

谷子 f103 基因的克隆及其特性分析

薛静 于静娟 赵倩 朱登云 敖光明

(中国农业大学 农业生物技术国家重点实验室,北京 100094)

摘要 本文首次报道从谷子 (*Setaria italica*) 未成熟种子 cDNA 文库中克隆到一个新基因 f103。序列分析表明,该基因由 819 个核苷酸组成,推测其编码的蛋白质有 130 个氨基酸,分子质量为 14 113 u。Southern 杂交结果显示 f103 基因以单拷贝存在于谷子基因组中,并且在玉米基因组中存在同源基因。Northern 杂交表明 f103 基因在谷子茎、幼穗及未成熟种子中表达,在叶片组织中未检测到其表达活性。生物学软件分析结果显示,F103 蛋白是亲水蛋白,定位于核内,肽链上有多个可以被磷酸化修饰的位点。这些特点可能是 F103 蛋白生理功能的结构基础。

关键词 谷子; f103 基因; cDNA 文库; 序列分析

中图分类号 Q 78

文章编号 1007-4333(2004)01-0007-04

文献标识码 A

Cloning of f103 in *Setaria italica* and its characters

Xue Jing, Yu Jingjuan, Zhao Qian, Zhu Dengyun, Ao Guangming

(State Key Laboratory of Agrobiotechnology, China Agricultural University, Beijing 100094, China)

Abstract A new gene (f103) was isolated from *Setaria italica* cDNA library. The sequenced and characterized. f103 gene was composed of 819 nt. The encoded putative polypeptide had a molecular weight of 14 113 u and contained 130 amino acid residues. Southern blotting analysis indicated that f103 was one copy in the genome of millet and had homologous gene in maize. Northern blot analysis showed that f103 mRNA accumulated in stems, inflorescence, developing seeds, but not in leaves. The analysis by a biology software indicated that the hydrophobic domain was dominant in the structure of the deduced protein which had several phosphor-sites and might be a nuclear targeting protein. All these would be the structural basis for the function of F103 protein.

Key words millet; f103 gene; cDNA library; sequence analysis

目前克隆基因的方法有多种,较为直接、简便的是构建 cDNA 文库后利用不同类型探针(蛋白质或核酸)筛选目的基因。获得一段 DNA 序列功能信息最简单的方法是将该 DNA 序列与 GenBank 公布的基因序列进行同源性比较,但仅通过同源性检索分析不足以确定基因的功能,多数情况下还需进一步的实验数据。例如,通过研究基因的时空表达模式确定其在细胞或发育上的功能、研究蛋白质的亚细胞定位及翻译后修饰以及通过基因敲除(knock-out)进行功能丧失(loss-of-function)分析或通过基因的过量表达进行功能获得(gain-of-function)分析,从而确定目的基因与表型性状之间的关系,进而推断出该基因在细胞或发育中的作用^[1]。

本实验利用高赖氨酸蛋白基因 *sb401* 的核苷酸序列^[2]为探针,筛选谷子未成熟种子 cDNA 文库,共获得 5 个阳性克隆。将它们进行序列测定,并进行 GenBank 同源序列查询。其中一个序列在 GenBank 中(具有很低的同源性)没有发现与其相似的序列,将其命名为 f103,同时对其特性进行了研究。

1 材料与方法

1) 材料。谷子 (*Setaria italica*) 3661,由中国农科院品种资源研究所提供。谷子未成熟种子 cDNA 文库为本实验室保存。

2) 酶和试剂。限制性内切酶购自 GIBCO 公司,其他试剂为进口或国产分析纯。

收稿日期: 2003-11-11

基金项目: 国家转基因植物研究与产业化开发专项(J00-B-003-2)

