

植物转录因子分类、预测和数据库构建

Classification, Identification and Database Construction of Plant Transcription Factors

2025年5月14日

Linux生物信息技术基础课

罗静初

北京大学生物信息中心

luojc@pku.edu.cn

<https://abc.gao-lab.org/leb25/>



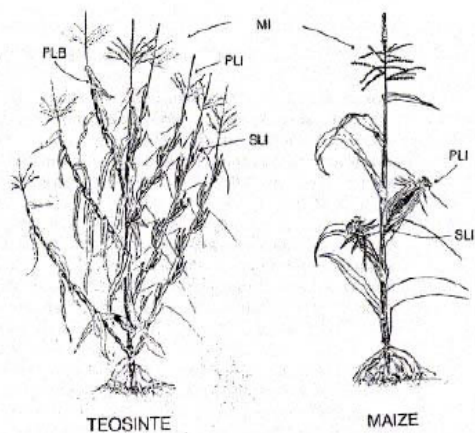
报告提纲

- 转录调控和转录因子
- 植物转录因子分类
- 基因组水平转录因子预测
- 植物转录因子数据库

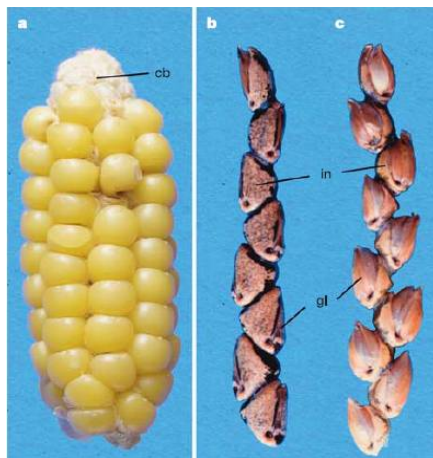
基因表达调控是决定生物多样性的的重要因素



转录因子与多种农作物性状有关



Doebley *et al.* (1997)
***Nature* 386:485**



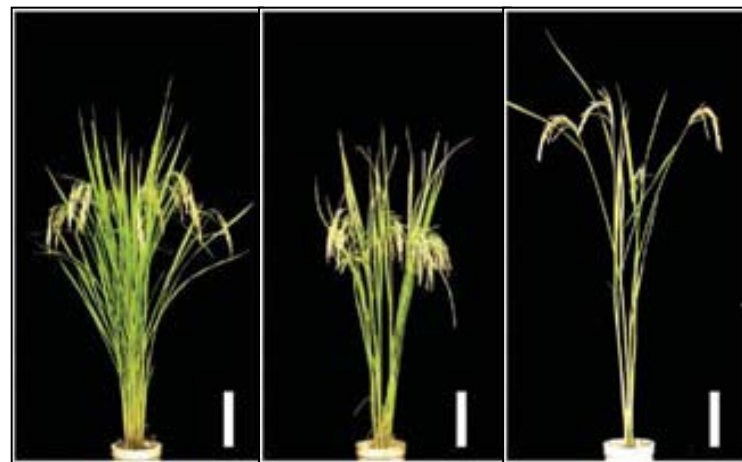
Wang *et al.* (2005)
***Nature*, 436:714**



Manning *et al* (2006) *Nature Genetics*, 38:948



Kinishi *et al* (2006) *Science*, 312:1392
Li *et al* (2006) *Science*, 312:1936

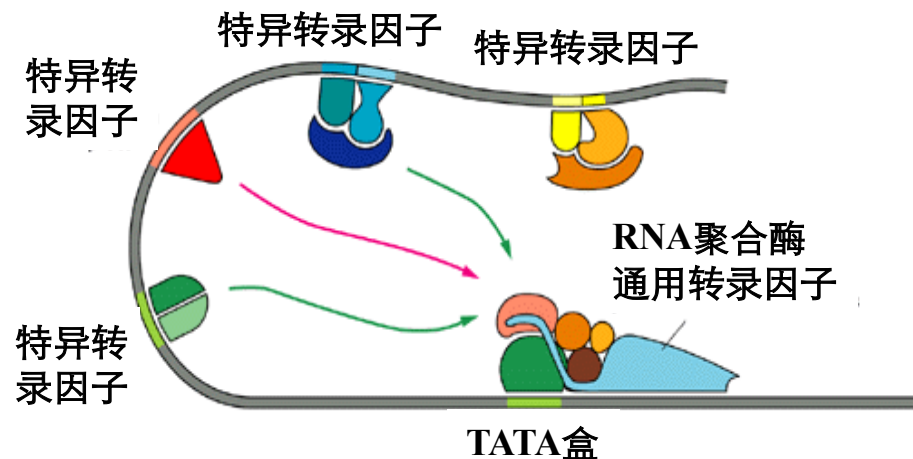


Jiao *et al* (2010) *Nature Genetics*, 42:541
Miura *et al* (2010) *Nature Genetics*, 42:545

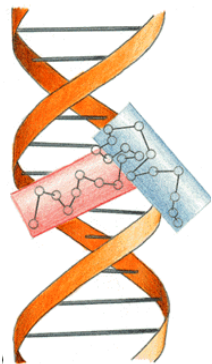
转录调控是基因表达主要调控机制之一



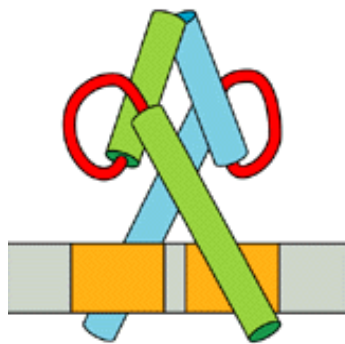
- 转录调控通过顺式作用元件和反式作用因子相互作用实现
- 反式作用因子通称转录因子
- 转录因子包括通用转录因子和特异转录因子
- 特异转录因子分为不同家族



