

郑丰5号 α -醇溶蛋白基因的克隆与序列分析

李玉阁 邢冉冉 崔一飞 李锁平

(河南大学 生命科学院,河南 开封 475004)

摘要: α -醇溶蛋白是小麦籽粒贮藏蛋白的重要组分,其组成与含量对小麦加工品质具有重要影响。利用PCR从郑丰5号基因组中克隆 α -醇溶蛋白基因,并对其序列进行分析。经克隆共获得32个 α -醇溶蛋白新基因(*ZF5A-1* ~ *ZF5A-32*, GenBank 注册序列号为JX828280 ~ JX828311),其中15个为假基因,17个(*ZF5A-1* ~ *ZF5A-17*)具有完整开放阅读框。17个 α -醇溶蛋白新基因中,除*ZF5A-1*、*ZF5A-3*、*ZF5A-6*、*ZF5A-9*、*ZF5A-10*、*ZF5A-11*、*ZF5A-15*编码的蛋白在特征Ⅱ区含有1个额外的半胱氨酸(C)外,其他10个基因编码的蛋白均具有 α -醇溶蛋白的典型结构。根据推断氨基酸序列中4种主要T细胞优势多肽的分布及多聚谷氨酰胺区的长度,推测*ZF5A-7*和*ZF5A-12*可能定位于6A染色体,*ZF5A-4*、*ZF5A-13*、*ZF5A-14*和*ZF5A-17*可能定位于6B染色体,而*ZF5A-1* ~ *ZF5A-3*、*ZF5A-5*、*ZF5A-6*、*ZF5A-8* ~ *ZF5A-11*、*ZF5A-15*和*ZF5A-16*可能定位于6D染色体。17个新克隆 α -醇溶蛋白基因及4个已知 α -醇溶蛋白基因编码的蛋白的二级结构预测结果表明: α -螺旋、 β -折叠的位置和核心序列是相对保守的,但不同蛋白 α -螺旋和 β -折叠的数量以及参与形成同一保守区域 α -螺旋和 β -折叠的氨基酸残基数却并不相同。克隆的17个 α -醇溶蛋白基因中,除*ZF5A-17*编码的蛋白缺少 α -螺旋(H_2)、*ZF5A-2*、*ZF5A-8*编码的蛋白在特征区Ⅰ均存在1个额外的 α -螺旋(H_{E1})、*GQ891685*和*ZF5A-15*编码的蛋白在多聚谷氨酰胺Ⅱ区存在1个额外的 α -螺旋(H_{E2})外,5个保守的 α -螺旋(H_1 ~ H_5)恒定出现在其他基因的2个谷氨酰胺重复区和特征区中;此外,在C-末端特征区大部分基因(61.11%)还形成1个 β -折叠结构(E)。郑丰5号中具有较多额外半胱氨酸、 α -螺旋和 β -折叠的 α -醇溶蛋白基因,可能与其良好的加工品质密切相关。

关键词: α -醇溶蛋白; 郑丰5号; 克隆; 序列分析

中图分类号: S512.1; Q78 文献标识码: A 文章编号: 1000-7091(2013)04-0046-07

Cloning and Sequencing of α -gliadin Genes from Common Wheat Cultivar Zhengfeng 5

LI Yu-ge, XING Ran-ran, CUI Yi-fei, LI Suo-ping

(College of Life Science, Henan University, Kaifeng 475004, China)

Abstract: The α -gliadins are the important components of seed storage proteins, and their composition and content play a major role in determining the processing quality of wheat. Using PCR technology, total 32 novel genes (Designated as *ZF5A-1* ~ *ZF5A-32*; GenBank No: JX828280 ~ JX828311) were cloned from the wheat cultivar Zhengfeng 5, including 17 full-ORF α -gliadin genes (Designated as *ZF5A-1* ~ *ZF5A-17*) and 15 pseudogenes. The sequence analysis of deduced amino acids showed that 10 had the typical structural characteristics of α -gliadin genes reported previously, and 7 (*ZF5A-1*, *ZF5A-3*, *ZF5A-6*, *ZF5A-9*, *ZF5A-10*, *ZF5A-11*, *ZF5A-15*) possessed an additional cysteine residue in the unique domain II. Based on the occurrence of the four major T cell immunogenic peptides and glutamine residues in the glutamine repeats, *ZF5A-7* and *ZF5A-12* could be assigned to chromosome 6A, *ZF5A-4*, *ZF5A-13*, *ZF5A-14* and *ZF5A-17* to chromosome 6B, and other 11 genes to chromosome 6D. The secondary structure predication of 17 cloned genes and other four genes in public database demonstrated that the numbers and distributions of α -helix and β -strand were variable in different mature α -gliadins. Generally, 5 α -helices (H_1 ~ H_5) usually presented in the two glutamine repeats and two unique domains for majority of α -gliadins, but the

收稿日期: 2013-05-06

基金项目: “十二五”农村领域国家科技计划项目(2011BAD07B00; 2012AA101105); 自然科学基金面上项目(31271713)

作者简介: 李玉阁(1977-), 女, 河南南阳人, 讲师, 在读博士, 主要从事麦类作物的遗传学研究。

通讯作者: 李锁平(1962-), 男, 河南商丘人, 教授, 博士, 主要从事小麦遗传育种研究。

putative proteins of *ZF5A-17* lost the α -helix (H_2), *ZF5A-2* and *ZF5A-8* had an additional α -helix in the unique domain I (H_{E1}) and *GQ891685* and *ZF5A-15* had an additional α -helix in the second glutamine repeat (H_{E2}). In addition, most of the genes (61.11%) had a β -strand (E) in the C-terminal unique domain II. High variations in numbers of cysteine residue, α -helices and β -strands in the α -gliadins of Zhengfeng 5 might be strongly related to its good rheological properties of dough.

