

西尔斯山羊草(*Aegilops searsii*) α -醇溶蛋白编码基因的克隆及原核表达

刘国娟¹, 马信¹, 尹华燕¹, 孙鑫¹, 杜旭焯¹, 王宏伟¹, 李安飞¹, 陈呈涛², 孔令让^{1*}

1. 山东农业大学农学院/作物生物学国家重点实验室, 山东 泰安 271018

2. 山东省新泰市天宝镇农业技术推广站, 山东 新泰 271200

摘要: 醇溶蛋白是小麦加工品质的重要影响因素, 主要决定面团的粘性和延展性。根据酸性条件下迁移率的不同, 可将醇溶蛋白分为 α -、 β -、 γ -、 ω -醇溶蛋白。 α -醇溶蛋白占贮藏蛋白的15%~30%, 是含量最丰富的一类贮藏蛋白, 同时, α -醇溶蛋白中含有一些致敏性的毒性多肽。克隆西尔斯山羊草 α -醇溶蛋白基因, 并进行序列分析, 通过构建原核表达载体, 使其在大肠杆菌中诱导表达融合蛋白, 并通过切胶纯化法, 获得纯度较高的目的蛋白。根据 α -醇溶蛋白基因编码区保守序列设计引物, 通过PCR扩增克隆得到 α -醇溶蛋白基因并进行序列分析。将目的基因KC421089连到表达载体pEASY-E1上, 在大肠杆菌BL21(DE3)中经IPTG诱导表达, 获得表达的融合蛋白。通过切胶纯化法, 获得纯度较高的目的蛋白。从小麦近缘植物西尔斯山羊草(*Aegilops searsii*)中首次克隆了7个 α -醇溶蛋白编码基因, 它们的编码区长度分布在849~954 bp之间, 可编码282~317个氨基酸。通过与已发表的其它物种的 α -醇溶蛋白氨基酸序列进行多重比对分析, 发现这些基因都具有 α -醇溶蛋白编码基因典型的结构特点, 同时还存在着碱基的缺失、插入以及SNPs。在克隆目的基因的基础上, 本研究还通过大肠杆菌体外表达和切胶纯化方法, 获得了纯度较高的目的蛋白, 实现了该基因的表达。克隆了7个 α -醇溶蛋白基因序列, 登录号为KC421089的基因可在原核系统中成功表达, 并通过切胶纯化法获得目的蛋白, 为进一步利用西尔斯山羊草改良小麦加工品质奠定了基础。

关键词: α -醇溶蛋白; 西尔斯山羊草; 序列分析; 原核表达

中图分类号: S5

文献标识码: A

文章编号: 1000-2324(2015)03-0321-05

Cloning and Prokaryotic Expression of α -Gliadin Genes from *Aegilops searsii*

LIU Guo-juan¹, MA Xin¹, YIN Hua-yan¹, SUN Xin¹, DU Xu-ye¹, WANG Hong-wei¹, LI An-fei¹, CHEN Cheng-tao², KONG Ling-rang^{1*}

1. Agronomy College, Shandong Agricultural University/State Key Laboratory of Crop Biology, Taian 271018, China

2. Agriculture Technology Generalization Station of Tianbao Town Xintai City Shandong Province, Xintai 271200, China

Abstract: Gliadins are the major influence factors of wheat processing quality, which determine dough extensibility. On the basis of the electrophoretic mobility by acidic polyacrylamide gel electrophoresis, gliadin can be separated into α -、 β -、 γ - and ω -gliadin. Among them, α -gliadins are the most abundant and accounted for 15%~30% of the wheat storage proteins. On the other hand, α -gliadins can cause different diseases. The present study aimed at cloning and analyzing the α -gliadin genes from *Aegilops searsii*, expressing it in *E.coli* and obtaining high purified proteins. The α -gliadin genes were amplified by PCR and then inserted the gene KC421089 into pEASY-E1. The recombinant plasmids were expressed in the *E.coli* BL21(DE3) and then the purified proteins were obtained by cutting the gel slices. Seven novel α -gliadin genes were cloned from *Aegilops searsii*. Their length of the open reading frames ranged from 849-954bp, encoding the putative proteins with 282-317 amino acid residues, respectively. A BLAST search showed that these sequences have the typical structure of α -gliadin genes and SNPs and In/Dels. Moreover, the target proteins were expressed by *E. coli* and highly purified proteins were obtained by cutting the gel slices. Seven novel α -gliadin genes were cloned from *Aegilops searsii*, and the gene KC421089 were expressed by *E. coli* and highly purified proteins were obtained by cutting the gel slices. This study laid a good foundation for wheat quality improvement.