作者简介: 薛静,硕士研究生;于静娟,副教授,博士生导师,通讯作者,主要从事植物基因工程的研究。E-mail: yujj@mail.cau.edu.cn

f103 在 GenBank 登陆号为:A Y327512。

3) cDNA 文库的筛选及克隆的获得。利用本实验室克隆的高赖氨酸蛋白基因 *sb401* cDNA^[2] 为探针,筛选谷子未成熟种子 cDNA 文库。经过 3 轮筛选后,得到 5 个阳性克隆。利用 helperphage 得到噬菌体质粒,选取其中一个克隆,命名为 *f103*,由博亚公司进行 DNA 测序。

4) 序列分析及同源性比较。使用 DNAMAN 软件由 *f103* 的 cDNA 序列推断其编码的氨基酸序列,并将其核苷酸序列及氨基酸序列与其他物种进行同源比较。

5) 探针的制备和标记。*EcoR* I 和 *Kpn* I 酶切重组质粒,电泳后回收质粒的插入片段,以其为模板,使用 Random Primer Labeling Kit (Promega 公司产品)以 [³²P]dCTP 标记探针。

6) Southern 杂交分析。从谷子叶片中提取基因组 DNA。使用限制性内切酶 *Hind* III, *Xho* I, *Pst* I, *Xba* I 和 *EcoR* I 完全酶解,经 0.8% 琼脂糖凝胶电泳后转移至尼龙膜上。参考分子克隆^[3]进行杂交,并按下列条件洗膜(2 ×SSC, 0.5% SDS, 30 min; 1 ×SSC, 0.5% SDS, 30 min; 0.5 ×SSC, 0.5% SDS, 30 min)。X 光片曝光 3 d。

7) Northern 杂交分析。提取谷子茎、叶、幼穗

及授粉后的 5, 15 和 25 d 未成熟种子等材料的总 RNA, 20 μg 样品进行甲醛变性电泳,真空转移至尼龙膜上,用随机引物法进行探针标记,杂交条件同 Southern 杂交。

2 结果

2.1 核苷酸序列及相应氨基酸序列的获得

利用马铃薯高赖氨酸蛋白基因 *sb401* cDNA 为探针筛选谷子未成熟种子的 cDNA 文库,对获得的其中 1 个阳性克隆进行序列测定,结果表明,插入片段由 819 个核苷酸组成。使用 DNAMAN 及 Omega 2.0 软件对此序列进行分析,该基因编码蛋白最长的开放阅读框为 130 个氨基酸,分子质量为 14 113 u,等电点 pI = 9.75。在此阅读框之前发现有终止密码子存在,因此推测 130 个氨基酸的阅读框是 *f103* 编码的蛋白质(图 1 和 2)。该蛋白质富碱性氨基酸(24.6%)和酸性氨基酸(21.5%),含有 14 个甘氨酸、13 个丙氨酸、12 个丝氨酸和 10 个脯氨酸,不含色氨酸和酪氨酸。

2.2 *f103* 序列的同源比较

将 *f103* 核苷酸序列和推断的氨基酸序列在 GenBank 数据库、Swiss-Prot 数据库和 EST 数据库

```

1      CGGCACGAGGGGAATAGCTCCGGCCAAGTAGGAAATTTAAAGCCCCCCTGTGTCTAGTTGG
61     AGCCATTGGATACATGGGCAGGCCCTCGAGGGAAGAGCAAGAAAGCGATCGAAGCCGCGAG
1      M G R P R G K S K K A I E A A S
121    CAACGACGACGAAGATGGCAGCGGGCGGAGGAGGCGCCGCCACCCCAAGAGGAGGGG
17     N D D E D G S G G E E A P P T P K R R G
181    AAGGCCTCAGACGAAGCCTCTCAAGGACGACGCCGACGAAGCCGAGGACAAGGACACCGC
37     R P Q T K P L K D D A D E A E D K D T A
241    CGAGGCCGAGGAAGACGACGACGCGCACAAAGCCGTCGTGCGGCAAGCAAGGACTC
57     E A E E D D A D G T K P V V R P S K D S
301    CACCAAGCAGAGCTCAGCTGAGGGTGGAGGCAAGAAGCGGCGGCGCGGAGAGGAGGA
77     T K Q S S A E G G G K K R R R R R E E E
361    GGAGGGTCACGTGAGGTCTCAAGTCCAATGGGTTCCGGCCGAACGGGAGCCGGCGGAA
97     E G H V R S S K S N G F R P N G S R R K
421    GAGCACTCCCCGGCAGCCGCGGAGGCTGGGGTTGAATGCAAGTGAACATTGCCGGATA
117    S T P R R A A E A G V E C K *
481    AGCTAGAGCTACTTGTGTGGTTTTTACTCCCCCTTTTCTCCCCATTTTACTGGGGGTT
541    TTAGCTTTTACCCATTTCTGTTACATAGATCAAACAACATGCATCTTGAAGCGTCTGGTA
601    GTTCGTAGTCCCATCTCTAGGAGTAGGAGCATGATCATGTTTCATGCCATTAAGTCTGGT
661    CGGTGGATGACATGCTAGTTATAGCTTGTGAGTTCGTTTGCATGAGGGTACTAGTGACTG
721    GTGGTTTGGGTTAAACTTGAAGTATAAGCAAGCATTCTCTGCGAACTGTTTATCTGT
781    GTTTTGTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