Key words: α -gliadin; Zhengfeng 5; Gene cloning; Sequence analysis

麦谷蛋白 (Glutenin) 和醇溶蛋白 (Gliadin) 是面筋蛋白的两大主要组分,二者均富含脯氨酸和谷氨酰胺,具有高度的异质性和复杂性,是影响小麦加工品质的重要因素^[1-2]。根据酸性 (pH 值 3.1) 聚丙烯酰胺凝胶电泳 (A-PAGE) 迁移率的不同,小麦醇溶蛋白又分为 α - β - γ - ω -醇溶蛋白 4 种。其中,由第六同源染色体组群 (6A、6B、6D) 短臂的 *Gli-2* 位点编码的 α -醇溶蛋白含量最丰富,占种子贮藏蛋白的 15% ~ 30%,是人类消费量最大的蛋白质^[3-6]。Khatkar 等^[7]通过向基础面粉中添加总醇溶蛋白和醇溶蛋白的不同组分 (α 、 β 、 γ 、 ω) 发现,醇溶蛋白及其组分会显著降低面筋强度、增大面包体积,而 α -醇溶蛋白对面筋强度的降低作用最小 ($\omega > \gamma > \beta > \alpha$),对面筋强度的增大效果最大。李敏等^[8]通过体外功能鉴定也发现,由 *GQ891685* 编码的含有 1 个额外半胱氨酸的 α -醇溶蛋白亚基 ACX71610 能显著增加面筋弹性,降低面筋强度。

醇溶蛋白是乳糜泻病 (Celiac disease, CD) 又称面筋蛋白过敏性肠病的外源诱导因子,而 α -醇溶蛋白因含有较多的能诱发 CD 发生的 T 细胞免疫肽段,是诱导 CD 发生的主要活力蛋白^[9-14]。目前,已发现大量分布于醇溶蛋白的 T 细胞免疫肽,其中 5 种具有明显较高的免疫活性,包括 *Glia- γ 1* (QQPQQSEFPQQQ)、*Glia- α 9* (PFPQPQLPY)、*Glia- α 2* (PQPQLPYPQ)、*Glia- α 20* (PFRPQQPYPQ) 和 *Glia- α* (QGSFQPSQQ),称为 T 细胞免疫优势多肽。除 *Glia- γ 1* (QQPQQSEFPQQQ) 分布于 γ -醇溶蛋白外,其他 4 种 T 细胞免疫优势多肽均分布于 α -醇溶蛋白^[11,13]。研究显示,这 4 种 T 细胞免疫优势多肽的数量和分布组合存在明显的基因组来源特异性^[10-11]。来源于一粒小麦 (*Triticum monococcum*, AA) 或来源于普通小麦 6A 染色体的 α -醇溶蛋白通常仅含有 *Glia- α 9* 和 *Glia- α 20*; 来源于拟斯脾尔脱山羊草 (*Aegilops speltoides*, SS) 或普通小麦 6B 染色体的 α -醇溶蛋白,大多不含有任何一种 T 细胞优势多肽或仅含有 *Glia- α* ; 而来源于粗山羊草 (*Aegilops tauchii*, DD) 或普通小麦 6D 染色体的 α -醇溶蛋白则可能含有 4 种 T 细胞免疫优势多肽 (不同蛋白的组合情况不同),而

且有些蛋白中甚至还分布有多个 *Glia- α 2*^[14]。因此,可根据 4 种主要 T 细胞优势多肽的分布对 α -醇溶蛋白基因进行染色体定位。

郑丰 5 号是河南省主要推广种植的优质弱筋小麦品种之一。为了解其 α -醇溶蛋白基因的组成特点,并为 α -醇溶蛋白基因的结构和功能分析提供参考,对郑丰 5 号的 α -醇溶蛋白基因进行克隆和序列分析,并采用 PSIPRED 3.3 对 α -醇溶蛋白的二级结构进行预测。

1 材料和方法

1.1 材料及试剂

优质弱筋小麦品种郑丰 5 号种子由河南省农业科学院小麦研究所胡琳研究员提供。菌株 *E. coli* DH5 α 购自上海莱枫生物科技有限公司, DNA 聚合酶及 PCR 相关试剂、琼脂糖凝胶回收试剂盒 (Ver 2.0)、克隆载体 pMD19-T 购自宝生物工程 (大连) 有限公司。引物由北京六合华大基因科技股份有限公司完成。

1.2 方法

1.2.1 植物材料培养及其基因组 DNA 的提取 随机选取 10 ~ 20 粒郑丰 5 号种子于温室中培养至长成黄化苗,采用改良的 CTAB 法^[15]从黄化苗嫩叶中提取基因组 DNA。

1.2.2 α -醇溶蛋白基因的克隆与鉴定 根据 GenBank 中已注册 α -醇溶蛋白基因序列,通过 Primer 5.0 和 Oligo 6.0 设计并优化得到 1 对简并引物 F: 5'-ATGAAGACCTTTCTCATCCT-3' 和 R: 5'-TCAGT TRGTACCGAAGATGCC-3',该引物可扩增从起始密码子 ATG 到终止密码子 TGA 之间的完整编码区。PCR 扩增体系、反应条件参照李玉阁等^[16]的方法。50 μ L PCR 反应体系中含 100 ng 左右的 DNA 模板、100 μ mol/L dNTP、2 \times GC Buffer I 25 μ L、0.5 μ mol/L 引物 F 和 R、1.5 U LA Taq 酶。PCR 反应条件为 94 $^{\circ}$ C 预变性 4 min; 然后 94 $^{\circ}$ C 变性 30 s, 62 $^{\circ}$ C 退火 45 s, 72 $^{\circ}$ C 延伸 60 s, 共 32 个循环; 最后 72 $^{\circ}$ C 延伸 15 min。PCR 产物经 2% 琼脂糖凝胶电泳,用 DNA 凝胶回收试剂盒 Ver 2.0 对目的片段进行回收

