

烟草 C2H2 锌指蛋白转录因子家族成员的鉴定与表达分析

杨明磊^{1,2}, 晁江涛¹, 王大伟¹, 胡军华^{1,2}, 吴华^{1,2}, 龚达平¹, 刘贯山¹

1. 中国农业科学院烟草研究所, 中国农业科学院烟草遗传改良与生物技术重点开放实验室, 青岛 266101;

2. 中国农业科学院研究生院, 北京 100081

摘要: C2H2 锌指蛋白转录因子家族在真核生物中具有重要的生物学功能, 广泛参与植物叶的发生、花器官的调控、侧枝的形成及逆境胁迫等生命过程。植物 C2H2 锌指蛋白不仅结合 DNA 和 RNA, 而且与蛋白质之间相互作用。本研究利用普通烟草(*Nicotiana tabacum*)基因组数据库, 运用 Blastp 比对, 结合 Pfam 和 SMART 分析, 鉴定了 118 条普通烟草 C2H2 锌指蛋白家族成员; 对烟草 C2H2 锌指蛋白家族进行了进化树分析、结构域分析、物理化学性质分析、染色体定位、基因结构分析、三维结构分析及组织表达分析等。结果表明: 不同成员的氨基酸长度差异较大; 系统进化及结构域分析显示, 所有 C2H2 家族成员可以被分为 5 个亚家族, 同一亚家族成员之间在结构域和理化性质上呈现较高一致性; 每个成员都含有 C2H2 结构域, 在数量上存在较大差异; 将所有基因家族成员定位在 22 条染色体上; 组织表达分析表明, 每个 C2H2 亚家族都有成员在不同组织中表达, 在叶及根中有些基因的表达量较高。

关键词: C2H2 锌指蛋白转录因子; 烟草; 基因家族

Genome-wide identification and expression profiling of the C2H2-type zinc finger protein transcription factor family in tobacco

Minglei Yang^{1,2}, Jiangtao Chao¹, Dawei Wang¹, Junhua Hu^{1,2}, Hua Wu^{1,2}, Daping Gong¹, Guanshan Liu¹

1. Key Laboratory of Tobacco Genetic Improvement and Biotechnology, Tobacco Research Institute of Chinese Academy of Agricultural Science, Qingdao 266101, China;

2. Graduate School of Chinese Academy of Agricultural Science, Beijing 100081, China

Abstract: C2H2 zinc finger protein transcription factor family members have important biological functions in eukaryotes. They not only bind DNA and RNA, but also interact with proteins. In this study, 118 members of the tobacco C2H2 zinc finger protein transcription factor family were identified from the *N. tabacum* genome database by using Pfam, SMART and Blastp. The analyses of phylogenetic tree, physical and chemical properties, chromosomal

收稿日期: 2015-10-21; 修回日期: 2015-12-25

基金项目: 国家烟草专卖局烟草基因组计划重大专项项目(编号: 110201301005(JY-05))和中国农业科学院烟草研究所青年科学基金项目(编号: 2015B05)[Supported by the Tobacco Genome Project of State Tobacco Monopoly Administration(No.110201301005(JY-05)) and the Science Foundation for Young Scholars of Institute of Tobacco Research of CAAS(No. 2015B05)]

作者简介: 杨明磊, 在读硕士, 专业方向: 烟草突变体鉴定与利用。E-mail: 787093482@qq.com

晁江涛, 助理研究员, 研究方向: 生物信息学。E-mail: 290594533@qq.com

杨明磊和晁江涛为并列第一作者。

通讯作者: 刘贯山, 研究员, 博士生导师, 研究方向: 烟草分子育种。E-mail: liuguanshan@caas.cn

DOI: 10.16288/j.yczz.15-440

网络出版时间: 2016/1/26 15:59:51

URL: <http://www.cnki.net/kcms/detail/11.1913.R.20160126.1559.002.html>

mapping, gene structures, protein three-dimensional structures and tissue expression patterns were performed. The results suggested that the peptide length of different subfamily members is significantly different. Phylogenetic and motif analysis revealed that the C2H2 zinc finger protein transcription factor family members can be divided into 5 subfamilies and each member has at least one C2H2 motif. Genes of the family members are distributed across the 22 chromosomes. C2H2 zinc finger protein transcription factor family members are expressed in different tissues although some have higher expression levels in leaves and roots. This study will be helpful for further analysis of the C2H2 zinc finger family proteins in other plants.

Keywords: C2H2 zinc finger protein transcription factor; tobacco; gene family

转录因子是生物体中调控基因转录的一类蛋白质,能直接结合或者间接作用于特定核苷酸序列,激活或抑制目标基因表达。锌指蛋白(Zinc finger protein, ZFP)转录因子是植物中最大的转录因子家族之一^[1],在植物形态发生、转录激活和逆境胁迫等多种生物学进程中起重要作用^[2]。锌指蛋白结构域是由锌原子作为核心、半胱氨酸或组氨酸包围锌原子形成的稳定的三维结构,有两个 β 折叠和一个 α 螺旋^[3],锌指之间由接头连接 β 折叠和 α 螺旋,接头内的氨基酸具有高度保守性。根据半胱氨酸和组氨酸的位置和数目可以将锌指转录因子分为 C2H2、C2HC、C2HC5、C2C2、CCCH、C3HC4、C4、C4HC3、C6 和 C8 等 10 种不同类型^[4,5]。C2H2 锌指转录因子(也称 C2H2 锌指蛋白)属于 TF 转录因子。RNA 聚合酶分为 RNA 聚合酶 I、RNA 聚合酶 II 及 RNA 聚合酶 III,TF 转录因子即与 RNA 聚合酶 II 相对应,是 RNA 聚合酶 II 转录 5S rRNA 所必需的,含有一个由大约 30 个氨基酸组成的锌指结构,该结构的保守序列为 X2-Cys-X(2-4)-Cys-X12-His-X(3-5)-His(Cys 为半胱氨酸,His 为组氨酸,X 代表任意氨基酸)^[6]。两个半胱氨酸及两个组氨酸与锌离子结合形成一个在蛋白质中相对独立的区域。与其他生物相比,植物 C2H2 锌指蛋白具有两个比较独特的结构:一是相邻两个锌指之间的氨基酸序列,数目较多且种类变化较大;二是与 DNA 的结合处存在一段高度保守的氨基酸序列 QALGGH,该序列是植物 C2H2 锌指蛋白所特有,在其他真核生物中尚未发现^[7,8]。在功能上,植物 C2H2 锌指蛋白不仅结合 DNA 和 RNA,而且与蛋白质相互作用,具有转录调控、RNA 代谢和其他多种生物学功能^[9-11]。Takatsuji 等^[12]在矮牵牛(*Petunia hybrida*)中发现了第一个植物特有的

C2H2 锌指蛋白,它参与 5-烯醇丙酮莽草酸-3-磷酸合酶基因的调控。随后,对于植物 C2H2 锌指蛋白的研究逐渐增多,在拟南芥(*Arabidopsis thaliana*)、水稻(*Oryza sativa* L.)、小麦(*Triticum aestivum* L.)等植物中都有该类基因的报道,如在矮牵牛中发现的 ZPT、水稻中发现的 RZF71 和 RZF5^[13]、拟南芥中发现的 SZT 基因^[14]等。SZT 编码一种双锌指结构的 C2H2 锌指蛋白,转入此基因的烟草对冷胁迫和高盐胁迫的耐受能力明显提高^[14]。目前已经在拟南芥、小麦、水稻等植物中分离到近百个 C2H2 锌指蛋白编码基因。

普通烟草(*Nicotiana tabacum*)是一种重要的经济作物,也是开展生物学研究重要的模式植物。目前研究的一些 C2H2 锌指蛋白编码基因如拟南芥中的 SZT 基因、矮牵牛中 ZPT2-3 基因都转入烟草中进行功能研究,但是对于烟草本身的 C2H2 锌指蛋白的研究还鲜见报道。本研究以普通烟草 C2H2 锌指转录因子家族为研究对象,采用生物信息学手段对其基因家族各成员的进化关系、基因结构、物理化学性质、保守结构域、三维结构、组织表达等进行分析,为烟草及其他植物中 C2H2 锌指蛋白转录因子家族的功能研究提供理论参考。

1 材料和方法

1.1 普通烟草 C2H2 锌指蛋白转录因子家族蛋白质序列的获得

利用 FGENESH 软件,基于已报道的植物 C2H2 锌指蛋白转录因子家族蛋白序列的相似性,预测烟草基因组 C2H2 锌指蛋白转录因子家族。烟草基因组数据库来源于茄科基因组数据库(<http://solgenomics.net/>)。

从 TAIR 数据库(<http://www.arabidopsis.org/>)下载拟南芥 C2H2 锌指蛋白全长序列, Pfam 数据库(<http://pfam.xfam.org/>)下载 C2H2 锌指蛋白种子序列(Pfam ID: PF00096)。将两类序列与预测得到的普通烟草 C2H2 锌指蛋白转录因子家族的序列中进行 Blastp 比对(e 值: $1e-15$), 获得普通烟草中相似性高的全长蛋白质序列, 人工去除冗余序列, 确定候选序列。然后将候选序列分别在 SMART 和 Pfam 中进行蛋白质结构域分析, 剔除不含有 C2H2 结构域的蛋白质序列, 最终鉴定出 118 条普通烟草 C2H2 锌指蛋白序列。利用在线工具 ExPASy Proteomics Server(<http://expasy.org/>)对其家族成员进行蛋白质理化性质分析^[15]。利用 WoLF PSORT 在线工具对各个基因家族成员的进行亚细胞定位分析^[16]。