Keywords: α -gliadin; *Aegilops searsii*; Sequence analysis; prokaryotic expression

醇溶蛋白是小麦种子中的主要贮藏蛋白之一, 约占小麦胚乳贮藏蛋白总量的50%~60%^[1], 是影响小麦加工品质的重要因素。在酸性电泳中, 根据迁移率大小可将其分为 α -、 β -、 γ -、 ω -醇溶蛋白^[2,3]。其中, α -醇溶蛋白占小麦籽粒总蛋白的15~30%^[4-6], 平均分子量为31 kD, 主要由位于第六同源染色体群短臂上的*Gli-2* (*Gli-A2*, *Gli-B2*和*Gli-D2*)位点控制^[7]。

乳糜泻(Celiac disease, CD)又称为小麦麸质过敏, 是小麦诱发的一种特殊过敏症状, 临床表现为

收稿日期: 2014-04-12

修回日期: 2014-05-07

基金项目: 国家转基因专项“优质转基因小麦新品种培育(2011ZX08002004-003)

作者简介: 刘国娟(1987-),女,在读硕士研究生,主要从事小麦及其近缘植物品质相关基因克隆及功能验证 E-mail:liuguojuan0814@126.com

*通讯作者: Author for correspondence. E-mail:lkong@sdau.edu.cn

肠黏膜损害和继发性吸收不良并伴有顽固腹泻、体重下降、营养不良、骨质疏松症、贫血、乏力、手足抽搐等症状,严重影响食品安全和人类健康^[8]。 α -醇溶蛋白基因编码的氨基酸序列中含有能够诱发 CD 的毒性多肽序列,这些多肽序列与病人的 T 细胞结合会产生小肠粘膜损伤和吸收不良症状。朱西平等根据推导氨基酸序列所具有的四种 CD 多肽结构,对普通小麦品种中克隆的 α -醇溶蛋白基因进行染色体定位,并用中国缺-四体对该染色体定位方法进行验证^[9]。目前,对普通小麦中 CD 多肽的研究已有些报道,并确认了一系列的毒性多肽,但是关于二倍体材料中 CD 抗原多肽的研究还相对较少。

对基因功能研究的方法主要有转基因、体外表达、基因沉默等。原核表达是一种重要的体外表达模式,在基因功能研究方面发挥了重要作用。Claudia 等^[10]和 Ferrante 等^[11]利用 *E. coli* 表达系统实现了普通小麦 γ -醇溶蛋白基因的原核表达,并通过体外掺粉实验对表达产物进行了体外功能验证。由于 α -醇溶蛋白序列具有较高的多态性,对蛋白的纯化及表达技术提出了较高的要求,因此对其进行体外功能鉴定的研究尚处于起步阶段,有必要对其进行进一步的研究。

西尔斯山羊草是小麦近缘种,是小麦育种改良的重要资源,本研究旨在从西尔斯山羊草中克隆出 α -醇溶蛋白的编码基因,并通过原核表达初步验证该基因功能,为今后进一步利用西尔斯山羊草改良小麦面粉加工品质奠定分子基础。

1 材料与方法

1.1 材料来源

西尔斯山羊草 Y2131(*Aegilops searsii*, S^sS^s, 2n=2x=14),由中国农业科学院作物科学研究所提供。

1.2 基因组 DNA 提取

采用 CTAB 法从西尔斯山羊草的幼叶中进行植物总 DNA 的提取^[12]。

1.3 基因的克隆及序列分析

根据 GenBank 公布的 α -醇溶蛋白基因的保守序列,设计引物扩增编码区全长,引物序列为, P1: 5'-ATG AAG ACC TTT CTC ATC CTT G-3', P2: 5'-TCA GTT RGT ACC RAA GAT GCC-3'。利用在线软件 <http://web.expasy.org/translate/> 进行氨基酸的翻译,并利用 MEGA 5.0 软件进行系统发育树的构建;同时,在利用 DNAMAN 软件进行序列比对分析的基础上,辨识 CD 毒性多肽的识别区域(参照朱西平等的方法进行^[8])。