螺旋-回折-螺旋
Helix-Turn-Helix



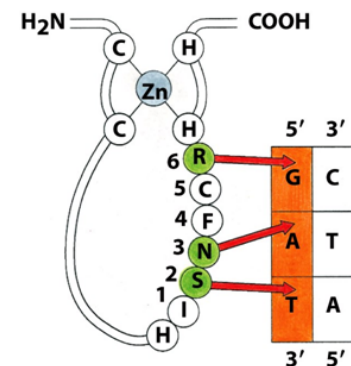
螺旋-回环-螺旋
Helix-Loop-Helix



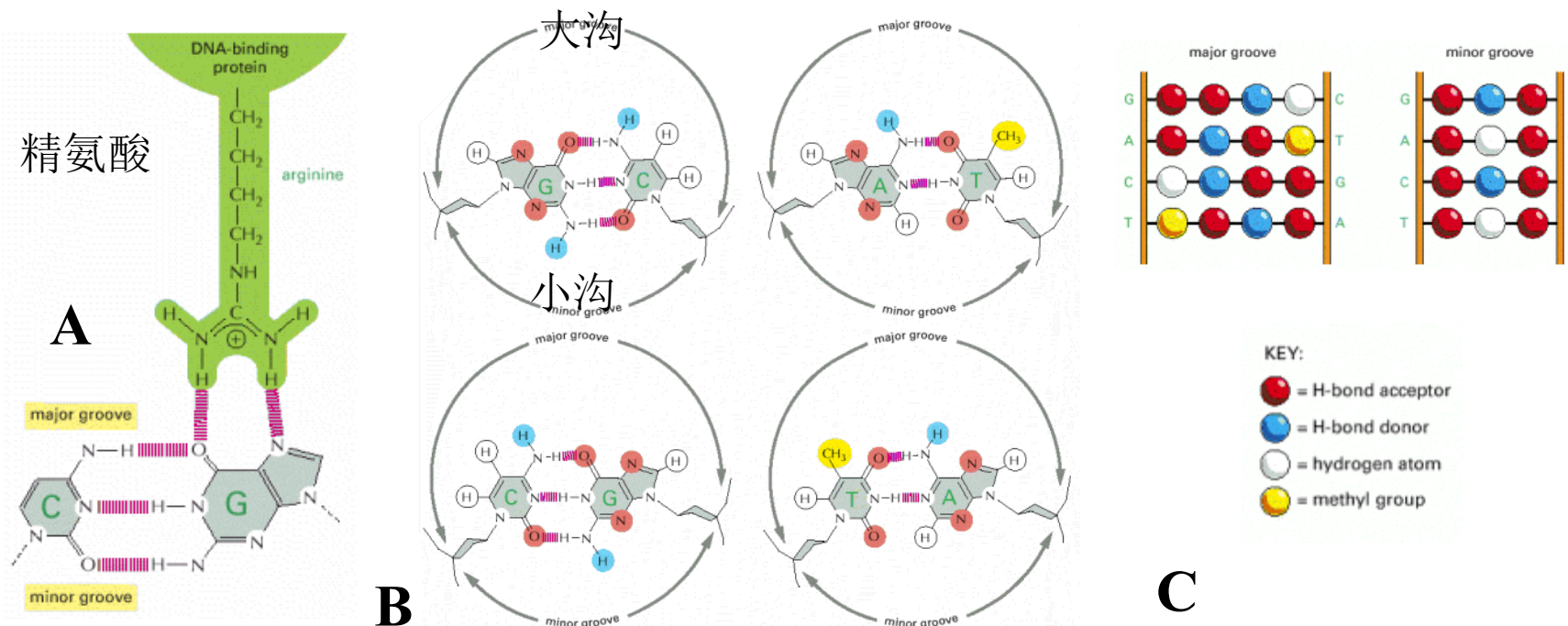
亮氨酸拉链
Leucine Zipper



锌指结构
Zinc Finger



转录因子识别位点与DNA顺式元件特异结合



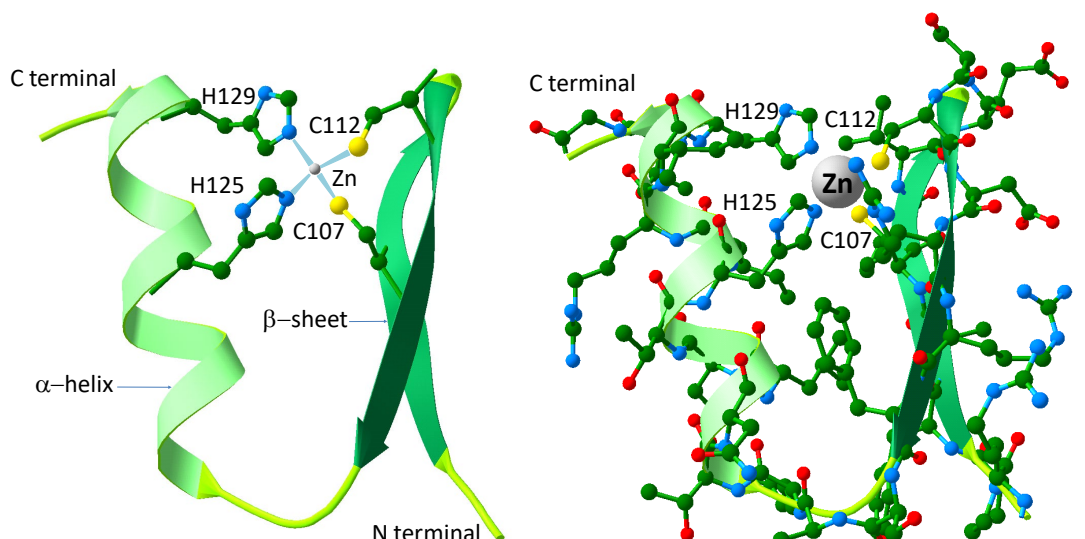
A. 精氨酸侧链胍基与鸟嘌呤通过氢键结合

B. DNA双螺旋中不同碱基配对方式具有不同识别模式

C. 不同序列模体具有不同结合位点

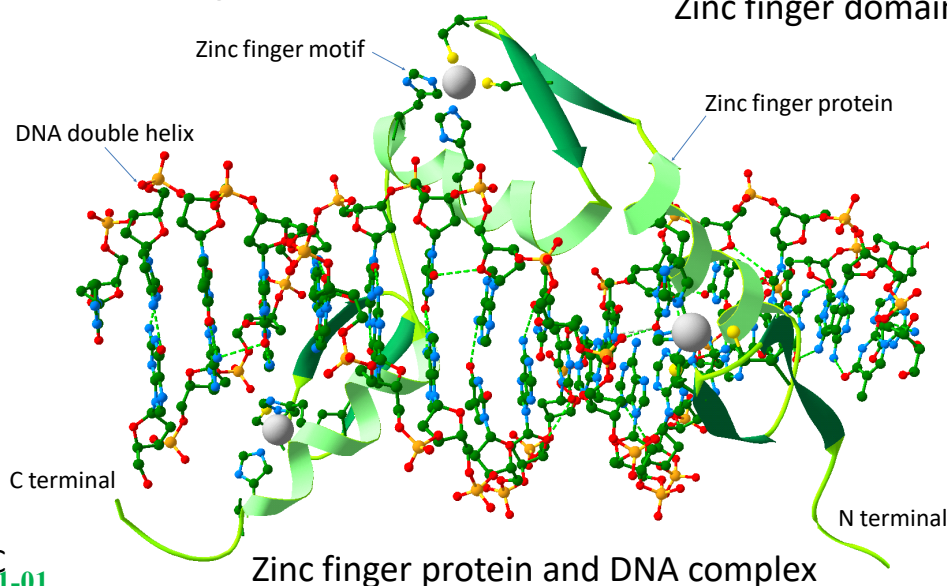
(Alberts *et al.* MBC, 2002)