1.2 多序列比对及系统进化树的构建

利用 MUSCLE^[17]软件对普通烟草和拟南芥的 C2H2 锌指蛋白转录因子家族成员进行多序列比对, 参数选为默认值。选取保守的 C2H2 结构域序列, 利用 BoxShade 将其结果可视化。将普通烟草 C2H2 锌指蛋白序列比对结果, 使用 MGEA5 构建邻接树(Neighbor-Joining, NJ), Bootstrap 检验次数设为 1000 次^[18]。根据进化树及结构域将 C2H2 锌指蛋白转录因子家族成员重新命名。

1.3 蛋白质全长序列结构域分析

将各个亚家族蛋白质全长序列, 利用 Multiple Expectation Maximization for Motif Elicitation (MEME, <http://meme.nbcr.net/meme/cgi-bin/meme.cgi>)^[19]进行蛋白序列基序分析。

1.4 染色体定位

利用 Perl 程序解析普通烟草基因组信息文件, 确定 C2H2 锌指蛋白编码基因的染色体位置, 利用 MapGene2 软件将 C2H2 锌指蛋白转录因子家族成员定位到普通烟草 24 条染色体上, 并绘制图谱。

1.5 普通烟草 C2H2 锌指蛋白三维结构的预测与分析

将普通烟草蛋白质全长序列进行多序列比对, 选取结构域较保守的蛋白序列为目标序列。将目标序列在 SWISS-MODEL(<http://swissmodel.expasy.org/interactive>)网站上查找三维结构模板, 根据覆盖率、一致性百分比、实验的可靠性选择最适三维结构模板,

使用 Pymol 软件标出保守结构域。

1.6 普通烟草 C2H2 锌指蛋白编码基因的组织表达分析

在 NCBI(<http://www.ncbi.nlm.nih.gov>)中下载烟草转录组测序数据, 利用 R 语言 Bioconductor 程序对转录组数据进行分析并提取 C2H2 锌指蛋白转录因子家族转录组数据。将数据均一化后, 使用 R 语言 gplots 程序包绘制热图。

2 结果与分析

2.1 普通烟草与拟南芥 C2H2 家族成员的进化分析及蛋白理化性质分析

利用 C2H2 锌指蛋白全长序列和种子序列在基因组中进行 Blastp 比对, 再经过序列和结构域筛选, 最终鉴定出 118 条普通烟草 C2H2 锌指蛋白序列。对这些全长蛋白序列进行系统发生分析, 拟南芥与烟草 C2H2 锌指蛋白转录因子家族成员分成 5 个亚家族(、 、 、 、), 每个亚家族中分别含有 56、24、24、6 和 8 个烟草 C2H2 锌指蛋白转录因子家族成员, 每个亚家族中都含有至少 3 个拟南芥 C2H2 锌指蛋白转录因子家族成员, 在每个亚家族中拟南芥家族成员在数量上也和普通烟草 C2H2 锌指蛋白转录因子家族成员的数量成对应关系(图 1, 图 2, 附表 1)。进化树分析显示, 普通烟草 C2H2 锌指蛋白转录因子家族成员中有 40 个支持率较高($>96\%$)的姊妹对。对获得的 C2H2 锌指蛋白转录因子家族成员的理化特征进行分析发现, 不同亚家族之间的氨基酸序列长度变化呈现不同规律, 第 IV 亚家族成员氨基酸序列长度变化较大, 在 222~1502aa 之间, 蛋白质分子质量介于 25.4~167.7 kDa 之间; 而第 、 、 、 亚家族成员氨基酸序列长度变化均相对较小, 在 107~572aa 之间, 蛋白质分子质量介于 12.4~61.5 kDa 之间。各成员的理论等电点变化范围较大, 介于 4.99~9.88 之间, 其中第 亚家族成员理论等电点介于 8.38~9.36 之间, 在碱性范围内, 说明该家族成员蛋白质序列含有较多的碱性氨基酸, 第 、 、 、 亚家族蛋白质序列等电点变化均较大。通过分析烟草 C2H2 锌指蛋白编码基因的结构, 发现第 、 亚家族分别有 8 个和 3 个成员含有一个内含子, 其他成员没有内含子; 第 亚家族成员基因全部含有 2~3 个内含子; 第 亚家族成员全部含有 3~6 个内含子; 第 亚家

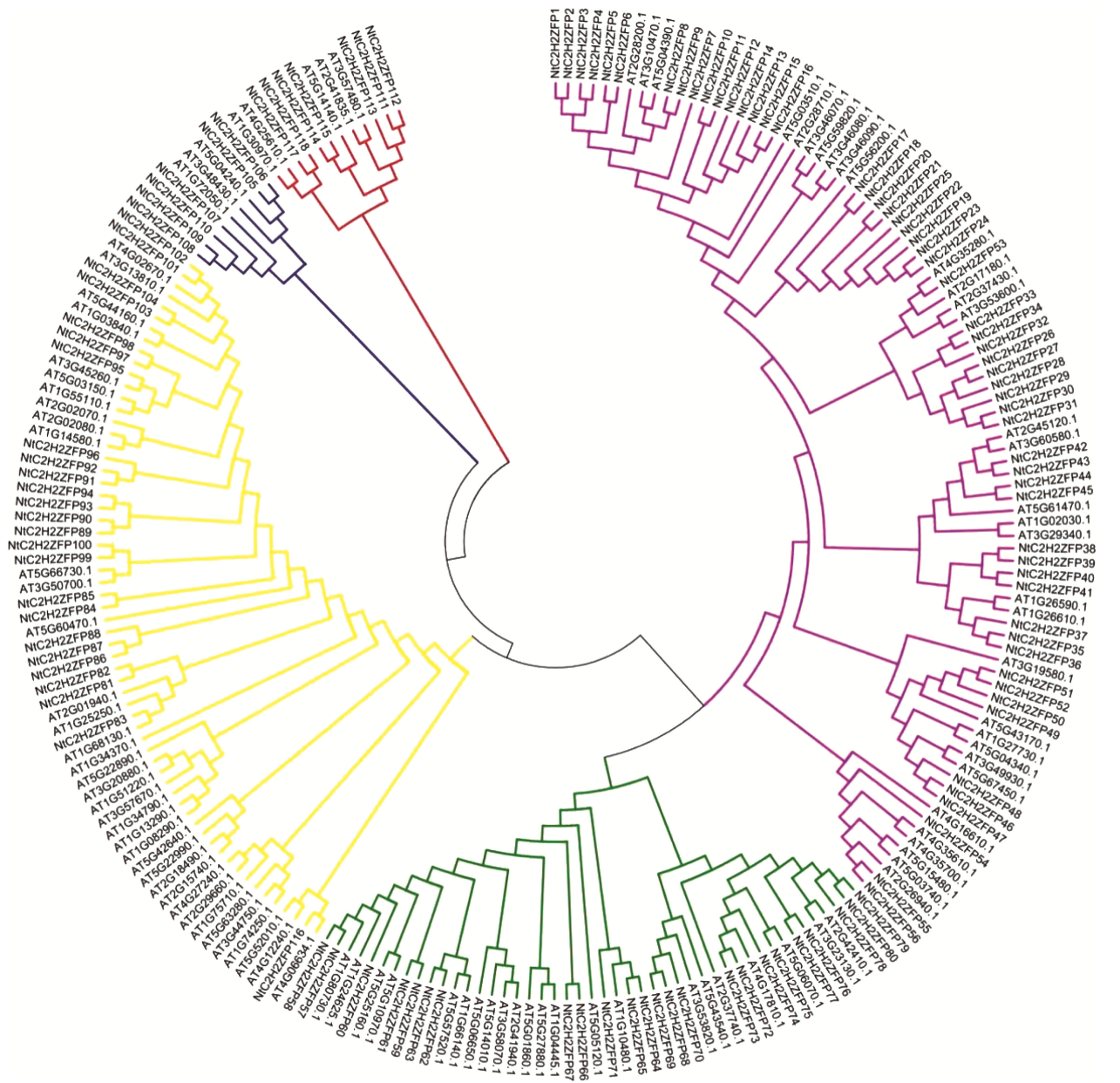


图 1 普通烟草与拟南芥 C2H2 锌指蛋白转录因子家族进化分析

Fig. 1 A phylogenetic tree of C2H2 zinc finger proteins transcription factor family in *N. tabacum* and *Arabidopsis*

