

拟南芥CBF转录因子亚家族的 生物信息学分析以及其功能、 演化方面的预测

报 告 人 施逸豪

组 员 田博书 李莉 宋靖慧

指导老师 罗静初

概要

1 / 背景

2 / 研究目的

3 / 研究方法与结果

4 / 分析与讨论

5 / 小结

6 / 致谢

拟南芥的环境适应性

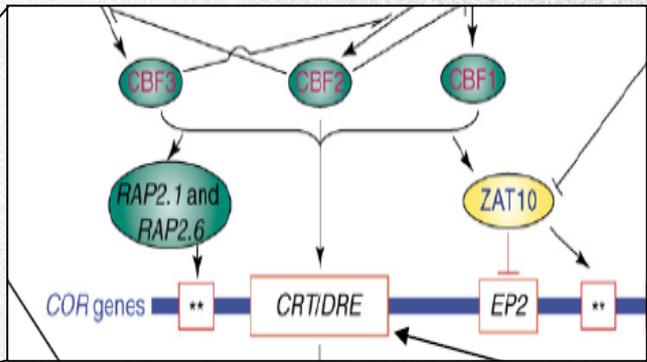
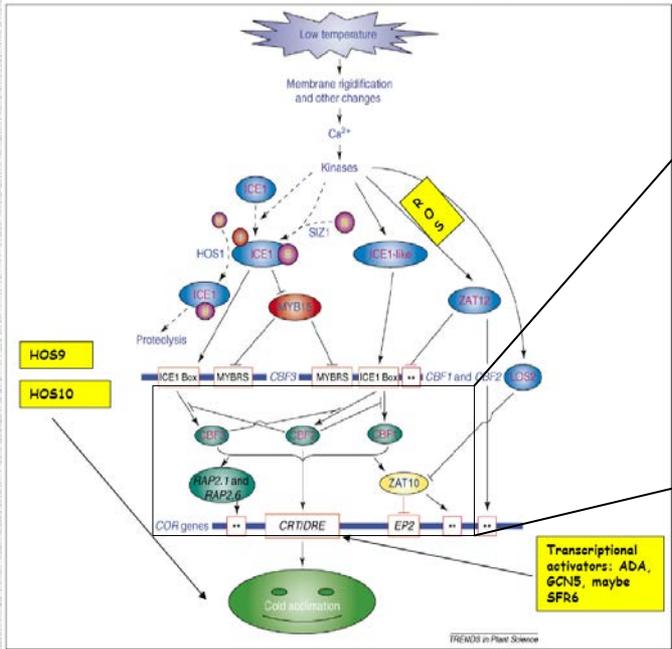
- 温度是影响植物分布的重要因素



野生拟南芥的地理分布(Koornneef et al., 2004)

CBF 亚家族与拟南芥对冷胁迫的响应

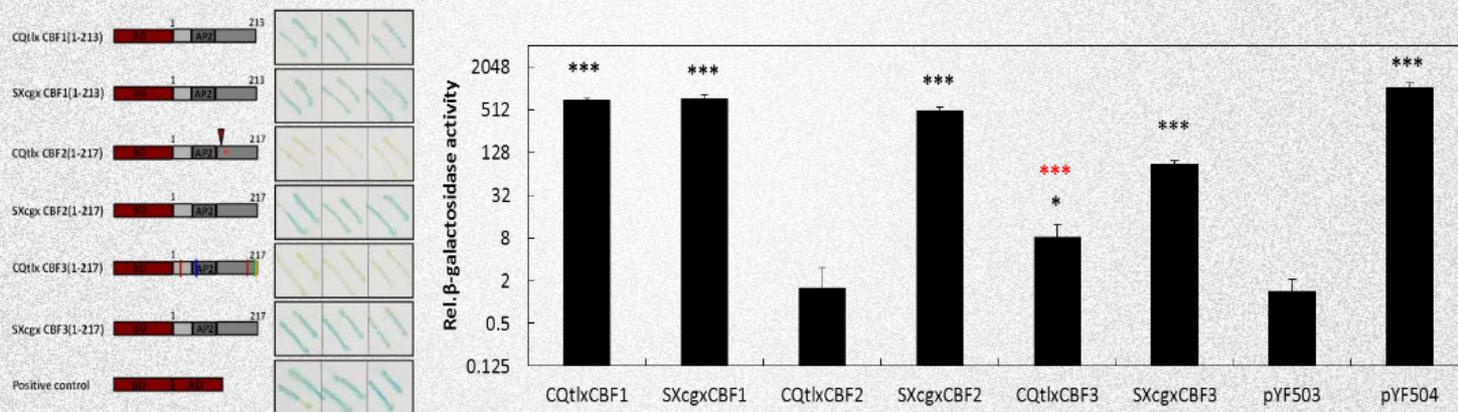
■ CBF 亚家族参与不依赖ABA途径的冷诱导调控网络



(Chinnusamy et al., 2007)

CBF的功能存在差异和分化

■ 中国居群的三个CBF转录激活活性存在适应性差异



国内两个典型居群（重庆铜梁、陕西城固）CBF转录因子的激活活性示意（Kang, 2013）

研究目的

- 能否从生物信息学的角度来预测3个*CBF* 基因的功能是否冗余
- 预测这三个基因对植物适应环境可能起到的作用以及它们在自然居群中的微演化模式

研究目的

- 能否从生物信息学的角度来预测3个*CBF* 基因的功能是否冗余
- 预测这三个基因对植物适应环境可能起到的作用以及它们在自然居群中的微演化模式

搜索 *CBF* 亚家族3个基因

■ 利用点阵图分析基因组重复序列 (DOTUP)