纯化,纯化产物连接到 pMD19-T 载体,并转化大肠杆菌感受态细胞 DH5 α 。通过菌液 PCR 鉴定阳性克隆,并送至北京六合华大基因科技股份有限公司进行双向测序。

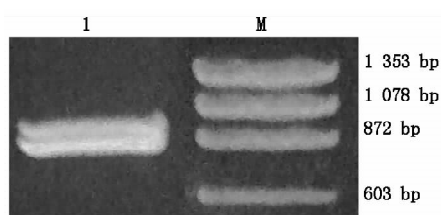
1.2.3 α -醇溶蛋白的序列分析 利用 Clustal W 1.8 对新克隆 α -醇溶蛋白及其近缘物种 α -醇溶蛋白氨基酸序列进行比对分析,用 Mega 4.0 构建 Neighbor-joining 树。根据 Van Herpen 等^[11]的方法,识别 α -醇溶蛋白基因推断氨基酸序列中所含有的 4 种主要 T 细胞优势多肽的数量和种类,并进行 α -醇溶蛋白基因的染色体定位。

1.2.4 α -醇溶蛋白的二级结构预测 采用 PSIPRED 3.3 (<http://bioinf.cs.ucl.ac.uk/psipred/psiform.html>) 对 α -醇溶蛋白的二级结构进行预测。

2 结果与分析

2.1 目的片段的 PCR 扩增、克隆与鉴定

扩增产物经 2% 琼脂糖凝胶电泳分离,在 850 ~ 1 000 bp 得到 2 条相邻的 DNA 条带(图 1),将目的条带回收纯化后与 pMD19-T 载体连接,并转化大肠杆菌 DH5 α ,通过菌液 PCR 鉴定及测序分析,共得到 32 条具有独特完整编码区的核苷酸序列(*ZF5A-1* ~ *ZF5A-32*)。经 NCBI BLAST 比对分析表明:32 个序列与已知的 α -醇溶蛋白基因序列的同源性介于 77% ~ 99%,均为 α -醇溶蛋白基因家族的新成员,GenBank 注册序列号为 JX828280 ~ JX828311。



1. 普通小麦郑丰 5 号; M. ϕ X174-Hae III digest DNA Marker.
1. Bread wheat variety Zhengfeng 5; M. ϕ X174-Hae III digest DNA Marker.

图 1 郑丰 5 号 α -醇溶蛋白基因的 PCR 扩增产物

Fig. 1 PCR amplification products of α -gliadin genes from common wheat cultivar Zhengfeng 5

2.2 α -醇溶蛋白基因的核苷酸及其编码氨基酸序列分析

开放阅读框分析表明,在 32 个基因中,17 个(*ZF5A-1* ~ *ZF5A-17*)具有完整的开放阅读框,可编码的氨基酸序列长度介于 284 ~ 324,其余 15 个(*ZF5A-18* ~ *ZF5A-32*)均为在编码区提前出现 1 个或多个终止子的假基因。17 个新基因与 4 个经典 α -醇溶蛋白基因的推断氨基酸序列比对结果(图 2)表明:17 个基因编码的蛋白均具有 α -醇溶蛋白的典

型结构,但无论在重复区长度、半胱氨酸和谷氨酰胺残基的数量上均存在一定的变异。在 N-末端重复区,相对于其他基因,*ZF5A-13*、*ZF5A-14* 和 *ZF5A-17* 编码的蛋白在 84 ~ 90 位有 1 个 PQPQLPYPQ 片段缺失,*ZF5A-15* 编码的蛋白在 83 ~ 89 位缺失片段 PQPQLPY,*ZF5A-7* 和 *ZF5A-12* 编码的蛋白在 87 ~ 93 位缺失片段 LPYPQPQ。而 *ZF5A-5*、*ZF5A-8* 和 *ZF5A-16* 编码的蛋白在 100 ~ 106 位有 1 个 QLPYPQP 片段插入。在 2 个特征区,除 7 个基因(*ZF5A-1*、*ZF5A-3*、*ZF5A-6*、*ZF5A-9*、*ZF5A-10*、*ZF5A-11*、*ZF5A-15*)的推断氨基酸序列在特征 II 区的第 6 个氨基酸残基 S \rightarrow C 替换而含有 1 个额外的半胱氨酸(C)外,其他 10 个基因的推断氨基酸均含有 6 个保守的半胱氨酸(C)。在 2 个多聚谷氨酰胺区,Ploy I 区谷氨酸的数量介于 10 ~ 21,Ploy II 介于 6 ~ 22。

进一步的比较分析发现,含有 7 个半胱氨酸的 2 个蛋白 *ZF5A-10* 和 *ZF5A-11* 与 GQ891685 氨基酸序列具有高度的相似度,其中,*ZF5A-11* 与 GQ891685 仅在信号肽区存在 1 个 P \rightarrow L 替换,在多聚谷氨酰胺 II 区缺失 3 个谷氨酰胺残基,而 *ZF5A-10* 除了上述 2 个差异外,与 GQ891685 相比,在 N-末端重复区还存在 1 个 Q \rightarrow R 替换。因此,推测 *ZF5A-10* 和 *ZF5A-11* 可能与优质弱筋小麦郑丰 5 号良好的面筋弹性密切相关。

