

蓖麻 *GATA* 基因家族的鉴定和特征分析*敖涛^{1,3}, 廖晓佳^{2,3}, 徐伟^{2,3}, 刘爱忠^{1**}

(1 中国科学院西双版纳热带植物园热带植物资源开放实验室, 云南 勐腊 666303; 2 中国科学院昆明植物研究所, 昆明 650201; 3 中国科学院大学, 北京 100049)

摘要: *GATA* 因子在植物中广泛参与了光响应调控、叶绿素合成以及碳、氮代谢等与作物产量息息相关的生物学过程。本研究基于蓖麻基因组信息, 鉴定了 19 个蓖麻 *GATA* 家族成员, 根据基因结构、蛋白结构和系统进化分析将其分为 4 个亚家族。理化性质分析发现 *GATA* 蛋白具有亲水性。基因和蛋白结构分析发现, 蓖麻与拟南芥的 *GATA* 成员在结构上非常保守。利用高通量 RNA-seq 测序数据分析蓖麻 *GATA* 基因在各个组织中的表达情况, 发现 *GATA* 基因主要在根和叶中表达, 在种子中表达较低。进一步通过半定量 RT-PCR 技术检测 *GATA* 基因在黑暗下的表达情况, 结果显示有三个 *GATA* 基因在黑暗处理后表达下调, 预示着蓖麻 *GATA* 基因参与了光响应调控过程。这些信息为蓖麻 *GATA* 基因家族的功能分析和提高蓖麻产量提供了理论基础。

关键词: 蓖麻; *GATA* 基因家族; 基因表达; 黑暗处理

中图分类号: Q 75

文献标志码: A

文章编号: 2095-0845(2015)04-453-10

Identification and Characterization of *GATA* Gene Family in Castor Bean (*Ricinus communis*)

AO Tao^{1,3}, LIAO Xiao-jia^{2,3}, XU Wei^{2,3}, LIU Ai-zhong^{1**}

(1 Key Laboratory of Tropical Plant Resource and Sustainable Use, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Science, Mengla, Yunnan 666303, China; 2 Kunming Institute of Botany, Chinese Academy of Science, Kunming 650201 China; 3 Graduate University of Chinese Academy of Science, Beijing 100049, China)

Abstract: *GATA* proteins are considered to be broadly involved in yield-associated biological process, such as photoresponse, chlorophyll biosynthesis, and carbon and nitrogen metabolism. Based on castor bean genome database, a total of 19 *GATA* genes were identified and classified into 4 subfamilies according to gene structure, protein structure and their phylogenetic relationships. Results exhibited that *GATA* factors were hydrophilic proteins. Analysis of gene structure and protein structure revealed the conserved structural features of *GATA* factors between castor bean and *Arabidopsis thaliana*. The high-throughput RNA-seq data were used to define the expressional profiles of *GATA* genes among tissues. The results showed that most of the castor *GATA* genes preferentially expressed in leaf and root in contrast to their expression in developing seeds. In particular, the expression of *GATA* genes responding to darkness treatment in leaves was detected using semi-quantitative RT-PCR. It was shown that expression of three genes was down-regulated under darkness treatment, which suggests a role for *GATA* genes of castor bean in light-mediated regulation. These results provide important theoretical basis to the functions identification of castor *GATA* genes and increase castor yields.

Key words: Castor bean; *GATA* gene family; Gene expression; Dark treatment

* 基金项目: 国家自然科学基金 (31401421)

** 通讯作者: Author for correspondence; E-mail: Liuiaizhong@mail.kib.ac.cn

收稿日期: 2014-11-11, 2015-03-27 接受发表

作者简介: 敖涛 (1989-) 男, 硕士研究生, 主要从事植物分子遗传育种研究。E-mail: 515979319@qq.com

GATA 因子是一类广泛存在于真核生物中的转录因子,在植物光响应调控、叶绿素合成、细胞分裂素响应以及碳、氮代谢等生物学过程中发挥重要作用,这些生物学过程与作物产量息息相关,对 GATA 家族的研究可以为作物增产提供一定的理论基础。GATA 因子因特异结合 WGATAR 基因序列 (W 为 T 或 A; R 为 G 或 A; Lowry 和 Atchley, 2000) 而得名,其 DNA 结合域由一个 IV 类锌指结构 (C-X₂-C-X₁₇₋₂₀-C-X₂-C) 和其后的基本区域组成 (Lowry 和 Atchley, 2000)。在动物中, GATA 因子含有两个典型的 C-X₂-C-X₁₇-C-X₂-C 锌指结构域,其中只有 C 端的锌指结构与 DNA 结合有关;动物的 GATA 转录因子在发育、分化和细胞增殖过程中发挥重要作用 (Patient 和 McGhee, 2002)。大部分酵母的 GATA 蛋白只含有一个单独的 C-X₂-C-X₁₇-C-X₂-C 或 C-X₂-C-X₁₈-C-X₂-C 结构域 (Teakle 和 Gilmartin, 1998); 酵母的 GATA 转录因子广泛参与了多种生物学过程,如氮代谢、光形态建成、昼夜节律和结合型转换等 (Scazzocchio, 2000)。通过核磁共振的方法,解析了两个 GATA 结合域的结构: 鸡 GATA1 蛋白的 C 端锌指结构 (Omichinski 等, 1993) 和构巢曲霉 AreA 蛋白的锌指结构 (Staric 等, 1998)。结果显示, GATA 结合域由两对反向平行的 β 折叠、一个 α 螺旋和一个非结构化的尾巴组成。

在植物中,第一个 GATA 转录因子是从烟草中鉴定到的,被命名为 NTL1,是粗糙脉胞菌 GATA 成员 NIT2 的同源物 (Daniel-Vedele 和 Caboche, 1993)。之后,Reyes 等 (2004) 从拟南芥和水稻基因组中分别鉴定到了 29 个和 28 个 GATA 转录因子,并发现大部分 GATA 蛋白只含有一个 C-X₂-C-X₁₈-C-X₂-C 锌指结构域,同时也有少数蛋白含有两个锌指结构域或 C-X₂-C-X₂₀-C-X₂-C 结构。根据系统进化关系、结构域组成和内含子-外显子构成,将拟南芥 GATA 家族成员分成了四个亚家族,并将水稻 GATA 家族成员分成了六个亚家族。在拟南芥中,亚家族 I 含有两个外显子,编码 GATA 结构域的基因序列位于第二个外显子上;亚家族 II 的基因由两到三个外显子组成,编码 GATA 结构域的 DNA 序列被内含子隔开,位于后两个外显子上;亚家族 III 的基因由七个外显子构成,编码锌指结构域的基