发现有3段序列彼此相似度都很高：

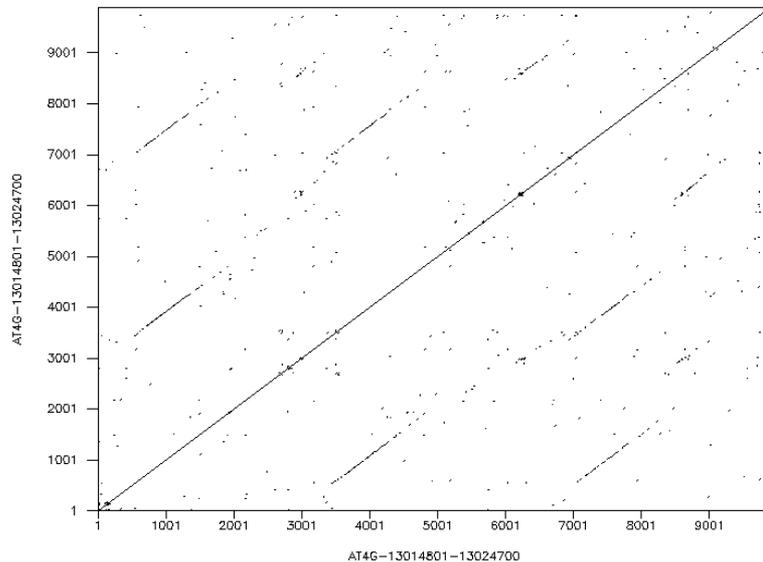
400-1400bp；

3500-4500bp；

7000-8000bp。

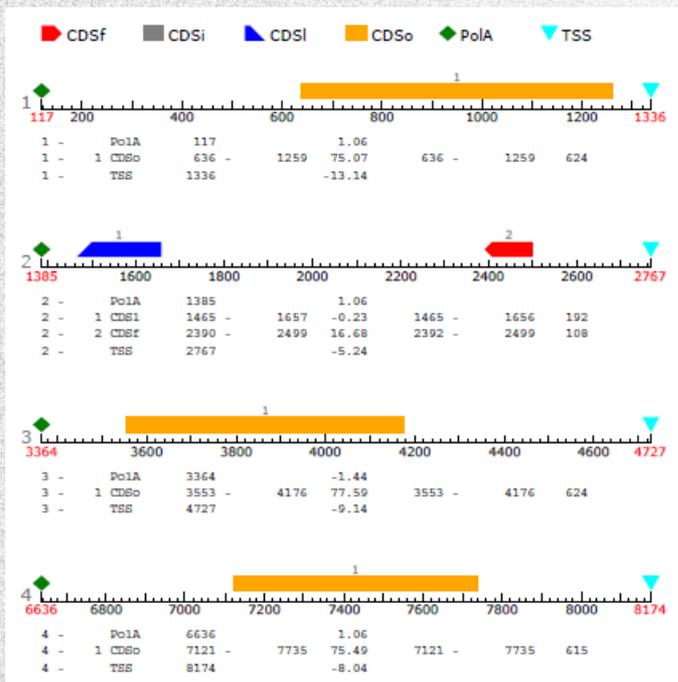
Dottup: fasta::727523:AT4G-13014801-13024700 vs fasta::7...

Sun 13 Apr 2014 10:45:51



研究方法与结果

■ 基因结构预测 (FGENESH)



共预测出4个基因，1、3、4相似度较高；为之前观测到的重复序列；

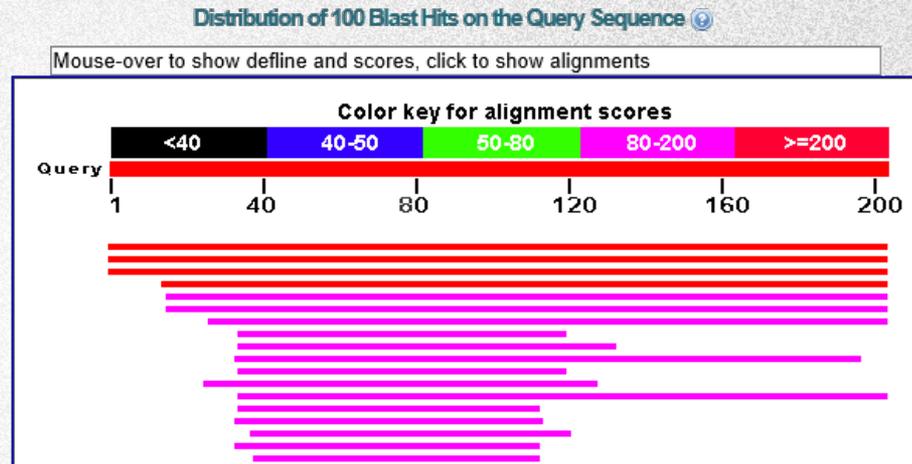
2经TAIR对比发现为假基因；

可以发现，这3个基因都以反向互补链为模板链，且都只有一个外显子，结构十分简单。

研究方法与结果

■ 搜索基因 (BLASTP)

可获得三个基因的注释信息；
Softberry预测的基因1是DREB1C/CBF2；
预测的3是DREB1A/CBF3；
预测的4是DREB1B/CBF1；
为3个与植物抗冻相关的转录因子；
在第四条染色体上串联重复排列。



Alignments Download GenPept Graphics Distance tree of results Multiple alignment

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	dehydration-responsive element-binding protein 1B [Arabidopsis thaliana]	432	432	100%	4e-155	100%	NP_567721.1
<input type="checkbox"/>	dehydration-responsive element-binding protein 1C [Arabidopsis thaliana]	373	373	100%	1e-131	86%	NP_567719.1
<input type="checkbox"/>	dehydration-responsive element-binding protein 1A [Arabidopsis thaliana]	348	348	100%	4e-122	86%	NP_567720.1