2.3 α -醇溶蛋白中 CD 诱发因子的分析

为对克隆基因进行染色体定位,对 17 个克隆基因的 4 种主要 T 细胞免疫优势多肽谷氨酰胺区的谷氨酰胺残基数量进行了统计分析(表 1)。从表 1 可以看出:克隆的 17 个基因中,*ZF5A-7*、*ZF5A-12* 编码的蛋白仅含有 T 细胞免疫优势多肽 Gli α -9 和 Gli α -20 而没有完整的 Gli α - α 、Gli α - α 2,所以被定位在 6A 染色体上。*ZF5A-13* 编码的蛋白仅含 T 细胞免疫优势多肽 Gli α - α ,*ZF5A-4*、*ZF5A-14* 和 *ZF5A-17* 不含有任何 1 种 T 细胞免疫优势多肽,所以被定位在 6B 染色体,其他 11 个基因编码的蛋白均以不同组合含有 1 ~ 4 种主要 T 细胞免疫优势多肽,且 *ZF5A-5*、*ZF5A-8* 和 *ZF5A-16* 编码的蛋白还含有 2 个 Gli α - α 2,所以被定位在 6D 染色体上。6A、6B、6D 染色体组含有的 4 种主要 T 细胞免疫优势多肽的数量分别为 4,1,34。可见,B 基因组来源的 α -醇溶蛋白,没有或具有很小的 CD 毒性,而 A、D 基因组,尤其是 D 基因组来源的 α -醇溶蛋白对 CD 患者来说具有较高的 CD 毒性。

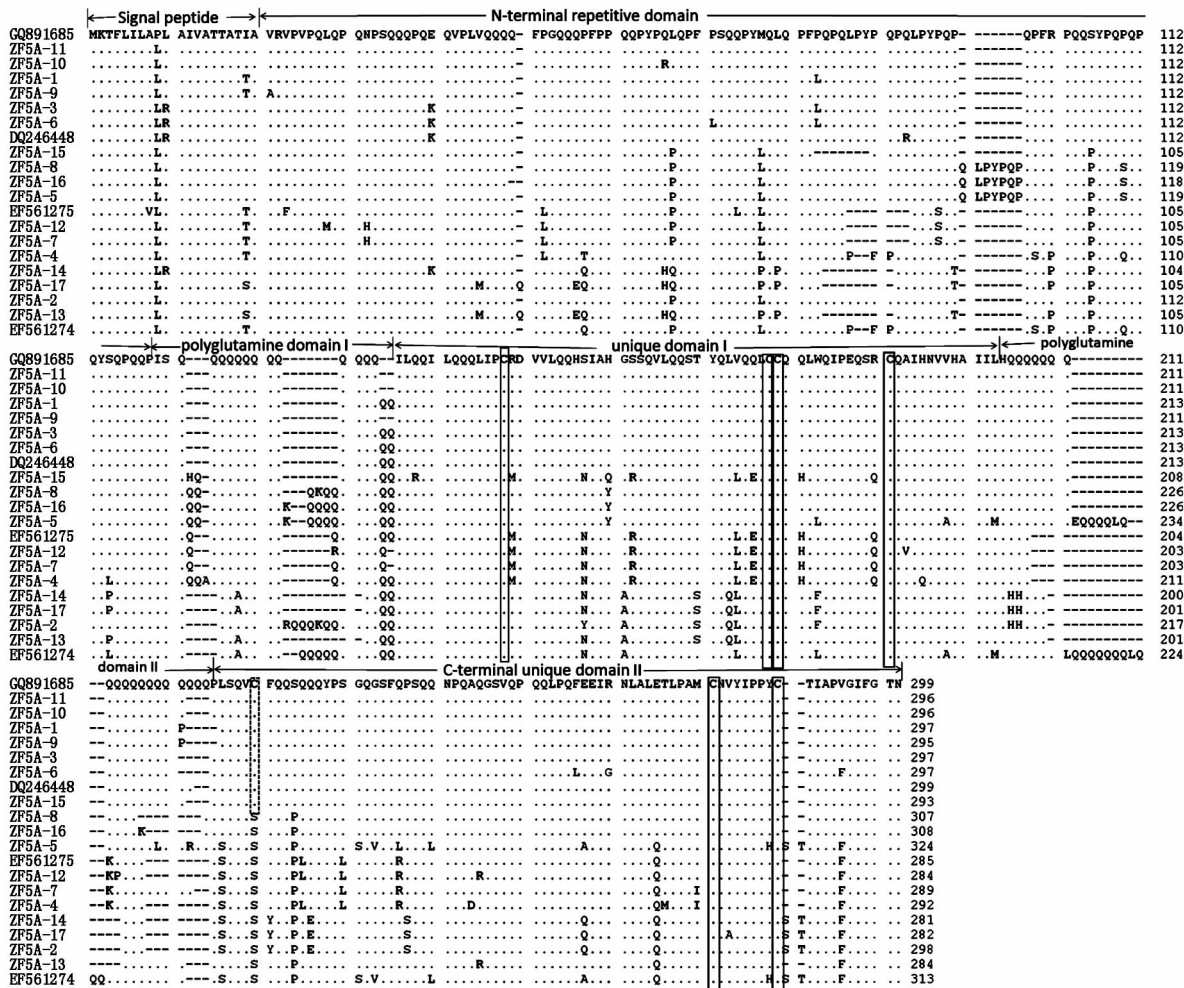


图 2 17 个新 α -醇溶蛋白基因与已知的 4 个经典 α -醇溶蛋白基因的推断氨基酸序列比对

Fig.2 Multiple alignment of the deduced amino acid sequences of 17 novel genes and 4 typical α -gliadin genes from GenBank

表 1 17 个新 α -醇溶蛋白基因编码的蛋白中 4 种主要 T 细胞优势多肽的分布和
多聚谷氨酰胺区的谷氨酸残基的数量及其染色体定位

Tab.1 Number of toxic epitopes presented in the 17 full-ORF α -gliadin proteins and their chromosomal locations