1.4 原核表达与纯化

根据目的基因的完整编码区,设计一对引物扩增不含有信号肽编码序列的 ORF,引物序列为, P3: 5'-GCA GTT AGA GTT CCA GTG CCA-3', P4: 5'-TCA GTT RGT ACC RAA GAT GCC-3',将目的片段回收纯化后连接到 pEASY-E1 载体上,转化至 Trans1-T1 感受态细胞中,并进行菌液测序。对测序正确的菌液进一步提取质粒 DNA,并将重组质粒转化到大肠杆菌 BL21(DE3)感受态细胞中,挑取单克隆,经菌落 PCR 验证正确后进行诱导表达,然后通过切胶纯化法得到目的蛋白^[13,14]。

2 结果与分析

2.1 α -醇溶蛋白的克隆及序列分析

利用一对保守引物 P1 和 P2,对西尔斯山羊草基因组 DNA 进行扩增,得到约 1000 bp 左右的目的条带(图 1),经测序共获得 7 个基因,通过在 Genbank 中的在线搜索,结果显示它们与已知的

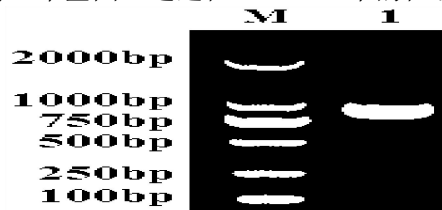


图 1 西尔斯山羊草 α -醇溶蛋白基因的 PCR 扩增
Fig.1 PCR amplification of the α -gliadin gene from *Aegilops searsii*

M: DL2000; 1: Y2131

α -醇溶蛋白基因具有很高的相似性, 同时与已报道的基因在序列上存在一定差异, 表明本试验得到的 α -醇溶蛋白编码基因均为新的 α -醇溶蛋白基因。随后, 将这些基因序列提交至 Genbank (登录号: KC421084-KC421090)。对核苷酸序列编码的氨基酸序列分析发现核苷酸序列长度为 849-954 bp (表 1), 编码氨基酸数目为 282~317 个, 等电点具有较高的多态性。通过与已发表的 α -醇溶蛋白序列进行多重比对分析(图 2), 这些基因都具有 α -醇溶蛋白编码基因典型的结构特点^[15], 并且都具有 6 个半胱氨酸残基, 同时还存在着碱基的缺失、插入以及 SNPs。

表 1 西尔斯山羊草 α -醇溶蛋白的特征分析

Table 1 Characterization of 7 novel α -gliadin from *Aegilops searsii*

序列号 Genbank No.	序列长度(bp) Length of gene (bp)	氨基酸数目 Deduced amino acids	分子量 (KD) Molecular weight (KD)	等电点(pI) pI
KC421084	849	282	32.40	7.64
KC421085	954	317	36.57	8.56
KC421086	909	302	34.82	7.65
KC421087	900	299	34.33	7.61
KC421088	936	311	35.56	8.24
KC421089	849	282	32.41	7.64
KC421090	924	307	35.07	8.24

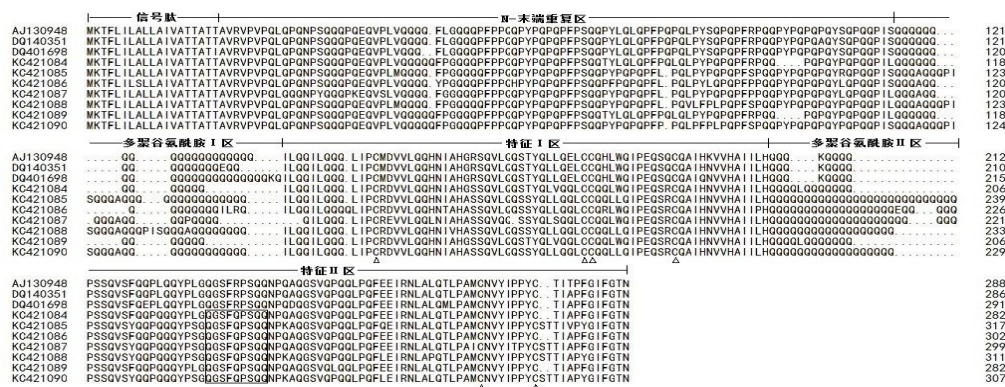


图 2 西尔斯山羊草 α -醇溶蛋白基因序列与斯卑尔脱小麦的 AJ130948, 圆锥小麦的 DQ140351, 栽培一粒小麦的 DQ401698 的推导氨基酸序列比较