研究方法的结果

■ 序列比对 (NEEDLE)

全长序列 :

LENGTH	SCORE	IDENTITY	SIMILARITY	GAPS
1718	3101.5	760/1718 (44.2%)	760/1718 (44.2%)	852/1718 (49.6%)
LENGTH	SCORE	IDENTITY	SIMILARITY	GAPS
1763	3632.0	971/1763 (55.1%)	971/1763 (55.1%)	623/1763 (35.3%)

编码区序列 :

LENGTH	SCORE	IDENTITY	SIMILARITY	GAPS
628	2365.0	543/628 (86.5%)	543/628 (86.5%)	8/628 (1.3%)
LENGTH	SCORE	IDENTITY	SIMILARITY	GAPS
628	2288.5	531/628 (84.6%)	531/628 (84.6%)	17/628 (2.7%)

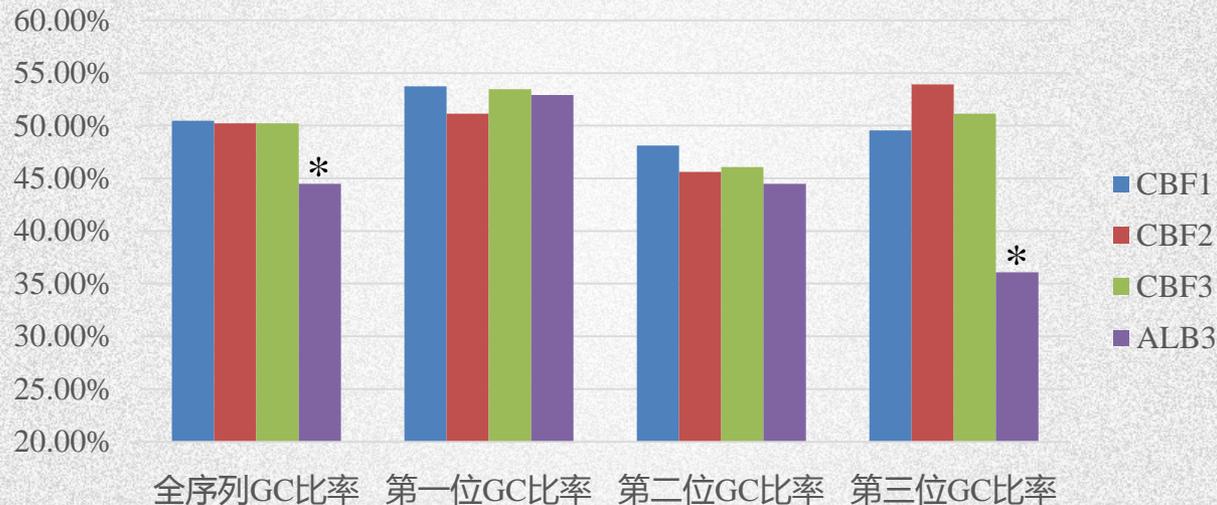
氨基酸序列 :

LENGTH	SCORE	IDENTITY	SIMILARITY	GAPS
208	939.0	180/208 (86.5%)	191/208 (91.8%)	2/208 (1.0%)
LENGTH	SCORE	IDENTITY	SIMILARITY	GAPS
208	922.0	178/208 (85.6%)	187/208 (89.9%)	5/208 (2.4%)

3个基因尤其是氨基酸序列彼此之间相似度都很高

CBFs理化性质的预测比较

■ 编码区密码子使用情况对比 (CUSP)



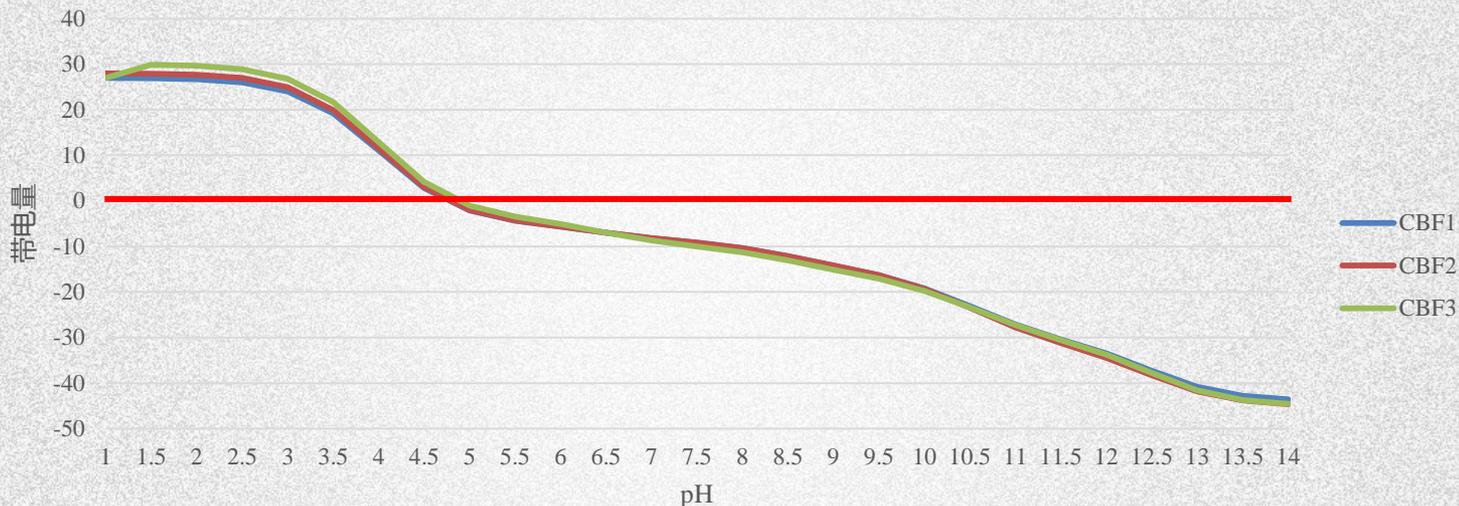
三个CBF基因编码区密码子各位点GC含量类似，与ALB3这种其他家族的基因存在差别