```

*终止密码子

图 1 谷子 cDNA *f103* 的核苷酸序列和推导的氨基酸序列

Fig. 1 Nucleotide sequence and deduced amino acid sequence of the millet cDNA clone *f103*

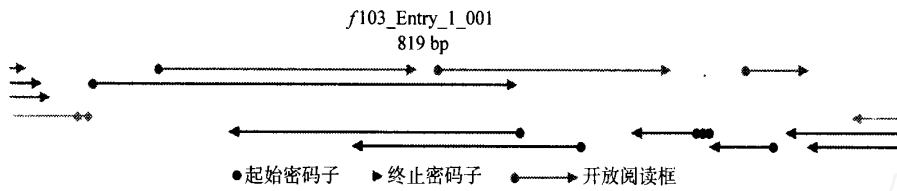


图 2 *f103* 开放阅读框分析

Fig. 2 ORF analysis of *f103*

中进行同源序列查询,结果见表 1。

综合查询结果,在 GenBank 数据库、Swiss-Prot 数据库查询显示 *f103* 基因仅与拟南芥 At2g42190 和拟南芥 At3g57930 有较低相似性。EST 数据库

查询显示,玉米、小麦、水稻等禾本科植物中含有与 *f103* 基因部分序列具有较高同源性的 EST 序列。由于 EST 序列不能提供基因功能方面的信息,*f103* 基因的功能还需进一步研究。

表 1 *f103* 数据库查询结果

Table 1 Blast results of *f103* in database

基 因	登陆号	同源性	数 据 库	方 法
At2g42190	AAB88644	34 %	All non-redundant GenBank CDS translations + PDB + SwissProt + PIR + PRF	blastx
At3g57930	NP-191352	31 %	All non-redundant GenBank CDS translations + PDB + SwissProt + PIR + PRF	blastx
Unknown in Arabidopsis	AAL87340	34 %	All non-redundant GenBank CDS translations + PDB + SwissProt + PIR + PRF	blastx
At2g42190 protein	AAB88644	34 %	Swiss - Prot + TrEMBL + TrEMBL - NEW	blastp
Wheat WHE4119_B09_D17ZS	BQ752790	23/ 33 (69 %)	GenBank + EMBL + DDBJ EST database	blastx
Maize 660035D06	AW360466	22/ 31 (70 %)	GenBank + EMBL + DDBJ EST database	blastx
Barely	AW982295	21/ 40 (52 %)	GenBank + EMBL + DDBJ EST database	blastx
HVSMEg0002M04f				
Rice OSJNEc04D03	CB653297	22/ 26 (84 %)	GenBank + EMBL + DDBJ EST database	blastx
Maize 605011D10	AI665454	23/ 30 (76 %)	GenBank + EMBL + DDBJ EST database	blastx