注: Δ 表示保守的半胱氨酸残基, 方框中的序列为 gliadin α 毒性多肽

Fig.2 The deduced amino acid sequences of *Aegilops searsii* α -gliadin genes compared with *Triticum spelta* derived sequence AJ130948; *Triticum turgidum* derived sequence DQ140351 and *Triticum monococcum* derived sequence DQ401698

Note: Δ represents conserved cysteine residues. The boxed letters are gliadin α toxic epitope

2.2 CD 毒性多肽的识别与分析

研究表明, 在 α -醇溶蛋白中共含有 4 种毒性多肽, 分别为 gliadin- α 9, gliadin- α 20, gliadin- α , gliadin- α 2, 它们的分布存在基因组特异性。A 基因组编码的 α -醇溶蛋白含有 gliadin- α 9 和 gliadin- α 20, B 基因组编码的 α -醇溶蛋白不含有或者仅含有 gliadin- α , 而来源于 D 基因组编码的 α -醇溶蛋白含有三种或者全部 4 种毒性多肽^[6]。对本实验中获得 α -醇溶蛋白的毒性多肽进行分析, 发现它们都只含有 gliadin- α 毒性多肽(表 2), 符合 B 染色体编码的 α -醇溶蛋白的特征。

表 2 西尔斯山羊草 α -醇溶蛋白具有的 T 细胞抗原表位数量及类型

Table 2 Number and type of the four T cell stimulatory toxic epitopes in the α -gliadin from *Aegilops searsii*

Genbank No.	gliadin- α	gliadin- α 2	gliadin- α 9	gliadin- α 20
KC421084	1	0	0	0
KC421085	1	0	0	0
KC421086	1	0	0	0
KC421087	1	0	0	0
KC421088	1	0	0	0
KC421089	1	0	0	0
KC421090	1	0	0	0

2.3 α -醇溶蛋白聚类分析

将克隆得到的核苷酸序列和 Genbank 中具有 α -醇溶蛋白完整编码阅读框的序列进行同源比对, 构建 Neighbor-joining 树(图 3)。结果显示, 共形成了 4 个大的分支, 分别是 A 基因组、D 基因组、S⁵基因组和 B 基因组编码的 α -醇溶蛋白基因。其中, 有两个 S⁵基因组基因和 D 基因组基因亲缘关系较近, 此外, 大多数来源于 S⁵基因组的 α -醇溶蛋白与来源于小麦 B 基因组的 α -醇溶蛋白亲缘关系相

对较近, 但是, 二者却位于两个不同的分支。

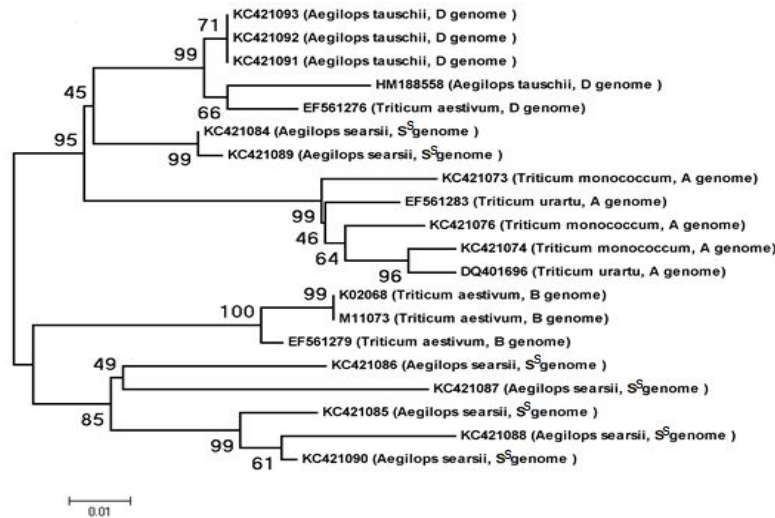


图 3 α -醇溶蛋白基因的系统进化分析
Fig.3 Phylogenetic analysis of α -gliadin genes.

