

对拟南芥转录因子 TCP 家族的生物信息学分析

[背景知识]

转录因子的结构与调控作用

转录因子 (transcription factor) 也称为反式作用因子 (trans-acting factor), 是指能够与真核基因启动子区顺式作用因子特异性结合的 DNA 结合蛋白。通过他们之间以及和其他相关蛋白之间的相互作用, 激活或抑制转录。近年来, 相继从高等植物中分离出一系列调控干旱、高盐、低温、激素、病原反应及发育等相关基因表达的转录因子。

植物转录因子通过其功能域 DNA 及其它蛋白间的相互作用, 激活或抑制诱导型基因的表达。转录因子的 DNA 结合区决定了它与顺式作用元件结合的特异性, 而转录调控区决定了它对基因表达起激活还是抑制作用。此外, 其自身活性还受到核定位及寡聚化的影响。近年来, 基因分子生物学研究领域的重点已经逐渐从功能基因转到启动子的顺式作用元件和转录因子及其调控机理上。对转录因子的结构与功能的分析鉴定, 是阐明在各种条件下基因表达调控机理的重要内容之一, 揭示转录因子之间及它们与 DNA 之间相互作用的具体机制, 就可以人为的控制特定基因的表达, 使植物基因转化能获得好的结果。

TCP family 的转录因子

最早发现 *cycloidea* (*cyc*) 和 *teosinte branched 1* (*tb1*) 基因编码结构相似的蛋白, 它们都形成不典型的螺旋-环-螺旋结构 (non-canonical basic-Helix-Loop-Helix (bHLH) structure), 在水稻 PCF DNA 结合蛋白中也发现这种 bHLH 结构的存在。这种 bHLH 结构在结构和功能上都与经典的 HLH 不同, 于是将含有这种保守结构域的转录因子定义为一个新的转录因子家族——TCP 转录因子家族。(图一) TB1, CYC 和 PCFs 是这个家族最早的成员, 除此之外,

electrophoretic mobility shift analyses), 发现寡核苷酸探针 GTGGTCCCAC 能够结合所有的蛋白, 突变核心序列任一个 G, C 碱基突变为 A 或 T, 都造成探针失去结合活性, 这暗示了核心序列的保守性。突变 3' 端 AC 和 5' 端 GT 中的任一个, 会造成某些 PCF 蛋白的结合活性下降, 如果两个都突变, 则所有 PCF 蛋白都失去了结合活性。对拟南芥基因组的搜索表明, 拟南芥中只有少数几个包含这个共有序列的启动子。

TCP 转录因子家族在植物中的作用

金鱼草中 *cyc* 与 *dich* 基因对花的发育有很大作用。金鱼草开唇形花, 两侧对称但沿背腹轴花瓣形状显著不同。*cyc* 和 *dich* 两个基因的同时突变会产生辐射对称的花。*cyc* 单个基因突变产生半辐射状的花, 原腹部的花瓣移到侧旁, 原侧瓣移到背部。*Dich* 单个基因的突变只影响背部花瓣的形状。这两个基因的功能有重叠但是对花的发育又有不同的作用。

玉米中 *Tb1* 基因主要影响叶腋中的芽的生长发育, 在驯化野生墨西哥类蜀黍的过程中曾起重要作用。*Tb1* 突变株具有显性表型, 侧旁的初生茎成倍增长并且末端产生穗状的雄花, 而野生型的侧枝带有停止发育的芽或是正在形成的雌花, 在墨西哥类蜀黍中, *tb1* 基因的表达水平被降至一半。这些都暗示, *tb1* 基因有抑制侧枝生长和雄花形成的功能。*Cyc* 突变株也支持这个假设, 其显性性状为背部的雄蕊发育被抑制, 可能是细胞周期受到抑制, 因为在受抑制的雄蕊中检测不到细胞周期的调控因子 *cyclin D3b*, *cyclins B1 B2*, *cdc2b*, *cdc2c*。

水稻中 *PCF1*, *PCF2* 转录因子是促使分生组织中 *PCAN* (proliferation cell nuclear antigen) 启动子的组织特异性表达所必需的, 以产生 DNA 复制和修复机器所需的一个组件。

我们的实验

我们以 TCP family domain 搜索拟南芥全基因组序列 (from MIPS), 预测共有 24 个含有 TCP 保守序列的基因, 我们对其中 6 个基因的 ORF 序列进行克隆。从拟南芥总 RNA 出发, 经过 RT-PCR 获得 6 个可能的 TCP 转录因子 ORF 序列。然

后用改进的酵母单杂交方法来鉴定他们是否具有转录激活活性。

[生物信息学分析]

1. 以下是我们进行研究的 TCP family 部分成员的 CDS 序列，我们通过 Gateway Cloning Technology 克隆到下列基因，测序得到如下结果：

>At1g68800

```
ATGTTTCCTTCTTTCATTACTCACATTCAAAGCCCTAATTCTCACCATCACTACTCTTCGCCTTCTTTT
CCTTTCTCTCCGATTTTCTTGAGAGTTTTGATGAATCCTTCTTGATAAACCAATCTTGTTACAGCAG
CAAGATGTAGCAGCAAATGTTGTTGAATCTCCTTGAAAATTTGCAAGAAGCTTGAGCTTAAGAAGAAG
AATGAGAAGTGTGTTGATGGAAGCACCTCACAAGAGGTTCAATGGAGAAGGACGGTCAAAAAAAGGGAC
AGGCATAGTAAGATCTGCACGGCTCAAGGTCCTAGAGACCGGAGGATGAGGCTGTCTCTCAGATTGCT
CGCAAGTTTTTCGATCTTCAAGACATGTTGGGTTTTGACAAGGCGAGCAAGACGATTGAATGGCTTTTC
TCCAAATCAAAGACTTCCATCAAACAACCTTAAAGAAAAGAGTGGCTGCATCGGAAGGAGGAGGAAAGGAT
GAACATCTCCAGGTTGATGAAAAGGAAAAGGATGAGACACTGAAGTTGAGAGTCTCAAAGAGAAGAACA
AAGACTATGGAGAGCTCTTTTAAGACTAAAGAGTCGAGAGAGAGAGCTAGAAAAGCGAGCAAGAGAGAGA
ACAATGGCAAAGATGAAGATGAGATTATTTGAGACCTCGAAAACAATTCAGATCCTCATCAAGAACT
AGAGAGATCAAGATAACCAATGGTGTACAATTAAGAAAAGGAAAATAAAGAACAAGAATGGAGTAAT
ACTAATGATGTTACATGGTAGAGTATCAAATGGATTCTGTGAGCATCATAGAGAAGTTTCTTGACTA
ACCAGTGACTCTAGCTCCTCTTCCATTTTTGGTGACTCCGAGGAATGTTACACAAGTCTTAGTTCAGTA
AGAGGTACAATTCAGCAGCAGGTAACAGCAATGTGTTAACTAAAAACCCTAATTGA
```