小鼠C2H2锌指蛋白 转录因子



A Zinc finger motif

B Zinc finger domain



C
1-01

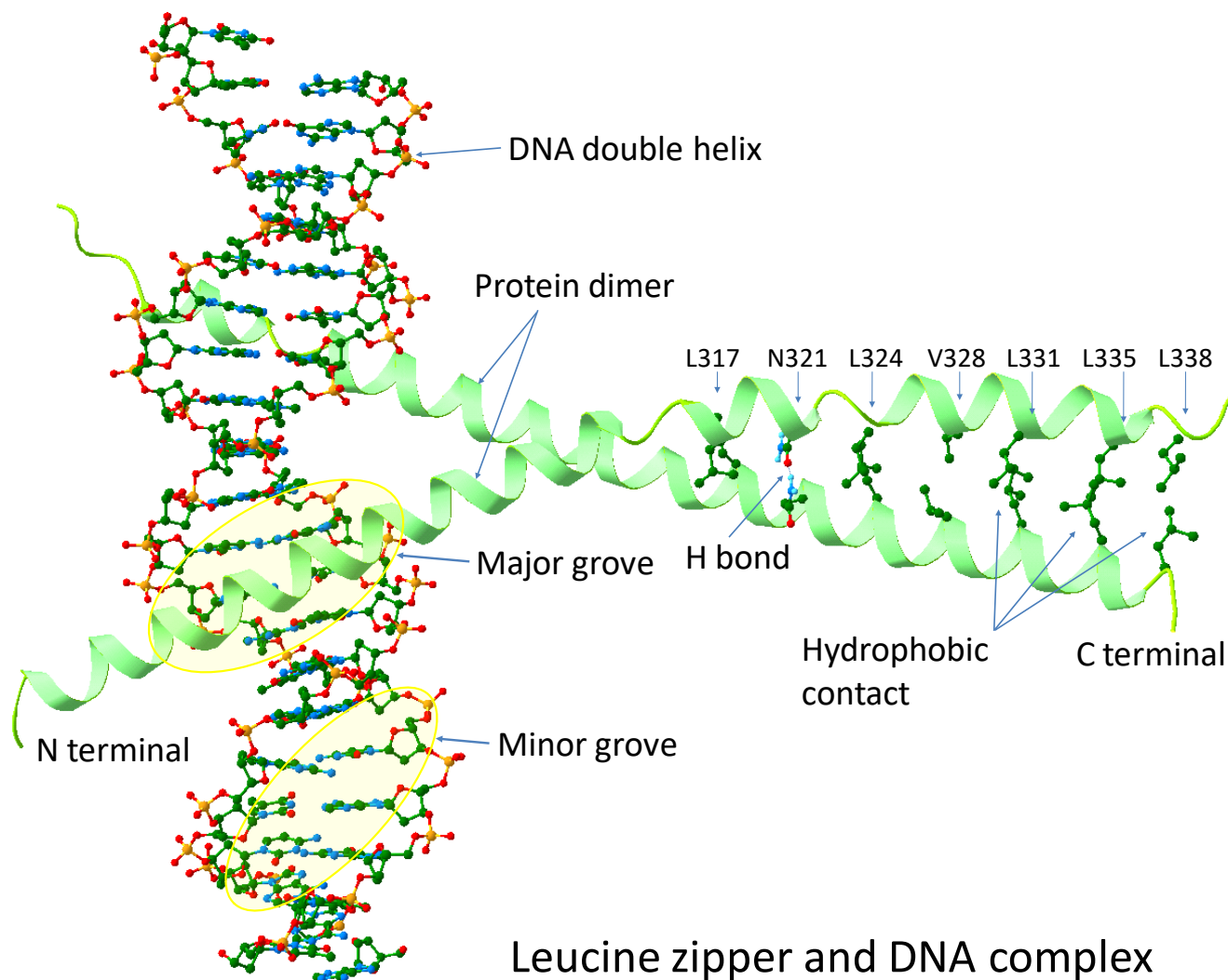
Zinc finger protein and DNA complex

A. 锌指结构模体，显示与锌原子结合的2个半胱氨酸和2个组氨酸侧链

B. 锌指结构模体，显示锌原子