α -醇溶蛋白 α -gliadin proteins	4 种主要的 T 细胞优势多肽 4 major T cell immunogenic peptides				多聚谷氨酰胺 I 区	多聚谷氨酰胺 II 区	染色体定位
					Polyglutamine	Polyglutamine	Chromosomal
	Glia- α 9	Glia- α 2	Glia- α 20	Glia- α	domain I	domain II	location
ZF5A-12	1	0	1	0	15	6	A
ZF5A-7	1	0	1	0	16	12	A
ZF5A-13	0	0	0	1	10	12	B
ZF5A-14	0	0	0	0	10	6	B
ZF5A-17	0	0	0	0	10	6	B
ZF5A-4	0	0	0	0	18	7	B
ZF5A-1	0	1	0	1	15	15	D
ZF5A-10	1	1	0	1	13	16	D
ZF5A-11	1	1	0	1	13	16	D
ZF5A-15	1	1	1	1	16	16	D
ZF5A-16	1	2	0	1	21	12	D
ZF5A-2	1	1	1	0	19	6	D
ZF5A-3	0	1	0	1	15	15	D
ZF5A-5	1	2	0	0	21	22	D
ZF5A-6	1	1	0	1	15	15	D
ZF5A-8	1	2	0	1	20	12	D
ZF5A-9	1	1	0	1	13	15	D

2.4 α -醇溶蛋白基因进化树分析

为进一步分析克隆基因与已知来源于 5 个优质小麦品种(藁城 8901、中优 9507、川农 16、陕 253、济南 177)的 29 个 α -醇溶蛋白基因间的同源性关系,对它们的推断氨基酸序列进行多序列比对分析,并构建 Neighbor-joining 树(图 3)。

从图 3 可以看出:来自普通小麦的 46 个 α -醇溶蛋白基因存在明显的基因组来源特异性,其可分为两大组群 Group I、Group II,Group II 又可分为 4 个亚组 Sub1、Sub2、Sub3 和 Sub4。Group I 中虽然包括了 5 个定位于 6D 染色体的基因,但 9 个定位于 6B 染色体的基因中除 ZF5A-4 和 EF561282 外,其他 7 个基因均分布于 Group I;而 Group II 中的 3 个亚组 Sub1、Sub2 和 Sub3 中全部是定位于 6D 染色体的基因,Sub4 中虽然包括 2 个定位于 6B 染色体和 4 个定位于 6D 染色体的基因,但所有的定位于 6A 染色体的 8 个基因全部分布于 Sub4。因此,Group I、Sub1~Sub3、Sub4 分别是代表 6B、6D 和 6A 染色体上的 α -醇溶蛋白基因。

进一步的分析发现,分布于 Sub3 的 14 个(包括 GQ891865) α -醇溶蛋白基因全部在特征区 II 因第 6 个氨基酸残基 S→C 转换而形成 1 个额外的半胱氨酸。本试验克隆的 7 个含有额外半胱氨酸的基因中,除 ZF5A-15 外,其他 6 个基因均位于此组。因此,推

测位于此亚组的基因可能与这些优质小麦良好的面筋弹性密切相关。



N-J 树上的分类名称分别代表分析基因的基因组来源(竖线前)、GenBank 登录号和品种名称。

The genome(before the vertical line), GenBank accession number and name of common wheat cultivars(after the vertical line) were indicated.

图 3 17 个新 α -醇溶蛋白基因与来源于普通小麦的 29 个 α -醇溶蛋白基因的同源性比对树状图

Fig. 3 Neighbor-joining tree based on deduced amino acid sequence of 17 novel α -gliadin genes in this work and 29 genes from four common wheat cultivars

组成的 α -螺旋(H)和 β -折叠(E)数量、位置及氨基酸组成

Tab. 2 Number sites and amino acid sequences of α -helix and β -strand in the 17 novel α -gliadin proteins and other 4 α -gliadin proteins in the public database