>At2g45680

```
ATGGCGACAATTCAGAAGCTTGAAGAAGTTGCAGGCAAAGATCAAACCTAAGAGCCGTTGATCTAACC
ATCATCAACGGCGTCAGAAACGTCGAAACTTCAAGACCTTTCCAAGTAAATCCCACAGTGAGTCTCGAG
CCCAAGGCGGAGCCGGTGTATGCCGTCGTTTTCAATGTCTTTAGCTCCACCGTCTTCGACAGGACCACCA
TTGAAGAGAGCTTCGACTAAAGACCGTCACACGAAGGTTGAAGGAAGAGGGAGAAGGATACGGATGCCT
GCCACGTGTGCGGCTAGGATTTTTCAATTAACCTCGAGAGTTAGGTCACAAATCCGACGGCGAAACGATT
CGGTGGTTGTTGGAGAACGCTGAGCCGGCGATTATAGCCGCCACGGGTACGGGAACGGTTCCCGCCATC
GCCATGTCGGTTAACGGAACCTTAAAAATCCCGACGACGACGAACGCTGATTCTGATATGGGTGAAAAT
CTGATGAAGAAGAAACGTAAACGACCTTCTAACAGTGAGTATATAGACATAAGCGACGCCGTTTCAGCT
TCCTCCGTTTTAGCTCCAATTGCCACGACGACAACGATCCAACCTCCGCAAGCTCTGGCATCATCCACT
GTGGCTCAGCAACTTCTGCCGCAAGGAATGTATCCGATGTGGGCTATTCCATCAAACGCAATGATTCCG
ACGGTCGGAGCTTTCTTCTTGATTCCACAAATCGTGGTCCGTCGAATCAGCCTCAGTTATTAGCTTTT
CCCCGCCCGCTGCTTCGCCGTCGTCTTACGTCGCCGCTGTTCAACAGGCTTCCACGATGGCTAGACCA
CCTCCTTTACAAGTTGTTCCAAGCAGCGGCTTTGTATCCGTTTCAGACGTTAGCGGTTTCAATTTATCA
AGAGCGACGTCGGTTATGGCTCCGAGCTCAAGCTCAGGCGTAACAACCGGTAGTTCATCGTCAATTGCA
ACAACAACGACGCACACGCTGAGAGACTTCTCCCTAGAGATATACGAGAAACAAGAGCTTACCAGTTC
ATGAGCACCACAACAGCACGGTCATCGAACCACTGA
```

>At3g02150

```
ATGAATATCGTCTCTTGAAAAGATGCAAACGACGAAGTTGCAGGCGGCGCTACGACAAGACGTGAAAGA
GAAGTAAAAGAGGATCAAGAAGAAAACCGAAGTCAGAGCCACCAGTGGCAAAACCGTAATTA AAAAGCAG
```

CCTACATCGATCTCTTCTTCTTCTTCTTCTCGTGGATGAAATCCAAGGATCCGAGGATTGTTAGGGTTTCA
CGGCCTTTGGAGGCAAAGACCGTCACAGCAAAGTGTGTACGTTACGTGGACTACGTGACAGACGCGTG
AGATTATCAGTCCCAACGGCTATTACAGCTCTACGATCTTCAAGAACGGCTCGGTGTTGACCAGCCTAGC
AAAGCCGTTGACTGGTTGCTTGATGCAGCTAAAGAGGAGATCGACGAGCTACCTCCGTTACCTATCTCG
CCGAAAATTTACGATCTTCAACCATCATCAGTCCTTCTTGAATCTTGGTCAACGGCCCCGGTCAAGAT
CCGACCCAACCTCGGGTTTAAAATCAATGGATGTGTACAAAAGTCTACTACTACTAGCCGCGAAGAAAAC
GATAGAGAGAAAGGAGAAAACGATGTCGTTTACACAAAACAATCATCATGTTGGGTCTTATGGAECTTAT
CACAACCTGGAACATCATCATCATCACCAACATTTGAGTTTACAGGCAGATTATCATAGTCATCAA
CTACATAGTCTTGTCCATTTCCATCACAAAATTTGGTATGTCCAATGACGACATCACCAACAACCTACA
ACTATAACAATCTTTGTTTCCATCATCATCGTCAGCTGGTTCAGGGACTATGGAGACATTAGATCCGAGG
CAAATGGTAAGCCATTTTCAAATGCCATTAATGGGTAAATTTCTTCATCTTCGCTCTCCCAAAACATTTCC
ACATTATATTCGTTGTTACATGGTAGTAGCAACAATGGTGGTAGAGACATTGATAATCGGATGTCCG
TCGGTCCAATTC AACCGAACTAATAGCACTACAACGGCTAACATGTCGAGGCATCTAGGCTCGGAGCGT
TGTACAAGTAGAGGAAGTGATCACCATATGTGA

>At3g45150

ATGGATTGAAAAATGGAATTAACAACAGCCAAAAGGCAAGAAGGACTCCAAAAGACCGCCATTTGAAA
ATTGGTGGCCGTGATCGTCGCATTCCGATCCCGCCGAGTGTGCTCCCAACTATTTAGATTGACAAAA
GAACTCGGCTTTAAAACCGATGGTGAAACTGTCAGTTGGCTCCTCCAGAATGCCGAGCCTGCCATTTTC
GCAGCCACGGGACATGGTGTACCACCACCTCCAATGAAGATATCCAGCCAAATAGGAATTTTCCTAGT
TACACCTTTAATGGTGATAATATTAGTAATAACGTTTTCCCTTGTACGGTTGTAAATACTGGTCATCGT
CAGATGGTGTTCGGTTTCTACAATGACAGATCATGCACCTTCAAATAATTACAGTACTATTAGTGAT
AATTACAATTCACCTTTAATGGTAATGCTACCGCCAGTGATACAACATCAGCAGCAACAACAACAGCC
ACAACCACAGTTTGA