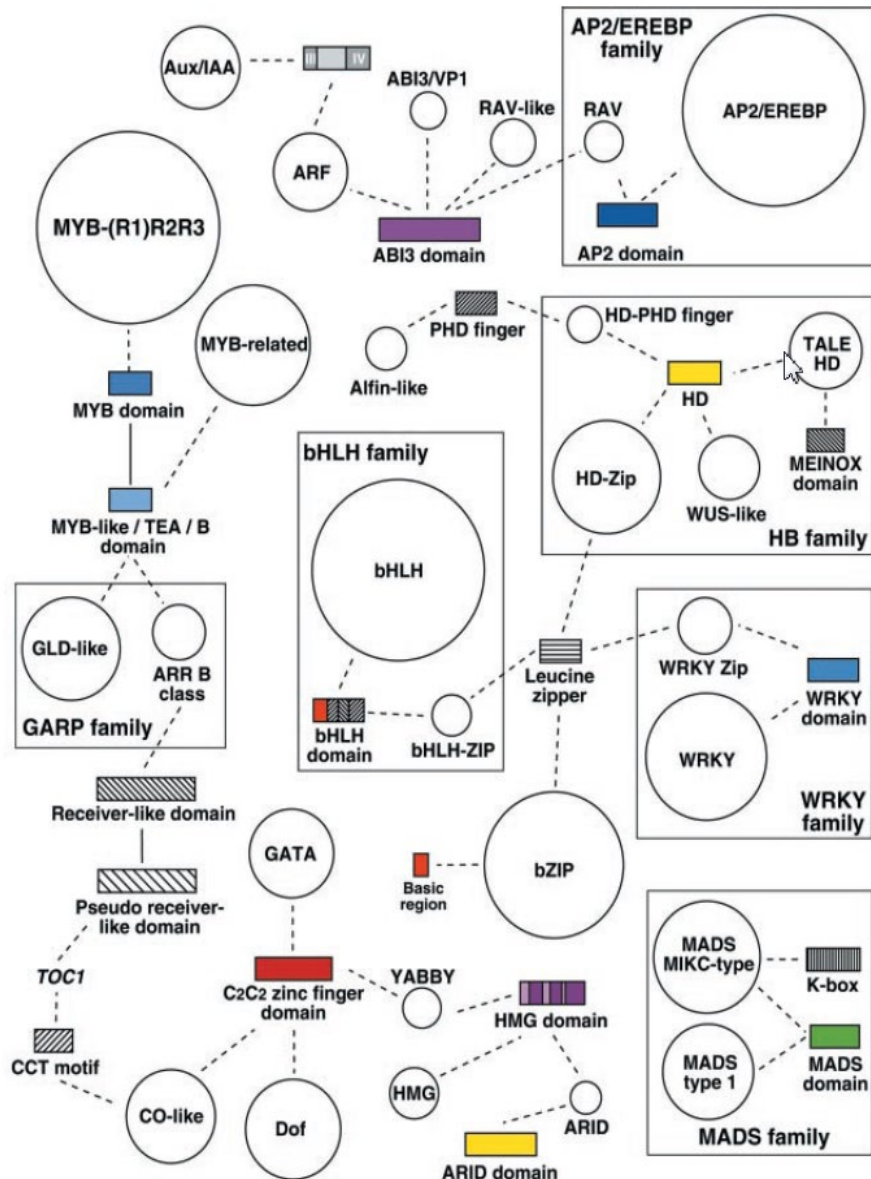
C. 锌指蛋白与DNA启动子结合复合物

亮氨酸拉链



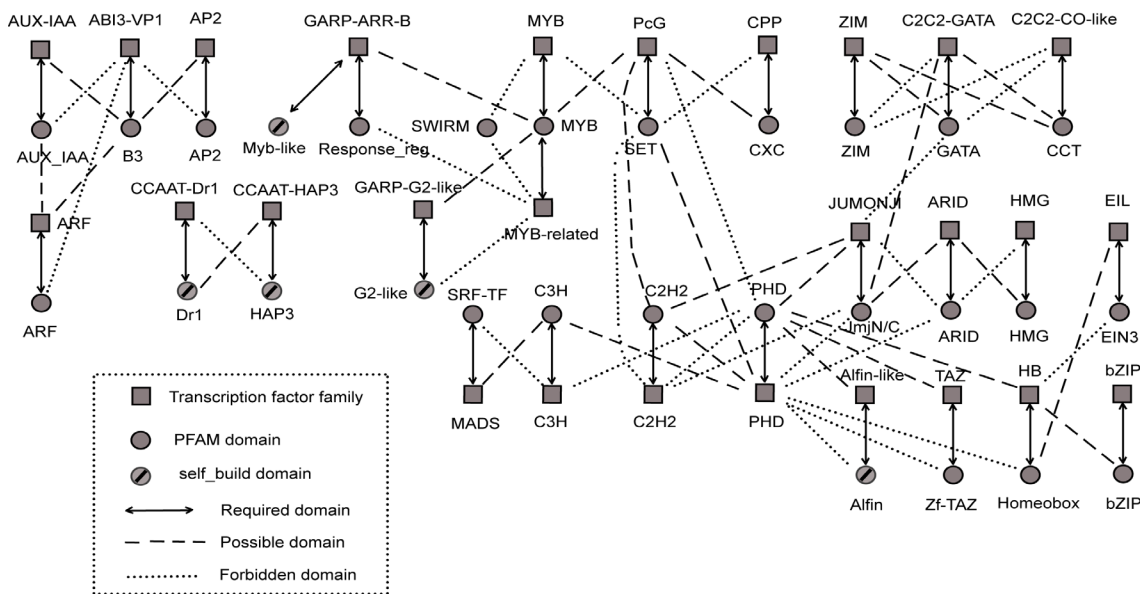
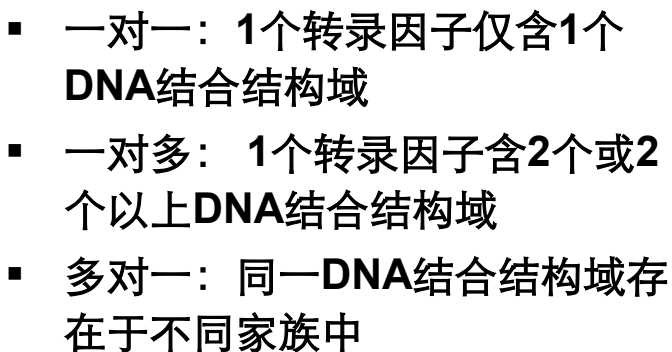
由两股**alpha**螺旋组成，螺旋**N**-端嵌入**DNA**大沟中，螺旋**C**-端亮氨酸成周期排列，通过疏水作用稳定结构。