研究方法与结果

■ CBF蛋白氨基酸组成分析

蛋白名称	CBF1	CBF2	CBF3
小氨基酸	53.052%	51.389%	52.315%
脂肪链氨基酸	22.535%	24.537%	25.926%
芳香族氨基酸	11.737%	11.574%	12.500%
非极性氨基酸	52.113%	51.852%	51.852%
极性氨基酸	47.887%	48.148%	48.148%
带电	26.761%	27.315%	28.704%
碱性	12.207%	12.500%	13.426%
酸性	14.554%	14.815%	15.278%

■ 三个CBF蛋白等电点分析 (IEP)



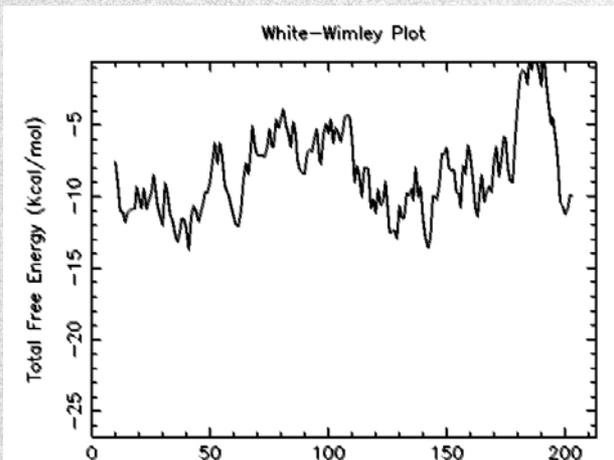
蛋白名称	CBF1	CBF2	CBF3
等电点	4.7395	4.7520	4.8673

三者等电点很近，能以等电状态存在于相同的酸性环境，因此其生化沉淀和电泳条件十分类似

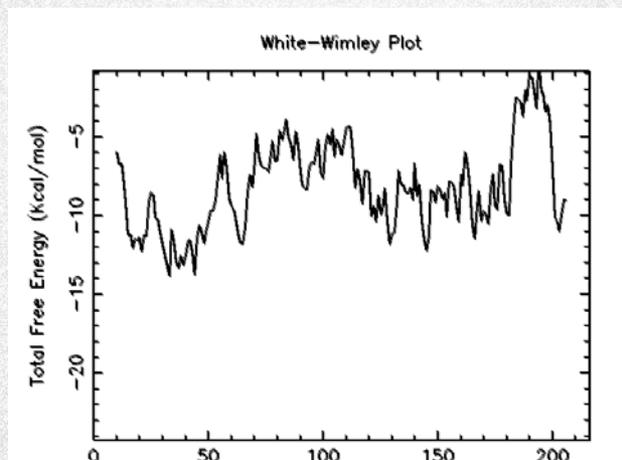
研究方法与结果

■ 蛋白亲水性分析(OCTANOL)