>At5g08330

ATGGCCGACAACGACGAGCAGTGAGTAACGGCATCATAGTCGAGCAGACGTCAAACAAAGGACCTCTT
AACGCCGTTAAGAAACCACCGTCTAAAAGATCGACACAGCAAAGTTGACGGAAGAGGAAGAAGGATTCGT
ATGCCAATCATTTGCGCAGCTCGAGTTTTTCAATTGACCAGAGAGTTAGGTCACAAGTCCGATGGTCAA
ACCATAGAGTGGCTTCTCCGTCAAGCTGAGCCTTCTATCATAGCCGCCACTGGAACCTGGCACTACTCCG
GCGAGTTTCTCCACTGCTTCTCTCTCACTTCTTCTCCGTTTACTCTCGGGAAACGTGTCGTCAGAGCG
GAGGAAGGAGAATCCGGCGGCGGAGGAGGAGGGTTAACAGTGGGACACACAATGGGGACTTCGTTA
ATGGGTGGTGGTGGTTCTGGTGGGTTTTGGGCTGTTCCGGCGAGGCCGATTTCCGACAAGTCTGGAGC
TTTGCAACCGGAGCTCCACCGGAAATGGTTTTTGGCAGCAGCAGCAACCAGCTACACTCTTCGTCCGC
CACCAGCAGCAACAGCAAGCTTCCGCCGCCGAGCAGCTGCAATGGGTGAGGCTTCAGCAGCTAGAGTT
GGGAATTATCTTCCGGGTCATCATCTCAATTTGCTTGCTTCTTGTCTGGTGGAGCTAACGGGTCGGGT
CGGAGGGAAGACGACCACGAACCACGTTGA

>At5g41030

ATGGTCATGGAGCCCAAGAAGAACCAAAAATCTACCAAGTTTCTTAAACCCATCACGACAGAATCAGGAC
AACGACAAGAAGAGGAAAACAAACAGAGGTTAAAGGTTTCGACATTGTGGTCCGGCGAAAAGAGGAAGAAG
AAGGAGAATGAAGAGGAAGACCAAGAAATTCAGATTCTTTATGAGAAGGAGAAGAAGAAACCAACAAA
GATCGTCACCTTAAAGTTGAAGGAAGAGGTCGTAGAGTTAGGTTACCTCCACTCTGTGCAGCAAGGATT
TATCAATTGACTAAAGAATTAGGTCACAAATCAGATGGTGAGACTCTTGAATGGTTGCTTCAACATGCT
GAGCCATCGATACTCTCTGCTACTGTAAATGGTATCAAACCCACTGAGTCTGTTGTTTCTCAACCTCCT
CTCACGGCTGATTTGATGATTTGTCATAGCGTTGAAGAAGCTTCAAGGACTCAAATGGAGGCAAATGGG
TTGTGGAGAAATGAAACAGGACAGACCATTGGAGGGTTTGATCTGAATTACGGAATTGGGTTTGATTTTC

AATGGTGTTCAGAGATTGGTTTTGGAGATAATCAAACGCCTGGACTTGAATTAAGGCTGTCTCAAGTT
GGGGTTTTGAATCCACAGGTTTTTCAACAAATGGGTAAAGAACAGTTCAGGGTTCTTCATCATCATTCA
CATGAAGATCAGCAGCAGAGTGCAGAGGAAAATGGTTCATAA

2. 以下是这些 TCP family 成员的全长 ORF ,
(<http://us.expasy.org/tools/#translate>) :

>At1g68800

MFPSFITHIQSPNSHHYSSPSFPFSSDFLESFDESFLINQFLLQQQDVAANVVESPWKFCKKLELKKK
NEKCVDGSTSQEVQWRRTVKKRDRHSKICTAQGPRDRMRLSLQIARKFFDLQDMLGFDKASKTIEWLF
SKSKTSIKQLKERVAASEGGGKDEHLQVDEKEKDETLKLRVSKRRTKTMESSFKTKESRERARKRERER
TMAKMKMRLFETSETISDPHQETREIKITNGVQLLEKENKEQWSNTNDVHMVEYQMDSVSIIEKFLGL
TSDSSSSSIFGDSEECYTSLSVVRGTISAAGNSNVLTKNPN

>At2g45680

MATIQLKEEVAGKDQTLRAVDLTIINGVRNVETSRPFQVNPTVSLEPKAEPVMPSFSMSLAPPSSTGPP
LKRASKDRHTKVEGRGRRIRMPATCAARIFQLTRELGHKSDGETIRWLLNAEPAIIAATGTGTVPAI
AMSVNGTLKIPTTTNADSDMGENLMKKKRKPSNSEYIDISDAVSASSGLAPIATTTTIQPPQALASST
VAQQLLPQGMYPMWAIPSNAMIPTVGAFFLIPQIAGPSNQPQLLAFPAAAAASPSSYVAAVQQASTMARP
PPLQVVPSSGFVSVSDVSGSNLSRATSVMAPSSSSGVTTGSSSSIATTTHTLRDFSLEIYEKQELHQF
MSTTTARSSNH

>At3g02150

MNIVSWKDANDEVAGGATRREREVKEDQEETEVRATSGKTVIKKQPTSISSSSSSWMKSKDPRIVRVS
RAFGGKDRHSKVCTLRGLRDRRVRLSVPTAIQLYDLQERLQVDQPSKAVDWLLDAAKEEIDELPPLPIS
PENFSIFNHHQSFNLGQRPGQDPTQLGFKINGCVQKSTTTREENDREKGENDVVYTNHHVGSYGTY
HNLEHHHHHHQHLSLQADYHSHQLHSLVPFPSQILVCPMTTSPTTTTIQSLFPSSSSAGSGTMETLDPR
QMVSHFQMPLMGNSSSSSSQNI STL YSLLHGSSSNNGGRDIDNRMSSVQFNRTNSTTTANMSRHLGSER
CTSRGSDHMM

>At3g45150

MDSKNGINNSQKARRTPKDRHLKIGGRDRRIRIPPSVAPQLFRLTKELGFKTDGETVSWLLQNAEPAIF
AATGHGVTTTSNEDIQPNRNFPSYTFNGDNISNNVFPCTVVNTGHRQMVFPVSTMDHAPSTNYSTISD
NYNSTFNGNATASDTTSAATTTATTTV





>At5g08330



MADNDGAVSNGIIVEQTSNKGPLNAVKKPPSKDRHSKVDGRGRRIRMPIICAAARVFQLTRELGHKSDGQ
TIEWLLRQAEPSIIAATGTGTPASFSTASLSTSSPFTLQKRVVRAEEGESGGGGGGGLTVGHTMGTSL
MGGGGSGGFVAVPARPDFGQVVSFATGAPPEMVFAQQQQPATLQVVRHQQQQASAAAAAAMGEASAARV
GNYLPGHHLNLLASLSGGANGSGRREDDHEPR

>At5g41030

MVMEPKKNQNLPSFLNPSRQNDNDKKRKQTEVKGFDIVVGEKRRKKKENEEDQEIQLYEKEKKKPNK
DRHLKVEGRGRRVRLPPLCAARIYQLTKELGHKSDGETLEWLLQHAEPSILSATVNGIKPTESVVSQPP
LTADLMICHSVVEASRTQMEANGLWRNETGQTIGGFDLNYGIGFDNGVPEIGFGDNQTPGLELRLSQV
GVLNPQVFQMQGKEQFRVLHHSHEDQQQSAEENG