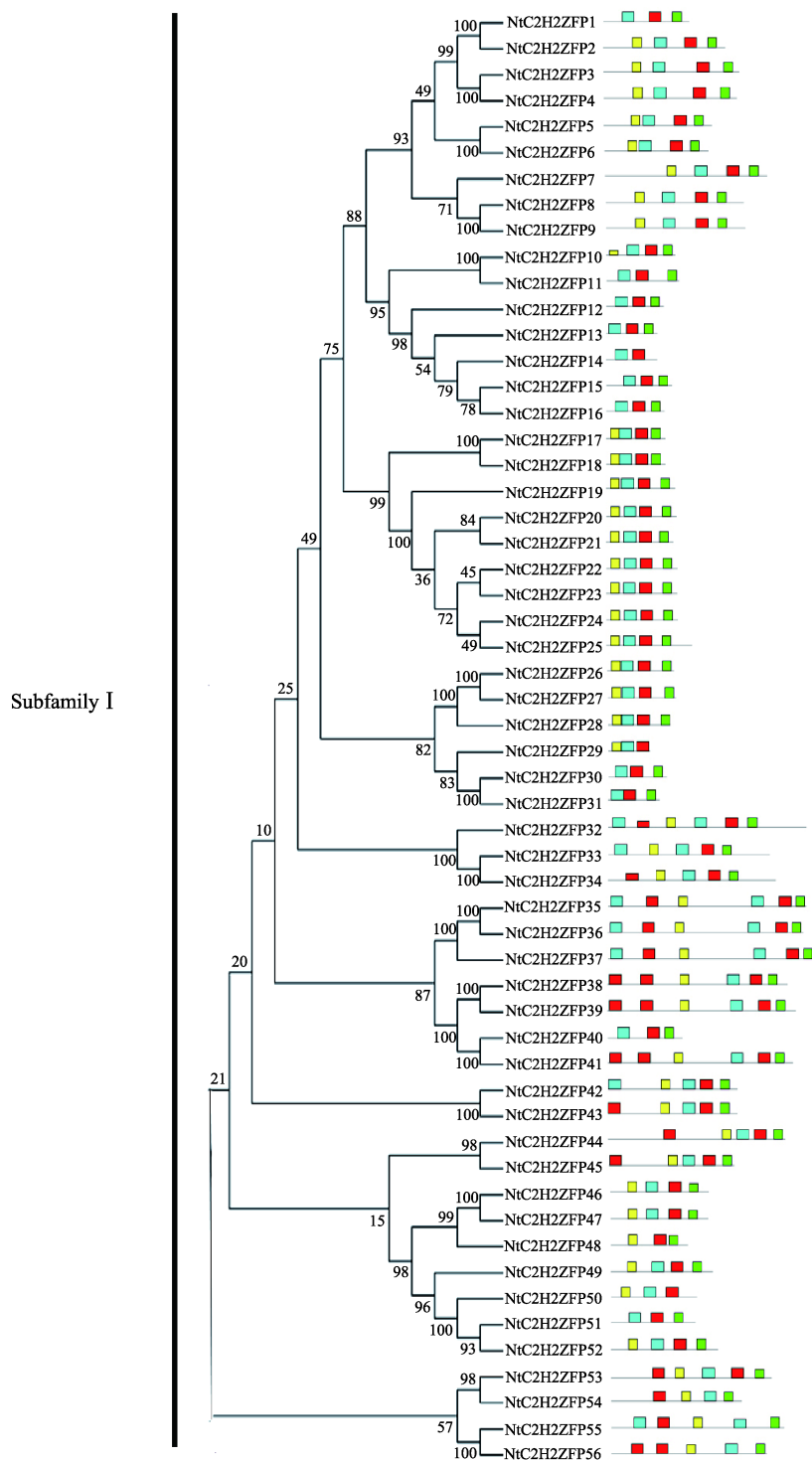
族各个成员基因的内含子数目差异较大, 在 0~6 个之间。亚细胞定位结果显示, 亚家族中的 *NtC2H2ZFP42* 和亚家族中的 *NtC2H2ZFP67* 定位于叶绿体, 亚家族中的 *NtC2H2ZFP43* 和亚家族中的 *NtC2H2ZFP78* 定位于细胞质, 亚家族中的 *NtC2H2ZFP112* 定位于液泡, 其他成员全部定位于细胞核。综上所述, 普通烟草同一亚家族蛋白成员及其编码基因在结构和特征上比较相似, 从侧面反映了系统发育分析的可靠性。

2.2 普通烟草 C2H2 锌指蛋白转录因子家族的结构域分析

利用 MEME 对 C2H2 锌指蛋白转录因子家族全长蛋白序列进行结构域分析可以发现, 所有家族成员蛋白质序列均含有 C2H2 结构域(Motif1 或/和 Motif3), 但不同亚家族所含结构域的种类和数量不同(图 1, 图 2)。第 1 亚家族所有成员都含有 Motif1 和 Motif3, 存在 2~4 个 C2H2 结构域, 绝大多数(除了两个成员外)

也含有 Motif6; 48 个成员含有 Motif5; 同一分支成员的结构域构成也具有较高的一致性。第 Ⅰ、Ⅱ 亚家族成员都只含有 1 个 C2H2 结构域(Motif1), 第 Ⅲ 亚家族的 *NtC2H2ZFP76*、*NtC2H2ZFP77*、*NtC2H2ZFP78*、*NtC2H2ZFP79* 和 *NtC2H2ZFP* 除了含有 Motif1, 还含有 Motif6; 第 Ⅳ 亚家族中还含有 Motif2 及 Motif4,

这两个结构域是该亚家族独有的。第 Ⅴ 亚家族各成员都只含有 C2H2 结构域, 但在数量上存在较大差异。综上所述, 普通烟草 C2H2 锌指蛋白转录因子家族所有成员除了预测出功能确定的 C2H2 结构域 (Motif1、Motif3) 以外, 还有 Motif2、Motif4 等一些新的结构域被预测出来, 同一个亚家族成员之间所含



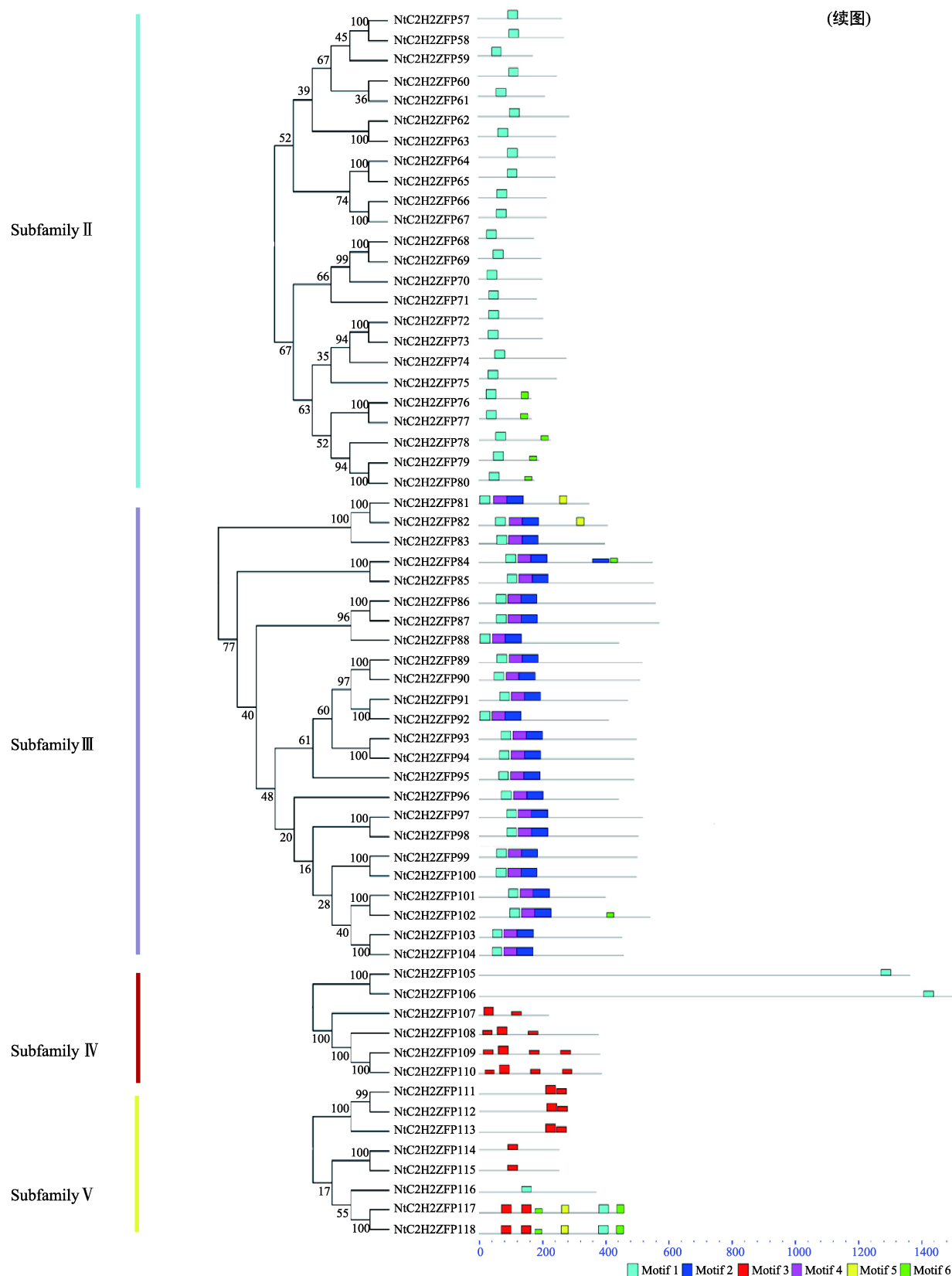


图 2 普通烟草 C2H2 锌指蛋白亚家族进化树及结构域分布

Fig. 2 Phylogenetic trees and domain distribution of subfamilies of C2H2 zinc finger proteins in *N. tabacum*