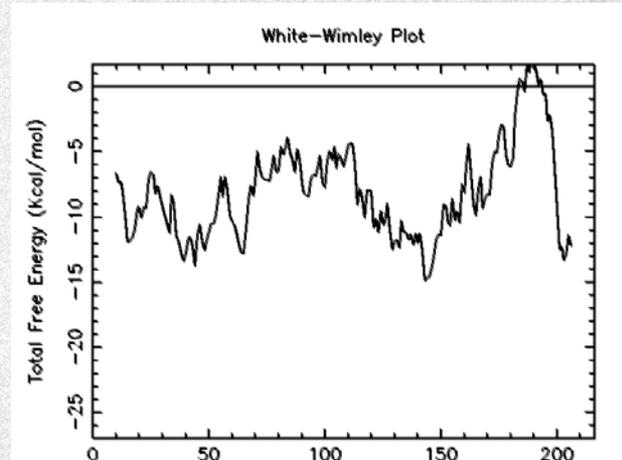
CBF1



CBF2



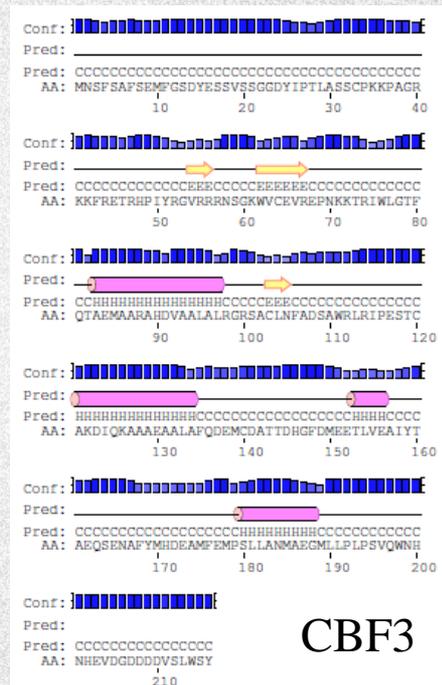
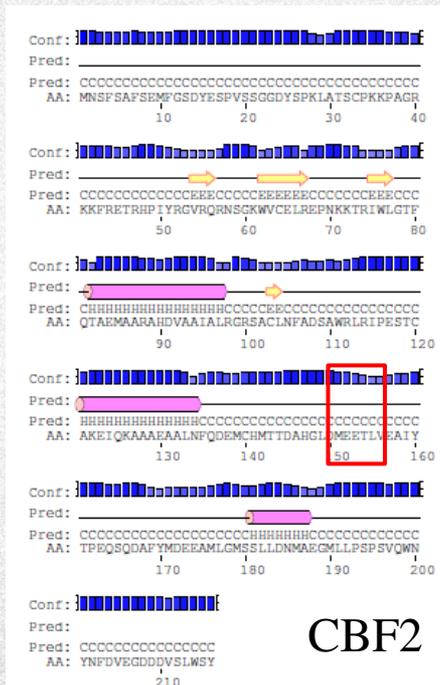
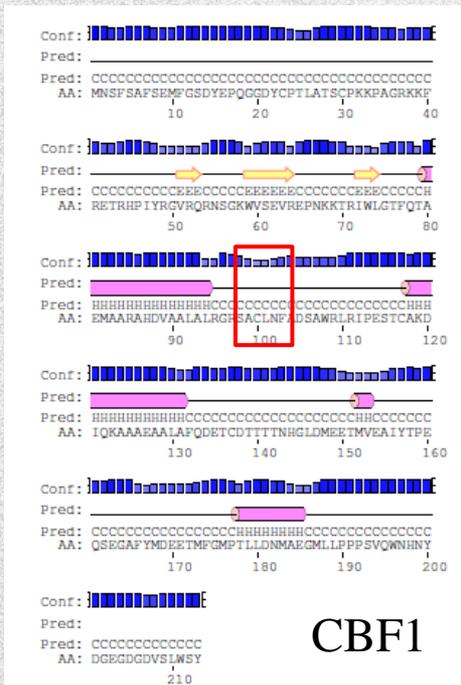
CBF3



三者疏水结构域很类似，暗示与其发生相互作用的分子类型也很类似

研究方法 与 结果

■ 二级结构预测(PSIPRED)



■ 启动子序列分析 (~700bp)

```
CBF1 -CATTTTTAACCGTCCCATCGAAATTGA---TAATTATCCAT-TACCAAATC-----TGATTA-----ATTTTTTTAAAAA--ATCAAG 612
CBF2 TT*CA***G*T*A*T**C*C*****A*ATTC**AAG*TTT*CCG*****A*TACT-**G*G*CT*TACCGTC*A*A**CGG*CG*CT*TT 636
CBF3 CG*AAC**G*T*TAAT----T****CA*ATA**TA**AT*A**TA*T*ATAATAA**AA*A*TATA-AC*****T*GG*C-*CA**T 623

CBF1 CTT---TTCTATATTGTAGTATTATTT-TTGGTTAAATAT-TAG-----GACATCTACTTCCAATACAAATACFACATGAGTATTAAA 533
CBF2 A**ATG*****C*T*TT**AC**G**CAC*****C*CTCTCCCA*TTCA**T*CTAG*AGG*TG*G***T*A**A*GGGAG** 546
CBF3 A*-----**--*CTC*C**AC*C*A---GA*G**T*GC*ATG-CAC**T--A*GTGAATGGAGA**G*AGA*A***C***G** 545

CBF1 A-TATC-----ATTTCACAGA--GATAT-TTATGTCTATTATGTT--ATAGACGGGTGA-CAATTAATGACAAATTTGTTTATT-CAT 459
CBF2 *T**CACAAGAAGACCA**TTTA*ATATCC*C+C*CAATCT*****CTA**C*****A*****C*****C*****-**G 457
CBF3 *A****TC-----G*****G**TCAT**G*C---AAT**T*ACC-----**AAA*C*-*****TAA*****AT*CGTGG 465

CBF1 AGGAAT-----TTAAAACGATTGTGAACAACAGCAGCCAGCCAA--CCACA-CAGGCACACACTCG-----ATAGAATTTAAGAACTCA 382
CBF2 GAAT**AATAA*G**T***G**ACCCT*CACCT**TACA*T**AT**TT*A**C*****T*ATACGC*A***G****T*A***** 367
CBF3 *AA**AATAT**T**T*G****CGT*C-CCT*-TA*A*T*ERG**--**TT**T**A*AAGAAGTTA*A**AA*TC*T*C*CT* 378

CBF1 TAAAGGTTAACGAGTGAAGAGTCAAAAGTCTCTTTTACAAGGGTCAAAGGACACCGTCAGA----CAGCGAGTGG-----AACCA---- 307
CBF2 *****AT*AT*A*A*C****G**T*AGTCAAAA*A*GTCT**TCT*****T*G****T-CTT*AT*****A-ATCCTT***-ACTA 281
CBF3 A*GGCA*C**C*****G****G**ACT**CCA*G*T***************TTCT**T**T*ACAGCCTTG*T*ATTA 288

CBF1 -----TCGTGGGATTGCTTCGCTATGTACTATAACCGTCTCATTCACAGAGACAAAACTCCGTGT--GCTCCCCACATATCCGTTATC 225
CBF2 CTCAT*T-ACA*****G****TAG*T*****G****AC*****C*****C*TCGC*TTT*TTT*GC*TC*A*AA 191
CBF3 CAAAAC*****C****A***G*T*CT**C*****G*****G*****C**TCGA*C**A**A*****--AA 201

CBF1 TCTCCTCCGGCCAATATAAAA-CACCA-ATTCTCACTCTCACTTTTATACTAACTACACACCTTAAAAAGAATCTACCTGAAAAGAAAA 137
CBF2 *A**T*****-***T*****A*****TGC*****CA*****-**T-***T*****A**A*T*-----********C--*G**C 110
CBF3 *A**T**********--*G*AGC*****CA**A***-C*-*****T**A**ACTT-----*CCTG*AT*AG***-G*-- 121

CBF1 AAAGAGAGAGATATAAATAGCTTTACCAAGACAGATATATCTTTTATTAATCCAAAAGACTGAGAACT-CTAGTACTACGTAC 48
CBF2 *G*A*****A*CT*T*-T**C*G*-----*CC*-*A*****A****T*TT***T*----- 42
CBF3 *G*T***T***A*****-T***T**T-----*CA*-----*A*****A****T**T***C*T***CTACT* 50

1
CBF1 FACTTAAACCTTATCCAGTTTCTTGAACAGAGTACTCT--GATCAATG
CBF2 *****TCT*****
CBF3 *-A*****TCT*****
```