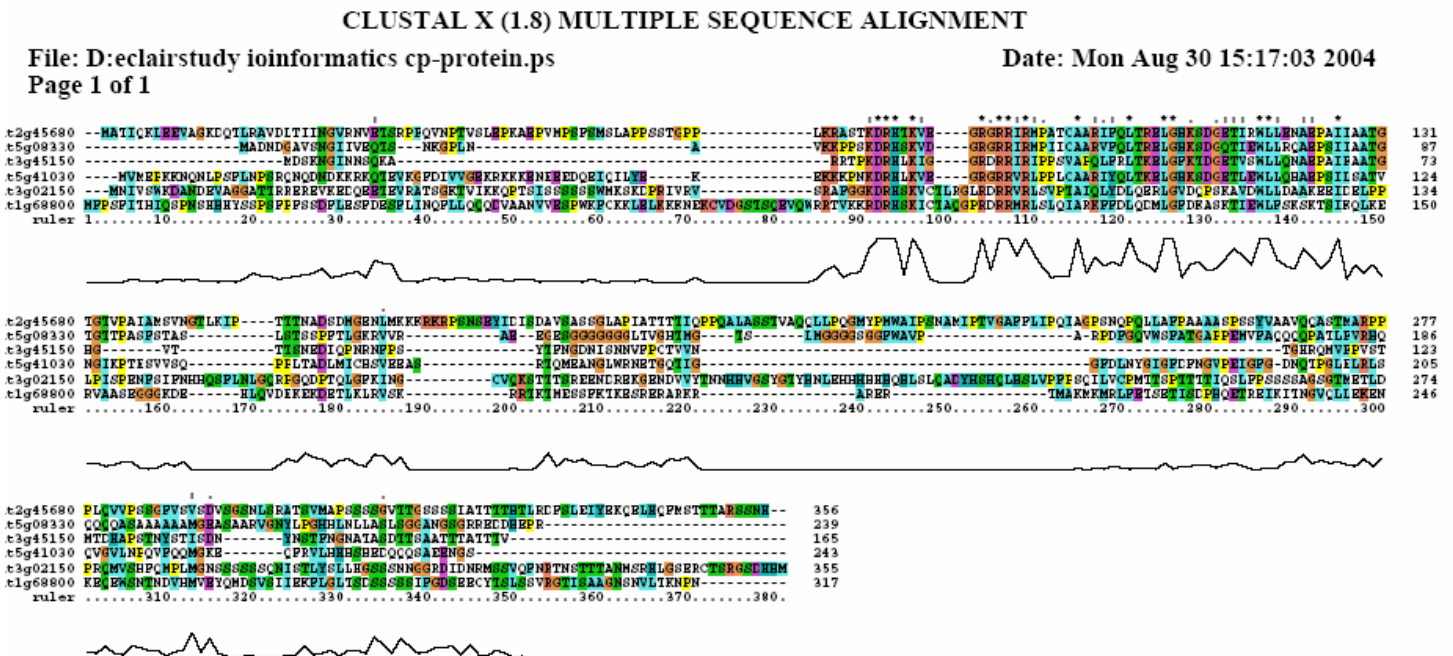
3. 首先对 ORF 蛋白序列进行结构域搜寻 (<http://smart.embl-heidelberg.de/>)
得到如下结果:

Name		保守序列																
At1g68800 	<table border="1"> <thead> <tr> <th>name</th> <th>begin</th> <th>end</th> <th>E-value</th> </tr> </thead> <tbody> <tr> <td>low complexity</td> <td>19</td> <td>38</td> <td>-</td> </tr> <tr> <td>low complexity</td> <td>59</td> <td>73</td> <td>-</td> </tr> <tr> <td>Pfam:TCP</td> <td>74</td> <td>283</td> <td>3.30e-87</td> </tr> </tbody> </table>	name	begin	end	E-value	low complexity	19	38	-	low complexity	59	73	-	Pfam:TCP	74	283	3.30e-87	VDGSTSQEVQWRRTVKKRDRHSK ICTAQGPRDRRMRLSLQIARKFF DLQDMLGFDKASKTIEWLFSKSK TSIKQLKERVAASEGGGKDEHLQ VDEKEKDETLKLRVSKRRRTKTME SSFKTKESRERARKRARERTMAK MKMRLFETSETISDPHQETREIK ITNGVQLLEKENKEQEWSNTNDV HMVEYQMDSVSIIEKFLGLTSDS SSS
name	begin	end	E-value															
low complexity	19	38	-															
low complexity	59	73	-															
Pfam:TCP	74	283	3.30e-87															
At2g45680 	<table border="1"> <thead> <tr> <th>name</th> <th>begin</th> <th>end</th> <th>E-value</th> </tr> </thead> <tbody> <tr> <td>Pfam:TCP</td> <td>59</td> <td>280</td> <td>4.80e-84</td> </tr> <tr> <td>low complexity</td> <td>281</td> <td>296</td> <td>-</td> </tr> <tr> <td>low complexity</td> <td>308</td> <td>328</td> <td>-</td> </tr> </tbody> </table>	name	begin	end	E-value	Pfam:TCP	59	280	4.80e-84	low complexity	281	296	-	low complexity	308	328	-	SLAPPSSTGPPLKRASTKDRHTKVEGR GRRIRMPATCAARIFQLTRELGHKSDG ETIRWLLENAEPATIAATGTGTVAIA MSVNGTLKIPTTTNADSDMGENLMKKK RKRPSNSEYIDISDAVSASSGLPIAT TTTIQPPQALASSTVAQQLLPQGMYPM WAIPSNAMIPTVGAFFLIPQIAGPSNQ PQLLAFPAAAAASSPYVAAVQQASTMA RPPPLQ
name	begin	end	E-value															
Pfam:TCP	59	280	4.80e-84															
low complexity	281	296	-															
low complexity	308	328	-															
At3g02150 	<table border="1"> <thead> <tr> <th>name</th> <th>begin</th> <th>end</th> <th>E-value</th> </tr> </thead> <tbody> <tr> <td>low complexity</td> <td>49</td> <td>56</td> <td>-</td> </tr> <tr> <td>Pfam:TCP</td> <td>58</td> <td>293</td> <td>7.10e-98</td> </tr> <tr> <td>low complexity</td> <td>307</td> <td>314</td> <td>-</td> </tr> </tbody> </table>	name	begin	end	E-value	low complexity	49	56	-	Pfam:TCP	58	293	7.10e-98	low complexity	307	314	-	MKSKDPRIVRVSRAFGGKDRHSKVCTL RGLRDRRRLSVPTAIQLYDLQERLGV DQPSKAVDWLLDAAKEEIDELPLPIS PENFIFNHHQSFLNLGQRPGDPTQL GFKINGCVQKSTTTTSREENDREKGEND VVYTNNHHVGSYGTYHNLEHHHHHHQH LSLQADYHSHQLHSLVPFSPQILVCPM TTSPTTTTIQSLFPSSSSAGSGTMETL DPRQMVSHFQMLMGNSSSS
name	begin	end	E-value															
low complexity	49	56	-															
Pfam:TCP	58	293	7.10e-98															
low complexity	307	314	-															
At3g45150 	<table border="1"> <thead> <tr> <th>name</th> <th>begin</th> <th>end</th> <th>E-value</th> </tr> </thead> <tbody> <tr> <td>Pfam:TCP</td> <td>1</td> <td>165</td> <td>4.10e-63</td> </tr> </tbody> </table>	name	begin	end	E-value	Pfam:TCP	1	165	4.10e-63	MDSKNGINNSQKARRTPKDRHLKIGGR DRRIRIPPSVAPQLFRLTKELGFKTDG ETVSWLLQNAEPAIFAATGHGVTTSN EDIQPNRNFPSTFNGDNISNNVFPCT VVNTGHRQMVFPVSTMTDHAPSTNYST ISDNYNSTFNGNATASDTTSAATTTAT TTV								
name	begin	end	E-value															
Pfam:TCP	1	165	4.10e-63															