2.3 *f103* 的 Southern blot 分析

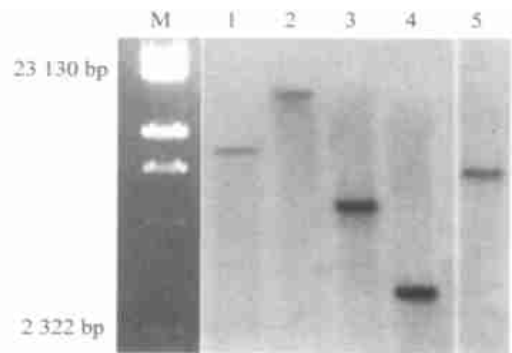
采用 5 种限制性内切酶 *Pst* , *EcoR* I, *Xba* , *Hind* 和 *Bam* H I 分别消化谷子基因组 DNA,以 $-^{32}$ P-dCTP 标记的 *f103* cDNA 为探针,进行基因组 Southern 杂交,结果 5 种不同的限制性内切酶酶解产物均只有 1 条杂交信号带(图 3),表明 *f103* 基因在谷子基因组中以单拷贝存在。

2.4 *f103* 的时空表达特异性

为确定 *f103* 基因在谷子中的发育表达模式,分别提取了谷子茎、叶、幼穗及授粉后 5, 15 和 25 d 的未成熟种子等材料的总 RNA,以 *f103* cDNA 为探针进行 Northern 杂交,结果 *f103* 基因在叶中不表达,在茎、幼穗、未成熟种子中均有杂交信号(图 4)。

2.5 *f103* 在不同植物中同源基因的 Southern 杂交分析

分别提取水稻、谷子、小麦、玉米、拟南芥、烟草

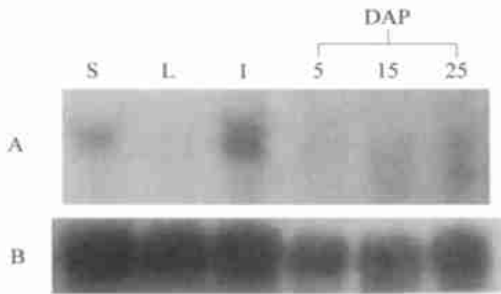


1—谷子基因组 DNA/*Pst* ; 2—*EcoR* I; 3—*Xba* I; 4—*Hind* ; 5—*Bam* H I; M—/*Hind* marker

图 3 *f103* 基因 Southern 杂交分析

Fig. 3 Southern Blotting analysis of millet genomic DNA with *f103* sequence as a probe

的基因组 DNA,用 *Hind* 进行酶切消化,以 $-^{32}$ P-dCTP 标记的 *f103* cDNA 为探针对其进行低严谨



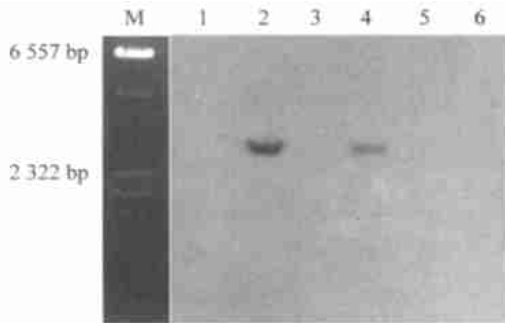
从谷子茎(S)、叶(L)、花序(I), 授粉后 5、15 和 25 d 的未成熟种子中提取 20 μ g 总 RNA, 电泳, 上样, 杂交。

A — 32 P-dCTP 标记 *f103* cDNA 为探针; B — 膜去探针后再以 28S rRNA 为探针。

图 4 Northern 杂交分析

Fig. 4 Northern Blotting analysis with the *f103* cDNA as a probe

条件下基因组 Southern 杂交, 结果只有玉米基因组 DNA 具有杂交信号, 而在水稻、小麦、拟南芥和烟草中则无任何杂交信号(图 5)。



M — *Hind* III 分子质量标准; 1 — 水稻 10 μ g; 2 — 谷子 15 μ g; 3 — 小麦 100 μ g; 4 — 玉米, 20 μ g; 5 — 拟南芥 5 μ g; 6 — 烟草 30 μ g