因序列位于第五个外显子;亚家族 IV 没有保守的基因结构,其锌指结构域位于氨基端 (Reyes 等, 2004)。除 GATA 结构域之外,亚家族 I 的成员还含有一个氨基端的酸性结构域,亚家族 III 的蛋白还含有一个 CCT 结构域,水稻的亚家族 V 还含有一个 ZnF PMZ 结构域 (Plant mutator transposase zinc finger domain) 和一个 FAR1 结构域 (Reyes 等, 2004), 这些结构域可能赋予了 GATA 家族成员更多的生物学功能。例如, CCT 结构域最先是在开花时间控制蛋白 CONSTANS (CO) 中鉴定到的,被认为与蛋白和蛋白的结合有关 (Robson 等, 2001; Griffiths 等, 2003); ZnF PMZ 结构域和 FAR1 结构域都在 MULE 转座子家族的转座酶中和光敏色素 A 信号转导蛋白 (FAR1 和 FHY3) 中被发现 (Lisch 等, 2001; Hudson 等, 2003), 暗示着亚家族 V 成员在转座过程和光敏色素 A 信号转导中的潜在功能。

GATA 元件存在于很多光调节基因和昼夜节律响应基因的调控区域,通过凝胶迁移实验和 DNase I 足迹实验发现 GATA 因子可以结合到这些 GATA 元件上,预示着 GATA 转录因子可能参与了光和昼夜节律相关的调控 (Jeong 和 Shih, 2003; Teakle 等, 2002; Schindler 和 Cashmore, 1990; Borello 等, 1993; Terzaghi 和 Cashmore, 1995)。之后,在拟南芥的研究中发现,很多 GATA 转录因子基因的表达可以响应光培养、暗培养以及昼夜节律的变化 (Manfield 等, 2007)。Luo 等 (2010) 发现, GATA2 可以直接调节许多光响应基因和油菜素类固醇响应基因的表达,并且在光形态发生中发挥重要作用。此外,大量报道显示, GATA 转录因子还参与了叶绿素水平、叶绿体大小和光合效率的调节 (Bi 等, 2005; Chiang 等, 2012; Hudson 等, 2011, 2013; An 等, 2014)。这些研究表明 GATA 转录因子在光相关的调控中发挥重要作用。除此之外, GATA 转录因子还参与了种子萌发 (Liu 等, 2005), 细胞分裂素应激 (Kiba 等, 2005; Naito 等, 2007), 花发育 (Marah 和 Irish, 2008), 分蘖 (Hudson 等, 2013) 以及碳、氮代谢 (Hudson 等, 2013; Wang 等, 2003; Scheible 等, 2004; Bi 等, 2005; Hudson 等, 2011) 等生物学过程。

目前,关于 GATA 家族的研究主要局限在拟

南芥和水稻等少数模式植物中, 其它作物中 *GATA* 转录因子的研究较少。蓖麻是世界十大油料作物之一, 蓖麻油是非常良好的生物柴油原料, 广泛应用于工业生产的各个领域, 然而蓖麻产量较低的现状严重制约了生物柴油产业的发展 (朱保葛等, 2007; 黄文霞等, 2006; 吕凤霞等, 1994), 因此提高蓖麻产量方面的研究具有重要的现实意义。目前关于提高蓖麻产量的分子机制研究较少, 对 *GATA* 家族的分析和研究尚未见报道。本研究基于蓖麻的基因组信息, 在蓖麻的基因组中鉴定了 19 个 *GATA* 家族成员, 并对其系统进化、理化性质、基因结构、结构域组成和时空表达进行了分析。此外, 我们还分析了蓖麻 *GATA* 基因在正常生长条件下和黑暗条件下的表达情况。本研究结果对深入研究 *GATA* 基因家族的生物学功能和探索蓖麻光相关调控的分子机制提供了一定的研究基础, 并为提高蓖麻产量提供了非常重要的理论依据和指导意义。

1 材料和方法

1.1 蓖麻 *GATA* 因子的鉴定, 分类和基因结构分析

以已报道的模式植物拟南芥 *GATA* 蛋白序列 (<http://www.arabidopsis.org/>) 为基序, 在蓖麻基因组数据库 (<http://castorbean.jcvi.org/index.php>) 中进行 BLASTP 比对, 设定阈值为 $p < e^{-10}$, 序列一致度大于 60%。对于得到的候选蛋白, 用 Pfam 数据库 (<http://pfam.sanger.ac.uk/>) 和 SMART 程序 (<http://smart.embl-heidelberg.de/>) 检测是否具有典型的 *GATA* 结构域。利用 GSDS (Gene Structure Display Server, <http://gsds.cbi.pku.edu.cn/>) 在线程序分析蓖麻 *GATA* 家族各成员的基因结构。根据系统进化关系、结构域组成和内含子-外显子构成, 将 *GATA* 家族成员分成了四个亚家族。

1.2 *GATA* 蛋白家族的系统进化树和蛋白理化性质分析

为了了解蓖麻和拟南芥 *GATA* 蛋白之间的亲缘关系, 利用 Clustal W 软件对蓖麻和拟南芥的 *GATA* 蛋白序列进行了比对, 利用 MEGA5.0 软件中的邻接算法 (Neighbor-Joining) 构建系统进化树, Bootstrap 值设为 1000 重复。用 ExPaSy 提供的在线 ProtParam 软件 (<http://web.expasy.org/protparam/>) 进行氨基酸数目、不稳定系数、理论等电点、脂肪族氨基酸指数和蛋白质疏水性的分析。

1.3 蓖麻 *GATA* 基因的表达分析

实验蓖麻品种为淄博 306 (由淄博农业科学研究院提供)。为了探索蓖麻 *GATA* 基因与光相关调控之间的关系, 挑选在自然条件下生长了两个月的蓖麻, 将其顶端第三片叶片用锡箔纸包裹 72 h 后, 取叶片进行后续实验, 对照组正常生长, 至少重复 3 次实验。用 Transgene 公司的 TRIzol plant RNA 抽提试剂盒提取叶片总 RNA, 利用 Nanodrop 测定 RNA 浓度和质量并用琼脂糖凝胶电泳检测 RNA 质量。随后取 3 μ g 总 RNA 用 Takara 公司的 PrimescriptTM II 1st strand cDNA Synthesis Kit 试剂盒进行反转录。RT-PCR 引物序列见表 1。用合成的 cDNA 为模板进行 PCR 扩增, 程序为 94 $^{\circ}$ C 预变性 4 min, 94 $^{\circ}$ C 变性 30 s, 55 $^{\circ}$ C 退火 30 s, 72 $^{\circ}$ C 延伸 35 s, 25 个循环, 72 $^{\circ}$ C 延伸 7 min。另外, 利用已公布的蓖麻表达谱数据 (Xu 等, 2013), 调查了 *GATA* 基因在不同组织包括叶、根、授粉后 20 天的种子 (种子 1)、授粉后 35 天种子 (种子 2) 和授粉后 40 天的胚乳中的表达情况。