结构域的数量、种类及分布情况具有高度的一致性。

2.3 普通烟草 C2H2 锌指蛋白保守结构域及其他结构域的序列特征分析

对拟南芥 C2H2 锌指蛋白序列保守结构域的序列分析发现, 多数成员蛋白质序列含有 X2-Cys-X(2-4)-Cys-X12-His-X(3-5)-His 结构域, 即 C2H2 结构域, 该结构域中存在一段高度保守的基序 QALGGH^[19], 其对于结合下游靶 DNA 序列具有至关重要的作用, 也是植物 C2H2 锌指蛋白所特有的氨基酸序列^[20]。通过多序列比对分析发现, 普通烟草 C2H2 锌指蛋白转录因子家族所有成员均含有该保守结构域 X2-Cys-X(2-4)-Cys-X12-His-X(3-5)-His, Motif1、Motif3 都属于 C2H2 结构域, 但任意氨基酸序列 X 存在较大差异(图 3A)。

蛋白质三维结构模板(图 3B)中标出的两个绿色部分为保守结构域 C2H2, 从三维结构模板中可以看出这两个结构域之间距离较远, 两者之间氨基酸数目较多, 此特征是植物所特有的^[1]。绿色部分三维空间上的中心位置为 Zn^{2+} 结合位点, Zn^{2+} 与两个 His 和两个 Cys 结合以维持整个锌指结构的稳定, 并使螺旋结构正常存在^[21]。

利用 MEME 软件对全长蛋白质序列预测, 得到 4 个功能未知的结构域(图 4), 对每个结构域进行分析发现, Motif2 中含有 50 个比较保守的氨基酸序列, 保守性相对较高。Motif4 中含有的保守氨基酸数目比 Motif2 少, 但其氨基酸序列保守性同样比较高。Motif5 与 Motif6 都只含有 21 个氨基酸序列, 只有少数位置上的氨基酸保守性较高。

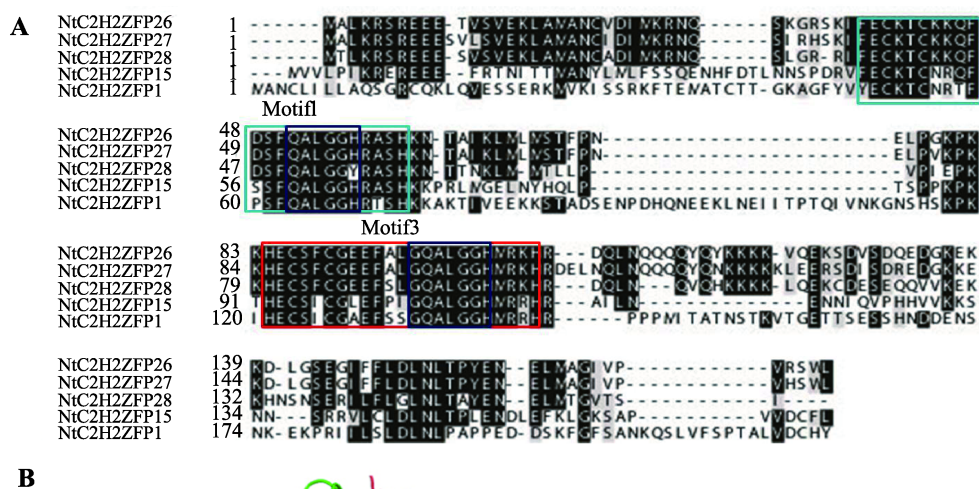
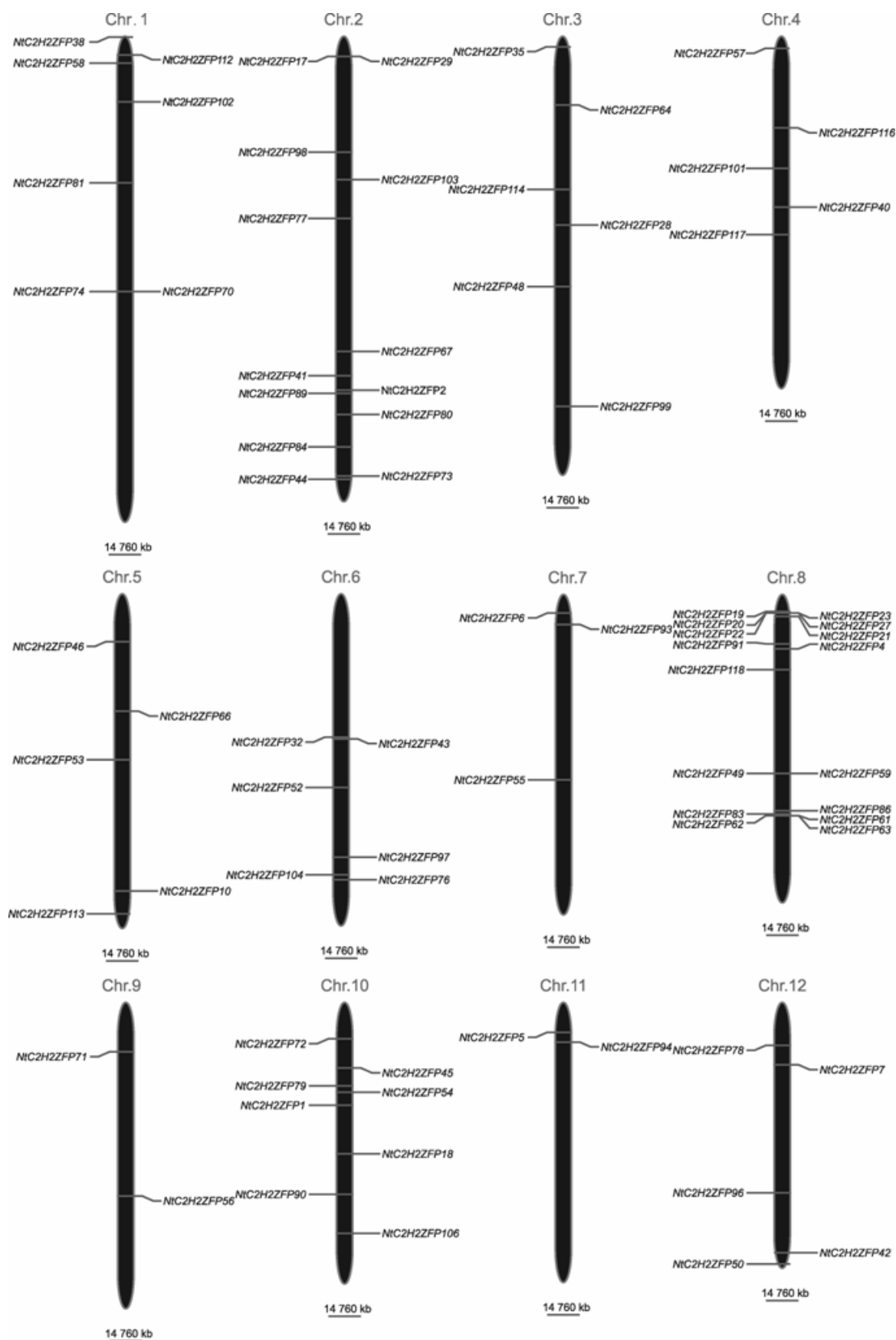