拟南芥转录因子家族分类

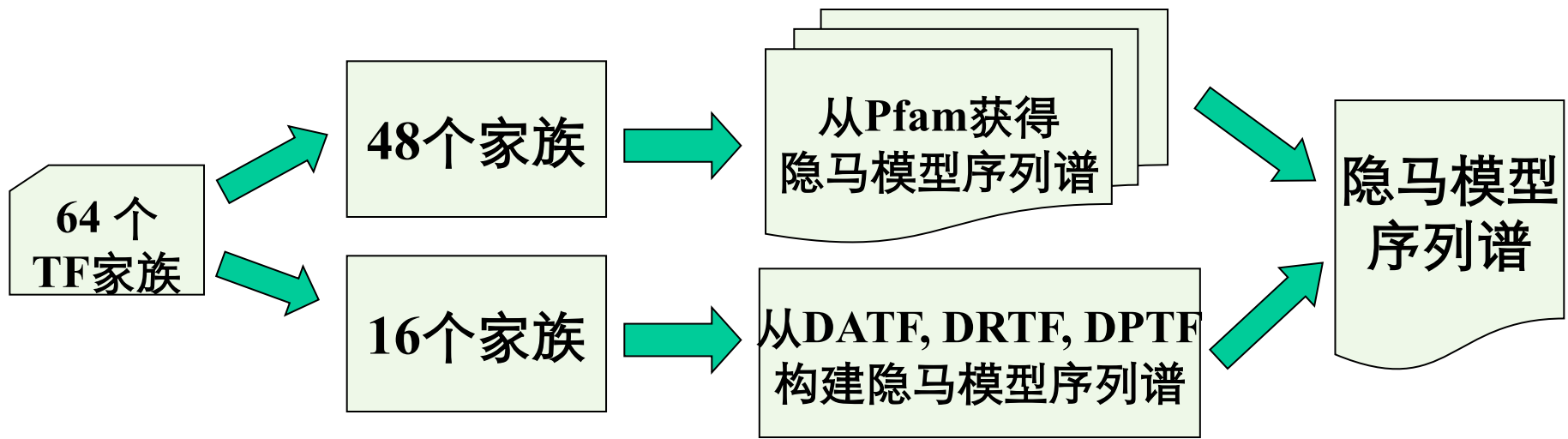


2000年，Riechmann等对拟南芥基因组转录因子进行分类，并和果蝇、线虫和酵母三个模式生物基因组转录因子进行比较。

Riechmann JL (2000) Science.



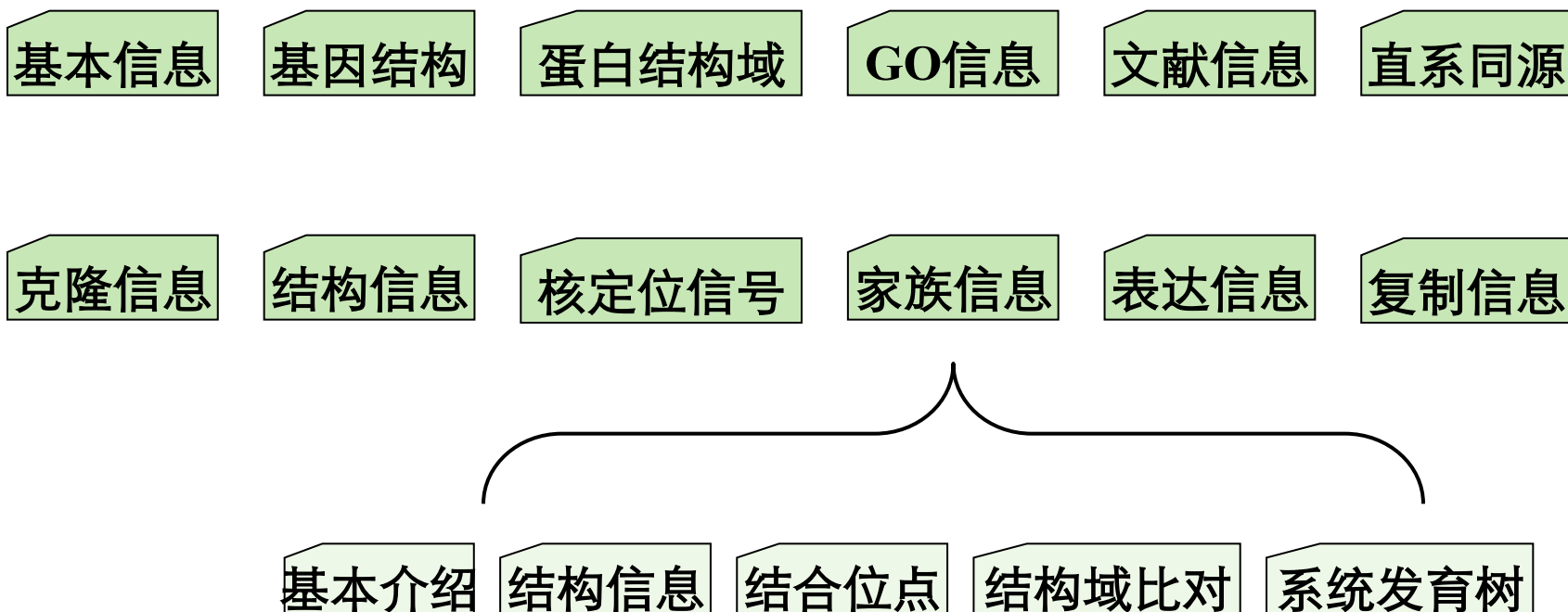
基因组水平植物转录因子预测

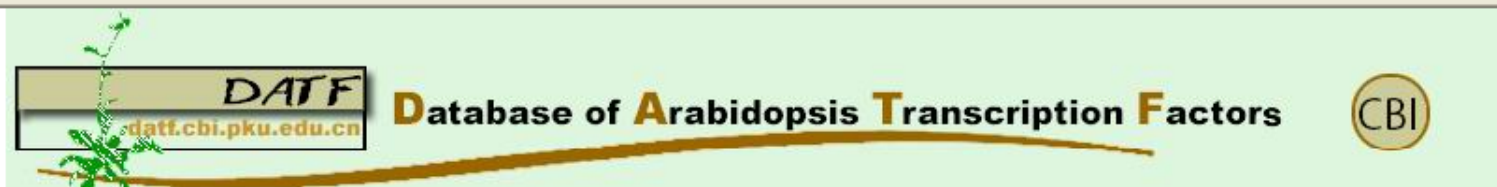


BLAST: 特异性低，阈值不能统一

HMMER: 特异性高，阈值易统一

植物转录因子注释





[Home](#) [Blast](#) [Search](#) [Download](#) [Help](#) ID (eg. AT1G01060) [Go](#)