2 结果与分析

2.1 蓖麻 *GATA* 蛋白的鉴定及理化性质分析

通过 BLASTP 比对以及 Pfam 和 SMART 结构分析, 本研究获得了 19 个蓖麻 *GATA* 同源蛋白 (表 2)。对蓖麻 *GATA* 蛋白进行理化性质分析发现, 19 个成员包含 122 到 542 个氨基酸数目不等,

表 1 RT-PCR 引物序列

Table 1 Primer for RT-PCR

基因名称 Gene name	正向引物序列 Forward primer sequences	反向引物序列 Reverse primer sequences	扩增长度 PCR product length/bp
30131. m007120	TTTCTCACCAGCGAACTTGC	CCGCTGGTTTTCTCTTTGCT	430
29904. m003012	GCGGAGGAAGAGCTTGAATG	AACGTGGGACTACTAGCAGG	485
29794. m003323	GCCAGTGAAATTTACAGTCAAGT	TGCTGCTTCTTCAACATCCTC	454
29912. m005322	CCCACCTCAAGAAGAAGTTGG	GTCTTCATGGCTGCTGTGTC	564
29838. m001723	CTGTTGGGTCTGACAAGTGC	GGTTCAGATTGAGGAGTCGTATTG	391
29838. m001724	TGGAGGTGGTGAGGAGTCTA	CITGTCCAGAACCCTGTGGTC	530
29596. m000720	CTGGTCCACATGGTCTAGAAGTTG	TGCACATTCTCCACACCAG	440
27940. m000323	AGGGGTTGTCCATGACTACAATC	TGAATAAGCCTCATTGTGTTGAGC	453
<i>Actin</i>	GAGCTTCGTGTGCTCCTGAG	TGGCCATCAGGAAGCTCATAG	461

表2 蓖麻 GATA 蛋白的理化性质分析
Table 2 Physical and chemical properties of castor bean GATA proteins

亚家族 Subfamily	基因座位 Gene loci	编码氨基酸 Coded amino acids	不稳定系数 Instability index	脂肪族氨基酸指数 Aliphatic index	等电点 pI	蛋白质疏水性 Hydrophobicity
亚家族 I Subfamily I	29634. m002055	398	67.90	60.48	6.03	-0.58
	30131. m007120	368	61.01	51.14	5.00	-0.73
	30146. m003524	338	52.36	51.75	6.22	-0.70
	30174. m008809	235	58.57	66.47	7.00	-0.62
	29904. m003012	205	61.45	58.39	10.06	-0.81
	27538. m000326	323	73.32	54.06	7.04	-0.80
	27494. m000044	359	65.55	62.17	6.60	-0.58
亚家族 II Subfamily II	29634. m002146	332	54.39	52.62	5.66	-0.70
	28035. m000366	161	49.38	48.51	9.57	-1.02
	30128. m008883	151	67.83	54.97	10.27	-0.83
	30076. m004701	122	40.33	60.82	9.53	-0.67
	29794. m003323	186	41.02	53.60	9.68	-0.70
	29912. m005322	312	54.52	58.21	9.16	-0.89
	29889. m003248	149	55.24	61.61	9.79	-0.98
亚家族 III Subfamily III	29596. m000719	327	44.86	70.64	4.84	-0.65
	29838. m001723	313	36.95	62.27	5.06	-0.66
	29838. m001724	311	45.07	61.77	5.07	-0.70
亚家族 IV Subfamily IV	29596. m000720	324	38.55	59.23	7.73	-0.71
	27940. m000323	542	53.97	73.36	6.53	-0.62

与已报道的物种拟南芥和水稻一致 (Reyes 等, 2004), 各成员之间的氨基酸数目保守性很低。对蛋白亲/疏水性进行预测, 发现 19 个蛋白的 GRA-VY 值都为负值, 即蓖麻的 GATA 蛋白都是亲水蛋白, 只是亲水程度有所差异。除 29838. m001723 和 29596. m000720 属于稳定蛋白之外, 其余 17 个成员均为不稳定蛋白。蓖麻 GATA 蛋白理化性质的分析可以为进一步的蛋白提纯、活性和功能研究提供理论基础。

2.2 蓖麻 GATA 蛋白的分类和特征分析

根据系统进化关系、结构域组成和内含子-外显子构成, 将 GATA 家族成员分成了四个亚家族, 亚家族 I 包含 7 个成员, 亚家族 II 也包含 7 个成员, 亚家族 III 包含 4 个成员, 亚家族 IV 只包含 1 个成员。进化树分析发现, 蓖麻 GATA 蛋白主要分为四个支, 且拓扑结构明显, 除 29889. m003248 之外, 各个亚家族的蛋白都分别聚在一支 (图 1: A)。对各亚家族进行结构域分析发现, 亚家族 I 和亚家族 II 成员只含有一个 GATA 结构域, 而除 GATA 结构域之外, 亚家族 III 的 4 个成员都含有一个额外的 CCT 结构域, 其中 3 个成员还含有

一个 TIFY 结构域, 亚家族 IV 的 27940. m000323 还含有一个额外的 ASXH 结构域 (图 1: B)。在拟南芥的报道中, 亚家族 I 的成员还含有一个氨基端的酸性结构域; 亚家族 III 只含有 GATA 结构域和 CCT 结构域, 而缺少 TIFY 结构域; 亚家族 IV 也缺少额外的 ASXH 结构域 (Reyes 等, 2004), 这些差异可能来源于检测方法和参数设置的不同。本研究重新用 Pfam 和 SMART 检测了拟南芥 GATA 蛋白的结构域组成, 在亚家族 I 中未检测到酸性结构域存在, 且在亚家族 III 的所有成员中都检测到了 CCT 结构域和 TIFY 结构域, 此外, 也在亚家族 IV 的所有成员中检测到了 ASXH 结构域 (数据未显示), 与蓖麻 GATA 蛋白的结构域组成一致。

对 GATA 家族成员的基因结构分析发现, 亚家族 I 含有两个外显子, 编码 GATA 结构域的基因序列位于第二个外显子上; 亚家族 II 的基因由两到四个外显子组成, 编码 GATA 结构域的 DNA 序列被内含子隔开, 位于后两个外显子上; 亚家族 III 的基因由七到十一个外显子构成, 编码锌指结构域的基因序列位于第五个外显子; 亚