图 3 普通烟草 C2H2 锌指蛋白保守结构域分析

Fig. 3 Conserved domain analysis of C2H2 zinc finger proteins in *N. tabacum*

A: 多序列比对; B: 蛋白质三维结构。图中淡蓝色和红色框区域分别为 motif1 与 motif3 保守结构域, 深蓝色框区域为 C2H2 特有的保守基序。



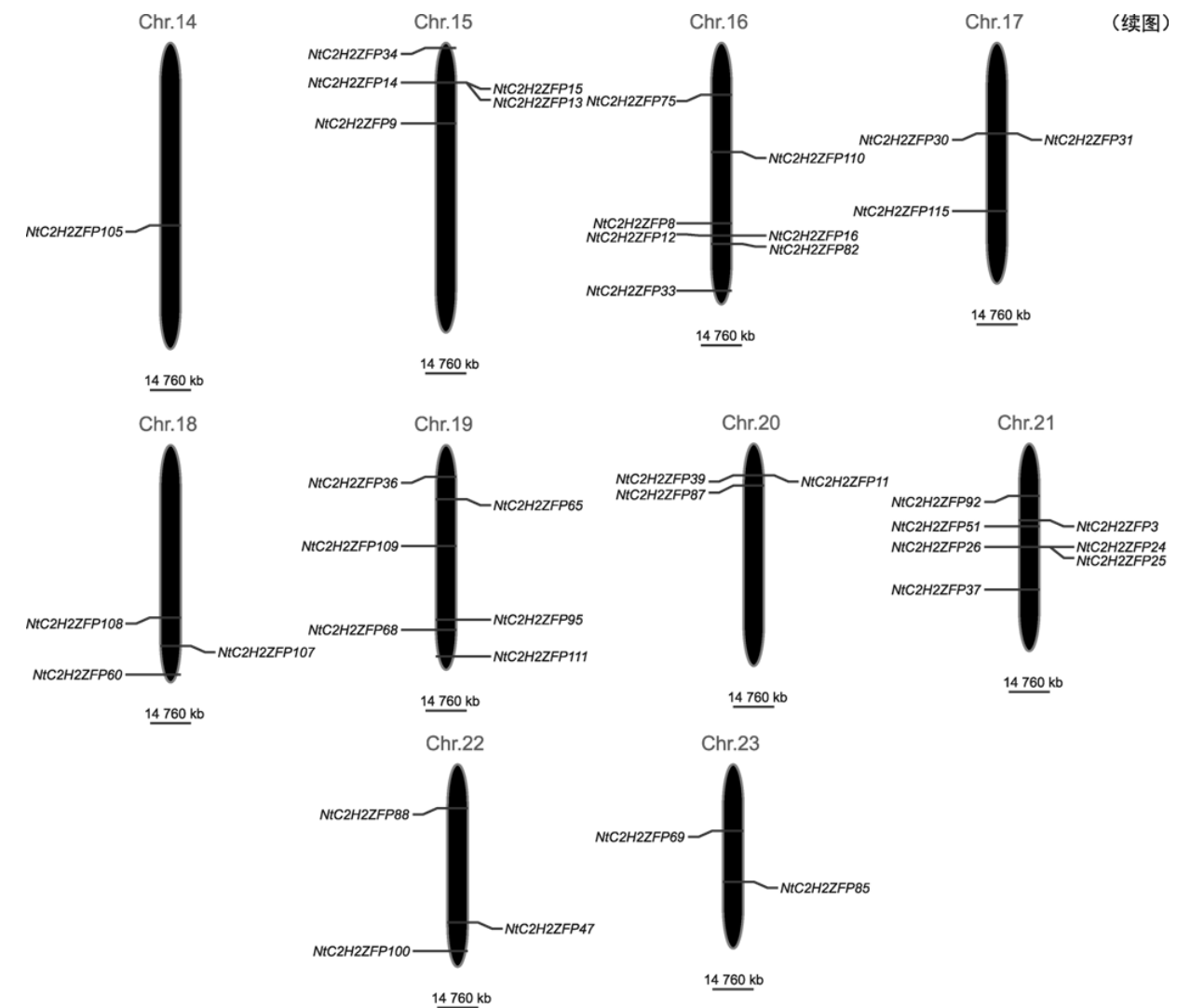


图 5 普通烟草 C2H2 锌指蛋白转录因子家族染色体定位
Fig. 5 Chromosomal locations of C2H2 zinc finger protein transcription factor family in *N. tabacum*

表 1 普通烟草各个亚家族成员分布情况

Table 1 Distribution analysis of every subfamily members in *N. tabacum*

染色体	亚家族 I	亚家族 II	亚家族 III	亚家族 IV	亚家族 V	染色体	亚家族 I	亚家族 II	亚家族 III	亚家族 IV	亚家族 V
1 号	1	3	2	1	0	12 号	3	1	1	0	0
2 号	5	4	4	0	0	14 号	0	0	0	1	0
3 号	3	1	1	0	1	15 号	5	0	0	0	0
4 号	1	1	1	0	2	16 号	4	1	1	1	0
5 号	3	1	0	0	1	17 号	2	0	0	0	1
6 号	3	1	2	0	0	18 号	0	1	0	2	0
7 号	2	0	1	0	0	19 号	1	2	1	1	1
8 号	8	4	3	0	1	20 号	2	0	1	0	0
9 号	1	1	0	0	0	21 号	6	0	1	0	0
10 号	4	2	1	1	0	22 号	1	0	2	0	0
11 号	1	0	1	0	0	23 号	0	1	1	0	0

2.5 普通烟草 C2H2 锌指蛋白转录因子家族的组织表达分析

从普通烟草高通量转录组测序数据库提取 C2H2 锌指蛋白转录因子家族数据, 利用 R 语言将数据进行可视化, 绘制热图(图 6)。由图 6 可知, C2H2 锌指蛋白转录因子家族中仅 10 个成员在候选检测的 10 个组织中均未表达外, 其余 108 个成员分别在不同的生长时期的不同组织中有表达, 表明 C2H2 锌指蛋白转录因子家族成员广泛参与了植物生长的各个生命过程^[12~14], 在不同的组织中均有表达^[22]。在各个组织中有表达的基因数目如下: 在种子有表达的家族成员有 94 个, 叶脉为 82 个, 茎中有 76 个, 幼根中有 94 个, 幼叶中有 80 个, 成熟根中有 92 个, 成熟叶中有 72 个, 花中有 91 个, 离体叶片中有 82 个, 愈伤组织中有 90 个。各个组织中都有较多的 C2H2 锌指蛋白转录因子家族成员在该组织的生长代谢过程中起作用。在不同的组织中各个成员的表达情况也明显不同, 具有一定的组织特异性。表达量较高的基因家族成员多存在于幼叶、离体叶、叶脉、幼根及成熟根中。进一步分析发现, 在幼叶中表达量较高的家族成员分别在成熟叶、离体叶中都有较高的表达量, 如 *NtC2H2ZFP84*、*NtC2H2ZFP85*、*NtC2H2ZFP86*、*NtC2H2ZFP87*、*NtC2H2ZFP88*、*NtC2H2ZFP95* 和 *NtC2H2ZFP102* 等均在幼叶、成熟叶、离体叶中有较高的表达量, 这 7 个基因同属于亚家族 I, 表明 C2H2 锌指蛋白转录因子家族在烟草叶片的发生中具有一定的作用。同时发现离体 24 h 的叶组织中 *NtC2H2ZFP52*、*NtC2H2ZFP47* 和 *NtC2H2ZFP46* 表达量非常高, 这 3 个基因同属于亚家族 II, 初步推测这 3 个基因可能与非生物胁迫有关^[23,24]。分析在幼根和成熟根中表达的基因发现, 在幼根中表达量较高的基因在成熟根中绝大多数也有表达而且表达量相对较高, 但相比在幼根中的表达量较低。同一亚家族中的成员在不同组织中的表达情况差异较大, 只有亚家族 IV 所有成员在各个组织中都有表达。

3 讨论

C2H2 锌指蛋白转录因子家族编码的转录因子

具有多种生物学功能, 其不仅与 DNA、RNA 相互作用还与蛋白质相互作用, 广泛参与了植物叶的发生、侧枝的形成、非生物逆境胁迫响应及花器官的形成等多种生命过程。

本研究运用生物信息学手段, 首次从普通烟草基因组中鉴定了 118 条 C2H2 锌指蛋白转录因子家族成员, 每个成员的蛋白质长度差别显著, 但是同一亚家族各成员蛋白质长度差异相对较小。对该基因家族各个成员的保守结构域的分析表明, 每个成员都包含了至少一个 C2H2 锌指结构域, 该结构域中存在植物特有的 QALGGH 基序, 该基序对于 C2H2 转录因子蛋白结合 DNA 序列具有重要的作用^[10,11]。两个保守结构域之间相隔较远。分析结果表明, 同一亚家族中的各个成员所含有结构域的数目和种类高度一致, 说明同一亚家族各个成员亲缘关系较近, 蛋白质序列相似度较高, 但不同亚家族之间差异明显, 亚家族 I 中含有的 Motif2 及 Motif4 在其他亚家族都不存在, 并且这两个结构域与 C2H2 保守结构域紧密相连, 表明不同亚家族成员可能具有不同的生物学功能。对各个亚家族成员的内含子数目统计分析, 同一亚家族内含子数目相对一致, 基因结构比较相似, 亚家族 I 中多不含有内含子, 说明 C2H2 锌指蛋白转录因子家族在进化上相对保守。根据 WoLF PSORT 在线工具对各个基因家族成员的亚细胞定位分析, 90% 的基因家族成员被预测在细胞核中起作用, 仅少数被预测在线粒体、液泡、叶绿体上起作用, 但是具体定位情况还需要分子生物学实验来验证。对 C2H2 锌指蛋白转录因子家族所有成员染色体定位分析发现, C2H2 锌指蛋白转录因子家族成员分布较广泛, 同一亚家族成员集中分布在某条染色体上的情况较少, 多分散于不同的染色体上, 这也从侧面说明 C2H2 锌指蛋白转录因子家族成员作用广泛。通过对普通烟草 C2H2 锌指蛋白转录因子家族表达分析发现, 该基因家族各个成员分别在不同组织均有表达, 说明该基因家族成员广泛参与了普通烟草的多种生物学进程, 但不同组织中的各成员表达量明显不同, 具有一定的组织特异性。