α -醇溶蛋白 α -gliadin proteins	多聚谷氨酰胺 I 区 Polyglutamine domain I		特征区 I Unique Domain I		多聚谷氨酰胺 II 区 Polyglutamine domain II		特征区 II Unique domain II	
	H ₁ (19~30)	H ₂ (0~11)	H _{E1}	H ₃ (12~23)	H ₄ (16~36)	H _{E2}	H ₅ (12~14)	E(2~3)
GQ891683	Q ₁₂ ILQ ₂ ILQ ₃	PCRDVVLQ		STYQLVQQLCCQQL	QSRCQAIHNVVHAILHQ ₈		PQFEEIRNLALET	NVY
GQ891685	Q ₁₃ ILQ ₂ ILQ ₂	PCRDVVLQ		VLQSQSTYQLVQQLCCQQLW	EQSRCQAIHNVVHAILHQ ₈	Q ₆	PQFEEIRNLALET	NVY
EF561287	Q ₁₀ ILQ ₂ ILQ ₃	CRDVVLQ		QSTYQLVQQLCCQQL	QSRCQAIHNVVHAILHQ ₄		PQFEEIRNLALET	NVY
EF561288	SQ ₃ AQ ₁₂ ILQ ₂ ILQ ₃	PCRDVVLQ		SQVLQSTYQLLQVWCCQQLL	RCQAIHNVVHAILHQ ₆ EQ ₁ Q ₆ LH		PQFAIRNLALET	NVY
ZF5A-1	Q ₁₁ ILQ ₂ ILQ ₃	PCRDVVLQ		STYQLVQQLCCQQL	QSRCQAIHNVVHAILHQ ₇		PQFEEIRNLALET	NVY
ZF5A-3	Q ₁₄ ILQ ₂ ILQ ₃	RDVV		TYQLVQQLCCQQL	QSRCQAIHNVVHAILHQ ₇		PQFEEIRNLALET	NVY
ZF5A-4	Q ₃ AQ ₁₅ ILQ ₂ ILQ ₂	CMDVVLQ		SQVLQSTYQLLQELCCQHL	SQCQAIQNVVHAILHQQQKQ		PQFEEIRNLALET	NVY
ZF5A-9	Q ₁₃ ILQ ₂ ILQ ₃	PCRDVVLQ		QSTYQLVQQLCCQQLW	QSRCQAIHNVVHAILHQ ₁₁		PQLEEIGNLALET	NVY
ZF5A-10	Q ₁₅ ILQ ₂ ILQ ₃	CRDVVLQ		QSTYQLVQQLCCQQL	RCQAIHNVVHAILHQ ₁₀		PQLEEIGNLALET	NVY
ZF5A-11	Q ₁₅ ILQ ₂ ILQ ₃	PCRDVVLQ		QSTYQLVQQLCCQQLW	RCQAIHNVVHAILHQ ₁₂		PQLEEIGNLALET	NVY
ZF5A-12	Q ₇ RQ ₅ ILQ ₂ ILQ ₃	PCMDVVLQ		HGRSQVLQSTYQLLQELCCQHL	QSQCVIHNVVHAILHQ ₁₀		PQFEEIRNLALET	NVY
ZF5A-13	Q ₃ AQ ₈ ILQ ₂ ILQ ₃	DVV		SYQQLQQLCCQQL	RCQAIHNVVHAILHQ ₈		PQFEEIRNLALET	NVY
ZF5A-14	Q ₂ AQ ₇ ILQ ₂ ILQ ₃	DVVL		SYQQLQQLCCQQLF	SRQAIHNVVHAILHQ		PQFEEIRNLALET	NVY
ZF5A-16	Q ₈ KQ ₁₀ ILQ ₂ ILQ ₃	CRDVVLQ		TYQLVQQLCCQQLW	CQAIHNVVHAILHQ ₁₀		LPQFEEIRNLALET	NVY
ZF5A-5	Q ₁ IKQ ₁₀ ILQ ₂ ILQ ₃	CRDVVLQ		TYQLVQQLCCQQLL	CQAIHNVVHAILHQ ₁₃	QVC	PQFEEIRNLALET	NVY
ZF5A-6	Q ₁₀ ILQ ₂ ILQ ₃	PCRDV		TYQLVQQLCCQQL	SRQAIHNVVHAILHQ ₅		PQLEEIGNLALET	NVY
ZF5A-7	Q ₁₅ ILQ ₂ ILQ ₃	LIPCMDVVLQ		VLQSQSTYQLLQELCCQHLW	QCQAIHNVVHAILHQ ₃ KQ ₆		LPQFEEIRNLALET	NVY
ZF5A-2	Q ₆ RQ ₂ KQ ₈ ILQ ₂ ILQ ₃	RDVV	HASSQ	QSSYQQLQQLCCQQLF	SRQAIHNVVHAILHHH		PQFEEIRNLALET	NVY
ZF5A-8	Q ₁₁ KQ ₈ ILQ ₂ ILQ ₃	CRDVVLQ	IAYGSSQVL	STYQLVQQLCCQQL	RCQAIHNVVHAILHQ ₉		PQFEEIRNLALET	NVY
ZF5A-15	Q ₁₅ ILRQILQ ₃	PCMDVVLQ		SQVLQSTYQLLQELCCQHL	EQSQCVIHNVVHAILHQ ₁₃		PQFEEIRNLALET	NVY
ZF5A-17	QAQ ₈ ILQ ₂ ILQ ₃			SYQQLQQLCCQQLF	SRQAIHNVVHAILHQ		PQFEEIRNLALET	NVY

2.5 α -醇溶蛋白的二级结构分析

采用 PSIPRED(version 3.3) 对克隆的 17 个 α -醇溶蛋白基因和来自弱筋小麦陕 253(GenBank 登录号: GQ891683、GQ891685) 和中优 9507(GenBank 登录号: EF561288、EF561287) 的 4 个 α -醇溶蛋白基因编码蛋白的二级结构进行了预测(表 2)。

从表 2 可以看出, 参与形成 α -螺旋(H) 和 β -折叠(E) 的核心序列及其位置是相对保守的, 但不同蛋白形成 α -螺旋(H) 和 β -折叠(E) 的数量却各不相同。而且除 α -螺旋 H_5 外, 不同蛋白中参与形成 α -螺旋 $H_1 \sim H_4$ 的氨基酸残基数也存在较大的差异。通常在 N-末端重复区不存在任何二级结构, 4~6 个 α -螺旋(H) 和 1 个 β -折叠(E) 主要出现在 2 个特征区和多聚谷氨酸区。在所预测的 21 个基因中, 除 ZF5A-17 编码的蛋白缺少 α -螺旋 H_2 , ZF5A-2 和 ZF5A-8 编码的蛋白在特征区 I 存在 1 个额外的 α -螺旋 H_{E1} , GQ891685 编码的蛋白 ACX71610 和 ZF5A-15 编码的蛋白在多聚谷氨酰胺 II 区存在 1 个额外的 α -螺旋 H_{E2} 外, 5 个保守的 α -螺旋($H_1 \sim H_5$) 恒定出现在其他 16 个预测基因中。此外, 除 7 个基因(ZF5A-2、ZF5A-5、ZF5A-6、ZF5A-7、ZF5A-12、ZF5A-15、ZF5A-17) 编码的蛋白无任何 β -折叠(E) 结构外, 其他 14 个基因编码的蛋白在 C-末端特征区形成 1 个 β -折叠(E) 结构。其中 GQ891685 编码的蛋白 ACX71610 具有最多的二级结构基本单元, 除 5 个保守的 α -螺旋和 β -折叠(E) 外, 还存在 1 个额外的 α -螺旋 H_{E2} 。而试验克隆的 7 个具有额外半胱氨酸的基因中, 除 ZF5A-6 外, 其他编码的蛋白均具有较多的二级结构单元, 含有 5 个保守的 α -螺旋和 β -折叠(E)。说明二级结构预测可为揭示 α -醇溶蛋白的分子功能提供参考。郑丰 5 号中较多含有额外半胱氨酸和二级结构单元的 α -醇溶蛋白可能与其良好的加工品质有关。