<p>At5g08330</p> 	<table border="1"> <thead> <tr> <th>name</th> <th>begin</th> <th>end</th> <th>E-value</th> </tr> </thead> <tbody> <tr> <td>Pfam:TCP</td> <td>15</td> <td>194</td> <td>2.90e-66</td> </tr> <tr> <td>low complexity</td> <td>213</td> <td>230</td> <td>-</td> </tr> </tbody> </table>	name	begin	end	E-value	Pfam:TCP	15	194	2.90e-66	low complexity	213	230	-	<p>EQTSNKGPLNAVKKPPSKDRHSKVDGRG RRIRMPIICAARVFLTRELGHKSDGQT IEWLLRQAEPSSIIAATGTGTPASFSTA SLSTSSPFTLGKRVVRAEEGESGGGGGG GLTVGHTMGTSLMGGGGSGGFVAVPARP DFGQVWSFATGAPPEMVFAQQQQPATLF VRHQQQQQASAA</p>
name	begin	end	E-value											
Pfam:TCP	15	194	2.90e-66											
low complexity	213	230	-											
<p>At5g41030</p> 	<table border="1"> <thead> <tr> <th>name</th> <th>begin</th> <th>end</th> <th>E-value</th> </tr> </thead> <tbody> <tr> <td>Pfam:TCP</td> <td>52</td> <td>227</td> <td>9.90e-67</td> </tr> </tbody> </table>	name	begin	end	E-value	Pfam:TCP	52	227	9.90e-67	<p>EDQEIQILYEKEKKKPNKDRHLKVE GRGRRVRLPPLCAARIYQLTKELGH KSDGETLEWLLQHAEPSILSATVNG IKPTESVVSQPPLTADLMICHVVEE ASRTQMEANGLWRNETGQTIGGFDL NYGIGDFDFNGVPEIGFG DNQTPGLELRLSQVGLNQPQVQQMGKE QFRVLH</p>				
name	begin	end	E-value											
Pfam:TCP	52	227	9.90e-67											

我们可以看到，克隆得到的全长 ORF 均符合预测，含有固定的 TCP 保守区。由此我们对得到的克隆进行了进一步的酵母单杂实验，以验证这些因子是否具有转录因子特有的激活活性和 DNA 结合活性。

3 我们对全长 ORF 通过软件 clustalx 进行多序列比对，得到如下结果：



由多序列比对图上我们可以看到，在大约 90~150AA 的地方有保守序列，即 SMART 分析中的 TCP DOMAIN.

4. 我们进行的酵母单杂实验数据如下(L12 及 L34 为 TCP family 特有的两个结合序

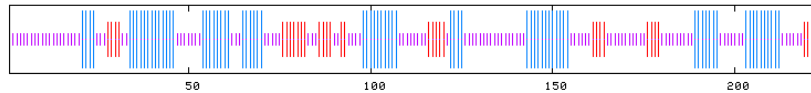
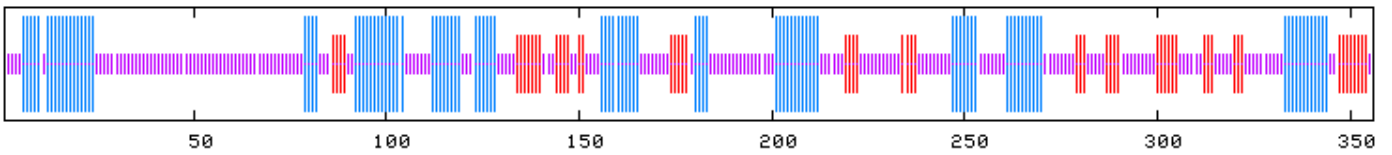
ARSSNH

eeeeec

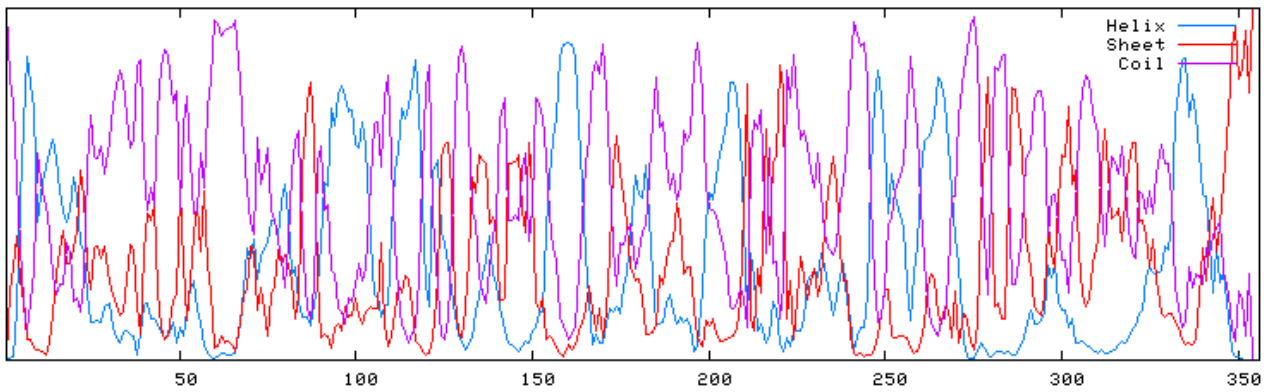
Sequence length : 356

GOR4 :

Alpha helix	(Hh)	:	104	is	29.21%
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)	:	57	is	16.01%
Beta turn	(Tt)	:	0	is	0.00%
Bend region	(Ss)	:	0	is	0.00%
Random coil	(Cc)	:	195	is	54.78%
Ambiguous states (?)		:	0	is	0.00%
Other states		:	0	is	0.00%