家族 IV 由八个外显子组成, 编码 *GATA* 结构域的 DNA 序列也被内含子隔开, 位于前两个外显子上 (图 1: C), 这些基因结构特征与拟南芥非常相似 (Reyes 等, 2004)。为了调查 *GATA* 蛋白的锌指结构域特征, 我们对蓖麻和拟南芥 *GATA* 蛋白的锌指结构域进行了序列比对, 并参考 Reyes 等 (2004) 的标准, 标示了锌指结构域的二级结构—四个 β 折叠和一个 α 螺旋 (图 2)。比对结果显示, 除亚家族 III 的锌指结构为 C-X₂₀-C-X₂-C 之外, 其余三个家族的锌指结构都为 C-X₂-C-X₁₈-C-X₂-C, 与拟南芥的报道一致 (Reyes 等, 2004)。此外, 如图 2 所示, 锌指结构域的很多氨基酸位点在所有蓖麻和拟南芥的 *GATA* 成员中是高度保守的, 如 Cys-3, Cys-6, Thr-13, Pro-14, Arg-17, Gly-19, Pro-20 以及第二个半胱氨酸对的侧翼序列 (LCNACG), 这些保守的氨基酸可能与 *GATA* 转录因子的 DNA 结合功能以及转录激活功能有关。为了进一步探索蓖麻和拟南芥 *GATA* 蛋白之间的系统进化关系, 用 19 个蓖麻 *GATA* 蛋白和 29 个拟南芥 *GATA* 蛋白序列进行系统进化树分析。如图 3 所示, 蓖麻和拟南芥的 *GATA* 蛋白被分为四个大支, 除 At5g26930 之外, 各个亚家族的成员都单独聚在一支, 说明 *GATA* 蛋白在蓖麻和拟南芥中是同源的。综上所述, 蓖麻与拟南芥的 *GATA* 成员在蛋

白结构、基因结构和进化关系上都非常同源, 预示着蓖麻的 *GATA* 蛋白与拟南芥 *GATA* 蛋白在植物中发挥着相似的功能。而 *GATA* 转录因子各亚家族基因结构和蛋白结构的差异预示着各亚家族间基因功能的分化。

2.3 蓖麻 *GATA* 基因的表达分析

我们首先利用已报道的蓖麻基因表达谱数据调查了 *GATA* 基因在根、叶、种子 1、种子 2 和胚乳中的表达情况。如图 4 所示, 除 30146. m003524, 30076. m004701 和 29889. m003248 之外, 其余 16 个基因在测序的五个组织中都检测到表达, 大部分基因表现出明显的组织偏爱性表达。几乎所有的基因在种子 2 和胚乳中表达量都很低, 只有 2 个基因 (29904. m003012 和 27940. m000323) 在种子 1 中有较高的表达; 在根中, 有 3 个基因 (27940. m000323, 27538. m000326 和 29904. m003012) 表达量较高; 而在叶中有 8 个基因 (30131. m007120, 29904. m003012, 29794. m003323, 29912. m005322, 29838. m001723, 29838. m001724, 29596. m000720 和 29912. m005322) 表达量较高, 且 30131. m007120, 29794. m003323 和 29912. m005322 为叶片偏爱性表达基因, 强烈预示着 *GATA* 基因家族成员, 特别是 3 个叶偏爱表达基因在叶中发挥着重要功能。大量文献报道 *GATA* 家族转录因子参与了光相关的调节过程, 而叶片是植物进行光合作用以及光

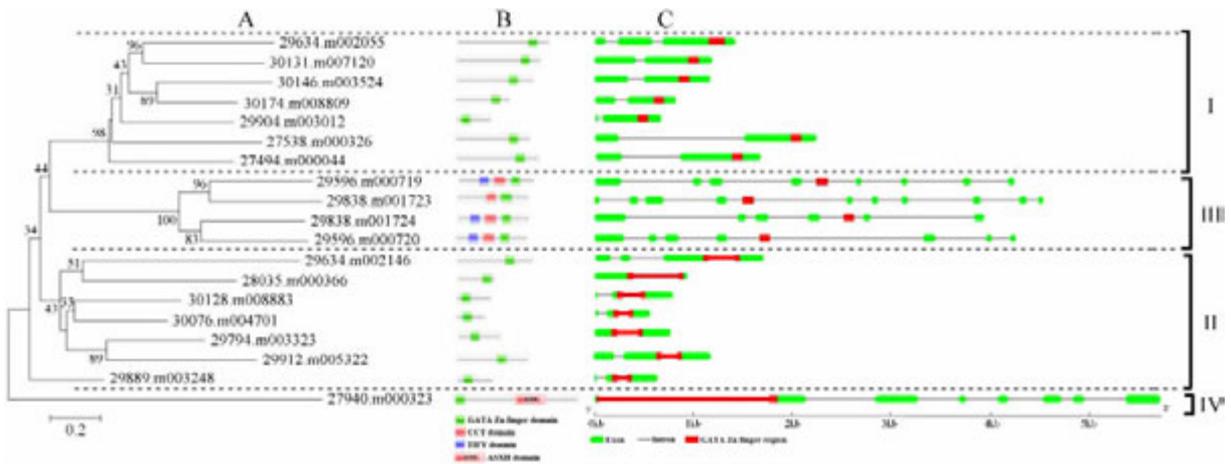


图 1 蓖麻 *GATA* 成员的基因结构和功能域组成

A. 蓖麻 *GATA* 蛋白的系统进化树分析; B. 蓖麻 *GATA* 蛋白的功能域组成; C. 蓖麻 *GATA* 家族成员的基因结构分析。I-IV 代表 *GATA* 的四个亚家族

Fig. 1 The gene structure and domain distribution of *GATA* members in castor bean