3 讨论

3.1 α -醇溶蛋白的变异、结构及其与功能的关系

α -醇溶蛋白是由多基因家族的 *Gli-2* 位点编码的, 对普通六倍体小麦而言, 不同品种间的基因拷贝数高达 25~35^[17] 甚至 100~150^[7]。因此, 尽管有大约 50%^[7] 甚至 87%^[11] 的 α -醇溶蛋白基因为假基因, 但 Lafandra 等^[18] 通过双向电泳分析仍发现, 在小麦品种 Cheyenne 中至少包括 16 个 α -醇溶蛋白基因。通过体外功能验证, 李敏等^[8] 发现, 由基因 GQ891685 编码的含有 1 个额外半胱氨酸残基的 α -醇溶蛋白亚基 ACX71610 能显著增加面筋弹性而降

低面筋强度。本研究从郑丰 5 号中成功克隆得到 17 个具有完整开放阅读框的 α -醇溶蛋白基因, 其中 7 个 α -醇溶蛋白基因(ZF5A-1、ZF5A-3、ZF5A-6、ZF5A-9、ZF5A-10、ZF5A-11 和 ZF5A-15) 与 GQ891685 一样, 其编码的蛋白均在特征区 II 含有 1 个额外的半胱氨酸残基, 尤其是 ZF5A-10、ZF5A-11 编码的蛋白氨基酸序列与 GQ891685 编码的蛋白氨基酸序列有很高的同源性, 因此推测, 这 7 个基因, 尤其是 ZF5A-10、ZF5A-11 可能与郑丰 5 号良好的面筋弹性密切相关。

蛋白质的生物学功能与蛋白质的二级结构及蛋白折叠过程有关^[19-20], 因此, 二级结构预测是洞悉和了解蛋白分子功能的方法之一, 可为揭示蛋白的分子功能提供一定的参考和线索^[19-21]。近年来基于计算机自动预测的各种方法, 如 GOR、PSIPRED、YASPIN 和 HNN 等得以迅速发展^[18-19]。但在互联网最常用的计算方法中, PSIPRED 法因其较高的预测准确率、输出结果的清晰易懂性等优势而成为蛋白质二级结构预测的最常用的方法^[19-21]。Xie 等^[22] 曾采用 PSIPRED 2.6 对 19 个克隆的 α -醇溶蛋白的二级结构进行预测, 结果表明, α -螺旋和 β -折叠在 α -醇溶蛋白中的位置相对保守, 但不同基因编码的蛋白形成的 α -螺旋和 β -折叠的数量不等。其中, β -折叠含量较高, 而且大多数 α -螺旋和 β -折叠都出现在 2 个特征区, 因此, Xie 等提出两个特征区, 尤其是 C-末端特征区是 α -醇溶蛋白的重要功能区。本研究对克隆的 17 个基因和其他 4 个基因的二级结构的预测结果与 Xie 等^[22] 的研究结果不完全一致。本研究的预测结果显示, α -醇溶蛋白中的 β -折叠含量并不高, 只出现在 C-末端特征区, 但 α -螺旋的含量较高, 可分布在 2 个多聚谷氨酰胺区(H_1 、 H_{E2}) 和特征区(H_2 、 H_{E1} 、 H_3 、 H_4 、 H_5)。其中特征区 I 的含量最高(H_2 、 H_{E1} 、 H_3 、 H_4)。另外, 不仅不同基因编码的蛋白形成的 α -螺旋和 β -折叠的数量不等, 而且参与形成特定保守区域的 α -螺旋的氨基酸残基的数量在不同基因编码的蛋白间也存在较大的差异。一般认为较多的 α -螺旋和 β -折叠有助于增加面筋的粘弹性^[22], 本研究比较分析发现, 分析的 21 个基因中, 已证明对品质有显著影响的基因 GQ891685 编码的蛋白具有最多的 α -螺旋(H_1 、 H_2 、 H_3 、 H_4 、 H_5 、 H_{E2}) 和 β -折叠(E), 而试验从郑丰 5 号中克隆的 17 个 α -醇溶蛋白基因中, 10 个基因含有所有保守的 α -螺旋和 β -折叠, 其他 7 个不含 β -折叠的基因中, 有 3 个含有额外的 α -螺旋 H_{E1} 或 H_{E2} 。郑丰 5 号中含有较多 α -螺旋和 β -折叠的 α -醇溶蛋白,

可能与郑丰 5 号良好的面筋品质密切相关。

3.2 α -醇溶蛋白的 CD 毒性肽分布、染色体定位及其与小麦的品质育种

分布于 α -醇溶蛋白的 4 种主要 T 细胞免疫优势多肽数量和分布组合存在明显的基因组特异性^[10-11]。这一方面说明,可根据 4 种主要 T 细胞免疫优势多肽的分布对 α -醇溶蛋白基因进行染色体定位;另一方面说明,对普通六倍体小麦而言,没有任何一个小麦品种对 CD 患者来说是绝对安全的^[15]。Koning^[14]的综合分析表明,面筋蛋白中 CD 毒性肽的含量对 CD 症的发生率有重要的影响,面筋蛋白中含有的 CD 毒性肽的种类和数量越多,CD 症的发病率也越高。本研究根据 4 种 T 细胞免疫优势多肽的分布特点,成功将克隆的 17 个具有完整开放阅读框的 α -醇溶蛋白基因定位在 6A、6B 和 6D 染色体上,而且这些 α -醇溶蛋白基因与其他来源于优质小麦的 α -醇溶蛋白基因的同源性分析也与此定位结果基本吻合。