三者启动子序列有一定的差异，不能确定功能完整的最小启动子区域是否相同；
PATA框外序列存在差异，因此三者表达模式上可能存
在不同

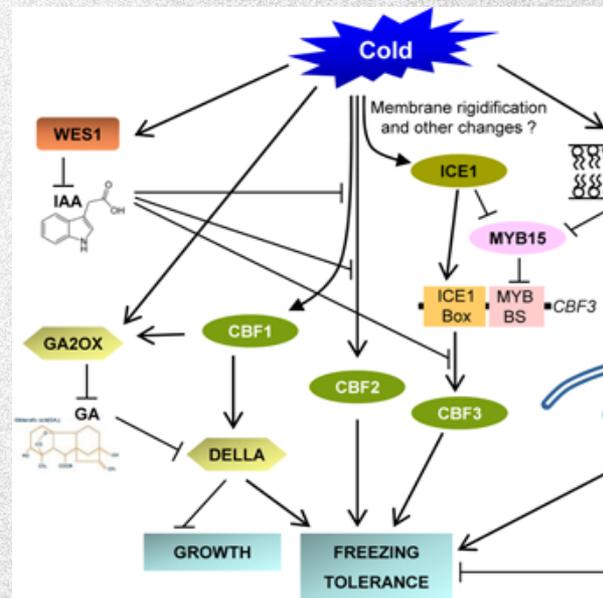
小结

- 三个CBF其基因编码区序列、蛋白氨基酸序列**相似**
度很高
- 三者的氨基酸组成方式、亲水性等理化性质也十分**相似**
- 二级结构预测结果**有差别**
- 启动子序列存在一定的**差异**

三个CBF的功能是否冗余？

分析与讨论

- 根据已有的研究结果，三者在功能上是存在差异的
- 从生信预测的结果来看并不矛盾
- 基因重复产生的同源基因在演化的过程中其功能发生了分化

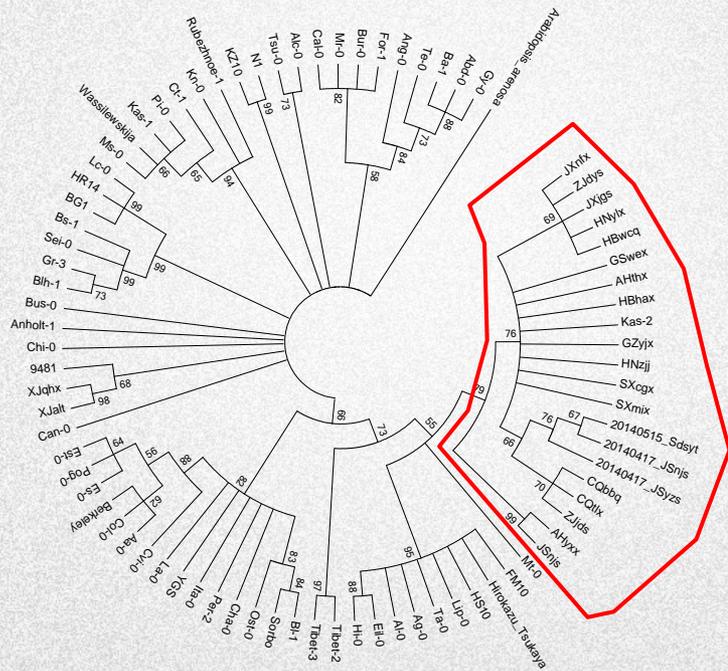


研究目的

- 能否从生物信息学的角度来预测3个*CBF* 基因的功能是否冗余
- **预测这三个基因对植物适应环境可能起到的作用以及它们在自然居群中的微演化模式**

自然居群的系统发育树构建

- 中国长江流域居群自成一支，分布于已知分布范围的最南端



已有实验结果（未发表）

- 长江流域居群的CBF1、CBF2转录激活活性极为**保守**，而CBF3的激活活性与居群所处生境温度**负相关**
- 重庆居群（CQtlx）等南方居群的CBF2编码区存在**移码突变**，提前终止，因此没有激活活性；其CBF3编码区存在一个点突变（G53→R53），也导致了活性的丧失
- 长江流域居群拥有一个重要的特点：CBF3的启动子存在大片段缺失和插入，使基因**无法正常表达**

中国居群的可能演化模式

- 长江流域居群来源于欧洲，从西部省市进入长江流域，由于其CBF3无法表达，在高温胁迫下得以保留而存活
- 而后居群向北扩散，低温胁迫开始发挥作用，其中CBF2活性正常的居群得以保留，而在陕西等地扩散开来。
- 三个CBF由于功能不一样，对于物种演化所起的作用也是不一样的

