDATTDHGFDMEETLVEAIYTAEQSENAFYMHDEAMFEMPSLLANMAEGMLLPLPSVQW
NHNHEVDGDDDDVSLWSY

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.

ruler: 1   100   200   300   400   500   600   700   800   900   1000

CBF3   AP2_^   (216 aa)

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

>CBF3
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
a	0.1026	0	0.0439	0
f	0.1633	0	0.0685	0
s	0.1187	0	0.0497	0
e	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
y	0.2686	0	0.0659	0
e	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
y	0.1924	0	0.0627	0
i	0.1332	0	0.0582	0
p	0.1247	0	0.0659	0
t	0.1137	0	0.0568	0
l	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
c	0.1336	0	0.0783	0
p	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.

Prot-Param

http://www.expasy.org/tools/protparam.html

氨基酸数目　Number of amino acids: 216

相对分子质量　Molecular weight: 24236.12

理论pI值　Theoretical pI: 5.08

Amino acid composition: [CSV format]

氨基酸组成

Ala (A)	27	12.5%
Arg (R)	14	6.5%
Asn (N)	8	3.7%
Asp (D)	15	6.9%
Cys (C)	5	2.3%
Gln (Q)	5	2.3%
Glu (E)	18	8.3%
Gly (G)	11	5.1%
His (H)	6	2.8%
Ile (I)	6	2.8%
Leu (L)	14	6.5%
Lys (K)	9	4.2%
Met (M)	10	4.6%
Phe (F)	10	4.6%
Pro (P)	9	4.2%
Ser (S)	19	8.8%
Thr (T)	10	4.6%
Trp (W)	5	2.3%
Tyr (Y)	6	2.8%
Val (V)	9	4.2%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%
(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

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

Atomic composition: 原子组成

Carbon	C	1055
Hydrogen	H	1615
Nitrogen	N	297
Oxygen	O	331
Sulfur	S	15

Formula: $C_{1055}H_{1615}N_{297}O_{331}S_{15}$　分子式
Total number of atoms: 3313　总原子数

Extinction coefficients: 消光系数

Extinction coefficients are in units of $M^{-1} cm^{-1}$, at 280 nm measured in water.

Ext. coefficient   36690
Abs 0.1% (=1 g/l)   1.514, assuming all pairs of Cys residues form cystines

Ext. coefficient   36440
Abs 0.1% (=1 g/l)   1.504, assuming all Cys residues are reduced

Estimated half-life: 半衰期

The N-terminal of the sequence considered is M (Met).

The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).
                          >20 hours (yeast, in vivo).
                          >10 hours (Escherichia coli, in vivo).

Instability index: 不稳定系数

The instability index (II) is computed to be 53.99
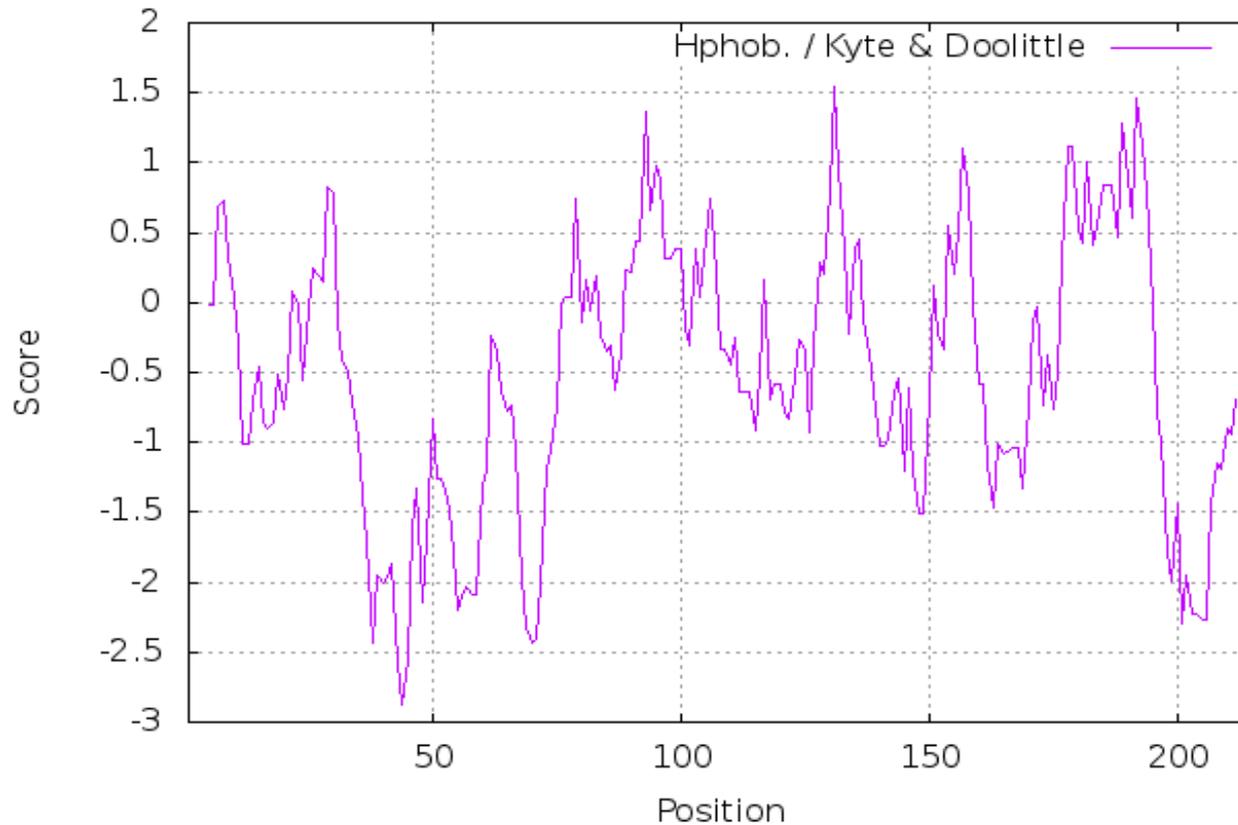This classifies the protein as unstable.