NtC2H2ZFP84、*NtC2H2ZFP85*、*NtC2H2ZFP86*、*NtC2H2ZFP87*、*NtC2H2ZFP88*、*NtC2H2ZFP95* 和

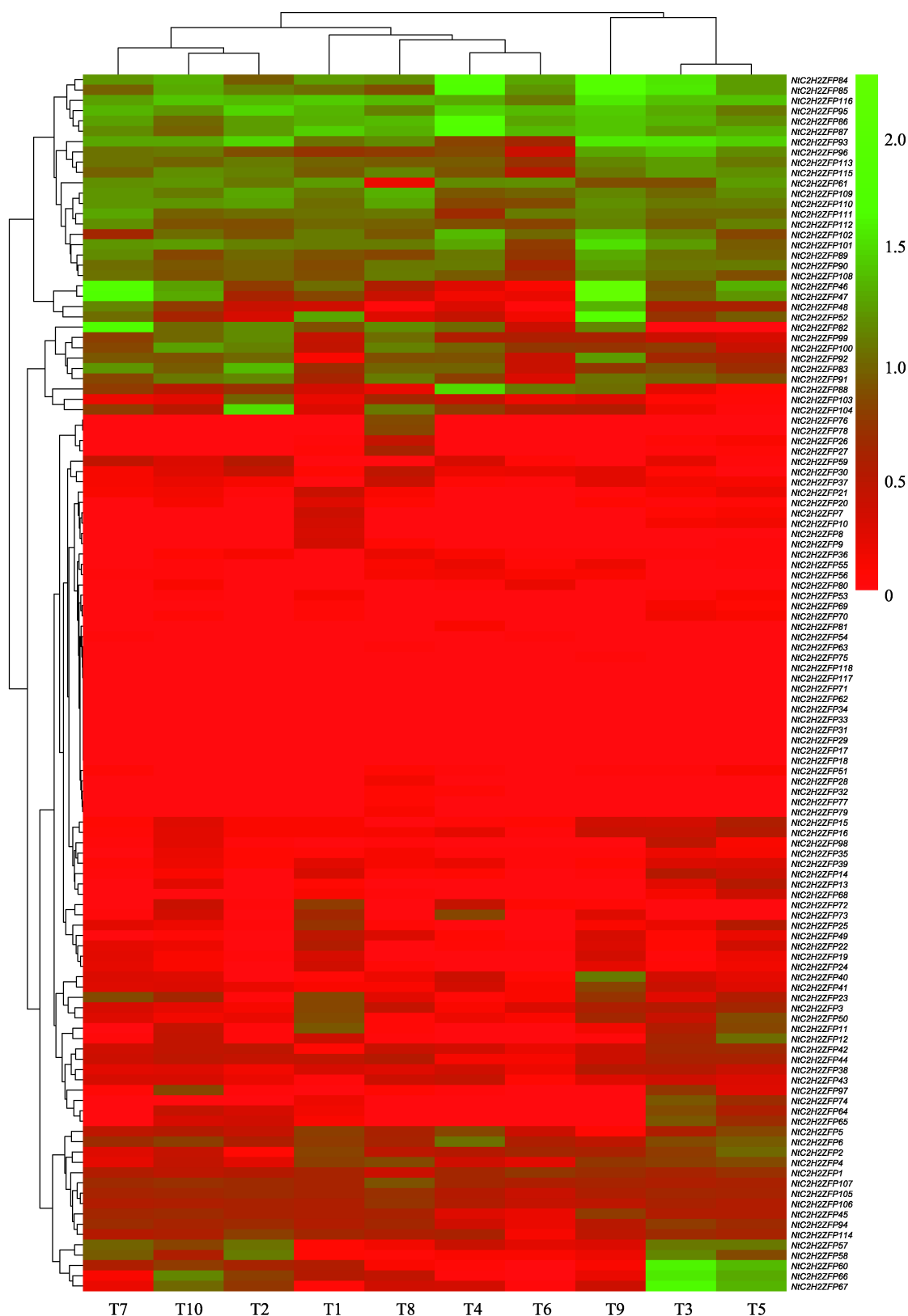


图 6 普通烟草 C2H2 锌指蛋白转录因子家族组织表达模式

Fig. 6 Tissue expression patterns of the C2H2 zinc finger protein transcription factor gene family in *N. tabacum*

T1: 种子; T2: 茎; T3: 幼根; T4: 幼叶; T5: 成熟根; T6: 成熟叶; T7: 叶脉; T8: 花芽; T9: 叶片(离体 24 h); T10: 愈伤组织。

NtC2H2ZFP102 这 7 个基因在幼叶、成熟叶、离体叶中均有较高的表达量, 并且同属于亚家族, 推测该基因家族的这些成员可能对叶的发生起到一定的作用。在离体叶片中, *NtC2H2ZFP52*、*NtC2H2ZFP47* 和 *NtC2H2ZFP46* 这 3 个基因具有非常高的表达量, 推测它们可能参与了非生物胁迫的响应, 但具体生物学功能还需要分子生物学实验验证。

附录:

附表见网络电子版 www.chinagene.cn。

参考文献(References):

- [1] Takatsuji H. Zinc-finger proteins: the classical zinc finger emerges in contemporary plant science. *Plant Mol Biol*, 1999, 39(6): 1073–1078. [DOI]
- [2] Laity JH, Lee BM, Wright PE. Zinc finger proteins: new insights into structural and functional diversity. *Curr Opin Struc Biol*, 2001, 11(1): 39–46. [DOI]
- [3] Hu W, Ma H. Characterization of a novel putative zinc finger gene *MIF1*: involvement in multiple hormonal regulation of *Arabidopsis* development. *Plant J*, 2006, 45(3): 399–422. [DOI]
- [4] Moore M, Ullman C. Recent developments in the engineering of zinc finger proteins. *Brief Funct Genomic Proteomic*, 2003, 1(4): 342–355. [DOI]
- [5] Berg JM, Shi YG. The galvanization of biology: a growing appreciation for the roles of zinc. *Science*, 1996, 271(5252): 1081–1085. [DOI]
- [6] Pabo CO, Peisach E, Grant RA. Design and selection of novel Cys₂His₂ zinc finger proteins. *Annu Rev Biochem*, 2001, 70: 313–340. [DOI]
- [7] Sakamoto H, Maruyama K, Sakuma Y, Meshi T, Iwabuchi M, Shinozaki K, Yamaguchi-Shinozaki K. *Arabidopsis* Cys₂/His₂-type zinc-finger proteins function as transcription repressors under drought, cold, and high-salinity stress conditions. *Plant Physiol*, 2004, 136(1): 2734–2746. [DOI]
- [8] Klug A, Schwabe JW. Protein motifs 5. Zinc fingers. *FASEB J*, 1995, 9(8): 597–604. [DOI]
- [9] Ciftci-Yilmaz S, Mittler R. The zinc finger network of plants. *Cell Mol Life Sci*, 2008, 65(7–8): 1150–1160. [DOI]
- [10] Gamsjaeger R, Liew CK, Loughlin FE, Crossley M, Mackay JP. Sticky fingers: zinc-fingers as protein-recognition motifs. *Trends Biochem Sci*, 2007, 32(2): 63–70. [DOI]
- [11] Lee BM, Xu J, Clarkson BK, Martinez-Yamout MA, Dyson HJ, Case DA, Gottesfeld JM, Wright PE. Induced fit and "Lock and Key" recognition of 5 S RNA by zinc fingers of transcription factor IIIA. *J Mol Biol*, 2006, 357(1): 275–291. [DOI]
- [12] Takatsuji H, Mori M, Benfey PN, Ren L, Chua NH. Characterization of a zinc finger DNA-binding protein expressed specifically in *Petunia* petals and seedlings. *EMBO J*, 1992, 11(1): 241–249. [DOI]
- [13] Guo SQ, Huang J, Jiang Y, Zhang HS. Cloning and characterization of *RZF71* encoding a C2H2-type zinc finger protein from rice. *Hereditas (Beijing)*, 2007, 29(5): 607–613. 郭书巧, 黄骥, 江燕, 张红生. 水稻 C2H2 型锌指蛋白基因 *RZF71* 的克隆与表达分析. *遗传*, 2007, 29(5): 607–613. [DOI]
- [14] Sakamoto H, Araki T, Meshi T, Iwabuchi M. Expression of a subset of the *Arabidopsis* Cys₂/His₂-type zinc-finger protein gene family under water stress. *Gene*, 2000, 248(1–2): 23–32. [DOI]
- [15] Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, de Castro E, Duvaud S, Flegel V, Fortier A, Gasteiger E, Grosdidier A, Hernandez C, Ioannidis V, Kuznetsov D, Liechti R, Moretti S, Mostaguir K, Redaschi N, Rossier G, Xenarios I, Stockinger H. ExPASy: SIB bioinformatics resource portal. *Nucl Acids Res*, 2012, 40(W1): W597–W603. [DOI]
- [16] Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai K. WoLF PSORT: protein localization predictor. *Nucl Acids Res*, 2007, 35(S2): W585–W587. [DOI]
- [17] Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, 2004, 5(1): 113. [DOI]
- [18] Zhang T, Zhang SX, Pei LL, Yu TF, Chen M, Li LC, Zhou YB, Ma YZ, Xu Q. Genome-wide analysis of ANK gene family and expression pattern of the *ANK25* gene in snap bean. *J Plant Gen Resour*, 2014, 15(6): 1334–1341. 张婷, 张双喜, 裴丽丽, 于太飞, 陈明, 李连城, 周永斌, 马有志, 徐强. 菜豆 ANK 基因家族鉴定及 *ANK25* 的表达模式分析. *植物遗传资源学报*, 2014, 15(6): 1334–1341. [DOI]
- [19] Englbrecht CC, Schoof H, Böhm S. Conservation, diversification and expansion of C2H2 zinc finger proteins in the *Arabidopsis thaliana* genome. *BMC Genomics*, 2004, 5(1): 39. [DOI]
- [20] Takatsuji H. Zinc-finger transcription factors in plants. *Cell Mol Life Sci*, 1998, 54(6): 582–596. [DOI]
- [21] Wolfe SA, Neklodova L, Pabo CO. DNA recognition by Cys₂His₂ zinc finger proteins. *Annu Rev Biophys Biomol Struct*, 2000, 29: 183–212. [DOI]
- [22] Huang F, Chi YJ, Meng QC, Gai JY, Yu DY. *GmZFP1* encoding a single zinc finger protein is expressed with enhancement in reproductive organs and late seed development in soybean (*Glycine max*). *Mol Biol Rep*, 2006, 33(4): 279–285. [DOI]
- [23] Mukhopadhyay A, Vij S, Tyagi AK. Overexpression of a zinc-finger protein gene from rice confers tolerance to cold, dehydration, and salt stress in transgenic tobacco. *Proc Natl Acad Sci USA*, 2004, 101(16): 6309–6314. [DOI]
- [24] Colasanti J, Yuan Z, Sundaresan V. The *indeterminate* gene encodes a zinc finger protein and regulates a leaf-generated signal required for the transition to flowering in maize. *Cell*, 1998, 93(4): 593–603. [DOI]