2.4 基因的诱导表达与检测

将登录号为 KC421089 的重组质粒转入宿主菌 BL21(DE3)中, 挑取单克隆培养, 培养物按一定比例接种于选择培养基, 当 OD₆₀₀ 达到 0.6 时, 加入 IPTG 诱导剂, 然后在 37 °C 条件下诱导表达 10 h. 用蛋白提取液提取总蛋白进行 SDS-PAGE 检测, 结果表明, 在 18.8~35KD 之间有一条诱导出的明显的蛋白条带, 与预测的蛋白分子量大小基本吻合。随后将其进行切胶纯化, 得到纯度较高的目的蛋白(图 4)。目的蛋白的成功诱导表达及纯化, 为研究西尔斯山羊草 α -醇溶蛋白基因对小麦加工品质的效应奠定了理论基础。

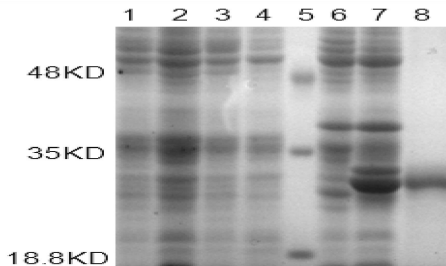


图 4 诱导表达产物和纯化产物的 SDS-PAGE 鉴定分析

注: 1:未诱导的 BL21(DE3);2:诱导后的 BL21(DE3);3:未诱导的转化空载体的菌株;4:诱导的转化空载体的菌株;
5:Marker;6:转化重组质粒后未经诱导的菌株;7:转化重组质粒经诱导的菌株;8:纯化的目的蛋白

Fig.4 Analysis of expressed and purified proteins by SDS-PAGE

Note: 1: Protein of BL21(DE3) without induction; 2: Protein of BL21(DE3) induced by IPTG; 3: BL21(DE3) with pEASY-E1 without induction; 4: BL21(DE3) with pEASY-E1 induced by IPTG; 5: Protein molecular weight marker; 6: Protein of recombinant plasmid without induction; 7: Protein of recombinant plasmid induced by IPTG; 8: Purified protein.

3 讨论

小麦面筋中常常含有对乳糜泻病人具有毒性的多肽, 这些多肽通过激活人体内免疫系统的 T 细胞, 使人体产生一定的炎症, 从而影响人们的健康。目前, 对预防 CD 的最有效措施是严格避免摄入含致敏源的食品。前人研究表明, 面筋中含有的 CD 毒性多肽的种类和数量越多, CD 的发病率也越高^[16]。Van Herpen 等人研究表明, 在小麦 B 基因组中克隆到的 α -醇溶蛋白基因序列一般不含有毒性多肽^[6], 而在本实验中克隆获得的来源于 S^s基因组的序列都含有 glia- α 毒性多肽, α -醇溶蛋白中的毒性多肽在 S^s基因组和小麦 B 基因组上存在较大差异, 因此, 在利用该材料进行小麦新品种培育时, 应避免含有毒性多肽 α -醇溶蛋白的引入。

前人研究表明, 乌拉尔图小麦是小麦 A 基因组的供体, 粗山羊草是小麦 D 基因组的供体^[17], 而关于小麦 B 基因组的供体目前还没有定论。一般认为拟斯卑尔脱山羊草单独或与其近缘植物共同作为小麦 B 基因组的供体^[15]。本研究结果表明, 栽培一粒小麦和乌拉尔图小麦亲缘关系较近, 粗山羊

草与小麦 D 基因组亲缘关系较近,而来源于西尔斯山羊草的 α -醇溶蛋白与来源于小麦 B 基因组的 α -醇溶蛋白尽管表现出一定的同源性,但却分布在两个不同的分支(图 3),说明 S^s基因组可能没有参与小麦 B 基因组的进化过程。

蛋白的体外表达为小麦种子贮藏蛋白的功能研究提供了一条有效途径^[11]。蛋白体外掺粉实验需要将外源的醇溶蛋白添加到基础面粉中,然后用揉混仪进行测定即可得知其对小麦品质的影响,为研究醇溶蛋白对面粉加工品质的影响提供了更快速、更简便的途径。目前,虽然通过克隆获得了大量的醇溶蛋白基因,但是对其进行功能验证的研究还相对较少。本研究通过大肠杆菌原核表达以及切胶纯化的方法,获得了纯度较高的目的蛋白,初步实现该目的基因的表达,所获得的表达产物可用于后续的体外掺粉实验,验证其对面粉加工品质的影响。