Aliphatic index: 60.69　脂肪族系数

Grand average of hydropathicity (GRAVY): -0.488
总平均亲水性

46

ProtScale

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

Analysis of hydrophilicity/hydrophobicity of CBF3

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

```
         10        20        30        40        50        60        70
          |         |         |         |         |         |         |
MNSFSAFSEMFGSDYESSVSSGGDYIPTLASSCPKKPAGRKKFRETRHPIYRGVRRRNSGKWVCEVREPN
ccccccccccccccccccccccccccceeecccccccccccccccccccccceeeeecccccccceeeeeccttt
KKTRIWLGTFQTAEMAARAHDVAALALRGRSACLNFADSAWRLRIPESTCAKDIQKAAAEAALAFQDEMC
cccceeeecccccchhhhhhhhhhhhhhhhttcccceeecccccccccccccccccchhhhhhhhhhhhhhhhccc
DATTDHGFDMEETLVEAIYTAEQSENAFYMHDEAMFEMPSLLANMAEGMLLPLPSVQWNHNHEVDGDDDD
cccccccccchhhhhhhhhhhhcchhhheehhhhhhhcchhhhhhhhhhhhhccccccccccccccccccccccc
VSLWSY
heeeee
```

α-helix, H
β-turn, T
random coil, C
extended strand, E


```
SOPMA :
   Alpha helix      (Hh) :    64 is   29.63%
   3₁₀  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  (?)  :     0 is    0.00%
   Other states          :     0 is    0.00%
```
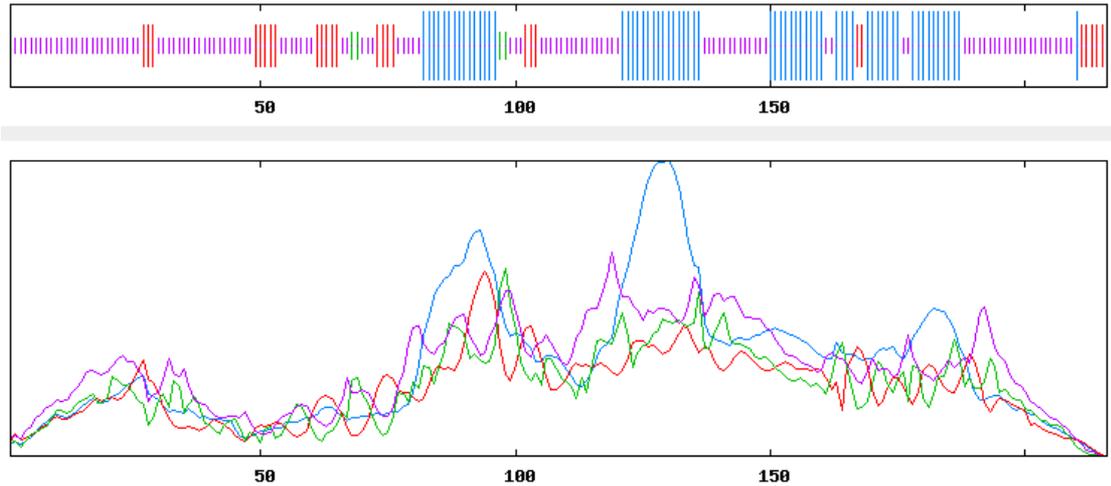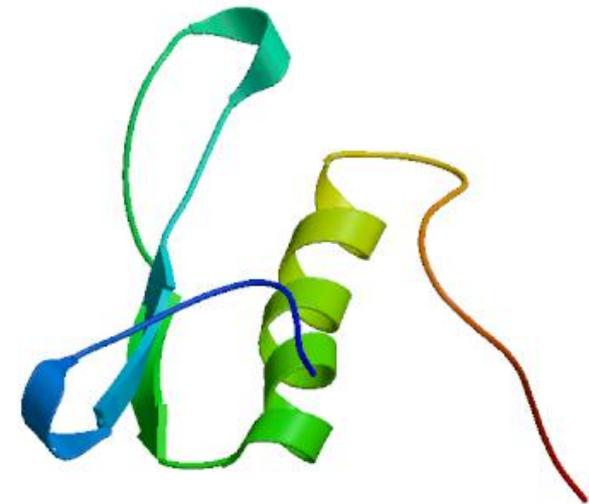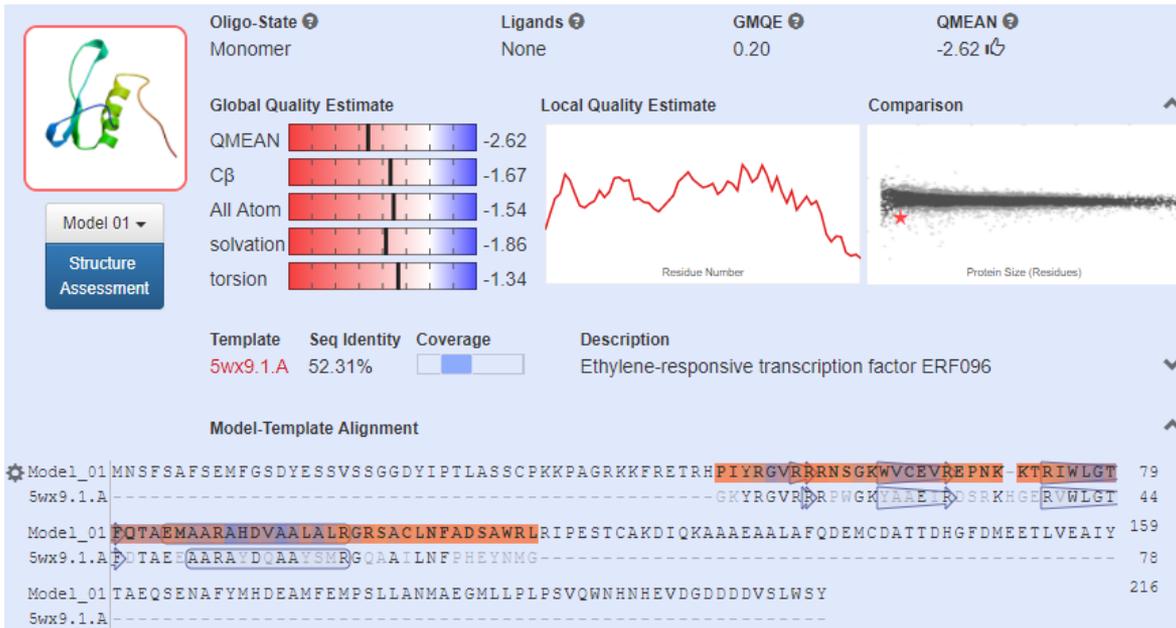
Prediction of secondary structure of CBF3 protein