依据 4 种主要 T 细胞免疫优势多肽的分布数量和组合,克隆的 17 个 α -醇溶蛋白基因中有 11 个(包括具有额外半胱氨酸的 7 个基因)定位于 6D 染色体,而同源性分析的结果也表明,尽管来源于普通小麦的 α -醇溶蛋白基因也具有明显的基因组特异性,但来源于 6D 染色体的 α -醇溶蛋白基因明显较 A、B 基因组来源的 α -醇溶蛋白基因具有更广泛的变异。此外,本研究中定位于 6D 染色体上的 α -醇溶蛋白含有最多的 T 细胞免疫多肽,这与前人的研究一致^[10-11, 22]。可见, D 基因组无论对小麦的加工品质改良还是筛选对 CD 患者无毒或毒性较小的小麦品种而言,都是极为重要的基因组。

参考文献:

- [1] Ma W, Apples R, Bekes F *et al.* Genetic characterization of dough rheological properties in a wheat doubled haploid population: additive genetic effects and epistatic interactions [J]. *Theoretical and Applied Genetics*, 2005, 111: 410 - 422.
- [2] Wieser H. Chemistry of gluten proteins [J]. *Food Microbiology*, 2007, 24: 115 - 119.
- [3] Chen F G, Xu C H, Chen M Z *et al.* A new alpha-gliadin gene family for wheat breeding: somatic introgression line II-42 derived from *Triticum aestivum* and *Agropyron elongatum* [J]. *Molecular Breeding*, 2008, 22: 675 - 685.
- [4] Gu Y Q, Crossman C, Kong X, *et al.* Genomic organization of the complex alpha-gliadin gene loci in wheat [J]. *Theoretical and Applied Genetics*, 2004, 109: 648 - 657.
- [5] Anderson O D, Greene F C. The α -gliadin gene family. II. DNA and protein sequence variation, subfamily structure, and origins of pseudogenes [J]. *Theoretical and Applied Genetics*, 1997, 95: 59 - 65.
- [6] D'ovidio R, Masci S. The low-molecular-weight glutenin subunits of wheat gluten [J]. *Journal of Cereal Science*, 2004, 39: 321 - 339.
- [7] Khatkar B S, Fido R J, Tatham S, *et al.* Functional properties of wheat gliadins II. Effects on dynamic rheological properties of wheat gluten [J]. *Journal of Cereal Science*, 2002, 35: 307 - 313.
- [8] 李敏, 高翔, 陈其皎, 等. 普通小麦中 α -醇溶蛋白基因(*GQ891685*)的克隆、表达及品质效应鉴定 [J]. *中国农业科学*, 2011, 43: 4765 - 4774.
- [9] Arentz-Hansen H, Kören R, Molberg Ø *et al.* The intestinal T cell response to α -gliadin in adult celiac disease is focused on a single deamidated glutamine targeted by tissue transglutaminase [J]. *The Journal of Experimental Medicine*, 2000, 191: 603 - 612.
- [10] Vader W, Steoeniak D, Kooy Y, *et al.* The HLA-DQ2 gene dose effect in celiac disease is directly related to the magnitude and breadth of gluten-specific T cell responses [J]. *Proceedings of the National Academy of Sciences of USA*, 2003, 100: 12390 - 12395.
- [11] Van Herpen T W J M, Goryunova S V, Van der schoot J, *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.
- [12] Vaccino P, Becker H A, Brandolini A *et al.* A catalogue of *Triticum monococcum* genes encoding toxic and immunogenic peptides for celiac disease patients [J]. *Molecular Genetics and Genomics*, 2009, 281: 289 - 300.
- [13] Ciccioppo R, Sabatino A D, Corazza G R. The immune recognition of gluten in coeliac disease [J]. *Clinical and Experimental Immunology*, 2005, 140: 408 - 416.
- [14] Koning F. Celiac disease: quantity matters [J]. *Seminars in Immunopathology*, 2012, 34: 541 - 549.
- [15] 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, 120(4): 354 - 356.
- [16] 李玉阁, 邢冉冉, 李锁平. 栽培一粒小麦 α -醇溶蛋白新基因的克隆与序列分析 [J]. *麦类作物学报*, 2012, 32(3): 387 - 392.
- [17] Okita T W, Cheesbrough V, Reeves C D. Evolution and heterogeneity of the alpha-/beta-type and gamma-type gliadin DNA sequences [J]. *The Journal of Biological Chemistry*, 1985, 260: 8203 - 8213.
- [18] Lafiandra D, Kasarda D D, Morris R. Chromosomal assignment of genes coding for the wheat gliadin protein components of the cultivars Cheyenne and Chinese Spring by two-dimensional (two-pH) electrophoresis [J]. *Theoretical and Applied Genetics*, 1984, 68(6): 531 - 539.
- [19] Palopoli L G, Simona E, Rombo A *et al.* Improving protein secondary structure predictions by prediction fusion [J]. *Journal of Computational Biology*, 2009, 3(1): 217 - 232.
- [20] Koswatta T J, Samaraweera P, Sumansingh V A. A simple comparison between specific protein secondary structure prediction tools [J]. *Tropical Agricultural Research*, 2011, 23(1): 91 - 98.
- [21] Link K, Simossis V A, Taylor R *et al.* A simple and fast secondary structure prediction method using hidden neural networks [J]. *Bioinformatics*, 2005, 21(2): 152 - 159.
- [22] Xie Z Z, Wang C Y, Wang K *et al.* Molecular characterization of the celiac disease epitope domains in α -gliadin genes in *Aegilops tauschii* and hexaploid wheats (*Triticum aestivum* L.) [J]. *Theoretical and Applied Genetics*, 2010, 121: 1239 - 1251.