新的 编码

■ 此位点

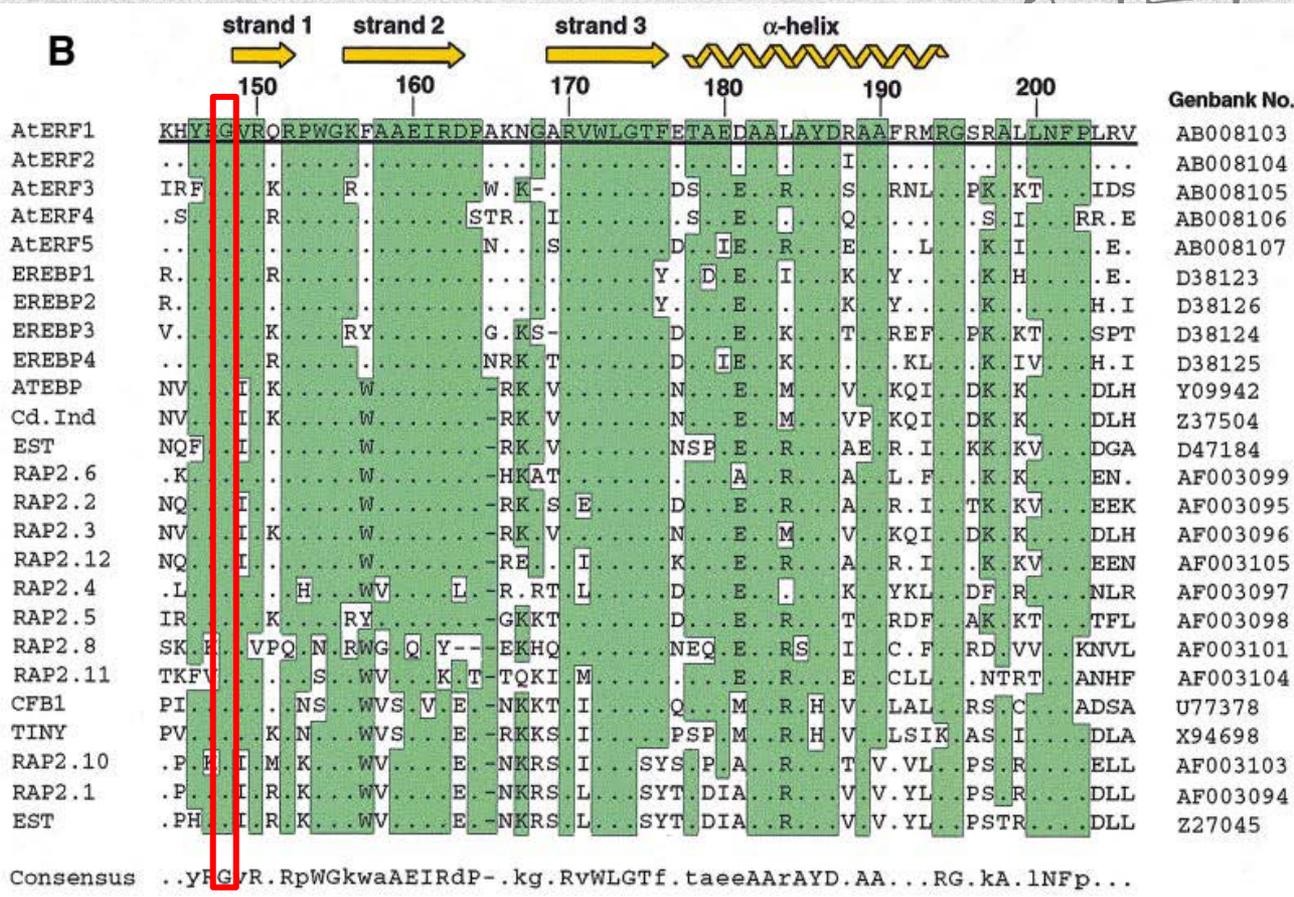
■ 氨基酸

响CBI

■ R可以

■ 氨基酸

作用



讨论

→R53)