A. Phylogenetic analysis of *GATA* proteins in castor bean; B. Domain distribution of castor bean *GATA* members;

C. Gene structure analysis of castor bean *GATA* genes. I-IV represent four subfamilies of *GATA* family

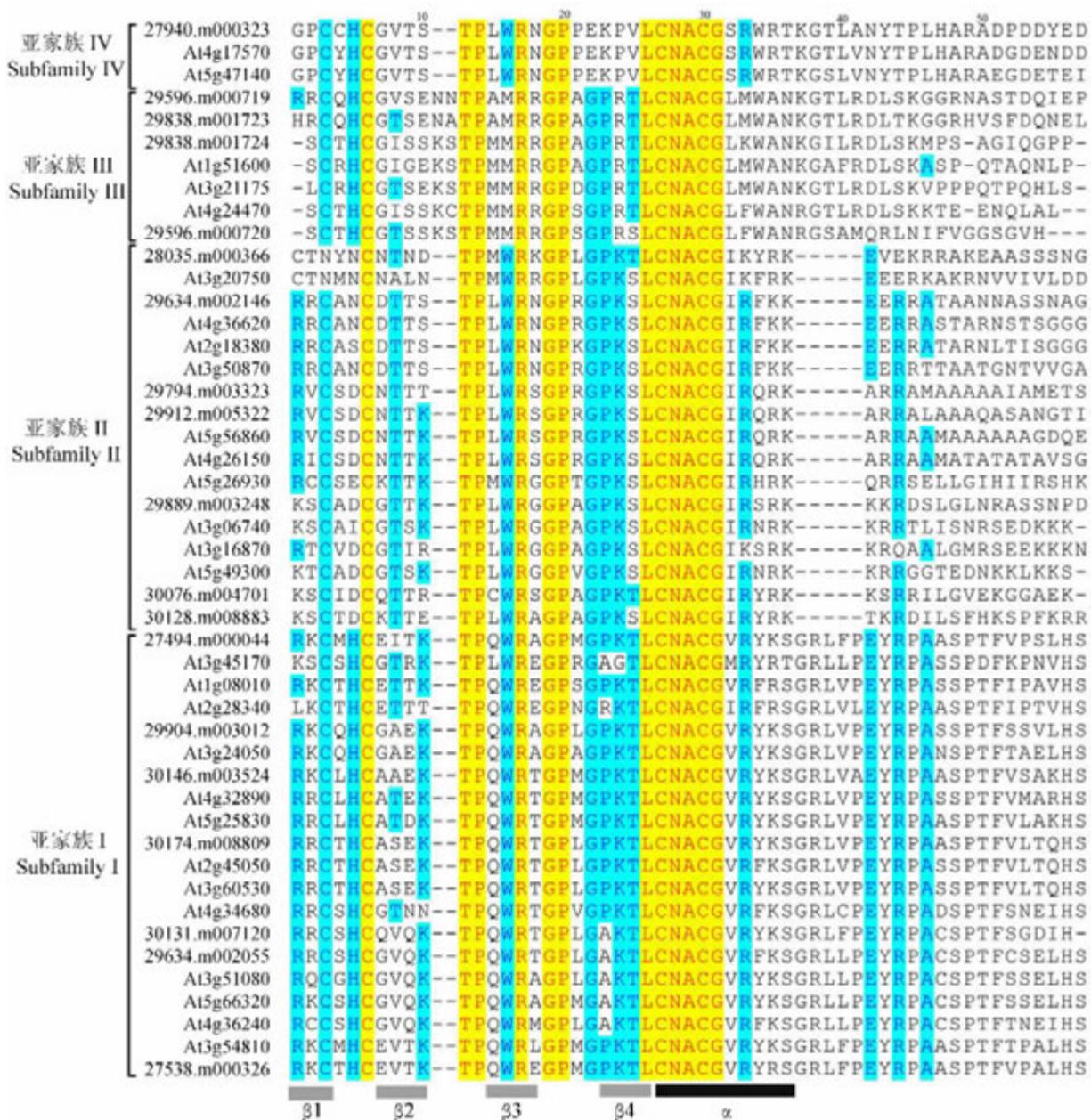


图 2 蓖麻与拟南芥 GATA 蛋白锌指结构域的氨基酸序列比对。底部的框标示锌指结构域的二级结构

Fig. 2 Amino acid sequence alignment of *Arabidopsis* and castor bean GATA zinc finger domains. The bottom part of the figure shows the secondary structure elements of zinc finger domains

应激的主要器官，故我们重点挑选了 8 个在叶片中表达较高的基因（表 1），探索它们于正常生长条件下和黑暗条件下在叶片中的表达情况。结果如图 5 所示，在黑暗处理后，亚家族 III 的两个基因（29838. *m001723* 和 29596. *m000720*）表现出明显下调，亚家族 II 的一个基因（29794. *m003323*）也表现出一定的下调趋势，暗示着这三个基因可能参与了光相关的调节。

3 讨论

本研究从蓖麻的基因组中鉴定了 19 个 GATA 家族成员，和拟南芥（29 个）、水稻（28 个）相比，蓖麻的 GATA 蛋白基因明显偏少。在已报道的 bZIP 转录因子家族（Jin 等，2013）和 AP2/ERF 转录因子家族（Xu 等，2013）中也发现，与其它被子植物相比，蓖麻含有较少的家族成员。在蓖麻中，GATA 家族成员被分为 I-IV 四个亚家族，