48

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

Willems P *et al*., Plant Journal, 2019

PSORT II Prediction

https://psort.hgc.jp/

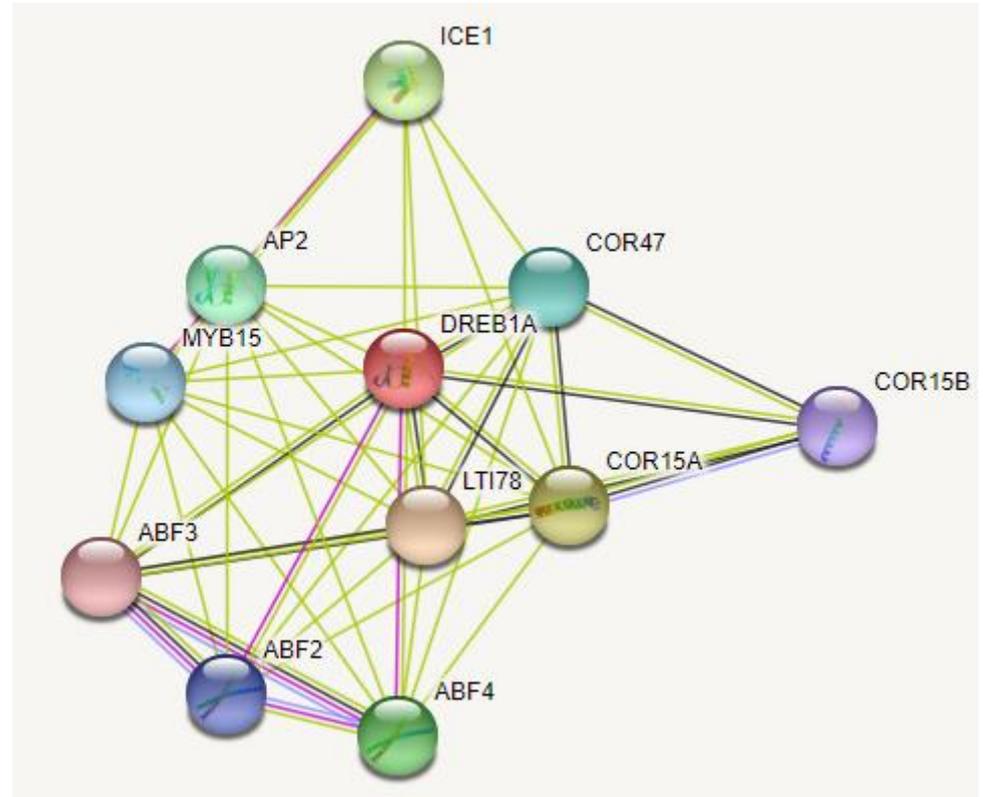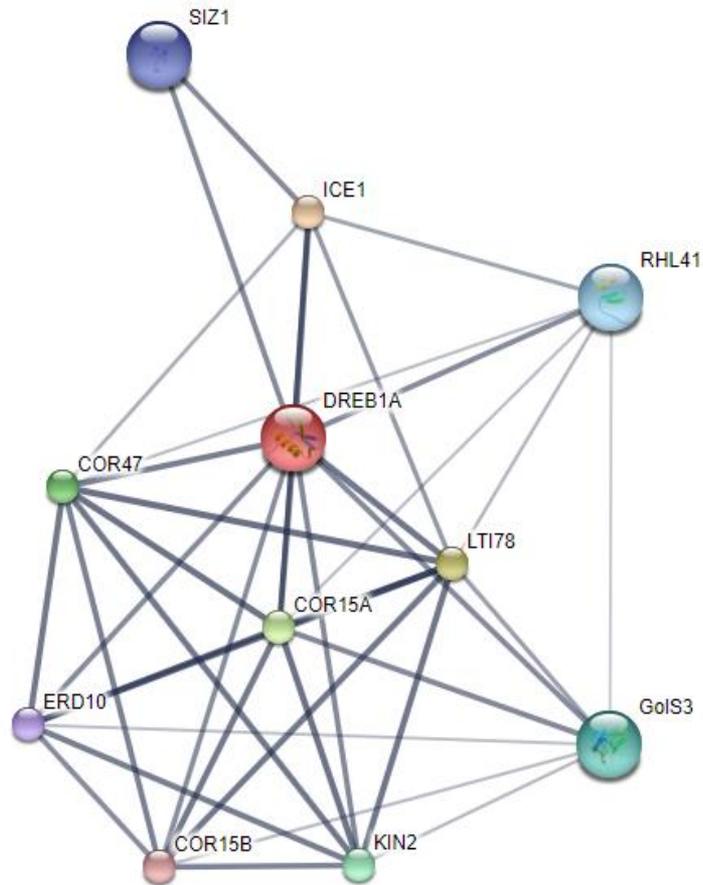## Results of the *k*-NN Prediction

$k = 9/23$

```
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