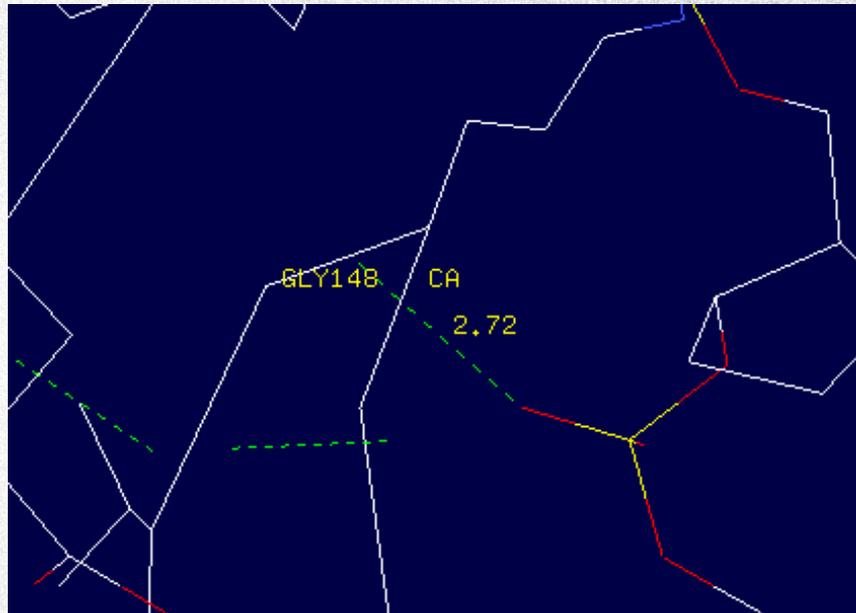
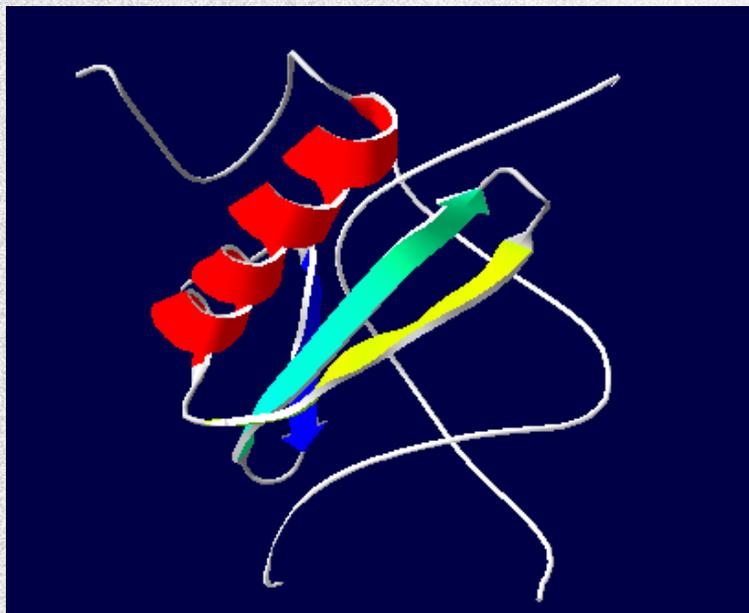
能会影

生

的相互

同源蛋白模拟相互作用

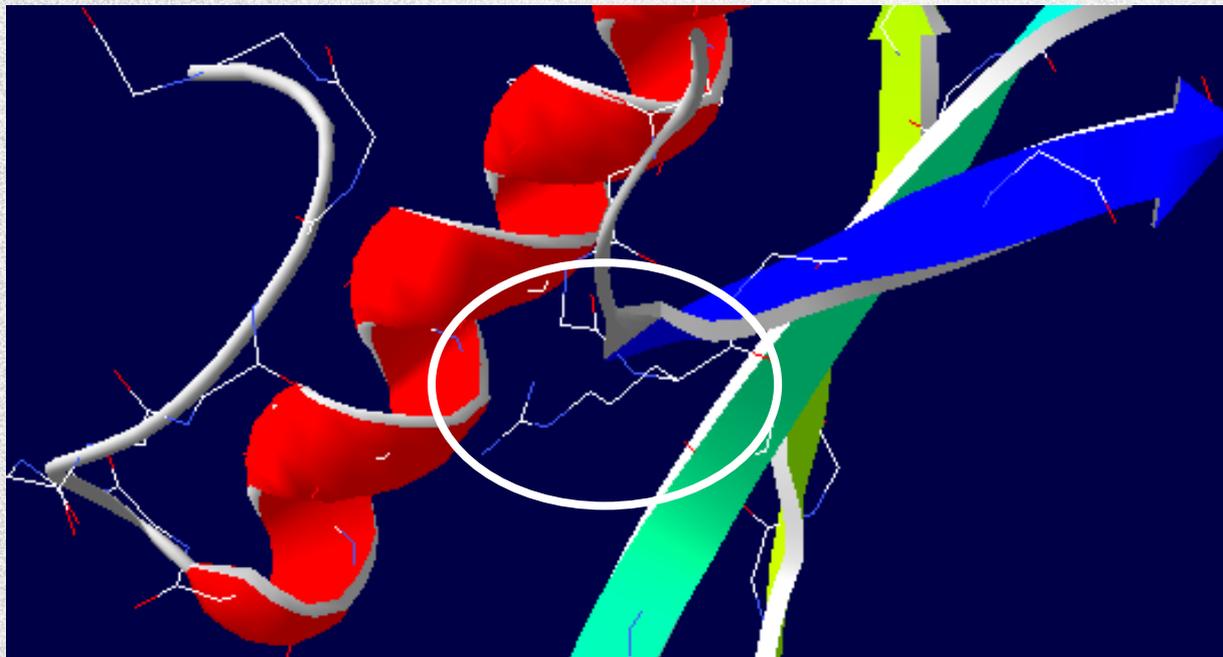
分析与讨论



模板：ERF1A，同属于AP2家族，对应CBF3第53位氨基酸的位点通过氢键与DNA发生相互作用

同源模建并构造点突变

分析与讨论



模拟突变后侧链无论朝向哪个方向，都可能对蛋白结构或者对蛋白与其他分子的相互作用有极大的影响

小结

- CBF转录因子亚家族中的三个成员是由**基因重复**产生的，但其功能在演化过程中发生了分化
- 三个CBF功能上的分化**反过来作用于演化过程**，尤其对中国拟南芥居群的微演化起着重要的作用
- 三者影响演化过程的方式是**不一样的**，CBF1相对**保守**，CBF2在自然居群中存在**移码突变**造成的提前终止，CBF3在自然居群中存在启动子的**缺失插入**，还存在单个氨基酸**位点突变**引起的活性完全丧失

致谢

- **感谢罗老师的指导**
- **感谢助教的耐心**
- **感谢组员们的无私付出**
- **感谢全班同学整个学期的陪伴和帮助**

THANKS

拟南芥CBF转录因子亚家族的生物信息学分析
以及其功能、演化方面的预测

报告人 施逸豪

组员 田博书 李莉 宋靖慧

Questions & Answers