图 5 植物中 *f103* 同源基因的分析

Fig. 5 Genomic Southern DNA analysis of *f103* homologous gene in some plants

3 讨论

sb401 (cDNA) 为本实验室克隆的高赖氨酸蛋白基因, 其编码的蛋白质赖氨酸含量很高, 氨基酸个数比达 16.7%。用该基因和种子特异表达启动子构建表达载体转化玉米, 可使转基因玉米种子中赖氨酸和蛋白质含量明显提高^[4,5]。利用该 cDNA 筛选谷子未成熟种子 cDNA 文库, 对得到的阳性克隆之一 *f103* 进行序列分析, 结果其编码的蛋白质赖氨酸含量氨基酸个数比为 11.5%, 但该 cDNA 核苷酸序列和由其推导的氨基酸序列与 *sb401* 同源性都不高, 这可能是由于筛选时采用较宽松的杂交条件, 而且个别区段连续的小段核苷酸相似性较高造成的。

利用 DNAMAN 和 OMEGA 2.0 等生物学软件对克隆的 *f103* 基因进行分析, 结果表明该基因编码蛋白质亲水性区域占优势, 是一个亲水性蛋白质。可能定位于核内^[6,7]。

利用 Omega 2.0 软件对该蛋白质可能的修饰位点进行分析, 发现该蛋白有多个位点可以发生磷酸化修饰, 个别位点可以发生糖基化修饰。利用 Net-Phos 2.0 预测该蛋白质的磷酸化位点, 发现存在多个丝氨酸和苏氨酸磷酸化位点。蛋白质的磷酸化和脱磷酸化在细胞信号传导中起重要作用^[8], 对于调节植物体生长和发育具有重要意义, *F103* 蛋白可能受激酶的调节而发挥作用。磷酸化、糖基化位点可能是 *F103* 蛋白完成其生理功能的结构基础。

另外, 我们已经将该基因和 35S 启动子构建表达载体, 转化烟草, 获得转基因烟草植株, 通过对转基因烟草的分子水平和细胞水平的检测, 将对其功能进行进一步研究。

参 考 文 献

- [1] Bouchez D, Höfte H. Functional genomics in plants[J]. *Plant physiol*, 1998, 118:725 ~ 732
- [2] Liu J Q, Seul U, Thompson R. Cloning and characterization of a pollen-specific cDNA encoding a glutamic-acid-rich protein (GARP) from potato *Solanum berthaultii* [J]. *Plant Mol Biol*, 1997, 33:291 ~ 300
- [3] 萨姆布鲁克 J, 菲里奇 E F, 曼尼阿蒂斯 T. 分子克隆实验指南[M]. 第 2 版. 金冬雁, 黎孟枫, 等译. 北京: 科学出版社, 1996. 483 ~ 490
- [4] 张秀君, 刘俊起, 赵倩, 等. 用基因枪将高赖氨酸基因导入玉米及转基因植株的检测[J]. *农业生物技术学报*, 1999, 7(4): 363 ~ 367
- [5] 孙学辉, 敖光明, 于静娟, 等. 高赖氨酸基因导入玉米自交系的研究[J]. *农业生物技术学报*, 2001, 9(2): 156 ~ 158
- [6] Nielsen H, Engelbrecht J, Brunak S, et al. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites[J]. *Prot Eng*, 1997, 10, 1 ~ 6
- [7] Emanuelsson O, Nielsen H, Brunak S, et al. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence[J]. *J Mol Biol*, 2000, 300: 1005 ~ 1016
- [8] 汪斌, 李维明. 植物激素的信号转导系统研究进展[J]. *福建农业大学学报*, 2001, 30(4): 433 ~ 438