保守区域



At3G45150

10 20 30 40 50 60 70

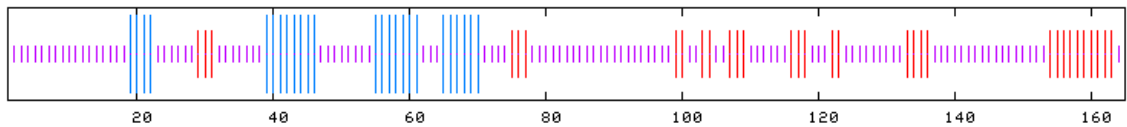
MDSKNG I NNSQKARRTPKDRHLK I GGRDRR I R I PPSVAPQLFRL TKELGFKTDGETVSWLLQNAEPA I FA
cccccccccccccccccccc hhhhccccceeecccccc hhhhhhhcccccccc hhhhhhhccc hhhh
ATGHGVTTTSNED I QPNRNFPSYTFNGDN I SNNVFCTVNTGHRQMVFPVSTMTDHAPSTNYST I SDNY
hccccceeecccccccccccccccccccc eeeceeeccccceeeccccceeecccccccc eeecccc
NSTFNGNATASDTTSAATTTATTTV

cccccccccccccccc eeeeeeeeeec

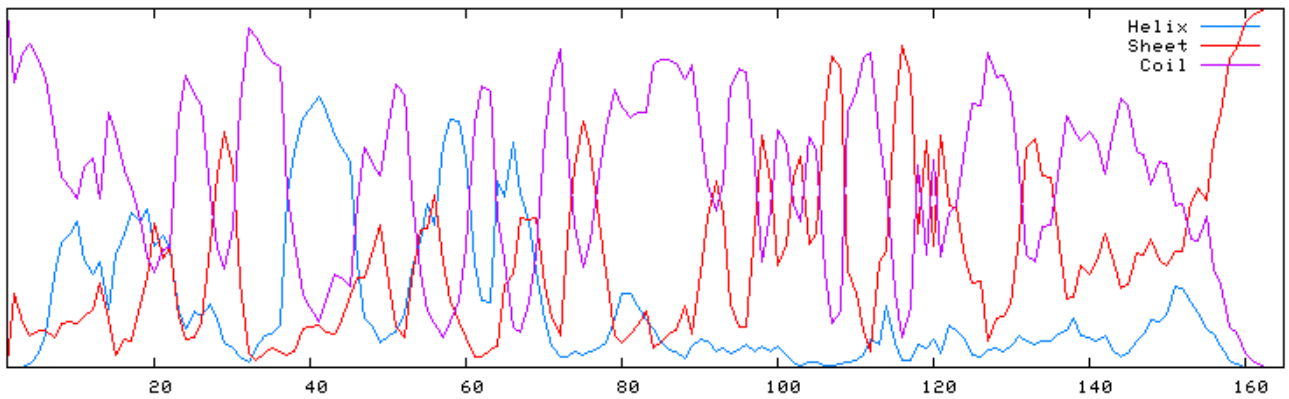
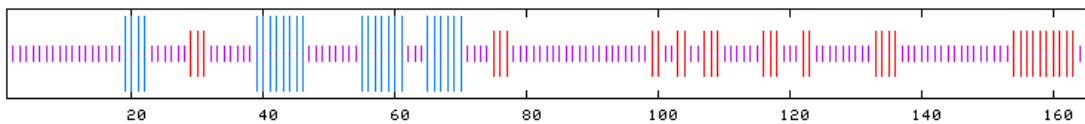
Sequence length : 165

GOR4 :

Alpha helix	(Hh)	:	25	is	15.15%
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)	:	32	is	19.39%
Beta turn	(Tt)	:	0	is	0.00%
Bend region	(Ss)	:	0	is	0.00%
Random coil	(Cc)	:	108	is	65.45%
Ambiguous states (?)		:	0	is	0.00%
Other states		:	0	is	0.00%



保守区



At5g08330

```

      10      20      30      40      50      60      70
      |      |      |      |      |      |      |
MADNDGAVSNGI | VEQTSNKGPLNAVKKPPSKDRHSKVDGRGRR | RMP | | CAARVFQLTRELGHKSDGQT
cccccccccccc eeeecccccccccccccccccccccccccccc eeee chhhhhhhhhhhhhhh cccccchh
|EWLLRQAEPS | |AATGTGTTASPFSTASLSTSSPFTLGKRVVRAEAGESGGGGGGGLTVGHTMGTSLMG
hhhhhccccceeeecccccccc eeeecccccccccccccccc eeeecccccccccccc eeeeeeeeeeee
GGGSGGFWAVPARPDFGQVWSFATGAPPEMVFAQQQQPATLFVRHQQQQASAAAAAAMGEASAARVGNV
ccccceeeecccccccc eeeecccc chhhhccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh ccc

```


ADLMI CHSVEEASRTQMEANGLWRNETGQT I GGFDLNYG I GDFNGVPE I GFGDNQTPGLELRLSQVGVL

cceeeccchhhhhhhhhhhhhccccccccceccccccccccccccccccccccccccccccccccccchhhhhcccccc

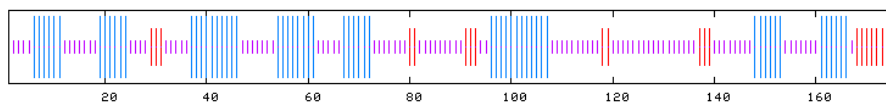
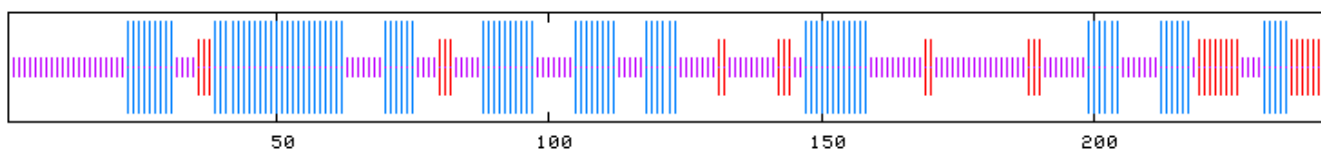
NPQVFQQMGKEQFRVLHHSHEDQQQSAEENG