[Basic Information](#) | [Clone Information](#) | [Gene Structure](#) | [Domain Structure](#) |
[3D Structure Hit](#) | [Features](#) | [GO Annotation](#) | [Cross Reference](#) | [Sequence](#) | [Details](#)

Basic Information [↑top](#)

Locus ID:	AT2G33810
Gene Model:	AT2G33810.1
Gene Family:	SBP
Gene Alias:	SPL3
Length:	Protein: 131aa CDS: 396bp Gene: 1072bp
Protein Properties:	MW: 15304 Da PI: 8.15
Description:	Symbol: SPL3 squamosa promoter-binding protein-like 3 (SPL3), identical to squamosa-promoter binding protein like 3 (Arabidopsis thaliana) GI:2462081; contains Pfam profile PF03110: SBP domain chr2:14312077-14313148 FORWARD Aliases: SQUAMOSA PROMOTER BINDING PROTEIN LIKE 3, T1B8.11, T1B8_11

Clone Information [↑top](#)

Cloned Or Not:	Cloned
Matched with reported cDNA	Match

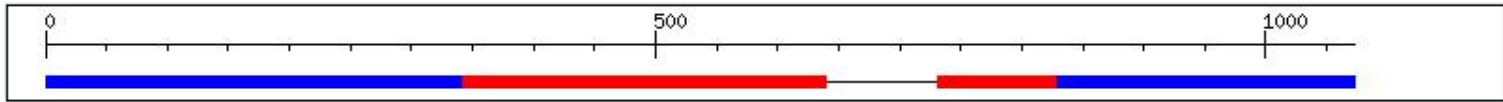


拟南芥基因组64个转录因子家族

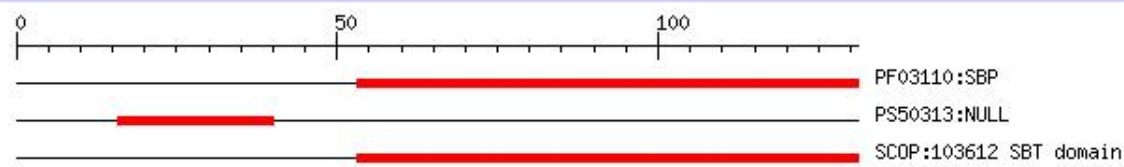
ABI3-VP1 (60)	Alfin (7)	AP2-EREBP (146)	ARF (23)
ARID (10)	AS2 (42)	AUX-IAA (29)	BBR-BPC (7)
BES1 (8)	bHLH (127)	bZIP (72)	C2C2-CO-like (37)
C2C2-DoF (36)	C2C2-GATA (26)	C2C2-YABBY (5)	C2H2 (134)
C3H (59)	CAMTA (6)	CCAAT-Dr1 (2)	CCAAT-HAP2 (10)
CCAAT-HAP3 (11)	CCAAT-HAP5 (13)	CPP (8)	E2F-DP (8)
EIL (6)	FHA (16)	GARP-ARR-B (10)	GARP-G2-like (43)
GeBP (21)	GIF (3)	GRAS (33)	GRF (9)
HB (87)	HMG (11)	HRT-like (2)	HSF (23)
JUMONJI (17)	LFY (1)	LIM (13)	LUG (2)
MADS (104)	MBF1 (3)	MYB (150)	MYB-related (49)
NAC (107)	Nin-like (14)	NZZ (1)	PcG (34)
PHD (56)	PLATZ (10)	S1Fa-like (3)	SAP (1)
SBP (16)	SRS (10)	TAZ (9)	TCP (23)
TLP (11)	Trihelix (26)	ULT (2)	VOZ (2)
Whirly (2)	WRKY (72)	ZF-HD (16)	ZIM (18)

Gene Structure ? ↑top

Gene Orientation: forward Gene Length: 1072
Gene Position on Chr, Start: 14312077 End: 14313148
Exon Position: 1-640 732-1072
Coding Region: 342-640 732-828

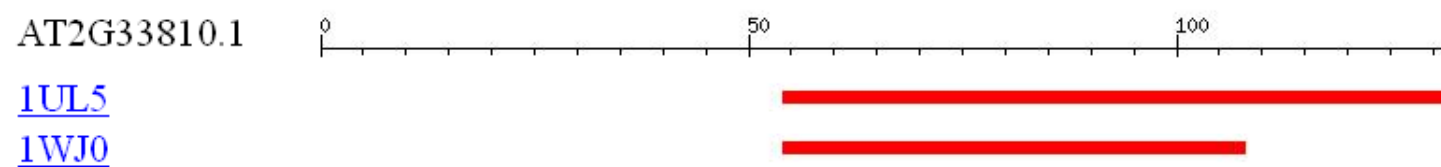


Domain Structure ? See Detail ↑top



3D Structure Hit ? ↑top

Mouse-over to show description, position and E-Value. Click to show alignments



Feature ? ↑top

UniGene: [At.25364](#)

- **Expression Information:** flower, leaf and root ; mixed floral buds, roots (9:1) ; mixture of silique and flower ; green siliques
- **Number of mRNA Sequence:** 7
- **Number of EST Sequence:** 16

Entry information - Microsoft Internet Explorer

文件(F) 编辑(E) 查看(V) 收藏(A) 工具(T) 帮助(H)

地址(D) http://datf.cbi.pku.edu.cn/entry-display.php?id=AT2G33810.1 转到

Feature ? ↑top

UniGene:
[At.25364](#)

- **Expression Information:** flower, leaf and root ; mixed floral buds, roots (9:1) ; mixture of silique and flower ; green siliques
- **Number of mRNA Sequence:** 7
- **Number of EST Sequence:** 16

Go Annotation ? ↑top

[GO:0003677](#)

- DNA binding

[GO:0003700](#)

- transcription factor activity

[GO:0005634](#)

- nucleus

[GO:0009908](#)

- flower development

[GO:0009911](#)

- positive regulation of flower development

Cross Reference ? ↑top

Genbank mRNA:

- **gi:** [30685845](#) **Accession Number:** [NM_128940.2](#)

Genbank Protein:

- **gi:** [18403387](#) **Accession Number:** [NP_565771.1](#)

Arabidopsis Database:

[TAIR](#) [TIGR](#) [MIPS](#) [SIGnal T-DNA Express](#)

PubMed:
[11118137](#)

Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddie J, Adam L,Pineda O, Ratcliffe OJ, Samaha RR, Creelman R, Pilgrim M, Broun P, Zhang JZ,Ghandehari D, Sherman BK, Yu G.
[Arabidopsis transcription factors: genome-wide comparative analysis amongeukaryotes.](#)
Science. 2000 Dec 15;290(5499):2105-10.