4 结论

4.1 α -醇溶蛋白基因的克隆和序列分析

本研究首次从西尔斯山羊草中克隆得到 7 个 α -醇溶蛋白基因,登录号分别为 KC421084-KC421090;序列分析表明,这 7 个基因序列都具有 α -醇溶蛋白基因家族的结构特征,并且都含有 6 个半胱氨酸残基(图 2),系统进化分析表明,栽培一粒小麦和乌拉尔图小麦亲缘关系较近,粗山羊草与小麦 D 基因组亲缘关系较近,而来源于西尔斯山羊草的 α -醇溶蛋白与来源于小麦 B 基因组的 α -醇溶蛋白尽管表现出一定的同源性,但却分布在两个不同的分支。

4.2 α -醇溶蛋白基因的原核表达

将登录号为 KC421089 的西尔斯山羊草 α -醇溶蛋白基因与原核表达载体连接,并成功地在 大肠杆菌中实现了表达,获得了与理论值相符的融合蛋白。此外,本实验还通过切胶纯化法获得了纯度较高的目的蛋白,对研究西尔斯山羊草 α -醇溶蛋白基因的品质效应提供了理论参考。

参考文献

- [1] 郝春燕,李建国,李雅轩,等.小麦醇溶蛋白基因克隆研究进展[J].首都师范大学学报,2006(2):67-70
- [2] Shewry P R, Halford N G, Lafandra D. Genetics of wheat gluten proteins[J]. Advances in Genetics, 2003(49):111-184
- [3] Wieser H. Chemistry of gluten proteins [J]. Food Microbiology, 2007(24):115-119
- [4] Chen F, Xu C, Xia G, et al. A new alpha-gliadin gene family for wheat breeding: somatic introgression line II-12 derived from *Triticum aestivum* and *Agropyron elongatum*[J]. Molecular Breeding, 2008(22):675-685
- [5] Gu Y Q, Crossman C, Kong X Y, et al. Genomic organization of the complex alpha-gliadin gene loci in wheat[J]. Theoretical and Applied Genetics, 2004(109):648-657
- [6] Van Herpen, Goryunova S V, et al. Alpha-gliadin genes from the A, B, and D genomes of wheat contain different sets of celiac disease epitopes[J]. BMC Genomics, 2006(7):1-13
- [7] Metakovsky E V, Akhmedov M G, Sozinov A A. Genetic analysis of gliadin-encoding genes reveals gene clusters as well as single remote genes[J]. Theoretical and Applied Genetics, 1986(73):278-285
- [8] Chiara B, Diana S, Armin A. Celiac disease: from gluten to autoimmunity[J]. Autoimmunity Reviews, 2008(7):644-650
- [9] 朱西平,李鑫,李雅轩,等.普通小麦及近缘粗山羊草 α -醇溶蛋白基因的克隆、定位与进化分析[J].作物学报,2010,36(4):580-589
- [10] Claudia G, Benitez C, Hélène R, et al. Cloning, bacterial expression, purification and structural characterization of N-terminal-repetitive domain of γ -Gliadin[J]. Protein Expression and Purification, 2006(46):358-366
- [11] Ferrante P, Masci S, D'Ovidio R, et al. A proteomic approach to verify in vivo expression of a novel α -gliadin containing an extra cysteine residue[J]. Proteomics, 2006(6):1908-1914
- [12] Stein N, Herren G, Keller B. A new DNA extraction method for high-throughput marker analysis in a large-genome species such as *Triticum aestivum*[J]. Plant Breeding, 2001(4):354-356
- [14] 高慎阳,查恩辉,王 珅,等.一种“高性价比”切胶纯化原核表达蛋白的方法[J].中国农学通报,2010,26(22):24-26
- [15] 于在江,马学恩,周建华.切胶纯化表达蛋白包涵体的可行性分析[J].生物技术,2007,17(3):46-48
- [16] Anderson O D, Litts J C, Greene F C. The α -gliadin gene family: I characterization of ten new wheat α -gliadin genomic clones, evidence for limited sequence conservation of flanking DNA, and southern analysis of the gene family[J]. Theoretical and Applied Genetics, 1997(95):50-58
- [17] Koining F. Celiac disease: quantity matters [J]. Seminars in immunopathology, 2012(4):541-549
- [18] 曹亚萍.小麦的起源、进化与中国小麦遗传资源[J].小麦研究,2008,29(3):1-10