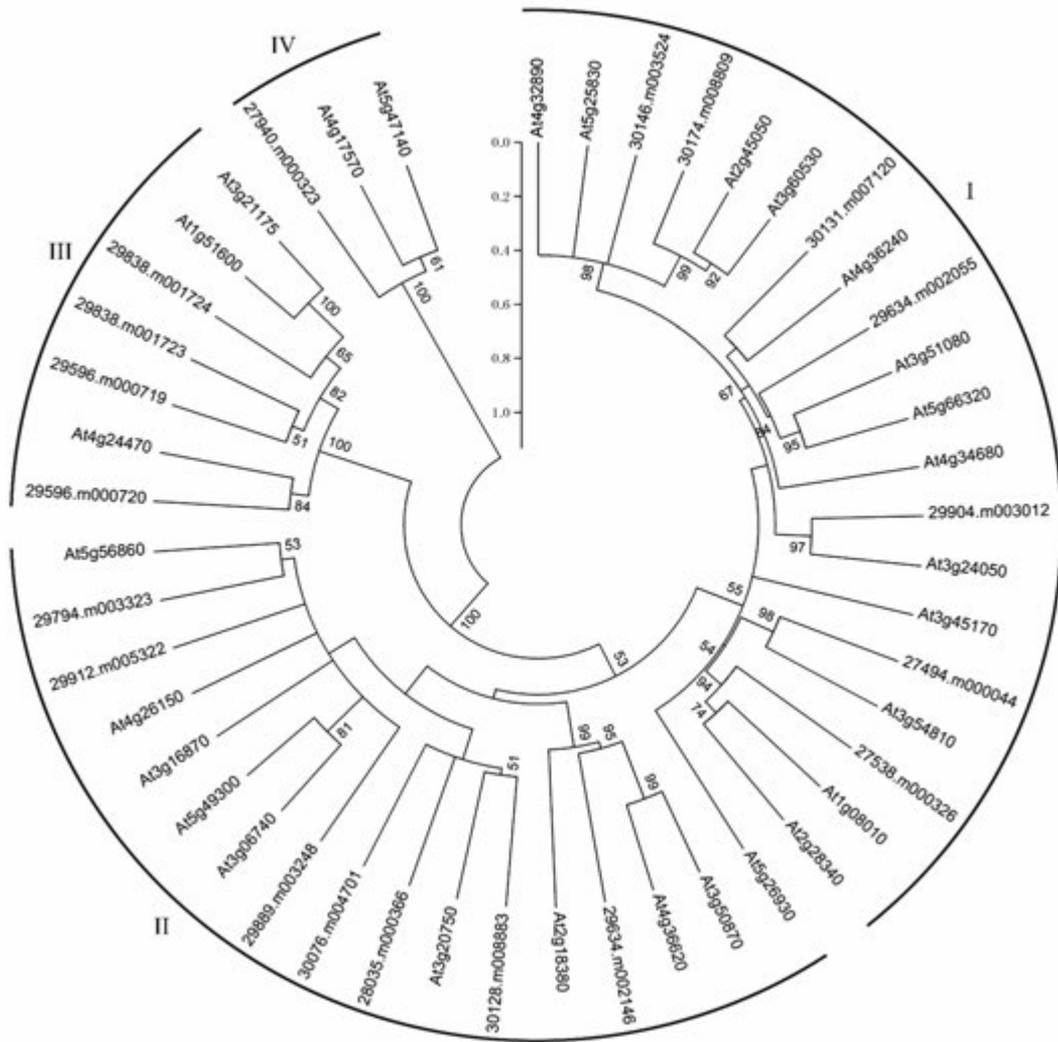


图 3 蓖麻与拟南芥 *GATA* 转录因子的系统进化树。I-IV 分别代表 *GATA* 家族的四个亚家族

Fig. 3 The phylogenetic tree of *GATA* transcription factors in castor bean and *Arabidopsis*.

I-IV represents four subfamilies of *GATA* family, respectively

和同为双子叶植物的拟南芥一致 (Reyes 等, 2004), 且蓖麻与拟南芥 *GATA* 成员在基因和蛋白结构上非常相似, 同源性也很高, 暗示着蓖麻和拟南芥 *GATA* 蛋白具有相似的功能。但是, 已报道的单子叶植物水稻中的 *GATA* 成员与拟南芥和蓖麻的 *GATA* 成员在基因结构和蛋白结构上具有很大的差异, 水稻的 *GATA* 蛋白被分为六个亚家族 (亚家族 I, II, III, V, VI, VII), 缺少亚家族 IV, 而亚家族 V 含有 ZnF PMZ 结构域和 FAR1 结构域, 亚家族 VI 含有两个 *GATA* 结构域; 此外, 亚家族 V, VI 和 VII 还具有异于拟南芥和蓖麻的基因结构, 如亚家族 VII 只含有一个外显子 (Reyes 等, 2004), 这些差异预示着 *GATA* 家族在单、双子

叶植物中可能存在结构和功能的分化, 有待进一步研究。

本研究在 *GATA* 蛋白中鉴定了两个新的结构域 (TIFY 结构域和 ASXH 结构域), TIFY 是一个植物特异的蛋白结构域, 被认为与蛋白结合有关, 广泛存在于 JAZ (JASMONATE ZIM-domain) 蛋白家族和 PEAPOD (PPD) 蛋白中, 可能广泛参与了茉莉酸相关的信号途径 (Bai 等, 2011; Vanholme 等, 2007; Chini 等, 2009; Melotto 等, 2008)。目前关于 ASXH 结构域的研究主要集中在动物中, 被认为与多梳型复合蛋白的结合有关, 含有 ASXH 结构域的 ASXL 1, ASXL 2 和 ASXL 3 通过与多梳复合蛋白结合而广泛参与了表观标记

(泛素化与甲基化) 和转录水平的调控, 进而参与调节骨骼和心脏发育、恶性骨髓疾病和肝癌、前列腺癌等疾病的形成 (Aravind 和 Iyer, 2012;

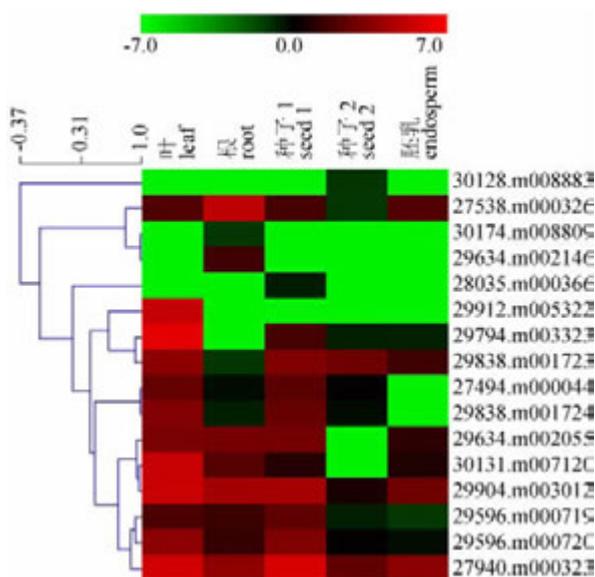


图 4 蓖麻 *GATA* 基因在不同组织中的表达谱热图
将 DGE 测序的数据用 \log_2 标准化后作图, 有颜色的标尺代表的是 \log_2 标准化后的表达数据

Fig. 4 Heatmap represents the expression profiles of castor bean *GATA* genes in different tissues
This heatmap was generated based on the DGE expression data (normalized \log_2). The color scale (representing \log_2 signal values) is shown at the top

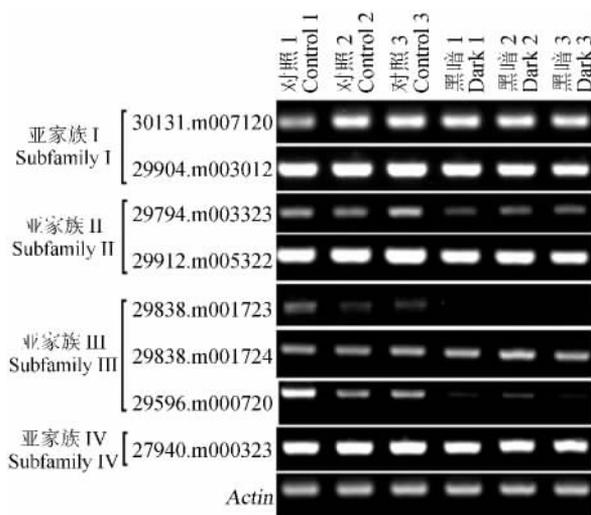


图 5 半定量 RT-PCR 检测蓖麻叶片中 *GATA* 基因在黑暗处理下的表达情况 (1, 2, 3 代表实验的三个重复)