Lannenpaa M, Janonen I, Holtta-Vuori M, Gardemeister M, Porali I,

开始

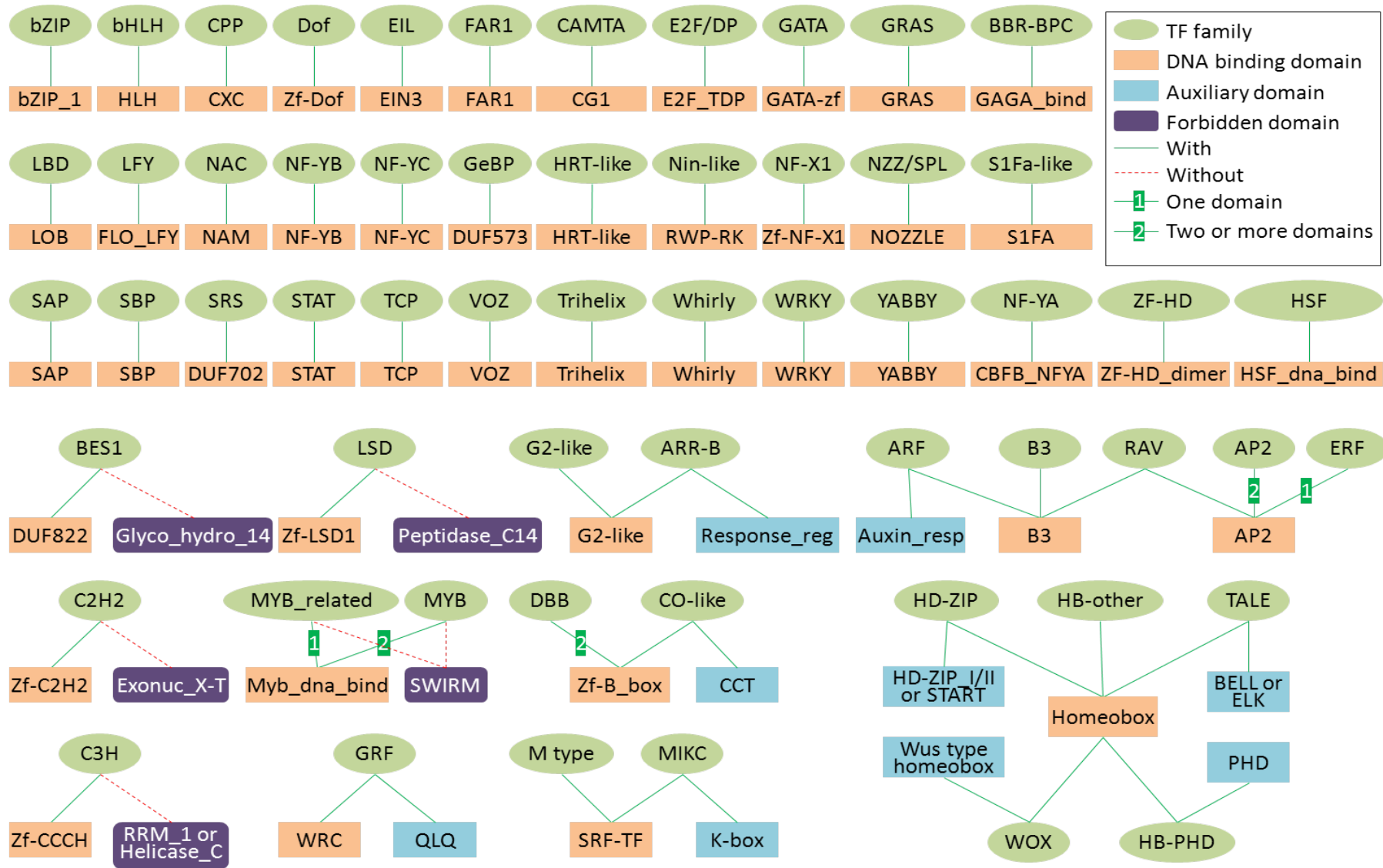
(4 unrea... 4 Inter... PTF Luo-JC_M... 1:26

植物转录因子数据库构建



- 植物转录因子数据库PlantTFDB构建
 - 5个模式植物：莱茵衣藻、江南卷柏、水稻、杨树、拟南芥
 - 17个经济作物：玉米、小麦、大豆、棉花、马铃薯、柑橘、松树、苜蓿等
- 转录因子注释
 - 家族水平：文献、多序列比对、系统发育树等
 - 基因水平：序列、结构、功能、文献等
- Web界面服务：浏览、检索、下载、相似性搜索

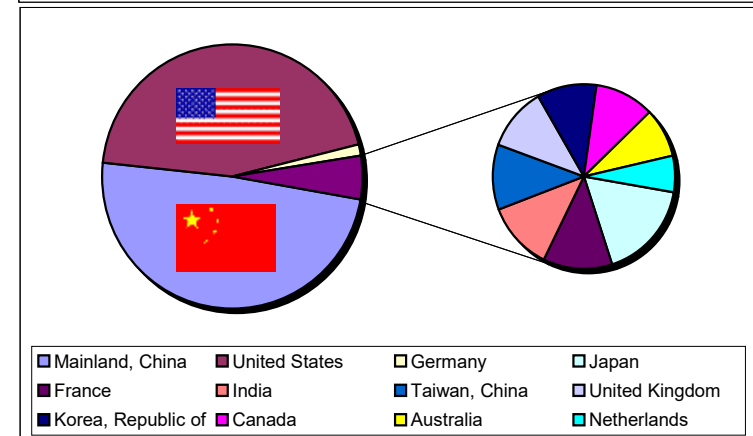
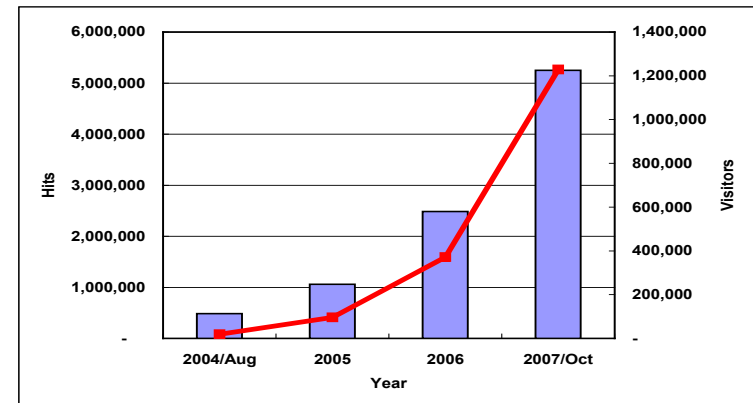
转录因子家族 (PlantTFDB Ver 2.0)