(责任编辑: 张红生)

附表1 普通烟草C2H2锌指蛋白家族理化特征分析

Supplementary Table 1 Characteristic analysis of the C2H2 zinc finger protein family in *N. tabacum*

基因名	亚家族	氨基酸数目	分子量(Da)	等电点	内含子数目	亚细胞定位	拟南芥同源基因	序列号
NtC2H2ZFP1	I	219	24186.2	8.58	0	Nucl	AT2G28200	Ntab-TN90_AYMY-SS14403
NtC2H2ZFP2	I	313	33949.6	7.13	0	Nucl	AT2G28200	Ntab-TN90_AYMY-SS1377
NtC2H2ZFP3	I	348	37876.9	7.65	0	Nucl	AT5G03510	Ntab-TN90_AYMY-SS1153
NtC2H2ZFP4	I	340	36996.8	6.42	0	Nucl	AT5G03510	Ntab-TN90_AYMY-SS17043
NtC2H2ZFP5	I	277	30086.3	7.63	0	Nucl	AT5G03510	Ntab-TN90_AYMY-SS17878
NtC2H2ZFP6	I	265	28911	7.61	0	Nucl	AT5G03510	Ntab-TN90_AYMY-SS238584
NtC2H2ZFP7	I	416	46448.5	5.48	0	Nucl	AT5G04390	Ntab-TN90_AYMY-SS11287
NtC2H2ZFP8	I	354	39796.2	7.13	0	Nucl	AT5G04390	Ntab-TN90_AYMY-SS789
NtC2H2ZFP9	I	356	40006.2	7.16	0	Nucl	AT5G04390	Ntab-TN90_AYMY-SS691
NtC2H2ZFP10	I	177	20611.6	8.85	0	Nucl	AT3G53600	Ntab-TN90_AYMY-SS196048
NtC2H2ZFP11	I	186	21310.2	9.1	0	Nucl	AT3G53600	Ntab-TN90_AYMY-SS10391
NtC2H2ZFP12	I	147	16639.1	9.5	0	Nucl	AT3G53600	Ntab-TN90_AYMY-SS15135
NtC2H2ZFP13	I	130	14624.8	9.38	0	Nucl	AT3G53600	Ntab-TN90_AYMY-SS142426
NtC2H2ZFP14	I	131	15029.3	9.65	0	Nucl	AT2G37430	Ntab-TN90_AYMY-SS13084
NtC2H2ZFP15	I	168	19264.1	8.99	0	Nucl	AT2G37430	Ntab-TN90_AYMY-SS694
NtC2H2ZFP16	I	148	16704.1	9.19	0	Nucl	AT2G37430	Ntab-TN90_AYMY-SS15135
NtC2H2ZFP17	I	151	16938.4	9.53	0	Nucl	AT2G37430	Ntab-TN90_AYMY-SS508
NtC2H2ZFP18	I	151	16938.4	9.53	0	Nucl	AT2G37430	Ntab-TN90_AYMY-SS509
NtC2H2ZFP19	I	177	19771	9.73	0	Nucl	AT2G37430	Ntab-TN90_AYMY-SS10312
NtC2H2ZFP20	I	179	20419.5	9.62	0	Nucl	AT2G37430	Ntab-TN90_AYMY-SS7910
NtC2H2ZFP21	I	173	19669.6	9.6	0	Nucl	AT2G37430	Ntab-TN90_AYMY-SS15863
NtC2H2ZFP22	I	181	20380.5	9.56	0	Nucl	AT2G37430	Ntab-TN90_AYMY-SS7912
NtC2H2ZFP23	I	181	20158.4	9.46	1	Nucl	AT2G37430	Ntab-TN90_AYMY-SS3197
NtC2H2ZFP24	I	183	20394.5	9.58	0	Nucl	AT2G37430	Ntab-TN90_AYMY-SS6996
NtC2H2ZFP25	I	220	24888.4	8.71	0	Nucl	AT2G37430	Ntab-TN90_AYMY-SS16418
NtC2H2ZFP26	I	171	19571.4	9.15	1	Nucl	AT4G35280	Ntab-TN90_AYMY-SS16419
NtC2H2ZFP27	I	176	20293.3	8.92	0	Nucl	AT4G35280	Ntab-TN90_AYMY-SS7913
NtC2H2ZFP28	I	161	18489.3	9.25	0	Nucl	AT4G35280	Ntab-TN90_AYMY-SS2413
NtC2H2ZFP29	I	107	12423.4	9.83	0	Nucl	AT4G35280	Ntab-TN90_AYMY-SS508
NtC2H2ZFP30	I	148	17098.7	9.68	0	Nucl	AT4G35280	Ntab-TN90_AYMY-SS17924
NtC2H2ZFP31	I	132	14966.2	9.2	1	Nucl	AT4G35280	Ntab-TN90_AYMY-SS17925
NtC2H2ZFP32	I	507	57194.1	6.12	0	Nucl	AT4G35280	Ntab-TN90_AYMY-SS1926
NtC2H2ZFP33	I	412	45822.6	5.95	1	Nucl	AT4G35280	Ntab-TN90_AYMY-SS12431
NtC2H2ZFP34	I	431	48129	6	1	Nucl	AT4G35280	Ntab-TN90_AYMY-SS10509
NtC2H2ZFP35	I	508	57834.2	6.32	0	Nucl	AT1G26610	Ntab-TN90_AYMY-SS1526
NtC2H2ZFP36	I	501	56788.2	8.12	0	Nucl	AT1G26610	Ntab-TN90_AYMY-SS1241
NtC2H2ZFP37	I	529	60362.3	8.93	0	Nucl	AT1G26610	Ntab-TN90_AYMY-SS1137
NtC2H2ZFP38	I	462	52322.7	8.72	0	Nucl	AT1G02030	Ntab-TN90_AYMY-SS5103
NtC2H2ZFP39	I	483	54345.9	8.76	0	Nucl	AT1G02030	Ntab-TN90_AYMY-SS15282
NtC2H2ZFP40	I	191	21351.8	7.11	0	Nucl	AT1G02030	Ntab-TN90_AYMY-SS1348