ccchhhhhceeeeeeeccccchhhhhheeeeeec

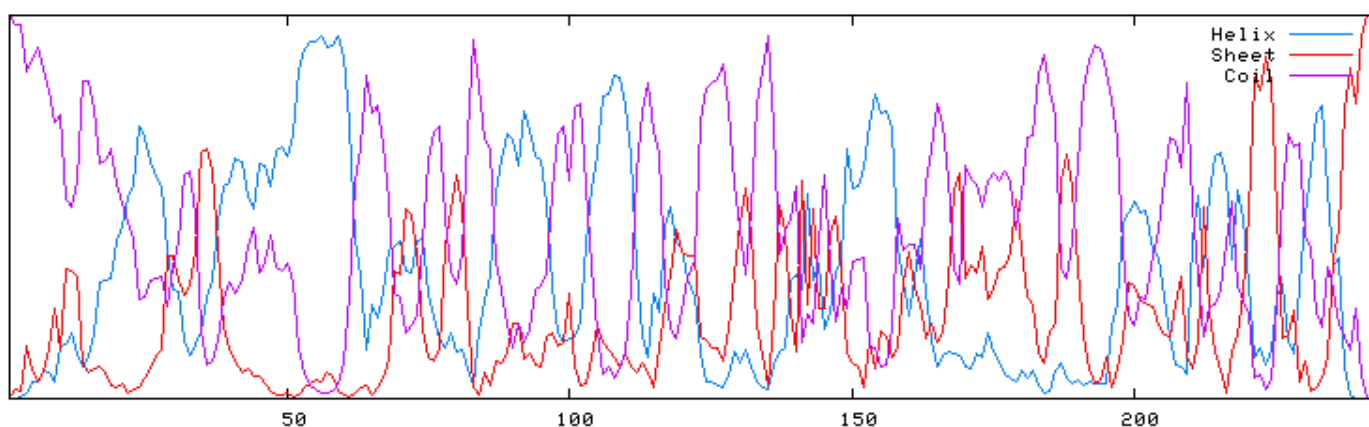
Sequence length : 243

GOR4 :

Alpha helix (Hh) :	92 is	37.86%
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) :	30 is	12.35%
Beta turn (Tt) :	0 is	0.00%
Bend region (Ss) :	0 is	0.00%
Random coil (Cc) :	121 is	49.79%
Ambiguous states (?) :	0 is	0.00%
Other states :	0 is	0.00%



保守区



由上图我们可以看到，这几个蛋白二级结构都非常相似，以 α 螺旋和随机 coil 为主。

6. 对蛋白进行三级结构预测 (SWISS MODEL), 由于序列较短, 结构简单, 没有三级结构。以下是 [HTTP://SWISSMODEL.EXPASY.ORG/](http://swissmodel.expasy.org/) 的回复 email:

Your modeling request could not be carried out.

Please look at the TraceLog file issued by the server.
The degree of similarity of your sequence with proteins of known 3D structure may be too low.

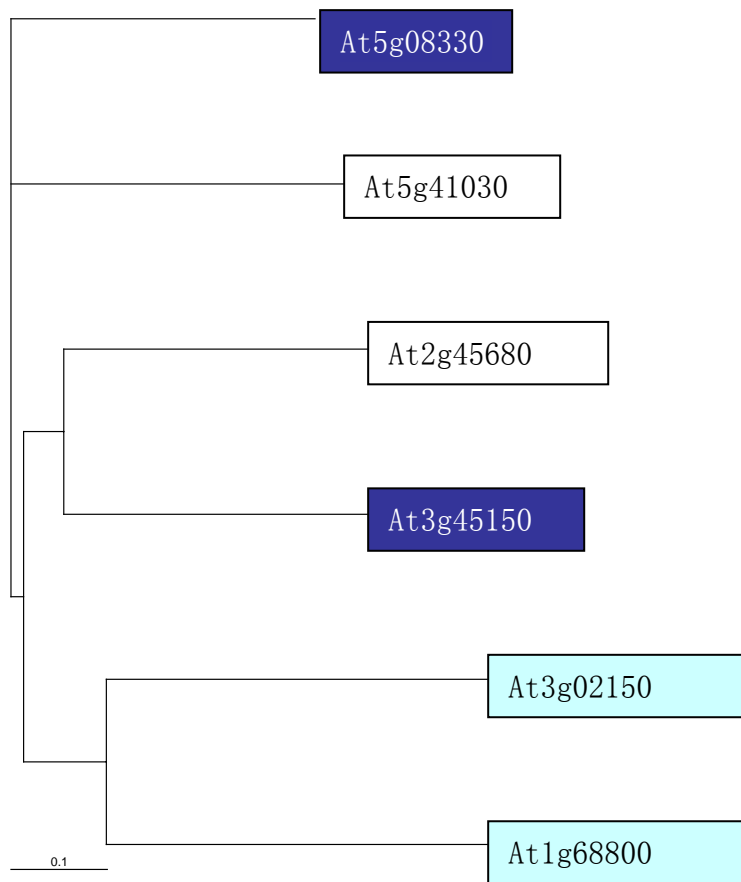
At present, SWISS-MODEL will generate models for sequences which respond to these criteria:

BLAST search P value : < 0.00001

Global degree of sequence identity (SIM) : > 25 %

Minimal projected model length = 25 aa.

7. CLUSTALX 建树:



建树的结果让我们发现, **At1g68800** 和 **At3g02150** 处于同一分支上, 他们都既没有激活活性也没有结合活性, 符合我们的实验结果; 同样既有激活活性又有和结合活性的

At3g45150 和 At5g08330 在进化关系上比较接近, At5g41030 和 At2g45680 与 At3g45150 和 At5g08330 在进化上如此接近却是既没有激活活性又没有结合活性, 这是为什么呢? 一是可能它们具有激活活性和结合活性的结构域在进化过程中丢失了, 但是其他部分仍然保存, 所以建树结果显示在进化上它们仍然与 At3g45150 和 At5g08330 非常接近; 另一种可能是我们的酵母双杂实验本身不够精细, 几个氨基酸的差别会在实验过程中导致结合序列无法识别, 如果改用 EMSA 可能会获得更可信的实验结果; 还有一个因素就是在转酵母的 IN VIVO 实验中, 插入质粒载体的序列的蛋白产物很容易就会产生变化甚至丢失, 这个也会导致蛋白原本的性能丢失, 或者导致表达量的巨大变化, 需要寻求更可靠的方法。将带有待测序列的载体质粒整合至酵母染色体基因组的方法可以保证在传代的过程中质粒不丢失。

[实验前景展望]

实验证明, 通过序列搜索得到的预测 TCP 转录因子家族中的成员并不都有转录激活活性和 DNA 结合活性。在拟南芥 TCP 家族全部成员全部检测完之后, 拟用凝胶阻滞实验 (EMSA) 作体外验证。同时, 正在进行的实验有筛选拟南芥 TCP 家族 knock-out 植株。由于已知 TCP 转录因子家族成员主要在分生组织中作用, 与细胞分化和生长有很大联系, 希望能够观察到 knock-out 株的表型, 进一步分析在拟南芥中 TCP 家族成员的作用机制。

[参考文献]