PlantTFDB 2 包括5种模式植物22种经济作物



类别	物种	学名	数量
模式生物	拟南芥	<i>Arabidopsis thaliana</i>	2290
	杨树	<i>Populus trichocarpa</i>	2576
	水稻	<i>Oryza sativa (indica)</i>	2025
		<i>Oryza sativa (japonica)</i>	2384
	小立碗藓	<i>Physcomitrella patens</i>	1170
	莱茵衣藻	<i>Chlamydomonas reinhardtii</i>	205
谷类	大麦	<i>Hordeum vulgare</i>	618
	玉米	<i>Zea mays</i>	764
	高粱	<i>Sorghum bicolor</i>	397
	甘蔗	<i>Saccharum officinarum</i>	1177
	小麦	<i>Triticum aestivum</i>	1127
水果	苹果	<i>Malus domestica</i>	1025
	葡萄	<i>Vitis vinifera</i>	867
	甜橙	<i>Citrus sinensis</i>	599
裸子植物	火炬松	<i>Pinus taeda</i>	950
	云杉	<i>Picea glauca</i>	440
经济作物	棉花	<i>Gossypium hirsutum</i>	1567
	马铃薯	<i>Solanum tuberosum</i>	1340
	大豆	<i>Glycine max</i>	1891
	向日葵	<i>Helianthus annuus</i>	513
	番茄	<i>Lycopersicon esculentum</i>	998
	百脉根	<i>Lotus japonicus</i>	457
	苜蓿	<i>Medicago truncatula</i>	1022



Guo et al, Bioinform 2005
Gao et al, Bioinform 2006
Zhu et al, Bioinform 2007
Guo et al, NAR, 2008

2010年—植物转录因子数据库第二版



Plant Transcription Factor Database

v2.0

Center for Bioinformatics, Peking University, China

[Previous version](#)

[Home](#) | [Blast](#) | [Search](#) | [Download](#) | [WebService](#) | [Help](#) | [About](#) | [Links](#)

(eg: SPL2)

Browse by Species

Arabidopsis lyrata	Arabidopsis thaliana	Arachis hypogaea
Artemisia annua	Brachypodium distachyon	Brassica napus
Brassica rapa	Carica papaya	Chlamydomonas reinhardtii
Chlorella sp. NC64A	Citrus sinensis	Coccomyxa sp. C-169
Cucumis sativus	Glycine max	Gossypium hirsutum
Helianthus annuus	Hordeum vulgare	Lotus japonicus
Malus x domestica	Manihot esculenta	Medicago truncatula
Micromonas pusilla CCMP1545	Micromonas sp. RCC299	Mimulus guttatus
Nicotiana tabacum	Oryza sativa subsp. indica	Oryza sativa subsp. japonica
Ostreococcus lucimarinus CCE9901	Ostreococcus sp. RCC809	Ostreococcus tauri
Panicum virgatum	Physcomitrella patens subsp. patens	Picea glauca
Picea sitchensis	Pinus taeda	Populus trichocarpa
Prunus persica	Raphanus sativus	Ricinus communis
Saccharum officinarum	Selaginella moellendorffii	Solanum lycopersicum
Solanum tuberosum	Sorghum bicolor	Theobroma cacao
Triticum aestivum	Vigna unguiculata	Vitis vinifera
Volvox carteri	Zea mays	

Browse by Family

AP2 (716)	ARF (646)	ARR-B (323)	B3 (1505)	BBR/BPC (218)	BES1 (247)
---------------------------	---------------------------	-----------------------------	---------------------------	-------------------------------	----------------------------



[Center for Bioinformatics](#), [Peking University](#)

Plant Transcription Factor Databases

Home
BLAST
Search
Download
Help
Links
About Us

Supported by



With the available genome sequence of model organisms, we predicted, annotated and analyzed all putative TFs at the genome scale for five plant species, Arabidopsis, Poplar, Oryza sativa, the green alga Chlamydomonas reinhardtii and the moss Physcomitrella patens. We are also working on the available EST sequences from 17 plant species including crops (maize, barley, wheat, etc), fruits (apple, orange, grape, etc), trees (pine, spruce, etc) and other economically important plants (cotton, potato, soybean, etc).

To provide comprehensive information for the putative TFs, we made extensive annotations at both family and gene levels. A brief introduction and key references are presented for each family. For each identified TF, functional domains, similar entries in various databases such as UniProt, RefSeq and TransFac are displayed as diagrams. Annotations with Gene Ontology, UniGene Expression are shown and putative TF orthologs among all these 22 species are listed. In addition, PlantTFDB has a simple and user-friendly interface to allow users to search by IDs or free texts, to make sequence similarity search using BLAST and to download all TF sequences.

Publication for more information and for citing this work:

[PlantTFDB: a comprehensive plant transcription factor database](#)

Species

[Arabidopsis](#)
[Rice](#)
[Poplar](#)
[Moss](#)
[Alga](#)

[Maize](#)
[Bread wheat](#)
[Barley](#)
[Sorghum](#)
[Sugarcane](#)
[Upland cotton](#)
[Soybean](#)
[Potato](#)
[Tomato](#)
[Apple](#)
[Sweet orange](#)
[Wine grape](#)
[Common sunflower](#)
[Barrel medicago](#)
[Lotus japonicus](#)
[Loblolly pine](#)
[White spruce](#)



www.cbi.pku.edu.cn