Fig. 5 Semi-quantitative RT-PCR analysis of *GATA* genes expression level in castor leaf under dark treatment (1, 2, 3 represent three repeat experiments)

Gelsi-Boyer 等, 2012; Katoh, 2013), ASXH 结构域在植物中的功能需要进一步研究。

表达分析发现, *GATA* 基因的表达具有明显的组织偏爱性, 有 8 个成员在叶中高表达, 3 个成员在根中高表达, 而大部分成员在种子中的表达很低; 在拟南芥中, *GATA* 基因在根和叶中表达量也很高 (Manfield 等, 2007), 这些结果预示着 *GATA* 转录因子可能广泛参与了根和叶相关的生物学过程。研究报道, 拟南芥的 *GATA* 成员 *GNC* 和 *CGA 1* 广泛参与了叶绿素水平、叶绿体大小、光合效率和碳、氮代谢的调节 (Bi 等, 2005; Hudson 等, 2011; Chiang 等, 2012)。这两个成员的水稻同源基因 *Os02g12790* 和杨树同源基因 *PdGNC* 也在调节叶绿素水平和碳、氮代谢中发挥重要作用 (Hudson 等, 2013; An 等, 2014)。在蓖麻中, 其同源基因 *27940.m000323* 在根和叶中表达量都很高, 预示着 *27940.m000323* 在调节叶绿素水平和碳、氮代谢中的潜在功能。为了进一步探索在叶片中高表达的 *GATA* 基因与光相关调控的联系, 我们调查了这些基因在黑暗条件和正常生长条件下的表达情况, 找到了三个在黑暗处理时表达下调的基因 (*29794.m003323*, *29838.m001723* 和 *29596.m000720*), 这三个基因的拟南芥同源基因在黑暗处理后都表现出明显的表达下调 (Manfield 等, 2007)。其中亚家族 II 的 *29794.m003323* 是拟南芥 *GNC* 和 *CGA 1* 基因的同源基因, 进一步预示着 *27940.m000323* 在调节叶绿素水平和光合效率中的潜在功能。*29838.m001723* 和 *29596.m000720* 属于亚家族 III, 目前关于植物中亚家族 III 的 *GATA* 成员的功能研究较少, 其在光相关调控中的功能需进一步探索。在所检测的三个蓖麻亚家族 III 成员中有两个在黑暗处理下表达下调, 且拟南芥中所有的亚家族 III 成员都表现出黑暗处理下调, 暗示着亚家族 III 的 *GATA* 成员可能在光相关调控中发挥重要作用。从结构上看, 蓖麻和拟南芥的亚家族 III 成员都含有 CCT 结构域, 且大部分还含有 TIFY 结构域, 这两个结构域都与蛋白之间的结合有关 (Chini 等, 2009; Melotto 等, 2008; Bai 等, 2011; Robson 等, 2001; Griffiths 等, 2003), 可能通过结合光相关通路蛋白而参与相关过程的调节。

蓖麻是一种非常重要的非食用油料作物, 蓖麻油是非常良好的生物柴油原料, 提高蓖麻的生产量是生物柴油产业快速发展的必经之路, 所以提高蓖麻产量方面的研究具有重要的现实意义。本研究鉴定了与蓖麻产量息息相关的 GATA 家族, 并对其系统进化、理化性质、基因结构、结构域组成、时空表达和黑暗处理后的表达情况进行了详尽分析, 为深入研究 GATA 基因家族的生物学功能, 探索蓖麻光相关调控的分子机制以及提高蓖麻产量提供了非常重要的理论基础和指导意义。

(参 考 文 献)

- 黄文霞, 何觉民, 陆建农, 2006. 国内外蓖麻杂种优势利用和蓖麻加工研究现状 [J]. 作物研究, (5): 541—544
- 吕凤霞, 刘恩魁, 刘永平, 1994. 蓖麻的开发利用前途广阔 [J]. 河北农业科技, 4 (2): 12
- An Y, Han X, Tang S *et al.*, 2014. Poplar GATA transcription factor PdGNC is capable of regulating chloroplast ultrastructure, photosynthesis, and vegetative growth in Arabidopsis under varying nitrogen levels [J]. *Plant Cell, Tissue and Organ Culture*
- Aravind L, Iyer LM, 2012. The HARE-HTH and associated domains [J]. *Cell Cycle*, 11 (1): 119—131
- Bai Y, Meng Y, Huang D *et al.*, 2011. Origin and evolutionary analysis of the plant-specific TIFY transcription factor family [J]. *Genomics*, 98 (2): 128—136
- Bi YM, Zhang Y, Signorelli T *et al.*, 2005. Genetic analysis of Arabidopsis GATA transcription factor gene family reveals a nitrate-inducible member important for chlorophyll synthesis and glucose sensitivity [J]. *The Plant Journal*, 44 (4): 680—692
- Borello U, Ceccarelli E, Giuliano G, 1993. Constitutive, light-responsive and circadian clock-responsive factors compete for the different 1 box elements in plant light-regulated promoters [J]. *The Plant Journal*, 4 (4): 611—619
- Chiang YH, Zubo YO, Tapken W *et al.*, 2012. Functional Characterization of the GATA Transcription Factors GNC and CGA1 Reveals Their Key Role in Chloroplast Development, Growth, and Division in Arabidopsis [J]. *Plant Physiology*, 160 (1): 332—348
- Chini A, Fonseca S, Chico JM *et al.*, 2009. The ZIM domain mediates homo- and heteromeric interactions between Arabidopsis JAZ proteins [J]. *The Plant Journal*, 59 (1): 77—87
- Daniel-Vedele F, Caboche M, 1993. A tobacco cDNA clone encoding a GATA-1 zinc finger protein homologous to regulators of nitrogen metabolism in fungi [J]. *Molecular and General Genetics*, 240 (3): 365—373
- Gelsi-Boyer V, Brecqueville M, Devillier R *et al.*, 2012. Mutations in ASXL1 are associated with poor prognosis across the spectrum of malignant myeloid diseases [J]. *Journal of Hematology & Oncology*, 5: 12
- Griffiths S, Dunford RP, Coupland G *et al.*, 2003. The evolution of CONSTANS-like gene families in barley, rice, and Arabidopsis [J]. *Plant Physiology*, 131 (4): 1855—1867
- Hudson D, Guevara DR, Hand AJ *et al.*, 2013. Rice cytokinin GATA transcription Factor1 regulates chloroplast development and plant architecture [J]. *Plant Physiology*, 162 (1): 132—144
- Hudson D, Guevara DR, Yaish MW *et al.*, 2011. GNC and CGA1 modulate chlorophyll biosynthesis and glutamate synthase (GLU1/Fd-GOGAT) expression in Arabidopsis [J]. *PLoS ONE*, 6 (11): 1—16
- Hudson ME, Lisch DR, Quail PH, 2003. The FHY3 and FAR1 genes encode transposase-related proteins involved in regulation of gene expression by the phytochrome A-signaling pathway [J]. *The Plant Journal*, 34 (4): 453—471
- Jeong MJ, Shih MC, 2003. Interaction of a GATA factor with cis-acting elements involved in light regulation of nuclear genes encoding chloroplast glyceraldehyde-3-phosphate dehydrogenase in Arabidopsis [J]. *Biochemical and Biophysical Research Communication*, 300 (2): 555—562
- Jin ZW, Xu W, Liu AZ, 2013. Genomic surveys and expression analysis of bZIP gene family in castor bean (*Ricinus communis* L.) [J]. *Planta*, 239 (2): 299—312
- Katoh M, 2013. Functional and cancer genomics of ASXL family members [J]. *British Journal of Cancer*, 109 (2): 299—306
- Kiba T, Naitou T, Koizumi N *et al.*, 2005. Combinatorial microarray analysis revealing Arabidopsis genes implicated in cytokinin Responses through the His → Asp Phosphorelay Circuitry [J]. *Plant Cell Physiology*, 46 (2): 339—355
- Lisch DR, Freeling M, Langham RJ *et al.*, 2001. Mutator transposase is widespread in the grasses [J]. *Plant Physiology*, 125 (3): 1293—1303
- Liu PP, Koizuka N, Martin RC *et al.*, 2005. The BME3 (Blue Mircropylar End 3) GATA zinc finger transcription factor is a positive regulator of Arabidopsis seed germination [J]. *The Plant Journal*, 44 (6): 960—971
- Lowry JA, Atchley WR, 2000. Molecular evolution of the GATA family of transcription factors: conservation within the DNA-binding Domain [J]. *Journal of Molecular Evolution*, 50 (2): 103—115
- Luo XM, Lin WH, Zhu SW *et al.*, 2010. Integration of Light- and Brassinosteroid-Signaling Pathways by a GATA Transcription Factor in Arabidopsis [J]. *Developmental Cell*, 19 (6): 872—883
- Manfield LW, Devlin PF, Jen CH *et al.*, 2007. Conservation, convergence, and divergence of light-responsive, circadian-regulated, and tissue-specific expression patterns during evolution of the Arabidopsis GATA gene family [J]. *Plant Physiology*, 143 (2): 941—958