1. Jiang CJ, Imamoto N, Matsuki R, Yoneda Y, Yamamoto N (1998) Functional characterization of a plant importin alpha homologue. Nuclear localization signal (NLS)-selective binding and mediation of nuclear import of nls proteins in vitro. *J Biol Chem* 273:24083-7
2. Washburn KB, Davis EA, Ackerman S (1997) Coactivators and TAFs of transcription activation in wheat. *Plant Mol. Biol.* 35:1037-1043
3. Hill A, Nantel A, Rock CD, Quatrano RS. (1996) A conserved domain of the viviparous-1 gene product enhances the DNA binding activity of the bZIP protein EmBP-1 and other transcription factors. *J.B.C.* 271:3366-3374
4. Goff SA, Cone KC, Chandler VL. (1992) Functional analysis of the transcriptional activator encoded by maize B gene: evidence for a direct functional interaction between two classes of regulatory protein. *Genes*

Dev. 6: 864-875

5. Lu J, Sloan SR. (2002) The basic helix-loop-helix domain of the E47 transcription factor requires other protein regions for full DNA binding activity. *Biochem Biophys Res Commun* 290:1521-1528

6. Liu L, White M J, MacRae T H. (1999) Transcription factors and their genes in higher plants. *Eur J Biochem*, 262: 247~257

7. Rubin GM, Yandell MD, Wortman JR, Gabor Miklos GL, Nelson CR, *et al* (2000) Comparative genomics of the eukaryotes. *Science*, 287: 2204-2215

8. Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddie J, Adam L, Pineda O, *et al* (2000) *Arabidopsis* Transcription Factors: Genome-Wide Comparative Analysis Among Eukaryotes. *Science*, 290:2105-2110

9. Luscombe NM, Austin SE, Berman HM, Thornton JM (2000) An overview of the structures of protein-DNA complex. *Genome Bio.* 1:1-37

10. Tautz D (2000) Evolution of transcriptional regulation *Current Opinion in Genetics & Development* 10:575 - 579

11. Pilar Cubas, Nick Lauter, John Doebley and Enrico Coen (1999) The TCP domain: a motif found in proteins regulating plant growth and development. *The Plant Journal* April;18(2):215-22

12. Shunichi Kosugi and Yuko Ohashi (2002) DNA binding and dimerization specificity and potential targets for the TCP protein family. *The Plant Journal* May;30(3):337-48

13. Kosugi, S., and Y. Ohashi. (1997) PCF1 and PCF2 specifically bind to *cis* elements in the rice proliferating cell nuclear antigen gene. *Plant Cell* 9:1607-1619

14. Kosugi, S., and Y. Ohashi. (2002) DNA binding and dimerization specificity and potential targets for the TCP protein family. *Plant J.* 30:337-348.

15. Cubas, P., N. Lauter, J. Doebley, and E. Coen. (1999). The TCP domain: a motif found in proteins regulating plant growth and development. *Plant J.* 18:215-222.

16. Luo D, Carpenter R, Copsey L, Vincent C, Clark J, Coen E. (1999) Control of organ asymmetry in flowers of *Antirrhinum*. *Cell*. Nov 12;99(4):367-76.

17. Doebley, J., Stec, A. and Hubbard, L. (1997) The evolution of apical dominance in maize. *Nature* 386, 485-488.

18. 刘强 张贵友 陈受宜, 植物转录因子的结构与调控作用 *科学通报* 第45卷第14期 2000年7月

19. Li JJ, Herskowitz I. Isolation of ORC6, (1993) a component of the yeast origin recognition complex by a one-hybrid system. *Science*;262:1870-1873

20. Inouye C, Remondelli P, Karin M, *et al.* (1994) Isolation of a cDNA encoding a metal response element binding protein using a novel expression cloning procedure: the one hybrid system. *DNA Cell Biology*, 13:731-742

21. Wei Z, Angerer LM. (1999) Identification of a new sea urchin Ets

protein, SpEts4, by yeast one-hybrid screening with the hatching enzyme promoter. *Mol Cell Biol*, 19: 1271-1278

22. Chew LJ, Huang F, Boutin JM. (1999) Identification of nuclear orphan receptors as regulators of expression of a neurotransmitter receptor gene. *J Biol Chem*, 274(41): 29366-29375

23. Wilson TE, Day ML, Pexton T, Padgett KA, Johnston M, Milbrandt J. (1992) In vivo mutational analysis of the NGFI-A zinc fingers. *J Biol Chem* ;267:3718-3724

24. Lisowsky T, Polosa PL, Sagliano A, Roberti M, Gadaleta MN, Cantatore P. (1999) Identification of human GC-box-binding zinc finger protein, a new Kruppel-like zinc finger protein, by the yeast one-hybrid screening with a GC-rich target sequence. *FEBS Lett* ;453:369-374

25. Mak KL, Longcor LC, Johnson SE, Lemercier C, To RQ, Konieczny SF. (1996) Examination of mammalian basic helix-loophelix transcription factors using a yeast one-hybrid system. *DNA Cell Biol* ;15:1-8

26. 廖名湘, 方福德, 酵母单杂交体系——一种研究 DNA-蛋白质相互作用的有效方法 *中国医学科学院学报* 第 22 卷第 4 期, 2000 年 8 月

27. Ye R, Yao QH, Xu ZH, Xue HW. (2004) Development of an efficient method for the isolation of factors involved in gene transcription during rice embryo development. *Plant J*. 2004 Apr;38(2):348-57

[致谢]

首先感谢为我们开设生物信息学这门课程的罗静初老师, 在课堂上, 他像所有我在北大遇到的其他老师一样, 给我们教授了许多非常使用的生物信息学分析方法以及分析思路; 但是他有着自己非常独特的地方, 那就是他极其严谨的治学态度和对学生的严格要求。一个学期的课程结束了, 我们从他那里学到的远远不止学习上的东西, 他的态度给我们这些将来会从事科学研究的人极大的鼓舞和指导, 感谢他给我们带来的这么多改变。

感谢我的导师朱玉贤老师给了我进入这个实验室的学习机会。虽然因为年级低, 很少能够得到他直接的指导, 但是一年多来的组会听他分析实验数据和实验前景, 以及多次听他与师兄师姐们分析讨论问题, 使我受益匪浅。

感谢我实验室的师兄师姐们, 尤其是我的师傅魏刚, 从基本实验技能到思考分析问题都是他们教会我的; 感谢小文、小强和大海, 他们是我的同学, 也是我在实验室的好伙伴, 我们相互关心相互鼓励, 共同进步。

最后, 感谢助教师兄, 很多次习题的指导, 他耐心的讲解, 帮助我们了解了很多软件的使用方法和思路。

对拟南芥转录因子 TCP 家族 的生物信息学分析

生化 03 硕 雷娟

10311039

2004. 8. 31