续附表1

基因名	亚家族	氨基酸数目	分子量(Da)	等电点	内含子数目	亚细胞定位	拟南芥同源基因	序列号
NtC2H2ZFP41	I	475	53832.3	8.73	0	Nucl	AT1G26610	Ntab-TN90_AYMY-SS8145
NtC2H2ZFP42	I	332	37331.5	5.71	0	Chlol	AT3G60580	Ntab-TN90_AYMY-SS16420
NtC2H2ZFP43	I	332	37260.5	5.58	0	Cyto	AT3G60580	Ntab-TN90_AYMY-SS1926
NtC2H2ZFP44	I	454	51499.2	6.37	0	Nucl	AT4G16610	Ntab-TN90_AYMY-SS7921
NtC2H2ZFP45	I	321	35973.4	8.02	1	Nucl	AT3G60580	Ntab-TN90_AYMY-SS11692
NtC2H2ZFP46	I	253	27033	7.65	0	Nucl	AT5G43170	Ntab-TN90_AYMY-SS8071
NtC2H2ZFP47	I	250	26766.8	8.14	0	Nucl	AT5G43170	Ntab-TN90_AYMY-SS14445
NtC2H2ZFP48	I	198	21658	7.66	0	Nucl	AT5G43170	Ntab-TN90_AYMY-SS1819
NtC2H2ZFP49	I	262	28564.7	8.86	1	Nucl	AT5G43170	Ntab-TN90_AYMY-SS28641
NtC2H2ZFP50	I	219	23575.9	9.03	0	Nucl	AT5G43170	Ntab-TN90_AYMY-SS171170
NtC2H2ZFP51	I	215	22360.8	8.99	1	Nucl	AT5G43170	Ntab-TN90_AYMY-SS171171
NtC2H2ZFP52	I	275	29090.5	8.48	0	Nucl	AT5G43170	Ntab-TN90_AYMY-SS6722
NtC2H2ZFP53	I	411	45368.1	7.52	0	Nucl	AT2G17180	Ntab-TN90_AYMY-SS998
NtC2H2ZFP54	I	334	36521.6	9.06	0	Nucl	AT4G16610	Ntab-TN90_AYMY-SS501
NtC2H2ZFP55	I	442	49195.8	6.55	0	Nucl	AT2G26940	Ntab-TN90_AYMY-SS2964
NtC2H2ZFP56	I	398	44563.4	9.16	0	Nucl	AT2G26940	Ntab-TN90_AYMY-SS18833
NtC2H2ZFP57	II	269	29533.7	8.91	0	Nucl	AT1G24625	Ntab-TN90_AYMY-SS15546
NtC2H2ZFP58	II	273	30183.3	8.74	0	Nucl	AT1G24625	Ntab-TN90_AYMY-SS16479
NtC2H2ZFP59	II	174	19293.7	9.88	0	Nucl	AT5G14010	Ntab-TN90_AYMY-SS10411
NtC2H2ZFP60	II	249	27976.8	8.85	0	Nucl	AT5G14010	Ntab-TN90_AYMY-SS3669
NtC2H2ZFP61	II	212	23501.6	7.14	1	Nucl	AT1G24625	Ntab-TN90_AYMY-SS11617
NtC2H2ZFP62	II	289	31885.1	7.74	0	Nucl	AT2G41940	Ntab-TN90_AYMY-SS11851
NtC2H2ZFP63	II	248	27328.1	8.38	0	Nucl	AT2G41940	Ntab-TN90_AYMY-SS11852
NtC2H2ZFP64	II	244	27831	9.11	1	Nucl	AT1G10480	Ntab-TN90_AYMY-SS1245
NtC2H2ZFP65	II	243	27450.3	8.63	0	Nucl	AT1G10480	Ntab-TN90_AYMY-SS1531
NtC2H2ZFP66	II	216	23460.9	7.78	1	Nucl	AT5G05120	Ntab-TN90_AYMY-SS3685
NtC2H2ZFP67	II	217	23424.9	8.44	0	Chlo	AT5G05120	Ntab-TN90_AYMY-SS9168
NtC2H2ZFP68	II	177	20023.4	5.06	0	Nucl	AT3G53820	Ntab-TN90_AYMY-SS17024
NtC2H2ZFP69	II	200	22529.4	5.74	0	Nucl	AT3G53820	Ntab-TN90_AYMY-SS4941
NtC2H2ZFP70	II	204	23012.4	5.42	0	Nucl	AT3G53820	Ntab-TN90_AYMY-SS759
NtC2H2ZFP71	II	186	21824.2	5.21	0	Nucl	AT3G53820	Ntab-TN90_AYMY-SS852
NtC2H2ZFP72	II	203	22473.5	4.99	0	Nucl	AT4G17810	Ntab-TN90_AYMY-SS11502
NtC2H2ZFP73	II	204	22664.9	5.3	0	Nucl	AT4G17810	Ntab-TN90_AYMY-SS11989
NtC2H2ZFP74	II	276	31511.1	8.8	0	Nucl	AT4G17810	Ntab-TN90_AYMY-SS16676
NtC2H2ZFP75	II	245	27133	9.85	0	Nucl	AT2G37740	Ntab-TN90_AYMY-SS12090
NtC2H2ZFP76	II	167	19235.7	9.56	0	Nucl	AT2G37740	Ntab-TN90_AYMY-SS8917
NtC2H2ZFP77	II	167	19292.9	9.56	0	Nucl	AT2G37740	Ntab-TN90_AYMY-SS16612
NtC2H2ZFP78	II	227	25450.6	9.3	0	Cyto	AT3G23130	Ntab-TN90_AYMY-SS408
NtC2H2ZFP79	II	192	22011.7	9.34	0	Nucl	AT3G23130	Ntab-TN90_AYMY-SS17250
NtC2H2ZFP80	II	177	20296.6	9.14	0	Nucl	AT3G23130	Ntab-TN90_AYMY-SS18018

续附表1

基因名	亚家族	氨基酸数目	分子量(Da)	等电点	内含子数目	亚细胞定位	拟南芥同源基因	序列号
NtC2H2ZFP81	III	352	40590.2	8.9	2	Nucl	AT1G68130	Ntab-TN90_AYMY-SS3893
NtC2H2ZFP82	III	409	46881.3	8.88	3	Nucl	AT1G68130	Ntab-TN90_AYMY-SS3727
NtC2H2ZFP83	III	399	44700.9	8.91	2	Nucl	AT2G01940	Ntab-TN90_AYMY-SS2188
NtC2H2ZFP84	III	550	60105.8	8.72	3	Nucl	AT2G02070	Ntab-TN90_AYMY-SS1361
NtC2H2ZFP85	III	555	61014.6	8.5	3	Nucl	AT2G02070	Ntab-TN90_AYMY-SS18129
NtC2H2ZFP86	III	561	61521.7	9.14	2	Nucl	AT2G02070	Ntab-TN90_AYMY-SS16886
NtC2H2ZFP87	III	572	61977.5	8.78	2	Nucl	AT2G02070	Ntab-TN90_AYMY-SS10052
NtC2H2ZFP88	III	444	49272.6	9.24	2	Nucl	AT2G02070	Ntab-TN90_AYMY-SS17945
NtC2H2ZFP89	III	518	56873.5	9.26	2	Nucl	AT5G03150	Ntab-TN90_AYMY-SS9247
NtC2H2ZFP90	III	511	56063.7	9.12	2	Nucl	AT5G03150	Ntab-TN90_AYMY-SS5641
NtC2H2ZFP91	III	472	52168.3	9.22	3	Nucl	AT5G03150	Ntab-TN90_AYMY-SS16141
NtC2H2ZFP92	III	411	45750.2	9.36	2	Nucl	AT5G03150	Ntab-TN90_AYMY-SS15939
NtC2H2ZFP93	III	500	55287.3	9.36	2	Nucl	AT5G03150	Ntab-TN90_AYMY-SS84
NtC2H2ZFP94	III	493	54292	9.3	2	Nucl	AT5G03150	Ntab-TN90_AYMY-SS588
NtC2H2ZFP95	III	492	54083.5	9.15	2	Nucl	AT5G03150	Ntab-TN90_AYMY-SS1773
NtC2H2ZFP96	III	443	49482.2	8.7	2	Nucl	AT3G45260	Ntab-TN90_AYMY-SS459
NtC2H2ZFP97	III	520	57087.7	9.24	3	Nucl	AT3G45260	Ntab-TN90_AYMY-SS13947
NtC2H2ZFP98	III	505	55506.6	8.8	3	Nucl	AT3G45260	Ntab-TN90_AYMY-SS1410
NtC2H2ZFP99	III	501	53718.2	8.53	3	Nucl	AT1G55110	Ntab-TN90_AYMY-SS5202
NtC2H2ZFP100	III	500	53681.3	8.38	3	Nucl	AT1G55110	Ntab-TN90_AYMY-SS8415
NtC2H2ZFP101	III	402	44935	6.44	3	Nucl	AT1G55110	Ntab-TN90_AYMY-SS10544
NtC2H2ZFP102	III	542	58889.7	8.77	2	Nucl	AT1G55110	Ntab-TN90_AYMY-SS3885
NtC2H2ZFP103	III	452	50359.2	8.92	2	Nucl	AT4G02670	Ntab-TN90_AYMY-SS5917
NtC2H2ZFP104	III	457	50700.7	8.95	2	Nucl	AT4G02670	Ntab-TN90_AYMY-SS5918
NtC2H2ZFP105	IV	1365	152809.2	6.58	5	Nucl	AT5G04240	Ntab-TN90_AYMY-SS1669
NtC2H2ZFP106	IV	1502	167713.6	6.51	6	Nucl	AT5G04240	Ntab-TN90_AYMY-SS1321
NtC2H2ZFP107	IV	222	25439.5	9.06	3	Nucl	AT3G29340	Ntab-TN90_AYMY-SS18420
NtC2H2ZFP108	IV	379	44159.6	8.61	5	Nucl	AT3G29340	Ntab-TN90_AYMY-SS16618
NtC2H2ZFP109	IV	385	44607	8.66	5	Nucl	AT3G29340	Ntab-TN90_AYMY-SS17392
NtC2H2ZFP110	IV	390	45302.9	8.83	5	Nucl	AT3G29340	Ntab-TN90_AYMY-SS4184
NtC2H2ZFP111	V	279	31120.4	8.78	1	Nucl	AT3G57480	Ntab-TN90_AYMY-SS7236
NtC2H2ZFP112	V	281	31168.8	9.06	2	Vacu	AT3G57480	Ntab-TN90_AYMY-SS312
NtC2H2ZFP113	V	278	31247.6	8.78	1	Nucl	AT3G57480	Ntab-TN90_AYMY-SS1022
NtC2H2ZFP114	V	255	28886.7	9.33	3	Nucl	AT5G14140	Ntab-TN90_AYMY-SS7680
NtC2H2ZFP115	V	255	28841.6	9.25	3	Nucl	AT5G14140	Ntab-TN90_AYMY-SS7683
NtC2H2ZFP116	V	372	42388	6.24	5	Nucl	AT4G06634	Ntab-TN90_AYMY-SS18625
NtC2H2ZFP117	V	456	51458.6	6.66	0	Nucl	AT2G26940	Ntab-TN90_AYMY-SS561
NtC2H2ZFP118	V	456	51458.6	6.66	0	Nucl	AT2G26940	Ntab-TN90_AYMY-SS562