- Mara CD, Irish VF, 2008. Two GATA transcription factors are downstream effectors of floral homeotic gene action in Arabidopsis [J]. *The Plant Physiology*, **147** (2): 707—718
- Melotto M, Mecey C, Niu Y *et al.*, 2008. A critical role of two positively charged amino acids in the Jas motif of Arabidopsis JAZ proteins in mediating coronatine- and jasmonoyl isoleucine-dependent interactions with the COII F-box protein [J]. *The Plant Journal*, **55** (6): 979—988
- Naito T, Kiba T, Koizumi N *et al.*, 2007. Characterization of a unique GATA family gene that responds to both light and cytokinin in Arabidopsis thaliana [J]. *Bioscience, Biotechnology, and Biochemistry*, **71** (6): 1557—1560
- Omichinski JG, Clore GM, Schaad O *et al.*, 1993. NMR structure of a specific DNA complex of Zn-containing DNA binding domain of GATA-1 [J]. *Science*, **261** (5120): 438—446
- Patient RK, McGhee JD, 2002. The GATA family (vertebrates and invertebrates) [J]. *Curr Opin Genet Dev*, **12** (4): 416—422
- Reyes JC, Muro-Pastor MI, Florencio FJ, 2004. The GATA family of transcription factors in Arabidopsis and rice [J]. *Plant Physiology*, **134** (4): 1718—1732
- Robson F, Costa MM, Hepworth SR *et al.*, 2001. Functional importance of conserved domains in the flowering-time gene *CONSTANS* demonstrated by analysis of mutant alleles and transgenic plants [J]. *The Plant Journal*, **28** (6): 619—631
- Scazzocchio C, 2000. The fungal GATA factors [J]. *Current Opinion in Microbiology*, **3** (2): 126—131
- Scheible WR, Morcuende R, Czechowski T *et al.*, 2004. Genome-wide reprogramming of primary and secondary metabolism, protein synthesis, cellular growth processes, and the regulatory infrastructure of Arabidopsis in response to nitrogen [J]. *Plant Physiology*, **136** (1): 2483—2499
- Schindler U, Cashmore AR, 1990. Photoregulated gene expression may involve ubiquitous DNA binding proteins [J]. *EMBO Journal*, **9** (11): 3415—3427
- Starich MR, Wikstrom M, Arst HN Jr *et al.*, 1998. The solution structure of a fungal AREA protein-DNA complex: an alternative binding mode for the basic carboxyl tail of GATA factors [J]. *Journal of Molecular Biology*, **277** (3): 605—620
- Teakle GR, Gilmartin PM, 1998. Two forms of type IV zinc-finger motif and their kingdom-specific distribution between the flora, fauna and fungi [J]. *Trends in Biochemical Sciences*, **23** (3): 100—102
- Teakle GR, Manfield IW, Graham JF *et al.*, 2002. Arabidopsis thaliana GATA factors: organization, expression and DNA-binding characteristics [J]. *Plant Molecular Biology*, **50** (1): 43—57
- Terzaghi WB, Cashmore AR, 1995. Light-regulated transcription [J]. *Annual Review of Plant Physiology and Plant Molecular Biology*, **46**: 445—474
- Vanholme B, Grunewald W, Bateman A *et al.*, 2007. The tify family previously known as ZIM [J]. *Trends in Plant Science*, **12** (6): 239—244
- Wang R, Okamoto M, Xing X *et al.*, 2003. Microarray analysis of the nitrate response in Arabidopsis roots and shoots reveals over 1 000 rapidly responding genes and new linkages to glucose, trehalose-6-phosphate, iron, and sulfate metabolism [J]. *Plant Physiology*, **132** (2): 556—567
- Xu W, Li F, Ling LZ *et al.*, 2013. Genome-wide survey and expression profiles of the AP2/ERF family in castor bean (*Ricinus communis* L.) [J]. *BMC Genomics*, **14** (13): 785
- Zhu BG (朱保葛), Li WB (李文彬), 2007. Progress in castor biotechnology [J]. *Chinese Biotechnology (中国生物工程杂志)*, **27** (10): 98—102